

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

WS 2016/2017

U. Mortensen

Version 08. 01. 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: Faktorenanalyse und die üblichen Approximationen (Hauptachsentransformation), multiple und kanonische Regression bzw. Korrelation, die Diskriminanz- und Korrespondenzanalyse. Deswegen wird der Begriff des Vektorraums nur für endlichdimensionale, reelle Vektoren definiert und die Diskussion von Eigenvektoren und Eigenwerten auf den Fall symmetrischer Matrizen beschränkt. Der Begriff der Determinante wird nur am Rande erwähnt. Das Hauptziel ist die Herleitung der Singularwertzerlegung einer Matrix, die bereits den Kern der Hauptachsenanalyse einer Datenmatrix darstellt. Im Anhang wird darüber hinaus als Beispiel eine Anwendung der Matrixrechnung die Methode der Kleinsten Quadrate behandelt, wobei die Relevanz der Singularwertzerlegung für die Diskussion des in der Praxis häufig auftretenden Falles korrelierender Prädiktoren erläutert wird.

Eine wesentlich ausführlichere Einführung und die Vektor- und Matrixrechnung wird im Skriptum *Elemente der Linearen Algebra für Multivariate Statistische Verfahren* gegeben.

Inhaltsverzeichnis

1	Vektoren und Vektorräume	5
1.1	Der Sinn der Vektor- und Matrixrechnung	5
1.2	Punkträume	7
1.3	Vektoren	9
1.4	Vektorräume	18
1.5	Lineare Abhängigkeit und Unabhängigkeit	19
1.6	Erzeugendensysteme und Basen von Vektorräumen	24
2	Matrizen	31
2.1	Definitionen	31
2.2	Operationen mit Matrizen	32
2.2.1	Multiplikation einer Matrix mit einem Vektor	32
2.2.2	Multiplikation einer Matrix mit einer Matrix	33
2.2.3	Varianz-Kovarianz-Matrizen	36
2.3	Der Rang einer Matrix	38
2.4	Die Inverse einer Matrix	42
2.5	Rotationen, Quadratische Formen und Eigenvektoren	44
2.5.1	Rotationen	45
2.5.2	Quadratische Formen und Eigenvektoren	46
2.5.3	Die Inverse und die Wurzel einer symmetrischen Matrix	54
2.5.4	Der Rayleigh-Quotient	55
2.5.5	Bestimmung einer Basis	58
2.5.6	Die Singularwertzerlegung einer Matrix	60
2.5.7	Ellipsoide und die Konfiguration der Fälle	63
2.5.8	Basiswechsel	65
2.5.9	Die Pseudoinverse einer Matrix	65
2.5.10	Faktorwerte und Ladungen	67
2.6	Lineare Gleichungssysteme	73
2.7	Projektionen	76
2.7.1	Orthogonale Projektion eines Vektors auf einen anderen	76
2.7.2	Projektionen auf Hauptachsen	77
3	Anhang	78

3.1	Steinerscher Austauschsatz und das Fundamental-Lemma	78
3.2	Zur Berechnung von Ellipsen für eine Punktekonfiguration	79
3.3	Die Differentiation von Vektoren	81
3.3.1	Die allgemeine Differentiationsformel	81
3.3.2	Die Differentiation quadratischer Formen	82
3.3.3	Die Kleinste-Quadrate-Schätzung für das Lineare Modell	82
3.3.4	Extrema unter Nebenbedingungen	84
3.4	Alternativer Beweis von Satz 2.19	86
3.5	Transformationen und Abbildungen	88
3.6	Determinanten	93
	Literatur	95
	Index	95

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. Ein Standarddatensatz besteht aus einer Tabelle X (d.h. einer *Matrix*) mit m Zeilen und n Spalten von Messwerten x_{ij} , $1 \leq i \leq m$ und $1 \leq j \leq n$, wobei die m Zeilen "Fälle" repräsentieren, das sind Objekte oder Personen oder Zeitpunkte, an bzw. zu denen Messungen von n Größen, den "Variablen", vorgenommen wurden; die Variablen werden durch die Spalten der Tabelle oder Matrix X repräsentiert. Hat man z.B. 100 Fälle und 7 Variablen, so hat man bereits 700 Messwerte zu inspizieren und zu interpretieren. Das ist schon wegen der großen Zahl keine leichte Aufgabe, die überdies erschwert wird durch die oft gefundene Tatsache, dass die gemessenen Variablen miteinander korrelieren, d.h. es existieren mehr oder weniger stark ausgeprägte Abhängigkeiten zwischen den x_{ij} . Diese bedeuten, dass ein Messwert nicht unabhängig von einem anderen interpretiert werden kann, und die Frage ist nun, in welcher Weise diese Abhängigkeiten zu charakterisieren und zu berücksichtigen sind.

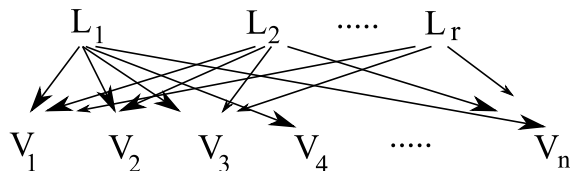
Ein Standardansatz für die Interpretation solcher Daten besteht in der Annahme der Existenz voneinander unabhängiger, latenter Variablen. Das sind Variablen, die nicht direkt gemessen wurden, von denen aber angenommen wird, dass sie auf die meisten der gemessenen Variablen einwirken und so die Korrelationen zwischen den gemessenen Variablen erzeugen (Abb. 1). Ein Messwert x_{ij} wird demnach anhand der Gleichung

$$x_{ij} = a_{1j}f_{i1} + a_{2j}f_{i2} + \dots + a_{rj}f_{ir} + e_{ij}, \quad r \leq \min(m, n) \quad (1.1)$$

interpretiert. Hierin sind die f_{i1}, \dots, f_{ir} Skalenwerte des i -ten Falles auf r voneinander unabhängigen Größen, eben den latenten Variablen, und die Koeffizienten a_{1j}, \dots, a_{rj} sind Gewichte oder "Ladungen", mit denen die r latenten Variablen in die Messungen der j -ten gemessenen Variablen eingehen. Der Term e_{ij} repräsentiert alle Effekte, die in die Messung x_{ij} eingehen und die nicht auf die latenten Variablen zurückzuführen sind. Ein wichtiger Aspekt dieses Ansatzes ist, dass die latenten Variablen alle nur *additiv* auf die gemessenen Größen wirken. Die Annahme der Additivität ist nicht selbstverständlich oder trivial, denn es ist denkbar, dass latente Variable existieren, die z.B. multiplikativ in die Messungen eingehen, wenn etwa der Effekt der einen latenten Variable proportional zum Effekt einer anderen latenten Variablen ist. Das Postulat der Additivität ist die einfachste Annahme¹, die man machen kann und man muß schauen, wie weit man damit kommt.

¹Man denke an das Prinzip des Ockhamschen Rasierers!

Abbildung 1: Latente Variablen L_1, L_2 und L_3 und gemessene Variablen V_1, \dots, V_n . Die f_{i1}, \dots, f_{ir} aus Gleichung (1.1) sind Messwerte der latenten Variablen L_1 bis L_r .



Der Ansatz (1.1) ist auch als Faktorenanalyse bekannt. Die Annahme latenter Variablen wird auch bei Verfahren wie der multiplen Regression und ihrer Verallgemeinerungen (kanonische und partielle Regression) so wie in Klassifikationsverfahren (Diskriminanzanalyse) gemacht. Der Ansatz (1.1) enthält $r \times n$ unbekannte Parameter a_{kj} plus $m \times r$ unbekannte Parameter f_{ik} , $k = 1, \dots, r$, und der Wert von r ist ebenfalls unbekannt. Da nur die Messungen x_{ij} gegeben sind, macht der Ansatz (1.1) nur Sinn, wenn diese unbekanntes Größen aus den x_{ij} geschätzt, d.h. berechnet werden können. So kann man vermuten, dass sich die f_{ik} aus den x_{ij} auf der Basis eines zu (1.1) analogen Ansatzes berechnen lassen:

$$f_{ik} = x_{i1}t_{1k} + x_{i2}t_{2k} + \dots + x_{in}t_{nk} \quad (1.2)$$

Jetzt müssen nur noch die $n \times r$ Koeffizienten t_{jk} gefunden werden, und die Frage ist, wie man sie findet. Außer der oben schon geforderten stochastischen Unabhängigkeit der f_{i1}, \dots, f_{ir} wird man zusätzliche Forderungen an diese Größen stellen, – etwa, dass die Varianz der f_{i1} , $i = 1, \dots, m$ maximal sein soll, d.h. dass sich die Fälle sich bezüglich der ersten latenten Variablen maximal unterscheiden sollen, dass die Varianz der f_{i2} die zweitgrößte sein soll, etc. Dazu muß man Nebenbedingungen formulieren, die sicherstellen, dass diese Varianzen endliche Größen sind. Schließlich müssen noch die Koeffizienten a_{jk} in Gleichung (1.1) bestimmt werden. Ohne auf Details einzugehen kann man sagen, dass die Rechnungen äußerst unübersichtlich werden. Spätestens hier zeigt sich, dass der Übergang zur Vektor- und Matrixschreibweise nicht nur die Übersichtlichkeit der Rechnungen erhöht, sondern auch das Auffinden der Antwort auf die Frage, wie die t_{jk} im Ansatz (1.2) zu bestimmen sind, erleichtert.

Bevor Vektoren und Matrizen eingeführt werden, muß ein kurzer Blick auf die Möglichkeit, sowohl die Fälle wie auch die Variablen durch Punkte in einem n -dimensionalen Raum zu repräsentieren, geworfen werden.

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 (wenn man etwa von Zentimetern zu Millimetern übergeht) und Punkte 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. Man rechnet leicht nach, dass dann der folgende Satz gilt:

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 $a, b \in \mathbb{R}$,

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

²Das sind die natürlichen Zahlen $\mathbb{N} = 0, 1, 2, \dots$, die rationalen Zahlen $\mathbb{Q} = \{p/q | p, q \in \mathbb{N}, q \neq 0\}$, und die irrationalen Zahlen, die sich nicht als Quotient p/q mit $p, q \in \mathbb{N}$ 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.

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 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) und 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 modellie-

ren, 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") mehrere Merkmale gemessen und die Korrelationen zwischen den Merkmalen berechnet. 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; für maximal drei Dimensionen (3 Variablen) kann er graphisch durch einen Pfeil repräsentiert werden (s. Abbildung 3). Die Länge des Vektors ist durch die Distanz $d(P, Q)$ gegeben, seine Orientierung ist durch das Verhältnis der Komponenten zueinander definiert; 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 SVMs (Support Vector Machines) erfordern mathematische Grundlagen, die in einem gesonderten Skript vorgestellt werden.

1.3 Vektoren

Es werden nur endliche Vektorräume über \mathbb{R} (die Menge der reellen Zahlen) betrachtet³.

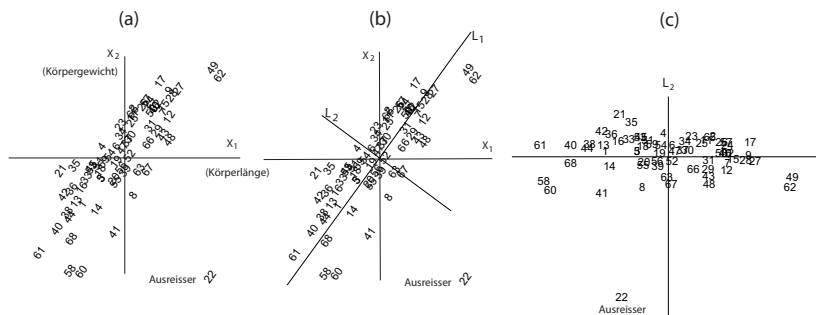
Definition 1.1 *Ein n -dimensionaler Vektor ist ein n -Tupel reeller Zahlen:*

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

die x_1, \dots, x_n heißen Komponenten des Vektors, und es kommt auf die Reihen-

³Die Lehrbuch-Definition des Begriffs des Vektorraums ist sehr viel allgemeiner, aber für die Zwecke dieses Skriptums ist die folgende Definition hinreichend.

Abbildung 2: Punktekonfiguration: (a) jeder Punkt entspricht einer Person, die Koordinaten sind (zentriert) (i) Körperlänge, (ii) Körpergewicht, (b) hypothetische "latente" Dimensionen L_1 und L_2 , L_1 ist durch maximale Ausdehnung der Konfiguration bestimmt, L_2 zu L_1 orthogonal, (c) Darstellung der Konfiguration im $L_1 \times L_2$ -System.



folge der Komponenten an. Es wird auch $\mathbf{x} \in \mathbb{R}^n$ geschrieben.⁴

Dementsprechend sind

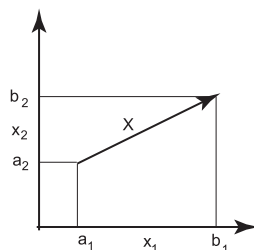
$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$$

verschiedene 3-dimensionale Vektoren.

Anmerkung: Die Bezeichnung 'n-dimensional' für einen Vektor mit n Komponenten kann als rein nominelle Definition verstanden werden. Andererseits legt der Ausdruck 'dimensional' oder auch 'Dimension' nahe, dass damit die Menge der möglichen Orientierungen eines Vektors bezeichnet werden soll. Dies tatsächlich so, muß aber spezifiziert werden. So kann man eine Menge 3-dimensionaler Vektoren haben, die aber alle in einer Ebene des 3-dimensionalen Raumes liegen. Innerhalb der Ebene sind immer noch beliebig viele Orientierungen möglich, – obwohl Orientierungen, die aus der Ebene hinausweisen, nicht vorkommen. Man kann sich aber den Begriff der Dimension veranschaulichen, indem mit einer Qualität (von was auch immer), oder, in der Psychologie, mit einer Fähigkeit in Verbindung bringt. Qualitäten oder Fähigkeiten können u. U. aus anderen Qualitäten oder Fähigkeiten zusammengesetzt sein: so wird die Intelligenz oft als Kombination von elementaren Fähigkeiten konzipiert. Eine Dimension kann dann als Qualität oder Fähigkeit definiert werden, die sich nicht aus den übrigen betrachteten Dimensionen zusammensetzen lassen. Diese Interpretation ist mit dem

⁴Es gibt eine allgemeinere Definition des Vektorbegriffs. Die hier gegebenen Definitionen (Vektor, Vektorraum, etc) beziehen sich auf die in der multivariaten Statistik benötigten Anwendungen der linearen Algebra.

Abbildung 3: Ein Vektor und seine Komponenten



Begriff der linearen Unabhängigkeit von Vektoren kompatibel, der in Abschnitt 1.5 eingeführt wird. \square

Ein Vektor heißt *gestürzt* oder *transponiert*, wenn er als Zeile angeschrieben wird:

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

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

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

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

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

gegeben, s. a. Abb. 3. 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.

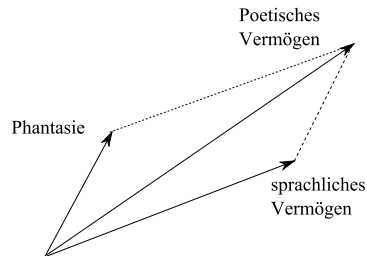
Addition von Vektoren Es seien \mathbf{x} und \mathbf{y} zwei n -dimensionale Vektoren. Dann heißt

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

die Summe der Vektoren \mathbf{x} und \mathbf{y} . Vektoren können also nur addiert werden, wenn sie dieselbe Anzahl von Komponenten haben. Abb. 4 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 erzeugen⁵. Nicht-additive Wechselwirkungen zwischen den Merkmalen werden durch die Wahl einer Vektorrepräsentation der Merkmale ausgeschlossen.

⁵Auch 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 4: 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. Man bemerke, dass \mathbf{y} als Addition von Vektoren analog zum in Abbildung 4 gezeigten Beispiel definiert ist; es handelt sich nicht um eine unziemliche Übertragung der Physik auf die Psychologie. Vielmehr sind die Addition von Kräften einerseits und die multiple Regression andererseits strukturgleiche Modelle. Dem *Modell* zufolge ist das "poetische Vermögen" eine additive Mischung von "verbalem Vermögen" und "Phantasie", die durch einen um einen Fehlervektor \mathbf{e} erweiterten Regressionsansatz

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

repräsentiert wird. \mathbf{e} repräsentiert zufällige Effekte und alle Aspekte von poetischem Vermögen, die nicht zu "verbalem Vermögen" oder "Phantasie" gehören. Die Komponenten x_{ij} beziehen sich auf ein Koordinatensystem, dass in Abb. 4 nicht eingezeichnet wurde; die Wahl eines speziellen Koordinatensystems ist für die Vektoraddition nicht wesentlich. Ob dieser Ansatz vernünftig ist, ist eine empirische Frage. \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*

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

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

$$\mathbf{x}\mathbf{y}' = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} (y_1, \dots, y_m) = \begin{pmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_m \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_m \\ \vdots & \vdots & \ddots & \vdots \\ x_n y_1 & x_n y_2 & \cdots & x_n y_m \end{pmatrix} \quad (1.19)$$

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

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. $\langle \mathbf{x}, \mathbf{y} \rangle$ und $\mathbf{x} \cdot \mathbf{y}$ sind alternative Schreibweisen für das Skalarprodukt $\mathbf{x}'\mathbf{y}$. 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.

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.

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

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

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

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

zentrierter Vektor. Offenbar ist

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

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:⁶ 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 5 (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

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

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 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*

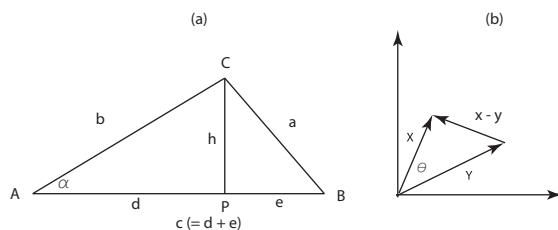
$$(\mathbf{x}' \mathbf{y})^2 \leq \|\mathbf{x}\|^2 \|\mathbf{y}\|^2. \quad (1.31)$$

⁶**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. 5 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 5: Zum Kosinussatz $a^2 = b^2 + c^2 - 2bc \cos \alpha$



Beweis: Es gilt

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

vergl. (1.29), 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

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

$$\left| \sum_{i=1}^n x_i y_i \right|^2 \leq \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2 \quad (1.33)$$

angeschrieben.

Es sei insbesondere $\mathbf{y} = \lambda \mathbf{x}$, $\lambda \in \mathbb{R}$. Dann ist $\theta = 0$ und $\cos^2 \theta = 1$, so dass in diesem Fall

$$(\mathbf{x}'\mathbf{y})^2 = \|\mathbf{x}\|\|\mathbf{y}\|, \text{ d.h. } \left| \sum_{i=1}^n x_i y_i \right|^2 = \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2 \quad (1.34)$$

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*⁷. Variablen X und Y , deren Korrelation gleich Null ist, werden durch orthogonale Vektoren \mathbf{x} und \mathbf{y} repräsentiert.

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

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} \exp \left[-\frac{1}{2(1-r^2)} (z_x^2 + z_y^2 - 2rz_xz_y) \right] \quad (1.35)$$

⁷Von griechisch *orthogonios*: ortho = rechts, gon = winklig

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.⁸

Definition 1.5 *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 .*

Beispiel 1.2 Es seien $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ irgendzwei parallele Vektoren; diese beiden Vektoren definieren eine Gerade, d.h. 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

$$\mathbf{n}'\mathbf{x} = a_1\mathbf{n}'\mathbf{x}_1 + a_2\mathbf{n}'\mathbf{x}_2 = 0, \tag{1.36}$$

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 30 geführt werden. Man erhält die

⁸Diese 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.

Ebenengleichung

$$\mathbf{n}'\mathbf{x} = n_1x_1 + n_2x_2 + \cdots + n_nx_n = 0. \quad (1.37)$$

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} . (1.37) ist die Gleichung einer Ebene im \mathbb{R}^n . \mathcal{E} geht durch den Nullpunkt des Koordinatensystems⁹. \mathcal{E} ist ein Teilraum des Vektorraums $V = \mathbb{R}^n$. Denn es seien

$$\mathbf{y}_1 = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 \quad (1.38)$$

$$\mathbf{y}_2 = b_1\mathbf{x}_1 + b_2\mathbf{x}_2 \quad (1.39)$$

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 (1.37). Auf die oben gemachte Voraussetzung, dass \mathbf{x}_1 und \mathbf{x}_2 nicht parallel sind, d.h. nicht dieselbe Orientierung (in Zeichen: \nparallel) haben, wird im Folgenden eingegangen.

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

$$\mathbf{x} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{n}, \quad \text{für alle } \mathbf{x} \in \mathbb{R}^3, \quad \mathbf{x} \neq \vec{0} \quad (1.40)$$

□

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

⁹Diese Definition läßt sich für Ebenen, die nicht durch den Nullpunkte gehen, verallgemeinern, aber diese allgemeine Definition wird im Folgenden nicht benötigt.

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:

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

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

$$a_1 = a_2 = \cdots = a_p = 0. \quad (1.42)$$

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 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

$$\mathbf{x}_p = a_1 \mathbf{x}_1 + \cdots + a_{p-1} \mathbf{x}_{p-1} \quad (1.43)$$

Man hat dann die

Definition 1.6 *Gilt (1.41) und sind nicht alle a_j gleich Null, so heißen die \mathbf{x}_j linear abhängig. Ist dagegen (1.42) 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

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 = a_1 \begin{pmatrix} x_{11} \\ x_{21} \end{pmatrix} + a_2 \begin{pmatrix} x_{12} \\ x_{22} \end{pmatrix} \quad (1.44)$$

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

$$y_1 = a_1 x_{11} + a_2 x_{12} \quad (1.45)$$

$$y_2 = a_1 x_{21} + a_2 x_{22} \quad (1.46)$$

mit $\mathbf{y} = (y_1, y_2)'$, $\mathbf{x}_1 = (x_{11}, x_{21})'$, $\mathbf{x}_2 = (x_{12}, x_{22})'$. Man findet

$$a_1 = \frac{y_1 x_{22} - y_2 x_{12}}{x_{11} x_{22} - x_{12} x_{21}} \quad (1.47)$$

$$a_2 = \frac{y_2 x_{11} - y_1 x_{21}}{x_{11} x_{22} - x_{12} x_{21}} \quad (1.48)$$

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

$$x_{11} x_{22} - x_{12} x_{21} \neq 0 \quad (1.49)$$

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.

Anmerkung: Der Ausdruck $x_{11} x_{22} - x_{12} x_{21}$ ist gleich der *Determinante* der (2×2) -Matrix X , s. Anhang, Abschnitt 3.6. Die Determinante einer Matrix ist nur von Null verschieden, wenn die Spaltenvektoren der Matrix linear unabhängig sind, – vergl. die folgenden Betrachtungen. \square

Nun betrachte man den Fall

$$x_{11} x_{22} - x_{12} x_{21} = 0. \quad (1.50)$$

Er impliziert

$$\frac{x_{21}}{x_{11}} = \frac{x_{22}}{x_{12}} \quad (1.51)$$

Aber $x_{21}/x_{11} = \tan \theta$, θ der Winkel, der die Orientierung von \mathbf{x}_1 angibt, und (1.51) besagt, dass dieser Winkel auch die Orientierung von \mathbf{x}_2 definiert. Daraus folgt, dass (1.50) dann erfüllt ist, wenn \mathbf{x}_1 und \mathbf{x}_2 parallel (kollinear) sind. Dann gilt $\mathbf{x}_2 = \lambda \mathbf{x}_1$ und die Gleichung (1.44) geht über in die Gleichung

$$\mathbf{y} = a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 = a_1 \mathbf{x}_1 + a_2 \lambda \mathbf{x}_1 = (a_1 + a_2 \lambda) \mathbf{x}_1, \quad (1.52)$$

aus der sofort hervorgeht, dass (1.44) nur dann eine Lösung hat, wenn \mathbf{y} tatsächlich dieselbe Orientierung wie \mathbf{x}_1 hat. 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 $a_1 + a_2 \lambda = k$, k eine Konstante, erfüllt ist, für die also

$$a_1 = k - a_2 \lambda \quad (1.53)$$

gilt.

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 (1.45) und (1.46) bei linearer Unabhängigkeit der $\mathbf{x}_1, \mathbf{x}_2$ für *alle* Vektoren \mathbf{y} eine durch (1.47) und (1.48) eine 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: Sind die $\mathbf{x}_j \in \mathbb{R}^n$, so existiert keine Menge von $p > n$ Vektoren $\mathbf{x}_1, \dots, \mathbf{x}_p$, 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 27).

Satz 1.3 *Es sei $\mathbf{y} = a_1\mathbf{x}_1 + \dots + a_p\mathbf{x}_p$. Sind die $\mathbf{x}_1, \dots, \mathbf{x}_p$ 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. \square

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*.

Satz 1.5 Die Einheitsvektoren $\mathbf{e}_1, \dots, \mathbf{e}_n$ sind paarweise orthogonal.

Beweis: Es gilt

$$\mathbf{e}'_j \mathbf{e}_k = \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases} \quad (1.54)$$

wie man unmittelbar verifiziert. \square

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

Satz 1.6 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 1.7 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

Beispiel 1.3 Lineare Abhängigkeit und Skalarprodukt: 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 $\cos \theta = r_{xy} = 1$.

Denn 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. $\cos \theta = 1$.

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 1.4 Lineare Abhängigkeit und Korrelationen: Gegeben seien zwei Merkmale, die durch die Vektoren \mathbf{x}_0 und $a\mathbf{x}_0$, $a \in \mathbb{R}$, repräsentiert werden, – wenn die Messungen der zu den Merkmalen korrespondierenden Variablen messfehlerfrei wären, was sie im Allgemeinen nicht sind. Den Messungen entsprechend hat man $\mathbf{x} = \mathbf{x}_0 + \varepsilon_1$, $\mathbf{y} = a\mathbf{x}_0 + \varepsilon_2$. Der Korrelationskoeffizient r_{xy} entspricht dann

$$-1 < r_{xy} = \cos \theta_{xy} = \frac{(\mathbf{x}_0 + \varepsilon_1)'(a\mathbf{x}_0 + \varepsilon_2)}{\|\mathbf{x}_0 + \varepsilon_1\| \|a\mathbf{x}_0 + \varepsilon_2\|} < 1, \quad \varepsilon_1, \varepsilon_2 \neq 0 \quad (1.55)$$

Nur für den Spezialfall $\varepsilon_1, \varepsilon_2 = 0$ würde man den Fall $r_{xy} = \cos \theta_{xy} = 1$ erhalten, und der hat die Wahrscheinlichkeit 0. Für die gemessenen Vektoren \mathbf{x}_j , $j = 1, \dots, b$ heißt dies, dass *rechnerisch* alle Vektoren linear unabhängig sind. \square

1.6 Erzeugendensysteme und Basen von Vektorräumen

Definition 1.7 Es sei $M = \{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ eine Menge von n -dimensionalen Vektoren. Es sei

$$\mathcal{L}(M) = \{\mathbf{x} | \mathbf{x} = a_1\mathbf{x}_1 + \dots + a_m\mathbf{x}_m\}, \quad (1.56)$$

d.h. $\mathcal{L}(M)$ sei die Menge aller Linearkombinationen der Vektoren aus M . Dann heißt $\mathcal{L}(M)$ die lineare Hülle von M , und $\mathbf{x}_1, \dots, \mathbf{x}_m$ 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 1.8 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

$$\mathbf{v} = a_1 \mathbf{b}_1 + a_2 \mathbf{b}_2 + \dots + a_n \mathbf{b}_n. \quad (1.57)$$

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

Anmerkungen:

1. 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 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.
2. Es sei $\mathbf{v} = (v_1, \dots, v_n)'$; nach (1.57) gilt dann für die i -te Komponente

$$v_i = a_1 b_{i1} + \dots + a_n b_{in}, \quad (1.58)$$

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 (1.57) 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 2.6). $\mathcal{L}(\mathbf{b}_1, \dots, \mathbf{b}_r)$ mit $r < n$ definiert einen Teilraum von V (vergl. Satz 1.8).

Definition 1.9 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 1.9 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 1.10 *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

$$\mathbf{b}'_j \mathbf{b}_k = \begin{cases} 0, & j \neq k \\ 1, & j = k \end{cases}, \quad j, k = 1, \dots, n \quad (1.59)$$

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 ; sie heißt die kanonische 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. (1.54). 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 1.11 *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 über Basiswechsel 2.5.8, Seite 65.

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 1.6 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 1.8 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 3.1.

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 linear unabhängige Menge von Vektoren. Dann ist sie bereits eine Basis, da man mit ihr 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

$$\mathbf{x} = a_1 \mathbf{b}_1 + \dots + a_n \mathbf{b}_n = \sum_{k=1}^n a_k \mathbf{b}_k. \quad (1.60)$$

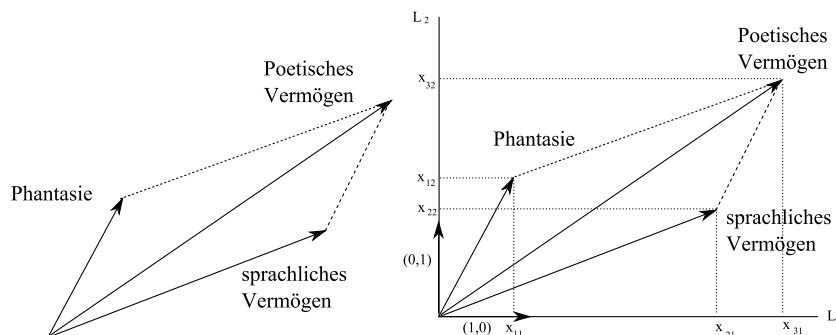
Für die Koeffizienten a_j ergibt sich eine einfache Darstellung. Man betrachte dazu das Skalarprodukt $\mathbf{x}' \mathbf{b}_j$:

$$\mathbf{x}' \mathbf{b}_j = \sum_{k=1}^n a_k \mathbf{b}'_k \mathbf{b}_j = a_j, \quad j = 1, \dots, n \quad (1.61)$$

denn

$$\mathbf{b}'_k \mathbf{b}_j = \begin{cases} 0, & j \neq k \\ 1, & j = k \end{cases} \quad (1.62)$$

Abbildung 6: 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.



(1.60) kann dann in der Form

$$\mathbf{x} = \sum_{k=1}^n (\mathbf{x}'\mathbf{b}_k)\mathbf{b}_k. \quad (1.63)$$

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 $\mathbf{x}'\mathbf{b}_k$ in Gleichung (1.63) 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 (1.63). \square

In Abbildung 6 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 werden¹⁰.

¹⁰Derartige 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

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 18 gerechtfertigt.

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

Definition 1.12 *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 r linear unabhängige Vektoren enthält; 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

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. Dies bedeutet, dass sich

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.

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 . 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 $\mathbf{x} = (x_1, \dots, x_n)'$ ein n -dimensionaler Vektor. Mit dem Produkt $A\mathbf{x}$ ist dann die Linearkombination

$$\mathbf{y} = A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \dots + x_n\mathbf{a}_n \quad (2.6)$$

gemeint, d.h. der m -dimensionale Vektor $\mathbf{y} = A\mathbf{x}$ ist eine Linearkombination der Spaltenvektoren von A , wobei die Komponenten x_j von \mathbf{x} als Koeffizienten der \mathbf{a}_j auftreten. \mathbf{y} ist ein m -dimensionaler Vektor, weil die \mathbf{a}_j m -dimensionale Vektoren sind. Man rechnet leicht nach, dass die Komponenten von \mathbf{y} gerade gleich den Skalarprodukten der *Zeilenvektoren* $\tilde{\mathbf{a}}_i$ von A mit \mathbf{x} sind:

$$y_i = \mathbf{x}'\tilde{\mathbf{a}}_i = \sum_{j=1}^n a_{ij}x_j. \quad (2.7)$$

Ebenso läßt sich das Produkt eines m -dimensionalen Zeilenvektors \mathbf{x}' mit der $(m \times n)$ -Matrix A definieren. Es seien $\tilde{\mathbf{a}}_1, \dots, \tilde{\mathbf{a}}_m$ die Zeilenvektoren von A . Dann sei

$$\mathbf{z}' = \mathbf{x}'A = x_1\tilde{\mathbf{a}}_1 + x_2\tilde{\mathbf{a}}_2 + \dots + x_m\tilde{\mathbf{a}}_m, \quad (2.8)$$

d.h. der Zeilenvektor \mathbf{z}' ist eine Linearkombination der Zeilenvektoren von A . Die Zeilenvektoren von A sind n -dimensionale Vektoren, also ist \mathbf{z} ein n -dimensionaler Vektor. Die j -te Komponente von \mathbf{z}' ist gleich dem Skalarprodukt von \mathbf{x} mit dem j -ten *Spaltenvektor* \mathbf{a}_j von A :

$$z_j = \mathbf{x}'\mathbf{a}_j = \sum_{i=1}^m x_i a_{ij}. \quad (2.9)$$

Man rechnet leicht nach, dass für (2.6)

$$\mathbf{y}' = (A\mathbf{x})' = \mathbf{x}'A' \quad (2.10)$$

folgt, d.h. der transponierte m -dimensionale Vektor \mathbf{y} erscheint jetzt als Linearkombination der Zeilenvektoren von A' , – d.h. natürlich wieder als Linearkombination der Spalten von A , jetzt nur als Zeilenvektor angeschrieben. Ebenso impliziert (2.8)

$$\mathbf{z} = (\mathbf{z}')' = A'\mathbf{x}, \quad (2.11)$$

d.h. die Komponente z_i von \mathbf{z} ist durch das Skalarprodukt

$$z_i = \tilde{\mathbf{a}}_i' \mathbf{x} \quad (2.12)$$

des i -ten Spaltenvektors von A' , also dem i -ten Zeilenvektor von A , mit \mathbf{x} definiert.

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 3.5 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 dass man insgesamt die Matrixgleichung

$$AB = C \quad (2.13)$$

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 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 .

Man überprüft durch einfaches Nachrechnen, dass dann

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

gilt. Diese Beziehung verdeutlicht die Aussage 2: Die Spalten von C' sind Linearkombinationen der Spalten von B' , d.h. die Linearkombinationen der Zeilenvektoren von B liefern die Zeilenvektoren von C' , also die Spalten von C .

Beispiel 2.1 Es seien A und B durch

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}, \quad B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

gegeben. Dann ist (durch Nachrechnen überprüfen!)

$$C = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} \end{pmatrix} = [b_{11}\mathbf{a}_1 + b_{21}\mathbf{a}_2, b_{12}\mathbf{a}_1 + b_{22}\mathbf{a}_2]$$

und

$$\begin{aligned} C' = B'A' &= \begin{pmatrix} b_{11}a_{11} + b_{21}a_{12} & b_{11}a_{21} + b_{21}a_{22} & b_{11}a_{31} + b_{21}a_{32} \\ b_{12}a_{11} + b_{22}a_{12} & b_{12}a_{21} + b_{22}a_{22} & b_{12}a_{31} + b_{22}a_{32} \end{pmatrix} \\ &= [a_{11}\tilde{\mathbf{b}}_1 + a_{12}\tilde{\mathbf{b}}_2, a_{21}\tilde{\mathbf{b}}_1 + a_{22}\tilde{\mathbf{b}}_2, a_{31}\tilde{\mathbf{b}}_1 + a_{32}\tilde{\mathbf{b}}_2], \end{aligned}$$

wobei $\tilde{\mathbf{b}}_1$ und $\tilde{\mathbf{b}}_2$ die Spaltenvektoren von B' , also die Zeilenvektoren von B sind. Man überzeuge sich, dass die Zeilen von C' gleich den Spalten von C sind. \square

Die Matrixmultiplikation ist im Allgemeinen nicht *kommutativ*, d.h. im Allgemeinen gilt

$$AB \neq BA. \quad (2.16)$$

Damit das Produkt AB gebildet werden kann, muß die Anzahl der Spalten von A gleich der Anzahl der Zeilen von B sein. Damit das Produkt BA gebildet werden kann, muß die Anzahl der Spalten von B gleich der Anzahl der Zeilen von A sein. Da A als $(m \times n)$ -Matrix, B als $(n \times p)$ -Matrix definiert wurde, folgt, dass $p = m$ sein muß, damit das Produkt BA überhaupt gebildet werden kann.

Die Matrixmultiplikation ist *assoziativ*: ist A eine $(m \times n)$ -Matrix, B eine $(n \times r)$ -Matrix und C eine $(r \times s)$ -Matrix, so gilt

$$(AB)C = A(BC). \quad (2.17)$$

Einsicht in diese Aussage erhält man durch Nachrechnen.

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 Diagonalzellen von Null verschiedene Elemente haben:

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

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 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 Matrix¹¹.

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

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

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.

Die Zentrierungsmatrix: Es sei

$$H = I - \frac{1}{m} \vec{1} \vec{1}' \quad (2.28)$$

H heißt *Zentrierungsmatrix*; $\vec{1} \vec{1}'$ ist das dyadische Produkt von $\vec{1}$ mit sich selbst. H ist offenbar symmetrisch und idempotent, denn

$$H' = (I - \frac{1}{m} \vec{1} \vec{1}')' = I' - (\frac{1}{m} \vec{1} \vec{1}')' = I - \frac{1}{m} \vec{1} \vec{1}' = H, \quad (2.29)$$

¹¹Nach Jørgen Pedersen Gram (1850 – 1916), dänischer Mathematiker.

und

$$\begin{aligned}
 HH &= H' = \left(I - \frac{1}{m} \vec{1}\vec{1}'\right) \left(I - \frac{1}{m} \vec{1}\vec{1}'\right) \\
 &= I - \frac{1}{m} I\vec{1}\vec{1}' - \frac{1}{m} I\vec{1}\vec{1}' + \frac{1}{m^2} \vec{1}\vec{1}'\vec{1}\vec{1}' \\
 &= I - \frac{1}{m} I\vec{1}\vec{1}' = H,
 \end{aligned} \tag{2.30}$$

denn

$$\vec{1}\vec{1}'\vec{1}\vec{1}' = \vec{1}(\vec{1}'\vec{1})\vec{1}' = m\vec{1}\vec{1}'.$$

Die Idempotenz von H erweist sich u.a. als nützlich, wenn Eigenschaften von Varianz-Kovarianzmatrizen betrachtet werden.

Nun sei X nicht zentriert (X ist die Matrix der "Rohwerte", kurz auch "Rohmatrix" genannt). Die Varianz-Kovarianzmatrix S läßt sich nun in der Form

$$S = \frac{1}{m} X'HX \tag{2.31}$$

schreiben. Die übliche Biaskorrektur für die Varianz-Kovarianzschätzungen erhält man, indem man diese Gleichung mit $m/(m-1)$ multipliziert:

$$\hat{S} = \frac{m}{m-1} S = \frac{1}{m-1} X'HX. \tag{2.32}$$

Für die Stichprobenkorrelation r_{jk} gilt

$$r_{jk} = \frac{s_{jk}}{s_j s_k}. \tag{2.33}$$

Es sei

$$D = \begin{pmatrix} s_1 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 \\ & & \ddots & \\ 0 & 0 & 0 & s_n \end{pmatrix}.$$

Für die Matrix der Korrelationen bzw. der Kovarianzen erhält man dann

$$R = D^{-1}SD^{-1}, \quad S = DRD. \tag{2.34}$$

2.3 Der Rang einer Matrix

Auf Seite 30 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 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 27 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 : $\mathbf{x}_j = U\mathbf{a}_j$. 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 2.1 *Es sei X eine $(m \times n)$ -Matrix mit dem Zeilenrang r und dem Spaltenrang s . Dann gilt*

$$r = s \tag{2.35}$$

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

$$X = UV', \tag{2.36}$$

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

$$r \leq \min(m, n) \tag{2.37}$$

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 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

$$X = [\mathbf{x}_1, \dots, \mathbf{x}_n] = UA', \quad (2.38)$$

wobei A' die zu U gehörende $(s \times n)$ -Matrix der Koeffizienten ist (s. Satz 1.3 über die Eindeutigkeit der Koeffizienten bei linear unabhängigen \mathbf{u}_j , Seite 22): die Spaltenvektoren $\mathbf{a}_j = (a_{1j}, \dots, a_{sj})'$ von A' erzeugen \mathbf{x}_j gemäß

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

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

$$X' = [\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_m] = VB' \quad (2.39)$$

B' ist die zu V korrespondierende Koeffizientenmatrix, analog zu A' in (2.38), und

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

Da U in (2.38) 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 34). 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 (2.39), 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 behauptet¹².

Nach (2.38) gilt $X = UA'$, und nach (2.39) 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 22), wenn die

¹²Das Zeichen \wedge steht für 'und'.

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 (2.37). □

Anmerkungen:

1. U ist eine $(m \times r)$ -Matrix und V ist eine $(n \times r)$ -Matrix; beide Matrizen haben r Spalten, die die latenten Variablen repräsentieren.
2. Für $r = \min(m, n)$ sagt man, X habe den *vollen Rang*.
3. Aus dem Beweis zu Satz 2.1 ergibt sich, dass sich *jede* $(m \times n)$ -Matrix X als Produkt der Form (2.36) darstellen läßt.
4. Die in Abschnitt 2.5.6 eingeführte Singularwertzerlegung von X ist ein Spezialfall von (2.36), der vielen faktorenanalytischen Ansätzen zur Interpretation von Datenmatrizen zugrunde liegt. □

Folgerungen: Aus (2.36) folgen sofort die Aussagen

$$\text{rg}(X'X) = \text{rg}(XX') = \text{rg}(X), \tag{2.40}$$

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 . In Abschnitt 2.5.6, Seite ?? (Satz ??) wird ein weiterer Beweis dieser Aussage gegeben. Die Aussagen (2.40) sind für die multivariate Statistik von Bedeutung: bei geeigneter Zentrierung entsprechen den Matrizen $X'X$ bzw. XX' Kovarianz- bzw. Korrelationsmatrizen, die demnach denselben Rang wie die Datenmatrix haben.

Satz 2.2 *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.*

$$\text{rg}(C) \leq \min[\text{rg}(A), \text{rg}(B)]. \tag{2.41}$$

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 2.35 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 2.3 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

$$\text{rg}(X) = \text{rg}(PXQ). \quad (2.42)$$

Beweis: Nach Satz 2.2, Gleichung (2.41) (Seite 41), 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 2.2 Es sei A eine $(n \times n)$ -Matrix. A^{-1} sei eine Matrix derart, dass

$$AA^{-1} = A^{-1}A = I_n, \quad (2.43)$$

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 ; die Bedingung für die Existenz einer Inversen ist

Satz 2.4 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 (2.41), Seite 41, $\text{rg}(AA^{-1}) \leq \min(\text{rg}(A), \text{rg}(A^{-1}))$ 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 22). 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 2.2 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

$$a\alpha + b\gamma = 1 \Rightarrow \alpha = \frac{1 - b\gamma}{a} \quad (2.44)$$

$$a\beta + b\delta = 0 \Rightarrow \beta = -\frac{b\delta}{a} \quad (2.45)$$

$$c\alpha + d\gamma = 0 \Rightarrow \gamma = -\frac{c\alpha}{d} \quad (2.46)$$

$$c\beta + d\delta = 1 \Rightarrow \delta = \frac{1 - c\beta}{d} \quad (2.47)$$

Durch Einsetzen etwa des Ausdrucks für γ (2.46) in (2.44) etc findet man

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \quad (2.48)$$

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

$$b = a(d/c), \quad d = c(b/a),$$

andererseits folgt auch

$$d/c = b/a = \lambda,$$

d.h. es 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. \square

Beispiel 2.3 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

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

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

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

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. □

Satz 2.5 *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*

$$(AB)^{-1} = B^{-1}A^{-1}. \tag{2.49}$$

Beweis: Sicherlich gilt $AB(AB)^{-1} = I_n$, I_n die $(n \times n)$ -Einheitsmatrix. Diese Beziehung wird genau dann erfüllt, wenn $(AB)^{-1} = B^{-1}A^{-1}$ gilt, denn dann hat man

$$AB(AB)^{-1} = ABB^{-1}A^{-1} = AA^{-1} = I_n.$$

□

2.5 Rotationen, Quadratische Formen und Eigenvektoren

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) ergeben sich zwei für die Anwendungen in der Multivariaten Statistik wichtige Spezialfälle:

1. A ist eine $(n \times n)$ -Matrix und es gelte $A\mathbf{x} = \mathbf{y}$ für einen beliebigen Vektor $\mathbf{x} \in \mathbb{R}^n$. Dann ist ebenfalls $\mathbf{y} \in \mathbb{R}^n$. Im Allgemeinen unterscheiden sich \mathbf{x} und \mathbf{y} durch ihre Orientierung und durch ihre Länge. Es gibt zwei 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

$$\mathbf{x} \neq \mathbf{y}, \quad \|\mathbf{y}\|^2 = \mathbf{y}'\mathbf{y} = \mathbf{x}'T'T\mathbf{x} = \mathbf{x}'\mathbf{x} = \|\mathbf{x}\|^2 \quad (2.50)$$

gelten.

Satz 2.6 Die Beziehung (2.50) 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 (2.50). Um diese Behauptung einzusehen, werde $U = T'T$ gesetzt. Nach (2.50) soll $\mathbf{x}'U\mathbf{x} = \mathbf{x}'\mathbf{x}$ für alle \mathbf{x} gelten. Dann liefert (3.11), Seite 82

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

und nochmalige Ableitung liefert $U = I$ (s. Gleichung (3.8), Seite 81), d.h. $T'T = I$. Dies bedeutet, dass die Spaltenvektoren von T orthonormal sind.

Es muß noch $T'T = I \Rightarrow TT' = I$ gezeigt werden. Da der Zeilenrang einer Matrix gleich ihrem Spaltenrang ist, muß $\text{rg}(T') = \text{rg}(T) = n$ gelten, d.h. die Spaltenvektoren \vec{t}_i von T' (das sind die Zeilenvektoren von T) sind linear unabhängig, so dass nur noch ihre Orthonormalität gezeigt werden muß. Es gilt $T = TI = T(T'T) = (TT')T = AT$, mit $A = TT'$, so dass $T = AT$. d.h. $T - AT = 0$, 0 die Nullmatrix. Dann folgt $(T - AT)' = T' - T'A' = T'(I - A) = 0$, da $A' = A$. Mit $B = I - A$ hat man $T'B = 0$. Sei \mathbf{b}_j der j -te Spaltenvektor von B ; dann gilt $T'\mathbf{b}_j = \vec{0}_j$, $\vec{0}_j$ der j -te Nullvektor von 0 . $\vec{0}_j$ wird als Linearkombination der linear unabhängigen Spalten von T' dargestellt, woraus $\mathbf{b}_j = \vec{0}$ für alle $j = 1, \dots, n$ folgt, so dass $B = I - A = 0$, also $I = A = TT'$. \square

Hinweis: Am Ende des Abschnitts 3.6, Seite 93, 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

Satz 2.7 *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

$$T = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}, \quad (2.51)$$

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 44 entsprechend hat man $T' = T^{-1}$. \square

2.5.2 Quadratische Formen und Eigenvektoren

Übersicht: Es wird zunächst gezeigt, dass bestimmten, d.h. positiv semidefiniten symmetrischen Matrizen (Definition siehe weiter unten) M Ellipsoide zugeordnet werden können; jedem Fall (z.B. einer Zeile in einer Matrix $(m \times n)$ -Matrix X mit $M = X'X$ mit m Fällen und n Variablen) entspricht ein Punkt in einem n -dimensionalen Raum, und jedem dieser Punkte entspricht ein Ellipsoid, auf dem der Punkt liegt. Alle Ellipsoide haben dieselbe Orientierung. Die Orientierung ist durch die Eigenvektoren von M gegeben. Die Hauptachsen der Ellipsoide repräsentieren eine mögliche Wahl für latente, den Daten unterliegende Variablen. \square

Definition 2.3 Es sei M eine symmetrische $(n \times n)$ -Matrix, $\mathbf{x} \in \mathbb{R}^n$, und es gelte

$$Q_M(\mathbf{x}) = \mathbf{x}'M\mathbf{x}. \quad (2.52)$$

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

Definition 2.4 Es sei $M \in \mathbb{R}^{n,n}$ eine symmetrische Matrix und $\mathbf{x}'M\mathbf{x}$ sei eine quadratische Form. Dann:

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 2.8 Es sei $X \in \mathbb{R}^{m,n}$ eine nicht zentrierte Matrix. Die Varianz-Kovarianzmatrix $S = \frac{1}{m}X'HX$, H die Zentrierungsmatrix, ist positiv semidefinit.

Beweis: Auf Seite 37 wurde gezeigt, dass H symmetrisch und idempotent ist, d.h. es gilt $HH = H'H = H^2 = H$, d.h. $mS = X'HX = X'H'HX$, und für einen beliebigen, n -dimensionalen Vektor $\mathbf{x} \neq \vec{0}$ folgt mit $\mathbf{y} = HX\mathbf{x}$

$$\mathbf{x}'S\mathbf{x} = \mathbf{x}'X'H'HX\mathbf{x} = \|\mathbf{y}\|^2 \geq 0,$$

d.h. S ist positiv semidefinit. □

Satz 2.9 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 \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 (2.52) aus, so erhält man

$$\mathbf{x}'M\mathbf{x} = \sum_{i=1}^n m_{ii}x_i^2 + 2 \sum_{i<j} m_{ij}x_i x_j = k \quad (2.53)$$

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

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

$$\mathbf{x}'M\mathbf{x} = m_{11}x_1^2 + m_{22}x_2^2 + 2m_{12}x_1x_2 = k > 0. \quad (2.54)$$

Die Menge der 2-dimensionalen Vektoren $\mathbf{x} = (x_1, x_2)'$, die dieser Gleichung genügen, definiert eine Ellipse (s. a. Satz 2.12, Seite 50).

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 (2.53) bzw. (2.54), 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

$$\mathbf{x}'M\mathbf{x} = \mathbf{y}'\Lambda\mathbf{y} = k > 0, \quad \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n). \quad (2.55)$$

$\mathcal{E}_x = \{\mathbf{x} | \mathbf{x}'M\mathbf{x} = k > 0\}$ 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 T eine Rotation repräsentiere. Offenbar gilt

$$\mathbf{x}'M\mathbf{x} = \mathbf{y}'TMT\mathbf{y} = \mathbf{y}'\Lambda\mathbf{y} = k,$$

woraus

$$T'MT = \Lambda \quad (2.56)$$

folgt. Denn \mathcal{E}_y ist ein achsenparalleles Ellipsoid, so dass auch $\mathbf{y}'TMT\mathbf{y} = k$ ein achsenparalleles Ellipsoid sein muß, d.h. $T'MT = D$ muß eine Diagonalmatrix sein, $D = \text{diag}(d_1, \dots, d_n)$. Dann folgt aber

$$\mathbf{y}'\Lambda\mathbf{y} = \mathbf{y}'D\mathbf{y} = k,$$

d.h.

$$\sum_{k=1}^n \lambda_k y_k^2 = \sum_{k=1}^n d_k y_k^2.$$

Differenziert man beide Seiten zweimal nach y_k , so erhält man $\lambda_k = d_k$, d.h. $\Lambda = D$, und das ist (2.56). \square

Da T als Rotationsmatrix angenommen wurde, folgt, dass T orthonormal ist. Deshalb folgt durch Multiplikation der Gleichung (2.56) von links mit T die Gleichung

$$MT = T\Lambda. \quad (2.57)$$

Diese Gleichung besagt, dass die Spaltenvektoren von T durch M so transformiert werden, dass sich nur ihre Länge, nicht aber ihre Orientierung verändert. 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*

$$M\mathbf{t} = \lambda\mathbf{t}, \quad \mathbf{t} \neq \vec{0}, \quad (2.58)$$

genügt. Dann heißt \mathbf{t} Eigenvektor¹³. von M und λ heißt der zu \mathbf{t} gehörende Eigenwert von M .

¹³Synonym sind auch die Ausdrücke 'latenter Vektor' und 'latenter Wert' und 'charakteristischer Vektor' und 'charakteristischer Wert'.

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 $\vec{0} = \lambda \vec{0}$ 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 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. Die Eigenwerte λ einer quadratischen Matrix ergeben sich u.a. aus dem *charakteristischen Polynom* der Matrix $A - \lambda I$, I die Einheitsmatrix, das im Anhang, Abschnitt 3.6 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 (2.87) 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 .

Satz 2.10 *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.*

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

Satz 2.11 *Es sei M eine reelle, symmetrische $(n \times n)$ -Matrix. Dann ist der Rang von M gleich der Anzahl von Null verschiedener Eigenwerte.*

Beweis: Der Beweis macht implizit von Satz 2.2, Seite 41, Gebrauch. Es gilt $M = T\Lambda T'$, und es seien $r \leq n$ Eigenwerte ungleich Null. Dann enthält Λ $n - r$ Spalten (und Zeilen), die nur Nullen enthalten. Die Spaltenvektoren von M sind Linearkombinationen der Spalten von $T\Lambda$, so dass für den j -ten Spaltenvektor \mathbf{v}_j von M

$$\mathbf{v}_j = \lambda_1 t_{j1} \mathbf{t}_1 + \cdots + \lambda_r t_{jr} \mathbf{t}_r + \underbrace{0 + \cdots + 0}_{n-r}$$

gilt. Es genügt demnach,

$$M = T_r \Lambda_r T_r' \quad (2.59)$$

zu schreiben, wobei T_r die Matrix der Eigenvektoren ist, die zu von Null verschiedenen Eigenwerten korrespondieren, die in der Matrix Λ_r zusammengefaßt werden. Die Matrix $T_r \Lambda_r$ besteht aus den Spaltenvektoren $\lambda_j \mathbf{t}_j$, $j = 1, \dots, r$, die orthogonal und damit linear unabhängig sind, mithin hat $T_r \Lambda_r$ den Rang r . (2.59) bedeutet, dass die Spaltenvektoren von M sich als Linearkombinationen der $\lambda_j \mathbf{t}_j$ darstellen lassen, d.h. die Spaltenvektoren von M sind Elemente der linearen Hülle $\mathcal{L}(\lambda_1 \mathbf{t}_1, \dots, \lambda_r \mathbf{t}_r)$, und somit ist $\text{rg}(M) \leq r$. Aber T_r ist orthonormal, so dass aus (2.59) $MT_r = T_r \Lambda_r$ folgt (Multiplikation von rechts mit T_r'). Dies heißt aber, dass sich die Spaltenvektoren $\lambda_j \mathbf{t}_j$ von $T_r \Lambda_r$ als Linearkombinationen der Spalten von M darstellen lassen, d.h. sie liegen in der linearen Hülle $\mathcal{L}(M)$ von M . Dies bedeutet, dass $r = \text{rg}(T_r \Lambda_r) \leq \text{rg}(M)$ sein muß. Es muß also $\text{rg}(M) \leq r$ und $\text{rg}(M) \geq r$ gelten, so dass $\text{rg}(M) = r$ folgt. □

Der Satz 2.10 sagt noch wenig aus über die Eigenschaften einer symmetrischen Matrix (es sind stets reelle Matrizen gemeint), die nur positive Eigenwerte implizieren (oder nur negative). Der folgende Satz gibt weitere Auskunft.

Satz 2.12 *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*

$$M = GG' \quad (2.60)$$

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 Matrizen T (orthonormal) und $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_j)$, $\lambda_j \geq 0$ für alle j (s. Satz 2.10), 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 2.12 spezifiziert die Bedingungen, die eine symmetrische Matrix M erfüllen muß, um eine Ellipse bzw. ein Ellipsoid zu definieren.

Zur Bedeutung der Eigenvektoren positiv semidefiniter Matrizen: Es sei insbesondere $\mathbf{y}_1 = y_1\mathbf{e}_1$, $\mathbf{e}_1 = (1, 0, \dots, 0)'$, so dass $\|\mathbf{y}_1\| = y_1\|\mathbf{e}_1\| = y_1 \in \mathbb{R}$. Der Vektor \mathbf{y}_1 definiert dann die erste Halbachse des durch Λ definierten achsenparallelen Ellipsoids. Dann folgt

$$\mathbf{x}_1 = T\mathbf{y}_1 = y_1T\mathbf{e}_1 = y_1 \begin{pmatrix} t_{11} \\ t_{21} \\ \vdots \\ t_{n1} \end{pmatrix} = y_1\mathbf{t}_1, \quad y_1 \in \mathbb{R}, \quad (2.61)$$

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}_k orthogonal sind, definieren die restlichen Vektoren \mathbf{t}_k , $k \neq 1$, 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 (2.61) 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:

$$\mathbf{x}_j = Ty_j \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = y_j \begin{pmatrix} t_{1j} \\ t_{2j} \\ \vdots \\ t_{nj} \end{pmatrix}, \quad \mathbf{x}_j \in \mathcal{E}_x, \quad \mathbf{y}_j = y_j\mathbf{e}_j \in \mathcal{E}_y \quad (2.62)$$

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.

Eine Rotationsmatrix S ist orthonormal, und wenn in einem bestimmten Kontext gefolgert wird, dass $S = T$ auch eine Matrix von Eigenvektoren ist, so folgt, dass T ebenfalls orthonormal ist. Nun sei umgekehrt bekannt, dass T eine Matrix von Eigenvektoren von $M' = M$ ist. Die Frage ist, ob nun auch folgt, dass T orthonormal ist, – es ist ja denkbar, dass man nicht über eine Rotation auf

die Eigenschaft der Spalten von T , Eigenvektoren zu sein, gekommen ist, und vielleicht gibt es auch nicht-orthogonale Eigenvektoren von M . Dazu wird der folgende Satz bewiesen:

Satz 2.13 $M \in \mathbb{R}^{n,n}$ 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

$$\mathbf{t}'_j \mathbf{t}_k = \begin{cases} 0, & j = k \\ \|\mathbf{t}_j\| \neq 0, & j \neq k \end{cases} \quad (2.63)$$

Beweis: Ist T eine Rotationsmatrix, ist M symmetrisch und gilt $MT = T\Lambda$, 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

$$M\mathbf{t}_j = \lambda_j \mathbf{t}_j \quad (2.64)$$

$$M\mathbf{t}_k = \lambda_k \mathbf{t}_k \quad (2.65)$$

Die Gleichung (2.64) werde von links mit \mathbf{t}'_k , die Gleichung (2.65) von links mit \mathbf{t}_j multipliziert. Es entstehen die Gleichungen

$$\mathbf{t}'_k M \mathbf{t}_j = \lambda_j \mathbf{t}'_k \mathbf{t}_j \quad (2.66)$$

$$\mathbf{t}'_j M \mathbf{t}_k = \lambda_k \mathbf{t}'_j \mathbf{t}_k. \quad (2.67)$$

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 (2.63) 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. Der folgende Satz macht hierüber eine Aussage.

Satz 2.14 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 λ_j korrespondierende Eigenvektoren.

Beweis: Der Beweis wird hier nicht gegeben, da er vom Begriff der Determinante Gebrauch macht, der zwar im Anhang (Abschnitt 3.6) kurz eingeführt, aber in diesem Skript nicht weiter diskutiert wird. \square

Spektraldarstellung von M : Die Orthonormalität von T bedeutet, dass man aus $MT = T\Lambda$ (Gleichungen (2.66) und (2.67)) durch Multiplikation von rechts mit T' die Beziehung

$$M = T\Lambda T' = \sum_{k=1}^n \lambda_k \mathbf{t}_k \mathbf{t}'_k. \quad (2.68)$$

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 (2.68) von M heißt Spektraldarstellung von M .

Bemerkungen:

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 $\mathbf{y} \in \mathcal{E}_y$.
2. Eine Matrix muß nicht symmetrisch sein, damit Eigenvektoren für sie existieren. Allerdings existieren nicht für jede Matrix Eigenvektoren. Dazu betrachte man die Matrix (2.51), d.h. so sei

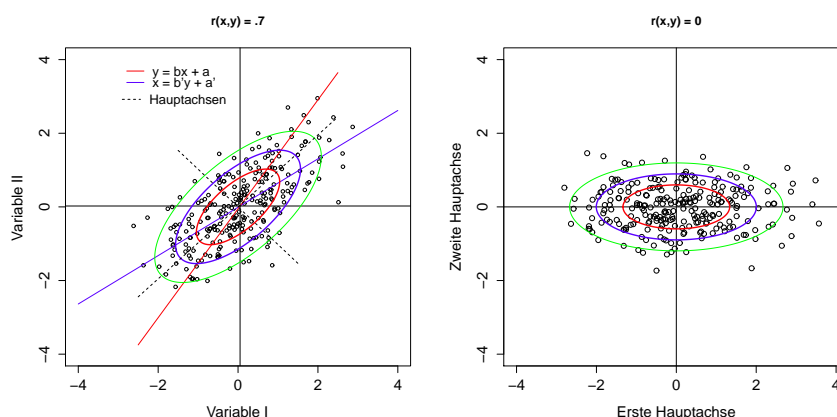
$$A = \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix}.$$

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

Abbildung 7: Links: Punktekonfiguration für $r_{xy} = .7$ mit Regressionsgeraden, Ellipsen und deren Hauptachsen; rechts: Die Hauptachsen als neue Koordinaten für die Punktekonfiguration. Zur Berechnung der Ellipsen s. Anhang 3.2, Seite 79.



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. \square

2.5.3 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 2.49, Seite 44)

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

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

$$M^{-1} = T\Lambda^{-1}T' = \sum_{k=1}^n \frac{1}{\lambda_k} \mathbf{t}_k \mathbf{t}_k' \quad (2.69)$$

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

Anmerkung: Die Beziehung (2.69) 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 (2.72) 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 nun¹⁴

$$M^{1/2} \stackrel{\text{Def}}{=} T\Lambda^{1/2}T' = \sum_{k=1}^n \sqrt{\lambda_k} \mathbf{t}_k \mathbf{t}'_k \quad (2.70)$$

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

$$(M^{1/2})' = (T\Lambda^{1/2}T')' = T\Lambda^{1/2}T', \quad (2.71)$$

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

$$(M^{1/2})^{-1} = M^{-1/2} = (T\Lambda^{1/2}T')^{-1} = T\Lambda^{-1/2}T' = \sum_{k=1}^n \frac{1}{\sqrt{\lambda_k}} \mathbf{t}_k \mathbf{t}'_k. \quad (2.72)$$

2.5.4 Der Rayleigh-Quotient

In Definition 2.3, Gleichung (2.52) (Seite 47) wurde der Begriff der quadratischen Form $\mathbf{x}'M\mathbf{x}$ eingeführt. Sie definiert ein n -dimensionales Ellipsoid, wenn M eine positiv definite (n, n) -Matrix ist (Definition 2.4, Seite 47) und dementsprechend $\mathbf{x} \in \mathbb{R}^n$ ist. In vielen Anwendungen ist es von Interesse, die *Orientierung* desjenigen Vektors zu kennen, in der das Ellipsoid seine größte Ausdehnung hat. Es gibt zwei Methoden, diesen Vektor zu bestimmen: (i) man benutzt die Differentialrechnung, indem man die quadratische Form nach \mathbf{x} differenziert und die Ableitung gleich Null setzt, wie bei der Bestimmung von Extremen üblich; dieses Vorgehen wird im Anhang, Abschnitt 3.3 besprochen; (ii) man bestimmt das Maximum des Rayleigh-Quotienten. Dieser Ansatz wird in diesem Abschnitt vorgestellt.

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

$$R(\mathbf{x}) = \frac{\mathbf{x}'M\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \frac{\mathbf{x}'M\mathbf{x}}{\|\mathbf{x}\|^2} \quad (2.73)$$

heißt Rayleigh-Quotient¹⁵ oder Rayleigh-Koeffizient.

¹⁴Mit dem Zeichen $\stackrel{\text{Def}}{=}$ soll ausgedrückt werden, dass der Ausdruck auf der linken Seite durch den auf der rechten Seite definiert wird.

¹⁵Nach dem britischen Physiker John William Strutt, Dritter Baron Rayleigh (1842–1919)

Anmerkung: Der Rayleigh-Quotient $R(\mathbf{x})$ hängt nicht von der Länge, sondern nur von der Orientierung des Vektors \mathbf{x} ab, denn es wird ja der normierte Vektor $\mathbf{x}/\|\mathbf{x}\|$ betrachtet:

$$R(\mathbf{x}) = \frac{\mathbf{x}'}{\|\mathbf{x}\|} M \frac{\mathbf{x}}{\|\mathbf{x}\|}.$$

Man könnte den Quotienten also auch in der Form $R(\mathbf{x}) = \mathbf{x}' M \mathbf{x}$ mit der Nebenbedingung $\mathbf{x}' \mathbf{x} = 1$ definieren. \square

Es gilt der folgende

Satz 2.15 (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

$$\max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}' M \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \max_j \lambda_j = \lambda_1, \quad (2.74)$$

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

$$\min_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}' M \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \min_j \lambda_j, \quad (2.75)$$

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. Da $\|\mathbf{t}_1\| = 1$ und $\mathbf{t}_1' M \mathbf{t}_1 = \lambda_1$ nimmt der Raleigh-Koeffizient den maximalen Wert für $\mathbf{x} = \mathbf{t}_1$ an. 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

Anmerkung: Es sei noch einmal darauf hingewiesen, dass die Aussage, dass $\max_{\mathbf{x}} R_M(\mathbf{x}) = \lambda_1$ mit $\mathbf{x} = \mathbf{t}_1$ ist, daraus folgt, dass $\mathbf{x} \in \mathcal{L}(\mathbf{t}_1, \dots, \mathbf{t}_n)$ beliebig gewählt werden kann, d.h. dass für alle möglichen Koeffizienten c_1, \dots, c_n Vektoren $\mathbf{x} = x_1 \mathbf{t}_1 + \dots + c_n \mathbf{t}_n$ resultieren derart, dass $R_M(\mathbf{x}) \leq \lambda_1$ gilt. Deshalb ist $\lambda_1 = \lambda_{\max}$. \square

Die Aussage (2.75) ergibt sich aus der folgenden Vervollständigung des Satzes von Courant-Fischer und ist eher ein Korollar zu diesem Satz.

Satz 2.16 *Es sei M wie in Satz 2.15 definiert. Dann gilt*

$$\max_{\mathbf{x} \perp \mathbf{t}_1, \dots, \mathbf{t}_k} \frac{\mathbf{x}' M \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \lambda_{k+1}, \quad k < n, \quad (2.76)$$

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 (2.74)

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

und analog zur Argumentation im Beweis zu Satz 2.15 folgt (2.76). \square

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

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

d.h.

$$\min_{\mathbf{x}} R_M(\mathbf{x}) = \lambda_n \quad (2.77)$$

so dass

$$\lambda_n \leq R_M(\mathbf{x}) \leq \lambda_1. \quad (2.78)$$

\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 (2.15) und (2.16) den Rayleigh-Quotienten $\mathbf{x}' M \mathbf{x} / \mathbf{x}' \mathbf{x}$ maximieren.*

2.5.5 Bestimmung einer Basis

Es sei $X = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ 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 2.1, 39, existieren stets Matrizen (m, r) - und (n, r) -Matrizen U und V ($U \in \mathbb{R}^{(m,r)}, V \in \mathbb{R}^{(n,r)}$) derart, dass $X = UV'$. Die Spaltenvektoren von X sind Linearkombinationen der m -dimensionalen Spaltenvektoren \mathbf{u}_k von U , und die Spaltenvektoren von X' sind Linearkombinationen der Spaltenvektoren von V . Die Spaltenvektoren von X sind Linearkombinationen der Spaltenvektoren von U , und V' ist die zugehörige Koeffizientenmatrix ($\mathbf{x}_j = U\tilde{\mathbf{v}}_j$, $\tilde{\mathbf{v}}_j$ der j -e Spaltenvektor von V'). Da die Spaltenvektoren von U linear unabhängig sind, liegen die Koeffizientenvektoren $\tilde{\mathbf{v}}_j$ und damit die Matrix V' eindeutig fest (Satz 1.3, Seite 22).

Es zeigt sich, dass die Wahl einer bestimmten Basis U die Wahl von V festlegt und umgekehrt.

Es gibt beliebig viele Möglichkeiten für eine Wahl von U bzw. V , so dass die Entscheidung für die Wahl einer bestimmten Basis nach irgendwelchen Kriterien erfolgen muß. So sei etwa X so definiert, dass die Zeilen von X zu "Fällen"¹⁶ korrespondieren und die Spalten von X zu "Variablen"; für jeden Fall i ($i = 1, \dots, m$) gibt es dann n Messungen x_{i1}, \dots, x_{in} ; die Koordinatenachsen repräsentieren die Variablen V_1, \dots, V_n . Der Zeilenvektor (x_{i1}, \dots, x_{in}) korrespondiert zum i -ten Zeilenvektor (u_{i1}, \dots, u_{ir}) von U ; es gilt

$$(x_{i1}, \dots, x_{in}) = (u_{i1}, \dots, u_{ir})V \quad (2.79)$$

Die erste Spalte von U enthält dann die Koordinaten u_{11}, \dots, u_{m1} der Fälle auf der ersten Achse eines durch die Basisvektoren U definierten Koordinatensystems, der zweite Spaltenvektor von U enthält die Koordinaten u_{12}, \dots, u_{m2} der Fälle auf der zweiten Achse des durch Basisvektoren von U definierten Koordinatensystems, etc.

Die Frage ist nun, welchen Annahmen bezüglich U und V gemacht werden können, ohne die Allgemeinheit unnötig einzuschränken. So kann man postulieren, dass V orthonormal ist. Die $\mathbf{v}_1, \dots, \mathbf{v}_r$ bilden dann eine orthonormale Basis für die Zeilenvektoren von X ; von dieser Basis kann man, wenn es von Vorteil sein sollte, zu irgendeiner anderen, insbesondere auch nicht orthogonalen Basis übergehen. Jedenfalls folgt dann für eine orthonormale Matrix V aus der Gleichung $X = UV'$ die Beziehung (Multiplikation von rechts mit V)

$$XV = U, \quad (2.80)$$

und weiter wegen $U' = V'X'$

$$U'U = V'X'XV'. \quad (2.81)$$

¹⁶Ein Fall ist eine Person oder ein ein Objekt, an dem Messungen vorgenommen werden, oder ein Zeitpunkt, zu dem die Messungen durchgeführt werden.

X habe den Rang $r \leq \min(m, n)$. Dann existieren r von Null verschiedene Eigenwerte λ_k , $k = 1, \dots, r$, und T_r sei die Matrix der zu diesen Eigenwerten korrespondierenden orthonormalen Eigenvektoren von $X'X$. Es sei Λ_r die Diagonalmatrix der von Null verschiedenen Eigenwerte von $X'X$; dann gilt $X'X = T_r \Lambda_r T_r'$. Man kann insbesondere $V = T_r$ setzen, denn bis jetzt ist nur gefordert worden, dass V orthonormal sein soll, und die Spalten von T_r sind orthonormal; für $r = n$ sind auch die Zeilen von $T = T_n$ orthonormal. Dann folgt aus (2.81)

$$U'U = V'X'XV' = T_r'X'XT_r = T_r'T_r\Lambda_rT_r'T_r = \Lambda_r, \quad (2.82)$$

d.h. U muß orthogonal sein. Aus (2.80) folgt

$$XT_r = U, \quad (2.83)$$

d.h. die Spaltenvektoren von U sind Linearkombinationen der Spalten von X . Multiplikation von rechts mit T_r' liefert

$$X = UT_r'. \quad (2.84)$$

T_r und Λ_r müssen numerisch bestimmt werden; die Statistikpakete enthalten entsprechende Programme. Ist X eine Datenmatrix, so findet man im Allgemeinen n *numerisch* von Null verschiedene Eigenwerte und die Frage ist, ob der "wahre" Rang r von X kleiner als n ist. Dies bedeutet, dass $n - r$ der Eigenwerte nur aufgrund von Rundungsfehlern oder wegen zufälliger Effekte von Null verschieden sind. Auf diese Frage wird hier nicht weiter eingegangen.

Die Lösung $U = XT_r$ für U ist für T_r spezifisch, weshalb im Folgenden eine spezifische Bezeichnung gewählt wird: $U = L_r$, wobei L_r an an den Ausdruck "latente Variable" erinnern soll. Dementsprechend hat man

$$X = L_rT_r'. \quad (2.85)$$

Die Spalten von L_r enthalten die Koordinaten der Fälle auf den Basisvektoren für die Spalten von X . Die Matrix X sei spaltenzentriert; ist $\vec{1}_m$ der m -dimensionale Einsvektor, dann ist $\vec{1}_m'X = \vec{0}' = (0, \dots, 0)$, so folgt

$$\vec{1}'XT_r = \vec{1}'L_r = \vec{0}', \quad (2.86)$$

d.h. die Spaltensummen von L_r sind ebenfalls alle gleich Null, so dass wegen (2.82) $\lambda_k = \mathbf{L}'_k \mathbf{L}_k = \|\mathbf{L}_k\|^2$ proportional zur Varianz der Komponenten von \mathbf{L}_k ist; der Proportionalitätsfaktor ist $1/m$ (oder $1/(m-1)$ bei kleinen Stichproben); die λ_k sind gleichzeitig die Eigenwerte von $X'X$. λ_1 ist dann proportional zur Varianz der Koordinaten der Fälle auf der ersten Achse des neuen Koordinatensystems L ; \mathbf{L}_1 repräsentiert dann eine latente Dimension, die maximal zwischen den Fällen diskriminiert.

2.5.6 Die Singularwertzerlegung einer Matrix

Die Betrachtungen des vorangegangenen Absatzes können in eine besondere Form gebracht werden. Es sei X wieder eine beliebige $(m \times n)$ -Matrix mit dem Rang $r \leq \min(m, n)$. Zunächst werden die Eigenvektorgleichungen für $X'X$ und XX' vorangestellt:

$$(X'X)T_r = T_r\Lambda_{r1} \quad (2.87)$$

$$(XX')Q_r = Q_r\Lambda_{r2}; \quad (2.88)$$

es werden zunächst also nur die zu Eigenwerten ungleich Null korrespondierenden Eigenvektoren betrachtet. Die Diagonalmatrizen Λ_{r1} und Λ_{r2} sind $(r \times r)$ -Matrizen, da $X'X$ und XX' denselben Rang haben (vergl. (2.40), Seite 41). Dann gilt der

Satz 2.17 *Es sei entsprechend (2.85) $XT_r = L_r$, wobei T_r die in (2.87) eingeführte Matrix der Eigenvektoren von $X'X$ sei. Die von Null verschiedenen Eigenwerte von $X'X$ und XX' sind identisch, d.h. es gilt $\Lambda_{r1} = \Lambda_{r2}$, und die normierten Spaltenvektoren von $S_r = L_r\Lambda_{r1}^{-1/2}$ sind die Eigenvektoren von XX' , d.h. es gilt $S_r = Q_r$.*

Beweis: Die Matrix L_r ist sicher orthogonal, denn

$$L_r'L_r = T_r'X'XT_r = T_r'T_r\Lambda_{r1}T_r'T_r = \Lambda_{r1} = \text{diag}(\lambda_1, \dots, \lambda_r). \quad (2.89)$$

Insbesondere gilt

$$\mathbf{L}_k'\mathbf{L}_k = \|\mathbf{L}_k\|^2 = \lambda_k, \quad k = 1, \dots, r \quad (2.90)$$

d.h. für die Länge von \mathbf{L}_k gilt $\|\mathbf{L}_k\| = \sqrt{\lambda_k}$. \mathbf{L}_k kann normiert werden, indem \mathbf{L}_k mit $1/\sqrt{\lambda_k}$ multipliziert wird: ist \mathbf{s}_k der normierte Vektor, so gilt

$$\mathbf{s}_k = \frac{1}{\sqrt{\lambda_k}}\mathbf{L}_k, \text{ bzw. } S_r = L_r\Lambda_{r1}^{-1/2}, \quad (2.91)$$

S_r die Matrix der normierten Vektoren. Dann läßt sich X in der Form

$$X = S_r\Lambda_{r1}^{1/2}T_r' \quad (2.92)$$

darstellen und für XX' erhält man den Ausdruck

$$XX' = S_r\Lambda_{r1}^{1/2}T_r'T_r\Lambda_{r1}^{1/2}S_r' = S_r\Lambda_{r1}S_r'. \quad (2.93)$$

S_r ist orthonormal, wegen der Normiertheit und der Orthogonalität von L_r . Dann folgt nach Multiplikation dieser Gleichung von rechts mit S_r

$$(XX')S_r = S_r\Lambda_{r1}, \quad (2.94)$$

d.h. S_r enthält die zu Eigenwerten ungleich Null korrespondierenden Eigenvektoren von XX' . Der Vergleich mit (2.88) zeigt, dass $Q_r = S_r$ und $\Lambda_{r2} = \Lambda_{r1}$.

□

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 (2.92) liefert bereits einen Ansatz dafür; sie ist als *Singularwertzerlegung* (SVD) der Matrix X bekannt. Bevor dieser Aspekt von (2.92) 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 der Matrix XX' , d.h. Q enthält auch die $n - r$ Eigenvektoren von XX' , die zu den Eigenwerten gleich Null von XX' korrespondieren, falls $r < \min(m, n)$. 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

$$\Sigma = \Lambda^{1/2} = \begin{pmatrix} \Lambda_r^{1/2} & 0 \\ 0 & 0 \end{pmatrix}, \quad (2.95)$$

wobei Λ_r die von Null verschiedenen Eigenwerte von $X'X$ bzw. XX' 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*

$$X = Q\Sigma T' \quad (2.96)$$

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 Singulärwerte.

Anmerkungen:

1. Nach Definition 2.6, Seite 51, repräsentiert die Matrix T der Eigenvektoren von $X'X$ eine Hauptachsentransformation der Spalten von X . Die Singularwertzerlegung von X ist äquivalent einer Hauptachsentransformation, denn $XT = Q\Lambda^{1/2} = L$. In Abschnitt 2.5.7 wird dieser Aspekt der SVD ausführlicher elaboriert.
2. Im Satz 2.1 (Seite 39) wurde ausgesagt, dass jede $(m \times n)$ -Matrix X als Produkt zweier Matrizen U und V' dargestellt werden kann, $X = UV'$, wobei $\text{rg}(X) = \text{rg}(U) = \text{rg}(V)$. Die SVD (2.96) ist ein solches Produkt; man kann entweder $U = Q$ und $V' = \Sigma T'$ bzw. $V = T\Sigma$ oder $U = Q\Sigma$ und $V = T$ setzen. Die Spaltenvektoren von X sind Linearkombinationen von U und die Zeilenvektoren sind Linearkombinationen der Zeilen von V' .
3. 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.

4. Die SVD ist nicht an eine Spaltenzentrierung oder Standardisierung der Matrix X gebunden; eine SVD kann für eine beliebige Matrix X bestimmt werden. \square

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

$$XT = Q\Sigma, \quad (2.97)$$

und von links mit Q' ergibt sich $Q'X = \Sigma T'$, oder

$$X'Q = T\Sigma, \quad (2.98)$$

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

$$X\mathbf{t}_k = \sigma_k\mathbf{q}_k \quad (2.99)$$

$$X'\mathbf{q}_k = \sigma_k\mathbf{t}_k, \quad k = 1, \dots, n \quad (2.100)$$

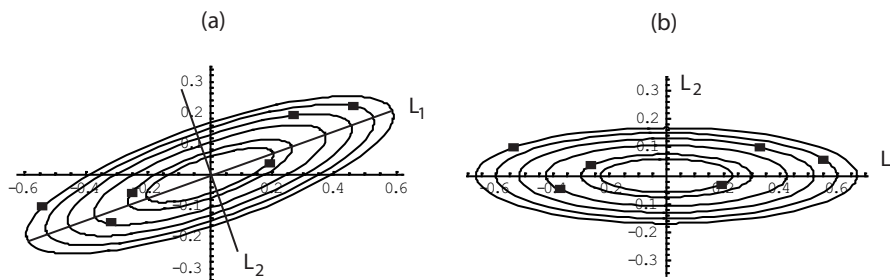
erhält. Diese beiden Gleichungen zeigen die Beziehungen zwischen den \mathbf{t}_k und den \mathbf{q}_k : X – als Transformationsmatrix gesehen – bildet \mathbf{t}_k auf $\sigma_k\mathbf{q}_k$ ab, und X' bildet \mathbf{q}_k auf $\sigma_k\mathbf{t}_k$ ab: \mathbf{q}_k ist eine Linearkombination der *Spaltenvektoren*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

Die SVD kann über die dyadischen Produkte der Spaltenvektoren von Q und T ausgedrückt werden:

$$X = \sigma_1\mathbf{q}_1\mathbf{t}'_1 + \sigma_2\mathbf{q}_2\mathbf{t}'_2 + \dots + \sigma_n\mathbf{q}_n\mathbf{t}'_n = \sum_{k=1}^n \sigma_k\mathbf{q}_k\mathbf{t}_k. \quad (2.101)$$

Anmerkung: In Abschnitt 3.3.4, Seite 84, wird der Satz von Courant-Fischer durch Differentiation der quadratischen Form (2.76), Seite 57 (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 $L'_1L_1 = \Lambda_1$, 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

Abbildung 8: Punktekonfiguration und Ellipsen. Im Anhang, Abschnitt 3.2 wird die Konstruktion dieser Ellipsen näher erläutert.



2.5.7 Ellipsoide und die Konfiguration der Fälle

Es soll noch gezeigt werden, dass zu jedem Punkt der Punktekonfiguration der Fälle ein Ellipsoid existiert, auf dem der jeweilige Punkt liegt; dieser Sachverhalt setzt nicht voraus, dass die Punktekonfiguration ellipsoidal ist, vergl. die Anmerkung zur Punktekonfiguration weiter unten. Aus $XT = L$ folgt $X = LT'$, wenn T eine Rotation repräsentiert, so dass $X' = TL'$ folgt. Es sei $\tilde{\mathbf{x}}_i$ der i -te Spaltenvektor von X' (d.h. der i -te Zeilenvektor von X) und $\tilde{\mathbf{L}}_i$ sei der korrespondierende Spaltenvektor von L' (der i -te Zeilenvektor von L). Dann folgt $\tilde{\mathbf{x}}_i = T\tilde{\mathbf{L}}_i$. Die Komponenten von $\tilde{\mathbf{L}}_i$ repräsentieren die Koordinaten des i -ten Falles im Raum der latenten Variablen, und die Komponenten von $\tilde{\mathbf{x}}_i$ sind die Messungen der n Variablen für den i -ten Fall. T' transformiert (d.h. rotiert) den Vektor $\tilde{\mathbf{x}}_i$ in den Vektor $\tilde{\mathbf{L}}_i$, d.h. $T\tilde{\mathbf{x}}_i = \tilde{\mathbf{L}}_i$.

Multipliziert man nun $X'X$ von links mit $\tilde{\mathbf{x}}_i'$ und von rechts mit $\tilde{\mathbf{x}}_i$, so erhält man wegen $X'X = T\Lambda T'$ und $T'\tilde{\mathbf{x}}_i = \tilde{\mathbf{L}}_i$ die Beziehung

$$\tilde{\mathbf{x}}_i'(X'X)\tilde{\mathbf{x}}_i = \tilde{\mathbf{x}}_i'T\Lambda T'\tilde{\mathbf{x}}_i = \tilde{\mathbf{L}}_i'\Lambda\tilde{\mathbf{L}}_i = k_i, \quad i = 1, \dots, m \quad (2.102)$$

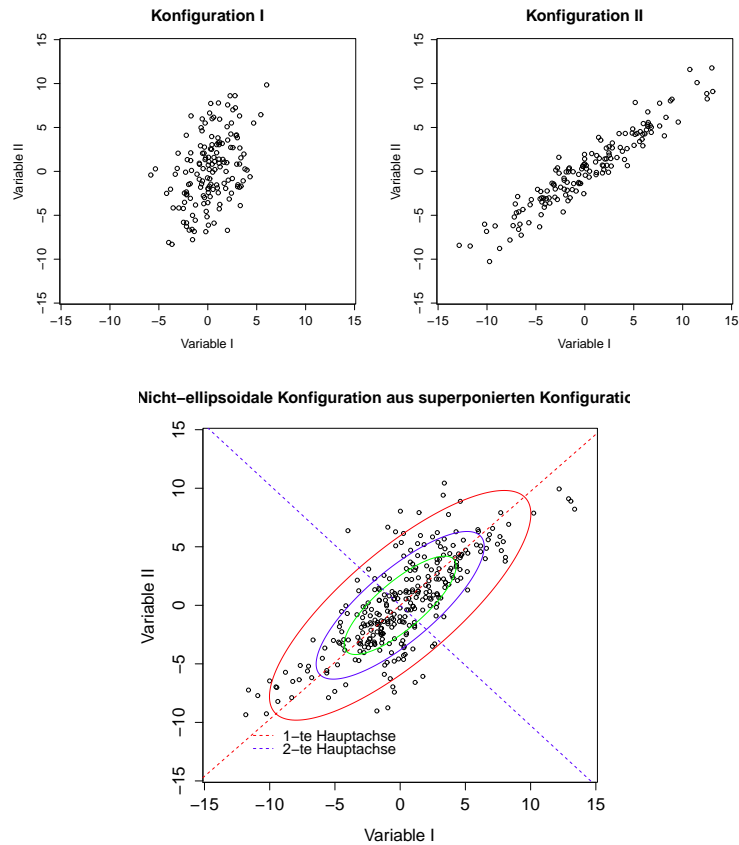
k_i ist eine für $\tilde{\mathbf{x}}_i$ charakteristische Konstante. $\mathbf{x}'(X'X)\mathbf{x} = k_i$ ist eine quadratische Form und die Menge der n -dimensionalen Vektoren, die dieser Gleichung genügen, bilden ein Ellipsoid. Da $\tilde{\mathbf{x}}_i$ ein Element dieser Menge ist, liegt der Endpunkt von $\tilde{\mathbf{x}}_i$ auf diesem Ellipsoid.

Rechnerisch ergibt sich der Wert von k_i , indem man $\tilde{\mathbf{x}}_i'(X'X)\tilde{\mathbf{x}}_i$ einfach ausrechnet. Damit hat man zwei Ellipsoide definiert (vergl. Abschnitt 2.5):

$$\mathcal{E}_{i1} = \{\tilde{\mathbf{x}}|\tilde{\mathbf{x}}'(X'X)\tilde{\mathbf{x}} = k_i\}, \quad \mathcal{E}_{i2} = \{\tilde{\mathbf{L}}|\tilde{\mathbf{L}}'\Lambda\tilde{\mathbf{L}} = k_i\}. \quad (2.103)$$

$\tilde{\mathbf{x}}_i$ und $\tilde{\mathbf{L}}_i$ sind spezielle Elemente von \mathcal{E}_{i1} bzw. \mathcal{E}_{i2} . \mathcal{E}_{i1} ist ein orientiertes, durch $X'X$ definiertes Ellipsoid, das die Punktekonfiguration der Fälle in den ursprünglichen Koordinaten ($\tilde{\mathbf{x}}_i$) beschreibt, und \mathcal{E}_{i2} ist ein achsenparalleles Ellipsoid, das die Fälle in den Koordinaten $\tilde{\mathbf{L}}_i$ beschreibt (s. Abschnitt 2.5.2). Die Spaltenvektoren von T definieren die Orientierung von \mathcal{E}_{i1} : \mathbf{t}_1 definiert die Orientierung der

Abbildung 9: Superponierte Punktekonfigurationen und Ellipsen



ersten Hauptachse, \mathbf{t}_2 die der zweiten Hauptachse, etc. Abbildung 8 illustriert die in (2.103) definierten Ellipsoide.

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 nicht ellipsoid ist kann stets eine Menge von Ellipsoiden gefunden werden, so dass jeder Fall auf einem Ellipsoid liegt, – einfach weil $X'X$ stets eine Menge von Ellipsoiden definiert. Dieser Fall kann eintreten, wenn sich die Stichprobe der Fälle aus Stichproben aus verschiedenen Populationen zusammensetzt. In Abbildung 9 wird dieser Sachverhalt illustriert.

2.5.8 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

$$\mathbf{c}_j = t_{1j}\mathbf{b}_1 + t_{2j}\mathbf{b}_2 + \dots + t_{rj}\mathbf{b}_r, \quad (2.104)$$

oder allgemein in Matrixform

$$C = BT \quad (2.105)$$

– 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

$$CT^{-1} = BT T^{-1} = B \quad (2.106)$$

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 2.6, Seite 45). 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 51). Abbildung 7 illustriert die Hauptachsentransformation für den 2-dimensionalen Fall.

2.5.9 Die Pseudoinverse einer Matrix

Es werde das Gleichungssystem $\mathbf{Ax} = \mathbf{y}$ 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 2.11 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

$$A'A\mathbf{x} = A'\mathbf{y} \Rightarrow \mathbf{x} = (A'A)^{-1}A'\mathbf{y}. \quad (2.107)$$

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

$$(A'A)^{-1}A'\mathbf{y} = \begin{cases} \mathbf{x}, & \mathbf{y} \in \mathcal{L}(A), \\ \hat{\mathbf{x}}, & \mathbf{y} \notin \mathcal{L}(A). \end{cases} \quad (2.108)$$

(vergl. Abschnitt 3.3.3, Gleichung (3.18), Seite 83).

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 . \square

Der folgende Ansatz, eine Pseudoinverse zu definieren, gilt auch für den Fall $\text{rg}(A) = r < \min(m, n)$ (Stewart (1973)). Es sei

$$A = Q\Sigma T', \quad \Sigma = \begin{pmatrix} \Lambda_r^{1/2} & 0 \\ 0 & 0 \end{pmatrix} \quad (2.109)$$

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

$$A^+ = A' = T \begin{pmatrix} \Lambda_r^{-1/2} & 0 \\ 0 & 0 \end{pmatrix} Q' \quad (2.110)$$

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. (2.110) definiert tatsächlich eine Pseudoinverse.

2.5.10 Faktorwerte und Ladungen

Die in Gleichung (2.96) definierte SVD von X kann in zwei möglichen Varianten interpretiert werden:

$$X = Q\Lambda^{1/2}T' = \begin{cases} LT', & L = Q\Sigma \\ QA', & A = T\Sigma, \end{cases} \quad \Sigma = \Lambda^{1/2}. \quad (2.111)$$

Definition 2.12 *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 der k -te Spaltenvektor von L , \mathbf{a}_k der k -te Spaltenvektor von A . Die Komponenten ℓ_{ik} von \mathbf{L}_k , $i = 1, \dots, m$, heißen Faktorwerte der Fälle auf der k -ten latenten Variablen, die Komponenten a_{jk} von \mathbf{a}_k , $j = 1, \dots, n$, heißen Ladungen der Variablen auf der k -ten latenten Variablen,.

Anmerkung: Gelegentlich werden in der Literatur auch die Komponenten des Spaltenvektors \mathbf{a}_k von Q als Faktorwerte bezeichnet; aus dem jeweiligen Kontext wird im Allgemeinen klar, was genau mit einem Faktorwert gemeint ist. \square

Die Gleichungen (2.99) und (2.100) auf Seite 62 zeigen die Beziehungen zwischen Ladungen und Faktorscores auf.

In der Definition 2.12 ist von den Ladungen der Variablen und den Faktorscores der Fälle die Rede. Es wird zunächst verdeutlicht, dass die Matrix A in der Tat die Variablen die Matrix L tatsächlich die Fälle charakterisiert.

Faktorwerte: Aus $X = LT'$ folgt wegen der Orthonormalität von T die Beziehung

$$XT = L, \quad \text{d.h. } X\mathbf{t}_k = \mathbf{L}_k, \quad k = 1, \dots, n \quad (2.112)$$

L ist eine $(m \times n)$ -Matrix; die k -te Spalte von L ist eine Linearkombination der Spaltenvektoren von X , wobei die Koeffizienten durch die Komponenten der k -ten Spalte von T gegeben sind:

$$\mathbf{L}_k = X\mathbf{t}_k = t_{1k}\mathbf{x}_1 + t_{2k}\mathbf{x}_2 + \dots + t_{nk}\mathbf{x}_n = \begin{pmatrix} \ell_{1k} \\ \ell_{2k} \\ \vdots \\ \ell_{mk} \end{pmatrix} \quad (2.113)$$

Der Spaltenvektor \mathbf{x}_j von X enthält die Messwerte der m Fälle für die j -te Variable, und die Komponenten von \mathbf{L}_k sind "Messwerte" der Fälle auf der k -ten latenten Variablen. Deswegen kann man sagen, dass L die Fälle repräsentiert; Q repräsentiert sie in normierter Form. Die Komponente ℓ_{ik} von \mathbf{L}_k ist der "Messwert", also der Faktorwert des i -ten Falls für die k -te latente Variable; nach Definition des Skalarprodukts hat man

$$\ell_{ik} = \tilde{\mathbf{x}}_i' \mathbf{t}_k = \sum_{j=1}^n x_{ij} t_{jk} = \|\tilde{\mathbf{x}}_i\| \|\mathbf{t}_k\| \cos \theta_{ik}, \quad (2.114)$$

Je kleiner der Winkel zwischen dem 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 $\cos \theta_{ik} = 1$.

Die Komponenten x_{ij} von $\tilde{\mathbf{x}}_i$ sind die Messwerte des i -ten Falls für die $j = 1, \dots, n$ Variablen. Die Komponente t_{jk} repräsentiert offenbar die j -te Variable auf der k -ten latenten Variablen; die Komponenten des k -ten Spaltenvektors \mathbf{t}_k von T repräsentieren die Variablen V_1 bis V_n auf der k -ten latenten Dimension. ℓ_{ik} wird um so größer, je mehr die Messwerte x_{ij} des i -ten Falles mit den Repräsentationen t_{jk} der Variablen auf der k -ten latenten Dimension "korrelieren" (wobei der Ausdruck 'korrelieren' eher metaphorisch gemeint ist, es handelt sich bei ℓ_{ik} ja nicht um eine Korrelation im strengen Sinne). ℓ_{ik} wird maximal, wenn $\tilde{\mathbf{x}}_i$ und \mathbf{t}_k dieselbe Orientierung haben (dann ist ja $\theta_{ik} = 0$).

Ladungen: Die Spaltenvektoren von \mathbf{t}_k von T enthalten zwar Repräsentationen der Variablen auf der k -ten latenten Dimension, die aber wegen der Normiertheit dieser Vektoren für die Interpretation der Daten weniger geeignet ist als die skalierte Form $A = T\Lambda^{1/2}$, wie in den folgenden Betrachtungen deutlich werden

wird. Es werde also der Fall $X = QA'$ betrachtet. Es folgt $X' = AQ'$ und die Multiplikation von rechts mit Q liefert wegen $Q'Q = I$

$$X'Q = A. \quad (2.115)$$

Diese Gleichung bedeutet, dass die Spalten von A Linearkombinationen der Spaltenvektoren $\tilde{\mathbf{x}}_i$ von X' , also der n -dimensionalen Zeilenvektoren von X sind. Es seien \mathbf{q}_k der k -te Spaltenvektor von Q und \mathbf{a}_k der k -te Spaltenvektor von A ; dann gilt

$$\mathbf{a}_k = X'\mathbf{q}_k = q_{1k}\tilde{\mathbf{x}}_1 + q_{2k}\tilde{\mathbf{x}}_2 + \cdots + q_{mk}\tilde{\mathbf{x}}_m = \begin{pmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{mk} \end{pmatrix}, \quad (2.116)$$

d.h. für die Ladung a_{jk} gilt – analog zu den Faktorscores ℓ_{ik} –

$$a_{jk} = \mathbf{x}'_j\mathbf{q}_k = \sum_{i=1}^m x_{ij}q_{ik} = \|\mathbf{x}_j\|\|\mathbf{q}_k\| \cos \theta_{jk}. \quad (2.117)$$

\mathbf{x}_j der j -te Spaltenvektor von X . Analog zu den Betrachtungen zu den Faktorwerten ℓ_{ik} ergibt sich, dass a_{jk} um so größer ist, je kleiner der Winkel θ_{jk} zwischen dem die j -te Variable repräsentierenden Vektor \mathbf{x}_j und dem die k -te latente Variable \mathbf{q}_k ist, d.h. je mehr die Messwerte x_{ij} der Fälle für die j -te Variable einerseits und der Repräsentation der Fälle auf der k -ten latenten Dimensionen mit einander "korrelieren". Ist $x_{ij} = z_{ij}$, d.h. ist $X = Z$, so bedeutet (2.117), dass a_{jk} tatsächlich gleich der Korrelation der j -ten gemessenen Variablen mit der k -ten latenten Dimension ist. (2.117) ist übrigens ein Beispiel für eine orthonormale Basisentwicklung (vergl. Gleichung (1.61), Seite 28).

Ladungen und die Korrelationen zwischen den Variablen: Aus $X = QA'$ folgt $X'X = AQ'QA'$, und wegen $Q'Q = I$ gilt

$$X'X = AA'. \quad (2.118)$$

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

$$R = \frac{1}{m}X'X = \frac{1}{m}AA', \quad X = Z \quad (2.119)$$

die Matrix der Korrelationen zwischen den Variablen. Es sei \mathbf{r}_k der k -te Spaltenvektor von R ; er ist eine Linearkombination der Spaltenvektoren von A , mit den Komponenten des k -ten Spaltenvektors $\tilde{\mathbf{r}}_k$ von A' (k -ter Zeilenvektor $(r_{k1}, r_{k2}, \dots, r_{kn})$ von A) als Koeffizienten:

$$\mathbf{r}_k = \frac{1}{m}A\tilde{\mathbf{r}}_k = \frac{1}{m}(r_{k1}\mathbf{a}_1 + r_{k2}\mathbf{a}_2 + \cdots + r_{kn}\mathbf{a}_n) = \frac{1}{m} \begin{pmatrix} r_{1k} \\ r_{2k} \\ \vdots \\ r_{nk} \end{pmatrix} \quad (2.120)$$

Die j -te Komponente r_{jk} von \mathbf{r}_k ist die Korrelation zwischen den Variablen V_j und V_k und ergibt sich als Skalarprodukt zwischen den Zeilenvektoren $\tilde{\mathbf{a}}_j$ und $\tilde{\mathbf{a}}_k$ von A :

$$r_{jk} = \frac{1}{m} \tilde{\mathbf{a}}_j' \tilde{\mathbf{a}}_k = \frac{1}{m} \sum_{u=1}^n a_{uj} a_{uk} = \frac{1}{m} \|\tilde{\mathbf{a}}_j\| \|\tilde{\mathbf{a}}_k\| \cos \theta_{jk}, \quad (2.121)$$

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 $\theta_{jj} = 0$)

$$r_{jj} = 1 = \frac{1}{m} \|\tilde{\mathbf{a}}_j\|^2 = \frac{1}{m} \sum_{u=1}^n a_{ju}^2, \quad j = 1, \dots, n \quad (2.122)$$

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

$$\|\mathbf{a}_k\|^2 = \sum_{jh=1}^n a_{jk}^2 = \sum_{j=1}^n \lambda_k t_{jk}^2 = \lambda_k \sum_{j=1}^n t_{jk}^2 = \lambda_k, \quad (2.123)$$

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: $\vec{1}$ 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

$$\|\mathbf{L}_k\|^2 = \sum_{i=1}^m \ell_{ik}^2 = \sum_{i=1}^m \lambda_k q_{ik}^2 = \lambda_k, \quad (2.124)$$

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 (2.123) hat man die Beziehung

$$\|\mathbf{L}_k\|^2 = \|\mathbf{a}_k\|^2 = \lambda_k, \quad k = 1, \dots, n \quad (2.125)$$

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

$$\pi_k = \frac{\lambda_k}{\sum_{j=1}^n \lambda_j} \quad (2.126)$$

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.

Anmerkung zu A: Auf Seite ?? wurde (??), d.h. $T'T = I \Rightarrow TT' = I$, bewiesen, d.h. orthogonale Spaltenvektoren einer Matrix implizieren orthogonale Zeilenvektoren. Wie oben bereits angemerkt wurde, repräsentieren die Elemente t_{jk} die Variablen (j) auf den latenten Variablen (k). Die Spalten von T entsprechen den möglichen latenten Variablen für X . Da $T'T = I$ die Eigenschaft $TT' = I$ impliziert, dass das Skalarprodukt zweier Zeilenvektoren von T gleich Null ist. Andererseits ist das Skalarprodukt zweier Zeilen von A entsprechend (2.121) gleich mr_{uv} , was natürlich daran liegt, dass die Spalten von A die mit $\sqrt{\lambda_j}$ skalierten Spaltenvektoren von T sind, d.h.

$$\mathbf{a}_k = \sqrt{\lambda_k} \mathbf{t}_k, \quad k = 1, \dots, n \quad (2.127)$$

Ein Zeilenvektor von A ist dann durch

$$\tilde{\mathbf{a}}_j = (\sqrt{\lambda_1} t_{j1}, \sqrt{\lambda_2} t_{j2}, \dots, \sqrt{\lambda_n} t_{jn})', \quad (2.128)$$

gegeben und man hat

$$r_{jk} = \frac{1}{m} \tilde{\mathbf{a}}_j' \tilde{\mathbf{a}}_k = \frac{1}{m} \sum_{u=1}^n \lambda_k t_{ju} t_{ku}. \quad (2.129)$$

Die Eigenwerte λ_k "verhindern" gewissermassen, dass $\tilde{\mathbf{a}}_u' \tilde{\mathbf{a}}_v = 0$.

Interpretation von Messwerten: Es sei x_{ij} der möglicherweise standardisierte Messwert beim i -ten Fall für die j -te Variable. Die SVD liefert einen Ausdruck für x_{ij} in Termen von latenten Variablen. Generell gilt $X = Q\Lambda^{1/2}T'$, so dass

$$x_{ij} = q_{i1} \sqrt{\lambda_1} t_{1j} + q_{i2} \sqrt{\lambda_2} t_{2j} + \dots + q_{in} \sqrt{\lambda_n} t_{nj} \quad (2.130)$$

geschrieben werden kann (vergl. auch (2.101)). In Anlehnung an faktorenanalytische Modelle kann man x_{ij} auch in der Form

$$x_{ij} = q_{i1} a_{j1} + q_{i2} a_{j2} + \dots + q_{in} a_{jn} \quad (2.131)$$

schreiben (vergl. (2.111), Seite 67). Es wird deutlich, dass sich die Faktorwerte q_{ik} und die Ladungen a_{jk} , $k = 1, \dots, n$, auf dieselben latenten Variablen beziehen. Natürlich kann auch die Form

$$x_{ij} = \ell_{i1}t_{1j} + \ell_{i2}t_{2j} + \dots + \ell_{in}t_{nj} \quad (2.132)$$

gewählt werden, wenn insbesondere die Interpretation der Fälle von Interesse ist. Hier werden die ℓ_{ik} als Faktorwerte bezeichnet, was anzeigt, dass der Begriff des Faktorwerte nicht ganz einheitlich verwendet wird. In jedem Fall ist damit die Ausprägung der k -ten latenten Variablen beim i -ten Fall gemeint, während t_{jk} bzw. a_{jk} als Maß für das Ausmaß, in dem die j -te Variable durch die k -te latente Variable bestimmt wird interpretiert werden kann. In jedem Fall ist x_{ij} ein Skalarprodukt von zwei Vektoren, von denen der eine die Ausprägung der latenten Variablen bei den Fällen und der andere die Maße, in dem die gemessenen Variablen die latenten Variablen erfassen repräsentieren. Weitere Details finden sich im Skriptum über die Hauptkomponentenanalyse.

x_{ij} ist ein Element der Matrix X ; analog zu (2.130) kann auch X als Summe von Matrizen dargestellt werden:

$$X = \sqrt{\lambda_1}\mathbf{q}_1\mathbf{t}'_1 + \sqrt{\lambda_2}\mathbf{q}_2\mathbf{t}'_2 + \dots + \sqrt{\lambda_n}\mathbf{q}_n\mathbf{t}'_n = \sum_{k=1}^n \sqrt{\lambda_k}\mathbf{q}_k\mathbf{t}'_k, \quad (2.133)$$

denn die $q_{ik}t_{1kj}$ sind ja Elemente der Matrizen, die als dyadische Produkte der \mathbf{q}_i und \mathbf{t}_k entstehen. Analog dazu findet man für $C = X'X$ wegen $C = T\Lambda T'$

$$C = \lambda_1\mathbf{t}_1\mathbf{t}'_1 + \lambda_2\mathbf{t}_2\mathbf{t}'_2 + \dots + \lambda_n\mathbf{t}_n\mathbf{t}'_n = \sum_{k=1}^n \lambda_k\mathbf{t}_k\mathbf{t}'_k. \quad (2.134)$$

Die Darstellung (2.134) verdeutlicht, dass wegen $\lambda_1 \geq \dots \geq \lambda_n$ die Summanden für größer werdenden k -Wert immer kleiner werden, so dass man u. U. die letzten Terme vernachlässigen kann (vergl. auch Satz 2.11, Seite 49), demzufolge der Rang von C (und damit von X) gleich der von Null verschiedenen Eigenwerte ist). Dies führt zum

Satz 2.18 (*Satz von Eckart & Young*) *Die Approximation*

$$X \approx X_r = Q_r\Lambda_r^{1/2}T'_r = \sum_{k=1}^r \sqrt{\lambda_k}\mathbf{q}_k\mathbf{t}'_k, \quad r < n \quad (2.135)$$

approximiert X im Sinne der Methode der Kleinsten Quadrate.

Beweis: Eckart & Young (1936); modernere und knappere Version: Golub & van Loan (2013), p. 79). Der Beweis macht vom Begriff der Norm einer Matrix Gebrauch, auf den in diesem Skript nicht eingegangen werden soll bzw. kann. \square

2.6 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

$$A\mathbf{x} = \mathbf{y}, \quad \mathbf{x} \in \mathbb{R}^n, \quad \mathbf{y} \in \mathbb{R}^m \quad (2.136)$$

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

$$A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \dots + x_n\mathbf{a}_n. \quad (2.137)$$

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*.

Definition 2.13 *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*

$$\text{kern}(A) = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \vec{0}\} \quad (2.138)$$

$$\mathcal{L}(A) = \{\mathbf{y} \in \mathbb{R}^m \mid A\mathbf{x} = \mathbf{y}\}. \quad (2.139)$$

kern(A) heißt Kern von A, und $\mathcal{L}(A)$ ist die lineare Hülle der Spaltenvektoren von A.

Anmerkungen:

1. In Definition 3.1, Punkte 4., Seite 89 ist der Begriff des Kerns einer Abbildung eingeführt worden. Die Matrix A definiert eine Abbildung, und (2.138) definiert damit den Kern einer Abbildung. $\text{kern}(A)$ 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 $b_1\mathbf{x}_1 + b_2\mathbf{x}_2 \in \text{kern}(A)$ ($b_1, b_2 \in \mathbb{R}$) gilt.
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 2.19 *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*

$$\mathcal{L}(A) = \mathcal{L}(\mathbf{q}_1, \dots, \mathbf{q}_r) \quad (2.140)$$

$$\text{kern}(A) = \mathcal{L}(\mathbf{t}_{r+1}, \dots, \mathbf{t}_n), \quad s = \min(m, n) \quad (2.141)$$

$$\text{rg}(\text{kern}(A) + \text{rg}(\mathcal{L}(A))) = \min(m, n) \quad (2.142)$$

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 (2.140) gelten muß.

Der Kern von A sind alle n -dimensionalen Vektoren \mathbf{x} , für die $A\mathbf{x} = \vec{0}$ gilt. Die Spaltenvektoren von T bilden eine Basis des \mathbb{R}^n , so dass allgemein

$$\mathbf{x} = c_1 \mathbf{t}_1 + \dots + c_r \mathbf{t}_r + c_{r+1} \mathbf{t}_{r+1} + \dots + c_n \mathbf{t}_n$$

geschrieben werden kann, wobei die $c_j \in \mathbb{R}$ geeignete 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. (2.99), Seite 62), 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 $c_1 = \dots = c_r = 0$. 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 Komplementärraums sein. Die $\mathbf{t}_{r+1}, \dots, \mathbf{t}_n$ sind eine Basis für diesen Komplementärraum, so dass man

$$\mathbf{x} = c_{r+1} \mathbf{t}_{r+1} + \dots + c_n \mathbf{t}_n,$$

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 (2.99), Seite 62, 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 (2.141).

Die Gleichung (2.142) 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 2.19 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 3.4. \square

Die allgemeine Lösungsmenge wird im folgenden Satz spezifiziert:

Satz 2.20 *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*

$$\mathcal{L} = \{\mathbf{x}_0 + \mathbf{x} \mid \mathbf{x} \in \mathcal{L}_{n-r}\} \quad (2.143)$$

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

$$A(\mathbf{x}_1 + \mathbf{x}) = A\mathbf{x}_1 = \mathbf{y}$$

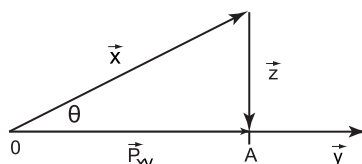
gelten, so dass $\mathbf{x}_1 + \mathbf{x} \in \mathcal{L}$ liegt. \square

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

$$\mathbf{x} = A^{-1}\mathbf{y}, \quad \mathbf{y} \in \mathcal{L}(A). \quad (2.144)$$

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

Abbildung 10: Orthogonale Projektion des Vektors \mathbf{x} auf einen Vektor \mathbf{y} bzw. auf eine Gerade



$A'A\mathbf{x} = A'\mathbf{y}$ folgt, und da $A'A$ den gleichen Rang wie A hat (s. (2.40), Seite 41) existiert die zu $A'A$ inverse Matrix $(A'A)^{-1}$, so dass

$$\mathbf{x} = (A'A)^{-1}A'\mathbf{y} \quad (2.145)$$

resultiert. Die Matrix $(A'A)^{-1}A'$ ist eine Pseudoinverse für A , s. Seite 65. Sind die Komponenten von \mathbf{y} Messwerte, so sind sie üblicherweise durch Messfehler kontaminiert, so dass $\mathbf{y} \notin \mathcal{L}(A)$. (2.145) liefert dann keine Lösung \mathbf{x} , die allen m Gleichungen genügt. (2.145) ist die bekannte Kleinste-Quadrate-Schätzung $\hat{\mathbf{x}}$ für \mathbf{x} , vergl. (3.18), Seite 83, und Abschnitt 2.5.9.

Für den Fall $r < n$ liefert (2.143) 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}

$$x_j = \frac{|A_j|}{|A|}, \quad j = 1, \dots, n \quad (2.146)$$

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 3.6, 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.

2.7 Projektionen

2.7.1 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 10. \mathbf{x} und \mathbf{y} schließen den Winkel θ ein, und es ist

$$\vec{P}_{yx} = a\mathbf{y} \quad (2.147)$$

der Vektor, der sich ergibt, wenn \mathbf{x} auf \mathbf{y} projiziert wird; in Abbildung 10 ist $a < 1$, aber $a > 1$ ist möglich. Nun ist einerseits

$$\mathbf{z} = \vec{P}_{yx} - \mathbf{x} = a\mathbf{y} - \mathbf{x}, \quad (2.148)$$

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

$$a = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{y}\|^2} \Rightarrow \vec{P}_{xy} = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{y}\|^2}\mathbf{y} \quad (2.149)$$

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

$$\|\vec{P}_{xy}\| = a\|\mathbf{y}\| \quad (2.150)$$

gegeben. Im folgenden Abschnitt wird dieser Aspekt von Projektionen weiter elaboriert.

2.7.2 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 der i -ten Zeile ξ'_i von X und der k -ten Spalte \mathbf{t}_k von T , d.h. es ist

$$\ell_{ik} = \xi'_i\mathbf{t}_k. \quad (2.151)$$

Hier wird der Zeilenvektor als *Spaltenvektor* aufgefasst, um die Schreibweise des Skalarprodukts beizubehalten. \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 ξ_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 10. 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 ξ_i und \mathbf{t}_k charakteristisch ist. Nach (2.149) gilt nun

$$a_{ik} = \frac{\xi'_i\mathbf{t}_k}{\|\mathbf{t}_k\|^2} = \xi'_i\mathbf{t}_k, \quad (2.152)$$

da ja $\|\mathbf{t}_k\| = 1$, und nach (2.150) hat man dann

$$\|\vec{P}_{ik}\| = a_{ik}\|\mathbf{t}_k\| = \xi'_i\mathbf{t}_k, \quad (2.153)$$

d.h. wegen (2.151) gilt

$$\ell_{ik} = \|\vec{P}_{ik}\|, \quad (2.154)$$

so dass die Koordinate ℓ_{ik} die Länge der Projektion des Vektors ξ_i auf die k -te Hauptachse des Ellipsoids ist.

3 Anhang

3.1 Steinerscher Austauschatz und das Fundamental-Lemma

Satz 3.1 *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*

$$\mathbf{v} = a_1 \mathbf{v}_1 + \dots + a_n \mathbf{v}_n \quad (3.1)$$

Jeder Vektor $\mathbf{v}_j \in \mathcal{B}$ mit $a_j \neq 0$ kann durch \mathbf{v} ausgetauscht werden, und

$$\{\mathbf{v}_1, \mathbf{v}_{j-1}, \mathbf{v}, \mathbf{v}_{j+1}, \dots, \mathbf{v}_n\} \quad (3.2)$$

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 \mathcal{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

$$b_1 a_1 \mathbf{v}_1 + (b_1 a_2 + b_2) \mathbf{v}_2 + \dots + (b_1 a_n + b_n) \mathbf{v}_n = \vec{0}.$$

Wegen der linearen Unabhängigkeit der \mathbf{v}_j folgt dann aber

$$b_1 a_1 = b_1 a_2 + b_2 = \dots = b_1 a_n + b_n = 0.$$

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

$$\mathbf{w} = c_1 \mathbf{v}_1 + \dots + c_n \mathbf{v}_n.$$

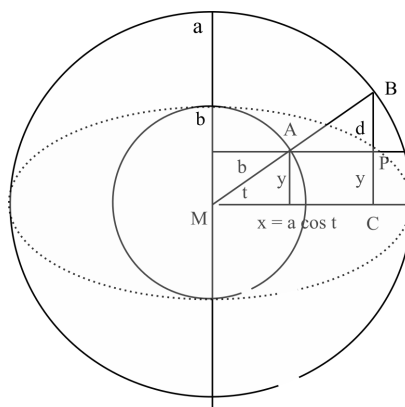
Aus (3.1) folgt

$$\mathbf{v}_1 = \frac{1}{a_1} (\mathbf{v} - a_2 \mathbf{v}_2 - \dots - a_n \mathbf{v}_n)$$

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).

Abbildung 11: Kreise mit den Radien a bzw. $b < a$ und zugehörige Ellipse (Menge der Punkte P) mit den Halbachsen a und b



Satz 3.2 *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 3.1 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 $m > n$, 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

3.2 Zur Berechnung von Ellipsen für eine Punktekongfiguration

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

$$\mathcal{E} : \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad x = x(t), \quad y = y(t) \quad (3.3)$$

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 (3.3) folgt. \square

Parameterdarstellung: Bei dieser Darstellung wird die Ellipse (3.3) als Menge der Punkte $P = P(t)$ für $0 \leq t \leq 2\pi$ definiert. Aus der Abbildung 11 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}

$$\mathcal{E} = \{\mathbf{p}(t) | \mathbf{p}(t) = (x(t), y(t))' = (a \cos t, b \sin t)\}. \quad (3.4)$$

$\mathbf{p}(t)$ der Vektor mit dem Anfangspunkt M und dem Endpunkt $P(t)$.

Die Gleichung (3.3) 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 (vergl. Abschnitt 2.5.5, Gleichung (??), Seite ??); die k -te Spalte \mathbf{L}_k hat die Länge $\sqrt{\lambda_k}$. Diese Forderung führt zu

$$a = \sqrt{\lambda_1}, \quad b = \sqrt{\lambda_2}. \quad (3.5)$$

In Bezug auf (3.4) 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 (3.3) äquivalent zu

$$\mathbf{y}'\Lambda^{-1}\mathbf{y} = 1. \quad (3.6)$$

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

$$y_{01} = \sqrt{\lambda_1}, \quad y_{02} = \sqrt{\lambda_2}.$$

3.3 Die Differentiation von Vektoren

3.3.1 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

$$\frac{d\mathbf{x}}{dx_j} = \begin{pmatrix} dx_1/dx_j \\ \vdots \\ dx_j/dx_j \\ \vdots \\ dx_n/dx_j \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{e}_j. \quad (3.7)$$

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 (3.7) 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

$$\frac{d\mathbf{y}}{d\mathbf{x}} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n] = A. \quad (3.8)$$

3.3.2 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)

$$\frac{\partial Q}{\partial x_j} = \mathbf{e}_j' C \mathbf{x} + \mathbf{x}' C \mathbf{e}_j \quad (3.9)$$

\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

$$\frac{\partial Q}{\partial x_j} = 2\mathbf{e}_j' C \mathbf{x}. \quad (3.10)$$

Fasst man die \mathbf{e}_j' zur Einheitsmatrix zusammen, so erhält man

$$\frac{\partial Q}{\partial \mathbf{x}} = 2C\mathbf{x}. \quad (3.11)$$

wegen $C' = C$. Damit hat man auch die Ableitung von $\|\mathbf{x}\|^2 = \mathbf{x}'\mathbf{x}$ nach \mathbf{x} gefunden, denn mit $C = I$ die Einheitsmatrix folgt aus (3.11)

$$\frac{\partial \mathbf{x}'\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{x}. \quad (3.12)$$

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 3.3.4 nachgelesen werden.

Zu maximieren sei

$$Q(\mathbf{x}) = \mathbf{x}'A\mathbf{x} + \lambda(\mathbf{x}'\mathbf{x} - a), \quad \mathbf{x}'\mathbf{x} - a = 0, \quad A' = A \quad (3.13)$$

Dann ist

$$\frac{dQ}{d\mathbf{x}} = 2A\mathbf{x} - 2\lambda\mathbf{x},$$

so dass

$$A\mathbf{x} = \lambda\mathbf{x}, \quad (3.14)$$

d.h. $Q(\mathbf{x})$ wird maximal, wenn \mathbf{x} ein Eigenvektor von A ist und $\lambda \in \mathbb{R}$ der zugehörige Eigenwert ist.

3.3.3 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

$$\mathbf{y} = X\mathbf{b} + \mathbf{e} \quad (3.15)$$

\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

$$\mathbf{e}'\mathbf{e} = (\mathbf{y} - X\mathbf{b})'(\mathbf{y} - X\mathbf{b}) = \mathbf{y}'\mathbf{y} - \mathbf{y}'X\mathbf{b} - \mathbf{b}'X'\mathbf{y} + \mathbf{b}'X'X\mathbf{b} \quad (3.16)$$

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 (3.8) ist $d(X\mathbf{b})/d\mathbf{b} = X$, und nach (3.11) 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

$$\left. \frac{d(\mathbf{e}'\mathbf{e})}{d\mathbf{b}} \right|_{\mathbf{b}=\hat{\mathbf{b}}} = 2(X'\mathbf{y} - (X'X)\mathbf{b}) = 0 \quad (3.17)$$

so dass $X'\mathbf{y} = (X'X)\hat{\mathbf{b}}$, d.h.

$$\hat{\mathbf{b}} = (X'X)^{-1}X'\mathbf{y}, \quad (3.18)$$

Einige Implikationen: Es sei $X = Q\Sigma T'$, $\Sigma = \Lambda^{-1/2}$, die SVD von X . Dann ist $X'X = T\Lambda T'$ und

$$(X'X)^{-1} = (T\Lambda T')^{-1} = (T')^{-1}\Lambda^{-1}T^{-1} = T\Lambda^{-1}T', \quad (3.19)$$

denn $T^{-1} = T'$ wegen der Orthonormalität von T . (3.18) impliziert dann

$$\hat{\mathbf{b}} = T\Lambda^{-1}T'T\Lambda^{1/2}Q'\mathbf{y} = T\Lambda^{-1/2}Q'\mathbf{y}. \quad (3.20)$$

Aber es ist $\mathbf{y} = X\mathbf{b} + \mathbf{e} = Q\Lambda^{1/2}Q'\mathbf{b} + \mathbf{e}$, so dass aus (3.20)

$$\hat{\mathbf{b}} = T\Lambda^{-1/2}Q'(Q\Lambda^{1/2}Q'\mathbf{b} + \mathbf{e}) = \mathbf{b} + T\Lambda^{-1/2}Q'\mathbf{e} \quad (3.21)$$

folgt, und schreibt man $T\Lambda^{-1/2}Q'$, indem man das dyadische Produkt anwendet, so ergibt sich (vergl. (2.101), Seite 62, wo der Fall $X = Q\Lambda^{1/2}T'$ betrachtet wird)

$$\hat{\mathbf{b}} = \mathbf{b} + \left(\sum_{k=1}^n \frac{\mathbf{t}_k \mathbf{q}'_k}{\sqrt{\lambda_k}} \right) \mathbf{e} \quad (3.22)$$

Der "wahre" Parametervektor \mathbf{b} und die Kleinste-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

$$\mathbf{c}_j = t_{1j}\lambda_1\mathbf{t}_1 + t_{2j}\lambda_2\mathbf{t}_2 + \dots + t_{nj}\lambda_n\mathbf{t}_n. \quad (3.23)$$

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:

$$\mathbf{y} = Q\Sigma T'\mathbf{b} + \mathbf{e} = Q\Sigma(T'\mathbf{b}) + \mathbf{e} = L\boldsymbol{\beta} + \mathbf{e}, \quad L = Q\Sigma \quad (3.24)$$

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.

3.3.4 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

$$f_x + f_y g' = 0 \tag{3.25}$$

$$g_x + g_y g' = 0 \tag{3.26}$$

Die bisher hergeleiteten Ableitungen enthalten noch die Ableitung g' von g . Um das Extremum zu 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

$$\begin{pmatrix} f_x \\ f_y \end{pmatrix} = \lambda \begin{pmatrix} g_x \\ g_y \end{pmatrix} \tag{3.27}$$

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 (3.27) zusammen mit der Bedingung $g(x, y) = 0$ führt sofort auf ein System von drei Gleichungen mit den drei Unbekannten x, y und λ :

$$f_x - \lambda g_x = 0 \tag{3.28}$$

$$f_y - \lambda g_y = 0 \tag{3.29}$$

$$g(x, y) = 0 \tag{3.30}$$

Diese Überlegungen müssen nicht immer explizit durchgeführt werden, denn sie implizieren die Möglichkeit, von vornherein die *Lagrange-Funktion* $L(x, y, \lambda)$ aufzustellen:

$$L(x, y, \lambda) = f(x, y) + \lambda g(x, y), \quad g(x, y) = 0. \tag{3.31}$$

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 (3.28), (3.29) und (3.30) heißen zusammen die *Lagrange-Multiplikatorenregel*, nach dem Mathematiker und Astronomen Jean-Louis Lagrange (1736 – 1813), der diese Regel 1788 herleitete.

Beispiel 3.1 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

$$\begin{aligned} -2x + \lambda 1 &= 0 \\ -\frac{2}{3}y + \lambda 1 &= 0 \\ x + y - 2 &= 0, \end{aligned}$$

woraus $x = 1/2$, $y = 3/2$ und $\lambda = -1$ folgt. \square

Beispiel 3.2 (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

$$\max_{\mathbf{x} \neq \vec{0}} \frac{\mathbf{x}' A \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \max_j \lambda_j = \lambda_1, \quad (3.32)$$

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

$$\min_{\mathbf{x} \neq \vec{0}} \frac{\mathbf{x}' A \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \min_j \lambda_j = \lambda_n \quad (3.33)$$

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

3.4 Alternativer Beweis von Satz 2.19

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 $U + V = \mathbb{R}$ und es folgt

$$\dim(U + V) = \dim(U) + \dim(V) = \dim \mathbb{R}^n = n. \quad (3.34)$$

Zur Bestimmung von $\dim(U)$ werde

$$A\mathbf{x} = x_1\mathbf{a}_1 + \cdots + x_r\mathbf{a}_r + x_{r+1}\mathbf{a}_{r+1} + \cdots + x_n\mathbf{a}_n = \vec{0}$$

betrachtet. Ohne Beschränkung der Allgemeinheit kann angenommen werden, dass die Spaltenvektoren $\mathbf{a}_1, \dots, \mathbf{a}_r$ die linear unabhängigen Vektoren von A sind. Schreibt man

$$x_1\mathbf{a}_1 + \cdots + x_r\mathbf{a}_r = -(x_{r+1}\mathbf{a}_{r+1} + \cdots + x_n\mathbf{a}_n),$$

und berücksichtigt man, dass die \mathbf{a}_{r+k} für $k = 1, \dots, n - r$ Linearkombinationen der \mathbf{a}_j , $j = 1, \dots, r$ sind, so dass

$$\mathbf{a}_{r+k} = \lambda_{k1}\mathbf{a}_1 + \cdots + \lambda_{kr}\mathbf{a}_r = \sum_{j=1}^r \lambda_{kj}\mathbf{a}_j \quad (3.35)$$

gelten muß, so hat man

$$\sum_{j=1}^r x_j\mathbf{a}_j = - \sum_{k=1}^{n-r} x_{r+k} \sum_{j=1}^r \lambda_{kj}\mathbf{a}_j. \quad (3.36)$$

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 (3.36) nimmt dann die Form

$$\sum_{k=1}^{n-r} x_{r+k} \sum_{j=1}^r \mu_{kj}\mathbf{a}_j = \sum_{j=1}^r \lambda_{kj}\mathbf{a}_j = - \sum_{j=1}^r x_j\mathbf{a}_j.$$

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

$$\mathbf{x}_k = (-\lambda_{k1}, \dots, -\lambda_{kr}, 0, \dots, 0, 1, 0, \dots, 0)', \quad k = 1, \dots, n-r \quad (3.37)$$

wobei die 1 an der $(r+k)$ -ten Stelle steht. In der Tat ist

$$A\mathbf{x}_k = -\lambda_{k1}\mathbf{a}_1 - \dots - \lambda_{kr}\mathbf{a}_r + \mathbf{a}_{r+k} = 0$$

wegen (3.35).

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 (3.34) (= 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

3.5 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:

$$f: \mathcal{M} \rightarrow \mathcal{N}, \quad x \mapsto y = f(x), \quad x \in \mathcal{M}, y \in \mathcal{N} \quad (3.38)$$

Man schreibt gelegentlich auch

$$f(\mathcal{M}) = \mathcal{N}. \quad (3.39)$$

$f(\mathcal{M})$ 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

$$\text{id}(\mathcal{M}) = \mathcal{M}. \quad (3.40)$$

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

$$I_n \mathbf{x} = \mathbf{x}, \quad \mathbf{x} \in \mathbb{R}^n. \quad (3.41)$$

Für eine Teilmenge von Abbildungen existiert die *inverse Abbildung* f^{-1} :

$$f(\mathcal{M}) = \mathcal{N}, \quad f^{-1}f(\mathcal{M}) = \mathcal{M} = f^{-1}(\mathcal{N}). \quad (3.42)$$

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

$$f(\mathcal{M}) = \{f(x), x \in \mathcal{M}\}. \quad (3.43)$$

Offenbar gilt $f(\mathcal{M}) \subseteq \mathcal{N}$.

Definition 3.1 *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 $\text{kern}(f) = \{x \in \mathcal{M} | f(x) = \vec{0}\}$ 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 $f^{-1}(\mathbf{y}) = \{\mathbf{x} \in \mathcal{M} | f(\mathbf{x}) = \mathbf{y}\}$ 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 3.2 *Es seien V und W Vektorräume und $f : V \rightarrow W$ sei eine Abbildung von V in W . f heißt linear bzw. homomorph, wenn*

$$f(\lambda\mathbf{v} + \mu\mathbf{w}) = \lambda f(\mathbf{v}) + \mu f(\mathbf{w}) \quad (3.44)$$

für $\lambda, \mu \in \mathbb{R}$ und für alle $\mathbf{v} \in V$ und $\mathbf{w} \in W$. Insbesondere heißt f isomorph, 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 3.1 von Abbildungen entspricht für $f \in \mathbb{R}^{m,n}$ der Nullvektor $\vec{0}$.

Satz 3.3 *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. \square

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

$$\lambda_f(\mathbf{x}_1) + \cdots + \lambda_n f(\mathbf{x}_n) = \vec{0}$$

gilt. Es wurde vorausgesetzt, dass f injektiv ist. Daraus folgt, dass

$$\lambda_1 \mathbf{x}_1 + \cdots + \lambda_n \mathbf{x}_n = \vec{0}$$

gelten muß, denn $\lambda_1 \mathbf{x}_1 + \cdots + \lambda_n \mathbf{x}_n$ ist ja das Urbild von f . Da die $\mathbf{x}_1, \dots, \mathbf{x}_n$ als linear unabhängig vorausgesetzt wurden, muß $\lambda_1 = \cdots = \lambda_n = 0$ gelten, und dann folgt sofort, dass auch die $f(\mathbf{x}_j)$ linear unabhängig sind. \square

Definition 3.3 *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 $f = A$. $\mathbf{e}_1, \dots, \mathbf{e}_m$ 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 3.3 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 $M\mathbf{y} = \mathbf{u}$ und $\mathbf{u} = \mathbf{u}$ folgt, dass $\mathbf{x} = \mathbf{y}$ ist, und aus $\mathbf{u} \neq \mathbf{v}$ folgt $\mathbf{x} \neq \mathbf{y}$. Die Frage nach der Injektivität ist also eine Frage nach der Eindeutigkeit der Abbildung. M definiert eine surjektive Abbildung, wenn für jeden Vektor $\mathbf{u} \in V_n$ eine Vektor $\mathbf{x} \in V_m$ existiert derart, dass $M\mathbf{x} = \mathbf{u}$, d.h. die Frage nach der Surjektivität ist die Frage, ob durch M alle Elemente von V_n bestimmt werden. M definiert eine bijektive Abbildung, wenn M eine sowohl injektive wie auch surjektive Abbildung definiert. Dies ist die Frage, ob eine surjektive Abbildung auch eindeutig ist. Offenbar hängen diese Eigenschaften von der Struktur der Matrix M ab. Was mit dem Begriff der Struktur einer Matrix genau gemeint ist, wird im Folgenden entwickelt.

Beispiel 3.4 Es sei

$$T = \begin{pmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{pmatrix} = \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix}. \quad (3.45)$$

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

3.6 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

$$\det(A) = |A| = \sum_{j=1}^n (-1)^{j+1} a_{1j} \det(A_{1j}), \quad (3.46)$$

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 (3.46) ist also rekursiv. Für eine (2×2) -Matrix erhält man dementsprechend

$$\begin{aligned} \det(A) &= \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = (-1)^{1+1} a_{11} a_{22} + (-1)^{2+1} a_{12} a_{21} \\ &= a_{11} a_{22} - a_{12} a_{21}. \end{aligned} \quad (3.47)$$

Für eine (3×3) -Matrix A erhält man

$$\begin{aligned} \det(A) &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = (-1)^{1+1} a_{11} (a_{22} a_{33} - a_{23} a_{32}) \\ &\quad + (-1)^{2+1} a_{12} (a_{11} a_{33} - a_{13} a_{31}) \\ &\quad + (-1)^{3+1} a_{13} (a_{11} a_{22} - a_{12} a_{21}), \end{aligned} \quad (3.48)$$

d.h.

$$\det(A) = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{11}a_{33} - a_{13}a_{31}) + a_{13}(a_{11}a_{22} - a_{12}a_{21}) \quad (3.49)$$

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

$$(A - \lambda I)\mathbf{t} = \begin{pmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{pmatrix} \mathbf{t} = \vec{0} \quad (3.50)$$

I die $(n \times n)$ -Einheitsmatrix. Dann ist

$$|A - \lambda I| = 0 = p(\lambda), \quad (3.51)$$

d.h. $p(\lambda)$ ist ein Polynom n -ten Grades in λ , wie man sich durch Anwendung der Definition (3.46) 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] Eckart, C., Young, G. (1936), The approximation of one matrix by another of lower rank. *Psychometrika*, 1 (3): 211–8. doi:10.1007/BF02288367.
- [2] Fischer, G.: Lineare Algebra. Braunschweig Wiesbaden 1997
- [3] Golub, G.H., van Loan, C. F.: Matrix computations. Baltimore 2013
- [4] 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)
- [5] Lange, M. (2009) A Tale of two Vectors. *Dialectica*, 63(4), 397–431
- [6] Lorenz, F.: Lineare Algebra I, II. Mannheim, 1988
- [7] Mardia, K.V., Kent, J. T., Bibby, J.M.: Multivariate Analysis. Academic Press, London, New York, Toronto 1979
- [8] 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.
- [9] Stewart, G. W.: Introduction to Matrix Computations. Academic Press, New York 1973

Index

- Abbildung
 - bijektiv, 89
 - Bild von, 89
 - homomorphe, 90
 - identische, 89
 - injektiv, 89
 - inverse, 89
 - isomorphe, 90
 - Kern einer, 89
 - lineare, 90
 - Rang der Abbildung, 91
 - surjektiv, 89
- achsenparallel, 48
- Automorphismus, 90
- Basis
 - eines Vektorraums, 25
 - kanonische, 26
 - orthonormale, 26
- Basisentwicklung eines Vektors
 - orthonormale, 29
- Basisvektoren, 25
- Cauchy-Schwarzsche Ungleichung, 16
- Cramersche Regel, 76
- Determinante, 21
- Dimension, 30
- dyadisches Produkt, 14
- Ebenengleichung, 19
- Eckart & Young
 - Satz von, 72
- Eigenvektoren, 45
- Einheitsvektor, 12
- Einsvektor, 12
- Endomorphismus, 90
- Erzeugendensystem, 24
 - minimales, 25
- Faktorwerte, 67
- Faser, 89
- Fundamental-Lemma, 22, 27, 78
- Gesamtvarianz der Daten, 71
- Gradientenvektor, 85
- Gram-Matrix, 37
- Hauptachsentransformation, 51, 65
- Homomorphismus, 90
- Identität, 89
- Inverse, 42, 83
 - generalisierte, 66
 - Moore-Penrose, 66
 - Pseudo, 66
- Isomorphismus, 90
- Kern, 73
- Kodimension, 30
- kollinear, 18, 21
- Koordinaten bezüglich einer Basis, 25
- Koordinaten eines Vektors bezüglich einer Basis, 25
- Kreuzproduktmatrix, 36
- längeninvariant, 45
- Ladung, 29
- Ladungen, 67
- Lagrange
 - Faktor, 85
 - Funktion, 85
 - Multiplikator, 85
 - sche Multiplikatorenregel, 85
- lineare Hülle, 24
- lineare Hülle (Gleich'syst.), 73
- Linearkombination, 13
- Matrix
 - Diagonal-, 31
 - gestürzte, transponierte, 31
 - Spaltenstandardisierung, 36
 - symmetrische, 31
- Metrik
 - City-Block, 8
 - euklidisch, 8
 - Manhattan, 8
 - Minkowski, 8

- negativ semidefinit, 47
- Normalenvektor, 18
- Nullvektor, 12
- Orientierungsinvarianz, 45
- Orthogonalität, 17
- Orthonormalbasis (ONB), 26
- orthonormale Basisentwicklung, 28
- Polynom, charakteristisches, 49, 94
- positiv semidefinit, 47, 50
- Produkt
 - äußeres, 14
 - inneres, 14
- Projektion
 - Vektor auf einen anderen, 76
- Pythagoras, 15
- quadratische Form, 47
- Rang
 - einer Kreuzproduktmatrix, 41
 - einer Vektormenge, 30
 - voller, 30
- Rayleigh-Quotient, 55
- Rohmatrix, 38
- Rotation, 27, 33, 46, 65
- Rotationsmatrix, 45
- Satz
 - von Courant-Fischer, 56
- Singularvektoren, 61
- Singularwerte, 61
- Singularwertzerlegung, 61
- Skalar, 14
- Skalarprodukt, 14
- Spaltenrang, 39
- Spaltenraum, 91
- Spaltenvektoren, 31
- Spektraldarstellung, 53
- Varianz-Kovarianzmatrizen
 - dyadisches Produkt, 37
- Vektor
 - charakteristischer, 48
 - latenter, 48
 - Norm, 15
 - normiert, 15
- Vektorabbildung, 33
- Vektoren
 - charakteristische, 45
- Vektorraum
 - n -dimensionaler, 25
- Vektortransformation, 33
- Wert
 - charakteristischer, 48
 - latenter, 48
- Zeilenrang, 39
- Zeilenvektoren, 31
- Zentrierungsmatrix, 37