

**Kurze Einführung in die Vektor- und Matrixrechnung
für die Multivariate Statistik**

U. Mortensen

Version 14. 07. 2018

Vorwort

Dieses Skript liefert keine allgemeine Einführung in die Lineare Algebra, sondern nur die Elemente der Vektor- und Matrixrechnung, die für die am häufigsten verwendeten multivariaten Verfahren notwendig sind: Faktoren- bzw. Hauptkomponentenanalyse, multiple und kanonische Regression bzw. Korrelation, die Diskriminanz- und Korrespondenzanalyse. Deswegen wird hauptsächlich der Begriff des Vektorraums für endlichdimensionale, reelle Vektoren definiert; im letzten Abschnitt (Funktionenräume) wird ein Ausblick auf Verallgemeinerungen gegeben.

Inhaltsverzeichnis

1	Vektoren und Vektorräume	5
1.1	Der Sinn der Vektor- und Matrixrechnung	5
1.2	Punkträume	7
1.3	Vektoren	10
1.4	Vektorräume	19
1.5	Lineare Abhängigkeit und Unabhängigkeit	22
1.6	Erzeugendensysteme und Basen von Vektorräumen	28
2	Matrizen	36
2.1	Definitionen	36
2.2	Operationen mit Matrizen	37
2.2.1	Multiplikation einer Matrix mit einem Vektor	38
2.2.2	Multiplikation einer Matrix mit einer Matrix	40
2.2.3	Beispiel: Varianz-Kovarianz-Matrizen	43
2.3	Der Rang einer Matrix	44
2.4	Die Inverse einer Matrix	49
2.5	Rotationen, Quadratische Formen und Eigenvektoren	52
2.5.1	Rotationen	53
2.5.2	Quadratische Formen und Eigenvektoren	55
2.5.3	Der Rayleigh-Quotient	66
2.5.4	Die Singularwertzerlegung einer Matrix	70
2.5.5	Die PCA als Anwendung der SVD	73
2.5.6	Basiswechsel	80
2.5.7	Die Inverse und die Wurzel einer symmetrischen Matrix	80
2.5.8	Die Pseudoinverse einer Matrix	82
2.5.9	Vektor- und Matrixnormen	83
2.5.10	Die Approximation von Matrizen	86
2.6	Basen und Transformationen von Basen	91
2.7	Eigenvektoren und Eigenwerte nichtsymmetrischer Matrizen	93
2.7.1	Der allgemeine Fall	93
2.7.2	Das generalisierte Eigenvektorproblem	99

2.7.3	Mehrfache Eigenwerte	102
2.8	Lineare Gleichungssysteme	103
2.9	Projektionen	107
2.9.1	Orthogonale Projektion eines Vektors auf einen anderen . .	107
2.9.2	Projektionsmatrizen	107
2.9.3	Projektionen auf Hauptachsen	110
2.9.4	Die Zentrierungsmatrix	111
3	Funktionsräume und PCA	114
3.1	Einführung	114
3.2	Karhunen-Loève-Entwicklung und PCA	116
4	Anhang	120
4.1	Steinerscher Austauschatz und das Fundamental-Lemma	120
4.2	Zur Berechnung von Ellipsen für eine Punktekonfiguration	122
4.3	Die Differentiation von Vektoren	123
4.3.1	Die allgemeine Differentiationsformel	123
4.3.2	Die Differentiation quadratischer Formen	124
4.3.3	Die Kleinste-Quadrate-Schätzung für das Lineare Modell .	125
4.3.4	Extrema unter Nebenbedingungen	127
4.4	Alternativer Beweis von Satz 2.31	130
4.5	Transformationen und Abbildungen	132
4.6	Determinanten	136
	Literatur	138
	Index	139

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

Bertrand Russell

1 Vektoren und Vektorräume

1.1 Der Sinn der Vektor- und Matrixrechnung

In der multivariaten Statistik werden große Datenmengen analysiert. So werden an hinreichend großen Stichproben von Personen oder Objekten ("Fälle") mehrere Variablen gemessen – wobei 'hinreichend' u. von der Anzahl der gemessenen Variablen abhängt – und es soll z.B. untersucht werden, welchen Abhängigkeiten zwischen den Variablen existieren. Diese Frage kann für sich genommen von Interesse sein, oder man hat das Ziel, bestimmte Variablen¹ auf der Basis der übrigen "vorauszusagen", oder die Fälle sollen anhand der Messungen bestimmten Klassen zugeordnet werden. Hat man z.B. 100 Fälle und 7 Variablen, so hat man bereits 700 Messwerte zu inspizieren und zu interpretieren. Betrachtet man nur die paarweisen Korrelationen zwischen den Variablen, so hat man bereits $\binom{7}{2} = 7 \cdot 6/2 = 21$ Korrelationen und die Relationen zwischen ihnen zu diskutieren; die Frage ist nun, in welcher Weise diese Abhängigkeiten zu interpretieren sind.

Ein Standardansatz für die Interpretation solcher Daten besteht in der Annahme der Existenz voneinander unabhängiger, latenter Variablen. Das sind Variablen, die nicht direkt gemessen wurden, von denen aber angenommen wird, dass sie auf die meisten der gemessenen Variablen einwirken und so die Korrelationen zwischen den gemessenen Variablen erzeugen. Man kann diese Annahme allgemein so formulieren, dass man sagt, jede der gemessenen Variablen V_j lasse sich als ein Funktion

$$V_j = f_j(L_1, \dots, L_r), \quad j = 1, \dots, n \quad (1.1)$$

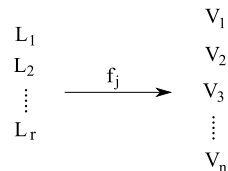
von noch unbekanntem, "latentem" Variablen L_1, \dots, L_r , $r \leq n$, darstellen (Abb. 1). Selbst wenn diese Annahme gilt, so weiß man noch nichts über die Form der Funktionen f_j . Die einfachste Annahme über die f_j ist, sie als lineare Funktionen anzunehmen:

$$V_j = a_{1j}L_1 + a_{2j}L_2 + \dots + a_{rj}L_r + e_j, \quad (1.2)$$

wobei hier der Einfachheit halber die Variablennamen mit den Messwerten für die Variablen gleichgesetzt wurden. Die a_{1j}, \dots, a_{rj} sind freie Parameter, die zusammen mit den latenten Variablen aus den Daten geschätzt werden müssen, und e_j

¹Es kann auch 'bestimmte Variable' heißen, dem Duden zufolge kann man zwischen der Pluralformen 'Variable' und 'Variablen' wählen. In diesem Text wird von der letzteren Form Gebrauch gemacht, es ist aber – einr alten Gewohnheit folgend – möglich, dass gelegentlich die erste Form gewählt wurde.

Abbildung 1: Latente Variablen L_1, L_2, \dots, L_r und gemessene Variablen V_1, \dots, V_n . $f_i = f_i(L_1, \dots, L_r) = V_i$, d.h. jede der gemessenen Variablen läßt sich durch eine Funktion f_j der latenten Variablen darstellen.



repräsentieren zufällige Messfehler. Die Annahme der Additivität der L_j ist nicht selbstverständlich oder trivial, denn es ist denkbar, dass latente Variable existieren, die z.B. multiplikativ in die Messungen eingehen, z.B. wenn der Effekt der einen latenten Variable proportional zum Effekt einer anderen latenten Variablen ist; Verallgemeinerungen des Ansatzes 1.2 sind im Prinzip möglich.

Der Versuch, latente Variablen mit elementarer Mathematik zu bestimmen, erweist sich als äußerst mühsam. Die Rechnungen werden unübersichtlich, man wird zum Beispiel auf Gleichungssysteme geführt, von denen nicht klar ist, ob sie überhaupt eine Lösung haben, und wenn ja, ob diese Lösung eindeutig ist oder nicht. Derartige Fragen lassen sich relativ einfach beantworten, wenn man sie durch Anwendung der Vektor- und Matrixrechnung angeht. Vektoren sind dabei ganze Spalten oder Zeilen von Zahlen, etwa von Messwerten für eine Variable, und Matrizen ergeben sich, indem man Vektoren neben- oder untereinander anschreibt. Man rechnet dann mit ganzen Blöcken von Zahlen. Die Rechenregeln für diese Blöcke sind so angelegt, dass sie die gewünschten Analysen der Daten ermöglichen. Abhängigkeiten zwischen den Messungen lassen sich durch einfache Eigenschaften von Matrizen ausdrücken, und die Berechnung von latenten Variablen oder Klassifikationen reduzieren sich auf übersichtliche Rechnungen mit Vektoren und Matrizen.

Der Begriff des Vektors geht auf den irischen Mathematiker William Rowland Hamilton (1839 – 1903) zurück, der sich mit bestimmten Vierergruppen von Zahlen, sogenannten Quaternionen, beschäftigte und fand, dass die Rechnungen mit diesen Gruppen vereinfacht wurden, wenn man für sie eben den Begriff des Vektors einführt, wobei Vektoren durch bestimmte Rechenregeln miteinander verknüpfen kann. Das Wort 'Vektor' kommt aus dem Lateinischen und bedeutet so viel wie Führer oder Fahrer, aber auf die Wortbedeutung muß zum Verständnis der Vektorrechnung nicht eingegangen werden. Ebenso muß auf den Begriff der Quaternionen hier nicht eingegangen werden, weil sie im Folgenden keine Rolle spielen. Die Physiker entdeckten dann, dass sich der Effekt verschiedener Kräfte auf Körper gut durch Vektoren darstellen läßt. Das Vektorparallelogramm zeigt, wie die resultierende Kraft sich nach den Regeln der Vektoraddition aus Teil-

kräften errechnen läßt. Aus Anwendungen wie diesen ergibt sich die Definition von Vektoren als *gerichteten Größen*. Die Anwendungen in der Physik und nicht zuletzt in der Statistik im Laufe des zwanzigsten Jahrhunderts haben wesentlich zu Entwicklung der Vektor- und Matrixrechnung beigetragen. Mittlerweile findet man kaum eine Darstellung multivariater statistischer Verfahren, die auf die Anwendung der Vektorrechnung verzichtet, – Darstellungen ohne den Gebrauch der Vektor- und Matrixrechnung werden so unübersichtlich, dass das Verständnis der Verfahren unnötig erschwert wird.

Die folgenden Darstellungen beziehen sich auf elementare Vektoren und dementsprechend auf elementare Matrizen. Elementare Vektoren sind als "*n*-Tupel" von Zahlen definiert, also auf eine Gruppe (x_1, x_2, \dots, x_n) von *n* Zahlen, wobei es auf die Reihenfolge, in der die Zahlen angeschrieben werden, ankommt. Dieselben Zahlen in verschiedener Anordnung angeschrieben definieren also verschiedene Vektoren. Aus mathematischer Sicht sind derartige *n*-Tupel allerdings Spezialfälle. So können Funktionen, die über einem bestimmten Intervall $[a, b]$ definiert sind, ebenfalls als Vektoren aufgefasst werden. Viele der im folgenden eingeführten Begriffe müssen dann allgemeiner definiert werden. Für die hier behandelten multivariaten Verfahren genügt es aber, sich auf "elementare" Vektoren zu beschränken. Die Kenntnis der elementaren Theorie der Vektoren und Matrizen erleichtert dann den Zugang zu den allgemeineren Begriffsbildungen.

1.2 Punkträume

Mit dem Symbol \mathbb{R} wird die Menge der reellen Zahlen bezeichnet². 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 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

$$\mathbb{R}^n = \{(x_1, x_2, \dots, x_n) | x_1, x_2, \dots, x_n \in \mathbb{R}\}, \quad n \in \mathbb{N} \quad (1.3)$$

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

Koordinatenachsen lassen sich skalieren, d.h. mit einem Faktor multiplizieren

²Das sind die natürlichen Zahlen $\mathbb{N} = 0, 1, 2, \dots$, die ganzen Zahlen $\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$ die rationalen Zahlen $\mathbb{Q} = \{p/q | p, q \in \mathbb{Z}, q \neq 0\}$, und 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.

(wenn man etwa von Zentimetern zu Millimetern übergeht) und die Koordinaten von Punkten lassen sich addieren: Sind $x = (x_1, \dots, x_n)$ und $y = (y_1, \dots, y_n)$ irgendzwei Punkte, so soll

1. $ax = (ax_1, ax_2, \dots, ax_n)$, für $a \in \mathbb{R}$, und
 2. $x + y = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n)$
- 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 $x = (x_1, \dots, x_n)$ und $y = (y_1, \dots, y_n)$ seien irgendzwei Punkte im \mathbb{R}^n . Die Punkte sind durch eine Distanz $d(x, y)$ voneinander getrennt. Was mit einer Distanz gemeint ist, wird durch die folgenden *Axiome* festgelegt:

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*).

Die Axiome definieren die allgemeinen Eigenschaften einer *Metrik*, die darüber hinaus durch die Art der Berechnung der $d(x, y)$ festgelegt wird, s. unten. 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.

Nach Euklid soll gelten, dass die kürzeste Verbindung zwischen zwei Punkten eine Gerade ist. Für irgendzwei Punkte x und y gilt dann

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} = \left[\sum_{i=1}^n (x_i - y_i)^2 \right]^{\frac{1}{2}}. \quad (1.4)$$

Dies ist der Satz des Pythagoras für den n -dimensionalen Raum. Ist die Metrik eines Raums durch (1.4) definiert, so heißt der Raum *euklidisch*.

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-Kontinuum) und definierte eine Metrik – die Minkowski-Metrik –, die durch

$$d(x, y) = \left[\sum_{i=1}^n |x_i - y_i|^p \right]^{\frac{1}{p}}, \quad 0 < p \in \mathbb{R} \quad (1.5)$$

definiert ist. Für $p = 2$ ergibt sich die euklidische Metrik (1.4), 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 dieser Geraden und den Koordinatenachsen gegeben. Für maximal drei Dimensionen (3 Variablen) kann der Vektor graphisch durch einen Pfeil repräsentiert werden (s. Abbildung 2). 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 Matrixrechnung 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

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

Definition 1.1 *Ein n -dimensionaler Vektor ist ein n -Tupel reeller Zahlen:*

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad (1.6)$$

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

Dementsprechend sind

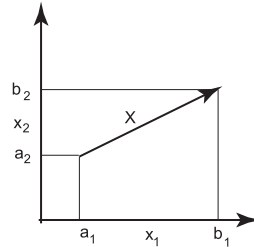
$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$$

verschiedene 3-dimensionale Vektoren.

Anmerkungen:

1. Vektoren werden im Folgenden wie in (1.6) durch fette Buchstaben bezeichnet; diese Bezeichnung 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 $\tilde{\vec{x}}$ und $\hat{\hat{x}}$ bzw. $\tilde{\tilde{x}}$ zeichnen sich durch sehr geringen graphischen Charme aus, insbesondere wenn transponierte Vektoren \vec{x}' (s. unten) betrachtet werden.

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



2. Ein Vektor \mathbf{x} wird wie in Gleichung (1.6) 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:

$$\mathbf{x}' = (x_1, x_2, \dots, x_n). \quad (1.7)$$

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

$$\mathbf{x} = (x_1, x_2, \dots, x_n)'. \quad \square$$

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, \dots, a_n , der Endpunkt habe die Koordinaten b_1, \dots, b_n . Dann sind die Komponenten durch

$$x_i = b_i - a_i, \quad i = 1, \dots, n \quad (1.8)$$

gegeben, s. a. Abb. 2. Die x_i definieren damit nur die Länge (s. unten) und die Orientierung, nicht aber den genauen Ort des repräsentierenden Pfeils. Dementsprechend ist auch die Schreibweise \vec{x} für einen Vektor üblich, also $\mathbf{x} = \vec{x}$.

Da ein Vektor durch ein Zahlentupel (x_1, \dots, 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, \dots, x_n) und der Vektor genau dann zusammen, wenn der Anfangspunkt in den Nullpunkt des Koordinatensystems gelegt werden kann.

Spezielle Vektoren: Es seien

$$\vec{0} = (0, 0, \dots, 0)' \quad (1.9)$$

$$\vec{1} = (1, 1, \dots, 1)' \quad (1.10)$$

$$\mathbf{e}_j = (0, \dots, 0, 1, 0, \dots, 0)', \quad j = 1, \dots, n \quad (1.11)$$

$\vec{0}$ 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 *Einheitsvektor*, seine Komponenten sind alle gleich Null bis auf die j -te Komponente, die gleich 1 ist.

Multiplikation mit einem Skalar: Es bedeutet

$$\lambda \mathbf{x} = \lambda \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{pmatrix}, \quad \lambda \in \mathbb{R} \quad (1.12)$$

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.

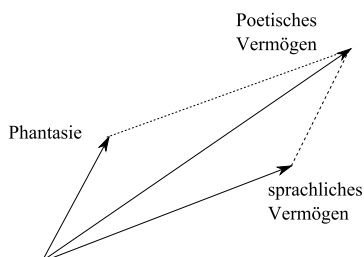
Addition von Vektoren Es seien \mathbf{x} und \mathbf{y} zwei n -dimensionale Vektoren. Dann heißt

$$\mathbf{z} = \mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} \quad (1.13)$$

die Summe der Vektoren \mathbf{x} und \mathbf{y} . Vektoren können also nur addiert werden, wenn sie dieselbe Anzahl von Komponenten haben. Abb. 3 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. Andererseits wird deutlich, dass die Annahme eines Vektormodells für einen bestimmten Objektbereich durchaus gewisse Beschränkungen impliziert: es ist ja möglich, dass sprachliches Vermögen und Phantasie *nicht* nur additiv (im Sinne einer Vektoraddition) zusammenwirken um ein komplexes Merkmal wie poetisches Vermögen zu erzeugen³. Nicht-additive Wechselwirkungen zwischen den Merkmalen werden durch die Wahl einer Vektorrepräsentation der Merkmale ausgeschlossen.

³Auch 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 3: Addition von Vektoren, analog zur Addition von Kräften in der Physik



Definition 1.2 *Die Summe*

$$\mathbf{x} = a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 + \cdots + a_p \mathbf{x}_p, \quad (1.14)$$

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

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:

$$Y_i = a_1 X_{i1} + a_2 X_{i2} + a_3 X_{i3} + e_i, \quad i = 1, \dots, m \quad (1.15)$$

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 $X_{ij}, j = 1, \dots, 3$ 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 der Methode der Kleinsten Quadrate so geschätzt, dass die Summe $\sum_i e_i^2$ der Fehlerquadrate minimal wird. Ausgeschrieben erhält man

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix} = a_1 \begin{pmatrix} X_{11} \\ X_{21} \\ \vdots \\ x_{m1} \end{pmatrix} + a_2 \begin{pmatrix} X_{12} \\ X_{22} \\ \vdots \\ x_{m2} \end{pmatrix} + a_3 \begin{pmatrix} X_{13} \\ X_{23} \\ \vdots \\ x_{m3} \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{pmatrix} \quad (1.16)$$

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

$$\mathbf{y} = a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 + a_3 \mathbf{x}_3 + \mathbf{e} \quad (1.17)$$

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. Das 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 *Zufallsvektoren*. Natürlich ist \mathbf{e} ebenfalls ein Zufallsvektor. Ist \mathbf{x} ein Zufallsvektor, so ist $\mathbb{E}(\mathbf{x}) = (\mathbb{E}(x_1), \dots, \mathbb{E}(x_n))'$ der zugehörige Erwartungswertvektor; die $\mathbb{E}(x_i)$ sind die Erwartungswerte der Komponenten.

Man bemerke, dass \mathbf{y} als Addition von Vektoren analog zum in Abbildung 3 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, das in Abb. 3 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. \square

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 1.3 *Es seien \mathbf{x} und \mathbf{y} n -dimensionale Vektoren. Dann heißt*

$$\mathbf{x}'\mathbf{y} = (x_1, x_2, \dots, x_n) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = x_1y_1 + \dots + x_ny_n = \sum_{i=1}^n x_iy_i \quad (1.18)$$

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

$$\mathbf{x}\mathbf{y}' = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} (y_1, \dots, y_m) = \begin{pmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_m \\ x_2y_1 & x_2y_2 & \cdots & x_2y_m \\ \vdots & \vdots & \ddots & \vdots \\ x_ny_1 & x_ny_2 & \cdots & x_ny_m \end{pmatrix} \quad (1.19)$$

dyadisches Produkt oder äußeres Produkt von \mathbf{x} und \mathbf{y} ; natürlich ist $m = n$ möglich.

Man beachte, dass jeweils einer der Vektoren, deren Produkt bestimmt werden soll, als Zeilenvektor angeschrieben wurde. Diese Schreibweise erleichtert die Einführung der Multiplikation von Matrizen. Es gibt andere Schreibweisen, bei denen dieser Aspekt nicht berücksichtigt wird, 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

$$\mathbf{x}'\mathbf{x} = \sum_{i=1}^n x_i^2. \quad (1.20)$$

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.

$$\sqrt{\mathbf{x}'\mathbf{x}} = \|\mathbf{x}\| \quad (1.21)$$

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

$$|\lambda| = \frac{1}{\|\mathbf{x}\|} > 0 \quad (1.22)$$

Der Vektor

$$\mathbf{y} = \left(\frac{1}{\|\mathbf{x}\|}x_1, \frac{1}{\|\mathbf{x}\|}x_2, \dots, \frac{1}{\|\mathbf{x}\|}x_n \right)' \quad (1.23)$$

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 1.4 weiter spezifiziert.

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

$$\mathbf{x} = X - \bar{x}\vec{1} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} - \begin{pmatrix} \bar{x} \\ \bar{x} \\ \vdots \\ \bar{x} \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad (1.24)$$

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

$$Kov = \frac{1}{n} \mathbf{x}'\mathbf{y} = \frac{1}{n} \sum_{i=1}^n x_i y_i = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{x})(Y_i - \bar{y}) \quad (1.25)$$

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

$$s_x^2 = \frac{1}{n} \mathbf{x}'\mathbf{x} = \frac{1}{n} \|\mathbf{x}\|^2 \quad (1.26)$$

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

$$r_{xy} = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} \quad (1.27)$$

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

Der Kosinussatz:⁴ Es gilt

$$\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\|\mathbf{x}\|\|\mathbf{y}\| \cos \theta \quad (1.28)$$

(vergl. Abbildung 4 (b)). Für $\theta = \pi/2$, also für einen Winkel von 90° , 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

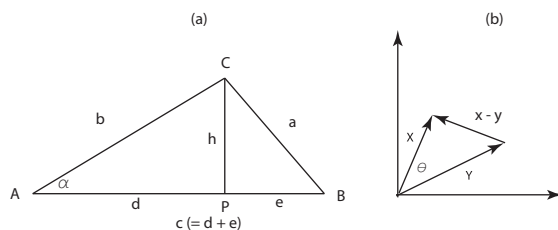
$$\begin{aligned} \|\mathbf{x} - \mathbf{y}\|^2 &= \sum_i (x_i - y_i)^2 = \sum_i x_i^2 + \sum_i y_i^2 - 2 \sum_i x_i y_i \\ &= \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\mathbf{x}'\mathbf{y}. \end{aligned}$$

⁴**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. 4 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 $a^2 = b^2 + c^2 - 2bc \cos \alpha$. \square

Abbildung 4: Zum Kosinussatz $a^2 = b^2 + c^2 - 2bc \cos \alpha$



Setzt man diesen Ausdruck für $\|\mathbf{x} - \mathbf{y}\|^2$ in (1.28) ein, so wird man auf die Beziehung

$$\mathbf{x}'\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\| \cos \theta \quad (1.29)$$

geführt. Mit Bezug auf (1.27) folgt daraus

$$\cos \theta = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = r_{xy}. \quad (1.30)$$

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 (1.30) als *Definition* des Kosinus des Winkels und damit des Winkels selbst zwischen zwei Vektoren aufgefasst werden; jedenfalls ist die Korrelation r_{xy} gleich dem Kosinus des Winkels θ zwischen den Vektoren \mathbf{x} und \mathbf{y} . Diese Gleichung impliziert die *Cauchy-Schwarzsche Ungleichung*:

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

$$(\mathbf{x}'\mathbf{y})^2 \leq \|\mathbf{x}\|^2 \|\mathbf{y}\|^2. \quad (1.31)$$

Beweis: Nach (1.29) gilt

$$\mathbf{x}'\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\| \cos \theta, \quad (1.32)$$

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 (1.31). \square

Korollar 1.1 *Es gilt*

$$\mathbf{x}'\mathbf{y} = \|\mathbf{y}\|\|\mathbf{x}\| \quad (1.33)$$

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

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. \square

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

$$\left| \sum_{i=1}^n x_i y_i \right|^2 \leq \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2 \quad (1.34)$$

angeschrieben. Für den Fall $\mathbf{y} = \lambda\mathbf{x}$, $\lambda \in \mathbb{R}$, hat man

$$(\mathbf{x}'\mathbf{y})^2 = \|\mathbf{x}\|\|\mathbf{y}\|, \text{ d.h. } \left| \sum_{i=1}^n x_i y_i \right|^2 = \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2 \quad (1.35)$$

gilt (s.a. (1.32)).

Für $\theta = \pi/2$ (90°) ist $\cos \theta = 0$; (1.29) impliziert dann $\mathbf{x}'\mathbf{y} = 0$; die Vektoren \mathbf{x} und \mathbf{y} sind dann *orthogonal*⁵. Variablen X und Y , deren Korrelation gleich Null ist, werden durch orthogonale Vektoren \mathbf{x} und \mathbf{y} repräsentiert.

Anmerkungen:

1. 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 Körperlänge 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 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.

⁵Von griechisch *orthogonios*: ortho = rechts, gon = winklig

2. 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 zufällige Veränderliche, die 2-dimensional normalverteilt seien, so dass die Dichte für $X = x$ und $Y = y$ durch

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} \exp\left[-\frac{1}{2(1-r^2)}(z_x^2 + z_y^2 - 2rz_xz_y)\right] \quad (1.36)$$

gegeben ist 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 Vektorräume

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

Definition 1.4 Es sei $V = \{\mathbf{x} | \mathbf{x} \in \mathbb{R}^n\}$ eine Menge n -dimensionaler Vektoren, und für beliebige $\mathbf{x}_1, \mathbf{x}_2 \in V$ gelte

1. $\mathbf{x} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 \in V$ mit $a_1, a_2 \in \mathbb{R}$,
 2. Für $\mathbf{x}_1, \mathbf{x}_2 \in V$ existiert das Skalarprodukt $\mathbf{x}'_1\mathbf{x}_2$.
- Dann heißt V n -dimensionaler Vektorraum.⁶

Anmerkung: Es sei M die Menge von Vektoren mit der Länge, etwa $\|\mathbf{x}_j\| = 1$. 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 $\mathbf{x} \in M$. \square

Auf Seite 15 wurde der Begriff der Norm eines Vektors eingeführt. Allgemein ist eine Norm⁷ eine Abbildung eines Vektorraums in die Menge der nicht-negativen reellen Zahlen \mathbb{R}_+ ; sie wird mit $\|\cdot\|$ bezeichnet:

$$\|\cdot\| : V \rightarrow \mathbb{R}_+, \quad \mathbf{x} \mapsto \|\mathbf{x}\| \in \mathbb{R} \text{ für } \mathbf{x} \in V. \quad (1.37)$$

also. Normen haben die folgenden Eigenschaften: für alle $\mathbf{x}, \mathbf{y} \in V$ und alle $a \in \mathbb{R}$ gelten

⁶Diese 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.

⁷Von lat. *norma* = Richtschnur

1. *Definitheit*: $\|\mathbf{x}\| = 0 \rightarrow \mathbf{x} = \vec{0}$,
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)

Man prüft leicht nach, dass die Länge eines Vektors die Eigenschaft einer Norm hat.

Definition 1.5 *Es sei V ein n -dimensionaler Vektorraum, und es seien $\mathbf{x}, \mathbf{y} \in V$. Dann heißt die Norm der Differenz*

$$\|\mathbf{x} - \mathbf{y}\| = \left(\sum_{i=1}^n (x_i - y_i)^2 \right)^{1/2} \quad (1.38)$$

die euklidische Distanz zwischen \mathbf{x} und \mathbf{y} , und V heißt auch euklidischer Vektorraum (vergl. (1.4)).

$\|\mathbf{x} - \mathbf{y}\|$ ist ein Vektor, dessen Länge der Distanz zwischen den Endpunkten von \mathbf{x} und \mathbf{y} entspricht. Als Übung überprüfe man, dass $\|\mathbf{x} - \mathbf{y}\|$ tatsächlich den Distanzaxiomen entspricht; die Metrik ist die *euklidische Metrik*. Dass hier die euklidische Metrik so hervorgehoben wird, liegt an der Definition des Skalarproduktes $\mathbf{x}'\mathbf{y}$, das in Punkt 2. der Definition 1.4 angenommen wurde. Es lässt sich eine andere Form des Skalarproduktes definieren:

$$\langle \mathbf{x}, \mathbf{y} \rangle_p = \sum_{i=1}^n |x_i y_i|^p, \quad 0 \leq p \in \mathbb{R}. \quad (1.39)$$

Eine Distanzfunktion ist dann durch die *Minkowski-Metrik* definiert,

$$d(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^n (x_i - y_i)^p \right)^{1/p} \quad (1.40)$$

(vergl. (1.5), Seite 9). Minkowski-Distanzen spielen in diesem Skript keine weitere Rolle.

Definition 1.6 *Es sei V ein n -dimensionaler Vektorraum und $U \subset V$ sei eine Teilmenge von Vektoren aus V . Für beliebige Vektoren $\mathbf{x}_1, \mathbf{x}_2 \in U$ gelte $\mathbf{x} = a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 \in U$, mit $a_1, a_2 \in \mathbb{R}$, und es existiere das Skalarprodukt $\mathbf{x}'_1 \mathbf{x}_2$. Dann heißt U Teilvektorraum von V .*

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 1.4 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.

Beispiel 1.2 Es seien $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ irgendzwei parallele Vektoren; diese beiden Vektoren definieren eine Gerade, deren Orientierung mit der der Vektoren identisch ist, d.h. sie definieren einen 1-dimensionalen Teilraum von $V = \mathbb{R}^n$. Denn es gilt nun $\mathbf{x}_2 = a\mathbf{x}_1$, $a \in \mathbb{R}$. Dann folgt mit $a_1, a_2 \in \mathbb{R}$

$$\mathbf{x} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 = a_1\mathbf{x}_1 + aa_2\mathbf{x}_1 = (a_1 + aa_2)\mathbf{x}_1, \quad a_1 + aa_2 \in \mathbb{R}$$

d.h. \mathbf{x} ist wieder parallel zu \mathbf{x}_1 und \mathbf{x}_2 , d.h. Es gibt Repräsentationen \mathbf{x}_1 und \mathbf{x}_2 des Vektors, die auf derselben Geraden liegen. Man sagt auch, die Vektoren seien *kollinear*. Die Gerade ist ein Teilraum des \mathbb{R}^n .

Es sei $n \geq 3$ und \mathbf{x}_1 und \mathbf{x}_2 seien nicht parallel. Sie definieren dann eine Ebene \mathcal{E} im \mathbb{R}^n und damit einen 2-dimensionalen Teilraum von \mathbb{R}^n . Denn es sei \mathbf{n} ein Vektor, der senkrecht auf \mathbf{x}_1 und \mathbf{x}_2 steht, d.h. der orthogonal zu diesen beiden Vektoren ist, und es sei \mathbf{x} eine Linearkombination von \mathbf{x}_1 und \mathbf{x}_2 , also $\mathbf{x} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2$. Dann folgt

$$\mathbf{n}'\mathbf{x} = a_1\mathbf{n}'\mathbf{x}_1 + a_2\mathbf{n}'\mathbf{x}_2 = 0, \quad (1.41)$$

da ja nach Definition von \mathbf{n} die Beziehungen $\mathbf{n}'\mathbf{x}_1 = \mathbf{n}'\mathbf{x}_2 = 0$ gelten. \mathbf{n} heißt *Normalenvektor* für \mathcal{E} ; die Orientierung von \mathbf{n} bestimmt die Orientierung der Ebene; der Nachweis, dass ein solcher Vektor existiert, wird auf Seite 34 geführt werden. Man erhält die

Ebenengleichung

$$\mathbf{n}'\mathbf{x} = n_1x_1 + n_2x_2 + \cdots + n_nx_n = 0. \quad (1.42)$$

Alle Punkte mit Koordinaten x_1, x_2, \dots, x_n , die für festen Vektor \mathbf{n} dieser Gleichung genügen, liegen in der Ebene \mathcal{E} . (1.42) ist die Gleichung einer Ebene im \mathbb{R}^n . \mathcal{E} geht durch den Nullpunkt des Koordinatensystems⁸. \mathcal{E} ist ein Teilraum des Vektorraums $V = \mathbb{R}^n$. Denn es seien

$$\mathbf{y}_1 = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 \quad (1.43)$$

$$\mathbf{y}_2 = b_1\mathbf{x}_1 + b_2\mathbf{x}_2 \quad (1.44)$$

irgendzwei Linearkombinationen von \mathbf{x}_1 und \mathbf{x}_2 . Dann ist auch die Linearkombination $\mathbf{y} = c_1\mathbf{y}_1 + c_2\mathbf{y}_2$ wieder ein Element von \mathcal{E} , denn

$$\mathbf{y} = c_1(a_1\mathbf{x}_1 + a_2\mathbf{x}_2) + c_2(b_1\mathbf{x}_1 + b_2\mathbf{x}_2) = d_1\mathbf{x}_1 + d_2\mathbf{x}_2,$$

mit

$$d_1 = c_1a_1 + c_2b_1, \quad d_2 = c_1a_2 + c_2b_2,$$

und $\mathbf{n}'\mathbf{y} = 0$, analog zu (1.42). Auf die oben gemachte Voraussetzung, dass \mathbf{x}_1 und \mathbf{x}_2 nicht parallel sind, d.h. nicht dieselbe Orientierung (in Zeichen: \parallel) haben, wird im Folgenden eingegangen.

⁸Diese Definition läßt sich für Ebenen, die nicht durch den Nullpunkte gehen, verallgemeinern, aber diese allgemeine Definition wird im Folgenden nicht benötigt.

Die Vektoren \mathbf{x}_1 und \mathbf{x}_2 bilden eine mögliche *Basis* für den Teilraum \mathcal{E} . Der Begriff der Basis wird noch explizit definiert werden, hier sei nur angemerkt, dass die Basis für einen Raum oder Teilraum nicht eindeutig ist, denn irgendzwei andere, nicht parallele Vektoren \mathbf{y}_1 und \mathbf{y}_2 aus \mathcal{E} erlauben ebenfalls, *alle* Vektoren aus \mathcal{E} zu erzeugen. Nun sei $\mathbf{x}_1 \nparallel \mathbf{x}_2$ und $\mathbf{n} \perp \mathbf{x}_1$ und $\mathbf{n} \perp \mathbf{x}_2$, wobei \nparallel für 'nicht parallel' und \perp für "ist orthogonal zu" stehen. Dann lassen sich alle 3-dimensionalen Vektoren \mathbf{x} als Linearkombinationen von \mathbf{x}_1 , \mathbf{x}_2 und \mathbf{n} darstellen, d.h es existieren Koeffizienten a_1, a_2 und a_3 derart, dass

$$\mathbf{x} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{n}, \quad \text{für alle } \mathbf{x} \in \mathbb{R}^3, \quad \mathbf{x} \neq \vec{0} \quad (1.45)$$

□

1.5 Lineare Abhängigkeit und Unabhängigkeit

Der Begriff der linearen Abhängigkeit bzw. Unabhängigkeit ist von zentraler Bedeutung für die lineare Algebra und deren Anwendung in der Multivariaten Statistik: eine Menge $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ 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. Eine formale Definition der Abhängigkeit ist allerdings notwendig, um diesen Begriff für die jeweiligen Betrachtungen umzusetzen.

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:

$$\vec{0} = a_1\mathbf{x}_1 + \dots + a_p\mathbf{x}_p. \quad (1.46)$$

Man kann die a_1, \dots, a_p als Unbekannte eines Gleichungssystems betrachten (s. unten). Es existiert stets eine Lösung für diese Gleichung:

$$a_1 = a_2 = \dots = a_p = 0. \quad (1.47)$$

Weil diese Lösung *stets* eine Lösung ist, kann man sie als triviale Lösung bezeichnen. Die Frage ist nun, ob auch andere Lösungen existieren, bei denen mindestens ein Koeffizient $a_j \neq 0$ ist, – dann ist mindestens ein weiterer Koeffizient ungleich Null, wie man sich leicht überlegt.

Es sei etwa $a_p \neq 0$, z.B. $a_p = 1$. Dann existiert mindestens ein weiterer Koeffizient $a_j \neq 0$, $1 \leq j \leq p - 1$ (wäre dies nicht so, würde entgegen der Voraussetzung $a_p = 0$ folgen), und man erhält

$$\mathbf{x}_p = a_1\mathbf{x}_1 + \dots + a_{p-1}\mathbf{x}_{p-1} \quad (1.48)$$

Man hat dann die

Definition 1.7 Gilt (1.46) und sind nicht alle a_j gleich Null, so heißen die Vektoren \mathbf{x}_j linear abhängig. Ist dagegen (1.47) die einzige Lösung für die a_j , so heißen die \mathbf{x}_j linear unabhängig.

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

Lineare Gleichungssysteme: Die Frage nach linearer Abhängigkeit bzw. Unabhängigkeit ist eng mit der Frage nach der Lösbarkeit linearer Gleichungssysteme verbunden. Der Einfachheit halber wird hier der Fall 2-dimensionaler Vektoren betrachtet.

Gegeben seien drei 2-dimensionale Vektoren \mathbf{y} , \mathbf{x}_1 und \mathbf{x}_2 . 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

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 = a_1 \begin{pmatrix} x_{11} \\ x_{21} \end{pmatrix} + a_2 \begin{pmatrix} x_{12} \\ x_{22} \end{pmatrix} \quad (1.49)$$

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

$$y_1 = a_1 x_{11} + a_2 x_{12} \quad (1.50)$$

$$y_2 = a_1 x_{21} + a_2 x_{22} \quad (1.51)$$

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

$$a_1 = \frac{y_1 x_{22} - y_2 x_{12}}{x_{11} x_{22} - x_{12} x_{21}} \quad (1.52)$$

$$a_2 = \frac{y_2 x_{11} - y_1 x_{21}}{x_{11} x_{22} - x_{12} x_{21}} \quad (1.53)$$

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

$$x_{11} x_{22} - x_{12} x_{21} \neq 0 \quad (1.54)$$

gegeben ist. Denn $x_{11} x_{22} - x_{12} x_{21} = 0$ würde bedeuten, dass durch 0 dividiert werden muß, damit man eine Lösung erhält, – und diese Operation macht bekanntlich keinen Sinn. Gleichwohl ist dieser Fall möglich, und seine Implikation ist für das Folgende von Bedeutung.

Es wird also der Fall

$$x_{11} x_{22} - x_{12} x_{21} = 0 \quad (1.55)$$

betrachtet. Er impliziert

$$\frac{x_{21}}{x_{11}} = \frac{x_{22}}{x_{12}} \quad (1.56)$$

Aber $x_{21}/x_{11} = \tan \theta$, θ der Winkel, der die Orientierung von \mathbf{x}_1 angibt, und (1.56) besagt, dass dieser Winkel identisch mit dem von \mathbf{x}_2 ist. Daraus folgt,

dass (1.55) dann erfüllt ist, wenn \mathbf{x}_1 und \mathbf{x}_2 parallel (kollinear) sind. Dann gilt $\mathbf{x}_2 = \lambda \mathbf{x}_1$ mit $\lambda \in \mathbb{R}$ und die Gleichung (1.49) geht über in die Gleichung

$$\mathbf{y} = a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 = a_1 \mathbf{x}_1 + a_2 \lambda \mathbf{x}_1 = (a_1 + a_2 \lambda) \mathbf{x}_1, \quad (1.57)$$

aus der sofort hervorgeht, dass (1.49) unter der Bedingung (1.55) keine Lösung hat, wenn \mathbf{y} nicht dieselbe Orientierung wie \mathbf{x}_1 hat. Gilt (1.55) und sind \mathbf{y} und \mathbf{x}_1 parallel, so liefern (1.52) und (1.53) keine eindeutige Lösung (a_1, a_2) , denn wegen (1.57) kann nur der Parameter $\alpha = a_1 + a_2 \lambda$ bestimmt werden. Diese Betrachtung illustriert die Tatsache, dass es nicht möglich ist, aus zwei Vektoren mit identischer Orientierung einen Vektor zu erzeugen, der eine andere Orientierung hat. a_1 und a_2 liegen nicht mehr eindeutig fest, denn $\mathbf{y} = (a_1 + a_2 \lambda) \mathbf{x}_1$ ist ja für alle a_1, a_2 erfüllt, für die $a_1 + a_2 \lambda = \alpha$, α eine Konstante, erfüllt ist, für die also

$$a_1 = \alpha - a_2 \lambda \quad (1.58)$$

gilt. 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. In Satz 1.7, Seite 27, wird über die Beziehung zwischen dem Winkel zwischen zwei Vektoren und ihrer linearen Abhängigkeit bzw. Unabhängigkeit eine allgemeine Aussage gemacht.

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 ist

$$|X| = x_{11}x_{22} - x_{12}x_{21}$$

die *Determinante* der (2×2) -Matrix X , s. Anhang, Abschnitt 4.6. 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 folgenden Betrachtungen zeigen. Diese Aussage gilt allgemein für $(n \times n)$ -Matrizen und damit für Systeme von linearen Gleichungen mit n Unbekannten. Für die folgenden Betrachtungen ist aber der Begriff der Determinante nicht notwendig, weshalb an dieser Stelle nicht weiter auf ihn eingegangen wird. \square

In diesem Beispiel werden 2-dimensionale Vektoren $\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}$ betrachtet. Die Frage ist jetzt noch, ob es einen 2-dimensionalen Vektor \mathbf{y} geben kann, der *nicht* als Linearkombination von \mathbf{x}_1 und \mathbf{x}_2 dargestellt werden kann. Einen solchen Vektor gibt es nicht, d.h. sind irgendzwei nicht-parallele, also linear unabhängige Vektoren $\mathbf{x}_1, \mathbf{x}_2$ gegeben, so läßt sich *jeder* 2-dimensionale Vektor \mathbf{y} als Linearkombination dieser beiden Vektoren darstellen. Diese Aussage folgt einfach daraus, dass das Gleichungssystem (1.50) und (1.51) bei linearer Unabhängigkeit

der $\mathbf{x}_1, \mathbf{x}_2$ für *alle* Vektoren \mathbf{y} eine durch (1.52) und (1.53) gegebene Lösung für die Koeffizienten a_1 und a_2 liefert. Es gibt also keine Menge von drei oder mehr 2-dimensionalen Vektoren, die insgesamt linear unabhängig sind.

Dies ist ein Spezialfall des allgemeinen Falls: Es sei $\mathbf{x}_j \in \mathbb{R}^n$ für $j = 1, \dots, p \leq n$. Sei weiter $\mathbf{y} \in \mathbb{R}^n$ und die Frage sei, ob \mathbf{y} als Linearkombination der \mathbf{x}_j dargestellt werden kann, so dass

$$\mathbf{y} = a_1\mathbf{x}_1 + \dots + a_p\mathbf{x}_p; \quad (1.59)$$

Die Frage ist, ob die Koeffizienten a_1, \dots, a_p existieren derart, dass (1.59) gilt. Diese Vektorgleichung entspricht einem linearen Gleichungssystem mit den a_1, \dots, a_p als Unbekannten: für die i -te Komponente von \mathbf{y} hat man die Gleichung

$$y_i = a_1x_{i1} + a_2x_{i2} + \dots + a_px_{ip}, \quad i = 1, \dots, n \quad (1.60)$$

Zunächst kann man feststellen, dass keine Menge $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ von $p > n$ n -dimensionalen Vektoren, die linear unabhängig sind, d.h. es können jeweils maximal n Vektoren aus einer gegebenen Menge von Vektoren $\mathbf{x}_j \in \mathbb{R}^n$ linear unabhängig sein. Diese Aussage wird gelegentlich auch als *Fundamental-Lemma* der linearen Algebra bezeichnet (Koecher (1997), p. 21). Der Hintergrund für diese Bezeichnung ist, dass die Herleitung vieler Ergebnisse der Linearen Algebra von diesem Lemma Gebrauch macht. Das Lemma wird weiter unten noch ausführlicher diskutiert (vergl. auch Kommentar 2 zu Satz 1.10 auf Seite 31).

Dann stellt sich die Frage, unter welchen Bedingungen es eine eindeutige Lösung für die Unbekannten Koeffizienten a_1, \dots, a_p gibt. Darüber gibt der folgende Satz Auskunft; dieser Satz erweist sich als nützlich, wenn Beweise weiterer Aussagen der Vektor- und Matrixrechnung geführt werden sollen.

Satz 1.3 *Es werde die Vektorgleichung (1.59) betrachtet. Sind die Vektoren $\mathbf{x}_1, \dots, \mathbf{x}_p$ linear unabhängig, so sind die a_1, \dots, a_p eindeutig bestimmt.*

Beweis: Die lineare Unabhängigkeit der \mathbf{x}_j impliziert $a_1\mathbf{x}_1 + \dots + a_p\mathbf{x}_p = \vec{0}$ nur dann, wenn $a_1 = \dots = a_p = 0$. Nun gelte

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

Dann ist

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

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

Der Fall der linearen Abhängigkeit der \mathbf{x}_j wird in Abschnitt 2.8 behandelt.

Satz 1.4 Die Vektoren $\mathbf{x}_1, \dots, \mathbf{x}_n$, $\mathbf{x}_j \neq \vec{0}$ für alle j , seien paarweise orthogonal. Dann sind sie linear unabhängig.

Beweis: Es sei

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

Dann gilt

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

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. \square

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

Korollar 1.2 Die Einheitsvektoren $\mathbf{e}_1, \dots, \mathbf{e}_n$ sind paarweise orthogonal.

Beweis: Es gilt

$$\mathbf{e}'_j\mathbf{e}_k = \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases} \quad (1.61)$$

wie man unmittelbar verifiziert. \square

Nach Satz 1.4 sind die \mathbf{e}_j , $j = 1, \dots, n$ linear unabhängig.

Satz 1.5 Es sei \mathbf{x} ein beliebiger n -dimensionaler Vektor. Dann ist \mathbf{x} als Linearkombination der \mathbf{e}_j darstellbar.

Beweis: Es ist

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

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

Der folgende Satz in diesem Abschnitt erweist sich als nützlich für manche Beweise bzw. Herleitungen:

Satz 1.6 Es sei $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ eine Menge von Vektoren aus dem \mathbb{R}^n . Ist einer von ihnen der Nullvektor, so sind die \mathbf{x}_j , $j = 1, \dots, p$ linear abhängig.

Beweis: Die Vektoren $\mathbf{x}_1, \dots, \mathbf{x}_{p-1}$ seien linear unabhängig, und $\mathbf{x}_p = \vec{0}$. Weiter sei

$$a_1\mathbf{x}_1 + \dots + a_{p-1}\mathbf{x}_{p-1} + a_p\mathbf{x}_p = \vec{0}$$

für $a_1 = \dots = a_{p-1} = 0$, $a_p \neq 0$, da $a_p\vec{0} = \vec{0}$ auch für $a_p \neq 0$. Also ist die Menge $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ von Vektoren linear abhängig. \square

Der folgende Satz charakterisiert die Beziehung zwischen dem Skalarprodukt zweier Vektoren und ihrer linearen Abhängigkeit bzw. Unabhängigkeit.

Satz 1.7 *Gegeben seien zwei n -dimensionale Vektoren \mathbf{x} und \mathbf{y} . Der Winkel zwischen ihnen sei θ . \mathbf{x} und \mathbf{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 ($\mathbf{x} \parallel \mathbf{y}$), so dass ein $a \in \mathbb{R}$ existiert derart, dass $\mathbf{y} = a\mathbf{x}$, 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 1.5) und es ist

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

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

Nun sei $\mathbf{x} \not\parallel \mathbf{y}$ (\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 $|\cos \theta| < 1$. \square

Beispiel 1.3 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 messfehlr repräsentieren. Der Korrelationskoeffizient r_{xy} entspricht dann

$$-1 < r_{xy} = \cos \theta_{xy} = \frac{(\mathbf{x}_0 + \vec{\varepsilon}_1)'(a\mathbf{x}_0 + \vec{\varepsilon}_2)}{\|\mathbf{x}_0 + \vec{\varepsilon}_1\|\|a\mathbf{x}_0 + \vec{\varepsilon}_2\|} < 1, \quad \varepsilon_1, \varepsilon_2 \neq 0 \quad (1.62)$$

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

bekanntem Sachverhalt, dass die Wahrscheinlichkeit, einen *bestimmten* Wert einer stetigen zufälligen Veränderlichen zu messen stets gleich Null ist. Für die gemessenen Vektoren \mathbf{x}_j , $j = 1, \dots, b$ heißt dies, dass *rechnerisch* alle mit Messfehlern behafteten Vektoren linear unabhängig sind. \square

1.6 Erzeugendensysteme und Basen von Vektorräumen

Definition 1.8 *Es sei $M = \{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ eine Menge von n -dimensionalen Vektoren. Es sei*

$$\mathcal{L}(M) = \{\mathbf{x} \mid \mathbf{x} = a_1 \mathbf{x}_1 + \dots + a_m \mathbf{x}_m\}, \quad (1.63)$$

d.h. $\mathcal{L}(M)$ sei die Menge aller Linearkombinationen der Vektoren aus M . $\mathcal{L}(M)$ heißt die lineare Hülle von M , und die $\mathbf{x}_1, \dots, \mathbf{x}_m$ bilden ein Erzeugendensystem für $\mathcal{L}(M)$.

Es sei $M \subseteq V$; M muß kein Teilraum sein, aber es gilt der

Satz 1.8 *Es sei V ein Vektorraum und $M = \{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ sei eine Teilmenge von Vektoren aus V . Dann gilt $\mathcal{L}(M) \subseteq V$, d.h. $\mathcal{L}(M)$ ist ein Teilvektorraum von V oder gleich V .*

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

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

Dann folgt

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

d.h. $\lambda \mathbf{x} + \mu \mathbf{y}$ ist ebenfalls eine Linearkombination der \mathbf{x}_j und damit Element von $\mathcal{L}(M)$. \square

Definition 1.9 *Es sei V_n ein n -dimensionaler Vektorraum. Die Menge $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ aus V_n heißt Basis von V_n , wenn gilt*

- (i) *die $\mathbf{b}_1, \dots, \mathbf{b}_n$ sind linear unabhängig,*
- (ii) *$V = \mathcal{L}(\mathcal{B})$, dh. V_n ist die lineare Hülle von $\mathbf{b}_1, \dots, \mathbf{b}_n$. Die Teilmenge $\mathcal{B}_r = \{\mathbf{b}_1, \dots, \mathbf{b}_r\}$ mit $r < n$ bildet eine Teilbasis von \mathcal{L} .*
- (iii) *Es sei $\mathbf{v} \in V_n$ und es gelte*

$$\mathbf{v} = a_1 \mathbf{b}_1 + a_2 \mathbf{b}_2 + \dots + a_n \mathbf{b}_n. \quad (1.64)$$

Die Koeffizienten a_1, \dots, a_n heißen Koordinaten von \mathbf{v} bezüglich \mathcal{B} .

Anmerkungen:

1. Hier werden nur endlich-dimensionale Vektorräume betrachtet. Die Definition einer Basis bedeutet dann, dass immer nur endlich viele Vektoren eine Basis bilden. Diese Basen werden auch als *Hamel-Basen*⁹ bezeichnet, im Unterschied zum Beispiel der *Schauder-Basis*¹⁰, die im Zusammenhang mit unendlich-dimensionalen (Funktionen-)Räumen betrachtet wird. Funktionenräume sind Mengen von Funktionen, die denselben Definitionsbereich haben und unter bestimmten Bedingungen Vektorräume bilden, s. Abschnitt 3.
2. Wegen (ii) ist eine Basis ein Erzeugendensystem für V_n . Ein Erzeugendensystem muß aber nicht nur aus linear unabhängigen Vektoren bestehen; so können die Vektoren $\mathbf{x}_1, \dots, \mathbf{x}_m$, $m > n$ ein Erzeugendensystem für V_n sein, wenn $\mathcal{L}(\mathbf{x}_1, \dots, \mathbf{x}_m) = V_n$. Nach Definition der Basis $\mathbf{b}_1, \dots, \mathbf{b}_n$ eines V_n werden aber die $\mathbf{x}_1, \dots, \mathbf{x}_n$ Elemente von $\mathcal{L}(\mathbf{b}_1, \dots, \mathbf{b}_n)$ sein. Dementsprechend heißt eine Basis auch *minimales Erzeugendensystem*: wegen (ii) muß ein Erzeugendensystem mindestens n linear unabhängige Vektoren enthalten, damit *alle* Vektoren des V_n damit erzeugt werden können, andererseits kann eine Basis nicht mehr als n Elemente enthalten, denn mehr als n n -dimensionale Vektoren sind linear abhängig.
3. Es sei $\mathbf{v} = (v_1, \dots, v_n)'$; nach (1.64) gilt dann für die i -te Komponente

$$v_i = a_1 b_{i1} + \dots + a_n b_{in}, \quad (1.65)$$

wobei die b_{ij} die i -ten Komponenten der \mathbf{b}_j sind. Wählt man eine andere Basis als die $\mathbf{b}_1, \dots, \mathbf{b}_n$, so müssen auch die Koeffizienten a_1, \dots, a_n entsprechend gewählt werden, damit die v_i berechnet werden können, d.h. diese Koeffizienten hängen von der gewählten Basis ab. Dies erklärt den Ausdruck *Koordinaten von \mathbf{v} bezüglich \mathcal{B}* für die Koeffizienten a_j . \square

Der Begriff der Basis ist intuitiv schon am Ende des Beispiels 1.2 eingeführt worden. $V = \mathcal{L}(\mathcal{B})$ bedeutet, dass *jeder* Vektor aus V als Linearkombination der *Basisvektoren* $\mathbf{b}_1, \dots, \mathbf{b}_n$ darstellbar ist, d.h. für eine gegebene Basis $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ existieren für jeden Vektor $\mathbf{v} \in \mathcal{L}$ Koeffizienten a_1, \dots, a_n (also ein Vektor $\mathbf{a} = (a_1, \dots, a_n)'$) derart, dass die Darstellung eines Vektors $\mathbf{v} \in V$ wie in (1.64) möglich ist: diese Darstellung von \mathbf{v} entspricht einem linearen System von n Gleichungen mit den n Unbekannten a_1, \dots, a_n , für das es nur eine Lösung gibt, wenn die $\mathbf{b}_1, \dots, \mathbf{b}_n$ linear unabhängig sind, und es existiert keine Menge von $p > n$ n -dimensionalen Vektoren, die linear unabhängig sind (s.a. Abschnitt 2.8). $\mathcal{L}(\mathbf{b}_1, \dots, \mathbf{b}_r)$ mit $r < n$ definiert einen Teilraum von V (vergl. Satz 1.8).

⁹Georg Hamel (1877 – 1954), deutscher Mathematiker

¹⁰Juliusz Pawel Schauder (1899 – 1943), polnischer Mathematiker.

Definition 1.10 *Es sei V ein Vektorraum mit der Basis $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$. Dann heißt V n -dimensionaler Vektorraum; man schreibt auch V_n , um die Anzahl der Vektoren in einer Basis von V anzuzeigen. n heißt Dimension des Vektorraums.*

Anmerkung: In Definition 1.10 wird der Begriff des n -dimensionalen Vektorraums durch die Anzahl der Basisvektoren definiert. Wie bereits angedeutet existiert ein Zusammenhang zwischen der Anzahl n der Komponenten der Vektoren eines Vektorraums V und der Anzahl der Basisvektoren, die notwendig sind, um alle Vektoren von V zu erzeugen. Dieser Zusammenhang wird im Folgenden elaboriert. Zuvor wird aber der Begriff der orthogonalen Basis eingeführt. \square

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 wichtig, weshalb eine eigene Definition dafür eingeführt wird:

Definition 1.11 *Es sei V ein n -dimensionaler Vektorraum. Eine Basis*

$$\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$$

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

$$\mathbf{b}'_j \mathbf{b}_k = \begin{cases} 0, & j \neq k \\ 1, & j = k \end{cases}, \quad j, k = 1, \dots, n \quad (1.66)$$

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

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

Satz 1.9 *Die n -dimensionalen Einheitsvektoren $\mathbf{e}_1, \dots, \mathbf{e}_n$ mit*

$$\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)'$$

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

Beweis: Die Einheitsvektoren sind linear unabhängig, denn $\vec{0} = \lambda_1 \mathbf{e}_1 + \lambda_2 \mathbf{e}_2 + \dots + \lambda_n \mathbf{e}_n$ ist nur möglich für $\lambda_1 = \dots = \lambda_n = 0$; 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. (1.61). Die Vektoren $\mathbf{e}_1, \dots, \mathbf{e}_n$ bilden deshalb eine orthonormale Basis des V_n . Da stets

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

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

Definition 1.12 Die Basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ heißt kanonische Basis des V_n .

Eine beliebige orthonormale Basis läßt sich als Rotation der kanonischen Basis herleiten, vergl. den Abschnitt 2.6 über Basiswechsel, Seite 91.

Satz 1.10 Es seien $\mathbf{x}_1, \dots, \mathbf{x}_n$ linear unabhängige n -dimensionale Vektoren. Dann lassen sich alle Vektoren des \mathbb{R}^n als Linearkombinationen dieser Vektoren erzeugen.

Beweis: Nach Satz 1.5 kann jeder Vektor \mathbf{x} als Linearkombination der Einheitsvektoren $\mathbf{e}_1, \dots, \mathbf{e}_n$ dargestellt werden, und damit auch die \mathbf{x}_j ; $j = 1, \dots, n$. Dementsprechend hat man

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

Da die \mathbf{x}_j als Linearkombinationen der \mathbf{e}_k dargestellt werden können, kann auch die Linearkombination \mathbf{x} der \mathbf{x}_j wieder als Linearkombination der \mathbf{e}_k dargestellt werden. \square

Der Satz 1.10 bedeutet,

1. dass die Bedingung (ii) von Definition 1.9 keine Einschränkung darstellt; jede Menge von n linear unabhängigen n -dimensionalen Vektoren kann als Basis des $V_n = \mathbb{R}^n$ gewählt werden.
2. *Fundamental-Lemma*: Gegeben seien m n -dimensionale Vektoren $\mathbf{x}_1, \dots, \mathbf{x}_m$ und es sei $m > n$. Dann sind die \mathbf{x}_j , $j = 1, \dots, m$ linear abhängig. Denn angenommen, die ersten n dieser Vektoren seien linear unabhängig. Nach Satz 1.10 lassen sich dann *alle* Vektoren des \mathbb{R}^n als Linearkombinationen der $\mathbf{x}_1, \dots, \mathbf{x}_n$ darstellen, und das heißt eben auch die $\mathbf{x}_{n+1}, \dots, \mathbf{x}_m$. So existieren Koeffizienten a_1, \dots, a_n derart, dass $\mathbf{x}_{n+1} = \mathbf{v} = a_1 \mathbf{x}_1 + \dots + a_n \mathbf{x}_n$. Wären die Vektoren \mathbf{x}_j , $j = 1, \dots, n + 1$ linear unabhängig, so wäre die Darstellung des Nullvektors

$$a_1 \mathbf{x}_1 + \dots + a_n \mathbf{x}_n + a_{n+1} \mathbf{x}_{n+1} = \vec{0}$$

nur möglich für den Fall $a_1 = \dots = a_{n+1} = 0$. Aber $\mathbf{v} + a_{n+1} \mathbf{x}_{n+1} = \vec{0}$ ist eben wegen $\mathbf{v} = \mathbf{x}_{n+1}$ – möglich für $1 = -a_{n+1} \neq 0$, d.h. \mathbf{x}_{n+1} kann nicht linear unabhängig von den $\mathbf{x}_1, \dots, \mathbf{x}_n$ sein. Dies bedeutet, dass eine Basis des \mathbb{R}^n nie mehr als n Basisvektoren enthält. Einen Beweis auf der Basis des Austauschsatzes von Steiner, auf den hier nicht weiter eingegangen werden soll, findet man im Anhang, Abschnitt 4.1.

Die Frage ist nun, ob jeder Vektorraum auch eine Basis haben muß. Man könnte die Möglichkeit betrachten, dass alle Vektoren eines Vektorraums linear abhängig sind. Dazu läßt sich eine Aussage beweisen:

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

Beweis: "Endlich erzeugt" soll heißen, dass für $\mathcal{L}(\mathbf{x}_1, \dots, \mathbf{x}_n) = V$ die Bedingung $n < \infty$ erfüllt ist. Nach dem obigen Fundamental-Lemma ist jede Menge von $n + 1$ Vektoren eines n -dimensionalen Vektorraums linear abhängig. Sei nun $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ eine Menge von n -dimensionalen Vektoren. Dann gilt

- (i) Ein einzelner Vektor, etwa $\mathbf{b}_1 \neq \vec{0}$, ist linear unabhängig, denn $\lambda \mathbf{b}_1 = \vec{0}$ nur dann, wenn $\lambda_1 = 0$,
- (ii) $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ sei eine Menge von linear unabhängigen Vektoren. Dann ist sie eine Basis, da man mit den \mathbf{b}_j , $j = 1, \dots, n$, alle Vektoren des Vektorraums erzeugen kann.
- (iv) Ist $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ linear abhängig, so gilt

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

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

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

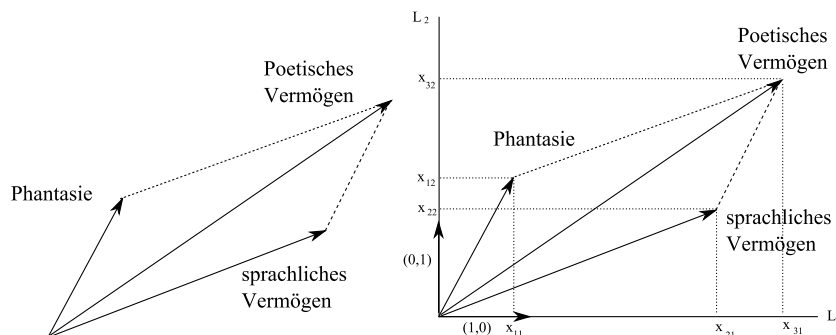
Sind die $\mathbf{b}_2, \dots, \mathbf{b}_n$ linear unabhängig, so bilden sie eine Basis des Vektorraums, und $\mathcal{L}(\mathbf{b}_2, \dots, \mathbf{b}_n) = V$. Ist $\{\mathbf{b}_2, \dots, \mathbf{b}_n\}$ linear abhängig, so kann man die Betrachtung wiederholen; ist $\{\mathbf{b}_3, \dots, \mathbf{b}_n\}$ linear unabhängig, so bildet diese Menge eine Basis und $\mathcal{L}(\mathbf{b}_3, \dots, \mathbf{b}_n) = V$, 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 $\mathbf{b}_1, \dots, \mathbf{b}_n$ seien orthonormale Basisvektoren. Dann existieren Koordinaten a_1, \dots, a_n derart, dass

$$\mathbf{x} = a_1 \mathbf{b}_1 + \dots + a_n \mathbf{b}_n = \sum_{k=1}^n a_k \mathbf{b}_k. \quad (1.67)$$

Abbildung 5: 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.



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

$$\mathbf{b}'_j \mathbf{x} = \sum_{k=1}^n a_k \mathbf{b}'_j \mathbf{b}_k = a_j, \quad j = 1, \dots, n \quad (1.68)$$

denn

$$\mathbf{b}'_j \mathbf{b}_k = \begin{cases} 0, & j \neq k \\ 1, & j = k \end{cases} \quad (1.69)$$

(1.67) kann dann in der Form

$$\mathbf{x} = \sum_{k=1}^n (\mathbf{b}'_k \mathbf{x}) \mathbf{b}_k. \quad (1.70)$$

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 (1.70) die Korrelation zwischen dem Vektor \mathbf{x} und dem Basisvektor \mathbf{b}_k . In der Faktorenanalyse und in Approximationen der Faktorenanalyse wird eine *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 (1.70). \square

In Abbildung 5 wird noch einmal die Vektorrepräsentation bestimmter kognitiver Fähigkeiten gezeigt: links ergibt sich das poetische Vermögen als Vektoraddition (Linearkombination) der zwei Fähigkeiten 'sprachliches Vermögen' und '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 werden¹¹.

Teilräume Es sei $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ eine orthonormale Basis des V_n . Eine echte Teilmenge der Basisvektoren definiert dann einen Teilraum des V_n . Ohne Einschränkung der Allgemeinheit seien $\mathbf{v}_1, \dots, \mathbf{v}_k$, $k < n$ ausgewählt worden, es sei $\mathcal{L}_k = \mathcal{L}(\mathbf{v}_1, \dots, \mathbf{v}_k)$ die lineare Hülle dieser Teilbasis. \mathcal{L}_k ist sicherlich ein Vektorraum (Satz 1.8). Dann existiert ein Vektor $\mathbf{n} \in V_n$, der orthogonal zu allen Vektoren aus \mathcal{L}_k ist.

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

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

Dann folgt

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

da $\mathbf{v}'_{k+1} \mathbf{v}_j = 0$ für $j = 1, \dots, k$. 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 21 gerechtfertigt.

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

Beweis:¹² 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

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

¹¹Derartige 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, zumal nicht klar ist, ob die Interaktion der Prozesse eine derartige additive Modellierung überhaupt zuläßt. Dieses Problem kann hier nicht diskutiert werden. Es kann aber gesagt werden, dass die Vektoralgebra ebenfalls zu den mathematischen Mitteln gehört, mit denen dynamische Prozesse modelliert werden, worauf aber in diesem Skript nicht eingegangen werden kann.

¹²Nach 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. \square

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

Definition 1.13 *Es sei V ein Vektorraum und S eine Teilmenge von V . Dann ist der Rang von S gleich der Dimension des von S erzeugten Unterraums $\mathcal{L}(S)$. 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 Unterraums $\mathcal{L}(S)$, und $n - r$ heißt die *Kodimension* des Unterraums $\mathcal{L}(S)$. 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. \square

Definition 1.14 *Es sei V ein n -dimensionaler Vektorraum und $U \subset V$ sei ein Teilraum von V . Weiter sei¹³*

$$U^\perp = \{\mathbf{x} \in V \mid \mathbf{x}'\mathbf{u} = 0, \forall \mathbf{u} \in U\} \quad (1.71)$$

Dann heißt U^\perp das orthogonale Komplement von U .

Beispiel 1.4 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 $\mathbf{b} \perp \mathbf{a}$, $\mathbf{b} \in V$. \square

Beispiel 1.5 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. \square

Definition 1.15 *Es sei V ein n -dimensionaler Vektorraum und U und W seien zwei Teilräume von V . V heißt orthogonale Summe von U und W , in Zeichen $V = U \perp W$, wenn*

- (i) Für jeden Vektor $\mathbf{v} \in V$, $\mathbf{v} = \mathbf{u} + \mathbf{w}$, $\mathbf{u} \in U$, $\mathbf{w} \in W$, und
- (ii) $\mathbf{u}'\mathbf{w} = 0$ für alle $\mathbf{u} \in U$, $\mathbf{w} \in W$.

¹³ \forall steht für "für alle"

Satz 1.13 *Es seien $U, W \subset V$, V ein n -dimensionaler Vektorraum. Die folgenden Aussagen (i) und (ii) sind äquivalent:*

(i) *W ist das orthogonale Komplement von U , d.h. $W = U^\perp$,*

(ii) *$V = U \perp W$, also V ist die orthogonale Summe von U und W . Dann gilt*

$$\dim V = \dim U + \dim U^\perp. \quad (1.72)$$

Beweis: (i) \Rightarrow (ii): Zu zeigen ist $V = U \perp U^\perp$. Es sei $\{\mathbf{b}_1, \dots, \mathbf{b}_r\}$ eine orthonormale Basis des Teilraums U ; sie kann nach Satz 1.12 zu einer orthonormalen Basis $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ von V erweitert werden. Nach Definition von U^\perp ist $U^\perp = \mathcal{L}(\mathbf{b}_{r+1}, \dots, \mathbf{b}_n)$, und da $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ eine Basis von V ist, folgt Aussage (i). Da die \mathbf{b}_j , $j = 1, \dots, n$ paarweise orthonormal sind, folgt die Aussage (ii) und damit (1.72).

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

Bisher ist nur gezeigt worden, dass jede beliebige Menge von n linear unabhängigen n -dimensionalen Vektoren als Basis zur Erzeugung aller Vektoren des \mathbb{R}^n verwendet werden kann, bzw. dass $r < n$ linear unabhängige n -dimensionale Vektoren einen Teilvektorraum erzeugen. Gegeben ist üblicherweise eine Menge von n m -dimensionalen Vektoren \mathbf{x}_j , $j = 1, \dots, n$, deren Komponenten Meßwerte sind; für diese Vektoren wird eine Basis von möglicherweise $r < n$ linear unabhängigen m -dimensionalen Vektoren gesucht. Da nur die \mathbf{x}_j gegeben sind, müssen die Basisvektoren aus den \mathbf{x}_j errechnet werden, d.h. dass sich die Basisvektoren als Linearkombinationen der \mathbf{x}_j ergeben müssen. Die Lösung des Problems, Basisvektoren zu bestimmen, läßt sich leichter beschreiben, wenn vom Begriff der Matrix Gebrauch gemacht werden kann.

2 Matrizen

2.1 Definitionen

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

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

Schreibt man diese Vektoren spaltenweise nebeneinander, so entsteht die Matrix

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

X heißt auch $(m \times n)$ -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 $\mathbf{x}_1, \dots, \mathbf{x}_n$ heißen die *Spaltenvektoren* von X . Die Schreibweise

$$X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \quad (2.2)$$

erweist sich oft als nützlich. Die Zeilen $(x_{i1}, x_{i2}, \dots, x_{in})$ heißen die *Zeilenvektoren* von X , $i = 1, \dots, m$. Eine Matrix wird *gestürzt* oder *transponiert*, indem die Zeilenvektoren als Spaltenvektoren angeschrieben werden; man schreibt X' dafür:

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

X' ist also eine $(n \times m)$ -Matrix. Die $\tilde{\mathbf{x}}_i$, $i = 1, \dots, m$ 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

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

Symmetrische Matrizen sind notwendig quadratisch. Eine Matrix heißt *Diagonalmatrix*, wenn alle Elemente gleich Null sind bis auf r Diagonalelemente x_{ii} ; eine Diagonalmatrix ist im Allgemeinen quadratisch, und $r \leq \min(m, n)$.

2.2 Operationen mit Matrizen

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 seien zwei $(m \times n)$ -Matrizen. Dann ist die Summe $X + Y$ durch

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

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

2.2.1 Multiplikation einer Matrix mit einem Vektor

Es sei

$$A = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n]$$

eine $(m \times n)$ -Matrix, d.h. die Spaltenvektoren \mathbf{a}_j seien m -dimensional, und $\mathbf{u} = (u_1, \dots, u_n)'$ sei ein n -dimensionaler Vektor. Der m -dimensionale Vektor \mathbf{y} sei eine Linearkombination der $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$, also der Spaltenvektoren von A , mit den Koeffizienten von \mathbf{u} als Koeffizienten:

$$\mathbf{y} = u_1 \mathbf{a}_1 + u_2 \mathbf{a}_2 + \dots + u_n \mathbf{a}_n. \quad (2.6)$$

Dann sei

$$A\mathbf{u} = \mathbf{y} \quad (2.7)$$

eine Schreibweise für (2.6). Wann immer ein Ausdruck der Form $A\mathbf{u}$ auftritt, soll er also als kompakte Schreibweise für eine Linearkombination der *Spaltenvektoren* von A mit den Koeffizienten u_1, \dots, u_n aufgefasst werden, die die Komponenten des Vektors \mathbf{u} sind. Diese Linearkombination ist ein m -dimensionaler Vektor \mathbf{y} . Ein Beispiel ist

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix},$$

und

$$A\mathbf{u} = u_1 \mathbf{a}_1 + u_2 \mathbf{a}_2 = u_1 \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix} + u_2 \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \mathbf{y}. \quad (2.8)$$

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

$$y_1 = u_1 a_{11} + u_2 a_{12}, \quad y_2 = u_1 a_{21} + u_2 a_{22}, \quad y_3 = u_1 a_{31} + u_2 a_{32}$$

gegeben sind, d.h. die Ausdrücke für y_i sind durch das Skalarprodukt des Vektors \mathbf{u} mit dem i -ten *Zeilenvektor* $\tilde{\mathbf{a}}_i$ (hier als Spaltenvektor angeschrieben) gegeben, d.h.

$$y_i = \tilde{\mathbf{a}}_i' \mathbf{u} = \mathbf{u}' \tilde{\mathbf{a}}_i, \quad i = 1, 2, 3 \quad (2.9)$$

Man kann nun auch von (2.9) ausgehen und das Produkt $A\mathbf{u} = \mathbf{y}$ durch (2.9) definieren, d.h. man definiert den Ausdruck $A\mathbf{u}$, indem man festlegt, dass die Skalarprodukte der Zeilenvektoren von A mit \mathbf{u} die Komponenten eines Vektors \mathbf{y} bedeuten sollen, so dass $A\mathbf{u} = \mathbf{y}$ gilt. Dann *folgt*, dass \mathbf{y} stets auch als Linearkombination der Spaltenvektoren von A mit den Komponenten von \mathbf{u} als

Koeffizienten interpretiert werden kann. Die Definitionen von $\mathbf{y} = \mathbf{A}\mathbf{u}$ (i) als Linearkombination der Spaltenvektoren von A und (ii) der Komponenten y_i von \mathbf{y} als Skalarprodukte der Zeilenvektoren von A mit \mathbf{u} sind also äquivalent. Üblicherweise wird (ii) als Definition von $\mathbf{A}\mathbf{u} = \mathbf{y}$ angegeben; der explizite Fokus auf \mathbf{y} als Linearkombination der Spaltenvektoren von A erweist sich aber bei vielen Überlegungen als außerordentlich nützlich. Im Übrigen wird je nach Fragestellung von beiden Definitionen Gebrauch gemacht werden.

So spielt im Folgenden auch das Produkt

$$\mathbf{v}'A = \mathbf{z}' \quad (2.10)$$

eine wichtige Rolle, wobei \mathbf{v} ein m -dimensionaler Vektor ist (A ist eine $(m \times n)$ -Matrix). Bei der Definition dieser Gleichung wird zunächst auf die Definition (ii) des Produkts $\mathbf{A}\mathbf{u} = \mathbf{y}$ zurückgegriffen: Die Komponenten von \mathbf{z}' sind die Skalarprodukte von \mathbf{v} mit den Spaltenvektoren von A , d.h. es soll $z_1 = \mathbf{v}'\mathbf{a}_1, \dots, z_n = \mathbf{v}'\mathbf{a}_n$ gelten, so dass \mathbf{z} ein n -dimensionaler Vektor ist. $\mathbf{v}'A$ muß also ein Zeilenvektor sein, – weshalb auch \mathbf{z}' statt \mathbf{z} geschrieben wurde. Man überlegt sich leicht, dass \mathbf{z} eine Linearkombination der Zeilenvektoren von A ist, mit den Komponenten von \mathbf{v} als Koeffizienten. Ist A wie im obigen Beispiel eine (3×2) -Matrix, so hat man

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

also

$$\mathbf{z} = \begin{pmatrix} v_1a_{11} + v_2a_{21} + v_3a_{31} \\ v_1a_{12} + v_2a_{22} + v_3a_{32} \end{pmatrix},$$

und

$$z_j = \mathbf{v}'\mathbf{a}_j, \quad j = 1, 2 \quad (2.11)$$

Zusammengefasst hat man

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

Nun sei $\mathbf{y} = \mathbf{A}\mathbf{u}$ und es wird ein Ausdruck für den transponierten Vektor \mathbf{y}' gesucht. Es gilt der

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

$$\mathbf{y}' = (\mathbf{A}\mathbf{u})' = \mathbf{u}'A'. \quad (2.12)$$

Beweis: \mathbf{y} ist eine Linearkombination der Spaltenvektoren von A , d.h. \mathbf{y}' muß

eine Linearkombination der Zeilenvektoren von A' sein. Mit Bezug¹⁴ auf (2.10) muß also $\mathbf{y}' = \mathbf{u}'A'$ gelten, wobei A in (2.10) durch A' ersetzt werden muß. \square

Man vergewissere sich der Korrektheit dieser Aussage anhand von (2.8).

Zusammenfassend kann man sagen, dass die Multiplikation einer Matrix A mit einem Vektor \mathbf{x} stets einen Vektor \mathbf{y} ergibt, der sich im Allgemeinen hinsichtlich seiner Länge, seiner Orientierung und möglicherweise auch in der Anzahl seiner Komponenten von \mathbf{x} unterscheidet. Es gibt verschiedene Redeweisen: die Matrix A transformiert den Vektor \mathbf{x} in den Vektor \mathbf{y} (*Vektortransformation*), oder A bildet \mathbf{x} auf dem Vektor \mathbf{y} ab (*Vektorabbildung*); die Matrix A definiert dann eine Abbildung oder Transformation des Vektors \mathbf{x} . Diese verschiedenen Verbalisierungen korrespondieren zu bestimmten möglichen Vorstellungen über das, was die Multiplikation einer Matrix mit einem Vektor bewirkt. So kann sich z.B. nur die Orientierung, nicht aber die Länge von \mathbf{y} von der des Vektors \mathbf{x} unterscheiden, so dass man sagen kann, $A\mathbf{x}$ bedeute eine Rotation von \mathbf{x} . Oder $A\mathbf{x}$ bewirkt nur eine Veränderung der Länge von \mathbf{x} . Rotation und reine Längenveränderung können, anschaulich gesprochen, als Transformationen von \mathbf{x} aufgefasst werden. Es sind gleichermaßen Abbildungen von Vektoren aus einem Vektorraum auf Vektoren aus demselben Vektorraum. Das gleiche gilt für Transformationen, bei denen \mathbf{y} sowohl eine andere Länge, als auch eine andere Orientierung als \mathbf{x} , aber dieselbe Anzahl von Komponenten hat. Unterscheidet sich auch die Anzahl der Komponenten von \mathbf{x} von der des Vektors \mathbf{y} , so bildet A Vektoren \mathbf{x} aus einem Vektorraum auf Vektoren aus einem anderen Vektorraum ab. Man kann die Eigenschaften von Matrizen als Eigenschaften der Abbildungen von Vektoren auf andere Vektoren diskutieren. In Abschnitt 4.5 wird auf diese Aspekte der Multiplikation von Matrizen mit Vektoren ausführlicher eingegangen.

2.2.2 Multiplikation einer Matrix mit einer Matrix

Es seien A eine $(m \times n)$ - 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

$$C = AB \tag{2.13}$$

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

¹⁴Eine Idee wäre, $\mathbf{y}' = A'\mathbf{u}$ zu versuchen, aber man sieht sofort, dass diese Gleichung nicht gelten kann: (i) A' ist eine $(n \times m)$ -Matrix und \mathbf{u} ist n -dimensional, so dass für $m \neq n$ das Produkt $A'\mathbf{u}$ gar nicht berechenbar ist, (ii) für den Spezialfall $m = n$ könnte ein Vektor berechnet werden, aber \mathbf{y}' wäre dann eine Linearkombination der Spalten von A' , entgegen der Definition von \mathbf{y}' als Linearkombination der Zeilenvektoren von A' .

die Beziehung

$$c_{ij} = \tilde{\mathbf{a}}_i' \mathbf{b}_j, \quad i = 1, \dots, m; \quad j = 1, \dots, p \quad (2.14)$$

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. 39.

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

$$C' = (AB)' = B' A' \quad (2.15)$$

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

$$(AB)C = A(BC) \quad (\text{Assoziativitat}) \quad (2.16)$$

Beweis: Zu (2.15): 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 mu $C' = B' A'$ gelten.

Zu (2.16): Es sei $U = AB$, $V = BC$, so dass $F = (AB)C = UC$, $G = A(BC) = AV$. U ist eine $(m \times p)$ -Matrix, 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

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

Nun ist $\mathbf{v}_j = B \mathbf{c}_j$, so dass $\mathbf{g}_j = A B \mathbf{c}_j = U \mathbf{c}_j$ folgt, was wiederum $\mathbf{f}_j = \mathbf{g}_j$ und damit $F = G$ bedeutet, – und das ist (2.16). \square

Die in (2.16) ausgedruckte 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 *kommutativ*, d.h. es gilt stets auch $ab = ba$.

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

$$AB \neq BA. \quad (2.17)$$

Damit $AB = BA$ gilt mussen einige *notwendige*, wenn auch noch nicht *hinreichende* Bedingungen erfullt sein. Zunachst einmal mu $p = m$ 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 $(n \times n)$ -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 (2.17) 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 $j = 1, \dots$. Diese Gleichheit kann für spezielle Matrizen gegeben sein, etwa falls $B = A^{-1}$ die Inverse von A ist (s. Abschnitt 2.4), oder wenn $B = A'$ und A eine orthonormale Matrix (Rotationsmatrix) ist, s. Abschnitt 2.5.1. Irgendzwei $(n \times n)$ 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.

Spezialfall: 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 Diagonalelementen von Null verschiedene Elemente haben:

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}, \quad \tilde{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_m \end{pmatrix} \quad (2.18)$$

Man schreibt dafür auch kurz

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n), \quad \tilde{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_m) \quad (2.19)$$

Dann ist

$$X\Lambda = \begin{pmatrix} \lambda_1 x_{11} & \lambda_2 x_{12} & \cdots & \lambda_n x_{1n} \\ \lambda_1 x_{21} & \lambda_2 x_{22} & \cdots & \lambda_n x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 x_{m1} & \lambda_2 x_{m2} & \cdots & \lambda_n x_{mn} \end{pmatrix}, \quad (2.20)$$

bzw.

$$\tilde{\Lambda}X = \begin{pmatrix} \lambda_1 x_{11} & \lambda_1 x_{12} & \cdots & \lambda_1 x_{1n} \\ \lambda_2 x_{21} & \lambda_2 x_{22} & \cdots & \lambda_2 x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_m x_{m1} & \lambda_m x_{m2} & \cdots & \lambda_m x_{mn} \end{pmatrix}. \quad (2.21)$$

$X\Lambda$ ist eine Matrix, deren Spaltenvektoren gleich dem j -ten Spaltenvektor \mathbf{x}_j , multipliziert mit dem korrespondierenden Diagonalelement λ_j von Λ multipliziert worden sind. Dies ist gleichbedeutend mit einer Längenskalierung der Spaltenvektoren von X . $\tilde{\Lambda}$ ist eine Matrix, deren Zeilenvektoren mit dem korrespondierenden Diagonalelement von $\tilde{\Lambda}$ multipliziert worden sind. Dies ist gleichbedeutend mit einer Längenskalierung der Zeilenvektoren von X .

2.2.3 Beispiel: Varianz-Kovarianz-Matrizen

Kovarianzen und Korrelationen: Es sei $X = (x_{ij})$ eine spaltenzentrierte ($m \times n$)-Matrix, d.h. x_{ij} ist die Abweichung der i -ten Messung der j -ten Variablen vom Mittelwert der Messungen der j -ten Variablen. Dann ist

$$\frac{1}{m} X'X = \left(\frac{1}{m} \sum_{i=1}^m x_{ij}x_{ik} \right) \quad (2.22)$$

die Matrix der Kovarianzen der Variablen;

$$\frac{1}{m} \sum_{i=1}^m x_{ij}x_{ik}$$

ist die Kovarianz zwischen der j -ten und der k -ten Variablen. Die Matrix $X'X$ ist symmetrisch, denn gemäß (2.15) hat man

$$(X'X)' = X'X$$

Die Diagonalelemente von $(1/n)X'X$ sind gerade die Varianzen s_j^2 der Variablen, $j = 1, \dots, n$. Man kann die Varianzen in einer Diagonalmatrix S zusammenfassen:

$$S = \begin{pmatrix} s_1^2 & 0 & \cdots & 0 \\ 0 & s_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_n^2 \end{pmatrix}, \quad S^{-1/2} = \begin{pmatrix} \frac{1}{s_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{s_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{s_n} \end{pmatrix} \quad (2.23)$$

Hier wurde gleich die Matrix $S^{-1/2}$ mit eingeführt; sie bezeichnet die Diagonalmatrix, in deren Diagonalzellen die Reziprokwerte der Standardabweichungen s_j stehen. Die Schreibweise $S^{-1/2}$ ist an die Schreibweise $x^{1/2}$ für \sqrt{x} und $x^{-1/2}$ für $1/\sqrt{x}$ angelehnt. Offenbar gilt nun (Überprüfung durch Nachrechnen!)

$$Z = XS^{-1/2} \quad (2.24)$$

Z ist die Matrix der (spalten-)standardisierten Messwerte. Nach (2.20) bedeutet ja die Multiplikation einer Matrix von rechts mit einer Diagonalmatrix eine Längenskalierung der Spaltenvektoren von S . Die Elemente z_{ij} von Z sind durch

$$z_{ij} = \frac{X_{ij} - \bar{x}_j}{s_j}, \quad i = 1, \dots, m; \quad j = 1, \dots, n \quad (2.25)$$

definiert.

Matrizen der Art $X'X$ oder XX' sind als "Kreuzproduktmatrizen" bekannt; der Grund dafür ist, dass $X'X$ eine Matrix ist, deren Elemente die Skalarprodukte der Spaltenvektoren $\mathbf{x}'_j \mathbf{x}_k$, $j, k = 1, \dots, n$ sind, und die Elemente von XX' sind die Skalarprodukte $\tilde{\mathbf{x}}'_i \tilde{\mathbf{x}}_j$, $i, j = 1, \dots, m$ der Zeilenvektoren von X . Der Ausdruck 'Kreuzprodukt' ist Jargon für den Ausdruck 'Skalarprodukt'. Es gibt allerdings einen offiziellen Ausdruck für diese Art von Matrizen:

Definition 2.1 Es seien $\mathbf{x}_1, \dots, \mathbf{x}_n$ eine Menge von Vektoren mit gleicher Dimensionalität. Es sei A die $(n \times n)$ -Matrix der Skalarprodukte $\mathbf{x}'_j \mathbf{x}_k$, $j, k = 1, \dots, n$ der Vektoren. Dann heißt A Gram-Matrix oder Gramsche Matrix¹⁵.

XX' ist eine Gram-Matrix in Bezug auf die Zeilenvektoren von X .

In (2.24) sind die Skalierungsfaktoren die Reziprokwerte der Standardabweichungen. Dann ist $Z = (z_{ij})$ die Matrix der (spalten-)standardisierten Meswerte, und

$$R = \frac{1}{m} Z'Z = \frac{1}{m} S^{-1/2} X'XS^{-1/2} \quad (2.26)$$

ist die Matrix der Korrelationen zwischen den Variablen. Natürlich ist R symmetrisch, denn $(Z'Z)' = Z'Z$.

Varianz-Kovarianzmatrizen: Die Matrix der Kovarianzen

$$s_{jk} = \frac{1}{m} \sum_{i=1}^m (X_{ij} - \bar{x}_j)(X_{ik} - \bar{x}_k)$$

läßt sich als Summe von dyadischen Produkten darstellen. Es sei $\tilde{\mathbf{x}}_i$ der i -te Zeilenvektor der Matrix X . Das dyadische Produkt von $\tilde{\mathbf{x}}_i$ mit sich selbst ist

$$\tilde{\mathbf{x}}_i \tilde{\mathbf{x}}'_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{in} \end{pmatrix} (x_{i1}, x_{i2}, \dots, x_{in}) = (x_{ij} x_{ik}) = \begin{pmatrix} x_{i1}x_{i1} & x_{i1}x_{i2} & \cdots & x_{i1}x_{in} \\ x_{i2}x_{i1} & x_{i2}x_{i2} & \cdots & x_{i2}x_{in} \\ \vdots & \vdots & \ddots & \vdots \\ x_{in}x_{i1} & x_{in}x_{i2} & \cdots & x_{in}x_{in} \end{pmatrix}.$$

Dabei ist $x_{ij}x_{ik} = (X_{ij} - \bar{x}_j)(X_{ik} - \bar{x}_k)$ das i -te Element in der Summe für die Kovarianz s_{jk} , so dass

$$S = \frac{1}{m} \sum_{i=1}^m \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}'_i = \frac{1}{m} \sum_{i=1}^m (X_{ij} - \bar{x}_j)(X_{ik} - \bar{x}_k)' \quad (2.27)$$

die Varianz-Kovarianzmatrix der Variablen ist. Für die Matrix der Korrelationen leitet man sich leicht eine analoge Darstellung her. Die Schreibweise (2.27) erweist sich oft als nützlich bei der Anwendung multivariater Verfahren.

2.3 Der Rang einer Matrix

Auf Seite 35 ist der Begriff des Ranges eines Vektorraums bzw. eines Teilraums eines Vektorraums eingeführt worden. Es sei nun $X = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ eine $(m \times n)$ -Matrix. Die lineare Hülle $\mathcal{L}_s = \mathcal{L}(\mathbf{x}_1, \dots, \mathbf{x}_n)$ ist ein Vektorraum und habe den Rang s , d.h. die als Linearkombinationen der Spaltenvektoren \mathbf{x}_j erzeugten Vektoren seien als Linearkombinationen von s linear unabhängigen, m -dimensionalen

¹⁵Nach Jörgen Pedersen Gram (1850 – 1916), dänischer Mathematiker.

Vektoren $\mathbf{u}_1, \dots, \mathbf{u}_s$ darstellbar, die zu einer Matrix $U = [\mathbf{u}_1, \dots, \mathbf{u}_s]$ zusammengefasst werden können. Die $\mathbf{u}_1, \dots, \mathbf{u}_s$ bilden eine Basis für \mathcal{L}_s , und es gibt beliebig viele Basen für \mathcal{L}_s , und $\mathbf{u}_1, \dots, \mathbf{u}_s$ kann beliebig gewählt werden. Zur Erinnerung: in Kommentar 2. zu Satz 1.10, Seite 31 wurde angemerkt, dass s nicht größer als n sein kann.

Die Zeilenvektoren von X sind die n -dimensionalen Spaltenvektoren $\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_m$ von X' . Die lineare Hülle $\mathcal{L}_z = \mathcal{L}(\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_m)$ habe den Rang $r < m$, d.h. die $\tilde{\mathbf{x}}_i$ seien als Linearkombinationen von r linear unabhängigen Vektoren $\mathbf{v}_1, \dots, \mathbf{v}_r$ darstellbar, die zu einer Matrix $V = [\mathbf{v}_1, \dots, \mathbf{v}_r]$ zusammengefasst werden können. $\mathbf{v}_1, \dots, \mathbf{v}_r$ ist ebenfalls eine beliebig gewählte Basis für \mathcal{L}_z . s heißt *Spaltenrang* von X , und r ist der *Zeilenrang* von X . Man beachte, dass die Spaltenvektoren von X m -dimensional, die Zeilenvektoren von X aber n -dimensional sind, d.h. die Zeilen- und Spaltenvektoren sind Elemente aus Vektorräumen mit verschiedener Dimensionalität.

Die Darstellungen der Spaltenvektoren \mathbf{x}_j von X und der Spaltenvektoren $\tilde{\mathbf{x}}_i$ von X' sind nicht unabhängig von einander. So sei $X = UA'$, $U = [\mathbf{u}_1, \dots, \mathbf{u}_s]$, und A' die Matrix ($s \times n$)-der Koeffizienten: der Spaltenvektor \mathbf{a}_j von A ist der Koeffizientenvektor für \mathbf{x}_j , d.h. $\mathbf{x}_j = U\mathbf{a}_j$. Dann folgt aber $X' = AU'$, d.h. die Spaltenvektoren $\tilde{\mathbf{x}}_i$ sind Linearkombinationen der Spaltenvektoren von A , und U' ist nun die Koeffizientenmatrix. A ist eine ($n \times s$)-Matrix. Dem obigen Ansatz zufolge soll aber $X' = VB$ gelten, wobei B die zugehörige Matrix der Koeffizienten für die $\mathbf{v}_1, \dots, \mathbf{v}_r$ ist. Man könnte nun einfach folgern, dass die oben angenommene Matrix V gleich der Matrix A ist und $B = U'$ gilt, und dass somit $r = s$ gilt, aber diese Gleichsetzung setzt voraus, dass die Spaltenvektoren von A linear unabhängig sind. Diese Eigenschaft muß noch nachgewiesen werden. Dieser Nachweis wird im Beweis des folgenden Satzes geliefert.

Satz 2.3 *Es sei X eine $(m \times n)$ -Matrix mit dem Zeilenrang r und dem Spaltenrang s . Dann gilt*

$$r = s \quad (2.28)$$

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

$$X = UV', \quad (2.29)$$

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

$$r \leq \min(m, n) \quad (2.30)$$

Beweis: X besteht aus den n m -dimensionalen Spaltenvektoren $\mathbf{x}_1, \dots, \mathbf{x}_n$, bzw. aus den m n -dimensionalen Zeilenvektoren $\tilde{\mathbf{x}}'_1, \dots, \tilde{\mathbf{x}}'_m$:

$$X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] = \begin{bmatrix} \tilde{\mathbf{x}}'_1 \\ \tilde{\mathbf{x}}'_2 \\ \vdots \\ \tilde{\mathbf{x}}'_m \end{bmatrix}.$$

Die Spaltenvektoren \mathbf{x}_j können als Linearkombinationen von s beliebig gewählten, aber linear unabhängigen m -dimensionalen Vektoren $\mathbf{u}_1, \dots, \mathbf{u}_s$ angeschrieben werden. Fasst man die \mathbf{u}_k , $k = 1, \dots, s$ zu einer Matrix $U = [\mathbf{u}_1, \dots, \mathbf{u}_s]$ zusammen, so erhält man

$$X = [\mathbf{x}_1, \dots, \mathbf{x}_n] = UA', \quad (2.31)$$

wobei A' die zu U gehörende $(s \times n)$ -Matrix der Koeffizienten ist: Für den j -ten Spaltenvektor \mathbf{x}_j hat man

$$\mathbf{x}_j = a_{1j}\mathbf{u}_1 + \dots + a_{sj}\mathbf{u}_s = U\tilde{\mathbf{a}}_j,$$

und $\tilde{\mathbf{a}}_j$ ist der j -te Spaltenvektor von A' . Dies ist der j -te Zeilenvektor von A . Da die $\mathbf{u}_1, \dots, \mathbf{u}_s$ linear unabhängig sind, liegen die a_{1j}, \dots, a_{sj} eindeutig fest (Satz 1.3, S. 25), und damit liegt für eine gewählte Matrix U die Matrix A eindeutig fest.

Die Spaltenvektoren $\tilde{\mathbf{x}}_i$ von X' (also die Zeilenvektoren von X) können als Linearkombinationen von r linear unabhängigen Spaltenvektoren $\mathbf{v}_1, \dots, \mathbf{v}_r$ dargestellt werden, wobei \mathbf{v}_k , $1 \leq k \leq r$ zunächst beliebig gewählt werden können unter der Einschränkung, dass sie linear unabhängig sind. Fasst man sie zu einer Matrix V zusammen, so erhält man

$$X' = [\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_m] = VB' \quad (2.32)$$

B' ist die zu V korrespondierende Koeffizientenmatrix, analog zu A' in (2.31), und

$$\tilde{\mathbf{x}}_i = b_{1i}\mathbf{v}_1 + \dots + b_{ir}\mathbf{v}_r = V\mathbf{b}_i.$$

Da U in (2.31) s Spalten hat, muß A' s Zeilen haben. Die Zeilenvektoren von X können aber auch als Linearkombinationen der Zeilenvektoren von A' aufgefasst werden (s. Kommentar 2. zur Matrixmultiplikation, Seite 41). Dann folgt, dass der Zeilenrang r von X nicht größer als s sein kann (es gibt eben nur s Zeilen in A'), d.h. es muß $r \leq s$ gelten.

Umgekehrt folgt aus (2.32), dass die Zeilenvektoren von X' (also die Spaltenvektoren von X) auch als Linearkombinationen der Zeilenvektoren von B' dargestellt werden können. Da V nach Voraussetzung r Spalten hat, muß B' r Zeilen haben. Daraus folgt, dass der Zeilenrang s (d.h. der Spaltenrang von X) von X' nicht größer als r sein kann (es gibt eben nur r Zeilen in B'), d.h. es muß $s \leq r$ gelten. Beide Ergebnisse zusammen liefern

$$r \leq s \wedge s \leq r \Rightarrow r = s,$$

wie behauptet¹⁶.

¹⁶Das Zeichen \wedge steht für 'und'.

Nach (2.31) gilt $X = UA'$, und nach (2.32) gilt $X' = VB'$, also $X = BV'$. U und B haben notwendig denselben Rang r ; B repräsentiert möglicherweise eine von U verschiedene Basis von \mathcal{L}_s . Da U und B beliebig gewählt werden können, kann man insbesondere $B = U$ wählen. Man hat dann $X = UA' = UV'$, und da die Linearkombinationen für die \mathbf{x}_j eindeutig sind (Satz 1.3, Seite 25), wenn die Spaltenvektoren von U linear unabhängig sind, hat man mit der Wahl von U auch die von A festgelegt, denn es muß nun $A' = V'$ gelten, womit nachgewiesen ist, dass X als Produkt UV' zweier Matrizen mit gleichem Rang dargestellt werden kann.

U und V haben wegen $r = s$ jeweils r linear unabhängige Spaltenvektoren. Es sei $m > n$. Da der Rang von X gleich der Maximalzahl der linear unabhängigen Vektoren von $\mathcal{L}(\mathbf{x}_1, \dots, \mathbf{x}_n)$ und von $\mathcal{L}(\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_m)$ ist, folgt $r \leq n$. Analog folgt für $m < n$, dass $r \leq m$ sein muß. Zusammengefaßt ergibt sich die Aussage

$$r \leq \min(m, n).$$

also (2.30). □

Definition 2.2 *Es seien \mathbf{x}_j , $j = 1, \dots, n$ die Spaltenvektoren von X , und $\tilde{\mathbf{x}}_i$, $i = 1, \dots, m$, seien die Zeilenvektoren von X (als Spaltenvektoren von X' aufgefasst). Dann heißt die lineare Hülle $\mathcal{L}(\mathbf{x}_1, \dots, \mathbf{x}_n)$ der Spaltenraum von X , und die lineare Hülle $\mathcal{L}(\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_m)$ 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 2.3 ergibt sich, dass sich *jede* $(m \times n)$ -Matrix X als Produkt der Form (2.29) darstellen läßt.
4. Die in Abschnitt 2.5.4 eingeführte Singularwertzerlegung von X ist ein Spezialfall von (2.29), der vielen faktorenanalytischen Ansätzen zur Interpretation von Datenmatrizen zugrunde liegt. □

Korollar 2.1 *Aus (2.29) folgt die Aussage*

$$\text{rg}(X'X) = \text{rg}(XX') = \text{rg}(X), \tag{2.33}$$

Beweis: Denn $X'X = V'U'UV = V'C$, $C = U'UV$, d.h. die Spalten von $X'X$ sind Linearkombinationen der r linear unabhängigen n -dimensionalen Spaltenvektoren von V' , und $XX' = UVV'U = UD$, $D = VV'U$, und die Spalten von XX' sind Linearkombinationen der r linear unabhängigen m -dimensionalen Spaltenvektoren von U . □

Zur Übung und zum späteren Gebrauch soll gleich noch der folgende Satz bewiesen werden:

Satz 2.4 *Es sei \mathbf{x} ein n -dimensionaler und \mathbf{y} ein m -dimensionaler Vektor. Das dyadische Produkt \mathbf{xy}' ist eine $(m \times n)$ -Matrix mit dem Rang 1.*

Beweis: Die Inspektion der Matrix \mathbf{xy}' (vergl. (1.19), (S. 14) 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. \square

Korollar 2.2 *Aus (2.29) folgt, dass jedes Element x_{ij} von X als Skalarprodukt*

$$x_{ij} = \tilde{\mathbf{u}}_i' \tilde{\mathbf{v}}_j \quad (2.34)$$

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

Beweis: (2.34) ist eine unmittelbare Folgerung aus der Gleichung $X = UV'$ (vergl. auch Gleichung (2.14), Seite 41). \square

Die Bedeutung von (2.34) besteht darin, dass der Messwert x_{ij} stets als Skalarprodukt von zwei r -dimensionalen "latenten" Vektoren dargestellt werden kann. Diese Darstellung ist allerdings nicht eindeutig, da es ja beliebig viele Möglichkeiten gibt, die Teilbasis U bzw. V zu wählen. Gabriel (1971) weist darauf hin, dass man von den $\tilde{\mathbf{u}}_i$ und den $\tilde{\mathbf{v}}_j$ als von den *Zeileneffekten* bzw. den *Spalteneffekten* sprechen könne, wenn X eine Datenmatrix sei. Dazu denke man an eine zweidimensionale Varianzanalyse (ANOVA¹⁷), wobei x_{ij} eine Messung für die i -te Stufe des ersten Faktors A (der ersten unabhängigen Bedingung) und der j -ten Stufe des zweiten Faktors B (der zweiten unabhängigen Bedingung) sei. Dann gilt die für eine ANOVA übliche *Strukturgleichung*

$$x_{ij} = \mu + \mu_i(A) + \mu_j(B) + \mu_{ij}(A \times B) + e_{ij}, \quad (2.35)$$

Hierin ist μ der allgemeine Erwartungswert der Messwerte, $\mu_i(A)$ ist der Haupteffekt der i -ten Stufe von A, $\mu_j(B)$ ist der Haupteffekt der j -ten Stufe von B und $\mu_{ij}(A \times B)$ ist der Wechselwirkungseffekt. Dabei steht A für die Fälle, B für die Variablen. e_{ij} ist ein zufälliger Effekt ("Fehler"), in dem alle nicht kontrollierten Effekte zusammengefasst werden. Die Haupteffekte gehen additiv in x_{ij} ein. Gleichung (2.34) sagt, dass für den Messwert x_{ij} keine additiven Haupteffekte angenommen werden und dass statt dessen nur eine multiplikative Wechselwirkung

$$\mu_{ij}(A \times B) = \tilde{\mathbf{u}}_i' \tilde{\mathbf{v}}_j \quad (2.36)$$

postuliert wird.

Die Aussage (2.33) sind für die multivariate Statistik von Bedeutung: bei geeigneter Zentrierung entsprechen den Matrizen $X'X$ bzw. XX' Kovarianz- bzw.

¹⁷Analysis Of Variance

Korrelationsmatrizen, die demnach denselben Rang wie die Datenmatrix haben. Gleichung (2.34) ist die Basis für graphische Repräsentationen der durch U und V definierten "latenten Struktur" von X .

Satz 2.5 *Es sei A eine $(m \times n)$ -Matrix, B sei eine $(n \times p)$ -Matrix. Dann ist das Produkt $C = AB$ eine $(m \times p)$ -Matrix. Der Rang von C ist kleiner oder gleich dem kleineren der Ränge von A und B , d.h.*

$$\text{rg}(C) \leq \min[\text{rg}(A), \text{rg}(B)]. \quad (2.37)$$

Beweis: A habe den Rang s . Die Spaltenvektoren von C sind Linearkombinationen der Spaltenvektoren von A , d.h. $C \subseteq \mathcal{L}(A) = \mathcal{L}(\mathbf{a}_1, \dots, \mathbf{a}_s)$, $\mathbf{a}_1, \dots, \mathbf{a}_s$ 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 $C' \subseteq \mathcal{L}(B') = \mathcal{L}(\mathbf{b}_1, \dots, \mathbf{b}_r)$, $\mathbf{b}_1, \dots, \mathbf{b}_r$ 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 2.28 sind dann auch die Spaltenvektoren von C als Linearkombinationen von maximal r linear unabhängigen Vektoren darstellbar. Sei umgekehrt $s \leq r$; analog zur vorangehenden Argumentation ist dann ist der Rang von C höchstens gleich s , d.h. $\text{rg}(C) \leq \min(\text{rg}(A), \text{rg}(B))$. \square

Satz 2.6 *X sei eine (m, n) -Matrix mit dem Rang $\text{rg}(X) = r$, P sei eine (m, m) -Matrix mit dem Rang $\text{rg}(P) = m$, und Q sei eine (n, n) -Matrix mit dem Rang $\text{rg}(Q) = n$. Dann gilt*

$$\text{rg}(X) = \text{rg}(PXQ). \quad (2.38)$$

Beweis: Nach Satz 2.5, Gleichung (2.37) (Seite 49), gilt $\text{rg}(PX) \leq \text{rg}(X)$. Es sei $B = PX$, so dass $\text{rg}(B) \leq \text{rg}(X)$. Nach Voraussetzung existiert P^{-1} (denn P hat vollen Rang). Dann folgt ebenso $\text{rg}(P^{-1}B) \leq \text{rg}(B) \leq \text{rg}(X)$, d.h. aber wegen $P^{-1}B = X$

$$\text{rg}(X) \leq \text{rg}(B) \leq \text{rg}(X),$$

woraus $\text{rg}(B) = \text{rg}(PX) = \text{rg}(X)$ folgt. Es sei $C = XQ$; auf analoge Weise folgt $\text{rg}(C) = \text{rg}(XQ) = \text{rg}(X)$. Dann folgt aber auch $\text{rg}(X) = \text{rg}(PXQ)$. \square

2.4 Die Inverse einer Matrix

Definition 2.3 *Es sei A eine $(n \times n)$ -Matrix. A^{-1} sei eine Matrix derart, dass*

$$AA^{-1} = A^{-1}A = I_n, \quad (2.39)$$

I_n die $(n \times n)$ -Einheitsmatrix. Dann heißt A^{-1} die zu A inverse Matrix.

Ist A eine $(m \times n)$ -Matrix und ist $m \neq n$, so existiert keine Inverse für A ; eine notwendige, aber nicht hinreichende Bedingung für die Existenz einer Inversen ist, dass A quadratisch ist.

Die Gleichung (2.39) ist ein Beispiel für den Fall einer kommutativen Matrixmultiplikation: mit $B = A^{-1}$ gilt $AB = BA$.

Satz 2.7 Die Inverse A^{-1} zur $(n \times n)$ -Matrix A existiert genau dann, wenn A den Rang n hat.

Beweis: Es sei A eine $(n \times n)$ -Matrix und es existiere A^{-1} . Dann gilt $AA^{-1} = I_n$. Da $\text{rg}(I_n) = n$ folgt $\text{rg}(AA^{-1}) = n$, und da nach (2.37), Seite 49, $\text{rg}(AA^{-1}) \leq \min(\text{rg}(A), \text{rg}(A^{-1}))$ gilt, folgt $\text{rg}(A) = \text{rg}(A^{-1}) = n$, d.h. die Zeilen- bzw. Spaltenvektoren von A und A^{-1} sind linear unabhängig.

Nun sei $\text{rg}(A) = n$, d.h. die Spaltenvektoren \mathbf{a}_j , $j = 1, \dots, n$ sind linear unabhängig. Dann enthält die lineare Hülle $\mathcal{L}(A) = \mathcal{L}(\mathbf{a}_1, \dots, \mathbf{a}_n)$ alle n -dimensionalen Vektoren, mithin auch die Einheitsvektoren \mathbf{e}_j und es existiert ein n -dimensionaler Vektor \mathbf{b}_j derart, dass $A\mathbf{b}_j = \mathbf{e}_j$; wegen der linearen Unabhängigkeit der \mathbf{a}_j ist \mathbf{b}_j eindeutig bestimmt (s. Satz 1.3, Seite 25). Das bedeutet aber, dass eine Matrix $B = [\mathbf{b}_1, \dots, \mathbf{b}_n]$ existiert derart, dass $AB = I_n$, und da B wegen der linearen Unabhängigkeit der Spaltenvektoren von A eindeutig bestimmt ist, folgt $B = A^{-1}$. \square

Beispiel 2.1 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

$$a\alpha + b\gamma = 1 \Rightarrow \alpha = \frac{1 - b\gamma}{a} \quad (2.40)$$

$$a\beta + b\delta = 0 \Rightarrow \beta = -\frac{b\delta}{a} \quad (2.41)$$

$$c\alpha + d\gamma = 0 \Rightarrow \gamma = -\frac{c\alpha}{d} \quad (2.42)$$

$$c\beta + d\delta = 1 \Rightarrow \delta = \frac{1 - c\beta}{d} \quad (2.43)$$

Durch Einsetzen etwa des Ausdrucks für γ (2.42) in (2.40) etc findet man

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \quad (2.44)$$

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 (2.44) 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. S. 23) 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. \square

Beispiel 2.2 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 sicher nicht orthogonal: $1 \cdot 3 + 2 \cdot 1 = 5 \neq 0$, aber sie 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:

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

Man findet nun leicht

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

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

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

Anmerkung: Es sei A eine $(n \times r)$ -Matrix mit $r < n$, und die Spaltenvektoren von A seien orthonormal. Dann ist $A'A = I_n$. Es ist aber $A'A \neq AA'$, so dass in diesem Fall $A' \neq A^{-1}$ ist, – die inverse Matrix existiert in diesem Fall nicht, weil A nicht quadratisch ist. \square

Satz 2.8 *Es seien A und B zwei $(n \times n)$ -Matrizen mit vollem Rang, so dass die Inversen Matrizen A^{-1} und B^{-1} existieren. Dann gilt*

$$(AB)^{-1} = B^{-1}A^{-1}. \quad (2.45)$$

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

2.5 Rotationen, Quadratische Formen und Eigenvektoren

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 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, wobei die Transformation als Multiplikation $T\mathbf{x} = \mathbf{y}$ der *Transformationsmatrix* T mit \mathbf{x} dargestellt werden kann. Die Aufgabe ist also, eine derartige Matrix T zu bestimmen. Es werden zunächst einige allgemeine Betrachtungen angestellt.

Es sei A eine beliebige $(m \times n)$ -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 . Es ergeben sich zwei für die Anwendungen in der Multivariaten Statistik wichtige Spezialfälle:

1. Es gelte insbesondere $\mathbf{x} \neq \mathbf{y}$, aber $\|\mathbf{x}\| = \|\mathbf{y}\|$, 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.

2.5.1 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 $(n \times n)$ -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

$$\mathbf{x} \neq \mathbf{y}, \quad \|\mathbf{y}\|^2 = \mathbf{y}'\mathbf{y} = \mathbf{x}'T'T\mathbf{x} = \mathbf{x}'\mathbf{x} = \|\mathbf{x}\|^2 \quad (2.46)$$

gelten.

Satz 2.9 Die Beziehung (2.46) gilt genau dann, wenn die Spaltenvektoren von T orthonormal sind, so dass $T'T = I$ gilt, wobei I die $(n \times n)$ -Einheitsmatrix ist. Die Beziehung $T'T = I$ impliziert die Beziehung $TT' = I$, d.h. die Orthonormalität der Spaltenvektoren von T impliziert die Orthonormalität der Zeilenvektoren von T .

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 $\mathbf{x}'T'T\mathbf{x} = \mathbf{x}'I\mathbf{x} = \mathbf{x}'\mathbf{x} = \|\mathbf{x}\|^2$.

Die Beziehung $T'T = I$ ist auch notwendig für die Gültigkeit von (2.46). Um diese Behauptung einzusehen, werde $U = T'T$ gesetzt. Nach (2.46) soll $\mathbf{x}'U\mathbf{x} = \mathbf{x}'\mathbf{x}$ für alle \mathbf{x} gelten. Dann liefert (4.11) aus dem Anhang, Abschnitt 4.3.2, Seite 125

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

und nochmalige Ableitung liefert $U = I$ (s. Gleichung (4.8), Seite 124), d.h. $T'T = I$. Dies bedeutet, dass die Spaltenvektoren von T orthonormal sind.

Es muß noch $TT' = I$ gezeigt werden. Da der Zeilenrang einer Matrix gleich ihrem Spaltenrang ist, muß $\text{rg}(T') = \text{rg}(T) = n$ gelten, d.h. die Spaltenvektoren $\tilde{\mathbf{t}}_i$ von T' (das sind die Zeilenvektoren von T) sind linear unabhängig, so dass nur noch ihre Orthonormalität gezeigt werden muß. Multiplikation von $T'T = I$ von links mit T liefert $T(T'T) = TI$, und wegen der Assoziativität der Matrixmultiplikation folgt

$$T(T'T) = (TT')T = T \Rightarrow T'(TT') = T' \Rightarrow T'(TT' - I) = 0,$$

0 die Nullmatrix. Es sei $B = TT' - I$ und \mathbf{b}_j sei der j -te Spaltenvektor von B . Dann ist $T'\mathbf{b}_j = \vec{0}_j$, $\vec{0}_j$ der j -te Spaltenvektor von 0, und man hat

$$b_{1j}\tilde{\mathbf{t}}_1 + b_{2j}\tilde{\mathbf{t}}_2 + \cdots + b_{nj}\tilde{\mathbf{t}}_n = \vec{0}_j, \quad j = 1, \dots, n$$

die Koeffizienten b_{kj} sind die Komponenten von \mathbf{b}_j . Aus der linearen Unabhängigkeit der Spaltenvektoren $\tilde{\mathbf{t}}_k$ von T' folgt, dass

$$b_{1j} = b_{2j} = \cdots = b_{nj} = 0, \quad j = 1, \dots, n$$

d.h. aber $\mathbf{b}_j = \vec{0}_j$. Das heißt wiederum, dass $TT' - I = 0$, also $TT' = I$, womit gezeigt ist, dass die Zeilenvektoren von T ebenfalls orthonormal sind. \square

Hinweis: Am Ende des Abschnitts 4.6, Seite 136, findet sich ein sehr kurzer Beweis von $T'T = I \Rightarrow TT' = I$, der allerdings den Begriff der Determinante einer Matrix voraussetzt, – der im Folgenden aber nicht benötigt wird. Der hier gegebene Nachweis dieser Aussage macht nur vom elementaren Begriff der linearen Unabhängigkeit Gebrauch. \square

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

$$T^{-1} = T' \tag{2.47}$$

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

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 $\mathbf{u}'\mathbf{v} = \mathbf{x}'T'T\mathbf{y} = \mathbf{x}'\mathbf{y}$, $T'T = I$. \square

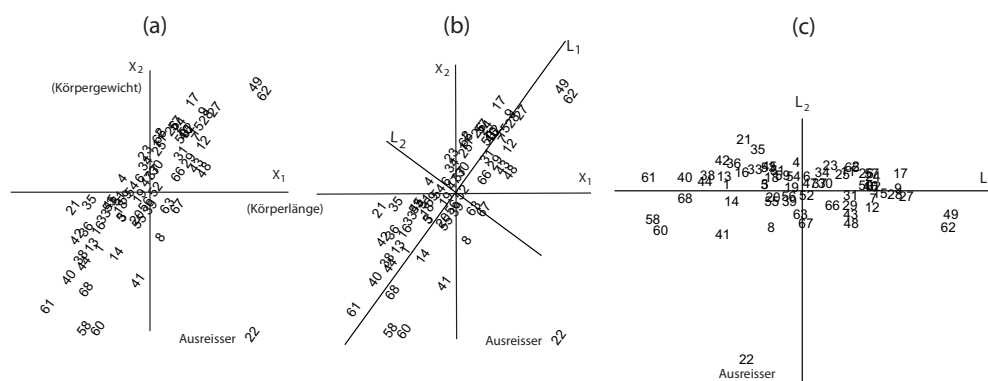
T läßt sich durch trigonometrische Betrachtungen zur Rotation (etwa von Koordinatensystemen) herleiten; im 2-dimensionalen Fall erhält man für T den Ausdruck

$$T = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}, \tag{2.48}$$

wobei θ der Rotationswinkel ist. Eine gegebene $(n \times n)$ -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 52 entsprechend hat man $T' = T^{-1}$. \square

Rotationen und Koordinatentransformationen Es sei D eine Menge von Vektoren \mathbf{x}_i , $i = 1, \dots, m$. Die \mathbf{x}_i können zum Beispiel Personen repräsentieren,

Abbildung 6: Konfiguration von Personen 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 2.4, Seite 108).



wobei die Komponenten x_{i1}, \dots, x_{im} von \mathbf{x}_i Messungen von n Variablen sind. Die Komponenten x_{ij} können als Koordinaten von durch die Endpunkte der Vektoren repräsentierten Punkten interpretiert werden, die wiederum die Personen in einem Koordinatensystem repräsentieren, das durch die Variablen definiert wird. Nun sei $\mathbf{y}_i = T\mathbf{x}_i$ und T sei eine Rotationsmatrix. Die Komponenten y_{ij} sind dann die Koordinaten derselben Punkte in einem anderen Koordinatensystem. Abbildung 6 zeigt eine Punktekonfiguration im ursprünglichen und im rotierten Koordinatensystem. Die Koordinaten der Punkte im rotierten System sind die Projektionen der Punkte auf diese Achsen im ursprünglichen System. Diese Achsen sind die 'Hauptachsen' der Konfiguration, d.h. die erste Hauptachse ist durch die Orientierung der maximalen Ausdehnung der Konfiguration definiert, und die zweite 'Hauptachse' ist im Allgemeinen die Orientierung der zweitgrößten Ausdehnung der Konfiguration gegeben; mehr darüber im folgenden Text.

2.5.2 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 zu-

nächst gezeigt, dass bestimmten symmetrischen Matrizen M – insbesondere also Varianz-Kovarianz-Matrizen – Ellipsoide zugeordnet werden können. Dabei sind die Orientierungen der Ellipsoide identisch und durch die "Eigenvektoren" von M gegeben. Jedem Eigenvektor von M entspricht ein "Eigenwert". Den Hauptachsen der Ellipsoide entsprechen mögliche latente Dimensionen, und die Eigenwerte enthalten Informationen über die Anzahl der latenten Dimensionen.

Latente Variablen werden durch (Basis-)Vektoren $\mathbf{L}_1, \dots, \mathbf{L}_r$ repräsentiert, die die Darstellung der Datenvektoren \mathbf{x}_j – die Spaltenvektoren von X , wenn die Spalten von X die gemessenen Variablen repräsentieren – als Linearkombinationen $\mathbf{x}_j = a_{1j}\mathbf{L}_1 + a_{2j}\mathbf{L}_2 + \dots + a_{jr}\mathbf{L}_r$ ermöglichen. Da aber nur die Matrix X gegeben ist, müssen die \mathbf{L}_k ($k = 1, \dots, r$) aus den Daten errechnet werden. Formal bedeutet dies, dass man Vektoren \mathbf{t}_k bestimmt derart, dass $X\mathbf{t}_k = \mathbf{L}_k$. Ist X (spalten-)zentriert, so gilt $\vec{1}'X = (0, \dots, 0) = \vec{0}'$, d.h. die Spaltensummen von X sind gleich Null; alle Betrachtungen übertragen sich auf zeilenzentrierte Datenmatrizen). Dann folgt $\vec{1}'X\mathbf{t}_k = \vec{1}'\mathbf{L}_k = 0$, 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 . 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.

Definition 2.4 *Es sei M eine symmetrische $(n \times n)$ -Matrix, $\mathbf{x} \in \mathbb{R}^n$, und es gelte*

$$Q_M(\mathbf{x}) = \mathbf{x}'M\mathbf{x}. \quad (2.49)$$

Dann heißt $Q_M(\mathbf{x})$ quadratische Form, und

1. M heißt positiv semidefinit, wenn $\mathbf{x}'M\mathbf{x} \geq 0$
 2. M heißt negativ semidefinit, wenn $\mathbf{x}'M\mathbf{x} \leq 0$
 3. M heißt positiv definit bzw. elliptisch, wenn $\mathbf{x}'M\mathbf{x} > 0$, und
 4. M heißt negativ definit bzw. hyperbolisch, wenn $\mathbf{x}'M\mathbf{x} < 0$
- jeweils für alle $\mathbf{x} \in \mathbb{R}^n$ gilt.

Satz 2.10 *Es sei $X \in \mathbb{R}^{m,n}$ eine zentrierte Matrix. Die Varianz-Kovarianzmatrix*

$$S = \frac{1}{m}X'X$$

ist 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 S positiv-semidefinit. \square

Satz 2.11 Es sei $M \in \mathbb{R}^{n,n}$ eine symmetrische, positiv semidefinite Matrix. Dann definiert die Menge $\mathcal{E}_x = \{\mathbf{x} | \mathbf{x}' M \mathbf{x} = k, \mathbf{x} \in \mathbb{R}^n, k \in \mathbb{R} \text{ eine Konstante}\}$ ein n -dimensionales Ellipsoid, wobei die Anfangspunkte der Vektoren \mathbf{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 (2.49) aus, so erhält man

$$\mathbf{x}' M \mathbf{x} = \sum_{i=1}^n m_{ii} x_i^2 + 2 \sum_{i < j} m_{ij} x_i x_j = k \quad (2.50)$$

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

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

$$\mathbf{x}' M \mathbf{x} = m_{11} x_1^2 + m_{22} x_2^2 + 2m_{12} x_1 x_2 = k > 0. \quad (2.51)$$

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

Anmerkung: Die in Satz 2.10 eingeführte Varianz-Kovarianzmatrix S ist symmetrisch positiv-definit, also definiert sie eine Menge von Ellipsoiden. \square

Spezialfall: Insbesondere sei $M = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ eine Diagonalmatrix. Dann sind die Ellipsoide $\mathbf{x}' \Lambda \mathbf{x} = k$ *achsenparallel*, d.h. die Hauptachsen der Ellipsoide sind parallel zu den Achsen des Koordinatensystems; diese Aussage folgt sofort aus (2.50) bzw. (2.51), 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 $(n \times n)$ -Matrix, und es gelte

$$\mathbf{x}' M \mathbf{x} = \mathbf{y}' \Lambda \mathbf{y} = k > 0, \quad \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n). \quad (2.52)$$

$\mathcal{E}_x = \{\mathbf{x} | \mathbf{x}' M \mathbf{x} = k > 0\}$ 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 $(n \times n)$ -Matrix T eine Rotation repräsentiere. Offenbar gilt

$$\mathbf{x}' M \mathbf{x} = \mathbf{y}' T' M T \mathbf{y} = \mathbf{y}' \Lambda \mathbf{y} = k, \quad (2.53)$$

woraus

$$T' M T = \Lambda \quad (2.54)$$

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

Da T als Rotationsmatrix angenommen wurde, folgt, dass T orthonormal ist. Deshalb folgt durch Multiplikation der Gleichung (2.54) von links mit T die Gleichung

$$MT = T\Lambda. \quad (2.55)$$

Diese Gleichung besagt, dass für die Spaltenvektoren \mathbf{t}_k , $k = 1, \dots, n$, 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. (2.20), S. 42). 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 2.5 *Es sei M eine beliebige $(n \times n)$ -Matrix und $\mathbf{t} \in \mathbb{R}^n$ sei ein Vektor, der der Beziehung*

$$M\mathbf{t} = \lambda\mathbf{t}, \quad \mathbf{t} \neq \vec{0}, \quad (2.56)$$

genügt. Dann heißt \mathbf{t} Eigenvektor¹⁸ von M und λ heißt der zu \mathbf{t} gehörende Eigenwert von M .

Anmerkungen:

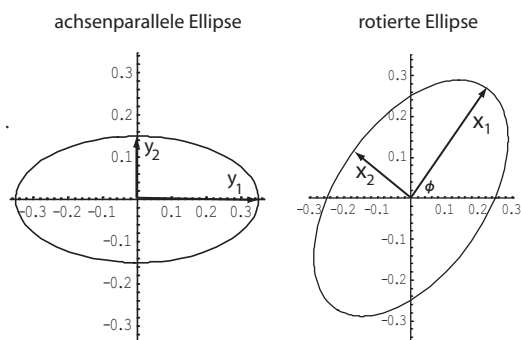
1. In der Definition wurde *nicht* vorausgesetzt, dass M symmetrisch ist, d.h. Eigenvektoren können auch für nicht-symmetrische Matrizen existieren. Die folgenden Betrachtungen beschränken sich aber auf symmetrische Matrizen M .
2. Der Nullvektor $\vec{0}$ ist *kein* Eigenvektor. Wäre $\vec{0}$ ein Eigenvektor, 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 kein sinnvoller Begriff.
3. Da T eine Rotationsmatrix ist, ist T orthonormal. Also sind die Eigenvektoren der symmetrischen Matrix M orthonormal.
4. Nach (2.56) kann man für irgendeinen Eigenvektor \mathbf{t} mit zugehörigem Eigenwert λ $M\mathbf{t} = \lambda\mathbf{t}$ schreiben, woraus

$$M\mathbf{t} - \lambda\mathbf{t} = (M - \lambda I)\mathbf{t} = 0 \quad (2.57)$$

folgt, wobei I die $(n \times n)$ -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,

¹⁸Synonym sind auch die Ausdrücke 'latenter Vektor' und 'latenter Wert' und 'charakteristischer Vektor' und 'charakteristischer Wert'.

Abbildung 7: Ellipsen in verschiedenen Orientierungen; die Hauptachsen sind die skalierten Eigenvektoren der zugehörigen Matrix M .



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 4.6) $|M - \lambda I| = 0$ gilt. Entwickelt man diese Determinante (vergl. (4.52), Seite 136), so ergibt sich ein Polynom n -ten Grades in λ , d.h.

$$f(\lambda) = |M - \lambda I| = \alpha_0 + \alpha_1 \lambda + \alpha_2 \lambda^2 + \dots + \alpha_n \lambda^n, \quad (2.58)$$

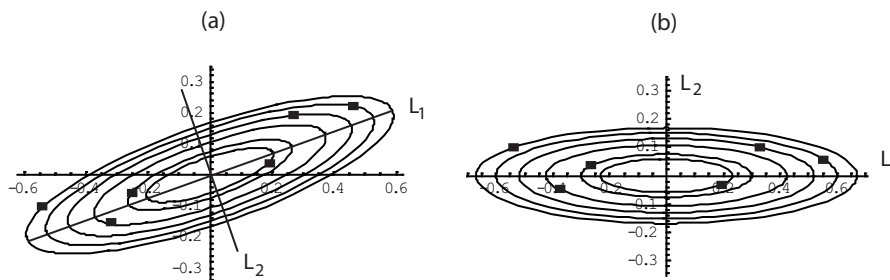
wobei die α_j reelle Zahlen sind, die durch die Elemente der Matrix $M - \lambda I$ definiert sind. Die Nullstellen dieses Polynoms, also diejenigen Werte für λ , für die $f(\lambda) = 0$ gilt, sind die Eigenwerte von M . $f(\lambda)$ heißt *charakteristisches Polynom* der Matrix $A - \lambda I$, I die Einheitsmatrix, das im Anhang, Abschnitt 4.6 vorgestellt wird, worauf aber im Folgenden kein Bezug genommen wird. Der Begriff des charakteristischen Polynoms wird hier nur der Vollständigkeit wegen erwähnt. \square

Die Gleichung (2.80) 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 positiv semidefiniter Matrizen: Der Vektor \mathbf{y}_1 definiere 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 kann $\mathbf{y}_1 = y_1 \mathbf{e}_1$, $\mathbf{e}_1 = (1, 0, \dots, 0)'$ geschrieben werden, so dass $\|\mathbf{y}_1\| = y_1 \|\mathbf{e}_1\| = y_1 \in \mathbb{R}$. Dann folgt

$$\mathbf{x}_1 = T \mathbf{y}_1 = T(y_1 \mathbf{e}_1) = y_1 T \mathbf{e}_1 = y_1 \begin{pmatrix} t_{11} \\ t_{21} \\ \vdots \\ t_{n1} \end{pmatrix} = y_1 \mathbf{t}_1, \quad y_1 \in \mathbb{R}, \quad (2.59)$$

Abbildung 8: Punktekonfiguration und Ellipsen. Im Anhang, Abschnitt 4.2 wird die Konstruktion dieser Ellipsen näher erläutert.



d.h. \mathbf{x}_1 ist proportional zum ersten Eigenvektor von M , der in der ersten Spalte von T steht. \mathbf{t}_1 definiert 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 Ellipsoids.

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 (2.59) ist die Orientierung dieser Halbachse durch den ersten Eigenvektor \mathbf{t}_1 von M gegeben. Analoge Interpretationen ergeben sich für die übrigen Halbachsen des durch M definierten Ellipsoids:

$$\mathbf{x}_j = T\mathbf{y}_j \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = y_j \begin{pmatrix} t_{1j} \\ t_{2j} \\ \vdots \\ t_{nj} \end{pmatrix}, \quad \mathbf{x}_j \in \mathcal{E}_x, \quad \mathbf{y}_j = y_j\mathbf{e}_j \in \mathcal{E}_y \quad (2.60)$$

y_j ist die Länge der jeweiligen Halbachse.

Definition 2.6 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 Ellipsoid \mathcal{E}_y . T und T' heißen deshalb Hauptachsentransformationen.

Beispiel 2.3 Rotation der Fälle Es sei X eine spaltenzentrierte $(m \times n)$ -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 8. Denn es sei $C = X'X$, $\tilde{\mathbf{x}}_i' C \tilde{\mathbf{x}}_i = k_i$. Dann ist

$$\mathcal{E}_{x,i} = \{\mathbf{x} | \mathbf{x} \in \mathbb{R}^n, \mathbf{x}' C \mathbf{x} = k_i > 0\} \quad (2.61)$$

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 2.3, Seite 45, kann X stets als Produkt $X = UV'$ zweier Matrizen U und V' dargestellt werden, wobei die Spaltenvektoren von V linear unabhängig 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.

$$\tilde{\mathbf{x}}_i = V\tilde{\mathbf{u}}_i, \quad \tilde{\mathbf{u}}_i = V'\tilde{\mathbf{x}}_i, \quad (2.62)$$

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

$$\mathcal{E}_{u,i} = \{\tilde{\mathbf{u}} | \tilde{\mathbf{u}}' \Lambda \tilde{\mathbf{u}} = k_i > 0\} \quad (2.63)$$

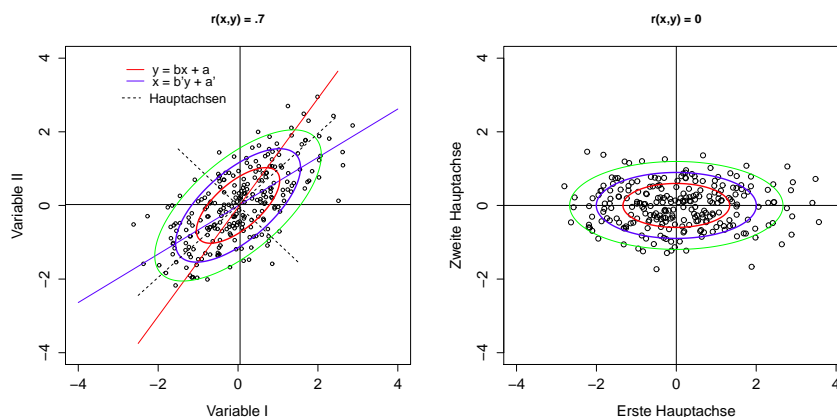
wobei $k_i = \tilde{\mathbf{u}}_i' \Lambda \tilde{\mathbf{u}}_i$. Die Ellipsoide $\mathcal{E}_{u,i}$ haben für die verschiedenen i alle dieselbe Orientierung, da Λ für alle Ellipsoide identisch ist, und da Λ eine Diagonalmatrix ist sind die Ellipsoide $\mathcal{E}_{u,i}$ achsenparallel. Abbildung 9 (S. 62) illustriert die Rotation der Punktekonfigurationen. \square

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. 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 10 wird dieser Sachverhalt illustriert.

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

Satz 2.12 *Es sei M eine reelle, symmetrische $(n \times n)$ -Matrix. M ist positiv semidefinit dann und nur dann, wenn die Eigenwerte λ_j größer als bzw. mindestens gleich Null sind.*

Abbildung 9: 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 4.2, Seite 122.



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

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

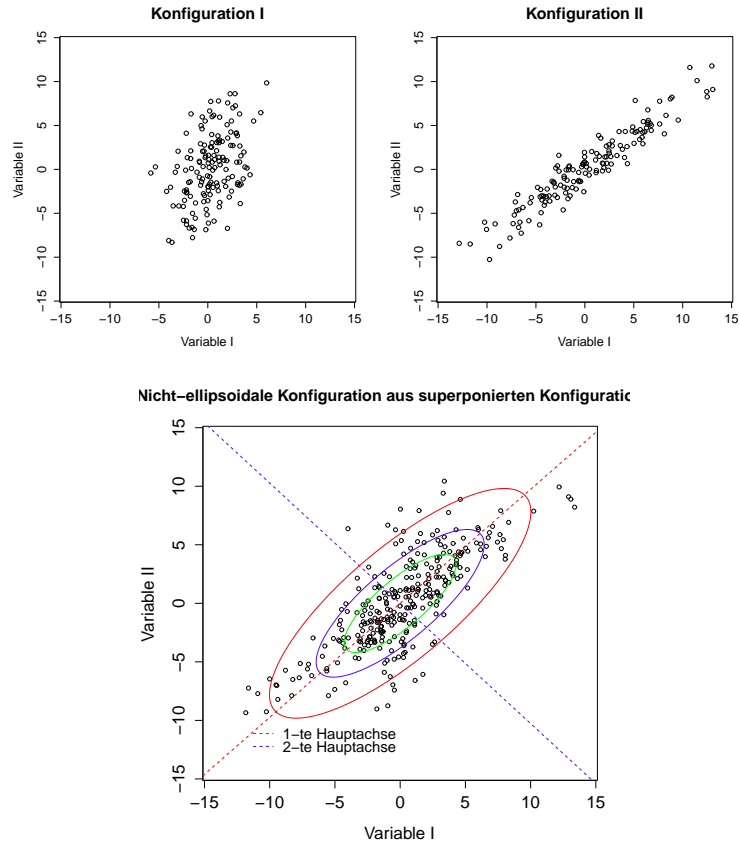
Da \mathbf{x} beliebig gewählt werden kann, kann insbesondere $\mathbf{x} = T\mathbf{e}_j$ gewählt werden, also $\mathbf{y}_j = \mathbf{e}_j$ der j -te Einheitsvektor, $j = 1, \dots, n$. Dann ist $q = \lambda_j$, und $q \geq 0$ genau dann, wenn $\lambda_j \geq 0$.

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

Um den folgenden Satz zu motivieren, sei wieder an eine $(m \times n)$ -Datenmatrix erinnert; X habe den Rang $r \leq \min(m, n)$. Dann hat die symmetrische, positiv-definite Varianz-Kovarianz-Matrix $S = X'HX$ ebenfalls den Rang r . Die Anzahl der von Null verschiedenen Eigenwerte von S gibt Auskunft über den Rang von S und damit von X :

Satz 2.13 *Es sei M eine beliebige reelle, symmetrische $(n \times n)$ -Matrix. Dann ist der Rang von M gleich der Anzahl von Null verschiedener Eigenwerte.*

Abbildung 10: Superponierte Punktekonfigurationen und Ellipsen



Beweis: Es sei $\text{rg}(M) = r \leq n$. Nach Satz 2.3, Seite 45, existieren dann zwei Matrizen U und V jeweils vom Rang r derart, dass $M = UV'$. Andererseits gilt $M = T\Lambda T'$, wobei T und Λ ($n \times n$)-Matrizen sind. Es seien $r \leq n$ Eigenwerte ungleich Null. Dann enthält Λ $n-r$ Spalten (und Zeilen), die nur Nullen enthalten, und das Produkt $T\Lambda$ enthält ebenfalls $n-r$ Spalten, die nur Nullen enthalten. Es sei T_r die Matrix, die aus T entsteht, wenn die aus Nullen bestehenden Spalten weggelassen werden, und Λ_r sei die Diagonalmatrix, deren Diagonale nur die von Null verschiedenen Eigenwerte enthält. Dann kann $U = T_r\Lambda_r$ und $V = T_r$ gesetzt werden, und U muß den Rang r haben, da die Spaltenvektoren von T_r und damit auch von $T_r\Lambda_r$ orthogonal sind. Dementsprechend hat auch V den Rang r , und nach Satz 2.3 hat dann $UV' = M$ den Rang r . \square

Der Satz 2.12 sagt noch wenig über die Eigenschaften einer reellen, symmetrischen Matrix aus. Der folgende Satz gibt weitere Auskunft.

Satz 2.14 *Es sei M eine symmetrische $(n \times n)$ -Matrix vom Rang $r \leq n$. Dann ist M genau dann positiv semidefinit, wenn eine $(n \times r)$ -Matrix G existiert derart, dass*

$$M = GG'. \quad (2.64)$$

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 $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, $\lambda_j \geq 0$ für $j = 1, \dots, n$ (s. Satz 2.12), mit $M = T\Lambda T'$. Es sei

$$\Lambda^{1/2} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_r}, \underbrace{0, \dots, 0}_{n-r}).$$

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. \square

Der Satz 2.14 spezifiziert die Bedingungen, die eine symmetrische Matrix M erfüllen muß, um eine Ellipse bzw. ein Ellipsoid zu definieren.

Eine Rotationsmatrix S ist orthonormal. Nun sei umgekehrt bekannt, dass $S = T$ eine Matrix von Eigenvektoren von $M' = M$ ist. Die Frage ist, ob nun auch folgt, dass S orthonormal ist, – es ist ja denkbar, dass man *nicht* über eine Rotation auf die Eigenschaft der Spalten von S , Eigenvektoren zu sein, gekommen ist, und vielleicht gibt es auch nicht-orthogonale Eigenvektoren von M . Dazu wird der folgende Satz bewiesen:

Satz 2.15 *$M \in \mathbb{R}^{n,n}$ sei symmetrisch und habe die Eigenvektoren $\mathbf{t}_1, \dots, \mathbf{t}_n$. Sind \mathbf{t}_j und \mathbf{t}_k mit zugehörigen Eigenwerten $\lambda_j \neq \lambda_k$ irgendzwei Eigenvektoren von M , so sind \mathbf{t}_j und \mathbf{t}_k orthogonal, d.h. es gilt*

$$\mathbf{t}_j'\mathbf{t}_k = \begin{cases} 0, & j \neq k \\ \|\mathbf{t}_j\| \neq 0, & j = k \end{cases} \quad (2.65)$$

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 eine Matrix von Eigenvektoren und für irgendzwei Eigenvektoren \mathbf{t}_j und \mathbf{t}_k gelte $\lambda_j \neq \lambda_k$. Dann sind \mathbf{t}_j und \mathbf{t}_k orthogonal. Denn dann gilt

$$M\mathbf{t}_j = \lambda_j\mathbf{t}_j \quad (2.66)$$

$$M\mathbf{t}_k = \lambda_k\mathbf{t}_k \quad (2.67)$$

Die Gleichung (2.66) werde von links mit \mathbf{t}'_k , die Gleichung (2.67) von links mit \mathbf{t}_j multipliziert. Es entstehen die Gleichungen

$$\mathbf{t}'_k M \mathbf{t}_j = \lambda_j \mathbf{t}'_k \mathbf{t}_j \quad (2.68)$$

$$\mathbf{t}'_j M \mathbf{t}_k = \lambda_k \mathbf{t}'_j \mathbf{t}_k. \quad (2.69)$$

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. \square

Anmerkung: In (2.65) 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. Ist bereits bekannt, dass T auch eine Rotationsmatrix ist, so wird die Normiertheit der \mathbf{t}_j gewissermaßen gleich mitgeliefert. \square

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 2.16 *Es sei λ_j ein Eigenwert der symmetrischen Matrix M mit der Mehrfachheit m , d.h. es gelte $\lambda_j = \lambda_{j+1} = \dots = \lambda_{j+m}$. Dann existieren m orthogonale, zu diesen Eigenwerten korrespondierende Eigenvektoren.*

Beweis: Der Beweis wird hier nicht gegeben, da er vom Begriff der Determinante bzw. von Aussagen über die Anzahl von Nullstellen von Polynomen Gebrauch macht; Determinanten werden zwar im Anhang (Abschnitt 4.6) kurz eingeführt, aber in diesem Skript nicht weiter diskutiert. Auch Polynome werden hier nicht weiter betrachtet. \square

Spektraldarstellung von M : Die Orthonormalität von T bedeutet, dass man aus $MT = T\Lambda$ (Gleichungen (2.68) und (2.69)) durch Multiplikation von rechts mit T' die Beziehung

$$M = T\Lambda T' = \sum_{k=1}^n \lambda_k \mathbf{t}_k \mathbf{t}'_k. \quad (2.70)$$

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 $T = [\mathbf{t}_1, \dots, \mathbf{t}_n]$ und $T\Lambda = [\lambda_1 \mathbf{t}_1, \dots, \lambda_n \mathbf{t}_n]$, so dass

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

Definition 2.7 Die Darstellung (2.70) von M heißt Spektraldarstellung von M .

Bemerkung 2.1 Da die Matrix T der Eigenvektoren einer symmetrischen Matrix M stets orthonormal ist, kann sie als eine Rotationsmatrix betrachtet werden, die die Vektoren $\mathbf{y} \in \mathcal{E}_y$ in die Vektoren $\mathbf{x} \in \mathcal{E}_x$ rotiert, wobei Λ die Diagonalmatrix der Eigenwerte von M ist. Umgekehrt rotiert T' die $\mathbf{x} \in \mathcal{E}_x$ in die Vektoren $\mathbf{y} \in \mathcal{E}_y$.

Bemerkung 2.2 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, worauf hier aber nicht weiter eingegangen wird). Dazu betrachte man die Matrix (2.48), Seite 54, 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} nur dann parallel zu \mathbf{x} für diejenigen Werte von ϕ ist, für die $\cos \phi = 1$ und $\sin \phi = 0$ ist, 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.

2.5.3 Der Rayleigh-Quotient

In Definition 2.4, Gleichung (2.49) (Seite 56) wurde der Begriff der quadratischen Form $\mathbf{x}'M\mathbf{x}$ eingeführt. Sie definiert ein n -dimensionales Ellipsoid, wenn M eine symmetrische, positiv definite (n, n) -Matrix ist (Definition 2.4, Seite 56) und dementsprechend $\mathbf{x} \in \mathbb{R}^n$ ist. Wie es scheint, hat die erste Hauptachse der Ellipsoide stets eine Orientierung, die der maximalen Ausdehnung der Punktekongfiguration entspricht.

Diese Vermutung ist berechtigt. In der Tat kann man von der Forderung ausgehen, dass eine latente Dimension bestimmt werden soll derart, dass sie die Orientierung der maximalen Ausdehnung der Konfiguration hat, die zweite latente Dimension soll dann orthogonal zur ersten orientiert sein, etc. Generell kann man von der Darstellung $X = UV'$ ausgehen. Allerdings gibt es beliebig viele Möglichkeiten, die Matrizen U bzw. V zu wählen (hat man eine Matrix gewählt, so liegt die jeweils andere Matrix fest). Es werde die

Annahme: V ist die Matrix der Eigenvektoren von $X'X$. □

gemacht. Wie in Beispiel 2.3, Seite 60, schon gezeigt wurde, ist die Matrix U dann orthogonal. Der erste Spaltenvektor \mathbf{u}_1 von U

$$\vec{1}'X = (\vec{1}'U)V' = \vec{0}',$$

d.h. $\vec{1}'U = \vec{0}'$, die Spaltensummen von U sind gleich Null, d.h. die Summe der Koordinaten der Punkte auf der ersten Geraden ist gleich Null. Die Summe der Quadrate u_{i1}^2 ist dann proportional zur Varianz der Koordinaten auf der ersten Achse. Die Gerade hat die Orientierung der maximalen Ausdehnung, wenn diese Varianz maximal wird. Man kann also die Orientierung der maximalen Ausdehnung bestimmen, indem man diese Varianz als Funktion des Vektors \mathbf{v}_1 maximiert. Nun folgt aus $U = XV$ für die erste Spalte von U die Gleichung $\mathbf{u}_1 = X\mathbf{v}_1$, \mathbf{v}_1 der erste Spaltenvektor von V . Die Varianz der Koordinaten auf der ersten Achse ist aber proportional zu $\|\mathbf{u}_1\|^2 = \mathbf{u}_1'\mathbf{u}_1$, so dass man

$$\|\mathbf{u}_1\|^2 = \mathbf{v}_1' C \mathbf{v}_1, \quad C = X'X \tag{2.71}$$

zu maximieren hat. $\mathbf{v}_1' C \mathbf{v}_1$ ist aber der Ausdruck einer quadratischen Form. Er entspricht dem als *Rayleigh-Quotienten* definierten Ausdruck:

Definition 2.8 *Es sei M eine symmetrische Matrix. Der Quotient*

$$R(\mathbf{x}) = \frac{\mathbf{x}'M\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \frac{\mathbf{x}'M\mathbf{x}}{\|\mathbf{x}\|^2} \tag{2.72}$$

heißt Rayleigh-Quotient¹⁹ oder Rayleigh-Koeffizient.

Man beachte, dass der Vektor \mathbf{x} in (2.72) beliebige Länge haben kann, aber wegen

$$\frac{\mathbf{x}'M\mathbf{x}}{\|\mathbf{x}\|^2} = \frac{\mathbf{x}'}{\|\mathbf{x}\|} M \frac{\mathbf{x}}{\|\mathbf{x}\|}$$

geht analog zu (2.71) nur der normierte Vektor \mathbf{x} in die Definition des Rayleigh-Quotienten ein. Der Maximierung von $\|\mathbf{u}_1\|^2$ entspricht die Maximierung eines Rayleigh-Quotienten in Abhängigkeit von der Orientierung von \mathbf{v}_1 bzw. allgemein $\mathbf{x}/\|\mathbf{x}\|$. Eine Methode zur Maximierung besteht darin, $R(\mathbf{x})$ als Funktion

¹⁹Nach dem britischen Physiker John William Strutt, Dritter Baron Rayleigh (1842–1919)

der Komponenten von \mathbf{x} zu maximieren, wobei die Nebenbedingung $\mathbf{x}'\mathbf{x} = 1$ erfüllt sein muß; dieses Vorgehen wird im Anhang, Abschnitt 4.3 besprochen. In diesem Abschnitt wird ein anderer Weg besprochen, der ohne die Anwendung der Differentialrechnung auskommt. Zur Erinnerung: $\mathbf{v}'_1 C \mathbf{v}_1$ bzw. $\mathbf{x}' M \mathbf{x}$ definieren, als quadratische Formen, Ellipsoide. Gemäß dem folgenden Satz von Courant-Fischer haben die Ellipsoide, die durch eine symmetrische, positiv-semidefinite Matrix M (also insbesondere durch eine Varianz-Kovarianz-Matrix $C = X'X$) definiert werden, eine Orientierung, die der maximalen Ausdehnung der zu M korrespondierenden Punktekonfiguration entspricht:

Satz 2.17 (Satz von Courant-Fischer) *Es sei M eine symmetrische, positiv definite Matrix; es gilt $M = T\Lambda T'$, T die Matrix der Eigenvektoren und*

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n), \quad \lambda_1 \geq \dots \geq \lambda_n$$

die Diagonalmatrix der zugehörigen Eigenwerte von M . Dann ist

$$\max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}' M \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \max_j \lambda_j = \lambda_1, \quad (2.73)$$

und der Vektor \mathbf{x} , für den das Maximum angenommen wird, ist der zu λ_1 korrespondierende Eigenvektor \mathbf{t}_1 . Weiter gilt

$$\min_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}' M \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \min_j \lambda_j, \quad (2.74)$$

mit dem zugehörigen Eigenvektor \mathbf{t}_{\min} .

Beweis: Sei $T = [\mathbf{t}_1, \dots, \mathbf{t}_n]$ die Matrix der Eigenvektoren von M mit den zugehörigen Eigenwerten $\lambda_1 \geq \dots \geq \lambda_n$, so dass $MT = T\Lambda$, $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$. Die Eigenvektoren \mathbf{t}_j , $j = 1, \dots, n$ sind eine orthonormale Basis des $V_n = \mathcal{L}(\mathbf{t}_1, \dots, \mathbf{t}_n)$. Es sei nun

$$\mathbf{x} = c_1 \mathbf{t}_1 + \dots + c_n \mathbf{t}_n$$

für beliebige Koeffizienten c_1, \dots, c_n . Dann ist

$$\mathbf{x}' M \mathbf{x} = (c_1 \mathbf{t}_1 + \dots + c_n \mathbf{t}_n)' M (c_1 \mathbf{t}_1 + \dots + c_n \mathbf{t}_n) = \sum_{j=1}^n c_j^2 \mathbf{t}'_j M \mathbf{t}_j = \sum_{j=1}^n c_j^2 \lambda_j,$$

denn $c_j c_k \mathbf{t}'_j M \mathbf{t}_k = c_j c_k \lambda_k \mathbf{t}'_j \mathbf{t}_k = 0$ wegen $\mathbf{t}'_j \mathbf{t}_k = 0$ für $j \neq k$. Es 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 $\max_{\mathbf{x}} R_M(\mathbf{x}) = \lambda_1$, 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 $R_M(\mathbf{x}) = \max$ für $\mathbf{x} = \mathbf{t}_1$. \square

Die Aussage (2.74) ergibt sich aus der folgenden Vervollständigung des Satzes von Courant-Fischer und ist eher ein Korollar zu diesem Satz.

Satz 2.18 *Es sei M wie in Satz 2.17 definiert. Dann gilt*

$$\max_{\mathbf{x} \perp \mathbf{t}_1, \dots, \mathbf{t}_k} \frac{\mathbf{x}' M \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \lambda_{k+1}, \quad k < n, \quad (2.75)$$

für $\mathbf{x} = \mathbf{t}_{k+1}$ der $(k+1)$ -te Eigenvektor. (\perp steht für "ist orthogonal zu".)

Beweis: Es sei wieder $T = [\mathbf{t}_1, \dots, \mathbf{t}_n]$ die Matrix der Eigenvektoren von A . Dann existieren reelle Zahlen y_1, \dots, 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 $\mathbf{x} \perp \mathbf{t}_1, \dots, \mathbf{t}_k$ gelten. Dann muß aber

$$\mathbf{v}'_k \mathbf{x} = y_1 \mathbf{t}'_k \mathbf{t}_1 + y_2 \mathbf{t}'_k \mathbf{t}_2 + \dots + y_n \mathbf{t}'_k \mathbf{t}_n = y_k = 0$$

gelten, 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 $y_1 = \dots = y_k = 0$. Dann folgt aus (2.73)

$$\frac{\mathbf{x}' M \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \frac{\sum_{j=k+1}^n \lambda_j y_j^2}{\sum_{j=k+1}^n y_j^2},$$

und analog zur Argumentation im Beweis zu Satz 2.17 folgt (2.75). \square

Anmerkung: Für $k = 0$ betrachtet man $\max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}' M \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \lambda_1$, und für $k = n - 1$ erhält man

$$\max_{\mathbf{x} \perp \mathbf{t}_1, \dots, \mathbf{t}_{n-1}} \frac{\mathbf{x}' M \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \lambda_n,$$

d.h.

$$\min_{\mathbf{x}} R_M(\mathbf{x}) = \lambda_n \quad (2.76)$$

so dass

$$\lambda_n \leq R_M(\mathbf{x}) \leq \lambda_1. \quad (2.77)$$

\square

Die folgende Definition liefert einen kurzen Bezug auf den Satz von Courant-Fischer:

Definition 2.9 *Es sei M eine beliebige symmetrische $(n \times n)$ -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 gemäß den Sätzen (2.17) und (2.18) den Rayleigh-Quotienten $\mathbf{x}' M \mathbf{x} / \mathbf{x}' \mathbf{x}$ maximieren.*

2.5.4 Die Singularwertzerlegung einer Matrix

Es sei X eine beliebige $(m \times n)$ Matrix mit dem Rang $r \leq \min(m, n)$. Gesucht ist eine Basis einerseits für die Spaltenvektoren von X , andererseits für die Zeilenvektoren von X . Nach Satz 2.3, Seite 45, existieren stets Matrizen (m, r) - und (n, r) -Matrizen U und V derart, dass $X = UV'$. Da die Spaltenvektoren von U linear unabhängig sind, liegen die Koeffizientenvektoren \tilde{v}_j und damit die Matrix V' eindeutig fest (Satz 1.3, Seite 25). Die Wahl einer bestimmten Basis U legt die Wahl von V fest und umgekehrt. Sind die Spaltenvektoren von U orthonormal, so liefert die Multiplikation von $X = UV'$ von links mit U' die Beziehung $U'X = V'$ oder $V = X'U$, d.h. die Spaltenvektoren von V sind Linearkombinationen der Spalten von X' bzw. der Zeilenvektoren von X . Die Frage ist also, wie man zu einer orthonormalen (Teil-)Basis der Spaltenvektoren von X kommt.

Generell gilt, dass es beliebig viele Möglichkeiten für eine Wahl von U bzw. V gibt, so dass die Entscheidung für die Wahl einer bestimmten Basis nach irgendwelchen Kriterien erfolgen muß. Die Frage ist nun, welchen Annahmen bezüglich U und V gemacht werden können, ohne die Allgemeinheit unnötig einzuschränken.

Es sei T die Matrix der Eigenvektoren von $X'X$, und Q sei die Matrix der Eigenvektoren von XX' . Nach Korollar 2.1, Gleichung (2.33), Seite 47, gilt

$$\text{rg}(X'X) = \text{rg}(XX') = r.$$

Die Anzahl der von Null verschiedenen Eigenwerte von $X'X$ und XX' ist dann identisch. Es seien

$$\Lambda_1 = \text{diag}(\lambda_1, \dots, \lambda_n) \quad (2.78)$$

$$\Lambda_2 = \text{diag}(\mu_1, \dots, \mu_m) \quad (2.79)$$

Λ_1 enthält r Eigenwerte ungleich Null von $X'X$, und Λ_2 enthält ebenfalls r Eigenwerte ungleich Null von XX' . Dann gilt

$$(X'X)T = T\Lambda_1 \quad (2.80)$$

$$(XX')Q = Q\Lambda_2; \quad (2.81)$$

wobei Q die Eigenvektoren von XX' sind. Es gilt der

Satz 2.19 *Die von Null verschiedenen Eigenwerte von $X'X$ und XX' sind identisch, d.h. es gilt $\Lambda_{r1} = \Lambda_{r2} = \Lambda_r$, wobei Λ_{r1} und Λ_{r2} nur die von Null verschiedenen Eigenwerte von $X'X$ bzw. XX' enthalten.*

Beweis: Es sei $X = U'V = UT_r'$, wobei T_r die zu Eigenwerten ungleich Null von $X'X$ korrespondierenden Eigenvektoren von $X'X$ enthält. Dann folgt

$$U'U = T_r'X'XT_r = T'(T_r\Lambda_{r1}T_r')T = \Lambda_{r1},$$

d.h. U ist orthogonal. Insbesondere hat man für den j -ten Spaltenvektor \mathbf{u}_j von U

$$\mathbf{u}'_j \mathbf{u}_j = \|\mathbf{u}_j\|^2 = \lambda_j.$$

Dann ist

$$\mathbf{q}_j = \mathbf{u}_j \lambda_j^{-1/2} \quad (2.82)$$

normiert, so dass $Q = U \Lambda_{r1}^{-1/2}$ die normierten Spaltenvektoren von U enthält. Dann ist $U = Q_r \Lambda_{r1}^{1/2}$, und man hat

$$X = Q_r \Lambda_{r1}^{1/2} T'_r. \quad (2.83)$$

Die Spaltenvektoren von Q_r sind orthonormal, und man hat weiter

$$X X' = Q_r \Lambda_{r1}^{1/2} T'_r T_r \Lambda_{r1} Q'_r = Q_r \Lambda_{r1} Q_r$$

und wegen der Orthonormalität der Spalten von Q_r folgt

$$(X X') Q_r = Q_r \Lambda_{r1}. \quad (2.84)$$

Dies heißt aber, dass Q_r die zu Eigenwerten ungleich Null korrespondierenden Eigenvektoren von $X X'$ enthalten muß, und Λ_{r1} ist die Diagonalmatrix der von Null verschiedenen Eigenwerte von $X X'$, so dass $\Lambda_{r1} = \Lambda_{r2}$ folgt. \square

Das Ziel vieler Datenanalysen ist, eine Menge von Basisvektoren zu finden, aus denen sich z.B. die Spaltenvektoren einer Datenmatrix als Linearkombinationen ergeben. Die Gleichung (2.83) liefert bereits einen Ansatz dafür; sie ist als *Singularwertzerlegung* (SVD²⁰) der Matrix X bekannt, s. die ausführliche Definition 2.86 weiter unten. Bevor dieser Aspekt von (2.83) weiter elaboriert wird, soll die Singularwertzerlegung in der Form vorgestellt werden, in der sie üblicherweise formuliert wird. Es sei Q die $(m \times m)$ -Matrix aller Eigenvektoren von $X X'$, also auch der Eigenvektoren, die zu den Eigenwerten gleich Null von $X X'$ korrespondieren. Analog dazu sei T die Matrix aller Eigenvektoren von $X' X$, also einschließlich der Eigenvektoren, die zu Eigenwerten gleich Null von $X' X$ korrespondieren, falls $r < \min(m, n)$. Ferner sei

$$\Sigma = \Lambda^{1/2} = \begin{pmatrix} \Lambda_r^{1/2} & 0 \\ 0 & 0 \end{pmatrix}, \quad (2.85)$$

wobei Λ_r die von Null verschiedenen Eigenwerte von $X' X$ bzw. $X X'$ enthält und die Nullen die $n - r$ Zeilen bzw. Spalten von Nullen repräsentieren, falls $r < \min(m, n)$. Dann hat man die

Definition 2.10 *Die Darstellung*

$$X = Q \Sigma T' \quad (2.86)$$

²⁰Singular Value Decomposition

heißt Singularwertzerlegung (*SVD = Singular Value Decomposition*) von X . Die Spaltenvektoren von Q heißen Linkssingulärvektoren, die von T heißen Rechtssingulärvektoren, und die Diagonalelemente $\sigma_j = \sqrt{\lambda_j}$ von Σ heißen Singularwerte.

Anmerkungen:

1. Nach Definition 2.6, Seite 60, repräsentiert die Matrix T der Eigenvektoren von $X'X$ eine Hauptachsentransformation der Spalten von X . Die Singularwertzerlegung von X ist äquivalent einer Hauptachsentransformation, denn T bewirkt eine Transformation der Spaltenvektoren von X in die Spaltenvektoren von $L = Q\Lambda^{1/2}$, d.h. L enthält die Koordinaten der Fälle auf den Hauptachsen der zu XX' korrespondierenden Ellipsoide und nach Definition 2.6, Seite 60, ist dies gerade die Hauptachsentransformation.
2. Die englische Bezeichnung für 'Singularwertzerlegung' ist *singular value decomposition*, abgekürzt SVD; diese Abkürzung ist auch im Deutschen üblich. Ein in der Psychologie häufig gebrauchter Ausdruck für die SVD ist 'Grundstruktur' einer Matrix (engl. basic structure). Der Ausdruck 'Singularwertzerlegung' ist allgemein in allen Wissenschaften, in denen eine Zerlegung von Matrizen gewünscht wird (Biologie, Medizin, Geologie, Klimaforschung, Archäologie, etc) gebräuchlich, weshalb auch hier von dieser Bezeichnung Gebrauch gemacht wird.
3. Die SVD ist nicht an eine Spaltenzentrierung oder Standardisierung der Matrix X gebunden; eine SVD kann für eine beliebige Matrix X bestimmt werden. □

Darstellung der SVD über das dyadische Produkt: Die Zerlegung $X = Q\Sigma T'$ kann in der Form

$$X = \sum_{k=1}^n \sigma_k \mathbf{q}_k \mathbf{t}'_k = \sum_{k=1}^n \sqrt{\lambda_k} \mathbf{q}_k \mathbf{t}'_k, \quad m \geq n \tag{2.87}$$

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

$$x_{ij} = \sqrt{\lambda_1} q_{i1} t_{j1} + \dots + \sqrt{\lambda_n} q_{in} t_{jn} \tag{2.88}$$

gegeben, $\sqrt{\lambda_1} q_{i1} t_{j1}$ ist der Beitrag der ersten latenten Dimension, etc.

Üblicherweise werden die Eigenwerte λ_k in Λ so angeordnet, 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

$$\sigma_1 > \sigma_2 > \dots > \sigma_n.$$

Die Matrix X wird in (2.87) 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:

$$\hat{X}_r = \sum_{k=1}^r \sigma_k \mathbf{q}_k \mathbf{t}'_k \approx X, \quad \sigma_k = \sqrt{\lambda_k} \quad (2.89)$$

"Hinreichend" klein bedeutet, dass man die σ_k mit $k > r$ als eben nur "zufällig" von Null abweichend betrachtet. Darauf wird in Abschnitt 2.5.5 noch eingegangen.

Zwischen den Eigenvektoren \mathbf{q}_k von XX' und den Eigenvektoren \mathbf{t}_k von $X'X$ bestehen die folgenden Beziehungen: Durch Multiplikation der SVD $X = Q\Sigma T'$ von rechts mit T erhält man

$$XT = Q\Sigma, \quad (2.90)$$

und die Multiplikation der SVD von links mit Q' ergibt $Q'X = \Sigma T'$, oder

$$X'Q = T\Sigma, \quad (2.91)$$

so dass man für die Eigenvektoren \mathbf{q}_k und \mathbf{t}_k die Beziehungen

$$X\mathbf{t}_k = \sigma_k \mathbf{q}_k \quad (2.92)$$

$$X'\mathbf{q}_k = \sigma_k \mathbf{t}_k, \quad k = 1, \dots, n \quad (2.93)$$

erhält. Diese beiden Gleichungen zeigen die Beziehungen zwischen den \mathbf{t}_k und den \mathbf{q}_k : X – als Transformationsmatrix gesehen – bildet \mathbf{t}_k auf $\sigma_k \mathbf{q}_k$ ab, und X' bildet \mathbf{q}_k auf $\sigma_k \mathbf{t}_k$ ab: \mathbf{q}_k ist eine Linearkombination der *Spaltenvektoren* von X , mit den Komponenten von \mathbf{t}_k als Koeffizienten, und \mathbf{t}_k ist eine Linearkombination der *Zeilenvektoren* von X , mit den Komponenten von \mathbf{q}_k als Koeffizienten.

Anmerkung: In Abschnitt 4.3.4, Seite 127, wird der Satz von Courant-Fischer durch Differentiation der quadratischen Form (2.75), Seite 69 (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. \square

2.5.5 Die PCA als Anwendung der SVD

Die Hauptkomponentenanalyse (engl. Principal Component Analysis (PCA); PCA ist die gängige Abkürzung) ist eine Methode, einen als eine $(m \times n)$ -Matrix X

gegebenen Datensatz in einer der Faktorenanalyse ähnlichen Weise zu komprimieren; die Methode geht auf Karl Pearson (1901) und Harold Hotelling (1933) zurück. 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.

Es wird angenommen, dass die Zeilen von X die "Fälle" repräsentieren, – das sind Personen oder Objekte, an denen die Messungen von n Variablen vorgenommen werden, die von den Spalten von X repräsentiert werden. Es ist auch möglich, dass die Fälle Zeitpunkte sind, zu denen die n Variablen bei einem Objekt oder einer Person vorgenommen werden. Dem klassischen, auf Hotelling (1933) zurückgehenden Ansatz zufolge wird zunächst eine latente Variable \mathbf{Y}_1 als Transformation $\mathbf{Y}_1 = X\mathbf{t}$ der Spaltenvektoren \mathbf{x}_j gesucht derart, dass $\mathbf{Y}'_1\mathbf{Y}_1$ maximal wird. Dazu wird $\mathbf{t}'X'X\mathbf{t} = \mathbf{Y}'_1\mathbf{Y}_1$ als Funktion des Vektors \mathbf{t} maximiert, wobei allerdings eine Nebenbedingung bezüglich \mathbf{t} eingeführt werden muß. Die wird durch die Annahme $\mathbf{t}'\mathbf{t} = 1$ definiert. Die Maximierung muß hier nicht mehr nachvollzogen werden, aufgrund des Satzes von Courant-Fischer weiß man bereits, dass $\mathbf{t} = \mathbf{t}_1$ der erste, d.h. zum maximalen Eigenwert korrespondierende Eigenvektor von $X'X$ sein muß, d.h. $\mathbf{Y}'_1\mathbf{Y}_1 = \lambda_1$. Dann wird die zweite Hauptkomponente \mathbf{y}_2 bestimmt, indem ein Vektor \mathbf{t}_2 gesucht wird, der $\mathbf{Y}'_2\mathbf{Y}_2$ maximiert, wieder unter der Nebenbedingung $\mathbf{t}'_2\mathbf{t}_2 = 1$ und $\mathbf{Y}'_1\mathbf{Y}_2 = 0$, d.h. \mathbf{y}_2 soll orthogonal zu \mathbf{Y}_1 sein. Dann erweist sich \mathbf{t}_2 als Eigenvektor von $X'X$, der zum zweitgrößten Eigenwert $\mathbf{Y}'_2\mathbf{Y}_2 = \lambda_2$ korrespondiert. Dieses Vorgehen führt auf die SVD, so dass man gleich von der SVD Gebrauch machen kann.

Im Folgenden wird angenommen, dass X spaltenzentriert oder sogar spaltenstandardisiert ist. $C = \frac{1}{m}X'X$ ist dann eine Kovarianz- bzw. eine Korrelationsmatrix.

Die folgenden Gleichungen drücken zwei Varianten mit den im Zusammenhang mit der PCA üblichen Bezeichnungen für die Matrizen aus:

$$X = Q\Lambda^{1/2}T' = LT', \quad L = Q\Lambda^{1/2} \quad (2.94)$$

$$= QA', \quad A = T\Lambda^{1/2} \quad (2.95)$$

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 Allgemeinen die Eigenwerte *numerisch*²¹ 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 $(m \times n)$ -Matrix, T bzw. A ist eine $(n \times n)$ -Matrix, und Λ ist eine $(n \times n)$ -Diagonalmatrix. Die Spalten von Q bzw. L sind Basisvektoren für die Spalten-

²¹Messungen 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.

vektoren \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 2.11 *Es seien*

$$\mathbf{L}_k = \begin{pmatrix} \ell_{1k} \\ \ell_{2k} \\ \vdots \\ \ell_{mk} \end{pmatrix}, \quad \mathbf{a}_k = \begin{pmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{nk} \end{pmatrix},$$

\mathbf{L}_k ist der k -te Spaltenvektor von L , \mathbf{a}_k ist der k -te Spaltenvektor von A (s. (2.95)). Die Komponenten ℓ_{ik} von \mathbf{L}_k , $i = 1, \dots, m$, heißen Faktorwerte (factor scores) der Fälle auf der k -ten latenten Variablen, die Komponenten a_{jk} von \mathbf{a}_k , $j = 1, \dots, n$, heißen Ladungen (factor loadings) der Variablen auf der k -ten latenten Variablen,.

Die Bezeichnung L für die Matrix $Q\Lambda^{1/2}$ soll an den Ausdruck 'latent' erinnern, A ist die in der Literatur übliche Bezeichnung 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 einem Faktorwert gemeint ist. \square

Aus (2.94) und (2.95) erhält man sofort die Beziehungen

$$X = LT' = QA' \quad (2.96)$$

$$X' = AQ' = TL' \quad (2.97)$$

Da T orthonormal ist, folgt aus (2.96) sofort der Hotellingsche Ansatz $XT = L$, d.h. die Spaltenvektoren von L sind Linearkombinationen der Spaltenvektoren \mathbf{x}_j von X . Andererseits sind nach (2.96) 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 Spaltenvektoren \mathbf{t}_k von T ; insbesondere gilt

$$\mathbf{x}_j = L\tilde{\mathbf{t}}_j = Q\tilde{\mathbf{a}}_j \quad (2.98)$$

$$\tilde{\mathbf{x}}_i = A\tilde{\mathbf{q}}_i = T\tilde{\mathbf{L}}_i \quad (2.99)$$

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. $\tilde{\mathbf{a}}_i$ repräsentieren 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 (2.99) beschreibt die in (2.53), Seite 57 angegebene Rotation von achsenparallelen Ellipsoiden in orientierte Ellipsoide und umgekehrt. Denn (2.99)

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

$$\ell_{ik} = \tilde{\mathbf{x}}'_i \mathbf{t}_k = \sum_{j=1}^n x_{ij} t_{jk} = \|\tilde{\mathbf{x}}_i\| \|\mathbf{t}_k\| \cos \theta_{ik}. \quad (2.100)$$

$$a_{jk} = \mathbf{x}'_j \mathbf{q}_k = \sum_{i=1}^m x_{ij} q_{ik} = \|\mathbf{x}_j\| \|\mathbf{q}_k\| \cos \phi_{jk}. \quad (2.101)$$

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

$$x_{ij} = \tilde{\mathbf{q}}'_i \tilde{\mathbf{a}}_j = \|\tilde{\mathbf{q}}_i\| \|\tilde{\mathbf{a}}_j\| \cos \varphi_{ij} = \sum_{k=1}^n \sqrt{\lambda_k} q_{ik} t_{kj} \quad (2.102)$$

und x_{ij} wird maximal (relativ zur Länge der Vektoren $\tilde{\mathbf{q}}_i$ und $\tilde{\mathbf{a}}_j$), 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

$$X'X = AA'. \quad (2.103)$$

Es gelte $X = Z$, d.h. die Matrix X sei spaltenstandardisiert. Dann ist

$$R = \frac{1}{m} X'X = \frac{1}{m} AA', \quad X = Z \quad (2.104)$$

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 :

$$r_{jk} = \frac{1}{m} \tilde{\mathbf{a}}_j' \tilde{\mathbf{a}}_k = \frac{1}{m} \sum_{u=1}^n a_{uj} a_{uk} = \frac{1}{m} \|\tilde{\mathbf{a}}_j\| \|\tilde{\mathbf{a}}_k\| \cos \theta_{jk}, \quad (2.105)$$

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 $\theta_{jj} = 0$)

$$r_{jj} = \frac{1}{m} \|\tilde{\mathbf{a}}_j\|^2 = \frac{1}{m} \sum_{u=1}^n a_{ju}^2 = 1, \quad j = 1, \dots, n \quad (2.106)$$

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 k -te Spaltenvektor von A . Dann gilt

$$\|\mathbf{a}_k\|^2 = \sum_{jh=1}^n a_{jk}^2 = \sum_{j=1}^n \lambda_k t_{jk}^2 = \lambda_k \sum_{j=1}^n t_{jk}^2 = \lambda_k, \quad (2.107)$$

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. Andererseits ist die Matrix Q spaltenzentriert, wenn X spaltenzentriert ist: it $\vec{1}$ 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

$$\|\mathbf{L}_k\|^2 = \sum_{i=1}^m \ell_{ik}^2 = \sum_{i=1}^m \lambda_k q_{ik}^2 = \lambda_k, \quad (2.108)$$

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} (der Proportionalitätsfaktor ist $1/m$). λ_k korrespondiert also zur Varianz der Koordinaten (Scores) der Fälle auf der k -ten latenten Dimension. Zusammen mit (2.107) hat man die Beziehung

$$\|\mathbf{L}_k\|^2 = \|\mathbf{a}_k\|^2 = \lambda_k, \quad k = 1, \dots, n \quad (2.109)$$

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

$$\pi_k = \frac{\lambda_k}{\sum_{j=1}^n \lambda_j} \quad (2.110)$$

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 (2.86) ist äquivalent zu

$$X = \sigma_1 \mathbf{q}_1 \mathbf{t}'_1 + \sigma_2 \mathbf{q}_2 \mathbf{t}'_2 + \cdots + \sigma_n \mathbf{q}_n \mathbf{t}'_n = \sum_{k=1}^n \sigma_k \mathbf{q}_k \mathbf{t}'_k \quad \sigma_k = \sqrt{\lambda_k}. \quad (2.111)$$

Zusammen mit (2.110) kann (2.111) 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 2.20 (*Satz von Eckart & Young*) *Die Approximation*

$$X \approx X_r = Q_r \Lambda_r^{1/2} T_r' = \sum_{k=1}^r \sqrt{\lambda_k} \mathbf{q}_k \mathbf{t}'_k, \quad r < n \quad (2.112)$$

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 2.5.10 angeboten, vergl. insbesondere den Beweis zu Satz 2.23, S. 88. Dort wird der Begriff der Matrixnorm vorausgesetzt, vergl. Abschnitt 2.5.9. \square

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 (2.100) und (2.101) 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 Ausdruck 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 (2.94) und (2.95) lassen sich sofort die Gleichungen

$$L = Q\Lambda^{1/2} = XT, \quad A = T\Lambda^{1/2} = X'Q \quad (2.113)$$

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:

$$a_{jk} = \sqrt{\lambda_k} t_{jk} = x_{1j}q_{1k} + x_{2j}q_{2k} + \cdots + x_{mj}q_{mk}, \quad (2.114)$$

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., $i = 1, \dots, m$. 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

$$\ell_{ik} = \sqrt{\lambda_k} q_{ik} = x_{i1}t_{1k} + x_{i2}t_{2k} + \cdots + x_{in}t_{nk}. \quad (2.115)$$

ℓ_{ik} ist demnach ein gewogener Mittelwert der t_{1k}, \dots, t_{nk} , also den Ausprägungen der Variablen auf der k -ten latenten Variablen, mit den für den i -ten Fall spezifischen Gewichtungen x_{ij} , $j = 1, \dots, n$. 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.

2.5.6 Basiswechsel

Gegeben seien die m -dimensionalen Vektoren $\mathbf{x}_1, \dots, \mathbf{x}_n$, $\mathcal{L}_x = \mathcal{L}(\mathbf{x}_1, \dots, \mathbf{x}_n)$ sei die lineare Hülle dieser Vektoren und $\mathcal{B}_b = \{\mathbf{b}_1, \dots, \mathbf{b}_r\}$, $r \leq n$ sei eine Basis für \mathcal{L}_x , d.h. alle Vektoren von \mathcal{L}_x können als Linearkombinationen der Vektoren \mathbf{b}_j , $j = 1, \dots, r$ dargestellt werden. Oft ist es von Interesse, zu einer anderen Basis $\mathcal{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_r\}$ überzugehen, – etwa, um zu einer besseren Interpretation von Daten zu gelangen. Die \mathbf{c}_j sind Elemente von $\mathcal{L}(\mathcal{B}_b) = \mathcal{L}_x$ und als Linearkombinationen der Basisvektoren aus \mathcal{B}_b darstellbar, d.h. es existieren Koeffizienten t_{kj} derart, dass

$$\mathbf{c}_j = t_{1j}\mathbf{b}_1 + t_{2j}\mathbf{b}_2 + \dots + t_{rj}\mathbf{b}_r, \quad (2.116)$$

oder allgemein in Matrixform

$$C = BT \quad (2.117)$$

– genau dann werden ja die Spaltenvektoren von C als Linearkombinationen der Spaltenvektoren von B dargestellt. B und C haben jeweils r Spalten, also muß T eine $(r \times r)$ -Matrix sein. Weiter gilt

$$\text{rg}(C) \leq \min(\text{rg}(B), \text{rg}(T)).$$

Es ist aber $\text{rg}(C) = \text{rg}(B)$ so dass

$$r \leq \min(r, \text{rg}(T))$$

folgt, was wiederum bedeutet, dass $\text{rg}(T) = r$. Dann aber existiert die zu T inverse Matrix T^{-1} , so dass

$$CT^{-1} = BTT^{-1} = B \quad (2.118)$$

resultiert. Wird also der Übergang von einer Basis \mathcal{B} zu einer Basis \mathcal{C} durch eine Transformationsmatrix T definiert, so ist die Transformationsmatrix für den Übergang von C zu B durch die zu T inverse Matrix T^{-1} gegeben.

Die Basisvektoren in \mathcal{C} haben im Allgemeinen eine andere Orientierung und eine andere Länge als Basisvektoren in \mathcal{B} .

Nun werde gefordert, dass $\|\mathbf{c}_j\| = \|\mathbf{b}_j\|$ für alle j , d.h. die Transformation T bzw. T^{-1} soll die Längen invariant lassen. Dann repräsentiert T eine Rotation und ist folglich orthonormal (s. Satz 2.9, Seite 53). Die Längen der Basisvektoren von \mathcal{C} , also der Spaltenvektoren von C , sind dann identisch mit denen von B , d.h. für $\mathbf{c}_j = T\mathbf{b}_j$ gilt $\|\mathbf{c}_j\|^2 = \|\mathbf{b}_j\|^2$. T ist eine Hauptachsentransformation (s. Definition 2.6, Seite 60). Abbildung 9 (S. 62) illustriert die Hauptachsentransformation für den 2-dimensionalen Fall.

2.5.7 Die Inverse und die Wurzel einer symmetrischen Matrix

Die Inverse von M : Es sei $M' = M$ eine $(n \times n)$ -Matrix und M habe vollen Rang, so dass $\text{rg}(M) = n$. 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 2.45, Seite 52)

$$M^{-1} = (T\Lambda T')^{-1} = (T')^{-1}\Lambda^{-1}T^{-1},$$

wobei $\Lambda^{-1} = \text{diag}(\lambda_1^{-1}, \dots, \lambda_N^{-1})$. Die Orthonormalität von T impliziert $T' = T^{-1}$, so dass $(T')^{-1} = (T^{-1})^{-1} = T$, so dass die Inverse von M durch

$$M^{-1} = T\Lambda^{-1}T' = \sum_{k=1}^n \frac{1}{\lambda_k} \mathbf{t}_k \mathbf{t}'_k \quad (2.119)$$

gegeben ist. Die $\mathbf{t}_k \mathbf{t}'_k$ sind die dyadischen Produkte der Eigenvektoren von M .

Anmerkung: Die Beziehung (2.119) ist von Bedeutung u.a. bei der Interpretation von Regressionsparametern: die Stichprobenvarianzen der Schätzungen der Regressionsparameter $\mathbf{b} = (b_1, \dots, b_n)'$ 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 (2.122) zeigt, dass diese Varianzen groß werden, wenn es kleine Eigenwerte λ_k gibt.

Die Wurzel von M : Es sei $\Lambda^{1/2} = \text{diag}(\lambda_1^{1/2}, \dots, \lambda_n^{1/2})$. Dann gilt sicherlich

$$M = T\Lambda^{1/2}\Lambda^{1/2}T'.$$

Es sei nun²²

$$M^{1/2} \stackrel{\text{Def}}{=} T\Lambda^{1/2}T' = \sum_{k=1}^n \sqrt{\lambda_k} \mathbf{t}_k \mathbf{t}'_k \quad (2.120)$$

$\Lambda^{1/2}T'$ 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

$$M^{1/2}M^{1/2} = T\Lambda^{1/2}T'T\Lambda^{1/2}T' = T\Lambda T'.$$

Weiter folgt

$$(M^{1/2})' = (T\Lambda^{1/2}T')' = T\Lambda^{1/2}T', \quad (2.121)$$

d.h. $M^{1/2}$ ist symmetrisch, und

$$(M^{1/2})^{-1} = M^{-1/2} = (T\Lambda^{1/2}T')^{-1} = T\Lambda^{-1/2}T' = \sum_{k=1}^n \frac{1}{\sqrt{\lambda_k}} \mathbf{t}_k \mathbf{t}'_k. \quad (2.122)$$

²²Mit dem Zeichen $\stackrel{\text{Def}}{=}$ soll ausgedrückt werden, dass der Ausdruck auf der linken Seite durch den auf der rechten Seite definiert wird.

2.5.8 Die Pseudoinverse einer Matrix

Es werde das Gleichungssystem $A\mathbf{x} = \mathbf{y}$ betrachtet, wobei A eine $(m \times n)$ -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 2.12 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. \square

Für das Gleichungssystem $A\mathbf{x} = \mathbf{y}$ kann nun leicht eine Pseudoinverse gefunden werden, wenn $\text{rg}(A) = n$ 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

$$A'A\mathbf{x} = A'\mathbf{y} \Rightarrow \mathbf{x} = (A'A)^{-1}A'\mathbf{y}. \quad (2.123)$$

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

$$(A'A)^{-1}A'\mathbf{y} = \begin{cases} \mathbf{x}, & \mathbf{y} \in \mathcal{L}(A), \\ \hat{\mathbf{x}}, & \mathbf{y} \notin \mathcal{L}(A). \end{cases} \quad (2.124)$$

(vergl. Abschnitt 4.3.3, Gleichung (4.18), Seite 126).

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 $\text{rg}(A) = r < \min(m, n)$ (Stewart (1973)). Es sei

$$A = Q\Sigma T', \quad \Sigma = \begin{pmatrix} \Lambda_r^{1/2} & 0 \\ 0 & 0 \end{pmatrix} \quad (2.125)$$

die Darstellung von A durch die SVD. Q ist die $(m \times m)$ -Matrix der orthonormalen Eigenvektoren von AA' , T ist die $(n \times n)$ -Matrix der orthonormalen Eigenvektoren von $A'A$, und Λ_r ist eine $(r \times r)$ -Matrix der von Null verschiedenen Eigenwerte von AA' bzw. $A'A$. Σ ist eine $(m \times n)$ -Matrix, deren Elemente bis auf die Diagonalzellen von Λ_r gleich Null sind. Dann ist

$$A^+ = A' = T \begin{pmatrix} \Lambda_r^{-1/2} & 0 \\ 0 & 0 \end{pmatrix} Q' \quad (2.126)$$

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. (2.126) definiert tatsächlich eine Pseudoinverse.

2.5.9 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 kann allgemeiner definiert werden.

Die Norm eines Vektors definiert, in welchem Sinne von der "Größe" 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"²³. 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:

²³Um sich eine inhaltliche Vorstellung zu machen, stelle man sich vor dass die Komponenten x_i von \mathbf{x} Maße für Begabungen M_1, \dots, M_n repräsentieren. Dann ist $\|\mathbf{x}\|$ ein mögliches Maß für die Gesamtbegabung einer Person.

Definition 2.13 Es seien $\mathbf{x} \in \mathbb{R}^n$ n -dimensionale Vektoren. Eine Vektornorm ist eine Abbildung $f : \mathbb{R}^n \rightarrow \mathbb{R}$ (d.h. es wird einem Vektor \mathbf{x} eine bestimmte reelle Zahl zugeordnet), die den Bedingungen

1. $f(\mathbf{x}) \geq 0$,
2. $f(\mathbf{x} + \mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y})$,
3. $f(a\mathbf{x}) = af(\mathbf{x})$, für $a \in \mathbb{R}$

genügt. Dann heißt f eine Vektornorm. f wird durch $f(\mathbf{x}) = \|\mathbf{x}\|$ notiert. Der Einheitsvektor in Bezug auf eine Norm $\|\cdot\|$ ist derjenige Vektor, für den $\|\mathbf{x}\| = 1$ gilt.

Von besonderem Interesse sind die p -Normen

$$\|\mathbf{x}\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p} = \left(\sum_{j=1}^n |x_j|^p \right)^{1/p}. \quad (2.127)$$

Für $p = 1$ erhält man die 1-Norm

$$\|\mathbf{x}\|_1 = (|x_1| + \cdots + |x_n|) = \sum_{j=1}^n |x_j|. \quad (2.128)$$

und für $p = 2$ die euklidische Norm

$$\|\mathbf{x}\|_2 = (|x_1|^2 + \cdots + |x_n|^2)^{1/2} = \left(\sum_{j=1}^n |x_j|^2 \right)^{1/2} = \sqrt{\mathbf{x}'\mathbf{x}}. \quad (2.129)$$

Für $p = \infty$ schließlich findet man die Maximum-Norm

$$\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|. \quad (2.130)$$

Die Maximum-Norm ergibt sich aus der p -Norm für $p \rightarrow \infty$. 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 $|x_j|/|x_k| \leq 1$ für alle $j \neq k$ folgt $\lim_{p \rightarrow \infty} |x_j|^p/|x_k|^p \rightarrow 0$ für $j \neq k$ und $|x_j|^p/|x_k|^p = 1$ für $j = k$, so dass

$$\lim_{p \rightarrow \infty} \|\mathbf{x}\|_p = x_k = x_{\max}.$$

Matrixnormen: Der Begriff der Norm kann auch auf Matrizen angewendet werden:

Definition 2.14 Es sei $\mathbb{R}^{m \times n}$ die Menge der reellen $(m \times n)$ -Matrizen²⁴. Eine Matrixnorm ist eine Abbildung $\|\cdot\| : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}_+$, \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. $\|A + B\| \leq \|A\| + \|B\|$

gilt. Zusammen mit der Norm $\|\cdot\|$ wird der Vektorraum der $(m \times n)$ -Matrizen dann zu einem normierten Vektorraum²⁵ $(\mathbb{R}^{m \times n}, \|\cdot\|)$.

Es gibt verschiedene Normen, von denen hier einige als Beispiel genannt werden:

1. Die *Frobenius-Norm*.

$$\|A\|_F = \left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2}. \quad (2.131)$$

Für diese Norm wird auch der Name *Schur-Norm* oder *Hilbert-Schmidt-Norm* verwendet.

2. Die *p-Norm*: sie ist definiert durch

$$\|A\|_p = \max_{\mathbf{x} \neq \vec{0}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p} \quad (2.132)$$

Da $A\mathbf{x} = \mathbf{y}$ ein Vektor ist, ist $\|A\mathbf{x}\|_p$ eigentlich eine Vektornorm; allgemein heißen Normen der Form

$$\|A\| = \max_{\mathbf{x} \neq \vec{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}. \quad (2.133)$$

durch eine Vektornorm induzierte Normen.

Die Frobenius- und die *p-Norm* sind die am häufigsten vorkommenden Matrixnormen. Für $\|A\|_p$ gilt, wenn A eine $(m \times n)$ -Matrix ist,

$$\|A\|_p = \sup_{\mathbf{x} \neq \vec{0}} \left\| \left(A \frac{\mathbf{x}}{\|\mathbf{x}\|} \right) \right\|_p = \max_{\|\mathbf{x}\|_p=1} \|A\mathbf{x}\|_p. \quad (2.134)$$

²⁴Diese Definition ist etwas vereinfacht formuliert, eigentlich muß es heißen: es sei $\mathbb{K} = \mathbb{R}$ der Körper der reellen Zahlen und $\mathbb{K}^{m \times n} = \mathbb{R}^{m \times n}$ die Menge der reellen $(m \times n)$ -Matrizen, etc

²⁵Hier 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!

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 2.17 findet man

$$\|A\|_2 = \max_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2 = \sqrt{\lambda_{\max}}, \quad (2.135)$$

wobei λ_{\max} der maximale Eigenwert von $A'A$ ist. Für die Frobenius-Norm findet man

Satz 2.21 *Es sei A eine $(m \times n)$ -Matrix. Für die Frobenius-Norm $\|A\|_F$ gilt*

$$\|A\|_F^2 = \text{spur}(AA') = \sum_{i=1}^n \lambda_i \quad (2.136)$$

wobei $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ 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 $j = 1, \dots, n$. 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 $j = 1, \dots, n$. 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 (2.136) gezeigt. \square

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

$$\text{spur}(A'A) = \sum_{i=1}^n \lambda_i. \quad (2.137)$$

2.5.10 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 2.22 *Es seien A und A_k $(m \times n)$ -Matrizen, $m \geq n$, und es seien die Matrizen A und A_k durch*

$$A = Q\Lambda^{1/2}P' = \sum_{j=1}^n \sqrt{\lambda_j} \mathbf{q}_j \mathbf{p}_j', \quad A_k = Q\Lambda_k^{1/2}P' = \sum_{j=1}^k \sqrt{\lambda_j} \mathbf{q}_j \mathbf{p}_j' \quad (2.138)$$

definiert, wobei $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ mit $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ und $\Lambda_k = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k)$ mit $k < n$ sei. Dann gilt

$$\|A - A_k\|_2 = \sqrt{\lambda_{k+1}}. \quad (2.139)$$

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

$$A - A_k = Q\Sigma P' - Q\Sigma_k P' = Q(\Sigma - \Sigma_k)P' = Q\Sigma^* P',$$

$\Sigma^* = \text{diag}(\underbrace{0, \dots, 0}_k, \sigma_{k+1}, \dots, \sigma_n)$, $\sigma_j = \sqrt{\lambda_j}$. Da $\|A\|_2 = \sigma_{\max} = \sigma_1$ im Falle geordneter Singularwerte $\sigma_{k+1} \geq \dots \geq \sigma_n$, 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 , 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 (2.139) mit der Aussage (2.138) von Satz 2.17. Wie die Gleichung (2.88), 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 2.4, Seite 48). 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 2.23 *Es seien A und B ($m \times n$)-Matrizen mit $m > n$, wobei A den Rang r und B den Rang $k < r$ habe. Weiter sei*

$$A_k = Q\Lambda_k^{1/2}P' = \sum_{j=1}^k \sqrt{\lambda_j} \mathbf{q}_j \mathbf{p}'_j, \quad (2.140)$$

Λ_k die zu den ersten²⁶ k Eigenvektoren korrespondierenden Eigenwerte von X enthält. Dann gilt

$$\min_{B \in \mathbb{R}^{m,n}, \text{rg}(B)=k} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1} = \sqrt{\lambda_{k+1}} \quad (2.141)$$

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 2.24 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

$$\min_{B \in \mathbb{R}^{m,n}, \text{rg}(B)=k} \|A - B\|_2 = \|A - A_k\|_2.$$

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

$$\|(A - B_{\min})\mathbf{b}\| \leq \|(A - A_k)\mathbf{b}\| \leq \sigma_{k+1} \|\mathbf{b}\|.$$

²⁶Es wird angenommen, dass die Eigenwerte der Größe nach geordnet sind, $\lambda_1 > \lambda_2 > \dots > \lambda_k$

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 2.5, Gleichung (2.37) (Seite 49) ist aber $\text{rg}(B_{\min}P_{k+1}) \leq \min(\text{rg}(B_{\min}), \text{rg}(P_{k+1})) = k$, da ja $\text{rg}(B_{\min}) = k$ nach Voraussetzung. Dann folgt aber

$$\text{rg}(B_{\min}P_{k+1}) + \dim(\text{kern}(B_{\min}P_{k+1})) = k + 1,$$

d.h.

$$\dim(\text{kern}(B_{\min}P_{k+1})) \geq k + 1 - k = 1.$$

Also enthält $\text{kern}(B_{\min}P_{k+1})$ mindestens einen Vektor \mathbf{x} mit $B_{\min}P_{k+1}\mathbf{x} = \vec{0}$. Es sei also $\mathbf{x} \in \text{kern}(B_{\min}P_{k+1})$; 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 (2.141). \square

Satz 2.24 (Satz von Schmidt-Mirsky) *Es sei A und B_{\min} ($m \times n$)-Matrizen mit $m > n$, wobei A den Rang r und B den Rang $k < r$ habe, B_{\min} sei wie in Satz 2.23 definiert und A_k sei wie in (2.140) definiert. Dann gilt*

$$\|A - B_{\min}\|_F = \sqrt{\sum_{j=k+1}^n \lambda_j}, \quad (2.142)$$

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 $\text{spur}((\mathbf{y}_j - A'\mathbf{x}_j)(\mathbf{y}_j - A'\mathbf{x}_j)) \geq 0$, $\text{spur}(A'\mathbf{x}_j \mathbf{x}_j' A) = \|A'\mathbf{x}_j\|^2$ 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 $P_1 = [|\mathbf{p}_1| \dots |\mathbf{p}_k| 0]$, $P_2 = [0|\mathbf{p}_{k+1}| \dots |\mathbf{p}_n|]$, so dass $P = [P_1|P_2]$. Analog dazu sei $\Sigma_1 = \text{diag}(\sigma_1, \dots, \sigma_k)$, $\Sigma_2 = \text{diag}(\sigma_{k+1}, \dots, \sigma_n)$. Dann hat man

$$\begin{aligned} \|A'\mathbf{x}_j\|^2 &= \|Q\Sigma P'\mathbf{x}_j\|_F^2 = \|\Sigma P'\mathbf{x}_j\|_F^2 = \\ &= \|\Sigma_1 P'_1 \mathbf{x}_j\|_F^2 + \|\Sigma_2 P'_2 \mathbf{x}_j\|_F^2 + \lambda_k - \lambda_k + \lambda_k (\|P'\mathbf{x}_j\|^2 - \|P'_1 \mathbf{x}_j\|_F^2 - \|P'_2 \mathbf{x}_j\|_F^2) \\ &= \lambda_k + (\underbrace{\|\Sigma_1 P'_1 \mathbf{x}_j\|_F^2}_{(1)} - \underbrace{\lambda_k \|P'_1 \mathbf{x}_j\|_F^2}_{(2)}) - \lambda_k (1 - \|P'\mathbf{x}_j\|_F^2) \end{aligned}$$

Der Term (1) ist positiv, ebenso (2), da P orthonormal, und \mathbf{x}_j ist ebenfalls orthonormal. Dann folgt

$$\begin{aligned} \sum_{j=1}^k \|A'\mathbf{x}_j\|^2 &\leq k\lambda_k + \sum_{j=1}^k (\|\Sigma_1 P'_1\|^2 - \lambda_k \|P'_1 \mathbf{x}_j\|^2) \\ &= k\lambda_k + \sum_{j=1}^k \sum_{i=1}^k (\lambda_i - \lambda_j) |\mathbf{v}'_j \mathbf{x}_j|^2 \\ &\leq \sum_{i=1}^k (\lambda_k + (\lambda_i - \lambda_k)) = \sum_{j=1}^k \lambda_j. \end{aligned}$$

□

2.6 Basen und Transformationen von 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 , $j = 1, \dots, n$ als Linearkombination von Basisvektoren repräsentieren will. Bevor auf diese Frage näher eingegangen wird, sollen ein paar grundsätzliche Sachverhalte geklärt werden.

Es seien $\mathcal{B} = (\mathbf{b}_1, \dots, \mathbf{b}_n)$ und $\mathcal{C} = (\mathbf{c}_1, \dots, \mathbf{c}_n)$ 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

$$\mathbf{c}_j = t_{1j}\mathbf{b}_1 + \dots + t_{nj}\mathbf{b}_n = B\mathbf{t}_j, \quad j = 1, \dots, n \quad (2.143)$$

$$\mathbf{b}_j = s_{1j}\mathbf{c}_1 + \dots + s_{nj}\mathbf{c}_n = C\mathbf{s}_j \quad (2.144)$$

wobei B eine Matrix ist, die entsteht, wenn man die \mathbf{b}_j spaltenweise zu einer Matrix zusammenfasst, und C ist analog die Matrix, deren Spaltenvektoren die \mathbf{c}_j sind. $\mathbf{t}_j = (t_{1j}, \dots, t_{nj})'$ und $\mathbf{s}_j = (s_{1j}, \dots, s_{nj})'$. Ist T die Matrix mit den \mathbf{t}_j als Spaltenvektoren und S die Matrix mit den \mathbf{s}_j als Spaltenvektoren, so sind (2.143) und (2.144) äquivalent zu

$$C = BT \quad (2.145)$$

$$B = CS. \quad (2.146)$$

Setzt man die rechte Seite von (2.146) für B in (2.145) ein, so erhält man $C = CST$. C hat vollen Rang, so dass C^{-1} existiert und es folgt $C^{-1}C = C^{-1}CST = ST = I$. Da C und B Basen repräsentieren, müssen beide Matrizen den vollen Rang n haben, so dass $\text{rg}(C) = \text{rg}(BT)$. Nach Satz 2.5, Seite 49, gilt

$$\text{rg}(C) \leq \min(\text{rg}(B), \text{rg}(T)) \leq \text{rg}(B).$$

Aus der Definition von B und C als Matrizen, deren Spalten Basisvektoren sind, folgt natürlich schon, dass $\text{rg}(B) = \text{rg}(C)$ und damit $\text{rg}(B) = \text{rg}(C) = \text{rg}(T)$, woraus wiederum die Existenz von T^{-1} folgt. Dann hat man aber $STT^{-1} = T^{-1}$, also

$$S = T^{-1}. \quad (2.147)$$

Analog folgt die Existenz von S^{-1} .

Formal sind die verschiedenen Basen eines Vektorraums einander äquivalent. Will man Daten interpretieren, so muß man sich für eine der Basen entscheiden. Die Entscheidung muß anhand von Kriterien geschehen, die sich aus dem Begriff

des Vektorraums oder dem der Basis eines Vektorraumes selbst nicht herleiten lassen. Hier soll noch auf die Suche nach einer Basis für eine gegebene $(m \times n)$ -Datenmatrix X eingegangen werden.

Formal gesehen sind die Spaltenvektoren von X eine Stichprobe von n m -dimensionalen Vektoren aus einem m -dimensionalen Vektorraum, bzw. eine Stichprobe von m n -dimensionalen Vektoren aus einem n -dimensionalen Vektorraum. Für $m > n$ kann nur eine Teilbasis mit maximal n m -dimensionalen Basisvektoren gewählt werden. Solche Teilbasen sollen mit \mathcal{B}_m^n bezeichnet werden: der untere Index m gibt die Dimensionalität der Vektoren an, der obere Index n mit $n < m$ die Anzahl der linear unabhängigen Vektoren, die in der Teilbasis enthalten sein müssen. Natürlich gibt es viele solche Teilbasen, aber man muß aus der Menge dieser Teilbasen eine Basis \mathcal{B}_m^n auswählen, deren lineare Hülle die Datenvektoren enthält: $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathcal{L}(\mathcal{B}_m^n)$. Dass nicht jede Teilbasis \mathcal{B}_m^n die Datenvektoren erklären kann macht man sich anschaulich klar, wenn man an die Menge der 2-dimensionalen Teilräume – also die Menge der Ebenen – in einem 3-dimensionalen Raum denkt: wenn zwei Ebenen nicht parallel sind, schneiden sich und die Schnittfläche ist eine Gerade, also ein 1-dimensionaler Teilraum. Zwei Basen, von denen die eine die Vektoren aus der Ebene E_1 generiert und die zweite die aus der Ebene E_2 können 3-dimensionale Datenvektoren nur dann als Linearkombinationen erzeugen, wenn die Datenvektoren eben in diesem 1-dimensionalen Teilraum liegen.

Man wird also nur solche Basen wählen können, die in der linearen Hülle $\mathcal{L}_x = \mathcal{L}(\mathbf{b}_1, \dots, \mathbf{b}_n)$ liegen. Sind \mathcal{B}_1 und \mathcal{B}_2 zwei Basen aus \mathcal{L}_x , so wird man sie wieder ineinander überführen können und es gelten die Beziehungen (2.145) und (2.146). Tatsächlich ist es so, dass man eine Basis für die Datenvektoren aus eben diesen Vektoren errechnen muß, – im allgemeinen hat man ja keine Information über die Basis außer der, die in den Daten steckt. Das bedeutet, dass man die Basisvektoren als Linearkombinationen der Datenvektoren berechnen muß, wobei zu berücksichtigen sein wird, dass die Anzahl r der benötigten Basisvektoren kleiner als n sein kann.

Analoge Betrachtungen gelten für die m n -dimensionalen Zeilenvektoren der Matrix X . Die Vektoren in einer Basis sind linear unabhängig, aber nicht notwendig orthogonal. Im Prinzip kann man irgendeine Basis zur "Erklärung" eines Datensatzes wählen, allerdings hat die Wahl einer orthogonalen Basis zumindest den Vorteil, dass man annehmen kann, dass die zu den Vektoren der Basis korrespondierenden Merkmale unkorreliert sind und damit unabhängig voneinander interpretiert werden können. Deswegen wird zunächst nur die Bestimmung einer orthogonalen Basis besprochen, von der man dann, wenn es gewünscht wird, zu einer nicht-orthogonalen Basis übergehen kann.

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 2.5.2 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

$$(A - \lambda I)\mathbf{u} = \vec{0} \tag{2.148}$$

geschrieben werden. Diese Gleichung beschreibt ein homogenes Gleichungssystem: In ausgeschriebener Form hat man

$$(a_{11} - \lambda)u_1 + a_{12}u_2 + \cdots + a_{1n}u_n = 0 \tag{2.149}$$

$$a_{21}u_1 + (a_{22} - \lambda)u_2 + \cdots + a_{2n}u_n = 0 \tag{2.150}$$

$$\vdots \tag{2.151}$$

$$a_{n1}u_1 + a_{n2}u_2 + \cdots + (a_{nn} - \lambda)u_n = 0 \tag{2.152}$$

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

$$|A - \lambda I| = \begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ & & \ddots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{vmatrix} = 0 \quad (2.153)$$

Entwickelt man die Determinante, so ergibt sich ein Polynom $P(\lambda)$ in λ vom Grad n :

$$|A - \lambda I| = (-1)^n [\lambda^n - \beta_1 \lambda^{n-1} + \beta_2 \lambda^{n-2} + \cdots + (-1)^{n-1} \beta_{n-1} \lambda + (-1)^n \beta_n] = 0. \quad (2.154)$$

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 . Wie aus dem Fundamentalsatz der Algebra bekannt ist, existieren genau n Nullstellen $\lambda_1, \dots, \lambda_n$, 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 2.25 *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

$$(A - \bar{\lambda} I)\bar{\mathbf{u}} = 0, \quad (2.155)$$

und dies heißt, dass $\bar{\lambda}$ ebenfalls ein Eigenwert von A ist. \square

Links- und Rechtseigenvektoren: Es sei A eine nicht notwendig symmetrische $(n \times n)$ -Matrix, und für einen n -dimensionalen Vektor \mathbf{u} gelte die Beziehung

$$A\mathbf{u} = \lambda\mathbf{u}. \quad (2.156)$$

Dann ist \mathbf{u} ein Eigenvektor von A , und λ ist der zugehörige Eigenwert.

Es sei $B = A'$; gilt

$$B\mathbf{v} = \mu\mathbf{v}, \quad (2.157)$$

so ist \mathbf{v} ein Eigenvektor von B und μ der zugehörige 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 (2.156) heißt dementsprechend *Rechtseigenvektor*. Wegen (2.157) übertragen sich alle Aussagen über Rechtseigenvektoren auf Linkseigenvektoren, was allerdings 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 (2.157) $B\mathbf{v} = A\mathbf{v} = \mu\mathbf{v}$, d.h. ein gegebener Linkseigenvektor entspricht einem Rechtseigenvektor. Im Falle $A' \neq A$ gilt der

Satz 2.26 *Es sei A eine quadratische, nicht-symmetrische Matrix. Es gelte einerseits $\mathbf{v}'A = \mu\mathbf{v}'$, andererseits $A\mathbf{u} = \lambda\mathbf{u}$ mit $\lambda \neq \mu$. Dann folgt $\mathbf{u}'\mathbf{v} = 0$, 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. \square

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 2.27 *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}_2 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

$$a_1\mathbf{u}_1 + a_2\mathbf{u}_2 = 0 \tag{2.158}$$

Multiplikation von links mit A führt dann auf $a_1A\mathbf{u}_1 + a_2A\mathbf{u}_2 = 0$, d.h. auf

$$a_1\mu\mathbf{u}_1 + a_2\lambda\mathbf{u}_2 = 0. \tag{2.159}$$

Multipliziert man (2.158) mit λ und subtrahiert (2.158) dann von (2.159), so erhält man

$$a_1(\lambda - \mu)\mathbf{u}_1 + a_2(\lambda - \lambda)\mathbf{u}_2 = 0,$$

woraus wegen $\lambda - \mu \neq 0$ sofort $a_1 = 0$ folgt. Auf analoge Weise folgt $a_2 = 0$, 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

$$a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + \cdots + a_r \mathbf{v}_r + a_{p+1} \mathbf{v}_{p+1} = 0 \quad (2.160)$$

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 (2.160) mit A führt dann unter Berücksichtigung dieser Beziehung auf die Gleichung

$$a_1 \lambda_1 \mathbf{u}_1 + a_2 \lambda_2 \mathbf{u}_2 + \cdots + a_r \lambda_r \mathbf{u}_r + a_{p+1} \lambda_{p+1} \mathbf{u}_{p+1} = 0. \quad (2.161)$$

Multipliziert man (2.160) mit λ_{p+1} und subtrahiert die Gleichung dann von (2.161), so erhält man

$$a_1(\lambda_1 - \lambda_{p+1}) \mathbf{u}_1 + \cdots + a_p(\lambda_p - \lambda_{p+1}) \mathbf{u}_p = 0,$$

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 (2.161) $a_{p+1} \lambda_{p+1} \mathbf{u}_{p+1} = 0$. Daraus folgt wegen $\lambda_{p+1} \neq 0$ dann $a_{p+1} = 0$, so dass die $\mathbf{u}_1, \dots, \mathbf{u}_{p+1}$ ebenfalls linear unabhängig sind. \square

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}

$$A = V\Lambda V^{-1}. \quad (2.162)$$

Durch Multiplikation von rechts mit V und von links mit V^{-1} erhält man hieraus

$$V^{-1}AV = \Lambda. \quad (2.163)$$

Mit diesen beiden Gleichungen hat man einen Spezialfall einer Beziehung zwischen Matrizen, die durch die folgende Definition charakterisiert wird:

Definition 2.15 *Es seien A und B zwei $(n \times n)$ -Matrizen und es existiere eine nichtsinguläre Matrix C derart, dass*

$$B = C^{-1}AC \quad (2.164)$$

gilt. Dann heißen A und B ähnlich.

Offenbar bedeuten (2.162) und (2.163), dass A und Λ ähnlich sind. Da Λ eine Diagonalmatrix ist, heißt A auch *diagonalisierbar*. Damit eine $(n \times n)$ -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

$$A(\phi) = \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix} \quad (2.165)$$

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 (2.156) zurück: so dass sich die Eigenwerte von A als Nullstellen des Polynoms ergeben. Speziell für die Matrix (2.165) erhält man

$$|A - \lambda I| = 4\lambda^2 - 4\lambda \cos \phi + 1 = 0; \quad (2.166)$$

auf die Herleitung des Polynoms wird hier verzichtet, da es nur auf die Implikationen von (2.166) ankommt. Man findet die Nullstellen

$$\lambda_1 = \frac{1}{2}(\cos \phi + i \sin \phi), \quad \lambda_2 = \frac{1}{2}(\cos \phi - i \sin \phi), \quad i = \sqrt{-1} \quad (2.167)$$

Die in (2.165) definierte Matrix A hat also zwei komplexe Eigenwerte, reelle 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° stattfindet.

Es ist also möglich, dass für eine beliebig gewählte quadratische Matrix keine reellen Eigenwerte existieren, dass man aber komplexwertige Eigenwerte finden kann, die als Paare konjugiert komplexer Zahlen auftreten²⁷. Nun hätte man noch gerne die zugehörigen Eigenvektoren bestimmt. Für A findet man zwei:

$$\mathbf{u}_1 = \begin{pmatrix} -i \\ 1 \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} i \\ 1 \end{pmatrix}. \quad (2.168)$$

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 \bar{A} enthalten 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 Aussage $a_{ij} = -a_{ji}$, so heißt A *schief-symmetrisch*.

Ein wichtiger Fall ist durch die Gleichung

$$A = \bar{A}' \quad (2.169)$$

definiert; in diesem Fall heißt A *hermitesch*²⁸. Ist $A = \bar{A}$, so sind die Elemente von A alle reell und (2.169) bedeutet einfach, dass A symmetrisch ist. Das 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.

²⁷Zwei 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.

²⁸Nach dem französischen Mathematiker Charles Hermite (1822 – 1901)

2.7.2 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 Nachhermessungen systematisch voneinander unterscheiden. Die Berechnung der Kanonischen Korrelationen kann hier zu Antworten führen. An dieser Stelle soll auf die rein formalen Aspekte derartiger Methoden eingegangen werden.

Definition 2.16 *Es seien A und B symmetrische, positiv semidefinite Matrizen. Dann repräsentiert*

$$A\mathbf{w} = \lambda B\mathbf{w} \quad (2.170)$$

das generalisierte Eigenvektorproblem.

Setzt man $B = I$, I die entsprechende Einheitsmatrix, so reduziert sich das generalisierte Eigenvektorproblem auf das bekannte einfache Eigenvektorproblem.

Generalisierter Rayleigh-Quotient: Der *generalisierte Rayleigh-Quotient* ist durch

$$\rho(\mathbf{w}) = \frac{\mathbf{w}'A\mathbf{w}}{\mathbf{w}'B\mathbf{w}} \quad (2.171)$$

definiert.

Es ist klar, dass es bei diesem Ausdruck nicht auf die Länge der \mathbf{w} ankommt, denn setzt man bei (2.171) $\mu\mathbf{w}$, $\mu \in \mathbb{R}$ für \mathbf{w} ein, so kürzt sich μ sofort heraus. Deswegen kann man für die Länge von \mathbf{w} fordern, dass sie der Bedingung $\mathbf{w}'B\mathbf{w} = 1$ genügt. Man kann dann (2.171) in

$$\rho(\mathbf{w}) = \mathbf{w}'A\mathbf{w}, \quad \mathbf{w}'B\mathbf{w} = 1 \quad (2.172)$$

umformulieren.

Die Matrizen A und B sind als symmetrisch vorausgesetzt worden, – aber $B^{-1}A$ ist deswegen nicht notwendig ebenfalls symmetrisch. Dies bedeutet, dass die Folgerungen für den symmetrischen Fall $C\mathbf{w} = \lambda\mathbf{w}$, C eine symmetrische Matrix, nicht mehr zutreffen müssen. So kann man z.B. nicht mehr folgern, dass die verschiedenen Eigenvektoren \mathbf{w}_j , die der Gleichung (2.173) genügen, notwendig orthogonal zueinander sein müssen.

Rückführung auf den symmetrischen Fall: B ist als symmetrisch und positiv semidefinit vorausgesetzt worden. Dann kann man die Wurzel $B^{1/2} = P\Lambda^{1/2}P'$ von B bestimmen, – offenbar ist

$$B^{1/2}B^{1/2} = B = P\Lambda^{1/2}P'\Lambda^{1/2}P' = P\Lambda P',$$

denn $P'P = I$ die Einheitsmatrix. Die Gleichung (2.170) führt durch Multiplikation von links mit B^{-1} auf

$$B^{-1}A\mathbf{w} = \lambda\mathbf{w}. \quad (2.173)$$

Multipliziert man diese Gleichung von links mit $B^{1/2}$, so erhält man

$$B^{1/2}B^{-1}A\mathbf{w} = B^{-1/2}A\mathbf{w} = \lambda B^{1/2}\mathbf{w}.$$

Es sei \mathbf{v} die Transformation von \mathbf{w} mit $B^{-1/2}$, d.h. es sei $\mathbf{w} = B^{-1/2}\mathbf{v}$. Dann erhält man

$$B^{-1/2}AB^{-1/2}\mathbf{v} = \lambda\mathbf{v}. \quad (2.174)$$

Die Matrix $B^{-1/2}AB^{-1/2}$ ist aber symmetrisch: $(B^{-1/2}AB^{-1/2})' = B^{-1/2}AB^{-1/2}$, wegen $A' = A$ und $(B^{1/2})' = B^{1/2}$, und damit hat man mit (2.174) ein Eigenwert- und Eigenvektorproblem der bekannten Art für symmetrische Matrizen. Die Eigenvektoren \mathbf{v}_j symmetrischer Matrizen sind bekanntlich orthonormal. Weiter ist

$$\mathbf{v}'A\mathbf{v} = \mathbf{w}'B^{1/2}B^{-1/2}AB^{-1/2}B^{1/2}\mathbf{w} = (B^{1/2}\mathbf{w})'B^{-1/2}(B^{1/2}\mathbf{w})$$

und

$$\mathbf{w}'B\mathbf{w} = \mathbf{w}'B^{1/2}B^{-1/2}BB^{-1/2}B^{1/2}\mathbf{w} = (B^{1/2}\mathbf{w})'(B^{1/2}\mathbf{w}) = \|B^{1/2}\mathbf{w}\|^2,$$

so dass der generalisierte Rayleigh-Quotient die Form

$$\rho = \frac{\mathbf{w}'A\mathbf{w}}{\mathbf{w}'B\mathbf{w}} = \frac{(B^{1/2}\mathbf{w})'B^{-1/2}AB^{-1/2}(B^{1/2}\mathbf{w})}{\|B^{1/2}\mathbf{w}\|^2} \quad (2.175)$$

annimmt. Damit ist der generalisierte Rayleigh-Quotient auf den üblichen Rayleigh-Quotienten für symmetrische Matrizen zurückgeführt. In anderer Formulierung kann man sagen, dass die Lösung für den generalisierten 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 2.28 *Ein beliebiger Vektor \mathbf{v} kann als Linearkombination der \mathbf{w}_j , $j = 1, \dots, k$ angeschrieben werden. Für die Eigenvektoren des generalisierten Eigenvektorproblems $A\mathbf{w} = \lambda B\mathbf{w}$ gelten die Relationen*

$$\mathbf{w}'_i B \mathbf{w}_j = \delta_{ij}, \quad \mathbf{w}'_i A \mathbf{w}_j = \delta_{ij} \lambda_j, \quad \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (2.176)$$

Beweis: Es war $\mathbf{v} = B^{1/2}\mathbf{w}$, und als Lösungen von (2.174) 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. (2.173)) 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 (2.170) gilt ja $A\mathbf{w} = \lambda B\mathbf{w}$ und deshalb $(1/\lambda_i)A\mathbf{w} = B\mathbf{w}$. Damit gilt (2.176) 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 (2.176) für den Fall $i = j$. \square

Die Maximierung von (2.176) (Maximierung unter Nebenbedingungen, S. Anhang) führt auf die Gleichung (2.173).

Satz 2.29 *Für den generalisierten Rayleigh-Quotienten gilt*

$$\rho_1 \leq \rho \leq \rho_2, \quad (2.177)$$

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 68).

Satz 2.30 *Gilt $A\mathbf{v} = \lambda B\mathbf{v}$ und sind λ und \mathbf{v} die Eigenwerte und Eigenvektoren für den generalisierten Rayleigh-Quotienten, so kann A gemäß*

$$A = \sum_{j=1}^r \lambda_j B \mathbf{v}_j (B \mathbf{v}_j)' \quad (2.178)$$

zerlegt werden.

Beweis: Für eine beliebige symmetrische Matrix C mit der Matrix P der Eigenvektoren und der Diagonalmatrix $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ der Eigenwerte gilt bekanntlich $C = \sum_j \lambda_j \mathbf{p}_j \mathbf{p}'_j$. Die Matrix $B^{-1/2} A B^{-1/2}$ ist symmetrisch, mithin gilt

$$B^{-1/2} A B^{-1/2} = \sum_{j=1}^r \lambda_j \mathbf{v}_j \mathbf{v}'_j.$$

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. \square

2.7.3 Mehrfache Eigenwerte

Es sei A eine $(n \times n)$ -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 $(n \times n)$ -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 2.17 *Es sei V ein Vektorraum und es sei $V_\lambda = \{\mathbf{v} \in V | A\mathbf{v} = \lambda\mathbf{v}\}$. 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_λ ist der Kern von $(A - \lambda I)$. \square

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

$$A(a\mathbf{v} + b\mathbf{w}) = aA\mathbf{v} + bA\mathbf{w} = a\lambda\mathbf{v} + b\lambda\mathbf{w} = \lambda\mathbf{u}.$$

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 2.18 *Die Matrix A definiere eine Abbildung f des Vektorraums V in sich selbst, d.h. $f : V \rightarrow V$, 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 $H(A, \lambda)$*

$$H(A, \lambda) = \{\mathbf{v} \in V | (A - \lambda I)^r(\mathbf{v}) = 0\}. \quad (2.179)$$

Die Elemente von $H(A, \lambda)$ heißen die Hauptvektoren. $v \in V$ ist Hauptvektor der Stufe p , wenn $(A - \lambda I)^p v \neq \vec{0}$.

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 2.19 Die Matrix A definiere eine Abbildung eines Vektorraums in sich selbst: $f: V \rightarrow V$, und es sei $U \subseteq V$. Gilt $f(U) \subseteq U$, d.h. ist die Menge der Vektoren Au wieder eine Teilmenge von U , so heißt U invarianter Teilraum von V , oder einfach f -invariant²⁹.

Anmerkung: Alle Eigenräume sowie alle Haupträume sind invariante Teilräume. □

2.8 Lineare Gleichungssysteme

Es sei A eine $(m \times n)$ -Matrix, \mathbf{x} ein n -dimensionaler Vektor, und \mathbf{y} ein m -dimensionaler Vektor, und es gelte

$$A\mathbf{x} = \mathbf{y}, \quad \mathbf{x} \in \mathbb{R}^n, \quad \mathbf{y} \in \mathbb{R}^m \quad (2.180)$$

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 $A = [\mathbf{a}_1, \dots, \mathbf{a}_n]$, $\mathbf{x} = (x_1, \dots, x_n)'$, so hat man

$$A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \dots + x_n\mathbf{a}_n = \mathbf{y} \quad (2.181)$$

Die Komponenten x_1, \dots, 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 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*.

²⁹oder invariant unter f

Definition 2.20 *Es sei A eine $(m \times n)$ -Matrix mit dem Rang $r = \text{rg}(A) \leq \min(m, n)$. und $A\mathbf{x} = \mathbf{y}$ sei ein System von Gleichungen. Weiter sei*

$$\text{kern}(A) = \{\mathbf{x} \in \mathbb{R}^n | A\mathbf{x} = \vec{0}\} \quad (2.182)$$

$$\mathcal{L}(A) = \{\mathbf{y} \in \mathbb{R}^m | A\mathbf{x} = \mathbf{y}\}. \quad (2.183)$$

$\text{kern}(A)$ heißt Kern von A , und $\mathcal{L}(A)$ ist die lineare Hülle der Spaltenvektoren von A .

Anmerkungen:

1. $\text{kern}(A)$ ist ein (Teil-)Vektorraum: sind die Vektoren \mathbf{x}_1 und \mathbf{x}_2 Elemente aus $\text{kern}(A)$, so rechnet man leicht nach, dass dann auch $b_1\mathbf{x}_1 + b_2\mathbf{x}_2 \in \text{kern}(A)$ ($b_1, b_2 \in \mathbb{R}$) gilt³⁰.
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. \square

Der folgende Satz gilt für beliebige $(m \times n)$ -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 2.31 *Es sei A eine (m, n) -Matrix mit der SVD $A = Q\Sigma T'$, wobei Q aus den Spaltenvektoren $\mathbf{q}_1, \dots, \mathbf{q}_n$ und T aus den Spaltenvektoren $\mathbf{t}_1, \dots, \mathbf{t}_n$ bestehe; ist $\text{rg}(A) = r \leq \min(m, n)$, so sind r Singularwerte σ_k größer als Null und die restlichen sind gleich Null. Dann gilt*

$$\mathcal{L}(A) = \mathcal{L}(\mathbf{q}_1, \dots, \mathbf{q}_r) \quad (2.184)$$

$$\text{kern}(A) = \mathcal{L}(\mathbf{t}_{r+1}, \dots, \mathbf{t}_s), \quad s = \min(m, n) \quad (2.185)$$

$$\text{rg}(\text{kern}(A) + \text{rg}(\mathcal{L}(A))) = \min(m, n) \quad (2.186)$$

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 (2.184) gelten muß.

³⁰In Definition 4.1, Punkte 4., Seite 133 wird der Begriff des Kerns einer Abbildung eingeführt. Die Matrix A definiert eine Abbildung, und (2.182) definiert damit den Kern einer Abbildung.

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. (2.92), Seite 73), 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 $\mathbf{t}_1, \dots, \mathbf{t}_r$ aufgespannten Vektorraums sein, sondern muß ein Element des $(n - r)$ -dimensionalen Komplementärums sein. Die $\mathbf{t}_{r+1}, \dots, \mathbf{t}_n$ sind eine Basis für diesen Komplementärums, 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 (2.92), Seite 73, 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 $\mathbf{t}_{r+1}, \dots, \mathbf{t}_n$, und dies ist die Aussage von (2.185).

Die Gleichung (2.186) ist eine unmittelbare Folge der vorangegangenen Argumente: $\mathcal{L}(A)$ hat den Rang r und $\text{kern}(A)$ hat den Rang $n - r$, so dass die Summe der Ränge gleich n sein muß. \square

Anmerkung: Der Satz 2.31 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 $\text{kern}(A)$ 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 4.4. \square

Die allgemeine Lösungsmenge wird im folgenden Satz spezifiziert:

Satz 2.32 *Es sei $A\mathbf{x} = \mathbf{y} \in \mathcal{L}(A)$, $\text{rg}(A) = r$, und insbesondere sei $\mathbf{x} = \mathbf{x}_0$ eine bestimmte Lösung, so dass $A\mathbf{x}_0 = \mathbf{y}$ gilt. Der Kern $\text{kern}(A)$ besteht aus dem $(n - r)$ -dimensionalen Teilraum $\mathcal{L}_{n-r} = \mathcal{L}(\mathbf{t}_{r+1}, \dots, \mathbf{t}_n)$ des V_n . Dann ist die Menge der Lösungsvektoren durch*

$$\mathcal{L} = \{\mathbf{x}_0 + \mathbf{x} \mid \mathbf{x} \in \mathcal{L}_{n-r}\} \quad (2.187)$$

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. \square

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

$$\mathbf{x} = A^{-1}\mathbf{y}, \quad \mathbf{y} \in \mathcal{L}(A). \quad (2.188)$$

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 $(m \times n)$ -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. (2.33), Seite 47) existiert die zu $A'A$ inverse Matrix $(A'A)^{-1}$, so dass

$$\mathbf{x} = (A'A)^{-1}A'\mathbf{y} \quad (2.189)$$

resultiert. Die Matrix $(A'A)^{-1}A'$ ist eine Pseudoinverse für A , s. Seite 82. Sind die Komponenten von \mathbf{y} Messwerte, so sind sie üblicherweise durch Messfehler kontaminiert, so dass $\mathbf{y} \notin \mathcal{L}(A)$. (2.189) liefert dann keine Lösung \mathbf{x} , die allen m Gleichungen genügt. (2.189) ist die bekannte Kleinste-Quadrate-Schätzung $\hat{\mathbf{x}}$ für \mathbf{x} , vgl. (4.18), Seite 126, und Abschnitt 2.5.8.

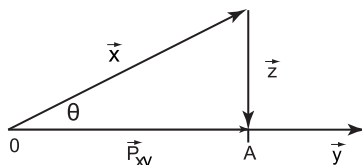
Für den Fall $r < n$ liefert (2.187) den Lösungsraum.

Cramersche Regel: Diese Regel wird hier nur der Vollständigkeit wegen genannt. Es sei $A\mathbf{x} = \mathbf{y}$ ein Gleichungssystem, wobei A eine $(n \times n)$ -Matrix sei. Dann gilt für die j -te Komponente x_j von \mathbf{x}

$$x_j = \frac{|A_j|}{|A|}, \quad j = 1, \dots, n \quad (2.190)$$

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 4.6, 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.

Abbildung 11: Orthogonale Projektion des Vektors \mathbf{x} auf einen Vektor \mathbf{y} bzw. auf eine Gerade



2.9 Projektionen

2.9.1 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

$$\vec{P}_{yx} = ay \quad (2.191)$$

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

$$\mathbf{z} = \vec{P}_{yx} - \mathbf{x} = ay - \mathbf{x}, \quad (2.192)$$

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

$$a = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{y}\|^2} \Rightarrow \vec{P}_{xy} = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{y}\|^2}\mathbf{y} \quad (2.193)$$

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

$$\|\vec{P}_{xy}\| = a\|\mathbf{y}\| \quad (2.194)$$

gegeben.

2.9.2 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 $(n \times n)$ Projektionsmatrix und I sei die $(n \times n)$ -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 2.33 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. \square

Anmerkung: P sei eine $(n \times n)$ -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 2.13, Seite 62), hat P den Rang $\text{rg}(P) = n - 1$. \square

Beispiel 2.4 Das lineare Modell Es werde das Allgemeine Lineare Modell (ALM)

$$\mathbf{y} = X\mathbf{b} + \mathbf{e} \tag{2.195}$$

betrachtet. Der Parametervektor \mathbf{b} wird mit der Methode der Kleinsten Quadrate geschätzt; man findet (s. Abschnitt 4.3.3 im Anhang); die Schätzung ist durch

$$\hat{\mathbf{b}} = (X'X)^{-1}X'\mathbf{y} \tag{2.196}$$

gegeben, und man hat³¹

$$\hat{\mathbf{y}} = X\hat{\mathbf{b}} = X(X'X)^{-1}X'\mathbf{y} = P\mathbf{y}. \tag{2.197}$$

³¹Wegen dieser Beziehung – $\hat{\mathbf{y}} = P\mathbf{y}$ – findet man gelegentlich auch den Ausdruck *Hut-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.

Man hat demnach

$$\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{e}} = P\mathbf{y} + \hat{\mathbf{e}}, \quad (2.198)$$

und da $\mathbf{y} - P\mathbf{y} = \hat{\mathbf{e}}$ folgt

$$(I - P)\mathbf{y} = \hat{\mathbf{e}}, \quad (2.199)$$

wobei $\hat{\mathbf{e}}$ der Fehlervektor ist, wenn \mathbf{b} durch die KQ-Schätzung $\hat{\mathbf{b}}$ ersetzt wird. Offenbar gilt

$$\hat{\mathbf{y}}'\hat{\mathbf{e}} = 0, \quad (2.200)$$

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 (2.197) eingeführte Matrix $P = X(X'X)^{-1}X'$ ist eine Projektionsmatrix, denn es gilt

$$(X(X'X)^{-1}X')' = X(X'X)^{-1}X', \quad (\text{Symmetrie}) \quad (2.201)$$

$$X(X'X)^{-1}X'X(X'X)^{-1}X' = X(X'X)^{-1}X', \quad (\text{Idempotenz}) \quad (2.202)$$

Nach (2.197) 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 (2.197) 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 i -ten Falls) von $\hat{\mathbf{y}}$ durch das Skalarprodukt des i -ten Zeilenvektors von P mit den Datenvektor \mathbf{y} gegeben:

$$\hat{y}_i = p_{i1}y_1 + p_{i2}y_2 + \cdots + p_{im}y_m. \quad (2.203)$$

Das Element p_{ij} läßt sich direkt interpretieren als das Ausmaß, den die Messung y_j via die KQ-Schätzungen der Regressionsparameter auf \hat{y}_i hat. Man spricht von der *Hebelwirkung* (leverage) der Messung y_i auf die Schätzungen \hat{y}_i (vergl. Abb. 6, S. 55). 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

$$p_{ii} = \sum_{j=1}^m p_{ij}^2 = p_{ii}^2 + \sum_{i \neq j} p_{ij}^2 \quad (2.204)$$

Man sieht leicht, dass diese Beziehung nicht gelten kann, wenn die $p_{ij} > 1$ sein können, d.h. es folgt, dass

$$0 \leq p_{ij} \leq 1. \quad (2.205)$$

Weiter gilt, dass wegen der Symmetrie von P die Summe der Eigenwerte gleich der Summe der Diagonalelemente p_{ii} ist, und diese Summe ist gleich der Summe der Eigenwerte. Da die Eigenwerte entweder den Wert 0 oder 1 haben, folgt

$$\sum_{i=1}^m p_{ii} = n \quad (2.206)$$

und n ist der Rang der $(m \times n)$ -Datenmatrix X , – diese muß vollen Rang haben, da sonst die Inverse $(X'X)^{-1}$ nicht existieren würde, und dann würde P nicht existieren. Aus (2.204) 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. \square

2.9.3 Projektionen auf Hauptachsen

Es sei X eine $(m \times n)$ -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

$$\ell_{ik} = \tilde{\mathbf{x}}_i' \mathbf{t}_k. \quad (2.207)$$

\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 (2.193) gilt nun

$$a_{ik} = \frac{\tilde{\mathbf{x}}_i' \mathbf{t}_k}{\|\mathbf{t}_k\|^2} = \tilde{\mathbf{x}}_i' \mathbf{t}_k, \quad (2.208)$$

da ja $\|\mathbf{t}_k\| = 1$, und nach (2.194) hat man dann

$$\|\vec{P}_{ik}\| = a_{ik} \|\mathbf{t}_k\| = \tilde{\mathbf{x}}_i' \mathbf{t}_k, \quad (2.209)$$

d.h. wegen (2.207) gilt

$$\ell_{ik} = \|\vec{P}_{ik}\|, \quad (2.210)$$

so dass die Koordinate ℓ_{ik} die Länge der Projektion des Vektors $\tilde{\mathbf{x}}_i$ auf die k -te Hauptachse des Ellipsoids ist.

2.9.4 Die Zentrierungsmatrix

X sei eine $(m, \times n)$ -Matrix von Messwerten; es gebe m Fälle und n Variablen. Um die Varianz-Kovarianzmatrix zu berechnen, muß X "zentriert" werden, d.h. von den Elementen einer Spalte von X muß der Mittelwert der Spalte subtrahiert werden, so dass die zentrierte Matrix $X_c = (x_{ij} - \bar{x}_j)$, $i = 1, \dots, m$, $j = 1, \dots, n$ entsteht, und $S = X_c' X_c / m$ ist die Varianz-Kovarianzmatrix.

Definition 2.22 Es sei $\vec{1} = (1, 1, \dots, 1)'$ 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

$$H_m = I - \frac{1}{m} \vec{1}\vec{1}'. \quad (2.211)$$

Zentrierungsmatrix.

H_M hat die Form

$$H_m = \begin{pmatrix} 1 - \frac{1}{m}, & -\frac{1}{m}, & -\frac{1}{m}, & \dots, & -\frac{1}{m} \\ -\frac{1}{m}, & 1 - \frac{1}{m}, & -\frac{1}{m}, & \dots, & -\frac{1}{m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{m}, & -\frac{1}{m}, & -\frac{1}{m}, & \dots, & 1 - \frac{1}{m} \end{pmatrix} \quad (2.212)$$

Ist X eine nichtzentrierte $(m \times n)$ -Matrix und ist X_{sc} die korrespondierende spaltenzentrierte Matrix, X_{zc} die korrespondierende zeilenzentrierte Matrix so gilt

$$X_{sc} = H_m X, \quad X_{zc} = X H_n. \quad (2.213)$$

Es gilt der

Satz 2.34 Die Zentrierungsmatrix hat die folgenden Eigenschaften:

1. H_m ist idempotent (eine Matrix M ist idempotent, wenn $MM = M$),
2. H_m ist symmetrisch und positiv-semidefinit, d.h. H_m ist eine Projektionsmatrix.
3. H_m hat einen Eigenwert $\lambda = 0$ und den Eigenwert 1 mit der Multiplizität $m - 1$,
4. H_m ist singulär.
5. H_m hat den Kern der Dimension 1 längs $\vec{1}$,

Beweis: Es gilt

$$\begin{aligned} H_m H_m &= \left(I - \frac{1}{m} \vec{1} \vec{1}'\right) \left(I - \frac{1}{m} \vec{1} \vec{1}'\right) \\ &= I - \frac{1}{m} I \vec{1} \vec{1}' - \frac{1}{m} I \vec{1} \vec{1}' + \frac{1}{m^2} \vec{1} \vec{1}' \vec{1} \vec{1}' \\ &= I - \frac{1}{m} I \vec{1} \vec{1}' = H_m, \end{aligned}$$

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

$$H_m' = \left(I - \frac{1}{m} \vec{1} \vec{1}'\right)' = I' - \left(\frac{1}{m} \vec{1} \vec{1}'\right)' = I - \frac{1}{m} \vec{1} \vec{1}' = H_m, \quad (2.214)$$

d.h. H_m ist symmetrisch. Damit ist H_m gemäß Definition 2.21 eine Projektionsmatrix. Nach Satz 2.33 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 $\text{rg}(H_m) = m - 1$, 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

$$-\frac{x_1}{m} - \dots - \frac{x_{j-1}}{m} + x_j \left(1 - \frac{1}{m}\right) - \frac{x_{j+1}}{m} - \dots - \frac{x_m}{m} = x_j - \frac{1}{m} \sum_{i=1}^m x_i = x_j - \bar{x}$$

Dann ist

$$\mathbf{x}' H_m \mathbf{x} = x_1^2 - x_1 \bar{x} + x_2^2 - x_2 \bar{x} + \dots - x_m^2 - x_m \bar{x} = \sum_{i=1}^m x_i^2 - m \bar{x}^2 = k,$$

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

$$\begin{pmatrix} 1 - \frac{1}{m}, & -\frac{1}{m}, & -\frac{1}{m}, & \dots, & -\frac{1}{m} \\ -\frac{1}{m}, & 1 - \frac{1}{m}, & -\frac{1}{m}, & \dots, & -\frac{1}{m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{m}, & -\frac{1}{m}, & -\frac{1}{m}, & \dots, & 1 - \frac{1}{m} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

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, $\mathbf{x} = (x, x, \dots, x)'$. Da sich die Orientierung eines Vektors nicht ändert, wenn er mit einem Skalar multipliziert wird, ist

$$\frac{1}{x}\mathbf{x} = (1, 1, \dots, 1)',$$

d.h. die Orientierung der \mathbf{x} aus dem Kern von H_m ist identisch mit der von $\vec{1} = (1, 1, \dots, 1)'$. Der Kern ist damit ein 1-dimensionaler Teilraum des \mathbb{R}^m . \square

Anmerkung: Es sei λ ein Eigenwert von H_m und \mathbf{p} sei der zugehörige Eigenvektor, so dass $H_m\mathbf{p} = \lambda\mathbf{p}$. Ist $\lambda = 1$, so folgt für den zugehörigen Eigenvektor $\mathbf{p} = (p_1, \dots, p_m)'$, dass $\sum_i p_i = 0$.

Denn ausgeschrieben bedeutet $H_m\mathbf{p} = \lambda\mathbf{p}$

$$\begin{pmatrix} 1 - \frac{1}{m} & -\frac{1}{m} & -\frac{1}{m} & \dots & -\frac{1}{m} \\ -\frac{1}{m} & 1 - \frac{1}{m} & -\frac{1}{m} & \dots & -\frac{1}{m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{m} & -\frac{1}{m} & -\frac{1}{m} & \dots & 1 - \frac{1}{m} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_m \end{pmatrix} = \lambda \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_m \end{pmatrix}$$

Für die i -te Komponente von $\lambda\mathbf{p}$ hat man dann

$$p_i - \frac{1}{m} \sum_{j=1}^m p_j = \lambda p_i, \quad i = 1, \dots, m$$

d.h. $p_i(1 - \lambda) = \bar{p}$, also $p_i = p_0$ eine Konstante. Für die Lösung $\lambda = 1$ gilt

$$-\frac{1}{m} \sum_{i=1}^m p_i + p_j = \bar{p} + p_j = p_j \Rightarrow \bar{p} = 0.$$

\square

Nun sei X nicht zentriert (X ist die Matrix der "Rohwerte", kurz auch "Rohmatrix" genannt). Die (spalten-)zentrierte Matrix X_c ist dann durch

$$X_c = H_m X \tag{2.215}$$

gegeben, und die Varianz-Kovarianzmatrix S läßt sich in der Form

$$S = \frac{1}{m} X_c' X_c = \frac{1}{m} X' H_m' H_m X = X' H_m X \tag{2.216}$$

schreiben, denn es gilt ja $H_m' = H_m$ und $H_m H_m = H_m$. Die übliche Biaskorrektur für die Varianz-Kovarianzschätzungen erhält man, indem man diese Gleichung mit $m/(m-1)$ multipliziert:

$$\hat{S} = \frac{m}{m-1} S = \frac{1}{m-1} X' H X. \tag{2.217}$$

Betrachtet man die spalten- oder zeilenzentrierte Matrix X_{sc} bzw. X_{zc} , so fokussiert man auf die Zusammenhänge zwischen den Variablen oder Fällen, weil

$X'_{sc}X_{sc}/m$ bzw. $X_{zc}X'_{zs}/n$ die Kovarianzmatrizen für die Variablen bzw. die Fälle sind. Es werden aber auch nichtzentrierte Datenmatrizen X betrachtet; $X'X$ und XX' sind dann keine Kovarianzmatrizen. Der Vergleich der Analysen von zentrierten und nichtzentrierten Matrizen ist nicht ganz einfach und kann hier nicht diskutiert werden. Relevante Arbeiten zu diesem Thema sind Cadima & Jolliffe (2009) und Honeine (2014).

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,

$$\langle f, g \rangle = \int_D f(t)g(t)dt, \quad f, g \in \mathcal{F} \quad (3.1)$$

(d.h. existiert das Integral), so hat man über

$$\|f\|^2 = \int_D |f(t)|^2 dt < \infty \quad (3.2)$$

auch eine Norm $\|f\|$ erklärt und \mathcal{F} ist ein Vektorraum. Es sei $\{\phi_1, \phi_2, \dots\}$ eine Menge von Funktionen aus \mathcal{F} derart, dass

$$f(t) = \sum_{i=1}^{\infty} a_i \phi_i(t) \quad (3.3)$$

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 $\phi_i, i = 1, 2, 3, \dots$ orthogonal, wenn

$$\langle \phi_i, \phi_j \rangle = \int_D \phi_i(t)\phi_j(t)dt = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (3.4)$$

erfüllt ist; δ_{ij} heißt *Kronecker-Delta*.

Beispiel 3.1 Fourier-Reihen³² Eine bekannte Reihenentwicklung einer Funktion $f(t)$ ist die Fourier-Reihe

$$f(t) = a_0 + \sum_{n=1}^{\infty} (a_n \cos(nt) + b_n \sin(nt)), \quad (3.5)$$

wobei die Koeffizienten a_n und b_n durch

$$a_0 = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) dt \quad (3.6)$$

$$a_n = \frac{1}{\pi} \int_{-\infty}^{\infty} f(t) \cos(nt) dt \quad (3.7)$$

$$b_n = \frac{1}{\pi} \int_{-\infty}^{\infty} f(t) \sin(nt) dt \quad (3.8)$$

gegeben sind. Die cos- und sin-Funktionen sind orthogonal:

$$\int_{-T/2}^{T/2} \cos(m\omega_0 t) \cos(n\omega_0 t) dt = \begin{cases} 0, & m \neq n \\ T/2, & m = n \neq 0 \end{cases} \quad (3.9)$$

$$\int_{-T/2}^{T/2} \sin(m\omega_0 t) \sin(n\omega_0 t) dt = \begin{cases} 0, & m \neq n \\ T/2, & m = n \neq 0 \end{cases} \quad (3.10)$$

$$\int_{-T/2}^{T/2} \sin(m\omega_0 t) \cos(n\omega_0 t) dt = 0, \quad \text{für alle } m \text{ und } n \quad (3.11)$$

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). \square

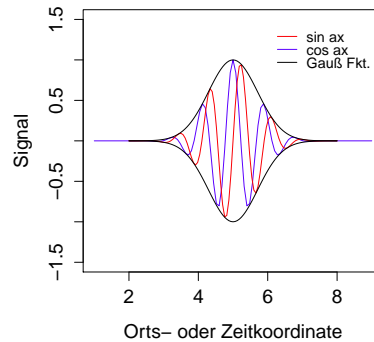
cos- und sin-Funktionen haben einen Nachteil: sie sind auf $-\infty, \infty$ definiert. Es ist aber oft notwendig, örtlich und zeitlich begrenzte Signale zu repräsentieren. Für diesen Zweck haben sich andere Basisfunktionen als geeigneter erwiesen. So können Gauß-Funktionen $\exp(-(t - t_n)^2)$, $n = 1, 2, \dots$ verwendet werden (Gishasby & O'Neill (1994), Calcaterra (2008)), oder Gabor-Funktionen:

$$f_n(t) = \exp\left(-\frac{(x - x_n)^2}{2\sigma^2}\right) \phi(k\omega_0 t), \quad (3.12)$$

wobei ϕ eine Sinus- oder Kosinus-Funktion ist (Gabor (1948)), vergl. Abbildung 12. Weitere Basisfunktionen sind Polynome (Hermite- oder Laguerre-Polynome);

³²Joseph Fourier, (1768 – 1830), französischer Mathematiker und Physiker

Abbildung 12: Gabor-Funktion



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-Loève-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 Ereignisse 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 repräsentieren.

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 (Pfad, 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

$$\mathbb{E}(X_t) = \int_{\Omega} X(t, \omega) dP(\omega) = m(t), \quad \text{Var}(X_t) = \mathbb{E}[(X_t - \mathbb{E}(X_t))^2] \quad (3.13)$$

definiert; für jeden Wert $t \in D$ wird über alle möglichen Trajektorien $X(t)$ gemittelt. $m(t)$ wird auch die Mittelwertfunktion des Prozesses genannt.

Es sei Y_t ein stochastischer Prozess mit der Mittelwertfunktion $\mathbb{E}(Y_t)$. Dann heißt

$$X_t = \{X_t\}_{t \in D}, \quad X_t = Y_t - m(t) \quad (3.14)$$

zentrierter stochastischer Prozess.

Definition 3.1 *Der stochastische Prozess heißt stetig im quadratischen Mittel, wenn*

$$\mathbb{E}[(X_{t+\varepsilon} - X_t)^2] = 0. \quad (3.15)$$

Es sei X_t ein zentrierter stochastischer Prozess; dann ist

$$R_X(s, t) = \mathbb{E}(X(s)X(t)), \quad s, t \in D \quad (3.16)$$

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*

$$\hat{X}(t) = \sum_{i=1}^n a_i \phi_i(t), \quad (3.17)$$

und die

$$a_i = \int_a^b X(t)\phi_i(t) \quad (3.18)$$

sind zufällige Veränderliche³³ mit

$$\mathbb{E}(a_i) = 0, \quad \mathbb{E}(a_i a_j) = \delta_{ij} \lambda_i \quad (3.19)$$

mit

$$\delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} \quad (\text{Kronecker-Delta})$$

sind; die ϕ_i sind die orthonormalen Eigenfunktionen von $C(s, t)$, mit den λ_i als dazu korrespondierenden Eigenwerten. Weiter gilt

$$\lim_{n \rightarrow \infty} \hat{X}(t) = X(t). \quad (3.20)$$

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

$$\mathbb{E}(a_i) = \mathbb{E} \left[\int_a^b X_t \phi_i(t) \right] = \int_a^b \mathbb{E}[X_t] \phi_i(t) dt = 0, \quad (3.21)$$

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 und ϕ_j sind orthonormal, $i \neq j$. Dann folgt

$$\begin{aligned} \mathbb{E}[a_i a_j] &= \mathbb{E} \left[\int_a^b \int_a^b X_s X_t \phi_i(s) \phi_j(t) ds dt \right] \\ &= \int_a^b \int_a^b \mathbb{E}[X_s X_t] \phi_i(s) \phi_j(t) ds dt \\ &= \int_a^b \int_a^b k(s, t) \phi_i(s) \phi_j(t) ds dt \\ &= \int_a^b \phi_i(s) \left(\int_a^b k(s, t) \phi_j(t) dt \right) ds \\ &= \lambda_i \int_a^b \phi_i(s) \phi_j(s) ds \\ &= \delta_{ij} \lambda_i, \end{aligned} \quad (3.22)$$

δ_{ij} das Kronecker-Delta. Es sei

$$\varepsilon_n(t) = \mathbb{E} \left[\left(X(t) - \sum_{i=1}^n a_i \phi_i(t) \right)^2 \right]. \quad (3.23)$$

³³ a_i ist zufällig weil $X(t)$ eine zufällige Funktion aus der Menge der zufälligen Funktionen ist, die den stochastischen Prozess X_t definieren.

Dann folgt

$$\varepsilon_n(t) = \mathbb{E}[X^2(t)] - 2\mathbb{E}\left[X(t) \sum_{i=1}^n a_i \phi_i(t)\right] + \mathbb{E}[a_i a_j \phi_i(t) \phi_j(t)] \quad (3.24)$$

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

$$\lim_{n \rightarrow \infty} \varepsilon_n(t) = 0$$

d.h. $\hat{X}(t) \rightarrow X(t)$. □

$X(t)$ **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 $i = 1, 2, \dots, k$ 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, \dots, t_N bestimmt wird. Man erhält dann den Vektor

$$X = (X_1, \dots, X_N)'$$

für den die $(N \times N)$ -Matrix C der Autokorrelationen berechnet werden kann. Man erhält man dann die Gleichung

$$C\vec{\phi}_j = \lambda_j\vec{\phi}_j, \quad j = 1, \dots, N \quad (3.25)$$

$\vec{\phi}_j$ 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, \dots, t_N , und natürlich sind die Eigenvektoren $\vec{\phi}_j$ paarweise orthonormal. Gesucht ist die Entwicklung für $X = (X(t_1), X(t_2), \dots, X(t_N))'$ als Reihe:

$$X = \sum_{j=1}^N a_j \phi_j(t),$$

mit $\phi_j = (\phi_j(t_1), \dots, \phi_j(t_N))'$. Die Koeffizienten a_j erhält man durch orthogonale Basisentwicklung:

$$\phi_j' X = a_j, \quad j = 1, \dots, N \quad (3.26)$$

Natürlich ist es von Interesse, weniger als N Eigenfunktionen zu wählen, – gesucht ist ja die sparsamste Repräsentation der Daten. Also betrachtet man

$$\hat{X} = \sum_{j=1}^r a_j \phi_j(t), \quad r < N \quad (3.27)$$

Die Güte der Abschätzung \hat{X} läßt sich durch den Quotienten

$$\frac{\sum_{j=1}^r \lambda_j}{\sum_{j=1}^N \lambda_j} \geq \alpha \quad (3.28)$$

bewerten, wobei α ein festgesetzter Anteil an erklärter Varianz ist.

4 Anhang

4.1 Steinerscher Austauschatz und das Fundamental-Lemma

Satz 4.1 *Es sei V ein Vektorraum mit der Basis $\mathcal{B} = (\mathbf{v}_1, \dots, \mathbf{v}_n)$. Für $\vec{0} \neq \mathbf{v} \in V$ gibt es dann die Darstellung*

$$\mathbf{v} = a_1 \mathbf{v}_1 + \dots + a_n \mathbf{v}_n \quad (4.1)$$

Jeder Vektor $\mathbf{v}_j \in \mathcal{B}$ mit $a_j \neq 0$ kann durch \mathbf{v} ausgetauscht werden, und

$$\{\mathbf{v}_1, \mathbf{v}_{j-1}, \mathbf{v}, \mathbf{v}_{j+1}, \dots, \mathbf{v}_n\} \quad (4.2)$$

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

$$b_1\mathbf{v} + b_2\mathbf{v}_2 + \dots + b_n\mathbf{v}_n = \vec{0}.$$

Da zu zeigen ist, dass $\{\mathbf{v}, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ eine Basis ist, muß gezeigt werden, dass die $\mathbf{v}, \mathbf{v}_2, \dots, \mathbf{v}_n$ linear unabhängig sind, d.h. die Darstellung des Vektors $\vec{0}$ nur möglich ist, wenn $b_1 = \dots = b_n = 0$. Setzt man den obigen Ausdruck der Linearkombination von \mathbf{v} ein und vereinfacht, so erhält man den Ausdruck

$$b_1a_1\mathbf{v}_1 + (b_1a_2 + b_2)\mathbf{v}_2 + \dots + (b_1a_n + b_n)\mathbf{v}_n = \vec{0}.$$

Wegen der linearen Unabhängigkeit der \mathbf{v}_j folgt dann aber

$$b_1a_1 = b_1a_2 + b_2 = \dots = b_1a_n + b_n = 0.$$

Da $a_1 \neq 0$ folgt $b_1 = 0$ und also $b_1 = \dots = b_n = 0$, d.h. die Vektoren $\{\mathbf{v}, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ sind linear unabhängig.

Jetzt muß nur noch gezeigt werden, dass $\{\mathbf{v}, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ ein Erzeugendensystem für V ist. Dazu sei $\mathbf{w} \in V$ ein beliebiger Vektor aus V . Da $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ nach Voraussetzung eine Basis ist, existieren Koeffizienten $c_1, \dots, c_n \in \mathbb{R}$ derart, dass

$$\mathbf{w} = c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n.$$

Aus (4.1) folgt

$$\mathbf{v}_1 = \frac{1}{a_1}(\mathbf{w} - a_2\mathbf{v}_2 - \dots - a_n\mathbf{v}_n)$$

Setzt man diesen Ausdruck für \mathbf{v}_1 in den Ausdruck für \mathbf{w} ein, so sieht man, dass \mathbf{w} als Linearkombination der $\{\mathbf{v}, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ dargestellt werden kann, so dass diese Vektoren ebenfalls eine Basis für V bilden. \square

Der folgende Satz wird gelegentlich als *Fundamental-Lemma der Linearen Algebra* bezeichnet (etwa in Koecher (1997), p. 21).

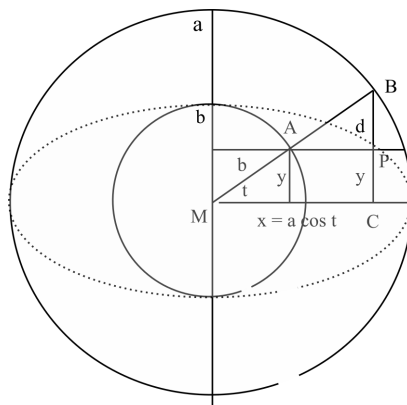
Satz 4.2 *Der Vektorraum V habe eine Basis $\mathcal{B} = (\mathbf{v}_1, \dots, \mathbf{v}_n)$. Dann ist jede Teilmenge $M \subset V$ mit $m > n$ Elementen linear abhängig. Zwei verschiedene Basen von V haben stets dieselbe Anzahl von Elementen aus V .*

Beweis: Es sei $\mathcal{B} = (\mathbf{v}_1, \dots, \mathbf{v}_n)$ eine Basis von V , und es sei $M = \{\mathbf{w}_1, \dots, \mathbf{w}_m\} \subset V$, $m > n$. Es werde angenommen, die $\mathbf{w}_1, \dots, \mathbf{w}_m$ seien linear unabhängig. Nach dem Austauschsatz 4.1 kann einer der Basisvektoren aus \mathcal{B} etwa durch den Vektor $\mathbf{v} = \mathbf{w}_1$ ausgetauscht werden. Man erhält dann die Basis $\mathbf{w}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$, und \mathbf{w}_2 kann als Linearkombination

$$\mathbf{w}_2 = b_1\mathbf{w}_1 + a_2\mathbf{v}_2 + \dots + a_n\mathbf{v}_n$$

ausgedrückt werden. So fährt man weiter fort, indem man \mathbf{v}_2 durch \mathbf{w}_2 austauscht, etc, so dass man schließlich die $\mathbf{v}_1, \dots, \mathbf{v}_n$ durch die $\mathbf{w}_1, \dots, \mathbf{w}_n$ ersetzt hat. Aber

Abbildung 13: Kreise mit den Radien a bzw. $b < a$ und zugehörige Ellipse (Menge der Punkte P) mit den Halbachsen a und b



$m > n$, und \mathbf{w}_m kann nun als Linearkombination der $\mathbf{w}_1, \dots, \mathbf{w}_n$ ausgedrückt werden. Aber das heißt, dass die $\mathbf{w}_1, \dots, \mathbf{w}_m$ 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. \square

4.2 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

$$\mathcal{E} : \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad x = x(t), \quad y = y(t) \quad (4.3)$$

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 (4.3) folgt. \square

Parameterdarstellung: Bei dieser Darstellung wird die Ellipse (4.3) als Menge der Punkte $P = P(t)$ für $0 \leq t \leq 2\pi$ definiert. Aus der Abbildung 13 liest man 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}

$$\mathcal{E} = \{\mathbf{p}(t) | \mathbf{p}(t) = (x(t), y(t))' = (a \cos t, b \sin t)\}. \quad (4.4)$$

$\mathbf{p}(t)$ der Vektor mit dem Anfangspunkt M und dem Endpunkt $P(t)$.

Die Gleichung (4.3) 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 $x = 0$). 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

$$a = \sqrt{\lambda_1}, \quad b = \sqrt{\lambda_2}. \quad (4.5)$$

In Bezug auf (4.4) 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 (4.3) äquivalent zu

$$\mathbf{y}'\Lambda^{-1}\mathbf{y} = 1. \quad (4.6)$$

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

$$y_{01} = \sqrt{\lambda_1}, \quad y_{02} = \sqrt{\lambda_2}.$$

4.3 Die Differentiation von Vektoren

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

$$\frac{d\mathbf{x}}{dx_j} = \begin{pmatrix} dx_1/dx_j \\ \vdots \\ dx_j/dx_j \\ \vdots \\ dx_n/dx_j \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{e}_j. \quad (4.7)$$

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 (4.7) tritt u.a. dann auf, wenn eine Größe in Abhängigkeit von einem Vektor $\mathbf{b} = (b_1, \dots, b_p)'$ 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 $A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \dots + x_n\mathbf{a}_n$, so dass

$$\frac{\partial \mathbf{y}}{\partial x_j} = \frac{\partial A\mathbf{x}}{\partial x_j} = \mathbf{a}_j, \quad j = 1, \dots, n$$

Dementsprechend erhält man

$$\frac{d\mathbf{y}}{d\mathbf{x}} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n] = A. \quad (4.8)$$

4.3.2 Die Differentiation quadratischer Formen

Es wird die quadratische Form $Q(\mathbf{x}) = \mathbf{x}'C\mathbf{x}$ betrachtet, wobei $C' = C$ eine $(n \times n)$ -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)

$$\frac{\partial Q}{\partial x_j} = \mathbf{e}_j' C \mathbf{x} + \mathbf{x}' C \mathbf{e}_j \quad (4.9)$$

\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

$$\frac{\partial Q}{\partial x_j} = 2\mathbf{e}'_j C \mathbf{x}. \quad (4.10)$$

Fasst man die \mathbf{e}'_j zur Einheitsmatrix zusammen, so erhält man

$$\frac{\partial Q}{\partial \mathbf{x}} = 2C\mathbf{x}. \quad (4.11)$$

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 (4.11)

$$\frac{\partial \mathbf{x}'\mathbf{x}}{\partial \mathbf{x}} = 2I\mathbf{x} = 2\mathbf{x}. \quad (4.12)$$

Die Maximierung von $\mathbf{x}'A\mathbf{x}$ unter der Nebenbedingung $\mathbf{x}'\mathbf{x} = a \in \mathbb{R}$. Die allgemeine Theorie der Extremwertbestimmung unter Nebenbedingungen kann in Abschnitt 4.3.4 nachgelesen werden.

Zu maximieren sei

$$Q(\mathbf{x}) = \mathbf{x}'A\mathbf{x} + \lambda(\mathbf{x}'\mathbf{x} - a), \quad \mathbf{x}'\mathbf{x} - a = 0, \quad A' = A \quad (4.13)$$

Dann ist

$$\frac{dQ}{d\mathbf{x}} = 2A\mathbf{x} - 2\lambda\mathbf{x},$$

so dass

$$A\mathbf{x} = \lambda\mathbf{x}, \quad (4.14)$$

d.h. $Q(\mathbf{x})$ wird maximal, wenn \mathbf{x} ein Eigenvektor von A ist und $\lambda \in \mathbb{R}$ der zugehörige Eigenwert ist.

4.3.3 Die Kleinste-Quadrate-Schätzung für das Lineare Modell

Es sei X eine $(m \times n)$ -Matrix mit vollem 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

$$\mathbf{y} = X\mathbf{b} + \mathbf{e} \quad (4.15)$$

\mathbf{e} ist ein Fehlervektor, und \mathbf{b} soll so bestimmt werden, dass der Fehler minimiert 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

$$\mathbf{e}'\mathbf{e} = (\mathbf{y} - X\mathbf{b})'(\mathbf{y} - X\mathbf{b}) = \mathbf{y}'\mathbf{y} - \mathbf{y}'X\mathbf{b} - \mathbf{b}'X'\mathbf{y} + \mathbf{b}'X'X\mathbf{b} \quad (4.16)$$

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 (4.8) ist $d(X\mathbf{b})/d\mathbf{b} = X$, und nach (4.11) 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

$$\left. \frac{d(\mathbf{e}'\mathbf{e})}{d\mathbf{b}} \right|_{\mathbf{b}=\hat{\mathbf{b}}} = 2(X'\mathbf{y} - (X'X)\mathbf{b}) = 0 \quad (4.17)$$

so dass $X'\mathbf{y} = (X'X)\hat{\mathbf{b}}$, d.h.

$$\hat{\mathbf{b}} = (X'X)^{-1}X'\mathbf{y}, \quad (4.18)$$

Die Projektionsmatrix: Es sei $\hat{\mathbf{y}} = X\hat{\mathbf{b}}$, so dass

$$\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{e}} = X\hat{\mathbf{b}} + \hat{\mathbf{e}}, \quad (4.19)$$

wobei $\hat{\mathbf{e}}$ der Vektor der Fehler ist, wenn \mathbf{y} durch $\hat{\mathbf{y}}$ "vorhergesagt" wird. Man hat dann wegen (4.18)

$$\hat{\mathbf{y}} = X\hat{\mathbf{b}} = X(X'X)^{-1}X'\mathbf{y} = P\mathbf{y}; \quad (4.20)$$

darin ist

$$P = X(X'X)^{-1}X' \quad (4.21)$$

die *Projektionsmatrix*. Offenbar gilt

$$P' = P \quad (4.22)$$

$$PP = P \quad (4.23)$$

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

$$\hat{\mathbf{y}}'\hat{\mathbf{e}} = 0, \quad (4.24)$$

d.h. $\hat{\mathbf{y}}$ und $\hat{\mathbf{e}}$ sind orthogonal. Denn wegen (4.22) und (4.23) gilt

$$\hat{\mathbf{y}}'\hat{\mathbf{e}} = (P\mathbf{y})'\hat{\mathbf{e}} = \mathbf{y}'P'(\mathbf{y} - \hat{\mathbf{y}}) = \mathbf{y}'P\mathbf{y} - \mathbf{y}'PP\mathbf{y} = 0.$$

ppp

Einige Implikationen: Es sei $X = Q\Sigma T'$, $\Sigma = \Lambda^{-1/2}$, die SVD von X . Dann ist $X'X = T\Lambda T'$ und

$$(X'X)^{-1} = (T\Lambda T')^{-1} = (T')^{-1}\Lambda^{-1}T^{-1} = T\Lambda^{-1}T', \quad (4.25)$$

denn $T^{-1} = T'$ wegen der Orthonormalität von T . (4.18) impliziert dann

$$\hat{\mathbf{b}} = T\Lambda^{-1}T'T\Lambda^{1/2}Q'\mathbf{y} = T\Lambda^{-1/2}Q'\mathbf{y}. \quad (4.26)$$

Aber es ist $\mathbf{y} = X\mathbf{b} + \mathbf{e} = Q\Lambda^{1/2}Q'\mathbf{b} + \mathbf{e}$, so dass aus (4.26)

$$\hat{\mathbf{b}} = T\Lambda^{-1/2}Q'(Q\Lambda^{1/2}Q'\mathbf{b} + \mathbf{e}) = \mathbf{b} + T\Lambda^{-1/2}Q'\mathbf{e} \quad (4.27)$$

folgt, und schreibt man $T\Lambda^{-1/2}Q'$, indem man das dyadische Produkt anwendet, so ergibt sich (vergl. (2.111), Seite 78, wo der Fall $X = Q\Lambda^{1/2}T'$ betrachtet wird)

$$\hat{\mathbf{b}} = \mathbf{b} + \left(\sum_{k=1}^n \frac{\mathbf{t}_k \mathbf{q}'_k}{\sqrt{\lambda_k}} \right) \mathbf{e} \quad (4.28)$$

Der wahre Parametervektor \mathbf{b} und die Kleinst-Quadrate-Schätzung $\hat{\mathbf{b}}$ unterscheiden 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 $(t_{1j}, \dots, t_{kj}, \dots, t_{nj})'$ gegeben und man hat

$$\mathbf{c}_j = t_{1j}\lambda_1\mathbf{t}_1 + t_{2j}\lambda_2\mathbf{t}_2 + \dots + t_{nj}\lambda_n\mathbf{t}_n. \quad (4.29)$$

Sind nur die ersten Eigenwerte $\lambda_1, \lambda_2, \dots$ "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:

$$\mathbf{y} = Q\Sigma T'\mathbf{b} + \mathbf{e} = Q\Sigma(T'\mathbf{b}) + \mathbf{e} = L\boldsymbol{\beta} + \mathbf{e}, \quad L = Q\Sigma \quad (4.30)$$

mit L als neuer Prädiktormatrix und $\boldsymbol{\beta} = T'\mathbf{b}$ als neuem Parametervektor; die Spaltenvektoren von L sind orthogonal und also unkorreliert. Im Skriptum über Regressionsverfahren wird dieser Ansatz ausführlicher diskutiert.

4.3.4 Extrema unter Nebenbedingungen

Es sei $f(x_1, \dots, x_n)$ eine Funktion der Variablen x_1, \dots, x_n . Gesucht sind diejenigen Werte x_{0j} von x_j , $j = 1, \dots, n$, für die f ein Maximum annimmt, wobei aber

die Nebenbedingung $g(x_1, \dots, x_n) = k$, k ein Konstante berücksichtigt werden sooo, d.h. der Vektor $\mathbf{x}_0 = (x_{01}, \dots, x_{0n})'$ soll so bestimmt werden, dass auch $g(x_{01}, \dots, x_{0n}) = k$ 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-dimensionalen 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

$$f_x + f_y g' = 0 \tag{4.31}$$

$$g_x + g_y g' = 0 \tag{4.32}$$

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

$$\begin{pmatrix} f_x \\ f_y \end{pmatrix} = \lambda \begin{pmatrix} g_x \\ g_y \end{pmatrix} \tag{4.33}$$

schreiben kann. $\lambda \in \mathbb{R}$ ist ein neuer, freier Parameter, der sogenannte *Lagrange-Faktor* oder *Lagrange-Multiplikator*. Er drückt einfach aus, dass man nur etwas

über die Orientierung, nicht aber über die Länge der Gradientenvektoren am Ort des Maximums weiß. Die Vektorgleichung (4.33) zusammen mit der Bedingung $g(x, y) = 0$ führt sofort auf ein System von drei Gleichungen mit den drei Unbekannten x, y und λ :

$$f_x - \lambda g_x = 0 \quad (4.34)$$

$$f_y - \lambda g_y = 0 \quad (4.35)$$

$$g(x, y) = 0 \quad (4.36)$$

Diese Überlegungen müssen nicht immer explizit durchgeführt werden, denn sie implizieren die Möglichkeit, von vornherein die *Lagrange-Funktion* $L(x, y, \lambda)$ aufzustellen:

$$L(x, y, \lambda) = f(x, y) + \lambda g(x, y), \quad g(x, y) = 0. \quad (4.37)$$

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 (4.34), (4.35) und (4.36) heißen zusammen die *Lagrange-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

$$f_x = -2x, \quad f_y = -\frac{2}{3}y, \quad g_x = 1, \quad g_y = 1,$$

und man erhält das Gleichungssystem

$$\begin{aligned} -2x + \lambda &= 0 \\ -\frac{2}{3}y + \lambda &= 0 \\ x + y - 2 &= 0, \end{aligned}$$

woraus $x = 1/2$, $y = 3/2$ und $\lambda = -1$ folgt. □

Beispiel 4.2 (Satz von Courant-Fisher). Es sei A eine symmetrische, positiv-definite $n \times n$ -Matrix mit den Eigenwerten $\lambda_1 \geq \lambda_2 \leq \dots \geq \lambda_n$. Dann gilt

$$\max_{\mathbf{x} \neq \vec{0}} \frac{\mathbf{x}' A \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \max_j \lambda_j = \lambda_1, \quad (4.38)$$

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

$$\min_{\mathbf{x} \neq \vec{0}} \frac{\mathbf{x}' A \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \min_j \lambda_j = \lambda_n \quad (4.39)$$

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. \square

4.4 Alternativer Beweis von Satz 2.31

Es sei \mathbf{a}_{zi} der i -te Zeilenvektor von A , $i = 1, \dots, m$. 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:

$$U = \{\mathbf{x} \in \mathbb{R}^n | A\mathbf{x} = \vec{0}\} = \text{kern}(A), \quad V = \{\mathbf{x} \in \mathbb{R}^n | A\mathbf{x} = \mathbf{y} \neq \vec{0}\}.$$

$U = \text{kern}(A)$ 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 $U + V = \mathbb{R}^n$ und es folgt

$$\dim(U + V) = \dim(U) + \dim(V) = \dim \mathbb{R}^n = n. \quad (4.40)$$

Zur Bestimmung von $\dim(U)$ werde

$$A\mathbf{x} = x_1\mathbf{a}_1 + \dots + x_r\mathbf{a}_r + x_{r+1}\mathbf{a}_{r+1} + \dots + x_n\mathbf{a}_n = \vec{0}$$

betrachtet. Ohne Beschränkung der Allgemeinheit kann angenommen werden, dass die Spaltenvektoren $\mathbf{a}_1, \dots, \mathbf{a}_r$ die linear unabhängigen Vektoren von A sind. Schreibt man

$$x_1\mathbf{a}_1 + \dots + x_r\mathbf{a}_r = -(x_{r+1}\mathbf{a}_{r+1} + \dots + x_n\mathbf{a}_n),$$

und berücksichtigt man, dass die \mathbf{a}_{r+k} für $k = 1, \dots, n - r$ Linearkombinationen der \mathbf{a}_j , $j = 1, \dots, r$ sind, so dass

$$\mathbf{a}_{r+k} = \lambda_{k1}\mathbf{a}_1 + \dots + \lambda_{kr}\mathbf{a}_r = \sum_{j=1}^r \lambda_{kj}\mathbf{a}_j \quad (4.41)$$

gelten muß, so hat man

$$\sum_{j=1}^r x_j \mathbf{a}_j = - \sum_{k=1}^{n-r} x_{r+k} \sum_{j=1}^r \lambda_{kj} \mathbf{a}_j. \quad (4.42)$$

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

$$\mathbf{x} = \mathbf{x}_{r+k} = (x_1, \dots, x_r, 0, \dots, 0, 1, 0, \dots, 0)', \quad k = 1, \dots, n-r$$

setzen, wobei die 1 an der $(r+k)$ -ten Stelle stehen soll, also

$$\begin{aligned} \mathbf{x}_{r+1} &= (x_1, \dots, x_r, 1, 0, \dots, 0)', \\ \mathbf{x}_{r+2} &= (x_1, \dots, x_r, 0, 1, 0, \dots, 0)', \\ \mathbf{x}_{r+3} &= (x_1, \dots, x_r, 0, 0, 1, 0, \dots, 0)' \\ &\vdots \end{aligned}$$

Die Gleichung (4.42) 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

$$\mathbf{x}_k = (-\lambda_{k1}, \dots, -\lambda_{kr}, 0, \dots, 0, 1, 0, \dots, 0)', \quad k = 1, \dots, n-r \quad (4.43)$$

wobei die 1 an der $(r+k)$ -ten Stelle steht. In der Tat ist

$$A\mathbf{x}_k = -\lambda_{k1}\mathbf{a}_1 - \dots - \lambda_{kr} + \mathbf{a}_{r+k} = 0$$

wegen (4.41).

Die \mathbf{x}_k sind linear unabhängig, wie man sofort sieht, denn

$$\mu_1 \mathbf{x}_1 + \dots + \mu_k \mathbf{x}_k = 0$$

impliziert $\mu_i = 0$ für $i = 1, \dots, n-r$. 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 $\text{rg}(A) = r$ angekommen wurde, der Rang von $\text{kern}(A)$ kann also nicht größer als $n-r$ sein. Wegen (4.40) (= Dimensionssatz) folgt weiter, dass $\text{rg}(V) = \mathcal{L}(A) = r$ ist. \square

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 $\mathbf{a}_1, \dots, \mathbf{a}_n$ linear unabhängig und $\sum_j x_j \mathbf{a}_j = \vec{0}$ nur dann, wenn $x_1 = \dots = x_n = 0$. \square

4.5 Transformationen und Abbildungen

Dieser Abschnitt enthält einige grundsätzliche Betrachtungen über Produkte von Matrizen und Vektoren, die einerseits das Verständnis der Vektor- und Matrixrechnung vertiefen, andererseits für das Verständnis der unmittelbaren Anwendung der Matrixrechnung auf Fragen der multivariaten Statistik nicht unbedingt notwendig sind und deshalb übersprungen werden können.

Das Produkt $X\mathbf{u} = \mathbf{v}$, X eine $(m \times n)$ -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:

$$f: \mathcal{M} \rightarrow \mathcal{N}, \quad x \mapsto y = f(x), \quad x \in \mathcal{M}, \quad y \in \mathcal{N} \quad (4.44)$$

Man schreibt gelegentlich auch

$$f(\mathcal{M}) = \mathcal{N}. \quad (4.45)$$

$f(\mathcal{M})$ heißt das *Bild* von \mathcal{M} in \mathcal{N} , und \mathcal{M} ist das *Urbild* von $f(\mathcal{M})$. Man schreibt auch $\text{Im}f = \mathcal{N}$.

Eine spezielle Abbildung ist die *Identität* oder identische Abbildung

$$\text{id}(\mathcal{M}) = \mathcal{M}. \quad (4.46)$$

Die Einheitsmatrix I_n der Spalten bzw. Zeilen aus den n -dimensionalen Einheitsvektoren $\mathbf{e}_1, \dots, \mathbf{e}_n$ bestehen, spezifiziert die identische Abbildung, denn sicherlich gilt

$$I_n \mathbf{x} = \mathbf{x}, \quad \mathbf{x} \in \mathbb{R}^n. \quad (4.47)$$

Für eine Teilmenge von Abbildungen existiert die *inverse Abbildung* f^{-1} :

$$f(\mathcal{M}) = \mathcal{N}, \quad f^{-1}f(\mathcal{M}) = \mathcal{M} = f^{-1}(\mathcal{N}). \quad (4.48)$$

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, dass verschiedenen Elementen $x, x' \in \mathcal{M}$ der gleiche Wert $y \in \mathcal{N}$ 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

$$f(\mathcal{M}) = \{f(x), x \in \mathcal{M}\}. \quad (4.49)$$

Offenbar gilt $f(\mathcal{M}) \subseteq \mathcal{N}$.

Definition 4.1 *Es sei $f : \mathcal{M} \rightarrow \mathcal{N}$. 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 $\text{kern}(f) = \{x \in \mathcal{M} | f(x) = 0\}$ heißt Kern der Abbildung f ; man schreibt für den Kern auch $\text{kern}(f) = f^{-1}(\vec{0})$.
5. Es sei $f(\mathcal{M}) = \mathcal{N}$, d.h. $f(\mathbf{x}) = \mathbf{y}$ für $\mathbf{x} \in \mathcal{M}$, $\mathbf{y} \in \mathcal{N}$. Dann heißt $f^{-1}(\mathbf{y}) = \{x \in \mathcal{M} | f(x) = \mathbf{y}\}$ die Faser über $\mathbf{y} \in \mathcal{N}$.

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 Definition der Faser offenbar voraus, dass die Inverse existiert. \square

Beispiele: $f : \mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto ax + b$ für $a, b \in \mathbb{R}$ fest gewählte Konstante. 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.

\mathbb{R} bezeichnet die Menge der reellen Zahlen. 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, \dots, 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 $(m \times n)$ -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 $(m \times n)$ -Matrix ist. Die Schreibweise $f : V_m \rightarrow V_n$ 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 $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$ 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 4.2 Es seien V und W Vektorräume und $f: V \rightarrow W$ sei eine Abbildung von V in W . f heißt linear bzw. Homomorphismus, wenn

$$f(\lambda \mathbf{v} + \mu \mathbf{w}) = \lambda f(\mathbf{v}) + \mu f(\mathbf{w}) \quad (4.50)$$

für $\lambda, \mu \in \mathbb{R}$ und für alle $\mathbf{v} \in V$ und $\mathbf{w} \in W$. Insbesondere heißt f Isomorphismus, wenn f bijektiv ist; man sagt auch, f definiere einen Homomorphismus bzw. Isomorphismus für f bijektiv. f definiert einen Endomorphismus, wenn $V = W$, und 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 4.1 von Abbildungen entspricht für $f \in \mathbb{R}^{m,n}$ der Nullvektor $\vec{0}$.

Satz 4.3 Es sei $f(\mathcal{M}) = \mathcal{N}$. Dann gilt

1. f ist surjektiv genau dann, wenn $\text{Im } f = \mathcal{N}$
2. f ist injektiv genau dann, wenn $\text{kern } f = \{\vec{0}\}$.
3. f sei injektiv und die Vektoren $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathcal{M}$ seien linear unabhängig. Dann sind auch die Bilder $f(\mathbf{x}_1), \dots, f(\mathbf{x}_n)$ linear unabhängig.

Anmerkung: Die Schreibweise $\text{kern } f = \{\vec{0}\}$ bedeutet, dass $\text{kern } f$ 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, betrachte 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 $\mathbf{v} - \mathbf{u} \in \text{kern } f$.

Um 3. einzusehen sei angenommen, dass

$$\lambda_f(\mathbf{x}_1) + \dots + \lambda_n f(\mathbf{x}_n) = \vec{0}$$

gilt. Es wurde vorausgesetzt, dass f injektiv ist. Daraus folgt, dass

$$\lambda_1 \mathbf{x}_1 + \dots + \lambda_n \mathbf{x}_n = \vec{0}$$

gelten muß, denn $\lambda_1 \mathbf{x}_1 + \dots + \lambda_n \mathbf{x}_n$ ist ja das Urbild von f . Da die $\mathbf{x}_1, \dots, \mathbf{x}_n$ als linear unabhängig vorausgesetzt wurden, muß $\lambda_1 = \dots = \lambda_n = 0$ gelten, und dann folgt sofort, dass auch die $f(\mathbf{x}_j)$ linear unabhängig sind. □

Definition 4.3 Es sei f eine Abbildung $f: \mathbb{R}^m \rightarrow \mathbb{R}^n$; dann heißt die Dimensionalität des Bildes $\text{Im } f$ der Rang; man schreibt $\text{rg}(f) = \dim \in f$.

f sei durch eine Matrix $A \in \mathbb{R}^{m,n}$ definiert, so dass $A: \mathbb{R}^m \rightarrow \mathbb{R}^n$ und $\mathbf{x} \mapsto \mathbf{y} = A\mathbf{x}$. Dann ist $f = A \cdot \mathbf{e}_1, \dots, \mathbf{e}_m$ ist die kanonische Basis von \mathbb{R}^m , d.h. die n m -dimensionalen Spaltenvektoren von A können als Linearkombinationen $A\mathbf{e}_1, A\mathbf{e}_2, \dots, A\mathbf{e}_n$ geschrieben werden. Dann ist das Bild der durch A definierten Abbildung die lineare Hülle

$$\text{Im}A = \mathcal{L}(A\mathbf{e}_1, A\mathbf{e}_2, \dots, A\mathbf{e}_n).$$

Demnach wird $\text{Im}A$ auch der *Spaltenraum* von A bezeichnet. Der Begriff des Ranges einer Matrix A wird in Abschnitt 2.3 noch ausführlich diskutiert.

Beispiel 4.3 Es sei $f: \mathbb{R}^3 \rightarrow \mathbb{R}^4$; 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

$$\text{kern}(f) = \{(x_1, x_2, x_3)' | x_1 = -2x_2\} = \{(-2x_2, x_2, x_3)'\}.$$

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

$$\begin{pmatrix} x_1 + 2x_2 \\ 0 \\ \frac{1}{2}x_1 + x_2 \\ 0 \end{pmatrix} = \frac{x_1}{2} \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix} = \left(\frac{x_1}{2} + x_2\right) \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix},$$

d.h. die Spalten von A haben die Form $(2c, 0, c, 0)'$ mit $c \in \mathbb{R}$ (Barrantes Campos (2012), p. 231). \square

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 $M\mathbf{y} = \mathbf{v}$ 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 $\mathbf{u} \in V_n$ ein 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

$$T = \begin{pmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{pmatrix} = \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix}. \quad (4.51)$$

T definiert eine Abbildung $\mathbb{R}^2 \rightarrow \mathbb{R}^2$:

$$T\mathbf{x} = \begin{pmatrix} x_1 \cos \phi + x_2 \sin \phi \\ x_1 \sin \phi - x_2 \cos \phi \end{pmatrix} = x_1 \begin{pmatrix} \cos \phi \\ \sin \phi \end{pmatrix} + x_2 \begin{pmatrix} \sin \phi \\ -\cos \phi \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

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. \square

4.6 Determinanten

Die Determinante $\det(A) = |A|$ einer $(n \times n)$ -Matrix A ist eine Funktion, die der Matrix A eine reelle Zahl $a \geq 0$ zuordnet; $\det(A)$ und $|A|$ sind zwei äquivalente Schreibweisen für die Determinante. Es ist

$$\det(A) = |A| = \sum_{j=1}^n (-1)^{j+1} a_{1j} \det(A_{1j}), \quad (4.52)$$

wobei A_{1j} eine $((n-1) \times (n-1))$ -Matrix ist, die aus A entsteht, indem man in A die erste Zeile und die j -te Spalte streicht. Die Definition (4.52) ist also rekursiv.

Für eine (2×2) -Matrix erhält man dementsprechend

$$\begin{aligned} \det(A) &= \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = (-1)^{1+1}a_{11}a_{22} + (-1)^{2+1}a_{12}a_{21} \\ &= a_{11}a_{22} - a_{12}a_{21}. \end{aligned} \quad (4.53)$$

Für eine (3×3) -Matrix A erhält man

$$\begin{aligned} \det(A) &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = (-1)^{1+1}a_{11}(a_{22}a_{33} - a_{23}a_{32}) \\ &\quad + (-1)^{2+1}a_{12}(a_{11}a_{33} - a_{13}a_{31}) \\ &\quad + (-1)^{3+1}a_{13}(a_{11}a_{22} - a_{12}a_{21}), \end{aligned} \quad (4.54)$$

d.h.

$$\det(A) = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{11}a_{33} - a_{13}a_{31}) + a_{13}(a_{11}a_{22} - a_{12}a_{21}) \quad (4.55)$$

Die folgenden Aussagen lassen sich beweisen:

1. $\det(I) = 1$, I die Einheitsmatrix,
2. $\det(AB) = \det(A)\det(B)$, $A, B \in \mathbb{R}^{n,n}$
3. $\det(A') = \det(A)$,
4. $\det(cA) = c^n \det(A)$, $c \in \mathbb{R}$,
5. $\det(A) \neq 0$ dann und nur dann, wenn A vollen Rang hat.

Eigenwerte: Es sei $A \in \mathbb{R}^{n,n}$ und es gelte $AT = T\Lambda$, d.h. T ist die Matrix der Eigenvektoren von A und Λ ist die Matrix der zugehörigen Eigenwerte. Dann ist insbesondere $A\mathbf{t} = \lambda\mathbf{t}$, \mathbf{t} ein Eigenvektor mit dem zugehörigen Eigenwert λ . Es folgt $A\mathbf{t} - \lambda\mathbf{t} = \vec{0}$, oder

$$(A - \lambda I)\mathbf{t} = \begin{pmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{pmatrix} \mathbf{t} = \vec{0} \quad (4.56)$$

I die $(n \times n)$ -Einheitsmatrix. Dann ist

$$|A - \lambda I| = 0 = p(\lambda), \quad (4.57)$$

d.h. $p(\lambda)$ ist ein Polynom n -ten Grades in λ , wie man sich durch Anwendung der Definition (4.52) der Determinante auf $A - \lambda I$ klarmachen kann. $p(\lambda)$ heißt *charakteristisches Polynom*; es hat soviele Nullstellen, wie es von Null verschiedene Eigenwerte λ von A gibt.

Beweis der Aussage $T'T = I \Rightarrow TT' = I$: Es gilt $|A'A| = |A'| |A| = |I| = 1$, wegen $A'A = I$ (Anwendung von 2. und 1.). Weiter gilt $|AA'| = |A| |A'| = |I| = 1$ (Kommutativität der Multiplikation reeller Zahlen), also folgt $AA' = I$. \square

Literatur

- [1] Calcaterra, C. (2008) Liner combinations of Gaussinas with a single variance are dense in L^2 . *Proceeding of the World Congress on Engineering*, Vol II, WCE 2008
- [2] Cadima, J., Jolliffe, I. (2009) On relationships between uncentered and column-cedntered principal component analysis. *Pak. J. Statistics*, 25(4), 473 – 503
- [3] Eckart, C., Young, G. (1936), The approximation of one matrix by another of lower rank. *Psychometrika*, 1 (3): 211–8. doi:10.1007/BF02288367.
- [4] Fischer, G.: Lineare Algebra. Braunschweig Wiesbaden 1997
- [5] Gabor, D. (1948) A new microscopic principle. *Nature*, 161, 777–778
- [6] Gabriel, K.R. (1971) The biplot display of matrices with applpication to principal component analysis. *Biometrika* 58 (3), 453 – 467
- [7] Golub, G.H., van Loan, C. F.: Matrix computations. Baltimore 2013
- [8] Goshtaby, A., O’Neill, W. (1994) Curve fitting by a sum of Gaussians. *CV-GIP: GRaphical Models and Image Processing*, 56(4). 281–288
- [9] Hoaglin, D,C, Welsch, R.E. (1978) The hat matrix in Regression and ANOVA. *The American Statistician*, 32(1), 17 – 22
- [10] Honeine, P. (2014) An eigenanalysis of data centering in machine learning. *Statistics Machine Learning (stat.ML)* arXiv:1407.2904fl, 10 Jul 2014.
- [11] 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)
- [12] Karhunen, K. (1947), Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys.*, 37, 1–79.
- [13] Koecher, M.: Lineare Algebra und analytische Geometrie. Springer-Verlag, Berlin, Heidelberg, New York 1997
- [14] Lange, M. (2009) A Tale of two Vectors. *Dialectica*, 63(4), 397–431
- [15] Loève, M. (1978), Probability theory. Vol. II, 4th ed. Graduate Texts in Mathematics, 46. Springer-Verlag.
- [16] Lorenz, F.: Lineare Algebra I, II. Mannheim, 1988
- [17] Mardia, K.V., Kent, J. T., Bibby, J.M.: Multivariate Analysis. Academic Press, London, New York, Toronto 1979

- [18] Papoulis, A.: Probability, random variables, and stochastic processes. Tokyo 1965
- [19] 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.
- [20] Shaw-Taylor, J. Christianini, N.: Kernel Methods for Pattern Analysis. Cambridge University Press, Cambridge 2004
- [21] Stewart, G. W.: Introduction to Matrix Computations. Academic Press, New York 1973
- [22] Wong, E.: Stochastic processes in information and dynamical systems. New York 1971

Index

- Abbildung
 - bijektiv, 133
 - Bild von, 132
 - homomorphe, 134
 - identische, 132
 - injektiv, 133
 - inverse, 132
 - isomorphe, 134
 - Kern einer, 133
 - lineare, 134
 - Rang der Abbildung, 134
 - surjektiv, 133
- achsenparallel, 57
- Allgemeines Lineares Modell, 108
- assoziativ, 41
- Assoziativität, 41
- Ausreißer, 109
- Autokorrelation, 117
- Automorphismus, 134
- Basis
 - eines Vektorraums, 28
 - orthonormale, 30
- Basisentwicklung eines Vektors
 - orthonormale, 33
- Basisfunktion, 20, 114
- Basisvektoren, 29
- Cauchy-Schwarzsche Ungleichung, 17
- Cramersche Regel, 106
- Determinante, 24, 51
- diagonalisierbar, 97
- Dimension, 35
- Distanz
 - euklidische, 20
- dyadisches Produkt, 14
- Ebenengleichung, 21
- Eckart & Young
 - Satz von, 78
- Eigenraum, 102
- Eigenvektor, 52, 58
 - Links-, 95
 - Rechts-, 95
- Eigenwert
 - komplexer, 98
 - Nullstellen eines Polynoms, 97
- Einflußmatrix, 109
- Einheitsvektor, 12
- Einsvektor, 12
- Ellipsoid, 56
- Endomorphismus, 134
- Erzeugendensystem, 28
 - minimales, 29
- Faktorwerte, 75
- Faser, 133
- Fundamental-Lemma, 25, 31, 121
- Funktionsraum, 114
- Gesamtvarianz der Daten, 78
- Gleichung, charakteristische, 94
- Gradientenvektor, 128
- Gram-Matrix, 44
- Hat-Matrix, 108
- Hauptachsentransformation, 60, 80
- Hauptraum, 102
- Hauptvektor, 103
- Hebelwirkung, 55, 109
- Homomorphismus, 134
- idempotent, 111
- Identität, 132
- influence matrix, 109
- Inverse, 49, 126
 - generalisierte, 82
 - Moore-Penrose, 82
 - Pseudo, 82
- Isomorphismus, 134
- Kern, 104
- Kodimension, 35
- kollinear, 21, 24

kommutativ, 41
 Komplement
 orthogonales, 35
 Koordinaten bezüglich einer Basis, 28, 29
 Korrelation
 kanonische, 99
 Kreuzproduktmatrix, 43

 längeninvariant, 52
 Ladung, 33
 Ladungen, 75
 Lagrange
 -Faktor, 128
 -Funktion, 129
 -Multiplikator, 128
 sche Multiplikatorenregel, 129
 latente Struktur (einer Matrix), 49
 leverage, 55, 109
 lineare Gleichungssysteme, 23
 lineare Hülle, 28
 lineare Hülle (Gleich'syst.), 104
 Linearkombination, 13

 Matrix
 assoziierte, 98
 charakteristische Gleichung, 93
 Diagonal-, 37
 gestürzte, transponierte, 37
 hermitesch, 98
 imaginär, 98
 konjugierte, 98
 schief-symmetrisch, 98
 Spaltenstandardisierung, 43
 symmetrische, 37
 Matrixnorm, 85
 Matrizen
 ähnliche, 96
 ähnliche, 97
 Metrik
 City-Block, 9
 euklidisch, 8
 euklidische, 20
 Manhattan, 9
 Minkowski, 9, 20

 negativ semidefinit, 56
 Norm
 p -, 85
 -Matrix, 85
 allgemein, 19
 definierende Eigenschaften, 19
 euklidische, 84
 Frobenius-, 85
 Hilbert-Schmidt-, 85
 induzierte, 85
 Maximum, 84
 Schur-, 85
 Vektor-, 15
 Normalenvektor, 21
 Nullvektor, 12

 Orientierungsinvarianz, 52
 orthogonale Summe, 35
 orthogonales Komplement, 35
 Orthogonalität, 18
 Orthonormalbasis (ONB), 30
 orthonormale Basisentwicklung, 32

 Pfad, 116
 Polynom, charakteristisches, 59, 94, 137
 positiv semidefinit, 56, 64
 Produkt
 äußeres, 14
 dyadisches, 87
 inneres, 14
 Projektion
 Vektor auf einen anderen, 107
 Projektionsmatrix, 107, 126
 Pythagoras, 15

 quadratische Form, 56

 Rang
 einer Kreuzproduktmatrix, 47
 einer Vektormenge, 35
 voller, 35
 Rayleigh-Quotient, 67
 generalisierter, 99

Rohmatrix, 113
 Rotation, 31, 40, 54, 80
 Rotationsmatrix, 52

 Satz
 Eckart-Young, 88
 von Schmidt-Mirsky, 88, 89
 Satz von Courant-Fischer, 68
 Schur-Norm, 85
 singular, 59
 Singularvektoren, 72
 Singularwerte, 72
 Singularwertzerlegung, 71, 72
 Skalar, 15
 Skalarprodukt, 14
 Skalarprodukt als Ähnlichkeitsmaß, 18
 Spalteneffekte, 48
 Spaltenrang, 45
 Spaltenraum, 47, 135
 Spaltenvektoren, 37
 Spektraldarstellung, 65
 stetig (im quadrat. Mittel), 117
 stochastischer Prozess
 zentrierter, 117
 Strukturgleichung, 48
 SVD, 72

 Teilraum
 invarianter, 103
 Trajektorie, 116

 Varianz-Kovarianzmatrizen, 113
 als dyadisches Produkt, 44
 Varianzanalyse
 Beziehung zur Matrixanalyse, 48
 Vektor
 charakteristischer, 52, 58
 Einheits-, 12
 Eins-, 12
 latenter, 58
 nicht messfehlerfreier, 27
 Norm, 15
 normiert, 15
 Null-, 12
 Zeilen-, 15
 Zufalls-, 14
 Vektorabbildung, 40
 Vektornorm, 84, 85
 Vektorraum
 n -dimensionaler, 19
 n -dimensionaler, 30
 euklidischer, 20
 normierter, 85
 Vektortransformation, 40

 Wert
 charakteristischer, 58
 latenter, 58

 Zeileneffekte, 48
 Zeilenrang, 45
 Zeilenraum, 47
 Zeilenvektoren, 37
 Zeilenvektor, 15
 Zentrierungsmatrix, 111