

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

U. Mortensen

Letzte Änderung: 29. 03. 2024

Inhaltsverzeichnis

1	Vektoren und Vektorräume	5
1.1	Zur Bedeutung und Geschichte der Vektor- und Matrixrechnung . . .	5
1.2	Vektoren und ihre Verknüpfungen	7
1.2.1	Der Begriff des Vektors	7
1.2.2	Linearkombinationen von Vektoren	10
1.2.3	Produkte von Vektoren	14
1.2.4	Zentrierte Vektoren, Varianzen und Kovarianzen	21
1.2.5	Länge, Orientierung und Kosinus-Ähnlichkeit von Vektoren	22
1.2.6	Vektorprojektionen	23
1.2.7	Die Cauchy-Schwarzsche Ungleichung	25
1.2.8	Lineare Abhängigkeit und Unabhängigkeit von Vektoren . .	25
1.3	Vektorräume	30
1.3.1	Definition eines Vektorraums	30
1.3.2	Erzeugendensysteme und Basen von Vektorräumen	38
1.4	Lineare Gleichungssysteme I	55
2	Matrizen	60
2.1	Definitionen	60
2.2	Operationen mit Matrizen	62
2.2.1	Addition und Multiplikation mit einem Skalar	62
2.2.2	Multiplikation einer Matrix mit einem Vektor	62
2.2.3	Multiplikation einer Matrix mit einer Matrix	64
2.3	Der Rang einer Matrix	66
2.4	Die Inverse einer Matrix	74
2.5	Die Beziehungen zwischen verschiedenen Basen	78
3	Bestimmung einer Basis: die Hauptachsentransformation	79
3.1	Rotationen	79
3.2	Die Rotation einer Punktekonfiguration	82
3.2.1	Annahmen und Implikationen: die Eigenstruktur von $X'X$.	82
3.2.2	Eigenschaften der Eigenstruktur symmetrischer Matrizen .	85
3.2.3	Zur Eindeutigkeit von Eigenvektoren	91
3.2.4	Zur Geometrie der Eigenvektoren symmetrischer Matrizen .	93

3.2.5	Rang und Anzahl der Eigenwerte ungleich Null	95
3.2.6	Inverse und Wurzel einer symmetrischen Matrix	99
3.2.7	Die Singularwertzerlegung einer Matrix	100
3.2.8	Die multivariate Gauss-Verteilung	103
3.3	Eigenvektoren und Eigenwerte nichtsymmetrischer Matrizen	106
3.3.1	Der allgemeine Fall	106
3.3.2	Mehrfache Eigenwerte	111
3.3.3	Das generalisierte Eigenvektorproblem	113
3.4	Lineare Gleichungssysteme II	115
3.5	Weitere Befunde	119
3.5.1	Die Zentrierungsmatrix	119
3.5.2	Die Pseudoinverse einer Matrix	122
3.5.3	Vektor- und Matrixnormen	124
3.5.4	Die Approximation von Matrizen	127
3.6	Anwendung: Latente Variablen	131
3.6.1	Explorieren (Hauptkomponenten und SVD)	131
3.6.2	Diskriminieren und klassifizieren	139
3.7	Projektionen	144
3.7.1	Projektion auf eine Ebene im \mathbb{R}^3	144
3.7.2	Der allgemeine Fall: Projektionsmatrizen	145
3.7.3	Beispiel: das Allgemeine Lineare Modell	146
3.7.4	Projektionen auf Hauptachsen	149
4	Funktionsräume und PCA	149
4.1	Einführung	149
4.2	Karhunen-Loève-Entwicklung und PCA	151
5	Anhang	156
5.1	Punkträume	156
5.2	Skalen und Abbildungen	159
5.3	Zur geometrischen Definition des Skalarprodukts	161
5.4	Elementarmatrizen und elementare Operationen	162
5.4.1	Lineare Gleichungen und Gauß-Algorithmus	164
5.4.2	Herleitung einer Rotationsmatrix	166

5.4.3	Die Cholesky-Zerlegung	169
5.5	Zur Berechnung von Ellipsen für eine Punktekonfiguration	170
5.6	Die Differentiation von Vektoren	172
5.6.1	Die allgemeine Differentiationsformel	172
5.6.2	Die Differentiation quadratischer Formen	173
5.6.3	Die Kleinste-Quadrate-Schätzung für das Lineare Modell	174
5.6.4	Extrema unter Nebenbedingungen	176
5.7	Der Rayleigh-Koeffizient und der Satz von Courant-Fischer	178
5.8	Alternativer Beweis von Satz 3.20	180
5.9	Vektortransformationen und Abbildungen	182
5.10	Determinanten	187
5.10.1	Transformationen und Volumen	196
5.10.2	Transformationen und Orientierung	200
5.10.3	Die Transformation zufälliger Veränderlicher	202
	Literatur	210
	Index	212

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

Bertrand Russell

1 Vektoren und Vektorräume

1.1 Zur Bedeutung und Geschichte der Vektor- und Matrixrechnung

Auf einem Tisch liegt ein Stein, und er soll verschoben werden. Dazu muß eine Kraft K aufgewendet werden, die in die gewünschte Richtung wirkt. Die Kraft kann durch einen Pfeil repräsentiert werden, dessen Länge das Ausmaß der aufgewendeten Kraft und dessen Orientierung die Richtung, in der die Kraft wirkt, repräsentiert. Dieser Pfeil repräsentiert einen Vektor (lat. für Führer, Fahrer, Träger) – Kraft ist eine vektorielle Größe. Ganz allgemein gilt: Vektoren sind (oder repräsentieren) gerichtete Größen. Vektoren werden durch einen Buchstaben repräsentiert mit entweder einem Pfeil darüber oder in Fettdruck: \vec{x} oder \mathbf{x} , und durch eine Anzahl von Komponenten, im hier betrachteten Fall sind es zwei, x_1 und x_2 , aus denen sich die Länge und die Orientierung des Vektors berechnen lassen:

$$\vec{x} = \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Im Folgenden wird hauptsächlich die Schreibweise \mathbf{x} verwendet werden.

Der eben beschriebene Vektor heißt 2-dimensional, weil er in einer Ebene liegt, so dass zwei Komponenten genügen, um die Länge und Orientierung zu berechnen. – eine ausführliche Beschreibung wird im folgenden Abschnitt gegeben. Natürlich kann auch eine im 3-dimensionalen Raum wirkende Kraft durch einen Vektor mit nun drei Komponenten repräsentiert werden, und wieder bilden die Länge des Vektors die Größe der Kraft und die Orientierung des Vektors die Richtung, in der die Kraft wirkt, ab.

Das Prinzip kann auf $n > 3$ Dimensionen verallgemeinert werden, n eine beliebige natürliche Zahl: an einem Objekt können n Messungen x_j , $j = 1, \dots, n$, bestimmt werden, so dass das Objekt durch einen Punkt P mit n Koordinaten x_j abgebildet werden kann, – die Messungen sind jetzt die Komponenten eines Vektors \mathbf{x} . Das Objekt kann aber auch durch einen Vektor mit dem Anfangspunkt im Nullpunkt des Koordinatensystems und dem Endpunkt P repräsentiert werden. Die Länge des Vektors ist ein Maß für die Gesamtausprägung der Merkmale, und die Orientierung gibt die Mischung der gemessenen Merkmale bei diesem Objekt an. Man kann Typen von Objekten durch die Mischung von Merkmalen definieren. Die Orientierung eines Vektors signalisiert dann den Typ des repräsentierten Objekts.

Es zeigt sich, dass man gewisse Operationen, die im folgenden Abschnitt im

Detail eingeführt werden, mit Vektoren durchführen kann:

1. Man kann Vektoren \mathbf{x} mit einem "Skalar" $a \in \mathbb{R}$ multiplizieren, so dass ein Vektor $\mathbf{y} = a\mathbf{x}$ entsteht. Diese Operation bedeutet eine Skalierung der Länge von \mathbf{x} .
2. Man kann Vektoren addieren und voneinander subtrahieren; das Resultat ist wieder ein Vektor.
3. Man kann Produkte von Vektoren bilden.

Das heißt, dass man nun mit ganzen Spalten oder Zeilen von Zahlen operieren kann wie mit einzelnen Zahlen in der "normalen" Algebra. Schreibt man Vektoren gleicher Dimensionalität spalten- oder zeilenweise neben- oder untereinander an, so entstehen Matrizen, auf die sich die Rechenregeln 1. – 3. übertragen. Damit kann man mit ganzen "Tabellen" – eine Matrix ist eine Art von Tabelle – rechnen. Auf diese Weise vereinfachen sich relativ unübersichtliche Rechnungen, wie sie bei der Berechnung von Regressionsparametern oder von "latenten Dimensionen" notwendig werden, beträchtlich. Vor allem werden sie transparenter, Abhängigkeiten zwischen verschiedenen Parametern werden deutlicher. Wie dies im Einzelnen geschieht wird in den folgenden Abschnitten deutlich werden. \square

Der Beginn der Vektorrechnung wird oft mit der Publikation von H.G. Graßmanns¹ *Lineale Ausdehnungslehre*² (1844), einem Lehrbuch der Geometrie, und W.R. Hamiltons³ *Lectures on Quaternions* (1853), einem Buch über bestimmte Vierergruppen von Zahlen, angesetzt. Graßmanns und Hamiltons Arbeiten sind unabhängig voneinander entstanden, aber es war Hamilton, der den Ausdruck 'Vektor' einführte. Wie Barrantes (2012) anmerkt liegen die Ursprünge der Vektorrechnung deutlich weiter zurück in der Geschichte der Mathematik und wurzeln in Bemühungen, eine systematische Theorie über das Lösen von Gleichungen mit mehreren Unbekannten zu schaffen. Erste Arbeiten zu diesem Thema wurden von den Babyloniern schon 1700 v. Chr. und den Chinesen (Shang Cang: *Mathematik in neun Büchern*, 180 v. Chr.) verfasst (Artmann und Törner (1991)). So hatte Leibniz die Aufgabe des Königs von Hannover, das Abpumpen von Grundwasser in den Bergwerken des Harzes zu organisieren, und in diesem Zusammenhang legte Leibniz erste Grundlagen für die Theorie der Determinanten (Hirsch (2000)); er fand auch bereits die später nach Cramer⁴ benannte Regel (Cramersche Regel), die Cramer 1750 wohl ohne Kenntnis der Leibnizschen Arbeit publizierte. In der Tat beginnen viele Lehrbücher über die Vektor- und Matrixrechnung (Lineare Algebra) mit einem Kapitel über lineare Gleichungssysteme. Wohl spätestens seit man physikalische Größen wie etwa die Kraft durch Vektoren repräsentierte sah man die allgemeine Nützlichkeit des Vektorbegriffs. Eine Kraft hat eine bestimmte Größe und wirkt in eine bestimmte Richtung, und noch heute werden Vektoren oft als Objekte, die sowohl eine Größe wie auch eine Orientierung haben ("gerichtete Größen"), eingeführt; in diesem Sinne verwen-

¹Hermann Günther Graßmann (1609 – 1877), deutscher Mathematiker

²Nicht: Lineare

³William Rowan Hamilton (1805 - 1865), irischer Mathematiker

⁴Gabriel Cramer (1704 – 1752), Genfer Mathematiker

te der schottische Physiker James Clerk Maxwell (1831 – 1879) den Vektorbegriff auch in seiner Entwicklung der Elektrodynamik⁵. Dass Vektoren auch in anderen Wissenschaften, z. B. der Statistik, angewendet werden können hat mit dazu beigetragen, dass die Vektor- und Matrixrechnung mehr als nun eine periphere mathematische Übung wurde.

1.2 Vektoren und ihre Verknüpfungen

1.2.1 Der Begriff des Vektors

Es werden zunächst nur endlich-dimensionale Vektoren über \mathbb{R} (die Menge der reellen Zahlen) betrachtet.

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

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad (1.1)$$

wobei es auf die Reihenfolge der x_1, \dots, x_n ankommt. Die $x_i, i = 1, \dots, n$ heißen die Komponenten des Vektors. (s. Abbildung 1), Seite 8). Es wird auch $\mathbf{x} \in \mathbb{R}^n$ geschrieben.

Dementsprechend sind

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

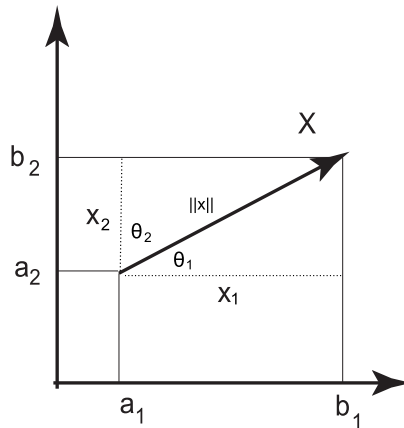
verschiedene 3-dimensionale Vektoren.

Anmerkungen:

1. Die Komponenten x_1, \dots, x_n bilden ein "n-Tupel" – eine Zusammenfassung – von Zahlen. Die Schreibweise \mathbb{R}^n ergibt sich, wenn man die Menge der möglichen n -Tupeln betrachtet. Für x_1 kann eine beliebige Zahl aus \mathbb{R} , der Menge aller reellen Zahlen, gewählt werden. Für x_2 ebenfalls, so dass man für die ersten beiden Zahlen x_1 und x_2 $x_1 \times x_2$ Möglichkeiten hat, was auch in der Form $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ geschrieben werden kann. Für x_3 kann man wieder jede beliebige Zahl aus \mathbb{R} wählen, dass man jedes 2-Tupel mit jedem möglichen x_3 kombinieren kann, man erhält $(\mathbb{R} \times \mathbb{R}) \times \mathbb{R} = \mathbb{R} \times \mathbb{R} \times \mathbb{R} = \mathbb{R}^3$, Möglichkeiten, etc. Allgemein ergeben sich für n Komponenten \mathbb{R}^n Möglichkeiten. $\mathbf{x} \in \mathbb{R}^n$ heißt dann, dass \mathbf{x} eine Tupel aus dem Cartesischen Produkt $R \times R \times \dots \times \mathbb{R} = \mathbb{R}^n$ möglichen n -Tupeln ist.
2. Vektoren werden im Folgenden wie in (5.24) durch fette Buchstaben bezeichnet (mit zwei Ausnahmen); diese Bezeichnung ist mittlerweile allgemein üblich.

⁵ *Treatise on Electricity and Magnetism* (1873)

Abbildung 1: Ein 2-dimensionaler Vektor \mathbf{x} und seine Komponenten $x_1 = b_1 - a_1$ und $x_2 = b_2 - a_2$; $\|\mathbf{x}\|$ bezeichnet die Länge von \mathbf{x} . θ_1 und θ_2 sind die Winkel zwischen \mathbf{x} und x_1 bzw. x_2



Eine alternative Schreibweise ist \vec{x} ; der Pfeil deutet an, dass mit x ein Vektor gemeint ist. Vor dieser Schreibweise wird ebenfalls Gebrauch gemacht. Die Fettschreibweise erleichtert Schreibweisen wie $\hat{\mathbf{x}}$, die Schätzungen eines Vektors \mathbf{x} bezeichnen, oder $\tilde{\mathbf{x}}$, die Vektoren bezeichnen, die durch die die Zeilen einer Matrix definiert sind. Die Schreibweise \vec{x} wird dann leicht unübersichtlich: $\hat{\vec{x}}$ oder $\tilde{\vec{x}}$ und $\hat{\hat{x}}$ bzw. $\tilde{\tilde{x}}$ zeichnen sich durch sehr geringen graphischen Charme aus, insbesondere wenn transponierte Vektoren \vec{x}' (s. unten) betrachtet werden.

3. Ein Vektor \mathbf{x} wird wie in Gleichung (5.24) stets als Spalte oder Kolumne angeschrieben, es sei denn, er wird ausdrücklich als *transponiert* oder *gestürzt* gekennzeichnet; in diesem Fall wird er als Zeile angeschrieben:

$$\mathbf{x}' = (x_1, x_2, \dots, x_n). \quad (1.2)$$

Analog dazu kann man einen Spaltenvektor als transponierten Zeilenvektor schreiben:

$$\mathbf{x} = (\mathbf{x}')' = (x_1, x_2, \dots, x_n)' \quad (1.3)$$

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

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

für einen Spaltenvektor, denn stürzt man eine Zeile, so wird sie zur Spalte. \square (daher der Pfeil in der Schreibweise \vec{x})

Vektoren werden graphisch oft durch Pfeile repräsentiert, s. Abbildung 1. Pfeile haben eine Länge und eine Orientierung, so dass von Vektoren auch als von *gerichteten Größen* gesprochen wird. Der Anfangspunkt des Pfeils habe die

Koordinaten a_1, \dots, a_n , der Endpunkt habe die Koordinaten b_1, \dots, b_n . Es sei nun

$$x_i = b_i - a_i, \quad i = 1, \dots, n \quad (1.4)$$

die x_i sind die *Komponenten* des Vektors. Wie man der Abb. 1 entnehmen kann werden sowohl die Länge wie auch die Orientierung durch die Komponenten definiert, nicht aber der Ort des repräsentierenden Pfeils, auf den es demnach nicht ankommt. Die Länge ist im 2-dimensionalen Fall offenbar dem Satz des Pythagoras entsprechend durch

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2} \quad (1.5)$$

gegeben. Offenbar gelten die Beziehungen

$$\cos \theta_1 = x_1 / \|\mathbf{x}\|, \quad \cos \theta_2 = x_2 / \|\mathbf{x}\|,$$

so dass sich für die Komponenten die Beziehungen

$$x_1 = \|\mathbf{x}\| \cos \theta_1, \quad x_2 = \|\mathbf{x}\| \cos \theta_2 \quad (1.6)$$

ergeben. Demnach kann

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \|\mathbf{x}\| \cos \theta_1 \\ \|\mathbf{x}\| \cos \theta_2 \end{pmatrix} = \|\mathbf{x}\| \begin{pmatrix} \cos \theta_1 \\ \cos \theta_2 \end{pmatrix}; \quad (1.7)$$

geschrieben werden. Die Rechtfertigung für den Ausdruck auf der rechten Seite, bei dem $\|\mathbf{x}\|$ vor die Klammer gezogen wird, wird im folgenden Abschnitt 1.2.2 gegeben. Die rechte Seite verdeutlicht die Definition eines Vektors als gerichteter Größe: die Komponenten sind durch die "Größe" $\|\mathbf{x}\|$ und die durch die Winkel θ_1 und θ_2 definierte Orientierung definiert. Diese Definitionen werden später in allgemeinerer Form wiederholt.

Spezielle Vektoren: Es seien

$$\vec{0} = (0, 0, \dots, 0)' \quad (1.8)$$

$$\vec{1} = (1, 1, \dots, 1)' \quad (1.9)$$

$$\mathbf{e}_j = (0, \dots, 0, 1, 0, \dots, 0)', \quad j = 1, \dots, n \quad (1.10)$$

$\vec{0}$ heißt der *Nullvektor*, $\vec{1}$ heißt *Einsvektor*; gelegentlich wird auch $\vec{1}_n$ bzw. $\vec{0}_n$ geschrieben, um anzudeuten, dass der Vektor n Komponenten hat. \mathbf{e}_j ist der j -te *kanonische Einheitsvektor*, seine Komponenten sind alle gleich Null bis auf die j -te Komponente, die gleich 1 ist (vergl. auch die Definition des Einheitsvektors auf Seite 16).

Anmerkung: Der Nullvektor kann nicht durch einen Pfeil dargestellt werden, denn er hat die Länge $\|\vec{0}\| = \sqrt{0^2 + 0^2 + \dots + 0^2} = 0$, und damit auch keine Orientierung. \square

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 Punktbegriff 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.

1.2.2 Linearkombinationen von Vektoren

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

$$a\mathbf{x} = a \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_n \end{pmatrix}, \quad a \in \mathbb{R} \quad (1.11)$$

a heißt *Skalar*, weil a ein Wert der "Skala", d.h. der reellen Zahlen zwischen $-\infty$ und ∞ ist.

Es sei $\mathbf{u} = a\mathbf{x}$, $\mathbf{v} = -a\mathbf{x}$ mit $a \neq 0$. \mathbf{u} und \mathbf{v} haben dann entgegengesetzte Orientierungen. Dazu werde der Einfachheit halber $a = 1$ angenommen; dann ist

$$\mathbf{u} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} -x_1 \\ -x_2 \\ \vdots \\ -x_n \end{pmatrix}. \quad (1.12)$$

Man veranschaulicht sich den Sachverhalt, dass \mathbf{u} und \mathbf{v} entgegengesetzte Orientierungen haben, am einfachsten, indem man in Abbildung 1 $a_1 = a_2 = 0$ setzt, d.h. den Anfangspunkt des Vektors in den Ursprung des Koordinatensystems legt. Dann ist $-x_1 = -b_1$, $-x_2 = -b_2$ und \mathbf{v} zeigt in die entgegengesetzte Richtung von \mathbf{u} . Abb. 2 zeigt die Vektoren $\mathbf{x} = (a, b)$ und $-\mathbf{x} = (-a, -b)'$. Die beiden Vektoren liegen offenbar auf einer Geraden, zeigen aber in entgegengesetzte Richtungen; der Einfachheit halber liegen die Anfangspunkte der Vektoren im Nullpunkt des Koordinatensystems. Offenbar gilt $\cos \theta_1 = a/\|\mathbf{x}\|$, $\cos \phi_2 = b/\|\mathbf{x}\|$, so dass

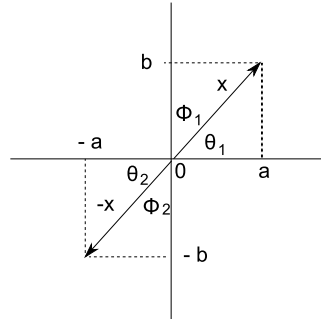
$$\mathbf{x} = (a, b)' = \|\mathbf{x}\|(\cos \theta_1, \cos \phi_1)',$$

und $\cos \theta_2 = -a/\|\mathbf{x}\|$, $\cos \phi_2 = -b/\|\mathbf{x}\|$, also

$$-\mathbf{x} = (-a, -b)' = \|\mathbf{x}\|(-\cos \theta_2, -\cos \phi_2)' = \|\mathbf{x}\|(-\cos \theta_1, -\cos \phi_1)' \quad (1.13)$$

wegen $\theta_1 = \theta_2$ und $\phi_1 = \phi_2$.

Abbildung 2: Vektoren \mathbf{x} und $-\mathbf{x}$



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

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

die Summe der Vektoren \mathbf{x} und \mathbf{y} . Vektoren können also nur addiert werden, wenn sie dieselbe Anzahl von Komponenten haben. Analog dazu ist die Differenz zweier Vektoren definiert:

$$\mathbf{u} = \mathbf{x} - \mathbf{y} = \begin{pmatrix} x_1 - y_1 \\ x_2 - y_2 \\ \vdots \\ x_n - y_n \end{pmatrix} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \quad (1.15)$$

Es sei $\mathbf{v} = \mathbf{y} - \mathbf{x}$, d.h. $\mathbf{v} = -\mathbf{u}$. \mathbf{u} und \mathbf{v} haben entgegengesetzte Orientierungen. vergl. Abb. 4.

Anmerkung: Es sei daran erinnert, dass Vektoren nur durch ihre Länge und Orientierung, nicht aber durch ihre Position definiert sind. Summe und Differenz zweier Vektoren \mathbf{x} und \mathbf{y} sollten also nicht automatisch so repräsentiert werden, als fielen die Anfangspunkte der beiden Vektoren zusammen, – dieser Fall ist ein Spezialfall, für die Addition erscheint der durch die Summe definierte Vektor wie in einem Kräfteparallelogramm in der Physik, und für die Differenz zweier Vektoren definierte Vektor erscheint als Verbindung der Endpunkte der Vektoren, vergl. Abb. 4. Grundsätzlich können aber die Summen- und Differenzen irgendwo positioniert werden, alle Positionen sind im Prinzip gleichwertig.

□

Summe und Differenz können als Spezialfälle der allgemeinen Vektorgleichung

$$\mathbf{y} = a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 \quad (1.16)$$

Abbildung 3: Addition von Vektoren, weitere Erläuterung s. Text

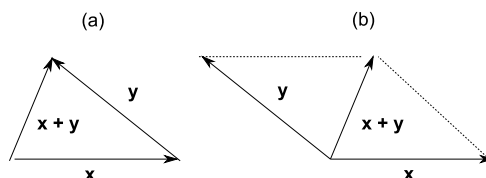
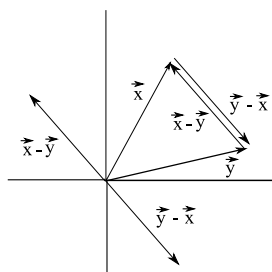


Abbildung 4: Subtraktion von Vektoren; die Differenzen $\mathbf{x} - \mathbf{y}$ und $\mathbf{y} - \mathbf{x}$ sind Vektoren mit entgegengesetzter Orientierung



aufgefasst werden, mit $\mathbf{x}_1 = \mathbf{x}$, $\mathbf{x}_2 = \mathbf{y}$, und $\mathbf{y} = \mathbf{z}$ bzw. $\mathbf{y} = \mathbf{u}$. Für $a_1 = a_2 = 1$ erhält man die einfache Summe, für $a_1 = 1$, $a_2 = -1$ erhält man die einfache Differenz.

Die Gleichung (1.16) kann verallgemeinert werden:

Definition 1.2 Es seien $\mathbf{x}_1, \dots, \mathbf{x}_p$ n -dimensionale Vektoren und a_1, \dots, a_p seien Skalare. Die Summe

$$\mathbf{y} = a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 + \dots + a_p \mathbf{x}_p, \quad (1.17)$$

heißt Linearkombination der $\mathbf{x}_1, \dots, \mathbf{x}_p$.

Anmerkungen:

- *Lineare Funktionen:* Der Ausdruck 'Linearkombination' verweist darauf, dass der Vektor \mathbf{y} eine *lineare Funktion* der \mathbf{x}_j , $j = 1, \dots, n$ ist. Der Vektor \mathbf{y} sei eine Funktion $\mathbf{y} = L(\mathbf{x}_1, \dots, \mathbf{x}_p)$ der \mathbf{x}_j . L heißt *linear*, wenn
 1. $L(a_j \mathbf{x}_j) = a_j L(\mathbf{x}_j)$, (L ist *homogen*)
 2. $L(\mathbf{x}_1 + \dots + \mathbf{x}_p) = L(\mathbf{x}_1) + \dots + L(\mathbf{x}_p)$ (L ist *additiv*)
 gilt. Eine Linearkombination ist offenbar eine lineare Funktion.
- *Linearkombination als Abbildung:* Die Koeffizienten a_1, \dots, a_p können zu einem p -dimensionalen Vektor $\mathbf{a} = (a_1, \dots, a_p)'$ zusammengefasst werden.

Dann kann eine Linearkombination $\mathbf{y} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \dots + a_p\mathbf{x}_p$ als *Abbildung* $\mathbf{a} \mapsto \mathbf{y}$ des Vektors \mathbf{a} auf den Vektor $\mathbf{y} = (y_1, \dots, y_n)'$ bzw. als *Transformation* des Vektors \mathbf{a} in den Vektor \mathbf{y} verstanden werden, wobei die Abbildung bzw. Transformation durch die Vektoren $\mathbf{x}_1, \dots, \mathbf{x}_p$ definiert ist. Man beachte, dass n und p nicht denselben Wert haben müssen.

Beispiel 1.1 Vektoren als Linearkombination von Einheitsvektoren Es sei $\mathbf{x} = (x_1, x_2, \dots, x_n)'$. Dann gilt

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \dots + x_n\mathbf{e}_n, \quad (1.18)$$

wie man leicht durch Nachrechnen überprüft, d.h. ein beliebiger Vektor kann stets als Linearkombination der Einheitsvektoren \mathbf{e}_j , $j = 1, \dots, n$ aufgefasst werden. \square

Beispiel 1.2 Multiple Regression Es wird angenommen, dass der Wert einer Variablen Y durch drei Prädiktorvariablen X_1 , X_2 und X_3 bis auf einen zufälligen Fehler "vorhergesagt" werden kann:

$$Y_i = a_1X_{i1} + a_2X_{i2} + a_3X_{i3} + e_i, \quad i = 1, \dots, m \quad (1.19)$$

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

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix} = a_1 \begin{pmatrix} X_{11} \\ X_{21} \\ \vdots \\ x_{m1} \end{pmatrix} + a_2 \begin{pmatrix} X_{12} \\ X_{22} \\ \vdots \\ x_{m2} \end{pmatrix} + a_3 \begin{pmatrix} X_{13} \\ X_{23} \\ \vdots \\ x_{m3} \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{pmatrix} \quad (1.20)$$

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

$$\mathbf{y} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{x}_3 + \mathbf{e} \quad (1.21)$$

schreiben kann. \mathbf{y} wird hier als Linearkombination der \mathbf{x}_j und \mathbf{e} dargestellt. Die Vektorschreibweise erscheint hier zunächst als vereinfachte Schreibweise; die "operativen Implikationen" dieser Schreibweise zeigen sich im Folgenden. Da die Komponenten von \mathbf{x}_1 , \mathbf{x}_2 und \mathbf{y} Messwerte sind, die üblicherweise mit zufälligen Fehlern behaftet sind, heißen diese Vektoren auch *zufällige Vektoren*. Natürlich ist \mathbf{e} ebenfalls ein Zufallsvektor. Zufällige Vektoren werden in Abschnitt 3.2.8 ausführlicher behandelt. \square

1.2.3 Produkte von Vektoren

Gegeben sei eine Linearkombination von n m -dimensionalen Vektoren

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

Die Komponenten y_i von \mathbf{y} sind demnach als *Summe* von Produkten definiert:

$$y_i = a_1x_{i1} + a_2x_{i2} + \cdots + a_nx_{in} = \sum_{j=1}^n a_jx_{ij}, \quad i = 1, \dots, m \quad (1.22)$$

Man kann die Koeffizienten a_j als Komponenten eines Vektors \mathbf{a} auffassen, die $x_{i1}, x_{i2}, \dots, x_{in}$ als Komponenten des Zeilenvektors⁶ einer Matrix X , deren Spaltenvektoren aus den $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ bestehen, die wiederum die Spaltenvektoren $\tilde{\mathbf{x}}_i$ der transponierten Matrix X' sind:

$$\tilde{\mathbf{x}}_i' = (x_{i1}x_{i2}, \dots, x_{in}), \quad 1 \leq i \leq m$$

und die Komponente y_i von \mathbf{y} als "Produkt" der Vektoren $\tilde{\mathbf{x}}_i$ und \mathbf{a} bezeichnen. Zentrale Größen in der multivariaten Statistik wie Varianzen, Kovarianzen und Korrelationen zwischen Variablen sind als Summen $\sum_i x_i y_i$ von Produkten definiert, so dass die folgende allgemeine Definition gerechtfertigt ist:

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

$$\mathbf{x}'\mathbf{y} = (x_1, x_2, \dots, x_n) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \sum_{i=1}^n x_i y_i \quad (1.23)$$

das Skalarprodukt *oder* inneres Produkt *von \mathbf{x} und \mathbf{y} .*

(ii) *Nun sei \mathbf{y} m -dimensional, d.h. \mathbf{x} und \mathbf{y} müssen nicht notwendig dieselbe Anzahl von Komponenten haben. Dann heißt*

$$\mathbf{x}\mathbf{y}' = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} (y_1, \dots, y_m) = \begin{pmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_m \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_m \\ \vdots & \vdots & \ddots & \vdots \\ x_n y_1 & x_n y_2 & \cdots & x_n y_m \end{pmatrix} \quad (1.24)$$

dyadisches Produkt *oder* tensorielles Produkt *von \mathbf{x} und \mathbf{y} .*

Anmerkungen:

1. Man überprüft anhand von (1.23) sofort, dass zwar $\mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x}$ gilt, aber anhand von (1.24), dass

$$\mathbf{x}'\mathbf{y} \neq \mathbf{x}\mathbf{y}'. \quad (1.25)$$

⁶Vergl. Die Gleichung (2.1), Seite 61 zur Definition von $\tilde{\mathbf{x}}_i$

2. **Alternative Definition:** Ein alternative, aber äquivalente Definition des Skalarprodukts wird auf Seite 18, Gleichung (1.43) gegeben; sie setzt die verallgemeinerte Definition der Länge eines Vektors sowie die Definition des von zwei Vektoren gebildeten Winkels voraus. Diese Begriffe werden im Folgenden aus der Definition (1.23) hergeleitet.

3. Der Ausdruck *Skalarprodukt* reflektiert den Sachverhalt, dass das Skalarprodukt zweier Vektoren eine einzelne Zahl ist, – eben ein *Skalar*, d.h. dass das Produkt einen Wert s auf der "Skala" $-\infty < s < \infty$ annimmt.

Die Schreibweise $\mathbf{x}'\mathbf{y}$ suggeriert, wie in (1.23) angezeigt, dass das Skalarprodukt als Produkt eines Zeilen- und eines Spaltenvektors angeschrieben werden kann; bei der Einführung von Produkten einer Matrix mit einem Vektor oder mit einer weiteren Matrix wird der Sinn dieser Regel deutlich werden. Analoge Bemerkungen gelten für das dyadische Produkt \mathbf{xy}' .

4. Statt der Schreibweise $\mathbf{x}'\mathbf{y}$ wird auch die Schreibweise $\mathbf{x} \cdot \mathbf{y}$ (im Englischen auch dot product genannt)

$$\mathbf{x} \cdot \mathbf{y} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} x_1y_1 \\ x_2y_2 \\ \vdots \\ x_ny_n \end{pmatrix} \quad (1.26)$$

verwendet. Im Folgenden wird die in (1.23) eingeführte Schreibweise $\mathbf{x}'\mathbf{y}$ verwendet, weil sie sich, wie schon angedeutet, im Zusammenhang mit der Matrixrechnung als nützlich erweist.

Mit dem in (1.26) definierten Vektor \mathbf{u} läßt sich das Skalarprodukt auch in der Form

$$\mathbf{x}'\mathbf{y} = \mathbf{x} \cdot \mathbf{y} = \mathbf{u}'\vec{1} \quad (1.27)$$

schreiben, wobei $\vec{1}$ der in (1.9) definierte Einsvektor ist. Die Summe der Komponenten eines Vektors kann also als Skalarprodukt des Vektors mit dem Einsvektor geschrieben werden.

Eigenschaften des Skalarprodukts Das Skalarprodukt hat die folgenden allgemeinen Eigenschaften:

1. *Kommutativität:* Es gilt

$$\mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x}, \quad (1.28)$$

denn es ist sicherlich $\sum_i x_i y_i = \sum_i y_i x_i$.

2. *Homogenität:* Aus der Definition des Produktes eines Vektors mit einem Skalar $a \in \mathbb{R}$ folgt

$$(a\mathbf{x})'\mathbf{y} = \mathbf{x}'(a\mathbf{y}) = a(\mathbf{x}'\mathbf{y}), \quad (1.29)$$

denn $\sum_i (ax_i)y_i = \sum_i x_i(ay_i) = a \sum_i x_i y_i$.

3. *Distributivität*: Für $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ gilt

$$\mathbf{x}'(\mathbf{y} + \mathbf{z}) = \mathbf{x}'\mathbf{y} + \mathbf{x}'\mathbf{z}, \quad (1.30)$$

$$(\mathbf{x} + \mathbf{y})'\mathbf{z} = \mathbf{x}'\mathbf{z} + \mathbf{y}'\mathbf{z}, \quad (1.31)$$

denn es gilt

$$\mathbf{x}'(\mathbf{y} + \mathbf{z}) = \sum_{i=1}^n x_i(y_i + z_i) = \sum_{i=1}^n x_i y_i + \sum_{i=1}^n x_i z_i = \mathbf{x}'\mathbf{y} + \mathbf{x}'\mathbf{z}.$$

und (1.31) folgt analog. Bei diesem Nachweis wird von der algebraischen Definition des Skalarprodukts Gebrauch gemacht; dass das Distributivgesetz auch gilt, wenn man die geometrische Definition zugrunde legt, kann man wegen der Äquivalenz der beiden Definitionen erwarten, der Nachweis verläuft aber etwas anders und kann im Anhang, Abschnitt 5.3, Seite 161 nachgelesen werden.

Beispiel 1.3 Einheitsvektoren Für die kanonischen \mathbf{e}_j gilt, wie man leicht nachrechnet,

$$\mathbf{e}_j' \mathbf{e}_k = \delta_{jk} = \begin{cases} 0, & j \neq k \\ 1, & j = k \end{cases} \quad (1.32)$$

δ_{jk} ist das *Kronecker-Delta*⁷, d.h. δ_{jk} ist eine Funktion, die abhängig von bestimmten Bedingungen (hier: $j \neq k$ oder $j = k$) entweder den Wert 0 oder den Wert 1 annimmt. Es seien $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$; die Vektoren können in der Form

$$\mathbf{x} = x_1 \mathbf{e}_1 + \cdots + x_n \mathbf{e}_n, \quad \mathbf{y} = y_1 \mathbf{e}_1 + \cdots + y_n \mathbf{e}_n$$

angeschrieben werden. Wegen der Distributivität des Skalarprodukts hat man dann

$$\begin{aligned} \mathbf{x}'\mathbf{y} &= (x_1 \mathbf{e}_1 + \cdots + x_n \mathbf{e}_n)'(y_1 \mathbf{e}_1 + \cdots + y_n \mathbf{e}_n) \\ &= \sum_{j=1}^n \sum_{k=1}^n x_j y_k \mathbf{e}_j' \mathbf{e}_k = \sum_{j=1}^n x_j y_j, \end{aligned} \quad (1.33)$$

wegen (1.32). □

Die folgenden Begriffe ergeben sich unmittelbar aus der Definition des Skalarprodukts und werden für weitere Kommentare zum Begriff des Skalarprodukts benötigt.

1. **Die Norm eines Vektors** Für $\mathbf{y} = \mathbf{x}$ ergibt sich das Skalarprodukt

$$\mathbf{x}'\mathbf{x} = \sum_{i=1}^n x_i^2 \quad (1.34)$$

⁷Leopold Kronecker (1823–1891), Mathematiker

Die rechte Seite entspricht dem Quadrat der Länge $\|\mathbf{x}\|$ des Vektors \mathbf{x} (vergl. Abbildung 1), d.h. es gilt

$$\mathbf{x}'\mathbf{x} = \|\mathbf{x}\|^2, \quad \text{so dass } \|\mathbf{x}\| = \sqrt{\mathbf{x}'\mathbf{x}}. \quad (1.35)$$

(Satz des Pythagoras).

2. Skalierung eines Vektors, normierte Vektoren Es sei $a \in \mathbb{R}$. Dann folgt

$$\|a\mathbf{x}\| = |a|\|\mathbf{x}\|, \quad (1.36)$$

wie man durch Einsetzen in (1.35) sieht. Die Multiplikation eines Vektors mit einem Skalar bedeutet die Skalierung der Länge des Vektors; für $a < 1$ erhält man eine Stauchung, d.h. eine Verkürzung des Vektors, für $a > 1$ eine Streckung oder Verlängerung. Es werde a so gewählt, dass mit $\mathbf{y} = a\mathbf{x}$ die Bedingung

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

erfüllt ist. Dann folgt

$$|a| = \frac{1}{\|\mathbf{x}\|} > 0 \quad (1.37)$$

Der Vektor

$$\mathbf{y} = \left(\frac{1}{\|\mathbf{x}\|}x_1, \frac{1}{\|\mathbf{x}\|}x_2, \dots, \frac{1}{\|\mathbf{x}\|}x_n \right)' = \frac{1}{\|\mathbf{x}\|}(x_1, x_2, \dots, x_n)' \quad (1.38)$$

heißt dann *normiert*, d.h. er hat die Länge 1. Die Länge $\|\mathbf{x}\|$ wird auch die *Norm* des Vektors \mathbf{x} genannt. Der Begriff der Norm wird in Abschnitt 1.3 weiter spezifiziert.

Satz 1.1 *Es sei θ der Winkel zwischen den Vektoren $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Dann gilt*

$$\cos \theta = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}. \quad (1.39)$$

Anmerkung: Für $n \leq 3$ ist die Bedeutung des Winkels zwischen zwei Vektoren klar. Für $n > 3$ kann (1.39) auch als *Definition* des Begriffs des Winkels zwischen zwei Vektoren gesehen werden.

Beweis: Die Aussage (1.39) ergibt sich aus einer Anwendung des Kosinussatzes:

$$\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\|\mathbf{x}\|\|\mathbf{y}\|\cos \theta \quad (1.40)$$

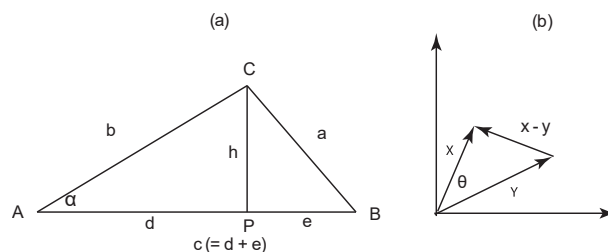
(vergl. Abbildung 5 (b))⁸.

⁸**Beweis:** Vergl. Abbildung 5: 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 $a^2 = b^2 + c^2 - 2bc \cos \alpha$



Es ist

$$\|\mathbf{x} - \mathbf{y}\|^2 = \sum_i (x_i - y_i)^2 = \sum_i x_i^2 + \sum_i y_i^2 - 2 \sum_i x_i y_i \quad (1.41)$$

$$= \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\mathbf{x}'\mathbf{y}. \quad (1.42)$$

Setzt man den Ausdruck auf der rechten Seite für $\|\mathbf{x} - \mathbf{y}\|^2$ in (1.40) ein, so erhält man

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

woraus sofort (1.39) folgt. \square

Alternative Definition des Skalarprodukts: Eine zu (1.23), Seite 14, alternative Definition des Skalarprodukts ist

$$\mathbf{x}'\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|\cos \theta, \quad (1.43)$$

wobei $\|\mathbf{x}\|$ und $\|\mathbf{y}\|$ die in (1.35), Seite 17 definierten Längen der Vektoren sind und θ der Winkel ist, der von den beiden Vektoren gebildet wird. (1.43) heißt auch *geometrische Definition* des Skalarprodukts, während (1.23) als *algebraische Definition* bekannt ist. Die Definitionen (1.23) und (1.43) sind äquivalent, d.h.

$$\mathbf{x}'\mathbf{y} = \sum_{i=1}^n x_i y_i \iff \mathbf{x}'\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|\cos \theta, \quad (1.44)$$

d.h. ist $\mathbf{x}'\mathbf{y} = \sum_i x_i y_i$, so folgt (1.43), und geht man von (1.43) aus, so folgt $\mathbf{x}'\mathbf{y} = \sum_i x_i y_i$. \square

Die algebraische Definition (1.23) des Skalarprodukts ist wesentlich für die Definition der Produkte Matrix \times Vektor und Matrix \times Matrix und ist die allgemeine Form der Ausdrücke für die in der multivariaten Statistik allgegenwärtigen Kovarianzen und Korrelationen (s. Abschnitt 1.2.4, Seite 21). Die Definition (1.43) wird oft bei Anwendungen der Vektorrechnung in der Physik oder der Technik als Definition des Skalarprodukts gewählt, s. z.B. Greiner (1981)⁹, Wylie

⁹Greiner, W.: Theoretische Physik, Band 1: Mechanik

(1975)¹⁰, aber auch in Lehrbüchern der Mathematik, etwa Courant (1963)¹¹. \square

Definition 1.4 Für $\theta = \pi/2$ (90°) bilden \mathbf{x} und \mathbf{y} einen rechten Winkel, sie heißen dann orthogonal¹². Gilt außerdem $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$, so heißen sie orthonormal.

Folgerung: Für $\theta = \pi/2$ gilt $\cos \theta = 0$, und für $\|\mathbf{x}\|, \|\mathbf{y}\| \neq 0$ folgt dann $\mathbf{x}'\mathbf{y} = 0$. Ebenso folgt aus $\mathbf{x}'\mathbf{y} = 0$ dann $\cos \theta = 0$, also $\theta = \pi/2$. $\mathbf{x}'\mathbf{y} = 0$ impliziert also die Orthogonalität bzw. die Orthonormalität von \mathbf{x} und \mathbf{y} . \square

Beispiel 1.4 Einheitsvektoren Nach Gleichung (1.32), Seite 16 sind die Skalarprodukte $\mathbf{e}'_j \mathbf{e}_k$ von Einheitsvektoren gleich Null für $j \neq k$, d.h. die \mathbf{e}_j und \mathbf{e}_k sind orthogonal, und da $\mathbf{e}'_j \mathbf{e}_j = 1$ für alle j folgt weiter, dass die Einheitsvektoren darüber hinaus auch orthonormal sind. Sie definieren demnach ein n -dimensionales Koordinatensystem, dessen Achsen n verschiedenen Orientierungen des \mathbb{R}^n angeben. Nach Gleichung (1.18), Seite 13, können die Komponenten $x_i, i = 1, \dots, n$ eines Vektors \mathbf{x} als Koordinaten des Endpunktes von \mathbf{x} auf diesen Achsen interpretiert werden. \square

Beispiel 1.5 Linearkombinationen und Skalarprodukte Es seien $\mathbf{x}_1, \dots, \mathbf{x}_n$ m -dimensionale Vektoren, und es sei $\mathbf{y} = a_1 \mathbf{x}_1 + \dots + a_n \mathbf{x}_n$ eine Linearkombination dieser Vektoren. Dann ist die Komponente y_i von \mathbf{y} durch das Skalarprodukt

$$y_i = \mathbf{a}' \tilde{\mathbf{x}}_i, \quad i = 1, 2, \dots, m \quad (1.45)$$

gegeben, wobei $\mathbf{a} = (a_1, a_2, \dots, a_n)'$ und $\tilde{\mathbf{x}}_i = (x_{i1}, x_{i2}, \dots, x_{in})'$ ist; x_{i1} ist die i -te Komponente von \mathbf{x}_1 , x_{i2} ist die i -te Komponente von \mathbf{x}_2 , etc. Denn es ist ja

$$\begin{aligned} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} &= \sum_{j=1}^n a_j \mathbf{x}_j = \begin{pmatrix} a_1 x_{11} + a_2 x_{12} + \dots + a_n x_{1n} \\ a_1 x_{21} + a_2 x_{22} + \dots + a_n x_{2n} \\ \vdots \\ a_1 x_{m1} + a_2 x_{m2} + \dots + a_n x_{mn} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{a}' \tilde{\mathbf{x}}_1 \\ \mathbf{a}' \tilde{\mathbf{x}}_2 \\ \vdots \\ \mathbf{a}' \tilde{\mathbf{x}}_m \end{pmatrix}. \end{aligned} \quad (1.46)$$

Von diesem Zusammenhang zwischen Linearkombination und Skalarprodukt wird u.a. bei der Definition der Matrixmultiplikation Gebrauch gemacht. \square

¹⁰Wylie, C.R.: Advanced Engineering Mathematics, Tokyo, Auckland etc 1975

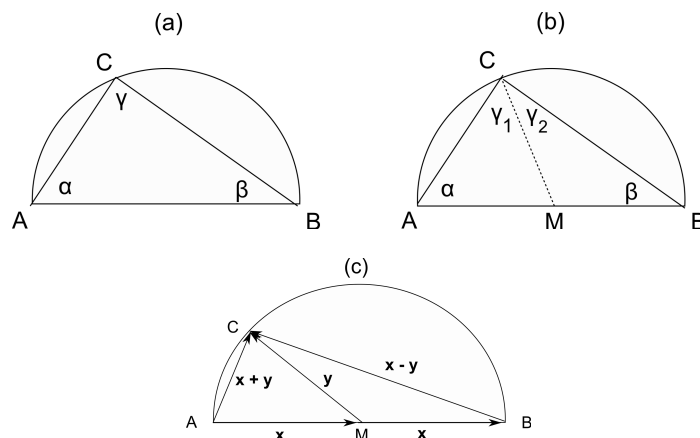
¹¹Courant, R. Vorlesungen über Differential- und Integralrechnung, Zweiter Band. Berlin 1963; Courant (p. 7) verweist darauf, dass eine Kraft \mathbf{K} , die einen Massenpunkt längs der gerichteten Strecke [Vektor] \mathbf{v} bewegt, dabei die *Arbeit* $\mathbf{K}'\mathbf{v} = \|\mathbf{K}\|\|\mathbf{v}\| \cos \theta$ leistet.

¹²Von griechisch *orthogonios*: ortho = rechts, gon = winklig

Anmerkungen: Für $\theta = \pi/2$, also für einen Winkel von 90° , folgt $\cos \theta = 0$ und (1.40) liefert den Satz des Pythagoras in Vektorschreibweise. 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 (1.39) als als *Definition* des Kosinus des Winkels und damit des Winkels selbst zwischen zwei Vektoren aufgefasst werden. \square

Beispiel 1.6 Der Satz des Thales¹³ Dieses Beispiel dient noch einmal als Übung bezüglich der Addition, Subtraktion und dem Skalarprodukt zweier Vektoren. Nach Thales ist der Winkel γ des im Halbkreis eingeschriebenen Dreiecks ABC stets ein rechter Winkel (90° , entsprechend $\pi/2$), unabhängig vom Ort C auf dem Halbkreis, s. Abb. 6 (a). Es gibt verschiedene Beweise für diese Aussage. Kann man die Aussagen (i) Die Winkelsumme im Dreieck beträgt 180° bzw π , und (ii) Die einander gegenüber liegenden Winkel eines gleichschenkligen Dreiecks sind stets gleich groß als bereits bewiesen voraussetzen, so gilt wegen $\overline{AC} = \overline{MC}$ (= Radius des Kreises) und $\gamma_1 = \alpha$, $\gamma_2 = \beta$, $\gamma = \gamma_1 + \gamma_2$, dass $\alpha + \beta + \gamma = 2\gamma = \pi$, also $\gamma = \pi/2$, vergl. Abb. 6 (b). Macht man von der

Abbildung 6: Der Satz des Thales



Vektorrechnung Gebrauch, so muß man die Aussagen (i) und (ii) nicht mehr voraussetzen; der folgende Beweis ist eine Übung in der Bildung von Vektorsummen und -differenzen (vergl. Abb. 1, Seite 8, Abb. 3, Seite 12, sowie Abb 4, Seite 12, sowie die Anmerkung nach Gleichung (1.15)), sowie der Bildung und Interpretation eines Skalarprodukts (Abb. 6, (c)). Man setze $\mathbf{x} = \overline{AM} = \overline{MB}$, $\mathbf{y} = \overline{MC}$, vergl. Abb. 6, und es gilt $\|\mathbf{x}\| = \|\mathbf{y}\|$. Weiter ist $\overline{AC} = \mathbf{x} + \mathbf{y}$, $\overline{BC} = \mathbf{x} - \mathbf{y}$. Dann folgt für das Skalarprodukt

$$(\mathbf{x} + \mathbf{y})'(\mathbf{x} - \mathbf{y}) = \mathbf{x}'\mathbf{x} - \mathbf{x}'\mathbf{y} + \mathbf{y}'\mathbf{x} - \mathbf{y}'\mathbf{y} = \|\mathbf{x}\|^2 - \|\mathbf{y}\|^2 = 0,$$

¹³Thales von Milet, 624 v.Chr. bis 546 v.Chr.; Thales hat als Erster strenge Beweise in die Mathematik eingeführt, und der hier genannte Satz war der erste streng bewiesene Satz.

also ist der Winkel zwischen den beiden Vektoren gleich 90° entsprechend (1.39), Seite 17, d.h. gemäß Definition 1.4 sind die Vektoren $\mathbf{x} + \mathbf{y}$ und $\mathbf{x} - \mathbf{y}$ orthogonal. \square

1.2.4 Zentrierte Vektoren, Varianzen und Kovarianzen

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

$$\bar{x} = \frac{1}{n} X' \vec{1} = \frac{1}{n} \vec{1}' X = \frac{1}{n} \sum_{i=1}^n X_i \quad (1.47)$$

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

$$\mathbf{x} = X - \bar{x} \vec{1} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} - \begin{pmatrix} \bar{x} \\ \bar{x} \\ \vdots \\ \bar{x} \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad (1.48)$$

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

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

und wenn \mathbf{y} ein ebenfalls zentrierter n -dimensionaler Vektor ist, so ist¹⁴

$$\text{Kov}_{xy} = \frac{1}{n} \mathbf{x}' \mathbf{y} = \frac{1}{n} \sum_{i=1}^n x_i y_i = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{x})(Y_i - \bar{y}) \quad (1.49)$$

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

$$s_x^2 = \frac{1}{n} \mathbf{x}' \mathbf{x} = \frac{1}{n} \|\mathbf{x}\|^2 \quad (1.50)$$

für die (Stichproben-)Varianz; s_y^2 ist analog definiert. Dann ist

$$r_{xy} = \frac{\mathbf{x}' \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|} = \frac{\text{Kov}(x, y)}{s_x s_y} \quad (1.51)$$

die Produkt-Moment-Korrelation der Messwerte X_i und Y_i . Mit Bezug auf (1.39) folgt daraus für zentrierte Vektoren

$$\cos \theta = r_{xy}. \quad (1.52)$$

Man beachte, dass sich die Faktoren $1/n$ bzw. $1/(n-1)$ in den Ausdrücken für die (Stichproben-)Varianzen und die -Kovarianz im Ausdruck für die Korrelation herauskürzen.

¹⁴Der Grund für die Wahl des Faktors $1/n$ statt $1/(n-1)$ ist, dass Varianzen und Kovarianzen dem Konzept nach Mittel- bzw. Erwartungswerte sind und es hier nur auf das jeweilige Konzept ankommt; der Faktor $1/(n-1)$ wird verwendet, um einen Bias bei den Stichprobenschätzungen auszugleichen.

1.2.5 Länge, Orientierung und Kosinus-Ähnlichkeit von Vektoren

Die Länge $\|\mathbf{x}\|$ eines Vektors \mathbf{x} wurde bereits in (1.34), Seite 16 über das Skalarprodukt eines Vektors mit sich selbst eingeführt; für den n -dimensionalen Fall ist nach dem verallgemeinerten Satz des Pythagoras die Definition der Länge durch

$$\|\mathbf{x}\| = \sqrt{\sum_{i=1}^n x_i^2}. \quad (1.53)$$

gegeben. Die Orientierung eines Vektors ist im n -dimensionalen Fall über (1.39), Seite 17 erklärt:

$$\cos \theta_j = \frac{\mathbf{x}'\mathbf{e}_j}{\|\mathbf{x}\|\|\mathbf{e}_j\|} = \frac{x_j}{\|\mathbf{x}\|},$$

und demnach ist

$$x_j = \|\mathbf{x}\| \cos \theta_j, \quad j = 1, \dots, n. \quad (1.54)$$

Jeder Vektor \mathbf{x} kann also in der Form

$$\mathbf{x} = \|\mathbf{x}\| \begin{pmatrix} \cos \theta_1 \\ \cos \theta_2 \\ \vdots \\ \cos \theta_n \end{pmatrix} = \|\mathbf{x}\| \mathbf{c}_x \quad (1.55)$$

angeschrieben werden, \mathbf{c}_x der zu \mathbf{x} gehörende Vektor der Richtungskosinus. Der Vektor \mathbf{c}_x definiert die Orientierung des Vektors, so dass \mathbf{c}_x auch *Orientierungsvektor* genannt wird.

Für den normierten Vektor $\mathbf{x}/\|\mathbf{x}\|$ und damit für den Orientierungsvektor \mathbf{c}_x gilt dann

$$\frac{\mathbf{x}}{\|\mathbf{x}\|} = \|\mathbf{c}_x\| = 1. \quad (1.56)$$

Zur Bedeutung von Orientierungsvektoren Es sei \mathbf{c} ein Orientierungsvektor, d.h. die Komponenten c_j von \mathbf{c} sind durch $c_j = \cos \theta_j$ definiert. Läßt man $a \in \mathbb{R}$ alle Werte im offenen Intervall $(-\infty, \infty)$ annehmen, so entsteht einer Gerade durch den Nullpunkt des Koordinatensystems mit einer durch \mathbf{c} definierten Orientierung. Repräsentieren die Achsen des Koordinatensystems irgendwelche Merkmale, so repräsentiert ein Punkt auf der Geraden bestimmte Ausprägungen $x_j = ac_j$, $j = 1, \dots, n$ dieser Merkmale, wobei die Ausprägungen in einem durch die Orientierung der Geraden bestimmten Verhältnis stehen. Die Gerade definiert gewissermassen das Mischungsverhältnis der Ausprägungen x_j ; es ist für alle Punkte auf der Geraden identisch.

Kosinus-Ähnlichkeit von Vektoren: Nach (1.56) gilt $\mathbf{x}/\|\mathbf{x}\| = \mathbf{c}_x$. Aus (1.39) folgt dann

$$\cos \theta = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = \mathbf{c}'_x \mathbf{c}_y \quad (1.57)$$

Das Skalarprodukt $\mathbf{c}'_x \mathbf{c}_y$ kann als Maß für die Ähnlichkeit der Vektoren \mathbf{x} und \mathbf{y} in Bezug auf ihre Orientierung interpretiert werden; man spricht von der *Kosinus-Ähnlichkeit* von \mathbf{x} und \mathbf{y} . Die Kosinus-Ähnlichkeit ist ein Maß für die Ähnlichkeit der Mischungsverhältnisse der Merkmale, die durch die Koordinatenachsen repräsentiert werden. Die Ähnlichkeit ist maximal für $\theta = 0$, wenn also \mathbf{x} und \mathbf{y} dieselbe Orientierung haben; dann ist $\cos \theta = 1$. In diesem Fall gilt für die Komponenten von \mathbf{x} und \mathbf{y} $y_i = ax_i$ mit $0 \neq a \in \mathbb{R}$ für alle i . In diesem Fall ist die Zusammensetzung der beiden Vektoren identisch in dem Sinne, dass

$$\frac{x_i}{x_{i'}} = \frac{y_i}{y_{i'}}, \quad i, i' = 1, \dots, m$$

Für $\cos \theta = 0$, also für $\theta = \pi/2$ ist diese Zusammensetzung maximal unterschiedlich.

Für $\cos \theta < 0$ hat man eine entgegengerichtete Ähnlichkeit. □

1.2.6 Vektorprojektionen

Orthogonale Projektion eines Punktes auf eine Gerade Gegeben sei eine Gerade $G : \mathbf{r}_0 + \lambda \mathbf{y}$, wobei \mathbf{r}_0 ein Orts- oder Stützvektor und \mathbf{y} ein Orientierungsvektor ist, und $\lambda \in \mathbb{R}$ ist ein Parameter; G wird erzeugt, wenn λ die Menge der reellen Zahlen durchläuft. \mathbf{y} bestimmt die Orientierung der Geraden im Raum (s. Abbildung 7). Weiter sei P ein Punkt mit den Koordinaten (x_1, x_2) die als Komponenten eines Vektors $\mathbf{x} = (x_1, x_2)'$ betrachtet werden können. Gesucht ist die orthogonale Projektion von P auf G , d.h. der Punkt $Q = Q(\mathbf{x})$ derart, dass der Vektor \mathbf{z} von P nach Q senkrecht auf G steht, so dass $\mathbf{z}'\mathbf{y} = 0$. Man hat also die folgenden Bedingungen:

1. Es existiert ein $\lambda = \lambda_0$ derart, $Q(\mathbf{x}) = \mathbf{r}_0 + \lambda_0 \mathbf{y}$
2. $\mathbf{z} = \mathbf{x} - (\mathbf{r}_0 + \lambda_0 \mathbf{y})$, $\mathbf{z}'\mathbf{y} = 0$.

Es muß der Wert von λ_0 bestimmt werden. Dazu werde das Skalarprodukt $\mathbf{z}'\mathbf{y}$ betrachtet:

$$\begin{aligned} \mathbf{z}'\mathbf{y} &= (\mathbf{x} - (\mathbf{r}_0 + \lambda_0 \mathbf{y}))'\mathbf{y} \\ &= \mathbf{x}'\mathbf{y} - \mathbf{r}'_0 \mathbf{y} - \lambda_0 \mathbf{y}'\mathbf{y} = 0 \end{aligned}$$

d.h.

$$\mathbf{x}'\mathbf{y} = \mathbf{r}'_0 \mathbf{y} + \lambda_0 \mathbf{y}'\mathbf{y},$$

so dass

$$\lambda_0 = \frac{(\mathbf{x} - \mathbf{r}_0)'\mathbf{y}}{\mathbf{y}'\mathbf{y}} \tag{1.58}$$

Setzt man diesen Ausdruck für λ_0 in den für $Q(x)$ ein, so erhält man

$$Q(\mathbf{x}) = \mathbf{r}_0 + \left(\frac{(\mathbf{x} - \mathbf{r}_0)'\mathbf{y}}{\mathbf{y}'\mathbf{y}} \right) \mathbf{y}. \tag{1.59}$$

Spezialfälle: (i) $\mathbf{r}_0 = \vec{0}$: Die Gerade läuft durch den Koordinatenursprung 0

Abbildung 7: Die Projektion eines Punktes auf eine Gerade

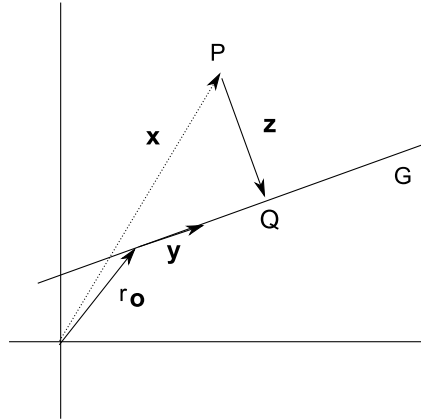
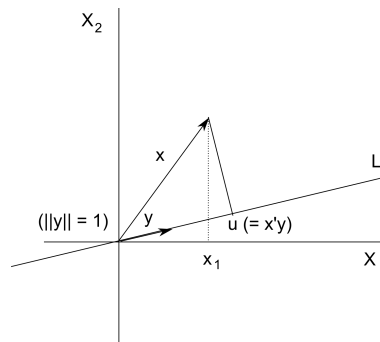


Abbildung 8: Neue Koordinate u als Projektion eines Vektors auf eine Gerade L



und man erhält

$$Q(\mathbf{x}) = \mathbf{x}_y = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{y}'\mathbf{y}}\mathbf{y} \quad (1.60)$$

Dies ist ebenfalls die Formel, die man erhält, wenn man die orthogonale Projektion \mathbf{x}_y eines Vektors \mathbf{x} auf einen Vektor \mathbf{y} bestimmt, denn für $\mathbf{r}_0 = \vec{0}$ fallen die Anfangspunkte von \mathbf{x} und \mathbf{y} zusammen, vergl. Abbildung 7. (ii) $\mathbf{r}_0 = \vec{0}$ und $\|\mathbf{y}\| = 1$: In diesem Fall hat man nach (1.58) und (1.60) wegen $\mathbf{y}'\mathbf{y} = 1$ ist $\lambda_0 = \mathbf{x}'\mathbf{y}$ und man erhält

$$\mathbf{x}_y = (\mathbf{x}'\mathbf{y})\mathbf{y} \quad (1.61)$$

Diese Beziehung bedeutet, dass das Skalarprodukt $\mathbf{x}'\mathbf{y}$ als Koordinate des Endpunktes von \mathbf{x} auf einer Geraden ("Achse") interpretiert werden kann, deren Orientierung durch die des normierten Vektors \mathbf{y} gegeben ist.

1.2.7 Die Cauchy-Schwarzsche Ungleichung

Die Gleichung (1.39), Seite 17 impliziert den

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

$$|\mathbf{x}'\mathbf{y}|^2 \leq \|\mathbf{x}\|^2 \|\mathbf{y}\|^2. \quad (1.62)$$

Beweis: Gleichung (1.39) impliziert

$$\mathbf{x}'\mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos \theta, \quad (1.63)$$

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 (1.62). \square

Korollar 1.1 *Es gilt*

$$\mathbf{x}'\mathbf{y} = \|\mathbf{y}\| \|\mathbf{x}\| \quad (1.64)$$

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

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

Die Cauchy-Schwarzsche Ungleichung (1.62) wird oft in der Form

$$\left| \sum_{i=1}^n x_i y_i \right|^2 \leq \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2 \quad (1.65)$$

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

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

d.h.

$$\left| \sum_{i=1}^n x_i y_i \right|^2 = \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2 = \lambda \left(\sum_{i=1}^n x_i^2 \right)^2. \quad (1.66)$$

(s.a. (1.63)).

1.2.8 Lineare Abhängigkeit und Unabhängigkeit von Vektoren

Der Begriff der Unabhängigkeit etwa von Variablen ist, wie die Betrachtungen zum Skalarprodukt gezeigt haben, mehrdeutig: gilt für zwei n -dimensionale Vektoren \mathbf{x} und \mathbf{y} $\mathbf{x}'\mathbf{y} = 0$ und sind also \mathbf{x} und \mathbf{y} orthogonal, so werden die

Variablen, die durch \mathbf{x} und \mathbf{y} repräsentiert werden, oft als unabhängig voneinander betrachtet, was aber nicht notwendig auch stochastisch unabhängig bedeutet, weil eine Korrelation gleich Null zwischen zwei Variablen X und Y noch nicht bedeutet, dass für die gemeinsame Dichtefunktion $f_{xy}(X, Y)$ die Beziehung $f_{xy}(X, Y) = f_x(X)f_y(Y)$ gilt. Der Fall $\mathbf{x}'\mathbf{y} \neq 0$ legt dann die Hypothese nahe, dass ein bestimmtes Ausmaß an Abhängigkeit zwischen X und Y existiert. In der linearen Algebra, die ja keine statistische Theorie ist, wird aber ein Abhängigkeits- bzw. Unabhängigkeitsbegriff eingeführt, der eine weitere Facette des Abhängigkeitsbegriffs liefert: selbst wenn $\cos \theta$ nur "unwesentlich" von 1 abweicht können zwei Vektoren dennoch im algebraischen Sinn unabhängig voneinander sein, wie gezeigt werden wird. Unter anderem für die Diskussion von Daten in Termen von unterliegenden, "latenten" Variablen ist dieser Sachverhalt von Bedeutung.

Gegeben sei eine Menge $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p\}$ von m -dimensionalen Vektoren $\neq \vec{0}$, also $\vec{0} \neq \mathbf{x}_k \in \mathbb{R}^m$ für alle $k = 1, \dots, p$. Zwischen diesen Vektoren können Abhängigkeiten existieren, die sich durch Linearkombinationen ausdrücken lassen; so könnte etwa \mathbf{x}_1 als Linearkombination

$$\mathbf{x}_1 = a_2\mathbf{x}_2 + \dots + a_p\mathbf{x}_p \quad (1.67)$$

repräsentierbar sein. Durch einfache Umformung können dann auch andere Vektoren \mathbf{x}_j aus der betrachteten Menge als Linearkombination der übrigen dargestellt werden, etwa \mathbf{x}_2 , falls $a_2 \neq 0$, man erhält

$$\mathbf{x}_2 = -\mathbf{x}_1 + \frac{a_3}{a_2}\mathbf{x}_3 + \dots + \frac{a_p}{a_2}\mathbf{x}_p,$$

etc. Sind derartige Darstellungen der Vektoren möglich, kann man von *linearer Abhängigkeit* der Vektoren sprechen. Von *linearer Unabhängigkeit* der Vektoren wäre dann die Rede, wenn man für die Darstellung von \mathbf{x}_1 in (1.67) keine Koeffizienten $a_k \neq 0$ finden kann (analog für die Darstellung von \mathbf{x}_2 etc), d.h. wenn derartige Koeffizienten nicht existieren. "Nicht existieren" ist hier ein intuitiver Begriff, der algebraisch ausgedrückt werden muss, etwa um die Frage beantworten zu können, ob die Vektoren der Menge $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ linear unabhängig sind. Dazu kann die obige Gleichung für \mathbf{x}_1 auf eine kanonische Form gebracht werden, indem man \mathbf{x}_1 auf beiden Seiten subtrahiert. Es entsteht die Gleichung

$$\begin{aligned} \vec{0} &= -\mathbf{x}_1 + a_2\mathbf{x}_1 + \dots + a_p\mathbf{x}_p \\ &= a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \dots + a_p\mathbf{x}_p, \quad a_1 = -1. \end{aligned}$$

Man kann die Gleichung

$$\vec{0} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \dots + a_p\mathbf{x}_p \quad (1.68)$$

ohne weitere Vorannahmen über die Werte der Koeffizienten a_j als eine allgemeine Repräsentation des Nullvektors $\vec{0}$ als Linearkombination der \mathbf{x}_j sehen. Dann wird deutlich, dass diese Repräsentation für beliebige \mathbf{x}_j stets möglich ist: man muß nur alle Koeffizienten a_j gleich Null setzen, d.h. $\mathbf{a} = (a_1, \dots, a_p)' = \vec{0}$ ist stets eine

Lösung für die Gleichung (1.68). $\mathbf{a} = \vec{0}$ heißt dementsprechend auch die *triviale Lösung* für die Gleichung (1.68). Im Falle der linearen Abhängigkeit existiert dann aber außerdem ein Vektor $\mathbf{a} \neq \vec{0}$, denn in diesem Fall sind ja zumindest einige der Koeffizienten $a_j \neq 0$. Das Prädikat "linear abhängig" ist die Negation des Prädikats "linear unabhängig", d.h. die Negation der Aussage "einige der a_j sind ungleich Null". Diese Negation bedeutet dann "alle Koeffizienten a_j in (1.68) sind gleich Null". Man hat dementsprechend die

Definition 1.5 *es gelte (1.68); sind nicht alle Komponenten a_j des Vektors $\mathbf{a} = (a_1, \dots, a_p)'$ gleich Null, so heißen die Vektoren \mathbf{x}_j linear abhängig. Ist dagegen $\mathbf{a} = \vec{0}$ die einzige Lösung für (1.68), so heißen die \mathbf{x}_j linear unabhängig.*

Anmerkung: Der Nullvektor $\vec{0}$ ist linear abhängig. Denn es sei $\vec{0} = a\vec{0}$, – diese Beziehung gilt für alle $0 \neq a \in \mathbb{R}$. Also ist $\vec{0}$ linear abhängig. \square

Satz 1.3 *Die n -dimensionalen Einheitsvektoren $\mathbf{e}_1, \dots, \mathbf{e}_n$ sind linear unabhängig.*

Beweis: Zu zeigen ist, dass $a_1\mathbf{e}_1 + \dots + a_n\mathbf{e}_n = \vec{0}$ nur dann gilt, wenn $a_j = 0$ für alle j . Tatsächlich gilt für die i -te Komponente des Nullvektors

$$a_1 0 + \dots + a_i 1 + \dots + a_i 0 = 0, \quad i = 1, \dots, n$$

wobei $a_i 1$ für den i -ten Einheitsvektor gilt. Da $1 \neq 0$ folgt $a_i = 0$ für alle i . \square

Satz 1.4 *Die Vektoren $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^n$, $\mathbf{x}_j \neq \vec{0}$ 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 wegen der vorausgesetzten Orthogonalität der \mathbf{x}_j und \mathbf{x}_k , $j \neq k$ folgt $\mathbf{x}'_j\mathbf{x}_k = 0$ für alle $j \neq k$. Dann muss aber auch

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

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

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

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

Beweis: Es gilt

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

wie man unmittelbar verifiziert. \square

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

Beweis: Es ist

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

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

Dieser Sachverhalt bedeutet, dass jede Komponente eines Vektors zu einer Orientierung ("Dimension") im Raum korrespondiert, die senkrecht auf den Orientierungen steht, die zu den übrigen Komponenten korrespondieren. Der Sachverhalt erklärt damit auch den Ausdruck "n-dimensionaler Vektor".

Satz 1.6 *Es seien $\mathbf{x}_1, \dots, \mathbf{x}_p$ n-dimensionale Vektoren, von denen einer der Nullvektor $\vec{0}$ ist. Dann sind die \mathbf{x}_j , $j = 1, \dots, p$ linear abhängig.*

Beweis: Ohne Einschränkung der Allgemeinheit kann $\mathbf{x}_1 = \vec{0}$ angenommen werden, – die Nummerierung der Vektoren ist ja beliebig. Es existieren reelle Koeffizienten a_1, \dots, a_n derart, dass

$$a_1 \vec{0} + a_2 \mathbf{x}_2 + \cdots + a_p \mathbf{x}_p = \vec{0}.$$

Für die triviale Lösung $a_1 = a_2 = \cdots = a_n = 0$ ist diese Gleichung stets erfüllt. Sie ist ebenfalls stets erfüllt, wenn $a_1 \neq 0$: sind einige der a_j mit $j \neq 1$ ungleich Null, so sind die $\mathbf{x}_2, \dots, \mathbf{x}_n$ linear abhängig und damit sind auch die $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ linear abhängig. Diese Aussage gilt auch, wenn $a_2 = \cdots = a_p = 0$ gewählt wird; dann sind ebenfalls *nicht alle* $a_j = 0$, und das heißt wiederum, dass die $\mathbf{x}_1, \dots, \mathbf{x}_p$ linear abhängig sind. \square

Lineare Unabhängigkeit: Die lineare Unabhängigkeit ist keine Art von statistischer Unabhängigkeit, sondern eine rein algebraische. Es seien $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Für den von diesen beiden Vektoren gebildeten Winkel θ gilt dann

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

$\cos \theta = 0$ (entsprechend $\rho_{xy} = 0$) gilt für $\theta = \pi/2$, d.h. \mathbf{x} und \mathbf{y} sind orthogonal.

Dazu werde angenommen, dass weder \mathbf{x} noch \mathbf{y} gleich dem Nullvektor sind. Die beiden Vektoren sind parallel (in Zeichen: $\mathbf{x} \parallel \mathbf{y}$), wenn ein Skalar $a \in \mathbb{R}$ existiert derart, dass $\mathbf{y} = a\mathbf{x}$ und \mathbf{x} und \mathbf{y} sind linear abhängig, denn dann $\vec{0} = a\mathbf{x} - \mathbf{y} = a_1\mathbf{x} + a_2\mathbf{y}$ mit $a_1 = a$, $a_2 = -1$. Gilt aber $\mathbf{y} \not\parallel \mathbf{x}$, so läßt sich keiner der beiden Vektoren als Linearkombination des anderen darstellen, so dass die beiden Vektoren linear unabhängig sind. Um sich leicht auf diesen Sachverhalt beziehen zu können, soll er als Satz formuliert werden:

Satz 1.7 \mathbf{x} und \mathbf{y} seien zwei n -dimensionale Vektoren $\neq \vec{0}$. Der Winkel zwischen ihnen sei θ . \mathbf{x} und \mathbf{y} sind linear abhängig genau dann, wenn $\theta = 0$, d.h. wenn $\cos \theta = 1$.

Beweis: Es gelte $\mathbf{x} \parallel \mathbf{y}$, d.h. es existiert ein $a \in \mathbb{R}$ mit $\mathbf{y} = a\mathbf{x}$ und

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

d.h. $\theta = 0$ und es gilt $\vec{0} = a_1\mathbf{x} + a_2\mathbf{x}_2$ mit $a_1 = a \neq 0$ und $a_2 = -1$, so dass \mathbf{x}, \mathbf{y} linear abhängig sind.

Sind \mathbf{x} und \mathbf{y} linear unabhängig, so folgt, dass kein $a \in \mathbb{R}$ mit $\mathbf{y} = a\mathbf{x}$ existiert¹⁵, somit folgt $\mathbf{x} \not\parallel \mathbf{y}$, also $\theta \neq 0$ und somit $|\cos \theta| < 1$. \square

Der Begriff der linearen Unabhängigkeit ist im Unterschied zu dem der Korrelation binär: für $|\cos \theta| = 1$ sind die Vektoren \mathbf{y} linear abhängig, und für $|\cos \theta| < 1$ sind sie linear unabhängig. Für die Korrelation ist die statistische Abhängigkeit klar, wenn $|r_{xy}| = 1$, und für $|r_{xy}| < 1$ nimmt die Abhängigkeit einen Wert auf einem Spektrum von "fast abhängig" an, wenn $|r_{xy}| = 1 - \varepsilon$, $0 < \varepsilon$ beliebig klein, und "unabhängig", wenn $r_{xy} = 0$. Es sei aber daran erinnert, dass "unabhängig" nur dann "stochastisch unabhängig" bedeutet, wenn bestimmte Annahmen über die Wahrscheinlichkeitsdichte erfüllt sind, wenn z.B. die 2-dimensionale Gauss-Verteilung gegeben ist.

Algebraisch gesehen sind also die Messungen X und Y abhängig nur dann, wenn $|r_{xy}| = 1$, d.h. $\theta = 0$, und unabhängig immer dann, wenn $|r_{xy}| < 1$, also $|\theta| \neq 0$. Andererseits wird man diese Interpretation von r_{xy} aber nicht wählen, denn bei der algebraischen Interpretation werden die Komponenten x_i und y_i als messfehlerfrei angenommen, während bei realen Messungen stets die Existenz von Messfehlern mitgedacht werden muß. Üblicherweise wird angenommen, dass der Messfehler additiv ist. Sind als \mathbf{x}_0 und \mathbf{y}_0 die "wahren" Vektoren, so sind die gemessenen Vektoren durch

$$\mathbf{x} = \mathbf{x}_0 + \boldsymbol{\varepsilon}_x, \quad \mathbf{y} = \mathbf{y}_0 + \boldsymbol{\varepsilon}_y \tag{1.70}$$

gegeben, wobei $\boldsymbol{\varepsilon}_x$ und $\boldsymbol{\varepsilon}_y$ "Fehler" vektoren sind, deren Komponenten ε_{x_i} und ε_{y_i} Messfehler bei den i -ten Messungen repräsentieren. Es gelte $\mathbf{y}_0 = a\mathbf{x}_0$, d.h. die

¹⁵modus tollens: $(p \rightarrow q) \Rightarrow (\neg q \rightarrow \neg p)$.

”wahren” Vektoren seien parallel. Der Korrelationskoeffizient r_{xy} ist dann

$$\begin{aligned} -1 < r_{xy1} = \cos \theta_{xy} &= \frac{(\mathbf{x}_0 + \boldsymbol{\varepsilon}_x)'(a\mathbf{x}_0 + \boldsymbol{\varepsilon}_y)}{\|\mathbf{x}_0 + \boldsymbol{\varepsilon}_x\| \|a\mathbf{x}_0 + \boldsymbol{\varepsilon}_y\|} \\ &= \frac{a\|\mathbf{x}_0\|^2 + \mathbf{x}_0'(\boldsymbol{\varepsilon}_y + a\boldsymbol{\varepsilon}_x) + \boldsymbol{\varepsilon}_x'\boldsymbol{\varepsilon}_y}{\|\mathbf{x}_0 + \boldsymbol{\varepsilon}_x\| \|a\mathbf{x}_0 + \boldsymbol{\varepsilon}_y\|} < 1, \end{aligned} \quad (1.71)$$

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

Nun seien \mathbf{x}_0 und \mathbf{x}_0 nicht parallel, so dass $\theta \neq 0$. Man erhält

$$\begin{aligned} -1 < r_{xy2} = \cos \theta_{xy} &= \frac{(\mathbf{x}_0 + \boldsymbol{\varepsilon}_x)'(\mathbf{y}_0 + \boldsymbol{\varepsilon}_y)}{\|\mathbf{x}_0 + \boldsymbol{\varepsilon}_x\| \|\mathbf{y}_0 + \boldsymbol{\varepsilon}_y\|} \\ &= \frac{\mathbf{x}_0'\mathbf{y}_0 + \mathbf{x}_0'\boldsymbol{\varepsilon}_y + \mathbf{y}_0'\boldsymbol{\varepsilon}_x + \boldsymbol{\varepsilon}_x'\boldsymbol{\varepsilon}_y}{\|\mathbf{x}_0 + \boldsymbol{\varepsilon}_x\| \|\mathbf{y}_0 + \boldsymbol{\varepsilon}_y\|} < 1 \end{aligned} \quad (1.72)$$

Gegeben ist nur der empirische Wert r_{xy} , und die Frage ist, ob $r_{xy} = r_{xy1}$ oder $r_{xy} = r_{xy2}$ gilt. Insbesondere im Fall $\theta \neq 0$, die Abweichung von Null aber nur klein ist, wird die Frage schwer zu entscheiden sein, denn der Wert von θ ist unbekannt und kann nur über die Formel $\cos \theta = \mathbf{x}'\mathbf{y}/(\|\mathbf{x}\|\|\mathbf{y}\|)$ aus den Daten geschätzt werden. Bei der Diskussion der Anzahl r benötigter latenter Variablen für eine gegebene Menge $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ von Vektoren wird darauf zurückgekommen.

Der Fall $\theta = 0$ spielt in der Statistik eine Rolle und hat deshalb einen eigenen Namen:

Definition 1.6 *Es seien p_1, p_2, \dots, p_r r Punkte in einem n -dimensionalen Raum. Die Punkte heißen kollinear, wenn sie alle auf einer Geraden in diesem Raum liegen.*

Wenn die Rede von kollinearen Vektoren ist, so ist Parallelität gemeint, da dann die Vektoren als auf einer Geraden liegend betrachtet werden können, so dass ihre Anfangs- und Endpunkte alle auf der Geraden liegen. Die Vektoren sind dann linear abhängig.

1.3 Vektorräume

1.3.1 Definition eines Vektorraums

Ein Vektorraum V ist eine Menge von Vektoren gleicher Dimensionalität, die abgeschlossen gegenüber der Bildung von Linearkombinationen ist, d.h. sind

$\mathbf{x}_1, \dots, \mathbf{x}_k \in V$, so ist auch $a_1\mathbf{x}_1 + \dots + a_k\mathbf{x}_k \in V$. Ist W eine Teilmenge von V und gilt die Abgeschlossenheit auch für W , so ist W ein Teilvektorraum. Üblicherweise wird noch die Möglichkeit, für irgendzwei Vektoren aus V oder W ein Skalarprodukt zu bilden genannt. Für die hier behandelten Vektoren ist diese Möglichkeit immer gegeben. Eine strengere Definition wird noch gegeben, davor soll aber illustriert werden, warum die Begriffe Vektorraum bzw. Teilvektorraum nützlich sind.

Eindimensionaler Teilraum: Man erinnere sich daran, dass der Nullvektor $\vec{0}$ für sich genommen linear abhängig ist (vergl. Anmerkung zur Definition 1.5, Seite 27. Nun sei $\mathbf{u} \neq \vec{0}$ ein beliebiger n -dimensionaler Vektor, und gelte $\vec{0} = a\mathbf{u}$, $a \in \mathbb{R}$. Da \mathbf{u} als verschieden vom Nullvektor $\vec{0}$ vorausgesetzt wurde kann diese Gleichung nur gelten, wenn $a = 0$. Also ist \mathbf{u} linear unabhängig. Nun seien $\mathbf{x}_1 = a_1\mathbf{u}$, $\mathbf{x}_2 = a_2\mathbf{u}$, $a_1, a_2 \neq 0$, und $\mathbf{x} = b_1\mathbf{u}_1 + b_2\mathbf{x}_2$ sei eine beliebige Linearkombination von \mathbf{x}_1 und \mathbf{x}_2 . Es folgt

$$\mathbf{x} = b_1\mathbf{u}_1 + b_2\mathbf{x}_2 = b_1a_1\mathbf{u} + b_2a_2\mathbf{u} = (b_1a_1 + b_2a_2)\mathbf{u}.$$

Offenbar haben $\mathbf{u}_1, \mathbf{u}_2$ und \mathbf{x} Linearkombinationen von \mathbf{u} und alle haben dieselbe Orientierung und können deshalb auf ein und derselben Geraden, deren Orientierung durch \mathbf{u} gegeben ist repräsentiert werden. Die Vektoren auf dieser Geraden bilden eine abgeschlossene Menge bezüglich der Bildung von Linearkombinationen und bilden deshalb einen 1-dimensionalen Teilraum in der Menge aller n -dimensionalen Vektoren. Es sei noch angemerkt, dass $a_1\mathbf{u}$ und $a_2\mathbf{u}$ linear abhängig sind. Denn aus

$$\vec{0} = a_1\mathbf{u} + a_2\mathbf{u} = (a_1 + a_2)\mathbf{u}, \quad \mathbf{u} \neq \vec{0}$$

folgt $a_1 + a_2 = 0$, d.h. $a_2 = -a_1$, d.h. für beliebige $0 \neq a_1 \in \mathbb{R}$ existiert stets ein $0 \neq a_2 \in \mathbb{R}$ derart, dass $\vec{0} = a_1\mathbf{u} + a_2\mathbf{u}$ erfüllt ist. Es genügt ein Vektor $\mathbf{u} \neq \vec{0}$, um alle Vektoren auf der durch \mathbf{u} definierten Geraden als Linearkombination von \mathbf{u} darzustellen.

Zweidimensionaler Teilraum: Nun seien \mathbf{u}_1 und \mathbf{u}_2 zwei n -dimensionale Vektoren, die nicht parallel sind, $\mathbf{u}_1 \nparallel \mathbf{u}_2$. \mathbf{u}_1 und \mathbf{u}_2 sind linear unabhängig. Denn angenommen, sie seien linear abhängig, so gilt

$$\vec{0} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2, \quad a_1, a_2 \neq 0$$

und es folgt

$$\mathbf{u}_2 = -\frac{a_1}{a_2}\mathbf{u}_1,$$

d.h. die beiden Vektoren sind parallel, entgegen der Voraussetzung, also sind sie linear abhängig. Man sieht leicht, dass die beiden Vektoren nun eine Ebene "aufspannen". Dazu werde eine beliebige Linearkombination

$$\mathbf{v} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2$$

betrachtet. Der Anschaulichkeit halber sei $n = 3$. Dann existiert stets ein Vektor \mathbf{n} , der senkrecht auf sowohl \mathbf{u}_1 wie auch \mathbf{u}_2 steht. Dann steht er auch senkrecht auf \mathbf{v} , denn

$$\mathbf{n}'\mathbf{v} = a_1\mathbf{n}'\mathbf{u}_1 + a_2\mathbf{n}'\mathbf{u}_2 = 0,$$

da nach Voraussetzung $\mathbf{n}'\mathbf{u}_1 + \mathbf{n}'\mathbf{u}_2 = 0$. Da dieser Befund für alle Linearkombinationen von \mathbf{u}_1 und \mathbf{u}_2 gilt heißt dies, dass die Ebene durch diese beiden erzeugt werden kann; die Orientierung der Ebene kann als durch \mathbf{n} definiert gelten, denn \mathbf{n} steht auf allen Punkten der Ebene senkrecht. Es sei nun \mathbf{w} ein Vektor, der nicht in dieser Ebene liegt. Da *alle* Linearkombinationen von \mathbf{u}_1 und \mathbf{u}_2 in der Ebene liegen, folgt, dass \mathbf{w} *nicht* als Linearkombination der Vektoren \mathbf{u}_1 und \mathbf{u}_2 werden kann. Da \mathbf{u}_1 und \mathbf{u}_2 beliebig bis auf die Forderung $\mathbf{u}_1 \nparallel \mathbf{u}_2$ gewählt werden können, folgt, dass man auch zwei Linearkombinationen \mathbf{v}_1 und \mathbf{v}_2 von $\mathbf{u}_1 \nparallel \mathbf{u}_2$ dazu benutzen kann, alle Vektoren der Ebene zu erzeugen, sofern nur $\mathbf{v}_1 \nparallel \mathbf{v}_2$ gilt, d.h. insbesondere können $\mathbf{u}_1 \nparallel \mathbf{u}_2$ als Linearkombinationen der $\mathbf{v}_1 \nparallel \mathbf{v}_2$ dargestellt werden. Man zeigt leicht, dass jede Linearkombination von Vektoren aus der Ebene wieder ein Vektor aus der Ebene ist; man kann sagen, dass die Menge der Linearkombinationen von Vektoren abgeschlossen bezüglich der Bildung von Linearkombinationen von Vektoren aus der Ebene ist. Man bemerke dass diese Betrachtungen zwar für 3-dimensionale Vektoren angestellt wurden, die Zahl 3 dabei aber gar keine Rolle spielt, bis auf die Notwendigkeit, einen zu $\mathbf{u}_1 \nparallel \mathbf{u}_2$ orthogonalen Vektor annehmen zu können. Die Betrachtungen gelten für alle $n > 2$.

Dreidimensionaler Teilraum: Ebenso kann man n -dimensionale Vektoren $\mathbf{u}_1, \mathbf{u}_2$ und \mathbf{u}_3 mit der Bedingung $n > 3$ betrachten; man kann dann einen Vektor \mathbf{n} finden, der auf allen drei Vektoren senkrecht steht, so dass $\mathbf{n}'\mathbf{u}_k = 0$ für $k = 1, 2, 3$. \mathbf{n} steht dann auf allen Linearkombinationen von $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ senkrecht; es gelingt nicht, einen Linearkombination der $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ zu finden, auf der \mathbf{n} nicht senkrecht stünde, d.h. die durch diese Vektoren gebildeten Linearkombinationen bilden eine Teilmenge aller n -Vektoren, und diese Teilmenge ist abgeschlossen bezüglich der Bildung von Linearkombinationen. Man kann diese Betrachtungen allgemein für $r \leq n$ linear unabhängige n -dimensionale Vektoren durchführen, das Prinzip wird deutlich geworden sein.

Es liegt nahe, zu vermuten, dass man n linear unabhängige Vektoren $\mathbf{u}_1, \dots, \mathbf{u}_n$ benötigt, um alle n -dimensionalen Vektoren als Linearkombinationen der \mathbf{u}_k , $k = 1, \dots, n$ darstellen zu können, die Menge V aller n -dimensionalen Vektoren ist dann ein n -dimensionaler Vektorraum.

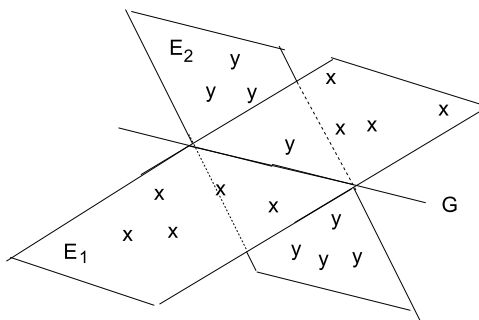
Abb. 9 Teilräume

Die formale Definition eines Vektorraums ist nun:

Definition 1.7 *Es sei V eine Menge von n -dimensionalen Vektoren aus \mathbb{R}^n , und für beliebige $\mathbf{x}_1, \mathbf{x}_2 \in V$ gelte*

1. *Es sei $\mathbf{u} \in V$ und $a \in \mathbb{R}$. Dann ist auch $a\mathbf{u} \in V$.*
2. *$\mathbf{x} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 \in V$ mit $a_1, a_2 \in \mathbb{R}$,*

Abbildung 9: Ausschnitte 2-dimensionaler Teilräume E_1 und E_2 sowie ein 1-dimensionaler Teilraum G eines 3-dimensionalen Vektorraums; x : Teilmenge von Fällen, die auf E_1 leben, y : Teilmenge von Fällen, die auf E_2 leben, und z : Teilmenge von Fällen, die auf dem 1-dimensionalen Durchschnitt G von E_1 und E_2 leben.



3. Für $\mathbf{x}_1, \mathbf{x}_2 \in V$ existiert das Skalarprodukt $\mathbf{x}'_1 \mathbf{x}_2$.

Dann heißt V n -dimensionaler Vektorraum über \mathbb{R} .

4. Es sei $U \subseteq V$ eine Teilmenge von Vektoren aus V . U heißt Teilvektorraum oder Untervektorraum von V , wenn U selbst wieder ein Vektorraum ist, d.h. wenn für die Elemente von U wieder die Forderungen 1. bis 3. gelten, wobei V durch U ersetzt wird. Die leere Menge $\{\emptyset\}$ und der Vektorraum V selbst heißen triviale Teilräume.

Anmerkungen:

1. Diese Definition ist nicht die allgemeine, in Lehrbüchern der Linearen Algebra gegebene Definition, weil sie sich nur auf den in der multivariaten statistischen Analyse dominierenden Vektorbegriff bezieht, d.h. auf Vektoren mit endlicher Dimensionalität n und reellen Komponenten. Eine erste Verallgemeinerung besteht darin, auch komplexwertige Komponenten zuzulassen, so dass $x_i = a_i + jb_i$ mit $a_i, b_i \in \mathbb{R}$ und $j = \sqrt{-1}$, kurz $x_i \in \mathbb{C}$. Die Rede ist dann von einem Vektorraum über \mathbb{C} . Eine wichtige Verallgemeinerung entsteht, wenn Funktionen f über einem Intervall $I = [a, b]$ betrachtet werden, $a, b \in \mathbb{R}$. Für jeden Wert $x \in I$ nimmt f einen Wert $f(x)$ an, und diese Werte können ebenfalls als Komponenten eines Vektors interpretiert werden. Man kann dann Linearkombinationen der Form $f = a_1 f_1 + a_2 f_2$ aus Funktionen f_1 und f_2 über I bilden und u. U. auch Skalarprodukte, die durch Integrale der Form $\lambda = \int_I \int_I f_j(x) f_k(x) dx$ definiert sind. Da die Anzahl der x -Werte aus I überabzählbar unendlich ist bilden die Funktionen einen unendlich-dimensionalen Vektorraum; in Abschnitt 4 wird ausführlicher auf diesen Fall eingegangen. Allgemein kann \mathbb{R} durch eine Menge K ersetzt werden, aus deren Elementen Linearkombinationen und Skalarprodukte definiert werden können (K bildet dann einen Körper). Eine allgemeine Definition des Begriffs 'Körper' findet man im Anhang. In der Definition 1.7 ist $K = \mathbb{R}$ gesetzt worden.

2. Ein Vektorraum ist eine spezielle *algebraische Struktur*, d.h. eine Menge von Elementen, für die bestimmte Verknüpfungen definiert sind. Die genannten Verknüpfungen (Addition und Bildung von Skalarprodukten) setzen voraus, dass die Vektoren der Menge dieselbe Anzahl von Komponenten haben.
3. Ein Beispiel für eine Menge \mathcal{M} von n -dimensionalen Vektoren, die *keinen* Vektorraum bilden ist die Menge $\mathcal{M} = \{\mathbf{x} | \mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\| = 1\}$. Diese Menge ist kein Vektorraum, denn die Linearkombinationen $\mathbf{x} = \sum_{j=1}^p a_j \mathbf{x}_j$, $p > 1$, haben, für beliebige Koeffizienten a_j , nicht mehr notwendig die Länge 1, d.h. es gilt nicht notwendig $\mathbf{x} \in \mathcal{M}$. \square

Wie in den einführenden Beispielen gezeigt wurde sind r -dimensionale Teilräume (Gerade: $r = 1$, Ebene: $r = 2$, etc), $r \leq n$, der Menge aller n -dimensionalen Vektoren Mengen von Vektoren, die sich als Linearkombinationen von jeweils r linear unabhängigen n -dimensionalen Vektoren ergeben. Diese r Vektoren liegen nicht eindeutig fest: bei einer Geraden kann man irgendeinen Vektor, der auf der Geraden liegt, wählen, um alle anderen Vektoren auf der Geraden zu erzeugen, bei einer Ebene kann man irgendzwei nicht parallele Vektoren aus der Ebene wählen, um alle Vektoren als Linearkombination dieser beiden Vektoren darzustellen, etc.

Hat man also n m -dimensionale Datenvektoren $\mathbf{x}_1, \dots, \mathbf{x}_n$, $n \leq m$, so bedeutet die Aussage, dass sie als Linearkombinationen von $r \leq n$ linear unabhängigen Vektoren dargestellt werden können, dass sie in einem r -dimensionalen Teilraum liegen. Sind die \mathbf{x}_k alle parallel, so das etwa $\mathbf{x}_j = a_j \mathbf{x}_1$ für $j = 2, \dots, n$, so liegen sie alle auf einer Geraden, d.h. in einem 1-dimensionalen Teilraum, etc. Die Frage ist, wie die Anzahl r der linear unabhängigen Vektoren $\mathbf{u}_1, \dots, \mathbf{u}_r$ und die \mathbf{u}_k , $1 \leq k \leq r$ bestimmt werden können. Bevor auf diese Frage in Abschnitt 1.3.2 eingegangen wird, sollen noch einige allgemeine Sachverhalte geklärt werden.

Auf Seite 17 wurde der Begriff der Norm eines Vektors eingeführt. Allgemein ist eine Norm¹⁶ eine Abbildung eines Vektorraums in die Menge der nicht-negativen reellen Zahlen \mathbb{R}_+ ; sie wird mit $\|\cdot\|$ bezeichnet, wobei der Punkt (\cdot) Platzhalter für einen Vektornamen ist:

$$\|\cdot\| : V \rightarrow \mathbb{R}_+, \quad \mathbf{x} \mapsto \|\mathbf{x}\| \in \mathbb{R} \text{ für } \mathbf{x} \in V. \quad (1.73)$$

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

1. *Definitheit*: $\|\mathbf{x}\| = 0 \rightarrow \mathbf{x} = \vec{0}$,
2. *absolute Homogenität*: $\|a\mathbf{x}\| = |a|\|\mathbf{x}\|$, $a \in \mathbb{R}$,
3. *Subadditivität*: $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ (Dreiecksungleichung).

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

¹⁶Von lat. *norma* = Richtschnur

leicht herleiten. Es ist

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

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

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

d.h. aber

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|, \quad (1.74)$$

also die Dreiecksungleichung. \square

Definition 1.8 *Es sei V ein Vektorraum, dessen Elemente n -dimensionale Vektoren seien, und es seien $\mathbf{x}, \mathbf{y} \in V$. Dann heißt die Norm der Differenz*

$$\|\mathbf{x} - \mathbf{y}\| = \left(\sum_{i=1}^n (x_i - y_i)^2 \right)^{1/2} \quad (1.75)$$

die euklidische Distanz zwischen \mathbf{x} und \mathbf{y} , und V heißt auch euklidischer Vektorraum (vergl. (5.2), Seite 157).

$\|\mathbf{x} - \mathbf{y}\|$ ist ein Vektor, dessen Länge der Distanz $d_e(x, y)$ zwischen den Endpunkten x und y der Vektoren \mathbf{x} und \mathbf{y} entspricht; $d_e(x, y)$ ist euklidisch, weil der Vektor $\mathbf{x} - \mathbf{y}$ die Distanz zwischen den Endpunkten von \mathbf{x} und \mathbf{y} durch eine Gerade repräsentiert¹⁷, und definiert die euklidische Metrix. Der Index e in der Bezeichnung d_e soll anzeigen, dass die euklidische Distanz gemeint ist, im Unterschied zur Distanz $d_p(x, y)$, der Minkowski-Distanz (5.3) (Seite 158). Es sei noch einmal darauf hingewiesen, dass eine Distanz über ein Skalarprodukt definiert ist: es sei

$$d_i = |x_i - y_i|, \quad i = 1, \dots, n \quad (1.76)$$

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

$$\|\mathbf{d}\| = (\mathbf{d}'\mathbf{d})^{1/2} = \|\mathbf{x} - \mathbf{y}\| = d_e(x, y). \quad (1.77)$$

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

$$\mathbf{d}'_p \mathbf{d}_p = \sum_{i=1}^n d_i^{p/2} d_i^{p/2} = \sum_{i=1}^n d_i^p = \sum_{i=1}^n |x_i - y_i|^p, \quad p > 1 \quad (1.78)$$

¹⁷Nach Euklid ist die Gerade die kürzeste Verbindung zwischen zwei Punkten.

und

$$d_p(x, y) = (\mathbf{d}'_p \mathbf{d}_p)^{1/p} = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{1/p}, \quad p > 1 \quad (1.79)$$

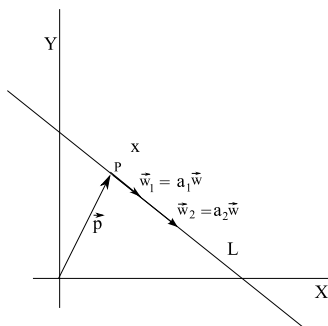
ist die Minkowski-Distanz zwischen x und y , die die Minkowski-Metrik definiert. Minkowski-Distanzen spielen in diesem Skript keine weitere Rolle, weshalb hier nicht weiter auf sie eingegangen wird.

Beispiel 1.7 Geraden im \mathbb{R}^n : Es ist intuitiv klar, dass für zwei verschiedene Punkte in einem 3-dimensionalen Raum genau eine Gerade existiert, auf die beiden Punkte liegen. Dieser Sachverhalt überträgt sich auf den allgemeinen \mathbb{R}^n . Es seien $\mathbf{p}, \mathbf{v} \in \mathbb{R}^n$; die Endpunkte dieser Vektoren seien zwei Punkte, durch die eine Gerade L läuft. Der durch die Differenz $\mathbf{w} = \mathbf{p} - \mathbf{v}$ definierte Vektor liegt dann auf (bzw. in) L , seine Orientierung ist identisch mit der von L . Für die Gerade ergibt sich dann die *Parameterdarstellung*

$$L: \mathbf{x} = \mathbf{u} + s\mathbf{v}, \quad s \in \mathbb{R}, \quad \mathbf{x}, \mathbf{u}, \mathbf{v} \in \mathbb{R}^2 \quad (1.80)$$

der Geraden. Hier ist \mathbf{u} der *Ortsvektor* und \mathbf{v} der *Richtungsvektor* der Geraden. ist s. Abbildung 10. Es sei angemerkt, dass die Gerade L und der Vektor \mathbf{u} *keinen* rechten Winkel bilden müssen. Es sei $\mathbf{n} = (n_1, n_2)'$ ein Vektor, der senkrecht

Abbildung 10: Gerade in Parameterform



auf der Geraden steht; dies ist der *Normalenvektor* der Geraden. Es gilt $\mathbf{n}'\mathbf{v} = 0$, da ja der Richtungsvektor \mathbf{v} und \mathbf{n} orthogonal sein müssen. Bildet man das Skalarprodukt von \mathbf{n} mit \mathbf{x} , so erhält man aus (1.80)

$$\mathbf{n}'\mathbf{x} = \mathbf{n}'\mathbf{u} + s\mathbf{n}'\mathbf{v} = \mathbf{n}'\mathbf{u}. \quad (1.81)$$

Schreibt man $\mathbf{n}'\mathbf{x}$ aus, so resultiert

$$n_1x_1 + n_2x_2 - d = 0, \quad d = \mathbf{n}'\mathbf{u}. \quad (1.82)$$

dies ist die zu (1.80) korrespondierende *Koordinatenform* der Geradengleichung. Die Koeffizienten von x_1 und x_2 sind gerade die Komponenten n_1, n_2 des Normalenvektors.

Es werde die Menge der Vektoren \mathbf{x} betrachtet, die durch (1.80) definiert werden. Diese Menge bildet einen Vektorraum. Denn es seien $\mathbf{x}_1 = \mathbf{u}_1 + s\mathbf{v}_1$ und $\mathbf{x}_2 = \mathbf{u}_2 + s\mathbf{v}_2$ irgend zwei Vektoren. Dann ist

$$\mathbf{x} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + s(a_1\mathbf{v}_1 + a_2\mathbf{v}_2),$$

und mit $\mathbf{x} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2$, $\mathbf{u} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2$ und $\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2$ ist \mathbf{x} wieder von der Form (1.80).

Es sei $\mathbf{u} = \vec{0}$. Die Menge der Vektoren $\mathbf{x}_s = s\mathbf{v}$ bildet einen Teilraum, denn

$$\mathbf{x} = a_1\mathbf{x}_s + a_2\mathbf{x}_s = s(a_1\mathbf{v} + a_2\mathbf{v}) = s(a_1 + a_2)\mathbf{v}$$

liegt auf derselben Geraden durch den Ursprung. □

Beispiel 1.8 Ebenen Es werde eine Ebene im \mathbb{R}^3 betrachtet; in *Parameterform* kann sie durch

$$\mathcal{E} = \{\mathbf{x} \mid \mathbf{x} = \mathbf{u} + s\mathbf{v}_1 + t\mathbf{v}_2, \quad \mathbf{x}, \mathbf{u}, \mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^3, \quad s, t \in \mathbb{R}\} \quad (1.83)$$

definiert werden: hier werden \mathbf{v}_1 und \mathbf{v}_2 als linear unabhängig vorausgesetzt. \mathcal{E} ist die Menge aller Punkte (Vektoren), die sich als Linearkombinationen von \mathbf{v}_1 und \mathbf{v}_2 darstellen lassen. Dabei ist \mathbf{u} der *Stütz- oder Ortsvektor* und $\mathbf{v}_1, \mathbf{v}_2$ sind nicht-parallele *Richtungsvektoren*. $\mathbf{v}_1, \mathbf{v}_2$ und \mathbf{x} liegen in der Ebene. Durch Variation der Parameter s und t kann jeder Punkt auf E dargestellt werden. Für $\mathbf{u} = \vec{0}$ geht die Ebene durch den Nullpunkt des Koordinatensystems.

Es sei $\mathbf{n} = (n_1, n_2, n_3)'$ wieder der Normalenvektor, d.h. \mathbf{n} stehe senkrecht auf der Ebene. \mathbf{n} definiert die Orientierung der Ebene. Es gelten die Beziehungen $\mathbf{n}'\mathbf{v}_1 = \mathbf{n}'\mathbf{v}_2 = 0$, – das Skalarprodukt zweier orthogonaler Vektoren ist gleich Null. Daraus folgt sofort, dass jede Linearkombination $\mathbf{v} = a_1\mathbf{u}_1 + a_2\mathbf{v}_2$ in der Ebene liegen muss, denn

$$a_1\mathbf{n}'\mathbf{v}_1 + a_2\mathbf{n}'\mathbf{v}_2 = \mathbf{n}'\mathbf{v} = 0, \quad (1.84)$$

d.h. \mathbf{n} steht auch senkrecht auf \mathbf{v} , was nur möglich ist, wenn \mathbf{v} in der Ebene liegt.

Nun folgt aus (1.83)

$$\mathbf{x} - \mathbf{u} = s\mathbf{v}_1 + t\mathbf{v}_2,$$

und multipliziert man von links mit \mathbf{n}' , so erhält man

$$\mathbf{n}'(\mathbf{x} - \mathbf{u}) = 0, \quad (1.85)$$

wegen $s\mathbf{n}'\mathbf{v}_1 + t\mathbf{n}'\mathbf{v}_2 = 0$. Nun ist \mathbf{u} ein konstanter Vektor, während \mathbf{x} variabel ist, d.h. \mathbf{x} ist ein Element der Menge aller Vektoren, deren Endpunkte in der Ebene liegen. Man kann also $\mathbf{n}'\mathbf{u} = d$ eine Konstante setzen, so dass aus (1.85)

$$n_1x_1 + n_2x_2 + n_3x_3 - d = 0 \quad (1.86)$$

folgt. Dies ist die Gleichung der Ebene in *Koordinatenform*. Die Komponenten des Normalenvektors \mathbf{n} sind also die Koeffizienten der Koordinatenform. Ist also die Gleichung der Ebene in Koordinatenform gegeben, so lassen sich die Komponenten des Normalenvektors direkt von den Koeffizienten der Komponenten x_i von \mathbf{x} ablesen.

Es sei $\mathbf{u} = \vec{0}$. Dann ist $E_0 : \mathbf{x} = s\mathbf{v}_1 + t\mathbf{v}_2$ für $s, t \in \mathbb{R}$ ein Teilraum des \mathbb{R}^3 . Denn mit $\mathbf{x} \in E_0$ ist auch $c\mathbf{x} \in E_0$:

$$c\mathbf{x} = cs\mathbf{v}_1 + ct\mathbf{v}_2, \quad \mathbf{n}'(c\mathbf{x}) = cs\mathbf{n}'\mathbf{v}_1 + ct\mathbf{n}'\mathbf{v}_2 = 0,$$

d.h. $c\mathbf{x}$ liegt in der von \mathbf{v}_1 und \mathbf{v}_2 aufgespannten Ebene. Sind \mathbf{x}_1 und \mathbf{x}_2 Elemente von E_0 , so liegt auch jede Linearkombination dieser beiden Vektoren in der Ebene:

$$\begin{aligned} \mathbf{x}_1 &= s_1\mathbf{v}_1 + t_1\mathbf{v}_2 \\ \mathbf{x}_2 &= s_2\mathbf{v}_1 + t_2\mathbf{v}_2 \\ \mathbf{x} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 &= (c_1s_1 + c_2s_2)\mathbf{n}'\mathbf{v}_1 + (c_1t_1 + c_2t_2)\mathbf{n}'\mathbf{v}_2 \end{aligned} \quad (1.87)$$

und

$$\mathbf{n}'\mathbf{x} = (c_1s_1 + c_2s_2)\mathbf{n}'\mathbf{v}_1 + (c_1t_1 + c_2t_2)\mathbf{n}'\mathbf{v}_2 = 0,$$

d.h. \mathbf{n} steht senkrecht auf \mathbf{x} , d.h. $\mathbf{x} \in E_0$.

Anmerkung: Die Einschränkung auf den Fall $n = 3$ ist für den Beweis nicht notwendig; man sieht leicht, dass er beliebiges $n \geq 2$ gilt: die Anzahl der Komponenten der Vektoren geht in die Argumentation nicht ein. \square

1.3.2 Erzeugendensysteme und Basen von Vektorräumen

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

1. *Es sei*

$$\mathcal{L}(M) = \{\mathbf{x} | \mathbf{x} = a_1\mathbf{x}_1 + \dots + a_p\mathbf{x}_p\}, \quad \mathcal{L}(\emptyset) = \vec{0}, \quad a_1, \dots, a_p \in \mathbb{R} \quad (1.88)$$

d.h. $\mathcal{L}(M)$ sei die Menge aller Linearkombinationen der Vektoren aus M .¹⁸ \emptyset bezeichnet die leere Menge. $\mathcal{L}(M)$ heißt die lineare Hülle von M .

2. *Es sei $V = \mathcal{L}(M)$. Dann heißt die Menge M Erzeugendensystem für V (die Vektoren von V werden als Linearkombinationen der Elemente von M erzeugt).*

3. *Ein Erzeugendensystem M heißt minimal, wenn kein Vektor $\mathbf{x}_j \in M$ existiert derart, dass¹⁹ $M \setminus \{\mathbf{x}_j\}$ kein Erzeugendensystem für V mehr ist.*

Anmerkung: Die lineare Hülle $\mathcal{L}(M)$ ist die Menge *aller* Linearkombinationen der $\mathbf{x}_1, \dots, \mathbf{x}_p$, das heißt die Koeffizienten a_k , $1 \leq k \leq p$ durchlaufen jeweils alle

¹⁸Statt des Ausdrucks 'lineare Hülle' ist auch der aus dem Englischen übernommene Ausdruck 'span' gebräuchlich, der zum deutschen Ausdruck 'aufspannen' korrespondiert.

¹⁹Sei $\mathbf{x}_j \in M$. $M \setminus \{\mathbf{x}_j\}$ ist die Menge M ohne den Vektor \mathbf{x}_j .

Werte zwischen $-\infty$ und $+\infty$, $-\infty < a_k < \infty$. Die Spezifikation $a_k \in \mathbb{R}$ bedeutet, dass jeder Koeffizient Werte aus der überabzählbaren Menge \mathbb{R} der reellen Zahlen annehmen kann, also Werte aus einer Menge, die mehr als unendlich viele Elemente enthält, wobei mit 'unendlich' die Mächtigkeit $|\mathbb{N}|$ der Menge \mathbb{N} gemeint ist. Man kann dann noch die Menge \mathbb{Q} der rationalen Zahlen p/q mit $p, q \in \mathbb{N}$ betrachten, und Cantor²⁰ hat gezeigt, dass $|\mathbb{Q}| = |\mathbb{N}|$ ist. Aber dann gibt es noch die Menge der *irrationalen*, d.h. der nicht durch einen Quotienten ("ratio") p/q darstellbaren Zahlen, wie etwa $\sqrt{2}$, π , e etc. Cantor hat gezeigt, dass die Menge aller Zahlen – natürliche plus rationale plus irrationale Zahlen – die Mächtigkeit $2^{|\mathbb{N}|}$ hat, diese Menge ist also *überabzählbar*. Die lineare Hülle ist also – wie der Vektorraum – eine "dicht gepackte", überabzählbare Menge von Vektoren. \square

Satz 1.8 *Es sei V ein Vektorraum und $M = \{\mathbf{x}_1, \dots, \mathbf{x}_p\} \subset V$. Dann gilt $\mathcal{L}(M) \subseteq V$. Für $\mathcal{L}(M) \subset V$ ist $\mathcal{L}(M)$ ein Teilvektorraum von V .*

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

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

Dann folgt, mit $\lambda, \mu \in \mathbb{R}$,

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

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

Definition 1.10 *Es sei $U \subset \mathbb{R}^n$ ein Teilraum des \mathbb{R}^n und es sei $\mathcal{B} = \{\mathbf{u}_1, \dots, \mathbf{u}_r\} \subset U$. \mathcal{B} heißt Basis von U , wenn gilt*

- (i) *die Vektoren $\mathbf{u}_1, \dots, \mathbf{u}_r$ sind linear unabhängig,*
- (ii) *$U = \mathcal{L}(\mathcal{B})$, dh. U ist die lineare Hülle der $\mathbf{u}_1, \dots, \mathbf{u}_r$, d.h. U wird von \mathcal{B} erzeugt.*

Anmerkungen:

1. Eine Basis ist ein minimales Erzeugendensystem.
2. \mathcal{B} erzeugt U im Sinne von Punkt 2. der Definition 1.9.
3. Ist \mathcal{B} eine Basis für $U \subset \mathbb{R}^n$ und ist $\mathbf{x} \in U$, so existieren Koeffizienten a_1, \dots, a_r derart, dass

$$\mathbf{x} = a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + \dots + a_r \mathbf{u}_r; \tag{1.89}$$

²⁰Georg Cantor (1845 – 1918), Mathematiker

die Koeffizienten a_j heißen die *Komponenten von \mathbf{x} in Bezug auf die Basis \mathcal{B}* .
 4. Wie bereits gezeigt wurde (Gleichung 1.18, Seite 13) kann jeder Vektor $\mathbf{x} \in \mathbb{R}^n$ als Linearkombination

$$\mathbf{x} = x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2 + \cdots + x_n \mathbf{e}_n \quad (1.90)$$

dargestellt werden. Es sei $\mathcal{B}_e = \{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$; nach dem eben Gesagten gilt $\mathcal{L}(\mathcal{B}_e) = \mathbb{R}^n$, d.h. \mathcal{B}_e ist eine Basis des \mathbb{R}^n . \mathcal{B}_e heißt auch *kanonische Basis* des \mathbb{R}^n ; sie spielt u.a. eine wichtige Rolle für den Beweis des Satzes 1.12 (s. unten). □

Die bisherigen Betrachtungen haben nur Eigenschaften einer Basis beschrieben, was aber noch nicht bedeutet, dass für einen beliebigen (Teil-)Vektorraum auch eine Basis existiert. Es läßt sich nun der folgende Satz beweisen:

Satz 1.9 *Es sei $V \neq \emptyset$ ein (Teil-)Vektorraum.*

(i) *Dann existiert stets eine Basis U_V für V .*

(ii) *Je zwei verschiedene Basen \mathcal{B}_1 und \mathcal{B}_2 von V haben stets gleich viele Elemente.*

Beweis: Es sei $A = \{\mathbf{a}_1, \dots, \mathbf{a}_s\} \subseteq V$ ein Erzeugendensystem von V .

Zu (i): Es sei $A_j = A \setminus \mathbf{a}_j$ die Menge, die aus A entsteht, wenn der Vektor \mathbf{a}_j aus A entfernt ist. Ist A_j kein Erzeugendensystem mehr, so ist A eine Basis von V . Ist A_j immer noch ein Erzeugendensystem von V , so kann ein weiterer Vektor \mathbf{a}_k aus A_j entfernt werden. Ist $(A_j \setminus \mathbf{a}_k)$ kein Erzeugendensystem mehr, so ist A_j eine Basis von V , etc. Das Verfahren stoppt für ein $1 \leq r \leq s$, so dass $U_V = \{\mathbf{a}_1, \dots, \mathbf{a}_r\} \subseteq V$ eine Basis von V ist.

Zu (ii): Es sei \mathcal{B}_1 eine Basis von V mit n Elementen; dann ist \mathcal{B}_1 ein Erzeugendensystem von V und jeweils $n + 1$ Elemente von V sind linear abhängig. Deshalb kann \mathcal{B}_2 nicht mehr als n Elemente enthalten. Argumentiert man in der umgekehrten Reihenfolge, so gilt dieselbe Schlußfolgerung: enthält \mathcal{B}_2 n Elemente, so ist jede Menge von $n + 1$ Elementen linear abhängig, also kann \mathcal{B}_1 nicht mehr als n Elemente enthalten.

Man macht sich leicht klar, dass, wenn \mathcal{B}_1 eine Basis von V mit n Elementen ist, \mathcal{B}_2 nicht weniger als n Elemente enthalten kann. Denn enthielte \mathcal{B}_2 nur $n - 1$ Elemente, so wäre bereits jede Menge mit n Elementen linear abhängig und \mathcal{B}_1 wäre keine Basis, sondern allenfalls ein Erzeugendensystem. □

Die Koeffizienten a_j in (1.89) sind gemäß dem folgenden allgemeinen Satz eindeutig bestimmt:

Satz 1.10 *\mathbf{y} sei eine Linearkombination der Vektoren $\mathbf{x}_1, \dots, \mathbf{x}_p$, $\mathbf{y}, \mathbf{x}_j \in \mathbb{R}^n$. Sind die Vektoren $\mathbf{x}_1, \dots, \mathbf{x}_p$ linear unabhängig, so sind die a_1, \dots, a_p eindeutig bestimmt.*

Beweis: Es werde angenommen, dass es zwei verschiedene Darstellungen von \mathbf{y}

als Linearkombination der $\mathbf{x}_1, \dots, \mathbf{x}_p$ gibt:

$$\mathbf{y} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \dots + a_p\mathbf{x}_p \quad (1.91)$$

$$\mathbf{y} = b_1\mathbf{x}_1 + b_2\mathbf{x}_2 + \dots + b_p\mathbf{x}_p \quad (1.92)$$

Subtrahiert man die Gleichung (1.92) von (1.91), so erhält man die Gleichung

$$\vec{0} = (a_1 - b_1)\mathbf{x}_1 + (a_2 - b_2)\mathbf{x}_2 + \dots + (a_p - b_p)\mathbf{x}_p \quad (1.93)$$

Nach Voraussetzung sind die $\mathbf{x}_1, \dots, \mathbf{x}_p$ linear unabhängig, so dass die Darstellung (1.93) des Nullvektors $\vec{0}$ nur möglich ist, wenn

$$a_1 - b_1 = a_2 - b_2 = \dots = a_p - b_p = 0$$

gilt, d.h. die Koeffizienten in den Gleichungen (1.91) und (1.92) sind identisch, d.h. die Darstellung von \mathbf{y} als Linearkombination der $\mathbf{x}_1, \dots, \mathbf{x}_p$ ist eindeutig. \square

Wie in Punkt 3. oben erwähnt wurde heißen die Koeffizienten a_j in (1.89) "Komponenten von \mathbf{x} in Bezug auf die Basis $\mathbf{u}_1, \dots, \mathbf{u}_r$ ". Es zeigt sich (s. unten), dass es für $U \subseteq \mathbb{R}^n$ nicht nur eine Basis \mathcal{B} gibt, und nach Satz 1.10 sind die Koeffizienten a_1, \dots, a_r für jeweils eine Basis \mathcal{B} spezifisch, da eindeutig bestimmt. Dieser Sachverhalt erklärt die Kennzeichnung der Koeffizienten als 'Komponenten von \mathbf{x} in Bezug auf \mathcal{B} '.

In der Definition 1.10 wird eine Basis als eine Menge \mathcal{B} von linear unabhängigen Vektoren erklärt. Wie gerade angedeutet gibt es für einen Vektorraum oder einen Teilraum eines Vektorraums stets beliebig viele Mengen von jeweils r linear unabhängigen Vektoren, die als Basis für einen Teilraum $U \subseteq \mathbb{R}^n$ dienen können, allen gemein ist die Anzahl r dieser Vektoren. Man veranschaulicht sich diesen Sachverhalt leicht am Beispiel einer Ebene in einem 3-dimensionalen Raum. Sind \mathbf{u}_1 und \mathbf{u}_2 irgendzwei 3-dimensionale, linear unabhängige (also nicht parallele) Vektoren aus der Ebene, so läßt sich jeder Vektor \mathbf{x} in der Ebene als Linearkombination von \mathbf{u}_1 und \mathbf{u}_2 darstellen (vergl. (??), Seite ?? für einen Nachweis dieser Behauptung). Nun seien \mathbf{x}_1 und \mathbf{x}_2 Linearkombinationen von \mathbf{u}_1 und \mathbf{u}_2 , und diese Vektoren seien nicht parallel. Dann sind sie linear unabhängig. Dann kann man auch \mathbf{x}_1 und \mathbf{x}_2 als Basis für die Vektoren der Ebene wählen und damit auch die ursprüngliche Basis $\mathbf{u}_1, \mathbf{u}_2$ als Linearkombination der $\mathbf{x}_1, \mathbf{x}_2$ darstellen. Offenbar kann man von einer Basis zu einer anderen übergehen.

Mögliche Werte von r : Die Frage ist nun, welche Werte r annehmen kann. Verschiedene Basen eines Vektorraums können offenbar nicht verschiedene Anzahlen r der jeweiligen Basisvektoren haben, da eine Basis ja ein minimales Erzeugendensystem ist, und es kann nur *eine* kleinste Anzahl von Basisvektoren geben. Dementsprechend kann man vermuten, dass aus der Darstellung (1.90) eines beliebigen Vektors $\mathbf{x} \in \mathbb{R}^n$ durch die kanonische Basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ nahelegt, dass der Wert von r zumindest nicht größer, vermutlich auch nicht kleiner als n sein kann. Gegeben sei insbesondere ein 3-dimensionaler Raum mit einer Ebene E als

Teilraum. Die Vektoren in E lassen sich alle als Linearkombinationen von irgendzwei nicht-parallelen, also linear unabhängigen Vektoren $\mathbf{u}_1, \mathbf{u}_2 \in E$ darstellen. Nun werde angenommen, es gäbe noch einen dritten Vektor \mathbf{u}_3 , der zusammen mit den Vektoren $\mathbf{u}_1, \mathbf{u}_2$ zur Darstellung von Vektoren $\mathbf{x} \in E$ herangezogen werden könnte oder sogar müßte, so dass

$$\mathbf{x} = a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + a_r \mathbf{u}_3 \in E$$

Aber eine derartige Darstellung ist sicher nicht möglich, da (i) *alle* Vektoren $\mathbf{x} \in E$ als Linearkombinationen der \mathbf{u}_1 und \mathbf{u}_2 darstellbar sind, und für $\mathbf{u}_3 \in E$ eben auch \mathbf{u}_3 , d.h. die $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ können nichtlinear unabhängig sein. Also muß $\mathbf{u}_3 \notin E$ gelten, damit die $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ linear unabhängig sein können. Für $a_3 \neq 0$ kann dann aber \mathbf{x} nicht mehr in E liegen, denn \mathbf{u}_3 erzwingt nun eine aus E hinausweisende Orientierung von \mathbf{x} . Für

$$\mathbf{x} = a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + a_r \mathbf{u}_3 \in \mathbb{R}^3$$

und $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ linear unabhängig kann aber *jeder* Vektor $\mathbf{x} \in \mathbb{R}^3$ eindeutig dargestellt werden (Satz 1.10), so dass es keinen Vektor \mathbf{u}_4 geben derart, dass $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4$ linear unabhängig sind: da ja die $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ schon *alle* $\mathbf{x} \in \mathbb{R}^3$ erzeugen, erzeugen sie notwendigerweise auch $\mathbf{u}_4 \in \mathbb{R}^3$, so dass man davon ausgehen kann, dass für den \mathbb{R}^3 jeweils $r = 3$ linear unabhängige Vektoren als Basis genügen. \square

Es wird jetzt explizit gezeigt werden, dass tatsächlich $r = n$ gelten muß, wenn *alle* Vektoren aus \mathbb{R}^n darstellbar sein sollen. Die Herleitung dieser Aussage folgt Stewart (1973). Zuerst wird ein Hilfssatz (*Steinitz'sches²¹ Lemma*, auch *Austauschsatz von Steinitz*) bewiesen:

Satz 1.11 Steinitz'sches Lemma: *Es sei $\mathcal{B}_u = \{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ eine Basis des Teilraums $U \subset \mathbb{R}^n$. Weiter sei $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_m\} \subset U$ eine Menge linear unabhängiger Vektoren. Dann existiert ein $k \in \mathbb{N}$ derart, dass die Vektoren*

$$\mathbf{u}_k, \mathbf{b}_2, \dots, \mathbf{b}_m$$

linear unabhängig sind.

Beweis: Es werde angenommen, dass die $\mathbf{u}_k, \mathbf{b}_2, \dots, \mathbf{b}_m$ linear abhängig sind. Dann folgt, dass

$$c_{k1} \mathbf{u}_k + c_2 \mathbf{b}_2 + \dots + c_m \mathbf{b}_m = \vec{0} \quad (1.94)$$

und nicht alle Koeffizienten c_{k1}, c_2, \dots, c_m sind gleich Null. Insbesondere muß $c_{k1} \neq 0$ sein, denn wäre $c_{k1} = 0$, so müßte $c_2 \mathbf{b}_2 + \dots + c_m \mathbf{b}_m = \vec{0}$ gelten und wegen der postulierten linearen Abhängigkeit der $\mathbf{u}_k, \mathbf{b}_2, \dots, \mathbf{b}_m$ dürfen nicht alle

²¹Ernst Steinitz (1871 – 1928), Mathematiker

Koeffizienten c_j gleich Null sein, entgegen der Voraussetzung der linearen Unabhängigkeit der \mathbf{b}_j . Dann aber impliziert (1.94), dass \mathbf{u}_k als Linearkombination der \mathbf{b}_j dargestellt werden kann:

$$\mathbf{u}_k = -\frac{1}{c_{k1}}(c_2\mathbf{b}_2 + \cdots + c_m\mathbf{b}_m) = d_{k2}\mathbf{b}_2 + \cdots + d_{km}\mathbf{b}_m, \quad d_{kj} = -\frac{c_j}{c_{k1}}. \quad (1.95)$$

Nun ist aber \mathcal{B}_u eine Basis für U und nach Voraussetzung ist $\mathbf{b}_1 \in U$, so dass Koeffizienten g_1, \dots, g_r existieren müssen derart, dass

$$\mathbf{b}_1 = \sum_{k=1}^m g_k \mathbf{u}_k = \sum_{k=1}^m g_k \sum_{j=2}^r d_{kj} \mathbf{b}_j = \sum_{j=2}^r \sum_{k=1}^m g_k d_{kj} \mathbf{b}_j. \quad (1.96)$$

\mathbf{b}_1 wird demnach als Linearkombination der $\mathbf{b}_2, \dots, \mathbf{b}_m$ dargestellt, im Widerspruch zur vorausgesetzten linearen Unabhängigkeit der $\mathbf{b}_1, \dots, \mathbf{b}_m$. Damit führt die obige Annahme, dass die $\mathbf{u}_k, \mathbf{b}_2, \dots, \mathbf{b}_m$ linear abhängig sind, zu einem Widerspruch, d.h. diese Annahme ist falsch, die $\mathbf{u}_k, \mathbf{b}_2, \dots, \mathbf{b}_m$ müssen linear unabhängig sein. \square

Satz 1.12 *Es sei $U = \mathbb{R}^n$. Für jede Basis $\mathcal{B}_u = \{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ von U gilt dann $r = n$.*

Beweis:²² (i) Es sei $U = \mathbb{R}^n$ und $\mathcal{B} = \{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ sei eine Basis. Zu zeigen ist, dass $r > n$ nicht möglich ist. Dazu werde angenommen, dass eine Basis \mathcal{B} mit $r > n$ Elementen existiert, z.B. $r = n + 1$, d.h. die $\mathbf{u}_1, \dots, \mathbf{u}_{n+1}$ seien linear unabhängig. Andererseits ist $\mathcal{B}_e = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ eine Basis des \mathbb{R}^n (s. Punkt 4., Seite 40). Nach Satz 1.11 kann nun eine Zahl $m_1 \in \mathbb{N}$ gefunden werden derart, dass \mathbf{u}_1 mit \mathbf{e}_{m_1} ausgetauscht werden kann und $\{\mathbf{e}_{m_1}, \mathbf{u}_2, \dots, \mathbf{u}_{n+1}\}$ ist linear unabhängig. Eine zweite Anwendung von Satz 1.11 liefert eine Zahl $m_2 \in \mathbb{N}$, $m_2 \neq m_1$, derart, dass \mathbf{u}_2 durch \mathbf{e}_{m_2} ersetzt werden kann, und

$$\{\mathbf{e}_{m_1}, \mathbf{e}_{m_2}, \mathbf{u}_3, \dots, \mathbf{u}_{n+1}\}$$

ist nach Voraussetzung linear unabhängig. So kann man fortfahren, bis die ersten n Vektoren $\mathbf{u}_1, \dots, \mathbf{u}_n$ durch die Einheitsvektoren $\mathbf{e}_{m_1}, \mathbf{e}_{m_2}, \dots, \mathbf{e}_{m_n}$ ersetzt worden sind. Aber $\{\mathbf{e}_{m_1}, \mathbf{e}_{m_2}, \dots, \mathbf{e}_{m_n}\}$ ist die kanonische Basis des \mathbb{R}^n (auf die Reihenfolge der \mathbf{e}_j kommt es ja nicht an), – jeder Vektor $\mathbf{x} \in \mathbb{R}^n$ kann als Linearkombination der \mathbf{e}_j dargestellt werden. Das heißt aber, dass \mathbf{u}_{n+1} als Linearkombination der \mathbf{e}_j repräsentiert werden kann, und die Menge $\{\mathbf{e}_{m_1}, \dots, \mathbf{e}_{m_n}, \mathbf{u}_{n+1}\}$ ist demnach linear abhängig, im Widerspruch zur obigen Annahme, d.h. die Anzahl der Basisvektoren des \mathbb{R}^n kann nicht größer als n sein.

²²Dieser Satz ist eine Konsequenz von Satz 1.11, also dem *Austauschsatz von Steinitz*; die Argumentation des folgenden Beweises, (i), ist der Kern des Beweises dieses Austauschsatzes; der vollständige Beweis des Austauschsatzes wird oft durch Anwendung des Prinzips der Vollständigen Induktion geführt.

(ii) Für $U = \mathbb{R}^n$ kann nicht $r < n$ gelten. Denn es sei $\mathcal{B} = \{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ eine Menge linear unabhängiger Vektoren mit $r < n$. Wiederholte Anwendung von Satz 1.11 führt auf $\{\mathbf{e}_{m_1}, \dots, \mathbf{e}_{m_r}\}$ mit $m_r < n$, und $\mathbf{x} \in U$ kann nicht vollständig als Linearkombination der $\mathbf{e}_{m_1}, \dots, \mathbf{e}_{m_r}$ dargestellt werden, weil $n - r$ Einheitsvektoren fehlen, so dass $r \geq n$ folgt. Nach (i) gilt also $r \leq n$, und nach (ii) $r \geq n$, d.h. es muß $r = n$ gelten. \square

Es existiert also keine Basis $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_r\}$ für \mathbb{R}^n mit $r > n$. Dies impliziert, dass jede Menge²³ von Vektoren $M = \{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ aus \mathbb{R}^n linear abhängig sein muß, wenn $p > n$. Der Einfachheit halber sei $p = n + 1$, und es werde angenommen, alle \mathbf{x}_j aus M wären linear unabhängig. Dann würden die ersten n Vektoren von M bereits eine Basis für \mathbb{R}^n bilden, weil jede Menge von n linear unabhängigen Vektoren aus \mathbb{R}^n eine Basis bildet. Dann kann aber \mathbf{x}_{n+1} als Linearkombination der ersten n Vektoren dargestellt werden, was bedeutet, dass M eine Menge von linear abhängigen Vektoren sein muß, was $r < p$ bedeutet.

Nun sei $p < n$ ²⁴. Es sei $\mathcal{L}_p = \mathcal{L}(\mathbf{x}_1, \dots, \mathbf{x}_p)$ die lineare Hülle der $\mathbf{x}_1, \dots, \mathbf{x}_p$. \mathcal{L}_p ist ein Teilraum des \mathbb{R}^n ; dann gilt ebenfalls $r \leq p$. Dies folgt sofort für den Fall, dass die \mathbf{x}_j linear unabhängig sind; dann bilden sie eine Basis für \mathcal{L}_p , so dass keine Basis $\{\mathbf{b}_1, \dots, \mathbf{b}_r\} \in \mathcal{L}$ mit $r > p$ existieren kann, denn es gibt ja nur die p Vektoren \mathbf{x}_k , und es können eben nur maximal p von ihnen linear unabhängig sein, d.h. der maximal mögliche Wert von r ist p , die \mathbf{x}_k bilden somit eine Basis für $\mathcal{L}(\mathbf{x}_1, \dots, \mathbf{x}_p)$. Sei $p > 1$. Unterscheiden sich die \mathbf{x}_j nur durch ihre Länge, so sind alle \mathbf{x}_j Linearkombinationen nur eines Vektors, d.h. Streckungen oder Stauchungen oder Identitäten, also $r = 1$ und \mathcal{L}_p ist eine Gerade im \mathbb{R}^n . Gilt $p > 2$ und liegen alle \mathbf{x}_j in einer Ebene, so genügen zwei linear unabhängige Vektoren, um alle Vektoren und damit auch die \mathbf{x}_j zu erzeugen, also $r = 2 < p$, \mathcal{L}_p ist eine Ebene, etc. (vergl. Beispiel 1.8, Seite 37).

Satz 1.13 *Es²⁵ sei V ein Vektorraum und M sei eine Teilmenge von V . Dann existiert ein Teilraum U von V mit $M \subseteq U$ und es gilt: Ist W ein Teilraum von V mit $M \subseteq W$, so folgt $U \subseteq W$ und es gilt*

$$U = \mathcal{L}(M), \tag{1.97}$$

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

Beweis: Es sei $U := \mathcal{L}(M)$. Nach Satz 1.8 ist U ein Teilraum von V . Sicherlich ist $M \subset \mathcal{L}(M)$, und $U = \{\vec{0}\}$ für $M = \emptyset$. Dann folgt für W ein Teilraum von V mit $M \subseteq W$, dass $U = \mathcal{L}(M) \subseteq W$.

²³Solche Mengen ergeben sich z.B. in Untersuchungen, in denen p Variablen bei n Fällen gemessen werden.

²⁴Beispiel: es werden $p < n$ Variablen bei n Fällen gemessen; dies ist wohl der häufigste Fall in empirischen Untersuchungen.

²⁵Lorenz (1988), Band I, p.33

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

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

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

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

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

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

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

Der folgende Satz erweist sich als nützlich und folgt leicht aus den vorangegangenen Resultaten (Fischer(1984), p. 86):

Satz 1.14 (Basisergänzungssatz) *Es sei V ein Vektorraum mit endlichem Erzeugendensystem, und es seien die linear unabhängigen Vektoren $\mathbf{w}_1, \dots, \mathbf{w}_r$ auf V gegeben. Dann existieren weitere linear unabhängige Vektoren $\mathbf{w}_{r+1}, \dots, \mathbf{w}_n$ derart, dass*

$$\mathcal{B} = \{\mathbf{w}_1, \dots, \mathbf{w}_r, \mathbf{w}_{r+1}, \dots, \mathbf{w}_n\}$$

eine Basis von V ist.

Beweis: Es sei $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ ein Erzeugendensystem. Es enthält eine Basis $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$, $r \leq n$. Nach dem Steinitzchen Lemma (1.11), 42, kann man dann

$$\mathbf{w}_{r+1} := \mathbf{u}_{r+1}, \dots, \mathbf{w}_n := \mathbf{u}_n$$

wählen; dies ist die gesuchte Ergänzung. \square

Anmerkung: Es sei $V = \mathbb{R}^n$. Man kann den Satz 1.14 auch knapp so formulieren: es sei eine Teilbasis $\mathcal{B}_r = \{\mathbf{w}_1, \dots, \mathbf{w}_r\}$, $r \leq n$ gegeben. Dann existiert stets eine Basis \mathcal{B} , die \mathcal{B}_r enthält. \square

Anmerkung: Die vorangegangenen Betrachtungen suggerieren, dass für einen Vektorraum stets eine Basis existiert, – tatsächlich aber wurden nur Eigenschaften von Mengen linear unabhängiger Vektoren aufgezählt, die eine Basis hat, falls sie denn existiert. Tatsächlich läßt sich beweisen, dass für jeden Vektorraum eine Basis existiert. Der Beweis ist allerdings begrifflich aufwändiger, als man denken

könnte, d.h. er sprengt den Rahmen dieses Skriptums. Es sei auf das Lehrbuch von Fischer (1997) verwiesen.

Siehe aber <https://mathepedia.de/Existenz.html>, in der ein Beweis gegeben wird – man muß dann nur hinzufügen, dass auf das Zornsche Lemma eingegangen werden müsste, was hier übergangen werden soll.

Definition 1.12 ²⁶ Es seien A_1, \dots, A_k , die Endpunkte der linear unabhängigen Vektoren $\mathbf{a}_j \in \mathbb{R}^n$, $j = 1, \dots, k$ mit Anfangspunkt 0. Dann ist

$$E_k^n = \{\mathbf{y} | \mathbf{y} = \sum_{j=1}^k b_j \mathbf{a}_j; \quad b_j \in \mathbb{R}, j = 1, \dots, k\} \quad (1.98)$$

eine k -dimensionale Ebene im \mathbb{R}^n durch den Ursprung des Koordinatensystems. Für $k = n - 1$ heißt die Ebene Hyperebene.

Für $n = 3$ und $k = 2$ erhält man eine Ebene. Der Begriff der Hyperebene ist eine Verallgemeinerung des Begriffs der Ebene.

Da E_k^n durch k linear unabhängige Vektoren aus \mathbb{R}^n aufgespannt wird ist die Hyperebene ein k -dimensionaler Teilraum des \mathbb{R}^n .

Beispiel 1.9 Es sei V ein Vektorraum und $\mathbf{x}_1 \in V$; $\mathcal{B} = \{\mathbf{x}_1\}$ ist die Basis für einen eindimensionalen Teilraum $V_1 = \{\mathbf{x} | \mathbf{x} = a\mathbf{x}_1, a \in \mathbb{R}\}$ von V .

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

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

$$\mathbf{n}'\mathbf{x} = a_1\mathbf{n}'\mathbf{x}_1 + a_2\mathbf{n}'\mathbf{x}_2 = 0, \quad (1.99)$$

wegen $\mathbf{n}'\mathbf{x}_1 = \mathbf{n}'\mathbf{x}_2 = 0$. Dies gilt für alle $a_1, a_2 \in \mathbb{R}$, d.h. \mathbf{n} ist ein Normalenvektor für die gesamte Ebene. Da \mathbf{n} eine bestimmte Orientierung hat, ist mit ihm auch die Orientierung der Ebene definiert. Diese Betrachtung gilt für einen beliebigen n -dimensionalen Vektorraum mit $n > 2$, vgl. Beispiel 1.8, Seite 37. \square

²⁶Graybill (1983), p. 68

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

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

Definition 1.13 *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

$$\mathbf{b}'_j \mathbf{b}_k = \begin{cases} 0, & j \neq k \\ 1, & j = k \end{cases}, \quad j, k = 1, \dots, n \quad (1.100)$$

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

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

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

Beweis: Die Einheitsvektoren sind linear unabhängig, denn $\vec{0} = \lambda_1\mathbf{e}_1 + \lambda_2\mathbf{e}_2 + \dots + \lambda_n\mathbf{e}_n$ ist nur möglich für $\lambda_1 = \dots = \lambda_n = 0$; für die i -te Komponente hat man nämlich $0 = \lambda_i 1$, woraus sofort $\lambda_i = 0$ folgt. Darüber hinaus sind die \mathbf{e}_i orthonormal, vergl. (1.69). 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 $\mathbf{e}_1, \dots, \mathbf{e}_n$. □

Definition 1.14 *Die Basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ heißt kanonische Basis des V_n .*

Eine beliebige orthonormale Basis läßt sich als Rotation der kanonischen Basis herleiten, vergl. den Abschnitt 2.5 über Basiswechsel, Seite 78.

Beispiel 1.10 Es sei V ein n -dimensionaler Vektorraum und $\mathcal{B} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ sei eine Basis für V , so dass für jeden Vektor $\mathbf{x} \in V$ die Darstellung

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

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

Nach Satz 1.5, Seite 28, kann jeder Vektor $\mathbf{x} = (x_1, \dots, x_n)' \in V$ als Linearkombination der n -dimensionalen Einheitsvektoren $\mathbf{e}_1, \dots, \mathbf{e}_n$ dargestellt werden, und damit können auch die Vektoren der Basis \mathcal{B} als Linearkombinationen der $\mathbf{e}_1, \dots, \mathbf{e}_n$ dargestellt werden. Es sei x_{kj} die k -te Komponente von $\mathbf{x}_j \in \mathcal{B}$. Dann hat man

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

Die gesuchten Koordinaten sind also durch

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

gegeben. □

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

Satz 1.16 *Jeder endlich erzeugte Vektorraum V hat eine Basis.*

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

- (i) Ein einzelner Vektor, etwa $\mathbf{b}_1 \neq \vec{0}$, ist linear unabhängig, denn $\lambda_1 \mathbf{b}_1 = \vec{0}$ nur dann, wenn $\lambda_1 = 0$,
- (ii) $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ sei eine Menge von linear unabhängigen Vektoren. Diese Vektoren bilden eine Basis von V , da alle Vektoren aus V als Linearkombinationen der \mathbf{b}_j dargestellt werden können.
- (iv) Ist $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ 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. \square

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

$$\mathbf{x} = a_1 \mathbf{b}_1 + \dots + a_n \mathbf{b}_n = \sum_{k=1}^n a_k \mathbf{b}_k. \quad (1.101)$$

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

$$\mathbf{b}'_j \mathbf{x} = \sum_{k=1}^n a_k \mathbf{b}'_j \mathbf{b}_k = a_j, \quad j = 1, \dots, n \quad (1.102)$$

denn

$$\mathbf{b}'_j \mathbf{b}_k = \begin{cases} 0, & j \neq k \\ 1, & j = k \end{cases} \quad (1.103)$$

(1.101) kann dann in der Form

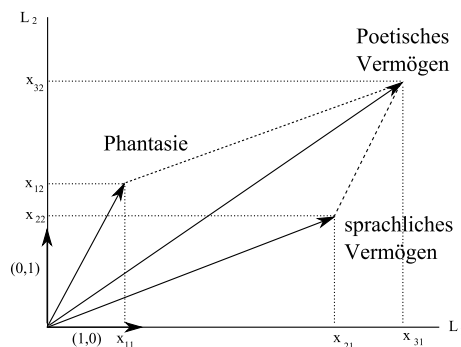
$$\mathbf{x} = \sum_{k=1}^n (\mathbf{b}'_k \mathbf{x}) \mathbf{b}_k. \quad (1.104)$$

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

Anmerkung: Bekanntlich kann unter bestimmten Normierungsbedingungen ein Skalarprodukt als Korrelation interpretiert werden. Dann bedeutet $\mathbf{b}'_k \mathbf{x}$ in Gleichung (1.104) die Korrelation zwischen dem Vektor \mathbf{x} und dem Basisvektor \mathbf{b}_k . In der Faktorenanalyse und in Approximationen der Faktorenanalyse wird die *Ladung* eines Items (d.h. die Koordinate des Items) auf einer latenten Dimension als Korrelation zwischen dem Item und der latenten Dimension interpretiert. Diese Interpretation beruht auf der orthonormalen Basisentwicklung (1.104). \square

In Abbildung 11 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

Abbildung 11: Poesie einerseits als Linearkombination von Vektoren, die Phantasie und Sprachvermögen repräsentieren, andererseits als Linearkombinationen der kanonischen Basis, die auch eine Darstellung der Merkmale Phantasie und sprachliches Vermögen als Linearkombinationen der \mathbf{e}_1 und \mathbf{e}_2 erlauben, die basalere Merkmale repräsentieren. Phantasie und sprachliches Vermögen sind selbst komplexe, d.h. multivariate Konzepte.



nicht parallel und deshalb linear unabhängig, sie bilden eine Basis im \mathbb{R}^2 . Es 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 werden²⁷.

Satz 1.17 *Es sei $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ eine beliebige Basis eines n -dimensionalen Vektorraums. Dann läßt sich aus \mathcal{B} eine Orthonormalbasis konstruieren.*

Beweis:²⁸ Zunächst wird \mathbf{b}_1 normiert, d.h. es sei

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

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

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

²⁷Derartige Repräsentationen von Kompetenzen bzw. Eigenschaften sind einerseits Standard, andererseits nicht unproblematisch, denn Ausdrücke wie 'Phantasie', 'sprachliches Vermögen' etc. bezeichnen komplexe Prozesse und 'poetisches Vermögen' steht für eine Interaktion dieser Prozesse, und der Ansatz, diese Interaktion durch einfache Addition darzustellen, kann eine zu große Vereinfachung bedeuten. Dieses Problem kann hier nicht diskutiert werden. Es sei aber angemerkt, dass die Vektoralgebra ebenfalls bei der Modellierung dynamischer Prozesse benötigt wird.

²⁸Nach Ehrhard Schmidt (1876 – 1956). s. Koecher (1997), p. 157

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

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

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

Anmerkung: Das im Beweis beschriebene Vorgehen zur Konstruktion einer orthogonalen Basis ist auch als Gram-Schmidt-Verfahren bekannt. \square

Orthogonale Teilräume

Satz 1.18 *Es sei $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ eine orthonormale Basis eines n -dimensionalen Vektorraums V , und es sei $\mathcal{B}_k = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$, $k < n$, und die lineare Hülle $\mathcal{L}_k = \mathcal{L}(\mathcal{B}_k)$ definiert einen Teilraum von V (Satz 1.8, Seite 39). Dann existiert ein Vektor $\mathbf{n} \in V$, der orthogonal zu allen Vektoren aus \mathcal{L}_k ist.*

Anmerkung: Der Vektor \mathbf{n} heißt auch *Normalenvektor*, vergl. die Gleichung (1.81), Seite 36.

Beweis: Es ist $V = \mathcal{L}(\mathbf{v}_1, \dots, \mathbf{v}_n)$, trivialerweise ist $\mathbf{v}_{k+1} \in V$. 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 $\mathbf{v}_{k+1} = \mathbf{n}$ ein Normalenvektor für alle Vektoren aus \mathcal{L}_k . \square

Satz 1.19 *V , \mathcal{B} und \mathcal{L}_k seien wie im vorangegangenen Satz definiert. Es sei $\mathcal{B}_c = \{\mathbf{v}_{k+1}, \dots, \mathbf{v}_n\}$ und $\mathcal{L}_c = \mathcal{L}(\mathcal{B}_c)$ sei der zu \mathcal{B}_c korrespondierende Teilraum von V . Weiter seien $\mathbf{x} \in \mathcal{L}_k$ und $\mathbf{y} \in \mathcal{L}_c$. Dann gilt $\mathbf{x} \perp \mathbf{y}$, d.h. \mathbf{x} und \mathbf{y} sind orthogonal.*

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

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

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

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

Definition 1.15 Es sei V ein n -dimensionaler Vektorraum und $U \subset V$ sei ein Teilraum von V . Weiter sei²⁹

$$U^\perp = \{x \in V \mid x'u = 0, \forall u \in U\} \quad (1.105)$$

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

Beispiel 1.11 Es sei $V = \mathbb{R}^2$ ein 2-dimensionaler Vektorraum und U sei ein 1-dimensionaler Teilraum von V , d.h. U sei eine Gerade durch den Ursprung des Koordinatensystems mit der Orientierung eines Vektors $\mathbf{a} \in V$; die Gerade werde mit $\mathbb{R}\mathbf{a}$ bezeichnet. Dann existiert ein Vektor $\mathbf{b} \in V$ mit $\mathbf{b} \perp \mathbf{a}$ und die Gerade $\mathbb{R}\mathbf{b}$ ist das zu U orthogonale Komplement von $\mathbb{R}\mathbf{a}$ durch den Ursprung. \square

Beispiel 1.12 Es sei $V = \mathbb{R}^3$ und U sei ein 2-dimensionaler Teilraum von V , d.h. eine Ebene in V . Es existieren stets zwei orthogonale Vektoren $\mathbf{b}_1, \mathbf{b}_2 \in U$, $\mathbf{b}_1 \perp \mathbf{b}_2$, und $\mathbf{n} \in V$ sei der Normalenvektor für U , d.h. \mathbf{n} stehe senkrecht auf U . Dann ist $U^\perp = \mathcal{L}(\mathbf{n})$, d.h. die Menge der zu \mathbf{n} parallelen Vektoren das orthogonale Komplement zur Ebene. \square

Kombination von Teilräumen:

Definition 1.16 Es sei W ein n -dimensionaler Vektorraum und U und V seien zwei Teilräume von W . W heißt orthogonale Summe von U und V , in Zeichen $W = U \perp V$, wenn

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

Beispiel 1.13 (Fortsetzung Beispiel 1.12) Es sei $W = \mathbb{R}^3$ und $\{\mathbf{b}_1, \mathbf{b}_2\}$, $\mathbf{b}_1 \perp \mathbf{b}_2$, sei eine 2-dimensionale Teilbasis von W . U sei die von $\mathbf{b}_1, \mathbf{b}_2$ aufgespannte Ebene, und \mathbf{n} sei der Normalenvektor für diese Ebene. Nun sei \mathbf{w} ein beliebiger Vektor aus W , $\mathbf{w} \notin U$ und $\mathbf{w} \nparallel \mathbf{n}$. Dann existieren Koeffizienten a_1, a_2, a_3 derart, dass

$$\mathbf{w} = a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + a_3\mathbf{n} = \mathbf{u} + \mathbf{v},$$

mit $\mathbf{u} = a_1\mathbf{b}_1 + a_2\mathbf{b}_2 \in U$ und $\mathbf{v} = a_3\mathbf{n} \in V$. \square

Satz 1.20 (Dimensionalitätssatz)³⁰ Es seien $U, V \subset W$, W ein n -dimensionaler Vektorraum. Die folgenden Aussagen (i) und (ii) sind äquivalent:

- (i) V ist das orthogonale Komplement von U , d.h. $V = U^\perp$,
- (ii) $W = U \perp V$, d.h. W ist die orthogonale Summe von U und V . Dann gilt

$$\dim W = \dim U + \dim V^\perp. \quad (1.106)$$

²⁹ \forall steht für "für alle"

³⁰Koecher, M.: Lineare Algebra und analytische Geometrie. Berlin 1997, p. 160

Beweis: (i) \Rightarrow (ii): Zu zeigen ist $W = U \perp U^\perp$. Es sei $\{\mathbf{b}_1, \dots, \mathbf{b}_r\}$ eine orthonormale Basis des Teilraums U ; sie kann nach Satz 1.17 zu einer orthonormalen Basis $\{\mathbf{b}_1, \dots, \mathbf{b}_r, \mathbf{b}_{r+1}, \dots, \mathbf{b}_n\}$ von W erweitert werden. Nach Definition von U^\perp ist $U^\perp = \mathcal{L}(\mathbf{b}_{r+1}, \dots, \mathbf{b}_n)$, und da $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ eine Basis von W ist, folgt Aussage (i). Da die $\mathbf{b}_j, j = 1, \dots, n$ paarweise orthonormal sind, folgt die Aussage (ii) und damit (1.106).

(ii) \Rightarrow (i): Nach (ii) der Definition 1.16 gilt $\mathbf{u}'\mathbf{v} = 0$ für $W = U \perp V$; Sei $\mathbf{x} \in W$. Dann $\mathbf{x} = \mathbf{u} + \mathbf{v}$ mit $\mathbf{u} \in U$ und $\mathbf{v} \in V$. Dann \mathbf{x}^\perp genau dann, wenn $\mathbf{x}'\mathbf{w} = 0$ für alle $\mathbf{w} \in U$, also $\mathbf{u}'\mathbf{w} = 0$ für alle $\mathbf{w} \in U$, d.h. $\mathbf{u} = \vec{0}$. \square

Es gilt der

Satz 1.21 *Es sei W ein Vektorraum und $U \subset W, V \subset W$ seien Teilräume von W . Dann gilt*

1. *Der Durchschnitt $U \cap V$ ist wieder ein Teilraum von W ,*
2. *Die Vereinigung $U \cup V$ ist dann und nur dann ein Teilraum von W , wenn entweder $U \subseteq V$ oder $V \subseteq U$.*

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

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

Beispiel 1.14 Als Beispiel für eine Vereinigung zweier Teilräume betrachte man den 2-dimensionalen Vektorraum, wobei U und V zwei nicht parallele Geraden seien. Sicherlich sind U und V Teilräume, aber ihre Vereinigung ist kein Teilraum: für $\mathbf{x}_1 \in U$ und $\mathbf{x}_2 \in V$ ist $\mathbf{x}_1 + \mathbf{x}_2 \notin U \cup V$. \square

Definition 1.17 *Es seien U und V Teilräume eines Vektorraums W . Dann heißt*

$$U + V = \{\mathbf{u} + \mathbf{v} \mid \mathbf{u} \in U, \mathbf{v} \in V\} \quad (1.107)$$

die Summe der Teilräume U und V .

Satz 1.22 Die Summe zweier Teilräume eines Vektorraums W ist wieder ein Teilraum von W .

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

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

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

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

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

Definition 1.18 Es sei V ein Vektorraum und S eine Teilmenge von V . Dann ist der Rang von S gleich der Dimension des von S erzeugten Teilraums $\mathcal{L}(S)$. Ist $V = V_n$ ein n -dimensionaler Vektorraum und ist $S \subset V_n$, so hat S den Rang $r < n$, wenn S von r linear unabhängigen Vektoren aufgespannt wird; für $r = n$ hat S den vollen Rang.

Anmerkungen:

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

2. Es sei $S = \{\mathbf{x}_1, \dots, \mathbf{x}_p\}$, $\mathbf{x}_j \in \mathbb{R}^n$ für $j = 1, \dots, p$. Dann existieren linear unabhängige Vektoren $\mathbf{u}_1, \dots, \mathbf{u}_r$, $\mathbf{u}_k \in \mathbb{R}^n$ und $r \leq p$ und Koeffizienten a_{1j}, \dots, a_{rj} derart, dass

$$\mathbf{x}_j = a_{1j}\mathbf{u}_1 + a_{2j}\mathbf{u}_2 + \dots + a_{rj}\mathbf{u}_r, \quad j = 1, \dots, p.$$

Dann ist $S \subset \mathcal{L}(\mathbf{u}_1, \dots, \mathbf{u}_r)$ und

$$\text{rg}(S) = \text{rg}(\mathcal{L}(\mathbf{u}_1, \dots, \mathbf{u}_r)) = r \leq p. \quad (1.108)$$

\square

Satz 1.23 Für die Summe der Teilräume gilt die Dimensionsformel

$$\dim(U + V) = \dim(U) + \dim(V) - \dim(U \cap V). \quad (1.109)$$

Beweis:³¹ Die Dimension eines (Teil-)Vektorraums ist gleich der Anzahl der Basisvektoren des Raums. $U \cap V$ ist ein Teilraum und habe die Basis $\mathbf{a}_1, \dots, \mathbf{a}_m$,

³¹Nach Lorenz I, p. 56

also $m = \dim(U \cap V)$. Diese Basis kann zu Basen von einerseits U , andererseits V ergänzt werden: $\mathbf{a}_1, \dots, \mathbf{a}_m, \mathbf{b}_1, \dots, \mathbf{b}_r$ sei die Basis von U , also $\dim(U) = m + r$, und $\mathbf{a}_1, \dots, \mathbf{a}_m, \mathbf{c}_1, \dots, \mathbf{c}_s$ sei die Basis von V , also $\dim(V) = m + 2$, und $\dim(U \cap V) = m$. Dann bildet

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

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

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

Zunächst muss gezeigt werden, dass auch

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

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

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

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

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

und das war zu zeigen. □

Korollar 1.3 *Es seien U und V Teilräume eines Vektorraums W . Aus (1.109) folgt*

$$\dim(U + V) \leq \dim(U) + \dim(V), \quad (1.110)$$

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

1.4 Lineare Gleichungssysteme I

Die Begriffe der linearen Abhängigkeit und Unabhängigkeit stehen in engem Zusammenhang mit der Frage nach den Lösungen linearer Gleichungssysteme. Am folgenden Beispiel eines Systems von zwei Gleichungen mit zwei Unbekannten wird der Zusammenhang recht explizit illustriert, woran sich eine Diskussion des allgemeinen Falls anschließt. Weitere Aspekte lineare Gleichungssysteme werden in Abschnitt 3.4 vorgestellt.

Gegeben seien drei 2-dimensionale Vektoren \mathbf{y} , \mathbf{x}_1 und \mathbf{x}_2 , alle $\neq \vec{0}$. Die Frage sei, ob sich \mathbf{y} als Linearkombination der Vektoren \mathbf{x}_1 und \mathbf{x}_2 darstellen lässt, d.h. ob Koeffizienten $a_1, a_2 \in \mathbb{R}$ existieren derart, dass

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 = a_1 \begin{pmatrix} x_{11} \\ x_{21} \end{pmatrix} + a_2 \begin{pmatrix} x_{12} \\ x_{22} \end{pmatrix} \quad (1.111)$$

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

$$y_1 = a_1 x_{11} + a_2 x_{12} \quad (1.112)$$

$$y_2 = a_1 x_{21} + a_2 x_{22} \quad (1.113)$$

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

$$a_1 = \frac{y_1 x_{22} - y_2 x_{12}}{x_{11} x_{22} - x_{12} x_{21}} \quad (1.114)$$

$$a_2 = \frac{y_2 x_{11} - y_1 x_{21}}{x_{11} x_{22} - x_{12} x_{21}} \quad (1.115)$$

Man kann zwei Fälle unterscheiden: (i) das homogene System $\mathbf{y} = \vec{0}$, bei dem $y_1 = y_2 = 0$, und (ii) das inhomogene System für den Fall $\mathbf{y} \neq \vec{0}$. Unter der Voraussetzung, dass der Nenner $x_{11} x_{22} - x_{12} x_{21}$ in den beiden Gleichungen ungleich null ist, folgt aus (1.114) und (1.115) sofort $a_1 = a_2 = 0$ als einzige Lösung, d.h. die Linearkombination $\vec{0} = a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2$ ist nur für die triviale Lösung möglich. Demnach sind die Vektoren \mathbf{x}_1 und \mathbf{x}_2 linear unabhängig. Wie sich herausstellt ist dieser Fall eine Implikation der Voraussetzung $x_{11} x_{22} - x_{12} x_{21} \neq 0$; umgekehrt impliziert die lineare Unabhängigkeit aber auch diese Voraussetzung.

Es sei $\mathbf{y} \in \mathcal{L}(\mathbf{x}_1, \mathbf{x}_2)$, d.h. \mathbf{y} sei eine Linearkombination der Vektoren \mathbf{x}_1 und \mathbf{x}_2 . Nach Satz 1.10 existiert eine eindeutige Lösung $\mathbf{a} = (a_1, a_2)'$ genau dann, wenn \mathbf{x}_1 und \mathbf{x}_2 linear unabhängig und damit nicht parallel sind. Es sei also³²

$$\Delta = x_{11} x_{22} - x_{12} x_{21}; \quad (1.116)$$

offenbar ist $\Delta \neq 0$ eine *notwendige Bedingung* für die Berechenbarkeit von a_1 und a_2 und damit für die Existenz einer Lösung $\mathbf{a} = (a_1, a_2)'$ für die Gleichungen (1.114) und (1.115). Es gelte insbesondere $\Delta = 0$. Dann folgt

$$\frac{x_{21}}{x_{11}} = \frac{x_{22}}{x_{12}} \quad (1.117)$$

und $x_{21}/x_{11} = \tan \theta$, θ der Winkel, der die Orientierung von \mathbf{x}_1 angibt³³, und (1.117) besagt, dass dieser Winkel identisch mit dem von \mathbf{x}_2 ist. Daraus folgt,

³² Δ ist der Buchstabe Delta aus dem griechischen Alphabet; er wird oft für die Bezeichnung einer Differenz verwendet.

³³In Bezug auf das Koordinatensystem, in dem die x_{ij} die Komponenten der Vektoren \mathbf{x}_1 und \mathbf{x}_2 sind.

dass $\Delta = 0$ dann erfüllt ist, wenn \mathbf{x}_1 und \mathbf{x}_2 parallel (kollinear) sind. Weiter gilt $\mathbf{x}_2 = \lambda \mathbf{x}_1$ mit $0 \neq \lambda \in \mathbb{R}$, und dies bedeutet, dass \mathbf{x}_1 und \mathbf{x}_2 linear abhängig sind, denn man hat nun $b_1 \mathbf{x}_1 + b_2 \mathbf{x}_1 = \vec{0}$ mit $b_1 = 1$ and $b_2 = -\lambda \neq 0$. Die lineare Abhängigkeit von \mathbf{x}_1 und \mathbf{x}_2 bedeutet, dass jeweils einer von ihnen als Linearkombination des anderen dargestellt werden kann, etwa $\mathbf{x}_2 = a \mathbf{x}_1$, $0 \neq a \in \mathbb{R}$. Dann

$$\mathbf{y} = a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 = a_1 \mathbf{x}_1 + a_2 a \mathbf{x}_1 = (a_1 + a a_2) \mathbf{x}_1 = \alpha \mathbf{x}_1,$$

und diese Beziehung bedeutet, dass $\alpha = a_1 + a a_2$ nur dann existiert, wenn \mathbf{y} parallel zu \mathbf{x}_1 (und \mathbf{x}_2) ist. Hat \mathbf{y} nicht dieselbe Orientierung wie \mathbf{x}_1 und \mathbf{x}_2 , so kann \mathbf{y} nicht als Linearkombination von \mathbf{x}_1 und \mathbf{x}_2 repräsentiert werden.

Nun gelte $\Delta \neq 0$. Dann sind \mathbf{x}_1 und \mathbf{x}_2 nicht parallel. Aus der linearen Abhängigkeit zweier Vektoren folgt, dass sie parallel sind, also folgt aus ihrer Nichtparallelität ihre lineare Unabhängigkeit, und die Koeffizienten a_1 und a_2 können aus den Gleichungen (1.114) und (1.115) bestimmt werden. Die a_1 und a_2 sind dann für einen beliebigen Vektor $\mathbf{y} \in \mathbb{R}^2$ eindeutig bestimmt, vergl. auch Satz 1.10, und es gilt $\mathcal{L}(\mathbf{x}_1, \mathbf{x}_2) = \mathbb{R}^2$. \square

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

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

zusammenfassen: die erste Spalte von X ist der Vektor \mathbf{x}_1 , die zweite Spalte ist der Vektor \mathbf{x}_2 . Dann heißt der Nenner $x_{11}x_{22} - x_{12}x_{21}$ in den Ausdrücken (1.112) und (1.113) die *Determinante* der (2×2) -Matrix X ; sie wird mit $|X|$ bezeichnet:

$$|X| = x_{11}x_{22} - x_{12}x_{21}, \quad (1.118)$$

s. Anhang, Abschnitt 5.10, Seite 187. Die Determinante einer (n, n) -Matrix mit reellen Elementen ist stets gleich einer reellen Zahl und ist nur dann von Null verschieden, wenn die Spaltenvektoren der Matrix linear unabhängig sind, wie die vorangegangenen Betrachtungen zeigen. Diese Aussage gilt allgemein für (n, n) -Matrizen und damit für Systeme von n linearen Gleichungen mit n Unbekannten, sofern $|X| \neq 0$. Wie die Inspektion der Gleichungen (1.114) und (1.115) nahelegt haben auch die Zähler in den Ausdrücken für a_1 und a_2 die Form einer Determinante. So ersetze man in der Matrix X die erste Spalte durch den Vektor \mathbf{y} und berechne die Determinante der so veränderten Matrix:

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

In analoger Weise bildet man die Matrix X_2 , indem man die zweite Spalte durch \mathbf{y} ersetzt und berechnet die Determinante von X_2 , $|X_2| = y_2 x_{11} - y_1 x_{21}$. Für die a_j kann

$$a_j = \frac{|X_j|}{|X|}, \quad j = 1, 2, \quad |X| \neq 0 \quad (1.119)$$

geschrieben werden. Diese Gleichung ist die *Cramersche Regel*, nach der lineare Gleichungen für den allgemeinen Fall von $n > 2$ Gleichungen mit n Unbekannten gelöst werden können: man ersetzt in der (n, n) -Matrix X die j -te Spalte durch \mathbf{y} und erhält für die j -te Unbekannte die Lösung (1.119) für $j = 1, \dots, n$, vorausgesetzt, dass $|X| \neq 0$, d.h. für den Fall, dass die Spaltenvektoren von X linear unabhängig sind. \square

Natürlich kann man auch die Linearkombinationen $\mathbf{y} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2$ für $\mathbf{y}, \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^3$ bzw. allgemein für $\mathbf{y}, \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ betrachten. Man erhält dann $n > 2$ Gleichungen für die zwei Unbekannten a_1 und a_2 . Sind die Vektoren \mathbf{x}_1 und \mathbf{x}_2 parallel, so kann die Cramersche Regel nicht mehr angewendet werden schon weil $X = [\mathbf{x}_1, \mathbf{x}_2]$ nicht mehr quadratisch ist und Determinanten nur für quadratische Matrizen berechnet werden können, s. Abschnitt 5.10, Seite 187. Es existiert nur dann eine Lösung, wenn \mathbf{y} parallel zu diesen beiden Vektoren ist, d.h. wenn \mathbf{y}, \mathbf{x}_1 und \mathbf{x}_2 auf einer Geraden im \mathbb{R}^n liegen, also kollinear sind. Sind \mathbf{x}_1 und \mathbf{x}_2 nicht parallel, so liegen alle Linearkombinationen $\mathbf{y} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2$ in einer Ebene im \mathbb{R}^n , wie in Abschnitt 1.3 gezeigt werden wird. Dies bedeutet, dass für einen gegebenen Vektor \mathbf{y} nur dann die Koeffizienten a_1 und a_2 bestimmt werden können, wenn \mathbf{y} in der von \mathbf{x}_1 und \mathbf{x}_2 *aufgespannten* Ebene liegen. In Beispiel 1.8, Seite 37, wird dieser Fall diskutiert.

Der allgemeine Fall Gegeben sei die Vektorgleichung

$$\mathbf{y} = a_1\mathbf{x}_1 + \dots + a_p\mathbf{x}_p, \quad (1.120)$$

wobei $\mathbf{y}, \mathbf{x}_1, \dots, \mathbf{x}_p$ n -dimensionale Vektoren seien. Gibt man die Koeffizienten a_1, \dots, a_p vor, so liefert (1.120) einen zu a_1, \dots, a_p korrespondierenden Vektor \mathbf{y} eine Linearkombination der \mathbf{x}_j , $1 \leq j \leq p$. Gibt man umgekehrt den Vektor \mathbf{y} vor, so ergibt sich die Frage, ob überhaupt Koeffizienten a_1, \dots, a_p existieren derart, dass \mathbf{y} als Linearkombination der \mathbf{x}_j dargestellt werden kann. Die Gleichung (1.120) entspricht einem linearen Gleichungssystem in den Unbekannten a_j , $j = 1, \dots, p$; die Gleichungen sind

$$\begin{aligned} y_1 &= a_1x_{11} + a_2x_{12} + \dots + a_px_{1p} \\ y_2 &= a_1x_{21} + a_2x_{22} + \dots + a_px_{2p} \\ &\vdots \\ y_n &= a_1x_{n1} + a_2x_{n2} + \dots + a_px_{np} \end{aligned} \quad (1.121)$$

Eine sehr allgemeine Antwort auf die Frage nach der Existenz eines "Lösungsvektors" $\mathbf{a} = (a_1, \dots, a_p)'$ ist, dass \mathbf{y} ein Element der linearen Hülle $\mathcal{L}(\mathbf{x}_1, \dots, \mathbf{x}_p)$ sein muß, damit \mathbf{a} existiert.

Es werden zunächst einige grundsätzliche Betrachtungen³⁴ zur Menge der Lösungen angestellt; lineare Gleichungssysteme werden in größerer Allgemeinheit in Abschnitt 3.4 besprochen, wenn einige Ergebnisse der Matrixalgebra zur Verfügung stehen.

³⁴vergl. Lorenz I (1988), p. 6

Dazu werden zwei mögliche Fälle unterschieden:

$$a_1 \mathbf{x}_1 + \cdots + a_p \mathbf{x}_p = \mathbf{y} = \begin{cases} = \vec{0}, & \text{homogenes System} \\ \neq \vec{0}, & \text{inhomogenes System} \end{cases} \quad (1.122)$$

Satz 1.24 Die Menge aller Lösungen eines inhomogenen Gleichungssystems ist die Vereinigung einer speziellen Lösung für das inhomogene System mit der Menge der Lösungen für das zugehörige homogene System.

Beweis: Für $\mathbf{y} \notin \mathcal{L}(\mathbf{x}_1, \dots, \mathbf{x}_p)$ hat das System keine Lösung, so dass auch nichts über mögliche Lösungen gesagt werden kann. Nun seien $\mathbf{a} = (a_1, \dots, a_p)'$ und $\mathbf{b} = (b_1, \dots, b_p)'$ zwei Lösungen des inhomogenen Systems (1.120), so dass die Gleichungen

$$a_1 \mathbf{x}_1 + \cdots + a_p \mathbf{x}_p = \mathbf{y} \quad (1.123)$$

$$b_1 \mathbf{x}_1 + \cdots + b_p \mathbf{x}_p = \mathbf{y} \quad (1.124)$$

gelten. Subtrahiert man die Gleichung (1.124) von der Gleichung (1.123), so erhält man

$$(a_1 - b_1) \mathbf{x}_1 + \cdots + (a_p - b_p) \mathbf{x}_p = \vec{0}, \quad (1.125)$$

d.h. der Vektor $\mathbf{c} = \mathbf{a} - \mathbf{b} = (a_1 - b_1, a_2 - b_2, \dots, a_p - b_p)'$ ist eine Lösung des zugehörigen homogenen Systems von Gleichungen. Daraus folgt aber, dass jede Lösung \mathbf{a} des inhomogenen Systems die Form

$$\mathbf{a} = \mathbf{b} + \mathbf{c} \quad (1.126)$$

hat, wobei \mathbf{c} eine Lösung des homogenen Systems ist.

Nun sei $\mathbf{c} = (c_1, c_2, \dots, c_p)'$ eine Lösung des homogenen Systems, d.h. es gilt

$$c_1 \mathbf{x}_1 + \cdots + c_p \mathbf{x}_p = \vec{0}$$

Dann ist jeder Vektor \mathbf{a} , der (1.126) erfüllt, eine Lösung des inhomogenen Systems, denn man hat nun

$$\begin{aligned} a_1 \mathbf{x}_1 + \cdots + a_p \mathbf{x}_p &= (b_1 + c_1) \mathbf{x}_1 + \cdots + (b_p + c_p) \mathbf{x}_p \\ &= (b_1 \mathbf{x}_1 + \cdots + b_p \mathbf{x}_p) + (c_1 \mathbf{x}_1 + \cdots + c_p \mathbf{x}_p) \\ &= \mathbf{y} + \vec{0} = \mathbf{y}, \end{aligned} \quad (1.127)$$

d.h. die eine Lösung für das System (1.120) ist eine Summe einer speziellen Lösung für das inhomogene und einer Lösung für das homogene System. \square

Es sei L die Menge aller Lösungen für das System (1.120), und L_0 sei die Menge aller Lösungen für das zugehörige homogene System. Gilt also $\mathbf{y} \in \mathcal{L}(\mathbf{x}_1, \dots, \mathbf{x}_p)$ und ist \mathbf{b} ein Lösungsvektor für das inhomogene System, so kann Satz 1.24, Seite 59 in der Form

$$L = \mathbf{b} + L_0 \quad (1.128)$$

angeschrieben werden. Man hat nun die

Folgerungen:

1. Sind die Vektoren \mathbf{b} und \mathbf{c} Elemente von L_0 , so ist auch die Summe $\mathbf{b} + \mathbf{c} \in L_0$,
2. ist $\mathbf{b} \in L_0$, so auch $\lambda \mathbf{b} \in L_0$ mit $\lambda \in \mathbb{R}$, wie man leicht nachrechnet.

Damit hat man eine Aussage über die möglichen Lösungen des homogenen Systems: sie müssen den genannten Folgerungen 1. und 2. genügen; das Skalarprodukt der Elemente von L_0 existiert sicherlich. Damit ist L_0 aber als Teilraum des \mathbb{R}^p charakterisiert (vergl. Definition 1.7), Seite 32. Vektoren, die diesen Aussagen nicht genügen, sind keine Lösungen des homogenen Systems.

Die Aussage (1.128) ist eine allgemeine Aussage über die Lösungsmenge, die aber noch weiter zu spezifizieren ist. Dabei gibt es drei Fälle zu betrachten:

1. $n < p$, die Zahl n der Gleichungen ist kleiner als die der Unbekannten,
2. $n = p$, es gibt so viele Gleichungen wie Unbekannte, und
3. $n > p$, es gibt mehr Gleichungen als Unbekannte.

Der Standardansatz zur Diskussion dieser Fälle besteht darin, die Vektoren \mathbf{x}_j spaltenweise zu einer Matrix X zusammenzufassen und darauf die elementaren Umformungen anzuwenden; dieser Ansatz wird der Vollständigkeit halber im Anhang, Abschnitt 5.4.1 vorgestellt. Die elementaren Umformungen werden im Folgenden aber nicht weiter benötigt, so dass hier nicht weiter darauf eingegangen werden soll. Die Implikationen der Fälle 1. bis 3. werden in Abschnitt 3.4 diskutiert, wenn bestimmte Ergebnisse der Matrixalgebra zur Verfügung stehen.

Anmerkung: Die Bedingung $p \leq n$ ist notwendig; es wird gezeigt werden, dass der Fall $p > n$ nicht mit der linearen Unabhängigkeit der $\mathbf{x}_1, \dots, \mathbf{x}_p$ verträglich ist, d.h. es können nicht mehr als n Vektoren aus \mathbb{R}^n linear unabhängig sein, vergl. Satz 1.12, Seite 43. \square

Betrachtet man (1.123) also als lineares Gleichungssystem mit $\mathbf{y} \in \mathcal{L}(\mathbf{x}_1, \dots, \mathbf{x}_p)$, so bedeutet Satz 1.10 die Eindeutigkeit der Lösung \mathbf{a} des Systems.

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

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] = \begin{bmatrix} \tilde{\mathbf{x}}'_1 \\ \tilde{\mathbf{x}}'_2 \\ \vdots \\ \tilde{\mathbf{x}}'_m \end{bmatrix} \quad (2.1)$$

Dabei sind die $\mathbf{x}_j = (x_{1j}, x_{2j}, \dots, x_{mj})'$, $j = 1, \dots, n$ die *Spaltenvektoren* von X , und die $\tilde{\mathbf{x}}'_i = (x_{i1}, x_{i2}, \dots, x_{in})$, $i = 1, \dots, m$, sind die *Zeilenvektoren* von X . Die Schreibweise $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ soll anzeigen, dass X aus den nebeneinander angeschriebenen Spaltenvektoren \mathbf{x}_j , $j = 1, \dots, n$, besteht, die Schreibweise

$$\begin{bmatrix} \tilde{\mathbf{x}}'_1 \\ \tilde{\mathbf{x}}'_2 \\ \vdots \\ \tilde{\mathbf{x}}'_m \end{bmatrix}$$

bedeutet, dass X ebenfalls als durch untereinander geschriebene Zeilenvektoren gebildet aufgefasst werden kann. Die $\tilde{\mathbf{x}}_i$ sind die Spaltenvektoren der *gestürzten* oder *transponierten* Matrix

$$X' = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{m1} \\ x_{12} & x_{22} & \cdots & x_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{mn} \end{pmatrix} = [\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_m] = \begin{bmatrix} \mathbf{x}'_1 \\ \mathbf{x}'_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} \quad (2.2)$$

die aus X hervorgeht, wenn man die Spaltenvektoren von X als Zeilenvektoren anschreibt. Weil die $\tilde{\mathbf{x}}_i$ als Spaltenvektoren von X' definiert sind, werden die Zeilen von X als transponierte Vektoren $\tilde{\mathbf{x}}'_i$ notiert.

X heißt auch (m, n) -Matrix; gelegentlich wird einfach $X_{m,n}$ dafür geschrieben, oder $X = (x_{ij})_{m,n}$ oder $X = (x_{ij})$, wenn klar ist, dass $1 \leq i \leq m$, $1 \leq j \leq n$. Eine andere Schreibweise ist $X \in \mathbb{R}^{m,n}$, womit angedeutet wird, dass die Elemente von X reelle Zahlen sind, denn man kann auch Matrizen betrachten, deren Elemente komplexe Zahlen sind. Wenn X eine (m, n) -Matrix ist, so ist X' eine (n, m) -Matrix.

X heißt *quadratisch*, wenn $m = n$. Die Matrix X heißt *symmetrisch*, wenn

$$X' = X, \quad x_{ij} = x_{ji}, \quad 1 \leq i, j \leq n \quad (2.3)$$

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

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{pmatrix}. \quad (2.4)$$

Λ das große, λ das kleine Lambda, und $x_{ii} = \lambda_i$.

2.2 Operationen mit Matrizen

2.2.1 Addition und Multiplikation mit einem Skalar

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

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

2.2.2 Multiplikation einer Matrix mit einem Vektor

Gegeben seien n m -dimensionale Vektoren $\mathbf{x}_1, \dots, \mathbf{x}_n$, und es werde die Linearkombination

$$\mathbf{y} = a_1 \mathbf{x}_1 + \dots + a_n \mathbf{x}_n \quad (2.6)$$

$$y_i = a_1 x_{i1} + \dots + a_n x_{in} = \tilde{\mathbf{x}}_i' \mathbf{a} = \mathbf{a}' \tilde{\mathbf{x}}_i, \quad i = 1, \dots, m \quad (2.7)$$

betrachtet. Die Vektoren \mathbf{x}_j lassen sich spaltenweise zu einer Matrix $X = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ zusammenfassen, und die Koeffizienten a_j , $j = 1, \dots, n$ können zu einem Vektor $\mathbf{a} = (a_1, a_2, \dots, a_n)'$ zusammengefasst werden. Für (2.6) wird nun die Schreibweise

$$\mathbf{y} = X \mathbf{a}, \quad \mathbf{a} \in \mathbb{R}^n, \quad \mathbf{y} \in \mathbb{R}^m \quad (2.8)$$

eingeführt. Mit $X \mathbf{a}$ ist im Folgenden also stets eine Linearkombination der Spaltenvektoren der Matrix X gemeint, wobei die Komponenten von \mathbf{a} die Koeffizienten sind. Für $\mathbf{y} = X \mathbf{a}$ sind dann die Komponenten von \mathbf{y} durch das Skalarprodukt (2.7) gegeben, d.h. die i -te Komponente von \mathbf{y} kann auch

$$y_i = (x_{i1} a_1 + x_{i2} a_2 + \dots + x_{in} a_n) = \sum_{j=1}^n x_{ij} a_j \quad (2.9)$$

angeschrieben werden. Man kann das Produkt $\mathbf{y} = X \mathbf{a}$ auch durch (2.9) definieren; dann folgt, dass $\mathbf{y} = X \mathbf{a}$ einer Linearkombination der Spalten von X entspricht. Die beiden Definitionen des Produkts $X \mathbf{a}$ sind also äquivalent. Man sagt, dass (2.8) die Multiplikation einer Matrix mit einem Spaltenvektor $\mathbf{a} \in \mathbb{R}^n$ von rechts bedeutet.

Bevor auf die Multiplikation einer Matrix mit einem Vektor *von links* eingegangen wird, soll eine Implikation des Produkts einer Matrix mit einem Vektor von rechts gezeigt werden. Aus (2.8) folgt $\mathbf{y}' = (X\mathbf{a})'$, und es gilt die Beziehung

$$(X\mathbf{a})' = \mathbf{a}'X', \quad (2.10)$$

denn

$$\begin{aligned} \mathbf{y}' &= (X\mathbf{a})' & (2.11) \\ &= \left(\begin{array}{c} \tilde{\mathbf{x}}_1' \mathbf{a} \\ \tilde{\mathbf{x}}_2' \mathbf{a} \\ \vdots \\ \tilde{\mathbf{x}}_m' \mathbf{a} \end{array} \right)' = (\mathbf{a}'\tilde{\mathbf{x}}_1, \mathbf{a}'\tilde{\mathbf{x}}_2, \dots, \mathbf{a}'\tilde{\mathbf{x}}_m) \\ &= \mathbf{a}'[\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_m] \\ &= \mathbf{a}'X' = [\mathbf{a}'\tilde{\mathbf{x}}_1, \mathbf{a}'\tilde{\mathbf{x}}_2, \dots, \mathbf{a}'\tilde{\mathbf{x}}_m] \quad \square & (2.12) \end{aligned}$$

Um die Multiplikation einer Matrix X mit einem Vektor *von links* zu erklären werden die Linearkombinationen der Zeilenvektoren von X betrachtet. Die Zeilenvektoren von X sind die Spaltenvektoren $\tilde{\mathbf{x}}_i$ der Transponierten X' von X . Es sei also

$$\mathbf{z} = X'\mathbf{b}, \quad \mathbf{z} \in \mathbb{R}^n, \quad \mathbf{b} \in \mathbb{R}^m \quad (2.13)$$

Wendet man hierauf die Beziehung (2.10) an, wobei X durch X' und \mathbf{a} durch \mathbf{b} ersetzt wird, so erhält man

$$\mathbf{z}' = \mathbf{b}'X. \quad (2.14)$$

Damit wird erklärt, was die Multiplikation einer Matrix X von links mit einem Vektor bedeutet: \mathbf{z}' ist der Zeilenvektor

$$\mathbf{z}' = \mathbf{b}'X = (\mathbf{b}'\mathbf{x}_1, \mathbf{b}'\mathbf{x}_2, \dots, \mathbf{b}'\mathbf{x}_n), \quad (2.15)$$

nach (2.14) also eine Linearkombination der Zeilenvektoren von X .

Das folgende Beispiel illustriert die hergeleiteten Beziehungen:

$$\begin{aligned} \mathbf{y} = X\mathbf{a} &= \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} x_{11}a_1 + x_{12}a_2 \\ x_{21}a_1 + x_{22}a_2 \end{pmatrix} \\ \mathbf{y}' = (X\mathbf{a})' &= (x_{11}a_1 + x_{12}a_2, x_{21}a_1 + x_{22}a_2) \end{aligned} \quad (2.16)$$

Berechnet man nun $\mathbf{y}' = \mathbf{a}'X'$ gemäß (2.10):

$$\mathbf{y}' = \mathbf{a}'X' = (a_1x_{11} + a_2x_{12}, a_1x_{21} + a_2x_{22}) \quad (2.17)$$

Vergleicht man den Zeilenvektoren der beiden Ausdrücke (2.16) und (2.17) für \mathbf{y}' auf den rechten Seiten, so sieht man, dass sie identisch sind bis auf die Reihenfolge der Terme in den Produkten: in (2.17) hat man a_1x_{11} statt $x_{11}a_1$ etc., aber auf diese Reihenfolgen kommt es nicht an, d.h. es gilt tatsächlich $\mathbf{y}' = \mathbf{a}'X'$. Zusammenfassend hat man die Interpretationen;

1. Das Produkt $\mathbf{y} = X\mathbf{a}$ repräsentiert eine Linearkombination der Spaltenvektoren von X ,
2. Das Produkt $\mathbf{z}' = \mathbf{b}'X$ repräsentiert eine Linearkombination der Zeilenvektoren von X .

Anmerkung: Die Gleichung $\mathbf{y} = X\mathbf{a}$ wurde oben als Kurzschreibweise für die Linearkombination der Spaltenvektoren von X bezeichnet. Tatsächlich kann man sie als *Transformation* auffassen: die Matrix X liefert eine Transformation des Vektors \mathbf{a} in den Vektor \mathbf{y} . Gleichbedeutend damit ist die Interpretation von X als Abbildung des Vektors \mathbf{a} auf den Vektor \mathbf{y} . Ist X eine (m, n) -Matrix mit $m \neq n$, so wird \mathbf{a} aus einem n -dimensionalen Vektorraum \mathbb{R}^n auf einen Vektor \mathbf{y} aus einem m -dimensionalen Vektorraum \mathbb{R}^m abgebildet, und für $m = n$ wird \mathbf{a} auf einen Vektoren \mathbf{y} aus demselben Vektorraum abgebildet. Viele Beziehungen zwischen Vektoren und Matrizen lassen sich als Eigenschaften dieser Abbildungen herleiten, worauf in Abschnitt 5.9 ausführlicher eingegangen wird. \square

2.2.3 Multiplikation einer Matrix mit einer Matrix

Es seien A eine (m, n) - und B eine (n, p) -Matrix mit den Spaltenvektoren \mathbf{b}_j . Dann ist

$$\mathbf{c}_j = A\mathbf{b}_j, \quad \mathbf{c}_j \in \mathbb{R}^m$$

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, p) - Matrix

$$C = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_p] = [A\mathbf{b}_1, A\mathbf{b}_2, \dots, A\mathbf{b}_p] = A[\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_p] = AB,$$

d.h. man erhält die Matrixgleichung

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

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

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

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

Es gilt:

1. Die Spaltenvektoren von C sind Linearkombinationen der Spaltenvektoren von A ,
2. Die Zeilenvektoren von C sind Linearkombinationen der Zeilenvektoren von B , vergl. die Anmerkung 2, Seite 64.

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

$$AB \neq BA. \quad (2.20)$$

Satz 2.1 *Es sei A eine (m, n) -Matrix, B sei eine (n, p) , und C eine (m, p) -Matrix, und es sei $C = AB$. Dann gilt*

$$C' = (AB)' = B' A' \quad (2.21)$$

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

$$(AB)C = A(BC) \quad (\text{Assoziativität}) \quad (2.22)$$

Beweis: Zu (2.21): Mit $C = AB = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n]$, erhält man wegen $\mathbf{c}_j = A\mathbf{b}_j \Rightarrow \mathbf{c}'_j = \mathbf{b}'_j A'$ (s. Abschnitt 2.2.2)

$$C' = \begin{bmatrix} \mathbf{c}'_1 \\ \mathbf{c}'_2 \\ \vdots \\ \mathbf{c}'_n \end{bmatrix} = \begin{bmatrix} \mathbf{b}'_1 A' \\ \mathbf{b}'_2 A' \\ \vdots \\ \mathbf{b}'_n A' \end{bmatrix} = \begin{bmatrix} \mathbf{b}'_1 \\ \mathbf{b}'_2 \\ \vdots \\ \mathbf{b}'_n \end{bmatrix} A' = B' A'.$$

Zu (2.22): Es sei $U = AB$, $V = BC$, so dass $F = (AB)C = UC$, $G = A(BC) = AV$. U ist eine $(m \times p)$ -Matrix, V ist eine $(n \times q)$ -Matrix, und F und G sind beide $(m \times q)$ -Matrizen. Es seien \mathbf{f}_j und \mathbf{g}_j die j -ten Spaltenvektoren von F bzw. G . Offenbar gilt

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

Nun ist $\mathbf{v}_j = B\mathbf{c}_j$, so dass $\mathbf{g}_j = AB\mathbf{c}_j = U\mathbf{c}_j$ folgt, was wiederum $\mathbf{f}_j = \mathbf{g}_j$ und damit $F = G$ bedeutet, – und das ist (2.22). \square

Die in (2.22) ausgedrückte Eigenschaft der Matrixmultiplikation wird als *assoziativ* bezeichnet. Sie ist aus der Multiplikation reeller Zahlen bekannt: Sind $0 \neq a, b, c \in \mathbb{R}$, so gilt $(ab)c = a(bc)$. Die Multiplikation reeller Zahlen ist auch *kommutativ*, d.h. es gilt stets auch $ab = ba$.

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

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

Man schreibt dafür auch kurz

$$\Lambda_n = \text{diag}(\lambda_1, \dots, \lambda_n), \quad \Lambda_m = \text{diag}(\lambda_1, \dots, \lambda_m) \quad (2.24)$$

Dann bedeutet

- (i) das Produkt $X\Lambda_n$ eine Längenskalierung der Spaltenvektoren von X ,
- (ii) das Produkt $\Lambda_m X$ eine Längenskalierung der Zeilenvektoren von X :

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

bzw.

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

Wie man sieht, ist der j -te Spaltenvektor von $X\Lambda_n$ gleich dem j -ten Spaltenvektor \mathbf{x}_j von X , multipliziert mit λ_j , und der i -te Zeilenvektor von $\Lambda_m X$ ist der i -te Zeilenvektor $\tilde{\mathbf{x}}_i$ von X multipliziert mit λ_i . Dies bedeutet eine Längenskalierung der \mathbf{x}_j bzw. der $\tilde{\mathbf{x}}_i$. Diagonalmatrizen sind Transformationsmatrizen, die die Orientierung der zu transformierenden Vektoren invariant lassen und nur die Länge der Vektoren verändern. In Abschnitt 3.1 wird eine Transformationsmatrix vorgestellt, die die Längen invariant läßt und die Orientierung verändert.

2.3 Der Rang einer Matrix

Es sei X eine (m, n) -Matrix, $X = [\mathbf{x}_1, \dots, \mathbf{x}_n]$, $\mathbf{x}_j \in \mathbb{R}^m$, $j = 1, \dots, n$, die Spaltenvektoren von X , und $X' = [\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_m]$, $\tilde{\mathbf{x}}_i \in \mathbb{R}^n$, $i = 1, \dots, m$, die Zeilenvektoren von X . Es werde angenommen, dass die Spaltenvektoren \mathbf{x}_j als Linearkombinationen von nicht mehr als $s \leq n$ linear unabhängigen (l.u.) m -dimensionalen Vektoren $\mathbf{u}_1, \dots, \mathbf{u}_s$ dargestellt werden können; die Anzahl s heißt dann der *Spaltenrang* der Matrix X . Weiter werde angenommen, dass die Zeilenvektoren $\tilde{\mathbf{x}}_i$ als Linearkombinationen von r l.u. n -dimensionalen Vektoren $\mathbf{v}_1, \dots, \mathbf{v}_r$ dargestellt werden können; r heißt der *Zeilenrang* von X . Fasst man die \mathbf{u}_k zu einer (m, s) -Matrix U und die \mathbf{v}_j zu einer (n, r) -Matrix V zusammen, so kann man

$$X = UA \quad (2.27)$$

$$X' = VB \quad (2.28)$$

schreiben, wobei $A = [\mathbf{a}_1, \dots, \mathbf{a}_n]$ eine (s, n) -Matrix ist, deren Spaltenvektoren \mathbf{a}_j Koeffizientenvektoren zur Darstellung der \mathbf{x}_j sind, d.h. $\mathbf{x}_j = U\mathbf{a}_j$, und $B = [\mathbf{b}_1, \dots, \mathbf{b}_m]$ ist eine (r, m) -Matrix, deren Spaltenvektoren \mathbf{b}_i die Koeffizientenvektoren zur Darstellung der Spaltenvektoren $\tilde{\mathbf{x}}_i$ von X' sind, d.h. $\tilde{\mathbf{x}}_i = V\mathbf{b}_i$.

Es gilt der folgende

Satz 2.2 (Rangatz) *Es sei X eine (m, n) -Matrix³⁵ mit dem Zeilenrang r und dem Spaltenrang s . Dann gilt*

$$r = s \quad (2.29)$$

d.h. der Zeilenrang ist stets gleich dem Spaltenrang, so dass nur vom Rang r der Matrix X gesprochen wird. Es gilt

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

Beweis: Nach (2.27) und (2.28) gilt

$$X = UA = B'V', \quad (2.31)$$

d.h. die Spaltenvektoren \mathbf{x}_j von X sind als Linearkombinationen der s Spaltenvektoren von U als auch der r Spaltenvektoren $\tilde{\mathbf{b}}_k$ von B' darstellbar. Nach Voraussetzung werden s l.u. Vektoren zur Darstellung der \mathbf{x}_j benötigt, d.h. die \mathbf{x}_j können nicht mit weniger als s linear unabhängigen m -dimensionalen Vektoren erzeugt werden. Dann kann r nicht kleiner als s sein, d.h. es folgt $r \geq s$.

Andererseits gilt

$$X' = VB = A'U' \quad (2.32)$$

d.h. die Spaltenvektoren $\tilde{\mathbf{x}}_i$ sind sowohl als Linearkombinationen der r Spaltenvektoren von V als auch der s Spaltenvektoren $\tilde{\mathbf{a}}_k$ von A' darstellbar. Nach Voraussetzung werden r l.u. Vektoren zur Darstellung der $\tilde{\mathbf{x}}_i$ benötigt, d.h. s kann nicht kleiner als r sein, d.h. es folgt $r \leq s$. Es muß also sowohl $r \geq s$ wie auch $r \leq s$ gelten, so dass $r = s$ folgt.

Es sei $m > n$. Da der Zeilenrang r gleich dem Spaltenrang s ist kann der Rang der Matrix nicht größer als n sein, so dass $r \leq \min(m, n)$. Die Argumentation für den Fall $m \leq n$ ist analog, d.h. es gilt (2.30). \square

Anmerkungen:

1. Der Beweis wird üblicherweise durch Anwendung elementarer Umformungen auf X geführt, s. Abschnitt 5.4.
2. Die (m, n) -Matrix X hat den Rang $r = 0$ dann und nur dann, wenn $X = 0$ die Nullmatrix ist, d.h. wenn $x_{ij} = 0$ für alle i und j ist.
3. X sei eine (m, n) -Datenmatrix, das Element X_{ij} von X sei die Messung der j -ten Variablen ($j = 1, \dots, n$) für den i -ten Fall ($i = 1, \dots, m$). Es sei s die Anzahl der linear unabhängigen Vektoren $\mathbf{u}_1, \dots, \mathbf{u}_s$, die zur Darstellung der Spaltenvektoren \mathbf{x}_j von X benötigt werden, und r sei die Anzahl der linear unabhängigen Vektoren $\mathbf{v}_1, \dots, \mathbf{v}_r$, die zur Darstellung der Zeilenvektoren $\tilde{\mathbf{x}}_i$ – also der Repräsentation der Fälle – benötigt werden. Es gilt stets $r = s$, d.h. die Anzahl der "latenten Vektoren" für (i) die Variablen und (ii) die Fälle ist notwendig identisch.

³⁵ X ist beliebig insofern, als es sich bei X nicht um eine Datenmatrix handeln muss, wie die einführenden Betrachtungen nahelegen könnten!

4. Es gelten die Beziehungen

$$X = UA \Rightarrow X' = A'U' \quad (2.33)$$

$$X' = VB \Rightarrow X = B'V' \quad (2.34)$$

Die Matrizen U und V können im Prinzip unabhängig voneinander gewählt werden, es kommt nur darauf an, dass sie denselben Rang haben. Hat man aber U gewählt, so liegt wegen des Satzes 1.10, Seite 40 die Matrix A und damit A' eindeutig fest. Die Implikation von (2.33) ist, dass A' und damit A ebenfalls den Rang r haben muß, d.h. dass die Spaltenvektoren von A' linear unabhängig sein müssen. Denn wären sie linear abhängig, so wären sie und damit die Spaltenvektoren $\tilde{\mathbf{x}}_i$ von X' als Linearkombinationen von weniger als r linear unabhängigen Vektoren $\mathbf{b}_1, \dots, \mathbf{b}_r$, $r_0 < r$, darstellbar, was wegen Satz 2.2 nicht sein kann. Man bemerke, dass die Spaltenvektoren von U Basisvektoren für die Darstellung der Spaltenvektoren \mathbf{x}_j von X sind, die Zeilenvektoren von U - also die Spaltenvektoren von U' aber die Koeffizientenvektoren für die $\tilde{\mathbf{x}}_i$ sind. In analoger Weise folgt, dass B' denselben Rang wie V haben muß, etc. Man kann die Matrix A umbenennen und $A = V'$ setzen, und erhält dann die Gleichungen

$$X = UV', \quad X' = VU'. \quad (2.35)$$

Die Gleichungen in (2.35) können für jede Matrix X angeschrieben werden, sie drücken einfach den Sachverhalt aus, dass die Spalten- und Zeilenvektoren von X jeweils als Linearkombinationen linear unabhängiger Vektoren dargestellt werden können, und man könnte die Gleichungen als *kanonische Darstellungen* von X bezeichnen, vergl. auch das Korollar 2.2, Seite 69 unten. Die Komponenten der Spaltenvektoren von U sind die Koordinaten der Fälle auf den latenten Variablen. Man überlegt sich leicht, dass die Zeilenvektoren von V die gemessenen und die Spaltenvektoren von V die latenten Variablen repräsentieren. Die Komponenten einer Zeile von V sind demnach die Koordinaten der Variablen auf den latenten Dimensionen. \square

Satz 2.3 *Es sei A eine (m, n) -Matrix. Es gelten die folgenden Aussagen:*

1. *B sei eine $(n \times p)$ -Matrix. Dann gilt*

$$\text{rg}(AB) \leq \min[\text{rg}(A), \text{rg}(B)]. \quad (2.36)$$

2. *Es sei Q eine (m, m) -Matrix und P eine (n, n) -Matrix, und sowohl P wie Q mögen vollen Rang haben, d.h. $\text{rg}(Q) = m$, $\text{rg}(P) = n$. Dann gilt*

$$\text{rg}(PAQ) = \text{rg}(A), \quad (2.37)$$

d.h. das Produkt von A von links mit einer Matrix mit vollem Rang und die Multiplikation von A von rechts mit einer Matrix mit vollem Rang hat denselben Rang wie A .

3. *A und B seien zwei (m, n) -Matrizen. Dann gilt*

$$\text{rg}(A + B) \leq \text{rg}(A) + \text{rg}(B) \quad (2.38)$$

Beweis: Zu 1. A habe den Rang s . Die Spaltenvektoren von $C = AB$ sind Linearkombinationen der Spaltenvektoren von A , d.h. $C \subseteq \mathcal{L}(A) = \mathcal{L}(\mathbf{a}_1, \dots, \mathbf{a}_s)$, $\mathbf{a}_1, \dots, \mathbf{a}_s$ eine Basis der linearen Hülle $\mathcal{L}(A)$ von A (s. Definition 1.9, Seite 38) 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 2.29 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 der Rang von C höchstens gleich s , d.h. $\text{rg}(C) \leq \min[\text{rg}(A), \text{rg}(B)]$.

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

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

Andererseits ist $P^{-1}B = A$, und wieder nach 1. folgt (zur Definition der Inversen Matrix P^{-1} s. Abschnitt 2.4)

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

d.h.

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

Es sei $C = AQ$; auf analoge Weise folgt $\text{rg}(C) = \text{rg}(AQ) = \text{rg}(A)$. Dann folgt aber auch $\text{rg}(A) = \text{rg}(PAQ)$.

Zu 3. Der Rang einer Matrix ist gleich der Dimension des von den Spalten der Matrix aufgespannten Teilraums, d.h. $\text{rg}(A) = \dim \mathcal{L}(A)$. Es sei $C = A + B$; dann ist die j -te Spalte \mathbf{c}_j von C durch $\mathbf{a}_j + \mathbf{b}_j$ gegeben, \mathbf{a}_j die j -te Spalte von A und \mathbf{b}_j die j -te Spalte von B . \mathbf{c}_j ist ein Element der Summe der Teilräume $\mathcal{L}(A)$ und $\mathcal{L}(B)$. Für die Dimensionalität der Summe zweier Teilräume gilt nach (2.38)

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

und das heißt (2.38). □

Korollar 2.1 *Es sei A eine (m, n) -Matrix mit dem Rang $\text{rg}(A) \leq \min(m, n)$, und P sei eine (m, m) -Matrix mit $\text{rg}(P) = m$. Dann gilt $\text{rg}(PA) = \text{rg}(A)$. Ist Q eine (m, n) -Matrix mit dem Rang $\text{rg}(Q) = n$, so folgt $\text{rg}(AQ) = \text{rg}(A)$.*

Beweis: Die Aussagen folgen direkt aus (2.37) für entweder $Q = I_n$ oder $P = I_m$, (vergl. auch den Beweis zu Satz 2.3, 3.). □

Korollar 2.2 *Es gelte $X = UA$ mit $\text{rg}(X) = \text{rg}(U)$. Dann gilt auch $\text{rg}(A) = r$.*

Beweis: Nach Satz 2.3, Gleichung (2.36) gilt

$$r = \text{rg}(X) = \text{rg}(UA) \leq \min(\text{rg}(U), \text{rg}(A)).$$

Angenommen, es sei $\text{rg}(A) = \text{rg}(V) = s < r$. Dann folgt $r = \text{rg}(UA) = s < r$, so dass die Annahme $\text{rg}(V) = s < r$ auf einen Widerspruch führt, – ebenso wie die Annahme $\text{rg}(A) = s > r$. Also muß $\text{rg}(A) = r$ gelten. \square

Die Koordinaten als Projektionen: Es sei $X = UV'$ eine (m, n) -Matrix, mit U eine (m, r) -Matrix und V eine orthonormale (m, r) -Matrix. Dann ist u_{ik} die Koordinate des i -ten Falls auf der k -ten latenten Dimension, und v_{jk} ist die Koordinate der j -ten gemessenen Variablen auf der k -ten Dimension:

$$u_{ik} = \tilde{\mathbf{x}}_i' \mathbf{v}_k, \quad k = 1, \dots, n \quad (2.39)$$

$$v_{jk} = \frac{1}{\lambda_k} \mathbf{x}'_j \mathbf{u}_k. \quad (2.40)$$

Die Beziehung (2.39) folgt sofort aus $X = UV'$, denn dann $XV = U$, und zu zeigen ist, dass es sich bei u_{ik} um eine Projektion handelt. das Skalarprodukt des k -ten Spaltenvektors von V und dem i -ten Zeilenvektors von X . Allgemein ist die Projektion \mathbf{x}_y eines Vektors \mathbf{x} auf einen Vektor \mathbf{y} nach (1.60), Seite 24, durch

$$\mathbf{x}_y = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{y}'\mathbf{y}}\mathbf{y}$$

gegeben Setzt man hier $\mathbf{x} = \tilde{\mathbf{x}}_i$ und $\mathbf{y} = \mathbf{v}_k$, so ist $\|\mathbf{y}\| = \|\mathbf{v}_k\| = 1$ und man erhält (2.39) entsprechend (1.61), Seite 24.

(2.40) erhält man aus $X' = VU'$. Multiplikation von rechts mit U liefert $X'U = VU'U = V\Lambda$ und damit

$$X'U\Lambda^{-1} = V, \quad \Lambda^{-1} = \text{diag}(1/\lambda_1, \dots, 1/\lambda_r).$$

v_{jk} ist dann das Skalarprodukt des j -ten Zeilenvektors von X' mit dem k -ten Spaltenvektor von U , multipliziert mit $1/\lambda_k$. Die Argumentation bezüglich der Projektion ist analog zu der der Fälle. (2.39) und (2.40) zeigen, dass die latenten Variablen für die Fälle einerseits und die gemessenen Variablen andererseits identisch sind.

Spalten- und Zeilenräume Gelegentlich ist vom Zeilen- bzw. Spaltenraum einer Matrix die Rede. Diese Begriffe sind wie folgt definiert:

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

Repräsentation der Elemente x_{ij} von X : Die Darstellung $X = UV'$ für eine beliebige Matrix X impliziert eine bestimmte Repräsentation der Elemente x_{ij} von X . Denn $X = UV'$ bedeutet ja

$$x_{ij} = (u_{i1}, u_{i2}, \dots, u_{ir}) \begin{pmatrix} v_{j1} \\ v_{j2} \\ \vdots \\ v_{jr} \end{pmatrix} = \tilde{\mathbf{u}}_i' \tilde{\mathbf{v}}_j, \quad (2.41)$$

d.h. x_{ij} kann stets als Skalarprodukt des i -ten Zeilenvektors (i -ter Fall) von U und des j -ten Spaltenvektors (j -te Variable) von V' dargestellt werden, so dass

$$x_{ij} = \|\tilde{\mathbf{u}}_i\| \|\tilde{\mathbf{v}}_j\| \cos \theta_{ij}, \quad (2.42)$$

gilt. $|x_{ij}|$ wird demnach maximal relativ zu den Längen von $\tilde{\mathbf{x}}_i$ und $\tilde{\mathbf{v}}_j$, wenn $|\cos \theta_{ij}| = 1$ ist. In diesem Fall ist $\theta_{ij} = 0$, d.h. die Vektoren $\tilde{\mathbf{u}}_i$ und $\tilde{\mathbf{v}}_j$ sind parallel. Die Definition von x_{ij} durch ein Skalarprodukt illustriert noch einmal, dass Vektoren $\tilde{\mathbf{u}}_i$ und $\tilde{\mathbf{v}}_j$ Elemente desselben Vektorraums sind, d.h. die Konfiguration der Fälle und die der Variablen werden durch dieselben "latenten" Variablen beschrieben. Aus diesem Sachverhalt ergibt sich die Möglichkeit, Fälle und Variablen simultan in einem (latenten) Koordinatensystem, also in einem *Biplot* (Gabriel (1971)) darzustellen.

Da Skalarprodukte als Ähnlichkeitsmaße interpretiert werden können (vergl. Anmerkung 1, Seite 25), ist x_{ij} ein Maß für die Ähnlichkeit (Kosinus-Ähnlichkeit) der Vektoren $\tilde{\mathbf{u}}_i$ (Ausprägungen der latenten Variablen beim i -ten Fall) und $\tilde{\mathbf{v}}_j$ (Anteile der latenten Variablen an den gemessenen Variablen). \square

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

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

Beweis: Die Inspektion der Matrix $\mathbf{x}\mathbf{y}'$ (vergl. (1.24), (S. 14) zeigt, dass die Spalten der Matrix durch $y_1\mathbf{x}, y_2\mathbf{x}, \dots, y_n\mathbf{x}$ gegeben sind, wobei y_1, \dots, y_n die Komponenten des Vektors \mathbf{y} sind. Die Spaltenvektoren sind also alle parallel zueinander und sind deshalb Elemente eines 1-dimensionalen Teilraums des \mathbb{R}^n , d.h. die zu $\mathbf{x}'\mathbf{y}$ korrespondierende Matrix hat den Rang 1. \square

Definition 2.2 *Es sei A eine (m, n) -Matrix. Die Menge*

$$\mathcal{N}(A) = \{\mathbf{x} | A\mathbf{x} = \mathbf{0}, \mathbf{x} \in \mathbb{R}^n\}$$

heißt Nullraum oder Kern von A . Die Dimensionalität $\dim(\mathcal{N})$ des Nullraums wird auch Nullität (nullity) genannt.

Es sei darauf hingewiesen, dass \mathcal{N} in Bezug auf die *Spalten* von A definiert ist, so dass im Allgemeinen $\mathcal{N}(A) \neq \mathcal{N}(A')$ sein wird, – es sei denn, $A = A'$ ist eine symmetrische Matrix.

Satz 2.5 *Es sei A eine (m, n) -Matrix vom Rang $r \leq \min(m, n)$. Es sei s der Rang von $\mathcal{N}(A)$. Dann gilt*

$$r + s = \min(m, n). \quad (2.43)$$

Beweis: Die Behauptung des Satzes kann auf die Aussage über orthogonale Komplemente (Definition 1.15, Seite 52) und Satz 1.20, Seite 52, zurückgeführt werden. Denn sei V das orthogonale Komplement von $\mathcal{N}(A)$. Gemäß (1.106), Seite 52, gilt dann die Beziehung (2.43). \square

Anmerkung: Es gelte insbesondere $r = \min(m, n)$; dann folgt aus (2.43) $s = 0$, d.h. der Nullraum hat den Rang 0. Der Nullraum enthält dann nur den Nullvektor $\vec{0}$ (s. Anmerkung 3. auf Seite 67). Für $\min(m, n) = n$ heißt dies, dass $A\mathbf{x} = \vec{0}$ nur die Lösung $\mathbf{x} = \vec{0}$ hat, und dies ist der Fall, wenn die Spaltenvektoren linear unabhängig sind. Für den Fall $m = \min(m, n)$ gilt die analoge Argumentation für $\mathbf{x}'A = \vec{0}'$, $\mathbf{x} \in \mathbb{R}^m$. \square

Der Nullraum (Kern) $\mathcal{N}(A)$ einer Matrix A ist ein Vektorraum, denn für \mathbf{x}_1 und \mathbf{x}_2 aus \mathcal{N} folgt

$$A\alpha_1\mathbf{x}_1 + A\alpha_2\mathbf{x}_2 = \alpha_1A\mathbf{x}_1 + \alpha_2A\mathbf{x}_2 = 0 + 0 = 0.$$

Beispiel 2.1 Es sei A eine (m, n) -Matrix mit dem Rang n , d.h. die Spaltenvektoren von A seien linear unabhängig. $A\mathbf{x} = \vec{0}$ ist ein lineares Gleichungssystem mit den Komponenten von \mathbf{x} als Unbekannten. Es gilt $A\mathbf{x} = \vec{0}$ genau dann, wenn $\mathbf{x} = \vec{0}$, und $\mathcal{N}(A) = \{\vec{0}\}$. Die Gleichung $A\mathbf{x} = \mathbf{y}$ hat dann eine (falls $\mathbf{y} \in \mathcal{L}(A)$) oder gar keine Lösung (falls $\mathbf{y} \notin \mathcal{L}(A)$). \square

Beispiel 2.2 Es sei

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \quad A' = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

Der Spaltenrang von A ist offenbar $r_s = 2$, – die beiden ersten Spaltenvektoren von A sind gerade die 2-dimensionalen Einheitsvektoren \mathbf{e}_1 und \mathbf{e}_2 , die wegen ihrer Orthogonalität linear unabhängig sind, und der dritte Spaltenvektor ist gleich der Summe $\mathbf{e}_1 + \mathbf{e}_2$, ist also eine Linearkombination der ersten beiden Spaltenvektoren. Dann ist der Spaltenrang von A' ebenfalls gleich 2, dh. $r = \text{rg}(A) = 2$. Für den Nullraum folgt aus (2.43) mit $n = \min(m, n) = 2$, dass der Nullraum die Dimension $n - 3 = 2 - 1 = 1$ hat, also

$$\dim(\mathcal{N}(A)) = 3 - 2 = 1, \quad \dim(\mathcal{N}(A')) = 2 - 2 = 0.$$

Der Nullraum von A ist also ein eindimensionaler Teilraum des \mathbb{R}^3 :

$$A\mathbf{x} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 + 0 + x_3 \\ 0 + x_2 + x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

woraus

$$x_1 = -x_3, \quad x_2 = -x_3 \Rightarrow x_1 = x_2$$

folgt. Man kann also eine Komponente frei wählen, etwa $x_3 = x$, und findet

$$\mathbf{x} = \begin{pmatrix} -x \\ -x \\ x \end{pmatrix} = x \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}, \quad x \in \mathbb{R}$$

Der Nullraum besteht demnach aus allen Vektoren, die parallel zum Vektor $(-1, -1, 1)$ sind, bzw. die auf der Geraden im \mathbb{R}^3 liegen, die durch diesen Vektor definiert ist. Der Nullraum von A' ist gleich dem Nullvektor $\vec{0}$:

$$A'\mathbf{x} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 + 0 \\ 0 + x_2 \\ x_1 + x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

d.h.

$$x_1 = 0, \quad x_2 = 0 \Rightarrow \mathbf{x} = \vec{0}$$

□

Es sei X eine (m, n) -Matrix. Üblicherweise repräsentieren die m -Zeilen die "Fälle", d.h. Personen oder Objekte, an denen jeweils n Messungen vorgenommen wurden. Die Messungen sind Messungen von n Variablen, die von den Spalten der Matrix repräsentiert werden. X_{ij} ist die Messung der j -ten Variablen beim i -ten Fall. Die Matrix X heißt *spaltenzentriert*, wenn von den Messwerten X_{ij} der j -ten Variablen (den Messwerten in der j -ten Spalte von X) der Mittelwert \bar{x}_j subtrahiert wurde: $x_{ij} = X_{ij} - \bar{x}_j$. Benennt man X um in die Matrix der x_{ij} -Werte, so heißt X *spaltenzentriert*. Werden die Messwerte X_{ij} *spaltenstandardisiert*, so gehen die X_{ij} über in $z_{ij} = (X_{ij} - \bar{x}_j)/s_j$, s_j die Streuung der Messwerte in der j -ten Spalte der Datenmatrix.

Es sei X eine (m, n) -Matrix; dann heißen die Produkte $X'X$ und XX' *Kreuzproduktmatrizen*. Repräsentieren die Spalten von X Variablen und ist X spaltenzentriert, so ist

$$C = \frac{1}{m} X'X \tag{2.44}$$

eine Kovarianzmatrix³⁶. Ist X spaltenstandardisiert, d.h. gilt $X = Z$, so ist

$$R = \frac{1}{m} Z'Z, \tag{2.45}$$

R ist die Matrix der Korrelationen zwischen den Variablen

Die Betrachtungen für eine zeilenzentrierte bzw. zeilenstandardisierte Matrix sind analog. Formal, d.h. ohne Bezug auf die Bedeutung der Variablen, ist eine derartige Standardisierung immer möglich. Man muss allerdings bedenken, dass eine Mittelung über verschiedene Variablen analog zu einer Mittelung von Äpfeln und Birnen ist.

³⁶Üblicherweise wird durch $m - 1$ geteilt, um den Bias bei der Schätzung von Varianzen und Kovarianzen auszugleichen. Hier wird durch m geteilt, um den Charakter der Varianzen und Kovarianzen als Mittelwerte zu verdeutlichen.

Satz 2.6 *Es sei X eine (m, n) -Matrix mit dem Rang $\text{rg}(X) = r \leq \min(m, n)$. Dann gilt für den Rang der Kreuzproduktmatrizen*

$$\text{rg}(X'X) = \text{rg}(XX') = \text{rg}(X). \quad (2.46)$$

Beweis:³⁷ (i) $\text{rg}(X) = \text{rg}(X'X)$. Es sei $\mathbf{u} \in \mathbb{R}^n$ und es gelte $X\mathbf{u} = \vec{0}$, so dass $\mathbf{u} \in \mathcal{N}(X)$, $\mathcal{N}(X)$ der Nullraum von X . Dann gilt auch $X'X\mathbf{u} = \vec{0}$, so dass ebenfalls $\mathbf{u} \in \mathcal{N}(X'X)$. Sei nun $\mathbf{v} \in \mathcal{N}(X'X)$, d.h. $X'X\mathbf{v} = \vec{0}$. Dann folgt $\mathbf{v}'X'X\mathbf{v} = \|X\mathbf{v}\|^2 = 0$, woraus $X\mathbf{v} = \vec{0}$ und damit $\mathbf{v} \in \mathcal{N}(X)$ folgt. Also haben X und $X'X$ denselben Nullraum und damit, nach Satz 2.5, denselben Rang.

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

Der Rang einer Diagonalmatrix läßt sich direkt an den Diagonalelementen ablesen:

Satz 2.7 *Es sei $A = \text{diag}(d_1, \dots, d_n)$ eine Diagonalmatrix, d.h. für die Elemente a_{ij} von A gelte $a_{ij} = 0$ für $i \neq j$ und $a_{ii} = d_i$, $i = 1, \dots, n$. Dann ist $\text{rg}(A) = r \leq n$, r die Anzahl der von Null verschiedenen Diagonalelemente.*

Beweis: Es sei $r = n$, d.h. alle Diagonalelemente sind ungleich Null. Die Spaltenvektoren von A können in der Form $\mathbf{a}_j = d_j \mathbf{e}_j$ geschrieben werden, wobei \mathbf{e}_j der j -te Einheitsvektor ist, d.h. die Komponenten von \mathbf{e}_j sind alle gleich Null bis auf die j -te, die gleich d_j ist. Da die n Einheitsvektoren $\mathbf{e}_1, \dots, \mathbf{e}_n$ linear unabhängig sind, ist $r = \text{rg}(A) = n$.

Nun sei $r < n$. Die ersten r Vektoren \mathbf{a}_j haben die Form $\mathbf{a}_j = \lambda_j \mathbf{e}_j$ und sind daher linear unabhängig, also kann der Rang von A nicht kleiner als r sein. A enthält aber $n - r$ Nullvektoren. Erweitert man die r Spaltenvektoren $\neq \vec{0}$ auch nur um einen Nullvektor, so sind nach Satz 1.6, Seite 28, diese $r + 1$ Vektoren linear abhängig, so dass der Rang von A nicht größer als r sein kann. Damit ist gezeigt, dass der Rang von A gleich der Anzahl der von Null verschiedenen Diagonalelemente ist. □

Anmerkung: Die Aussage des Satzes folgt leicht, wenn der Begriff der Determinante einer Matrix vorausgesetzt werden kann, siehe die Folgerungen 6 bis 8, Seite 189. □

2.4 Die Inverse einer Matrix

In der Arithmetik kennt man die vier Grundrechenarten Addition, Subtraktion, Multiplikation und Division. Die Division $\frac{a}{b}$ einer reellen Zahl $a \in \mathbb{R}$ durch eine reelle Zahl $b \neq 0$ läßt sich als Multiplikation von a mit der zur Zahl b inversen Zahl $b^{-1} = 1/b$ schreiben: $\frac{a}{b} = ab^{-1}$. Für die inverse Zahl gilt $b^{-1}b = bb^{-1} = 1$.

Für zwei Matrizen A und B sind ebenfalls die Operationen Addition, Subtraktion und der Multiplikation erklärt, sofern die beiden Matrizen bestimmte

³⁷Seber, p. 385

Bedingungen erfüllen: für die Addition und die Subtraktion müssen A und B dieselbe Anzahl von Zeilen und Spalten haben, und um das Produkt AB bilden zu können, muss die Anzahl der Spalten von A gleich der Anzahl der Zeilen von B sein. Die Analogie zur Arithmetik wäre komplett, wenn man für Matrizen auch eine Division definieren könnte. Sie wäre wie in der Arithmetik als Multiplikation mit einer zu einer gegebenen Matrix A inversen Matrix A^{-1} erklärt, so dass $A^{-1}A = AA^{-1} = I$, I die Einheitsmatrix, die zur Zahl 1 in der Arithmetik korrespondiert. Für das Lösen von Matrixgleichungen erweist sich diese "Division" als sehr nützlich. Es zeigt sich, dass eine Reihe von Bedingungen erfüllt sein müssen, damit eine Matrix A^{-1} existiert. Diese Bedingungen werden zuerst elaboriert.

Gegeben sei die (m, n) -Matrix A . A definiert eine Abbildung $A: \mathbb{R}^m \rightarrow \mathbb{R}^n$, d.h. sie bildet gemäß der Gleichung $A\mathbf{x} = \mathbf{y}$ n -dimensionale Vektoren \mathbf{x} auf m -dimensionale Vektoren \mathbf{y} ab; diese Beziehung gilt also für alle $\mathbf{x} \in \mathbb{R}^m$ und damit für alle $\mathbf{y} \in \mathbb{R}^n$. Die Abbildung soll umkehrbar eindeutig sein, d.h. es soll eine Matrix B existieren derart, dass $B\mathbf{y} = \mathbf{x}$. Setzt man diesen Ausdruck für \mathbf{x} in die Gleichung $A\mathbf{x} = \mathbf{y}$ ein, so erhält man

$$A\mathbf{x} = AB\mathbf{y} = \mathbf{y}.$$

Da $\mathbf{y} \in \mathbb{R}^n$ folgt, dass B m Spalten haben muss, und da $\mathbf{x} \in \mathbb{R}^m$ folgt, dass B n Zeilen haben muss, d.h. B muss eine (n, m) -Matrix sein, und die Beziehung $AB\mathbf{y} = \mathbf{y}$ impliziert³⁸, dass das Matrixprodukt $AB = I_m$ die (m, m) -Einheitsmatrix ist. Andererseits kann man die Existenz einer Matrix C annehmen, die von der Beziehung $A\mathbf{x} = \mathbf{y}$ ausgeht:

$$C A \mathbf{x} = \mathbf{x} = C \mathbf{y}.$$

Damit das Matrixprodukt CA gebildet werden kann, muss C m Spalten und n Zeilen haben. Man kann davon ausgehen, dass $CA = I_n$ die (n, n) -Einheitsmatrix ist. Wegen $AB = I_m$ heißt B die *Rechtsinverse* von A , und wegen $CA = I_n$ heißt C die *Linksinverse* von A . Fordert man weiter, dass

$$AB = CA = I, \text{ d.h. } I_m = I_n \tag{2.47}$$

gelten soll, so folgt, dass $m = n$ und demnach $I = I_n$ eine (n, n) -Einheitsmatrix ist. Multipliziert man nun (5.80) von links mit C , so erhält man

$$\underbrace{CA}_{I_n} B = CI \Rightarrow B = C, \tag{2.48}$$

d.h. die Linksinverse ist gleich der Rechtsinversen und man hat die Definition

³⁸Aus $AB\mathbf{y} = \mathbf{y}$ folgt noch nicht, dass $AB = I_m$ ist. Es folgt zunächst, dass $AB\mathbf{y} - \mathbf{y} = (AB - I_m)\mathbf{y} = \vec{0}$ ist, und diese Gleichung repräsentiert ein homogenes Gleichungssystem $D\mathbf{y} = \vec{0}$ mit $D = AB - I_m$, für das möglicherweise nur ein Lösungsvektor \mathbf{y} existiert. Aber die Abbildung $A\mathbf{x} = \mathbf{y}$ soll für alle $\mathbf{x} \in \mathbb{R}^m$ und damit für die zum jeweiligen \mathbf{x} korrespondierenden Vektoren $\mathbf{y} \in \mathbb{R}^n$ gelten. Man kann demnach die Derivierte $\frac{dD}{d\mathbf{y}} = \frac{dI}{d\mathbf{y}} = 0$ betrachten (s. Gleichung (5.42), Seite 172), derzufolge $\frac{dD}{d\mathbf{y}} = D = AB - I = 0$, also $D = 0$ gilt, woraus $AB = I$ folgt.

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

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

$$A^{-1}A = AA^{-1} = I_n, \quad (2.49)$$

I_n die (n, n) -Einheitsmatrix. \square

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

Beispiel 2.3 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

$$a\alpha + b\gamma = 1 \Rightarrow \alpha = \frac{1 - b\gamma}{a} \quad (2.50)$$

$$a\beta + b\delta = 0 \Rightarrow \beta = -\frac{b\delta}{a} \quad (2.51)$$

$$c\alpha + d\gamma = 0 \Rightarrow \gamma = -\frac{c\alpha}{d} \quad (2.52)$$

$$c\beta + d\delta = 1 \Rightarrow \delta = \frac{1 - c\beta}{d} \quad (2.53)$$

Durch Einsetzen etwa des in Gleichung (2.52) gegebenen Ausdrucks für γ in die Gleichung (2.50) etc findet man schließlich

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \quad (2.54)$$

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

Es werde nun angenommen, dass $ad - bc = 0$ ist, d.h. es gelte $ad = bc$. Dann ist der Ausdruck $1/(ad - bc)$ in (2.54) nicht definiert, d.h. A^{-1} existiert nicht. Man hat dann einerseits

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

andererseits folgt auch

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

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

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

die Spaltenvektoren von A sind linear abhängig. Man rechnet leicht nach, dass umgekehrt die lineare Abhängigkeit der Spaltenvektoren die Gleichung $ad - bc = 0$ impliziert. Die Voraussetzung $ad - bc \neq 0$ gilt also genau dann, wenn die Spalten- und damit auch die Zeilenvektoren von A linear unabhängig sind. Es sei angemerkt, dass $ad - bc = |A|$ die Determinante der Matrix A ist (vergl. (1.118), S. 57); allgemein gilt, dass die Inverse einer quadratischen Matrix nur dann existiert, wenn die Determinante der Matrix ungleich Null ist, und dies ist der Fall, wenn A vollen Rang hat. \square

Beispiel 2.4 Es sei R eine (2×2) -Korrelationsmatrix,

$$R = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}. \quad (2.55)$$

Die Inverse R^{-1} ergibt sich aus (2.54)

$$R^{-1} = \begin{pmatrix} \frac{1}{1-r^2} & -\frac{r}{1-r^2} \\ -\frac{r}{1-r^2} & \frac{1}{1-r^2} \end{pmatrix} = \frac{1}{1-r^2} \begin{pmatrix} 1 & -r \\ -r & 1 \end{pmatrix} \quad (2.56)$$

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

$$\lim_{r \rightarrow 0} R^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \lim_{r \rightarrow 1} R^{-1} = \begin{pmatrix} \infty & -\infty \\ -\infty & \infty \end{pmatrix} \quad (2.57)$$

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

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

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

Beispiel 2.5 Es werde der Fall einer Matrix A mit den Zeilen $(1, 3)$ und $(2, 1)$ betrachtet; gesucht ist die zu A inverse Matrix A^{-1} :

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

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

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

Dann ist

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

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

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

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

$$(AB)^{-1} = B^{-1}A^{-1}. \quad (2.58)$$

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

2.5 Die Beziehungen zwischen verschiedenen Basen

Es sei $S \subseteq \mathbb{R}^n$ ein Teilraum des n -dimensionalen Vektorraums, und es seien $\mathcal{B} = (\mathbf{b}_1, \dots, \mathbf{b}_r)$ und $\mathcal{C} = (\mathbf{c}_1, \dots, \mathbf{c}_r)$, $r \leq n$, zwei Mengen von linear unabhängigen Vektoren aus S , und es gelte $S = \mathcal{L}(\mathcal{B}) = \mathcal{L}(\mathcal{C})$, d.h. \mathcal{B} und \mathcal{C} Basen von S ; für $r = n$ sind es Basen des \mathbb{R}^n . Natürlich gilt

$$\mathcal{C} \subset \mathcal{L}(\mathcal{B}), \quad \mathcal{B} \subset \mathcal{L}(\mathcal{C})$$

Weiter sei $B = [\mathbf{b}_1, \dots, \mathbf{b}_r]$ die (n, r) -Matrix mit den Spaltenvektoren \mathbf{b}_j , und $C = [\mathbf{c}_1, \dots, \mathbf{c}_r]$ die (n, r) -Matrix mit den Spaltenvektoren \mathbf{c}_j , $j = 1, \dots, r$. Dann existieren sicherlich r r -dimensionale Vektoren $\mathbf{t}_1, \dots, \mathbf{t}_r$ derart, dass

$$B\mathbf{t}_k = \mathbf{c}_k, \quad k = 1, \dots, r$$

oder, mit $T = [\mathbf{t}_1, \dots, \mathbf{t}_r]$, T eine (r, r) -Matrix,

$$BT = C, \quad (2.59)$$

und da $\text{rg}(B) = \text{rg}(C)$ und $\text{rg}(C) = \text{rg}(BT) \leq \min[\text{rg}(B), \text{rg}(T)]$, folgt $\text{rg}(T) = r$. Dann existiert die Inverse T^{-1} und

$$C = BT \tag{2.60}$$

und da $CT^{-1} = BTT^{-1}$ folgt

$$B = CT^{-1} \tag{2.61}$$

$$C = BT \tag{2.62}$$

$$B = CT^{-1} \tag{2.63}$$

Insbesondere sei \mathcal{B} eine Orthogonalbasis, d.h. die Spaltenvektoren von B seien paarweise orthogonal, während \mathcal{C} keine Orthogonalbasis ist. Man kann also von einer Orthogonalbasis zu einer beliebigen anderen Basis übergehen und umgekehrt kann man von einer beliebigen, nicht-orthogonalen Basis zu einer Orthogonalbasis.
gggg

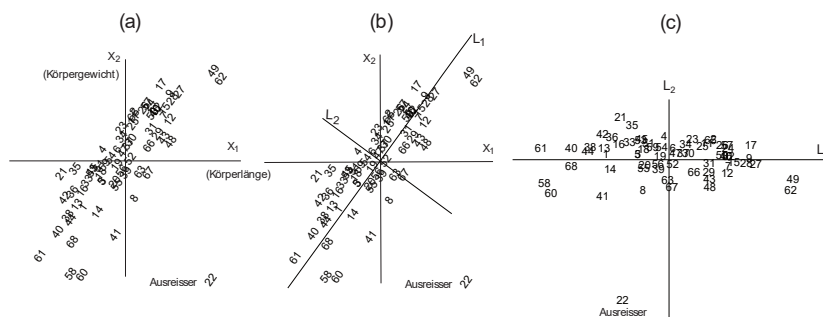
3 Bestimmung einer Basis: die Hauptachsentransformation

3.1 Rotationen

Es sei X eine (m, n) -Datenmatrix; die Komponenten der n m -dimensionalen Spaltenvektoren $\mathbf{x}_1, \dots, \mathbf{x}_n$ sind Messungen von im allgemeinen korrelierenden Variablen. Gesucht ist eine der Matrizen U oder V mit jeweils linear unabhängigen Spaltenvektoren, die X gemäß $X = UV'$ bzw. $X' = VU'$ repräsentieren.

Es existieren beliebig viele Matrizen U und dazu gehörige Matrizen V , die dieser Beziehung genügen. Man muß also Kriterien spezifizieren, nach denen U bzw. V gewählt werden. Inhaltliche Kriterien, die den Spaltenvektoren \mathbf{u}_k oder \mathbf{v}_k bestimmte Bedeutungen zuordnen, liegen oft nicht vor, so dass man ein eher formales Kriterium wählt. Dem üblicherweise gewählten Kriterium zufolge wählt man eine Achse L_1 derart, dass die Projektionen b_{i1} , $i = 1, \dots, m$, der Endpunkte der $\tilde{\mathbf{x}}_i$, also der Spaltenvektoren von X' auf L_1 eine maximale Varianz haben, dann wählt man eine zweite Achse L_2 , die orthogonal zu L_1 ist, und für die Projektionen b_{i2} der Endpunkte eine zweitmaximale Varianz haben, etc. L_1 repräsentiert dann ein Merkmal, hinsichtlich dessen sich die Fälle maximal unterscheiden, L_2 repräsentiert ein Merkmal, hinsichtlich dessen sie sich zweitmaximal unterscheiden, etc. Normalerweise bestimmt man n , solche Achsen, – ob alle Achsen L_1, \dots, L_n oder nur die ersten $r < n$ für die Interpretation der Daten benötigt werden, ist dann eine Frage, die nach der Bestimmung der L_k und der zugehörigen Vektoren \mathbf{u}_k mit den Komponenten b_{ik} gestellt wird. Die Transformation der $\tilde{\mathbf{x}}_i$, die für die Bestimmung der L_k vorgenommen werden muß, ist eine Rotation, Abbildung 12 illustriert diesen Prozess, wobei zentrierte oder standardisierte Messwerte vorausgesetzt wurden, so dass der Ursprung sowohl des alten wie des

Abbildung 12: Rotation einer 2-dimensionalen Punktekonfiguration von Fällen



neuen Koordinatensystems im Zentrum der Punktekonfiguration der Fälle liegt. Aus Gründen, die im Folgenden elaboriert werden, spricht man auch von einer *Hauptachsentransformation*. Wegen $X' = VU'$ muß die Matrix V so bestimmt werden, dass sie eine Rotation der $\tilde{\mathbf{u}}_i$ in die $\tilde{\mathbf{x}}_i$ bewirkt. Die inverse Transformation erlaubt dann die Rotation der $\tilde{\mathbf{x}}_i$ in die \mathbf{u}_i . Zunächst werden Eigenschaften dieser *Rotationsmatrix* hergeleitet.

Es sei $\mathbf{x} \in \mathbb{R}^n$ ein beliebiger n -dimensionaler Vektor und T sei eine (n, n) -Matrix derart, dass $\mathbf{y} = T\mathbf{x} \in \mathbb{R}^n$; man kann die Matrix T als eine Abbildung interpretieren, die jedem Vektor $\mathbf{x} \in \mathbb{R}^n$ einen Vektor $\mathbf{y} \in \mathbb{R}^n$ zuordnet (vergl. Abschnitt 5.9, Seite 182); man schreibt auch $T: \mathbf{x} \rightarrow \mathbf{y}$, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. T lasse die Länge von \mathbf{x} invariant, so dass sich \mathbf{y} und \mathbf{x} nur durch ihre Orientierungen voneinander unterscheiden. Dementsprechend soll

$$\|\mathbf{y}\|^2 = \mathbf{y}'\mathbf{y} = \mathbf{x}'T'T\mathbf{x} = \mathbf{x}'\mathbf{x} = \|\mathbf{x}\|^2 \quad (3.1)$$

gelten.

Satz 3.1 *Es seien $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ und T sei eine (n, n) -Matrix. Die Beziehung (3.1) gilt genau dann, wenn T orthonormal ist, d.h. wenn*

$$T'T = TT' = I \quad (3.2)$$

gilt.

Beweis: (i) $T'T = I$ ist hinreichend, denn

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

(ii) $T'T = I$ ist notwendig: Um diese Behauptung einzusehen, werde $A = T'T$ gesetzt, wobei zunächst noch offen sei, ob $A = I$ gilt. Nach (3.1) soll $\mathbf{x}'A\mathbf{x} = \mathbf{x}'\mathbf{x}$ für alle $\mathbf{x} \in \mathbb{R}^n$ gelten, d.h. \mathbf{y} kann als Funktion des Vektors \mathbf{x} aufgefasst werden³⁹.

³⁹Deswegen der Hinweis auf die Interpretation von T als einer Abbildung von Vektoren $\mathbf{x} \in \mathbb{R}^n$ auf Vektoren $\mathbf{y} \in \mathbb{R}^n$

Dann liefert (5.45) aus dem Anhang, Abschnitt 5.6.2, Seite 174

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

und die nochmalige Ableitung liefert $A = I$ (s. Gleichung (5.42), Seite 172), d.h. $T'T = I$. Dies bedeutet, dass die Spaltenvektoren von T orthonormal sind.

(iii) $T'T = I \Rightarrow TT' = I$: Es gilt

$$\begin{aligned} T'T &= I \\ \Rightarrow T'TT' &= IT' = T'I \\ \Rightarrow T'TT' - T'I &= 0 \\ \Rightarrow T' \underbrace{(TT' - I)}_{B=[\mathbf{b}_1, \dots, \mathbf{b}_n]} &= 0 \end{aligned}$$

wobei 0 die (n, n) -Null-Matrix ist, deren Elemente alle gleich 0 sind. Da der Zeilenrang einer Matrix gleich ihrem Spaltenrang ist sind die Spaltenvektoren von T' linear unabhängig, so dass $T'\mathbf{b}_j = \vec{0}_j \Rightarrow \mathbf{b}_j = \vec{0}_j$ für $j = 1, \dots, n$, also folgt $B = TT' - I = 0$, d.h. $TT' = I$.

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

□

Anmerkungen:

1. **Alternativer Beweis** Einen alternativen Beweis für die Aussage $TT' = I$ findet man in Beispiel 5.1 auf Seite 173 im Anhang. Ein weiterer Beweis setzt den Begriff der Determinante voraus, s. Beispiel 5.6, Seite 191 im Anhang.

2. **Bedeutung von TT'** Die Beziehung $T'T = TT' = I$ für orthonormale Matrizen bedeutet, dass im Falle von orthonormalen Matrizen T die jeweilige Inverse durch

$$T^{-1} = T' \tag{3.3}$$

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

3. **Die Elemente der Rotationsmatrix** sind Richtungskosinus. Nach (1.55), Seite 22, gilt

$$\mathbf{x} = \|\mathbf{x}\|(\cos \theta_1, \dots, \cos \theta_n)'$$

und da die Spaltenvektoren \mathbf{t}_j einer Rotationsmatrix T die Länge $\|\mathbf{t}_j\| = 1$ haben, muss

$$\mathbf{t}_j = (\cos \theta_{1j}, \dots, \cos \theta_{nj})' \tag{3.4}$$

gelten. Im 2-dimensionalen Fall hat man dementsprechend

$$T = \begin{pmatrix} \cos \theta_{11} & \cos \theta_{12} \\ \cos \theta_{21} & \cos \theta_{22} \end{pmatrix} \tag{3.5}$$

Wie aus Abbildung 1, Seite 8 hervorgeht ist aber $\cos \theta_2 = \sin \theta_1$ und (3.5) kann in der Form (3.8) (s. unten) geschrieben werden.

4. Invarianz der Skalarprodukte Es sei X eine (m, n) -Datenmatrix. Aus $X = UV'$ folgt $X' = VU'$, d.h. die Spaltenvektoren $\tilde{\mathbf{x}}_i$, $i = 1, \dots, m$, die die Fälle repräsentieren, werden als Linearkombination der Spaltenvektoren von V dargestellt. Ist V eine orthonormale Matrix, so geht $\tilde{\mathbf{x}}_i$ durch Rotation aus $\tilde{\mathbf{u}}_i$ hervor, und umgekehrt impliziert $X' = VU'$ die Beziehung $V'X' = U'$, dass die $\tilde{\mathbf{u}}_i$ Linearkombinationen der Spaltenvektoren $\tilde{\mathbf{x}}_i$ sind, insbesondere gehen die $\tilde{\mathbf{u}}_i$ durch Rotation aus den $\tilde{\mathbf{x}}_i$ hervor, wobei jetzt V' die Rotationsmatrix ist. Die Rotation der $\tilde{\mathbf{x}}_i$ bzw. der $\tilde{\mathbf{u}}_i$ läßt die Skalarprodukte zwischen den $\tilde{\mathbf{x}}_i$ einerseits bzw. der $\tilde{\mathbf{u}}_i$ andererseits invariant:

$$\tilde{\mathbf{x}}_i' \tilde{\mathbf{x}}_k = \tilde{\mathbf{u}}_i' V' V \tilde{\mathbf{u}}_k = \tilde{\mathbf{u}}_i' \tilde{\mathbf{u}}_k, \quad i, k = 1, \dots, m \quad (3.6)$$

und da Rotationen von Vektoren ihre Länge invariant lassen ($i = k$) folgt, dass eine Rotation die interne Struktur der Punktekonfiguration invariant läßt. Dies gilt für Rohdaten, zentrierte oder standardisierte Werte, wobei sich natürlich die jeweiligen Matrizen U und V voneinander unterscheiden. \square

3.2 Die Rotation einer Punktekonfiguration

3.2.1 Annahmen und Implikationen: die Eigenstruktur von $X'X$

Die Punktekonfiguration der Fälle wird zunächst in einem durch die gemessenen Variablen definierten Koordinatensystem repräsentiert. Latente Variablen definieren ein alternatives Koordinatensystem für die Konfiguration der Punkte, deren interne Struktur – d.h. die Distanzen zwischen den Punkten – üblicherweise unverändert bleibt. Der Übergang zum neuen ("latenten") Koordinatensystem ist dann eine Rotation:

1. Die zu jedem Punkt korrespondierenden Vektoren \mathbf{x}_i werden alle um ein und denselben Winkel θ rotiert und gehen in Vektoren \mathbf{y}_i , $i = 1, \dots, m$ über. Die Vektoren $\mathbf{y}_1, \dots, \mathbf{y}_m$ repräsentieren dann die rotierte Konfiguration. Die alten Koordinatenachsen X_1, \dots, X_n werden dann in L_1, \dots, L_n umbenannt.
2. Die ursprünglichen Achsen werden um einen Winkel $-\theta$ rotiert; die Koordinaten der Punkte auf den neuen Achsen müssen bestimmt werden.⁴⁰

Die beiden Ansätze sind äquivalent. Sind $\mathbf{x}_i = \tilde{\mathbf{x}}_i$ die Zeilenvektoren von X , repräsentieren sie also die Fälle, so bilden die $\mathbf{y}_i = \tilde{\mathbf{y}}_i = T\tilde{\mathbf{x}}_i$ die rotierte Konfiguration der Fälle. Die inverse Transformation ist

$$\tilde{\mathbf{x}}_i = T' \tilde{\mathbf{y}}_i, \quad i = 1, \dots, m \quad (3.7)$$

⁴⁰Die Rotation der Vektoren heißt auch *Alibi-Transformation*, die der Koordinatenachsen *Alias-Transformation*

wobei für den 2-dimensionalen Fall

$$T = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}, \quad (3.8)$$

gilt (eine andere Herleitung als die oben angedeutete wird im Anhang, Seite 166 gegeben.).

Gegeben sei also eine (m, n) -Matrix X vom Rang $\text{rg}(X) = r \leq \min(m, n)$, z.B. eine Datenmatrix, bei der die m Zeilen die untersuchten Fälle und die n Spalten die bei jedem Fall gemessenen Variablen repräsentieren. Nach Satz 2.2, Seite 66 existieren stets eine (m, r) -Matrix U und eine (n, r) -Matrix V derart, dass $X = UV'$. Die Spaltenvektoren von X sind Linearkombinationen der Spaltenvektoren von U , und da sie linear unabhängig sind, wird nach Satz 1.10, Seite 40 mit einer Wahl von U auch die Matrix V eindeutig bestimmt. $X = UV'$ impliziert $X' = VU'$, die Spaltenvektoren $\tilde{\mathbf{x}}_i$ sind demnach Linearkombinationen der Spaltenvektoren von V , d.h. die Spaltenvektoren $\tilde{\mathbf{u}}_i$ von U' werden in die Vektoren $\tilde{\mathbf{x}}_i$ transformiert. Es werde nun angenommen, dass diese Transformation eine Rotation ist. Dann muß V eine orthonormale Matrix sein. Es muß allerdings noch der Rotationswinkel bestimmt werden. Diese Wahl kann möglicherweise auf der Basis von Annahmen über die gemessenen Variablen oder die Fälle bestimmt werden, was aber eher die Ausnahme sein wird. Eine beliebte Alternative zu diesem Vorgehen ist, zu fordern, dass die Spaltenvektoren von U orthogonal sein sollen, so dass $U'U = \Lambda$ eine Diagonalmatrix ist. Um sich leicht auf die Annahmen beziehen zu können, werden sie wie folgt zusammengefasst:

Annahmen:

$$\text{A1: } V'V = I, \text{ Rotation} \quad (3.9)$$

$$\text{A2: } U'U = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n), \quad \lambda_k = \mathbf{u}'_k \mathbf{u}_k = \|\mathbf{u}_k\|^2, \quad (3.10)$$

Es wird im Folgenden gezeigt werden, dass (i) der Punktekonfiguration der Fälle eine Menge von Ellipsoiden mit identischer Orientierung entspricht, und dass (ii) die Matrix V eine Rotation des ursprünglichen, durch die \mathbf{x}_j gegebenen Koordinatensystems in das durch die Hauptachsen dieser Ellipsoide definiertes Koordinatensystem bedeutet. Man sagt dementsprechend, dass V eine *Hauptachsen-Transformation* (HAT) sei. Die gängige Abkürzung ist allerdings PCA, für englisch *Principal Component Analysis*. Abbildung 12 illustriert die Rotation.

Der folgende Satz liefert eine Charakterisierung der Matrix V , die sie berechenbar macht, weshalb dann wegen $XV = U$ auch U bestimmt werden kann. Es gilt also der

Satz 3.2 *Es sei $X = UV'$ eine beliebige (m, n) -Matrix, so dass $C = X'X$ eine (n, n) -Matrix ist; U ist eine (m, r) -Matrix, wenn r der Rang von X ist, und V ist eine (n, r) -Matrix. Die Annahmen A1 und A2 implizieren*

$$CV = V\Lambda, \quad \text{d.h. } C\mathbf{v}_k = \lambda_k \mathbf{v}_k, \quad k = 1, \dots, n \quad (3.11)$$

Beweis: Nach A1 gilt $V'V = VV' = I$, I die (n, n) -Einheitsmatrix. Aus $X = UV'$ folgt dann $XV = U$ und A2 impliziert $V'X'XV = U'U = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, so dass $VV'X'XV = X'XV = V\Lambda$. Nach Multiplikation von links mit V erhält man $X'XV = C\Lambda$. \square

Wählt man einen beliebigen Vektor $\mathbf{x} \in \mathbb{R}^n$ und bildet die Transformation $C\mathbf{x} = \mathbf{y} \in \mathbb{R}^n$, so wird sich \mathbf{y} typischerweise nicht nur in der Länge, sondern auch bezüglich der Orientierung von \mathbf{x} unterscheiden. Die Spaltenvektoren \mathbf{v}_k von V haben aber nach Satz 3.2 die Eigenschaft, bei einer Transformation durch die Matrix C ihre Orientierung beizubehalten. Dies bedeutet, dass die \mathbf{v}_k für die Matrix C charakteristisch sind. Dies rechtfertigt einen speziellen Namen für die \mathbf{v}_k :

Definition 3.1 *Es seien M eine (n, n) -Matrix und $\mathbf{v} \in \mathbb{R}^n$, und es gelte (3.11), d.h. es gelte für einen k -ten Spaltenvektor \mathbf{v}_k von V*

$$M\mathbf{v}_k = \lambda_k\mathbf{v}_k, \quad \lambda_k \in \mathbb{R}, \quad \mathbf{v}_k \neq \vec{0}, \quad k = 1, \dots, n \quad (3.12)$$

Dann heißt \mathbf{v}_k Eigenvektor oder charakteristischer Vektor von M und λ_k ist der zu \mathbf{v} gehörige Eigenwert oder charakteristische Wert⁴¹. Das Paar (V, Λ) wird auch als Eigenstruktur der Matrix M bezeichnet, wobei $V = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n]$ die Matrix der Eigenvektoren und $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ die Diagonalmatrix der zugehörigen Eigenwerte von M sind.

Aus (3.12) folgen die oft verwendeten Beziehungen

$$M = V\Lambda V' \quad (3.13)$$

$$\Lambda = V'MV \quad (3.14)$$

Die Beziehung (3.14) wird oft als *Diagonalisierung* von M bezeichnet.

Anmerkungen:

1. Symmetrie von M In der Gleichung (3.11) ist die Matrix C symmetrisch, aber in der Definition 3.1 der Begriffe Eigenvektor und Eigenwert wird die Symmetrie von M nicht gefordert, denn es zeigt sich, dass Eigenvektoren auch für nicht-symmetrische Matrizen existieren. Auf diesen allgemeinen Fall wird in Abschnitt 3.3.2 zurückgekommen.

2. Hat man die Matrix V bestimmt, wo lässt sich wegen $XV = U$ auch die Matrix U berechnen, womit eine (von mehreren möglichen) Lösungen des Problems der Bestimmung latenter Variablen im Prinzip gefunden wurde, – auf weitere Details wird in einem späteren Abschnitt (Abschnitt 3.6) näher eingegangen. Allerdings kann V nur mit numerischen Mitteln bestimmt werden kann, auf die hier nicht eingegangen werden kann⁴². \square

⁴¹Man findet auch die Ausdrücke *latenter Vektor* und *latenter Vektor* für den Eigenvektor und Eigenwert.

⁴²Das Prinzip der Berechnung von Eigenvektoren und zugehörigen Eigenwerten ist nicht schwierig, die aktuelle Berechnung "per Hand" ist allerdings langwierig, und insbesondere für größere Matrizen müssen die Verfahren zur Berechnung optimiert werden, vgl. Golub & Van Loan (2023). Die Statistikprogramme enthalten diese speziellen Subroutinen.

3.2.2 Eigenschaften der Eigenstruktur symmetrischer Matrizen

Orthogonalität der Eigenvektoren In Gleichung (3.11) hat V zunächst die Bedeutung einer Rotationsmatrix, d.h. V ist orthonormal. Der Schluß von $V'X'XV = \Lambda$ auf die Gleichung $X'XV = V\Lambda$, die die Eigenvektoreigenschaft der Spaltenvektoren von V zeigt, legt nahe, dass die Eigenvektoren von $C = X'X$ stets orthonormal sind. Tatsächlich läßt sich zeigen, dass unter bestimmten Bedingungen die Symmetrie einer Matrix M bereits hinreichend für die Orthogonalität der Eigenvektoren von M ist:

Satz 3.3 *Es sei M eine symmetrische (n, n) -Matrix und V eine (n, n) -Matrix der Eigenvektoren von M . Eigenvektoren, die zu verschiedenen großen Eigenwerten ungleich Null korrespondieren, sind orthogonal.*

Beweis: Es gelte $M\mathbf{v}_j = \lambda_j\mathbf{v}_j$, $M\mathbf{v}_k = \lambda_k\mathbf{v}_k$, $j \neq k$, und es sei $\lambda_j \neq \lambda_k$. Die letztere Gleichung werde von links mit \mathbf{v}'_j multipliziert: $\mathbf{v}'_j M\mathbf{v}_k = \lambda_k\mathbf{v}'_j\mathbf{v}_k$, und die erstere werde von links mit \mathbf{v}'_k multipliziert: $\mathbf{v}'_k M\mathbf{v}_j = \lambda_j\mathbf{v}'_k\mathbf{v}_j$, woraus wegen der Symmetrie von M

$$(\mathbf{v}'_k M\mathbf{v}_j)' = \mathbf{v}'_j M\mathbf{v}_k = (\lambda_j\mathbf{v}'_k\mathbf{v}_j)' = \lambda_j\mathbf{v}'_j\mathbf{v}_k \quad (3.15)$$

folgt, so dass man die Gleichungen

$$\mathbf{v}'_j M\mathbf{v}_k = \lambda_k\mathbf{v}'_j\mathbf{v}_k \quad (3.16)$$

$$\mathbf{v}'_j M\mathbf{v}_k = \lambda_j\mathbf{v}'_j\mathbf{v}_k \quad (3.17)$$

$$0 = (\lambda_k - \lambda_j)\mathbf{v}'_j\mathbf{v}_k \quad (3.18)$$

erhält, wobei sich (3.18) aus (3.16) durch Subtraktion von (3.17) ergibt. Wegen $\lambda_k - \lambda_j \neq 0$ folgt $\mathbf{v}'_j\mathbf{v}_k = 0$ für $j \neq k$, d.h. verschiedene Eigenvektoren sind orthogonal. \square

Die Eigenvektoren nicht-symmetrischer Matrizen können, müssen aber nicht orthogonal sein. Da Orthogonalität lineare Unabhängigkeit impliziert bilden $r \leq n$ Eigenvektoren einer symmetrischen Matrix eine Teilbasis bzw. im Fall von $r = n$ eine komplette Basis des \mathbb{R}^n .

Quadratische Formen und Ellipsoide Die Bedeutung der Matrix U ist nach dem bisher Gesagten klar: die Komponenten des Spaltenvektors \mathbf{u}_k sind die Koordinaten der Fälle auf Geraden, die latente Variablen repräsentieren. Da die Komponenten der Spaltenvektoren $\tilde{\mathbf{v}}_j$ von V' die Koeffizienten für die Darstellung der \mathbf{x}_j als Linearkombinationen der \mathbf{u}_k sind, repräsentieren die Komponenten von $\tilde{\mathbf{v}}_j$ die gemessene j -te Variable auf den verschiedenen latenten Variablen. Ein Spaltenvektor \mathbf{v}_k von V ist ein Eigenvektor von $C = X'X$ und definiert, wie im Folgenden gezeigt wird, die Orientierung der Spaltenvektoren \mathbf{u}_k von U . Um diesen Sachverhalt herleiten zu können werden die folgenden Begriffe benötigt.

Definition 3.2 *Es sei M eine symmetrische (n, n) -Matrix, und es gelte*

$$Q(\mathbf{x}) = \mathbf{x}'M\mathbf{x}, \quad \text{für alle } \mathbf{x} \in \mathbb{R}^n \quad (3.19)$$

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

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

Anmerkung: $Q(\mathbf{x})$ ist ein Skalar, denn $\mathbf{x}'M$ ist ein Zeilenvektor, etwa $\mathbf{x}'M = \mathbf{y}'$, so dass $\mathbf{x}'M\mathbf{x} = \mathbf{y}'\mathbf{x} = Q(\mathbf{x}) \in \mathbb{R}$. \square

Im Folgenden werden nur positiv-semidefinite Matrizen betrachtet. Die Ausführlichkeit der Definition 3.19 soll zeigen, dass eine symmetrische Matrix nicht notwendig positiv-semidefinit ist und eine Reihe von Aussagen nur für positiv-semidefinite Matrizen gilt. .

Definition 3.3 Es sei X eine (m, n) -Matrix. Dann heißt $C = X'X$ Gram-Matrix oder Gramsche Matrix⁴³.

Satz 3.4 Es sei M eine Gramsche Matrix. Dann ist M positiv-semidefinit.

Beweis: C ist eine (n, n) -Matrix, und es sei $\mathbf{v} \in \mathbb{R}^n$. Weiter sei $\mathbf{w} = X\mathbf{v}$. Dann ist

$$\mathbf{v}'X'X\mathbf{v} = \mathbf{w}'\mathbf{w} = \|\mathbf{w}\|^2 \geq 0,$$

(eine Summe von Quadraten ist stets größer oder gleich Null). \square

Folgerung: Die symmetrische (n, n) -Matrix M sei positiv-semidefinit. Dann gilt für Diagonalelemente m_{ii} von M

$$m_{ii} \geq 0. \tag{3.20}$$

Denn $\mathbf{x}'M\mathbf{x} \geq 0$ soll für alle $\mathbf{x} \in \mathbb{R}^n$ gelten, also auch für die n -dimensionalen Einheitsvektoren \mathbf{e}_i , $i = 1, \dots, n$. Es ist $M\mathbf{e}_i = \mathbf{m}_i$ die i -te Spalte von M , und $\mathbf{e}_i'M\mathbf{e}_i = \mathbf{e}_i'\mathbf{m}_i = m_{ii} \geq 0$ wegen der vorausgesetzten Positiv-Semidefinitheit. \square

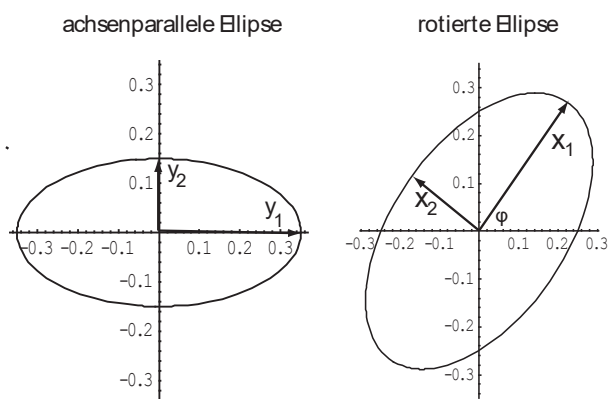
Ellipsoide: Die symmetrische Matrix M sei positiv-semidefinit. Multipliziert man die quadratische Form $\mathbf{x}'M\mathbf{x}$ aus so erhält man

$$\mathbf{x}'M\mathbf{x} = \sum_i m_{ii}x_i^2 + 2 \sum_{i < j} m_{ij}x_i x_j, \quad \mathbf{x} \neq \vec{0} \tag{3.21}$$

wobei x_i und x_j die Komponenten des Vektors \mathbf{x} und m_{ij} , $i, j = 1, \dots, n$, sind die Elemente von M . Der Ausdruck 'quadratische Form' ergibt sich aus dem Sachverhalt, dass (3.21) eine Verallgemeinerung der Formel für $(a + b)^2 = a^2 + b^2 + 2ab$, $a, b \in \mathbb{R}$ ist; man kann sich davon überzeugen, indem man den Ausdruck (3.21) für den Fall $n = 2$ explizit aufschreibt.

⁴³Nach Jörgen Pedersen Gram (1850 – 1916), dänischer Mathematiker.

Abbildung 13: Ellipsen in verschiedenen Orientierungen; die Hauptachsen sind die skalierten Eigenvektoren der zugehörigen Matrix M .



Man betrachte alle Vektoren $\mathbf{x} \neq \vec{0}$ mit Anfangspunkt im Ursprung des Koordinatensystems, die dieser Gleichung für einen bestimmten Wert von $k > 0$ genügen. Aus (3.21) folgt, dass die Endpunkte der \mathbf{x} im 2-dimensionalen Fall auf einer Ellipse, im allgemeinen n -dimensionalen Fall auf einem n -dimensionalen Ellipsoid \mathcal{E}_k liegen:

$$\mathcal{E}_k = \{\mathbf{x} | \mathbf{x}'M\mathbf{x} = k \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^n, k = \text{konstant} > 0\}, \quad (3.22)$$

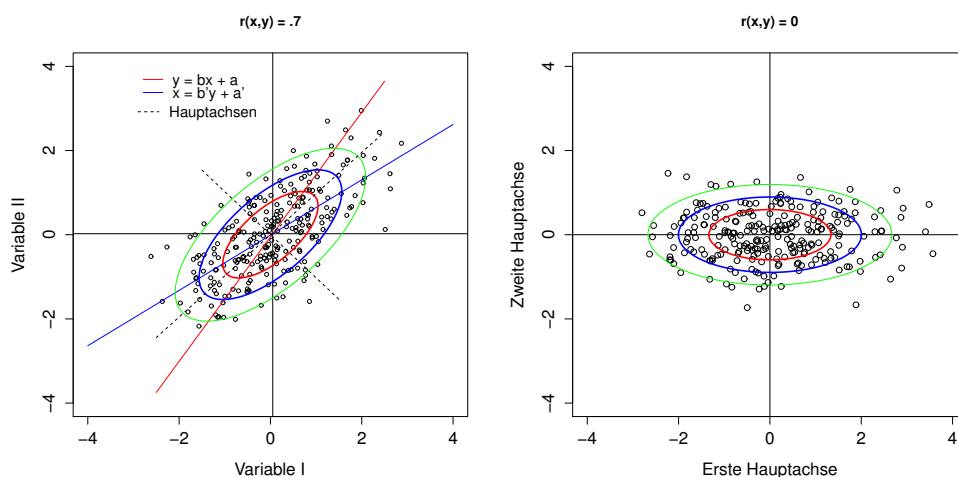
vergl. Abbildung 13. Gilt $m_{ij} = 0$ für alle $i \neq j$, d.h. ist M eine Diagonalmatrix, so ist \mathcal{E} *achsenparallel*, d.h. die Hauptachsen des Ellipsoids haben die Orientierung der Koordinatenachsen. Sind die $m_{ij} \neq 0$ für $i \neq j$, so ist das Ellipsoid nicht achsenparallel, d.h. die Hauptachsen des Ellipsoids sind nicht parallel zu den Koordinatenachsen.

Geometrische Bedeutung der Eigenvektoren: (Neuer Text 08. 10. 2023) Es sei $X = UV'$ und $M = X'X$, V sei die Matrix der Eigenvektoren von M , so dass

$$\mathbf{v}_k' M \mathbf{v}_k = \lambda_k = \|\mathbf{u}_k\|^2$$

und es gelte $\|\mathbf{u}_1\| \geq \dots \geq \|\mathbf{u}_n\|$. Die \mathbf{v}_k haben alle dieselbe Länge $\|\mathbf{v}_k\| = 1$, so dass man folgern kann, dass der Wert von $\|\mathbf{u}_k\|$ eine Funktion der Orientierung von \mathbf{v}_k ist. Die quadratische Form $\mathbf{x}'M\mathbf{x} = Q(\mathbf{x})$ definiert für $Q(\mathbf{x}) =$ eine Konstante ein Ellipsoid, dessen erste Hauptachse die maximale Ausdehnung des Ellipsoids repräsentiert, die zweite Hauptachse die zweitmaximale Ausdehnung, etc. Man kann jetzt nach dem Vektor \mathbf{x}_{\max} fragen, dessen Orientierung mit der der ersten Hauptachse identisch ist. Die Länge $\|\mathbf{x}_{\max}\|$ von \mathbf{x}_{\max} hängt vom Wert der Konstanten ab. Es ist klar, dass man nicht einfach nach dem Vektor fragen kann, für den $\mathbf{x}'M\mathbf{x} = Q(\mathbf{x})$ maximal ist, einfach weil $\mathbf{x}'M\mathbf{x} = \|\mathbf{x}\|^2$ maximal für $\|\mathbf{x}\| = \infty$ ist. Man muß sich also auf die Orientierung von \mathbf{x}_{\max} beschränken

Abbildung 14: Links: Punktekonfiguration für $r_{xy} = .7$ mit Regressionsgeraden, Ellipsen und deren Hauptachsen; Die Orientierungen der Regressionsgeraden sind im Allgemeinen nicht identisch mit der Orientierung der ersten Hauptachse, deren Orientierung die der maximalen Ausdehnung der Punktekonfiguration ist (s. Satz 3.5, Seite 89 in Abschnitt 3.2; man beachte auch die Anmerkungen zur Rolle der Maßeinheiten am Ende dieses Abschnitts). Da zentrierte Messwerte betrachtet werden gilt $a = a' = 0$. Rechts: Die Hauptachsen als neue Koordinatenachsen für die Punktekonfiguration. Die Komponenten der \mathbf{y}_i sind die Koordinaten in Bezug auf diese Achsen. Die Regressionskoeffizienten sind gleich Null, die Regressionsgeraden fallen mit der ersten Hauptachse zusammen. Zur Berechnung der Ellipsen s. den Anhang 5.5, Seite 170.



und \mathbf{c}_{max} bestimmen, d.h. man muß den in der folgenden Definition eingeführten Quotienten maximieren:

Definition 3.4 Es sei M eine symmetrische (n, n) - Matrix, und $\mathbf{x} \in \mathbb{R}^n$. Dann heißt

$$R(\mathbf{x}) = \frac{\mathbf{x}' M \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \mathbf{c}'_x M \mathbf{c}_x, \quad \mathbf{c}_x = \frac{\mathbf{x}}{\|\mathbf{x}\|} \quad (3.23)$$

der Rayleigh-Quotient⁴⁴ (s. (??), Seite ??), wobei \mathbf{c}_c der zu \mathbf{x} korrespondierende Vektor der Richtungskoeffizienten ist.

Die Frage nach \mathbf{x}_{max} wird allgemein durch den Satz von Courant-Fischer (auch Mini-max-Theorem von Courant-Fischer genannt) beantwortet, dessen ausführliche Formulierung im Anhang, Abschnitt 5.7 mit ebenfalls ausführlichem Beweis gegeben wird. Diesem Satz zufolge gilt

$$\lambda_{min} \leq R(\mathbf{x}) \leq \lambda_{max}, \quad (3.24)$$

⁴⁴Gelegentlich auch *Rayleigh-Koeffizient* genannt

d.h. der maximal mögliche Wert von $R(\mathbf{x})$ ist $\lambda_1 = \lambda_{\max}$, mit $\mathbf{c}_{\max} = \mathbf{c}_1 = \mathbf{v}_1$ dem zu λ_1 korrespondierendem Eigenvektor, der zweitgrößte Wert ist λ_2 mit zugehörigem Eigenvektor \mathbf{v}_2 , der \mathbf{c}_2 definiert, etc., wenn die Eigenwerte der Größe nach durchnummeriert werden. Hier wird zur Illustration nur der Beweis des ersten Teils des Courant-Fischer-Theorems gegeben; er kommt – wie der allgemeine Beweis des Courant-Fischer-Theorems – ohne Anwendung der Differentialrechnung mit Nebenbedingungen⁴⁵ aus:

Satz 3.5 *M sei symmetrisch und für die Eigenwerte gelte $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. Der Rayleigh-Quotient $R(\mathbf{x})$ wird maximal, wenn $\mathbf{c}_x = \mathbf{v}_1$, \mathbf{v}_1 der zu λ_1 korrespondierende Eigenvektor von M ist.*

Beweis: Es sei $V = [\mathbf{v}_1, \dots, \mathbf{v}_n]$ die Matrix der Eigenvektoren von M mit den zugehörigen Eigenwerten $\lambda_1 \geq \dots \geq \lambda_n$, so dass $MV = V\Lambda$, $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, und damit $M = V\Lambda V'$. Dann folgt

$$R(\mathbf{x}) = \frac{\mathbf{x}'M\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \frac{\mathbf{x}'V\Lambda V'\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \frac{\mathbf{w}'\Lambda\mathbf{w}}{\mathbf{w}'\mathbf{w}}, \quad \mathbf{w} = V'\mathbf{x}, \text{ bzw. } \mathbf{x} = V\mathbf{w}$$

$$\frac{\mathbf{x}'M\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \frac{\mathbf{w}'\Lambda\mathbf{w}}{\mathbf{w}'\mathbf{w}} = \frac{\sum_i w_i^2 \lambda_i}{\sum_i w_i^2} \leq \frac{\sum_i \lambda_1 w_i^2}{\sum_i w_i^2} = \frac{\lambda_1 \sum_i w_i^2}{\sum_i w_i^2} = \lambda_1.$$

Es sind also alle λ_i durch λ_1 ersetzt worden, und wegen $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ folgt, dass der Wert von λ_1 die obere Grenze für $R(\mathbf{x})$ ist. Es sei $\mathbf{x}_{\max} \in \mathbb{R}^n$ der Vektor, für den $\mathbf{x}'_{\max} M \mathbf{x}_{\max} = \lambda_1$. Weiter gilt sicher $\mathbf{v}'_1 M \mathbf{v}_1 = \lambda_1$, \mathbf{v}_1 der zu λ_1 korrespondierende Eigenvektor von M . Subtrahiert man die erste von der zweiten Gleichung, so erhält man $(\mathbf{x}_{\max} - \mathbf{v}_1)' M (\mathbf{x}_{\max} - \mathbf{v}_1) = 0$, d.h. es muss $\mathbf{x}_{\max} = \mathbf{v}_1$ gelten⁴⁶. Das heißt aber

$$R(\mathbf{v}_1) = \mathbf{v}'_1 M \mathbf{v}_1 = \lambda_1 \mathbf{v}'_1 \mathbf{v}_1 = \lambda_1,$$

wegen der Normiertheit der Eigenvektoren. Damit ist gezeigt worden, dass für $\mathbf{v} = \mathbf{v}_1$ der Rayleigh-Quotient $R(\mathbf{v}_1)$ maximal ist. \square

Es wird nun gezeigt, dass \mathbf{u}_1 unter den Bedingungen **A1** und **A2** die maximal mögliche Länge hat:

Korollar 3.1 *(zu Satz 3.5) Setzt man $M = C = X'X$, so erhält man (3.14) entsprechend*

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

Für einen beliebigen Vektor $\mathbf{v} \in \mathbb{R}^n$ von V heißt dies

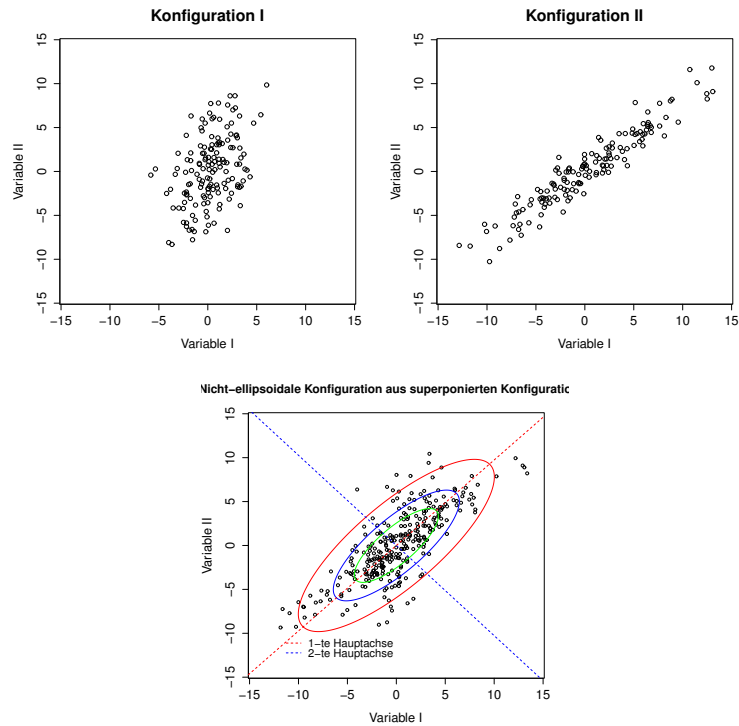
$$\mathbf{v}'C\mathbf{v} \leq \lambda_1 = \lambda_{\max}$$

und nach Satz 3.5 wird $\mathbf{v}'C\mathbf{v}$ maximal als Funktion von \mathbf{v} , wenn $\mathbf{v} = \mathbf{v}_1$, also $\mathbf{v}_1'C\mathbf{v}_1 = \lambda_1 = \|\mathbf{u}_1\|^2$.

⁴⁵Den klassischen Beweis (Extremwertbestimmung unter Nebenbedingungen) findet man im Anhang, Abschnitt 5.6

⁴⁶Für $\mathbf{x} \in \mathbb{R}^n$ und $M \neq 0$ ist $\mathbf{x}'M\mathbf{x} = 0$ genau dann, wenn $\mathbf{x} = \vec{0}$, vergl. Gleichung (3.21), Seite 86.

Abbildung 15: Superponierte Punktekongfigurationen und Ellipsen



Der zu $\lambda_{max} = \lambda_1$ korrespondierende Eigenvektor \mathbf{v}_1 repräsentiert also die Orientierung der ersten Hauptachse der Ellipsoide durch die Daten gegebenen gegebenen Konfiguration der Fälle, d.h. die Orientierung von \mathbf{u}_1 . Nach dem Courant-Fischer-Theorem repräsentieren die übrigen Eigenvektoren die Orientierungen der restlichen \mathbf{u}_k

Teilräume vereinigt in einem Teilraum höherer Dimension: Es sei noch angemerkt, dass die Punktekongfiguration der Fälle keinesfalls ein Ellipsoid sein muß, um durch eine Schar von Ellipsoiden (für jeden Fall ein Ellipsoid) beschrieben werden zu können, d.h. die Daten müssen nicht multivariat Gauß-verteilt sein, wie gelegentlich behauptet wird. In Abbildung 15 wird eine Punktekongfiguration gezeigt, die aus zwei Subpopulationen besteht, und jede von ihnen durch ein Ellipsoid repräsentiert wird, wobei sich die Ellipsoide sich durch die Orientierung ihrer Hauptachsen unterscheiden. Gleichwohl kann die Gesamtstichprobe durch eine Menge von Ellipsoiden mit identischer Orientierung repräsentiert werden, s. Abbildung 15. Im Prinzip hat man es hier mit zwei Teilräumen zu tun (vergl. Abbildung 9, Seite 33, die in einem Teilraum höherer Dimension zusammengefasst werden).

3.2.3 Zur Eindeutigkeit von Eigenvektoren

In Anmerkung 3 zur Definition von Eigenvektoren und Eigenwerten wurde die Frage nach der Eindeutigkeit der Matrix V von Eigenvektoren einer symmetrischen Matrix M gestellt. Man kann die Frage in etwas anderer Weise formulieren. Sei M eine symmetrische (n, n) -Matrix und es gelte $MV = V\Lambda$, d.h. die Spaltenvektoren \mathbf{v}_k der (n, n) -Matrix V sind Eigenvektoren von V , und $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ ist die Diagonalmatrix der Eigenwerte von V . Da die Eigenvektoren im Falle ungleicher Eigenwerte paarweise orthogonal sind bilden sie eine Orthogonalbasis des \mathbb{R}^n , d.h. $\mathcal{L}(V) = \mathcal{L}(\mathbf{v}_1, \dots, \mathbf{v}_n) = \mathbb{R}^n$, \mathcal{L} die lineare Hülle der Eigenvektoren. Sollte es eine zweite Matrix W von Eigenvektoren $\mathbf{w}_1, \dots, \mathbf{w}_n$ geben, so können die \mathbf{w}_k als Linearkombinationen der $\mathbf{v}_1, \dots, \mathbf{v}_n$ dargestellt werden, d.h. es existiert dann eine Matrix T derart, dass $W = VT$ und $\mathbf{w}_k = V\mathbf{t}_k$, \mathbf{t}_k der k -te Spaltenvektor von T . Von T wird im Folgenden nur gefordert, dass sie denselben Rang wie V hat, also $\text{rg}(W) = \text{rg}(V)$, so dass die Spaltenvektoren von W ebenfalls den \mathbb{R}^n aufspannen. Es wird aber nicht gefordert, dass T auch orthonormal ist, d.h. die Spaltenvektoren von W spannen zwar den \mathbb{R}^n auf, sind aber nur linear unabhängig und nicht notwendig auch orthogonal. Ein Spezialfall ist die Matrix $T_a = \text{diag}(a, \dots, a)$, d.h. T_a ist eine Diagonalmatrix, deren Diagonalelemente alle gleich $a \in \mathbb{R}$ sind. Die Spaltenvektoren von $W_a = VT_a$ sind bis auf eine Längentransformation identisch mit denen von V .

Satz 3.6 *Es sei M eine symmetrische (n, n) -Matrix und V sei eine Matrix von Eigenvektoren von M . Für die Eigenwerte gelte $\lambda_j \neq \lambda_k$ für $j \neq k$. Dann ist die Matrix V der Eigenvektoren von M eindeutig bestimmt bis auf eine Längenskalierung der Spaltenvektoren von V . Für den Fall identischer Eigenwerte, $\lambda_1 = \dots = \lambda_n = \lambda$ gibt es unendlich viele Matrizen $W = VT$, $T \neq T_a$, die der Gleichung $MW = W\Lambda$ genügen.*

Beweis: (i) Ungleiche Eigenwerte Es sei M eine symmetrische (n, n) -Matrix, und es gelte $MV = V\Lambda_v$, Λ_v die Diagonalmatrix der Eigenwerte mit $\lambda_j \neq \lambda_k$ für $j \neq k$. Weiter sei $W = VT$, $T \neq T_a$, T_a wie oben definiert, die zueinander korrespondierenden Spalten \mathbf{v}_k und \mathbf{w}_k von V und W sollen nicht parallel zueinander sein. Dann soll also

$$MV = V\Lambda_v \quad (3.25)$$

$$MW = W\Lambda_w \quad (3.26)$$

gelten⁴⁷. Aus (3.25) folgt $M = V\Lambda_v V'$, und die Multiplikation von (3.26) von links mit V' liefert

$$V'MW = V'V\Lambda_v V'W = V'W\Lambda_w,$$

d.h.

$$\Lambda_v V'W = V'W\Lambda_w. \quad (3.27)$$

⁴⁷Diese Bedingung schließt auch die triviale Transformation $T = I$, I die Identitätsmatrix (d.h. $a = 1$) aus!

Es sei $A = V'W$, so dass $\Lambda_V A = A\Lambda_W$. $\Lambda_V A$ bedeutet eine Skalierung der Zeilenvektoren $\tilde{\mathbf{a}}_i$ von A , und $A\Lambda_W$ bedeutet eine Skalierung der Spaltenvektoren von A . Es sei (a_{i1}, \dots, a_{in}) der i -te Zeilenvektor von A ; (3.27) bedeutet dann

$$(\lambda_i a_{i1}, \lambda_i a_{i2}, \dots, \lambda_i a_{in}) = (\mu_1 a_{i1}, \mu_2 a_{i2}, \dots, \mu_n a_{in}).$$

Zwei Vektoren sind identisch, wenn die korrespondierenden Komponenten identisch sind. Betrachtet man also die Gleichungen $\lambda_i a_{i1} = \mu_1 a_{i1}$, $\lambda_i a_{i2} = \mu_2 a_{i2}$, \dots , $\lambda_i a_{in} = \mu_n a_{in}$, so sieht man, dass sich die Komponenten a_{i1}, a_{i2}, \dots herauskürzen, – insgesamt (also für $i = 1, \dots, m$) hat man man die Gleichungen

$$\begin{array}{cccc} \lambda_1 = \mu_1, & \lambda_1 = \mu_2, & \cdots & \lambda_1 = \mu_n \\ \lambda_2 = \mu_1, & \lambda_2 = \mu_2, & \cdots & \lambda_2 = \mu_n \\ \vdots & \vdots & \cdots & \vdots \\ \lambda_n = \mu_1, & \lambda_n = \mu_2, & \cdots & \lambda_n = \mu_n \end{array}$$

erhält, und damit

$$\lambda_1 = \dots = \lambda_n = \lambda, \quad \mu_1 = \dots = \mu_n = \mu, \quad \lambda = \mu, \quad (3.28)$$

$V \neq W = VT$, $T \neq T_a$ impliziert also nicht nur $\Lambda_V = \Lambda_W := \Lambda$, sondern darüber hinaus, dass die Diagonalelemente von Λ identisch sind, entgegen der Annahme. Wenn also M verschiedene Eigenwerte hat folgt, dass V und W *nicht* verschieden sein können: $\neg(V \neq W) \equiv (V = W)$ ⁴⁸. Dies aber heißt, dass Eigenvektoren von M bis auf eine Längentransformation eindeutig bestimmt sind.

(ii) Identische Eigenwerte Nun werde der Fall $\lambda_1 = \dots = \lambda_n = \lambda$ betrachtet. Es sei V eine orthonormale Matrix derart, dass $MV = V\Lambda$, $\Lambda = \text{diag}(\lambda, \dots, \lambda)$. Die Spaltenvektoren von V sind offenbar Eigenvektoren von M . Angenommen, $W = VT$, $T \neq T_a$ für beliebiges $a \in \mathbb{R}$, und W genüge der Bedingung $MW = W\Lambda$; wie unter (i) gezeigt wurde, impliziert die Annahme der Existenz einer Matrix $W \neq V$, dass die zu W gehörige Diagonalmatrix der Eigenwerte dieselbe wie die zu V gehörige ist.

Voraussetzungsgemäß existiert eine Transformationsmatrix $T \neq T_a$ derart, dass $W = TV$, so dass $MW = MTV$. Für den k -ten Spaltenvektor \mathbf{w}_k von W gilt dann

$$\mathbf{w}_k = V\mathbf{t}_k = t_{1k}\mathbf{v}_1 + t_{2k}\mathbf{v}_2 + \dots + t_{nk}\mathbf{v}_n,$$

t_{1k}, \dots, t_{nk} die Komponenten des k -ten Spaltenvektors \mathbf{t}_k von T und $\mathbf{v}_1, \dots, \mathbf{v}_n$ die Spaltenvektoren von V . Zu zeigen ist, dass \mathbf{w}_k ein Eigenvektor von M ist,

⁴⁸– ist das Negationszeichen $\neg(V \neq W)$ heißt, die Aussage $V \neq W$ gilt nicht.

dass also $M\mathbf{w}_k = \lambda\mathbf{w}_k$ gilt. Es ist

$$\begin{aligned}
M\mathbf{w}_k = MT\mathbf{v}_k &= M(t_{1k}\mathbf{v}_1 + \cdots + t_{nk}\mathbf{v}_n) \\
&= t_{1k}M\mathbf{v}_1 + \cdots + t_{nk}M\mathbf{v}_n \\
&= t_{1k}\lambda\mathbf{v}_1 + \cdots + t_{nk}\lambda\mathbf{v}_n \\
&= \lambda(t_{1k}\mathbf{v}_1 + \cdots + t_{nk}\mathbf{v}_n) \\
&= \lambda\mathbf{w}_k
\end{aligned} \tag{3.29}$$

Damit ist gezeigt, dass die Linearkombination der Eigenvektoren $\mathbf{v}_1, \dots, \mathbf{v}_n$ ebenfalls ein Eigenvektor ist, der zum Eigenwert λ korrespondiert. Da T beliebig gewählt werden kann folgt, dass es beliebig viele Eigenvektormatrizen außer V gibt. \square

Anmerkungen:

1. Ist T orthonormal, und $W = VT$, so gilt $W'W = T'V'VT = T'T = I$, d.h. die Spaltenvektoren von W sind wieder orthonormal. Wie eingangs festgestellt wurde muß die Matrix T nicht orthogonal sein, und (3.29) zeigt, dass die \mathbf{w}_k gleichwohl Eigenvektoren sind. Für diese Eigenschaft ist es im Falle identischer Eigenwerte von M hinreichend, dass die Eigenvektoren von M linear unabhängig sind.

2. In der multivariaten Statistik ist M häufig durch die Kovarianz- bzw. Korrelationsmatrix $X'X$ gegeben, wobei X entweder spaltenzentriert oder spaltenstandardisiert ist. Die Messungen X_{ij} (i -ter Fall, j -te Variable) sind normalerweise fehlerbehaftet, so dass die Eigenwerte von $X'X$ im Allgemeinen verschieden sind. \square

Auf den Fall identischer Eigenwerte wird in Abschnitt 3.3.2 zurückgekommen.

3.2.4 Zur Geometrie der Eigenvektoren symmetrischer Matrizen

Mit der Rotation der empirisch gegebenen Vektoren $\tilde{\mathbf{x}}_i$ in die Vektoren $\tilde{\mathbf{u}}_i$ werden auch die Hauptachsen der für die Punktekonfiguration spezifischen Ellipsoide rotiert, weshalb A1 und A2 eine *Hauptachsentransformation* definieren. Sie entspricht dem Übergang vom X -Koordinatensystem zum "latenten" U -Koordinatensystem, das durch die transformierten (d.h. rotierten) Hauptachsen der Ellipsoide der Datenkonfiguration definiert ist.

Es seien

$$K_x = \{\tilde{\mathbf{x}}_i, 1 \leq i \leq m\}, \quad K_u = \{\tilde{\mathbf{u}}_i, 1 \leq i \leq m\} \tag{3.30}$$

die Punktekonfigurationen der Fälle im X -, bzw. im U -Koordinatensystem (das auch als (L_1, \dots, L_n) -Koordinatensystem der latenten Variablen aufgefasst werden kann).

Satz 3.7 *Es mögen die Annahmen A1 und A2 gelten. Es ist $X' = VU'$ und damit $\tilde{\mathbf{x}}_i = V\tilde{\mathbf{u}}_i$, $i = 1, \dots, m$, V ist die Matrix der Eigenvektoren von $C = X'X$ und Λ ist die Diagonalmatrix der zugehörigen Eigenwerte. Dann folgt*

$$\tilde{\mathbf{x}}_i' C \tilde{\mathbf{x}}_i = \tilde{\mathbf{u}}_i' \Lambda \tilde{\mathbf{u}}_i = k_i \in \mathbb{R}, \quad i = 1, \dots, m \tag{3.31}$$

Abbildung 16: Punktekonfiguration und Ellipsen. (a) Die Punktekonfiguration der Fälle im X -Koordinatensystem, (b) dieselbe Konfiguration im U - bzw. (L_1, L_2) -Koordinatensystem. Die Berechnungen beruhen *nicht* auf der Annahme der 2-dimensionalen Normalverteilung. Im Anhang, Abschnitt 5.5 wird die Konstruktion dieser Ellipsen näher erläutert.

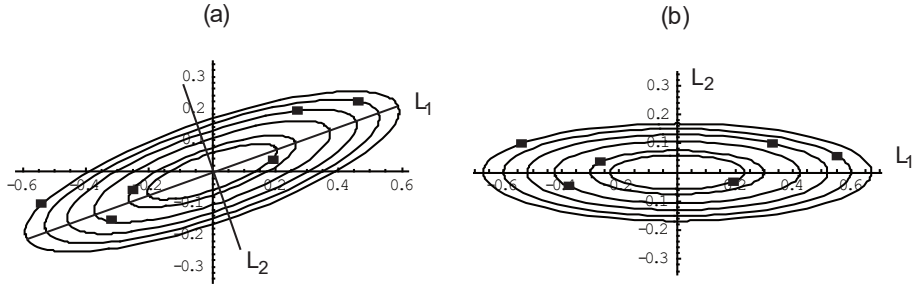
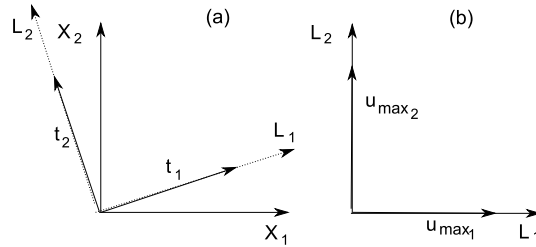


Abbildung 17: Orientierung der maximalen Ausdehnung der Konfiguration, (a) im (X_1, X_2) -, (b) im (L_1, L_2) -Koordinatensystem. Die Vektoren $\mathbf{t}_1, \mathbf{t}_2, \mathbf{u}_{\max_1}$ und \mathbf{u}_{\max_2} haben die Länge 1.



d.h. jeder Punkt aus K_x bzw. K_u liegt auf einem durch C bzw. Λ definierten Ellipsoid \mathcal{E}_{x,k_i} bzw. \mathcal{E}_{u,k_i}

$$\mathcal{E}_{x,k_i} = \{\tilde{\mathbf{x}} | \tilde{\mathbf{x}}' C \tilde{\mathbf{x}} = k_i\}, \quad \tilde{\mathbf{x}}_i \in \mathcal{E}_{x,k_i} \quad (3.32)$$

$$\mathcal{E}_{u,k_i} = \{\tilde{\mathbf{u}} | \tilde{\mathbf{u}}' \Lambda \tilde{\mathbf{u}} = k_i\}, \quad \tilde{\mathbf{u}}_i \in \mathcal{E}_{u,k_i} \quad (3.33)$$

Beweis: Aus $\tilde{\mathbf{x}}_i = V \tilde{\mathbf{u}}_i$ folgt $\tilde{\mathbf{x}}_i' = \tilde{\mathbf{u}}_i' V'$ und

$$\tilde{\mathbf{x}}_i' C \tilde{\mathbf{x}}_i = \tilde{\mathbf{u}}_i' V' C V \tilde{\mathbf{u}}_i = k_i$$

Wegen $V' C V = \Lambda$ (vergl. (??)) hat man dann

$$\tilde{\mathbf{u}}_i' V' C V \tilde{\mathbf{u}}_i = \tilde{\mathbf{u}}_i' \Lambda \tilde{\mathbf{u}}_i = k_i,$$

d.h. es gilt (3.31). □

Die Ellipsoide \mathcal{E}_{x,k_i} für $i = 1, \dots, m$ haben alle dieselbe Orientierung, s.a. Abbildung 17. Die Orientierung der Ellipsoide und damit der Konfigurationen ist durch ihre jeweiligen Hauptachsen gegeben. Nach Satz 3.5, Seite 89 ist die Orientierung der maximalen Ausdehnung durch $\tilde{\mathbf{x}} = \mathbf{t}_1$ gegeben, die für die zweitgrößte Ausdehnung durch $\tilde{\mathbf{x}} = \mathbf{t}_2$, etc. Wegen $V'\tilde{\mathbf{x}} = \tilde{\mathbf{u}}$ hat man also

$$V'\mathbf{t}_1 = \tilde{\mathbf{u}}_{\max_1} = \mathbf{e}_1 = (1, 0, 0, \dots, 0)' \quad (3.34)$$

$$V'\mathbf{t}_2 = \tilde{\mathbf{u}}_{\max_2} = \mathbf{e}_2 = (0, 1, 0, \dots, 0)' \quad (3.35)$$

$$\begin{aligned} & \vdots \\ V'\mathbf{t}_n &= \tilde{\mathbf{u}}_{\max_n} = \mathbf{e}_n = (0, 0, 0, \dots, 1)' \quad (3.36) \end{aligned}$$

Da die $\tilde{\mathbf{u}}_k$, $k = 1, \dots, n$ die Projektionen der Fälle auf die latenten Achsen L_1, \dots, L_n sind, sind die Ellipsoide im durch die L_k definierten Koordinatensystem "achsenparallel", d.h. die Hauptachsen sind parallel zu den Koordinatenachsen L_k .

3.2.5 Rang und Anzahl der Eigenwerte ungleich Null

Es gilt der

Satz 3.8 *Es sei X eine (m, n) -Matrix mit dem Rang $r \leq \min(m, n)$. Weiter sei $C = X'X$, und V sei die (n, n) -Matrix der Eigenvektoren und Λ sei die (n, n) -Matrix der Eigenwerte $\lambda_k \geq 0$ von C . Dann ist die Anzahl der Eigenwerte ungleich Null von Λ gleich r .*

Beweis: Es gilt $V'CV = \Lambda$ und nach Satz 2.3, Gleichung (2.37), S. 68, gilt

$$\text{rg}(V'CV) = \text{rg}(\Lambda) = \text{rg}(C),$$

und nach von Satz 2.7, Seite 74 ist $\text{rg}(\Lambda)$ gleich der Anzahl der Diagonalelemente (hier der Eigenwerte) ungleich Null. \square

Satz 3.9 *Es sei X eine (m, n) -Matrix mit dem Rang $r \leq \min(m, n)$. Dann haben $X'X$ und XX' dieselben von Null verschiedenen Eigenwerte $\lambda_1, \dots, \lambda_r$.*

Beweis: Allgemein gilt $X = UV'$ (vergl. (2.35), Seite 68). X habe den Rang $r \leq \min(m, n)$, und U ist eine (m, r) - und V ist eine (n, r) -Matrix. Dann ist $X' = VU'$, d.h. die Spaltenvektoren von X' sind Linearkombinationen der Spaltenvektoren von V . Es sei insbesondere V die Matrix der Eigenvektoren von $X'X$; V ist dann eine orthonormale Basis für die Spaltenvektoren von X' . Dann folgt $XV = U$ und

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

Λ die Diagonalmatrix der r von Null verschiedenen Eigenwerte, d.h. U ist orthogonal und die Eigenwerte sind gleich den Quadraten der Längen der Spaltenvektoren

von U . Dann enthält $Q = U\Lambda^{-1/2}$ die normierten Vektoren, d.h. $U = Q\Lambda^{1/2}$. Man hat nun

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

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

Anmerkung: Geht man zunächst vom Fall $r = n$ aus, so sind V und Λ (n, n) -Matrizen. Ist dann tatsächlich $r < n$, so sind nur r Diagonalelemente von Λ ungleich Null, die restlichen $n - r$ Diagonalelemente sind gleich Null. Das heißt auch, dass von den n Spaltenvektoren von Λ $n - r$ Nullvektoren sind; analog dazu sind $n - r$ Zeilenvektoren Nullvektoren. Dies heißt aber *nicht*, dass V ebenfalls $n - r$ Nullvektoren enthält! Es heißt nur, dass man die r benötigten orthogonalen Vektoren \mathbf{v}_k , $k = 1, \dots, r$ stets zu einer vollständigen Basis von n orthogonalen Vektoren erweitern kann. Da Λ nur r Spalten- (und Zeilen-)vektoren enthält, die nicht gleich dem Nullvektor sind, bedeutet dies, dass tatsächlich nur die ersten r Spaltenvektoren des Produkts $V\Lambda$ ungleich dem Nullvektor sind. Insofern sind die zu Eigenwerten gleich Null korrespondierenden Vektoren in V redundant. Trotzdem macht man i. A. von der Schreibweise $CV = V\Lambda$ mit V und C (n, n) -Matrizen Gebrauch, da er den Fall $r < n$ als Spezialfall enthält.

Andererseits kann man natürlich auch $V\Lambda V' = V_r\Lambda_r V_r'$ schreiben, wobei Λ_r eine (r, r) -Diagonalmatrix ist, deren Diagonalelemente nur die von Null verschiedenen Eigenwerte enthält, und V_r ist die (n, r) -Matrix der Eigenvektoren von M , die zu den von Null verschiedenen Eigenwerten korrespondieren. Von dieser Möglichkeit wird allerdings hauptsächlich dann Gebrauch gemacht, wenn angezeigt werden soll, dass X mit dem *numerischen* Rang n durch eine Matrix von kleinerem Rang $r < n$ approximiert werden soll; hierauf wird später noch ausführlich eingegangen werden. \square

Satz 3.10 *Es sei M eine symmetrische, positiv-semidefinite (n, n) -Matrix mit dem Rang r , und V sei die Matrix der Eigenvektoren \mathbf{v}_k von M und Λ sei die Diagonalmatrix der zugehörigen Eigenwerte λ_k , $\lambda_j \neq \lambda_k$ für $j \neq k$. m_{ii} , $i = 1, \dots, n$ seien die Diagonalelemente von M . Dann gilt*

$$\sum_{i=1}^n m_{ii} = \sum_{k=1}^n \lambda_k. \quad (3.37)$$

Anmerkung: Die Summe $\sum_{i=1}^n m_{ii}$ heißt auch *Spur* der Matrix M . Auf Satz 3.37 wird einem anderen Zusammenhang noch einmal zurückgekommen (vergl. Satz 3.24, Seite 126). \square

Beweis: Die Aussage (3.37) folgt aus dem Sachverhalt, dass die Eigenvektoren reeller, symmetrischer Matrizen stets orthonormal sind. Man hat aus $MV = V\Lambda$ (Definition der Eigenvektoren) die Darstellung $M = M\Lambda V'$. $M\Lambda$ ist die Matrix

der mit den Eigenwerten skalierten Spaltenvektoren von M , d.h.

$$M\Lambda = [\lambda_1 \mathbf{v}_1, \lambda_2 \mathbf{v}_2, \dots, \lambda_n \mathbf{v}_n].$$

m_{ii} muß dann gleich dem Skalarprodukt des i -ten Zeilenvektors von $V\Lambda$ mit dem i -ten Spaltenvektor $\tilde{\mathbf{v}}_i$ von V' , – d.h. mit dem i -ten Zeilenvektor von V sein, also

$$m_{ii} = (\lambda_1 v_{i1}, \dots, \lambda_n v_{in}) \begin{pmatrix} v_{i1} \\ v_{i2} \\ \vdots \\ v_{in} \end{pmatrix} = \sum_{k=1}^n \lambda_k v_{ik}^2.$$

Dann ist

$$\sum_{i=1}^n m_{ii} = \sum_{i=1}^n \sum_{k=1}^n \lambda_k v_{ik}^2 = \sum_{k=1}^n \lambda_k \sum_{i=1}^n v_{ik}^2 = \sum_{k=1}^n \lambda_k,$$

denn $\sum_{i=1}^n v_{ik}^2 = 1$ wegen der Normiertheit der Eigenvektoren. \square

Es sei X eine spaltenzentrierte Datenmatrix. Dann ist $C = \frac{1}{m} X'X$ die symmetrische Matrix der Varianzen (in der Diagonale von C) und Kovarianzen zwischen den Variablen, die für die Spalten von X stehen. Das Element c_{jj} von C ist die Varianz

$$s_j^2 = \frac{1}{m} \sum_{i=1}^m x_{ij}^2, \quad x_{ij} = (X_{ij} - \bar{x}_j)^2$$

der j -ten Variablen. $V_x = \sum_{j=1}^n c_{jj}$ wird auch als *Gesamtvarianz* der Daten bezeichnet. Nach (3.37) gilt also

$$V_x = \sum_{j=1}^n c_{jj} = \sum_{k=1}^n \lambda_k \quad (3.38)$$

Wegen der allgegenwärtigen Messfehler sind im Allgemeinen alle $\lambda_k \neq 0$, obwohl der "wahre" Wert von r kleiner als n sein kann; in dem Fall sind die letzten $n - r$ Eigenwerte gleich Null. Man kann versuchen, den kleinstmöglichen Wert von r so zu bestimmen, dass V_x gut durch die Summe der r größten Eigenwerte approximiert wird.

Spezialfall: Die Matrix X sei spaltenstandardisiert, so dass $\frac{1}{m} X'X = R$ die Matrix der Korrelationen zwischen den Variablen ist. Dann ist $m_{ii} = r_{ii} = 1$ für alle i und man hat

$$\sum_{i=1}^n r_{ii} = n = \sum_{k=1}^n \lambda_k, \quad (3.39)$$

die Summe der Eigenwerte ist gleich der Anzahl der Variablen.

Der folgende Satz spezifiziert die Bedingungen, die eine symmetrische Matrix M erfüllen muss, um positiv semidefinit zu sein, d.h. um eine Ellipse bzw. ein Ellipsoid zu definieren.

Satz 3.11 *Es sei M eine symmetrische (n, n) -Matrix vom Rang $r \leq n$. Dann ist M genau dann positiv semidefinit, wenn eine (n, r) -Matrix G existiert derart, dass*

$$M = GG'. \quad (3.40)$$

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

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

so dass M positiv semidefinit ist.

(2) \Leftarrow : Aus der Symmetrie von M folgt die Existenz der orthonormalen Matrix V und der Diagonalmatrix $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, $\lambda_j \geq 0$ für $j = 1, \dots, n$, mit $M = V\Lambda V'$. Es sei

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

Dann kann man

$$M = V\Lambda^{1/2}\Lambda^{1/2}V' = (V\Lambda^{1/2})(V\Lambda^{1/2})'$$

schreiben. Streicht man in $V\Lambda^{1/2}$ alle Spalten, die nur Nullen enthalten, so erhält man eine Matrix $G = V_r\Lambda_r^{1/2}$ und M ist in der Form $M = GG'$ darstellbar. \square

Bemerkung 3.1 Eine Matrix muss nicht symmetrisch sein, damit Eigenvektoren für sie existieren, und andererseits existieren nicht für jede symmetrische Matrix Eigenvektoren mit *reellwertigen* Komponenten (d.h. es ist möglich, dass Eigenvektoren mit komplexwertigen Komponenten existieren, vgl. Satz 3.14, 107). Dazu betrachte man die Matrix (3.8), Seite 83, d.h.

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

Dann ist

$$T\mathbf{x} = x_1 \begin{pmatrix} \cos \phi \\ \sin \phi \end{pmatrix} + x_2 \begin{pmatrix} -\sin \phi \\ \cos \phi \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \mathbf{y}.$$

\mathbf{x} ist ein Eigenvektor von T , wenn $\mathbf{y} = \lambda\mathbf{x}$, d.h. wenn sie dieselbe Orientierung haben, also wenn

$$\frac{x_2}{x_1} = \frac{y_2}{y_1}$$

gilt. Es ist aber

$$\frac{y_2}{y_1} = \frac{x_1 \sin \phi + x_2 \cos \phi}{x_1 \cos \phi - x_2 \sin \phi},$$

so dass \mathbf{y} parallel zu \mathbf{x} nur für spezielle Werte von ϕ ist, für die $\cos \phi = 1$ und $\sin \phi = 0$ gilt, also z.B. für $\phi = 0$, so dass $T = I$ mit den Spaltenvektoren $(1, 0)'$ und $(0, 1)'$. Dies ist der gewissermaßen triviale Fall, bei dem gar keine Rotation erzeugt wird. Man findet allerdings komplexwertige Eigenvektoren mit zugehörigen komplexwertigen Eigenwerten, – für $\phi = \pi/4$ etwa findet man die Eigenvektoren $(i, 1)'$ und $(-i, 1)'$ mit den Eigenwerten $(1+i)/\sqrt{2}$ und $(1-i)/\sqrt{2}$, mit $i = \sqrt{-1}$, wie man durch Nachrechnen bestätigt. Komplexe Eigenvektoren und -werte werden allerdings im Folgenden keine Rolle spielen.

3.2.6 Inverse und Wurzel einer symmetrischen Matrix

Es sei M eine symmetrische (n, n) -Matrix mit vollem Rang, d.h. $\text{rg}(M) = n$, so dass die Inverse M^{-1} von M existiert. Die Spektraldarstellung von M sei $M = V\Lambda V'$. Es gelten die Beziehungen

$$M = V\Lambda V' = \sum_{k=1}^n \lambda_k \mathbf{v}_k \mathbf{v}_k' \quad (3.41)$$

$$M^{-1} = V\Lambda^{-1}V' = \sum_{k=1}^n \frac{1}{\lambda_k} \mathbf{v}_k \mathbf{v}_k' \quad (3.42)$$

Die Gleichung (3.41) ergibt sich aus $M = V\Lambda V'$ einfach durch Ausmultiplizieren. Die Gleichung (3.42) ergibt sich wie folgt: Es ist (vergl. Satz 2.58, Seite 78)

$$M^{-1} = (V\Lambda V')^{-1} = (V')^{-1} \Lambda^{-1} V^{-1},$$

wobei $\Lambda^{-1} = \text{diag}(\lambda_1^{-1}, \dots, \lambda_n^{-1})$. Die Orthonormalität von V impliziert $V' = V^{-1}$, so dass $(V')^{-1} = (V^{-1})^{-1} = V$, so dass die Inverse von M durch (3.42) gegeben ist. Die $\mathbf{v}_k \mathbf{v}_k'$ sind die dyadischen Produkte der Eigenvektoren von M .

Beispiel 3.1 Varianz von Parameterschätzungen: Die Beziehung (3.42) ist von Bedeutung u.a. bei der Interpretation von Regressionsparametern. Für die Schätzung des Parametervektors $\mathbf{b} = (b_1, \dots, b_n)'$ einer multiplen Regression $\mathbf{y} = X\mathbf{b} + \mathbf{e}$, X die zentrierte Matrix der Prädiktorwerte, gilt

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

(s. Anhang, Abschnitt 5.6.3). Die Varianz-Kovarianzmatrix $D(\hat{\mathbf{b}})$ ist (Seber (1977), p. 48) und wegen (3.42) durch

$$D(\hat{\mathbf{b}}) = \sigma^2(X'X)^{-1} = \sigma^2 V\Lambda^{-1}V' = \sigma^2 \sum_{k=1}^n \frac{1}{\lambda_k} \mathbf{v}_k \mathbf{v}_k'$$

gegeben). Offenbar werden Stichprobenvarianzen der Schätzungen der Regressionsparameter "groß", wenn es "kleine" Eigenwerte λ_k gibt. \square

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 nun⁴⁹

$$M^{1/2} \stackrel{\text{Def}}{=} T\Lambda^{1/2}T' = \sum_{k=1}^n \sqrt{\lambda_k} \mathbf{t}_k \mathbf{t}_k' \quad (3.43)$$

⁴⁹Mit dem Zeichen $\stackrel{\text{Def}}{=}$ soll ausgedrückt werden, dass der Ausdruck auf der linken Seite durch den auf der rechten Seite definiert wird.

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

$$(M^{1/2})' = (T\Lambda^{1/2}T')' = T\Lambda^{1/2}T', \quad (3.44)$$

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

$$(M^{1/2})^{-1} = M^{-1/2} = (T\Lambda^{1/2}T')^{-1} = T\Lambda^{-1/2}T' = \sum_{k=1}^n \frac{1}{\sqrt{\lambda_k}} \mathbf{t}_k \mathbf{t}'_k. \quad (3.45)$$

3.2.7 Die Singularwertzerlegung einer Matrix

Die im Folgenden dargestellte Zerlegung einer Matrix gilt für beliebige (m, n) -Matrizen, und die Betrachtung von Datenmatrizen erleichtert möglicherweise den intuitiven Zugang. Es geht um die Frage, wie latente Variablen zur Erklärung von beobachteten Daten bestimmt werden können. Die Daten seien in einer (m, n) -Matrix X zusammengefasst worden, bei der die Komponenten der Spaltenvektoren \mathbf{x}_j die Messungen x_{ij} der j -ten Variablen ($1 \leq j \leq n$) bei $i = 1, \dots, m$ Fällen repräsentieren. Die Komponenten der Zeilenvektoren $\tilde{\mathbf{x}}_i$ sind die Messungen x_{ij} , $j = 1, \dots, n$, beim i -ten Fall für alle n Variablen.

Die Betrachtungen der vorangegangenen Abschnitte können in einem Satz⁵⁰ über die *Singularwertzerlegung* der Matrix X zusammengefasst werden, der für die multivariate Analyse von Daten von zentraler Bedeutung ist:

Satz 3.12 *Es sei X eine (m, n) -Matrix mit dem Rang $r \leq \min(m, n)$. Dann gilt die Singularwertzerlegung (SVD⁵¹)*

$$X = Q\Sigma V', \quad (3.46)$$

wobei Σ eine (m, n) -Matrix der Form

$$\Sigma = \begin{pmatrix} \Lambda_r^{1/2} & 0_{12} \\ 0_{21} & 0_{22} \end{pmatrix} \quad (3.47)$$

mit $\Lambda_r^{1/2} = \text{diag}(\lambda_1^{1/2}, \dots, \lambda_r^{1/2})$ ist eine (r, r) -Diagonalmatrix, und λ_j , $j = 1, \dots, r$ sind die von Null verschiedenen Eigenwerte von $X'X$ bzw. XX' . 0_{12} ist eine $(r, n - r)$ -Matrix, 0_{21} ist eine $(m - r, r)$ -Matrix und 0_{22} ist eine $(m - r, n - r)$ -Matrix; die Elemente von 0_{12} , 0_{21} und 0_{22} sind alle gleich Null. Q ist eine (m, m) -Matrix, deren Spaltenvektoren die Eigenvektoren von XX' sind, und V ist eine (n, n) -Matrix, deren Spaltenvektoren die Eigenvektoren von $X'X$ sind. Die von Null verschiedenen Eigenwerte sind für XX' und $X'X$ identisch.

⁵⁰Der Satz hat aber eine längere Vorgeschichte, vergl. Stewart (1993).

⁵¹Übliche Abkürzung für Singularwertzerlegung, von engl. Singular Value Decomposition.

Anmerkungen: Die Dimensionen der Matrizen Q und V sind so gewählt worden, dass allen möglichen Fällen ($m > n$, $m = n$ und $m < n$) Genüge getan wird. Die Diagonalmatrix Σ enthält die Wurzeln aus den Eigenwerten; sie heißen auch *Singularwerte*; $\sigma_j = \sqrt{\lambda_j}$, $j = 1, \dots, r$. Die Spaltenvektoren von Q heißen *Linkseigenvektoren*, die von V heißen *Rechtseigenvektoren* von X . Ein anderer, häufig gebrauchter Ausdruck für die SVD ist 'Grundstruktur' einer Matrix (engl. basic structure). \square

Beweis: Generell gilt $X = UV'$; wegen $X' = VU'$ kann angenommen werden, dass die Spaltenvektoren \mathbf{v}_k , $k = 1, \dots, n$ von V linear unabhängig sind, so dass die Spaltenvektoren $\tilde{\mathbf{x}}_i$, $i = 1, \dots, m$ als Linearkombinationen der \mathbf{v}_k dargestellt werden können, so dass $\tilde{\mathbf{x}}_i = V\tilde{\mathbf{u}}_i$. Insbesondere kann V als orthonormale Matrix gewählt werden, so dass $XV = U$ folgt. Insbesondere kann V so gewählt werden, dass $U'U = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, wobei $\lambda_k = \|\mathbf{u}_k\|^2 \geq 0$, $k = 1, \dots, n$, \mathbf{u}_k der k -te Spaltenvektor von U . Es sei $\mathbf{q}_k = (1/\|\mathbf{u}_k\|)\mathbf{u}_k$ für $\|\mathbf{u}_k\| > 0$, so dass $\|\mathbf{q}_k\| = 1$; offenbar ist $1/\|\mathbf{u}_k\| = 1/\sqrt{\lambda_k} = \lambda_k^{-1/2}$. Es sei

$$\Lambda^{-1/2} := \text{diag}(\lambda_1^{-1/2}, \dots, \lambda_n^{-1/2});$$

dann sei

$$Q = U\Lambda^{-1/2} = [\mathbf{q}_1, \dots, \mathbf{q}_n],$$

d.h. die Spaltenvektoren \mathbf{q}_k von Q sind die auf die Länge 1 normierten Spaltenvektoren von U . Es folgt $U = Q\Lambda^{1/2}$ und man hat

$$X = UV' = Q\Lambda^{1/2}V'. \quad (3.48)$$

Dann folgt $X'X = V\Lambda^{1/2}Q'Q\Lambda^{1/2}V' = V\Lambda V'$ und $X'XV = V\Lambda$, so dass die V die Eigenvektoren von $X'X$ sind. Weiter folgt $XX' = Q\Lambda^{1/2}V'V\Lambda^{1/2}Q' = Q\Lambda Q'$, so dass $XX'Q = Q\Lambda$, d.h. die Spaltenvektoren von Q sind die Eigenvektoren von Q und die von Null verschiedenen Eigenwerte von $X'X$ und XX' sind identisch. \square

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

Eine Anwendung der SVD wird in Abschnitt 3.6.1 vorgestellt.

Zur Beziehung zwischen den Eigenvektoren von XX' und $X'X$: Nach Multiplikation der SVD $X = Q\Sigma V'$ von rechts mit V erhält man

$$XV = Q\Sigma, \quad (3.49)$$

und die Multiplikation der SVD von links mit Q' ergibt $Q'X = \Sigma V'$, oder

$$X'Q = V\Sigma, \quad (3.50)$$

so dass man für die Eigenvektoren \mathbf{q}_k aus Q und \mathbf{v}_k aus V die Beziehungen

$$X\mathbf{v}_k = \sigma_k\mathbf{q}_k \quad (3.51)$$

$$X'\mathbf{q}_k = \sigma_k\mathbf{v}_k, \quad k = 1, \dots, n \quad (3.52)$$

erhält. Diese beiden Gleichungen zeigen die Beziehungen zwischen den \mathbf{v}_k und den \mathbf{q}_k : $\sigma_k \mathbf{q}_k$ ist eine Linearkombination der Spalten von X mit den Komponenten von \mathbf{v}_k als Koeffizienten, und $\sigma_k \mathbf{q}_k$ ist eine Linearkombination der Zeilen von X (Spalten von X') mit den Komponenten von \mathbf{q}_k als Koeffizienten.

Darstellung der SVD über das dyadische Produkt: Die Zerlegung $X = Q\Sigma V'$ kann in der Form

$$X = \sum_{k=1}^n \sigma_k \mathbf{q}_k \mathbf{v}'_k, \quad \sigma_k = \sqrt{\lambda_k}, \quad m \geq n \quad (3.53)$$

dargestellt werden, wobei \mathbf{q}_j die Spaltenvektoren von Q und \mathbf{v}_j die Spaltenvektoren von V sind. $\mathbf{q}_j \mathbf{v}'_j$ ist das dyadische Produkt dieser Vektoren. X kann also als Summe von Matrizen aufgefasst werden, die jeweils eine Dimension repräsentieren. Das Element x_{ij} ist demnach durch die Summe

$$x_{ij} = \sigma_1 q_{i1} v_{j1} + \cdots + \sigma_n q_{in} v_{jn} \quad (3.54)$$

gegeben, $\sigma_1 q_{i1} v_{j1}$ ist der Beitrag der ersten latenten Dimension, etc.

Üblicherweise werden die Eigenwerte λ_k in Λ so angeordnet, dass λ_1 der größte und λ_n der kleinste Eigenwert ist, – numerisch sind üblicherweise schon aufgrund von Rundungsfehlern alle $\lambda_k \neq 0$, auch wenn der "wahre" Rang von X kleiner als $\min(m, n)$ ist. Man hat also

$$\sigma_1 > \sigma_2 > \cdots > \sigma_n.$$

Die Matrix X wird in (3.53) als Summe der als dyadische Produkte der Eigenvektoren \mathbf{q}_k von XX' und der Eigenvektoren \mathbf{v}_k von $X'X$ definierten Matrizen, jeweils gewichtet mit $\sigma_k = \sqrt{\lambda_k}$, dargestellt. Betrachtet man die σ_k für einen bestimmten Wert $r < k$ als "hinreichend" klein, kann man X durch eine Matrix \hat{X}_r approximieren:

$$\hat{X}_r = \sum_{k=1}^r \sigma_k \mathbf{q}_k \mathbf{v}'_k \approx X, \quad \sigma_k = \sqrt{\lambda_k}, \quad r < n \quad (3.55)$$

"Hinreichend" klein bedeutet, dass man die σ_k mit $k > r$ als eben nur "zufällig" von Null abweichend betrachtet. Darauf wird in Abschnitt 3.6.1 noch eingegangen.

Anmerkung: In Abschnitt 5.6.4, Seite 176, wird der Satz von Courant-Fischer durch Differentiation der quadratischen Form (5.69), Seite 178 (hier also $T'X'XT = L'L = \Lambda$) unter der Nebenbedingung $\mathbf{t}'\mathbf{t} = 1$ bewiesen; dieser Herleitung entnimmt man leicht, dass es nur eine Lösung für T gibt, eben die Matrix der Eigenvektoren von $X'X$, denn die Ableitung $Q(\mathbf{t}) = d(\mathbf{t}'X'X\mathbf{t})/d\mathbf{t} = 0$ hat nur eine Lösung für \mathbf{t} . Es gibt also keine Rotation $T_1 \neq T$ derart, dass $XT_1 = L_1$ mit $L'_1 L_1 = \Lambda_1$, Λ_1 eine Diagonalmatrix. Wählt man demnach eine von T verschiedene Matrix T_1 , um die Vektoren von X zu rotieren, so repräsentieren die zu T_1 korrespondierenden $\mathbf{L}_k^{(1)}$ ein Koordinatensystem, in Bezug auf das die Konfiguration der Fälle nicht mehr achsenparallel ist, d.h. die latenten Variablen sind nicht mehr unkorreliert. \square

3.2.8 Die multivariate Gauss-Verteilung

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

Ein zufälliger Vektor wird etwa mit

$$\vec{X} = \mathbf{X} = (X_1, X_2, \dots, X_n)' \quad (3.56)$$

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

$$X = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{pmatrix} = (X_{ij}), \quad i = 1, \dots, m; j = 1, \dots, n \quad (3.57)$$

eine zufällige Matrix. Datenmatrizen sind Beispiele für zufällige Matrizen. Zufällige Vektoren werden mit einer multivariaten Verteilung bzw. Dichtefunktion $f(x_1, \dots, x_n)$ assoziiert; in Abschnitt 5.10.3 wird insbesondere die multivariate Normalverteilung besprochen.

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

$$F(x_1, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) \quad (3.58)$$

charakterisiert.

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

$$\mathbb{E}(X) = \begin{cases} \int_{-\infty}^{\infty} xf(x)dx, & X \text{ ist stetig} \\ \sum_i p_i X_i, & X \text{ ist diskret} \end{cases} \quad (3.59)$$

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

$$\mathbb{E}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i \mathbb{E}(X_i); \quad (3.60)$$

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

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

$$\mathbb{E}(\mathbf{X}) = (\mathbb{E}(X_1), \mathbb{E}(X_2), \dots, \mathbb{E}(X_n))'. \quad (3.61)$$

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

$$\boldsymbol{\mu} = \mathbb{E}(\mathbf{X}) = (\mu_1, \mu_2, \dots, \mu_n)', \quad (3.62)$$

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

Mit

$$\mathbf{X} - \boldsymbol{\mu} = (X_1 - \mu_1, X_2 - \mu_2, \dots, X_n - \mu_n)' \quad (3.63)$$

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

$$\boldsymbol{\Sigma} = \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'] = (\mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)']) = (\sigma_{ij}) \quad (3.64)$$

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

$$\sigma_{ii} = \sigma_i^2 = \mathbb{E}[(X_i - \mu_i)^2] \quad (3.65)$$

Satz 3.13 *Es sei $\mathbf{Z} = (Z_1, \dots, Z_n)'$ ein zufälliger Vektor, dessen Komponenten paarweise unabhängig sind mit $\mathbb{E}(Z_i) = 0$, $\text{Var}(Z_i) = 1$, $1 \leq i \leq n$. Weiter sei A eine (n, n) -Matrix und $\mathbf{X} = A\mathbf{Z} + \boldsymbol{\mu}$. Dann ist $\boldsymbol{\Sigma} := AA'$ die Matrix der Kovarianzen zwischen den Komponenten von \mathbf{X} .*

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

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

wegen

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

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

Es sei $\mathbf{X} = (X_1, \dots, X_n)'$ ein zufälliger Vektor, dessen Komponente X_j die Ausprägung einer messbaren Variablen. Weiter sei $\mu_j = \mathbb{E}(X_j)$ der Erwartungswert von X_j , und $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$ sei der Vektor der Erwartungswerte. $\boldsymbol{\Sigma}$ sei die (n, n) -Matrix der Varianzen und Kovarianzen der X_j . Die Variablen variieren zufällig und nicht notwendig unabhängig voneinander. $f(X_1, \dots, X_n)$ sei die gemeinsame Dichtefunktion, die die zufällige Variation charakterisiert. f ist insbesondere die n -dimensionale Normal- oder Gaußverteilung, wenn

$$f(X_1, \dots, X_n) = A \exp \left[-\frac{1}{2} (\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \right], \quad A = \frac{1}{(2\pi)^{\frac{n}{2}} |\boldsymbol{\Sigma}^{-1}|} \quad (3.66)$$

mit $-\infty < X_j < \infty$ für alle j . A ist eine Normierungskonstante. Eine Herleitung dieser Formel findet man u.a. im Skriptum *Die n-dimensionale Normalverteilung*⁵².

In Bezug auf die spaltenzentrierte Datenmatrix X_D (X_D , damit es nicht zu Verwechslungen mit den zufälligen Veränderlichen X kommt) korrespondieren die Zeilenvektoren $\tilde{\mathbf{x}}_i$ zu den Vektoren $X - \boldsymbol{\mu}$. Die $\tilde{\mathbf{x}}_i$ sind Elemente einer Stichprobe aus der Menge der kontinuierlich variierend gedachten $X - \boldsymbol{\mu}$. Betrachtet man die $\tilde{\mathbf{x}}_i$ als Realisierungen des multivariat Gauss-verteilten Zufallsvektors $\tilde{\mathbf{x}}$, so kann man kürzer $\tilde{\mathbf{x}}'\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{x}}$ statt $(X - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(X - \boldsymbol{\mu})$ schreiben. Die Matrix C erscheint als Schätzung für die Varianz-Kovarianzmatrix $\boldsymbol{\Sigma}$.

Im Abschnitt 3.2 wurde gezeigt, dass $C = X'X$ eine Menge von Ellipsoiden definiert, deren Hauptachsen der maximalen Ausdehnung, der zweitmaximalen Ausdehnung etc entsprechen; jeder einen Fall repräsentierende Punkt liegt auf einer dieser Ellipsoide. Die Definition der n -dimensionalen Gauß-Verteilung zeigt, dass $f(x_1, \dots, x_n) = k$ eine Konstante genau dann, wenn

$$(X - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(X - \boldsymbol{\mu}) = k_0 \quad (3.67)$$

k_0 eine Konstante. Das Ellipsoid ist ein Ort gleicher Wahrscheinlichkeit. Die in Abschnitt ?? betrachteten Ellipsoide waren aber durch

$$(X - \boldsymbol{\mu})'\boldsymbol{\Sigma}(X - \boldsymbol{\mu}) = \tilde{\mathbf{x}}'C\tilde{\mathbf{x}} = k_1, \quad \boldsymbol{\Sigma} = C \quad (3.68)$$

definiert; $\tilde{\mathbf{x}} = (X_1 - \mu_1, \dots, X_n - \mu_n)'$, wobei der Index i für einen bestimmten Fall unterdrückt wurde, $\tilde{\mathbf{x}}$ wird hier als Zufallsvektor aufgefasst. Die Beschreibung der Konfiguration anhand dieser Ellipsoide gilt unabhängig von der Wahrscheinlichkeitsverteilung der Daten, sie bezieht sich auf geometrische Aspekte der Konfiguration: ihre Ausdehnungen und ihre Orientierung. Die Konfiguration muss selbst nicht ellipsoid sein (vergl. Abbildung 14.)

Der Unterschied zwischen den beiden Definitionen liegt in der definierenden Matrix: bei der Gauß-Verteilung ist es $\boldsymbol{\Sigma}^{-1}$, bei den nur beschreibenden Ellipsoiden ist es die Matrix $\boldsymbol{\Sigma}$. Es läßt sich leicht zeigen, dass die Orientierung der Ellipsoide in beiden Fällen identisch ist. So habe $\boldsymbol{\Sigma}$ die Eigenvektoren V und die zugehörigen Eigenwerte $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, d.h. es gelte $\boldsymbol{\Sigma}V = V\Lambda$. Dann folgt $\boldsymbol{\Sigma} = V\Lambda V'$, und nach den Regeln für das Rechnen mit Inversen folgt dann, sofern die Inverse $\boldsymbol{\Sigma}^{-1}$ überhaupt existiert,

$$\boldsymbol{\Sigma}^{-1} = (V\Lambda V')^{-1} = (V')^{-1}\Lambda^{-1}V^{-1} = V\Lambda^{-1}V', \quad (3.69)$$

denn wegen der Orthonormlität von V hat man $V^{-1} = V'$, $(V')^{-1} = (V^{-1})^{-1} = V$. Die Inverse $\boldsymbol{\Sigma}^{-1}$ hat also dieselben Eigenvektoren wie $\boldsymbol{\Sigma}$, die die Orientierung der Ellipsoide bestimmen, so dass die Orientierung der Ellipsoide identisch sein muss. Die Ellipsoide unterscheiden sich nur durch die Skalierung der Achsen: einmal durch Λ , das andere mal durch Λ^{-1} .

⁵²<http://www.uwe-mortensen.de/2dnormalverteilungA.pdf>

Wie schon in Abschnitt ?? angemerkt wurde, bedeutet die Existenz der Ellipsoide $\tilde{\mathbf{x}}'C\tilde{\mathbf{x}} = k_i$ nicht die Gauß-Verteilung der Daten, denn es wird ja nichts über die Häufigkeitsverteilung der $\tilde{\mathbf{x}}$ bzw. über die Wahrscheinlichkeitsdichte $f(X_1, \dots, X_n)$ ausgesagt.

3.3 Eigenvektoren und Eigenwerte nichtsymmetrischer Matrizen

3.3.1 Der allgemeine Fall

Der Begriff des Eigenvektors und der des zugehörigen Eigenwerts ergab sich in Abschnitt 3.2 bei der Betrachtung einer Koordinatentransformation auf eine natürliche Art und Weise für den Spezialfall symmetrischer Matrizen. Für nichtsymmetrische quadratische Matrizen können ebenfalls Eigenvektoren existieren, die aber nicht notwendig reell sind; so sei A eine orthonormale Matrix. Das Produkt von A mit einem Vektor \mathbf{x} liefert einen Vektor \mathbf{y} , der sich von \mathbf{x} möglicherweise durch eine Rotation unterscheidet: so sei

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

Dann ist

$$A\mathbf{x} = x_1 \begin{pmatrix} \cos \phi \\ \sin \phi \end{pmatrix} + x_2 \begin{pmatrix} -\sin \phi \\ \cos \phi \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \mathbf{y},$$

und \mathbf{y} ist nur parallel zu \mathbf{x} für diejenigen Werte von ϕ , für die $\cos \phi = 1$ und $\sin \phi = 0$ ist, also z.B. für $\phi = 0$, so dass $A = I$ mit den Spaltenvektoren $(1, 0)'$ und $(0, 1)$. Dies ist der gewissermaßen triviale Fall, bei dem gar keine Rotation erzeugt wird. Man findet allerdings komplexwertige Eigenvektoren mit zugehörigen komplexwertigen Eigenwerten, – für $\phi = \pi/4$ etwa findet man die Eigenvektoren $(i, 1)'$ und $(-i, 1)'$ mit den Eigenwerten $(1+i)/\sqrt{2}$ und $(1-i)/\sqrt{2}$, mit $i = \sqrt{-1}$, wie man durch Nachrechnen bestätigt. Wie komplexe Eigenvektoren und -werte zu deuten sind, wird später noch besprochen werden.

Charakteristische Gleichung einer Matrix: Es sei A eine (n, n) -Matrix und $\mathbf{u} \in \mathbb{R}^n$ derart, dass $A\mathbf{u} = \lambda\mathbf{u}$. Dann folgt

$$(A - \lambda I)\mathbf{u} = \vec{0}, \tag{3.70}$$

I die (n, n) -Einheitsmatrix. Diese Gleichung beschreibt ein homogenes Gleichungssystem mit den Komponenten u_i des Vektors \mathbf{u} als Unbekannten: In ausgeschriebener Form hat man

$$(a_{11} - \lambda)u_1 + a_{12}u_2 + \dots + a_{1n}u_n = 0 \tag{3.71}$$

$$a_{21}u_1 + (a_{22} - \lambda)u_2 + \dots + a_{2n}u_n = 0 \tag{3.72}$$

$$\vdots \tag{3.73}$$

$$a_{n1}u_1 + a_{n2}u_2 + \dots + (a_{nn} - \lambda)u_n = 0 \tag{3.74}$$

Solche Gleichungssysteme haben nur dann mindestens eine nicht-triviale Lösung (d.h. eine Lösung, die nicht gleich dem Nullvektor $\vec{0}$ ist), wenn die Koeffizientenmatrix nicht vollen Rang hat, d.h. wenn ihre Determinante verschwindet, so dass

$$|A - \lambda I| = \begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ & & \ddots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{vmatrix} = 0 \quad (3.75)$$

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

$$\begin{aligned} |A - \lambda I| &= (-1)^n [\lambda^n - \beta_1 \lambda^{n-1} + \beta_2 \lambda^{n-2} + \cdots + (-1)^{n-1} \beta_{n-1} \lambda + (-1)^n \beta_n] \\ &= 0. \end{aligned} \quad (3.76)$$

Diese Gleichung heißt *charakteristische Gleichung* der Matrix A , wenn man den Faktor $(-1)^n$ wegläßt, und das Polynom auf der rechten Seite heißt *charakteristisches Polynom* von A . Die Gleichung hat, wie aus der Theorie der Polynome bekannt ist, insgesamt n Lösungen für λ , die Nullstellen von A , die wiederum gleich den möglichen Eigenwerten $\lambda_1, \dots, \lambda_n$ von A sind. Die Eigenwerte müssen nicht alle verschieden sein und können durch komplexe Zahlen $\lambda = \alpha + i\beta$, $i = \sqrt{-1}$, gegeben sein. Es gilt dabei

Satz 3.14 *Ist ein Eigenwert λ der quadratischen Matrix mit reellen Elementen komplex, so existiert ein zweiter Eigenwert $\bar{\lambda}$, der zu λ konjugiert komplex ist, dh gilt $\lambda = \alpha + i\beta$, so ist auch $\bar{\lambda} = \alpha - i\beta$ ein Eigenwert von A .*

Beweis: Es ist $(A - \lambda I)\mathbf{u} = 0$. Der Übergang zu konjugiert komplexen Zahlen führt zu $(\bar{A} - \bar{\lambda}I)\bar{\mathbf{u}} = 0$. Aber A ist als reell vorausgesetzt worden, also folgt

$$(A - \bar{\lambda}I)\bar{\mathbf{u}} = 0, \quad (3.77)$$

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

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

$$A\mathbf{u} = \lambda\mathbf{u}. \quad (3.78)$$

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

Es sei $B = A'$; gilt

$$B\mathbf{v} = \mu\mathbf{v}, \quad (3.79)$$

so ist \mathbf{v} ein Eigenvektor von B und μ der zugehörige Eigenwert. Es ist

$$(B\mathbf{v})' = \mathbf{v}'B' = \mathbf{v}'A = \mu\mathbf{v}'.$$

\mathbf{v} heißt auch *Linkseigenvektor* von A ; \mathbf{u} in (3.78) heißt dementsprechend *Rechtseigenvektor*. Wegen (3.79) übertragen sich alle Aussagen über Rechtseigenvektoren auf Linkseigenvektoren, was allerdings nicht bedeutet, dass Links- und Rechtseigenvektoren notwendig identisch sind. Notwendig identisch sind sie nur für den Spezialfall symmetrischer Matrizen. Denn wenn $A' = B = A$ gilt, so folgt aus (3.79) $B\mathbf{v} = A\mathbf{v} = \mu\mathbf{v}$, d.h. ein gegebener Linkseigenvektor entspricht einem Rechtseigenvektor. Im Falle $A' \neq A$ gilt der

Satz 3.15 *Es sei A eine quadratische, nicht-symmetrische Matrix. Es gelte einerseits $\mathbf{v}'A = \mu\mathbf{v}'$, andererseits $A\mathbf{u} = \lambda\mathbf{u}$ mit $\lambda \neq \mu$. Dann folgt $\mathbf{u}'\mathbf{v} = 0$, d.h. die Links- und Rechtseigenvektoren sind orthogonal zueinander.*

Beweis: Multiplikation von $\mathbf{v}'A = \mu\mathbf{v}'$ von rechts mit \mathbf{u} und von $A\mathbf{u} = \lambda\mathbf{u}$ von links mit \mathbf{v}' liefert

$$\begin{aligned}\mathbf{v}'A\mathbf{u} &= \mu\mathbf{v}'\mathbf{u} = \lambda\mathbf{v}'\mathbf{u} \\ \mathbf{v}'A\mathbf{u} &= \lambda\mathbf{v}'\mathbf{u} = \mu\mathbf{v}'\mathbf{u}\end{aligned}$$

Da $\mathbf{v}'A\mathbf{u} - \mathbf{v}'A\mathbf{u} = 0$ folgt $\lambda\mathbf{v}'\mathbf{u} - \mu\mathbf{v}'\mathbf{u} = (\lambda - \mu)\mathbf{v}'\mathbf{u} = 0$, woraus wegen $\lambda - \mu \neq 0$ die Behauptung $\mathbf{v}'\mathbf{u} = 0$ folgt, d.h. \mathbf{u} und \mathbf{v} sind orthogonal. \square

Im Falle nicht-symmetrischer Matrizen sind Links- und Rechtseigenvektoren also verschieden, da sie ja orthogonal zueinander sind. Dieses Resultat bedeutet nicht, dass auch die Rechts- und Linkseigenvektoren untereinander orthogonal zueinander sind. Aber die Gültigkeit des folgenden Satzes läßt sich zeigen:

Satz 3.16 *Es sei A eine nicht-symmetrische, quadratische Matrix mit mehr als einem Rechtseigenvektor. Die Rechtseigenvektoren sind linear unabhängig, sofern die zugehörigen Eigenwerte verschieden sind.*

Beweis: Es seien \mathbf{u}_1 und \mathbf{u}_2 zwei Rechtseigenvektoren von A mit zugehörigen Eigenwerten $\mu \neq \lambda$. Angenommen, sie seien linear abhängig; dann existieren Koeffizienten a_1 und a_2 ungleich Null derart, dass

$$a_1\mathbf{u}_1 + a_2\mathbf{u}_2 = 0 \tag{3.80}$$

Multiplikation von links mit A führt dann auf $a_1A\mathbf{u}_1 + a_2A\mathbf{u}_2 = 0$, d.h. auf

$$a_1\mu\mathbf{u}_1 + a_2\lambda\mathbf{u}_2 = 0. \tag{3.81}$$

Multipliziert man (3.80) mit λ und subtrahiert die entstehende Gleichung dann von (3.81), so erhält man

$$a_1(\lambda - \mu)\mathbf{u}_1 = 0,$$

woraus wegen der Voraussetzung $\lambda - \mu \neq 0$ sofort $a_1 = 0$ folgt. Auf analoge Weise folgt $a_2 = 0$, d.h. \mathbf{u}_1 und \mathbf{u}_2 sind linear unabhängig.

Diese Aussage gilt für irgendzwei Rechtseigenvektoren von A . Hat man also insgesamt drei Eigenvektoren, so sind sie paarweise linear unabhängig, so dass man sagen könnte, sie seien insgesamt linear unabhängig. Das Argument ist aber intuitiv, und ein strenger Beweis ist einer intuitiven Betrachtung stets vorzuziehen. Dieser ergibt sich durch das Prinzip der vollständigen Induktion. Es gebe also $r > 2$ linear unabhängige Eigenvektoren, so dass

$$a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + \cdots + a_r \mathbf{u}_r = 0 \text{ genau dann, wenn } a_1 = \cdots = a_r = 0.$$

Es ist zu zeigen, dass dann auch $r + 1$ Eigenvektoren linear unabhängig sind, so dass

$$a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + \cdots + a_r \mathbf{v}_r + a_{p+1} \mathbf{v}_{p+1} = 0 \quad (3.82)$$

gilt mit $a_1 = a_2 = \cdots = a_{p+1} = 0$ als einziger Lösung. Da die \mathbf{u}_j Eigenvektoren sind, gilt $A\mathbf{u}_j = \lambda_j \mathbf{u}_j$. Multiplikation von (3.82) mit A führt dann unter Berücksichtigung dieser Beziehung auf die Gleichung

$$a_1 \lambda_1 \mathbf{u}_1 + a_2 \lambda_2 \mathbf{u}_2 + \cdots + a_r \lambda_r \mathbf{u}_r + a_{p+1} \lambda_{p+1} \mathbf{u}_{p+1} = 0. \quad (3.83)$$

Multipliziert man (3.82) mit λ_{p+1} und subtrahiert die Gleichung dann von (3.83), so erhält man

$$a_1(\lambda_1 - \lambda_{p+1})\mathbf{u}_1 + \cdots + a_p(\lambda_p - \lambda_{p+1})\mathbf{u}_p = 0,$$

und wegen der vorausgesetzten linearen Unabhängigkeit der \mathbf{v}_j , $1 \leq j \leq r$ hat man einerseits $a_1 = \cdots = a_p = 0$ und wegen der ebenso vorausgesetzten Ungleichheit der λ_j folgt dann aus (3.83) $a_{p+1} \lambda_{p+1} \mathbf{u}_{p+1} = 0$. Daraus folgt wegen $\lambda_{p+1} \neq 0$ dann $a_{p+1} = 0$, so dass die $\mathbf{u}_1, \dots, \mathbf{u}_{p+1}$ ebenfalls linear unabhängig sind. \square

Der Beweis gilt für eine beliebige quadratische Matrix, also auch für A' und damit für die Rechtseigenvektoren von A' , die aber die Linkseigenvektoren von A sind, so dass deren lineare Unabhängigkeit ebenfalls nachgewiesen ist. Gilt der Spezialfall $A' = A$, ist A also symmetrisch, so folgt sofort, dass in diesem Fall die Linkseigenvektoren gleich den Rechtseigenvektoren sind, und wie bereits gezeigt wurde gilt dann nicht nur die lineare Unabhängigkeit der Eigenvektoren, sondern darüber hinaus auch die Orthogonalität der Eigenvektoren.

Im Folgenden werden nur die Rechtseigenvektoren betrachtet und es wird der Kürze wegen nur von Eigenvektoren geredet; alle Aussagen übertragen sich auf die Linkseigenvektoren. Zunächst soll die Beziehung zwischen einer quadratischen Matrix A und ihren Eigenvektoren und Eigenwerten auf eine andere Art dargestellt werden, die Aufschluss über die Anzahl und Art der Eigenvektoren und -eigenwerte gibt.

Ähnliche Matrizen: Es sei V die Matrix der Eigenvektoren einer beliebigen quadratischen Matrix A . Da die Spaltenvektoren von V linear unabhängig sind,

folgt die Existenz der zu V inversen Matrix V^{-1} . Aus $AV = V\Lambda$, Λ die Matrix der zugehörigen Eigenwerte, folgt dann durch Multiplikation von rechts mit V^{-1}

$$A = V\Lambda V^{-1}. \quad (3.84)$$

Durch Multiplikation von rechts mit V und von links mit V^{-1} erhält man hieraus

$$V^{-1}AV = \Lambda. \quad (3.85)$$

Mit diesen beiden Gleichungen hat man einen Spezialfall einer Beziehung zwischen Matrizen, die durch die folgende Definition charakterisiert wird:

Definition 3.5 *Es seien A und B zwei (n, n) -Matrizen und es existiere eine nichtsinguläre Matrix C derart, dass*

$$B = C^{-1}AC \quad (3.86)$$

gilt. Dann heißen A und B ähnlich.

Offenbar bedeuten (3.84) und (3.85), dass A und Λ ähnlich sind. Da Λ eine Diagonalmatrix ist, heißt A auch *diagonalisierbar*. Damit eine (n, n) -Matrix diagonalisierbar ist, muss also die Matrix V^{-1} existieren, und diese Matrix existiert, wenn A vollen Rang hat, denn dann hat A n linear unabhängige Eigenvektoren.

Komplexe Eigenwerte und -vektoren: Es ist bisher stets vorausgesetzt worden, dass für eine gegebene quadratische Matrix Eigenwerte und -vektoren existieren. Die Frage ist aber, ob für eine beliebige quadratische Matrix überhaupt Eigenvektoren existieren müssen. Gegeben sei etwa die Matrix

$$A(\phi) = \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix} \quad (3.87)$$

Ein Eigenvektor \mathbf{v} von A muss die Bedingung $A\mathbf{v} = \mathbf{w} = \lambda\mathbf{v}$ erfüllen, d.h. der Vektor \mathbf{w} muss parallel zu \mathbf{v} sein, er darf sich nur in der Länge von \mathbf{v} unterscheiden. Aber für $\phi \neq 0$ und $\phi \neq \pi$ bewirkt A eine Rotation des Vektors \mathbf{v} , \mathbf{w} kann also nicht parallel zu \mathbf{v} sein. A hat mit Ausnahme spezieller ϕ -Werte zumindest keinen reellen Eigenvektor. Um die Situation allgemein zu klären, geht man noch einmal auf die Gleichung (3.78) zurück: so dass sich die Eigenwerte von A als Nullstellen des Polynoms ergeben. Speziell für die Matrix (3.87) erhält man

$$|A - \lambda I| = 4\lambda^2 - 4\lambda \cos \phi + 1 = 0; \quad (3.88)$$

auf die Herleitung des Polynoms wird hier verzichtet, da es an dieser Stelle nur auf die Implikationen von (3.88) ankommt. Man findet die Nullstellen

$$\lambda_1 = \frac{1}{2}(\cos \phi + i \sin \phi), \quad \lambda_2 = \frac{1}{2}(\cos \phi - i \sin \phi), \quad i = \sqrt{-1} \quad (3.89)$$

Die in (3.87) definierte Matrix A hat also zwei komplexe Eigenwerte, reelle Eigenwerte ergeben sich nur für solche ϕ -Werte, für die $\sin \phi = 0$ ist, also etwa für $\phi = 0$, wenn gar keine Rotation der Vektoren stattfindet, oder für $\phi = \pi/2$, wenn eine Rotation um 90° stattfindet.

Es ist also möglich, dass für eine beliebig gewählte quadratische Matrix keine reellen Eigenwerte existieren, dass man aber komplexwertige Eigenwerte finden kann, die als Paare konjugiert komplexer Zahlen auftreten⁵³. Nun hätte man noch gerne die zugehörigen Eigenvektoren bestimmt. Für A findet man zwei:

$$\mathbf{u}_1 = \begin{pmatrix} -i \\ 1 \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} i \\ 1 \end{pmatrix}. \quad (3.90)$$

Natürlich ergibt sich die Frage der Deutung von komplexen Eigenwerten und Eigenvektoren. Diese treten etwa bei der Analyse dynamischer Systeme und dementsprechend bei allgemeinen Diskussionen von Zeitreihenproblemen auf. Hier sollen zunächst noch bestimmte Typen von Matrizen eingeführt werden.

Typen von Matrizen: In der multivariaten Statistik spielen symmetrische Matrizen mit reellen Elementen eine zentrale Rolle, es ist aber trotzdem sinnvoll, auch den allgemeinen Fall einer Matrix mit möglicherweise komplexwertigen Elementen zu betrachten.

Sind die Elemente einer Matrix A komplex, d.h. von der Form $z = x + iy$ mit $i = \sqrt{-1}$, so heißt \bar{A} die zu A konjugierte Matrix; die Elemente von A enthalten die zu z konjugierten komplexen Elemente $\bar{z} = x - iy$. Sind nur die Imaginärteile iy der Elemente einer Matrix A von Null verschieden, so heißt A *imaginär*; in diesem Fall gilt $A = -\bar{A}$. Die Transponierte \bar{A}' einer Matrix A heißt die mit A *assoziierte* Matrix.

Für symmetrische Matrizen gilt $A' = A$, d.h. $a_{ij} = a_{ji}$ für alle i, j . Gilt für eine Matrix die Aussage $a_{ij} = -a_{ji}$, so heißt A *schief-symmetrisch*.

Ein wichtiger Fall ist durch die Gleichung

$$A = \bar{A}' \quad (3.91)$$

definiert; in diesem Fall heißt A *hermitesch*⁵⁴. Ist $A = \bar{A}$, so sind die Elemente von A alle reell und (3.91) bedeutet einfach, dass A symmetrisch ist. Da der reelle Fall ein Spezialfall ist, gelten alle Aussagen über hermitesche Matrizen auch für reelle symmetrische Matrizen, so dass es Sinn macht, bestimmte Aussagen allgemein für hermitesche Matrizen zu machen.

3.3.2 Mehrfache Eigenwerte

Es sei A eine (n, n) -Matrix und es werde die Gleichung $A\mathbf{v} = \lambda\mathbf{v}$ betrachtet: λ ist ein Eigenwert von A und \mathbf{v} der zugehörige Eigenvektor. Es gibt maximal

⁵³Zwei komplexe Zahlen z und \bar{z} heißen konjugiert komplex, wenn sie sich nur im Vorzeichen des Imaginärteils unterscheiden, wenn also $z = x + iy$ und $\bar{z} = x - iy$ gilt.

⁵⁴Nach dem französischen Mathematiker Charles Hermite (1822 – 1901)

n verschiedene Eigenwerte, d.h. es ist möglich, dass einige Eigenwerte mehrfach vorkommen (multiple Eigenwerte, *multiplicity, repeated eigenvalues*). Ein einfaches Beispiel ist die (m, n) -Identitätsmatrix I : für jeden n -dimensionalen Vektor \mathbf{x} gilt $I\mathbf{x} = \mathbf{x}$, d.h. jeder Vektor \mathbf{x} ist ein Eigenvektor von I , und alle haben den Eigenwert $\lambda = 1$.

Definition 3.6 *Es sei V ein Vektorraum und es sei $V_\lambda = \{\mathbf{v} \in V | A\mathbf{v} = \lambda\mathbf{v}\}$. Dann heißt V_λ der Eigenraum von A zum Eigenwert λ .*

Bemerkung: Aus $A\mathbf{v} = \lambda\mathbf{v}$ folgt $(A - \lambda I)\mathbf{v} = \vec{0}$. Diese Gleichung ist ein lineares Gleichungssystem in \mathbf{v} , d.h. in den Komponenten von \mathbf{v} als Unbekannten. Bekanntlich (s. Definition 2.2, Seite 71, und Definition 3.10, Seite 116) heißt die Menge der Vektoren \mathbf{x} , die der Gleichung $A\mathbf{x} = \vec{0}$ genügen, der Kern von A : $\text{kern}(A) = \{\mathbf{x} | A\mathbf{x} = \vec{0}\}$. Dementsprechend ist $\text{kern}(A - \lambda I) = V_\lambda$, d.h. der Eigenraum V_λ ist der Kern von $(A - \lambda I)$. \square

Da zu jedem Eigenwert λ ein Eigenvektor \mathbf{v} korrespondiert, enthält V_λ zumindest ein Element. Da mit \mathbf{v} auch $a\mathbf{v}$, $a \neq 0$ ein Eigenvektor ist, ist V_λ zumindest ein 1-dimensionaler Teilraum von V . Die Frage ist, ob V_λ stets ein Teilraum von V ist. Man sieht dies leicht ein: sind $\mathbf{v} \neq \mathbf{w}$ aus V_λ , so ist mit $a \in \mathbb{R}$, $b \in \mathbb{R}$ auch $\mathbf{u} = a\mathbf{v} + b\mathbf{w} \in V_\lambda$. Denn wegen $A\mathbf{v} = \lambda\mathbf{v}$, $A\mathbf{w} = \lambda\mathbf{w}$ hat man auch

$$A(a\mathbf{v} + b\mathbf{w}) = aA\mathbf{v} + bA\mathbf{w} = a\lambda\mathbf{v} + b\lambda\mathbf{w} = \lambda(a\mathbf{v} + b\mathbf{w}) = \lambda\mathbf{u}.$$

Die Eigenwerte von A sind die Nullstellen des Polynoms, das durch die Determinante

$$P_A = |A - \lambda I| = 0$$

definiert ist. Mehrfache Eigenwerte gibt es demnach dann, wenn dieses Polynom mehrfache Nullstellen hat. Man kann nun zeigen, dass, wenn λ eine m -fache Nullstelle von P_A ist, dann die Dimension des Eigenraums V_λ kleiner, höchstens gleich m ist, d.h. es gibt maximal m linear unabhängige Vektoren in V_λ (der Beweis für diese Aussage wird hier übergangen (vergl. Fischer (1984), Kapitel 4).

Der Begriff des Hauptraums ist eine Verallgemeinerung des Begriffs des Eigenraums:

Definition 3.7 *Die Matrix A definiere eine Abbildung f des Vektorraums V in sich selbst, d.h. $f : V \rightarrow V$, und λ sei ein Eigenwert von A (d.h. von f), und $r(\lambda)$ sei die algebraische Vielfachheit von λ . Der Kern der r -fachen Hintereinanderschaltung von $A - \lambda I$ heißt Hauptraum zu λ $H(A, \lambda)$*

$$H(A, \lambda) = \{\mathbf{v} \in V | (A - \lambda I)^r(\mathbf{v}) = \mathbf{0}\}. \quad (3.92)$$

Die Elemente von $H(A, \lambda)$ heißen die Hauptvektoren. $\mathbf{v} \in V$ ist Hauptvektor der Stufe p , wenn $(A - \lambda I)^p \mathbf{v} \neq \vec{0}$.

Anmerkung: Alle Eigenvektoren sind Hauptvektoren der Stufe $p = 1$. □

Der in der folgenden Definition eingeführte Begriff des invarianten Teilraums ist eine weitere Verallgemeinerung des Begriffs des Eigenraums:

Definition 3.8 Die Matrix A definiere eine Abbildung eines Vektorraums in sich selbst: $f: V \rightarrow V$, und es sei $U \subseteq V$. Gilt $f(U) \subseteq U$, d.h. ist die Menge der Vektoren Au wieder eine Teilmenge von U , so heißt U invarianter Teilraum von V , oder einfach f -invariant⁵⁵.

Anmerkung: Alle Eigenräume sowie alle Haupträume sind invariante Teilräume. □

3.3.3 Das generalisierte Eigenvektorproblem

Eine Reihe von statistischen Fragestellungen führt auf das *generalisierte Eigenvektorproblem*, so etwa die Frage, ob zwei, an m "Fällen" erhobene Datensätze die gleiche oder eine ähnliche latente Struktur haben oder nicht. So kann man an m Personen (Patienten, etc) Messungen von n Variablen vor und nach einer Intervention (etwa einer Therapie) erheben. Die Frage nach einer Veränderung durch die Intervention (Therapieerfolg) führt auf die Frage, ob sich Vorher- und Nachhermessungen systematisch voneinander unterscheiden. Die Berechnung der Kanonischen Korrelationen kann hier zu Antworten führen. An dieser Stelle kann nur auf die rein formalen Aspekte derartiger Methoden eingegangen werden.

Definition 3.9 Es seien A und B symmetrische, positiv semidefinite Matrizen. Dann repräsentiert

$$A\mathbf{w} = \lambda B\mathbf{w} \tag{3.93}$$

das generalisierte Eigenvektorproblem.

Der generalisierte Rayleigh-Quotient: Der *generalisierte Rayleigh-Quotient* ist durch

$$\rho(\mathbf{w}) = \frac{\mathbf{w}'A\mathbf{w}}{\mathbf{w}'B\mathbf{w}} \tag{3.94}$$

definiert. Wie beim schon behandelten Fall $B = I$ die Einheitsmatrix ergibt sich die Frage nach dem maximalen Wert von $\rho(\mathbf{w})$. In Abschnitt 3.6.2 wird eine Anwendung dieses Quotienten vorgestellt.

Dazu werde vorausgesetzt, dass die Inverse B^{-1} existiert. Da B als symmetrisch und positiv semidefinit vorausgesetzt worden ist, kann man die Wurzel $B^{1/2} = P\Lambda^{1/2}P'$ von B bestimmen, – offenbar ist

$$B^{1/2}B^{1/2} = B = P\Lambda^{1/2}P'P\Lambda^{1/2}P' = P\Lambda P',$$

⁵⁵oder invariant unter f

denn $P'P = I$ die Einheitsmatrix, und P die Matrix der Eigenvektoren von B . Es sei $\mathbf{v} = B^{1/2}\mathbf{w}$. Dann ist $\mathbf{w} = B^{-1/2}\mathbf{v}$ und man erhält

$$\rho(\mathbf{w}) = \frac{\mathbf{w}'A\mathbf{w}}{\mathbf{w}'B\mathbf{w}} = \frac{\mathbf{v}'B^{-1/2}AB^{-1/2}\mathbf{v}}{\mathbf{v}'\mathbf{v}} = \frac{\mathbf{v}'M\mathbf{v}}{\mathbf{v}'\mathbf{v}}. \quad (3.95)$$

$M = B^{-1/2}AB^{-1/2}$ ist symmetrisch (warum?) und die Maximierung des generalisierten Rayleigh-Quotienten ist auf den einfachen Fall zurückgeführt worden. Der Vektor \mathbf{w} des ursprünglichen Problems ergibt sich aus der Lösung für (3.95) gemäß $\mathbf{w} = B^{-1/2}\mathbf{v}$.

Die Gleichung (3.93) führt durch Multiplikation von links mit B^{-1} auf

$$B^{-1}A\mathbf{w} = \lambda\mathbf{w}, \quad (3.96)$$

d.h. \mathbf{w} ist ein Eigenvektor der nicht-symmetrischen Matrix $BV^{-1}A$, und λ ist der zugehörige Eigenwert. Multipliziert man diese Gleichung von links mit $B^{1/2}$, so erhält man

$$B^{1/2}B^{-1}A\mathbf{w} = B^{-1/2}A\mathbf{w} = \lambda B^{1/2}\mathbf{w}.$$

und nochmalige Multiplikation von links mit $B^{-1/2}$ führt zu

$$B^{-1/2}AB^{-1/2}\mathbf{v} = \lambda\mathbf{v}. \quad (3.97)$$

Damit hat man mit (3.97) ein Eigenwert- und Eigenvektorproblem der bekannten Art für symmetrische Matrizen.

Die Eigenvektoren \mathbf{v}_j symmetrischer Matrizen sind bekanntlich orthonormal. Die Lösung für den generalisierten Rayleigh-Quotienten durch die Lösung für den gewöhnlichen Rayleigh-Quotienten in einem *transformierten Raum* gegeben ist. Man kommt damit zu der Aussage (Shaw-Taylor & Christianini (2004), p. 162)

Satz 3.17 *Ein beliebiger Vektor \mathbf{v} kann als Linearkombination der \mathbf{w}_j , $j = 1, \dots, k$ angeschrieben werden. Für die Eigenvektoren des generalisierten Eigenvektorproblems $A\mathbf{w} = \lambda B\mathbf{w}$ gelten die Relationen*

$$\mathbf{w}'_i B \mathbf{w}_j = \delta_{ij}, \quad \mathbf{w}'_i A \mathbf{w}_j = \delta_{ij} \lambda_i, \quad \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (3.98)$$

Beweis: Es war $\mathbf{v} = B^{1/2}\mathbf{w}$, und als Lösungen von (3.97) sind die \mathbf{v}_j orthonormal. Es $i \neq j$ und $\lambda_j \neq 0$. Dann folgt, wegen $A\mathbf{w} = \lambda B\mathbf{w}$ (vergl. (3.96)) und damit $B\mathbf{w} = (1/\lambda)A\mathbf{w}$,

$$0 = \mathbf{v}'_i \mathbf{v}_j = \mathbf{w}'_i B^{1/2} B^{1/2} \mathbf{w}_j = \mathbf{w}'_i B \mathbf{w}_j = \frac{1}{\lambda_j} \mathbf{w}'_i A \mathbf{w}_j;$$

nach (3.93) gilt ja $A\mathbf{w} = \lambda B\mathbf{w}$ und deshalb $(1/\lambda_i)A\mathbf{w} = B\mathbf{w}$. Damit gilt (3.98) für den Fall $i \neq j$.

Nun sei $i = j$; es ist $1 = \mathbf{v}_i' \mathbf{v}_i = \mathbf{w}_i' B^{1/2} B^{1/2} \mathbf{w}_i$, also

$$\lambda_i = \lambda_i \mathbf{v}_i' \mathbf{v}_i = \lambda_i \mathbf{w}_i' B^{1/2} B^{1/2} \mathbf{w}_i = \lambda_i \mathbf{w}_i' B \mathbf{w}_i = \mathbf{w}_i A \mathbf{w}_i,$$

und dies ist (3.98) für den Fall $i = j$. \square

Die Maximierung von (3.98) (Maximierung unter Nebenbedingungen, S. Anhang) führt auf die Gleichung (3.96).

Satz 3.18 *Für den generalisierten Rayleigh-Quotienten gilt*

$$\rho_1 \leq \rho \leq \rho_2, \quad (3.99)$$

und ρ_1, ρ_2 sind durch die Eigenvektoren definiert, die zum kleinsten bzw. größten Eigenwert korrespondieren.

Der Beweis ergibt sich analog zum Beweis für den Rayleigh-Quotienten für symmetrische Matrizen (Satz von Courant-Fisher, Seite 178).

Satz 3.19 *Gilt $A\mathbf{v} = \lambda B\mathbf{v}$ und sind λ und \mathbf{v} die Eigenwerte und Eigenvektoren für den generalisierten Rayleigh-Quotienten, so kann A gemäß*

$$A = \sum_{j=1}^r \lambda_j B \mathbf{v}_j (B \mathbf{v}_j)' \quad (3.100)$$

zerlegt werden.

Beweis: Für eine beliebige symmetrische Matrix C mit der Matrix P der Eigenvektoren und der Diagonalmatrix $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ der Eigenwerte gilt bekanntlich $C = \sum_j \lambda_j \mathbf{p}_j \mathbf{p}_j'$. Die Matrix $B^{-1/2} A B^{-1/2}$ ist symmetrisch, mithin gilt

$$B^{-1/2} A B^{-1/2} = \sum_{j=1}^r \lambda_j \mathbf{v}_j \mathbf{v}_j'.$$

Multipliziert man von links mit $B^{1/2}$ und von rechts ebenfalls mit $B^{1/2}$, so folgt

$$A = \sum_{j=1}^r \lambda_j B^{1/2} (B^{1/2} \mathbf{v}_j \mathbf{v}_j)' = \sum_j \lambda_j B \mathbf{w}_j (B \mathbf{w}_j)',$$

und das war zu zeigen. \square

3.4 Lineare Gleichungssysteme II

In Abschnitt 1.4, Seite 55 wurde zur Illustration der Darstellbarkeit von Vektoren als Linearkombination anderer Vektoren bereits ein System von zwei Gleichungen mit zwei Unbekannten betrachtet; man kann das System in der Form $X\mathbf{a} = \mathbf{y}$

schreiben, mit $X = [\mathbf{x}_1, \mathbf{x}_2]$ und $\mathbf{a} = [a_1, a_2]$. Die Lösung für die unbekanntenen Koeffizienten a_1 und a_2 ist eindeutig, wenn die Vektoren \mathbf{x}_1 und \mathbf{x}_2 linear unabhängig sind; sind sie linear abhängig, so ist die Lösung nicht mehr eindeutig. Nicht betrachtet wurde der Fall $\mathbf{y} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{x}_3$, wobei \mathbf{y}, \mathbf{x}_1 und \mathbf{x}_3 2-dimensionale Vektoren sind. Dieser Fall liegt vor, wenn man weniger Gleichungen als Unbekannte hat. Die Lösungsgesamtheit hängt dann vom Rang der Matrix $X = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3]$ ab. Im Folgenden wird ein allgemeiner Lösungsansatz vorgestellt, wobei allerdings die Notation geändert wird. Die Matrix X wird in A umbenannt, und der Vektor \mathbf{a} der Unbekannten wird in \mathbf{x} umbenannt, – damit wird die für Gleichungssysteme übliche Schreibweise verwendet.

Es sei also A eine (m, n) -Matrix, \mathbf{x} ein n -dimensionaler Vektor, und \mathbf{y} ein m -dimensionaler Vektor, und es gelte

$$A\mathbf{x} = \mathbf{y}, \quad \mathbf{x} \in \mathbb{R}^n, \quad \mathbf{y} \in \mathbb{R}^m \quad (3.101)$$

Formal bedeutet diese Gleichung, dass \mathbf{y} eine Linearkombination der Spaltenvektoren von A ist oder sein soll. \mathbf{x} ist dann der Vektor der Koeffizienten: gilt $A = [\mathbf{a}_1, \dots, \mathbf{a}_n]$, $\mathbf{x} = (x_1, \dots, x_n)'$, so hat man

$$A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \dots + x_n\mathbf{a}_n = \mathbf{y} \quad (3.102)$$

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: Das Gleichungssystem heißt

homogen, wenn $\mathbf{y} = \vec{0}$,
inhomogen, wenn $\mathbf{y} \neq \vec{0}$.

Definition 3.10 *Es sei A eine (m, 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*

$$\text{kern}(A) = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \vec{0}\} \quad (3.103)$$

$$\mathcal{L}(A) = \{\mathbf{y} \in \mathbb{R}^m \mid A\mathbf{x} = \mathbf{y}\}. \quad (3.104)$$

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

Anmerkungen:

1. $\text{kern}(A)$ ist ein (Teil-)Vektorraum: sind die Vektoren \mathbf{x}_1 und \mathbf{x}_2 Elemente aus $\text{kern}(A)$, so rechnet man leicht nach, dass dann auch $b_1\mathbf{x}_1 + b_2\mathbf{x}_2 \in \text{kern}(A)$ ($b_1, b_2 \in \mathbb{R}$) gilt⁵⁶.

⁵⁶In Definition 5.4, Punkt 4., Seite 183 wird der Begriff des Kerns einer Abbildung eingeführt. Die Matrix A definiert eine Abbildung, und (3.103) definiert damit den Kern einer Abbildung.

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, 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 3.20 *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*

$$\mathcal{L}(A) = \mathcal{L}(\mathbf{q}_1, \dots, \mathbf{q}_r) \quad (3.105)$$

$$\text{kern}(A) = \mathcal{L}(\mathbf{t}_{r+1}, \dots, \mathbf{t}_n), \quad s = \min(m, n) \quad (3.106)$$

$$\text{rg}(\text{kern}(A)) + \text{rg}(\mathcal{L}(A)) = \min(m, n) \quad (3.107)$$

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 (3.105) 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 $c_j \in \mathbb{R}$ geeignet gewählte Koeffizienten sind. Dann gilt

$$A\mathbf{x} = A \sum_{j=1}^n c_j \mathbf{t}_j = \sum_{j=1}^n c_j A\mathbf{t}_j = \sum_{j=1}^r c_j \sigma_j \mathbf{q}_j = \vec{0},$$

(vergl. (3.51), Seite 101), 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 Komplementärtraums sein. Die $\mathbf{t}_{r+1}, \dots, \mathbf{t}_n$ sind eine Basis für diesen Komplementärtraum, so dass man

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

ansetzen kann, und

$$\mathbf{Ax} = A \sum_{j=r+1}^n c_j \mathbf{t}_j = \sum_{j=r+1}^n c_j A \mathbf{t}_j = \sum_{j=r+1}^n c_j \sigma_j \mathbf{q}_j = \vec{0},$$

wegen (3.51), Seite 101, und weil $\sigma_j = 0$ für $j > r$, falls $r < \min(m, n)$. Alle Linearkombinationen von Vektoren $\mathbf{x} \neq \vec{0}$ mit $\mathbf{Ax} = \vec{0}$ sind Linearkombinationen der $\mathbf{t}_{r+1}, \dots, \mathbf{t}_n$, und dies ist die Aussage von (3.106).

Die Gleichung (3.107) 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 3.20 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 Beweis, in dem ein anderer Satz von Basisvektoren verwendet wird, findet man im Anhang, Abschnitt 5.8. \square

Die allgemeine Lösungsmenge wird im folgenden Satz spezifiziert:

Satz 3.21 *Es sei $\mathbf{Ax} = \mathbf{y} \in \mathcal{L}(A)$, $\text{rg}(A) = r$, und insbesondere sei $\mathbf{x} = \mathbf{x}_0$ eine bestimmte Lösung, so dass $\mathbf{Ax}_0 = \mathbf{y}$ 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*

$$\mathcal{L} = \{\mathbf{x}_0 + \mathbf{x} \mid \mathbf{x} \in \mathcal{L}_{n-r}\} \quad (3.108)$$

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 $\mathbf{Ax} = \mathbf{y} \in V_n^r = \mathcal{L}(A)$ folgt sofort die Lösung

$$\mathbf{x} = A^{-1}\mathbf{y}, \quad \mathbf{y} \in \mathcal{L}(A). \quad (3.109)$$

Der Fall $m > n = r$: In diesem Fall gibt es mehr Gleichungen als Unbekannte; das Gleichungssystem ist überbestimmt. Im Allgemeinen wird man keinen

Lösungsvektor \mathbf{x} finden, der allen Gleichungen exakt genügt. Dies ist z.B. bei der multiplen Regression der Fall, da man üblicherweise eine größere Anzahl m von Fällen als unbekannte Regressionsparameter hat. Die (m, n) -Matrix $A = X$ der Prädiktoren hat aber im Allgemeinen den vollen Rang $r = n$, so dass man eine Lösung finden könnte, indem man von links mit A' multipliziert, so dass $A'A\mathbf{x} = A'\mathbf{y}$ folgt, und da $A'A$ den gleichen Rang wie A hat (s. (2.46), Seite 74) existiert die zu $A'A$ inverse Matrix $(A'A)^{-1}$, so dass

$$\mathbf{x} = (A'A)^{-1}A'\mathbf{y} \quad (3.110)$$

resultiert. Die Matrix $(A'A)^{-1}A'$ ist eine Pseudoinverse für A , s. Seite 122. Sind die Komponenten von \mathbf{y} Messwerte, so sind sie üblicherweise durch Messfehler kontaminiert, so dass $\mathbf{y} \notin \mathcal{L}(A)$. (3.110) liefert dann keine Lösung \mathbf{x} , die allen m Gleichungen genügt. (3.110) ist die bekannte Kleinste-Quadrate-Schätzung $\hat{\mathbf{x}}$ für \mathbf{x} , vgl. (5.52), Seite 175, und Abschnitt 3.5.2.

Für den Fall $r < n$ liefert (3.108) den Lösungsraum.

Cramersche Regel: Diese Regel wird hier nur der Vollständigkeit wegen genannt. Es sei $A\mathbf{x} = \mathbf{y}$ ein Gleichungssystem, wobei A eine (n, n) -Matrix sei. Dann gilt für die j -te Komponente x_j von \mathbf{x}

$$x_j = \frac{|A_j|}{|A|}, \quad j = 1, \dots, n \quad (3.111)$$

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 5.10, 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.

3.5 Weitere Befunde

3.5.1 Die Zentrierungsmatrix

Die Zentrierung einer Datenmatrix kann durch eine Matrixmultiplikation bewerkstelligt werden:

Definition 3.11 *Es sei $\vec{1} = (1, 1, \dots, 1)'$ ein m -dimensionaler Vektor, dessen Komponenten alle gleich 1 sind, und $\vec{1}\vec{1}'$ ist das dyadische Produkt von $\vec{1}$ mit sich selbst. Dann heißt*

$$H_m = I - \frac{1}{m} \vec{1}\vec{1}'. \quad (3.112)$$

Zentrierungsmatrix.

H_M hat die Form

$$H_m = \begin{pmatrix} 1 - \frac{1}{m}, & -\frac{1}{m}, & -\frac{1}{m}, & \cdots, & -\frac{1}{m} \\ -\frac{1}{m}, & 1 - \frac{1}{m}, & -\frac{1}{m}, & \cdots, & -\frac{1}{m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{m}, & -\frac{1}{m}, & -\frac{1}{m}, & \cdots, & 1 - \frac{1}{m} \end{pmatrix} \quad (3.113)$$

Ist X eine nichtzentrierte (m, n) -Matrix und ist X_{sc} die korrespondierende spaltenzentrierte Matrix, X_{zc} die korrespondierende zeilenzentrierte Matrix so gilt

$$X_{sc} = H_m X, \quad X_{zc} = X H_m. \quad (3.114)$$

Es gilt der

Satz 3.22 Die Zentrierungsmatrix hat die Eigenschaften:

1. H_m ist idempotent (eine Matrix M ist idempotent, wenn $MM = M$),
2. H_m ist symmetrisch und positiv-semidefinit,
3. H_m hat einen Eigenwert $\lambda = 0$ und den Eigenwert 1 mit der Multiplizität $m - 1$, d.h. H_m hat den Rang $\text{rg}(H_m) = m - 1$,

Beweis: Es gilt

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

denn

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

d.h. H_m ist idempotent. Weiter gilt

$$H_m' = \left(I - \frac{1}{m} \vec{1} \vec{1}'\right)' = I' - \left(\frac{1}{m} \vec{1} \vec{1}'\right)' = I - \frac{1}{m} \vec{1} \vec{1}' = H_m, \quad (3.115)$$

d.h. H_m ist symmetrisch. Damit ist H_m gemäß Definition 3.188, Seite 145, eine Projektionsmatrix. Nach Satz 3.30 (Seite 146) hat H_m dann die Eigenwerte 0 und 1. Da H_m reell und symmetrisch ist, ist der Rang von H_m gleich der Anzahl der von Null verschiedenen Eigenwerte, mithin ist der Rang von H_m gleich $\text{rg}(H_m) = m - 1$, d.h. H_m hat keinen vollen Rang und ist damit singulär. Nach Definition 3.2 ist H_m positiv-semidefinit, wenn für $\mathbf{x} \in \mathbb{R}^m$ die Relation $\mathbf{x}' H_m \mathbf{x} = k \geq 0$ gilt. $\mathbf{x}' H_m$ ist ein Zeilenvektor; die j -te Komponente von $\mathbf{x}' H_m$ ist

$$-\frac{x_1}{m} - \cdots - \frac{x_{j-1}}{m} + x_j \left(1 - \frac{1}{m}\right) - \frac{x_{j+1}}{m} - \cdots - \frac{x_m}{m} = x_j - \frac{1}{m} \sum_{i=1}^m x_i = x_j - \bar{x}$$

Dann ist

$$\mathbf{x}'H_m\mathbf{x} = x_1^2 - x_1\bar{x} + x_2^2 - x_2\bar{x} + \cdots - x_m^2 - x_m\bar{x} = \sum_{i=1}^m x_i^2 - m\bar{x}^2 = k,$$

so dass

$$\frac{1}{m}\mathbf{x}'H_m\mathbf{x} = \frac{k}{m} = \frac{1}{m}\sum_{i=1}^m x_i^2 - \bar{x}^2 = \frac{1}{m}\sum_{i=1}^m (x_i - \bar{x})^2 \geq 0,$$

also ist H_m positiv-semidefinit.

Der Kern von H_m ist $\ker(H_m) = \{\mathbf{x} | H_m\mathbf{x} = \vec{0}\}$. Man hat

$$\begin{pmatrix} 1 - \frac{1}{m} & -\frac{1}{m} & -\frac{1}{m} & \cdots & -\frac{1}{m} \\ -\frac{1}{m} & 1 - \frac{1}{m} & -\frac{1}{m} & \cdots & -\frac{1}{m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{m} & -\frac{1}{m} & -\frac{1}{m} & \cdots & 1 - \frac{1}{m} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Für die i -te Komponente von $\vec{0}$ findet man dann

$$x_i - \bar{x} = 0 \Rightarrow x_i = \bar{x}$$

für alle i , d.h. die Komponenten von \mathbf{x} sind identisch, $\mathbf{x} = (x, x, \dots, x)'$. Da sich die Orientierung eines Vektors nicht ändert, wenn er mit einem Skalar multipliziert wird, ist

$$\frac{1}{x}\mathbf{x} = (1, 1, \dots, 1)',$$

d.h. die Orientierung der \mathbf{x} aus dem Kern von H_m ist identisch mit der von $\vec{1} = (1, 1, \dots, 1)'$. Der Kern ist damit ein 1-dimensionaler Teilraum des \mathbb{R}^m . \square

Es sei X eine unzentrierte Datenmatrix, X_c sei die zugehörige spaltenzentrierte Matrix. Dann gilt

$$X_c = H_m X, \quad X_c' X_c = X H_m' H_m X = X' H_m^2 X = X' H_m X, \quad (3.116)$$

wegen der Symmetrie und Idempotenz von H_m .

Der Rang zentrierter Datenmatrizen: Nach (3.114) ist $X_{sc} = H_m X$, so dass wegen (2.36), Seite 68,

$$\text{rg}(X_{sc}) \leq \min[\text{rg}(H_m), \text{rg}(X)]. \quad (3.117)$$

Satz 3.23 *Es ist $\text{rg}(H_m) = m$.*

Beweis: Zu zeigen ist, dass die Darstellung des Nullvektors $\vec{0}$ als Linearkombination der Spaltenvektoren von H_m nur möglich ist, wenn alle Koeffizienten $a_i = 0$ gilt. Allgemein hat man

$$a_1 \begin{pmatrix} 1 - 1/m \\ -1/m \\ \vdots \\ -1/m \end{pmatrix} + a_2 \begin{pmatrix} -1/m \\ 1 - 1/m \\ \vdots \\ -1/m \end{pmatrix} + \cdots + a_m \begin{pmatrix} -1/m \\ -1/m \\ \vdots \\ 1 - 1/m \end{pmatrix} = \vec{0}$$

Mit $S = \sum_i a_i$ gilt dann speziell für die i -te Komponente von $\vec{0}$

$$0 = a_i \left(1 - \frac{1}{m}\right) - \frac{m-1}{m} (S - a_i),$$

d.h.

$$\frac{m-1}{m} a_i = \frac{m-1}{m} (S - a_i) \Rightarrow a_i = S - a_i,$$

so dass

$$a_i = \frac{1}{2} S, \quad i = 1, \dots, m$$

Andererseits hat man

$$\frac{m-1}{m} \sum_i a_i = (m-1)S + \frac{m-1}{m} \sum_i a_i,$$

woraus $0 = (m-1)S$, d.h. $S = 0$ folgt. Dann folgt $a_i = \frac{1}{2}S = 0$ für alle i . Damit folgt, dass die Spaltenvektoren von H_m linear unabhängig sind, d.h. H_m hat den vollen Rang m . \square

Bei Datenmatrizen stehen die m Zeilen für die m Fälle und die n Spalten für die gemessenen n Variablen, und man hat $m \geq n$. Die Gleichung (3.117) impliziert dann

$$\text{rg}(X_{sc}) = \text{rg}(X) \leq n. \quad (3.118)$$

3.5.2 Die Pseudoinverse einer Matrix

Es werde das Gleichungssystem $A\mathbf{x} = \mathbf{y}$ betrachtet, wobei A eine (m, n) -Matrix sei mit $m > n$. \mathbf{x} ist ein n -dimensionaler Vektor, dessen Komponenten die Unbekannten sind. Man hat jetzt mehr Gleichungen als Unbekannte. \mathbf{y} ist ein m -dimensionaler Vektor. Ist $\mathbf{y} \in \mathcal{L}(A)$, d.h. ist \mathbf{y} eine Linearkombination der Spaltenvektoren von A ($\mathcal{L}(A)$ ist die lineare Hülle der Spaltenvektoren von A), so existiert \mathbf{x} , andernfalls – wenn \mathbf{y} keine Linearkombination der Spaltenvektoren von A ist ($\mathbf{y} \notin \mathcal{L}(A)$), so existiert \mathbf{x} nicht.

Man kann nun eine Pseudoinverse (auch: *generalisierte Inverse*) A^+ definieren:

Definition 3.12 Die Matrix A^+ heißt Pseudoinverse oder Moore-Penrose-Inverse, wenn sie die folgenden Bedingungen (Moore-Penrose-Bedingungen) erfüllt:

1. $AA^+A = A$,
2. $A^+AA^+ = A^+$
3. $(AA^+)^* = AA^+$
4. $(A^+A)^* = A^+A$.

Anmerkung: Die Bedingungen 3. und 4. beziehen sich auf Matrizen mit komplexwertigen Elementen; der Stern definiert die konjugiert komplexe Zahl einer komplexen Zahl. In diesem Skript werden keine komplexen Matrizen und Vektoren

betrachtet, so dass 3. und 4. nicht weiter berücksichtigt werden müssen, die beiden Punkte sind nur der Vollständigkeit halber mit aufgeführt worden. \square

Für das Gleichungssystem $A\mathbf{x} = \mathbf{y}$ kann nun leicht eine Pseudoinverse gefunden werden, wenn $\text{rg}(A) = n$ ist. Dann hat $A'A$ ebenfalls den Rang n , d.h. $A'A$ hat vollen Rang, so dass die Inverse $(A'A)^{-1}$ existiert. Man hat dann nach Multiplikation von $A\mathbf{x} = \mathbf{y}$ von links mit A' die Gleichung

$$A'A\mathbf{x} = A'\mathbf{y} \Rightarrow \mathbf{x} = (A'A)^{-1}A'\mathbf{y}. \quad (3.119)$$

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

$$(A'A)^{-1}A'\mathbf{y} = \begin{cases} \mathbf{x}, & \mathbf{y} \in \mathcal{L}(A), \\ \hat{\mathbf{x}}, & \mathbf{y} \notin \mathcal{L}(A). \end{cases} \quad (3.120)$$

(vergl. Abschnitt 5.6.3, Gleichung (5.52), Seite 175).

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

$$A = Q\Sigma T', \quad \Sigma = \begin{pmatrix} \Lambda_r^{1/2} & 0 \\ 0 & 0 \end{pmatrix} \quad (3.121)$$

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

$$A^+ = A' = T \begin{pmatrix} \Lambda_r^{-1/2} & 0 \\ 0 & 0 \end{pmatrix} Q' \quad (3.122)$$

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. (3.122) definiert tatsächlich eine Pseudoinverse.

3.5.3 Vektor- und Matrixnormen

Ein Vektor \mathbf{x} ist normiert, wenn $\|\mathbf{x}\| = 1$, wobei $\|\mathbf{x}\|$ die Länge im Sinne des Satzes von Pythagoras ist, weshalb auch von *Euklidischer Norm* die Rede ist (s. Skalierung eines Vektors, Seite 17). Diese Definition einer Norm charakterisiert einen Spezialfall, im Folgenden wird eine allgemeinere Definition gegeben.

Die Norm eines Vektors definiert, in welchem Sinne von der "Größe" eines Vektors gesprochen werden soll, – die übliche euklidische Norm $\|\mathbf{x}\| = (\sum_i x_i^2)^{1/2}$ definiert die Länge des Vektors \mathbf{x} als seine "Größe"⁵⁷. Ebenso kann eine *Matrixnorm* definiert werden. Dieser Begriff erweist sich als nützlich, wenn bestimmte Maxima oder Minima gefunden werden sollen, etwa die Varianzen von Projektionen einer Punktekonfiguration auf bestimmte Dimensionen, oder die Güte der Approximation an eine Datenmatrix. Es wird zuerst der Begriff der Vektornorm spezifiziert:

Definition 3.13 *Es seien $\mathbf{x} \in \mathbb{R}^n$ n -dimensionale Vektoren. Eine Vektornorm ist eine Abbildung $f : \mathbb{R}^n \rightarrow \mathbb{R}$ (d.h. es wird einem Vektor \mathbf{x} eine bestimmte reelle Zahl zugeordnet), die den Bedingungen*

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

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

Von besonderem Interesse sind die p -Normen

$$\|\mathbf{x}\|_p = (|x_1|^p + \dots + |x_n|^p)^{1/p} = \left(\sum_{j=1}^n |x_j|^p \right)^{1/p}. \quad (3.123)$$

Für $p = 1$ erhält man die 1-Norm

$$\|\mathbf{x}\|_1 = (|x_1| + \dots + |x_n|) = \sum_{j=1}^n |x_j|. \quad (3.124)$$

⁵⁷Um sich eine inhaltliche Vorstellung zu machen, stelle man sich vor dass die Komponenten x_i von \mathbf{x} Maße für Begabungen M_1, \dots, M_n repräsentieren. Dann ist $\|\mathbf{x}\|$ ein mögliches – nicht notwendig sinnvolles – Maß für die Gesamtbegabung einer Person.

und für $p = 2$ die euklidische Norm

$$\|\mathbf{x}\|_2 = (|x_1|^2 + \dots + |x_n|^2)^{1/2} = \left(\sum_{j=1}^n |x_j|^2 \right)^{1/2} = \sqrt{\mathbf{x}'\mathbf{x}}. \quad (3.125)$$

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

$$\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|. \quad (3.126)$$

Die Maximum-Norm ergibt sich aus der p -Norm für $p \rightarrow \infty$. Es sei $x_k = x_{\max}$ die maximale Komponente von \mathbf{x} . Dann ist

$$\|\mathbf{x}\|_p = \left(|x_k|^p \sum_{j=1}^n \frac{|x_j|^p}{|x_k|^p} \right)^{1/p}.$$

Wegen $|x_j|/|x_k| \leq 1$ für alle $j \neq k$ folgt $\lim_{p \rightarrow \infty} |x_j|^p/|x_k|^p \rightarrow 0$ für $j \neq k$ und $|x_j|^p/|x_k|^p = 1$ für $j = k$, so dass

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

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

Definition 3.14 *Es sei $\mathbb{R}^{m \times n}$ die Menge der reellen (m, n) -Matrizen⁵⁸. Eine Matrixnorm ist eine Abbildung $\|\cdot\| : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}_+$, \mathbb{R}_+ die Menge der reellen Zahlen größer oder gleich Null, und $A \mapsto \|A\|$ derart, dass*

1. $\|A\| = 0$ genau dann, wenn $A = 0$ die Nullmatrix ist,
2. $\|\lambda X\| = \lambda \|A\|$,
3. $\|A + B\| \leq \|A\| + \|B\|$

gilt. Zusammen mit der Norm $\|\cdot\|$ wird der Vektorraum der (m, n) -Matrizen dann zu einem normierten Vektorraum⁵⁹ $(\mathbb{R}^{m \times n}, \|\cdot\|)$.

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

1. Die *Frobenius-Norm*.

$$\|A\|_F = \left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2}. \quad (3.127)$$

⁵⁸Diese Definition ist etwas vereinfacht formuliert, eigentlich muss es heißen: es sei $\mathbb{K} = \mathbb{R}$ der Körper der reellen Zahlen und $\mathbb{K}^{m \times n} = \mathbb{R}^{m \times n}$ die Menge der reellen (m, n) -Matrizen, etc

⁵⁹Hier wird vom *allgemeinen* Begriff des Vektorraums Gebrauch gemacht, demzufolge auch Matrizen als "Vektoren" aufgefasst werden können, so dass auch Mengen von Matrizen einen Vektorraum bilden können. Der Begriff des Vektorraums bezieht sich ja eigentlich nur auf die Kombination bzw. Verknüpfungen von Elementen einer Menge!

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

2. Die p -Norm: sie ist definiert durch

$$\|A\|_p = \max_{\mathbf{x} \neq \vec{0}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p} \quad (3.128)$$

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

$$\|A\| = \max_{\mathbf{x} \neq \vec{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}. \quad (3.129)$$

durch eine Vektornorm induzierte Normen.

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

$$\|A\|_p = \sup_{\mathbf{x} \neq \vec{0}} \left\| \left(A \frac{\mathbf{x}}{\|\mathbf{x}\|} \right) \right\|_p = \max_{\|\mathbf{x}\|_p=1} \|A\mathbf{x}\|_p. \quad (3.130)$$

Speziell für $p = 2$ ist mit $\mathbf{y} = A\mathbf{x}$ die Norm $\|A\mathbf{x}\|_2$ durch die Norm $\|\mathbf{y}\|_2 = (\mathbf{y}'\mathbf{y})^{1/2} = \|\mathbf{y}\|$ gegeben, und nach dem Courant-Fischer Theorem 5.4 findet man

$$\|A\|_2 = \max_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2 = \sqrt{\lambda_{\max}}, \quad (3.131)$$

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

Satz 3.24 *Es sei A eine (m, n) -Matrix. Für die Frobenius-Norm $\|A\|_F$ gilt*

$$\|A\|_F^2 = \text{spur}(AA') = \sum_{i=1}^n \lambda_i \quad (3.132)$$

wobei $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ die Eigenwerte von $A'A$ sind.

Beweis: Auf A kann die SVD angewendet werden: $A = Q\Lambda^{1/2}P'$, $\lambda_j \geq 0$ für $j = 1, \dots, n$. Dann ist $AA' = Q\Lambda^{1/2}P'P\Lambda^{1/2}Q' = Q\Lambda Q'$, und die Diagonalelemente von $Q\Lambda Q'$ sind von der Form $\sum_i \lambda_i q_{ji}^2$ für $j = 1, \dots, n$. Die Spur von $A'A$ ist die Summe dieser Diagonalelemente, d.h.

$$\text{spur}(AA') = \sum_{j=1}^n \sum_{i=1}^n \lambda_i q_{ji}^2 = \sum_{i=1}^n \lambda_i \sum_{j=1}^n q_{ij}^2.$$

Aber $\sum_{j=1}^n q_{ij}^2 = 1$ für alle i , da Q orthonormal ist, d.h. die Eigenvektoren haben alle die Länge 1. Damit ist (3.132) gezeigt. \square

Anmerkung: $A = Q\Lambda^{1/2}P'$ impliziert $A'A = P\Lambda P'$ und wegen der Orthonormalität der Spaltenvektoren von P folgt in analoger Weise

$$\text{spur}(A'A) = \sum_{i=1}^n \lambda_i. \quad (3.133)$$

3.5.4 Die Approximation von Matrizen

Der folgende Satz macht eine Aussage über die Güte der Approximation einer Matrix A durch eine Matrix mit kleinerem Rang. So sei etwa $A = X$ eine Datenmatrix mit dem Rang n und man will versuchen, X durch eine Matrix X_r mit dem Rang $r < n$ zu approximieren, d.h. durch möglichst wenige latente Variable zu "erklären".

Satz 3.25 *Es seien A und A_k (m, n) -Matrizen, $m \geq n$, und es seien die Matrizen A und A_k durch*

$$A = Q\Lambda^{1/2}P' = \sum_{j=1}^n \sqrt{\lambda_j} \mathbf{q}_j \mathbf{p}'_j, \quad A_k = Q\Lambda_k^{1/2}P' = \sum_{j=1}^k \sqrt{\lambda_j} \mathbf{q}_j \mathbf{p}'_j \quad (3.134)$$

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

$$\|A - A_k\|_2 = \sqrt{\lambda_{k+1}}. \quad (3.135)$$

Beweis: Es ist $A = Q\Sigma P'$, $A_k = Q\Sigma_k P'$, wobei $\Sigma = \Lambda^{1/2}$, $\Sigma_k = \Lambda_k^{1/2}$, Λ die Diagonalmatrix der Eigenwerte von $A'A$, Λ_k die Diagonalmatrix der ersten k Eigenwerte. Dann ist

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

$\Sigma^* = \text{diag}(\underbrace{0, \dots, 0}_k, \sigma_{k+1}, \dots, \sigma_n)$, $\sigma_j = \sqrt{\lambda_j}$. Da $\|A\|_2 = \sigma_{\max} = \sigma_1$ im Falle geordneter Singularwerte $\sigma_{k+1} \geq \dots \geq \sigma_n$, folgt

$$\|A - A_k\|_2 = \sigma_{k+1} = \sqrt{\lambda_{k+1}},$$

da nun σ_{k+1} der maximale Singularwert ist. □

Anmerkungen:

1. Die Approximation wird trivialerweise immer besser, je größer der Wert von k ist, da ja der Wert von λ_{k+1} mit größer werdendem k immer kleiner wird. Der nichttriviale Teil der Aussage ist, dass $\|A - A_k\|_2$ gerade dem Wert von $\sqrt{\lambda_{k+1}}$ entspricht.

2. Bei der Approximation von A durch A_k wurde von der SVD von A Gebrauch gemacht. Die Gleichung $A = Q\Lambda^{1/2}P'$ ist insofern trivial, als die SVD stets gilt. Dass man A durch A_k approximiert, wobei A_k nur durch die ersten k Terme der SVD definiert ist, kann zur Frage führen, ob es eine andere Repräsentation für A_k gibt, die nicht auf der SVD beruht, aber besser ist in dem Sinne, dass $\|A - A_k\|_2 < \sqrt{\lambda_{k+1}}$. Eine solche gibt es nicht, wie noch gezeigt werden wird.
3. Man vergleiche die Aussage (3.135) mit der Aussage (3.134) von Satz 5.4. Wie die Gleichung (3.54), also die SVD von A , zeigt, ist A additiv durch Matrizen aufgebaut, die jeweils als dyadisches Produkt der Singularvektoren \mathbf{q}_j und \mathbf{p}_j definiert sind und die jeweils den Rang 1 haben (vergl. Satz 2.4, Seite 71). Der Rang $\text{rg}(A) \leq \min(m, n)$ ist durch die Anzahl der von Null verschiedenen Eigenwerte λ_j und damit durch die Anzahl der von Null verschiedenen $\sigma_k \mathbf{q}_k \mathbf{p}_k$ gegeben. Da die ersten k Eigenwerte von A und A_k identisch sind, enthält die Differenz $\Lambda^{1/2} - \Lambda_k^{1/2}$ nur Nullen, und der erste von Null verschiedene Wert in der Diagonalen ist $\sigma_{k+1} = \sqrt{\lambda_{k+1}}$. Da die Eigenwerte λ_j in Λ der Größe nach angeordnet sind, ist σ_{k+1} nun der größte Singularwert für $A - A_k$. \square

Im Folgenden bedeutet $\min_{\text{rg}(B)=k} \|X - B\|$ bzw. $\min_{\text{rg}(B)=k} \|X - B\|_F$ diejenige Matrix B , die (i) den Rang $\text{rg}(B) = k$ hat und die (ii) den Wert für die Norm $\|X - B\|$ bzw. $\|X - B\|_F$ minimiert. Es kann nun der folgende Satz bewiesen werden:

Satz 3.26 (Satz von Eckart & Young) *Es seien A eine (m, n) -Matrix mit dem Rang r , und $A = Q\Lambda^{1/2}P'$ sei die SVD von A . Weiter sei*

$$A_k = Q\Lambda_k^{1/2}P' = \sum_{j=1}^k \sqrt{\lambda_j} \mathbf{q}_j \mathbf{p}'_j, \quad (3.136)$$

wobei Λ_k die Diagonalmatrix mit den zu den ersten⁶⁰ k Eigenvektoren korrespondierenden Eigenwertev von A ist. Dann gilt

$$\min_{B \in \mathbb{R}^{m,n}, \text{rg}(B)=k} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1} = \sqrt{\lambda_{k+1}} \quad (3.137)$$

Anmerkung: Dieser Satz wird gelegentlich als Satz von Eckart & Young bezeichnet, weil Eckart & Young (1936) eine derartige Aussage vorgestellt haben, allerdings nicht mit dem hier folgenden Beweis (vergl. Golub & van Loan (2013), p. 79). Tatsächlich hat schon Schmidt (1907) diese Aussage hergeleitet, und Mirsky (1960) hat diesen und den folgenden Satz 3.27 in allgemeiner Weise bewiesen, so dass auch zusammenfassend vom Schmidt-Mirsky-Theorem gesprochen wird.

⁶⁰Es wird angenommen, dass die Eigenwerte der Größe nach geordnet sind, $\lambda_1 > \lambda_2 > \dots > \lambda_k$

Beweis: Zur Vereinfachung werde

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

gesetzt. Zu zeigen ist, dass

$$\|A - B_{\min}\| = \|A - A_k\|_2$$

gilt. Dazu werde angenommen, dass

$$\|A - B_{\min}\| < \|A - A_k\|_2.$$

Die Ungleichung bleibt bestehen, wenn beide Seiten mit dem gleichen Faktor (> 0) multipliziert werden. Für alle n -dimensionalen Vektoren $\mathbf{b} \neq \vec{0}$ gilt dann

$$\|A - B_{\min}\| \|\mathbf{b}\| < \|A - A_k\|_2 \|\mathbf{b}\| = \sigma_{k+1} \|\mathbf{b}\|.$$

Dann gilt auch

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

Insbesondere kann dann \mathbf{b} als Linearkombination der ersten $k+1$ (Eigen-)Vektoren von P gewählt werden: sind also gerade die ersten $k+1$ Spalten von P die Spalten von P_{k+1} , so sei $\mathbf{b} = P_{k+1}\mathbf{x} = x_1\mathbf{p}_1 + \cdots + x_{k+1}\mathbf{p}_{k+1}$. Es ist

$$B_{\min}\mathbf{b} = B_{\min}P_{k+1}\mathbf{x},$$

$B_{\min}P_{k+1}$ ist also eine $(m \times (k+1))$ -Matrix. Nach Satz 2.3, Gleichung (2.36) (Seite 68) ist aber $\text{rg}(B_{\min}P_{k+1}) \leq \min(\text{rg}(B_{\min}), \text{rg}(P_{k+1})) = k$, da ja $\text{rg}(B_{\min}) = k$ nach Voraussetzung. Dann folgt aber

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

d.h.

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

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

$$\|A\mathbf{b} - B_{\min}\mathbf{b}\| = \|AP_{k+1}\mathbf{x}\| < \|(AP_{k+1}\mathbf{x} - A_kP_{k+1}\mathbf{x})\| \leq \sigma_{k+1} \|\mathbf{b}\|.$$

Es ist aber

$$\|A\| \|P_{k+1}\mathbf{x}\| < \|(A - A_k)P_{k+1}\| \|\mathbf{x}\| \leq \sigma_{k+1} \|\mathbf{b}\|,$$

d.h.

$$\sigma_1 < \sigma_{k+1},$$

im Widerspruch zu $\sigma_1 \geq \sigma_{k+1}$. Damit gilt (3.137). \square

Satz 3.27 (Satz von Schmidt-Mirsky) Es sei A und B_{\min} (m, n)-Matrizen mit $m > n$, wobei A den Rang r und B den Rang $k < r$ habe, B_{\min} sei wie in Satz 3.26 definiert und A_k sei wie in (3.136) definiert. Dann gilt

$$\|A - B_{\min}\|_F = \sqrt{\sum_{j=k+1}^n \lambda_j}, \quad (3.138)$$

wobei $\|\cdot\|_F$ die Frobenius-Norm ist.

Beweis: Die Anwendung der SVD auf $A - A_k$ liefert

$$\|A - A_k\|^2 = \|Q(\Lambda^{1/2} - \Lambda_k^{1/2})P'\|^2 = \sum_{j=1}^n \lambda_j - \sum_{j=k+1}^n \lambda_j = \|A\|_F^2 - \sum_{j=k+1}^n \lambda_j.$$

B_{\min} kann als Summe von durch dyadische Produkte definierte Matrizen definiert werden, also

$$B_{\min} = \sum_{j=1}^n \mathbf{x}_j \mathbf{y}_j',$$

wobei die \mathbf{x}_j m -dimensionale und die \mathbf{y}_j n -dimensionale Vektoren sind. Da auch für B_{\min} eine Singularwertzerlegung gilt, können für \mathbf{x}_j und \mathbf{y}_j die jeweils mit $\sqrt{\sigma_j}$ multiplizierten Links- und Rechtssingulärvektoren von B_{\min} gewählt werden, d.h. man kann orthogonale Vektoren wählen. Zu zeigen ist dann, dass

$$\|A - \sum_{j=1}^k \mathbf{x}_j \mathbf{y}_j\| \geq \|A\|^2 - \sum_{j=1}^k \lambda_j.$$

Nach Definition der Frobenius-Norm hat man

$$\begin{aligned} \|A - \sum_{j=1}^k \mathbf{x}_j \mathbf{y}_j\|_F^2 &= \text{spur} \left((A - \sum_{j=1}^k \mathbf{x}_j \mathbf{y}_j)' (A - \sum_{j=1}^k \mathbf{x}_j \mathbf{y}_j) \right) \\ &= \text{spur} \left(A'A + \sum_{j=1}^k (\mathbf{y}_j - A'\mathbf{x}_j)(\mathbf{y}_j - A'\mathbf{x}_j)' - \sum_{j=1}^k A'\mathbf{x}_j \mathbf{x}_j A \right) \end{aligned}$$

Es ist $\text{spur}((\mathbf{y}_j - A'\mathbf{x}_j)(\mathbf{y}_j - A'\mathbf{x}_j)) \geq 0$, $\text{spur}(A'\mathbf{x}_j \mathbf{x}_j' A) = \|A'\mathbf{x}_j\|^2$ und es ist zu zeigen, dass

$$\sum_{j=1}^k \|A'\mathbf{x}_j\|^2 \leq \sum_{j=1}^k \lambda_j.$$

Die SVD von A sei $A = Q\Sigma P'$, und es sei $P_1 = [|\mathbf{p}_1\rangle \dots |\mathbf{p}_k\rangle |0\rangle]$, $P_2 = [|0\rangle |\mathbf{p}_{k+1}\rangle \dots |\mathbf{p}_n\rangle]$, so dass $P = [P_1 | P_2]$. Analog dazu sei $\Sigma_1 = \text{diag}(\sigma_1, \dots, \sigma_k)$, $\Sigma_2 = \text{diag}(\sigma_{k+1}, \dots, \sigma_n)$.

Dann hat man

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

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

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

□

3.6 Anwendung: Latente Variablen

Datenmatrizen X lassen sich oft "erklären", indem man etwa die Spaltenvektoren von X als Linearkombinationen der Vektoren einer (Teil-)Basis darstellt; diese Basisvektoren liefern u. U. eine Interpretation der Kovarianzen oder Korrelationen zwischen den gemessenen Variablen. Man muß sich dann für eine bestimmte Basis oder Teilbasis aus der Menge der möglichen entscheiden, und im Allgemeinen fällt dann die Wahl auf die Hauptachsentransformation: für rtdzr latente Variable trennt dann maximal zwischen den Fällen, die zweite zweitmaximal, etc., was u. U. die Interpretierbarkeit erleichtert. Dieser Ansatz besteht in einer direkten Anwendung der SVD, s. Abschnitt 3.6.1.

Eine andere Fragestellung, die auf die Berechnung latenter Variablen führt, ist die Zuordnung von Fällen zu bestimmten Kategorien auf der Basis von Messungen einer Reihe von Variablen. Diese Messungen können wieder zu einer Matrix X zusammengefasst werden, und die latenten Variablen werden so bestimmt, dass die Projektionen der Fälle auf die latenten Variablen nach Maßgabe maximaler Trennung der Kategorien erfolgt. Der Standardansatz für diese Fragestellung wird in Abschnitt 3.6.2 vorgestellt.

3.6.1 Explorieren (Hauptkomponenten und SVD)

Die Hauptkomponentenanalyse (engl. Principal Component Analysis (PCA); PCA ist die gängige Abkürzung) ist eine Methode, einen als eine (m, n) -Matrix X gegebenen Datensatz in einer der Faktorenanalyse ähnlichen Weise zu komprimieren;

die Methode geht auf Karl Pearson (1901) und Harold Hotelling (1933) zurück. Gewöhnlich wird X als zentriert vorausgesetzt (spaltenzentriert, wenn die Spalten gemessene Variablen repräsentieren und man an der Analyse der Kovarianzen oder Korrelationen zwischen den Variablen interessiert ist, zeilenzentriert, wenn man an einer Analyse der Kovarianzen oder Korrelationen der "Fälle" interessiert ist, – sofern diese Analyse Sinn macht⁶¹). Sie besteht im Wesentlichen aus einer Koordinatentransformation, nämlich dem Übergang von den ursprünglichen Koordinaten zu den Hauptachsen des Ellipsoids, das die Punktekonfiguration der Fälle definiert. Insbesondere Hotellings Ansatz bestand darin, die gesuchten "latenten Variablen" oder "latenten Dimensionen" so zu bestimmen, dass sich die Fälle auf der ersten latenten Variablen maximal unterscheiden, d.h. dass die Varianz der Koordinaten der Fälle auf der ersten latenten Dimension maximal ist, auf der zweiten latenten Dimension soll die Varianz der Koordinaten zweitmaximal sein, etc.

Die PCA ergibt sich unmittelbar aus der SVD: nach Satz 3.46 gilt $X = Q'\Lambda^{1/2}V'$, wobei Q und V die Eigenvektoren von XX' bzw. $X'X$ sind; Q und T sind also orthonormal. Dann folgt $XV = Q\Lambda^{1/2} = U$; in der faktoreanalytischen Literatur oft wird U i. A. mit F (für 'Faktor') bezeichnet, aber der hier betrachtete Ansatz unterscheidet sich von dem der Faktorenanalyse, so dass eine andere Bezeichnung naheliegt, z.B. L für *latente Variable*. Die i -te Zeile \tilde{x}_i von X enthält die Messungen des i -ten Falles für die n Variablen, und die i -te Zeile von L enthält die Koordinaten des i -ten Falles auf "latenten Variablen" oder "latenten Dimensionen", deren Orientierung durch die Hauptachsen der durch $X'X$ definierten Ellipsoide gegeben ist. Aus der Herleitung der SVD geht hervor, dass die Hotellingsche Forderung bezüglich der Koordinaten der Fälle durch die Wahl von L erfüllt wird; $L_1'L_1$ ist maximal (vergl. die Anmerkungen auf Seite 101). Darüber hinaus liefert die SVD die bei Hotelling nicht weiter diskutierte Beziehung zwischen der Repräsentation der Fälle einerseits und der Variablen andererseits (R- und Q-Analyse, Cattell (1966)). Bei der R-Analyse werden die Korrelationen zwischen den Variablen, bei der Q-Analyse werden die Korrelationen zwischen den Fällen analysiert. Die Frage ist, worin sich die latenten Dimensionen bei diesen beiden Analysearten unterscheiden, bzw. welche Beziehungen zwischen den Ergebnissen dieser Analysen existieren. Die SVD gibt diese Beziehungen in kompakter Form an, wie im Folgenden gezeigt wird.

Für die folgenden Betrachtungen wird angenommen, dass X spaltenzentriert oder sogar spaltenstandardisiert ist. $C = \frac{1}{m}X'X$ ist dann eine Kovarianz- bzw. eine Korrelationsmatrix für die Variablen. Die Analyse von Kovarianzmatrizen macht Sinn, wenn die gemessenen Variablen alle dieselbe Maßeinheit haben. Sind die Maßeinheiten verschieden (wie bei den Variablen Länge und Gewicht) sollten

⁶¹Sind die Fälle z.B. Personen, an denen verschiedene Variablen gemessen wurden, so bedeutet die Berechnung einer Korrelation zwischen Fällen das Mitteln über die Variablen, was eine sinnfreie Übung sein kann! Die Herleitung der PCA aus der SVD macht überdies deutlich, dass die Analyse der Korrelationen zwischen den Fällen nicht notwendig ist, da die "faktorielle Struktur" der Fälle von der SVD gleich mitgeliefert wird.

die Variablen standardisiert werden.

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

$$X = Q\Lambda^{1/2}V' = LV', \quad L = Q\Lambda^{1/2} \quad (3.139)$$

$$= QA', \quad A = V\Lambda^{1/2} \quad (3.140)$$

Es werde $m > n$ vorausgesetzt, d.h. man habe mehr Fälle als Variablen; diese Voraussetzung ist für die folgenden Betrachtungen nicht essentiell, entspricht aber den normalerweise vorliegenden Bedingungen. Der Rang von X sei $r \leq \min(m, n)$; es wird der Einfachheit halber $r = n = \min(m, n)$ angenommen, weil im Allgemeinen die Eigenwerte *numerisch*⁶² alle ungleich Null sind; der Fall, dass der wahre Wert von r kleiner als $\min(m, n)$ ist, wird später behandelt. Q bzw. L ist eine (m, n) -Matrix, V bzw. A ist eine (m, n) -Matrix, und Λ ist eine (m, n) -Diagonalmatrix. Die Spalten von Q bzw. L sind Basisvektoren für die Spaltenvektoren \mathbf{x}_j von X , und die Zeilen von Q bzw. L repräsentieren die Fälle. Die Spaltenvektoren von V bzw. A sind Basisvektoren für die Zeilenvektoren von X .

Definition 3.15 *Es seien*

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

\mathbf{L}_k ist der k -te Spaltenvektor von L , \mathbf{a}_k ist der k -te Spaltenvektor von A (s. (3.140)). Die Komponenten ℓ_{ik} von \mathbf{L}_k , $i = 1, \dots, m$, heißen Faktorwerte (factor scores) der Fälle auf der k -ten latenten Variablen, die Komponenten a_{jk} von \mathbf{a}_k , $j = 1, \dots, n$, heißen Ladungen (factor loadings) der Variablen auf der k -ten latenten Variablen,.

Wie oben schon angemerkt wurde soll die Bezeichnung L für die Matrix $Q\Lambda^{1/2}$ an den Ausdruck 'latent' erinnern, A ist die in der Literatur übliche Bezeichnung für die Matrix der Faktorladungen.

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

Aus (3.139) und (3.140) erhält man sofort die Beziehungen

$$X = LV' = QA' \quad (3.141)$$

$$X' = AQ' = VL' \quad (3.142)$$

⁶²Messungen haben stets nur eine endliche Genauigkeit, Berechnungen können stets nur mit endlich vielen Dezimalstellen durchgeführt werden, etc, so dass die Elemente x_{ij} von X i. A. von den "wahren" Werten abweichen.

Da V orthonormal ist, folgt aus (3.141) sofort der Hotellingsche Ansatz $XT = L$, d.h. die Spaltenvektoren von L sind Linearkombinationen der Spaltenvektoren \mathbf{x}_j von X . Andererseits sind nach (3.141) die Spaltenvektoren \mathbf{x}_j von X Linearkombinationen der Spaltenvektoren \mathbf{L}_k von L bzw. \mathbf{q}_k von Q , und die Spaltenvektoren $\tilde{\mathbf{x}}_i$ von X' sind Linearkombinationen der Spaltenvektoren \mathbf{a}_k von A bzw. der Spaltenvektoren \mathbf{t}_k von T ; insbesondere gilt

$$\mathbf{x}_j = L\tilde{\mathbf{t}}_j = Q\tilde{\mathbf{a}}_j \quad (3.143)$$

$$\tilde{\mathbf{x}}_i = A\tilde{\mathbf{q}}_i = T\tilde{\mathbf{L}}_i \quad (3.144)$$

wobei $\tilde{\mathbf{t}}_j$ der j -te Spaltenvektor von T' (Zeilenvektor von T) und $\tilde{\mathbf{q}}_i$ der i -te Spaltenvektor von Q' (Zeilenvektor von Q) ist. Die Spaltenvektoren \mathbf{t}_k von V bzw. \mathbf{a}_k von A repräsentieren, ebenso wie die Spaltenvektoren \mathbf{L}_k bzw. \mathbf{q}_k , latente Variablen (auch: latente Dimensionen); auf Fragen der inhaltlichen Interpretation wird am Ende dieses Abschnitts eingegangen. Die Zeilenvektoren $\tilde{\mathbf{v}}_i$ bzw. $\tilde{\mathbf{a}}_i$ repräsentieren Koordinaten der Variablen auf den latenten Variablen, und die Zeilenvektoren $\tilde{\mathbf{L}}_i$ oder $\tilde{\mathbf{q}}_i$ von L bzw. Q repräsentieren die Fälle auf latenten Variablen.

Die Gleichung (3.144) beschreibt die Rotation von achsenparallelen Ellipsoiden in orientierte Ellipsoide und umgekehrt. Denn (3.144) impliziert $\tilde{\mathbf{x}}'_i = \tilde{\mathbf{L}}'_i V'$, und die Multiplikation von rechts mit $X'X = V\Lambda V'$ und noch einmal mit $\tilde{\mathbf{x}}_i$ liefert

$$\tilde{\mathbf{x}}'_i(X'X)\tilde{\mathbf{x}}_i = \tilde{\mathbf{L}}'_i V'V\Lambda V'\tilde{\mathbf{L}}_i = \tilde{\mathbf{L}}'_i\Lambda\tilde{\mathbf{L}}_i$$

d.h. T rotiert die Vektoren $\tilde{\mathbf{L}}_i$ des achsenparallelen Ellipsoids in die Vektoren $\tilde{\mathbf{x}}_i$ des orientierten Ellipsoids. In diesem Sinne ist die SVD äquivalent einer Hauptachsentransformation.

Die Faktorenscores ℓ_{ik} und die Ladungen a_{jk} sind die Skalarprodukte

$$\ell_{ik} = \tilde{\mathbf{x}}'_i \mathbf{v}_k = \sum_{j=1}^n x_{ij} v_{jk} = \|\tilde{\mathbf{x}}_i\| \|\mathbf{v}_k\| \cos \theta_{ik}. \quad (3.145)$$

$$a_{jk} = \mathbf{x}'_j \mathbf{q}_k = \sum_{i=1}^m x_{ij} q_{ik} = \|\mathbf{x}_j\| \|\mathbf{q}_k\| \cos \phi_{jk}. \quad (3.146)$$

Wie im Kommentar auf Seite ?? angemerkt wurde können Skalarprodukte als Ähnlichkeitsmaße ("Korrelationen", wie gelegentlich gesagt wird, wobei dieser Ausdruck nicht wörtlich zu nehmen ist) für Paare von Vektoren interpretiert werden. ℓ_{ik} als Maß für die Ausprägung beim i -ten Fall ist demnach die Ähnlichkeit zwischen den Vektoren $\tilde{\mathbf{x}}_i$ und \mathbf{v}_k ; Die Komponenten x_{ij} von $\tilde{\mathbf{x}}_i$ ist der Messwert des i -ten Falls für die j -te Variable, und die Komponenten v_{jk} können als Maß für die Ausprägung, mit der die j -te Variable die k -te Dimension erfasst gesehen werden. Das Skalarprodukt ℓ_{ik} wird maximal relativ zu den Längen $\|\tilde{\mathbf{x}}_i\|$ und $\|\mathbf{v}_k\|$, wenn der Winkel θ_{ik} zwischen den Vektoren gleich Null ist. Die Vektoren haben dann dieselbe Orientierung, d.h. es gilt $x_{ij} = \alpha v_{jk}$. $\alpha \in \mathbb{R}$, für alle

$j = 1, \dots, n$. $\ell_{ik} = 0$ wenn $\cos \theta_{ik} = 0$, d.h. wenn $\tilde{\mathbf{x}}_i$ und \mathbf{v}_k orthogonal zueinander sind.

Offenbar beziehen sich die ℓ_{ik} und die a_{jk} stets auf dieselben latenten Dimensionen L_1, \dots, L_r , $r \leq n$ ($m > n$ vorausgesetzt). Die Catellschen R - und Q -Analysen führen grundsätzlich auf dieselben latenten Variablen.

Eine analoge Betrachtung gilt für die Ladungen a_{jk} . Als Skalarprodukte $\mathbf{x}'_j \mathbf{q}_k$ sind die a_{jk} ein Maß für die Ähnlichkeit der Vektoren \mathbf{x}_j und \mathbf{q}_k . Hier wird gewissermaßen über die Produkte $x_{ij} q_{ik}$ über alle $i = 1, \dots, m$ gemittelt, und q_{ik} kann als Maß gesehen werden, in dem die k -te latente Dimension beim i -ten Fall ausgeprägt ist. Für den Maximalwert von a_{jk} gelten dieselben Betrachtungen wie für die ℓ_{ik} . Die Ladung a_{jk} ist gleich Null wenn \mathbf{x}_j und \mathbf{q}_k orthogonal zueinander sind. Formal sind ℓ_{ik} und a_{jk} orthonormale Basisentwicklungen, (vergl 49, Seite 49).

Weiter sind sie Vektorprojektionen im Sinne der Gleichung (1.61), Seite 24: $\ell_{ik} = \|\tilde{P}_{ik}\|$, und \tilde{P}_{ik} ist der Projektionsoperator für den Vektor $\tilde{\mathbf{x}}_i$ auf die Gerade oder Achse \tilde{L}_k im \mathbb{R}^n , deren Orientierung durch den k -ten Eigenvektor \mathbf{t}_k von $X'X$ gegeben ist. Analog dazu ist $a_{jk} = \|\tilde{P}_{jk}\|$, $\|\tilde{P}_{jk}\|$ der Projektionsoperator des Vektors \mathbf{x}_j auf die Gerade oder Achse L_k im \mathbb{R}^m , deren Orientierung durch den k -ten Eigenvektor \mathbf{q}_k von XX' ist.

Repräsentation von Variablen und Fällen Die Datenmatrix X ist eine (m, n) -Matrix. Die Fälle werden durch die Zeilenvektoren von X , also durch n -dimensionale Vektoren repräsentiert, die Variablen werden durch die Spaltenvektoren, also durch m -dimensionale Vektoren abgebildet. Im Raum der latenten Variablen werden aber sowohl die Fall- wie auch die Variablenvektoren im selben n -dimensionalen Raum repräsentiert. Denn die Zeilenvektoren $\tilde{\mathbf{L}}_i$ bzw. $\tilde{\mathbf{q}}_i$ sind ebenso wie die Zeilenvektoren $\tilde{\mathbf{a}}_j$ bzw. $\tilde{\mathbf{t}}_j$ im normierten Fall jeweils n -dimensionale Vektoren, – d.h. sowohl die Fälle wie auch die Variablen werden durch n -dimensionale Vektoren repräsentiert. der Messwert x_{ij} ist dann ein Skalarprodukt eines Fall- und eines Variablenvektors:

$$x_{ij} = \tilde{\mathbf{q}}'_i \tilde{\mathbf{a}}_j = \|\tilde{\mathbf{q}}_i\| \|\tilde{\mathbf{a}}_j\| \cos \varphi_{ij} \quad (3.147)$$

$$= \tilde{\mathbf{L}}'_i \tilde{\mathbf{v}}_j = \|\tilde{\mathbf{L}}_i\| \|\tilde{\mathbf{v}}_j\| \cos \varphi_{ij} = \sum_{k=1}^n \sqrt{\lambda_k} q_{ik} v_{kj} \quad (3.148)$$

und x_{ij} wird maximal (relativ zur Länge der Vektoren $\tilde{\mathbf{q}}_i$ und $\tilde{\mathbf{a}}_j$ bzw. $\tilde{\mathbf{L}}_i$ und $\tilde{\mathbf{v}}_j$), wenn der Winkel φ_{ij} zwischen diesen Vektoren gleich Null ist, wenn also das Profil $\tilde{\mathbf{q}}_i$ der Werte des i -ten Falls auf den latenten Variablen proportional zum Profil $\tilde{\mathbf{a}}_j$ der Werte der Variablen auf den latenten Variablen ist. Diese Gleichung wurde für den allgemeinen Fall $X = UV'$ bereits auf Seite 71, Gleichung (2.42) vorgestellt, um zu argumentieren, dass der Zeilen- und Spaltenraum durch dieselben latenten Variablen definiert ist und für den Fall $r = 2$ den exakten Biplot definiert. Natürlich kann man auch $x_{ij} = \tilde{\mathbf{L}}'_i \tilde{\mathbf{v}}_j$ betrachtet werden, – die Interpretation ist analog; es ist der Winkel zwischen den Vektoren $\tilde{\mathbf{q}}_i$ und $\tilde{\mathbf{v}}_j$, der für gegebene Längen der Vektoren den Wert von x_{ij} bestimmt.

Ladungen und die Korrelationen zwischen den Variablen: Die Skalierung $\mathbf{a}_k = \sqrt{\lambda_k} \mathbf{v}_k$ der Spaltenvektoren \mathbf{v}_k von V erweist sich als vorteilhaft, wenn man insbesondere an einer Analyse der Variablen interessiert ist. So folgt aus $X = QA'$ die Beziehung $X'X = AQ'QA'$, und wegen $Q'Q = I$ gilt

$$X'X = AA'. \quad (3.149)$$

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

$$R = \frac{1}{m} X'X = \frac{1}{m} AA', \quad X = Z \quad (3.150)$$

die Matrix der Korrelationen zwischen den Variablen. Die Korrelation zwischen der j -ten und der k -ten Variablen ergibt sich als Skalarprodukt zwischen den Zeilenvektoren $\tilde{\mathbf{a}}_j$ und $\tilde{\mathbf{a}}_k$ von A :

$$r_{jk} = \frac{1}{m} \tilde{\mathbf{a}}_j' \tilde{\mathbf{a}}_k = \frac{1}{m} \sum_{u=1}^n a_{uj} a_{uk} = \frac{1}{m} \|\tilde{\mathbf{a}}_j\| \|\tilde{\mathbf{a}}_k\| \cos \theta_{jk}, \quad (3.151)$$

θ_{jk} der Winkel zwischen den Vektoren $\tilde{\mathbf{a}}_j$ und $\tilde{\mathbf{a}}_k$; je kleiner der Winkel, desto größer ist der Absolutbetrag $|r_{jk}|$. Für $j = k$ erhält man ($\cos \theta_{jj} = 1$ wegen $\theta_{jj} = 0$)

$$r_{jj} = \frac{1}{m} \|\tilde{\mathbf{a}}_j\|^2 = \frac{1}{m} \sum_{u=1}^n a_{ju}^2 = 1, \quad j = 1, \dots, n \quad (3.152)$$

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

$$\|\mathbf{a}_k\|^2 = \sum_{j=1}^n a_{jk}^2 = \sum_{j=1}^n \lambda_k v_{jk}^2 = \lambda_k \sum_{j=1}^n v_{jk}^2 = \lambda_k, \quad (3.153)$$

denn $\sum_{j=1}^n v_{jk}^2 = 1$, da die Spaltenvektoren von V ja normiert sind. $\|\mathbf{a}_k\|^2$ ist die Summe der Quadrate der Ladungen der Variablen auf der k -ten latenten Dimension, – und sie ist gleich dem k -ten Eigenwert von $X'X$. Auch wenn X spaltenzentriert ist, so ist der Mittelwert der a_{jk} allerdings nicht gleich Null, so dass λ_k nicht proportional der Varianz der Ladungen auf der k -ten Dimension ist; der Proportionalitätsfaktor ist $1/m$. Andererseits ist die Matrix Q spaltenzentriert, wenn X spaltenzentriert ist: ist $\vec{1}$ ein m -dimensionaler Vektor, dessen Komponenten alle gleich 1 sind, so ist $\vec{1}'X = \vec{0}'$, d.h. die Spaltensummen von X sind alle gleich Null. Dann hat man aber

$$\vec{1}'X = \vec{1}'L\Lambda^{1/2}V' = \vec{0}'$$

wegen $\vec{1}'L = \vec{0}'$. Für den k -ten Spaltenvektor \mathbf{L}_k erhält man dann

$$\|\mathbf{L}_k\|^2 = \sum_{i=1}^m \ell_{ik}^2 = \sum_{i=1}^m \lambda_k q_{ik}^2 = \lambda_k, \quad (3.154)$$

wegen $\sum_{i=1}^m q_{ik}^2 = 1$ (die Spalten von Q sind ja normiert). Da die Summe der ℓ_{ik} gleich Null ist, ist $\sum_i \ell_{ik}^2$ proportional zur Varianz der ℓ_{ik} . λ_k korrespondiert also zur Varianz der Koordinaten (Scores) der Fälle auf der k -ten latenten Dimension. Zusammen mit (3.153) hat man die Beziehung

$$\|\mathbf{L}_k\|^2 = \|\mathbf{a}_k\|^2 = \lambda_k, \quad k = 1, \dots, n \quad (3.155)$$

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.

Der Biplot im allgemeinen Fall Wie oben schon im Zusammenhang mit der Gleichung (3.147) erwähnt wurde, liefert (3.147) eine exakte Darstellung des Biplots, d.h. der simultanen Repräsentation der Fälle und der Variablen in einem Koordinatensystem, dessen Achsen die latenten Variablen darstellen. Allerdings kann es sein, dass die Analyse der Daten einen Wert $r > 2$ darlegen. In diesem Fall liefert der Biplot nur eine approximative Darstellung.

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

$$\pi_k = \frac{\lambda_k}{\sum_{j=1}^n \lambda_j} \quad (3.156)$$

der Anteil der Varianz der k -ten latenten Variablen an der Gesamtvarianz. π_k kann zur Diskussion der Frage, wieviele latente Variable zur Interpretation der Daten benötigt werden, herangezogen werden.

Die Matrix X kann auf der Basis der SVD $X = Q\Lambda^{1/2}V'$ über die dyadischen Produkte der Spaltenvektoren von Q und T ausgedrückt werden:

$$X = \sum_{k=1}^n \sigma_k \mathbf{q}_k \mathbf{v}'_k \quad \sigma_k = \sqrt{\lambda_k}. \quad (3.157)$$

(vergl.(3.53), Seite 102). Zusammen mit (3.156) kann (3.157) benutzt werden, um den Wert des Ranges r von X abzuschätzen: Terme mit "hinreichend" kleinen λ_k -Werten können u.U. vernachlässigt werden. Man hat dazu den

Satz 3.28 (Satz von Eckart & Young) *Die Approximation*

$$X \approx X_r = Q_r \Lambda_r^{1/2} v_r' = \sum_{k=1}^r \sqrt{\lambda_k} \mathbf{q}_k v_k', \quad r < n \quad (3.158)$$

approximiert X im Sinne der Methode der Kleinsten Quadrate.

Beweis: Bekannt wurde diese Aussage (samt Beweis) durch die Arbeit von Eckart & Young (1936); eine modernere Version des Beweises wird in Abschnitt 3.5.4 angeboten, vergl. insbesondere den Beweis zu Satz 3.26, S. 128. Dort wird der Begriff der Matrixnorm vorausgesetzt, vergl. Abschnitt 3.5.3. \square

Fragen der Interpretation: Es wurde weiter oben gesagt, dass die Spaltenvektoren der Matrizen L und A – also die skalierten Versionen von A und T – "latente" Variablen repräsentieren. Rein formal sind diese Spaltenvektoren Basisvektoren für die Teilräume des \mathbb{R}^m bzw. des \mathbb{R}^n , in denen die m -dimensionalen Vektoren \mathbf{x}_j bzw. die n -dimensionalen Vektoren $\tilde{\mathbf{x}}_i$ liegen. Die Frage ist, ob man ihnen eine inhaltliche Bedeutung zuordnen kann, und wenn ja, ob die latenten Variablen für die \mathbf{x}_j eine andere Bedeutung als die für die Vektoren $\tilde{\mathbf{x}}_i$ haben oder nicht.

Die Gleichungen (3.145) und (3.146) für die Koordinaten ℓ_{ik} des i -ten Falls auf der k -ten latenten Variablen und a_{jk} der j -ten gemessenen Variablen auf der k -ten latenten Variablen zeigen, dass es sich bei diesen Koordinaten um "Korrelationen" (um einen zwar saloppen, aber auf den Kern dieser Größen zielenden Ausdruck zu gebrauchen) eines Falles oder einer Variablen mit latenten Variablen handelt. ℓ_{ik} und a_{jk} werden maximal, wenn die Vektoren $\tilde{\mathbf{x}}_i$ und \mathbf{t}_k einerseits und \mathbf{x}_j und \mathbf{q}_k parallel sind. In diesen Fällen entsprechen der i -Fall bzw. die j -te Variable den jeweiligen latenten Variablen und damit repräsentieren sie eine mögliche Interpretation der latenten Variablen. In der Praxis werden diese Maximalkorrelationen kaum vorkommen, aber man kann dann die Interpretation nach Maßgabe derjenigen Fälle oder Variablen vornehmen, die hoch mit den latenten Variablen "korrelieren". Im einen Fall wird die Interpretation in Termen hoch korrelierender Fälle, im anderen Fall durch hoch korrelierender Variablen erfolgen.

Die beiden Interpretationen werden nicht unabhängig voneinander sein. Aus den Gleichungen (3.139) und (3.140) lassen sich sofort die Gleichungen

$$L = Q\Lambda^{1/2} = XV, \quad A = V\Lambda^{1/2} = X'Q \quad (3.159)$$

herleiten, aus denen hervorgeht, dass Q und V bzw. L und A in Abhängigkeit voneinander definiert werden. Dann ist a_{jk} die Ausprägung der j -ten Variablen auf der k -ten latenten Dimension:

$$a_{jk} = \sqrt{\lambda_k} v_{jk} = x_{1j} q_{1k} + x_{2j} q_{2k} + \cdots + x_{mj} q_{mk}, \quad (3.160)$$

d.h. a_{jk} ist eine Art gewogener Mittelwert der normierten Ausprägungen *der Fälle* auf der k -ten latenten Variablen; die "Gewichte" sind die Messungen x_{ij} für die

j -te Variable., $i = 1, \dots, m$. Man könnte $a_{jk} = \bar{q}_k$ schreiben um zu betonen, dass hier Werte derselben latenten Variablen gemittelt werden. Eine analoge Aussage gilt für die Scores ℓ_{ik} : Es ist $L = XV$, d.h. für die Komponenten ℓ_{ik} – der Score für den i -ten Fall auf der k -ten latenten Variablen – gilt

$$\ell_{ik} = \sqrt{\lambda_k} q_{ik} = x_{i1}v_{1k} + x_{i2}v_{2k} + \dots + x_{in}v_{nk}. \quad (3.161)$$

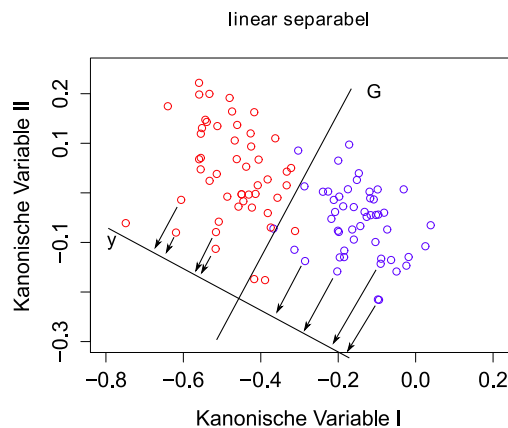
ℓ_{ik} ist demnach ein gewogener Mittelwert der v_{1k}, \dots, v_{nk} , also den Ausprägungen der Variablen auf der k -ten latenten Variablen, mit den für den i -ten Fall spezifischen Gewichtungen x_{ij} , $j = 1, \dots, n$. Die Ausprägung des i -ten Falls auf der k -ten latenten Dimension ist ein für den i -ten Fall spezifischer Mittelwert der Ausprägungen der Variablen auf der k -ten latenten Dimension. Die latenten Variablen sind Merkmale, aus denen sich die gemessenen Variablen zusammensetzen und die auch dazu dienen, die Fälle zu charakterisieren.

Beispiele findet man in <http://www.uwe-mortensen.de/fakanalysews0506b.pdf>, p. 91.

3.6.2 Diskriminieren und klassifizieren

Wie in den einführenden Bemerkungen schon angedeutet, soll eine Lösung für die Aufgabe, Fälle Gruppen oder Kategorien zuzuordnen gefunden werden. Betrachtet wird die Konfiguration der Fälle, und gesucht sind latente Variablen derart, dass die Projektion der Punkte auf die zu diesen Variablen korrespondierenden Koordinatenachsen ("Diskriminanzfunktionen") maximal zwischen den Gruppen oder Kategorien trennt. s. Abbildung 18. Gegeben ist eine $(m \times n)$ Matrix X , de-

Abbildung 18: Klassifikation nach Fisher (1936) (I): Ω_1 blau, Ω_2 rot, eine mögliche Trennlinie G , eine Projektionsgerade Y



ren Zeilen Fälle und deren Zeilen Prädiktorvariablen repräsentieren. Eine – sagen

wir: die erste – der gesuchten Achsen läßt sich durch einen m -dimensionalen Vektor \mathbf{y} darstellen, dessen Komponenten die gesuchten Projektionen sind. \mathbf{y} ist eine Linearkombination der Spaltenvektoren von X , d.h. es gilt allgemein $\mathbf{y} = X\mathbf{u}$. \mathbf{u} ist ein Koeffizientenvektor, mit dem der gewünschte Vektor \mathbf{y} erzeugt wird. Die Bestimmung von \mathbf{y} ist demnach die Bestimmung von \mathbf{u} .

Da die Komponenten von \mathbf{y} Koordinaten auf einer Koordinatenachse darstellen, kann man sie zweifach indizieren: die j -te Komponente von \mathbf{y} sei die Koordinate y_{ik} des i -ten Falls in der k -ten Kategorie. \bar{y}_k sei der Mittelwert der Koordinaten der k -ten Gruppe. \mathbf{u} soll so bestimmt werden, dass die Varianz der \bar{y}_k , $k = 1, \dots, K$ (es gebe K Gruppen oder Kategorien) maximal wird relativ zur durchschnittlichen Varianz innerhalb der Gruppen. Die Varianz der $\mathbf{a}y_k$ wird durch eine Quadratsumme QS_{zw} (zw für "zwischen") definiert, und die durchschnittliche Varianz innerhalb der Gruppen wird durch eine Quadratsumme QS_{inn} (inn für "innerhalb") definiert. Es soll

$$\lambda = \frac{QS_{zw}}{QS_{inn}} \quad (3.162)$$

maximiert werden; λ ist als *Diskriminanzkoeffizient* bekannt. Die Varianz aller Komponenten von \mathbf{y} wird durch eine Quadratsumme QS_{tot} (tot für "total") definiert, und wie aus der Varianzanalyse bekannt gilt

$$QS_{ges} = QS_{inn} + QS_{zw}. \quad (3.163)$$

Da \mathbf{u} bestimmt werden muss, muss der Quotient (3.162) aus Funktion des unbekanntens Vektors \mathbf{u} angeschrieben werden. Wegen $\mathbf{y} = X\mathbf{u}$ muss also \mathbf{y} durch $X\mathbf{u}$ ersetzt werden.

Zunächste eine kleine Vorbetrachtung. Es seien $\mathbf{u} = (u_1, \dots, u_n)'$ und $\mathbf{x} = (x_1, \dots, x_n)'$ zwei n -dimensionale Vektoren. Für das Quadrat des Skalarprodukts $(\mathbf{x}'\mathbf{u})^2$ gilt

$$(\mathbf{x}'\mathbf{u})^2 = (x_1u_1 + \dots + x_nu_n)^2 = (x_1u_1)^2 + \dots + (x_nu_n)^2 + \sum_{i \neq j} x_i x_j u_i u_j. \quad (3.164)$$

Der Ausdruck rechts erinnert an eine quadratische Form (vergl. (3.21)), Seite 86). In der Tat läßt sich das Produkt $u_i u_j$ in der Summe rechts als das (i, j) -te Element des dyadischen Produkts $\mathbf{u}\mathbf{u}'$ interpretieren:

$$\mathbf{u}\mathbf{u}' = \begin{pmatrix} u_1^2 & u_1u_2 & \cdots & u_1u_n \\ u_2u_1 & u_2^2 & \cdots & u_2u_n \\ \vdots & \vdots & \ddots & \vdots \\ u_nu_2 & u_nu_2 & \cdots & u_n^2 \end{pmatrix}$$

Man rechnet leicht nach, dass nun

$$(\mathbf{x}'\mathbf{u})^2 = \mathbf{x}'\mathbf{u}\mathbf{u}'\mathbf{x} \quad (3.165)$$

gilt.

Es sei X eine (m, n) -Matrix; die m Fälle seien in K Gruppen mit den Umfängen n_1, n_2, \dots, n_K aufgeteilt, $m = \sum_{k=1}^K n_k$. Für jeden Fall werden Messwerte bei insgesamt n Variablen bestimmt. Die Matrix X läßt sich dann wie in (3.166) anschreiben

$$\mathbf{y} = \begin{pmatrix} Y_{11} \\ Y_{21} \\ \vdots \\ Y_{n_1 1} \\ Y_{12} \\ Y_{22} \\ \vdots \\ Y_{n_2 2} \\ \vdots \\ Y_{1K} \\ Y_{2K} \\ \vdots \\ Y_{n_K K} \end{pmatrix}, \quad X = \begin{pmatrix} X_{111} & X_{112} & \cdots & X_{11p} \\ X_{211} & X_{212} & \cdots & X_{21p} \\ \vdots & \vdots & \cdots & \vdots \\ X_{n_1 11} & X_{n_1 12} & \cdots & X_{n_1 1p} \\ X_{121} & X_{122} & \cdots & X_{12p} \\ X_{221} & X_{222} & \cdots & X_{22p} \\ \vdots & \vdots & \cdots & \vdots \\ X_{n_2 21} & X_{n_2 22} & \cdots & X_{n_2 2p} \\ \vdots & \vdots & \cdots & \vdots \\ X_{1K1} & X_{1K2} & \cdots & X_{1Kp} \\ X_{2K1} & X_{2K2} & \cdots & X_{2Kp} \\ \vdots & \vdots & \cdots & \vdots \\ X_{n_K K1} & X_{n_K K2} & \cdots & X_{n_K Kp} \end{pmatrix}, \quad (3.166)$$

Die Indizierung der Elemente von X bezieht sich (i) auf einen Fall in einer gegebenen Gruppe, (ii) auf die gegebene Gruppe, und (iii) auf eine Variable. Es soll eine Linearkombination \mathbf{y} der Spaltenvektoren von X bestimmt werden, die eine möglichst gute Separierung der Gruppen gestattet. Dies bedeutet, dass ein Vektor \mathbf{u} gesucht wird derart, dass

$$\mathbf{y} = X\mathbf{u}. \quad (3.167)$$

Man muss natürlich spezifizieren, was mit "möglichst gute Separierung der Gruppen" gemeint ist. Dem Ansatz (3.167) zufolge wird für jeden Fall eine Komponente von \mathbf{y} bestimmt; die Indizierung der Komponenten von \mathbf{y} werde dabei wie in (3.166) angegeben vorgenommen, so dass gruppenspezifische Mittelwerte \bar{y}_k der Komponenten von \mathbf{y} bestimmt werden können:

$$\bar{y}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} Y_{ik}, \quad k = 1, \dots, K \quad (3.168)$$

Die "möglichst gute Separierung der Gruppen" soll nun bedeuten, dass die \bar{y}_k sich maximal voneinander unterscheiden sollen. Dies bedeutet, dass die Varianz der \bar{y}_k so groß wie möglich sein soll, was natürlich die Einführung geeigneter Nebenbedingungen erfordert (die maximale Varianz wäre ohne Nebenbedingungen unendlich), auf die später eingegangen wird. Wie aus der Varianzanalyse bekannt läßt sich nun die Gesamtvarianz, bzw. die ihr entsprechende Quadratsumme QS_{ges} , in eine Quadratsumme QS_{inn} und eine Quadratsumme QS_{zw} "zwischen" den Gruppen

aufteilen; diese entspricht der Varianz der Mittelwerte der Gruppen. Man hat

$$QS_{ges} = \sum_{k=1}^K \sum_{i=1}^{n_k} (Y_{ik} - \bar{y})^2 \quad (3.169)$$

$$QS_{inn} = \sum_{k=1}^K \sum_{i=1}^{n_k} (Y_{ik} - \bar{y}_k)^2, \quad QS_{zw} = \sum_{k=1}^K n_k (\bar{y}_k - \bar{y})^2 \quad (3.170)$$

und es gilt (3.163), wie man leicht nachrechnet. Die Komponente Y_{ik} von \mathbf{y} ist das Skalarprodukt der Zeilenvektoren $\tilde{\mathbf{x}}_{ik}$ und \mathbf{u} , also $Y_{ik} = \tilde{\mathbf{x}}'_{ik} \mathbf{u}$. Der Index i bezeichnet stets den i -ten Fall in der k -ten Gruppe. Gleichung (3.168) bedeutet dann

$$\bar{y}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \tilde{\mathbf{x}}'_{ik} \mathbf{u}.$$

Es werde $y_{ik} = Y_{ik} - \bar{y}_k$ gesetzt. Dann ist

$$QS_{inn} = \sum_{k=1}^K \sum_{i=1}^{n_k} y_{ik}^2 = \sum_{k=1}^K \sum_{i=1}^{n_k} (\tilde{\mathbf{x}}'_{ik} \mathbf{u})^2 = \sum_{k=1}^K \sum_{i=1}^{n_k} (\mathbf{u}' \tilde{\mathbf{x}}_{ik})^2,$$

d.h. (vergl. (3.164))

$$QS_{inn} = \sum_{k=1}^K \sum_{i=1}^{n_k} \mathbf{u}' \tilde{\mathbf{x}}_{ik} \tilde{\mathbf{x}}'_{ik} \mathbf{u} = \mathbf{u}' \underbrace{\left(\sum_{k=1}^K \sum_{i=1}^{n_k} \tilde{\mathbf{x}}_{ik} \tilde{\mathbf{x}}'_{ik} \right)}_W \mathbf{u}. \quad (3.171)$$

$$QS_{inn} = \mathbf{u}' W \mathbf{u}. \quad (3.172)$$

Für QS_{zw} findet man

$$QS_{zw} = \sum_{k=1}^K n_k (u_1(\bar{x}_{k1} - \bar{x}_1) + \cdots + u_p(\bar{x}_{kp} - \bar{x}_p))^2,$$

und

$$QS_{zw} = \sum_{k=1}^K n_k \mathbf{u}' (\bar{\mathbf{x}}_{k\cdot} - \bar{\mathbf{x}}) (\bar{\mathbf{x}}_{k\cdot} - \bar{\mathbf{x}})' \mathbf{u} = \mathbf{u}' \underbrace{\left(\sum_{k=1}^K n_k (\bar{\mathbf{x}}_{k\cdot} - \bar{\mathbf{x}}) (\bar{\mathbf{x}}_{k\cdot} - \bar{\mathbf{x}})' \right)}_B \mathbf{u},$$

d.h.

$$QS_{zw} = \mathbf{u}' B \mathbf{u} \quad (3.173)$$

Die Maximierung von QS_{zw} , d.h. der Maximierung der Gruppenmittelwerte \bar{y}_k , kann nun als Maximierung des Quotienten

$$\lambda(\mathbf{u}) := \frac{QS_{zw}}{QS_{inn}} = \frac{\mathbf{u}' B \mathbf{u}}{\mathbf{u}' W \mathbf{u}} \quad (3.174)$$

definiert werden. Offenbar entspricht $\lambda(\mathbf{u})$ dem aus der Varianzanalyse bekannten F -Wert, der hier allerdings als Funktion von \mathbf{u} maximiert werden soll. Andererseits ist der Quotient auf der rechten Seite der auf Seite 113, Gleichung (3.94), definierte generalisierte Rayleigh-Quotient.

Eine Möglichkeit, den Maximalwert von $\lambda(\mathbf{u})$ zu bestimmen, besteht darin, $\lambda(\mathbf{u})$ nach \mathbf{u} zu differenzieren und die Ableitung gleich Null zu setzen. Eine andere besteht darin, $\lambda(\mathbf{u})$ in einen Rayleigh-Quotienten (vergl. (3.23), Seite 88) umzuformen und diesen zu maximieren.

Dazu muss man nur berücksichtigen, dass B und W symmetrische Matrizen sind. Für W hat man die Darstellung $W = P\Lambda P'$, wobei P die Matrix der Eigenvektoren von W ist und Λ die Diagonalmatrix der Eigenwerte von W . Dann ist $W^{1/2} = P\Lambda^{1/2}$. Weiter sei $\mathbf{v} = W^{1/2}\mathbf{u}$; dann ist $\mathbf{u} = W^{-1/2}\mathbf{v}$ und $\lambda(\mathbf{u})$ kann in der Form

$$\lambda = \frac{\mathbf{u}'B\mathbf{u}}{\mathbf{u}'W^{1/2}W^{1/2}\mathbf{u}} = \frac{\mathbf{v}'W^{-1/2}BW^{-1/2}\mathbf{v}}{\mathbf{v}'\mathbf{v}}$$

geschrieben werden. Setzt man $A = W^{-1/2}BW^{-1/2}$, so erhält man

$$\lambda = \frac{\mathbf{v}'A\mathbf{v}}{\mathbf{v}'\mathbf{v}}, \quad (3.175)$$

d.h. λ entspricht einem Rayleigh-Quotienten. Nach dem Satz von Courant-Fisher wird λ maximal, wenn $\mathbf{v} = \mathbf{t}$, \mathbf{t} der Eigenvektor von A , der zum maximalen Eigenwert λ_1 von A korrespondiert, d.h. es muss $A\mathbf{t} = \lambda_{\max}\mathbf{t}$ und damit

$$W^{-1/2}BW^{-1/2}\mathbf{t} = \lambda_{\max}\mathbf{t} \quad (3.176)$$

gelten. Multiplikation von links mit $W^{-1/2}$ liefert dann

$$W^{-1}BW^{-1/2}\mathbf{t} = \lambda_{\max}W^{-1/2}\mathbf{t}.$$

\mathbf{t} ist ein spezieller Vektor für $\mathbf{v} = W^{1/2}\mathbf{u}$; setzt man $\mathbf{t} = W^{1/2}\mathbf{u}_t$, so folgt $\mathbf{t}W^{-1/2} = \mathbf{u}_t$ und man erhält

$$W^{-1}B\mathbf{u}_t = \lambda_{\max}\mathbf{u}_t. \quad (3.177)$$

\mathbf{u}_t ist also ein Eigenvektor von $W^{-1}B$; λ_{\max} , der maximale Wert von λ , ist der zugehörige Eigenwert. $\mathbf{y}_t = X\mathbf{u}_t$ ist der Vektor, dessen Komponenten die Projektionen der Fälle auf eine Diskriminanzdimension sind, auf der die Gruppenmittelwerte maximal separiert sind.

Die Gleichung (3.177) charakterisiert das generalisierte Eigenwertproblem, vergl. Gleichung (3.93), Seite 113, – man muss Gleichung (3.177) nur von links mit W multiplizieren, und (3.175) ist ein generalisierter Rayleigh-Quotient, vergl. Gleichung (3.94), ebenfalls Seite 113. Existiert für (3.177) mehr als nur ein Eigenvektor \mathbf{u}_t , so gibt es mehr als nur einen Vektor \mathbf{y} , d.h. mehr als nur eine Achse, die zwischen den Klassen oder Gruppen diskriminiert.

Satz 3.29 Sind $\mathbf{y}_j = X\mathbf{u}_j$ und $\mathbf{y}_k = X\mathbf{u}_k$ zwei verschiedene Vektoren, so gilt

$$\mathbf{y}'_j \mathbf{y}_k = 0, \quad j \neq k \quad (3.178)$$

Beweis: Es ist

$$\begin{aligned} \mathbf{y}'_j \mathbf{y}_k &= \mathbf{u}'_j X' X \mathbf{u}_k = \mathbf{v}_j W^{-1/2} (X' X) W^{-1/2} \mathbf{v}_k \\ &= \mathbf{v}_j W^{-1/2} W W^{-1/2} \mathbf{v}_k = \mathbf{v}'_j \mathbf{v}_k = 0 \end{aligned}$$

denn \mathbf{v}_j und \mathbf{v}_k sind Eigenvektoren der symmetrischen Matrix $W^{-1/2} B W^{-1/2}$ und deswegen orthonormal, und W ist eine Schätzung für $X' X$, Gleichung (3.171). \square

3.7 Projektionen

3.7.1 Projektion auf eine Ebene im \mathbb{R}^3

Ebenen im \mathbb{R}^3 sind bereits in Beispiel 1.8, Seite 37 betrachtet worden. Demnach ist eine Ebene durch

$$\mathbb{E} = \{\mathbf{x} \mid \mathbf{x} = \mathbf{u} + s\mathbf{b}_1 + t\mathbf{b}_2, s, t \in \mathbb{R}\} \quad (3.179)$$

definiert, wobei $\mathbf{x}, \mathbf{u}, \mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^3$ und $s, t \in \mathbb{R}$ Parameter sind. \mathbf{u} ist ein Stütz- oder Ortsvektor und $\mathbf{v}_1, \mathbf{v}_2$ sind nichtparallele Richtungsvektoren. Sie sind Basisvektoren für den 2-dimensionalen Teilraum des \mathbb{R}^3 , der durch (3.180) definiert wird. $B = [\mathbf{b}_1, \mathbf{b}_2]$ ist eine Basis für diesen 2-dimensionalen Teilraum des \mathbb{R}^3 (vergl. (1.99), Seite 46). Die allgemeine Darstellung der Ebene ist

$$\mathbb{E}_2^3 = \{\mathbf{p} \in \mathbb{R}^3 \mid \mathbf{p} = \mathbf{p}_0 + c_1\mathbf{b}_1 + c_2\mathbf{b}_2, \quad \mathbf{b}_1, \mathbf{b}_2 \text{ nicht parallel}, c_1, c_2 \in \mathbb{R}\} \quad (3.180)$$

(\mathbb{E}_2^3 bezeichne einen 2-dimensionalen Teilraum des 3-dimensionalen Vektorraums.) Für $\mathbf{p}_0 = \vec{0}$ geht die Ebene durch den Nullpunkt des Koordinatensystems. \mathbf{p}_0 verbindet den Nullpunkt mit dem Punkt P in der Ebene. Es gelte $\mathbf{p}_0 = \vec{0}$. $\mathbf{x} = (x_1, x_2, x_3)'$ sei ein Ortsvektor, dessen Endpunkt x außerhalb der Ebene liege, und A sei der Fuß- oder Lotpunkt der orthogonalen Projektion von x auf E . Das Linienelement $\overline{0, A}$ werde durch den Vektor $\mathbf{y}_0 = \vec{P}_{xy}$ repräsentiert. Es gelte

1. $\mathbf{y}_0 = c_1\mathbf{b}_1 + c_2\mathbf{b}_2$
2. $(\mathbf{y}_0 - \mathbf{x})' \mathbf{b}_1 = 0$
3. $(\mathbf{y}_0 - \mathbf{x})' \mathbf{b}_2 = 0$.

(vergl. Beispiel 1.8, Seite 37, mit leicht veränderter Notation.) Setzt man 1. in 2. und 3. ein, so ergibt sich

$$\begin{aligned} (c_1\mathbf{b}_1 + c_2\mathbf{b}_2 - \mathbf{x})' \mathbf{b}_1 &= c_1\mathbf{b}'_1 \mathbf{b}_1 + c_2\mathbf{b}'_2 \mathbf{b}_1 - \mathbf{x}' \mathbf{b}_1 = 0 \\ (c_1\mathbf{b}_1 + c_2\mathbf{b}_2 - \mathbf{x})' \mathbf{b}_2 &= c_1\mathbf{b}'_1 \mathbf{b}_2 + c_2\mathbf{b}'_2 \mathbf{b}_2 - \mathbf{x}' \mathbf{b}_2 = 0 \end{aligned}$$

Damit hat man ein System von zwei Gleichungen in den zwei Unbekannten c_1 und c_2 . Bisher ist nur vorausgesetzt worden, dass die Vektoren \mathbf{b}_1 und \mathbf{b}_2 nicht parallel sind. Jetzt werde angenommen, dass die beiden Vektoren orthonormal sind, d.h. es soll $\mathbf{b}'_1\mathbf{b}_1 = \mathbf{b}'_2\mathbf{b}_2 = 1$ und $\mathbf{b}'_1\mathbf{b}_2 = 0$ gelten. Dann vereinfacht sich das System der Gleichungen zu

$$c_1 = \mathbf{x}'\mathbf{b}_1 \quad (3.181)$$

$$c_2 = \mathbf{x}'\mathbf{b}_2 \quad (3.182)$$

und man erhält

$$\mathbf{y}_0 = (\mathbf{x}'\mathbf{b}_1)\mathbf{b}_1 + (\mathbf{x}'\mathbf{b}_2)\mathbf{b}_2, \quad (3.183)$$

also eine orthonormale Basisentwicklung von \mathbf{y}_0 , s. Gleichung (1.100), Seite 47.

3.7.2 Der allgemeine Fall: Projektionsmatrizen

Es werde allgemein der \mathbb{R}^n und eine k -dimensionale Ebene E_n^k betrachtet, wobei

$$E_n^k = \left\{ \mathbf{y} \mid \mathbf{y} = \sum_{j=1}^k c_j \mathbf{b}_j, \quad c_j \in \mathbb{R}, \quad j = 1, \dots, k < n \right\} \quad (3.184)$$

sei und die $\mathbf{b}_1, \dots, \mathbf{b}_k$ linear unabhängige Vektoren aus \mathbb{R}^n sind, d.h. die \mathbf{b}_j , $1 \leq j \leq k$ bilden eine Teilbasis des \mathbb{R}^n . Setzt man $B = [\mathbf{b}_1, \dots, \mathbf{b}_k]$ und $\mathbf{c} = (c_1, \dots, c_k)'$, so hat man

$$\mathbf{y} = B\mathbf{c}. \quad (3.185)$$

E_n^k ist eine k -dimensionale Ebene im \mathbb{R}^n durch den Ursprung des Koordinatensystems. Weiter sei $\mathbf{x} \in \mathbb{R}^n$ mit dem Anfangspunkt im Ursprung des Koordinatensystems und Endpunkt außerhalb der Ebene E_n^k . Gesucht ist die orthogonale Projektion von \mathbf{x} auf die Ebene; der Fuß- oder Lotpunkt A sei der Endpunkt eines Vektors $\mathbf{y}_0 \in E_n^k$, vergl. Abbildung 7, Seite 24. Weiter ist

$$(\mathbf{y}_0 - \mathbf{x})'\mathbf{b}_j = 0, \quad j = 1, \dots, k \quad (3.186)$$

d.h.

$$\mathbf{y}'_0\mathbf{b}_j = \mathbf{x}'\mathbf{b}_j, \quad \text{oder} \quad \mathbf{y}'_0 B = \mathbf{x}' B \quad (3.187)$$

d.h. $B'\mathbf{y}_0 = B'\mathbf{x}$, d.h. $B'B\mathbf{c} = B'\mathbf{x}$ nach (3.185), so dass

$$\mathbf{c} = (B'B)^{-1}B'\mathbf{x}$$

folgt und die Multiplikation von links mit B liefert, wieder wegen (3.185),

$$\mathbf{y}_0 = \underbrace{B(B'B^{-1})B'}_P \mathbf{x} = P\mathbf{x} \quad (3.188)$$

Die Matrix $P = B(B'B^{-1})B'$ heißt *Projektionsmatrix*. P transformiert den zu projizierenden Vektor \mathbf{x} direkt in den Vektor \mathbf{y}_0 . Es gelten die Aussagen

1. $P' = P$, P d.h. ist symmetrisch,
2. $PP = P$, d.h. P ist *idempotent*.

Es sei $P = B(B'B)^{-1}B'$, B eine (m, k) -Matrix mit $\text{rg}(B) = k$, so dass $(B'B)^{-1}$ existiert. Dann ist P symmetrisch, denn $(B(B'B)^{-1}B')' = B(B'B)^{-1}B'$, und idempotent, denn

$$(B(B'B)^{-1}B')(B(B'B)^{-1}B') = B(B'B)^{-1}B'B(B'B)^{-1}B' = B(B'B)^{-1}B'.$$

Also ist $B(B'B)^{-1}B'$ eine Projektionsmatrix. □

Da P als symmetrisch definiert ist, muss P quadratisch sein, also sei P eine (m, m) Projektionsmatrix. I sei die (m, m) -Identitätsmatrix. Dann ist $(I - P)$ ebenfalls eine Projektionsmatrix. Denn

$$(I - P)' = I' - P' = I - P, \quad (I - P)(I - P) = I - 2P + PP = I - P.$$

Weiter gilt der

Satz 3.30 P sei eine Projektionsmatrix. Dann hat P die Eigenwerte $\lambda = 0$ und $\lambda = 1$.

Beweis: Für die Eigenwerte und Eigenvektoren von P gilt $P\mathbf{v} = \lambda\mathbf{v}$. Da P symmetrisch ist, folgt, dass alle Eigenwerte größer oder gleich Null sind. Dann hat man wegen der Idempotenz von P

$$PP\mathbf{v} = \lambda P\mathbf{v} = \lambda^2\mathbf{v} = P\lambda = \lambda\mathbf{v}$$

wegen $PP = P$. Es folgt $\lambda\mathbf{v} = \mathbf{v}$ und damit $(\lambda^2 - \lambda)\mathbf{v} = \vec{0}$. Da \mathbf{v} ein Eigenvektor ist, muss $\mathbf{v} \neq \vec{0}$ sein, also folgt $\lambda^2 - \lambda = 0$ bzw. $\lambda^2 = \lambda$ bzw. $\lambda = \sqrt{\lambda}$. Eine Lösung ist sicherlich $\lambda = 0$. Eine zweite ist $\lambda = 1$. Eine weitere Lösung $\lambda \neq 0$ existiert nicht. Denn angenommen, es existiert ein $0 < \lambda = a \in \mathbb{R}$, $a \neq 1$. Für $a < 1$ folgt $a > \sqrt{a}$ und für $a > 1$ folgt $a < \sqrt{a}$, entgegen der Forderung $a = \sqrt{a}$. Also gibt es außer $\lambda = 1$ und der "trivialen" Lösung $\lambda = 0$ keine anderen Eigenwerte. □

Anmerkung: P sei eine (m, m) -Projektionsmatrix. Dann existieren m Eigenvektoren, – einer zum Eigenwert 0 und $m - 1$ mit dem korrespondierenden Eigenwert $\lambda = 1$. Da der Rang einer symmetrischen Matrix gleich der Anzahl der von Null verschiedenen Eigenwerte ist (Satz 3.8, Seite 95), hat P den Rang $\text{rg}(P) = m - 1$. □

3.7.3 Beispiel: das Allgemeine Lineare Modell

Das Allgemeine Lineare Modell (ALM) ist durch

$$\mathbf{y} = X\mathbf{b} + \mathbf{e} \tag{3.189}$$

definiert, wobei \mathbf{y} ein m -dimensionaler Vektor, dessen Komponenten Messungen einer abhängigen Variablen sind, X ist eine (m, n) -Matrix, deren Spaltenvektoren Messungen von unabhängigen Variablen bzw. die Indikatorvariablen experimenteller Bedingungen enthalten, \mathbf{b} ist ein n -dimensionaler Vektor von Regressionskoeffizienten, und \mathbf{e} ist ein m -dimensionaler Vektor von "Messfehlern"; sie repräsentieren tatsächliche Messfehler sowie den Effekt nicht berücksichtigter unabhängiger Variablen.

Der Koeffizienten- oder Parametervektor \mathbf{b} ist im Allgemeinen nicht bekannt und wird üblicherweise mit der Methode der Kleinsten Quadrate geschätzt: (3.189) liefert $\mathbf{y} - X\mathbf{b} = \mathbf{e}$ und man minimiert $(\mathbf{y} - X\mathbf{b})'(\mathbf{y} - X\mathbf{b}) = \mathbf{e}'\mathbf{e}$, d.h.

$$\|\mathbf{y} - X\mathbf{b}\|^2 = \|\mathbf{e}\|^2$$

bezüglich \mathbf{b} . Man findet (s. Abschnitt 5.6.3, Seite 174 (Anhang))

$$\hat{\mathbf{b}} = (X'X)^{-1}X'\mathbf{y}, \quad (3.190)$$

so dass

$$\mathbf{y} = X\hat{\mathbf{b}} + \hat{\mathbf{e}} = \hat{\mathbf{y}} + \hat{\mathbf{e}} \quad (3.191)$$

mit

$$\hat{\mathbf{y}} = X\hat{\mathbf{b}} = \underbrace{X(X'X)^{-1}X'}_P \mathbf{y} = P\mathbf{y}. \quad (3.192)$$

P erfüllt offenbar die Bedingungen 1. und 2. für eine Projektionsmatrix. Man hat demnach

$$\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{e}} = P\mathbf{y} + \hat{\mathbf{e}}, \quad (3.193)$$

Wegen $\hat{\mathbf{y}} = X\hat{\mathbf{b}}$ ist $\hat{\mathbf{y}} \in \mathcal{L}(X)$, $\mathcal{L}(X)$ die lineare Hülle der Spaltenvektoren von X , und \mathbf{y} ist ein Element des orthogonalen Komplements von $\mathcal{L}(X)$ (s. Definition 1.15, Seite 52). Wegen $\mathbf{y} - P\mathbf{y} = \hat{\mathbf{e}}$ folgt

$$(I - P)\mathbf{y} = \hat{\mathbf{e}}. \quad (3.194)$$

Zusammenfassend hat man mit (3.192)

$$\hat{\mathbf{y}} = P\mathbf{y} \quad (3.195)$$

$$\hat{\mathbf{e}} = (I - P)\mathbf{y}, \quad (3.196)$$

d.h. $I - P$ projiziert \mathbf{y} auf den zu $\hat{\mathbf{y}}$ orthogonalen Vektor $\hat{\mathbf{e}}$. P und $I - P$ projizieren einen Vektor also auf jeweils eine der beiden Katheten eines rechtwinkligen Dreiecks. Die Vektoren $\hat{\mathbf{y}}$ und $\hat{\mathbf{e}}$ sind in der Tat orthogonal:

$$\begin{aligned} \hat{\mathbf{y}}'\hat{\mathbf{e}} &= (P\mathbf{y})'(\mathbf{y} - P\mathbf{y}) = \mathbf{y}'P'\mathbf{y} - \mathbf{y}'P'\hat{\mathbf{y}} \\ &= \hat{\mathbf{y}}'P\mathbf{y} - \mathbf{y}'PP\mathbf{y} = \hat{\mathbf{y}}'P\mathbf{y} - \hat{\mathbf{y}}'P\mathbf{y} = 0 \end{aligned} \quad (3.197)$$

Demnach steht $\hat{\mathbf{e}}$ senkrecht auf $\hat{\mathbf{y}}$. Man findet

$$\mathbf{y}'\mathbf{y} = \|\mathbf{y}\|^2 = (\hat{\mathbf{y}} + \hat{\mathbf{e}})'(\hat{\mathbf{y}} + \hat{\mathbf{e}}) = \hat{\mathbf{y}}'\hat{\mathbf{y}} + \hat{\mathbf{e}}'\hat{\mathbf{e}} + 2\hat{\mathbf{y}}'\hat{\mathbf{e}}$$

und wegen (3.197) gilt dementsprechend der Satz des Pythagoras

$$\|\mathbf{y}\|^2 = \|\hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{e}}\|^2. \quad (3.198)$$

Die Einflußmatrix Nach (3.192) gilt $\hat{\mathbf{y}} = P\mathbf{y}$; $\hat{\mathbf{y}}$ ist also eine Projektion von \mathbf{y} auf die lineare Hülle $\mathcal{L}(X)$ mit $\hat{\mathbf{y}} \in \mathcal{L}(X)$, und $\hat{\mathbf{e}} \perp \mathcal{L}(X)$. $\|\hat{\mathbf{e}}\|$ ist die kürzeste Distanz zwischen dem Endpunkt von \mathbf{y} und $\mathcal{L}(X)$. Nach Gleichung (3.192) gilt $\hat{\mathbf{y}} = P\mathbf{y}$; die Komponenten von \mathbf{y} sind die tatsächlich gemessenen Werte, die Komponenten von $\hat{\mathbf{y}}$ sind die auf der Basis der Prädiktoren vorhergesagten Werte. P heißt deshalb auch *Einflußmatrix* (influence matrix). Die Gleichung $\hat{\mathbf{y}} = P\mathbf{y}$ gibt dann an, wie die gemessenen Werte die vorhergesagten Werte beeinflussen. So ist der Wert der i -ten Komponente (des i -ten Falls) von $\hat{\mathbf{y}}$ durch das Skalarprodukt des i -ten Zeilenvektors von P mit den Datenvektor \mathbf{y} gegeben:

$$\hat{y}_i = p_{i1}y_1 + p_{i2}y_2 + \cdots + p_{im}y_m. \quad (3.199)$$

Das Element p_{ij} läßt sich direkt interpretieren als das Ausmaß, den die Messung y_j via die KQ-Schätzungen der Regressionsparameter auf auf \hat{y}_i hat. Man spricht von der *Hebelwirkung* (leverage) der Messung y_i auf die Schätzungen \hat{y}_i (vergl. Abb. 12, S. 80). Damit lassen sich Ausreißer identifizieren, oder, anders formuliert, der Effekt von Ausreißern läßt sich damit charakterisieren. Dazu wird insbesondere das Element p_{ii} betrachtet. Da eine Projektionsmatrix idempotent und symmetrisch ist, gilt $PP = P$, und somit ist p_{ii} das Skalarprodukt der i -ten Zeile von P und der i -ten Spalte von P , so dass

$$p_{ii} = \sum_{j=1}^m p_{ij}^2 = p_{ii}^2 + \sum_{i \neq j} p_{ij}^2 \quad (3.200)$$

folgt. Man sieht leicht, dass diese Beziehung nicht gelten kann, wenn die $p_{ij} > 1$ sein können, so dass folgt, dass

$$0 \leq p_{ij} \leq 1. \quad (3.201)$$

Wegen der Symmetrie von P ist die Summe der Eigenwerte gleich der Summe der Diagonalelemente p_{ii} . Da die Eigenwerte entweder den Wert 0 oder 1 haben, folgt

$$\sum_{i=1}^m p_{ii} = n \quad (3.202)$$

und n ist der Rang der (m, n) -Datenmatrix X mit $n \leq m$, – diese muss vollen Rang haben, da sonst die Inverse $(X'X)^{-1}$ und damit P nicht existierte. Aus (3.200) folgt, dass $p_{ii} = 0$ oder $p_{ii} = 1$, wenn $p_{ij} = 0$ für alle i und j . Ist $p_{ii} = 0$, so wird \hat{y}_i durch keine andere Beobachtung y_j beeinflusst. Gilt andererseits $p_{ii} = 1$, so folgt $p_{ii} = y_i$, – in diesem Fall passt das Regressionsmodell perfekt, es ist fehlerfrei. Hoaglin & Welsch (1978) liefern Beispiele für derartige Analysen.

3.7.4 Projektionen auf Hauptachsen

Es sei X eine (m, n) -Matrix von Messwerten; x_{ij} sei der Messwert des i -ten Objects ("Person") für die j -te Variable ("Test"), $1 \leq i \leq m$ und $1 \leq j \leq n$. Für X gilt die Singularwertzerlegung $X = Q\Lambda^{1/2}T' = LP'$ mit $L = Q\Lambda^{1/2}$. Wegen der Orthonormalität von T folgt $XT = L$. Das Element ℓ_{ik} von L ist die Koordinate des i -ten Falls auf der k -ten latenten Dimension und ergibt sich als Skalarprodukt des i -ten Zeilenvektors $\tilde{\mathbf{x}}'_i$ von X ($\tilde{\mathbf{x}}_i$ ist Spaltenvektor von X') und der k -ten Spalte \mathbf{t}_k von T , d.h. es ist

$$\ell_{ik} = \tilde{\mathbf{x}}'_i \mathbf{t}_k. \quad (3.203)$$

ℓ_{ik} ergibt sich als Projektion des Vektors $\tilde{\mathbf{x}}_i$ auf eine Gerade durch den Nullpunkt des Koordinatensystems mit dem Orientierungsvektor \mathbf{t}_k , der als Eigenvektor von $X'X$ die Länge 1 hat, $\|\mathbf{t}_k\| = 1$. Tatsächlich ist nach (1.60), Seite 24,

$$\tilde{\mathbf{x}}_{i, \mathbf{t}_k} = \frac{\tilde{\mathbf{x}}'_i \mathbf{t}_k}{\mathbf{t}'_k \mathbf{t}_k} \mathbf{t}_k,$$

und

$$\|\tilde{\mathbf{x}}_{i, \mathbf{t}_k}\| = |\tilde{\mathbf{x}}'_i \mathbf{t}_k| \|\mathbf{t}_k\| = |\tilde{\mathbf{x}}'_i \mathbf{t}_k|,$$

und dies ist der Betrag des Skalarprodukts ℓ_{ik} in (3.203) (vergl. Gleichung (1.61), Seite 24).

4 Funktionenräume und PCA

4.1 Einführung

Vielfach ist die Untersuchung zeitlicher oder räumlicher Verläufe von Interesse; hier werden nur zeitliche Verläufe betrachtet, viele der zur Analyse dieser Verläufe eingeführten Begriffsbildungen übertragen sich auf räumliche Verläufe. Man betrachtet *Funktionenräume* \mathcal{F} , d.h. Mengen von Funktionen, die über einem bestimmten Bereich D definiert sind. Linearkombinationen von Funktionen lassen sich analog zu den bisher betrachteten Linearkombinationen definieren: sind f, g Elemente eines Funktionenraums \mathcal{F} , so soll auch $\lambda f + \mu g \in \mathcal{F}$ gelten; ist außerdem für alle $f, g \in \mathcal{F}$ außerdem das Skalarprodukt zweier Funktionen erklärt⁶³

$$\langle f, g \rangle = \int_D f(t)g(t)dt, \quad f, g \in \mathcal{F} \quad (4.1)$$

(d.h. existiert das Integral), so hat man über

$$\|f\|^2 = \int_D |f(t)|^2 dt < \infty \quad (4.2)$$

⁶³Die an $\mathbf{x}'\mathbf{y}$ angelehnte Schreibweise $f'g$ ist suboptimal, weil f' mit der ersten Ableitung df/dt der Funktion f verwechselt werden kann.

auch eine Norm $\|f\|$ erklärt und \mathcal{F} ist ein Vektorraum. Es sei $\{\phi_1, \phi_2, \dots\}$ eine Menge von Funktionen aus \mathcal{F} derart, dass

$$f(t) = \sum_{i=1}^{\infty} a_i \phi_i(t) \quad (4.3)$$

gilt, wobei die a_i für f spezifische Koeffizienten sind. Die ϕ_i heißen dann *Basisfunktionen*. Werden zur Darstellung von f unendlich viele Basisfunktionen benötigt, so heißt der Vektorraum *unendlich-dimensional*. Wie Basisvektoren sind Basisfunktionen linear unabhängig. Gilt für die Basisfunktionen ϕ_i , $i = 1, 2, 3, \dots$

$$\langle \phi_i, \phi_j \rangle = \int_D \phi_i(t) \phi_j(t) dt = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (4.4)$$

so heißen die ϕ_i *orthogonal*; δ_{ij} heißt *Kronecker-Delta*.

Beispiel 4.1 Fourier-Reihen⁶⁴ Eine bekannte Reihenentwicklung einer Funktion $f(t)$ ist die Fourier-Reihe

$$f(t) = a_0 + \sum_{n=1}^{\infty} (a_n \cos(nt) + b_n \sin(nt)), \quad (4.5)$$

wobei die Koeffizienten a_n und b_n durch

$$a_0 = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) dt \quad (4.6)$$

$$a_n = \frac{1}{\pi} \int_{-\infty}^{\infty} f(t) \cos(nt) dt \quad (4.7)$$

$$b_n = \frac{1}{\pi} \int_{-\infty}^{\infty} f(t) \sin(nt) dt \quad (4.8)$$

gegeben sind. Die cos- und sin-Funktionen sind orthogonal, denn wie gezeigt werden kann gilt

$$\int_{-T/2}^{T/2} \cos(m\omega_0 t) \cos(n\omega_0 t) dt = \begin{cases} 0, & m \neq n \\ T/2, & m = n \neq 0 \end{cases} \quad (4.9)$$

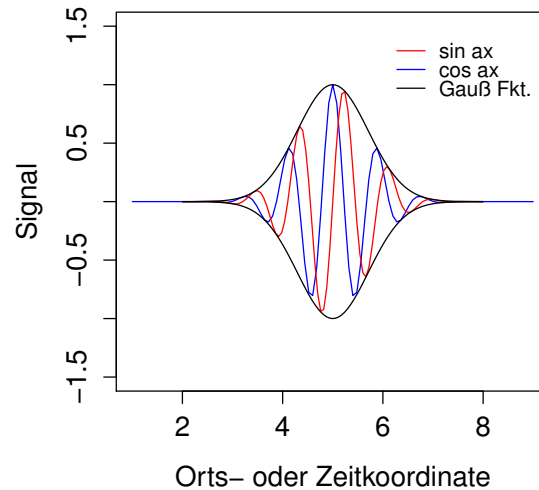
$$\int_{-T/2}^{T/2} \sin(m\omega_0 t) \sin(n\omega_0 t) dt = \begin{cases} 0, & m \neq n \\ T/2, & m = n \neq 0 \end{cases} \quad (4.10)$$

$$\int_{-T/2}^{T/2} \sin(m\omega_0 t) \cos(n\omega_0 t) dt = 0, \quad \text{für alle } m \text{ und } n \quad (4.11)$$

Ein wichtiges Anwendungsgebiet der Fourier-Entwicklung ist die Signalanalyse; hier eignet sich die Fourier-Entwicklung insbesondere dann, wenn das Signal $f(t)$ stationär ist. Für nicht-stationäre Signale sind andere Entwicklungen, etwa auf der Gabor-Wavelet-Basis, von größerem Nutzen (s. Abbildung 19 und die folgende Erläuterung). \square

⁶⁴Joseph Fourier, (1768 – 1830), französischer Mathematiker und Physiker

Abbildung 19: Gabor-Funktion



Wavelets cos- und sin-Funktionen haben den Nachteil, auf $(-\infty, \infty)$ definiert zu sein. Es ist aber oft notwendig, örtlich und zeitlich begrenzte Signale zu repräsentieren. Für diesen Zweck haben sich andere Basisfunktionen als die cos- und sin-Funktionen als geeigneter erwiesen. So können Gauß-Funktionen $\exp(-(t - t_n)^2)$, $n = 1, 2, \dots$ verwendet werden (Gishtasby & O'Neill (1994), Calcaterra (2008)), oder Gabor-Funktionen:

$$f_n(t) = \exp\left(-\frac{(x - x_n)^2}{2\sigma^2}\right)\phi(k\omega_0 t), \quad (4.12)$$

wobei ϕ eine Sinus- oder Kosinus-Funktion ist (Gabor (1948)), vergl. Abbildung 19. Weitere Basisfunktionen sind Polynome (Hermite- oder Laguerre-Polynome); die Diskussion der verschiedenen Möglichkeiten geht über den Rahmen dieses Skripts hinaus.

Unter Umständen lassen sich die Basisfunktionen auch empirisch bestimmen. Dies geschieht bei der Karhunen-Loève-Analyse, auf die im Folgenden eingegangen werden soll.

4.2 Karhunen-Loève-Entwicklung und PCA

Die KL-Entwicklung dient der Charakterisierung und Interpretation stochastischer Prozesse. Man stelle sich eine Untersuchung vor, bei der das Ergebnis eines experimentellen Durchgangs ein bestimmter Werteverlauf $\omega \in \Omega$ innerhalb eines Zeitintervalls $D = [a, b]$; statt eines Zeitintervalls kann natürlich auch ein Ortsbereich betrachtet werden, im Folgenden werden allerdings nur zeitliche Intervalle

betrachtet. Ω ist der Stichprobenraum, d.h. die Menge der möglichen Verläufe. Wie bei der Betrachtung zufälliger Veränderlicher werden 'Ereignisse', d.h. Klassen bestimmter Verläufe, durch Teilmengen von Ω , definiert; diese Teilmengen bilden eine Sigma-Algebra Σ , und die Wahrscheinlichkeiten, mit denen Ereignisse eintreten, werden durch ein Wahrscheinlichkeitsmaß P festgelegt. Mit (Ω, Σ, P) ist dann der Wahrscheinlichkeitsraum der Untersuchung festgelegt.

Jedem $\omega \in \Omega$ wird nun eine Funktion der Zeit zugeordnet: $\omega \mapsto X(t, \omega)$, $t \in D$. $X(t, \omega)$ als Funktion der Zeit bildet den Verlauf ω ab. $X(t, \omega)$ heißt auch *Pfad* oder *Trajektorie* des Prozesses. Die Schreibweise $X(t, \omega)$ ist üblich, kann aber verwirrend sein, da $X(t, \omega)$ einen Wert des Verlaufs zu einem Zeitpunkt t meinen könnte. Papoulis (1968), p. 280, macht diesen Punkt explizit, indem er die vier möglichen Interpretationen von $X(t, \omega)$ auflistet:

1. $X(t, \omega)$ kann eine ganze Familie von Funktionen der Zeit bedeuten,
2. $X(t, \omega)$ kann eine einzelne Funktion der Zeit repräsentieren,
3. Für einen fixen Wert von t kann $X(t, \omega)$ eine zufällige Veränderliche bedeuten,
4. Schließlich kann $X(t, \omega)$ für einen fixen Wert von t und festes $\omega \in \Omega$ eine einzelne Zahl repräsentieren.

Was jeweils gemeint ist wird durch den jeweiligen Zusammenhang bestimmt. Ein stochastischer Prozess ist dann eine Familie $X_t = \{X(t, \omega)\}_{t \in D}$ von zufälligen Funktionen (Pfad, Trajektorien) über dem Zeitbereich D . Um die Notation zu vereinfachen wird im Folgenden einfach $X(t)$ statt $X(t, \omega)$ geschrieben, und der Prozess wird kurz mit $X_t = \{X(t)\}_{t \in D}$ bezeichnet.

Erwartungswert und Varianz eines stochastischen Prozesses X_t sind durch

$$\mathbb{E}(X_t) = \int_{\Omega} X(t, \omega) dP(\omega) = m(t), \quad \text{Var}(X_t) = \mathbb{E}[(X_t - \mathbb{E}(X_t))^2] \quad (4.13)$$

definiert; für jeden Wert $t \in D$ wird über alle möglichen Trajektorien $X(t)$ gemittelt. $m(t)$ wird auch die *Mittelwertfunktion* des Prozesses genannt.

Es sei Y_t ein stochastischer Prozess mit der Mittelwertfunktion $\mathbb{E}(Y_t)$. Dann heißt

$$X_t = \{X_t\}_{t \in D}, \quad X_t = Y_t - m(t) \quad (4.14)$$

zentrierter stochastischer Prozess.

Definition 4.1 *Der stochastische Prozess heißt stetig im quadratischen Mittel, wenn*

$$\mathbb{E}[(X_{t+\varepsilon} - X_t)^2] = 0. \quad (4.15)$$

Es sei X_t ein zentrierter stochastischer Prozess; dann ist

$$R_X(s, t) = \mathbb{E}[X(s)X(t)], \quad s, t \in D \quad (4.16)$$

die *Autokorrelationsfunktion* von X_t .

Satz 4.1 *Es sei X_t ein stochastischer Prozess mit der Autokorrelationsfunktion $R_X(s, t)$. X_t ist stetig im quadratischen Mittel genau dann, wenn R_X stetig auf $D = [a, b] \times [a, b]$ ist.*

Beweis: Wong (1971). □

Der folgende Satz geht auf Loève (1945) und Karhunen (1947) zurück.

Satz 4.2 *Es sei $X_t = \{X(t) | t \in [a, b]\}$ mit $\mathbb{E}(X_t) = 0$ für alle $t \in [a, b]$ und stetiger Kovarianzfunktion $C(s, t)$, wobei die Funktionen $X(t)$ quadratintegrierbar seien. Dann gilt*

$$\hat{X}(t) = \sum_{i=1}^n a_i \phi_i(t), \quad (4.17)$$

und die

$$a_i = \int_a^b X(t) \phi_i(t) \quad (4.18)$$

sind zufällige Veränderliche⁶⁵ mit

$$\mathbb{E}(a_i) = 0, \quad \mathbb{E}(a_i a_j) = \delta_{ij} \lambda_i \quad (4.19)$$

mit

$$\delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} \quad (\text{Kronecker-Delta})$$

sind; die ϕ_i sind die orthonormalen Eigenfunktionen von $C(s, t)$, mit den λ_i als dazu korrespondierenden Eigenwerten. Weiter gilt

$$\lim_{n \rightarrow \infty} \hat{X}(t) = X(t). \quad (4.20)$$

Beweis: Der Ansatz (4.17) gilt für eine willkürliche Funktion $f(t)$, also insbesondere für eine zufällige Funktion $X(t)$ aus der Menge X_t ; da $X(t)$ zufällig ist, ist a_i eine zufällige Veränderliche. Es gilt

$$\mathbb{E}(a_i) = \mathbb{E} \left[\int_a^b X_t \phi_i(t) \right] = \int_a^b \mathbb{E}[X_t] \phi_i(t) = 0, \quad (4.21)$$

da ja $\mathbb{E}[X_t] = 0$. Nach Mercers Theorem gilt $C(s, t) = \sum_{i=1}^{\infty} \lambda_i \phi_i(s) \phi_i(t)$ und ϕ_i

⁶⁵ a_i ist zufällig weil $X(t)$ eine zufällige Funktion aus der Menge der zufälligen Funktionen ist, die den stochastischen Prozess X_t definieren.

und ϕ_j sind orthonormal, $i \neq j$. Dann folgt

$$\begin{aligned}
\mathbb{E}[a_i a_j] &= \mathbb{E} \left[\int_a^b \int_a^b X_s X_t \phi_i(s) \phi_j(t) ds dt \right] \\
&= \int_a^b \int_a^b \mathbb{E}[X_s X_t] \phi_i(s) \phi_j(t) ds dt \\
&= \int_a^b \int_a^b k(s, t) \phi_i(s) \phi_j(t) ds dt \\
&= \int_a^b \phi_i(s) \left(\int_a^b k(s, t) \phi_j(t) dt \right) ds \\
&= \lambda_i \int_a^b \phi_i(s) \phi_j(s) ds \\
&= \delta_{ij} \lambda_i,
\end{aligned} \tag{4.22}$$

δ_{ij} das Kronecker-Delta. Es sei

$$\varepsilon_n(t) = \mathbb{E} \left[\left(X(t) - \sum_{i=1}^n a_i \phi_i(t) \right)^2 \right]. \tag{4.23}$$

Dann folgt

$$\varepsilon_n(t) = \mathbb{E}[X^2(t)] - 2\mathbb{E} \left[X(t) \sum_{i=1}^n a_i \phi_i(t) \right] + \mathbb{E} [a_i a_j \phi_i(t) \phi_j(t)] \tag{4.24}$$

Es ist $\mathbb{E}(X^2(t)) = C(t, t)$, und

$$\begin{aligned}
\mathbb{E} \left[X(t) \sum_{i=1}^n a_i \phi_i(t) \right] &= \mathbb{E} \left[X(t) \sum_{i=1}^n \int_D X(s) \phi_i(s) ds \phi_i(t) \right] \\
&= \sum_{i=1}^n \left(\int_D \mathbb{E}[X(s) X(t)] \phi_i(s) ds \right) \phi_i(t) \\
&= \sum_{i=1}^n \left(\int_D C(s, t) \phi_i(s) ds \right) \phi_i(t) \\
&= \sum_{i=1}^n \lambda_i \phi_i^2(t),
\end{aligned}$$

und schließlich folgt analog

$$\mathbb{E} \left[\sum_{i=1}^n a_i a_j \phi_i(t) \phi_j(t) \right] = \sum_{i=1}^n \lambda_i \phi_i^2(t).$$

Also folgt

$$\varepsilon_n(t) = C(t, t) - \sum_{i=1}^n \lambda_i \phi_i(t) \phi_i(t),$$

und wegen Mercers Theorem folgt

$$\lim_{n \rightarrow \infty} \varepsilon_n(t) = 0$$

d.h. $\hat{X}(t) \rightarrow X(t)$. □

$X(t)$ **als Zeitreihe:** Erhebt man $X(t)$ als kontinuierliche Funktion über $D = [a, b]$, so muss man die Integralgleichung

$$\int_D k(s, t) \phi_i(t) dt = \lambda_i \phi_i(s)$$

für $i = 1, 2, \dots, k$ lösen, d.h. die Eigenfunktionen $\phi_i(t)$ für alle $t \in [a, b]$ bestimmen. Das ist im Allgemeinen sehr aufwendig. Einfacher wird die Aufgabe, wenn $X(t)$ für diskrete Werte t_1, \dots, t_N bestimmt wird. Man erhält dann den Vektor

$$X = (X_1, \dots, X_N)'$$

für den die (m, n) -Matrix C der Autokorrelationen berechnet werden kann. Man erhält man dann die Gleichung

$$C \vec{\phi}_j = \lambda_j \vec{\phi}_j, \quad j = 1, \dots, N \quad (4.25)$$

$\vec{\phi}_j$ ist jetzt ein N -dimensionaler (Eigen-)Vektor mit λ_j als zugehörigem Eigenwert. $\vec{\phi}_j$ repräsentiert die entsprechende Eigenfunktion an den Stellen t_1, \dots, t_N , und natürlich sind die Eigenvektoren $\vec{\phi}_j$ paarweise orthonormal. Gesucht ist die Entwicklung für $X = (X(t_1), X(t_2), \dots, X(t_N))'$ als Reihe:

$$X = \sum_{j=1}^N a_j \phi_j(t),$$

mit $\phi_j = (\phi_j(t_1), \dots, \phi_j(t_N))'$. Die Koeffizienten a_j erhält man durch orthogonale Basisentwicklung:

$$\phi_j' X = a_j, \quad j = 1, \dots, N \quad (4.26)$$

Natürlich ist es von Interesse, weniger als N Eigenfunktionen zu wählen, – gesucht ist ja die sparsamste Repräsentation der Daten. Also betrachtet man

$$\hat{X} = \sum_{j=1}^r a_j \phi_j(t), \quad r < N \quad (4.27)$$

Die Güte der Abschätzung \hat{X} läßt sich durch den Quotienten

$$\frac{\sum_{j=1}^r \lambda_j}{\sum_{j=1}^N \lambda_j} \geq \alpha \quad (4.28)$$

bewerten, wobei α ein festgesetzter Anteil an erklärter Varianz ist.

5 Anhang

5.1 Punkträume

Mit dem Symbol \mathbb{R} wird die Menge der reellen Zahlen bezeichnet⁶⁶. 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

$$\mathbb{R}^n = \{(x_1, x_2, \dots, x_n) | x_1, x_2, \dots, x_n \in \mathbb{R}\}, \quad n \in \mathbb{N} \quad (5.1)$$

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.

Anmerkung: Gelegentlich findet man für einen Punkt x auch die Schreibweise $x = (x_1 | x_2 | \dots | x_n)$, um ihn vom Vektor $\mathbf{x} = (x_1, x_2, \dots, x_n)'$ zu unterscheiden.

□

Koordinatenachsen lassen sich skalieren, d.h. mit einem Faktor multiplizieren (wenn man etwa von Zentimetern zu Millimetern übergeht) und die Koordinaten von Punkten lassen sich addieren: Sind $x = (x_1, \dots, x_n)$ und $y = (y_1, \dots, y_n)$ irgendzwei Punkte, so soll

1. $ax = (ax_1, ax_2, \dots, ax_n)$, für $a \in \mathbb{R}$, und

2. $x + y = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n)$

gelten. $x + y$ definiert also ebenfalls einen Punkt. Es gilt der folgende

Satz 5.1 *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 $a, b \in \mathbb{R}$,

⁶⁶Das ist die Vereinigung (i) der natürlichen Zahlen $\mathbb{N} = 0, 1, 2, \dots$, (ii) die ganzen Zahlen $\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$, (iii) die rationalen Zahlen $\mathbb{Q} = \{p/q | p, q \in \mathbb{Z}, q \neq 0\}$, und (iv) die irrationalen Zahlen, die sich nicht als Quotient p/q mit $p, q \in \mathbb{Z}$ darstellen lassen (lat. ratio = Bruch, Quotient). Bekannte irrationale Zahlen sind π , die Eulersche Zahl e (Basis des natürlichen Logarithmus), $\sqrt{2}$, etc. Die Dezimaldarstellung einer irrationalen Zahl ist nicht periodisch und bricht nicht ab. Es läßt sich zeigen, dass zwischen irgendzwei rationalen Zahlen stets beliebig viele (genauer: überabzählbar viele) irrationale Zahlen liegen. \mathbb{R} bildet ein Kontinuum, also eine lückenlos zusammenhängende Menge von Zahlen.

Beweis: Es genügt ein einfaches Nachrechnen. \square

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

Definition 5.1 Die Distanz $d(x, y)$ ist eine Funktion der Differenzen $x_i - y_i$, $i = 1, \dots, n$ der Koordinaten x_i, y_i der Punkte x und y ; genügen die Distanzen den folgenden Axiomen

1. $d(x, y) \geq 0$,
2. $d(x, y) = d(y, x)$ (*Reflexivität*)
3. $d(x, z) \leq d(x, y) + d(y, z)$ (*Dreiecksungleichung*),

so definiert d eine Metrik des Raumes, in dem die Punkte x und y liegen.

Die Distanzen zwischen den Punkten eines Raumes können auf verschiedene Weisen definiert sein, und dementsprechend ist ein Raum durch eine bestimmte Metrik definiert, die die geometrische Struktur des Raumes kennzeichnet. So ist nach Euklid ist die kürzeste Verbindung zwischen zwei Punkten eine Gerade. Für irgendzwei Punkte x und y gilt dann

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} = \left[\sum_{i=1}^n (x_i - y_i)^2 \right]^{\frac{1}{2}}. \quad (5.2)$$

Hier wird implizit angenommen, dass die x_i und y_i Koordinaten in einem rechtwinkligen Koordinatensystem sind, und $x_i - y_i$ ist die Differenz der Koordinaten der Punkte x und y auf der i -ten Koordinate. Die Gleichung (5.2) ist dann der Satz des Pythagoras für den n -dimensionalen Raum. Ist die Metrik eines Raumes durch (5.2) definiert, so heißt der Raum *euklidisch*. Die euklidische Metrik ist nicht die einzige mögliche Metrik, wie weiter unten illustriert wird.

Die Axiome in Definition 5.1 definieren die allgemeinen Eigenschaften einer Metrik. Die Forderung, dass eine Distanz $d(x, y)$ nicht negativ sein darf, entspricht dem umgangssprachlichen Begriff von "Distanz": die Länge eines Weges von Ort A zu Ort B kann nicht negativ sein. Die Forderung der Reflexivität $d(x, y) = d(y, x)$ stellt eine Einschränkung dar: will man in einer Stadt mit dem Auto von der Adresse A_1 zu Adresse A_2 fahren, so kann wegen eines Einbahnstrassensystems die Distanz $d(A_1, A_2)$ größer als die Distanz $d(A_2, A_1)$ sein. Die Dreiecksungleichung ist wiederum intuitiv einleuchtend. Sind A_1, A_2 und A_3 drei Adressen in einer Stadt, so ist es möglich, dass es zwischen A_1 und A_3 keinen direkten Weg gibt und man immer über die Adresse A_2 fahren muss; dann gilt eben $d(A_1, A_3) = d(A_1, A_2) + d(A_2, A_3)$. Es wird aber nie $d(A_1, A_3) > d(A_1, A_2) + d(A_2, A_3)$ gelten, d.h. es muss $d(A_1, A_3) \leq d(A_1, A_2) + d(A_2, A_3)$ gelten.

Euklidische Räume wurden lange Zeit als "natürliche" Räume betrachtet. Isaac Newton (1642 – 1727) nahm implizit eine euklidische Struktur des physikalischen Raumes an, und der Philosoph Immanuel Kant (1724 – 1804) erklärte,

der euklidische Raum sei eine notwendige Vorstellung a priori. So notwendig wie von Kant angenommen ist diese Vorstellung allerdings nicht, schon in der ersten Hälfte des 19-ten Jahrhunderts schlugen der ungarische Mathematiker János Bolyai (1802 – 1860), der russische Mathematiker Nikolai Iwanowitsch Lobatschewski (1792 – 1856) sowie der Mathematiker Carl Friederich Gauß (1777 – 1855) nicht-euklidische Geometrien vor. Der Mathematiker Bernhard Riemann (1826 – 1866) hielt 1850 seinen Habilitationsvortrag über eine nicht-euklidische Geometrie, die später für die Weiterentwicklung der Relativitätstheorie wichtig wurde. Der Mathematiker Hermann Minkowski (1864 – 1909) modifizierte ebenfalls im Zusammenhang mit der Relativitätstheorie die Begriffe von Raum und Zeit (Raum-Zeit-Kontinuum) und definierte eine Metrik – die später nach ihm benannte Minkowski-Metrik –, die durch

$$d(x, y) = \left[\sum_{i=1}^n |x_i - y_i|^p \right]^{\frac{1}{p}}, \quad 0 < p \in \mathbb{R} \quad (5.3)$$

definiert ist. Für $p = 2$ ergibt sich die euklidische Metrik (5.2). Die Minkowski-Metrik kann als Verallgemeinerung der euklidischen Metrik angesehen werden. Für $p = 1$ ergibt sich die *City-Block-* oder *Manhattan-Metrik*, weil man von A_1 nach A_2 gelangt, indem man rechtwinklig zueinander liegende Strassenabschnitte durchfahren oder durchlaufen muss. Die Minkowski-Metrik erlaubt es, *psychologische Distanzen*, etwa zwischen Begriffen, Stereotypen etc, zu modellieren, für die sich die euklidische Metrik oft als inadäquat erweist. Die Minkowski-Metrik wird im Zusammenhang mit der multidimensionalen Skalierung behandelt.

In vielen Untersuchungen werden an jeweils einer Person (allgemein: einem "Fall", und ein Fall muss keine Person sein) mehrere Merkmale gemessen und man berechnet die Korrelationen zwischen den Merkmalen. Man kann die Messungen als Koordinaten eines Punktes interpretieren, der dann den Fall repräsentiert. Auf diese Weise entsteht eine Punktekonfiguration ("Punktwolke"). Die Distanzen zwischen den Punkten reflektieren die Relationen zwischen den Fällen. Mit dem Distanzbegriff ist aber der Begriff der Orientierung nicht verbunden, der wiederum für Fragen der Interpretation von Interesse ist. Deswegen wird der Begriff des Vektors eingeführt. Sind P und Q zwei Punkte der Konfiguration, so sei die euklidische Distanz durch $d(P, Q)$ gegeben. Zu dieser Distanz korrespondiert ein Vektor \overrightarrow{PQ} , der durch eine Länge und eine Orientierung definiert ist; die Länge ist durch die euklidische Distanz $d(P, Q)$ gegeben, und die Orientierung der durch $d(P, Q)$ definierten Geraden ist durch die Winkel zwischen dieser Geraden und den Koordinatenachsen gegeben. Für maximal drei Dimensionen (3 Variablen) kann der Vektor graphisch durch einen Pfeil repräsentiert werden (s. Abbildung 1). Die Orientierung hängt von gewählten Koordinatensystem ab. Insbesondere lassen sich Abhängigkeiten zwischen Variablen leicht mittels des Vektorbegriffs darstellen und latente, also nicht direkt gemessene Variablen, zur Erklärung von Korrelationen zwischen Variablen bestimmen. Analog zum Punkt-raum kann dann der Begriff des Vektorraums eingeführt werden, der isomorph

zum jeweiligen Punktraum ist. Je nach Perspektive macht man in der multivariaten Analyse sowohl vom Begriff des Punkt- wie des Vektorraums Gebrauch. In den folgenden Abschnitten wird der Begriff des Vektors und der des Vektorraums eingeführt und einige Resultate aus der Vektor- und Matrixrechnung vorgestellt, soweit sie für die üblichen multivariaten Verfahren notwendig sind. Neuere Verfahren wie zum Beispiel die Klassifikation von Mustern oder Objekten anhand von "Support Vector Machines" erfordern mathematische Grundlagen, die in einem gesonderten Skript vorgestellt werden.

5.2 Skalen und Abbildungen

X sei eine (m, n) -Datenmatrix, bei der die n Spalten die gemessenen Variablen repräsentieren. Die Spaltenvektoren der Matrix V der Eigenvektoren von $X'X$ definieren die Orientierung der Geraden L_k , und die Komponenten der Vektoren \mathbf{u}_k sind die Projektionen der Fälle auf L_k , und die $\lambda_k = \mathbf{u}'_k \mathbf{u}_k$ repräsentieren die Varianz der Komponenten von \mathbf{u}_k . Ist λ_{k+1} "klein" im Verhältnis zu den $\lambda_1, \dots, \lambda_k$, so wird man die Dimensionen L_{k+1}, \dots, L_n als nicht mehr relevante Dimensionen ansehen. Als Eigenwerte von $X'X$ werden die λ_k von den Maßeinheiten der Variablen V_1, \dots, V_n abhängen. Sind die V_j Variablen von derselben Art, etwa Temperaturen, die z.B. an verschiedenen Orten (Variable) zu verschiedenen Zeitpunkten (Fälle) gemessen werden, so kann man sie alle auf eine Skala transformieren, wenn etwa an einigen Orten in Fahrenheit, an anderen Orten in Celsius gemessen wurde. Anders ist es, wenn die Variablen verschiedene Merkmale wie Körperlänge und Körpergewicht repräsentieren. Für eine gegebene Längenskala gibt es keine Gewichtsskala, die der Längenskala auf "natürliche" Weise entspräche: hat man die Länge in Zentimetern gemessen, so gibt es keinen Grund, weshalb das Gewicht in Gramm gemessen werden sollte, man kann es ebensogut in Kilogramm, Zentnern oder britischen Stones messen. Die verschiedenen Skalen unterscheiden sich durch einen Faktor. Man kann sich den Effekt der Skalen auf die Eigenwerte so veranschaulichen, dass man den Faktor a_1 etwa für die \mathbf{u}_1 -Skala gleich 1 setzt, $a_1 = 1$, und den für \mathbf{u}_2 gleich $a_2 \neq a_1$. Der Fall $a_1 = a_2$ reflektiert den Fall, dass Variablen gleichen Typs (z.B. Temperaturen) auf derselben Skala (z.B. Celsius) gemessen werden. Im Falle verschiedener Merkmale sollte man zu standardisierten Messungen $z = (x - \bar{x})/s_x$ übergehen: $x - \bar{x}$ und s_x haben stets dieselbe Maßeinheit, die sich im Ausdruck für z herauskürzt, d.h. z ist an keine Maßeinheit gekoppelt. Wie die Folgerung $zs = x - \bar{x}$ zeigt drückt der z -Wert die Abweichung $x - \bar{x}$ eines x -Werts vom Mittelwert \bar{x} als Anteil (der auch größer als 1 sein kann) der Standardabweichung s aus, alternativ dazu kann man sagen, dass $x - \bar{x}$ z s -Einheiten beträgt. Eine Veranschaulichung liefert die Tchebycheffsche Ungleichung

$$P(|x - \mu| \geq k) \leq \frac{\sigma^2}{k^2}, \quad k > 0 \quad (5.4)$$

wobei $\mu = \mathbb{E}(x)$ der Erwartungswert und $\sigma^2 = \mathbb{E}(x - \mu)^2$ die Varianz von x ist. Setzt man $k = z\sigma$, so erhält man

$$P(|x - \mu| > z\sigma) \geq \frac{\sigma^2}{z^2\sigma^2} = \frac{1}{z^2} \quad (5.5)$$

Für $z = 3$ erhält man

$$P(|x - \mu| \geq 3\sigma) \leq \frac{1}{9} \approx .11,$$

und für $z = 4$ erhält man

$$P(|x - \mu| \geq 4\sigma) \leq \frac{1}{16} \approx .03,$$

d.h. nahezu 90% der $x - \mu$ -Werte liegen innerhalb des Intervalls $(-3s, 3s)$, und 97% der Abweichung von μ liegen innerhalb des Intervalls $(-4s, 4s)$. Nahezu die gesamte Masse der Abweichungen $x - \mu$ korrespondiert zu z -Werten zwischen $z = -4$ und $z = 4$, – unabhängig von der speziellen Skala und unabhängig von der speziellen Wahrscheinlichkeitsverteilung bzw. -dichte. Die Eigenwerte λ_1 und λ_2 als Varianzen der Projektionen der Fälle auf L_1 bzw. L_2 sind unabhängig von den Einheiten, in denen X_1 und X_2 gemessen wurden, sie reflektieren aber, welcher Anteil der Gesamtvariation zu Lasten von L_1 bzw. von L_2 geht.

Die Abbildung 20 illustriert den Effekt verschiedener Maßeinheiten auf die Form der Punktekonfiguration. Betrachtet werden Konfigurationen von Fällen, die als Stichproben einer 2-dimensionalen Gaußverteilung mit den Erwartungswerten $\mu_1 = \mu_2 = 0$ und den Varianzen $\sigma_1^2 = \sigma_2^2 = 1$ gewonnen wurden (Package `mvrnorm`⁶⁷. Der Effekt unterschiedlicher Skalen wurde erzeugt durch Multiplikation der x - bzw. y -Werte mit einem Faktor fx bzw. fy . Die Tabelle 1 enthält die Eigenwerte für die verschiedenen Konfigurationen.

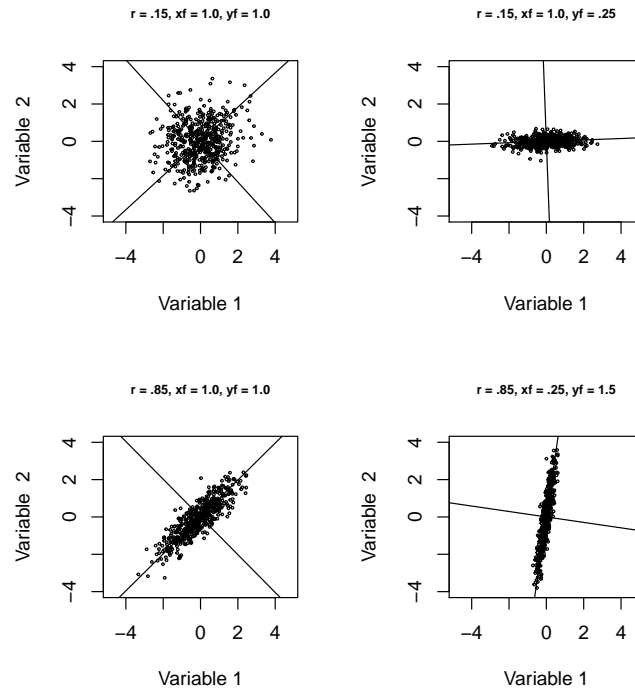
Tabelle 1: Eigenwerte für verschiedene fx, fy -Faktoren; λ_1, λ_2 Eigenwerte der Kovarianzmatrix

$r = .15$	$r = .15$	$r = .85$	$r = .85$
$fx = fy = 1$	$fx = 1, fy = .25$	$xf = fy = 1$	$fy = .25, yf = 1.5$
$\lambda_1 = 1.140$	$\lambda_1 = .980$	$\lambda_1 = 1.905$	$\lambda_1 = 2.330$
$\lambda_2 = .861$	$\lambda_2 = .063$	$\lambda_2 = .142$	$\lambda_2 = .016$
$\lambda_1/\lambda_2 = 1.323$	$\lambda_1/\lambda_2 = 15.660$	$\lambda_1/\lambda_2 = 13.443$	$\lambda_1/\lambda_2 = 142.998$

Werden die verschiedenen Variablen in verschiedenen Einheiten gemessen, so ergibt sich ein Problem. Ein einfache Variante dieser Situation liegt vor, wenn dieselbe Größe, aber auf verschiedenen Skalen gemessen wird. Ein Beispiel ist die Messung von Temperaturen an verschiedenen Orten; die "Fälle" seien Zeitpunkte.

⁶⁷R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Abbildung 20: Effekt verschiedener Skalen



Am Ort A werde die Temperatur auf einer Celsius-Skala, an Ort B auf einer Fahrenheit-Skala gemessen. Die beiden Skalen sind durch eine lineare Beziehung $F = aC + b$ aufeinander bezogen. Bekanntlich gilt dann $Var(F) = a^2 Var(C)$, mmit $a = 1.8, b = 32$, d.h. $Var(F) = 3.24 Var(C)$, oder $Var(C) = .31 Var(F)$

5.3 Zur geometrischen Definition des Skalarprodukts

In Gleichung (1.43) wurde der Ausdruck

$$\mathbf{x} \cdot \mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos \theta$$

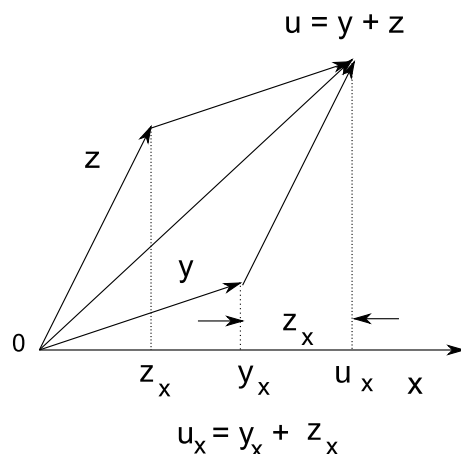
als alternative (geometrische) Definition des Skalarprodukts eingeführt. Zu zeigen ist, dass aus ihr folgt, dass auch $\|\mathbf{x}\| \|\mathbf{y}\| \cos \theta = \sum_i x_i y_i$ gilt. Dazu muß zuerst gezeigt werden, dass das Distributivgesetz für Skalarprodukte gilt, d.h. es soll

$$\mathbf{x}(\mathbf{y} + \mathbf{z}) = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z} \quad (5.6)$$

gelten, wobei $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ sind, vergl. Abbildung 21. Nach (1.61)), Seite 24 gilt

$$\|\mathbf{x}_y\| = |\mathbf{x}'\mathbf{y}| \|\mathbf{y}\| = |\mathbf{x}'\mathbf{y}|,$$

Abbildung 21: Distributivgesetz für die geometrische Definition des Skalarprodukts



d.h. \mathbf{x}_y ist gleich dem Skalarprodukt $\mathbf{x} \cdot \mathbf{y}$. Insbesondere werde die Projektion des Vektors $\mathbf{y} + \mathbf{z}$ auf den Vektor \mathbf{x} (Abb. 21) betrachtet. Wie die Abbildung zeigt, gilt

$$(\mathbf{y} + \mathbf{z})_x = \mathbf{x}_y + \mathbf{x}_z, \quad (5.7)$$

d.h. aber

$$\mathbf{x}(\mathbf{y} + \mathbf{z}) = \mathbf{x}(\mathbf{y} + \mathbf{z})_x = \mathbf{x}y_x + \mathbf{x}z_x = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z},$$

d.h. das Distributivgesetz gilt auch für die geometrische Definition des Skalarprodukts.

5.4 Elementarmatrizen und elementare Operationen

Es werden die *elementaren Spaltenumformungen* einer beliebigen (m, n) -Matrix U eingeführt:

1. Die Multiplikation eines beliebigen Zeilenvektors von $\tilde{\mathbf{u}}_i$ mit einer Zahl $0 \neq \lambda \in \mathbb{R}$,
2. Addition des Vektors $c\tilde{\mathbf{u}}_i$, $c \in \mathbb{R}$, zu einem beliebigen Zeilenvektor $\tilde{\mathbf{u}}_k$ von U ,
3. Vertauschung von irgendzwei Zeilenvektoren $\tilde{\mathbf{u}}_i$ und $\tilde{\mathbf{u}}_k$ von U .

Die elementaren Spaltenumformungen sind analog definiert (man muss nur den Ausdruck Zeilenvektor' durch den Ausdruck 'Spaltenvektor' ersetzen).

Satz 5.2 Die Anwendung elementarer Umformungen auf die Matrix M verändert nicht den Rang r von M .

Beweis: Einen knappen Beweis für diesen Satz findet man u.a. in Lorenz, Band I, (1988), p. 49.

Definition 5.2 Die⁶⁸ (n, n) -Matrix E_{ij} enthalte nur Nullen bis auf das (i, j) -te Element, das gleich 1 ist. Dann heißt E_{ij} eine Standardmatrix.

Definition 5.3 Es sei I_n die (n, n) -Einheitsmatrix. Eine Elementarmatrix entsteht, wenn auf I_n eine der möglichen elementaren Umformungen angewendet wird; dabei entstehen drei Typen von Elementarmatrizen:

1. $I_n \rightarrow Q_i^j(\lambda)$: das (i, j) -te Element von I_n wird durch $\lambda \in \mathbb{R}$ ersetzt, d.h.⁶⁹

$$Q_i^j(\lambda) = I_n + \lambda E_{ij} \quad (5.8)$$

2. $I_n \rightarrow P_i^j$: die i -te Zeile von I_n wird mit der j -ten Zeile vertauscht, d.h.

$$P_i^j = I_n - E_{ii} - E_{jj} + E_{ij} + E_{ji}, \quad i \neq j. \quad (5.9)$$

Die Subtraktion von E_{ii} und E_{jj} bewirkt die Ersetzung der Einsen an der i -ten und j -ten Diagonale von I_n , und die Addition von E_{ij} und E_{ji} bewirkt die Ersetzung der Null durch eine Eins an der (i, j) -ten und der (j, i) -ten Position von I_n .

3. $I_n \rightarrow S_i(\lambda)$: die Eins in der i -ten Diagonale von I_n wird durch $1 \neq \lambda \in \mathbb{R}$ ersetzt:

$$S_i(\lambda) = I_n + \lambda E_{ii} - E_{ii} = I_n + (\lambda - 1)E_{ii}. \quad (5.10)$$

Statt der elementaren Zeilenumformungen kann man die analog definierten Spaltenumformungen vornehmen.

Man überprüft nun direkt die Aussagen über elementare Umformungen einer (m, n) -Matrix A :

1. A_I entstehe aus A durch Multiplikation der i -ten Zeile von A mit $\lambda \in \mathbb{R}$. Dann gilt

$$A_I = S_i(\lambda)A.$$

2. A_{II} entstehe durch der j -ten Zeile zur i -ten Zeile. Dann gilt

$$A_{II} = Q_j^i A.$$

3. A_{III} entstehe aus A durch Addition des λ -fachen der j -ten Zeile zur i -ten Zeile. Dann gilt

$$A_{III} = Q_j^i(\lambda)A.$$

3. A_{IV} entstehe aus A durch Vertauschen der i -ten mit der j -ten Zeile. Dann gilt

$$A_{IV} = P_i^j A.$$

Will man Spalten- statt Zeilenumformungen vornehmen, muss A statt von links mit einer Elementarmatrix von rechts multipliziert werden.

⁶⁸Im Folgenden wird die in <https://de.wikipedia.org/wiki/Elementarmatrix> vorgestellte Notation übernommen.

⁶⁹Die Bezeichnung der Elementarmatrizen entspricht der in Fischer (1997), p.155 gewählten.

Die Elementarmatrizen sind invertierbar und ihre Inversen sind wieder Elementarmatrizen, d.h. es gelten die Gleichungen

$$(S_i(\lambda))^{-1} = S_i\left(\frac{1}{\lambda}\right), \quad (Q_i^j)^{-1} = Q_i^j(-1) \quad (5.11)$$

$$(Q_i^j(\lambda))^{-1} = Q_i^j(-\lambda), \quad (P_i^j)^{-1} = P_i^j \quad (5.12)$$

Zum Beweis multipliziere man die rechten Seiten der Gleichungen mit den linken: es ergibt sich stets die Einheitsmatrix.

C. F. Gauß hat ein allgemeines Verfahren gefunden, mit dem man lineare Gleichungssysteme $A\mathbf{x} = \mathbf{y}$ systematisch lösen kann. Dazu wird die Matrix A durch sukzessive elementare Umformungen z. B. in eine *obere Dreiecksmatrix* B überführt:

$$B = \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & b_{nn} \end{pmatrix} \quad (5.13)$$

Da die elementaren Umformungen durch Elementarmatrizen repräsentiert werden, heißt dies, dass Elementarmatrizen B_1, \dots, B_r existieren derart, dass

$$B = B_r B_{r-1} \cdots B_1 A. \quad (5.14)$$

Dieser Sachverhalt führt auf eine Methode zur Berechnung der inversen A^{-1} , sofern diese existiert. Darüber hinaus läßt sich von B der Rang der Matrix ablesen: er ist gleich der Anzahl der Zeilen von B , die nicht nur Nullen aufweisen.

5.4.1 Lineare Gleichungen und Gauß-Algorithmus

Gegeben sei ein lineares Gleichungssystem $A\mathbf{x} = \mathbf{y}$; gesucht ist eine möglichst systematische Art und Weise, einen Lösungsvektor \mathbf{x} zu bestimmen. In diesem Abschnitt soll das gaußsche Eliminationsverfahren (auch Gauß-Algorithmus vorgestellt werden. Die Idee des Verfahrens ist, die Koeffizientenmatrix A in eine Dreiecksmatrix zu transformieren, anhand der sehr schnell eine Lösung gefunden werden kann, sofern eine Lösung existiert. Zur Illustration werde ein System mit drei Unbekannten betrachtet:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= y_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= y_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= y_3 \end{aligned} \quad (5.15)$$

Bei der Bestimmung der Dreiecksmatrix mit den Elementen \tilde{a}_{ij} wird auch der Vektor \mathbf{y} in einen Vektor $\tilde{\mathbf{y}}$ transformiert:

$$\begin{aligned} \tilde{a}_{11}x_1 + \tilde{a}_{12}x_2 + \tilde{a}_{13}x_3 &= \tilde{y}_1 \\ \tilde{a}_{22}x_2 + \tilde{a}_{23}x_3 &= \tilde{y}_2 \\ \tilde{a}_{33}x_3 &= \tilde{y}_3 \end{aligned} \quad (5.16)$$

Für dieses System findet man sofort die Lösung für x_3 , nämlich $x_3 = \tilde{y}_3/\tilde{a}_{33}$, die man dann in die zweite Gleichung einsetzen kann, die dann nur noch die Unbekannten x_1 und x_2 enthält und die damit nach x_2 aufgelöst werden kann, und die Lösungen für x_3 und x_2 können dann in die erste Gleichung eingesetzt werden, um x_1 zu bestimmen.

Um die Koeffizienten \tilde{a}_{ij} zu bestimmen, macht man von den *elementaren Umformungen* Gebrauch. Demnach bleibt eine Gleichung ja korrekt, wenn man beide Seiten mit einem Faktor multipliziert. Ebenso kann man man zwei Gleichungen addieren; es entsteht wieder eine gültige Gleichung. Diese Operationen muss man so anwenden, dass man von den a_{ij} zu den \tilde{a}_{ij} gelangt.

Die Koeffizientenmatrix in (5.16) ist eine Dreiecksmatrix

$$\tilde{A} = \begin{pmatrix} \tilde{a}_{11} & \tilde{a}_{12} & \cdots & \tilde{a}_{1n} \\ 0 & \tilde{a}_{22} & \cdots & \tilde{a}_{2n} \\ & & \ddots & \\ 0 & 0 & \cdots & \tilde{a}_{nn} \end{pmatrix}. \quad (5.17)$$

Die elementaren Umformungen, die von A zu \tilde{A} führen, verändern den Rang nicht, d.h. \tilde{A} hat denselben Rang wie A . Berechnet man die Determinante von $\tilde{A} - \lambda I_n$, so erhält man

$$|\tilde{A}| = \begin{vmatrix} \tilde{a}_{11} - \lambda & \tilde{a}_{12} & \cdots & \tilde{a}_{1n} \\ 0 & \tilde{a}_{22} - \lambda & \cdots & \tilde{a}_{2n} \\ & & \ddots & \\ 0 & 0 & \cdots & \tilde{a}_{nn} - \lambda \end{vmatrix} = (\tilde{a}_{11} - \lambda)(\tilde{a}_{22} - \lambda) \cdots (\tilde{a}_{nn} - \lambda) = 0. \quad (5.18)$$

Daraus folgt unmittelbar, dass die \tilde{a}_{jj} die Eigenwerte von \tilde{A} sind.

Die Anwendung der elementaren Umformungen macht man sich am besten durch ein konkretes Beispiel klar:

$$\begin{aligned} 2x_1 - x_2 + 3x_3 &= 1 \quad (I) \\ 3x_1 + x_2 - 2x_3 &= 0 \quad (II) \\ x_1 + x_2 + x_3 &= 3 \quad (III) \end{aligned}$$

Die Koeffizientenmatrix ist

$$\left(\begin{array}{ccc|c} 2 & -1 & 3 & 1 \\ 3 & 1 & -2 & 0 \\ 1 & 1 & 1 & 3 \end{array} \right)$$

Man beginnt, indem man die erste Gleichung (I) noch einmal anschreibt, die zweite durch $3III - II$ ersetzt und die dritte durch $2III - I$

$$\begin{aligned} 2x_1 - x_2 + 3x_3 &= 1 \\ 0 + 2x_2 + 5x_3 &= 9 \\ 0 + 3x_2 + x_3 &= 5 \end{aligned}$$

Die Koeffizientenmatrix ist hier

$$\left(\begin{array}{ccc|c} 2 & -1 & 3 & 1 \\ 0 & 2 & 5 & 9 \\ 0 & 3 & 1 & 5 \end{array} \right)$$

Ersetzt man die Gleichung (5.19) durch $3(5.19) - 2(5.19)$, so erhält man

$$\begin{aligned} 2x_1 - x_2 + 3x_3 &= 1 \\ 0 + 2x_2 + 5x_3 &= 9 \\ 0 + 0 + 17x_3 &= 17 \end{aligned}$$

mit der Koeffizientenmatrix

$$\left(\begin{array}{ccc|c} 2 & -1 & 3 & 1 \\ 0 & 2 & 5 & 9 \\ 0 & 0 & 17 & 17 \end{array} \right) \quad (5.19)$$

Hieraus ergeben sich direkt die Lösungen: $17x_3 = 17 \Rightarrow x_3 = 1$, $2x_2 + 5x_3 = 2x_2 + 5 = 9 \Rightarrow x_2 = 4/2 = 2$ und $2x_1 - 2 + 3 = 2x_1 + 1 = 1 \Rightarrow x_1 = 0$.

Man kann die ursprüngliche Matrix A und die transformierte Matrix \tilde{A} einander gegenüber stellen:

$$\left(\begin{array}{ccc} 2 & -1 & 3 \\ 3 & 1 & -2 \\ 1 & 1 & 1 \end{array} \right), \quad \left(\begin{array}{ccc} 2 & -1 & 3 \\ 0 & 2 & 5 \\ 0 & 0 & 17 \end{array} \right)$$

und sieht die Transformation von A in eine Dreiecksmatrix direkt. Die hier vorgeführte Transformation hat einen gewissen ad hoc-Charakter und mag insofern nicht zufriedenstellend erscheinen, und natürlich existieren kanonische Darstellungen des Algorithmus (Golub & van Loan (2013), Kapitel 3), auf die hier aber nicht eingegangen werden soll bzw. kann, da diese Darstellungen auf der einen Seite recht lang sind und auf der anderen nicht benötigt werden, wenn man nur an der Anwendung interessiert ist, was im Allgemeinen der Fall sein wird; Statistikpakete enthalten entsprechende Module. Hier soll nur festgestellt werden, dass A und \tilde{A} natürlich denselben Rang haben.

Im folgenden Abschnitt wird anhand der SVD von A der Teilraum \mathbf{L}_{n-r} näher bestimmt.

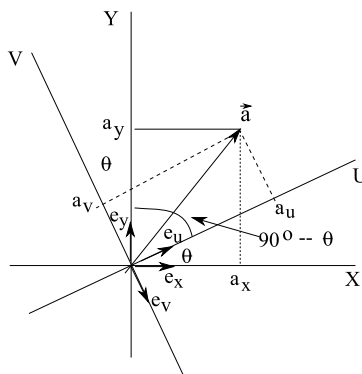
5.4.2 Herleitung einer Rotationsmatrix

Es sei (X, Y) ein Koordinatensystem für die Fälle, wobei X und Y gemessene Variablen seien. Sei

$$\mathbf{a} = \begin{pmatrix} a_x \\ a_y \end{pmatrix} = \begin{pmatrix} a_u \\ a_v \end{pmatrix},$$

wobei a_x, a_y die Komponenten von \mathbf{a} im (X, Y) - und a_u, a_v die Koordinaten im (U, V) -System seien. Das System (U, V) unterscheide sich vom System (X, Y)

Abbildung 22: Vektorrotation, $\angle(X, U) = \theta$



durch eine Rotation um den Winkel θ entgegen dem Uhrzeigersinn. Gesucht sind die Koordinaten (u, v) von \mathbf{a} im System (U, V) .

Es seien $\mathbf{e}_x = (1, 0)'$ und $\mathbf{e}_y = (0, 1)'$ die Einheitsvektoren im (X, Y) -System. Weiter seien $\mathbf{e}_u = (u_1, u_2)'$ und $\mathbf{e}_v = (v_1, v_2)'$ die Einheitsvektoren im (U, V) -System. Natürlich gilt $\|\mathbf{e}_x\| = \|\mathbf{e}_y\| = \|\mathbf{e}_u\| = \|\mathbf{e}_v\| = 1$. Für ein Skalarprodukt zweier Vektoren \mathbf{x} und \mathbf{y} gilt bekanntlich $\mathbf{x}'\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|\cos\theta$, θ der Winkel zwischen den beiden Vektoren (vergl. (??), Seite ??).

Für \mathbf{a} gelten dann die Darstellungen

$$\mathbf{a} = a_x \mathbf{e}_x + a_y \mathbf{e}_y, \quad (X, Y) - \text{System} \quad (5.20)$$

$$= a_u \mathbf{e}_u + a_v \mathbf{e}_v, \quad (U, V) - \text{System} \quad (5.21)$$

Zu bestimmen sind a_u, a_v und \mathbf{e}_u und \mathbf{e}_v , d.h. die Komponenten u_1, u_2, v_1, v_2 . Dazu berechne man die Skalarprodukte⁷⁰

$$\mathbf{e}'_x \mathbf{e}_u = (1, 0) \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = u_1 = \cos\theta \quad (5.22)$$

$$\mathbf{e}'_y \mathbf{e}_u = (0, 1) \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = u_2 = \cos(90^\circ - \theta) = \sin\theta \quad (5.23)$$

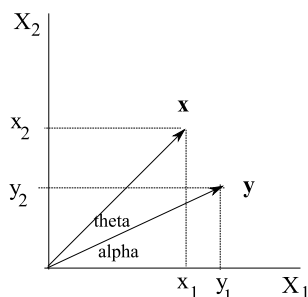
$$\mathbf{e}'_x \mathbf{e}_v = (1, 0) \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = v_1 = \cos(90^\circ + \theta) = -\sin\theta \quad (5.24)$$

$$\mathbf{e}'_y \mathbf{e}_v = (0, 1) \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = v_2 = \cos\theta \quad (5.25)$$

In (5.21) eingesetzt ergibt sich das Gleichungssystem in den zwei Unbekannten

⁷⁰Es wird von der bekannten Identität $\cos(\alpha \pm \beta) = \cos\alpha \cos\beta \mp \sin\alpha \sin\beta$ Gebrauch gemacht; 90° entspricht $\pi/2$ im Radians-System.

Abbildung 23: Rotation eines Vektors (Alibi-Transformation)



a_u und a_v

$$a_x = a_u \cos \theta - a_v \sin \theta \quad (5.26)$$

$$a_y = a_u \sin \theta + a_v \cos \theta \quad (5.27)$$

In Matrixform hat das System die Form

$$\mathbf{a} = \underbrace{\begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}}_{T'} \begin{pmatrix} a_u \\ a_v \end{pmatrix} = T' \begin{pmatrix} a_u \\ a_v \end{pmatrix} \quad (5.28)$$

T' ist orthonormal, wie man durch Nachrechnen bestätigt, so dass man die Lösung

$$T\mathbf{a} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} a_x \\ a_y \end{pmatrix} = \begin{pmatrix} a_u \\ a_v \end{pmatrix} \quad (5.29)$$

erhält.

Es soll nun $\mathbf{y} = T\mathbf{x}$ gelten $\mathbf{y} = (y_1, y_2)'$, $\mathbf{x} = (x_1, x_2)'$, und es gelte $\mathbf{x}'\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|\cos \theta$, d.h. die Vektoren \mathbf{x} und \mathbf{y} bilden den Winkel θ , und natürlich sei $\|\mathbf{x}\| = \|\mathbf{y}\|$. \mathbf{y} bilde mit der X -Achse den Winkel α . Es gelten die Beziehungen (vergl. Abb. 23)

$$x_1 = \frac{\cos(\alpha + \theta)}{r}, \quad x_2 = \frac{\sin(\alpha + \theta)}{r}, \quad y_1 = \frac{\cos \alpha}{r}, \quad y_2 = \frac{\sin \alpha}{r}$$

Wegen

$$\begin{aligned} \sin(\alpha \pm \theta) &= \sin \alpha \cos \theta \pm \cos \alpha \sin \theta \\ \cos(\alpha \pm \theta) &= \cos \alpha \cos \theta \mp \sin \alpha \sin \theta \end{aligned}$$

folgt

$$\begin{aligned} x_1 &= \frac{1}{r}(\cos \alpha \cos \theta - \sin \alpha \sin \theta) = y_1 \cos \theta - y_2 \sin \theta \\ x_2 &= \frac{1}{r}(\sin \alpha \cos \theta + \cos \alpha \sin \theta) = y_1 \sin \theta + y_2 \cos \theta \end{aligned}$$

d.h.

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \quad (5.30)$$

d.h.

$$\mathbf{x} = T' \mathbf{y} \quad (5.31)$$

und T wurde in (5.29) definiert.

5.4.3 Die Cholesky-Zerlegung

Die Anwendung des Gauß-Algorithmus impliziert, dass die Koeffizientenmatrix A des Gleichungssystems $A\mathbf{x} = \mathbf{y}$ in eine Dreiecksform überführt wird. Für den Spezialfall einer symmetrischen, positiv-definiten Matrix läßt sich zeigen, dass A stets als Produkt zweier Dreiecksmatrizen darstellbar ist; dieser Sachverhalt bedeutet, dass der Aufwand für das Lösen des Gleichungssystems drastisch reduziert wird; dies ist die Cholesky-Zerlegung⁷¹. In diesem Skript wird auf numerische Fragen kaum eingegangen, da die Programmpakete für multivariate Verfahren im Allgemeinen effiziente Algorithmen enthalten, um die sich der Anwender nicht weiter kümmern muss. In einigen theoretischen Herleitungen wird aber auf die Cholesky-Zerlegung eingegangen, so dass in diesem Absatz kurz auf sie eingegangen werden soll.

Satz 5.3 *Es sei A eine symmetrische, positiv-definite Matrix. Dann existiert eine untere Dreiecksmatrix L derart, dass*

$$A = LL' \quad (5.32)$$

Beweis: Statt eines allgemeinen Beweises wird der Satz anhand einer (3×3) -Matrix illustriert; das angewendete Prinzip überträgt sich sofort auf den allgemeinen $(n \times n)$ -Fall. Nach Behauptung existieren also L und L^* derart, dass

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} L_{11} & L_{21} & L_{31} \\ 0 & L_{22} & L_{32} \\ 0 & 0 & L_{33} \end{pmatrix}$$

Rechnet man das Produkt auf der rechten Seite aus, so erhält man

$$A = LL' = \begin{pmatrix} L_{11}^2 & L_{11}L_{21} & L_{11}L_{31} \\ L_{21}L_{11} & L_{21}^2 + L_{22}^2 & L_{21}L_{31} + L_{22}L_{32} \\ L_{31}L_{11} & L_{31}L_{21} + L_{32}L_{22} & L_{31}^2 + L_{32}^2 + L_{33}^2 \end{pmatrix}.$$

Das Element a_{ij} von A ist dann gleich dem Element in der i -ten Zeile und j -ten Spalte der Matrix auf der rechten Seite, also

$$a_{11} = L_{11}^2, \quad a_{22} = L_{21}^2 + L_{22}^2, \quad a_{33} = L_{31}^2 + L_{32}^2 + L_{33}^2,$$

⁷¹André-Louis Cholesky (1875 – 1918), französischer Mathematiker

und

$$a_{21} = L_{21}L_{11}, \quad a_{23} = L_{21}L_{31} + L_{22}L_{22}L_{32}, \quad a_{31} = L_{31}L_{11}$$

und wegen der vorausgesetzten Symmetrie von A gilt $a_{ij} = a_{ji}$, so dass nicht mehr Elemente bestimmt werden müssen. Wegen der vorausgesetzten Positiv-Definitheit gilt $a_{11} > 0$, also folgt $L_{11} = \sqrt{a_{11}}$. Dann folgt $L_{21} = a_{21}/L_{11} = a_{21}/\sqrt{a_{11}}$ und $L_{31} = a_{31}/\sqrt{a_{11}}$, etc. Auf diese Weise lassen sich die L_{ij} aus den a_{ij} berechnen. Das Prinzip kann leicht auf den Fall allgemeiner $(n \times n)$ -Matrizen übertragen werden. \square

Die LDL-Zerlegung: Eine etwas verallgemeinerte Form der Cholesky-Zerlegung ist die *LDL-Zerlegung* einer symmetrischen, positiv-definiten Matrix:

$$A = LDL', \quad (5.33)$$

wobei L eine untere Dreiecksmatrix und D eine Diagonalmatrix ist. Man kann $D = D^{1/2}D^{1/2}$ schreiben, wobei $D^{1/2}$ eine Diagonalmatrix ist, deren Diagonalelemente die Wurzeln aus den Diagonalelementen von D ist, d.h. die Diagonalelemente von D müssen größer als Null sein. Setzt man $G = LD^{1/2}$, so hat man

$$A = GG'. \quad (5.34)$$

Es sei $A\mathbf{x} = \mathbf{y}$ ein Gleichungssystem. Setzt man $G\mathbf{z} = \mathbf{y}$, $G'\mathbf{x} = \mathbf{z}$, so hat man $GG'\mathbf{x} = \mathbf{y}$. Man löst also das Gleichungssystem, indem man zunächst $G\mathbf{z} = \mathbf{y}$ und dann $G'\mathbf{x} = \mathbf{z}$ löst; da G und G' in Dreiecksform sind, sind diese beiden Gleichungssysteme schnell und einfach zu lösen. Man findet für die Elemente g_{ik} von G

$$g_{ik} = \begin{cases} 0, & i < k \\ \sqrt{a_{ik} - \sum_{j=1}^{k-1} g_{kj}^2}, & i = k \\ \frac{1}{a_{ik}} \left(a_{ik} - \sum_{j=1}^{k-1} g_{ij}g_{kj} \right), & i > k \end{cases} \quad (5.35)$$

Eine ausführliche Darstellung von Lösungen von linearen Gleichungssystemen durch Rückführung auf Dreiecksmatrizen wird in Golub & vanLoan (2013), Kap. 3 und 4, gegeben.

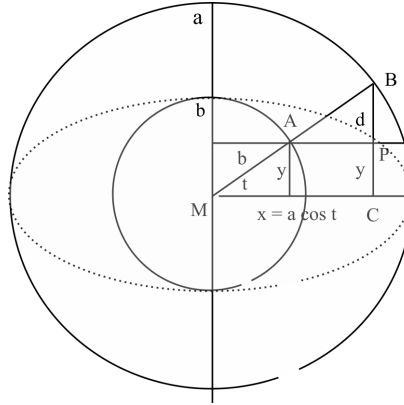
5.5 Zur Berechnung von Ellipsen für eine Punktekonfiguration

Gegeben seien zwei konzentrische Kreise mit den Radien a und $b < a$. Die Gerade \overline{MB} schneidet den kleineren Kreis im Punkt A . Für die Gerade \overline{BC} gilt $\overline{BC} = y + d$ mit $y = \overline{CP}$, und es ist $\overline{MC} = x$. t ist der Winkel, den die Gerade \overline{MA} bzw. $\overline{MB} = a$ mit der x -Achse bildet. Die Position des Punktes P hängt vom Wert des Parameters (Winkels) t ab, d.h. $P = P(t)$.

Behauptung: Die Punkte $P(t)$, $0 \leq t \leq 2\pi$, liegen auf einer Ellipse \mathcal{E} der Form

$$\mathcal{E}: \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad x = x(t), \quad y = y(t) \quad (5.36)$$

Abbildung 24: Kreise mit den Radien a bzw. $b < a$ und zugehörige Ellipse (Menge der Punkte P) mit den Halbachsen a und b



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 (5.36) folgt. \square

Parameterform: Bei dieser Darstellung wird die Ellipse (5.36) als Menge der Punkte $P = P(t)$ für $0 \leq t \leq 2\pi$ definiert. Aus der Abbildung 24 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}

$$\mathcal{E} = \{\mathbf{p}(t) | \mathbf{p}(t) = (x(t), y(t))' = (a \cos t, b \sin t)\}. \quad (5.37)$$

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

Die Gleichung (5.36) impliziert, dass die Länge der ersten Hauptachse durch $x_0 = a$ gegeben ist (man setzt $y = 0$), und die Länge der zweiten Hauptachse ist $y_0 = b$ (man setzt $x = 0$). Gesucht sind Werte für a und b derart, dass die erste Hauptachse der Ellipse der maximalen Ausdehnung der Punktekonfiguration entspricht; die k -te Spalte \mathbf{L}_k hat die Länge $\sqrt{\lambda_k}$. Diese Forderung führt zu

$$a = \sqrt{\lambda_1}, \quad b = \sqrt{\lambda_2}. \quad (5.38)$$

In Bezug auf (5.37) 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 (5.36) äquivalent zu

$$\mathbf{y}'\Lambda^{-1}\mathbf{y} = 1. \quad (5.39)$$

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

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

5.6 Die Differentiation von Vektoren

5.6.1 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 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

$$\frac{d\mathbf{x}}{dx_j} = \begin{pmatrix} dx_1/dx_j \\ \vdots \\ dx_j/dx_j \\ \vdots \\ dx_n/dx_j \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{e}_j. \quad (5.40)$$

Dann hat man

$$\frac{d\mathbf{x}}{d\mathbf{x}} = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n] = I, \quad (5.41)$$

I die (n, n) -Einheitsmatrix.

Der Fall (5.40) 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

$$\frac{d\mathbf{y}}{d\mathbf{x}} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n] = A. \quad (5.42)$$

Beispiel 5.1 Alternativer Beweis (I) für $TT' = I$: Für $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ gelte $T\mathbf{y} = \mathbf{x}$. Dann

$$\begin{aligned} T\mathbf{y} &= \mathbf{x} \\ T'T\mathbf{y} &= T'\mathbf{x} \\ \mathbf{y} &= T'\mathbf{x} \\ T\mathbf{y} &= TT'\mathbf{x} = \mathbf{x}, \quad \text{für alle } \mathbf{x} \end{aligned}$$

Dann folgt

$$\frac{dT'T'\mathbf{x}}{d\mathbf{x}} = TT' = \frac{d\mathbf{x}}{d\mathbf{x}} = I$$

nach (5.41) und (5.42). □

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.

5.6.2 Die Differentiation quadratischer Formen

Es wird die quadratische Form $Q(\mathbf{x}) = \mathbf{x}'C\mathbf{x}$ betrachtet, wobei $C' = C$ eine (n, n) -Matrix ist und $\mathbf{x} \in \mathbb{R}^n$ n -dimensionale Vektoren sind. Es soll bezüglich \mathbf{x} differenziert werden. Man differenziert zunächst nach einer Komponente x_j von \mathbf{x} und findet (Anwendung der Produktregel)

$$\frac{\partial Q}{\partial x_j} = \mathbf{e}_j' C \mathbf{x} + \mathbf{x}' C \mathbf{e}_j \tag{5.43}$$

\mathbf{e}_j der j -te Einheitsvektor. $\mathbf{e}_j' C \mathbf{x}$ und $\mathbf{x}' C \mathbf{e}_j$ sind Skalarprodukte und es folgt $\mathbf{e}_j' C \mathbf{x} = \mathbf{x}' C \mathbf{e}_j$, wegen $C' = C$, so dass

$$\frac{\partial Q}{\partial x_j} = 2\mathbf{e}_j' C \mathbf{x}. \tag{5.44}$$

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

$$\frac{\partial Q}{\partial \mathbf{x}} = 2IC\mathbf{x} = 2C\mathbf{x}. \quad (5.45)$$

Damit hat man auch die Ableitung von $\|\mathbf{x}\|^2 = \mathbf{x}'\mathbf{x}$ nach \mathbf{x} gefunden, denn mit $C = I$, I die Einheitsmatrix, folgt aus (5.45)

$$\frac{\partial \mathbf{x}'\mathbf{x}}{\partial \mathbf{x}} = 2I\mathbf{x} = 2\mathbf{x}. \quad (5.46)$$

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 5.6.4 nachgelesen werden.

Zu maximieren sei

$$Q(\mathbf{x}) = \mathbf{x}'A\mathbf{x} + \lambda(\mathbf{x}'\mathbf{x} - a), \quad \mathbf{x}'\mathbf{x} - a = 0, \quad A' = A \quad (5.47)$$

Dann ist

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

und $Q(\mathbf{x})$ nimmt ein Maximum für den Vektor \mathbf{x}_{\max} an, für den $dQ/d\mathbf{x}|_{\mathbf{x}_{\max}} = 0$, d.h. $2A\mathbf{x}_{\max} - 2\lambda\mathbf{x}_{\max} = \vec{0}$, so dass

$$A\mathbf{x}_{\max} = \lambda\mathbf{x}_{\max} = \vec{0}, \quad (5.48)$$

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

5.6.3 Die Kleinste-Quadrate-Schätzung für das Lineare Modell

Es sei X eine (m, 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

$$\mathbf{y} = X\mathbf{b} + \mathbf{e} \quad (5.49)$$

\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

$$\mathbf{e}'\mathbf{e} = (\mathbf{y} - X\mathbf{b})'(\mathbf{y} - X\mathbf{b}) = \mathbf{y}'\mathbf{y} - \mathbf{y}'X\mathbf{b} - \mathbf{b}'X'\mathbf{y} + \mathbf{b}'X'X\mathbf{b} \quad (5.50)$$

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 (5.42) ist $d(X\mathbf{b})/d\mathbf{b} = X$, und nach (5.45) 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

$$\left. \frac{d(\mathbf{e}'\mathbf{e})}{d\mathbf{b}} \right|_{\mathbf{b}=\hat{\mathbf{b}}} = 2(X'\mathbf{y} - (X'X)\mathbf{b}) = 0 \quad (5.51)$$

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

$$\hat{\mathbf{b}} = (X'X)^{-1}X'\mathbf{y}, \quad (5.52)$$

Einige Implikationen: Es sei $X = Q\Sigma V'$, $\Sigma = \Lambda^{1/2}$, die SVD von X . Dann ist $X'X = V\Lambda V'$ und

$$(X'X)^{-1} = (V\Lambda V')^{-1} = (V')^{-1}\Lambda^{-1}V^{-1} = V\Lambda^{-1}V', \quad (5.53)$$

denn $V^{-1} = V'$ wegen der Orthonormalität von V . (5.52) impliziert dann

$$\hat{\mathbf{b}} = V\Lambda^{-1}V'V\Lambda^{1/2}Q'\mathbf{y} = V\Lambda^{-1/2}Q'\mathbf{y}. \quad (5.54)$$

Aber es ist $\mathbf{y} = X\mathbf{b} + \mathbf{e} = Q\Lambda^{1/2}Q'\mathbf{b} + \mathbf{e}$, so dass aus (5.54)

$$\hat{\mathbf{b}} = V\Lambda^{-1/2}Q'(Q\Lambda^{1/2}Q'\mathbf{b} + \mathbf{e}) = \mathbf{b} + V\Lambda^{-1/2}Q'\mathbf{e} \quad (5.55)$$

folgt, und schreibt man $V\Lambda^{-1/2}Q'$, indem man das dyadische Produkt anwendet, so ergibt sich (vergl. (3.157), Seite 137, wo der Fall $X = Q\Lambda^{1/2}V'$ betrachtet wird)

$$\hat{\mathbf{b}} = \mathbf{b} + \left(\sum_{k=1}^n \frac{\mathbf{v}_k \mathbf{q}'_k}{\sqrt{\lambda_k}} \right) \mathbf{e} \quad (5.56)$$

Der wahre Parametervektor \mathbf{b} und die Kleinste-Quadrate-Schätzung $\hat{\mathbf{b}}$ unterscheiden sich also um den Vektor $V\Lambda^{-1/2}Q'\mathbf{e} = \left(\sum_{k=1}^n \mathbf{v}_k \mathbf{q}'_k / \sqrt{\lambda_k} \right) \mathbf{e}$. Der ist um so größer, je größer einerseits die Komponenten von $\mathbf{e}V\Lambda V'$ folgt, dass die Spaltenvektoren von C als Linearkombinationen der Spalten von $V\Lambda$ dargestellt werden können: Es sei \mathbf{c}_j die j -te Spalte von C , und v_{ik} sei das i -te Element des k -ten Eigenvektors in V . Dann ist der j -te Spaltenvektor von V' durch $(v_{1j}, \dots, v_{kj}, \dots, v_{nj})'$ gegeben und man hat

$$\mathbf{c}_j = v_{1j}\lambda_1\mathbf{v}_1 + v_{2j}\lambda_2\mathbf{v}_2 + \dots + v_{nj}\lambda_n\mathbf{v}_n. \quad (5.57)$$

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 $Q\Sigma V'$ einzusetzen:

$$\mathbf{y} = Q\Sigma V'\mathbf{b} + \mathbf{e} = Q\Sigma(V'\mathbf{b}) + \mathbf{e} = L\boldsymbol{\beta} + \mathbf{e}, \quad L = Q\Sigma \quad (5.58)$$

mit L als neuer Prädiktormatrix und $\boldsymbol{\beta} = V'\mathbf{b}$ als neuem Parametervektor; die Spaltenvektoren von L sind orthogonal und also unkorreliert. Im Skriptum über Regressionsverfahren wird dieser Ansatz ausführlicher diskutiert.

5.6.4 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 soll, 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 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

$$f_x + f_y g' = 0 \quad (5.59)$$

$$g_x + g_y g' = 0 \quad (5.60)$$

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

$$\begin{pmatrix} f_x \\ f_y \end{pmatrix} = \lambda \begin{pmatrix} g_x \\ g_y \end{pmatrix} \quad (5.61)$$

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

$$f_x - \lambda g_x = 0 \quad (5.62)$$

$$f_y - \lambda g_y = 0 \quad (5.63)$$

$$g(x, y) = 0 \quad (5.64)$$

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

$$L(x, y, \lambda) = f(x, y) + \lambda g(x, y), \quad g(x, y) = 0. \quad (5.65)$$

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 (5.62), (5.63) und (5.64) heißen zusammen die *Lagrange-sche Multiplikatorenregel*, nach dem Mathematiker und Astronomen Jean-Louis Lagrange (1736 – 1813), der diese Regel 1788 herleitete.

Beispiel 5.2 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. □

Beispiel 5.3 (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

$$\max_{\mathbf{x} \neq \vec{0}} \frac{\mathbf{x}' A \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \max_j \lambda_j = \lambda_1, \quad (5.66)$$

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

$$\min_{\mathbf{x} \neq \vec{0}} \frac{\mathbf{x}' A \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \min_j \lambda_j = \lambda_n \quad (5.67)$$

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 demnach maximal, wenn $\mathbf{x} = \mathbf{u}$ der erste Eigenvektor von A ist. \square

5.7 Der Rayleigh-Koeffizient und der Satz von Courant-Fischer

Satz 5.4 (Satz von Courant-Fischer) *Es sei M eine symmetrische, positiv definite Matrix mit $M = T\Lambda T'$, 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

$$\max_{\mathbf{x} \neq \vec{0}} R(\mathbf{x}) = \max_{\mathbf{x} \neq \vec{0}} \frac{\mathbf{x}' M \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \max_j \lambda_j = \lambda_1, \quad (5.68)$$

und der Vektor \mathbf{x} , für den das Maximum angenommen wird, ist der zu λ_1 korrespondierende Eigenvektor \mathbf{t}_1 , so dass $\max_{\mathbf{x} \neq \vec{0}} R(\mathbf{x}) = \mathbf{t}_1' M \mathbf{t}_1 = \lambda_1$. Weiter gilt

$$\max_{\mathbf{x} \perp \mathbf{t}_1, \dots, \mathbf{t}_k} \frac{\mathbf{x}' M \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \lambda_{k+1}, \quad k < n, \quad (5.69)$$

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

$$\min_{\mathbf{x} \neq \vec{0}} R(\mathbf{x}) = \frac{\mathbf{x}' M \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \min_j \lambda_j, \quad (5.70)$$

mit dem zugehörigen Eigenvektor \mathbf{t}_{\min} , d.h. $\min_{\mathbf{x} \neq \vec{0}} R(\mathbf{x}) = \mathbf{t}_{\min}' M \mathbf{t}_{\min} = \lambda_{\min}$.

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 , $j = 1, \dots, n$ bilden eine orthonormale Basis des $V_n = \mathcal{L}(\mathbf{t}_1, \dots, \mathbf{t}_n)$. Es sei nun \mathbf{x} ein beliebiger n -dimensionaler Vektor. Er kann stets als Linearkombination einer Basis dargestellt werden, insbesondere können die Spaltenvektoren \mathbf{t}_j von T gewählt werden⁷². Es gilt also insbesondere

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

für für geeignet gewählte Koeffizienten c_1, \dots, c_n . Dann ist

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

Alle Kreuzprodukte $c_j c_k \mathbf{t}_j' \mathbf{t}_k$ verschwinden, weil $\mathbf{t}_j' \mathbf{t}_k = 0$ für alle $j \neq k$, so dass

$$\frac{\mathbf{x}'M\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \frac{\sum_{j=1}^n c_j^2 \lambda_j}{\sum_{j=1}^n c_j^2}$$

und $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. Ersetzt man also auf der rechten Seite alle $\lambda_j \neq \lambda_1$ durch λ_1 , so folgt

$$\frac{\mathbf{x}'M\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \frac{\sum_{j=1}^n c_j^2 \lambda_j}{\sum_{j=1}^n c_j^2} \leq \frac{\lambda_1 \sum_{j=1}^n c_j^2}{\sum_{j=1}^n c_j^2} = \lambda_1,$$

so dass $\max_{\mathbf{x}} R(\mathbf{x}) = \lambda_1$, λ_1 der größte Eigenwert. Jetzt muss noch gezeigt werden, für welchen Vektor \mathbf{x} das Maximum angenommen wird. Es sei \mathbf{t}_1 der zu $\lambda_{\max} = \lambda_1$ korrespondierende Eigenvektor. Für $\mathbf{x} = \mathbf{t}_1$ folgt $\mathbf{t}_1' M \mathbf{t}_1 = \lambda_1$ und wegen $\mathbf{t}_1' \mathbf{t}_1 = 1$ sieht man, dass $R(\mathbf{x}) = \max$ für $\mathbf{x} = \mathbf{t}_1$.

Die Aussage (5.70) ergibt sich aus (5.69). Es sei wieder $T = [\mathbf{t}_1, \dots, \mathbf{t}_n]$ die Matrix der Eigenvektoren von M . 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 folgt aber

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

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 (5.68)

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

und analog zur Argumentation im Beweis zu Satz 5.4, Seite 178, folgt (5.69). \square

Anmerkungen:

⁷²s. die Anmerkung 1. auf Seite 179

1. Der Beweis geht davon aus, dass die Eigenvektoren $\mathbf{t}_1, \dots, \mathbf{t}_n$ als Basis des V_n gewählt werden können. Diese Wahl ist keine Einschränkung der Allgemeinheit, da man von einer beliebigen Basis V stets zu einer orthonormalen Basis übergehen kann, vergl. die Singularwertzerlegung (Abschnitt 3.2.7), angewendet auf V . Deswegen bedeutet dieser Übergang auch nur, einen Umweg zu gehen: man kann gleich mit den Eigenvektoren als Basis beginnen.
2. Für $k = 0$ betrachtet man $\max_{\mathbf{x} \neq \vec{0}} \frac{\mathbf{x}'M\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_1$, und für $k = n - 1$ erhält man

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

d.h.

$$\min_{\mathbf{x}} R(\mathbf{x}) = \lambda_n \quad (5.71)$$

so dass

$$\lambda_n \leq R(\mathbf{x}) \leq \lambda_1. \quad (5.72)$$

3. Es sei M eine beliebige symmetrische (m, n) -Matrix und $\mathbf{x} \in \mathbb{R}^n$. Weiter sei $\lambda = \mathbf{x}'M\mathbf{x}/\mathbf{x}'\mathbf{x}$. Die Werte λ heißen *maximal*, wenn sie den Rayleigh-Quotienten $\mathbf{x}'M\mathbf{x}/\mathbf{x}'\mathbf{x}$ maximieren.
4. Es sei T die Matrix der Eigenvektoren von $M = X'X$; für irgendzwei Eigenvektoren \mathbf{t}_j und \mathbf{t}_k , $j \neq k$, gilt also $\mathbf{t}_j'\mathbf{t}_k = 0$. Es sei $L = XT$, so dass $\mathbf{L}_j = X\mathbf{t}_j$, $j = 1, \dots, n$. Dann gilt

$$\mathbf{L}_j'\mathbf{L}_k = \mathbf{t}_j'X'X\mathbf{t}_k = \mathbf{t}_j'M\mathbf{t}_k = 0, \quad (5.73)$$

d.h. die Matrix L ist orthogonal. Denn $M = T\Lambda T'$, Λ die Diagonalmatrix der Eigenwerte von M . Dann hat man

$$\mathbf{t}_j'M\mathbf{t}_k = \mathbf{t}_j'T\Lambda T'\mathbf{t}_k = \mathbf{e}_j'\Lambda\mathbf{e}_k = \lambda_j\mathbf{e}_j'\mathbf{e}_k = 0, \quad j \neq k.$$

□

5.8 Alternativer Beweis von Satz 3.20

Es sei \mathbf{a}_{zi} der i -te Zeilenvektor von A , $i = 1, \dots, m$. Es sei $A\mathbf{x} = \vec{0}$; dies bedeutet, dass die Skalarprodukte $\mathbf{a}'_{ui}\mathbf{x} = 0$ für alle i , d.h. \mathbf{x} ist orthogonal zu allen Zeilenvektoren von A . Für die \mathbf{x} mit $A\mathbf{x} = \mathbf{y} \neq \vec{0}$ gilt diese Aussage nicht, d.h. \mathbb{R}^n wird in zwei Teilmengen U und V zerlegt:

$$U = \{\mathbf{x} \in \mathbb{R}^n | A\mathbf{x} = \vec{0}\} = \text{kern}(A), \quad V = \{\mathbf{x} \in \mathbb{R}^n | A\mathbf{x} = \mathbf{y} \neq \vec{0}\}.$$

$U = \text{kern}(A)$ ist ein Teilraum des \mathbb{R}^n . V ist ebenfalls ein Teilraum des \mathbb{R}^n , denn es sei $A\mathbf{x}_1 = \mathbf{y}_1$, $A\mathbf{x}_2 = \mathbf{y}_2$, $\mathbf{x}_1, \mathbf{x}_2 \notin \text{kern}(A)$. Dann ist, für $\lambda, \mu \in \mathbb{R}$ beliebig, $A\lambda\mathbf{x}_1 = \lambda\mathbf{y}_1$, $A\mu\mathbf{x}_2 = \mu\mathbf{y}_2$ und $A(\lambda\mathbf{x}_1 + \mu\mathbf{x}_2) = \lambda\mathbf{y}_1 + \mu\mathbf{y}_2 = \mathbf{y} \in \mathcal{L}(A)$, mithin

ist $\mathbf{x} = \lambda \mathbf{x}_1 + \mu \mathbf{x}_2 \in V$. Offenbar ist $U \cap V = \emptyset$, denn ein Vektor $\mathbf{x} \in \mathbb{R}^n$ kann nicht zugleich in U und in V sein. Also folgt $U + V = \mathbb{R}^n$ und es folgt

$$\dim(U + V) = \dim(U) + \dim(V) = \dim \mathbb{R}^n = n. \quad (5.74)$$

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

$$\mathbf{a}_{r+k} = \lambda_{k1} \mathbf{a}_1 + \cdots + \lambda_{kr} \mathbf{a}_r = \sum_{j=1}^r \lambda_{kj} \mathbf{a}_j \quad (5.75)$$

gelten muss, so hat man

$$\sum_{j=1}^r x_j \mathbf{a}_j = - \sum_{k=1}^{n-r} x_{r+k} \sum_{j=1}^r \lambda_{kj} \mathbf{a}_j. \quad (5.76)$$

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 (5.76) 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

$$\mathbf{x}_k = (-\lambda_{k1}, \dots, -\lambda_{kr}, 0, \dots, 0, 1, 0, \dots, 0)', \quad k = 1, \dots, n-r \quad (5.77)$$

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 (5.75).

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 $\text{rg}(A) = r$ angebrochen wurde, der Rang von $\text{kern}(A)$ kann also nicht größer als $n-r$ sein. Wegen (5.74) (= 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

5.9 Vektortransformationen 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, 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:

$$f: \mathcal{M} \rightarrow \mathcal{N}, \quad x \mapsto y = f(x), \quad x \in \mathcal{M}, y \in \mathcal{N} \quad (5.78)$$

Man schreibt gelegentlich auch

$$f(\mathcal{M}) = \mathcal{N}. \quad (5.79)$$

$f(\mathcal{M})$ heißt das *Bild* von \mathcal{M} in \mathcal{N} , und \mathcal{M} ist das *Urbild* von $f(\mathcal{M})$. Man schreibt auch $\text{Im}f = \mathcal{N}$ ⁷³.

Eine spezielle Abbildung ist die *Identität* oder identische Abbildung

$$\text{id}(\mathcal{M}) = \mathcal{M}. \quad (5.80)$$

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

$$I_n \mathbf{x} = \mathbf{x}, \quad \mathbf{x} \in \mathbb{R}^n. \quad (5.81)$$

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

$$f(\mathcal{M}) = \mathcal{N}, \quad f^{-1}f(\mathcal{M}) = \mathcal{M} = f^{-1}(\mathcal{N}). \quad (5.82)$$

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

$$f(\mathcal{M}) = \{f(x), x \in \mathcal{M}\}. \quad (5.83)$$

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

Definition 5.4 *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 $\text{kern}(f) = \{x \in \mathcal{M} | f(x) = \mathbf{0}\}$ 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

⁷³Im wohl als Abkürzung des lat. *imago* für 'Bild'.

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. \mathbb{R} bezeichnet die Menge der reellen Zahlen. f ist sicher injektiv, denn $f(x) = f(x')$ impliziert $ax + b = ax' + b$ und damit $x = x'$, wie man leicht nachrechnet. f ist auch surjektiv, denn für $y = ax + b$ existiert genau ein $x = (y - b)/a$ derart, dass $y = f(x)$. Da f sowohl injektiv wie surjektiv ist, ist f auch bijektiv.

Mit $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ wird die Menge der Paare $(x, y), x, y \in \mathbb{R}$, bezeichnet, allgemein mit

$$\mathbb{R}^m = \underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}}_{m\text{-mal}}$$

Die Menge der m -tupel $(x_1, x_2, \dots, x_m), x_j \in \mathbb{R}$, 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, n) -Matrizen bezeichnet. Alle diese Definitionen übertragen sich auf \mathbb{C} , die Menge der komplexen Zahlen $x + iy, x, y \in \mathbb{R}$ und $i = \sqrt{-1}$.

Die Schreibweise $M \in \mathbb{R}^{m,n}$ bedeutet, dass M eine (m, n) -Matrix ist. Die Schreibweise $f: V_m \rightarrow V_n$ 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 5.5 *Es seien V und W Vektorräume und $f: V \rightarrow W$ sei eine Abbildung von V in W .*

1. f heißt linear bzw. homomorph, wenn

$$f(\lambda\mathbf{v} + \mu\mathbf{w}) = \lambda f(\mathbf{v}) + \mu f(\mathbf{w}) \tag{5.84}$$

für $\lambda, \mu \in \mathbb{R}$ und für alle $\mathbf{v} \in V$ und $\mathbf{w} \in W$.

2. f heißt isomorph, wenn f bijektiv ist; man sagt auch, f definiere einen Homomorphismus bzw. Isomorphismus wenn f bijektiv ist.

3. f definiert einen Endomorphismus, wenn $V = W$, und

4. einen Automorphismus, wenn f bijektiv ist und außerdem $V = W$ gilt.

Es sei $M \in \mathbb{R}^{m,n}$; M definiert eine lineare, also homomorphe Abbildung, denn $M\mathbf{x} = \mathbf{y}$ erfüllt die Bedingungen einer linearen Abbildung. für $m \neq n$ ist f offenbar weder ein Endomorphismus noch ein Automorphismus.

Dem Begriff des Kerns in der allgemeinen Definition 5.4 von Abbildungen entspricht für $f \in \mathbb{R}^{m,n}$ der Nullvektor $\vec{0}$.

Satz 5.5 *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 $\ker f = \{\vec{0}\}$ bedeutet, dass $\ker f$ 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 $\mathbf{v} - \mathbf{u} \in \ker f$.

Um 3. einzusehen sei angenommen, dass

$$\lambda_1 f(\mathbf{x}_1) + \dots + \lambda_n f(\mathbf{x}_n) = \vec{0}$$

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

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

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

Definition 5.6 Es sei f eine Abbildung $f: \mathbb{R}^m \rightarrow \mathbb{R}^n$; dann heißt die Dimensionalität des Bildes $\operatorname{Im} f$ der Rang; man schreibt $\operatorname{rg}(f) = \dim \in f$.

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 $A\mathbf{e}_1, A\mathbf{e}_2, \dots, A\mathbf{e}_m$ geschrieben werden. Dann ist das Bild der durch A definierten Abbildung die lineare Hülle

$$\operatorname{Im} A = \mathcal{L}(A\mathbf{e}_1, A\mathbf{e}_2, \dots, A\mathbf{e}_m).$$

Demnach wird $\operatorname{Im} A$ auch der Spaltenraum von A bezeichnet. Der Begriff des Ranges einer Matrix A wird in Abschnitt 2.3 noch ausführlich diskutiert.

Beispiel 5.4 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

$$\text{kern}(f) = \{(x_1, x_2, x_3)' \mid x_1 = -2x_2\} = \{(-2x_2, x_2, x_3)'\}.$$

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 $\mathbf{u} \in V_n$ ein Vektor $\mathbf{x} \in V_m$ existiert derart, dass $M\mathbf{x} = \mathbf{u}$, d.h. die Frage nach der Surjektivität ist die Frage, ob durch M alle Elemente von V_n bestimmt werden. M definiert eine bijektive Abbildung, wenn M eine sowohl injektive wie auch surjektive Abbildung definiert. Dies ist die Frage, ob eine surjektive Abbildung auch eindeutig ist. Offenbar hängen diese Eigenschaften von der Struktur der Matrix M ab. Was mit dem Begriff der Struktur einer Matrix genau gemeint ist, wird im Folgenden entwickelt.

Beispiel 5.5 Es sei

$$T = \begin{pmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{pmatrix} = \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix}. \quad (5.85)$$

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

5.10 Determinanten

Der Begriff der Determinante ist bereits auf Seite 57 im Zusammenhang mit einem linearen Gleichungssystem aufgetaucht, mit dem die Koeffizienten von Linearkombinationen bestimmt wurden: die Ausdrücke für die Koeffizienten sind Quotienten, bei denen Zähler und Nenner Zahlen sind, die sich als Funktionen ganzer Matrizen ergeben. Der Ausdruck 'Determinante' bringt zum Ausdruck, dass es vom Wert bestimmter Determinanten abhängt, ob das Gleichungssystem eine Lösung hat oder nicht, sie "determiniert" in diesem Sinne die Existenz einer Lösung. Tatsächlich liegen die Anfänge der Determinantentheorie wohl bei Gerolamo Cardano (1501 – 1576), der den Spezialfall einer (2, 2)-Matrix behandelte; allgemeiner wurden sie dann von Gottfried Wilhelm Leibniz (1646 – 1716) im Zusammenhang mit Gleichungssystemen mit $n > 2$ Unbekannten untersucht. Es war allerdings der Genfer Mathematiker Gabriel Cramer (1704 – 1752), der die allgemeine Verwendung von Determinanten zur Lösung linearer Gleichungssysteme formulierte (Cramersche Regel), und die eigentliche Entwicklung der Determinantentheorie fand erst im 19-ten Jahrhundert statt. Die heutige axiomatische Definition des Begriffs der Determinante wurde von Karl Theodor Wilhelm Weierstraß (1815 – 1897) eingeführt.

Definition 5.7 *Es sei $A = [\mathbf{a}_1, \dots, \mathbf{a}_n]$ eine quadratische Matrix, d.g. die Spaltenvektoren \mathbf{a}_j von A seien n -dimensional. Die Determinante $|A|$ bzw. $\det(A)$ von A ist eine Zahl (im Folgenden werden nur Matrizen mit reellen Elementen betrachtet, so dass die Determinante eine reelle Zahl ist), wobei $|A|$ die folgenden Eigenschaften hat:*

1. $|A|$ ist multilinear, d.h. es gilt mit $j \neq k$
 - (i) $\det(\mathbf{a}_1, \dots, \mathbf{u} + \mathbf{v}, \dots, \mathbf{a}_n) = \det(\mathbf{a}_1, \dots, \mathbf{u}, \dots, \mathbf{a}_n) + \det(\mathbf{a}_1, \dots, \mathbf{v}, \dots, \mathbf{a}_n)$,
 - (ii) $\det(\mathbf{a}_1, \dots, \lambda \mathbf{a}_j, \dots, \mathbf{a}_n) = \lambda \det(\mathbf{a}_1, \dots, \mathbf{a}_j, \dots, \mathbf{a}_n)$
2. $|A|$ ist alternierend, d.h. gilt $\mathbf{a}_j = \mathbf{a}_k$ für irgendzwei Spaltenvektoren, so gilt

$$|A| = 0.$$

$$3. \det(\mathbf{e}_1, \dots, \mathbf{e}_n) = 1.$$

Dann heißt $\det(\mathbf{a}_1, \dots, \mathbf{a}_n)$ die Determinante der Matrix $A = [\mathbf{a}_1, \dots, \mathbf{a}_n]$.

Betrachtet man statt der Spaltenvektoren die Zeilenvektoren einer Matrix A , so ändert dies nicht den Wert der Determinante von A .

Permutationen: Eine Permutation der Zahlen $1, 2, \dots, n$ ist eine mögliche Anordnung dieser Zahlen: für die Zahlen $\{1, 2, 3\}$ ist $(3, 1, 2)$ eine mögliche Permutation. Für n Zahlen $1, 2, \dots, n$ enthält die Menge \mathcal{P} der möglichen Permutationen $n! = 1 \cdot 2 \cdot \dots \cdot n$ Elemente. Es sei $\tau = (i_1, i_2, \dots, i_n)$ ein Element von \mathcal{P} , wobei die $i_k, k = 1, \dots, n$ Elemente aus $\{1, 2, \dots, n\}$ sind. $\tau(i)$ und $\tau(k)$ definieren eine *Transposition*, wenn $\tau(i) = k$ und $\tau(k) = i$. Jedem $\tau \in \mathcal{P}$ kann ein Vorzeichen zugeordnet werden: unterscheiden sich $\tau(i)$ und $\tau(j)$ durch nur eine Transposition, so ist $\tau(j) = -1\tau(i)$. Benötigt man m Transpositionen, um $\tau(i)$ in die Folge $\{1, 2, \dots, n\}$ zu transponieren, so ist das Vorzeichen von $\tau(i)$ durch $\text{sgn}(\tau) = (-1)^m$ gegeben. Für jede Permutation τ existiert eine Inverse τ^{-1} derart, dass $\tau^{-1}\tau = \{1, 2, \dots, n\}$; τ^{-1} ist die auf τ angewandte Permutation, die die Permutation τ wieder in die "natürliche" Anordnung $1, 2, \dots, n$ überführt. Die Determinante $|A|$ einer (n, n) -Matrix $A = (a_{ij})$ ist dann durch

$$|A| = \sum_{\tau \in \mathcal{P}} \text{sgn}(\tau) a_{\tau(1),1} a_{\tau(2),2} \cdots a_{\tau(n),n} \quad (5.86)$$

gegeben (s. a. Leibnizregel und Laplacescher Entwicklungssatz weiter unten).

Folgerungen:

1. Die Matrix B gehe aus der Matrix A durch Vertauschung zweier Spalten hervor. Dann gilt $|B| = -|A|$.

Die Aussage ergibt sich aus der sukzessiven Anwendung der definierenden Eigenschaften der Determinante, insbesondere

- a. Addition des Vektors an j -ter Stelle zum Vektor an der k -ten Stelle,
- b. Subtraktion des neuen Vektors an der k -ten Stelle vom Vektor an der j -ten Stelle,
- c. Addition des Vektors an der j -ten Stelle zum Vektor an der k -ten Stelle,
- d. Ersetzung des Vektors an der j -ten Stelle durch sein Negatives.

2. Es sei $A = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n]$. Dann gilt

$$|[\lambda_1 \mathbf{a}_1, \lambda_2 \mathbf{a}_2, \dots, \lambda_n \mathbf{a}_n]| = \lambda_1 \lambda_2 \cdots \lambda_n |A|. \quad (5.87)$$

Für $\lambda_1, \dots, \lambda_n = \lambda$ ist dann

$$|[\lambda \mathbf{a}_1, \lambda \mathbf{a}_2, \dots, \lambda \mathbf{a}_n]| = |\lambda A| = \lambda^n |A| \quad (5.88)$$

Die Aussagen folgen sofort aus Definition 5.7, 1. (ii).

3. Die Matrix B entstehe aus A , indem der Vektor \mathbf{a}_j durch $\mathbf{a}_j + \lambda \mathbf{a}_k$ ersetzt wird, wobei \mathbf{a}_k ebenfalls ein Vektor von A ist und $\lambda \in \mathbb{R}$. Dann gilt $|A| = |B|$.

Denn

$$\begin{aligned} \det(\mathbf{a}_1, \dots, \mathbf{a}_j + \lambda \mathbf{a}_k, \dots, \mathbf{a}_k, \dots, \mathbf{a}_n) &= |A| + \lambda \det(\mathbf{a}_1, \dots, \mathbf{a}_k, \dots, \mathbf{a}_k, \dots, \mathbf{a}_n) \\ &= |A|, \end{aligned}$$

denn $\det(\mathbf{a}_1, \dots, \mathbf{a}_k, \dots, \mathbf{a}_k, \dots, \mathbf{a}_n) = 0$ wegen Folgerung 1

4. $|A| = [\mathbf{a}_1, \dots, \mathbf{a}_n]$ verändert ihren Wert nicht, wenn einer der Vektoren \mathbf{a}_j durch eine Linearkombination der übrigen ersetzt wird.

Diese Folgerung ist eine offenkundige Verallgemeinerung von Folgerung 3.

5. Ist einer der Vektoren von A der Nullvektor, so ist $|A| = 0$.

Die Matrix B ergebe sich aus der Matrix A , indem der Vektor \mathbf{a}_j durch $\lambda \mathbf{a}_j$ ersetzt wird; Für $\lambda = 0$ ist $\lambda \mathbf{a}_j = \vec{0}$, und gleichzeitig gilt dann $|B| = \lambda |A| = 0 \cdot |A| = 0$.

6. Es sei $A = \text{diag}(d_1, d_2, \dots, d_n)$ eine Diagonalmatrix, d.h. es sei $a_{ij} = 0$ für alle $i \neq j$ und $a_{ii} = d_i$, $i = 1, \dots, n$. Dann gilt

$$|A| = \prod_{i=1}^n d_i. \quad (5.89)$$

Denn nach (5.86) verschwinden wegen $a_{ij} = 0$ für $i \neq j$ alle Produkte, in denen Elemente a_{ij} mit $i \neq j$ auftreten, so dass nur der erste Term $a_{11}a_{22} \cdots a_{nn} = d_1 d_2 \cdots d_n$ übrig bleibt. $|A| = 0$, wenn mindestens einer der d_i -Werte gleich Null ist; in diesem Fall hat A einen Rang kleiner als n (s. die folgenden Folgerungen 7 und 8).

7. Es sei $A = [\mathbf{a}_1, \dots, \mathbf{a}_n]$ und die $\mathbf{a}_1, \dots, \mathbf{a}_n$ seien linear abhängig. Dann gilt $|A| = 0$.

Es sei \mathbf{a}_j eine Linearkombination der übrigen \mathbf{a}_k , d.h. es sei

$$\mathbf{a}_j = \sum_{k=1, k \neq j}^n a_k \mathbf{a}_k.$$

Dann kann \mathbf{a}_j durch

$$\mathbf{a}_j - \sum_{k=1, k \neq j}^n a_k \mathbf{a}_k = \vec{0}$$

ersetzt werden, wodurch sich der Wert der Determinante nicht ändert. Wenn aber einer der Vektoren der Nullvektor ist, so ist nach Folgerung 5 die Determinante gleich Null.

Anmerkung: Es sei $B = \lambda A$, $\lambda \in \mathbb{R}$. Findet man numerisch, dass $|B| \approx 0$, so kann man nicht folgern, dass die Spalten- oder Zeilenvektoren von B linear abhängig sind. Denn es sei $0 < |\lambda| < 1$, Dann wird λ^n "klein" und nach (?? muss $|B|$ dann ebenfalls "klein" sein, obwohl $|A|$ nicht "klein" ist.

8. Aus Folgerung 7 folgt unmittelbar, dass $|A| \neq 0$ impliziert, dass die \mathbf{a}_j nicht linear abhängig sind, d.h. dass sie linear unabhängig sind. Dies ist gleichbedeutend damit, dass $|A| \neq 0$ genau dann, wenn A vollen Rang hat. Ist A insbesondere eine Diagonalmatrix, so folgt, dass A genau dann den vollen Rang n hat, wenn alle $d_i \neq 0$.

9. Sind zwei der Vektoren von A identisch, so ist $|A| = 0$.

Die Vektoren von A sind in diesem Fall linear abhängig, und nach Folgerung 7 ist dann $|A| = 0$.

10. Es sei $C = [\mathbf{c}_1, \dots, \mathbf{c}_n]$ und die \mathbf{c}_j seien linear abhängig, so dass nach Folgerung 7 $|C| = 0$ folgt. Es sei $\mathcal{L}_C = \mathcal{L}(\mathbf{c}_1, \dots, \mathbf{c}_n)$ die lineare Hülle der \mathbf{c}_j . Die Vektoren $\mathbf{a}_1, \dots, \mathbf{a}_n$ seien Elemente aus \mathcal{L}_C und es sei $A = [\mathbf{a}_1, \dots, \mathbf{a}_n]$. Dann gilt $|A| = 0$.

Die \mathbf{a}_j sind Linearkombinationen der \mathbf{c}_j , und da die \mathbf{c}_j linear abhängig sind, sind auch die \mathbf{a}_j linear abhängig, so dass nach Folgerung 7 $|A| = 0$ sein muss.

11. Es sei A eine (n, n) -Matrix und A' sei die Transponierte von A . Dann gilt

$$|A'| = |A|. \quad (5.90)$$

Man hat die **Leibniz-Formel**

$$\begin{aligned} |A'| &= \sum_{\tau \in \mathcal{P}} \operatorname{sgn}(\tau) a_{1,\tau(1)} a_{2,\tau(2)} \cdots a_{n,\tau(n)} \\ &= \sum_{\tau \in \mathcal{P}} \operatorname{sgn}(\tau) a_{\tau^{-1}(1),1} \cdots a_{\tau^{-1}(n),n} \end{aligned}$$

Es sei $\sigma \in \mathcal{P}$ mit $\sigma = \tau^{-1}$; dann gilt $\operatorname{sgn}(\tau) = \operatorname{sgn}(\sigma)$. Da über alle Elemente von \mathcal{P} summiert wird, folgt

$$|A'| = \sum_{\sigma \in \mathcal{P}} \operatorname{sgn}(\sigma) a_{\sigma(1),1} \cdots a_{\sigma(n),n} = |A|.$$

12. **Produktsatz** Es seien A und B zwei (n, n) -Matrizen. Dann gilt

$$|AB| = |A||B|. \quad (5.91)$$

Es sei $C = AB$; der j -te Spaltenvektor \mathbf{c}_j von C ist dann durch $\mathbf{c}_j = A\mathbf{b}_j$ gegeben, \mathbf{b}_j der j -te Spaltenvektor von B , $j = 1, \dots, n$. Dann ist

$$|C| = \left| \left[\sum_{j=1}^n b_{j1} \mathbf{a}_j, \sum_{j=1}^n b_{j2} \mathbf{a}_j, \dots, \sum_{j=1}^n b_{jn} \mathbf{a}_j \right] \right|.$$

Die Aussage (5.91) folgt dann aus der wiederholten Anwendung der Eigenschaft 1. (i) der Definition 5.7 und der Folgerung 5.87.

Beispiel 5.6 Alternativer Beweis (II) von $TT' = I$, Satz 3.1, Seite 80: Nach (5.91) gilt $|AB| = |A||B|$, und nach (5.106) hat man $|T'T| = |I| = |T'||T| = |T||T'| = |TT'| = 1$, also $T'T = TT' = I$. \square

13. Es sei A eine (n, n) -Matrix mit vollem Rang. Dann gilt

$$|A^{-1}| = \frac{1}{|A|}. \quad (5.92)$$

Da A vollen Rang hat existiert die Inverse A^{-1} , so dass $A^{-1}A = I$, I die Einheitsmatrix. (5.91) impliziert dann $|A^{-1}A| = |A^{-1}||A| = |I| = 1$ (s. Definition 5.7, 3.). (5.92) folgt unmittelbar.

14. **Cramersche Regel**⁷⁴ Es sei A eine (n, n) -Matrix, und \mathbf{b} und \mathbf{x} seien n -dimensionale Vektoren. Sind A und \mathbf{b} gegeben und ist \mathbf{x} unbekannt, so ist $A\mathbf{x} = \mathbf{b}$ ein lineares Gleichungssystem. Es sei $A_j(\mathbf{b})$ die Matrix, die aus A entsteht, wenn der j -te Spaltenvektor von A durch \mathbf{b} ersetzt wird, und $I_j(\mathbf{x})$ sei die Matrix, wenn in der Einheitsmatrix I die j -te Spalte durch \mathbf{x} ersetzt wird. Dann folgt die *Cramersche Regel*,

$$x_j = \frac{|A_j(\mathbf{b})|}{|A|}. \quad (5.93)$$

wobei x_j die j -te Komponente von \mathbf{x} , also die j -te Unbekannte ist.

Man macht sich zunächst klar, dass $|I_j(\mathbf{x})| = x_j$ gilt: Für den Fall $n = 2$ hat man

$$\begin{vmatrix} x_1 & 0 \\ x_2 & 1 \end{vmatrix} = x_1 \cdot 0 - 0 \cdot x_2 = x_1, \quad \begin{vmatrix} 1 & x_2 \\ 0 & x_2 \end{vmatrix} = 1 \cdot x_2 - 0 \cdot x_1 = x_2.$$

Den Fall $n = 3$ illustriert man anhand der Sarrusschen Regel (s. unten), und für $n > 3$ kann man den Laplaceschen Entwicklungssatz (s. unten) heranziehen.

Weiter ist

$$A_j(\mathbf{b}) = [\mathbf{a}_1, \dots, \mathbf{a}_{j-1}, \mathbf{b}, \mathbf{a}_{j+1}, \dots, \mathbf{a}_n],$$

und weiter ist $A^{-1}\mathbf{a}_k = \mathbf{e}_k$ der k -te Einheitsvektor (wegen $A^{-1}A = I$). Dann folgt

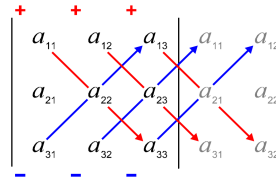
$$A^{-1}A_j(\mathbf{b}) = [\mathbf{e}_1, \dots, \mathbf{e}_{j-1}, A^{-1}\mathbf{b}, \mathbf{e}_{j+1}, \dots, \mathbf{e}_n] = I_j(\mathbf{b}) = x_j,$$

und

$$|A^{-1}A_j(\mathbf{b})| = |A^{-1}||A_j(\mathbf{b})| = x_j,$$

wegen (5.91), und wegen (5.92) folgt (5.93).

Abbildung 25: Zur Regel von Sarrus



15. **Regel von Sarrus** Es sei also A eine $(3, 3)$ -Matrix. Sarrus⁷⁵ fand eine Regel, nach der sich diese Formel für $\det(A)$ leicht finden läßt: man hängt an die Matrix die ersten beiden Spalten noch einmal an und verbindet dann zuerst die Diagonale von A , d.h. die Elemente a_{11} , a_{22} , a_{33} : sie bilden das erste Produkt, s. Abbildung 25. Dann verbindet man die Elemente a_{12} , a_{23} und a_{31} : sie bilden das zweite Produkt, und schließlich noch die Elemente a_{13} , a_{21} und a_{32} , sie bilden das dritte Produkt. Die ersten drei Produkte werden addiert. Dann folgen drei Produkte mit negativem Vorzeichen: man verbindet die Elemente von a_{13} bis a_{32} , deren Produkt das erste mit negativem Vorzeichen ist, dann folgt das Produkt $a_{11}a_{23}a_{12}$ und schließlich das Produkt $a_{12}a_{21}a_{33}$, vergl. Abbildung⁷⁶ 25. Für

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

hat man also

$$|A| = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{12} - a_{12}a_{21}a_{33}. \quad (5.94)$$

Die Regel von Sarrus ist sehr hilfreich, wenn man die Determinante einer $(3, 3)$ -Matrix "per Hand" berechnen muss, aber leider gilt sie *nur* für den Fall $n = 3$; für alle $n > 3$ gilt sie nicht.

16. **Entwicklung einer Determinante nach einer Spalte bzw. Zeile:** Der in (5.94) gegebene Ausdruck für $\det(A)$ ist durch eine Summe von Produkten aus jeweils n Elementen von A gegeben; der allgemeine Ausdruck ist nach (5.86)

$$|A| = \sum_{\tau \in \mathcal{P}} \operatorname{sgn}(\tau) a_{\tau(1),1} a_{\tau(2),2} \cdots a_{\tau(n),n}$$

τ ist eine der $n!$ Anordnungen (Permutationen) von n Elementen aus A , $\tau(k) = j$ ist eine der Zahlen $j \in \{1, 2, \dots, n\}$. Die Anordnung der Elemente

⁷⁴Gabriel Cramer (1704 – 1752), Genfer Mathematiker

⁷⁵Pierre Frédéric Sarrus (1798 – 1861), französischer Mathematiker

⁷⁶kopiert aus dem Wikipedia-Eintrag 'Regel von Sarrus'

innerhalb eines Produktes in (5.94) ist eine Konsequenz der Sarusschen Regel; da es auf die Anordnung der Elemente innerhalb eines Produktes nicht ankommt, kann diese so anordnen, dass die Komponente von \mathbf{a}_1 an erster, die von \mathbf{a}_2 an zweiter und die von \mathbf{a}_3 an dritter Stelle steht:

$$\begin{aligned} |A| &= a_{11}a_{22}a_{33} + a_{31}a_{12}a_{23} + a_{21}a_{32}a_{13} \\ &\quad - a_{31}a_{22}a_{13} - a_{11}a_{12}a_{23} - a_{21}a_{12}a_{33} \end{aligned} \quad (5.95)$$

Diese Anordnung entspricht der "kanonischen" Anordnung in (5.86), wie sie oben noch einmal angegeben wurde: die zweiten Indices zeigen immer die Reihenfolge 1, 2, 3.

Inspiziert man die Produkte, so fällt auf, dass die jeweils erste Zahl in einem Produkt, etwa a_{11} , die folgenden Zahlen so bestimmt, dass sie aus einer Restmatrix A_{11} kommen, die entsteht, wenn man die dem ersten Index entsprechende Zeile und die dem zweiten Index entsprechende Spalte aus A streicht. Für a_{11} erhält man die Restmatrix

$$A_{11} = \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix}, \quad |A_{11}| = a_{22}a_{33} - a_{23}a_{32} \quad (5.96)$$

Man sieht in (5.95), dass die Komponente a_{11} von \mathbf{a}_1 zweimal auftritt: $a_{11}a_{22}a_{33}$ und $a_{11}a_{12}a_{23}$. Die Teilprodukte $a_{22}a_{33}$ und $a_{12}a_{23}$ sind gerade die Summanden, die die Determinante von A_{11} definieren. Die Teilsumme $a_{11}a_{22}a_{33} - a_{11}a_{12}a_{23}$ in (5.95) läßt sich also zu

$$a_{11}a_{22}a_{33} - a_{11}a_{12}a_{23} = a_{11}|A_{11}| \quad (5.97)$$

zusammenfassen. Die zweite Komponente a_{21} von \mathbf{a}_1 taucht in (5.95) ebenfalls in zwei Summanden (Produkten) auf, und zwar als Differenz

$$a_{21}a_{32}a_{13} - a_{21}a_{12}a_{33} = -a_{21}|A_{21}|, \quad (5.98)$$

wobei jetzt

$$A_{21} = \begin{pmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{pmatrix}, \quad |A_{21}| = a_{12}a_{33} - a_{13}a_{32} \quad (5.99)$$

ist. A_{21} entsteht durch das Streichen der zweiten Zeile und der ersten Spalte aus A . Die Determinante $|A_{21}|$ geht hier mit negativem Vorzeichen ein, um die Differenz $a_{21}a_{32}a_{13} - a_{21}a_{12}a_{33}$ darzustellen. Allgemein wird das Vorzeichen für eine Differenz durch die Indices von A_{ji} gemäß $(-1)^{j+i}$ bestimmt: ist $j+i$ eine gerade Zahl, so ist das Vorzeichen positiv, ist $j+i$ ungerade, so ist das Vorzeichen negativ. Betrachtete man also die dritte Komponente a_{31} von \mathbf{a}_1 , so findet man die Differenz $a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13}$ im Ausdruck (5.95) für $|A|$. Für diese Differenz wird die dritte Zeile und wieder die erste Spalte aus A gestrichen; man erhält

$$A_{31} = \begin{pmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{pmatrix}, \quad |A_{31}| = a_{12}a_{23} - a_{13}a_{22}, \quad (5.100)$$

so dass

$$a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13} = a_{31}|A_{31}|. \quad (5.101)$$

Hier ist das Vorzeichen wieder positiv, weil $1 + 3$ eine gerade Zahl ist. Die Determinante $|A|$ ist die Summe der Differenzen (5.97), (5.98) und (5.101):

$$\begin{aligned} |A| &= a_{11}|A_{11}| - a_{21}|A_{21}| + a_{31}|A_{31}| \\ &= (-1)^{1+1}a_{11}|A_{11}| + (-1)^{2+1}a_{21}|A_{21}| + (-1)^{3+1}a_{31}|A_{31}| \\ &= \sum_{k=1}^3 (-1)^{k+1}a_{k1}|A_{k1}| \end{aligned} \quad (5.102)$$

(5.102) ist die *Entwicklung von $|A|$ nach der ersten Spalte von A* , eine andere Bezeichnung ist *Laplacescher Entwicklungssatz*⁷⁷. Die Matrizen A_{ji} heißen *Minoren* von A : sie entstehen durch Streichen einer Spalte und einer Zeile von A , sie sind also $(n-1, n-1)$ -Matrizen. Die Determinanten $|A_{ji}|$ der Minoren heißen *Kofaktoren* von A . Sie sind die Faktoren der Komponenten a_{kj} des Spaltenvektors \mathbf{a}_j , nach dem $|A|$ entwickelt wird.

17. **Zur Berechnung von A^{-1} :** A habe vollen Rang, so dass die Inverse A^{-1} existiert. Es ist $A^{-1}A = AA^{-1} = I$. Es sei \mathbf{y}_j der j -te Spaltenvektor von A^{-1} ; dann folgt $A\mathbf{y}_j = \mathbf{e}_j$, \mathbf{e}_j die j -te Spalte der Einheitsmatrix I . Die Beziehung $A\mathbf{y}_j = \mathbf{e}_j$ ist ein Gleichungssystem mit \mathbf{y}_j als unbekanntem Vektor, so dass nach der Cramerschen Regel die i -te Komponente von \mathbf{y}_j durch

$$y_{ij} = \frac{|A_j(\mathbf{e}_i)|}{|A|} \quad (5.103)$$

gegeben ist. Es sei A_{ij} die Matrix, die aus A durch Streichen der i -ten Zeile und der j -ten Spalte entsteht. Dann gilt

$$|A_j(\mathbf{e}_i)| = (-1)^{i+j}|A_{ij}| := \alpha_{ij}. \quad (5.104)$$

$|A_{ij}|$ heißt der *Kofaktor* oder *Minor* des Elements a_{ij} von A . Die Matrix $\text{Adj}(A)$ der α_{ij} -Werte heißt *Adjunkte* der Matrix A , und (5.103) impliziert dann

$$A^{-1} = \frac{1}{|A|}\text{Adj}(A). \quad (5.105)$$

Zu zeigen ist (5.104). $A_j(\mathbf{e}_i)$ entsteht, wenn die j -te Spalte von A durch ein Einheitsvektor \mathbf{e}_i ersetzt wird. Man kann die Determinante $|A_j(\mathbf{e}_i)|$ bestimmen, indem man sie nach der j -ten Spalte entwickelt:

$$|A_j(\mathbf{e}_i)| = \sum_{k=1}^n (-1)^{k+j} e_{ij} |A_{ij}|.$$

Aber $e_{ij} = 0$ für alle $i \neq k$, so dass nur ein Summand – der für $k = i$ – übrig bleibt, also hat man (5.104).

⁷⁷Pierre Simon Laplace (1749 – 1827), französischer Mathematiker

18. Es sei A eine orthonormale (n, n) -Matrix mit vollem Rang. Dann gilt

$$|A| = 1. \quad (5.106)$$

Es ist

$$(\det(A))^2 = \det(A) \det(A) = \det(A') \det(A) = \det(A'A) = \det(I) = 1,$$

und es ist $\sqrt{(\det(A))^2} = \det(A) = \sqrt{1} = 1$.

19. Es sei $A = \text{diag}(a_{11}, a_{22}, \dots, a_{nn})$ eine Diagonalmatrix. Dann gilt

$$|A| = |\text{diag}(a_{11}, a_{22}, \dots, a_{nn})| = \prod_{i=1}^n a_{ii}, \quad (5.107)$$

A kann in der Form $A = [a_{11}\mathbf{e}_1, a_{22}\mathbf{e}_2, \dots, a_{nn}\mathbf{e}_n]$ geschrieben werden. Nach Folgerung 2, Gleichung (5.87) gilt dann

$$|A| = a_{11}a_{22} \cdots a_{nn}|I| = \prod_{i=1}^n a_{ii}.$$

20. **Ähnliche Matrizen** Es seien M und X Matrizen mit vollem Rang, wobei M symmetrisch sei. Die Matrix N sei durch

$$X^{-1}MX = N$$

definiert. Die Matrizen M und N heißen dann *ähnlich*, und es gilt

$$|X^{-1}MX| = |N|. \quad (5.108)$$

Insbesondere gilt

$$|N| = \prod_{i=1}^n \lambda_i, \quad (5.109)$$

wobei $\lambda_1, \dots, \lambda_n$ die (positiven) Eigenwerte von M sind.

Nach dem Produktsatz und nach Folgerung 14, Gleichung (5.92) gilt

$$|X^{-1}MX| = |X^{-1}M||X| = |X^{-1}||M||X| = \frac{1}{|X|}|M||X| = |M| = |N|.$$

Insbesondere sei $X = P$, P die Matrix der (orthonormalen) Eigenvektoren von M . Dann gilt $P'MP = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, λ_j , $j = 1, \dots, n$ die zu den Eigenvektoren korrespondierenden Eigenwerte von M . Man hat dann $MP = P\Lambda$ und der Produktsatz liefert

$$|PM| = |P||M| = |P||\Lambda|.$$

Da P aber orthonormal ist folgt $|P| = 1$, so dass $|M| = |\Lambda| = \prod_{i=1}^n \lambda_i$ nach Folgerung 5.115, Gleichung (5.107).

21. (**Charakteristisches Polynom**) Es sei A eine (n, n) -Matrix; \mathbf{v} sei ein Eigenvektor von A und λ sei der zugehörige Eigenwert, so dass

$$A\mathbf{v} = \lambda\mathbf{v}$$

gilt. Es folgt

$$A\mathbf{v} - \lambda\mathbf{v} = (A - \lambda I)\mathbf{v} = \vec{0}.$$

Die Gleichung

$$|A - \lambda I| = \lambda^n - a_1\lambda^{n-1} + a_2\lambda^{n-2} + \cdots + (-1)^n a_n = 0 \quad (5.110)$$

heißt *charakteristische Gleichung* von A , das Polynom auf der rechten Seite heißt *charakteristisches Polynom* von A . Dabei gilt

$$a_1 = \sum_{i=1}^n a_{ii}, \quad a_n = |A|, \quad (5.111)$$

d.h. a_1 ist die Spur von A . Die Nullstellen des Polynoms sind gleich den Eigenwerten von A .

Das charakteristische Polynom kann z.B. durch die Anwendung des Laplace'schen Entwicklungssatzes auf $|A - \lambda I|$ hergeleitet werden; die Koeffizienten a_i ergeben sich aus den Elementen der Matrix $A - \lambda I$. Die Herleitung soll hier übergangen werden. Die Bestimmung der Eigenwerte und der Eigenvektoren wird im Allgemeinen von Programmen übernommen (Golub & van Loan (2013)).

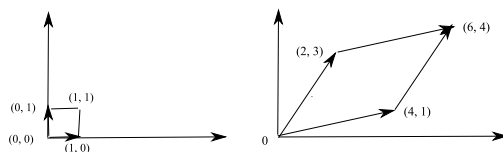
5.10.1 Transformationen und Volumen

Matrizen definieren Transformationen bzw. Abbildungen von Vektoren: ist A eine (m, n) -Matrix und \mathbf{x} ein n -dimensionaler Vektor, so ist $\mathbf{y} = A\mathbf{x}$ ein m -dimensionaler Vektor. \mathbf{y} ist eine Linearkombination der Spaltenvektoren von A , und man kann sagen, dass \mathbf{x} durch A in \mathbf{y} transformiert wird, oder dass der Vektor \mathbf{x} durch A auf den Vektor \mathbf{y} abgebildet wird. im Folgenden spielen insbesondere quadratische Matrizen eine Rolle; sie bilden n -dimensionale Vektoren auf n -dimensionale Vektoren ab; diese Matrizen definieren einen Endomorphismus, also eine Abbildung eines n -dimensionalen Vektorraums in sich selbst (s. Abschnitt 4.5). Es sei A also eine (n, n) -Matrix. Man kann eine Menge \mathcal{G} von Vektoren \mathbf{x} auswählen und die Menge $\mathcal{G}^* = \{\mathbf{y} | \mathbf{y} = A\mathbf{x}, \mathbf{x} \in \mathcal{G}\}$ bestimmen. Hat die Matrix vollen Rang, so existiert die Determinante von A , $|A| \neq 0$. Es wird nun die geometrische Bedeutung von $|A|$ betrachtet.

Der Einfachheit halber wird eine $(2, 2)$ -Matrix betrachtet. Es sei insbesondere \mathcal{G} ein Quadrat mit den Seitenlängen gleich 1. Dann ist \mathcal{G}^* durch ein Parallelogramm gegeben, s. Abbildung 26. Das Quadrat ist durch die Einheitsvektoren $\mathbf{e}_1 = (1, 0)'$ und $\mathbf{e}_2 = (0, 1)'$ definiert. Die Matrix A ist durch

$$A\mathbf{x} = \mathbf{y}, \quad A = [\mathbf{a}_1, \mathbf{a}_2] = \begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix} \quad (5.112)$$

Abbildung 26: Abbildung eines Quadrats auf ein Parallelogramm



gegeben. Wählt man $\mathbf{x}_1 = \mathbf{e}_1$ und $\mathbf{x}_2 = \mathbf{e}_2$, so erhält man

$$A\mathbf{e}_1 = \mathbf{a}_1 = \begin{pmatrix} 4 \\ 1 \end{pmatrix}, \quad A\mathbf{e}_2 = \mathbf{a}_2 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \quad (5.113)$$

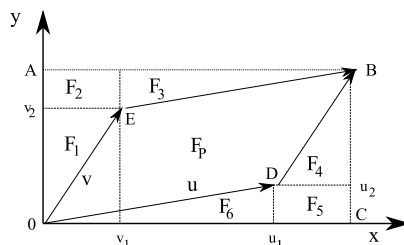
also gerade die Vektoren, die das Parallelogramm definieren. Wählt man $\mathbf{x} = \vec{0}$, so erhält man $\mathbf{y} = A\mathbf{x} = \vec{0}$; man bildet also den Punkt $(0,0)$ im linken Koordinatensystem auf den Punkt $(0,0)$ im rechten Koordinatensystem ab. Der Punkt $(6,4)$ des Parallelogramms ergibt sich durch die Summe der Vektoren $\mathbf{y}_1 = (4,1)$ und $\mathbf{y}_2 = (2,3)$:

$$\mathbf{y}_1 + \mathbf{y}_2 = \begin{pmatrix} 4 \\ 1 \end{pmatrix} + \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 6 \\ 4 \end{pmatrix}.$$

Es läßt sich nun zeigen, dass der Flächeninhalt des Parallelogramms durch die Determinante $|A|$ von A gegeben ist.

Das Quadrat hat sicherlich den Flächeninhalt 1. Der Flächeninhalt des Parallelogramms ergibt sich aus der bekannten Formel $F = hg$, wobei h die Höhe des Parallelogramms und g die Länge des Vektors $\mathbf{a}_2 = (3,2)'$ ist. h ist die Länge des Vektors \mathbf{z} in Abbildung ??, Seite ??, und nach Gleichung (??), Seite ??, läßt sich $h = \|\mathbf{z}\|$ ausrechnen. Für die Zwecke dieses Abschnitts ist es aber sinnvoller, den Flächeninhalt ohne diese Berechnungen zu bestimmen, s. Abbildung 27.

Abbildung 27: Durch die Vektoren $\mathbf{a}_1 = (u_1, u_2)'$, und $\mathbf{a}_2 = (v_1, v_2)'$ definiertes Parallelogramm



Satz 5.6 Es sei $|A| = a_{11}a_{22} - a_{12}a_{21}$. Der Flächeninhalt F_p des durch die Vektoren \mathbf{u} und \mathbf{v} definierten Parallelogramms (s. Abb. 27) ist durch

$$F_p = u_1v_2 - u_2v_1 = a_{11}a_{22} - a_{12}a_{21} \quad (5.114)$$

gegeben.

Beweis: Offenbar ist F_P gleich der Fläche $F_{tot} = (u_1 + v_1)(u_2 + v_2)$ des Rechtecks 0ABC minus der Summe der Flächen F_1 bis F_6 . Es ist

$$F_1 = F_4 = \frac{1}{2}v_1v_2 \quad (5.115)$$

$$F_2 = F_4 = v_1u_2 \quad (5.116)$$

$$F_3 = F_6 = \frac{1}{2}u_1u_2 \quad (5.117)$$

Einsetzen der Ausdrücke für F_1 bis F_6 liefert sofort

$$F_P = F_{tot} - (F_1 + \dots + F_6) = |u_1v_2 - v_1u_2|, \quad (5.118)$$

also den Absolutbetrag der Determinante von A . □

Die Aussage, dass der Flächeninhalt des durch A repräsentierten Parallelogramms gleich dem Absolutbetrag der Determinante von A ist, kann auf allgemeine (n, n) -Matrizen verallgemeinert werden, wobei noch zu klären ist, warum vom *Absolutbetrag* der Determinante die Rede ist. Im vorangegangenen Beispiel ist die Determinante positiv, und Flächeninhalte bzw. allgemein Volumen sind stets positive Zahlen. Vertauscht man allerdings die Spalten von A , so dass die Matrix $B = [\mathbf{a}_2, \mathbf{a}_1]$ entsteht, so findet man, dass die Determinante

$$|B| = a_{12}a_{21} - a_{11}a_{22} = -(a_{11}a_{22} - a_{12}a_{21}) = -|A|$$

negativ ist. In diesem Fall muss der Absolutbetrag von $|B|$ genommen werden, wenn von einem Volumen bzw. Flächeninhalt gesprochen werden soll. Auf die Bedeutung des Vorzeichens einer Determinante wird weiter unten eingegangen.

Das Parallelogramm entsteht, wenn alle Vektoren innerhalb einschließlich der Ränder des Quadrats in Abbildung 26 mit der Matrix A transformiert werden. Man kann nun ein beliebiges Gebiet G aus der xy -Ebene definieren und alle Vektoren \mathbf{x} aus G mit A transformieren und erhält eine Menge \mathcal{G}^* von Vektoren \mathbf{y} ,

$$\mathcal{G}^* = \{\mathbf{y} | \mathbf{y} = A\mathbf{x}, \mathbf{x} \in G\}.$$

Die Frage ist, was über das Volumen von \mathcal{G}^* gesagt werden kann.

Zunächst werde statt eines Quadrats mit der Seitenlänge 1 ein Quadrat mit der Seitenlänge $a \neq 1$ betrachtet. Dieses Quadrat ist das Gebiet G und hat den Flächeninhalt $F_G = a^2$. Die Einheitsvektoren \mathbf{e}_j gehen dann über in die Vektoren $a\mathbf{e}_j$, die nun auf die Vektoren

$$Aa\mathbf{e}_1 = a\mathbf{a}_1 = \begin{pmatrix} 4a \\ 1a \end{pmatrix}, \quad Aa\mathbf{e}_2 = a\mathbf{a}_2 = \begin{pmatrix} 2a \\ 3a \end{pmatrix}, \quad (5.119)$$

abgebildet werden; es entsteht ein Parallelogramm, dessen Seitenlänge und Diagonale um den Faktor a verändert werden. Der Flächeninhalt des Parallelogramms

beträgt nun nach (5.118) (die Komponenten der Vektoren \mathbf{u} und \mathbf{v} werden alle mit dem Faktor a multipliziert)

$$F_p = F_{\mathcal{G}^*} = a^2 |u_1 v_2 - v_1 u_2| = F_G |A|. \quad (5.120)$$

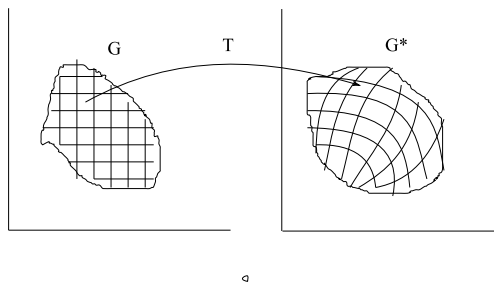
Für die Determinante $|A|$ folgt daraus sofort

$$|\det(A)| = \frac{F_{\mathcal{G}^*}}{F_G}, \quad (5.121)$$

d.h. im allgemeinen Fall ist $|A|$ gleich dem *Verhältnis* der Flächeninhalte bzw. im mehr als 2-dimensionalen Fall der Volumina der Gebiete \mathcal{G}^* und F_G . Die Redeweise, dass $|A|$ gleich dem Volumen eines Parallelepipeds ist, impliziert den Fall $F_G = 1$.

Das Gebiet G ist aber nicht notwendig ein Quadrat, sondern kann irgendein zusammenhängendes Gebiet sein (Abbildung 28). T in Abb. 28 kann irgendeine Transformation sein, die lineare Transformation, wie sie durch eine (n, n) -Matrix

Abbildung 28: Allgemeine Gebietstransformation



gegeben ist, ist ein Spezialfall. Das Volumen des Gebiets G kann abgeschätzt werden, indem man G in Quadrate unterteilt und die Summe der Fläche dieser Quadrate bildet. Die Approximation ist um so genauer, die kleiner die Quadrate und je größer ihre Anzahl. Im Grenzfall geht man zu infinitesimalen Flächen über und bestimmt die Fläche (das Volumen) von G durch ein Gebietsintegral. Analog dazu verfährt man mit der Fläche bzw. dem Volumen von $\mathcal{G}^* = T(G)$. Die zu den Quadraten in G korrespondierenden Flächen in \mathcal{G}^* können im Limit durch infinitesimale Quadrate angenähert werden, die eine lineare Transformation A der Quadrate in G sind, d.h. A ist eine Matrix, und man findet analog zu (5.120)

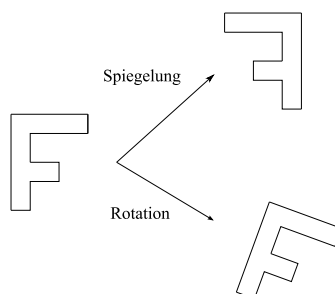
$$\text{vol}(G) = |\det(A)| \text{vol}(\mathcal{G}^*) \quad (5.122)$$

oder

$$|\det(A)| = \frac{\text{vol}(G)}{\text{vol}(\mathcal{G}^*)}. \quad (5.123)$$

Einen exakten Beweis findet man z.B. in Courant II (1963); einen moderner formulierten . allerdings dann auch abstrakteren Beweis findet man in Fischer (1997). Eine für die multivariate Statistik relevante Anwendung dieses Ergebnisses wird im folgenden Abschnitt betrachtet.

Abbildung 29: Transformation: Spiegelung (Änderung der Orientierung) und Rotation



5.10.2 Transformationen und Orientierung

Zum Abschluß soll noch kurz auf die Bedeutung des Vorzeichens einer Determinante eingegangen werden. Dazu werden die beiden Matrizen⁷⁸

$$A = \begin{pmatrix} -1 & -\frac{1}{4} \\ 1 & -\frac{1}{2} \end{pmatrix}, \quad B = \begin{pmatrix} -\frac{1}{4} & 1 \\ 1 & \frac{1}{2} \end{pmatrix}. \quad (5.124)$$

Man rechnet leicht nach, dass $|A| = 3/4$ und $|B| = -9/8$. Die Determinanten sind ungleich Null, weshalb die Spaltenvektoren \mathbf{a}_1 und \mathbf{a}_2 von A linear unabhängig sind, ebenso die Spaltenvektoren \mathbf{b}_1 und \mathbf{b}_2 von B , d.h. sowohl $\{\mathbf{a}_1, \mathbf{a}_2\}$ wie $\{\mathbf{b}_1, \mathbf{b}_2\}$ sind Basen des \mathbb{R}^2 , also kann man mit beiden Basen alle Vektoren des \mathbb{R}^2 erzeugen. Wie gezeigt wurde, kann man $|A|$ als Volumen bzw. als Quotient von Volumina deuten, aber $|B|$ nicht, da Volumina nicht negativ sein können. Allgemein gilt für Matrizen, dass sie Vektoren transformieren, d.h. gilt $A\mathbf{x} = \mathbf{y}$, so transformiert A den Vektor \mathbf{x} in den Vektor \mathbf{y} , und natürlich gilt die analoge Aussage für B . Die Transformation bewirkt allgemein eine Rotation und eine Veränderung der Länge von \mathbf{x} , oder eines von beiden (und in speziellen Fällen gar keine Veränderung). Nun können die Vektoren $\mathbf{x}_1, \dots, \mathbf{x}_p$ eine Figur abbilden, eben z.B. ein F . Die Matrix A rotiert alle \mathbf{x}_j und verändert die Längen, und die Figur, hier das F , wird als Ganzes "nur" rotiert. Wendet man dagegen die Matrix B auf die \mathbf{x}_j an, so wird das F nicht nur rotiert und eventuell verzerrt, sondern darüber hinaus gespiegelt, d.h. es wird die Orientierung der Figur oder allgemein der Vektorkonfiguration verändert (Abbildung 29). Die Spiegelung oder Veränderung der Orientierung wird durch das Vorzeichen der Determinante, hier von B , angezeigt. Man kann sagen, dass der Betrag der Determinante von B , $|\det(B)|$ ein Volumen repräsentiert, und das Vorzeichen der Determinante zeigt an, ob die Orientierung der Konfiguration durch die Transformation geändert wird.

⁷⁸Fischer (1997), p. 203)

Der Begriff der Orientierung kann mithilfe des *Vektorprodukts* illustriert werden, das im \mathbb{R}^3 definiert ist. Gegeben seien drei 3-dimensionale Vektoren \mathbf{x} , \mathbf{y} , und \mathbf{z} . Die Vektoren \mathbf{x} und \mathbf{y} liegen in einer Ebene (d.h. sie definieren eine Ebene im \mathbb{R}^3 , und \mathbf{z} steht senkrecht auf dieser Ebene. Ohne Beschränkung der Allgemeinheit kann man annehmen, dass diese Ebene durch die kanonischen Einheitsvektoren \mathbf{e}_1 und \mathbf{e}_2 definiert wird; dann steht \mathbf{e}_3 senkrecht auf dieser Ebene, hat also dieselbe Orientierung wie \mathbf{z} . Das Vektorprodukt ist dann definiert durch

$$\mathbf{x} \times \mathbf{y} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} \quad (5.125)$$

Entwickelt man diese Determinante nach der ersten Zeile, so erhält man

$$\begin{aligned} \mathbf{z} = \mathbf{x} \times \mathbf{y} &= \mathbf{e}_1 \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} - \mathbf{e}_2 \begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix} + \mathbf{e}_3 \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \\ &= (x_2y_3 - x_3y_2)\mathbf{e}_1 + (x_3y_1 - x_1y_3)\mathbf{e}_2 + (x_1y_2 - x_2y_1)\mathbf{e}_3 \\ &= \begin{pmatrix} x_2y_3 - x_3y_2 \\ x_3y_1 - x_1y_3 \\ x_1y_2 - x_2y_1 \end{pmatrix} \end{aligned} \quad (5.126)$$

$\mathbf{x} \times \mathbf{y}$ ist ein Vektor, der orthogonal zu den Vektoren \mathbf{x} und \mathbf{y} ist; so ist

$$\mathbf{x}'\mathbf{z} = \mathbf{x}'(\mathbf{x} \times \mathbf{y}) = x_1x_2y_3 - x_1x_3y_2 + x_2x_3y_1 - x_1x_2y_3 + x_3x_1y_2 - x_3x_2y_1 = 0,$$

und analog dazu findet man $\mathbf{y}'(\mathbf{x} \times \mathbf{y}) = 0$. Also steht \mathbf{z} senkrecht auf \mathbf{x} und \mathbf{y} und damit auf der durch diese beiden Vektoren definierten Ebene.

Man rechnet leicht nach, dass

$$\mathbf{x} \times \mathbf{y} = -(\mathbf{y} \times \mathbf{x}). \quad (5.127)$$

Es läßt sich nun zeigen, dass dem Vektor \mathbf{z} eine Drehrichtung zugeordnet werden kann; dieser Sachverhalt ergibt sich aus der Analyse der Bewegung eines Massepunktes auf einer Kreisbahn; die Details werden hier übergangen (man findet die Analyse Lehrbüchern der Physik, z.B. in Westphal (1959), p. 16)). Zerlegt man die bei Drehbewegungen auftretenden Kräfte, so zeigt \mathbf{z} die Orientierung der resultierenden Kraft und $\|\mathbf{z}\|$ ist die Größe der Kraft. Ist $\|\mathbf{z}\|$ die Kraft, mit der man eine Schraube in ein Holzstück dreht, so entspricht die Orientierung der üblichen Schraubendrehung; $-\mathbf{z}$ entspricht dann der Orientierung der Drehung beim Herausschrauben der Schraube.

Um den Begriff der Orientierung formal fassen zu können, wird der Begriff der *geordneten Basis* eingeführt; dies ist eine Basis $\mathbf{b}_1, \dots, \mathbf{b}_n$, bei der es auf die Anordnung der \mathbf{b}_j ankommt. Nun seien $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ und $\mathcal{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_n\}$ zwei geordnete Basen eines n -dimensionalen Vektorraumes V . Dann existiert stets eine Matrix A , die die Basis \mathcal{B} in die Basis \mathcal{C} überführt⁷⁹, so dass mit $B =$

⁷⁹Warum?

$[\mathbf{b}_1, \dots, \mathbf{b}_n]$ und $C = [\mathbf{c}_1, \dots, \mathbf{c}_n]$ die Beziehung $C = AB$ gilt. Dann heißen die Basen \mathcal{B} und \mathcal{C} *gleichorientiert*, wenn $|A| > 0$ ist. Gilt diese Bedingung *nicht*, so heißen sie *verschieden orientiert*; dann gilt $|A| < 0$.

Vertauscht man in einer Matrix zwei Vektoren miteinander, so ändert sich das Vorzeichen der Determinante. Hat man also $\mathcal{B}_1 = \{\mathbf{b}_1, \dots, \mathbf{b}_j, \mathbf{b}_{j+1}, \dots, \mathbf{b}_n\}$ und $\mathcal{B}_2 = \{\mathbf{b}_1, \dots, \mathbf{b}_{j+1}, \mathbf{b}_j, \dots, \mathbf{b}_n\}$, so sind $\mathcal{L}(\mathcal{B}_1)$ und $\mathcal{L}(\mathcal{B}_2)$ verschieden orientierte Vektorräume. insbesondere heißt eine geordnete Basis *positiv orientiert*, wenn sie dieselbe Orientierung wie die kanonische Basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ hat; andernfalls heißt sie *negativ orientiert*.

5.10.3 Die Transformation zufälliger Veränderlicher

Integrale lassen sich oft leichter bestimmen, wenn man die Integrationsvariable durch eine andere Variable ersetzt. Um die Substitution von Variablen diskutieren zu können, wird der Begriff der Differenzierbarkeit benötigt:

Definition 5.8 Eine Funktion $f(x)$ mit einer Variablen heißt differenzierbar in x , wenn eine Zahl $f'(\xi)$ existiert derart, dass

$$f(x + \xi) = f(x) + f'(x)\xi + o(\xi), \quad \lim_{\xi \rightarrow 0} \frac{o(\xi)}{\|\xi\|} = 0. \quad (5.128)$$

gilt. $o(\xi)$ ist eine Funktion von ξ , die schneller gegen Null geht als ξ . Dabei ist

$$f'(x) = \lim_{\xi \rightarrow 0} \frac{f(x + \xi) - f(x)}{\xi} \quad (5.129)$$

Der mehrdimensionale Fall ist analog definiert:

Definition 5.9 Es sei $U \subset \mathbb{R}^n$ eine offene Menge und es sei \mathbf{f} eine Abbildung $\mathbf{f}: U \rightarrow \mathbb{R}^m$. in $\mathbf{x} \in U$ total differenzierbar (auch einfach differenzierbar), wenn eine weitere lineare Abbildung $A: \mathbb{R}^n \rightarrow \mathbb{R}^m$ gibt derart, dass in einer Umgebung von \mathbf{x}

$$\mathbf{f}(\mathbf{x} + \boldsymbol{\xi}) = \mathbf{f}(\mathbf{x}) + A\boldsymbol{\xi} + o(\|\boldsymbol{\xi}\|) \quad (5.130)$$

gilt.

Hier ist $A = (a_{ij})$ eine (m, n) -Matrix und $\boldsymbol{\xi}$ ein n -dimensionaler Vektor; weiter ist $\mathbf{f} = (f_1, \dots, f_m)'$ ein m -dimensionaler Vektor. Offenbar ist \mathbf{f} differenzierbar, wenn die Komponenten f_i differenzierbar sind.

Satz 5.7 Es seien \mathbf{f} und A wie in Definition 5.9 definiert. Dann gilt (i) \mathbf{f} ist in \mathbf{x} stetig, (ii) die Komponenten f_i von \mathbf{f} sind in \mathbf{x} partiell differenzierbar, d.h. es gilt

$$\frac{\partial f_i}{\partial x_j} = a_{ij}. \quad (5.131)$$

Beweis: (vergl. Forster (2), p. 78) Es ist

$$\lim_{\boldsymbol{\xi} \rightarrow \vec{0}} A\boldsymbol{\xi} = \mathbf{f}'(\boldsymbol{\xi}),$$

und dies heißt, dass \mathbf{f} in \mathbf{x} stetig ist. Weiter gilt

$$f_i(\mathbf{x} + \boldsymbol{\xi}) = f_i(\mathbf{x}) + \sum_{j=1}^n a_{ij}\xi_j + o(\|\boldsymbol{\xi}\|), \quad i = 1, \dots, n$$

d.h.

$$f_i(\mathbf{x} + h\mathbf{e}_j) = f_i(\mathbf{x}) + ha_{ij} + o(\|h\mathbf{e}_j\|),$$

so dass

$$\frac{\partial f_i}{\partial x_j} = \lim_{h \rightarrow 0} \frac{f_i(\mathbf{x} + h\mathbf{e}_j) - f_i(\mathbf{x})}{h} = a_{ij} + o(\|h\mathbf{e}_j\|) = a_{ij}.$$

□

Die Implikation des Satzes ist, dass die Matrix A durch die Abbildung \mathbf{f} eindeutig bestimmt ist.

Definition 5.10 Die Matrix A heißt das Differential oder Funktionalmatrix oder Jacobi-Matrix⁸⁰ der Abbildung \mathbf{f} im Punkt \mathbf{x} .

Im mehrdimensionalen Fall wird also der Differentialquotient durch eine Matrix von partiellen Differentialquotienten ersetzt. Man hat die Schreibweisen

$$(D\mathbf{f})(\mathbf{x}) := J_{\mathbf{f}} := \frac{\partial(f_1, \dots, f_n)}{\partial(x_1, \dots, x_n)}(\mathbf{x}) := \begin{pmatrix} \frac{\partial f_1}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} \\ \vdots \\ \frac{\partial f_n}{\partial x_1} \\ \frac{\partial f_n}{\partial x_2} \\ \vdots \end{pmatrix}, \quad (5.132)$$

wobei ":= " bedeutet, dass der Ausdruck links von diesem Zeichen durch den rechts davon stehenden definiert wird. In ausgeschriebener Form hat die Funktional- oder Jacobi-Matrix die Gestalt

$$J_{\mathbf{f}} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \dots & \frac{\partial f_n}{\partial x_n} \end{pmatrix} \quad (5.133)$$

Die Differenzierbarkeit von \mathbf{f} bedeutet nach (5.130), dass in einer hinreichend kleinen Umgebung von \mathbf{x} linear approximiert werden kann. $A\boldsymbol{\xi}$ ist ein Vektor, der die infinitesimale Veränderung von \mathbf{f} im Punkt (\mathbf{x}) angibt.

Die Kettenregel: Im mehrdimensionalen Fall gilt

Satz 5.8 Es seien $U \subset \mathbb{R}^n$ und $V \subset \mathbb{R}^m$ offene Mengen und $\mathbf{f}: U \rightarrow \mathbb{R}^m$, $\mathbf{g}: V \rightarrow \mathbb{R}^k$ offene Mengen und es gelte $f: U \rightarrow \mathbb{R}^m$, $g: V \rightarrow \mathbb{R}^k$, $(\mathbf{U}) \subset \mathbf{V}$. \mathbf{f} sei in \mathbf{x} und \mathbf{g} sei in $\mathbf{y} := \mathbf{f}(\mathbf{x})$ differenzierbar. Es sei $\mathbf{g} \circ \mathbf{f}: U \rightarrow \mathbb{R}^k$ die aus \mathbf{g} und \mathbf{f} zusammengesetzte Abbildung; sie sei in \mathbf{x} differenzierbar und es gilt

$$D(\mathbf{g} \circ \mathbf{f})(\mathbf{x}) = (D\mathbf{g})\mathbf{f}(\mathbf{x}) \cdot D\mathbf{f}(\mathbf{x}). \quad (5.134)$$

⁸⁰Carl Gustav Jacobi (1804 – 1851), Mathematiker

Beweis: Siehe z.B. Forster (2), p.80.

Zur einführenden Illustration werde die Substitutionsregel bei der Integration einer Funktion von nur einer Veränderlichen erinnert. So sei $f(x)$ eine Funktion auf einem Intervall $I \subseteq \mathbb{R}$. Eine neue Variable u kann durch eine auf dem Intervall I invertierbare Funktion $\psi(u) = x$ eingeführt werden. Ist $F(x)$ die Stammfunktion von f , d.h. gilt $F(x) = \int f(x)dx$ ⁸¹. So kann F als Verkettung $F \circ \varphi = F(\varphi(u))$ gesehen werden. Man kann dann

$$F(\varphi(u)) = G(u). \quad (5.135)$$

und die Kettenregel führt auf

$$\frac{dF(\varphi(u))}{du} = \frac{dF(\varphi)}{d\varphi} \frac{d\varphi}{du} = f(\varphi)\varphi' = g(u). \quad (5.136)$$

und

$$\begin{aligned} \int_a^b f(\varphi(u))\varphi'(u)du &= F(\varphi(b)) - F(\varphi(a)) \\ &= \int_{\varphi(a)}^{\varphi(b)} f(x)dx, \quad x = \varphi(u) \end{aligned} \quad (5.137)$$

Diese Aussage "ist die Grundlage für die Einführung neuer Veränderlicher in ein Integral" (Courant I (1961), p. 183). Wie Courant I, p. 184, ausführt, folgt man aber der Gleichung (5.137) von rechts nach links: gegeben ist eine zu integrierende Funktion $f(x)$, Die Aufgabe, das Integral zu bestimmen, kann u. U. vereinfacht werden, indem man für x die Funktion $x = \varphi(u)$ und damit die neue Integrationsvariable u einführt. Man bestimmt also

$$G(u) = \int f(\varphi(u))\varphi'(u)du$$

und ersetzt dann u wieder durch x . Dazu muss φ umkehrbar sein, d.h. es muss die inverse Funktion $\psi(x) = u = \varphi^{-1}(x)$ existieren mit $\psi'(x) \neq 0$ für alle $x \in I$. Als *Grundformel* (Courant I, p. 184) erhält man dann

$$\int f(x)dx = \int f(\varphi(u))\varphi'(u)du, \quad u = \psi(x) \quad (5.138)$$

Man bestimmt also das Integral auf der linken Seite der Gleichung, indem man das Integral auf der rechten Seite findet und anschließend u über die Abbildung $u = \psi(x)$ wieder einführt. Es sei angemerkt, dass

$$\varphi' = \frac{1}{\psi'} \quad (5.139)$$

⁸¹Diese Schreibweise mag etwas lax sein, weil die Laufvariable x nicht vom Argument x in $F(x)$ unterschieden wird, ist aber als möglichst einfache, intuitive Schreibweise zu lesen (vergl. Courant I (1961), p. 182)

gilt; hier wird klar, warum $\psi' \neq 0$ für $x \in I$ gefordert werden muss.

Es ist wichtig, sich die Bedeutung des Faktors φ' in Gleichung (5.137) noch einmal klar machen: offenbar genügt es nicht, x einfach durch $\varphi(u)$ zu ersetzen, es muss eben noch der Faktor $\varphi'(u)$ hinzugefügt werden. Im Vergleich mit der rechten Seite von (5.137) erhält man die Beziehung

$$dx = \varphi'(u)du \quad (5.140)$$

Die Transformation impliziert i. A., dass x und u und damit auch dx und du auf verschiedenen Skalen laufen⁸² Bei nichtlinearen Funktionen φ bzw. ψ wird die Beziehung zwischen dx und dy aber nicht nur durch eine Konstante bestimmt; $\varphi'(u)$ kann ja eine Funktion von u sein.

In (5.138) repräsentiert $f(x)dx$ ein "infinitesimales Rechteck" der Breite dx und der Höhe $f(x)$. Beim Integral auf der rechten Seite werden aber nicht die infinitesimalen Rechtecke $f(\varphi(u))du$ aufsummiert, sondern die Rechtecke $f(\varphi(u))\varphi'(u)du$, was man so lesen kann, dass der Verlauf von f durch den u.U. von u abhängigen Faktor φ' modifiziert wird. Ist $\varphi' = a$ eine Konstante, so werden die $f(\varphi(u))$ -Werte vergrößert, wenn $a > 1$, und verkleinert, wenn $a < 1$ ist. Die Veränderung der Skala geht also mit einer Modifikation des Verlaufs von $f(\varphi(u))$ einher.

Im Übrigen kann man die infinitesimalen Flächen $f(x)dx$ etc als Spezialfall des allgemeinen Volumenbegriffs auffassen; diese Sichtweise wird weiter unten noch elaboriert.

Beispiel 5.7 Eine einfache, illustrierende Anwendung ist die Integration der Funktion $f(x) = \sin(2x)$. Man kann hier die Transformation $u = \psi(x) = 2x$ einführen; dann ist

$$\frac{du}{dx} = \frac{d\psi(x)}{dx} = \psi'(x) = 2, \quad \text{d.h. } du = 2dx, \text{ bzw. } dx = \frac{1}{2}du$$

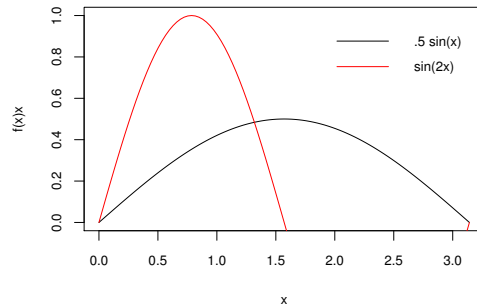
einer Veränderung auf der x -Skala entspricht eine doppelt so große Veränderung auf der u -Skala, und einer Veränderung auf der u -Skala entspricht die Hälfte dieser Veränderung auf der x -Skala. Diese Beziehungen übertragen sich auf die infinitesimalen Größen dx und du . Sind a und b die Grenzen des Integrals auf der u -Skala, so erhält man wegen $u/2 = x$ die Grenzen $a/2$ und $b/2$ auf der x -Skala; speziell für $a = 0$ und $b = \pi$ erhält man also

$$\begin{aligned} \int_0^{\pi/2} \sin(2x)dx &= \int_0^{\pi} \sin(u) \frac{1}{2}du = \frac{1}{2} \int_0^{\pi} \sin(u)du \\ &= -\frac{1}{2} \cos(\pi) + \frac{1}{2} \cos(0) = 1. \end{aligned} \quad (5.141)$$

Abbildung 30 illustriert das Prinzip: Die Funktion $\sin(2x)$ "überdeckt" nur die

⁸²Ein einfacher Fall sind die Celsius- und Fahrenheitskalen für die Temperatur, die durch eine lineare Transformation aufeinander bezogen sind: $F = aC + b$ mit $a = 1.8$, $b = 32$. Der Unterschied von einem Grad Celsius entspricht einem Unterschied von 1.8 Grad F. Differenziert man, so erhält man $dF/dC = a$, d.h. $dF = a \cdot dC$. d.h. der Faktor a bezieht auch die Differentiale aufeinander. Da dx und dy infinitesimale *Differenzen* sind, verschwindet die additive Konstante b .

Abbildung 30: Skalen und Fläche unter Funktionen



Hälfte des Intervalls, den die Funktion $.5 \sin(u)$ überdeckt, und damit die Flächen unter den Funktionen – also die Integrale über die entsprechenden Intervalle – gleich groß sind, müssen die Funktionswerte von $\sin(u)$ in diesem Fall mit dem Faktor $1/2$ multipliziert werden. Eine intuitive Plausibilitätsbetrachtung mag hier hilfreich sein: Geht man von der Riemannschen Definition des Integrals als Grenzwert einer Summe von Rechtecken aus, so gilt speziell für eine Zerlegung des Integrationsintervalls $[a, b]$ in gleichgroße Intervalle $\Delta x = (b - a)/n$

$$\int f(x)dx = \lim_{\Delta x_i \rightarrow 0} \lim_{n \rightarrow \infty} \sum_{i=1}^n f(t_i)\Delta x, \quad t_i \in [x_i - x_{i-1}] \quad (5.142)$$

d.h. für $\Delta x \rightarrow dx$ für $n \rightarrow \infty$. Gilt nun $f(t_i)\Delta x = f(u_i)\varphi'\Delta x$, so erhält man im Limit für die Rechtecke $du = \varphi'dx$, und φ' kann als Dehnung oder Verkürzung von Δx oder als Streckung oder Stauchung von $f(t_i)$ interpretiert werden, so dass $f(u)\varphi'du = f(x)dx$ gilt. So haben die Rechtecke von $\sin(u)$ nur die Breite $du = dx/2$, haben also nur den halben Flächeninhalt. Da man den Faktor $1/2$ auch auf die Höhe der Rechtecke beziehen kann, kann man ebensogut sagen, dass die Rechtecke für $\sin(u)$ nur die halbe Höhe der Rechtecke für die Funktion $\sin(2x)$ haben. \square

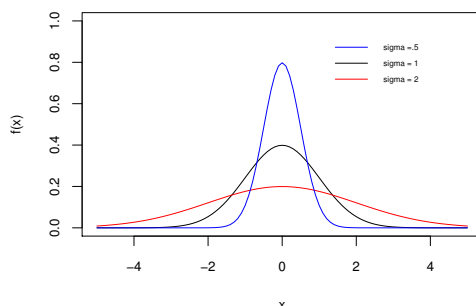
Beispiel 5.8 Es sei Z eine $N(0, 1)$ -verteilte zufällige Veränderliche, d.h. Z sei standardnormalverteilt, so dass $\mathbb{E}(Z) = 0$ und $Var(Z) = 1$. Es sei nun $X = \varphi(Z) = \sigma Z + \mu$. Die Verteilung von X wird sich also über die Verteilung von Z ergeben, denn $\varphi^{-1}(X) = (X - \mu)/\sigma = Z$. Dementsprechend ergibt sich der folgende Ansatz

$$P(X \leq x) = P(\sigma Z + \mu \leq x) = P(Z \leq \frac{x - \mu}{\sigma}),$$

so dass die Verteilungsfunktion F von X auf G , die Verteilungsfunktion von Z bezogen wird. Für die Dichten folgt

$$f(x) = \frac{dF(x)}{dx} = \frac{dG((x - \mu)/\sigma)}{dx} = f_Z\left(\frac{x - \mu}{\sigma}\right) \frac{1}{\sigma},$$

Abbildung 31: Skalen und Fläche: Gauss-Funktionen für verschiedene σ - (sigma)Werte unter Funktionen



wobei der Faktor $1/\sigma$ durch Anwendung der Kettenregel entsteht. Hier wird also der Verlauf von f um den konstanten Faktor $1/\sigma$ modifiziert: ist $\sigma > 1$, wird der Verlauf flacher, dafür wird der Bereich, in dem $f > \varepsilon$ für ein $\varepsilon > 0$ ist, größer; für $\sigma < 1$ gilt die Umkehrung. Wegen $X = \sigma Z + \mu$ gilt $\Delta Z = \sigma \Delta X$, d.h. eine Veränderung von Z um ΔZ bedeutet eine σ -fache Veränderung Δx von X . Setzt man $(x - \mu)/\sigma$ für z in die Dichtefunktion $f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}z^2)$ für Z ein und multipliziert, mit $1/\sigma$, so erhält man die Dichtefunktion für X :

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right). \quad (5.143)$$

Allgemein gelte nun $X = \varphi(Z)$, wobei φ^{-1} für alle X existiere und von Null verschieden sei. Dann ergibt sich

$$P(X \leq x) = P(\varphi(Z) \leq x) = P(Z \leq \varphi^{-1}(x)),$$

so dass

$$f(x) = f_Z(\varphi^{-1}(z)) \frac{d\varphi^{-1}(x)}{dx} = f_Z(\psi(x)) \frac{d\psi(x)}{dx} \quad (5.144)$$

f_Z ist hier wieder in der in (5.138) angegebenen allgemeinen Form. Während eine lineare Transformation von Z wieder auf die Gaußsche Dichte (5.143) führt, muss dies bei einer nichtlinearen Transformation φ nicht mehr der Fall sein. \square

Der mehrdimensionale Fall: Es seien $\mathbf{x} = (x_1, \dots, x_m)'$ und $\mathbf{y} = (y_1, \dots, y_n)'$ zufällige Vektoren, d.h. die x_i und y_i seien zufällige Veränderliche, und es gelte $\mathbf{y} = \mathbf{A}\mathbf{x}$, \mathbf{A} eine (n, n) -Matrix mit vollem Rang, so dass die Inverse \mathbf{A}^{-1} existiert. Gesucht ist die Verteilung von \mathbf{y} , wenn die von \mathbf{x} gegeben ist. Der Übersichtlichkeit halber wird das vorgehen am Fall $n = 2$ illustriert.

Gegeben sei ein Doppelintegral einer Funktion mit zwei unabhängigen Veränderlichen; eine Transformation der beiden Variablen kann oft die Aufgabe, das Integral zu bestimmen, erleichtern. Dazu werden die Funktionen

$$x = \varphi(u, v), \quad y = \psi(u, v) \quad (5.145)$$

eingeführt. Die Jacobi-Matrix ist

$$J = \frac{\partial(x, y)}{\partial(u, v)} = \begin{pmatrix} \frac{\partial\varphi}{\partial u} & \frac{\partial\varphi}{\partial v} \\ \frac{\partial\psi}{\partial u} & \frac{\partial\psi}{\partial v} \end{pmatrix} = \begin{pmatrix} \varphi_u & \varphi_v \\ \psi_u & \psi_v \end{pmatrix}. \quad (5.146)$$

Dann gilt der

Satz 5.9 Transformationssatz

$$\iint_G f(x, y) dx dy = \iint_{G^*} f(\varphi(u, v), \psi(u, v)) |\det(J_{\varphi\psi})| du dv \quad (5.147)$$

Beweis: Statt eines ausführlichen Beweises wird hier eine eher intuitive Betrachtung vorgestellt (Courant II, p. 231), einen vollständigen Beweis findet man zB in Courant II, pp. 222 – 225. Man denke sich das Gebiet \mathcal{G}^* in Rechtecke aufgeteilt: Geraden u_i parallel zur v -Achse und Geraden v_j parallel zur u -Achse. Die Differenz zwischen u_i und u_{i+1} betrage $u_{i+1} - u_i = \Delta u$, die zwischen v_{j+1} und v_j sei Δv für alle i und j . Das i -te Rechteck G_i hat dann die Fläche $\Delta u \Delta v$. in der xy -Ebene entspricht g_i ein Teilgebiet G_i^* , definiert durch die Abbildungen $\varphi(u_i, v_j)$, $\psi(u_i, v_j)$ etc. Im allgemeinen Fall sind die Gebiete G_i^* nicht mehr notwendig Rechtecke, denn die Verbindungslinien zwischen den Eckpunkten können nun krummlinig sein. Sind φ und ψ allerdings linear, so sind die G_i^* Parallelogramme. Man betrachte nun die Determinante

$$D := \begin{vmatrix} \varphi(u_i + h, v_j) - \varphi(u_i, v_j) & \varphi(u_i, v_j + k) - \varphi(u_i, v_j) \\ \psi(u_i + h, v_j) - \psi(u_i, v_j) & \psi(u_i, v_j + k) - \psi(u_i, v_j) \end{vmatrix}$$

Für hinreichend kleine Werte von h und k kann man dafür

$$\begin{vmatrix} \varphi_u(u_i, v_j) & \varphi_v(u_i, v_j) \\ \psi_u(u_i, v_j) & \psi_v(u_i, v_j) \end{vmatrix} hk \approx hkD \quad (5.148)$$

schreiben; hkD approximiert die Fläche von G_i^* . $f(\varphi(u_i, v_j), \psi(u_i, v_j))Dkh$ definiert dann ein Volumenelement, und die Summe über alle diese Elemente approximiert dann das Integral und man hat

$$\lim_{h \rightarrow 0, k \rightarrow 0} \sum_{i, j} f(\varphi(u_i, v_j), \psi(u_i, v_j)) Dkh = \iint_{\mathcal{G}^*} f(\varphi(u, v), \psi(u, v)) D du dv.$$

□

Um das folgende Beispiel vorzubereiten, wird das folgende Lemma bewiesen:

Lemma: A habe vollen Rang. Dann gilt

$$(A^{-1})'A^{-1} = (AA')^{-1}. \quad (5.149)$$

Beweis: Es ist⁸³ $(AA')^{-1} = (A')^{-1}A^{-1}$. Multiplikation von links mit A' liefert $A'(AA')^{-1} = A^{-1}$, und die Transponierung der Gleichung führt auf $((AA')^{-1})'A = (A^{-1})'$. Multiplikation von rechts mit A^{-1} liefert schließlich $(AA')^{-1} = (A^{-1})'A^{-1}$, und das war zu zeigen. \square

Beispiel 5.9 Die multivariate Normalverteilung Es sei $\mathbf{z} = (z_1, \dots, z_n)'$ ein zufälliger Vektor, dessen Komponenten stochastisch unabhängig $N(0, 1)$ -verteilt seien, d.h. es soll

$$f(\mathbf{z}) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\mathbf{z}'\mathbf{z}\right). \quad (5.150)$$

gelten. Weiter sei $\mathbf{x} = A\mathbf{z} + \boldsymbol{\mu}$, d.h. \mathbf{x} und $\boldsymbol{\mu}$ seien ebenfalls n -dimensionale Vektoren und A sei eine (n, n) -Matrix mit vollem Rang. Dann ist die Verteilung von \mathbf{x} durch

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right). \quad (5.151)$$

gegeben.

Beweis: Es ist $\mathbf{x} - \boldsymbol{\mu} = A\mathbf{z}$, so dass

$$(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})' = A\mathbf{z}\mathbf{z}'A'.$$

Dann ist $\mathbb{E}[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'] = A\mathbb{E}(\mathbf{z}\mathbf{z}')A'$, und

$$\mathbb{E}(\mathbf{z}\mathbf{z}') = (\mathbb{E}(z_i z_j)), \quad \mathbb{E}(z_i z_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases},$$

wegen der vorausgesetzten Unabhängigkeit der Komponenten von \mathbf{z} , d.h. $\mathbb{E}(\mathbf{z}\mathbf{z}') = I$ die Einheitsmatrix, so dass

$$\Sigma = \mathbb{E}[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'] = AA', \quad (5.152)$$

d.h. AA' ist gleich der Matrix der Varianzen und Kovarianzen der Komponenten von \mathbf{x} . Da A vollen Rang hat, existiert die zu A inverse Matrix A^{-1} , so dass

$$\mathbf{z} = A^{-1}(\mathbf{x} - \boldsymbol{\mu})$$

folgt, und nach (5.150) gilt

$$f(\mathbf{z}'\mathbf{z}) = f((\mathbf{x} - \boldsymbol{\mu})'(A^{-1})'A^{-1}(\mathbf{x} - \boldsymbol{\mu})|A^{-1}|)$$

denn $d\mathbf{x}/d\mathbf{z} = A^{-1}$, und wegen (5.149) und nach dem Produktsatz für Determinanten gilt mit $\Sigma = AA'$

$$|\Sigma| = |AA'| = |A||A'| = |A|^2$$

so dass $|A| = |\Sigma|^{1/2}$, also erhält man für die Dichte von \mathbf{x} die Dichte (5.151). \square

Kommentar: $|\Sigma|^{1/2}$ ist gleich der Funktionaldeterminante, die dem φ' im 1-dimensionalen Fall entspricht. \square

⁸³Wegen $(AB)^{-1} = B^{-1}A^{-1}$

Literatur

- [1] Barrantes Campos, H.: Elementos de álgebra lineal. *Editorio Unniversidad a Distancia*, San José, 2012
- [2] Cadima, J. Jolliffe, I. (2009) On relationships between uncentred and column-centred Principal Component Analysis. *PAk. J. Statist.*, 25(4), 473-503
- [3] Calcaterra, C. (2008) Linear combinations of Gaussians with a single variance are dense in L^2 . *Proceeding of the World Congress on Engineering*, Vol II, WCE 2008
- [4] Cangelosi, R., Goriely, A. (2007) Component retention in principal component analysis with application to cDNA microarray data, *Biology Direct*, 2(2), doi:10.1186/1745-6150-2-2
- [5] Cattell, R.B. The data box: its ordering of total resources in terms of possible relational systems. In: Cattell, R.B. (ed): *Handbook of multivariate experimental psychology*. Chicago 1966
- [6] Courant, R.: *Vorlesungen über Differential- und Integralrechnung, Erster Band: Funktionen mehrerer Veränderlicher*. Springer Verlag Berlin 1961
- [7] Courant, R.: *Vorlesungen über Differential- und Integralrechnung, Zweiter Band: Funktionen einer Veränderlichen*. Springer Verlag Berlin 1963
- [8] Courant, R. (1929) Über die Eigenwerte bei den Differentialgleichungen der mathematischen Physik. *Mathematische Zeitschrift* Band 7, Nr. 1–4, 1–57.
- [9] Dorier, J.L. (1995) A general outline of the genesis of vector space theory. *Historia Mathematica* 22, 227–261
- [10] Eckart, C., Young, G. (1936), The approximation of one matrix by another of lower rank. *Psychometrika*, 1 (3): 211–8. doi:10.1007/BF02288367.
- [11] Fischer, E. (1905) Über quadratische Formen mit reellen Koeffizienten. *Monatshefte für Mathematik und Physik*, Band 16, 234–249.
- [12] Fischer, G.: *Lineare Algebra*. Braunschweig Wiesbaden 1997
- [13] Forster, G.: *Analysis 2 – Differentialrechnung im \mathbb{R}^n , gewöhnliche Differentialgleichungen*. Springer- Spektrum, Wiesbaden 2017
- [14] Forster, G.: *Analysis 3 – Maß- und Integrationstheorie, Integralsätze im \mathbb{R}^n und Anwendungen*. Springer- Spektrum, Wiesbaden 2017
- [15] Gabor, D. (1948) A new microscopic principle. *Nature*, 161, 777–778
- [16] Gabriel, K.R. (1971) The biplot display of matrices with application to principal component analysis. *Biometrika* 58 (3), 453 – 467

- [17] Golub, G.H., van Loan, C. F.: Matrix computations. Baltimore 2013
- [18] Gumbel, E.J. (1961) Bivariate Logistic Distributions, *Journal of the American Statistical Association*, 56, No. 294. (Jun., 1961), 335-349
- [19] Goshtaby, A., O'Neill, W. (1994) Curve fitting by a sum of Gaussians. *CV-GIP: Graphical Models and Image Processing*, 56(4). 281–288
- [20] Graybill, F.: Matrices with Applications in Statistics. Belmont 1969
- [21] Hoaglin, D,C, Welsch, R.E. (1978) The hat matrix in Regression and ANOVA. *The American Statistician*, 32(1), 17 – 22
- [22] Honeine, P. (2014) An eigenanalysis of data centering in machine learning. *Statistics Machine Learning (stat.ML)* arXiv:1407.2904fl, 10 Jul 2014.
- [23] 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)
- [24] Karhunen, K. (1947), Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys.*, 37, 1–79.
- [25] Koecher, M.: Lineare Algebra und analytische Geometrie. Springer-Verlag, Berlin, Heidelberg, New York 1997
- [26] Lange, M. (2009) A Tale of two Vectors. *Dialectica*, 63(4), 397–431
- [27] Loève, M. (1978), Probability theory. Vol. II, 4th ed. Graduate Texts in Mathematics, 46. Springer-Verlag.
- [28] Lorenz, F.: Lineare Algebra I, II. Mannheim, 1988
- [29] Mardia, K.V., Kent, J. T., Bibby, J.M.: Multivariate Analysis. Academic Press, London, New York, Toronto 1979
- [30] Papoulis, A.: Probability, random variables, and stochastic processes. Tokyo 1965
- [31] 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.
- [32] Seber, G.A.F.: Linear regression analysis. John Wiley & Sons, New York, London, Sydney Toronto 1977
- [33] Shaw-Taylor, J. Christianini, N.: Kernel Methods for Pattern Analysis. Cambridge University Press, Cambridge 2004
- [34] Stewart, G. W.: Introduction to Matrix Computations. Academic Press, New York 1973

- [35] Stewart, G. W. (1993) On the early history of the singular value decomposition. *SIAM Review*, 35 (4), 551 – 566
- [36] Wong, E.: Stochastic processes in information and dynamical systems. New York 1971

Index

- \mathbf{c}_x Vektor der Richtungskosinus, 22
- Produkt
 - dyadisches und Kovarianz, 104
- Abbildung, 13
 - bijektiv, 183
 - Bild von, 183
 - homomorphe, 184
 - identische, 183
 - injektiv, 183
 - inverse, 183
 - isomorphe, 184
 - Kern einer, 183
 - lineare, 184
 - Rang der Abbildung, 185
 - surjektiv, 183
- abhängig
 - linear, 27
- absolute Homogenität, 34
- achsenparallel, s.a. orientiert, 87
- Adjunkte
 - einer quadratischen Matrix, 194
- algebraische Struktur, 34
- Algorithmus, Gauß-, 164
- assoziativ, 65
- Assoziativität, 65
- Ausreißer, 148
- Austauschsatz von Steinitz, 43
- Autokorrelation, 152
- Automorphismus, 184

- basic structure, 101
- Basis
 - eines Vektorraums, 39
 - geordnete, 201
 - gleichorientierte, 202
 - kanonische, 40
 - orthonormale, 47
- Basisentwicklung eines Vektors
 - orthonormale, 49
- Basisergänzungssatz, 45
- Basisfunktion, 150

- Basisfunktionen
 - orthogonal, 150
- Basiswechsel, 47
- Beweis
 - Existenz der Basis eines Vektorraums, 46
- Biplot, 71, 135

- charakteristische Gleichung, 196
- charakteristisches Polynom, 196
- Cholesky-Zerlegung, 169
- Cramersche Regel, 58, 119

- Definitheit, 34
- Determinante, 57, 77
- diagonalisierbar, 110
- Diagonalisierung, 84
- Diagonalmatrix, 61
- Differential, 203
- differenzierbar, total, 202
- Dimension (s.a. Rang), 54
- Dimensionalitätsatz, 52
- Diskriminanzkoeffizient, 140
- Distanz
 - Dreiecksungleichung, 157
 - euklidische, 35
 - Minkowski, 36
 - Reflexivität, 157
- Distributivität
 - Skalarprodukt, 16
- Dreiecksmatrix, 164
- Dreiecksungleichung, 34

- Ebene
 - k -dimensionale, 46
 - in einem Vektorraum, 37, 46
- Eckart & Young
 - Satz von, 128, 138
- Eigenraum, 112
- Eigenstruktur
 - einer $(n \times n)$ -Matrix, 84
- Eigenvektor
 - Links-, 108

Rechts-, 108
 Eigenwert, 84
 komplexer, 111
 Nullstellen eines Polynoms, 110
 Rang von Kreuzproduktmatrizen, 95
 Varianz der Projektionen auf latente Achsen, 137
 Eindeutigkeit
 von Linearkombinationen, 40
 Einflußmatrix, 148
 Einheitsvektor
 kanonischer, 9
 Einsvektor, 9
 Elementarmatrix, 163
 Eliminationsverfahren, gaußsches, 164
 Ellipsoid, 86
 Punktekonfiguration, 94
 Endomorphismus, 184, 196
 Erwartungswert, 103
 Erzeugendensystem, 38

 Faktorwerte, 133
 Faser, 183
 Funktion
 homogene, 12
 lineare, 12
 Funktionalmatrix, 203
 Funktionenraum, 149

 Gaußverteilung
 n -dimensionale, 104
 Gebietsintegral, 199
 Gerade
 Ortsvektor, 36
 Parametergleichung, 36
 Richtungsvektor, 36
 Gesamtvarianz
 Daten, 97
 Gesamtvarianz der Daten, 137
 Gleichung, charakteristische, 107
 Gleichungssystem
 homogen/inhomogen, 116
 homogenes, 59
 inhomogenes, 59
 lineares, 56

 Gradientenvektor, 177
 Gram-Matrix, 86
 Gram-Schmidt-Verfahren, 51
 Grundformel einf. einer neuen Variablen, 204
 Grundstruktur einer Matrix, 101

 Hauptachsentransformation, 80, 83, 93
 Hauptraum, 112
 Hauptvektor, 112
 Hebelwirkung, 148
 Homogenität
 Skalarprodukt, 15
 Homomorphismus, 184
 Hyperebene, 46

 idempotent, 120, 146
 Identität, 183
 influence matrix, 148
 Inverse, 76, 175
 generalisierte, 122
 Moore-Penrose, 122
 Pseudo, 122
 Isomorphismus, 184

 Jacobi-Matrix, 203

 kanonische Darstellung, 68
 Karhunen-Loève-Analyse, 151
 Kern, 71, 116
 einer Matrix, 112
 Kettenregel, 203
 Kodimension eines Teilraums, 54
 Kofaktor, 194
 Kofaktoren, 194
 kollinear, 30, 57
 kommutativ, 65
 Komplement
 orthogonales, 52
 Komponenten in Bezug auf eine Basis, 40
 Koordinaten
 als Projektionen, 70
 Koordinatenform
 Ebene, 38
 Gerade, 36

Korrelation
 kanonische, 113
 Produkt-Moment-, 21
 Stichproben-, 21
 Kosinus-Ähnlichkeit, 22
 Kosinussatz, 17
 Kovarianz
 Stichproben-, 21
 Kovarianzen
 von Linearkombinationen unabhängiger Variablen, 104
 Kovarianzmatrix, 73
 Kreuzproduktmatrix, 73
 Kreuzproduktmatrizen
 Rang von, 74
 Kronecker-Delta, 16, 150
 Körper, 33

 Ladung, 49
 Ladungen, 133
 Lagrange
 -Faktor, 177
 -Funktion, 177
 -Multiplikator, 177
 sche Multiplikatorenregel, 177
 latente Dimension, 132
 latente Variable, 132
 LDL-Zerlegung, 170
 Leibniz-Formel, 190
 leverage, 148
 linear abhängig, 27
 lineare Hülle, 38
 lineare Hülle (Gleich'syst.), 116
 Linearkombination, 12
 als Abbildung, 12
 lineare Gleichungssysteme, 55
 von Funktionen, 149
 Linkseigenvektoren, 101
 Linksinverse, 75
 Längenskalierung, 65

 Matrix
 Wurzel einer symmetrischen, 99
 assoziierte, 111
 charakteristische Gleichung, 106
 Diagonal-, 61
 Dreiecks, 164
 elliptisch, 86
 gestürzte, transponierte, 61
 Gram-, 86
 hermitesch, 111
 hyperbolisch, 86
 imaginär, 111
 inverse, 76
 inverse (2-dim), 76
 konjugierte, 111
 negativ semidefinit, 86
 positiv semidefinit, 86
 Präzisions, 77
 schief-symmetrisch, 111
 spaltenstandardisiert, 73
 spaltenzentriert, 73
 symmetrische, 61
 Transformations-, 92
 zufällige, 103
 ähnliche, 195
 Matrixnorm, 125
 Matrizen
 ähnliche, 110
 Mercers Theorem, 153, 155
 Metrik, 157
 City-Block, 158
 euklidisch, 157
 euklidische, 35, 157
 Manhattan, 158
 Minkowski, 36, 158
 Mini-max-Theorem
 von Courant-Fischer, 88
 Minor, 194
 Mittelwertfunktion, 152

 Norm
 p -, 126
 -Matrix, 125
 allgemein, 34
 definierende Eigenschaften, 34
 euklidische, 125
 Frobenius-, 125
 Hilbert-Schmidt-, 126
 induzierte, 126

- Maximum, 125
- Schur-, 126
- Vektor-, 16
- Normalenvektor, 46, 51
 - Ebene, 37
 - Gerade, 36
- Normalverteilung
 - multivariate, 209
- Nullität, 71
- Nullmatrix, 67
- Nullraum, 71
- Nullvektor, 9

- Operator
 - linearer, 103
- orientiert
 - negativ, 202
 - positiv, 202
- Orientierung
 - eines Vektorraums, 200
- Orientierungsvektor, 22
- orthogonal, 19
- orthogonales Komplement, 52
- orthonormal, 19
- Orthonormalbasis (ONB), 47
- orthonormale Basisentwicklung, 49
- Ortsvektor, 36

- Parameterform
 - Ebene, 37
 - Gerade, 171
- Permutation, 188
- Pfad, 152
- Polynom, charakteristisches, 107
- positiv semidefinit, 97
- Principal Component Analysis, 83
- Produkt
 - dyadisches, 14, 128
 - dyadisches und SVD, 127
 - inneres, 14
 - tensorielles, 14
- Produktsatz, 190
- Projektion
 - als Koordinate, 70
 - Vektor auf Gerade, 23
 - Vektor auf Vektor, 24
- Projektionsmatrix, 145
- Projektionsoperator
 - Faktorenscores, 135
 - Ladungen, 135
- Präzisionsmatrix, 77
- Punkt: Schreibweise, 156
- Punktekonfigurationen K_x, K_y , 93
- Pythagoras
 - Satz von, 17

- quadratische Form, 86

- Rang
 - Diagonalmatrix, 74
 - dyadisches Produkt, 71
 - einer Matrix, 67
 - einer Vektormenge, 54
 - Kreuzproduktmatrizen, 74
 - voller, 54
- Rangsatz, 67
- Rayleigh-Quotient, 88
 - generalisierter, 113
- Rechtseigenvektoren, 101
- Rechtsinverse, 75
- Richtungskosinus, 22
- Richtungsvektor, 36
- Richtungsvektoren
 - Ebene, 37
- Rotation, 47
- Rotationsmatrix, 80, 166

- Sarrussche Regel, 192
- Satz
 - des Thales, 20
 - Eckart-Young, 128
 - von Courant-Fischer, 88, 178
 - von Schmidt-Mirsky, 128, 130
 - von Steinitz, 42
- Schur-Norm, 126
- Schwarzsche Ungleichung, 25
- Singularwert, 101
- Singularwertzerlegung, 100
- Skalarprodukt, 14
 - als Koordinate, 24
 - alternative Definition, 15

- Skalierung
 - eines Vektors, 17, 65
- Spaltenrang, 66
- Spaltenraum, 70, 185
- Spaltenvektoren, 61
- span, 38
- Spur, 96, 126
- Standardmatrix, 163
- Steinitzches Lemma, 42
- stetig (im quadrat. Mittel), 152
- stochastischer Prozess
 - zentrierter, 152
- Stütz-/Ortsvektor
 - Ebene, 37
- Subadditivität, 34
- SVD, 100
- Teilraum
 - r -dimensionaler, 45
 - Dimension (Rang) des, 54
 - invarianter, 113
 - orthogonales Komplement, 52
 - orthonormale Basis, 51
- Teilraum, trivialer, 33
- Teilräume
 - Dimensionsformel, 54
 - Durchschnitt, 53
 - Kombination von, 52
 - orthogonale, 51
 - orthogonale Summe, 52
 - Summe zweier, 54
 - Vereinigung, 53
- Teilvektorraum, 33, 39
- Trajektorie, 152
- Transformation
 - Alias, 82
 - Alibi, 82
 - eines Spaltenvektors, 64
 - eines Vektors, 13
- Transformationsatz, 208
- Transposition, 188
- triviale Lösung, 27
- Umformungen, elementare, 60, 165
- Unabhängigkeit
 - lineare und Korrelation, 28
- unabhängig
 - linear, 27
- unendlich-dimensional, 150
- Unterraum, 54
- Untervektorraum, 33
- Varianz
 - Parameterschätzungen, 99
 - Stichproben-, 21
- Vektor
 - n -dimensionaler, 28
 - transformation, 13
 - charakteristischer, 84
 - Eins-, 9
 - gerichtete Größe, 8
 - Komponenten, 9
 - Länge, 9
 - Norm, 17
 - normiert, 17
 - Null-, 9
 - zentrierter, 21
 - zufälliger, 13, 103
- Vektornorm, 16, 124, 126
- Vektorprodukt, 201
- Vektorraum, 33
 - n -dimensionaler, 45
 - euklidischer, 35
 - normierter, 125
- Zeilenraum, 70
- Zeilenvektoren, 61
- Zentrierungsmatrix, 119
- Zufallsmatrix, 103
- Zufallsvektor, 103
- überabzählbar, 39