.

Einführung in die Vektor- und Matrixrechnung für die Multivariate Statistik[1]

U. Mortensen

[1]Letzte Korrektur: 28. 04. 2026

Inhaltsverzeichnis

1 Vektoren und Vektorräume **5**

1.1 Zur Bedeutung und Geschichte der Vektor- und Matrixrechnung . 5

1.2 Vektoren und ihre Verknüpfungen 8

1.2.1 Endlichdimensionale Vektoren 8

1.2.2 Linearkombinationen von Vektoren 12

1.2.3 Produkte von Vektoren 16

1.2.4 Mittelwerte, Varianzen, und Kovarianzen 24

1.2.5 Länge, Orientierung und Kosinus-Ähnlichkeit von Vektoren 25

1.2.6 Vektorprojektionen 28

1.2.7 Die Cauchy-Schwarzsche Ungleichung 30

1.2.8 Lineare Abhängigkeit und Unabhängigkeit von Vektoren . . 31

1.3 Vektorräume 36

1.3.1 Definition eines Vektorraumes 36

1.3.2 Erzeugendensysteme und Basen von Vektorräumen 44

2 Matrizen **60**

2.1 Definitionen 60

2.2 Operationen mit Matrizen 61

2.2.1 Addition und Multiplikation mit einem Skalar 61

2.2.2 Multiplikation einer Matrix mit einem Vektor 62

2.2.3 Multiplikation einer Matrix mit einer Matrix 63

2.3 Der Rang einer Matrix 66

2.4 Die Inverse einer Matrix 75

2.5 Die Beziehungen zwischen verschiedenen Basen 79

3 Bestimmung einer Basis: die Hauptachsentransformation **80**

3.1 Rotationen 80

3.2 Die Rotation einer Punktekonfiguration 83

3.2.1 Die Bestimmung einer Basis durch Rotation 83

3.2.2 Eigenstrukturen und Singularwertzerlegung 86

3.2.3 Quadratische Formen und Ellipsoide 90

3.2.4 Zur Geometrie der Eigenvektoren symmetrischer Matrizen . 95

3.2.5 Inverse und Wurzel einer symmetrischen Matrix 100

3.2.6 Die multivariate Gauß-Verteilung 101
3.3 Eigenvektoren und Eigenwerte nichtsymmetrischer Matrizen 104
3.3.1 Der allgemeine Fall 104
3.3.2 Mehrfache Eigenwerte 109
3.3.3 Ähnliche Matrizen 110
3.3.4 Das generalisierte Eigenvektorproblem 113
3.4 Lineare Gleichungssysteme 115
3.4.1 Allgemeine Aussagen 115
3.4.2 Die Regularisierung von Matrizen 120
3.5 Weitere Befunde 123
3.5.1 Die Zentrierungsmatrix 123
3.5.2 Die Pseudoinverse einer Matrix 126
3.5.3 Vektor- und Matrixnormen 127
3.5.4 Die Approximation von Matrizen 130
3.6 Bestimmung latenter Variablen 135
3.6.1 Explorieren (Hauptkomponenten und SVD) 135
3.6.2 Diskriminieren und klassifizieren 141
3.7 Projektionen 146
3.7.1 Projektion auf eine Ebene im $\mathbb{R}^3$ 146
3.7.2 Der allgemeine Fall: Projektionsmatrizen 147
3.7.3 Beispiel: das Allgemeine Lineare Modell 149
3.7.4 Projektionen auf Hauptachsen 151

4 Nichtlineare latente Strukturen 151
4.1 Nichtlineare Funktionen und Kernfunktionen 151
4.2 Kernreproduzierende Hilbert-Räume (RKHS) 155

5 Funktionenräume und PCA 164
5.1 Einführung 164
5.2 Karhunen-Loève-Entwicklung und PCA 165

6 Anhang 170
6.1 Punkträume 170
6.2 Allgemeine Vektorräume 173
6.2.1 Definition eines Vektorraums 173

6.3 Skalen und Abbildungen 175
6.4 Zur geometrischen Definition des Skalarprodukts 177
6.5 Elementarmatrizen und elementare Operationen 178
6.5.1 Lineare Gleichungen und Gauß-Algorithmus 180
6.5.2 Herleitung einer Rotationsmatrix 182
6.5.3 Die Cholesky-Zerlegung 185
6.6 Zur Berechnung von Ellipsen für eine Punktekonfiguration 186
6.7 Zur Eindeutigkeit von Eigenvektoren 188
6.8 Alternative Herleitung der Singularwertzerlegung 190
6.9 Die Differentiation von Vektoren 192
6.9.1 Die allgemeine Differentiationsformel 192
6.9.2 Die Differentiation quadratischer Formen 193
6.9.3 Die Kleinste-Quadrate-Schätzung für das Lineare Modell . 194
6.9.4 Extrema unter Nebenbedingungen 196
6.10 Der Rayleigh-Koeffizient und der Satz von Courant-Fischer 198
6.11 Alternativer Beweis von Satz 3.17 200
6.12 Vektortransformationen und Abbildungen 202
6.13 Determinanten 207
6.13.1 Transformationen und Volumen 216
6.13.2 Das Vektorprodukt, Transformationen und Orientierung . . 220
6.13.3 Die Transformation zufälliger Veränderlicher 223

Literatur **232**

Index **234**

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

1 Vektoren und Vektorräume

Dieser Text wurde für WissenschaftlerInnen aus den Bereichen Soziologie, Pscyhologie, Biologie, Medizinetc, geschrieben, die insbesondere multivariate statistische Verfahren anwenden möchten und die sich mit den notwendigen Begriffen aus der linearen Algebra vertraut machen wollen. Der Text fokussiert deswegen insbeosndere auf das Ziel, "latente Variable" zu bestimmen und folgt daher nicht der in der Mathematik üblichen Darstellung der linearen Algebra.

1.1 Zur Bedeutung und Geschichte der Vektor- und Matrixrechnung

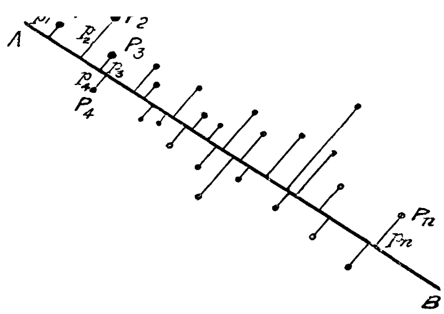
Gegeben seien Messungen von n Variablen an jeweils m "Fällen" (Objekte, Personen, ...). Eine der Standardaufgaben der mulivariaten Statistik ist, die Korrelationen zwischen den Variablen durch die Wirkung latenter, d.h. nicht direkt gemessener Variablen zu erklären, wobei die Anzahl der latenten Variablen so klein wie möglich sein sollte. Insbesondere wenn die Variablen in verschiedenen Maßeinheiten gemessen werden (z.B. Körperlänge und Körpergewicht) wird dabei von standardisierten Messwerten $z_{ii} = (x_{ij} - \bar{x}_j)/s_j$ ausgegangen, $\bar{x}_j$ das arithmetische Mittel der Messwerte x_{ij} für die j-te Variable und s_j die dazu korrespondierende Standardabweichung. Der übliche Ansatz ist[2]

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

Die f_{ik}, $1 \leq i \leq m$, $1 \leq k \leq r$, sind "Scores" (Faktorwerte) des i-ten Falls auf der k-ten latenten Dimension (Variablen), die Koeffizienten a_{kj} sind die Koordinaten der j-ten gemessenen Variablen auf der k-ten latenten Variablen ("Dimension"), wobei angenommen wird, dass die latenten Variablen unabhängig voneinander sind. 'Unabhängig' wird dabei mit 'unkorreliert' gleichgesetzt. e_{ij} repräsentiert einen Messfehler, der gelegentlich auch weggelassen wird, wobei man von $r = \min(m, n)$ ausgeht und untersucht, ob man für einen Wert $\hat{r} < \min(m, n)$ eine gute Approximation $\hat{z}_{ij}$ für die z_{ij} erhält. Der Ansatz geht auf den Mathematiker Karl Pearson (1901) zurück, der eine Schätzung der Parameter durch eine Anwendung der Methode der Kleinsten Quadrate versuchte, – er betrachtete ein "System von Punkten $P_1, \ldots, P_n$ und ihre Projektion auf eine Gerade" (vergl Abb. 1), oder auf eine Ebene, wobei er Gebrauch von der "gewöhnlichen" Algebra machte. Die Details seines Ansatzes ähneln denen bei der Schätzung von

[2] In der leicht modifizierten Schreibweise von Bortz (1999), p. 501, einem klassischen Lehrbuch der Statistik für Anwender.

Abbildung 1: Karl Pearson's Ansatz (Pearson (1901))



Regressionsparametern und können hier übergangen werden, es genügt, festzustellen, das stets von einer Punktekonfiguration der Fälle ausgegangen wird, d.h. von der Repräsentation der Fälle durch Punkte in einem Koordinatensystem, dessen Achsen die gemessenen Variablen abbilden.

Dies trifft auch auf den allgemeineren Ansatz von Hotelling (1933) zu. Zeigt sich, dass die z_{ij} gut durch Werte $\hat{z}_{ij}$ approximiert werden können, die auf einer Schätzung $\hat{r} < \min(m, n)$ beruhen, die Rede ist dann von einer 'Datenreduktion' oder 'Datenkompression'. Insbesondere Hotellings Ansatz hat die Darstellung des Verfahrens für Anwender bis weit in die 90er Jahre geprägt (z.B. Bortz, p. 509 etc). Dieser Ansatz hat allerdings Nachteile:

1. Die algebraischen bzw. geometrischen Grundlagen des Ansatzes bleiben undeutlich.

2. Die Forderung nach Unabhängigkeit der "latenten Variablen" wird mit der Forderung nach Unkorreliertheit dieser Variablen gleichgesetzt, womit implizit die Annahme bestimmter Wahrscheinlichkeitsverteilungen der Variablen vorausgesetzt wird.Dieser Sachverhalt könnte der Grund für Hotellings Bemerkung sein, dass er die multivariate Normalverteilung der Daten annimmt[3]. Hotelling elaboriert diese Annahme aber nicht weiter. Der im Rahmen der Linearen Algebra definierte Begriff der linearen Unabhängigkeit von Vektoren (der hier in Abschnitt 1.2.8 eingeführt wird) ist allgemeiner und ist rein algebraisch definiert.

3. Die Darstellung der Berechnung der Parameter steht i.A. im Zentrum der Darstellung, ohne dass der Modellcharakter des Ansatzes – die Beziehungen zwischen den Variablen werden als linear angenommen – geklärt wird.

Wird von der Vektor- und Matrixrechnung gesprochen, so ist die Lineare Algebra gemeint. Gegenstand der Linearen Algebra sind lineare Abbildungen in

[3]Bei dieser Verteilung bedeuten Korrelationen gleich Null zwischen den Variablen auchstochastische Unabhängigket der Variablen, weas bei anderen Verteilungen i.a. nicht der Fall ist.

Vektorräumen bzw. von einem Vektorraum in einen anderen, wobei unter einer linearen Abbildung z.B. eine Funktion f verstanden wird, für die

$$f(ax + by) = af(x) + bf(y) \tag{1.2}$$

gilt; dabei sind x und y Variablen, und a, und b sind Konstante. Die Elemente von Vektorräumen sind Vektoren, wobei Vektoren z.B. die in den folgenden Abschnitten ausführlich behandelten n-Tupel $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ reeller oder komplexer Zahlen sein können, wobei es auf die Anordnung der Zahlen ankommt. Ein Beispiel sind die Messungen x_i, $1 \leq i \leq n$, einer Variablen bei n Personen, oder die Messungen von n Variablen bei einer Person, und es gilt $n < \infty$; n heißt die Dimension des Vektors. Vektorräume sind Mengen V von Vektoren, für die die Beziehung (1.2) gilt (ausführliche Definition auf Seite 174): repräsentieren zum Beispiel x und y Vektoren $\mathbf{x}$ und $\mathbf{y}$, und gilt für beliebige $\mathbf{x}, \mathbf{y} \in V$

$$\mathbf{z} = f(a\mathbf{x} + b\mathbf{y}) = a\mathbf{x} + b\mathbf{y} \in V, \quad a, b \in \mathbb{R} \tag{1.3}$$

($\mathbb{R}$ ist die Menge der reellen Zahlen), wobei $\mathbf{z}$ wieder ein Vektor ist, d.h. ist V abgeschlossen in Bezug auf (i) die Multiplikation mit einer Zahl (a bzw b) sowie (ii) der Addition $a\mathbf{x} + b\mathbf{y}$, d.h. ist die Summe $a\mathbf{x} + b\mathbf{y}$ wieder ein Element von V, so ist die Menge V ein Vektorraum. Es zeigt sich, dass die Elemente von Vektorräumen nicht nur die hier als Beispiel angeführten n-dimensionalen Vektoren sein können. So können Matrizen, die durch Nebeneinanderschreiben von m-dimensionalen Vektoren in Spaltenform oder durch Untereinanderschreiben in Zeilenform von n-dimensionalen Vektoren entstehen ebenfalls Elemente von Vektorräumen sein, oder über einem Bereich D definierte reelle, stetige Funktionen, wobei die Dimensionalität nicht mehr endlich zu sein braucht, oder die Elemente von V können Lösungen von linearen Gleichungssystemen sein. Der Begriff des Vektors und des Vektorraumes sind also sehr viel allgemeiner als es die Fokussierung auf endlichdimensionale Vektoren $\mathbf{x} = (x_1, \ldots, x_n)$ vermuten läßt. Endlichdimensionale Vektoren spielen in der Multivariaten Analyse eine zentrale Rolle, weshalb sie im Zentrum dieses Textes stehen. Ein Vorteil des Übergangs zur Vektor- und damit auch der Matrixrechung ist, dass man mit Vektoren und Matrizen rechnen kann. Wie im Folgenden deutlich werden wird erleichtert dieser Sachverhalt die Konzeption und Berechung u.a. von latenten Variablen wesentlich.

Wie bereits die Gleichungen (1.2) und (1.3) nahelegen wird davon ausgegangen, dass die Beziehungen z.B. zwischen Variablen linear im Sinne von (1.2) sind. Für empirisch gegebene Messungen von Variablen muß diese Annahme allerdings keinesfalls gelten. In neueren Verfahren können auch nichtlineare Relationen zwischen Variablen berücksichtigt werden. In Abschnitt 4 wird auf diese Möglichkeiten eingegangen, wobei die betrachteten Vektorräume dann Funktionenräume sind.

Der Beginn der Vektorrechnung wird oft mit der Publikation von H.G. Graßmanns[4] *Lineale Ausdehnungslehre*[5] (1844), einem Lehrbuch der Geometrie, und W.R. Hamiltons[6] *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 gleichwohl 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 Cramer[7] 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 repäsentierte sah man die allgemeine Nützlichkeit des Vektorbegriffs. Eine Kraft hat eine bestimmte Größe und wirkt in eine bestimmte Richtung, was einer der Gründe für die Charakerisierung 'gerichtete Größe' für Vektoren sein mag; in diesem Sinne verwendete der schottische Physiker James Clerk Maxwell (1831 – 1879) den Vektorbegriff auch in seiner Entwicklung der Elektrodynamik[8].

In der Statistik wurde relativ spät von den Möglichkeiten der Vektor- und Matrixrechnung Gebrauch gemacht, eines der ersten Lehrbücher, in dem ausgiebig von der Vektor- und Matrixrechnung Gebrauch gemacht wurde, war Andersons *An Introduction to Multivariate statistical Analysis* (1958). Searle (1999) gibt eine detaillierte Geschichte der Vektor- und Matrixrechnung in der Statistik.

1.2 Vektoren und ihre Verknüpfungen

1.2.1 Endlichdimensionale Vektoren

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

[4]Hermann Günther Graßmann (1809 – 1877), deutscher Mathematiker
[5]Nicht: Lineare
[6]William Rowan Hamilton (1805 - 1865), irischer Mathematiker
[7]Gabriel Cramer (1704 – 1752), Genfer Mathematiker
[8] *Treatise on Electricity and Magnetism* (1873)

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

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \tag{1.4}$$

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 2), Seite 10). Es wird auch $\boldsymbol{x} \in \mathbb{R}^n$ geschrieben.*

Dementsprechend sind

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

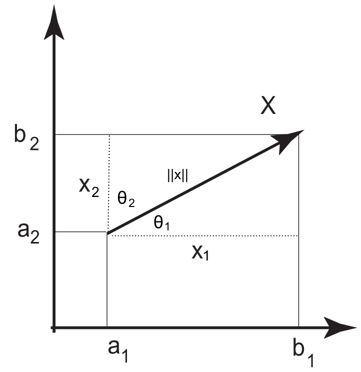
verschiedene 3-dimensionale Vektoren.

Anmerkungen:

1. 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 die Menge $(\mathbb{R} \times \mathbb{R}) \times \mathbb{R} = \mathbb{R} \times \mathbb{R} \times \mathbb{R} = \mathbb{R}^3$ möglicher Vektoren, etc. Allgemein ergibt sich für n Komponenten die Menge $\mathbb{R}^n$ möglicher Vektoren. $\mathbf{x} \in \mathbb{R}^n$ heißt dann, dass $\mathbf{x}$ ein Tupel aus dem*Cartesischen Produkt* $R \times R \times \cdots \times \mathbb{R} = \mathbb{R}^n$ ist.
2. Vektoren werden im Folgenden wie in (6.24) durch fette Buchstaben bezeichnet. Eine alternative Schreibweise ist $\vec{x}$, $\vec{y}$, etc, der Pfeil ber dem Buchstaben zeigt an, dass es sich um einen *vector* handelt, Sie wird hier nur als Ausnahme verwendet: für den *Nullvektor* $\vec{0} = (0, \dots, 0)$, dessen Komponenten alle gleich Null sind, und dem *Einsvektor* $\vec{1} = (1, \dots, 1)$, dessen Komponenten alle gleich 1 sind. Die Kennzeichnung durch fette Buchstgaben en); diese Bezeichung ist mittlerweile allgemein üblich. Die Fett-Schreibweise erleichtert Schreibweisen wie $\hat{\mathbf{x}}$, die Schätzungen eines Vektors $\mathbf{x}$ bezeichnen, oder $\tilde{\mathbf{x}}$, die Vektoren bezeichnen, die durch die die Zeilen einer Matrix definiert sind. Die Schreibweise $\vec{x}$ wird dann leicht unübersichtlich: $\hat{\vec{x}}$ oder $\vec{\hat{x}}$ und $\hat{\tilde{x}}$ bzw. $\tilde{\vec{x}}$ zeichnen sich durch sehr geringen graphischen Charme aus, insbesondere wenn transponierte Vektoren $\vec{\tilde{x}}'$ (s. unten) betrachtet werden.
3. Ein Vektor $\mathbf{x}$ wird wie in Gleichung (6.24) stets als Spalte oder Kolumne angeschrieben, es sei denn, er wird ausdrücklich als *transponiert* oder *gestürzt* gekennzeichnet; in diesem Fall wird er als Zeile geschrieben:

$$\mathbf{x}' = (x_1, x_2, \dots, x_n). \tag{1.5}$$

Abbildung 2: 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



Analog dazu kann man einen Spaltenvektor als transponierten Zeilenvektor schreiben:

$$\mathbf{x} = (\mathbf{x}')' = (x_1, x_2, \ldots, x_n)' \tag{1.6}$$

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

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

für einen Spaltenvektor, denn "stürzt" man eine Zeile, so wird sie zur Spalte.

□

Vektoren werden graphisch oft durch Pfeile repräsentiert, s. Abbildung 2. 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 $a_1, \ldots, a_n$, der Endpunkt habe die Koordinaten $b_1, \ldots, b_n$. Es sei nun

$$x_i = b_i - a_i, \quad i = 1, \ldots, n \tag{1.7}$$

die x_i sind die *Komponenten* des Vektors. Wie man der Abb. 2 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

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2} \tag{1.8}$$

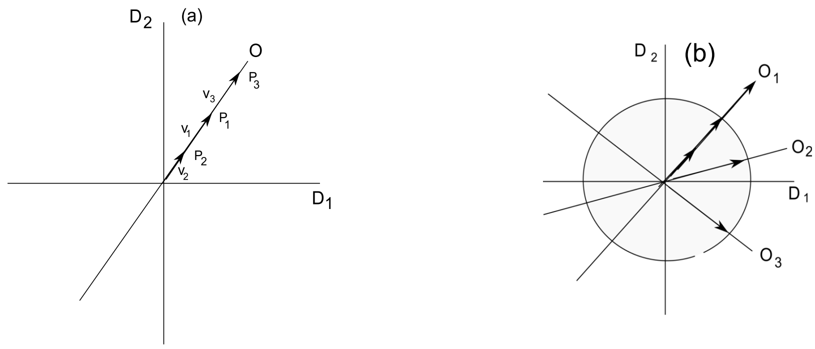
gegeben. Offenbar gelten die Beziehungen

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

so dass sich für die Komponenten die Beziehungen

$$x_1 = \|\mathbf{x}\|\cos\theta_1, \quad x_2 = \|\mathbf{x}\|\cos\theta_2 \tag{1.9}$$

Abbildung 3: Punkte P_i und Vektoren $\mathbf{v}_i$ mit verschiedenen Orientierungen und Grössen



ergeben. Demnach kann

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \|\mathbf{x}\| \cos\theta_1 \\ \|\mathbf{x}\| \cos\theta_2 \end{pmatrix} = \|\mathbf{x}\| \begin{pmatrix} \cos\theta_1 \\ \cos\theta_2 \end{pmatrix}; \tag{1.10}$$

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.

Anmerkung: Der Richtungsaspekt eines Vektors definiert eine Orientierung im jeweiligen Koordinatensystem. Dazu werde Abbildung 3 betrachtet. Das Ende des Pfeils v_1 ist der Punkt P_1, der den Fall 1 abbildet. Die Koordinaten des Falles P_1 seien x_1 und x_2. Die zwei Aspekte der Position von P_1 sind

(i) seine Distanz d_1 vom Nullpunkt, die nach dem Satz von Pythagoras durch $d_1 = \sqrt{x_1^2 + x_2^2}$ gegeben ist, und

(ii) das Verhältnis x_2/x_1, das das Verhältnis der Ausprägungen der Merkmale bzw. Variablen D_1 und D_2 beim Fall 1 beschreibt, – also das Mischungsverhältis dieser beiden Merkmale bei diesem Fall.

Legt man eine Gerade O (O für Orientierung), die einerseits durch den Nullpunkt des Koordinatensystems und andererseits durch den Punkt P_1 verläuft, so erhält man den geometrischen Ort aller Fälle, die durch dasselbe Verhältnis der Merkmale D_1 und D_2 charakterisiert sind; man kann von einem *Typus* von Fall reden, der durch dieses Verhältnis bestimmt wird. Dieser Typus wird durch die Orientierung der Geraden O, auf der die Punkte liegen, bestimmt. Die verschiedenen Punkte auf der Geraden unterscheiden sich durch die Ausprägung der Merkmale unter der Nebenbedingung des durch O definierten Verhältnisses der Merkmale.

$\square$

Spezielle Vektoren: Es seien

$$\vec{0} = (0, 0, \ldots, 0)' \tag{1.11}$$
$$\vec{1} = (1, 1, \ldots, 1)' \tag{1.12}$$
$$\mathbf{e}_j = (0, \ldots, 0, 1, 0, \ldots, 0)', \; j = 1, \ldots, n \tag{1.13}$$

$\vec{0}$ 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 18).

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

Da ein Vektor durch ein Zahlentupel $(x_1, \ldots, x_n)$ definiert ist, korrespondiert ein Vektor offenbar zu einem Punkt im n-dimensionalen Punktraum. Gleichwohl sind verschiedene Vorstellungen mit dem 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 $(x_1, \ldots, x_n)$ 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:

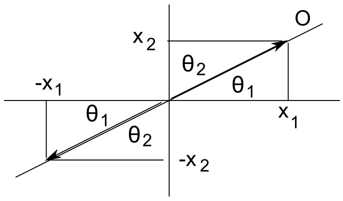
$$a\mathbf{x} = a \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_n \end{pmatrix}, \quad a \in \mathbb{R} \tag{1.14}$$

a heißt *Skalar*, weil a ein Wert der "Skala", d.h. der reellen Zahlen zwischen $-\infty$ und ∞ ist. Interpretiert man G als "Orientierung", d.h. als Mischungsverhältnis von Merkmalen (vergl. 1.2.1, Seite 11)

Ein wichtiger Spezialfall ist $a = -1$, mit dem ein negativer Vektor definiert werden kann:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad -\mathbf{x} = \begin{pmatrix} -x_1 \\ -x_2 \\ \vdots \\ -x_n \end{pmatrix}. \tag{1.15}$$

Abbildung 4: Vektoren $\mathbf{x}$ und $-\mathbf{x}$



Die Vektoren $\mathbf{x}$ und $-\mathbf{x}$ liegen auf derselben Geraden O, die ein Mischungsverhältnis von Merkmalen repräsentiert, und auf dieser Geraden repräsentieren sie verschiedene Richtungen (vergl. Abb. 4).

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

$$\mathbf{z} = \mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} \tag{1.16}$$

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:

$$\mathbf{u} = \mathbf{x} - \mathbf{y} = \begin{pmatrix} x_1 - y_1 \\ x_2 - y_2 \\ \vdots \\ x_n - y_n \end{pmatrix} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \tag{1.17}$$

Es ist

$$\mathbf{v} = -\mathbf{u} = -(\mathbf{x} - \mathbf{y}) = \mathbf{y} - \mathbf{x}$$

d.h $\mathbf{x} - \mathbf{y}$ und $\mathbf{y} - \mathbf{x}$ haben entgegengesetzte Orientierungen. vergl. Abb. 6.

Anmerkung: Es sei daran erinnert, dass Vektoren nur durch ihre Länge und Orientierung (bzw. Richtung), 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 Kräfteprallelogramm in der Physik, und der Summe für die Differenz zweier Vektoren definierte Vektor erscheint als Verbindung der Endpunkte der Vektoren, vergl. Abb. 6. Grundsätzlich können aber die Summen- und Differenzen irgendwo positioniert werden, alle Positionen sind im Prinzip gleichwertig. □

Linearkombinationen Summen und Differenzen von Vektoren können als Spe-

Abbildung 5: Addition von Vektoren, weitere Erläuterung s. Text

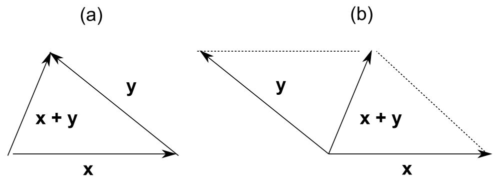
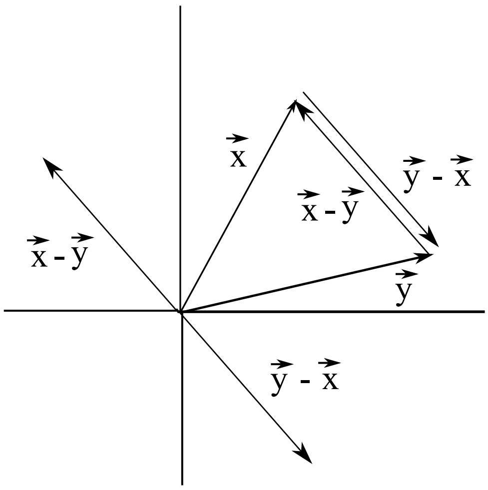


Abbildung 6: Subtraktion von Vektoren; die Differenzen $\mathbf{x} - \mathbf{y}$ und $\mathbf{y} - \mathbf{x}$ sind Vektoren mit entgegengesetzter Orientierung



zialfälle der allgemeinen Vektorgleichung

$$\mathbf{y} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2, \quad \mathbf{y}, \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n \tag{1.18}$$

aufgefasst werden. 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 (1.18) kann verallgemeinert werden:

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

$$\boldsymbol{y} = a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_p\boldsymbol{x}_p, \tag{1.19}$$

heißt Linearkombination *der $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$.*

Anmerkungen:

- *Lineare Funktionen*: Der Ausdruck 'Linearkombination' verweist darauf, dass der Vektor $\mathbf{y}$ eine *lineare Funktion* der $\mathbf{x}_j$, $j = 1, \ldots, n$ ist. Der Vektor

$\mathbf{y}$ sei eine Funktion $\mathbf{y} = L(\mathbf{x}_1, \ldots, \mathbf{x}_p)$ 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. $L(\mathbf{x}_1 + \cdots + \mathbf{x}_p) = L(\mathbf{x}_1) + \cdots + L(\mathbf{x}_p)$ (L ist additiv)
gilt. Eine Linearkombination ist offenbar eine lineare Funktion.

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

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

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n, \tag{1.20}$$

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

□

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:

$$Y_i = a_1 X_{i1} + a_2 X_{i2} + a_3 X_{i3} + e_i, \quad i = 1, \ldots, m \tag{1.21}$$

Für die i-te Person sind also die Messungen Y_i, X_{i1}, X_{i2} und X_{i3} gegeben und die Y_i sollen anhand der X_{ij}, $j = 1, \ldots, 3$ vorhergesagt werden; e_i ist ein Fehlerterm; er repräsentiert alle Effekte in Y_i, die nicht durch die drei Prädiktorvariablen definiert werden. Die Koeffizienten a_1, a_2 und a_3 werden im Allgemeinen mit der Methode der Kleinsten Quadrate so geschätzt, dass die Summe $\sum_i e_i^2$ der Fehlerquadrate minimal wird. Ausgeschrieben erhält man

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix} = a_1 \begin{pmatrix} X_{11} \\ X_{21} \\ \vdots \\ x_{m1} \end{pmatrix} + a_2 \begin{pmatrix} X_{12} \\ X_{22} \\ \vdots \\ x_{m2} \end{pmatrix} + a_3 \begin{pmatrix} X_{13} \\ X_{23} \\ \vdots \\ x_{m3} \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{pmatrix} \tag{1.22}$$

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

$$\mathbf{y} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{x}_3 + \mathbf{e} \tag{1.23}$$

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 3.2.6 ausführlicher behandelt. □

1.2.3 Produkte von Vektoren

Es gibt verschiedene Produkte von Vektoren. Die für die multivariate Statistik wichtigsten werden in der folgenden Definition vorgestellt:

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

$$\boldsymbol{x}'\boldsymbol{y} = (x_1, x_2, \ldots, x_n)\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \sum_{i=1}^{n} x_i y_i \tag{1.24}$$

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

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

$$\boldsymbol{x}\boldsymbol{y}' = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}(y_1, \ldots, y_m) = \begin{pmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_m \\ x_2y_1 & x_2y_2 & \cdots & x_2y_m \\ \vdots & \vdots & \ddots & \vdots \\ x_ny_1 & x_ny_2 & \cdots & x_ny_m \end{pmatrix} \tag{1.25}$$

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

Im Folgenden wird auf das Skalarprodukt fokussiert. Es spielt in der multivariaten Statistik eine wichtige Rolle, – so sind etwa Varianzen, Kovarianzen und Korrelationen Skalarprodukte. Auf das dyadische Produkt wird im Kapitel über die Matrixrechnung zurückgekommen. Ein drittes Produkt – das Vektorprodukt – kommt in der multivariaten Statistik nicht vor und wird der Vollständigkeit halber in Abschnitt 6.9 im Anhang vorgestellt.

Beispiel 1.3 Linearkombinationen und Skalarprodukte Es seien $\mathbf{x}_1, \ldots, \mathbf{x}_n$ m-dimensionale Vektoren, und es sei

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

eine Linearkombination dieser Vektoren. Für die Komponente y_i von $\mathbf{y}$ gilt dann

$$y_i = a_1x_{i1} + a_2x_{i2} + \cdots + a_nx_{in}, \quad i = 1, \ldots, m$$

y_i ist offenbar ein Skalarprodukt der Vektoren $\mathbf{a} = (a_1, a_2, \dots, a_n)'$ und $\tilde{\mathbf{x}}_i = (x_{i1}, x_{i2}, \dots, x_{in})'$[9]. Um an die Schreibweise (1.24) für Skalarprodukte anzuknüpfen kann man

$$y_i = \mathbf{a}'\tilde{\mathbf{x}}_i, \quad i = 1, 2, \dots, m \tag{1.26}$$

schreiben, oder ausführlicher

$$\begin{aligned} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \sum_{j=1}^{n} a_j \mathbf{x}_j \quad &= \begin{pmatrix} a_1 x_{11} + a_2 x_{12} + \cdots + a_n x_{1n} \\ a_1 x_{21} + a_2 x_{22} + \cdots + a_n x_{2n} \\ \vdots \\ a_1 x_{m1} + a_2 x_{m2} + \cdots + a_m x_{mn} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{a}'\tilde{\mathbf{x}}_1 \\ \mathbf{a}'\tilde{\mathbf{x}}_2 \\ \vdots \\ \mathbf{a}'\tilde{\mathbf{x}}_m \end{pmatrix}. \end{aligned} \tag{1.27}$$

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

Anmerkungen:

1. Man überprüft anhand von (1.24) sofort, dass zwar $\mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x}$ gilt, aber anhand von (1.25), dass
$$\mathbf{x}'\mathbf{y} \neq \mathbf{x}\mathbf{y}'. \tag{1.28}$$

2. **Alternative Definition:** Ein alternative, aber äquivalente Definition des Skalarprodukts wird auf Seite 21, Gleichung (1.49) 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 (1.24) 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 (1.24) 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 $\mathbf{x}\mathbf{y}'$.

4. Eine andere Schreibweise für das Skalarprodukt ist $\langle \mathbf{x}, \mathbf{y} \rangle$; sie erweist sich in bestimmten Betrachtungen als nützlich, z.B. wenn man zeigen will, dass die oben erwähnte alternative Definition die Definition $\mathbf{x}'\mathbf{y} = \sum_i x_i y_i$ impliziert.

[9]Die Schreibweise $\tilde{\mathbf{x}}$ für einen Vektor wird im Zusammenhang mit der Matrixrechnung ausführlich begründet.

5. Eine weitere Schreibweise für das Skalarprodukt ist $\mathbf{x} \cdot \mathbf{y}$ (im Englischen auch *dot product* genannt):

$$\mathbf{x} \cdot \mathbf{y} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} := \sum_{i=1}^{n} x_i y_i \tag{1.29}$$

verwendet; $\mathbf{x} \cdot \mathbf{y}$ ist nur eine andere Schreibweise für $\mathbf{x}'\mathbf{y}$. Im Folgenden wird die in (1.24) eingeführte Schreibweise $\mathbf{x}'\mathbf{y}$ verwendet, weil sie sich, wie schon angedeutet, im Zusammenhang mit der Matrixrechnung als nützlich erweist. Der Doppelpunkt vor dem Gleichheitszeichen soll aussagen, dass $\mathbf{x} \cdot \mathbf{y}$ durch die Summe $\sum_i x_i y_i$ definiert wird.

Eigenschaften des Skalarprodukts Das Skalarprodukt hat die folgenden allgemeinen Eigenschaften:
1. *Kommutativität:* Es gilt

$$\mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x}, \tag{1.30}$$

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

$$(a\mathbf{x})'\mathbf{y} = \mathbf{x}'(a\mathbf{y}) = a(\mathbf{x}'\mathbf{y}), \tag{1.31}$$

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

$$\mathbf{x}'(\mathbf{y} + \mathbf{z}) = \mathbf{x}'\mathbf{y} + \mathbf{x}'\mathbf{z}, \tag{1.32}$$

$$(\mathbf{x} + \mathbf{y})'\mathbf{z} = \mathbf{x}'\mathbf{z} + \mathbf{y}'\mathbf{z}, \tag{1.33}$$

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 (1.33) 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 6.4, Seite 177 nachgelesen werden.

Beispiel 1.4 Einheitsvektoren Für die kanonischen $\mathbf{e}_j$ gilt, wie man leicht nachrechnet,

$$\mathbf{e}_j'\mathbf{e}_k = \delta_{jk} = \begin{cases} 0, & j \neq k \\ 1, & j = k \end{cases} \tag{1.34}$$

δ_{jk} ist das *Kronecker-Delta*[10], 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

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

angeschrieben werden. Wegen der Distributivität des Skalarprodukts hat man dann

$$\begin{aligned} \mathbf{x}'\mathbf{y} &= (x_1\mathbf{e}_1 + \cdots x_n\mathbf{e}_n)'(y_1\mathbf{e}_1 + \cdots + y_n\mathbf{y}_n) \\ &= \sum_{j=1}^n \sum_{k=1}^n x_j y_k \mathbf{e}_j'\mathbf{e}_k = \sum_{j=1}^n x_j y_j, \end{aligned} \tag{1.35}$$

wegen (1.34). □

Der folgende Begriff ergibt sich unmittelbar aus der Definition des Skalarprodukts:

Definition 1.4 *Es seien $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. Für $\boldsymbol{y} = \boldsymbol{x}$ ergibt sich das Skalarprodukt*

$$\boldsymbol{x}'\boldsymbol{x} = \|\boldsymbol{x}\|^2 = \sum_{i=1}^n x_i^2, \tag{1.36}$$

d.h. das Quadrat der Länge von $\boldsymbol{x}$. Dann heißt die Länge

$$(\boldsymbol{x}'\boldsymbol{x})^{1/2} = \|\boldsymbol{x}\| \tag{1.37}$$

die Norm *des Vektors $\boldsymbol{x}$.*

Anmerkungen:

1. Redeweise: Es wird auch gesagt, dass die Norm eines Vektors durch das Skalarprodukt *induziert* wird

2. Die Definition (1.37) des Begriffs 'Norm' ist eigentlich ein Spezialfall, der auch als euklidische Norm bekannt ist. Allgemein ist mit der Norm eine "Größe" gemeint, bei der euklidischen Norm wird diese Größe durch die Länge des Vektors repräsentiert. So bezeichnet

$$\|\mathbf{x}\|_\infty = \max_{i=1,\ldots,n} |x_i| \tag{1.38}$$

   die *Maximumsnorm* eines Vektors $\mathbf{x}$; sie ist durch den größten Betrag der Komponenten des Vektors definiert.

3. Die Beziehung (1.36) wird auch *verallgemeinerter Satz des Pythagoras*) genannt.

[10] Leopold Kronecker (1823 –1891), Mathematiker

4. **Skalierung eines Vektors, normierte Vektoren** Es sei $a \in \mathbb{R}$. Dann folgt
$$\|a\mathbf{x}\| = |a|\|\mathbf{x}\|, \tag{1.39}$$
wie man durch Einsetzen in (1.36) 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
$$|a| = \frac{1}{\|\mathbf{x}\|} > 0 \tag{1.40}$$
Der Vektor
$$\mathbf{y} = \left(\frac{1}{\|\mathbf{x}\|}x_1, \frac{1}{\|\mathbf{x}\|}x_2, \ldots, \frac{1}{\|\mathbf{x}\|}x_n\right)' = \frac{1}{\|\mathbf{x}\|}(x_1, x_2, \ldots, x_n)' \tag{1.41}$$
heißt dann *normiert*, d.h. er hat die Länge 1.

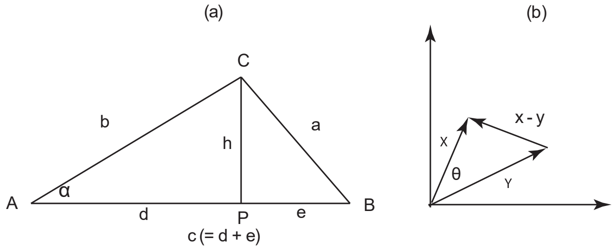
5. Die Dreiecksungleichung: Für $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ gilt
$$\|\mathbf{x} + \mathbf{y}\|^2 \leq (\|\mathbf{x}\| + \|\mathbf{y}\|)^2 \tag{1.42}$$
Denn
$$\begin{aligned} \|\mathbf{x} + \mathbf{y}\|^2 &= (\mathbf{x} + \mathbf{y})'(\mathbf{x} + \mathbf{y}) \\ &= \mathbf{x}'\mathbf{x} + \mathbf{x}'\mathbf{y} + \mathbf{y}'\mathbf{x} + \mathbf{y}'\mathbf{y} \\ &= \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2\mathbf{x}'\mathbf{y} \leq \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| \quad (1.43) \\ &\leq \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 \quad (1.44) \end{aligned}$$
wobei auf die Validität der Gleichung (1.43) weiter unten (s. Gleichung (1.50), Seite 21), noch eingegangen wird.

Über das Skalarprodukt wird also die Länge eines Vektors definiert. Darüber hinaus läßt sich der Winkel zwischen zwei Vektoren (Orientierungen) über das Skalarprodukt erklären:

Satz 1.1 *Es sei θ der Winkel zwischen den Vektoren $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. Dann gilt*
$$\cos\theta = \frac{\boldsymbol{x}'\boldsymbol{y}}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|}, \quad \|\boldsymbol{x}\| \neq 0,\ \|\boldsymbol{y}\| \neq 0 \tag{1.45}$$

Beweis: Die Aussage (1.45) ergibt sich aus einer Anwendung des Kosinussatzes: Mit
$$a = \|\mathbf{y} - \mathbf{x}\|,\ b = \|\mathbf{x}\|,\ c = \|\mathbf{y}\|$$

Abbildung 7: Zum Kosinussatz $a^2 = b^2 + c^2 - 2bc\cos\alpha$



ergibt sich (1.45) aus dem Kosinussatz $a^2 = b^2 + c^2 - 2bc\cos\alpha$:

$$\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta \tag{1.46}$$

(vergl. Abbildung 7 (b))[11]. Es ist

$$\begin{aligned}\|\mathbf{x} - \mathbf{y}\|^2 = \sum_i (x_i - y_i)^2 &= \sum_i x_i^2 + \sum_i y_i^2 - 2\sum_i x_i y_i & (1.47)\\ &= \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\mathbf{x}'\mathbf{y}. & (1.48)\end{aligned}$$

Setzt man den Ausdruck auf der rechten Seite für $\|\mathbf{x}-\mathbf{y}\|^2$ in (1.46) 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 (1.45) folgt. □

Anmerkung: Für $n \leq 3$ ist die Bedeutung des Winkels zwischen zwei Vektoren klar. Für $n > 3$ kann (1.45) auch als *Definition* des Begriffs des Winkels zwischen zwei Vektoren gesehen werden. □

Alternative Definition des Skalarprodukts: Aus (1.45) folgt sofort

$$\mathbf{x}'\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|\cos\theta. \tag{1.49}$$

Wegen $|\cos\theta| \leq 1$ hat man die Ungleichung

$$\mathbf{x}'\mathbf{y} \leq \|\mathbf{x}\|\|\mathbf{y}\|; \tag{1.50}$$

von dieser Beziehung wurde in Gleichung (1.43), Seite 20, Gebrauch gemacht. Der Ausdruck auf der rechten Seite von (1.49) ist die in der Anmerkung 2, Seite 17 genannte alternative Definition des Skalarprodukts. (1.49) besagt, dass

$$\sum_i x_i y_i = \|\mathbf{x}\|\|\mathbf{y}\|\cos\theta.$$

[11]**Beweis:** Vergl. Abbildung 7: 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. 7 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 $a^2 = b^2 + c^2 - 2bc\cos\alpha$. □.

Beginnt man mit dem Ausdruck $\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta$ als Definition des Skalarprodukts, so muß man erklärt haben, dass $\|\mathbf{x}\|$ und $\|\mathbf{y}\|$ die Längen der Vektoren $\mathbf{x}$ und $\mathbf{y}$ sind, – tatsächlich ist dieLänge eines Vektors bereits in Gleichung (1.8), Seite 10, für den Fall $n = 2$ erklärt worden, und diese Gleichung kann man sofort für allgemeinies n als verallgemeinerten Satz des Pythagoras erklären. Während die Herleitung von (1.49) zeigt, dass $\sum_i x_i y_i$ gleich $\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta$ ist, lässt sich nun umgekehrt zeigen, dass der Ausdruck $\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta$ gleich der Summe $\sum_i x_i y_i$ ist. Dazu geht man wieder vom Kosinussatz 1.46 aus. Aus dieser Gleichung folgt

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

Berücksichtigt man, dass ja nach (1.47)

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

gilt, so hat man

$$\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta = \sum_i x_i y_i.$$

Zusammenfassend hat man dass Resultat, dass die Definitionen des Skalarprudukts durch $\sum_i x_i y_i$ einerseits und durch $\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta$ andererseits äuqivalent sind, jede Definition impliziert die jeweils andere. □

In der multivariaten Statistik geht man im Allgemeinen von der Definition $\mathbf{x}'\mathbf{y} = \sum_i x_i y_i$ aus; wie in Abschnitt 1.2.4 gezeigt wird, erweisen sich z.B. zentrale Begriffe wie der Mittelwert, die Varianz und die Kovarianz im Sinne dieser Definition als Skalarprodukte. Die Definition durch den Ausdruck $\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta$ dominiert z.B. in der Physik, vergl. z.B. Greiner (1981)[12], Wylie (1975)[13], aber auch in Lehrbüchern der Mathematik, etwa Courant (1963)[14]. (1.49), heißt auch *geometrische Definition* des Skalarprodukts, während (1.24), Seite 16, als *algebraische Definition* bekannt ist. □

Die algebraische Definition (1.24) des Skalarprodukts ist wesentlich für die Definition der Produkte Matrix × Vektor und Matrix × Matrix (s. Abschnitt 1.2.4, Seite 24).

Definition 1.5 *Für $\theta = \pi/2$ (90^o) bilden $\boldsymbol{x}$ und $\boldsymbol{y}$ einen rechten Winkel, sie heißen dementsprechend* orthogonal[15]. *Gilt außerdem $\|\boldsymbol{x}\| = \|\boldsymbol{y}\| = 1$, so heißen sie* orthonormal.

Folgerung: Für $\theta = \pi/2$ gilt $\cos\theta = 0$, und für $\|\mathbf{x}\|, \|\mathbf{y}\| \neq \vec{0}$ 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 $\mathbf{y}$. □

[12]Greiner, W.: Theoretische Physik, Band 1: Mechanik

[13]Wylie, C.R.: Advanced Engineering Mathematics, Tokyo, Auckland etc 1975

[14]Courant, 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.

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

Beispiel 1.5 Einheitsvektoren Nach Gleichung (1.34), Seite 18 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 (1.20), Seite 15, können die Komponenten x_i, $i = 1, \ldots, n$ eines Vektors $\mathbf{x}$ als Koordinaten des Endpunktes von $\mathbf{x}$ auf diesen Achsen interpretiert werden. □

Anmerkungen: Für $\theta = \pi/2$, also für einen Winkel von 90^0, folgt $\cos\theta = 0$ und (1.46) 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 (1.45) als als *Definition* des Kosinus des Winkels und damit des Winkels selbst zwischen zwei Vektoren aufgefasst werden. □

Die Verknüpfungsregeln für n-dimensionale Vektoren ermöglichen es, Geometrie zu treiben, wie das folgende Beispiel zeigt.

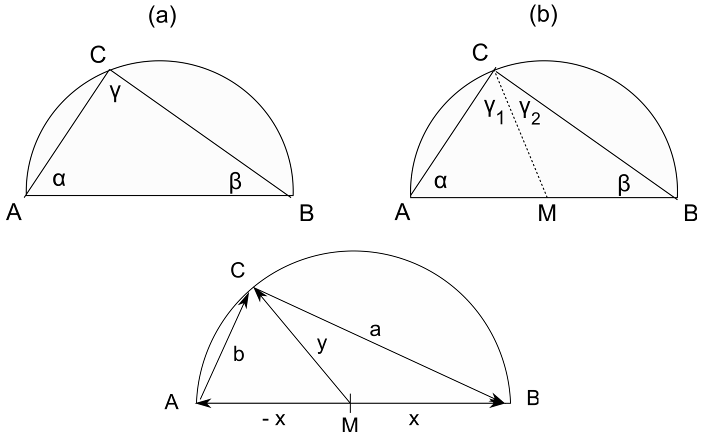
Beispiel 1.6 Der Satz des Thales[16] Dieses Beispiel dient noch einmal als Übung bezüglich der Addition, Subtraktion und dem Skalarprodukt zweier Vektoren. Nach Thales ist der Winkel γ des im Halbkreis eingeschriebenen Dreiecks ABC stets ein rechter Winkel (90^o , entsprechend $\pi/2$), unabhängig vom Ort C auf dem Halbkreis, s. Abb. 8 (a). Es gibt verschiedene Beweise für diese Aussage. Kann man die Aussagen
(i) Die Winkelsumme im Dreieck beträgt 180^o 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 $\gamma_1 = \alpha, \quad \gamma_2 = \beta, \quad \gamma = \gamma_1 + \gamma_2$, dass $\alpha + \beta + \gamma = 2\gamma = \pi$, also $\gamma = \pi/2$, vergl. Abb. 8 (b). Macht man von der 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. 2, Seite 10, Abb. 5, Seite 14, sowie Abb 6, Seite 14, sowie die Anmerkung nach Gleichung (1.17)), sowie der Bildung und Interpretation eines Skalarprodukts (Abb. 8, (c)). Man setze $\mathbf{x} = \overline{MB}$, $-\mathbf{x} = \overline{MA}$, $\mathbf{y} = \overline{MC}$, vergl. Abb. 8, und es gilt $\|\mathbf{x}\| = \|\mathbf{y}\|$. Weiter ist $a = \mathbf{x} + \mathbf{y}$, $b = \mathbf{x} - \mathbf{y}$. Dann folgt für das Skalarprodukt

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

also ist der Winkel zwischen den Vektoren $\mathbf{x}+\mathbf{y}$ und $\mathbf{x}-\mathbf{y}$ gleich 90^o entsprechend (1.45), Seite 20, d.h. gemäß Definition 1.5 sind diese Vektoren orthogonal. □

[16] Thales 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.

Abbildung 8: Der Satz des Thales



1.2.4 Mittelwerte, Varianzen, und Kovarianzen

Einige Grundbegriffe der Statistik, wie Mittelwert, Varianz bzw. Standardbweichung, Kovariannz und Korrelation werden durch Skalarprodukte definiert:

Das arithmetische Mittel: Es sei $X = (X_1, X_2, \ldots, X_n)'$ und $\vec{1} = (1, 1, \ldots, 1)' \in \mathbb{R}^n$. Dann ist

$$\mathbf{X}'\vec{1} = \sum_{i=1}^{n} X_i \cdot 1 = \sum_{i=1}^{n} X_i$$

d.h.

$$\frac{1}{n}\mathbf{X}'\vec{1} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{x} \tag{1.51}$$

ist das arithmetische Mittel $\bar{x}$ der Komponenten X_i von $\mathbf{X}$ läßt sich über das Skalarprodukt von $\mathbf{X}$ mit dem 1-Vektor $\vec{1}$ darstellen.

Die Zentrierung: Es sei $\mathbf{x} = (x_1, \ldots, x_n)'$ und die i-te Komponente x_i sei durch

$$x_i = X_i - \bar{x}, \quad i = 1, \ldots, n$$

definiert. Dann heißt $\mathbf{x}$ *zentrierter Vektor*; für zentrierte Vektoren gilt offenbar

$$\mathbf{x}'\vec{1} = \sum_{i=1}^{n} x_i = 0, \tag{1.52}$$

denn aus (1.51) folgt $\sum X_i = n\bar{x}$, so dass

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n}(X_i - \bar{x}) = \underbrace{\sum_{i=1}^{n} X_i}_{n\bar{x}} - n\bar{x} = 0$$

Varianz und Standardabweichung: $\mathbf{x} = (x_1, \dots, x_n)'$ sei ein zentrierter Vektor. Dann ist

$$\mathbf{x}'\mathbf{x} = \|\mathbf{x}\|^2 = \sum_{i=1}^{n} x_i^2,$$

so dass

$$s_x^2 = \frac{1}{n}\mathbf{x}'\mathbf{x} = \frac{1}{n}\|\mathbf{x}\|^2, \tag{1.53}$$

$$s_x = \frac{1}{\sqrt{n}}\|\mathbf{x}\| \tag{1.54}$$

gerade die (Stichproben-)Varianz und die (Stichproben-)Stndardbweichung der Komponenten x_i von $\mathbf{x}$ ist. s_x^2 ist über das Skalarprodukt von $\mathbf{x}$ mit sich selbst definiert[17].

Kovarianz und Korrelation: Es sei $\mathbf{y} = (y_1, \dots, y_n)'$ ebenfalls ein zentrierter Vektor. Dann ist

$$\mathrm{Kov}_{xy} = \frac{1}{n}\mathbf{x}'\mathbf{y} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{x})(Y_i - \bar{y}) \tag{1.55}$$

die Kovarianz der Messwerte X_i, Y_i; die (Stichproben-)Kovarianz ist also über das Skalarprodukt von $\mathbf{x}$ und $\mathbf{y}$ definiert. Weiter gilt

$$r_{xy} = \frac{\mathrm{Kov}(x,y)}{s_x s_y} = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}, \tag{1.56}$$

d.h. die Produkt-Moment-Korrelation der Messwerte X_i und Y_i ist durch Skalarprodukte definiert. Mit Bezug auf (1.45), Seite 20, folgt daraus für zentrierte Vektoren

$$\cos\theta = r_{xy}. \tag{1.57}$$

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.

1.2.5 Länge, Orientierung und Kosinus-Ähnlichkeit von Vektoren

Die Länge $\|\mathbf{x}\|$ eines Vektors $\mathbf{x}$ wurde bereits in (1.36), Seite 19 über das Skalarprodukt eines Vektors mit sich selbst eingeführt, und die Orientierung eines n-dimensionalen Vektors wurde über (1.45), Seite 20 erklärt. Der Winkel zwischen einem Vektor $\mathbf{x}$ und dem j-ten Einheitsvektor ist dann durch

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

[17] Der 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.

gegeben, und demnach erhält man für die j-te Kkomponente von $\mathbf{x}$

$$x_j = \|\mathbf{x}\| \cos\theta_j, \quad j = 1, \ldots, n. \tag{1.58}$$

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

$$\mathbf{x} = \|\mathbf{x}\| \begin{pmatrix} \cos\theta_1 \\ \cos\theta_2 \\ \vdots \\ \cos\theta_n \end{pmatrix} = \|\mathbf{x}\|\mathbf{c}_x \tag{1.59}$$

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

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

$$\frac{\mathbf{x}}{\|\mathbf{x}\|} = \|\mathbf{c}_x\| = 1. \tag{1.60}$$

Zur Bedeutung von Orientierungsvektoren: Gegeben seien n m-dimensionale Vektoren $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Die Komponenten eines Vektors $\mathbf{x}_j$, $1 \leq jleqn$, seien Messungen einer Variablen – der j-ten – an m "Fällen" (Personen oder Objekten), d.h. x_{ij}, die i-te Komponente des j-ten Vektors, sei der Messwert der j-ten Variablen beim i-ten Fall. Schreibt man die Vektoren nebeneinander an, so entsteht eine (m, n)-Tabelle (im Folgenden auch Matrix genannt). Man ist zunächst an den Korrelationen zwischen den Variablen interessiert. Dazu werden die Vektoren standardisiert, um den Effekt verschiedener Maßeinheiten bei den verschiedenen Variablen zu eliminieren. Man geht also von den x_{ij} zu den z_{ij} über:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, \quad \bar{x}_j = \frac{1}{m}\sum_{i=1}^{m} x_{ij}, \quad s_j = \sqrt{\frac{1}{m}(x_{ij} - \bar{x}_j)^2} \tag{1.61}$$

Für die Zwecke dieses Abschnitts genügt es, den Fall $m = n = 2$ zu betrachten, die Verallgemeinerung auf den allgemeinen all $m > 2$, $n > 2$ ist dann direkt durchführbar. Zunächst werde die Tabelle der z_{ij}-Werte vorgestellt: Es gibt hier

Tabelle 1: Datentabelle

	Variablen	
Fälle	$\mathbf{z}_r$	$\mathbf{z}_s$
$\tilde{\mathbf{z}}_i$	z_{ir}	z_{is}
$\tilde{\mathbf{z}}_j$	z_{jr}	z_{js}

insgesamt vier Vektoren: die *Spaltenvektoren* $\mathbf{z}_j$ und $\mathbf{z}_{j'}$, und die *Zeilenvektoren*

$\tilde{\mathbf{z}}_i$ und $\tilde{\mathbf{z}}_{i'}$. Es können nun zwei Skalarprodukte gebildet werden:

$$\mathbf{z}_r'\mathbf{z}_s = z_{ir}z_{is} + z_{jr}z_{js} \tag{1.62}$$

$$\tilde{\mathbf{z}}_i'\tilde{\mathbf{z}}_j = z_{ir}z_{js} + z_{is}z_{js} \tag{1.63}$$

Um die Unterschiede zwischen den Skalarprodukten sichtbarer zu machen seien die Gleichungen noch einmal in ausführlicher Form angeschrieben:[18]

$$\mathbf{z}_r'\mathbf{z}_s = \frac{(x_{ir}-\bar{x}_r)}{s_r}\frac{(X_{is}-\bar{x}_s)}{s_s} + \frac{(x_{jr}-\bar{x}_r)}{s_r}\frac{(X_{js}-\bar{x}_s)}{s_s} \tag{1.64}$$

$$= \|\mathbf{z}_r\|\|\mathbf{z}_s\|\cos\phi_{rs} \tag{1.65}$$

und

$$\tilde{\mathbf{z}}_i'\tilde{\mathbf{z}}_j = \frac{(x_{ir}-\bar{x}_r)}{s_r}\frac{(X_{jr}-\bar{x}_r)}{s_s} + \frac{(x_{is}-\bar{x}_s)}{s_r}\frac{(x_{js}-\bar{x}_s)}{s_s} \tag{1.66}$$

$$= \|\tilde{\mathbf{z}}_i\|\|\tilde{\mathbf{z}}_j\|\cos\varphi_{ij} \tag{1.67}$$

Bis auf den Faktor $1/m$ (im Falle von m Fällen) entspricht das Skalarprodukt $\mathbf{z}_r'\mathbf{z}_s$ einem Produkt-Moment-Korrelationskoeffizienten, das Skalarprodukt $\tilde{\mathbf{z}}_i'\tilde{\mathbf{z}}_j$ aber nicht. Andererseits folgt aus dem Vorangegangenen

$$\frac{\mathbf{z}_r'\mathbf{z}_s}{\|\mathbf{z}_r\|\|\mathbf{z}_s\|} = \cos\phi_{rs} = r_{rs} \tag{1.68}$$

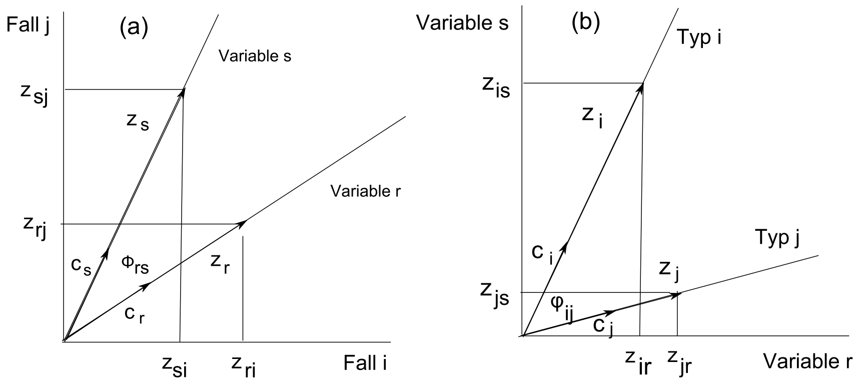
$$\frac{\tilde{\mathbf{z}}_i'\tilde{\mathbf{z}}_j}{\|\tilde{\mathbf{z}}_i\|\|\tilde{\mathbf{z}}_j\|} = \cos\varphi_{ij} \neq r_{ij} \tag{1.69}$$

Es ist klar, dass diese Formeln für irgendeine Anzahl m von Fällen verallgemeinert werden können, – hier sollen nur die Unterschiede zwischen (1.64) und (1.66) anhand des Spezialfalls $m = 2$ illustriert werden. Sowohl (1.68) wie (1.69) sind Ähnlichkeitsmaße für die Orientierungen der jeweiligen Vektoren, beide Maße nehmen Werte zwischen -1 und +1 an. Es soll deshalb kurz auf die Bedeutung der Vektororientierung eingegangen werden; der Spezialfall $m = 2$ erlaubt eine graphische Veranschaulichung. Abb. 9 (a) illustriert die Kosinusähnlichkeit von Variablen. Das Koordinatensystem wird durch Fälle definiert, die Messwerte der Variablen wurden z-standardisiert. Die Darstellung (a) von Variablen in einem durch Fälle definierten Koordinatensystem ist ein wenig unüblich, führt aber auf die vertraute Formel für den Korrelationskoefizienten. Die Vektoren $\mathbf{c}_r$ und $\mathbf{c}_s$ definieren die von den Fällen bestimmten Orientierungen, d.h. schlicht von den normierten, standardisierten Einheiten der Stichprobe von Fällen. Der Vektor $\mathbf{Z}_r$ ist die Repräsentation der r-ten Variablen, repräsentiert als Abweichung vom Stichprobenmittelwert (also gemittelt über die Fälle) in Einheiten der Standardabweichung für diese Variable.

Abb. 9 (b) zeigt die Repräsention der Fälle in einem von den Variablen definierten Koordinatensystem. Sie entspricht der Punktekonfiguration von Fällen,

[18] Die Längen $\|\mathbf{z}_r\|$ und $\|\mathbf{z}_s\|$ sind nicht gleich 1, sondern im allgemeinen Fall – gleich m, wie man leicht nachrechnet.

Abbildung 9: Typen und Skalarprodukte: (a) Kosinusähnlichkeit von Variablen (Korrelation), (b) Kosinusähnlichkeit von Fällen

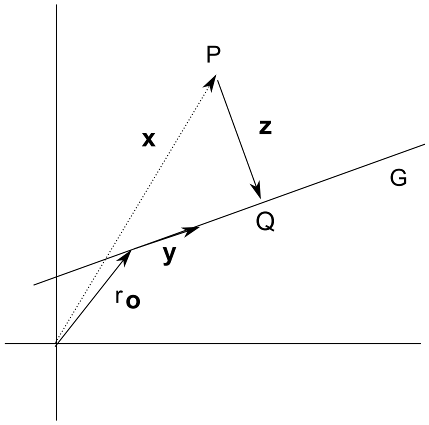


wie man sie aus der Regressionsrechnung kennt. Der Punkt bei dieser Repräsentation ist, dass die Skalarprodukte für Paare von Fällen *keine* Korrrelationskoeffizienten sind. Der Ausdruck 'Typ' für die Fälle soll sagen, dass es sich der Fall durch eine spezielle Mischung von Merkmalsausprägungen handelt. Handelt es sich bei den Fällen um Personen, so entspricht der Ausdruck 'Typ' zumindest im Prinzip dem Begriff 'Typ', wie er in der Persönlichkeitspsycholgie verwendet wird, allerdings eher als Bezeichung einer Art von idealisierten Person, der einzelne Personen mehr oder weniger gut zugeordnet werden können. Auf ähnliche Art wird auch in anderen Wissenschaften vorgegangen: Tonscherben können für eine spezielle Kultur in einem speziellen Zeitabschnitt typisch sein, wenn sie sich bis auf stochastishe Effekte ähneln, etc. Hier meint 'Typ' einfach eine spezielle Merkmalskombination, die durch die Orientierungsvektoren gekennzeichnet werden. Speziell im 2-dimensionalen Fall kann man die Mischung durch das Verhältnis der Ausprägungen – repräsentiert in Termen von Standardabweichhungen) kennzeichnen (vergl. Abbildung 2, Seite 10, und die dortige erste Einführung der Richtungskosinus). Die durch die Vektoren $\mathbf{c}_i$ und $\mathbf{c}_j$ definierten Geraden definierten Mengen von Fällen, die alle durch die jeweilige Mischung charakterisiert werden. In Abschnitt 3.6 wird auf die Darstellungen (a) und (b) von Abb. 9 im Rahmen eines statistischen Analyse näher eingegangen.

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 10). 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

Abbildung 10: Die Projektion eines Punktes auf eine Gerade



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. $\mathbf{z} = \mathbf{x} - (\mathbf{r}_o + \lambda_0\mathbf{y})$, $\mathbf{z}'\mathbf{y} = 0$.

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

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

d.h.

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

so dass

$$\lambda_0 = \frac{(\mathbf{x} - \mathbf{r}_0)'\mathbf{y}}{\mathbf{y}'\mathbf{y}} \tag{1.70}$$

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

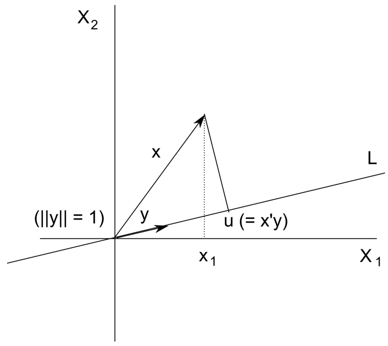
$$Q(\mathbf{x}) = \mathbf{r}_0 + \left(\frac{(\mathbf{x} - \mathbf{r}_0)'\mathbf{y}}{\mathbf{y}'\mathbf{y}}\right)\mathbf{y}. \tag{1.71}$$

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

$$Q(\mathbf{x}) = \mathbf{x}_\mathbf{y} = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{y}'\mathbf{y}}\mathbf{y} \tag{1.72}$$

Dies ist ebenfalls die Formel, die man erhält, wenn man die orthogonale Projektion $\mathbf{x}_\mathbf{y}$ 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 10. (ii) $\mathbf{r}_0 = \vec{0}$

Abbildung 11: Neue Koordinate u als Projektion eines Vektors auf eine Gerade L



und $\|\mathbf{y}\| = 1$: In diesem Fall hat man nach (1.70) und (1.72) wegen $\mathbf{y}'\mathbf{y} = 1$ ist $\lambda_0 = \mathbf{x}'\mathbf{y}$ und man erhält

$$\mathbf{x}_{\mathbf{y}} = (\mathbf{x}'\mathbf{y})\mathbf{y} \tag{1.73}$$

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 (1.45), Seite 20 impliziert den

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

$$|\boldsymbol{x}'\boldsymbol{y}|^2 \leq \|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2. \tag{1.74}$$

Beweis: Gleichung (1.45) impliziert

$$\mathbf{x}'\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|\cos\theta, \tag{1.75}$$

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 (1.74). □

Korollar 1.1 *Es gilt*

$$\boldsymbol{x}'\boldsymbol{y} = \|\boldsymbol{y}\|\|\boldsymbol{x}\| \tag{1.76}$$

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

Beweis: Offenbar ist $\mathbf{x}'\mathbf{y} = \|\mathbf{y}\|\|\mathbf{x}\|$ genau dann, wenn $\theta = 0$. Dann sind $\mathbf{x}$ und $\mathbf{y}$ parallel, d.h. sie unterscheiden sich 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 (1.74) wird oft in der Form

$$\left|\sum_{i=1}^n x_i y_i\right|^2 \leq \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2 \tag{1.77}$$

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.

$$\left|\sum_{i=1}^n x_i y_i\right|^2 = \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2 = \lambda\left(\sum_{i=1}^n x_i^2\right)^2. \tag{1.78}$$

(s.a. (1.75)).

1.2.8 Lineare Abhängigkeit und Unabhängigkeit von Vektoren

Der Begriff der Unabhängigkeit etwa von Variablen ist, wie die Betrachtungen zum Skalarpprodukt 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 Variaben 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; für kontinuierliche Variable gilt diese Beziehung nur für die 2-dimensionale Gauss-Dichte im Fall $r_{xy} = 0$. Bei den folgenden Betrachtungen werden aber keine Annahmen über Wahtscheinlichkeitsverteilungen gemacht. In der linearen Algebra, die ja keine statistische Theorie ist, wird ein Abhängigkeits- bzw. Unabhängigkeitsbegriff eingeführt, der weder stochastische Unabhängigkeit noch Unabhängigkeit im Sinne einer Null-Korrelation bedeutet: selbst wenn $\cos\theta$ nur "unwesentlich" von 1 abweicht können zwei Vektoren dennoch im algebraischen Sinn unabhängig voneinander sein, wie gezeigt werden wird. Dieser Sachverhalt ist von Bedeutung unter anderem für die Diskussion von Daten in Termen von unterliegenden, "latenten" Variablen.

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

$$\mathbf{x}_1 = a_2\mathbf{x}_2 + \ldots + a_p\mathbf{x}_p \tag{1.79}$$

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

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

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 (1.79) 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 $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ 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

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

Man kann die Gleichung

$$\vec{0} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \ldots + a_p\mathbf{x}_p \tag{1.80}$$

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. $\mathbf{a} = (a_1, \ldots, a_p)' = \vec{0}$ ist stets eine Lösung für die Gleichung (1.80). $\mathbf{a} = \vec{0}$ heißt dementsprechend auch die *triviale Lösung* für die Gleichung (1.80). Im Falle der linearen Abhängigkeit existiert dann aber außerdem ein Vektor $\mathbf{a} \neq \vec{0}$, für den nicht alle Koeffizienten a_j gleich Null sind, d.h. . Das Prädikat "linear abhängig" ist die Negation des Prädikats "linear unabhängig", es gilt also "nicht nicht alle Koeffizienten sind gleich Null", d.h. "alle Koeffizienten a_j in (1.80) sind gleich Null". Man hat dementsprechend die

Definition 1.6 *Es gelte (1.80); sind nicht alle Komponenten a_j des Vektors $\boldsymbol{a} = (a_1, \ldots, a_p)'$ gleich Null, so heißen die Vektoren $\boldsymbol{x}_j$* linear abhängig. *Ist dagegen $\boldsymbol{a} = \vec{0}$ die einzige Lösung für (1.80), so heißen die $\boldsymbol{x}_j$* 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. □

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

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

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

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

Satz 1.4 *Die Vektoren $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^n$, $\boldsymbol{x}_j \neq \vec{0}$ für alle j, seien paarweise orthogonal. Dann sind sie linear unabhängig.*

Beweis: Es sei
$$\vec{0} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_n\mathbf{x}_n.$$
Dann gilt
$$\mathbf{x}_j'\vec{0} = 0 = a_1\mathbf{x}_j'\mathbf{x}_1 + a_2\mathbf{x}_j'\mathbf{x}_2 + \cdots + a_j\mathbf{x}_j'\mathbf{x}_j + \cdots + a_n\mathbf{x}_j'\mathbf{x}_n$$
und 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. □

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

Satz 1.5 *Es sei $\boldsymbol{x}$ ein beliebiger n-dimensionaler Vektor. Dann ist $\boldsymbol{x}$ als Linearkombination der Einheitsvekoren $\mathbf{e}_j$ $j = 1\ldots, n$, 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. □

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

Beweis: Ohne Einschränkung der Allgemeinheit kann $\mathbf{x}_1 = \vec{0}$ angenommmen werden, – die Nummerierung der Vektoren ist ja beliebig. Es existieren reelle Koeffizienten $a_1, \ldots, a_n$ 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 $\mathbf{x}_2, \ldots, \mathbf{x}_n$ linear abhängig und damit sind auch die $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ 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 $\mathbf{x}_1, \ldots, \mathbf{x}_p$ linear abhängig sind. $\square$

Lineare Unabhängigkeit und Korrelationen: 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

$$\cos\theta = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}, \quad \|\mathbf{x}\| \neq 0, \; \|\mathbf{y}\| \neq 0,$$

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

Die Vektoren $\mathbf{x}$ und $\mathbf{y}$ sind parallel (in Zeichen: $\mathbf{x}\|\mathbf{y}$), wenn ein Skalar $a \in \mathbb{R}$ existiert derart, dass $\mathbf{y} = a\mathbf{x}$. Dann sind $\mathbf{x}$ und $\mathbf{y}$ linear abhängig, denn $\vec{0} = a\mathbf{x} - \mathbf{y} = a_1\mathbf{x} + \mathbf{a}_2\mathbf{y}$ mit $a_1 = a$, $a_2 = -1$. Gilt aber $\mathbf{y} \nparallel \mathbf{x}$, so läßt sich keiner der beiden Vektoren als Linearkombinbation 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 $\mathbf{x}\|\mathbf{y}$, 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 $\mathbf{y} = a\mathbf{x}$ existiert[19], somit folgt $\mathbf{x} \nparallel \mathbf{y}$, 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 $\underline{,}\mathbf{y}$ 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 für die Wahrscheinlichkeitsdichte $f(x, y)$ die Bedingung

$$f(x, y) = f_x(x) f_y(y)$$

[19] *modus tollens*: $(p \to q) \Rightarrow (\neg q \to \neg p)$.

erfüllt ist; die rechte Seite entspricht der *Produktregel* für stochastisch unabhängige zufällige Ereignisse A und B: $p(A\&B) = p(A)p(B)$. Diese Bedingung ist für die 2-dimensionale Gauß-Verteilung

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y(1-r^2)} \exp\left[-\frac{1}{2(1-r^2)}(z_x^2 + z_y^2 - 2rz_xz_y)\right], \tag{1.81}$$

erfüllt mit $z_x = (x-\mu_x)/\sigma_x$, $z_y = (y-\mu_y)/\sigma_y$, ρ der Korrelationskoefiizient. Für $r = 0$ folgt $f(x,y) = f_x(x)f_y(y)$, also die "Produktregel" für stochastisch unabhängige zufällige Veränderliche.

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 $|\theta| \neq 1$. 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

$$\mathbf{x} = \mathbf{x}_0 + \boldsymbol{\varepsilon}_x, \quad \mathbf{y} = \mathbf{y}_0 + \boldsymbol{\varepsilon}_y \tag{1.82}$$

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

$$\begin{aligned} -1 < r_{xy1} = \cos\theta_{xy} &= \frac{(\mathbf{x}_0 + \boldsymbol{\varepsilon}_x)'(a\mathbf{x}_0 + \boldsymbol{\varepsilon}_y)}{\|\mathbf{x}_0 + \boldsymbol{\varepsilon}_x\|\|a\mathbf{x}_0 + \boldsymbol{\varepsilon}_y\|} \\ &= \frac{a\|\mathbf{x}_0\|^2 + \mathbf{x}_0'(\boldsymbol{\varepsilon}_y + a\boldsymbol{\varepsilon}_x) + \boldsymbol{\varepsilon}_x'\boldsymbol{\varepsilon}_y}{\|\mathbf{x}_0 + \boldsymbol{\varepsilon}_x\|\|a\mathbf{x}_0 + \boldsymbol{\varepsilon}_y\|} < 1, \end{aligned} \tag{1.83}$$

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 $\mathbf{x}_j$, $j = 1, \ldots, n$ 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

$$\begin{aligned} -1 < r_{xy_2} = \cos\theta_{xy} &= \frac{(\mathbf{x}_0 + \boldsymbol{\varepsilon}_x)'(\mathbf{y}_0 + \boldsymbol{\varepsilon}_y)}{\|\mathbf{x}_0 + \boldsymbol{\varepsilon}_x\|\|\mathbf{y}_0 + \boldsymbol{\varepsilon}_y\|} \\ &= \frac{\mathbf{x}_0'\mathbf{y}_0 + \mathbf{x}_0'\boldsymbol{\varepsilon}_y + \mathbf{y}_0'\boldsymbol{\varepsilon}_x + \boldsymbol{\varepsilon}_x'\boldsymbol{\varepsilon}_y}{\|\mathbf{x}_0 + \boldsymbol{\varepsilon}_x\|\|\mathbf{y}_0 + \boldsymbol{\varepsilon}_y\|} < 1 \end{aligned} \tag{1.84}$$

Gegeben ist nur der empirische Wert r_{xy}, und die Frage ist, ob $r_{xy} = r_{xy_1}$ oder $r_{xy} = r_{xy_2}$ 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 $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ von Vektoren wird darauf zurückgekommen.

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

Definition 1.7 *Es seien $p_1, p_2, \ldots, p_r$ 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 Vektorraumes

Die Begriffe Vektorraum und Teilvektorraum spielen eine zentrale Rolle bei der Bestimmung der "latenten Dimensionen" für eine gegebene Menge von Datenvektoren. Zwei einfache Beispiele sollen diese Begriffe illustrieren.

Eindimensionaler Teilraum: Man erinnere sich daran, dass der Nullvektor $\vec{0}$ für sich genommen linear abhängig ist (vergl. Anmerkung zur Definition 1.6, Seite 32). Ein einzelner Vektor $\mathbf{u}$ dagegen ist linear unabhängig: es sei $\mathbf{u} \neq \vec{0}$ ein beliebiger n-dimensionaler Vektor, und es gelte $\vec{0} = a\mathbf{u}$, $a \in \mathbb{R}$. Da diese Gleichung nur gelten kann, wenn $a = 0$, folgt, dass $\mathbf{u}$ linear unabhängig ist.

Es seien $\mathbf{u}_1$ und $\mathbf{u}_2$ zwei Vektoren aus $\mathbb{R}^n$ mit identischer Orientierung, d.h. es gelte

$$\mathbf{u}_j = \|\mathbf{u}_j\|\mathbf{c}, \quad j = 1, 2$$

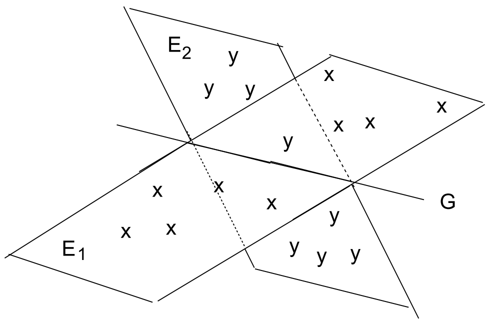
Offenbar gilt

$$\mathbf{v} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 = a_1\|\mathbf{u}_1\|\mathbf{c} + a_2\|\mathbf{u}_2\|\mathbf{c} = (a_1\|\mathbf{u}_1 + a_2\|\mathbf{u}_2\|)\mathbf{c} = b\mathbf{c}, \tag{1.85}$$

wobei $b = a_1\|\mathbf{u}_1 + a_2\|\mathbf{u}_2\|$, d.h. $\mathbf{v}$ hat ebenfalls die Orientierung $\mathbf{c}$. Dies bedeutet, dass alle Linearkombinationen $\mathbf{v}$ von $\mathbf{u}_1$ und $\mathbf{u}_2$ Elemente der Menge $\mathcal{G}$ der Vektoren mit der Orientierung $\mathbf{c}$ sind. Statt des Wortes 'Menge' wird auch das Wort *Raum* wird von $\mathcal{G}$ als von einem eindimensionalen Raum gesprochen, – auch wenn $\mathbf{u}_1$ und $\mathbf{u}_2$ n-dimensionale Vektoren mit beliebigem $n < \infty$ sind.

Offenbar ist $\mathcal{G}_\mathbf{c}$ eine Teilmenge von n-dimensionalen Vektoren aus der Menge aller n-dimensionalen Vektoren. Für $n = 3$ kann man $\mathcal{G}_\mathbf{c}$ durch eine Gerade in diesem Raum repräsentieren, und die Gerade ist dann ein Teilraum des $\mathbb{R}^3$. Der Ausdruck 'Teilraum' wird dann auch für den allgemeinen Fall $\mathbb{R}^n$ beibehalten.

Abbildung 12: 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.



Zweidimensionaler Teilraum: Nun seien $\mathbf{u}_1 \neq \vec{0}$ und $\mathbf{u}_2 \neq \vec{0}$ zwei nicht parallele, also linear unabhängige (s, Satz 1.7, Seite 34) n-dimensionale Vektoren; der Anschaulichkeit wegen mag man zunächst den Fall $n = 3$ betrachten. Um zu sehen, dass die beiden Vektoren eine Ebene $\mathcal{E}$ "aufspannen", werde eine beliebige Linearkombination

$$\mathbf{v} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 \tag{1.86}$$

der $\mathbf{u}_1$ und $\mathbf{u}_2$ betrachtet, d.h. die Koeffizienten a_1 und a_2 werden beliebig aus $\mathbb{R}$ gewählt. Dabei entsteht eine Menge von Vektoren, die "zweidimensional" ist, weil sie von zwei Vektoren $\mathbf{u}_1$ und $\mathbf{u}_2$ erzeugt – eben aufgespannt – werden. Dass die Vektoren $\mathbf{v}$ dieser Menge eine Ebene bilden, macht man sich klar, wenn man bedenkt, dass für $n > 2$ stets ein Vektor $\mathbf{n}$ existiert, der senkrecht auf sowohl $\mathbf{u}_1$ wie auch $\mathbf{u}_2$ steht, wobei $\mathbf{u}_1, \mathbf{u}_2$ beliebig aus der Ebene gewählt werden können; für den Fall $n = 3$ ist dieser Sachverhalt anschaulich vorstellbar. $\mathbf{n}$ heißt *Normalenvektor* für $\mathcal{E}$. Man sieht leicht, dass $\mathbf{n}$ auch senkrecht auf $\mathbf{v}$ steht, denn

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

da ja nach Voraussetzung $\mathbf{n}'\mathbf{u}_1 = \mathbf{n}'\mathbf{u}_2 = 0$. Da die a_1, a_2 alle Werte aus $\mathbb{R}$ annehmen dürfen, liegt es nahe, zu vermuten, dass alle Vektoren der Ebene $\mathcal{E}$ als Linearkombinationen von irgendzwei linear unabhängigen Vektoren $\mathbf{u}_1 \in \mathcal{E}$ und $\mathbf{u}_2 \in \mathcal{E}$ erzeugt werden können, indem man a_1 und a_2 die Menge $\mathbb{R}$ der reellen Zahlen durchlaufen läßt. Natürlich kann man ebensogut $-\mathbf{n}$ betrachten. Da eine Ebene den Raum in zwei Teile teilt, kann man mit $\mathbf{n}$ und $-\mathbf{n}$ kennzeichnen, welche Seite der Ebene man jeweils meint, falls das von Interesse sein sollte.

$\mathcal{E}$ ist eine Teilmenge *aller* n-dimensionalen Vektoren, und wieder ist hierfür der Ausdruck 'Teilraum' üblich, in Anlehnung an den Spezialfall $n = 3$.

Im Folgenden geht es primär um die Bestimmung von "latenten" Vektoren[20], die die Darstellung von empirisch gegebenen Vektoren erlauben; diese Vektoren sind unabhängig im Sinne der linearen Unabhängigkeit. Es wird zunächst eine Definition des Begriffs des Vektorraum gegeben, allerdings nicht in der Allgemeinheit, die in Lehrbüchern der Linearen Algebra üblich ist, sondern nur in hier benötigten Form; die allgemeine Definition wird im Anhang, Abschnitt 6.2.1, Seite 173, vorgestellt. Die allgemeine Definition erfordert die Einführung weiterer Begriffe aus der Algebra, die im Folgenden aber nicht benötigt werden. Auf die allgemeine Definition wird zurückgegriffen, wenn Methoden der Datenanalyse behandelt werden, die einen verallgemeinerten Vektorbegriff voraussetzen.

Definition 1.8 *Ein* Vektorraum über $\mathbb{R}$ *ist eine nichtleere Menge V von n-dimensionalen, für die gilt:*
(1) Der Nullvektor $\vec{0}$ ist Element von V
(2) Summenbildung: für beliebige $\boldsymbol{x}, \boldsymbol{y} \in V \Rightarrow \boldsymbol{v} = \boldsymbol{x} + \boldsymbol{y} \in V$, und
(3) Multiplikation mit einem Skalar: $a \in \mathbb{R}$, $\boldsymbol{x} \in V \Rightarrow a\boldsymbol{x} \in V$,
(4) Es sei $U \subseteq V$ eine Teilmenge von Vektoren aus V. U heißt Teilvektoraum *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.

Folgerung: V sei ein Vektorraum, und es gelte $\mathbf{x}, \mathbf{y} \in V$, $a_1, a_2 \in \mathbb{R}$. Dann ist auch $\mathbf{v} = a_1\mathbf{x} + a_2\mathbf{y} \in V$, d.h. für irgendzwei Elemente von V ist auch die Linearkombination der Elemente ein Element von V.

Anmerkung: Der Kern der Definition des Vektorraums V ist die Abgeschlossenheit von V in Bezug auf die Bildung von Linearkombinationen. Es zeigt sich, dass die Axiome (1) bis (4) auch auf andere mathematische Objekte wie Matrizen, Polynome, allgemein bestimmte Arten von Funktionen zutreffen, wobei aber das Skalarprodukt für diese Objekte nicht definiert ist. Vektorräume, für deren Elemente ein Skalarprodukt definiert ist – Skalarprodukträume – , bilden also eine Unterklasse der allgemeinen Vektorräume; dies ist für die hier betrachteten Vektorräume der Fall. Die im Folgenden betrachteten Vektorräume sind demnach Skalarprodukträume. □

Wie in den einführenden Beispielen gezeigt wurde sind r-dimensionale Teilraume (Gerade: $r = 1$, Ebene: $r = 2$, etc), $r \leq n$, Mengen von Vektoren, die sich als Linearkombinationen von jeweils r linear unabhängigen n-dimensionalen Vektoren $\mathbf{u}_j$ ergeben. Diese 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.

[20]Das ist der Jargon der Multivariaten Analyse.

(Teil-)Vektorräume spielen eine Rolle bei der Diskussion von Lösungen für lineare Gleichungssysteme. So sei

$$\mathcal{L} = \{\mathbf{v}|\mathbf{v} = a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n\}, \tag{1.87}$$

d.h. $\mathcal{L}$ sei die Menge *aller* Vektoren, die sich als Linearkombinationen der $\mathbf{x}_j$ ergeben. Diese Menge entspricht der Menge aller möglichen Tupel $(a_1, \ldots, a_n) \in \mathbb{R}^n$). Ist nun ein Vektor $\mathbf{v}$ gegeben, von dem man aus irgendwelchen Gründen weiß, dass er ein Element von $\mathcal{L}$ ist, so kann man nach den Koeffizienten $a_1, \ldots, a_n$ fragen, für die $\mathbf{v} = a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n$ gilt, d.h. man muß das Gleichungssystem

$$\begin{aligned} v_1 &= a_1x_{11} + a_2x_{12} + \cdots + a_nx_{1n} \\ v_2 &= a_1x_{21} + a_2x_{22} + \cdots + a_nx_{2n} \\ &\vdots \\ v_n &= a_1x_{m1} + a_2x_{m2} + \cdots + a_nx_{mn} \end{aligned} \tag{1.88}$$

mit den Unbekannten $a_1, \ldots, a_n$ lösen. Allgemein kann man sagen, dass man stets eine Lösung für einen beliebigen Vektor $\mathbf{v} \in \mathbb{R}^n$ finden kann, wenn das Gleichungssystem die Form

$$\mathbf{v} = a_1\mathbf{u}_1 + \cdots + a_n\mathbf{u}_n, \quad \mathbf{u}_j \in \mathbb{R}^n \tag{1.89}$$

hat, also ein System von n Gleichungen mit n Unbekannten repräsentiert und die $\mathbf{u}_1, \ldots, \mathbf{u}_n$ linear unabhängig sind, – dann spannen sie den $\mathbb{R}^n$ auf, $\mathcal{L}(\mathbf{u}_1, \ldots, \mathbf{u}_n)$ ist kein Teilraum eines größeren Vektorraums mehr, sondern enthält eben *alle* Vektoren $\mathbf{v}$. Wieder gilt, dass im Falle der linearen Abhängigkeit nur dann eine Lösung existiert, wenn V ein Element des von den $\mathbf{u}_j$ erzeugten, entsprechenden Teilraums ist. Diese Betrachtungen sind eher intuitiv und werden in Abschnitt 3.4 über lineare Gleichungssysteme ausführlicher diskutiert.

Die Rolle von Skalarprodukten: In der Definition von Vektorräumen wird *nicht* gefordert, dass für irgendzwei Elemente eines Vektoraums V ein Skalarprodukt berechnet werden kann. Aus der Definition des Skalarprodukts in Abschnitt 1.2.3 folgt, dass für die in Definition 6.2 spezifizierten Vektorräume die Skalarprodukte für irgendzwei Vektoren existieren. Derartige Vektorräume sind Spezialfälle in der Menge der Vektorräume; zur leichten Referenz wird dafür eine Extradefinition gegeben:

Definition 1.9 *Ein Vektoraum V, in dem ein Skalarprodukt der Elemente definiert ist, heißt* Skalarproduktraum. *Ist der Vektorraum die Menge $\mathbb{R}^n$ der n-dimensionalen Vektoren, so heißt er auch* euklidischer Vektorraum.

Alle hier betrachteten (Teil-)Vektorräume sind Skalarprodukträume. Für Skalarprodukträume ist die Länge, die Orientierung, und der Winkel zwischen den Elementen definiert.

Auf Seite 20 wurde der Begriff der Norm eines Vektors eingeführt. Allgemein ist eine Norm[21] eine Abbildung eines Vektorraums in die Menge der nichtnegativen reellen Zahlen $\mathbb{R}_+$; sie wird mit $\|\cdot\|$ bezeichnet, wobei der Punkt $(\cdot\ )$ Platzhalter für einen Vektornamen ist:

$$\|\cdot\| : V \to \mathbb{R}_+, \quad \mathbf{x} \mapsto \|\mathbf{x}\| \in \mathbb{R} \text{ für } \mathbf{x} \in V. \tag{1.90}$$

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

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

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

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

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

d.h. aber

$$\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|, \tag{1.91}$$

also die Dreiecksungleichung. □

Definition 1.10 *Es sei V ein Vektorraum, dessen Elemente n-dimensionale Vektoren seien, und es seien $\boldsymbol{x}, \boldsymbol{y} \in V$. Dann heißt die Norm der Differenz*

$$\|\boldsymbol{x} - \boldsymbol{y}\| = \left(\sum_{i=1}^{n} (x_i - y_i)^2 \right)^{1/2} \tag{1.92}$$

die euklidische Distanz *zwischen $\boldsymbol{x}$ und $\boldsymbol{y}$, und V heißt auch* euklidischer Vektorraum *(vergl. Definition 1.9, Seite 39, und (6.2), Seite 171).*

$\|\mathbf{x} - \mathbf{y}\|$ ist ein Vektor, dessen Länge der Distanz $d_e(x, y)$ zwischen den Endpunkten x und y der Vektoren $\mathbf{x}$ und $\mathbf{y}$ entspricht; $d_e(x, y)$ ist euklidisch, weil der Vektor $\mathbf{x} - \mathbf{y}$ die Distanz zwischen den Endpunkten von $\mathbf{x}$ und $\mathbf{y}$ durch eine Gerade repräsentiert[22], und definiert die euklidische Metrik. Der Index e in der

[21] Von lat. *norma* = Richtschnur

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

Bezeichnung d_e soll anzeigen, dass die euklidische Distanz gemeint ist, im Unterschied zur Distanz $d_p(x,y)$, der Minkowski-Distanz (6.3) (Seite 172). Es sei noch einmal darauf hingewiesen, dass eine Distanz über ein Skalarprodukt definiert ist: es sei

$$d_i = |x_i - y_i|, \quad i = 1, \ldots, n \tag{1.93}$$

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

$$\|\mathbf{d}\| = (\mathbf{d}'\mathbf{d})^{1/2} = \|\mathbf{x} - \mathbf{y}\| = d_e(x,y). \tag{1.94}$$

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

$$\mathbf{d}_p'\mathbf{d}_p = \sum_{i=1}^n d_i^{p/2} d_i^{p/2} = \sum_{i=1}^n d_i^p = \sum_{i=1}^n |x_i - y_i|^p, \quad p > 1 \tag{1.95}$$

und

$$d_p(x,y) = (\mathbf{d}_p'\mathbf{d}_p)^{1/p} = \left(\sum_{i=1}^n |x_i - y_i|^p\right)^{1/p}, \quad p > 1 \tag{1.96}$$

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 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*

$$L\colon \mathbf{x} = \mathbf{u} + s\mathbf{v}, \quad s \in \mathbb{R}, \ \mathbf{x}, \mathbf{u}, \mathbf{v} \in \mathbb{R}^2 \tag{1.97}$$

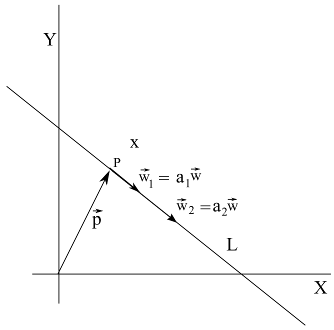
der Geraden. Hier ist $\mathbf{u}$ der *Ortsvektor* und $\mathbf{v}$ der *Richtungsvektor* der Geraden. ist s. Abbildung 13. Es sei angemerkt, dass die Gerade L und der Vektor $\mathbf{u}$ *keinen* rechten Winkel bilden müssen. Es sei $\mathbf{n} = (n_1, n_2)'$ ein Vektor, der senkrecht auf der Geraden 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 (1.97)

$$\mathbf{n}'\mathbf{x} = \mathbf{n}'\mathbf{u} + s\mathbf{n}'\mathbf{v} = \mathbf{n}'\mathbf{u}. \tag{1.98}$$

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

$$n_1x_1 + n_2x_2 - d = 0, \quad d = \mathbf{n}'\mathbf{u}. \tag{1.99}$$

Abbildung 13: Gerade in Parameterform



dies ist die zu (1.97) 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 (1.97) 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 (1.97).

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

$$\mathcal{E} = \{\mathbf{x}|\ \mathbf{x} = \mathbf{u} + s\mathbf{v}_1 + t\mathbf{v}_2, \quad \mathbf{x}, \mathbf{u}, \mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^3,\ s, t \in \mathbb{R}\} \tag{1.100}$$

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 nichtparallele *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

$$a_1\mathbf{n}'\mathbf{v}_1 + a_2\mathbf{n}'\mathbf{v}_2 = \mathbf{n}'\mathbf{v} = 0, \tag{1.101}$$

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

Nun folgt aus (1.100)

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

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

$$\mathbf{n}'(\mathbf{x} - \mathbf{u}) = 0, \tag{1.102}$$

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 (1.102)

$$n_1x_1 + n_2x_2 + n_3x_3 - d = 0 \tag{1.103}$$

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 $\mathbf{x} \in \mathbb{E}_0$ 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. $\underset{\sim}{\mathbf{x}}$ 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:

$$\begin{aligned} \mathbf{x}_1 &= s_1\mathbf{v}_1 + t_1\mathbf{v}_2 \\ \mathbf{x}_2 &= s_2\mathbf{v}_1 + t_2\mathbf{v}_2 \\ \mathbf{x} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 &= (c_1s_1 + c_2s_2)\mathbf{n}'\mathbf{v}_1 + (c_1t_1 + c_2t_2)\mathbf{n}'\mathbf{v}_2 \end{aligned} \tag{1.104}$$

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 für 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 1.11 *Es sei V ein Vektorraum und $M = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p\} \subset V$ eine Menge von Vektoren aus V.*
1. Es sei

$$\mathcal{L}(M) = \{\boldsymbol{x} | \boldsymbol{x} = a_1\boldsymbol{x}_1 + \cdots + a_p\boldsymbol{x}_p\}, \quad \mathcal{L}(\emptyset) = \vec{0}, \; a_1, \ldots, a_p \in \mathbb{R} \tag{1.105}$$

d.h. $\mathcal{L}(M)$ sei die Menge aller Linearkombinationen der Vektoren aus M.[23] *$\emptyset$ 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 $\boldsymbol{x}_j \in M$ existiert derart, dass*[24] *$M \setminus \{\boldsymbol{x}_j\}$ kein Erzeugendensystem für V mehr ist.*

Anmerkung: Die lineare Hülle $\mathcal{L}(M)$ ist die Menge *aller* Linearkombinationen der $\mathbf{x}_1, \ldots, \mathbf{x}_p$, das heißt die Koeffizienten a_k, $1 \leq k \leq p$ durchlaufen jeweils alle 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 Cantor[25] 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 $M = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p\} \subset V$. Dann gilt $\mathcal{L}(M) \subseteq V$. Für $\mathcal{L}(M) \subset V$ ist $\mathcal{L}(M)$ ein Teilvektorraum von V.*

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

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

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

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

[23]Statt des Ausdrücks 'lineare Hülle' ist auch der aus dem Englischen übernommene Audruck 'span' gebräuchlich, der zum deutschen Ausdruck 'aufspannen' korrespondiert.
[24]Sei $\mathbf{x}_j \in M$. $M \setminus \{\mathbf{x}_j\}$ ist die Menge M ohne den Vektor $\mathbf{x}_j$.
[25]Georg Cantor (1845 – 1918), Mathematiker

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 $\mathcal{L}(\mathcal{M})$ nur ein Teilvektorraum (einfach: Teilraum) von V sein kann. □

Definition 1.12 *Es sei $U \subset \mathbb{R}^n$ ein Teilraum des $\mathbb{R}^n$ und es sei $\mathcal{B} = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_r\} \subset U$. $\mathcal{B}$ heißt* Basis *von U, wenn gilt*

(i) die Vektoren $\boldsymbol{u}_1, \dots, \boldsymbol{u}_r$ sind linear unabhängig,

(ii) die $\boldsymbol{u}_1, \dots, \boldsymbol{u}_r$ bilden ein minimales Ereugendendsystem (s. Defintion 1.11, 3.),

(iii) $U = \mathcal{L}(\mathcal{B})$, dh. U ist die lineare Hülle der $\boldsymbol{u}_1, \dots, \boldsymbol{u}_r$, d.h. U wird von $\mathcal{B}$ erzeugt.

Anmerkungen:
1. $\mathcal{B}$ erzeugt U im Sinne von Punkt 2. der Definition 1.11.
2. 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

$$\mathbf{x} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_r\mathbf{u}_r; \tag{1.106}$$

die Koeffizienten a_j heißen die *Komponenten von $\boldsymbol{x}$ in Bezug auf die Basis $\mathcal{B}$*.
3. Wie bereits gezeigt wurde (Gleichung 1.20, Seite 15) kann jeder Vektor $\mathbf{x} \in \mathbb{R}^n$ als Linearkombination

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n \tag{1.107}$$

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 Satzs 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 $A_j = A \backslash \mathbf{a}_j$ die Menge, die aus A entsteht, wenn der Vektor $\mathbf{a}_j$ aus A entfernt wird. 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. 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 (1.106) sind gemäß dem folgenden allgemeinen Satz eindeutig bestimmt:

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

Beweis: Es werde angenommen, dass es zwei verschiedene Darstellungen von $\mathbf{y}$ als Linearkombination der $\mathbf{x}_1, \ldots, \mathbf{x}_p$ gibt:

$$\mathbf{y} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_p\mathbf{x}_p \tag{1.108}$$

$$\mathbf{y} = b_1\mathbf{x}_1 + b_2\mathbf{x}_2 + \cdots + b_p\mathbf{x}_p \tag{1.109}$$

Subtrahiert man die Gleichung (1.109) von (1.108), so erhält man die Gleichung

$$\vec{0} = (a_1 - b_1)\mathbf{x}_1 + (a_2 - b_2)\mathbf{x}_2 + \cdots + (a_p - b_p)\mathbf{x}_p \tag{1.110}$$

Nach Voraussetzung sind die $\mathbf{x}_1, \ldots, \mathbf{x}_p$ linear unabhängig, so dass die Darstellung (1.110) des Nullvektors $\vec{0}$ nur möglich ist, wenn

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

gilt, d.h. die Koeffizienten in den Gleichungen (1.108) und (1.109) sind identisch, d.h. die Darstellung von $\mathbf{y}$ als Linearkombination der $\mathbf{x}_1, \ldots, \mathbf{x}_p$ ist eindeutig. □

Wie in Punkt 3. oben erwähnt wurde heißen die Koeffizienten a_j in (1.106) "Komponenten von $\mathbf{x}$ in Bezug auf die Basis $\mathbf{u}_1, \ldots, \mathbf{u}_r$". 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 $a_1, \ldots, a_r$ 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 1.12 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. (1.101), Seite 43 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 (vergl. Satz 1.9, Seite 45). Dementsprechend kann man vermuten, dass aus der Darstellung (1.107) eines beliebigen Vektors $\mathbf{x} \in \mathbb{R}^n$ durch die kanonische Basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ 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

$$\mathbf{x} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 \in E$$

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 $\mathbf{u}_3$ als Linearkombination von $\mathbf{u}_1$ und $\mathbf{u}_2$ gewonnen werden kann, d.h. die $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ können nicht linear unabhängig sein. Damit also die $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ linear unabhängig sein können, muß $\mathbf{u}_3 \notin E$ gelten. Für $a_3 \neq 0$ kann dann aber $\mathbf{x}$ nicht mehr in E liegen, denn $\mathbf{u}_3$ erzwingt 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 kann erart, 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 R^3 jeweils $r = 3$ linear unabhängige Vektoren als Basis genügen. □

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 (*Steinitzsches*[26] *Lemma*, auch *Austauschsatz von Steinitz*) bewiesen:

[26] Ernst Steinitz (1871 – 1928), Mathematiker

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

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

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

$$c_{k1}\mathbf{u}_k + c_2\mathbf{b}_2 + \cdots + c_m\mathbf{b}_m = \vec{0} \tag{1.111}$$

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 $c_2\mathbf{b}_2 + \cdots + c_m\mathbf{b}_m = \vec{0}$ gelten und wegen der postulierten linearen Abhängigkeit der $\mathbf{u}_k, \mathbf{b}_2, \dots, \mathbf{b}_m$ dürfen nicht alle Koeffizienten c_j gleich Null sein, entgegen der Voraussetzung der linearen Unabhängigkeit der $\mathbf{b}_j$. Dann aber impliziert (1.111), dass $\mathbf{u}_k$ als Linearkombination der $\mathbf{b}_j$ dargestellt werden kann:

$$\mathbf{u}_k = -\frac{1}{c_{k1}}(c_2\mathbf{b}_2 + \cdots + c_m\mathbf{b}_m) = d_{k2}\mathbf{b}_2 + \cdots + d_{km}\mathbf{b}_m, \quad d_{kj} = -\frac{c_j}{c_{k1}}. \tag{1.112}$$

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

$$\mathbf{b}_1 = \sum_{k=1}^{m} g_k\mathbf{u}_k = \sum_{k=1}^{m} g_k \sum_{j=2}^{r} d_{kj}\mathbf{b}_j = \sum_{j=2}^{r}\sum_{k=1}^{m} g_k d_{kj}\mathbf{b}_j. \tag{1.113}$$

$\mathbf{b}_1$ 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. □

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

Beweis:[27] (i) Es sei $U = \mathbb{R}^n$ und $\mathcal{B} = \{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ sei eine Basis von U. 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 45). 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

[27]Dieser 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.

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

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

ist nach Voraussetzung linear unabhängig. So kann man fortfahren, bis die ersten n Vektoren $\mathbf{u}_1, \ldots, \mathbf{u}_n$ durch die Einheitsvektoren $\mathbf{e}_{m_1}, \mathbf{e}_{m_2}, \ldots, \mathbf{e}_{m_n}$ ersetzt worden sind. Aber $\{\mathbf{e}_{m_1}, \mathbf{e}_{m_2}, \ldots, \mathbf{e}_{m_n}\}$ 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 $\{\mathbf{e}_{m_1}, \ldots, \mathbf{e}_{m_n}, \mathbf{u}_{n+1}\}$ 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.

(ii) Für $U = \mathbb{R}^n$ kann nicht $r < n$ gelten. Denn es sei $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ eine Menge linear unabhängiger Vektoren mit $r < n$. Wiederholte Anwendung von Satz 1.11 führt auf $\{\mathbf{e}_{m_1}, \ldots, \mathbf{e}_{m_r}\}$ mit $m_r < n$, und $\mathbf{x} \in U$ kann nicht vollständig als Linearkombination der $\mathbf{e}_{m_1}, \ldots, \mathbf{e}_{m_r}$ 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 $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_r\}$ für $\mathbb{R}^n$ mit $r > n$. Dies impliziert, dass jede Menge[28] von Vektoren $M = \{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ 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 $p < n$[29]. Es sei $\mathcal{L}_p = \mathcal{L}(\mathbf{x}_1, \ldots, \mathbf{x}_p)$ die lineare Hülle der $\mathbf{x}_1, \ldots, \mathbf{x}_p$. $\mathcal{L}_p$ 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 $\mathcal{L}_p$ und es existiert keine Basis $\{\mathbf{b}_1, \ldots, \mathbf{b}_r\} \in \mathcal{L}$ mit $r > p$, denn es gibt ja nur die p Vektoren $\mathbf{x}_k$, von denen nur maximal p von ihnen linear unabhängig sein können. deshalb gilt $r \leq p$ und die $\mathbf{x}_k$ bilden somit eine Basis für $\mathcal{L}_p$. Es 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, so ist $r = 1$ 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 $r = 2 < p$ (vergl. Beispiel 1.8, Seite 42).

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

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

Satz 1.13 *Es[30] 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*

$$U = \mathcal{L}(M), \tag{1.114}$$

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 $U = \mathcal{L}(M) \subseteq W$.

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. □

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

Definition 1.13 *Es sei $V = \{\boldsymbol{x} | \boldsymbol{x} \in \mathbb{R}^n\}$, $\mathcal{B}_r = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_r\} \subset V$, und die $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_r$ seien linear unabhängig. Für $r < n$ heißt $\mathcal{L}(\mathcal{B}_r)$* r-dimensionaler Teilraum *von V; er 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 $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r$ auf V gegeben. Dann existieren weitere linear unabhängige Vektoren $\boldsymbol{w}_{r+1}, \ldots, \boldsymbol{w}_n$ derart, dass*

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

eine Basis von V ist.

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

Beweis: Es sei $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ ein Erzeugendensystem. Es enthält eine Basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$, $r \leq n$. Nach dem Steinitzschen Lemma (1.11), 48, kann man dann

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

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 $\mathcal{B}_r = \{\mathbf{w}_1, \ldots, \mathbf{w}_r\}$, $r \leq n$ gegeben. Dann existiert stets eine Basis $\mathcal{B}$, die $\mathcal{B}_r$ enthält. $\square$

Anmerkung: Die vorangeganenen 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 1.14 [31] *Es seien $A_1, \ldots, A_k$, die Endpunkte der linear unabhängigen Vektoren $\boldsymbol{a}_j \in \mathbb{R}^n$, $j = 1, \ldots, k$ mit Anfangspunkt 0. Dann ist*

$$E_k^n = \{\boldsymbol{y} | \boldsymbol{y} = \sum_{j=1}^{k} b_j \boldsymbol{a}_j; \quad b_j \in \mathbb{R}, j = 1, \ldots, k\} \tag{1.115}$$

eine k-dimensionale Ebene *im $\mathbb{R}^n$ durch den Urspung 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 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 1.12). 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,

[31] Graybill (1983), p. 68

die durch Geraden mit verschiedener Orientierung in dieser Ebene repräsentiert werden.

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

$$\mathbf{n}'\mathbf{x} = a_1\mathbf{n}'\mathbf{x}_1 + a_2\mathbf{n}'\mathbf{x}_2 = 0, \tag{1.116}$$

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 $n > 2$, vergl. Beispiel 1.8, Seite 42. $\square$

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

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

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

$$\boldsymbol{b}_j'\boldsymbol{b}_k = \begin{cases} 0, & j \neq k \\ 1, & j = k \end{cases}, \quad j, k = 1, \ldots, n \tag{1.117}$$

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

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

Satz 1.15 *Die n-dimensionalen Einheitsvektoren $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ mit*

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

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

Beweis: Die Einheitsvektoren sind linear unabhängig, denn $\vec{0} = \lambda_1 \mathbf{e}_1 + \lambda_2 \mathbf{e}_2 + \cdots + \lambda_n \mathbf{e}_n$ ist nur möglich für $\lambda_1 = \cdots = \lambda_n = 0$; für die i-te Komponente hat man nämlich $0 = \lambda_i 1$, woraus sofort $\lambda_i = 0$ folgt. Darüber hinaus sind die $\mathbf{e}_i$ orthonormal, vergl. (1.107). Die Vektoren $\mathbf{e}_1, \ldots, \mathbf{e}_n$ bilden deshalb eine orthonormale Basis des V_n. Da stets

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

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

Definition 1.16 *Die Basis* $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ *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 79.

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

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

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

Nach Satz 1.5, Seite 33, kann jeder Vektor $\mathbf{x} = (x_1, \ldots, x_n)' \in V$ als Linearkombination der n-dimensionalen Einheitsvektoren $\mathbf{e}_1, \ldots, \mathbf{e}_n$ dargestellt werden, und damit können auch die Vektoren der Basis $\mathcal{B}$ als Linearkombinationen der $\mathbf{e}_1, \ldots, \mathbf{e}_n$ dargestellt werden. Es sei x_{kj} die k-te Komponente von $\mathbf{x}_j \in \mathcal{B}$. Dann hat man

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

Die gesuchten Koordinaten sind also durch

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

gegeben. □

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

Satz 1.16 *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

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

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

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

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

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

$$\mathbf{x} = a_1 \mathbf{b}_1 + \cdots + a_n \mathbf{b}_n = \sum_{k=1}^{n} a_k \mathbf{b}_k. \tag{1.118}$$

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

$$\mathbf{b}_j' \mathbf{x} = \sum_{k=1}^{n} a_k \mathbf{b}_j' \mathbf{b}_k = a_j, \quad j = 1, \dots, n \tag{1.119}$$

denn

$$\mathbf{b}_j' \mathbf{b}_k = \begin{cases} 0, & j \neq k \\ 1, & j = k \end{cases} \tag{1.120}$$

(1.118) kann dann in der Form

$$\mathbf{x} = \sum_{k=1}^{n} (\mathbf{b}_k'\mathbf{x})\mathbf{b}_k. \tag{1.121}$$

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 (1.121) 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 (1.121). □

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

Beweis:[32] 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

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

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. □

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

Orthogonale Teilräume

Satz 1.18 *Es sei $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ eine orthonormale Basis eines n-dimensionalen Vektorraums V, und es sei $\mathcal{B}_k = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$, $k < n$, und die lineare Hülle $\mathcal{L}_k = \mathcal{L}(\mathcal{B}_k)$ definiert einen Teilraum von V (Satz 1.8, Seite 44). Dann existiert ein Vektor $\boldsymbol{n} \in V$, der orthogonal zu allen Vektoren aus $\mathcal{L}_k$ ist.*

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

Anmerkung: Der Vektor $\mathbf{n}$ heißt auch *Normalenvektor*, vergl. die Gleichung (1.98), Seite 41.

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

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

Dann folgt

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

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

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

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

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

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. □

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 1.17 *Es sei V ein n-dimensionaler Vektorraum und $U \subset V$ sei ein Teilraum von V. Weiter sei*[33]

$$U^{\perp} = \{\boldsymbol{x} \in V | \boldsymbol{x}'\boldsymbol{u} = 0, \forall \boldsymbol{u} \in U\} \tag{1.122}$$

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. □

[33] $\forall$ steht für "für alle"

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. □

Kombination von Teilräumen:

Definition 1.18 *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 $\boldsymbol{w} \in W$, $\boldsymbol{w} = \boldsymbol{u} + \boldsymbol{v}$, $\boldsymbol{u} \in U$, $\boldsymbol{v} \in V$, und
(ii) $\boldsymbol{u}'\boldsymbol{v} = 0$ für alle $\boldsymbol{u} \in U$, $\boldsymbol{v} \in V$ gilt.

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 $\mathbf{v} = a_3\mathbf{n} \in V$. □

Satz 1.20 *(Dimensionalitätssatz)*[34] *Es seien $U, V \subset W$, W ein n-dimensionaler Vektorraum. Die folgenden Aussagen (i) und (ii) sind äquivalent:*
(i) V ist das orthogonale Komplement von U, d.h. $V = U^\perp$,
(ii) $W = U \perp V$, d.h. W ist die orthogonale Summe von U und V. Dann gilt

$$\dim W = \dim U + \dim V^\perp. \tag{1.123}$$

Beweis: (i) ⇒(ii): Zu zeigen ist $W = U \perp U^\perp$. Es sei $\{\mathbf{b}_1, \ldots, \mathbf{b}_r\}$ eine orthonormale Basis des Teilraums U; sie kann nach Satz 1.17 zu einer orthonormalen Basis $\{\mathbf{b}_1, \ldots, \mathbf{b}_r, \mathbf{b}_{r+1}, \ldots, \mathbf{b}_n\}$ von W erweitert werden. Nach Definition von $U^\perp$ ist $U^\perp = \mathcal{L}(\mathbf{b}_{r+1}, \ldots, \mathbf{b}_n)$, und da $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ eine Basis von W ist, folgt Aussage (i). Da die $\mathbf{b}_j$, $j = 1, \ldots, n$ paarweise orthonormal sind, folgt die Aussage (ii) und damit (1.123).

(ii) ⇒ (i): Nach (ii) der Definition 1.18 gilt $\mathbf{u}'\mathbf{v} = 0$ für $W = U \perp V$; Sei $\mathbf{x} \in W$. Dann $\mathbf{x} = \mathbf{u} + \mathbf{v}$ mit $\mathbf{u} \in U$ und $\mathbf{v} \in V$. Dann $\mathbf{x}^\perp$ genau dann, wenn $\mathbf{x}'\mathbf{w} = 0$ für alle $\mathbf{w} \in U$, also $\mathbf{u}'\mathbf{w} = 0$ für alle $\mathbf{u} \in U$, d.h $\mathbf{u} = \vec{0}$. □

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 $V \subseteq U$.

[34] Koecher, M.: Lineare Algebra und analytische Geometrie. Berlin 1997, p. 160

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

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

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

Definition 1.19 *Es seien U und V Teilräume eines Vektorraums W. Dann heißt*

$$U + V = \{\boldsymbol{u} + \boldsymbol{v} | \boldsymbol{u} \in U, \, \boldsymbol{v} \in V\} \tag{1.124}$$

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. □

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

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

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

$$\operatorname{rg}(S) = \operatorname{rg}(\mathcal{L}(\mathbf{u}_1, \ldots, \mathbf{u}_r))) = r \leq p. \tag{1.125}$$

□

Satz 1.23 *Für die Summe der Teilräume gilt die* Dimensionsformel

$$\dim(U + V) = \dim(U) + \dim(V) - \dim(U \cap V). \tag{1.126}$$

Beweis:[35] Die Dimension eines (Teil-)Vektorraums ist gleich der Anzahl der Basisvektoren des Raums. $U \cap V$ ist ein Teilraum und habe die Basis $\mathbf{a}_1, \ldots, \mathbf{a}_m$, also $m = \dim(U \cap V)$. Diese Basis kann zu Basen von einerseits U, andererseits V ergänzt werden: $\mathbf{a}_1, \ldots, \mathbf{a}_m, \mathbf{b}_1, \ldots, \mathbf{b}_r$ sei die Basis von U, also $\dim(U) = m + r$, und $\mathbf{a}_1, \ldots, \mathbf{a}_m, \mathbf{c}_1, \ldots, \mathbf{c}_s$ sei die Basis von V, also $\dim(V) = m + 2$, und $\dim(U \cap V) = m$. Dann bildet

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

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

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

Zunächst muss gezeigt werden, dass auch

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

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

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

[35] Nach Lorenz I, p. 56

nur dann, wenn die Koeffizienten λ_i, μ_i, ν_i alle gleich Null sind. Es ist $\mathbf{a} \in U \cap V$, $\mathbf{b} \in U$, $\mathbf{c} \in V$. Dann folgt aus der vorangehenden Gleichung $\mathbf{b} = -\mathbf{a} - \mathbf{c} \in U \cap V$. Aus der Wahl der $\mathbf{b}_1, \dots, \mathbf{b}_r$ 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 1.2 *Es seien U und V Teilräume eines Vektorraums W. Aus (1.126) folgt*

$$\dim(U + V) \leq \dim(U) + \dim(V), \tag{1.127}$$

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

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

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

Dabei sind die 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

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

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 $1 \le i \le m$, $1 \le j \le n$. 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

$$X' = X, \quad x_{ij} = x_{ji}, \;\; 1 \le i, j \le n \tag{2.3}$$

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

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

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

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

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

Fasst man (1) und (2) zusammen, so erhält man

$$Z = \lambda X + \mu Y, \quad \lambda, \mu \in \mathbb{R} \tag{2.6}$$

mit Z, X und Y (m, n)-Matrizen als Verallgemeinerung des Begriffs der Linearkombination von Vektoren.

2.2.2 Multiplikation einer Matrix mit einem Vektor

Gegeben sei eine (m, n)-Matrix X mit den Spaltenvektoren $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ und den Zeilenvektoren $\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_m$, und $\mathbf{a} \in \mathbb{R}^n$ sei ein Vektor aus $\mathbb{R}^n$. . Das Produkt von X mit dem Spaltenvektor $\mathbf{a}$ *von rechts* steht dann für eine Linearkombination der Spaltenvektoren von X

$$\begin{aligned} \mathbf{y} &= X\mathbf{a} & (2.7)\\ &= a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n. & (2.8) \end{aligned}$$

Die Gleichung (2.8) verdeutlicht, dass die Anzahl der Komponenten von $\mathbf{a}$ stets gleich der Anzahl von Spalten von X sein muß, damit $X\mathbf{a}$ ein sinnvoller Ausdruck ist. Die Anzahl der Komponenten von $\mathbf{y}$ ist stets gleich der Anzahl der Zeilen von X.

Ein Produkt von X *von links* mit einem *Spalten*vektor $\mathbf{b}$ ist allerdings nicht definiert: $\mathbf{b}$ eine 1-spaltige Matrix, und man müsste analog zu den Gleichngen (2.7) und (2.8) $\mathbf{b}X$ als Linearkombination von $\mathbf{b}$ mit Koeffizienten aus der Matrix X interpretieren können, – was offenbar keinen Sinn macht. Was aber betrachtet werden kann ist das analog zu (2.7) definierte Produkt

$$\begin{aligned} \mathbf{z} &= X'\mathbf{b} & (2.9)\\ &= b_1\tilde{\mathbf{x}}_1 + \cdots + b_m\tilde{\mathbf{x}}_m. & (2.10) \end{aligned}$$

Das Produkt der transponierten Matrix X' von rechts mit einem Vektor $\mathbf{b}$ entspricht einer Linearkombination der *Zeilen*vektoren von X. Aus der weiter unten bewiesenen Gleichung (2.15), Seite 64, derzufolge für die Transponierte des Produkts AB zweier Matrizen A und B die Beziehung $(AB)' = B'A'$ gilt, folgt dann die Gleichung

$$\mathbf{z}' = (X'\mathbf{b})' = \mathbf{b}'X. \tag{2.11}$$

$\mathbf{b}$ wird wieder als 1-spaltige Matrix aufgefasst, wobei die Anzahl der Komponenten $\mathbf{b}$ nun gleich der Anzahl m von Zeilen von X ist. Dies heißt, dass die Multiplikation einer Matrix X *von links* mit einem Vektor erklärt ist, wenn der Vektor als Zeilenvektor angeschrieben wird. Das Resultat ist ein Zeilenvektor $\mathbf{z}'$, der sich als Linearkombination der Zeilenvektoren von X ergibt. Letzlich wird die Definition von $\mathbf{b}'X$ auf die Definition (2.8) des Produkts $\mathbf{y} = X\mathbf{a}$ zurückgeführt.

Zusammenfassend hat man die Aussagen

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: Transformation und Abbbildung 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 6.12 ausführlicher eingegangen wird. □

Zwei spezielle Transformationen spielen in der multivariaten Statistik eine besondere Rolle: es seien $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.
1. Der Vektor $\mathbf{x}$ wird auf den Vektor $\mathbf{y}$ abgebildet, der sich von $\mathbf{x}$ nur durch seine Orientierung unterscheidet, d.h. die Transformation ist eine Rotation um einen Winkel θ oder
2. Der Vektor $\mathbf{x}$ wird auf den Vektor $\mathbf{y}$ abgebildet und $\mathbf{y}$ unterscheidet sich von $\mathbf{x}$ nur durch seine Länge, d.h. die Transformation ist eine Längenskalierung.

Sei M eine (m, n)-Matrix und gilt $\mathbf{y} = M\mathbf{x}$, so ist $\mathbf{y} \in \mathbb{R}^m$. Da in den beiden betrachteten Fällen $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, folgt, dass $m = n$ gelten muß, d.h. M muß quadratisch sein. Im Fall 1. muß

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

gelten . In Abschnitt 3.1 wird gezeigt werden, dass dann $M'M = I_n$ gilt, I_n die (n, n)-Einheitsmatrix, d.h. verschiedene Spaltenvektoren von M sind orthogonal und alle Spalgtenvektoren haben die Länge 1.

Im Fall 2. gilt für $\mathbf{y} = M\mathbf{x}$, dass $\mathbf{y} = \lambda\mathbf{x}$, λ ein Skalar, und damit $\|\mathbf{y}\| = \lambda\|\mathbf{x}\|$; $\mathbf{y}$ und $\mathbf{y}$ unterscheiden sich in der Länge und nicht in Bezug auf die Orientierung, und wieder gilt $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, woraus wieder folgt, dass M quadratisch sein muß. Es sind nur spezielle, für M charakteristische Vektoren, die der Beziehung $M\mathbf{x} = \lambda\mathbf{x}$ genügen, weshalb die Vektoren $\mathbf{x}$, die dieser Bedingung genügen, *Eigenvektoren* oder auch *charakteristische Vektoren* von M heißen, und λ ist der jeweils zugehörige *Eigenwert* von M. Eigenvektoren und Eigenwerte werden in Abschnitt 3.2.1 eingeführt.

2.2.3 Multiplikation einer Matrix mit einer Matrix

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

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

d.h. man erhält die Matrixgleichung

$$C = AB. \tag{2.12}$$

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

$$c_{ij} = \tilde{\mathbf{a}}_i'\mathbf{b}_j, \quad i = 1, \ldots, m; \; j = 1, \ldots, p \tag{2.13}$$

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 62.

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,

$$U = AB \neq V = BA. \tag{2.14}$$

denn die Spaltenvektoren $\mathbf{u}_j \in \mathbb{R}^m$ von U sind Linearkombinationen der Spaltenvektoren von A, während die Spaltenvektoren $\mathbf{v}_k \in \mathbb{R}^n$ von V Linearkombinationen der Spaltenvektoren von B sind. Für $m \neq n$ sind die $\mathbf{u}_j$ und $\mathbf{v}_k$ sogar Elemente von Vektorräumen verschiedener Dimensionalität. Es gibt allerdings wichtige Ausnahmen, die insbesondere die Identität $m = n$ voraussetzen.

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*

$$C' = (AB)' = B'A' \tag{2.15}$$

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

$$(AB)C = A(BC) \quad \textit{(Assoziativität)} \tag{2.16}$$

Beweis: Zu (2.15): Mit $C = AB = [\mathbf{c}_1, \mathbf{c}_2, \ldots \mathbf{x}_n]$, 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 (2.16): Es sei $U = AB$, $V = BC$, so dass $F = (AB)C = UC$, $G = A(BC) = AV$. U ist eine $(m \times p)$-Matrix, V ist eine $(n \times q)$-Matrix, und F und G sind beide $(m \times q)$-Matrizen. Es seien $\mathbf{f}_j$ und $\mathbf{g}_j$ die j-ten Spaltenvektoren von F bzw. G. Offenbar gilt

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

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

Anmerkung: Nach Anmerkung 2, Seite 64, gilt für das Produkt $C = AB$, dass die Zeilenvektoren von C Linearkombinationen der Zeilenvektoren von B sind. Die Gleichung (2.15), d.h. $C' = B'A'$, drückt diesen Sachverhalt in kompakter Form aus: die Zeilenvektoren von C sind ja gerade die Spaltenvektoren von C', die sich gemäß der Anmerkung 1 als Linearkombinationen der Spaltenvektoren von B' ergeben, – also den Zeilenvektoren von B. . □

Die in (2.16) ausgedrückte Eigenschaft der Matrixmultiplikation wird als *assoziativ* bezeichnet. Sie ist aus der Multiplikation reeller Zahlen bekannt: Sind $0 \neq a, b, c \in \mathbb{R}$, so gilt $(ab)c = a(bc)$. Die Multiplikation reeller Zahlen ist auch *kommmutativ*, d.h. es gilt stets auch $ab = ba$.

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

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}, \quad \tilde{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_m \end{pmatrix} \tag{2.17}$$

Man schreibt dafür auch kurz

$$\Lambda_n = \mathrm{diag}(\lambda_1, \ldots, \lambda_n), \quad \Lambda_m = \mathrm{diag}(\lambda_1, \ldots, \lambda_m) \tag{2.18}$$

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:

$$X\Lambda_n = \begin{pmatrix} \lambda_1 x_{11} & \lambda_2 x_{12} & \cdots & \lambda_n x_{1n} \\ \lambda_1 x_{21} & \lambda_2 x_{22} & \cdots & \lambda_n x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 x_{m1} & \lambda_2 x_{m2} & \cdots & \lambda_n x_{mn} \end{pmatrix}, \tag{2.19}$$

bzw.

$$\Lambda_m X = \begin{pmatrix} \lambda_1 x_{11} & \lambda_1 x_{12} & \cdots & \lambda_1 x_{1n} \\ \lambda_2 x_{21} & \lambda_2 x_{22} & \cdots & \lambda_2 x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_m x_{m1} & \lambda_m x_{m2} & \cdots & \lambda_m x_{mn} \end{pmatrix}. \tag{2.20}$$

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,

$$X = [\mathbf{x}_1, \ldots, \mathbf{x}_n],\ \mathbf{x}_j \in \mathbb{R}^m, \quad X' = [\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_m],\ \tilde{\mathbf{x}}_i \in \mathbb{R}^n$$

Es wird $n \leq m$ vorausgesetzt, der Fall $n > m$ kann in analoger Weie behandelt werden. Die Spaltenvektoren von X sind Elemente des $\mathbb{R}^m$, die Zeilenvektoren von X sind aus $\mathbb{R}^n$. Die lineare Hälle $\mathcal{L}(X) = \mathcal{L}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ heißt *Spaltenraum* von X, und die lineare Hülle $\mathcal{L}(X') = \mathcal{L}(\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_m)$ heißt *Zeilenraum* von X. Ein Frage von theoretischem und praktischem Interesse ist, ob Zeilen- und Spaltenraum verschiedene Dimensionen haben können oder nicht.

Nach Satz 1.8, Seite 44, ist die lineare Hülle $\mathcal{L}(X)$ ein Teilvektorraum des $\mathbb{R}^m$ und nach Satz 1.9, Seite 45, existiert stets eine Basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_s\}$ für $\mathcal{L}(X)$, mit $s \leq n$, d.h. die Vektoren $\mathbf{x}_j$ können als Linearkombinationen der $\mathbf{u}_k$ dargestellt werden. Fasst man die $\mathbf{u}_k$ spaltenweise zu einer (m,s)-Matrix U_s zusammen, so kann man

$$X = U_s A_s, \quad \mathbf{x}_j = U_s \mathbf{a}_j, \quad j = 1, \ldots, n \tag{2.21}$$

schreiben, wobei A_s eine (s,n)-Matrix ist, deren Spalten $\mathbf{a}_j$ die Koeffizientenvektoren für die Darstellung der $\mathbf{x}_j$ als Linearkombinationen der $\mathbf{u}_k$ sind.

Eine analoge Betrachtung gilt für die Spaltenvektoren $\tilde{\mathbf{x}}_i$ von X', d.h. es existiert eine Basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ von $\mathcal{L}(X')$, so dass, wenn die $\mathbf{v}_j$ spaltenweise zu einer Matrix V_r zusammengefasst werden,

$$X' = V_r B_r, \quad \tilde{\mathbf{x}}_i = V_r \mathbf{b}_i, \quad i = 1, \ldots, m \tag{2.22}$$

gilt. V_r ist eine (m,r)-Matrix, und B_r ist eine (r,m)-Matrix, deren Spaltenvektoren $\mathbf{b}_i$ die Koeffizientenvektoren für die Darstellung der $\tilde{\mathbf{b}}_i$ sind. Dann folgt[36]

$$\begin{aligned} X &= U_s A_s = B_r' V_r' \qquad (2.23)\\ X' &= V_r B_r = A_s' U_s' \qquad (2.24) \end{aligned}$$

d.h. die Spaltenvektoren von X sind Linearkombinationen sowohl der Spaltenvektoren von U_s wie von B_r', und die Spaltenvektoren von X', d.h. die Zeilenvektoren von X sind Linearkombinationen der Spaltenvektoren von V_r wie der Spaltenvektoren von A_s. Nach Voraussetzung sind die Spaltenvektoren von V_r, wie auch die Spaltenvektoren von U_s, linear unabhängig, – es ist allerdings noch nicht klar,

[36] s. Abschnitt 1.3.2 und Satz 2.1, Seite 64

ob auch die *Zeilen*-vektoren von V_r und U_s, also die Spaltenvektoren von V_r' und U_s' ebenfalls linear unabhängig sein müssen; dass sie es sind, wird weiter unten gezeigt. Weiter sei daran erinnert, dass die Basen U_s und V_r *minimale* Erzeugendensysteme sind, d.h. s ist die kleinste Anzahl linear unabhängiger Vektoren, die zur Darstellung der Spaltenvektoren von X notwendig sind, und r ist, analog dazu, die kleinste Anzahl linear unabhängiger Vektoren zur Darstellung von X'. Für die Gleichung (2.23) bedeutet dies, dass die Anzahl r der Spaltenvektoren von B_r größer sein könnte als s, und in (2.24) kann die Anzahl s der Spalten von U_s' größer als r sei. Dies bedeutet, dass die Gleichungen

$$\mathcal{L}(X) \subseteq \mathcal{L}(B_r'), \tag{2.25}$$

$$\mathcal{L}(X') \subseteq \mathcal{L}(A_s'), \tag{2.26}$$

aus (2.23) und (2.24) folgen. Dann folgt der

Satz 2.2 *(***Rangsatz***) Es sei $X_{m,n}$ eine beliebige (m,n)-Matrix mit dem Zeilenrang r und dem Spaltenrang s. Dann gilt*

$$r = s \tag{2.27}$$

d.h. der Zeilenrang ist stets gleich dem Spaltenrang, so dass nur vom Rang *r der Matrix $X_{m,n}$ gesprochen wird. Es gilt*

$$r \leq \min(m,n) \tag{2.28}$$

Beweis: Aus dem Kommentar zu den Gleichungen (2.25) und (2.26) folgt sofort

$$r \leq s, \quad s \leq r,$$

also (2.27).

Es muß noch gezeigt werden, dass $r \leq \min(m,n)$. Es sei $n \leq m$. Der Spaltenrang s von X ist höchstens gleich n, und da $s = r$, r der Rang von X, folgt $r \leq \min(m,n)$. Nun sei $m \leq n$. Der Zeilenrang von X ist höchstens gleich m, und da der Zeilenrang von X gleich dem Rang r von X ist, gilt ebenfalls $r \leq \min(m,n)$. □

Anmerkungen:

1. Wie in der Einführung zu Satz 2.2 über den Zeilen- und Spaltenrang gesagt wurde, ist die Aussage über der Rang einer Matrix eine Aussage über die *kleinstmögliche* Anzahl von linear unabhängigen Vektoren, die jeweils zur Darstellung der Zeilen- bzw. Spaltenvektoren einer Matri X benötigt werden. Dies schließt nicht aus, dass gelegentlich eine vollständige Basis des $\mathbb{R}^n$ für die Darstellung der Zeilenvektoren, bw. des $\mathbb{R}^m$ für die Darstellung der Spaltenvektoren gewählt wird. Da mit einer vollständigen Basis des $\mathbb{R}^m$ *alle* Vektoren des $\mathbb{R}^m$ dargestellt werden können, werden auch die Vektoren

des Teilraums $\mathcal{L}(X)$ dargestellt: eine analoge Aussage gilt natürlich für die Vektoren von $\mathcal{L}(X')$.

So sei $\mathbf{x} \in \mathbb{R}^m$, und $\mathcal{B}_m = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ sei eine Basis des $\mathbb{R}^m$; dann kann $\mathbf{x}$ als Linearkombination

$$\mathbf{x} = a_1\mathbf{b}_1 + \cdots + a_m\mathbf{b}_m$$

dargestellt werden. Weiter sei $\mathcal{E} = \mathcal{L}(\{\mathbf{x}_1, \mathbf{x}_2\})$, $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^m$, eine von $\mathbf{x}_1$ und $\mathbf{x}_2$ aufgespannten Ebene, und $\mathbf{x} \in \mathcal{E}$. Dann kann $\mathbf{x}$ als Linearkombination

$$\mathbf{x} = v_1\mathbf{x}_1 + v_2\mathbf{x}_2$$

repräsentiert werden, bzw. als Linerkombination von irgendzwei nicht-parallelen Vektoren aus $\mathcal{E}$.

2. Wegen der allgemeinen Beziehung $(AB)' = B'A'$ folgt

$$X = UV' \Rightarrow X' = VU' \tag{2.29}$$

und natürlich $X' = VU' \Rightarrow X = UV'$, womit eine enge Beziehung zwischen den Zeilen- und Spaltenvektoren von X verdeutlicht wird: sind die Spaltenvektoren $\mathbf{x}_j$ von X Elemente der linearen Hülle $\mathcal{L}(U)$ der Spaltenvektoren $\mathbf{u}_1, \ldots, \mathbf{u}_r \in \mathbb{R}^m$, $r \leq \min(m,n)$, so sind die Zeilenvektoren von X, d.h. die Spaltenvektoren von X', Elemente der linearen Hülle $\mathcal{L}(V)$, d.h. der Spaltenvektoren $\mathbf{v}_1, \ldots, \mathbf{v}_r \in \mathbb{R}^n$. Die Beziehung (2.29) gilt allgemein: es sei $r_0 = \min(m,n)$, so dass $r \leq r_0$ gilt. Dann kann zur Darstellung der Spaltenvektoren $\mathbf{x}_j$ von X stets eine Menge von linear unabhängigen Vektoren $\mathbf{u}_1, \ldots, \mathbf{u}_s$ mit $s \leq r_0$ herangezogen werden, d.h. s kann größer als der Rang r von X sein. Denn sei $\{\mathbf{b}_1, \ldots, \mathbf{b}_{r_0}\}$ eine Menge linear unabhängiger Vektoren aus $\mathbb{R}^m$ derart, dass $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \in \mathcal{L}(bb_1, \ldots, \mathbf{b}_{r_0})$; dann gilt

$$\mathcal{L}(\mathbf{u}_1), \ldots, \mathbf{u}_r) \subseteq \mathcal{L}(\mathbf{b}_1, \ldots, \mathbf{b}_{r_0})$$

d.h. die $\mathbf{x}_j$ sind auch Linearkombinationen der $\mathbf{b}_1, \ldots, \mathbf{b}_{r_0}$.

Eine analoge Aussage gilt für die Spaltenvektoren $\mathbf{v}_1, \ldots, \mathbf{v}_{r_0}$ von V für die Darstellung $X' = VU'$.

Beispiel: Es sei X ein (m,n)-Matrix mit dem Rang $r < \min(m,n)$. Dann existieren Matrizen V und U' derart, dass $X' = VU'$, so dass die Spaltenvektoren $\tilde{\mathbf{x}}_i$ von X' Elemente der linearen Hülle $\mathcal{L}(V)$ der Spaltenvektoren $\mathbf{v}_k$ von V sind, $\mathbf{v}_k \in \mathbb{R}^n$ sind. Es sei insbesondere $\{\mathbf{v}1, \ldots, \mathbf{v}_n\}$ eine Basis des $\mathbb{R}^n$. Dann sind die $\tilde{\mathbf{x}}_i \in \mathcal{L}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$. So kann man stets eine (n,n)-Matrix V wählen derart, dass $V'V = I_n$, I_n die (n,n)-Einheitsmatrix. Die Spaltenvektoren $\mathbf{v}_k$ von V sind dann orthonormal, damit sind sie linear unabhängig und bilden eine Basis des $\mathbb{R}^n$ und $X' = VU'$ ist eine mögliche Darstellung der Spaltenvektoren von X'. □

Es sei daran erinnert, dass wegen Satz 1.10, Seite 46, für eine bestimmte Wahl von U die Matrix V' und damit auch V eindeutig bestimmt ist, und umgekehrt; U ist für eine gegebene Wahl von V eindeutig bestimmt.

3. Der Beweis von Satz 2.2 wird üblicherweise durch Anwendung elementarer Umformungen auf X geführt, s. Abschnitt 6.5

4. 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. □

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

$$rg(AB) \leq \min[rg(A), rg(B)]. \tag{2.30}$$

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. $rg(Q) = m$, $rg(P) = n$. Dann gilt

$$rg(PAQ) = rg(A), \tag{2.31}$$

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

$$rg(A + B) \leq rg(A) + rg(B) \tag{2.32}$$

Beweis: Zu 1. A habe den Rang s. Die Spaltenvektoren von $C = AB$ sind Linearkombinationen der Spaltenvektoren von A, d.h. $C \subseteq \mathcal{L}(A) = \mathcal{L}(\mathbf{a}_1, \ldots, \mathbf{a}_s)$, $\mathbf{a}_1, \ldots, \mathbf{a}_s$ eine Basis der linearen Hülle $\mathcal{L}(A)$ von A (s. Definition 1.11, Seite 44) so dass C höchstens den Spaltenrang s hat. Die Matrix B habe den Rang r, und die Zeilenvektoren von C sind Linearkombinationen der Zeilenvektoren von B, so dass $C' \subseteq \mathcal{L}(B') = \mathcal{L}(\mathbf{b}_1, \ldots, \mathbf{b}_r)$, $\mathbf{b}_1, \ldots, \mathbf{b}_r$ eine Basis von $\mathcal{L}(B')$, so dass C höchstens den Rang r haben kann. Es sei $r \leq s$; dann hat C höchstens den Rang r, denn nach Gleichung (2.27) sind dann auch die Spaltenvektoren von C als Linearkombinationen von maximal r linear unabhängigen Vektoren darstellbar. Sei umgekehrt $s \leq r$; analog zur vorangegehenden Argumentation ist dann der Rang von C höchstens gleich s, d.h. $\mathrm{rg}(C) \leq \min[\mathrm{rg}(A), \mathrm{rg}(B)]$.

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

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

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

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

d.h.

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

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

Zu 3. Der Rang einer Matrix ist gleich der Dimension des von den Spalten der Matrix aufgespannten Teilraums, d.h. $\mathrm{rg}(A) = \dim \mathcal{L}(A)$. Es sei $C = A + B$; dann ist die j-te Spalte $\mathbf{c}_j$ von C durch $\mathbf{a}_j + \mathbf{b}_j$ gegeben, $\mathbf{a}_j$ die j-te Spalte von A und $\mathbf{b}_j$ die j-te Spalte von B. $\mathbf{c}_j$ ist ein Element der Summe der Teilräume $\mathcal{L}(A)$ und $\mathcal{L}(B)$. Für die Dimensionalität der Summe zweier Teilräume gilt nach (2.32)

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

und das heißt (2.32). □

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

Beweis: Die Aussagen folgen direkt aus (2.31) für entweder $Q = I_n$ oder $P = I_m$. □

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

Beweis: Nach Satz 2.3, Gleichung (2.30) 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, – also kann nur $\text{rg}(A) = s \geq r$ gelten. □

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

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

Beweis: Die Inspektion der Matrix $\mathbf{xy}'$ (vergl. (1.25), (S. 16) zeigt, dass die Spalten der Matrix durch $y_1\mathbf{x}, y_2\mathbf{x}, \ldots, y_n\mathbf{x}$ gegeben sind, wobei $y_1, \ldots, y_n$ 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. □

Der Nullraum:

Definition 2.1 *Es sei A eine (m,n)-Matrix. Die Menge*

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

heißt Nullraum *oder* Kern *von A. Die Dimensionalität* $\dim(\mathcal{N})$ *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*

$$r + s = \min(m,n). \tag{2.33}$$

Beweis: Die Behauptung des Satzes kann auf die Aussage über orthogonale Komplemente (Definition 1.17, Seite 56) und Satz 1.20, Seite 57, zurückgeführt werden. Denn sei V das orthogonale Komplement von $\mathcal{N}(A)$. Gemäß (1.123), Seite 57, gilt dann die Beziehung (2.33). □

Anmerkung: Es gelte inbesondere $r = \min(m,n)$; dann folgt aus (2.33) $s = 0$, d.h. der Nullraum hat den Rang 0. Der Nullraum enthält dann nur den Nullvektor $\vec{0}$ (s. Anmerkung 3. auf Seite 69). 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 $\mathbf{x}'A = \vec{0}'$, $\mathbf{x} \in \mathbb{R}^m$. □

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

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

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 Gleichungsystem 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 genau eine (falls $\mathbf{y} \in \mathcal{L}(A)$) oder gar keine Lösung (falls $\mathbf{y} \notin \mathcal{L}(A)$). □

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. $r = \mathrm{rg}(A) = \mathrm{rg}(A') = 2$. Für den Nullraum folgt aus (2.33) 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.

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

. □

Der Rang von Kreuzproduktmatrizen:

Definition 2.2 *es sei X eine (m, n)-Matrix, Dann heißen die Produkte $X'X$ und XX'* Gram-Matrizen *(oder auch* Gramsche Matrizen*).*[37]

Unter Anwendern multivariater statistischer Verfahren ist statt des Ausdrucks Gram-Matrix bzw. Gramsche Matrix der Ausdruck 'Kreuzprodukt'[38] üblich. Die Kreizprodukte $X'X$ und XX' sind in der multivariaten Analyse von speziellem Interesse, z.B. sind Kovarianz- und Korrelationsmatrizen Kreuzprodukte.

Es gelten die folgenden Aussagen:

Satz 2.6 *Es sei X eine (m, n)-Matrix mit dem Rang $r \leq \min(m, n)$, so dass $X = UV'$. rür U und V sollen die Annahmen A1: $V'V = I$, und A2: $U'U = \Lambda = diag(\lambda_1, \ldots, \lambda_n)$ (Seite* **??***gelten, d.h. und . Dann gilt (i) Die von Null verschiedenen Eigenwerte der Kreuzprodukte $X'X$ und XX' sind identisch (ii) Es sei X eine (m, n)-Matrix mit dem Rang $rg(X) = r \leq \min(m, n)$. Dann gilt für den Rang der Kreuzproduktmatrizen*

$$rg(X'X) = rg(XX') = rg(X). \tag{2.34}$$

[37]Nach dem dänischen Mathematiker Jørgen Pedersen Gram (1850 – 1916)
[38]Kreuzprodukt

Beweis:

Beweis:[39] (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. □

Ist X eine Datenmatrix, so repräsentieren üblicherweise die m-Zeilen die "Fälle", d.h. Personen oder Objekte, an denen jeweils Messungen von n jeweils vorgenommen wurden, 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.

Repräsentieren die Spalten von X Variablen und ist X spaltenzentriert, so ist

$$C = \frac{1}{m} X'X \tag{2.35}$$

eine Kovarianzmatrix[40]. Ist X spaltenstandardisiert, d.h. gilt $X = Z$, so ist

$$R = \frac{1}{m} Z'Z, \tag{2.36}$$

R ist die Matrix der Korrelationen zwischen den Variablen.

Kovarianzmatrizen und dyadische Produkte In vielen Darstellungen multivariater Verfahren wird von einer alternativen, auf dem dyadischen Produkt beruhenden Darstellung von Kovarianz- und Korrelationsmatrizen Gebrauch gemacht. Es seien $x_{ij} = X_{ij} - \bar{x}_j$, $x_{ik} = X_{ik} - \bar{x}_k$; dann ist

$$c_{jk} = \frac{1}{n} \sum_{i=1}^{n} x_{ij} x_{ik}$$

die Kovarianz zwischen der j-ten und der k-ten Variablen. Weiter werde für die Zwecke dieses Abschnitt der i-te Zeilenvektor von X mit $\tilde{\mathbf{x}}_i$ bezeichnet (abweichend von der sonst in diesem Text befolgten Regel, derzufolge $\tilde{\mathbf{x}}_i$ der *Spalten-*

[39] Seber, p. 385

[40] Üblicherweise 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.

vektor von X' bezeichnet). Dann ist

$$\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i' = \begin{pmatrix} x_{i1}x_{i1}, & x_{i1}x_{i2}, & \cdots & x_{i1}x_{in} \\ xi2x_{i1}, & x_{i2}x_{i2} & \cdots & x_{i2}x_{in} \\ \vdots & \vdots & \ddots & \vdots \\ xinx_{i1}, & x_{in}x_{i2}, & \cdots & x_{in}x_{nn} \end{pmatrix} \tag{2.37}$$

das dyadische Produkt von $\tilde{\mathbf{x}}_i$ mit sich selbst, und das Element $x_{ij}x_{ik}$ dieser Matrix ist gerade der i-te Summand der Summe $\sum_i x_{ij}x_{ik}$, d.h.

$$C = \frac{1}{n}\sum_{i=1}^{n} \tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i \tag{2.38}$$

ist die Matrix der (Stichproben-)Kovarianzen. Diese Darstellung von C hat für Herleitungen von Aussagen Vorteile, vergl. Seite 102, Gleichung (3.64).

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

Satz 2.7 *Es sei D eine (m, n)-Matrix mit dem Rang $r \leq \min(m, n)$, und den den Elementen $d_{ij} = 0$ für $i \neq j$, $d_{ij} = \lambda_i \neq 0$ für $i = j \leq r$, und $d_{ij} = 0$ für $i > r$, $j > r$, d.h. die Anzahl der von Null verschiedenen Diagonalelemente ist gleich dem Rang von D.*

$$D = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \lambda_r & 0 \\ 0 & 0 & \vdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix} \tag{2.39}$$

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 $\mathbf{e}_1, \ldots, \mathbf{e}_n$ 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 enhält aber $n - r$ Nullvektoren. Erweitert man die r Spaltenvektoren $\neq \vec{0}$ auch nur um einen Nullvektor, so sind nach Satz 1.6, Seite 33, 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 vorausgesezt werden kann, siehe die Folgerungen 6 bis 8, Seite 209. □

2.4 Die Inverse einer Matrix

In der Artihmetik kennt man die vier Grundrechenarten Addition, Subtraktion, Multiplikation und Division. Die Divsion $\frac{a}{b}$ 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: $\frac{a}{b} = ab^{-1}$. Es 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 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 gegebenene 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 $A\colon \mathbb{R}^m \to \mathbb{R}^n$, 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 $\mathbf{x} \in \mathbb{R}^n$ und damit für alle $\mathbf{y} \in \mathbb{R}^m$. 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 $\mathbf{y} \in \mathbb{R}^m$ folgt, dass B m Spalten haben muss, und da $\mathbf{x} \in \mathbb{R}^n$ folgt, dass B n Zeilen haben muss, d.h. B muss eine (n, m)-Matrix sein, und die Beziehung $AB\mathbf{y} = \mathbf{y}$ impliziert[41], 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:

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

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

$$AB = CA = I, \text{ d.h. } I_m = I_n \tag{2.40}$$

[41]Aus $AB\mathbf{y} = \mathbf{y}$ 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 repäsentiert 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* $\mathbf{x} \in \mathbb{R}^n$ und damit für die zum jeweiligen $\vec{x}$ korrespondierenden Vektoren$\mathbf{y} \in \mathbb{R}^m$ gelten. Man kann demnach die Derivierte $\frac{dD}{d\mathbf{y}} = \frac{dI}{d\mathbf{y}} = 0$ betrachten (s. Gleichung (6.55), Seite 193), derzufolge $\frac{dD}{d\mathbf{y}} = D = AB - I = 0$, also $D = 0$ gilt, woraus $AB = I$ folgt.

gelten soll, so folgt, dass $m = n$ und demnach $I = I_n$ eine (n, n)-Einheitsmatrix ist. Multpliziert man nun (2.40) von links mit C, so erhält man

$$\underbrace{CA}_{I_n} B = CI \Rightarrow B = C, \tag{2.41}$$

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

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

$$A^{-1}A = AA^{-1} = I_n, \tag{2.42}$$

I_n die (n, n)-Einheitsmatrix. □

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

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

$$a\alpha + b\gamma = 1 \Rightarrow \alpha = \frac{1 - b\gamma}{a} \tag{2.43}$$

$$a\beta + b\delta = 0 \Rightarrow \beta = -\frac{b\delta}{a} \tag{2.44}$$

$$c\alpha + d\gamma = 0 \Rightarrow \gamma = -\frac{c\alpha}{d} \tag{2.45}$$

$$c\beta + d\delta = 1 \Rightarrow \delta = \frac{1 - c\beta}{d} \tag{2.46}$$

Durch Einsetzen etwa des in Gleichung (2.45) gegebenen Ausdrucks für γ in die Gleichung (2.43) und Auflösen nach α etc findet man die Beziehungen

$$\alpha = \frac{d}{ad - bc} \tag{2.47}$$

$$\beta = -\frac{b}{ad - bc} \tag{2.48}$$

$$\gamma = -\frac{c}{ad - bc} \tag{2.49}$$

$$\delta = \frac{a}{ad - bc}, \tag{2.50}$$

so dass man schließlich

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{2.51}$$

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

Anmerkung: Der Nenner $ad - bc$ ist die *Determinante* der Matrix A; sie wird mit $|A|$ notiert, d.h. $|A| = ad - bc$. Determinanten werden in Abschnitt 6.13 behandelt, wo auch die *Cramersche Regel* (Seite 119) zur Lösung von linearen Gleichungssystemen vorgestellt wird. □

Es werde nun angenommen, dass $ad - bc = 0$ ist, d.h. es gelte $ad = bc$. Dann ist der Ausdruck $1/(ad - bc)$ in (2.51) nicht definiert, d.h. A^{-1} existiert nicht. Man hat dann einerseits

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

andererseits folgt auch

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

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

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

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

Beispiel 2.4 Die Gleichungen (**??**) und (**??**), Seite **??**, sind ein System von Gleichungen mit zwei Unbekannten, $\mathbf{y} = A\mathbf{x}$, mit

$$A = \begin{pmatrix} \|\mathbf{u}_1\|^2, & \mathbf{u}_1'\mathbf{u}_2 \\ \mathbf{u}_2'\mathbf{u}_1, & \|\mathbf{u}_2\|^2 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} \mathbf{u}_1'\mathbf{v} \\ \mathbf{u}_1'\mathbf{v} \end{pmatrix} \tag{2.52}$$

Die inverse Matrix A^{-1} ist hier durch

$$A^{-1} = \frac{1}{\mathbf{u}_1\|^2\|\mathbf{u}_2\|^2 - (\mathbf{u}_1'\mathbf{u}_2)^2} \begin{pmatrix} \|\mathbf{u}_2\|^2, & -\mathbf{u}_1'\mathbf{u}_2 \\ -\mathbf{u}_1'\mathbf{u}_2, & \|\mathbf{u}_1\|^2. \end{pmatrix} \tag{2.53}$$

Man rechnet leicht nach, dass die Komponenten von $\mathbf{x} = A^{-1}\mathbf{y}$ durch die Gleichungen (**??**) und (**??**), Seite **??**, gegeben sind. □

Beispiel 2.5 Es sei R eine (2×2)-Korrelationsmatrix,

$$R = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}. \tag{2.54}$$

Die Inverse R^{-1} ergibt sich aus (2.51)

$$R^{-1} = \begin{pmatrix} \frac{1}{1-r^2} & -\frac{r}{1-r^2} \\ -\frac{r}{1-r^2} & \frac{1}{1-r^2} \end{pmatrix} = \frac{1}{1-r^2} \begin{pmatrix} 1 & -r \\ -r & 1 \end{pmatrix} \tag{2.55}$$

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

$$\lim_{r \to 0} R^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \ \lim_{r \to 1} R^{-1} = \begin{pmatrix} \infty & -\infty \\ -\infty & \infty \end{pmatrix} \tag{2.56}$$

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

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

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

Beispiel 2.6 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:

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

Dann ist

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

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

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

Das Prinzip läßt sich auf eine (n,n)-Matrix A anwenden. A habe die Elemente a_{ij}, $i, = 1,\ldots,n$. Es werde angenommen, dass die inverse Matrix A^{-1} existiert; die Elemente von A^{-1} seien $\hat{a}_{ij}$. Dann soll gelten

$$AA^{-1} = \begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{n1} & \cdots & a_{nj} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} \hat{a}_{11} & \cdots & \hat{\hat{a}}_{1j} & \cdots & \hat{a}_{1n} \\ \hat{a}_{11} & \cdots & \hat{\hat{a}}_{1j} & \cdots & \hat{a}_{1n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \hat{a}_{n1} & \cdots & \hat{a}_{nj} & \cdots & \hat{a}_{nn} \end{pmatrix} = I_n \tag{2.57}$$

mit

$$I_n = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Die Matrix A^{-1} ergibt sich also us den Lösungen der Gleichungssysteme

$$\begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{n1} & \cdots & a_{nj} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} \hat{a}_{1j} \\ \hat{a}_{2j} \\ \vdots \\ \hat{a}_{nn} \end{pmatrix} = \mathbf{e}_j, \quad j = 1,\ldots,n \tag{2.58}$$

mit den Unbekannten $\hat{a}_{1j}, \hat{a}_{2j}, \ldots, \hat{a}_{nj}$. Im Prinzip ergibt sich die Lösung wie im Falle der $(2,2)$-Matrix, allerdings wendet dazu den Gauß-Jordan-Algorithmus an, auf den an dieser Stelle nicht eingegangen werden soll, weil er für das Verständnis des Folgenden nicht benötigt wird. Wichtig ist allerdings, dass die Matrix A den vollen Rang $r = n$ haben muß, damit die Inverse A^{-1} berechnet werden kann, d.h. die Spalten-, und damit auch die Zeilenvektoren von A müssen linear unabhängig sein. Man hat dann den Satz

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*

$$(AB)^{-1} = B^{-1}A^{-1}. \tag{2.59}$$

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

2.5 Die Beziehungen zwischen verschiedenen Basen

Es sei $S \subseteq \mathbb{R}^n$ ein Teilraum des n-dimensionalen Vektorraums, und es seien $\mathcal{B} = (\mathbf{b}_1,\ldots,\mathbf{b}_r)$ und $\mathcal{C} = (\mathbf{c}_1,\ldots,\mathbf{c}_r)$, $r \leq n$, 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 $B = [\mathbf{b}_1, \ldots, \mathbf{b}_r]$ die (n, r)-Matrix mit den Spaltenvektoren $\mathbf{b}_j$, und $C = [\mathbf{c}_1, \ldots, \mathbf{c}_r]$ die (n, r)-Matrix mit den Spaltenvektoren $\mathbf{c}_j$, $j = 1, \ldots, r$. Dann existieren sicherlich r r-dimensionale Vektoren $\mathbf{t}_1, \ldots, \mathbf{t}_r$ derart, dass

$$B\mathbf{t}_k = \mathbf{c}_k, \quad k = 1, \ldots, r$$

oder, mit $T = [\mathbf{t}_1, \ldots, \mathbf{t}_r]$, T eine (r, r)-Matrix,

$$BT = C, \tag{2.60}$$

und da $\mathrm{rg}(B) = \mathrm{rg}(C)$ und $\mathrm{rg}(C) = \mathrm{rg}(BT) \leq \min[\mathrm{rg}(B), \mathrm{rg}(T)]$, folgt $\mathrm{rg}(T) = r$. Dann existiert die Inverse T^{-1} und

$$C = BT \tag{2.61}$$

und da $CT^{-1} = BTT^{-1}$ folgt

$$B = CT^{-1}.$$

Zusammengefasst hat man

$$\begin{aligned} C &= BT && (2.62)\\ B &= CT^{-1}. && (2.63) \end{aligned}$$

Definition 2.4 *Die Matrizen T und T^{-1} heißen* Übergangsmatrizen *für den Übergang von der Basis $\mathcal{B}$ zu $\mathcal{C}$ bzw. von $\mathcal{C}$ zu $\mathcal{B}$.*

Übergangsmatrizen sind Koordinatenmatrizen; die Komponenten des Spaltenvektors $\mathbf{t}_k$ von T sind die Koordinaten des Vektors $\mathbf{c}_k$ in Bezug auf die Basis $\mathcal{B}$, und die Komponenten $\mathbf{t}_k^*$ von T^{-1} sind die Koordinaten des Vektors $\mathbf{b}_j$ in Bezug auf die Basis $\mathcal{C}$.

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 Orthgonalbasis übergehen.

3 Bestimmung einer Basis: die Hauptachsentransformation

3.1 Rotationen

Eine Möglichkeit, latente Dimensionen zu bestimmen, besteht in der Rotation von Koordinatenachsen bzw. von Vektoren. Generell ist die Transformation eines

Vektors $\mathbf{x}$ in einen Vektor $\mathbf{y}$ durch die Gleichung $\mathbf{y} = T\mathbf{x}$ definiert, wobei T die Transformationsmatrix ist. Ist $\mathbf{x} \in \mathbb{R}^n$ und $\mathbf{y} \in \mathbb{R}^m$, so ist T eine (m, n)-Matrix. Soll der Vektor $\mathbf{x}$ nur rotiert werden, so ist $\mathbf{y}$ ebenfalls ein n-dimensionaler Vektor, und darüber hinaus unterscheidet sich $\mathbf{y}$ von $\mathbf{x}$ nur bezüglich der Orientierung, nicht aber hinsichtlich der Länge. Dementsprechend ist T eine (n, n)-Matrix.

Es sei also $\mathbf{x} \in \mathbb{R}^n$ ein beliebiger n-dimensionaler Vektor und T sei eine (n, n)-Matrix derart, dass $\mathbf{y} = T\mathbf{x} \in \mathbb{R}^n$. Die Invarianz der Länge bedeutet $\|\mathbf{y}\| = \|\mathbf{x}\|$, d.h.

$$\|\mathbf{y}\|^2 = \mathbf{y}'\mathbf{y} = \mathbf{x}'T'T\mathbf{x} = \mathbf{x}'\mathbf{x} = \|\mathbf{x}\|^2. \tag{3.1}$$

Es gilt dann der

Satz 3.1 *Es seien $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ und T sei eine (n, n)-Matrix. Die Beziehung (3.1) gilt genau dann, wenn T orthonormal ist, d.h. wenn*

$$T'T = TT' = I \tag{3.2}$$

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 werden. Dann liefert (6.58) aus dem Anhang, Abschnitt 6.9.2, Seite 194

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

und die nochmalige Ableitung liefert $A = I$ (s. Gleichung (6.55), Seite 193), d.h. $T'T = I$. Dies bedeutet, dass die Spaltenvektoren von T orthonormal sind.

(iii) $T'T = I \Rightarrow TT' = I$: Es gilt

$$\begin{aligned}
T'T &= I \\
\Rightarrow T'TT' &= IT' = T'I \\
\Rightarrow T'TT' - T'I &= 0 \\
\Rightarrow T' \underbrace{(TT' - I)}_{B=[\mathbf{b}_1,\ldots,\mathbf{b}_n]} &= 0
\end{aligned}$$

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[42] $T'\mathbf{b}_j = \vec{0}_j \Rightarrow \mathbf{b}_j = \vec{0}_j$ für $j = 1, \ldots, n$, also folgt $B = TT' - I = 0$, d.h. $TT' = I$.

[42] vergl. Satz 1.10, Seite 46

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

$\square$

Anmerkungen:
1. **Alternativer Beweis** Einen alternativen Beweis für die Aussage $TT' = I$ findet man in Beispiel 6.1 auf Seite 193 im Anhang. Ein weiterer Beweis setzt den Begriff der Determinante voraus, s. Beispiel 6.6, Seite 211 im Anhang.

2. **Zur Bedeutung von $TT' = I$:** Dass $T'T = I$ auch die Aussage $TT' = I$ – und umgekehrt – impliziert heißt, dass die Orthonormalität der Spaltenvektoren die Orthonormalität der Zeilenvektoren von T impliziert, und umgekehrt.

Die Beziehung $T'T = TT' = I$ bedeutet überdies, dass die Inverse T^{-1} durch die Transponierte T' gegeben ist:

$$T^{-1} = T'. \tag{3.3}$$

3. **Die Elemente der Rotationsmatrix** sind Richtungskosinus. Nach (1.59), Seite 26, gilt

$$\mathbf{x} = \|\mathbf{x}\|(\cos\theta_1, \ldots, \cos\theta_n)'$$

und da die Spaltenvektoren $\mathbf{t}_j$ einer Rotationsmatrix T die Länge $\|\mathbf{t}_j\| = 1$ haben, muss

$$\mathbf{t}_j = (\cos\theta_{1j}, \ldots, \cos\theta_{nj})' \tag{3.4}$$

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 2, Seite 10 hervorgeht ist aber $\cos\theta_2 = \sin\theta_1$ und (3.5) kann in der Form (**??**) (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 $\tilde{\mathbf{x}}_i$, $i = 1, \ldots, m$, 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:

$$\tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_k = \tilde{\mathbf{u}}_i'V'V\tilde{\mathbf{u}}_k = \tilde{\mathbf{u}}_i'\tilde{\mathbf{u}}_k, \quad i, k = 1, \ldots, m \tag{3.6}$$

und da Rotationen von Vektoren ihre Länge invariant lassen 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 Die Bestimmung einer Basis durch Rotation

Ausgangspunkt für die folgenden Betrachungen ist wieder die allgemeine Repräsentation einer (m,n)-Matrix X gemäß

$$X = UV', \quad \mathbf{x}_j = U\tilde{\mathbf{v}}_j, \quad j = 1,\ldots,n \tag{3.7}$$

d.h. die Repräsentation der Spaltenvektoren $\mathbf{x}_j$ von X als Linearkombinationen linear unabhängiger Vektoren $\mathbf{u}_k$ (Spaltenvektoren der Matrix U) und der Koeffizientenvektoren $\tilde{\mathbf{v}}_j$ (Spaltenvektoren von V'). Dabei kann, unabhängig von Rang $r \leq \min(m,n)$ von X, U eine (m,n)-Matrix und V eine (n,n)-Matrix sein, zumal wenn r noch nicht bekannt ist und noch geschätzt werden muß, oder man wählt für einen bestimmten Wert r eine (m,r)-Matrix U und eine korrespondierende (n,r)-Matrix V. (3.7) impliziert

$$X' = VU', \quad \tilde{\mathbf{x}}_i = V\tilde{\mathbf{u}}_i, \quad i = 1,\ldots,m \tag{3.8}$$

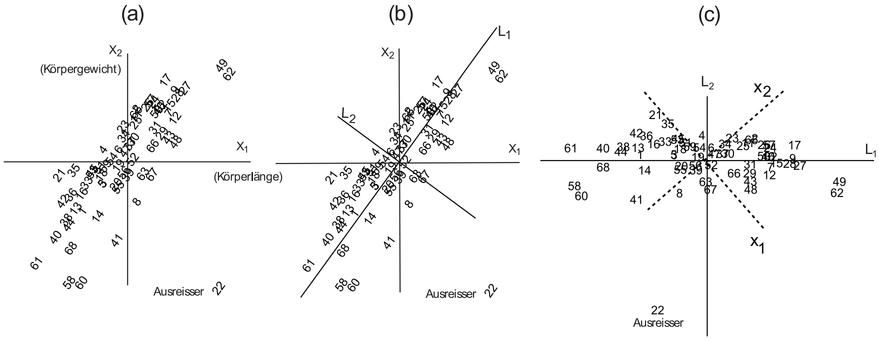
d.h. die Spaltenvektoren $\tilde{\mathbf{x}}_i$ (Zeilenvektoren von X) werden als Linearkombinationen der Spaltenvektoren von V dargestellt. Die Endpunkte der $\tilde{\mathbf{x}}_i$ bilden die Punktekonfiguration der Fälle. Es gibt, etwas salopp gesagt, unendlich viele Möglichkeiten, eine Matrix U und die zugehörige Matrix V zu wählen. Eine erste Möglichkeit, eine Wahl zu treffen, besteht darin, dass die Vektoren $\tilde{\mathbf{x}}_i$ und $\tilde{\mathbf{u}}_i$ sich nur durch ihre Orientierung voneinander unterscheiden, wobei der Winkel zwischen den Orientierungen von $\tilde{\mathbf{x}}_i$ und $\tilde{\mathbf{u}}_i$ fäüe alle i identisch ist; dann gilt $\|\tilde{\mathbf{x}}_i\| = \|\tilde{\mathbf{v}}_i\|$ für alle i. Wie im vorangehenden Abschnitt gezeigt wurde, bedeutet diese Annahme, dass die Matrix V in (3.8) orthonormal ist, d.h. es soll $V'V = I$ gelten. Die Spaltenvektoren (und ebenso die Zeilenvektoren) von V sind demnach linear unabhängig und bilden jeweils eine Basis des $\mathbb{R}^n$[43]. Aber damit ist der Rotationswinkel noch nicht festgelegt, – auch hier gibt es unendlich viele Möglichkeiten. Wie im Folgenden gezeigt wird, kann man diesen Winkel durch eine Annahme über die Matrix U festlegen. Gilt $V'V = I_n$, so führt die Multiplikation der allgemeinen Gleichung $X = UV'$ von rechts mit V auf die Beziehung $XV = U$, d.h. die Spaltenvektoren von U ergeben sich als Linearkombination der Datenvektoren $\mathbf{x}_j$. Da aber auch $U' = V'X'$ gilt, sieht man, dass sich die $\tilde{\mathbf{x}}_i$ durch Rotation aus den $\tilde{\mathbf{x}}_i$ ergeben:

$$\tilde{\mathbf{u}}_i = V'\tilde{\mathbf{x}}_i, \quad i = 1,\ldots,m \tag{3.9}$$

hier ist V' die Rotationsmatrix. Da $\tilde{\mathbf{x}}_i$ ein n-dimensionaler Vektor ist, muß V' n Zeilen haben, damit die Multiplikation $V'\tilde{\mathbf{x}}_i$ "funktioniert", d.h. V muß eine (n,n)-Matrix sein, und U muß dann eine (m,n)-Matrix sein. Der Versuch, $r < \min(m,n)$ Basisvektoren für die $\mathbf{x}_j$ bzw. die $\tilde{\mathbf{x}}_i$ zu bestimmen, scheint damit von vorn herein ausgeschlossen zu werden. Es zeigt sich aber, dass dies nicht der Fall ist, wie im Folgenden deutlich werden wird.

[43] Dieser Fall wurde in in den Anmerkungen 1 und 2 zu Satz 2.2, Seite 67, diskutiert.

Abbildung 14: Rotation einer 2-dimensionalen Punktekonfiguration von Fällen



Man betrachte die Annahmen bzw.Forderungen

Annahmen:[44] Es seien U eine (m,n)- und V eine (n,n)-Matrix. Dann gelte

$$\begin{aligned} \text{A1: } V'V &= I, \ \text{(Rotation)} & (3.10)\\ \text{A2: } U'U &= \Lambda = \text{diag}(\lambda_1,\dots,\lambda_n), \quad \lambda_k = \mathbf{u}_k'\mathbf{u}_k = \|\mathbf{u}_k\|^2, & (3.11) \end{aligned}$$

Es wird im Folgenden gezeigt werden, dass der Punktekonfiguration der Fälle eine Menge von Ellipsoiden mit identischer Orientierung entspricht, wobei die Orientierungen der Hauptachsen der Ellipsoide durch die Orientierungen der Spaltenvektoren von V gegeben sind. Man sagt dementsprechend, dass V eine *Hauptachsentransformation* (HAT) definiert. Die gängige Abkürzung ist allerdings PCA[45], Englisch für *Principal Component Analysis.* Abbildung 14 illustriert die Rotation.

Der folgende Satz liefert eine Charakterisierung der Matrix V:

Satz 3.2 *Es sei X beliebige (m,n)-Matrix, es gelte $X = UV'$, wobei U eine (m,n)-Matrix und V eine (n,n)-Matrix ist. Die Annahmen A1 und A2 implizieren*

$$CV = V\Lambda, \quad \text{d.h. } C\boldsymbol{v}_k = \lambda_k \boldsymbol{v}_k, \quad k = 1,\dots,n \tag{3.12}$$

mit $C = X'X$.

Beweis: Nach A1 gilt $V'V = VV' = I$, I_n 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 $VV'CV = CV = V\Lambda$. □

[44] Man kann gleichermaßen von den Annahmen A1*: $V'V = \Lambda = \text{diag}(\lambda_1,\dots,\lambda_n)$, und A2*: $U'U = I$ ausgehen; die Annahmen A1 und A2 korrespondieren zum üblichen Vorgehen, zunächst von der Rotation der Punktekonfiguration der Fälle auszugehen.

[45] u.a. weil viele Statistikprogramme wir z.B. R, die Verfahren auf Englisch benannt und beschrieben werden.

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 legt nahe, dass die $\mathbf{v}_k$ für die Matrix C charakteristisch sind – in Abschnitt 6.7 wird gezeigt, dass für den Fall, dass nicht alle Eigenwerte identisch sind, die Matrix der Eigenvektoren eindeutig betimmt ist – , weshalb ein spezieller Name für die $\mathbf{v}_k$ als gerechtfertigt erscheint:

Definition 3.1 *Es seien M eine (n, n)-Matrix und $\boldsymbol{v} \in \mathbb{R}^n$, und es gelte (3.12), d.h. es gelte für einen k-ten Spaltenvektor $\boldsymbol{v}_k$ von V*

$$M\boldsymbol{v}_k = \lambda_k \boldsymbol{v}_k, \quad \lambda_k \in \mathbb{R}, \; \boldsymbol{v}_k \neq \vec{0}, \; k = 1, \ldots, n \tag{3.13}$$

Dann heißt $\boldsymbol{v}_k$ Eigenvektor *oder* charakteristischer Vektor *von M und λ_k ist der zu $\boldsymbol{v}$ gehörige* Eigenwert *oder* charakteristische Wert[46]*. Mit $V = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n]$ die Matrix der Eigenvektoren und $\Lambda = diag(\lambda_1, \ldots, \lambda_n)$ die Diagonalmatrix der Eigenwerte wird (3.13) zu*

$$MV = V\Lambda. \tag{3.14}$$

Das Paar (V, Λ) wird auch als Eigenstruktur *der Matrix M bezeichnet.*

Aus (3.13) folgen die oft verwendeten Beziehungen

$$M = V\Lambda V' \tag{3.15}$$

$$\Lambda = V'MV \tag{3.16}$$

Die Beziehung (3.16) wird oft als *Diagonalisierung* von M bezeichnet.

Anmerkungen:
1. Symmetrie von M: In der Gleichung (3.12) 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 gefunden wurde, – auf weitere Details wird in Abschnitt 3.6 näher eingegangen. Allerdings kann V nur mit numerischen Mitteln bestimmt werden kann, auf die hier nicht eingegangen werden kann[47]. □

[46] Man findet auch die Ausdrücke *latenter Vektor* und *latenter Vektor* für den Eigenvektor und Eigenwert.

[47] Das 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 Berechung optimiert werden, vergl. Golub & Van Loan (2023). Die Statistikprogramme enthalten diese speziellen Subroutinen.

3.2.2 Eigenstrukturen und Singularwertzerlegung

Orthogonalität der Eigenvektoren In Gleichung (3.12) 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, also auch, wenn V nicht als Rotationsmatrix eingeführt wurde. Tatsächlich läßt sich zeigen, dass unter bestimmten, milden 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* verschieden 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

$$(\mathbf{v}_k' M\mathbf{v}_j)' = \mathbf{v}_j' M\mathbf{v}_k = (\lambda_j \mathbf{v}_k'\mathbf{v}_j)' = \lambda_j \mathbf{v}_j'\mathbf{v}_k \tag{3.17}$$

folgt, so dass man die Gleichungen

$$\begin{aligned}
\mathbf{v}_j' M\mathbf{v}_k &= \lambda_k \mathbf{v}_j'\mathbf{v}_k && (3.18)\\
\mathbf{v}_j' M\mathbf{v}_k &= \lambda_j \mathbf{v}_j'\mathbf{v}_k && (3.19)\\
0 &= (\lambda_k - \lambda_j)\mathbf{v}_j'\mathbf{v}_k && (3.20)
\end{aligned}$$

erhält, wobei sich (3.20) aus (3.18) durch Subtraktion von (3.19) 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. □

Anmerkungen:

1. Die Orthogonalität der Eigenvektoren wird hier nur für Eigenvektoren, die zu Eigenwerten ungleich Null korrespondieren, gezeigt. Sind nur $r < n$ Eigenwerte ungleich Null könnte man vermuten, dass es noch $n - r$ Eigenvektoren gibt, die nicht orthogonal sind. Es ist allerdings so, dass die r orthogonalen Eigenvektoren eine Basis eines r-dimensionalen Teilraums $\mathcal{L}(\mathbf{v}_1, \ldots, \mathbf{v}_r)$ des $\mathbb{R}^n$ bilden. Nach Satz 1.19, Seite 56, kann aber eine r-dimensionale Teilbasis um weitere $n - r$ orthonormale Vektoren ergänzt werden; gemäß der Anmerkung zu Satz 1.19 bilden sie das orthogonale Komplement zu $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$. Berücksichtigt man, dass die Eigenvektoren für den Fall ungleicher Eigenwerte eindeutig bestimmt sind (s. Abschnitt 6.7), so kann man immer von einer orthonormalen (n,n)-Matrix V von Eigenvektoren sprechen.

2. Die Eigenvektoren nicht-symmetrischer Matrizen können, müssen aber nicht orthogonal sein; in Abschnitt 3.3.1 wird gezeigt, dass sie zumindest linear unabhängig sind. 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 $\mathbb{R}^n$. $\square$

Die Singularwertzerlegung (SVD)[48] Es sei X eine beliebige (m, n)-Matrix vom Rang $\text{rg}(X) = r \leq \min(m, n)$. Dann kann X in der Form $X = U_r V_r'$ dargestellt werden, wobei $U_r = [\mathbf{u}_1, \dots, \mathbf{u}_r]$, $V_r = [\mathbf{v}_1, \dots, \mathbf{v}_r]$, und die $\mathbf{u}_k$, und $\mathbf{v}_k$ sind jeweils linear unabhängig. Es mögen insbesondere die Annahmen A1: $V_r'V_r = I_r$, und A2: $U_r'U_r = \Lambda_r = \text{diag}(\lambda_1, \dots, \lambda_r$, $\lambda_k \neq 0$, gelten. Offenbar gilt $\lambda_k = \|\mathbf{u}_k\|^2$, so dass $\|\mathbf{u}_k\| = \lambda^{1/2}$ die Länge von $\mathbf{u}_k$ ist. Dann ist

$$\frac{1}{\|\mathbf{u}_k\|} = \lambda_k^{-1/2}, \quad k = 1, \dots, r$$

und mit

$$\Lambda_r^{-1/2} = \text{diag}(\lambda_1^{-1/2}, \dots, \lambda_r^{-1/2})$$

erhält man für die Normierung der $\mathbf{u}_k$ die Gleichung

$$Q_r = U_r \Lambda_r^{-1/2} \tag{3.21}$$

(vergl. Gleichung (2.19), Seite 65). Dann folgt[49] $U_r = Q_r \Lambda_r^{1/2}$ und man erhält für X die **Singularwertzerlegung:**

$$X = Q_r \Lambda_r^{1/2} V_r'. \tag{3.22}$$

von X; eine alternative, aber natürlich äquivalente Darstellung wird weiter unten geliefert (vergl. (6.52)). Die $\lambda_k^{1/2}$ werden auch *Singularwerte* genannt, die Spaltenvektoren $\mathbf{q}_k$ von Q_r heißen *Linkseigenvektoren* von X, und die $\mathbf{v}_k$ sind die *Rechtseigenvektoren* von X.

Anmerkung: Nach Satz 3.23, Seite 132, liefert die Beziehung (3.22) für den Fall, dass X "numerisch" den Rang $\min(m, n)$ hat, die beste Approximation $X \approx Q_r \Lambda_r^{1/2} V_r'$ für X, wenn ein spezieller Wert $r < \min(m, n)$ angenommen wird. Diese Situation ist gegeben, wenn die Elemente von X empirische und damit messfehlerbehaftete Messungen sind: die zufälligen Messfehler erzeugen Irregularitäten in den x_{ij}-Werten, die numerisch den Rang $\min(m, n)$ von X bedeuten, d.h. $\lambda_k \neq 0$ für alle k. $\square$

Ränge von Kreuzproduktmatrizen: Aus (3.22) folgen sofort die Beziehungen

$$X'X = V_r \Lambda_r^{1/2} Q_r' Q_r \Lambda_r^{1/2} V_r = V_r \Lambda_r V_r',$$

[48] von engl. **S**ingular **V**alue **D**ecomposition

[49] Multiplikation der Gleichung von rechts mit $\Lambda^{1/2}$.

d.h. $X'XV_r = V_r\Lambda_r$, also die bereits bekannte Tatsache, dass die Spaltenvektoren $\mathbf{v}_k$ von V_r die zu Eigenwerten ungleich Null korrespondierenden Eigenvektoren von $X'X$ sind, und

$$XX' = Q_r\Lambda_r^{1/2}V_r'V_r\Lambda_r^{1/2}Q_r' = Q_r\Lambda_rQ_r',$$

so dass $XX'Q_r = Q_r\Lambda_r$ folgt, d.h Q_r enthält die m-dimensionalen, zu Eigenwerten ungleich Null korrespondierenden Eigenvektoren[50] von XX'. Offenbar sind die von Null verschiedenen Eigenwerte von $X'X$ und XX' identisch. Nach Satz 2.3, Aussage (2.31), Seite 69, gelten

$$\begin{aligned} \mathrm{rg}(X'X) &= \mathrm{rg}(V_r\Lambda_rV_r') = \mathrm{rg}(\Lambda_r) = r, \qquad (3.23)\\ \mathrm{rg}(XX') &= \mathrm{rg}(Q_r\Lambda_rQ_r') = \mathrm{rg}(\Lambda_r) = r. \qquad (3.24)\end{aligned}$$

Die Produkte $X'X$ und XX' haben also denselben Rang, was nicht ganz trivial ist, denn schließlich gehören die $\mathbf{x}_j$ und die $\tilde{\mathbf{x}}_i$ Vektorräumen mit verschiedener Dimensionalität an.

Alternative Schreibweise: Die Matrix XX' ist eine (m,m)-Matrix, und die Matrix $X'X$ ist eine (n,n)-Matrix. Nach dem Basisergänzungssatz 1.14, Seite 50, lässt sich die (m,r)-Matrix Q_r zu einer (m,m)-Matrix Q_m von linear unabhängigen Vektoren vervollständigen, wobei insbesondere orthonormale Vektoren gewählt werden können. In gleicher Weise läßt sich die (n,r)-Matrix V_r zu einer orthonormalen (n,n)-Matrix V_n vervollständigen. Erweitert man die (r,r)-Diagonalmatrix Λ_r durch Hinzunahme von $m-r$ $n-r$-dimensionalen Zeilen-Nullvektoren und $n-r$ m-dimensionalen Spalten-Nullvektoren zu einer (m,n)-Matrix, deren Elemente bis auf die ersten r-Diagonalelemente alle gleich Null sind, so kann (3.22) in der Form

$$X = Q\Sigma V' \qquad (3.25)$$

mit

$$\Sigma = \quad \Lambda^{1/2} = \begin{pmatrix} \Lambda_r^{1/2} & 0_{12} \\ 0_{21} & 0_{22} \end{pmatrix} \qquad (3.26)$$

geschrieben werden, wobei 0_{12} eine $(r, n-r)$-Matrix ist, 0_{21} eine $(m-r, r)$-Matrix und 0_{22} eine $(m-r, n-r)$-Matrix ist; die Elemente von 0_{12}, 0_{21} und 0_{22} sind alle gleich Null.(6.52) ist die in vielen Lehrbüchern gewählte Darstellung der Singularwertzerlegung; im Deutschen ist auch der Ausdruck *Grundstruktur* gebräuchlich. Aus (3.22) ergeben sich zwischen den Eigenvektoren in V_r und Q_r die Beziehungen

$$\begin{aligned} V_r &= X'Q_r\Lambda_r^{-1/2}, \qquad (3.27)\\ Q_r &= XV_r\Lambda_r^{-1/2} \qquad (3.28)\end{aligned}$$

[50] d.h. die $\mathbf{u}_k$ sind Eigenvektoren von XX', allerdings nicht normierte.

ergeben (eine Längenskalierung bedeutet die Postmultiplikation mit einer Diagonalmmatrix, vergl. Gleichung 2.19, Seite 65). Für die einzelnen Eigenvektoren ergiben sich daraus die Gleichungen

$$\mathbf{v}_k = \lambda_k^{-1/2} X' \mathbf{q}_k = X'(\lambda_k^{-1/2} \mathbf{q}_k) \tag{3.29}$$

$$\mathbf{q}_k = \lambda_k^{-1/2} X \mathbf{v}_k = X(\lambda_k^{-1/2} \mathbf{v}_k) \tag{3.30}$$

Eine alternative Herleitung, die von den Eigenstrukturen von $X'X$ und XX' ausgeht, findet man in Abschnitt 6.8, Seite 190 im Anhang.

Darstellung der SVD über das dyadische Produkt: Die Zerlegung $X = Q\Sigma V'$ kann in der Form

$$X = \sum_{k=1}^{n} \sigma_k \mathbf{q}_k \mathbf{v}_k', \quad \sigma_k = \sqrt{\lambda_k},\ m \geq n \tag{3.31}$$

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

$$x_{ij} = \sigma_1 q_{i1} v_{j1} + \cdots + \sigma_n q_{in} v_{jn} \tag{3.32}$$

gegeben, $\sigma_1 q_{i1} v_{j1}$ ist der Beitrag der ersten latenten Dimension, etc.

Üblicherweise werden die Eigenwerte λ_k in Λ so angordnet, dass λ_1 der größte und λ_n der kleinste Eigenwert ist, – numerisch sind üblicherweise schon aufgrund von Rundungsfehlern alle $\lambda_k \neq 0$, auch wenn der "wahre" Rang von X kleiner als $\min(m, n)$ ist. Man hat also

$$\sigma_1 > \sigma_2 > \cdots > \sigma_n.$$

Die Matrix X wird in (3.31) 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:

$$\hat{X}_r = \sum_{k=1}^{r} \sigma_k \mathbf{q}_k \mathbf{v}_k' \approx X, \quad \sigma_k = \sqrt{\lambda_k}, \quad r < n \tag{3.33}$$

"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.

3.2.3 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 in einem r-dimensionalen Raum, 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*[51]

$$Q(\boldsymbol{x}) = \boldsymbol{x}' M \boldsymbol{x}, \quad \textit{für alle}\, \boldsymbol{x} \in \mathbb{R}^n \tag{3.34}$$

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

Anmerkung: $Q(\mathbf{x})$ ist ein Skalar, denn $\mathbf{x}'M$ ist ein Zeilenvektor, etwa $\mathbf{x}'M = \mathbf{y}'$, so dass $\mathbf{x}'M\mathbf{x} = \mathbf{y}'\mathbf{x} = Q(\mathbf{x}) \in \mathbb{R}$. □

Im Folgenden werden nur positiv-semidefinite Matrizen betrachtet. Die Ausführlichkeit der Definition 3.34 soll zeigen, dass eine symmetrische Matrix nicht notwendig positiv-semidefinit ist und eine Reihe von Aussagen nur für positiv-semidefinite Matrizen gilt.

Satz 3.4 *Es sei M eine Gramsche Matrix. Dann ist M positiv-semidefinit.*

Beweis: C ist eine (n,n)-Matrix, und es sei $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{v} \neq \vec{0}$. 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). □

Folgerung: Die symmetrische (n,n)-Matrix M sei positiv-semidefinit. Dann gilt für Diagonalelemente m_{ii} von M

$$m_{ii} \geq 0. \tag{3.35}$$

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 $\mathbf{e}_i$, $i = 1, \ldots, n$. 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. □

[51] Q hat nichts mit der in der Singularwertzerleung auftretenden Matrix Q zu tun, sondern steht für 'quadratisch'.

Ellipsoide: Die symmetrische Matrix M sei positiv-semidefinit. Multipliziert man die quadratische Form $\mathbf{x}'M\mathbf{x}$ aus so erhält man

$$\mathbf{x}'M\mathbf{x} = \sum_i m_{ii}x_i^2 + 2\sum_{i<j} m_{ij}x_i x_j, \quad \mathbf{x} \neq \vec{0} \tag{3.36}$$

wobei x_i und x_j die Komponenten des Vektors $\mathbf{x}$ und m_{ij}, $i,j = 1,\ldots,n$, sind die Elemente von M. Der Ausdruck 'quadratische Form' ergibt sich aus dem Sachverhalt, dass (3.36) eine Verallgemeinerung der Formel für $(a+b)^2$ ist, also

$$(a+b)^2 = a^2 + b^2 + 2ab, \quad a, b \in \mathbb{R}$$

ist; man kann sich davon überzeugen, indem man den Ausdruck (3.36) für den Fall $n = 2$ explizit aufschreibt.

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 (3.36) 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:

$$\mathcal{E}_k = \{\mathbf{x}|\mathbf{x}'M\mathbf{x} = k \in \mathbb{R},\ \mathbf{x} \in \mathbb{R}^n,\ k = \text{konstant} > 0\}, \tag{3.37}$$

vergl. Abbildung 15.

Anmerkung: Der Fall $\mathbf{x}'M\mathbf{x} = 0$ impliziert nicht notwendig $\mathbf{x} = \vec{0}$; in Abschnitt **??** wird ausführlich auf diesen Spezialfall eingegangen. □

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: 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 $\|\mathbf{u}_1\| \geq \cdots \geq \|\mathbf{u}_n\|$. 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 $Q(\mathbf{x}) = k$ eine Konstante ein Ellipsoid, dessen erste Hauptachse die maximale Ausdehnung des Ellipsoids repräsentiert, die zweite Hauptachse die zweitmaximale Ausdehung, 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 und $\mathbf{c}_{max}$ bestimmen, d.h. man muß den in der folgenden Definition eingeführten Quotienten maximieren:

Abbildung 15: Ellipsen in verschiedenen Orientierungen; die Hauptachsen sind die skalierten Eigenvektoren der zugehörigen Matrix M.

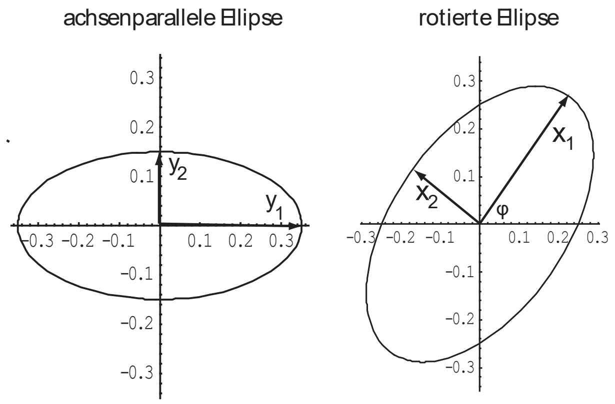
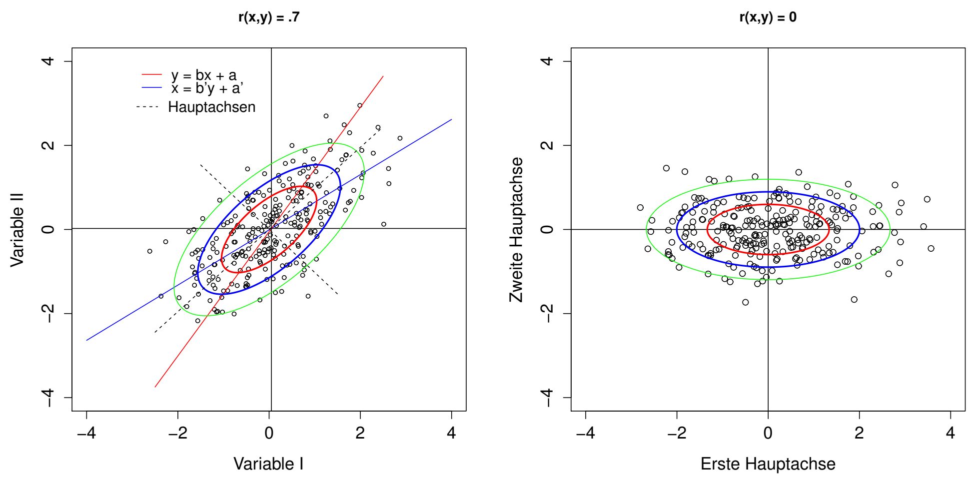


Abbildung 16: Links: Punktekonfiguration für $r_{xy} = .7$ mit Regressionsgeraden, Ellipsen und deren Hauptachsen; Die Orientierungen der Regressionsgeraden sind im Allgemeinen nicht identisch mit der Orientierung der ersten Hauptachse, deren Orientierung die der maximalen Ausdehung der Punktekonfiguration ist (s. Satz **??**, Seite 93 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 6.6, Seite 186.



Definition 3.3 *Es sei M eine symmetrische (n,n)- Matrix, und $\boldsymbol{x} = \|\boldsymbol{x}\|\boldsymbol{c}_x \in \mathbb{R}^n$. Dann heißt*

$$R(\boldsymbol{x}) = \frac{\boldsymbol{x}'M\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}} = \boldsymbol{c}_x'M\boldsymbol{c}_x = R(\boldsymbol{c}_x), \quad \boldsymbol{c}_x = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} \tag{3.38}$$

der Rayleigh-Quotient[52]*, wobei $\boldsymbol{c}_c$ der zu $\boldsymbol{x}$ korrespondierende Vektor der Richtungkoeffizienten ist.*

Der Rayleigh-Quotient bezieht sich offenbar nicht auf einen Vektor beliebiger Länge, sondern auf den Orientierungsvektor $\mathbf{c}_x$, für den stets $\|\mathbf{c}_x\| = 1$ gilt, denn

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

Wenn nun gefragt wird, für welchen Vektor $\mathbf{x}$ eine quadratische Form $Q(\mathbf{x}) = \mathbf{x}'M\mathbf{x}$ einen endlichen maximalen Wert annimmt, so ist die Frage nicht besonders sinnvoll, denn ohne weitere Einschränkung wird $Q(\mathbf{x})$ maximal, wenn $\|\mathbf{x}\| = \infty$. Da $\mathbf{x}'M\mathbf{x} = k$, k eine Konstante und M positiv-semidefinit, ein Ellipsoid definiert, so macht die Frage nach dem Maximum von $\mathbf{x}'M\mathbf{x}$ nur Sinn, wenn sie sich auf die Orientierung von $\mathbf{x}$ bezieht, weshalb in (3.38) der normierte Vektor $\mathbf{x}/\|\mathbf{x}\| = \mathbf{c}_x$ betrachtet wird.

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 6.10 mit ebenfalls ausführlichem Beweis gegeben wird. Diesem Satz zufolge gilt

$$\lambda_{\min} \leq R(\mathbf{x}) \leq \lambda_{\max}, \tag{3.39}$$

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 Nebenbedingungen[53] aus:

Satz 3.5 *M sei symmetrisch und positiv-semidefinit, und für die Eigenwerte gelte $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Der Rayleigh-Quotient $R(\boldsymbol{x})$ wird maximal, wenn $\boldsymbol{c}_x = \boldsymbol{v}_1$, $\boldsymbol{v}_1$ der zu λ_1 korrespondierende Eigenvektor von M ist.*

Beweis: Es sei $V = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ die Matrix der Eigenvektoren von M mit den zugehörigen Eigenwerten $\lambda_1 \geq \cdots, \geq \lambda_n$, so dass $MV = V\Lambda$, $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$,

[52] Gelegentlich auch *Rayleigh-Koeffizient* genannt

[53] Den klassischen Beweis (Extremwertbestimmung unter Nebenbedingen) findet man im Anhang, Abschnitt 6.9

und damit $M = V\Lambda V'$. Dann folgt $\mathbf{c}_x' M \mathbf{c}_x = \mathbf{c}_x' V\Lambda V' \mathbf{c}_x$. Es werde $\mathbf{w} = V'\mathbf{c}_x = (w_1, w_2, \ldots, w_n)'$ gesetzt. Dann folgt

$$\mathbf{c}_x' V\Lambda V' \mathbf{c}_x = \mathbf{w}'\Lambda\mathbf{w} = \sum_{i=1}^n \lambda_i w_i^2.$$

Ersetzt man die λ_i, $i > 1$ alle durch λ_1, so erhält man die Ungleichung

$$\sum_{i=1}^n \lambda_i w_i^2 \leq \lambda_1 \sum_{i=1}^n w_i^2.$$

Es ist aber

$$\sum_{i=1}^n w_i^2 = \mathbf{w}'\mathbf{w} = \mathbf{c}_x' VV' \mathbf{c}_x = \mathbf{c}_x'\mathbf{c}_x = 1,$$

da ja $VV' = I$ die Einheitsmatrix (vergl. (3.2), Seite 81). Mithin gilt

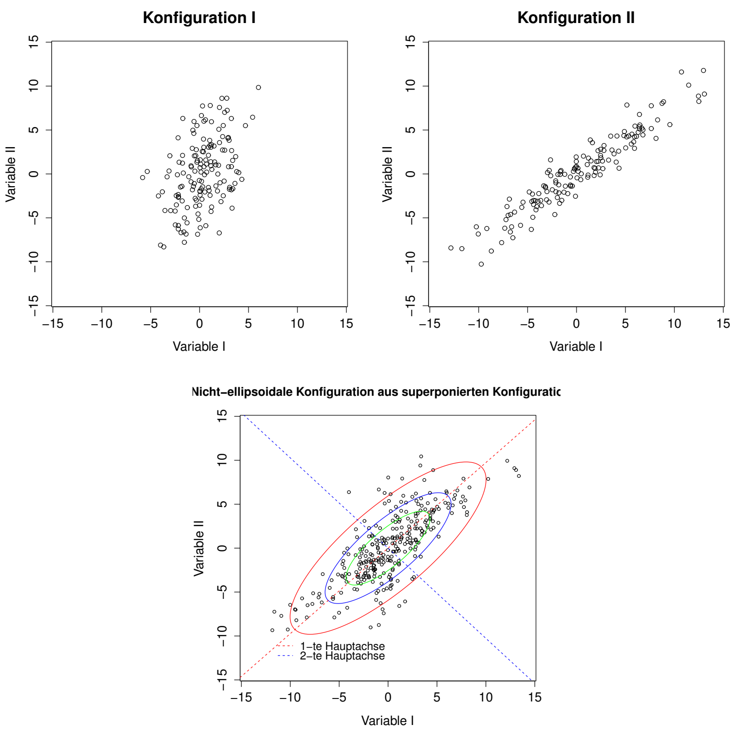
$$\mathbf{c}_x' M \mathbf{c}_x \leq \lambda_1.$$

Es folgt, dass der Wert von λ_1 die obere Grenze für $R(\mathbf{c}_x)$ ist. Nun ist aber $\mathbf{v}_1' M \mathbf{v}_1 = \lambda_1$, $\mathbf{v}_1$ der zu λ_1 korrespondierende Eigenvektor von M. Das heißt aber, dass der Eigenvektor $\mathbf{v}_1$ die Orientierung der größten Ausdehung der Konfiguraion der Fälle kennzeichnet ($\lambda_1 = \|\mathbf{u}_1\|^2$, und $\|\mathbf{u}_1\|$ ist die Länge von $\mathbf{u}_1$). Damit ist gezeigt worden, dass für $\mathbf{v} = \mathbf{v}_1$ der Rayleigh-Koeffizient $R(\mathbf{v}_1)$ maximal ist. □

Anmerkungen:
1. Ist also V die Matrix der Eigenvektoren von $C = X'X$, und berechnet man die Matrix U gemäß $XV = U$, so hat der erste Spaltenvektor $\mathbf{u}_1$ von U die maximale Länge $\lambda_1^{1/2} = \|\mathbf{u}_1\|$, und dem Satz von Courant-Fischer folgend ist $\mathbf{u}_2$ der zweitlängste Spaltenvektor von U, etc.
2. 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 $\mathbf{u}_k$.
3. In Abschnitt 6.9.4, Seite 196, wird der Satz von Courant-Fischer durch Differentiation der quadratischen Form (6.82), Seite 198 (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. □

Teilräume vereinigt in einem Teilraum höherer Dimension: Es sei noch angemerkt, dass die Punktekonfiguration der Fälle keinesfalls ein Ellipsoid sein

Abbildung 17: Superponierte Punktekonfigurationen und Ellipsen

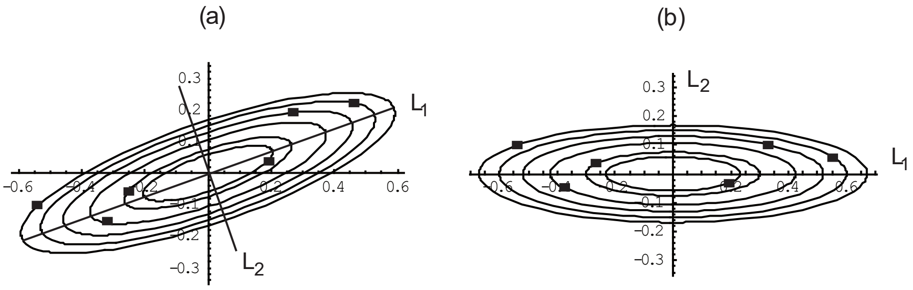


muß, um durch eine Schar von Ellipsoiden (für jeden Fall ein Ellipsoid) beschrieben werden zu können, d.h. die Daten müssen nicht multivariat Gauß-verteilt sein, wie gelegentlich behauptet wird. In Abbildung 17 wird eine Punktekonfiguration gezeigt, die aus zwei Subpopulationen besteht, und jede von ihnen durch ein Ellipsoid repräsentiert wird, wobei sich die Ellipsoide durch die Orientierung ihrer Hauptachsen unterscheiden. Gleichwohl kann die Gesamtstichprobe durch eine Menge von Ellipsoiden mit identischer Orientierung repräsentiert werden derart, dass zu jedem Punkt der Konfiguration ein Ellipsoid existiert und die zueinander korrespondierenden Hauptachsen dieser Ellipsoide dieselbe Orientierung haben, s. Abbildung 17. Im Prinzip hat man es hier mit zwei Teilräumen zu tun (vergl. Abbildung 12, Seite 37, die in einem Teilraum höherer Dimension zusammengefasst werden).

3.2.4 Zur Geometrie der Eigenvektoren symmetrischer Matrizen

Mit der Transformation (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 "laten-

Abbildung 18: 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 6.6 wird die Konstruktion dieser Ellipsen näher erläutert.



ten" U-Koordinatensystem, das durch die transformierten (d.h. rotierten) Hauptachsen der Ellipsoide der Datenkonfiguration definiert ist.

Es seien

$$K_x = \{\tilde{\mathbf{x}}_i,\, 1 \le i \le m\}, \quad K_u = \{\tilde{\mathbf{u}}_i,\, 1 \le i \le m\} \tag{3.40}$$

die Punktekonfigurationen der Fälle im X-, bzw. im U-Koordinatensystem (das auch als $(L_1, \ldots, L_n)$-Koordinatensystem der latenten Variablen aufgefasst werden kann).

Satz 3.6 *Es mögen die Annahmen A1 und A2 gelten. Es ist $X' = VU'$ und damit $\tilde{\boldsymbol{x}}_i = V\tilde{\boldsymbol{u}}_i$, $i = 1, \ldots, m$, V ist die Matrix der Eigenvektoren von $C = X'X$ und Λ ist die Diagonalmatrix der zugehörigen Eigenwerte. Dann folgt*

$$\tilde{\boldsymbol{x}}_i' C \tilde{\boldsymbol{x}}_i = \tilde{\boldsymbol{u}}_i' \Lambda \tilde{\boldsymbol{u}}_i = k_i \in \mathbb{R}, \quad i = 1, \ldots, m \tag{3.41}$$

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. $\mathcal{E}_{u,k_i}$

$$\mathcal{E}_{x,k_i} = \{\tilde{\boldsymbol{x}} | \tilde{\boldsymbol{x}}' C \tilde{\boldsymbol{x}} = k_i\}, \quad \tilde{\boldsymbol{x}}_i \in \mathcal{E}_{x,k_i} \tag{3.42}$$

$$\mathcal{E}_{u,k_i} = \{\tilde{\boldsymbol{u}} | \tilde{\boldsymbol{u}}' \Lambda \tilde{\boldsymbol{u}} = k_i\}, \quad \tilde{\boldsymbol{u}}_i \in \mathcal{E}_{u,k_i} \tag{3.43}$$

Beweis: Aus $\tilde{\mathbf{x}}_i = V\tilde{\mathbf{u}}_i$ folgt $\tilde{\mathbf{x}}_i' = \tilde{\mathbf{u}}_i' V'$ und

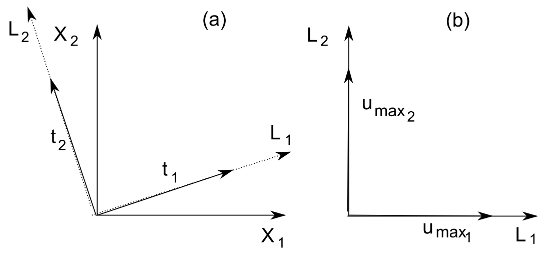
$$\tilde{\mathbf{x}}_i' C \tilde{\mathbf{x}}_i = \tilde{\mathbf{u}}_i' V' C V \tilde{\mathbf{u}}_i = k_i$$

Wegen $V'CV = \Lambda$ (vergl. (3.15), Seite 85) 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 (3.41). □

Abbildung 19: Orientierung der maximalen Ausdehung 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.



Die Ellipsoide $\mathcal{E}_{x,k_i}$ für $i = 1, \ldots, m$ haben alle dieselbe Orientierung, s.a. Abbildung 18. Die Orientierung der Ellipsoide und damit der Konfigurationen ist durch ihre jeweiligen Hauptachsen gegeben. Nach Satz 3.5, Seite 93 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

$$V'\mathbf{t}_1 = \tilde{\mathbf{u}}_{\max_1} = \mathbf{e}_1 = (1, 0, 0, \ldots, 0)' \tag{3.44}$$

$$V'\mathbf{t}_2 = \tilde{\mathbf{u}}_{\max_2} = \mathbf{e}_2 = (0, 1, 0, \ldots, 0)' \tag{3.45}$$

$$\vdots$$

$$V'\mathbf{t}_n = \tilde{\mathbf{u}}_{\max_n} = \mathbf{e}_n = (0, 0, 0, \ldots, 1)' \tag{3.46}$$

Da die $\tilde{\mathbf{u}}_k$, $k = 1, \ldots, n$ die Projektionen der Fälle auf die latenten Achsen $L_1, \ldots, L_n$ sind, sind die Ellipsoide im durch die L_k definierten Koordinatensystem "achsenparallel", d.h. die Hauptachsen sind parallel zu den Koordinatenachsen L_k.

Satz 3.7 *Es sei M eine symmetrische, positiv-semidefinite (n, n)-Matrix mit dem Rang r, V sei die Matrix der Eigenvektoren $\boldsymbol{v}_k$ von M und Λ sei die Diagonalmatrix der zugehörigen Eigenwerte λ_k, $\lambda_j \neq \lambda_j$ für $j \neq k$. m_{ii}, $i = 1, \ldots, n$ seien die Diagonalelemente von M. Dann gilt*

$$\sum_{i=1}^{n} m_{ii} = \sum_{k=1}^{n} \lambda_k. \tag{3.47}$$

Anmerkung: Die Summe $\sum_{i=1}^{n} m_{ii}$ heißt auch *Spur* der Matrix M. Auf Satz 3.47 wird einem anderen Zusammenhang noch einmal zurückgekommen (vergl Satz 3.21, Seite 130). □

Beweis: Die Aussage (3.47) folgt aus dem Sachverhalt, dass die Eigenvektoren reeller, symmetrischer Matrizen stets orthonormal sind. Man hat aus $MV = V\Lambda$ (Definition der Eigenvektoren) die Darstellung $M = V\Lambda V'$. $V\Lambda$ ist die Matrix der mit den Eigenwerten skalierten Spaltenvektoren von V, d.h.

$$V\Lambda = [\lambda_1\mathbf{v}_1, \lambda_2\mathbf{v}_2, \dots, \lambda_n\mathbf{v}_n].$$

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 (3.47) gilt also

$$V_x = \sum_{j=1}^{n} c_{jj} = \sum_{k=1}^{n} \lambda_k \tag{3.48}$$

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

$$\sum_{i=1}^{n} r_{ii} = n = \sum_{k=1}^{n} \lambda_k, \tag{3.49}$$

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 seín, d.h. um eine Ellipse bzw. ein Ellipsoid zu definieren.

Satz 3.8 *Es sei M eine symmetrische (n,n)- Matrix vom Rang $r \leq n$. Dann ist M genau dann positiv semidefinit, wenn eine (n,r)-Matrix G existiert derart, dass*

$$M = GG'. \tag{3.50}$$

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 $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$, $\lambda_j \geq 0$ für $j = 1, \ldots, n$, mit $M = V\Lambda V'$. Es sei

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

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, vergl. Satz 3.10, 105). Dazu betrachte man die Matrix

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

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.

3.2.5 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

$$M = V\Lambda V' = \sum_{k=1}^{n} \lambda_k \mathbf{v}_k \mathbf{v}_k' \tag{3.51}$$

$$M^{-1} = V\Lambda^{-1}V' = \sum_{k=1}^{n} \frac{1}{\lambda_k} \mathbf{v}_k \mathbf{v}_k' \tag{3.52}$$

Die Gleichung (3.51) ergibt sich aus $M = V\Lambda V'$ einfach durch Ausmultiplizieren. Die Gleichung (3.52) ergibt sich wie folgt: Es ist (vergl. Satz 2.59, Seite 79)

$$M^{-1} = (V\Lambda V')^{-1} = (V')^{-1}\Lambda^{-1}V^{-1},$$

wobei $\Lambda^{-1} = \text{diag}(\lambda_1^{-1}, \ldots, \lambda_n^{-1})$. Die Orthonormalität von V impliziert $V' = V^{-1}$, so dass $(V')^{-1} = (V^{-1})^{-1} = V$, so dass die Inverse von M durch (3.52) gegeben ist. Die $\mathbf{v}_k\mathbf{v}_k'$ sind die dyadischen Produkte der Eigenvektoren von M.

Beispiel 3.1 Varianz von Parameterscchätzungen: Die Beziehung (3.52) ist von Bedeutung u.a. bei der Interpretation von Regressionsparametern. Für die Schätzung des Parametervektors $\mathbf{b} = (b_1, \ldots, b_n)'$ 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 6.9.3). Die Varianz-Kovarianzmatrix $D(\hat{\mathbf{b}})$ ist (Seber (1977), p. 48) und wegen (3.52) 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. □

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

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

Es sei nun[54]

$$M^{1/2} =_{\text{Def}} T\Lambda^{1/2}T' = \sum_{k=1}^{n} \sqrt{\lambda_k}\mathbf{t}_k\mathbf{t}_k' \tag{3.53}$$

[54]Mit dem Zeichen $=_{\text{Def}}$ 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

$$(M^{1/2})' = (T\Lambda^{1/2}T')' = T\Lambda^{1/2}T', \tag{3.54}$$

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

$$(M^{1/2})^{-1} = M^{-1/2} = (T\Lambda^{1/2}T')^{-1} = T\Lambda^{-1/2}T' = \sum_{k=1}^{n} \frac{1}{\sqrt{\lambda_k}} \mathbf{t}_k \mathbf{t}_k'. \tag{3.55}$$

3.2.6 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

$$\vec{X} = \mathbf{X} = (X_1, X_2, \ldots, X_n)' \tag{3.56}$$

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

$$X = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{pmatrix} = (X_{ij}), \quad i = 1, \ldots, m; j = 1, \ldots, n \tag{3.57}$$

eine zufällige Matrix. Datenmatrizen sind Beispiele für zufällige Matrizen. Zufällige Vektoren werden mit einer multivariaten Verteilung bzw. Dichtefunktion $f(x_1, \ldots, x_n)$ assoziiert; in Abschnitt 6.13.3 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

$$F(x_1, \ldots, x_n) = P(X_1 \leq x_1, X \leq x_2, \ldots, X_n \leq x_n) \tag{3.58}$$

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 it $\mathbb{E}$ bezeichnet:

$$\mathbb{E}(X) = \begin{cases} \int_{-\infty}^{\infty} xf(x)dx, & X \text{ ist stetig} \\ \\ \sum_i p_i X_i, & X \text{ ist diskret} \end{cases} \tag{3.59}$$

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

$$\mathbb{E}(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i \mathbb{E}(X_i); \tag{3.60}$$

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:

$$\mathbb{E}(\mathbf{X}) = (\mathbb{E}(X_1), \mathbb{E}(X_2), \ldots, \mathbb{E}(X_n))'. \tag{3.61}$$

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

$$\boldsymbol{\mu} = \mathbb{E}(\mathbf{X}) = (\mu_1, \mu_2, \ldots, \mu_n)', \tag{3.62}$$

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

Mit

$$\mathbf{X} - \boldsymbol{\mu} = (X_1 - \mu_1, X_2 - \mu_2, \ldots, X_n - \mu_n)' \tag{3.63}$$

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

$$\mathbf{\Sigma} = \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'] = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)'] = (\sigma_{ij}) \tag{3.64}$$

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

$$\sigma_{ii} = \sigma_i^2 = \mathbb{E}[(X_i - \mu_i)^2] \tag{3.65}$$

Satz 3.9 *Es sei $\boldsymbol{Z} = (Z_1, \ldots, Z_n)'$ ein zufälliger Vektor, dessen Komponenten paarweise unabhängig sind mit $\mathbb{E}(Z_i) = 0$, $Var(Z_i) = 1$, $1 \leq i \leq n$. Weiter sei A eine (n,n)-Matrix und $\boldsymbol{X} = A\boldsymbol{Z} + \boldsymbol{\mu}$. Dann ist $\mathbf{\Sigma} := AA'$ die Matrix der Kovarianzen zwischen den Komponenten von $\boldsymbol{X}$.*

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

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

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. □

Es sei $\mathbf{X} = (X_1, \ldots, X_n)'$ 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 $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$ sei der Vektor der Erwartungswert. $\mathbf{\Sigma}$ sei die (n, n)-Matrix der Varianzen und Kovarianzen der X_j. Die Variablen variieren zufällig und nicht notwendig unabhängig voneinander. $f(X_1, \ldots, X_n)$ sie die gemeinsame Dichtefunktion, die die zufällige Variation charakterisiert. f ist insbesondere die
n-dimensionale Normal- oder Gaußverteilung, wenn

$$f(X_1, \ldots, X_n) = A \exp\left[-\frac{1}{2}(X - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(X - \boldsymbol{\mu})\right], \quad A = \frac{1}{(2\pi)^{\frac{n}{2}}|\mathbf{\Sigma}^{-1}|} \tag{3.66}$$

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*[55].

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 $\tilde{\mathbf{x}}'\mathbf{\Sigma}^{-1}\tilde{\mathbf{x}}$ statt $(X - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(X - \boldsymbol{\mu})$ schreiben. Die Matrix C erscheint als Schätzung für die Varianz-Kovarianzmatrix $\mathbf{\Sigma}$.

Im Abschnitt 3.2.2 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 $f(x_1, \ldots, x_n) = k$ eine Konstante genau dann, wenn

$$(X - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(X - \boldsymbol{\mu}) = k_0 \tag{3.67}$$

k_0 eine Konstante. Das Ellipsoid ist ein Ort gleicher Wahrscheinlichkeit. Die in Abschnitt 3.2.2 betrachteten Ellipsoide sind aber durch

$$(X - \boldsymbol{\mu})'\mathbf{\Sigma}(X - \boldsymbol{\mu}) = \tilde{\mathbf{x}}'C\tilde{\mathbf{x}} = k_1, \quad \mathbf{\Sigma} = C \tag{3.68}$$

definiert; $\tilde{\mathbf{x}} = (X_1 - \mu_1, \ldots, X_n - \mu_n)'$, wobei der Index i für einen bestimmten Fall unterdrückt wurde, $\tilde{\mathbf{x}}$ wird hier als Zufallsvektor aufgefasst. Die Beschreibung der

[55] http://www.uwe-mortensen.de/2dnormalverteilungA.pdf

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 16.)

Der Unterschied zwischen den beiden Definitionen liegt in der definierenden Matrix: bei der Gauß-Verteilung ist es $\mathbf{\Sigma}^{-1}$, bei den nur beschreibenden Ellipsoiden ist es die Matrix $\mathbf{\Sigma}$. Es läßt sich leicht zeigen, dass die Orientierung der Ellipsoide in beiden Fällen identisch ist. So habe $\mathbf{\Sigma}$ die Eigenvektoren V und die zugehörigen Eigenwerte $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$, d.h. es gelte $\Sigma V = V\Lambda$. Dann folgt $\mathbf{\Sigma} = V\Lambda V'$, und nach den Regeln für das Rechnen mit Inversen folgt dann, sofern die Inverse $\mathbf{\Sigma}^{-1}$ überhaupt existiert,

$$\mathbf{\Sigma}^{-1} = (V\Lambda V')^{-1} = (V')^{-1}\Lambda^{-1}V^{-1} = V\Lambda^{-1}V', \tag{3.69}$$

denn wegen der Orthonormalität von V hat man $V^{-1} = V'$, $(V')^{-1} = (V^{-1})^{-1} = V$. Die Inverse $\mathbf{\Sigma}^{-1}$ hat also dieselben Eigenvektoren wie $\mathbf{\Sigma}$, 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}.

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

$$\begin{aligned} (a_{11} - \lambda)u_1 + a_{12}u_2 + \cdots + a_{1m}u_n &= 0 && (3.71)\\ a_{21}u_1 + (a_{22} - \lambda)u_2 + \cdots + a_{2n}u_n &= 0 && (3.72)\\ &\vdots && (3.73)\\ a_{n1}u_1 + a_{n2}u_2 + \cdots + (a_{nn} - \lambda)u_n &= 0 && (3.74) \end{aligned}$$

Solche Gleichungssysteme haben nur dann mindestens eine nicht-triviale Lösung (d.h. eine Lösung, die nicht gleich dem Nullvektor $\vec{0}$ ist), wenn die Koeffizientenmatrix nicht vollen Rang hat, d.h. wenn ihre Determinante verschwindet, so dass

$$|A - \lambda I| = \begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ & & \ddots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{vmatrix} = 0 \tag{3.75}$$

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

$$\begin{aligned} |A - \lambda I| &= (-1)^n[\lambda^n - \beta_1\lambda^{n-1} + \beta_2\lambda^{n-2} + \cdots + (-1)^{n-1}\beta_{n-1}\lambda + (-1)^n\beta_n] \\ &= 0. \end{aligned} \tag{3.76}$$

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 $\lambda_1, \ldots, \lambda_n$ 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 3.10 *Ist ein Eigenwert λ der quadratischen Matrix mit reellen Elementen komplex, so existiert ein zweiter Eigenwert $\bar{\lambda}$, der zu λ konjugiert komplex ist, dh gilt $\lambda = \alpha + i\beta$, so ist auch $\bar{\lambda} = \alpha - i\beta$ ein Eigenwert von A.*

Beweis: Es ist $(A - \lambda I)\mathbf{u} = 0$. Der Übergang zu konjugiert komplexen Zahlen führt zu $(\bar{A} - \bar{\lambda} I)\bar{\mathbf{u}} = 0$. Aber A ist als reell vorausgesetzt worden, also folgt

$$(A - \bar{\lambda} I)\bar{\mathbf{u}} = 0, \tag{3.77}$$

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

$$A\mathbf{u} = \lambda\mathbf{u}. \tag{3.78}$$

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

Es sei $B = A'$; gilt

$$B\mathbf{v} = \mu\mathbf{v}, \tag{3.79}$$

so ist $\mathbf{v}$ ein Eigenvektor von B und μ der zughörige Eigenwert. Es ist

$$(B\mathbf{v})' = \mathbf{v}'B' = \mathbf{v}'A = \mu\mathbf{v}'.$$

$\mathbf{v}$ heißt auch *Linkseigenvektor* von A; $\mathbf{u}$ in (3.78) heißt dementsprechend *Rechtseigenvektor*. Wegen (3.79) übertragen sich alle Aussagen über Rechtseigenvektoren auf Linkseigenvektoren, was allerding nicht bedeutet, dass Links- und Rechtseigenvektoren notwendig identisch sind. Notwendig identisch sind sie nur für den Spezialfall symmetrischer Matrizen. Denn wenn $A' = B = A$ gilt, so folgt aus (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 3.11 *Es sei A eine quadratische, nicht-symmetrische Matrix. Es gelte einerseits $\boldsymbol{v}'A = \mu\boldsymbol{v}'$, andererseits $A\boldsymbol{u} = \lambda\boldsymbol{u}$ mit $\lambda \neq \mu$. Dann folgt $\boldsymbol{u}'\boldsymbol{v} = 0$, d.h. die Links- und Rechtseigenvektoren sind orthogonal zueinander.*

Beweis: Multiplikation von $\mathbf{v}'A = \mu\mathbf{v}'$ von rechts mit $\mathbf{u}$ und von $A\mathbf{u} = \lambda\mathbf{u}$ von links mit $\mathbf{v}'$ liefert

$$\begin{aligned} \mathbf{v}'A\mathbf{u} &= \mu\mathbf{v}'\mathbf{u} = \lambda\mathbf{v}'\mathbf{u} \\ \mathbf{v}'A\mathbf{u} &= \lambda\mathbf{v}'\mathbf{u} = \mu\mathbf{v}'\mathbf{u} \end{aligned}$$

Da $\mathbf{v}'A\mathbf{u} - \mathbf{v}'A\mathbf{u} = 0$ folgt $\lambda\mathbf{v}'\mathbf{u} - \mu\mathbf{v}'\mathbf{u} = (\lambda - \mu)\mathbf{v}'\mathbf{u} = 0$, woraus wegen $\lambda - \mu \neq 0$ die Behauptung $\mathbf{v}'\mathbf{u} = 0$ folgt, d.h. $\mathbf{u}$ und $\mathbf{v}$ sind orthogonal. □

Im Falle nicht-symmetrischer Matrizen sind Links- und Rechtseigenvektoren also verschieden, da sie ja orthogonal zueinander sind. Dieses Resultat bedeutet nicht, dass auch die Rechts- und Linkseigenvektoren untereinander orthogonal zueinander sind. Aber die Gültigkeit des folgenden Satzes läßt sich zeigen:

Satz 3.12 *Es sei A eine nicht-symmetrische, quadratische Matrix mit mehr als einem Rechtseigenvektor. Die Rechtseigenvektoren sind linear unabhängig, sofern die zugehörigen Eigenwerte verschieden sind.*

Beweis: Es seien $\mathbf{u}_1$ und $\mathbf{u}_1$ zwei Rechtseigenvektoren von A mit zugehörigen Eigenwerten $\mu \neq \lambda$. 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 $a_1 A\mathbf{u}_1 + a_2 A\mathbf{u}_2 = 0$, 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

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_r\mathbf{v}_r + a_{p+1}\mathbf{v}_{p+1} = 0 \tag{3.82}$$

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

$$a_1\lambda_1\mathbf{u}_1 + a_2\lambda_2\mathbf{u}_2 + \cdots + a_r\lambda_r\mathbf{u}_r + a_{p+1}\lambda_{p+1}\mathbf{u}_{p+1} = 0. \tag{3.83}$$

Multipliziert man (3.82) mit λ_{p+1} und subtrahiert die Gleichung dann von (3.83), so erhält man

$$a_1(\lambda_1 - \lambda_{p+1})\mathbf{u}_1 + \cdots + a_p(\lambda_p - \lambda_{p+1})u_p = 0,$$

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 $\mathbf{u}_1, \ldots, \mathbf{u}_{p+1}$ 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.

Komplexe Eigenwerte und - vektoren: Es ist bisher stets vorausgesetzt worden, dass für eine gegebene quadratische Matrix Eigenwerte und - vektoren existieren. Die Frage ist aber, ob für eine beliebige quadratische Matrix überhaupt Eigenvektoren existieren müssen. Gegeben sei etwa die Matrix

$$A(\phi) = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \tag{3.84}$$

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 (3.84) erhält man

$$|A - \lambda I| = 4\lambda^2 - 4\lambda\cos\phi + 1 = 0; \tag{3.85}$$

auf die Herleitung des Polynoms wird hier verzichtet, da es an dieser Stelle nur nur auf die Implikationen von (3.85) ankommt. Man findet die Nullstellen

$$\lambda_1 = \frac{1}{2}(\cos\phi + i\sin\phi), \quad \lambda_2 = \frac{1}{2}(\cos\phi - i\sin\phi), \;\; i = \sqrt{-1} \tag{3.86}$$

Die in (3.84) definierte Matrix A hat also zwei komplexe Eigenwerte, relle Eigenwerte ergeben sich nur für solche ϕ-Werte, für die $\sin\phi = 0$ ist, also etwa für $\phi = 0$, wenn gar keine Rotation der Vektoren stattfindet, oder für $\phi = \pi/2$, wenn eine Rotation um 90^o stattfindet.

Es ist also möglich, dass für eine beliebig gewählte quadratische Matrix keine rellen Eigenwerte existieren, dass man aber komplexwertige Eigenwerte finden kann, die als Paare konjugiert komplexer Zahlen auftreten[56]. Nun hätte man noch gerne die zugehörigen Eigenvektoren bestimmt. Für A findet man zwei:

$$\mathbf{u}_1 = \begin{pmatrix} -i \\ 1 \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} i \\ 1 \end{pmatrix}. \tag{3.87}$$

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.

[56]Zwei komplexe Zahlen z und $\bar{z}$ heißen konjugiert komplex, wenn sie sich nur im Vorzeichen des Imaginärteils unterscheiden, wenn also $z = x + iy$ und $\bar{z} = x - iy$ gilt.

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 konjugiert komplexen Elemente $\bar{z} = x - iy$. Sind nur die Imaginärteile iy der Elemente einer Matrix A von Null verschieden, so heißt A *imaginär*; in diesem Fall gilt $A = -\bar{A}$. Die Transponierte $\bar{A}'$ einer Matrix A heißt die mit A *assoziierte* Matrix.

Für symmetrische Matrizen gilt $A' = A$, d.h. $a_{ij} = a_{ji}$ für alle i, j. Gilt für eine Matrix die Aussage $a_{ij} = -a_{ji}$, so heißt A *schief-symmetrisch.*

Ein wichtiger Fall ist durch die Gleichung

$$A = \bar{A}' \tag{3.88}$$

definiert; in diesem Fall heißt A *hermitesch*[57]. Ist $A = \bar{A}$, so sind die Elemente von A alle reell und (3.88) 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 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 3.4 *Es sei V ein Vektorraum und es sei $V_\lambda = \{\boldsymbol{v} \in V | A\boldsymbol{v} = \lambda\boldsymbol{v}\}$. Dann heißt V_λ der* Eigenraum *von A zum Eigenwert λ.*

Bemerkung: Aus $A\mathbf{v} = \lambda\mathbf{v}$ folgt $(A - \lambda I)\mathbf{v} = \vec{0}$. Diese Gleichung ist ein lineares Gleichungssystem in $\mathbf{v}$, d.h. in den Komponenten von $\mathbf{v}$ als Unbekannten. Bekanntlich (s. Definition 2.1, Seite 70, und Definiton 3.9, 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 $V\lambda$ 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

[57]Nach dem französischen Mathematiker Charles Hermite (1822 – 1901)

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 3.5 *Die Matrix A definiere eine Abbildung f des Vektorraums V in sich selbst, d.h. $f : V \to V$, und λ sei ein Eigenwert von A (d.h. von f), und $r(\lambda)$ sei die algebraische Vielfachheit von λ. Der Kern der r-fachen Hintereinanderschaltung von $A - \lambda I$ heißt* Hauptraum zu λ $H(A, \lambda)$

$$H(A, \lambda) = \{\boldsymbol{v} \in V | (A - \lambda I)^r(\boldsymbol{v}) = 0\}. \tag{3.89}$$

Die Elemente von $H(A, \lambda)$ heißen die Hauptvektoren. *$v \in V$ ist* Hauptvektor der Stufe *p, wenn $(A - \lambda I)^p \boldsymbol{v} \neq \vec{0}$.*

Anmerkung: Alle Eigenvektoren sind Hauptvektoren der Stufe $p = 1$. □

Der in der folgenden Definition eingeführte Begriff des invarianten Teilraums ist eine weitere Verallgemeinerung des Begriffs des Eigenraums:

Definition 3.6 *Die Matrix A definiere eine Abbildung eines Vektorraums in sich selbst: $f\colon V \to V$, und es sei $U \subseteq V$. Gilt $f(U) \subseteq U$, d.h. ist die Menge der Vektoren $A\boldsymbol{u}$ wieder eine Teilmenge von U, so heißt U* invarianter Teilraum von V*, oder einfach* f-invariant[58].

Anmerkung: Alle Eigenräume sowie alle Haupträume sind invariante Teilräume.

□

3.3.3 Ähnliche Matrizen

Der Begriff *ähnliche Matrizen* erweist sich in bestimmten Zusammenhängen als nützlich.

[58] oder invariant unter f

Definition 3.7 *Es seien S und T zwei (n,n)-Matrizen und P sei eine (n,n)-Matrix, für die die Inverse P^{-1} existiere, und es gelte die Beziehung*

$$T = P^{-1}SP. \tag{3.90}$$

Dann heißen die Matrizen S und T ähnlich.

Der Begriff der ähnlichen Matrizen ergibt sich im Zusammenhang mit Basistransformationen, etwa für die Vektoren einer (m,n)-Matrix X. So sei etwa X eine Datenmatrix, und es soll untersucht werden, ob sie die Zeilen- oder Spaltenvektoren von X als Linearkombinationen

Die Vektoren $\mathbf{x}$ und $\mathbf{y}$ mögen die Komponenten bzw. Koordinaten x_i bzw. y_i in Bezug auf eine Basis U haben, in Bezug auf eine andere Basis V seien die Vektoren durch $\mathbf{x}^*, \mathbf{y}^*$ mit den Koordinaten x_i^*, y_i^* haben. Es gelte $V = PU$, P die Matrix, mit der U in V transformiert wird. P hat vollen Rang, d.h. $\text{rg}(P) = n$, so dass die Inverse P^{-1} existiert. Betrachtet wird eine Transformation T, die zur Transformation S in der Basis V korrespondiert. Es soll demnach gelten

$$\mathbf{y} = T\mathbf{x}, \quad \mathbf{y}^* = S\mathbf{x}^* \tag{3.91}$$

wobei T und S (n,n)-Matrizen sind. Gesucht ist die Beziehung zwischen T und S. Es gilt dann

$$\begin{aligned}
\mathbf{x}^* &= P\mathbf{x}, \quad \mathbf{y}^* = P\mathbf{y} && (3.92)\\
P\mathbf{y} &\overset{(3.91),(3.92)}{=} SP\mathbf{x} && (3.93)\\
\mathbf{y} &= \underbrace{P^{-1}SP}_{T}\mathbf{x} && (3.94)\\
\Rightarrow T &= P^{-1}SP && (3.95)
\end{aligned}$$

Die Transformation T kann also durch (i) die Transformation der Basis U in die Basis V durch P^{-1}, (ii) die Transformation $\mathbf{y}^*$ in $\mathbf{x}^*$ durch S, und (iii) folgende Transformation durch P erreicht werden.

Satz 3.13 *Es seien S und T ähnliche Matrizen. Dann gilt $rg(S) = rg(T)$ und S und T haben identische Eigenwerte.*

Beweis: (i) Rang: Es sei $\text{rg}(T) = r \leq n$. Es gilt $P^{-1}SP = (P^{-1}S)P$. Dem Rangsatz zufolge gilt $\text{rg}(P^{-1}SP) \leq \min(\text{rg}(P^{-1}S), \text{rg}(P))$. P hat vollen Rang, d.h. $\text{rg}(P) = n \geq r$. Also kann $(P^{-1}S) < r$ nur gelten, wenn $\text{rg}(S) < r$. Aber $T = P^{-1}TS$, also $\text{rg}(T) = \text{rg}(P^{-1}SP) = r$, so dass $\text{rg}(S) = r$.

(ii) Eigenwerte: Es gelte $S\mathbf{u} = \lambda\mathbf{u}$, $\mathbf{u}$ ein Eigenvektor von T, und λ der zugehörige Eigenwert; die Eigenwerte seien paarweise verschieden. Dann gilt auch:

$$\begin{aligned}
PTP^{-1} &= S\\
PTP^{-1}\mathbf{u} &= S\mathbf{u} = \lambda\mathbf{u}\\
\Rightarrow TP^{-1}\mathbf{u} &= P^{-1}\lambda\mathbf{u} = \lambda P^{-1}\mathbf{u}
\end{aligned}$$

d.h. $P^{-1}\mathbf{u}$ ist ein Eigenvektor von T mit zugehörigem Eigenwert λ, d.h. S und T haben dieselben Eigenwerte aber unterschiedliche Eigenvektoren. Für Matrizen mit identischen (mehrfachen) Eigenwerten lässt sich zeigen, dass dieses Resultat nicht nowendig gilt[59]. □

Beispiel: symmetrische Matrizen Es sei X eine (m,n)-Matrix U und eine (n,n)-Matrix V derart, dass $X = UV'$. Mit den Annahmen (3.10) und (3.11), Seite 84, erhält man daraus

$$V'X'XV = U'U = \Lambda = \mathrm{diag}(\lambda_1, \dots, \dots, \lambda_n) \tag{3.96}$$

Wegen $V'V = I$ gilt $V' = V^{-1}$, so dass $X'X$ und Λ ähnlich im Sinne der Definition 3.90. Man sagt auch, dass $X'X$ durch V *diagonalisiert* wird.

Eine Übung: Es sei X eine (m,n)-Matrix mit dem Rang $r \leq \min(m,n)$; man kann stets eine[60] (m,n)-Matrix U und eine (n,n)-Matrix V finden derart, dass $X = UV'$. Es mögen nun die Annahmen $V'V = I$ und $U'U = \Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$ gelten (vergl Seite 84), – sollte der Rang r von X kleiner als $\min(m,n)$ sein, so gilt $\lambda_{r+1}, \dots, \lambda_n = 0$. Es folgt $X' = VU'$; für einen Spaltenvektor $\tilde{\mathbf{x}}_j$ gilt $\tilde{\mathbf{x}}_j = V\tilde{\mathbf{u}}_j$; allgemein gelte

$$\tilde{\mathbf{x}} = V\tilde{\mathbf{u}}_x \quad \tilde{\mathbf{y}} = V\tilde{\mathbf{u}}_y. \tag{3.97}$$

Weitere seien T und S (n,n)-Matrizen derart, dass

$$\tilde{\mathbf{x}} = T\tilde{\mathbf{y}}, \quad \tilde{\mathbf{u}}_x = S\tilde{\mathbf{u}}_y. \tag{3.98}$$

Es soll gezeigt werden, dass eine spezielle Wahl von S die Matrix T impliziert, insbesondere soll (3.96) gezeigt werden.

$$\begin{aligned}
\tilde{\mathbf{x}} &\overset{(3.98)}{=} T\tilde{\mathbf{y}} \\
V\tilde{\mathbf{u}}_x &\overset{(3.97)}{=} TV\tilde{\mathbf{u}}_y \\
\tilde{\mathbf{u}}_x &\overset{(3.98)}{=} \underbrace{V'TV}_{S}\tilde{\mathbf{u}}_y \\
S &= V'TV && (3.99) \\
VSV' &= T && (3.100)
\end{aligned}$$

Nun werde von der Annahme A2 Gebrauch gemacht, d.h. es soll $U'U = \Lambda$, Λ eine Diagonalmatrix, gelten, und es werde $S = \Lambda$ gesetzt. Dann folgt aus (3.100) wegen $XV = U$

$$\begin{aligned}
T &= V\Lambda V', \\
T &= VU'UV' = VV'X'XVV', \text{ d.h.} \\
T &= X'X, && (3.101)
\end{aligned}$$

[59]vergl. SimilarAndNonsimilarMatrices.pdf
[60]tatsächlich beliebig viele Matrizen

d.h. zu $S = \Lambda$ korrespondiert die Transformation $\mathbf{x} = (X'X)\mathbf{y}$, wobei $\mathbf{y}$ ein beliebiger Vektor aus $\mathbb{R}^n$ sein kann. Wählt man speziell $\mathbf{y} = \mathbf{v}_k$, $\mathbf{v}_k$ der k-te Spaltenvektor von V, so folgt $\mathbf{y} = \lambda_k \mathbf{v}_k$, d.h. als Transformationsmatrix bewirkt $X'X$ nur eine Längenskalierung von $\mathbf{v}_k$, ohne Veränderung der Orientierung. Betrachtet man allgemein $(X'X)V$, so implizieren A1 und A2 die Beziehungen $X'X = VU'UV' = V\Lambda V'$, also $(X'X)V = V\Lambda$, d.h. die Rotationsmatrix V ist die Matrix der Eigenvektoren von $X'X$, und $X'X$ und Λ sind ähnliche Matrizen.

Natürlich kann man auch die Matrix T vorgeben und die zugehörige Matrix S bestimmen, gegeben die Annahmen A1 und A2. Es sei also $T = X'X$. Die Gleichung (3.99) liefert dann

$$S = V'(X'X)V = V'(VU'UV')V = \Lambda. \tag{3.102}$$

□

3.3.4 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 3.8 *Es seien A und B symmetrische, positiv semidefinite Matrizen. Dann repräsentiert*

$$A\boldsymbol{w} = \lambda B\boldsymbol{w} \tag{3.103}$$

das generalisierte Eigenvektorproblem.

Der generalisierte Rayleigh-Quotient: Der *generalisierte Rayleigh-Quotient* ist durch

$$\rho(\mathbf{w}) = \frac{\mathbf{w}'A\mathbf{w}}{\mathbf{w}'B\mathbf{w}} \tag{3.104}$$

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

$$B^{1/2}B^{1/2} = B = P\Lambda^{1/2}P'P\Lambda^{1/2}P' = P\Lambda P',$$

denn $P'P = I$ die Einheitsmatrix, und P die Matrix der Eigenvektoren von B.. Es sei $\mathbf{v} = B^{1/2}\mathbf{w}$. Dann ist $\mathbf{w} = B^{-1/2}\mathbf{v}$ und man erhält

$$\rho(\mathbf{w}) = \frac{\mathbf{w}'A\mathbf{w}}{\mathbf{w}'B\mathbf{w}} = \frac{\mathbf{v}'B^{-1/2}AB^{-1/2}\mathbf{v}}{\mathbf{v}'\mathbf{v}} = \frac{\mathbf{v}'M\mathbf{v}}{\mathbf{v}'\mathbf{v}}. \tag{3.105}$$

$M = B^{-1/2}AB^{-1/2}$ 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 (3.105) gemäß $\mathbf{w} = B^{-1/2}\mathbf{v}$.

Die Gleichung (3.103) führt durch Multiplikation von links mit B^{-1} auf

$$B^{-1}A\mathbf{w} = \lambda\mathbf{w}, \tag{3.106}$$

d.h. $\mathbf{w}$ ist ein Eigenvektor der nicht-symmetrischen Matrix $BV^{-1}A$, und λ ist der zugehörige Eigenwert. Multipliziert man diese Gleichung von links mit $B^{1/2}$, so erhält man

$$B^{1/2}B^{-1}A\mathbf{w} = B^{-1/2}A\mathbf{w} = \lambda B^{1/2}\mathbf{w}.$$

und nochmalige Multiplikation von links mit $B^{-1/2}$ führt zu

$$B^{-1/2}AB^{-1/2}\mathbf{v} = \lambda\mathbf{v}. \tag{3.107}$$

Damit hat man mit (3.107) ein Eigenwert- und Eigenvektorproblem der bekannten Art für symmetrische Matrizen.

Die Eigenvektoren $\mathbf{v}_j$ symmetrischer Matrizen sind bekanntlich orthonormal. Die Lösung für den generaliserten Rayleigh-Quotienten durch die Lösung für den gewöhnlichen Rayleigh-Quotienten in einem *transformierten Raum* gegeben ist. Man kommt damit zu der Aussage (Shaw-Taylor & Christianini (2004), p. 162)

Satz 3.14 *Ein beliebiger Vektor $\boldsymbol{v}$ kann als Linearkombination der $\boldsymbol{w}_j$, $j = 1, \ldots, k$ angeschrieben werden. Für die Eigenvektoren des generalisierten Eigenvektorproblems $A\boldsymbol{w} = \lambda B\boldsymbol{w}$ gelten die Relationen*

$$\boldsymbol{w}_i'B\boldsymbol{w}_j = \delta_{ij}, \quad \boldsymbol{w}_i'A\boldsymbol{w}_j = \delta_{ij}\lambda_i, \quad \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \tag{3.108}$$

Beweis: Es war $\mathbf{v} = B^{1/2}\mathbf{w}$, und als Lösungen von (3.107) 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. (3.106)) und damit $B\mathbf{w} = (1/\lambda)A\mathbf{w}$,

$$0 = \mathbf{v}_i'\mathbf{v}_j = \mathbf{w}_i'B^{1/2}B^{1/2}\mathbf{w}_j = \mathbf{w}_i'B\mathbf{w}_j = \frac{1}{\lambda_j}\mathbf{w}_iA\mathbf{w}_j;$$

nach (3.103) gilt ja $A\mathbf{w} = \lambda B\mathbf{w}$ und deshalb $(1/\lambda_i)A\mathbf{w} = B\mathbf{w}$. Damit gilt (3.108) 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

$$\lambda_i = \lambda_i\mathbf{v}_i'\mathbf{v}_i = \lambda_i\mathbf{w}_i'B^{1/2}B^{1/2}\mathbf{w}_i = \lambda_i\mathbf{w}_i'B\mathbf{w}_i = \mathbf{w}_iA\mathbf{w}_i,$$

und dies ist (3.108) für den Fall $i = j$. □

Die Maximierung von (3.108) (Maximierung unter Nebenbedingungen, S. Anhang) führt auf die Gleichung (3.106).

Satz 3.15 *Für den generalisierten Rayleigh-Quotienten gilt*

$$\rho_1 \leq \rho \leq \rho_2, \tag{3.109}$$

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 198).

Satz 3.16 *Gilt $A\boldsymbol{v} = \lambda B\boldsymbol{v}$ und sind λ und $\boldsymbol{v}$ die Eigenwerte und Eigenvektoren für den generalisierten Rayleigh-Quotienten, so kann A gemäß*

$$A = \sum_{j=1}^{r}\lambda_j B\boldsymbol{v}_j(B\boldsymbol{v}_j)' \tag{3.110}$$

zerlegt werden.

Beweis: Für eine beliebige symmetrische Matrix C mit der Matrix P der Eigenvektoren und der Diagonalmatrix $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ der Eigenwerte gilt bekanntlich $C = \sum_j \lambda_j\mathbf{p}_j\mathbf{p}_j'$. Die Matrix $B^{-1/2}AB^{-1/2}$ ist symmetrisch, mithin gilt

$$B^{-1/2}AB^{-1/2} = \sum_{j=1}^{r}\lambda_j\mathbf{v}_j\mathbf{v}_j'.$$

Multipliziert man von links mit $B^{1/2}$ und von rechts ebenfalls mit $B^{1/2}$, so folgt

$$A = \sum_{j=1}^{r}\lambda_j B^{1/2}(B^{1/2}\mathbf{v}_j\mathbf{v}_j)' = \sum_j \lambda_j B\mathbf{w}_j(B\mathbf{w}_j)',$$

und das war zu zeigen. □

3.4 Lineare Gleichungssysteme

3.4.1 Allgemeine Aussagen

Bei der Einführung des Begriffs des Vektoraums (Abschnitt 1.3.1) wurde darauf hingewiesen, das Linearkombinationen Systeme von linearen Gleichungen repräsentieren: gegeben seien etwa n m-dimensionale Vektoren $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Die lineare Hülle $\mathcal{L}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$, also die Menge aller Linearkombinationen der $\mathbf{x}_j$, ist ein (Teil-)Vektorraum des $\mathbb{R}^m$. Gibt man nun einen Vektor $\mathbf{v} \in \mathbb{R}^m$ vor, so ist

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

ein System von m Gleichungen mit den Unbekannten $a_1, \ldots, a_n$. Das System hat sicher eine Lösung $\mathbf{a} = (a_1, \ldots, a_n)'$, wenn $\mathbf{v} \in \mathcal{L}$, denn dann ist $\mathbf{v}$ ja ein Element der Menge der von den $\mathbf{x}_j$ erzeugten Vektoren. Wegen Satz 1.10, Seite 46, kann man außerdem sagen, dass der Lösungsvektor $\mathbf{a}$ eindeutig ist, wenn die $\mathbf{x}_j$ linear unabhängig sind, d.h. es gibt dann nur einen Lösungsvektor $\mathbf{a}$. Offen ist die Frage, wieviel Lösungsvektoren es gibt, wenn die $\mathbf{x}_j$ linear abhängig sind. Keine Lösung gibt es jedenfalls, wenn $\mathbf{v} \notin \mathcal{L}$.

Üblicherweise wird die Unbekannte in einer Gleichung mit x bezeichnet, bzw. mit dem Vektor $\mathbf{x}$, wenn es sich um eine System von Gleichungen handelt. Um eine leichte Vergleichbarkeit mit anderen Texte zu ermöglichen wird jezt eine entsprechende Umbennung vorgenommen: Die bis jetzt mit $\mathbf{x}_j$ bezeichneten Vektoren mit $\mathbf{a}_j$ bezeichnet. Betrachtet werden also Linearkombinationen

$$\mathbf{y} = x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n, \quad \mathbf{y}, \mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m, \tag{3.111}$$

und $\mathbf{x} = (x_1, \ldots, x_n)'$ ist der Lösungsvektor. Fasst man die $\mathbf{a}_j$ zu einer Matrix $A = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$ zusammen, so kann man das der Gleichung (3.111) entsprechende Gleichungssystem in der Form

$$A\mathbf{x} = \mathbf{y}, \quad \mathbf{x} \in \mathbb{R}^n, \;\; \mathbf{y} \in \mathbb{R}^m \tag{3.112}$$

schreiben; A ist eine (m, n)-Matrix. Die Antwort auf die Frage nach der Existenz und Anzahl von Lösungen hängt von der linearen Abhängigkeit bzw. Unabhängigkeit der Spaltenvektoren $\mathbf{a}_j$ von A ab. Man kann sagen, dass sie vom Rang r der Matrix A abhängt.

Zunächst werde zwischen zwischen zwei Arten von Gleichungssystemen unterschieden: Das Gleichungssystem heißt

homogen, wenn $\mathbf{y} = \vec{0}$,
inhomogen, wenn $\mathbf{y} \neq \vec{0}$.

Definition 3.9 *Es sei A eine (m, n)-Matrix mit dem Rang $r = rg(A) \leq \min(m, n)$. und $A\boldsymbol{x} = \boldsymbol{y}$ sei ein System von Gleichungen. Weiter sei*

$$\begin{aligned} kern(A) &= \{\boldsymbol{x} \in \mathbb{R}^n | A\boldsymbol{x} = \vec{0}\} & (3.113) \\ \mathcal{L}(A) &= \{\boldsymbol{y} \in \mathbb{R}^m | A\boldsymbol{x} = \boldsymbol{y}\}. & (3.114) \end{aligned}$$

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

Folgerungen:

1. Ist $m \geq n$ und gilt $r = n$ (A hat den Rang n), so hat das homogene System $A\mathbf{x} = \vec{0}$ nur die Lösung $\mathbf{x} = \vec{0}$, denn dann sind die Spaltenvektoren $\mathbf{a}_j$ von A linear unabhängig.

2. Is $m = n$ und $\text{rg}(A) = n$, so existiert die zu A inverse Matrix A^{-1} und aus $A\mathbf{x} = \mathbf{y}$ folgt die Lösung
$$\mathbf{x} = A^{-1}\mathbf{y}. \tag{3.115}$$

3. $\text{kern}(A)$ ist ein (Teil-)Vektorraum: sind die Vektoren $\mathbf{x}_1$ und $\mathbf{x}_2$ Elemente (Lösungen) aus $\text{kern}(A)$, so rechnet man leicht nach, dass dann auch $b_1\mathbf{x}_1 + b_2\mathbf{x}_2 \in \text{kern}(A)$ $(b_1, b_2 \in \mathbb{R})$ gilt[61].

4. 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 ist, d.h. $\mathbf{y} \in \mathcal{L}(A)$. □

Der folgende Satz gilt für beliebige (m, n)-Matrizen A.

Satz 3.17 *Es sei A eine (m, n)-Matrix mit der SVD $A = Q\Sigma T'$, wobei Q aus den Spaltenvektoren $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ und T aus den Spaltenvektoren $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n$ bestehe; ist $rg(A) = r \leq \min(m, n)$, so sind r Singularwerte σ_k größer als Null und die restlichen sind gleich Null. Dann gilt*

$$\begin{aligned} \mathcal{L}(A) &= \mathcal{L}(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_r) && (3.116)\\ \mathit{kern}(A) &= \mathcal{L}(\boldsymbol{t}_{r+1}, \ldots, \boldsymbol{t}_s), \quad s = \min(m, n) && (3.117)\\ \mathit{rg}(\mathit{kern}(A)) + \mathit{rg}(\mathcal{L}(A)) &= \min(m, n) && (3.118) \end{aligned}$$

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 (3.116) gelten muss.

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

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

geschrieben werden kann, wobei die $c_j \in \mathbb{R}$ geeignet gewählte Koeffizienten sind. Dann gilt

$$A\mathbf{x} = A\sum_{j=1}^{n} c_j\mathbf{t}_j = \sum_{j=1}^{n} c_j A\mathbf{t}_j = \sum_{j=1}^{r} c_j\sigma_j\mathbf{q}_j = \vec{0},$$

[61] In Definition 6.5, Punkt 4., Seite 203 wird der Begriff des Kerns einer Abbildung eingeführt. Die Matrix A definiert eine Abbildung, und (4.4) definiert damit den Kern einer Abbildung.

denn $\sigma_j = 0$ für $j > r$ (falls $r < \min(m, n)$). Wegen der linearen Unabhängigkeit der $\mathbf{q}_j$ kann diese Gleichung nur gelten, wenn $c_1 = \cdots = c_r = 0$. Dann kann $\mathbf{x} \neq \vec{0}$ kein Element des durch die $\mathbf{t}_1, \ldots, \mathbf{t}_r$ aufgespannten Vektorraums sein, sondern muss ein Element des $(n-r)$-dimensionalen Komplementärraums sein. Die $\mathbf{t}_{r+1}, \ldots, \mathbf{t}_n$ sind eine Basis für diesen Komplementärraum, so dass man

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

ansetzen kann, und

$$A\mathbf{x} = A\sum_{j=r+1}^{n} c_j\mathbf{t}_j = \sum_{j=r+1}^{n} c_j A\mathbf{t}_j = \sum_{j=r+1}^{n} c_j\sigma_j\mathbf{q}_j = \vec{0},$$

weil $\sigma_j = 0$ für $j > r$, falls $r < \min(m, n)$. Alle Linearkombinationen von Vektoren $\mathbf{x} \neq \vec{0}$ mit $A\mathbf{x} = \vec{0}$ sind Linearkombinationen der $\mathbf{t}_{r+1}, \ldots, \mathbf{t}_n$, und dies ist die Aussage von (3.117).

Die Gleichung (3.118) ist eine unmittelbare Folge der vorangegangenen Argumente: $\mathcal{L}(A)$ hat den Rang r und $\mathrm{kern}(A)$ hat den Rang $n-r$, so dass die Summe der Ränge gleich n sein muss. $\square$

Anmerkung: Der Satz 3.17 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 $\mathrm{kern}(A)$ 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 6.11, Seite 200. $\square$

Die allgemeine Lösungsmenge wird im folgenden Satz spezifiziert:

Satz 3.18 *Es sei $A\boldsymbol{x} = \boldsymbol{y} \in \mathcal{L}(A)$, $rg(A) = r$, und insbesondere sei $\boldsymbol{x} = \boldsymbol{x}_0$ eine bestimmte Lösung, so dass $A\boldsymbol{x}_0 = \boldsymbol{y}$ gilt. Der Kern kern(A) besteht aus dem $(n-r)$-dimensionalen Teilraum $\mathcal{L}_{n-r} = \mathcal{L}(\boldsymbol{t}_{r+1}, \ldots, \boldsymbol{t}_n)$ des V_n. Dann ist die Menge der Lösungsvektoren durch*

$$\mathcal{L} = \{\boldsymbol{x}_0 + \boldsymbol{x} | \boldsymbol{x} \in \mathcal{L}_{n-r}\} \tag{3.119}$$

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. □

1. Der Fall $m = n = r$**:** Ist $m = n = r$, r der Rang von A, so existiert die Inverse A^{-1} und aus $A\mathbf{x} = \mathbf{y} \in V_n^r = \mathcal{L}(A)$ folgt sofort die Lösung

$$\mathbf{x} = A^{-1}\mathbf{y}, \quad \mathbf{y} \in \mathcal{L}(A. \tag{3.120}$$

2. Der Fall $m > n = r$**:** Es gilt $A\mathbf{x} = \mathbf{y}$, A ist eine (m, r)-Matrix, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$. Offenbar gibt es mehr Gleichungen als Unbekannte, und es gibt zwei Fälle:

$$\mathbf{y} \begin{cases} \in \mathcal{L}(A) \\ \notin \mathcal{L}(A) \end{cases}$$

Es gelte $\mathbf{y} \in \mathcal{L}(A)$, $\text{rg}(A) = n$. In diesem Fall existiert $\mathbf{x}$, und man findet diesen Vektor, indem man die Gleichung $A\mathbf{x} = \mathbf{y}$ von links mit A' multipliziert: $A'A\mathbf{x} = A'\mathbf{y}$, und dann noch einmal von links mit $(A'A)^{-1}$, so dass man für $\mathbf{x}$ die Lösung

$$\mathbf{x} = (A'A)^{-1}A'\mathbf{y} \tag{3.121}$$

erhält $\text{rg}(A'A) = \text{rg}(A) = n$, d.h. $(A'A)^{-1}$ existiert.

3. Der Fall rg$(A) < \min(m, n)$**:** Es gelte $\mathbf{y} \in \mathcal{L}(A)$. In diesem Fall ist $\text{rg}(A'A) = r < \min(m, n)$ und die Inverse $(A'A)^{-1}$ existiert nicht. Für den Fall $r < n$ liefert (3.119) den Lösungsraum. Auch hier kann eine Pseudoinverse gefunden werden, vergl. (3.145), Seite 127.

Nun gelte **Der Fall 4.** $\mathbf{y} \notin \mathcal{L}(A), \mathbf{rg}(A) = n$**:**. In diesem Fall existiert kein Lösungsvektor $\mathbf{x}$, 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. (2.34), Seite 72) existiert die zu $A'A$ inverse Matrix $(A'A)^{-1}$, so dass

$$\hat{\mathbf{x}} = (A'A)^{-1}A'\mathbf{y} \tag{3.122}$$

resultiert. Die Matrix $(A'A)^{-1}A'$ ist eine *Pseudoinverse* für A, s. Seite 126.

Sind die Komponenten von $\mathbf{y}$ Messwerte, so sind sie üblicherweise durch Messfehler kontaminiert, so dass $\mathbf{y} \notin \mathcal{L}(A)$. (3.122) liefert dann keine Lösung $\mathbf{x}$, die allen m Gleichungen genügt. (3.122) ist die bekannte Kleinste-Quadrate-Schätzung $\hat{\mathbf{x}}$ für $\mathbf{x}$, vergl. (6.65), Seite 195, und Abschnitt **??**.

Cramersche Regel: Diese Regel wird hier nur der Vollständigkeit wegen genannt. Es sei $A\mathbf{x} = \mathbf{y}$ ein Gleichungssystem, wobei A eine (n, n)-Matrix sei. Dann gilt für die j-te Komponente x_j von $\mathbf{x}$

$$x_j = \frac{|A_j|}{|A|}, \quad j = 1, \ldots, n \tag{3.123}$$

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 6.13, 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.

3.4.2 Die Regularisierung von Matrizen

Gegeben sei ein lineares Gleichungssystem $\mathbf{y} = A\mathbf{x}$ mit $\mathbf{x}$; dabei ist $\mathbf{x}$ der Vektor, dessen Komponenten die Unbekannten des Systems sind und A ist eine (m,n)-Matrix. Gilt $m = n$ und hat A den vollen Rang $r = n$, so existiert die Inverse A^{-1} und man hat die Lösung $\mathbf{x} = A^{-1}\mathbf{x}$, und für den Fall $m > n$ und $\text{rg}(A) = n$ hat man nach (3.121) die Lösung

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

und

$$\hat{\mathbf{x}} = (A'A)^{-1}A'\mathbf{y}, \quad \mathbf{y} \notin \mathcal{L}(A)$$

nach (3.122); $\hat{\mathbf{x}}$ ist die beste Lösung im Sinne einer Approximation im Sinne der Methode der kleinsten Quadrate. $\hat{\mathbf{x}}$ ist die Lösung für die Regressionskoeffizienten für den Fall, dass $A = X$ die Matrix der Prädiktorvariablen ist, wenn man also die Regressionskoeffizienten für eine multiple Regression schätzen will.

Lösungen bzw. Approximationen könnnen also berechnet werden, wenn die Spalten von A linear unabhängig sind und dementsprechend die Inversen A^{-1} bzw. $(A'A)^{-1}$ berechnet werden können. Hat eine Matrix den vollen Rang, so sind die Eigenwerte von A bzw. $A'A$ von Null verschieden, $A'A$ ist positiv-definit. Ist $A'A$ aber positiv semi-definit, so ist zumindest ein Eigenwert gleich Null und die Inverse $(A'A)^{-1}$ existiert nicht. Aber schon wenn alle Eigenwerte numerisch von Null verschieden sind, aber einige der Eigenwerte "klein" im Vergleich zu anderen sind können die Berechnungen ungenau werden; für die Praxis, z.B. bei der Schätzung von Regressionsparametern, ist das ein störender Effekt, weil die auf diesen Berechnungen beruhenden Voraussagen von Werten abhängiger Variablen dann ungenau werden. Man sagt, das Problem (etwa die Schätzung des Vektors $\mathbf{b}$) sei "schlecht formuliert" ("ill posed").

Ein für die folgenden Betrachtungen relevantes Beispiel ist die multiple Regression, bei der eine abhängige Variable Y durch eine Linearkombination von "Prädiktorvariablen" $X_1, \ldots, X_n$ gemäß

$$Y = b_1X_1 + \cdots + b_nX_n + e$$

vorausgesagt werden soll; e steht hier wie üblich für "error" und die die $b_1, \ldots, b_n$ sind zu schätzende Parameter. Fasst man die X_j zu einer (m,n)-Matrix $X = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$ und die b_j zu einem Vektor $\mathbf{b} = (b_1, \ldots, b_n)'$ zusammen, so erhält man die Gleichung

$$\mathbf{y} = X\mathbf{b} + \mathbf{e},$$

$\mathbf{y}, \mathbf{e} \in \mathbb{R}^m$. Die KQ-Schätzung für $\mathbf{b}$ ist

$$\hat{\mathbf{b}} = (X'X)^{-1}X'\mathbf{Y}. \tag{3.124}$$

Üblicherweise ist $m > n$ und der Rang von X ist $r \leq \min(m, n)$. Wegen der unvermeidlichen Messfehler gilt numerisch $r = n$. Gibt es aber größere Korrelationen zwischen den X_j (Kollinearitäten), so sind einige der Eigenwerte von $X'X$ klein und das Schätzproblem ist 'ill-posed', was zu Fehlschätzungen für Y führt. Eine mögliche Lösung für dieses Problem ist die *Regularisierung* von $X'X$.

Tychonow-Regularisierung Ein Ansatz, stabile Lösungen für schlecht konditionierte Gleichungssysteme zu finden, geht auf Tychonow (1943)[62] zurück, dessen Arbeit aber erst 1963 in übersetzter Form vorlag; die allgemeine Version erschien in Tychonow & Arsenin (1977). Im hier gegebenen Zusammenhang hat man das Modell $\mathbf{Y} = X\mathbf{b} + \mathbf{e}$, und man möchte $\mathbf{b}$ so bestimmen, dass $\mathbf{e}'\mathbf{e}$ minimal wird, wobei $\mathbf{e}'\mathbf{e} = (\mathbf{Y} - X\mathbf{b})'(\mathbf{Y} - X\mathbf{b}) = \|\mathbf{Y} - X\mathbf{b}\|^2$. Ist X schlecht konditioniert, kann man nach Tychonow vom Ansatz

$$\|X\mathbf{b} - \mathbf{Y}\|^2 + \|\Gamma\mathbf{b}\|^2 \stackrel{!}{=} \min \tag{3.125}$$

ausgehen. Hierin ist Γ eine geeignet gewählte Matrix, die *Tychonow-Matrix.* $\|\Gamma\mathbf{b}\|^2$ heißt *Regularisierungsterm*, und der Ansatz (3.125) *Tychonow-Regularisierung.* Der Term $\|\Gamma\mathbf{b}\|^2$ heißt auch *Strafterm* (*penalty.*

Ridge-Regularisierung: Wohl unabhängig von Tychonow sind Hoerl & Kennard (1970) auf die gleiche Idee gekommen. Sie haben insbesondere $\Gamma = \lambda I$, $0 < \lambda \in \mathbb{R}$, I die Einheitsmatrix, gesetzt und sprechen dann von *Ridge-Regression.* Dazu wird der Begriff der *Verlustfunktion* (*Loss function*) eingeführt. Für die gewöhnliche KQ-Schätzung ist sie durch

$$Q(\mathbf{b}) = (\mathbf{y} - X\mathbf{b})'(\mathbf{y} - X\mathbf{b}) = \mathbf{e}'\mathbf{e} = \|\mathbf{y} - X\mathbf{b}\|^2$$

gegeben. Je größer im Durchschnitt die Differenzen $\mathbf{y} - X\mathbf{b}$ sind, desto größer ist der Verlust $Q(\mathbf{b})$.

Hoerl & Kennard haben durch Hinzufügen eines Strafterms ("penalty") eine andere Verlustfunktion eingeführt:

$$\min_{\mathbf{b}} Q_r(\mathbf{b}, \lambda) = \|\mathbf{y} - X\mathbf{b}\|^2 + \min_{\mathbf{b}} \lambda\|\mathbf{b}\|^2. \tag{3.126}$$

wobei Q_r für Q-ridge-regression steht. Multipliziert man den Ausdruck auf der rechten Seite aus, so erhält man

$$\begin{aligned} Q_r(\mathbf{b}, \lambda) &= \mathbf{b}'X'X\mathbf{b} - \mathbf{b}'X'\mathbf{y} - \mathbf{y}'X\mathbf{b} + \mathbf{y}'\mathbf{y} + \lambda\|\mathbf{b}\|^2 \\ &= \mathbf{b}'X'X\mathbf{b} - 2\mathbf{b}'X'\mathbf{y} + \mathbf{y}'\mathbf{y} + \lambda\|\mathbf{b}\|^2 \end{aligned}$$

Differenziert man bezüglich des Vektors $\mathbf{b}$, so erhält man

$$\frac{sQ_r}{d\mathbf{b}} = 2X'X\mathbf{b} - 2X'\mathbf{y} + 2\lambda\mathbf{b}.$$

[62] Andrey Nikolayevitch Tychonow (1906 – 1993), russischer Mathematiker

Es ist $dQ_r/d\mathbf{b} = 0$ für $\mathbf{b} = \hat{\mathbf{b}}_\lambda$, und man erhält

$$X'X\hat{\mathbf{b}}_\lambda + \lambda\hat{\mathbf{b}}_\lambda = (X'X + \lambda I_n)\hat{\mathbf{b}}_\lambda = X'\mathbf{y}, \tag{3.127}$$

I_n die (n,n)-Einheitsmatrix, $X'X$ ist eine (n,n)-Matrix. Es folgt

$$\hat{\mathbf{b}}_\lambda = (X'X + \lambda I_n)^{-1}X'\mathbf{y} \tag{3.128}$$

folgt. $\hat{\mathbf{b}}_\lambda$ ist die Ridge-Regression-Schätzung für $\mathbf{b}$. Für $\lambda = 0$ erhält man die Standard-KQ-Schätzung. Für $\lambda > 0$ ist die Matrix $X'X + \lambda I$ stets positiv-definit, also invertierbar[63]

Aus (3.127) folgt andererseits

$$\lambda\hat{\mathbf{b}}_\lambda = X'\mathbf{y} - X'X\hat{\mathbf{b}}_\lambda = X'(\mathbf{y} - X\hat{\mathbf{b}}),$$

d.h.

$$\hat{\mathbf{b}}_\lambda = \lambda^{-1}X'(\mathbf{y} - X\hat{\mathbf{b}}) = X'\lambda^{-1}(\mathbf{y} - X\hat{\mathbf{b}}). \tag{3.129}$$

Setzt man

$$\boldsymbol{\alpha} = \lambda^{-1}(\mathbf{y} - X\hat{\mathbf{b}}) \in \mathbb{R}^m \tag{3.130}$$

($\boldsymbol{\alpha}$ ist ein Vektor!), so kann (3.129) in der Form

$$\hat{\mathbf{b}}_\lambda = X'\boldsymbol{\alpha} \in \mathbb{R}^n \tag{3.131}$$

geschrieben werden. Setzt man diesen Ausdruck für $\hat{\mathbf{b}}$ in (3.129) ein, so erhält man

$$\lambda\boldsymbol{\alpha} = \mathbf{y} - XX'\boldsymbol{\alpha},$$

d.h.

$$\mathbf{y} = \lambda\alpha + XX'\boldsymbol{\alpha} = (XX' + \lambda I_m)\boldsymbol{\alpha} \tag{3.132}$$

hier ist I_m die (m,m)-Einheitsmatrix, denn XX' ist eine (m,m)-Matrix. Setzt man $XX' = G$, so erhält man für α den Ausdruck

$$\boldsymbol{\alpha} = (G + \lambda I_m)^{-1}\mathbf{y} \tag{3.133}$$

Wegen (3.131) erhält man die Schätzung $\hat{\mathbf{b}}_\lambda$ als Linearkombination der Spaltenvektoren von X'.

Man hat also zwei Ausdrücke für die Schätzung von $\hat{\mathbf{b}}_\lambda$: einerseits (3.128), andererseits (3.131) mit (3.133). Zum Vergleich werden sie noch einmal zusammen aufgeführt:

$$\begin{aligned} \hat{\mathbf{b}}_\lambda &= (X'X + \lambda I_n)^{-1}X'\mathbf{y}, \quad \text{primäre Lösung} && (3.134)\\ &= X'(XX' + \lambda I_m)^{-1}\mathbf{y}, \quad \text{duale Lösung} && (3.135) \end{aligned}$$

[63]Vergl. den Beweis für den Satz 3.8, Seite 99: Sei $C = X'X$, $\mathbf{a} \in \mathbb{R}^n$. Dann ist

$$\mathbf{a}'(C + \lambda I)\mathbf{a} = \mathbf{a}'C\mathbf{a} + \lambda\mathbf{a}'I\mathbf{a} \geq 0,$$

da $\mathbf{a}'C\mathbf{a} \geq 0$ (C ist positiv-semidefinit), und $\mathbf{a}'I\mathbf{a} = \|\mathbf{a}\|^2 \geq 0$ ebenfalls.

(3.134) ist natürlich (3.128), und (3.135) ist (3.131) mit (3.133) mit $G = XX'$. Die primäre Lösung (3.134) wird durch die (n, n)-Matrix $X'X$ bestimmt, während die *duale Lösung* (3.135) durch die (m, m)-Matrix XX' bestimmt wird. Der Unterschied zwischen den beiden Schätzungen besteht im Rechenaufwand: ist X eine (m, n)-Matrix, so ist im Fall $n < m$ die Matrix $X'X + \lambda I_n$ eine (n, n)-Matrix und die Inverse $(X'X + \lambda I_n)^{-1}$ ist mit weniger Aufwand zu berechnen als die von $XX' + \lambda I_m$, im Fall $m < n$ ist es umgekehrt. Unabhängig davon erweist sich die Lösung (3.131) aber von Vorteil, wenn der nichtlineare Fall betrachtet wird, wie im Folgenden erläutert wird.

3.5 Weitere Befunde

3.5.1 Die Zentrierungsmatrix

Die Zentrierung einer Datenmatrix kann durch eine Matrixmultiplikation bewerkstelligt werden:

Definition 3.10 *Es sei $\vec{1} = (1, 1, \ldots, 1)'$ ein m-dimensionaler Vektor, dessen Komponenten alle gleich 1 sind, und $\vec{1}\vec{1}'$ ist das dyadische Produkt von $\vec{1}$ mit sich selbst. Dann heißt*

$$H_m = I - \frac{1}{m}\vec{1}\vec{1}'. \tag{3.136}$$

Zentrierungsmatrix.

H_M hat die Form

$$H_m = \begin{pmatrix} 1 - \frac{1}{m}, & -\frac{1}{m}, & -\frac{1}{m}, & \cdots, & -\frac{1}{m} \\ -\frac{1}{m}, & 1 - \frac{1}{m}, & -\frac{1}{m}, & \cdots, & -\frac{1}{m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{m}, & -\frac{1}{m}, & -\frac{1}{m}, & \cdots, & 1 - \frac{1}{m} \end{pmatrix} \tag{3.137}$$

Ist X eine nichtzentrierte (m, n)-Matrix und ist X_{sc} die korrespondierende spaltenzentrierte Matrix, X_{zc} die korrespondierende zeilenzentrierte Matrix so gilt

$$X_{sc} = H_m X, \quad X_{zc} = X H_n. \tag{3.138}$$

Es gilt der

Satz 3.19 *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 $rg(H_m) = m - 1$,

Beweis: Es gilt

$$
\begin{aligned}
H_m H_m &= (I - \frac{1}{m}\vec{1}\vec{1}')(I - \frac{1}{m}\vec{1}\vec{1}') \\
&= I - \frac{1}{m}I\vec{1}\vec{1}' - \frac{1}{m}I\vec{1}\vec{1}' + \frac{1}{m^2}\vec{1}\vec{1}'\vec{1}\vec{1}' \\
&= I - \frac{1}{m}I\vec{1}\vec{1}' = H_m,
\end{aligned}
$$

denn

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

d.h. H_m ist idempotent. Weiter gilt

$$H_m' = (I - \frac{1}{m}\vec{1}\vec{1}')' = I' - (\frac{1}{m}\vec{1}\vec{1}')' = I - \frac{1}{m}\vec{1}\vec{1}' = H_m, \tag{3.139}$$

d.h. H_m ist symmetrisch. Damit ist H_m gemäß Definition 3.218, Seite 148, eine Projektionsmatrix. Nach Satz 3.26 (Seite 148) 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 $\mathrm{rg}(H_m) = m - 1$, 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

$$-\frac{x_1}{m} - \cdots - \frac{x_{j-1}}{m} + x_j(1 - \frac{1}{m}) - \frac{x_{j+1}}{m} - \cdots - \frac{x_m}{m} = x_j - \frac{1}{m}\sum_{i=1} x_i = x_j - \bar{x}$$

Dann ist

$$\mathbf{x}'H_m\mathbf{x} = x_1^2 - x_1\bar{x} + x_2^2 - x_2\bar{x} + \cdots - x_m^2 - x_m\bar{x} = \sum_{i=1}^m x_i^2 - m\bar{x}^2 = k,$$

so dass

$$\frac{1}{m}\mathbf{x}'H_m\mathbf{x} = \frac{k}{m} = \frac{1}{m}\sum_{i=1}^m x_i^2 - \bar{x}^2 = \frac{1}{m}\sum_{i=1}^m (x_i - \bar{x})^2 \geq 0,$$

also ist H_m positiv-semidefinit.

Der Kern von H_m ist $\ker(H_m) = \{\mathbf{x} | H_m\mathbf{x} = \vec{0}\}$. Man hat

$$\begin{pmatrix} 1 - \frac{1}{m}, & -\frac{1}{m}, & -\frac{1}{m}, & \cdots, & -\frac{1}{m} \\ -\frac{1}{m}, & 1 - \frac{1}{m}, & -\frac{1}{m}, & \cdots, & -\frac{1}{m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{m}, & -\frac{1}{m}, & -\frac{1}{m}, & \cdots, & 1 - \frac{1}{m} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Für die i-te Komponente von $\vec{0}$ findet man dann

$$x_i - \bar{x} = 0 \Rightarrow x_i = \bar{x}$$

für alle i, d.h. die Komponenten von $\mathbf{x}$ sind identisch, $\mathbf{x} = (x, x, \ldots, x)'$. Da sich die Orientierung eines Vektors nicht ändert, wenn er mit einem Skalar multipliziert wird, ist

$$\frac{1}{x}\mathbf{x} = (1, 1, \ldots, 1)',$$

d.h. die Orientierung der $\mathbf{x}$ aus dem Kern von H_m ist identisch mit der von $\vec{1} = (1, 1, \ldots, 1)'$. Der Kern ist damit ein 1-dimensionaler Teilraum des $\mathbb{R}^m$. □

Es sei X eine unzentrierte Datenmatrix, X_c sei die zugehörige spaltenzentrierte Matrix. Dann gilt

$$X_c = H_m X, \quad X_c' X_c = X H_m' H_m X = X' H_m^2 X = X' H_m X, \tag{3.140}$$

wegen der Symmetrie und Idempotenz von H_m.

Der Rang zentrierter Datenmatrizen: Nach (3.138) ist $X_{sc} = H_m X$, so dass wegen (2.30), Seite 69,

$$\mathrm{rg}(X_{sc}) \leq \min[\mathrm{rg}(H_m), \mathrm{rg}(X)]. \tag{3.141}$$

Satz 3.20 *Es ist rg*$(H_m) = m$.

Beweis: Zu zeigen ist, dass die Darstellung des Nullvektors $\vec{0}$ als Linearkombination der Spaltenvektoren von H_m nur möglich ist, wenn alle Koeffizienten $a_i = 0$ gilt. Allgemein hat man

$$a_1 \begin{pmatrix} 1 - 1/m \\ -1/m \\ \vdots \\ -1/m \end{pmatrix} + a_2 \begin{pmatrix} -1/m \\ 1 - 1/m \\ \vdots \\ -1/m \end{pmatrix} + \cdots + a_m \begin{pmatrix} -1/m \\ -1/m \\ \vdots \\ 1 - 1/m \end{pmatrix} = \vec{0}$$

Mit $S = \sum_i a_i$ gilt dann speziell für die i-te Komponente von $\vec{0}$

$$0 = a_i(1 - \frac{1}{m}) - \frac{m-1}{m}(S - a_i),$$

d.h.

$$\frac{m-1}{m} a_i = \frac{m-1}{m}(S - a_i) \Rightarrow a_i = S - a_i,$$

so dass

$$a_i = \frac{1}{2}S, \quad i = 1, \ldots, m$$

Andererseits hat man

$$\frac{m-1}{m} \sum_i a_i = (m-1)S + \frac{m-1}{m} \sum_i a_i,$$

woraus $0 = (m-1)S$, d.h. $S = 0$ folgt. Dann folgt $a_i = \frac{1}{2}S = 0$ für alle i. Damit folgt, dass die Spaltenvektoren von H_m linear unabhängig sind, d.h. H_m hat den vollen Rang m. □

Bei Datenmatrizen stehen die m Zeilen für die m Fälle und die n Spalten für die gemessenen n Variablen, und man hat $m \geq n$. Die Gleichung (3.141) impliziert dann

$$\mathrm{rg}(X_{sc}) = \mathrm{rg}(X) \leq n. \tag{3.142}$$

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 3.11 *Die Matrix A^+ heißt* Pseudoinverse *oder* Moore-Penrose-Inverse*, wenn sie die folgenden Bedingungen (Moore-Penrose-Bedingungen) erfüllt:*
1. $AA^+A = A$,
2. $A^+AA^+ = A^+$
3. $(AA^+)^* = AA^+$
4. $(A^+A)^* = A^+A$.

Anmerkung: Die Bedingungen 3. und 4. beziehen sich auf Matrizen mit komplexwertigen Elementen; der Stern definiert die konjugiert komplexe Zahl einer komplexen Zahl. In diesem Skript werden keine komplexen Matrizen und Vektoren betrachtet, so dass 3. und 4. nicht weiter berücksichtigt werden müssen, die beiden Punkte sind nur der Vollständigkeit halber mit aufgeführt worden. □

Für das Gleichungssystem $A\mathbf{x} = \mathbf{y}$ kann nun leicht eine Pseudoinverse gefunden werden, wenn $\mathrm{rg}(A) = n$ ist. Dann hat $A'A$ ebenfalls den Rang n, d.h. $A'A$ hat vollen Rang, so dass die Inverse $(A'A)^{-1}$ existiert. Man hat dann nach Multiplikation von $A\mathbf{x} = \mathbf{y}$ von links mit A' die Gleichung

$$A'A\mathbf{x} = A'\mathbf{y} \Rightarrow \mathbf{x} = (A'A)^{-1}A'\mathbf{y}. \tag{3.143}$$

Für $\mathbf{y} \in \mathcal{L}(A)$ ist $\mathbf{x}$ die exakte Lösung für das Gleichungssystem, und für $\mathbf{y} \notin \mathcal{L}(A)$ liefert $(A'A)^{-1}A'\mathbf{y}$ die Kleinste-Quadrate-Approximation $\hat{\mathbf{x}}$ für $\mathbf{x}$, d.h. man hat

$$(A'A)^{-1}A'\mathbf{y} = \begin{cases} \mathbf{x}, & \mathbf{y} \in \mathcal{L}(A), \\ \hat{\mathbf{x}}, & \mathbf{y} \notin \mathcal{L}(A). \end{cases} \tag{3.144}$$

(vergl. Abschnitt 6.9.3, Gleichung (6.65), Seite 195).

Die Matrix $(A'A)^{-1}A'$ ist eine Pseudoinverse für A. Um diese Behauptung einzusehen, genügt es, die Bedingungen 1. und 2. zu überprüfen. In Bezug auf 1. hat man

$$A((A'A)^{-1}A')A = A(A'A)^{-1}A'A = A,$$

und in Bezug auf 2. hat man

$$((A'A)^{-1}A')A((A'A)^{-1}A') = (A'A)^{-1}A'A(A'A)^{-1}A' = (A'A)^{-1}A',$$

d.h. $(A'A)^{-1}A'$ ist eine Pseudoinverse für A. □

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

$$A = Q\Sigma T', \quad \Sigma = \begin{pmatrix} \Lambda_r^{1/2} & 0 \\ 0 & 0 \end{pmatrix} \tag{3.145}$$

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

$$A^+ = A' = T \begin{pmatrix} \Lambda_r^{-1/2} & 0 \\ 0 & 0 \end{pmatrix} Q' \tag{3.146}$$

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. (3.146) 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 20). Diese Definition einer Norm charakterisiert einen Spezialfall, im Folgenden wird eine allgemeinere Definition gegeben.

Die Norm eines Vektors definiert, in welchem Sinne von der "Göße" 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"[64]. Ebenso kann eine *Matrixnorm*

[64]Um sich eine inhaltliche Vorstellung zu machen, stelle man sich vor dass die Komponenten x_i von $\mathbf{x}$ Maße für Begabungen $M_1, \ldots, M_n$ repräsentieren. Dann ist $\|\mathbf{x}\|$ ein *mögliches* –nicht notwendig sinnvolles – Maß für die Gesamtbegabung einer Person.

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 3.12 *Es seien $\boldsymbol{x} \in \mathbb{R}^n$ n-dimensionale Vektoren. Eine Vektornorm ist eine Abildung $f : \mathbb{R}^n \to \mathbb{R}$ (d.h. es wird einem Vektor $\boldsymbol{x}$ eine bestimmte reelle Zahl zugeordnet), die den Bedingungen*
1. $f(\boldsymbol{x}) \geq 0$,
2. $f(\boldsymbol{x}+\boldsymbol{y}) \leq f(\boldsymbol{x}) + f(\boldsymbol{y})$,
3. $f(a\boldsymbol{x}) = af(\boldsymbol{x})$, für $a \in \mathbb{R}$
genügt. Dann heißt f eine Vektornorm. *f wird durch $f(\boldsymbol{x}) = \|\boldsymbol{x}\|$ notiert. Der* Einheitsvektor *in Bezug auf eine Norm $\|\cdot\|$ ist derjenige Vektor, für den $\|\boldsymbol{x}\| = 1$ gilt.*

Von besonderem Interesse sind die p-Normen

$$\|\mathbf{x}\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p} = \left(\sum_{j=1}^n |x_j|^p\right)^{1/p}. \tag{3.147}$$

Für $p = 1$ erhält man die 1-Norm

$$\|\mathbf{x}\|_1 = (|x_1| + \cdots + |x_n|) = \sum_{j=1}^n |x_j|. \tag{3.148}$$

und für $p = 2$ die euklidische Norm

$$\|\mathbf{x}\|_2 = (|x_1|^2 + \cdots + |x_n|^2)^{1/2} = \left(\sum_{j=1}^n |x_j|^2\right)^{1/2} = \sqrt{\mathbf{x}'\mathbf{x}}. \tag{3.149}$$

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

$$\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|. \tag{3.150}$$

Die Maximum-Norm ergibt sich aus der p-Norm für $p \to \infty$. Es sei $x_k = x_{\max}$ die maximale Komponente von $\mathbf{x}$. Dann ist

$$\|\mathbf{x}\|_p = \left(|x_k|^p \sum_{j=1}^n \frac{|x_j|^p}{|x_k|^p}\right)^{1/p}.$$

Wegen $|x_j|/|x_k| \leq 1$ für alle $j \neq k$ folgt $\lim_{p\to\infty} |x_j|^p/|x_k|^p \to 0$ für $j \neq k$ und $|x_j|^p/|x_k|^p = 1$ für $j = k$, so dass

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

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

Definition 3.13 *Es sei* $\mathbb{R}^{m\times n}$ *die Menge der rellen* (m,n)*-Matrizen*[65]. *Eine* Matrixnorm *ist eine Abbildung* $\|\cdot\| : \mathbb{R}^{m\times n} \to \mathbb{R}_+$*,* $\mathbb{R}_+$ *die Menge der reellen Zahlen größer oder gleich Null, und* $A \mapsto \|A\|$ *derart, dass*

1. $\|A\| = 0$ *genau dann, wenn* $A = 0$ *die Nullmatrix ist,*
2. $\|\lambda X\| = \lambda\|A\|$*,*
3. $\|A + B\| \leq \|A\| + \|B\|$
gilt. Zusammen mit der Norm $\|\cdot\|$ *wird der Vektorraum der* (m,n)*-Matrizen dann zu einem* normierten Vektorraum[66] $(\mathbb{R}^{m\times n}, \|\cdot\|)$.

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

1. Die *Frobenius-Norm.*

$$\|A\|_F = \left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2\right)^{1/2}. \tag{3.151}$$

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

2. Die p-Norm: sie ist definiert durch

$$\|A\|_p = \max_{\mathbf{x}\neq\vec{0}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p} \tag{3.152}$$

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

$$\|A\| = \max_{\mathbf{x}\neq\vec{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}. \tag{3.153}$$

   durch eine Vektornorm induzierte Normen.

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

$$\|A\|_p = \sup_{\mathbf{x}\neq\vec{0}} \left\|\left(A\frac{\mathbf{x}}{\|\mathbf{x}\|}\right)\right\|_p = \max_{\|\mathbf{x}\|_p=1} \|A\mathbf{x}\|_p. \tag{3.154}$$

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

[66]Hier wird vom *allgemeinen* Begriff des Vektorraums Gebrauch gemacht, demzufolge auch Matrizen als "Vektoren" aufgefasst werden können, so dass auch Mengen von Matrizen einen Vektorraum bilden können. Der Begriff des Vektorraums bezieht sich ja eigentlich nur auf die Kombination bzw. Verknüpfungen von Elementen einer Menge!

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 6.5 findet man

$$\|A\|_2 = \max_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2 = \sqrt{\lambda_{\max}}, \tag{3.155}$$

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

Satz 3.21 *Es sei A eine (m, n)-Matrix. Für die Frobenius-Norm $\|A\|_F$ gilt*

$$\|A\|_F^2 = spur(AA') = \sum_{i=1}^{n} \lambda_j \tag{3.156}$$

wobei $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ die Eigenwerte von $A'A$ sind.

.

Beweis: Auf A kann die SVD angewendet werden: $A = Q\Lambda^{1/2}P'$, $\lambda_j \geq 0$ für $j = 1, \ldots, n$. Dann ist $AA' = Q\Lambda^{1/2}P'P\Lambda^{1/2}Q' = Q\Lambda Q'$, und die Diagonalelemente von $Q\Lambda Q'$ sind von der Form $\sum_i \lambda_i q_{ji}^2$ für $j = 1, \ldots, n$ Die Spur von $A'A$ ist die Summe dieser Diagonalelemente, d.h.

$$\text{spur}(AA') = \sum_{j=1}^{n}\sum_{i=1}^{n} \lambda_i q_{ji}^2 = \sum_{i=1}^{n} \lambda_i \sum_{j=1}^{n} q_{ij}^2.$$

Aber $\sum_{j=1}^{n} q_{ij}^2 = 1$ für alle i, da Q orthonormal ist, d.h. die Eigenvektoren haben alle die Länge 1. Damit ist (3.156) gezeigt. □

Anmerkung: $A = Q\Lambda^{1/2}P'$ impliziert $A'A = P\Lambda P'$ und wegen der Orthonormalität der Spaltenvektoren von P folgt in analoger Weise

$$\text{spur}(A'A) = \sum_{i=1}^{n} \lambda_j. \tag{3.157}$$

3.5.4 Die Approximation von Matrizen

Der folgende Satz macht eine Aussage über die Güte der Approximation einer (m, n)-Matrix A vom Rang $\text{rg}(A) = r \leq \min(m, n)$ durch eine (m, n)-Matrix A_k mit kleinerem Rang $k \leq r$. So sei etwa $A = X$ eine Datenmatrix mit dem Rang $n < m$ und man will versuchen, X durch eine Matrix X_k mit dem Rang $k < n$ zu approximieren, d.h. durch möglichst wenige latente Variable zu "erklären".

Satz 3.22 *Es seien A und A_k (m, n)-Matrizen, $m \geq n$, und es seien die Matrizen A und A_k durch*

$$A = Q\Lambda^{1/2}P' = \sum_{j=1}^{n} \sqrt{\lambda_j}\,\boldsymbol{q}_j\boldsymbol{p}_j', \quad A_k = Q_k\Lambda_k^{1/2}P_k' = \sum_{j=1}^{k} \sqrt{\lambda_j}\,\boldsymbol{q}_j\boldsymbol{p}_j' \tag{3.158}$$

definiert, wobei Q eine (m,n)-Matrix und V eine (n,n)-Matrix ist, $\Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_n)$ mit $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ ist, und Q_k ist eine (m,k)-Matrix, V_k und $\Lambda_k = diag(\lambda_1, \lambda_2, \cdots, \lambda_k)$ sind (k,k)-Matrizen mit $k < n$. Dann gilt

$$\|A - A_k\|_2 = \sqrt{\lambda_{k+1}}. \tag{3.159}$$

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

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

$\Sigma^* = \mathrm{diag}(\underbrace{0, \ldots, 0}_{k}, \sigma_{k+1}, \ldots, \sigma_n)$, $\sigma_j = \sqrt{\lambda_j}$. Da $\|A\|_2 = \sigma_{\max} = \sigma_1$ im Falle geordneter Singularwerte $\sigma_{k+1} \geq \cdots \geq \sigma_n$, folgt

$$\|A - A_k\|_2 = \sigma_{k+1} = \sqrt{\lambda_{k+1}},$$

da nun σ_{k+1} der maximale Singularwert ist. □

Anmerkungen:

1. Die Approximation wird trivialerweise immer besser, je größer der Wert von k ist, da ja der Wert von λ_{k+1} mit größer werdendem k immer kleiner wird. Der nichttriviale Teil der Aussage ist, dass $\|A - A_k\|_2$ gerade dem Wert von $\sqrt{\lambda_{k+1}}$ entspricht.

2. Bei der Approximation von A durch A_k wurde von der SVD von A Gebrauch gemacht. Die Gleichung $A = Q\Lambda^{1/2}P'$ ist insofern trivial, als die SVD stets gilt. Dass man A durch A_k approximiert, wobei A_k nur durch die ersten k Terme der SVD definiert ist, kann zur Frage führen, ob es eine andere Repräsentation für A_k gibt, die nicht auf der SVD beruht, aber besser ist in dem Sinne, dass $\|A - A_k\|_2 < \sqrt{\lambda_{k+1}}$. Eine solche gibt es nicht, wie noch gezeigt werden wird.

3. Man vergleiche die Aussage (3.159) mit der Aussage (3.158) von Satz 6.5. Wie die Gleichung (3.32), 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 70). Der Rang $\mathrm{rg}(A) \leq \min(m,n)$ ist durch die Anzahl der von Null verschiedenen Eigenwerte λ_j und damit durch die Anzahl der von Null verschiedenen $\sigma_k \mathbf{q}_k \mathbf{p}_k$ gegeben. Da die ersten k Eigenwerte von A und A_k identisch sind, enthält die Differenz $\Lambda^{1/2} - \Lambda_k^{1/2}$ nur Nullen, und der erste von Null verschiedene Wert in der Diagonalen ist $\sigma_{k+1} = \sqrt{\lambda_{k+1}}$. Da die Eigenwerte λ_j in Λ der Größe nach angeordnet sind, ist σ_{k+1} nun der größte Singularwert für $A - A_k$. □

Im Folgenden bedeutet $\min_{\mathrm{rg}(B)=k} \|X - B\|$ bzw $\min_{\mathrm{rg}(B)=k} \|X - B\|_F$ diejenige Matrix B, die (i) den Rang $\mathrm{rg}(B) = k$ 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 3.23 *(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*

$$A_k = Q\Lambda_k^{1/2}P' = \sum_{j=1}^{k} \sqrt{\lambda_j}\,\boldsymbol{q}_j\boldsymbol{p}_j', \tag{3.160}$$

wobei Λ_k die Diagonalmatrix mit den zu den ersten[67] *k Eigenvektoren korrespondierenden Eigenwertev von A ist. Dann gilt*

$$\min_{B\in\mathbb{R}^{m,n}, rg(B)=k} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1} = \sqrt{\lambda_{k+1}} \tag{3.161}$$

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 3.24 in allgemeiner Weise bewiesen, so dass auch zusammenfassend vom Schmidt-Mirsky-Theorem gesprochen wird.

Beweis: Zur Vereinfachung werde

$$\|A - B_{\min}\| = \min_{B\in\mathbb{R}^{m,n}, \mathrm{rg}(B)=k} \|A - B\|_2$$

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

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

[67]Es wird angenommen, dass die Eigenwerte der Größe nach geordnet sind, $\lambda_1 > \lambda_2 > \cdots > \lambda_k$

Insbesondere kann dann $\mathbf{b}$ als Linearkombination der ersten $k+1$ (Eigen-)Vektoren von P gewählt werden: sind also gerade die ersten $k+1$ Spalten von P die Spalten von P_{k+1}, so sei $\mathbf{b} = P_{k+1}\mathbf{x} = x_1\mathbf{p}_1 + \cdots + x_{k+1}\mathbf{p}_{k+1}$. Es ist

$$B_{\min}\mathbf{b} = B_{\min}P_{k+1}\mathbf{x},$$

$B_{\min}P_{k+1}$ ist also eine $(m\times(k+1))$-Matrix. Nach Satz 2.3, Gleichung (2.30) (Seite 69) ist aber $\mathrm{rg}(B_{\min}P_{k+1}) \leq \min(\mathrm{rg}(B_{\min})), \mathrm{rg}(P_{k+1})) = k$, da ja $\mathrm{rg}(B_{\min}) = k$ nach Voraussetzung. Dann folgt aber

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

d.h.

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

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

$$\|A\mathbf{b} - B_{\min}\mathbf{b}\| = \|AP_{k+1}\mathbf{x}\| < \|(AP_{k+1}\mathbf{x} - A_kP_{k+1}\mathbf{x}\| \leq \sigma_{k+1}\|\mathbf{b}\|.$$

Es ist aber

$$\|A\|\|P_{k+1}\mathbf{x}\| < \|(A - A_kP_{k+1}\|\|\mathbf{x}\| \leq \sigma_{k+1}\|\mathbf{b}\|,$$

d.h.

$$\sigma_1 < \sigma_{k+1},$$

im Widerspruch zu $\sigma_1 \geq \sigma_{k+1}$. Damit gilt (3.161). □

Satz 3.24 *(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 3.23 definiert und A_k sei wie in (3.160) definiert. Dann gilt*

$$\|A - B_{\min}\|_F = \sqrt{\sum_{j=k+1}^{n} \lambda_j}, \tag{3.162}$$

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

$$\begin{aligned}
\|A - \sum_{j=1}^{k} \mathbf{x}_j \mathbf{y}_j\|_F^2 &= \operatorname{spur}\left((A - \sum_{j=1}^{k} \mathbf{x}_j \mathbf{y}_j)'(A - \sum_{j=1}^{k} \mathbf{x}_j \mathbf{y}_j) \right) \\
&= \operatorname{spur}\left(A'A + \sum_{j=1}^{k} (\mathbf{y}_j - A'\mathbf{x}_j)(\mathbf{y}_j - A'\mathbf{x}_j)' - \sum_{j=1}^{k} A'\mathbf{x}_j \mathbf{x}_j A \right)
\end{aligned}$$

Es ist $\operatorname{spur}((\mathbf{y}_j - A'\mathbf{x}_j)(\mathbf{y}_j - A'\mathbf{x}_j)) \geq 0$, $\operatorname{spur}(A'\mathbf{x}_j \mathbf{x}_j' A) = \|A'\mathbf{x}_j\|$ und es ist zu zeigen, dass

$$\sum_{j=1}^{k} \|A'\mathbf{x}_j\|^2 \leq \sum_{j=1}^{k} \lambda_j.$$

Die SVD von A sei $A = Q\Sigma P'$, und es sei $P_1 = [|\mathbf{p}_1| \dots |\mathbf{p}_k|0]$, $P_2 = [|0|\mathbf{p}_{k+1}| \dots |\mathbf{P}_n|]$, so dass $P = [P_1|P_2]$. Analog dazu sei $\Sigma_1 = \operatorname{diag}(\sigma_1, \dots, \sigma_k)$, $\Sigma_2 = \operatorname{diag}(\sigma_{k+1}, \dots, \sigma_n)$. Dann hat man

$$\begin{aligned}
\|A'\mathbf{x}_j\| &= \|Q\Sigma P'\|_F^2 = \|\Sigma P'\mathbf{x}_j\|_F^2 = \\
&= \|\Sigma_1 P_1'\mathbf{x}_j\|_F^2 + \|\Sigma_2 P_2'\mathbf{x}_j\|_F^2 + \lambda_k - \lambda_k + \lambda_k(\|P'\mathbf{x}_j\| - \|P_1'\mathbf{x}_j\|_F^2 - \|P_2'\mathbf{x}_j\|_F^2) \\
&= \lambda_k + (\|\Sigma_1 \mathbf{P}_1'\mathbf{x}_j\|_F^2 - \underbrace{\lambda_k \|P_1'\mathbf{x}_j\|_F^2)}_{(1)} - \underbrace{\lambda_k(1 - \|P'\mathbf{x}_j\|_F^2)}_{(2)}
\end{aligned}$$

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

$$\begin{aligned}
\sum_{j=1}^{k} \|A'\mathbf{x}_j\|^2 &\leq k\lambda_k + \sum_{j=1}^{k} (\|\Sigma_1 P_1'\|^2 - \lambda_k \|P_1'\mathbf{x}_j\|^2) \\
&= k\lambda_k + \sum_{j=1}^{k} \sum_{i=1}^{k} (\lambda_i - \lambda_j) |\mathbf{v}_j' \mathbf{x}_j|^2 \\
&\leq \sum_{i=1}^{k} (\lambda_k + (\lambda_i - \lambda_k)) = \sum_{j=1}^{k} \lambda_j.
\end{aligned}$$

□

3.6 Bestimmung latenter Variablen

Datenmatrizen X lassen sich oft "erklären", indem man etwa die Spaltenvektoren $\mathbf{x}_j$ von X als Linearkombinationen der Vektoren einer (Teil-)Basis darstellt. Die Komponenten dieser Basisvektoren repräsentieren die Fälle, die Koeffizienten der Vektoren repräsentieren die die Variablen auf den "'latenten" Variablen. Das es unendlich[68] viele Basen für die $\mathbf{x}_j$ gibt, muß man sich für eine bestimmte Basis oder Teilbasis aus der Menge der möglichen entscheiden. Eine erste Wahl ist die Hauptachsentransformation: die Basisvektoren $\mathbf{u}_1, \ldots, \mathbf{u}_n$ ergeben sich als Linearkombinationen der $\mathbf{x}_j$, die so gewählt werden, dass die Länge $\|\mathbf{u}_1\|^2$ maximal, $\|\mathbf{u}_2\|^2$ zweit-maximal etc ist, und die $\mathbf{u}_k$ orthogonal sind. Dieser Ansatz ergibt sich aus 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)

Gesamtvarianz und Varianzanteile: Es sei $X_0 = [\mathbf{X}_1, \ldots, \mathbf{X}_n]$ eine (m, n)-Datenmatrix. Die Komponenten der m-dimensionalen Spaltenvektoren $\mathbf{X}_j$, $j = 1, \ldots, n$, sind Messungen von n Variablen $\xi_1, \ldots, \xi_n$) repräsentieren. Im Allgemeinen geht man nicht von diesen Rohdaten aus, sondern (i) von den spaltenzentrierten Messungen $x_{ij} = \mathbf{X}_{ij} - \bar{x}_j$, $j = 1, \ldots, n$, zusammengefasst in einer Matrix $X = (x_{ij})$, oder (ii) von spaltenstandardisierten Messungen $z_{ij} = (X_{ij} - \bar{x}_j)/s_j$, s_j die Standardabweichung der Messwerte in der j-ten Spalte. Die z_{ij} werden in einer Matrix $Z = (z_{ij})$ zusammenfasst. Dann hat man

$$C = \frac{1}{m-1} X'X, \quad R = \frac{1}{m-1} Z'Z \tag{3.163}$$

C die Matrix der Kovarianzen, R die Matrix der Korrelationen zwischen den Variablen[69].

In vielen Lehrbüchern (z.B. Jolliffe (1986)) wird die PCA in Bezug auf die Matrix X der zentrierten Messwerte hergeleitet, der Fall der spaltenstandardisierten Daten ergibt sich dann automatisch, wenn man statt der zentrierten Vektoren $\mathbf{x}_j$ die Vektoren $\mathbf{z}_j$ aus Ausgangspunkt nimmt. Es ist aber sinnvoll, von vorn herein auf die Unterschiede der beiden Fälle hinzuweisen. In vielen Untersuchungen werden Variablen mit verschiedenen Maßeinheiten betrachtet: z.B. werden Körpergrößen in Zentimetern oder Metern angegeben, Körpergewichte in Gramm, oder

[68] Sogar überabzählbar viele!

[69] Die Schätzung $(1/m)Z'Z$ hat bekanntlich einen Bias, die Schätzung $1/(m-1)$ nicht.

Kilogramm, etc. Die Wahl der Einheit ist im Prinzip beliebig, die verschiedenen Einheiten für eine Variable unterschieden sich durch Faktoren. Aber die spezielle Wahl eines solchen Faktors (d.h. der dazu korrespondierenden Einheit) betimmt die Werte der Kovarianzen c_{jk} in der Matrix C, was wiederum einen großen Einfluß auf die Anzahl der latenten Vektoren und die Repräsentation von Variablen und Fällen auf den latenten Variablen haben kann: die Repräsentation wird beliebig in dem Sinne, wie die Wahl der Einheiten beliebig ist. Diese Verzerrung wird ausgeschlossen, wenn man zu spaltenstandarisierten Messwerten übergeht. Ein Spezialfall ergibt sich, wenn die Maßeinheiten der Variablen identisch sind. Die Varianzen und Kovarianzen hängen zwar nach wie vor von der gewählten Einheit ab, aber die Relationen zwischen diesen Werten nicht. In diesem Fall wird den möglicherweise existierenden Unterschieden etwa zwischen den Varianzen der Variablen Rechnung getragen. Die Spaltenstandardisierung bewirkt dagegen eine Gleichsetzung der Varianzen: sie sind alle gleich 1, d.h $\mathbf{z}_j'\mathbf{z}_j/m = 1$ für alle j. Dieser Sachverhalt kann ebenfalls ein Verzerrung der Repräsentationen bedeuten.

Im Folgenden wird der Fall spaltenstandardisierter Variablen betrachtet, es wird aber auf Eigenschaften der Schätzungen von Faktorladungen und Faktorscores für den Fall von nur zentrierten Daten hingewiesen.

Eine spaltenstandardisierte Messung ist durch

$$z_{ij} = \frac{X_{ij} - \bar{x}_j}{s_j} \tag{3.164}$$

definiert; darin ist $\bar{x}_j$ der Mittelwert der Komponenten X_{ij} und s_j sei die (Stichproben-)Varianz der x_{ij}, d.h.[70]

$$\bar{x}_j = \frac{1}{m}\sum_{i=1}^m x_{ij}, \quad s_j^2 = \frac{1}{m-1}\sum_{i=1}^m (x_{ij} - \bar{x}_j)^2, \quad j = 1, \ldots, n.$$

Es sei

$$W = \frac{1}{\sqrt{m-1}} Z. \tag{3.165}$$

Dann gilt insbesondere für den j-ten Spaltenvektor von $\mathbf{w}_j$ von W,

$$\mathbf{w}_j = \frac{1}{\sqrt{m-1}}\mathbf{z}_j, \tag{3.166}$$

und mit $\vec{1} = (1, 1, \ldots, 1)'$

$$\mathbf{w}_j'\vec{1} = \frac{1}{\sqrt{m-1}}\frac{1}{s_j}\sum_{i=1}^m (x_{ij} - \bar{x}_j) = 0, \tag{3.167}$$

weil die Summe der Abweichungen vom Mittelwert stets gleich Null ist, d.h. der Mittelwert der Komponenten von $\mathbf{w}_j$ ist gleich Null. Es ist

$$\mathbf{w}_j'\mathbf{w}_{j'} = \frac{1}{m-1}\mathbf{z}_j'\mathbf{z}_{j'} = r_{jj'} \tag{3.168}$$

[70] Division durch $m-1$ statt durch m zur Vermeidung des Bias in der Schätzung der Stichprobenvarianz!

die Produkt-Moment-Korrelation zwischen den Variablen ξ_j und $\xi_{j'}$, oder allgemein

$$\frac{1}{m-1}Z'Z = W'W = R, \tag{3.169}$$

R die (n,n)-Matrix der Korrelationen zwischen den Variaben ξ_j

$$\mathbf{w}_j'\mathbf{w}_j = \|\mathbf{w}_j\|^2 = r_{jj} = 1, \tag{3.170}$$

denn

$$\|\mathbf{w}_j\|^2 = \frac{1}{m-1}\mathbf{z}_j'\mathbf{z}_j = \frac{1}{s_j^2}\underbrace{\frac{1}{m-1}\sum(x_{ij}-\bar{x}_j)^2}_{s_j^2} = \frac{1}{s_j^2}s_j^2 = 1. \tag{3.171}$$

Die SVD von W ist

$$W = Q\Lambda^{1/2}V'. \tag{3.172}$$

V ist die orthonormale (n,n)-Matrix der Eigenvektoren $\mathbf{v}_k$ von $W'W = R$. Die Spalten von V repräsentieren die maximal n "latenten Dimensionen", die Zeilen von V repräsentnieren die Variablen ξ_j. Der Eigenvektor $\mathbf{v}_k$ definiert die Orientierung der k-ten Hauptachse (der k-ten latenten Dimension) des durch $\mathbf{v}'R\mathbf{v} = konst. \in \mathbb{R}$ definierten Ellipsoids. Q ist die orthonormale (m,m)-Matrix der Eigenvektoren $\mathbf{q}_k$ von WW'. Die Spalten von Q stehen wieder für die latenten Dimensionen, die Zeilen von Q repräsentieren die "Fälle", an denen die Messungen vorgenommen wurden. Offenbar folgt

$$W'W = R = V\Lambda^{1/2}Q'Q\Lambda^{1/2}V' = V\Lambda V'. \tag{3.173}$$

Für die Variablen werden "Ladungen" und für die Fälle werden "Faktoren-Scores" (F-Scores) gemäß

$$W = Q\Lambda^{1/2}V' = \begin{cases} QA', & A = V\lambda^{1/2} \text{ Ladungen} \\ UV', & U = Q\Lambda^{1/2} \text{ F-Scores} \end{cases} \tag{3.174}$$

Es ist

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \vdots & a_{nn} \end{pmatrix} \tag{3.175}$$

Die Spaltenvektoren $\mathbf{a}_k$ von A und die Zeilenvektoren $\tilde{\mathbf{a}}_j$ (Spaltenvektoren von A') sind

$$\mathbf{a}_k = \begin{pmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{nk} \end{pmatrix}, \quad \tilde{\mathbf{a}}_j = \begin{pmatrix} a_{j1} \\ a_{j2} \\ \vdots \\ a_{jn} \end{pmatrix} \tag{3.176}$$

Nach (3.174) gilt $W = QA'$, so dass $Q'W = A'$, d.h. $A = W'Q$. Daraus ergibt sich für das Element a_{jk}

$$a_{jk} = \mathbf{w}_j'\mathbf{q}_k = \|\mathbf{w}_j\|\|\mathbf{q}_k\|\cos\phi_{jk}, \tag{3.177}$$

ϕ_{jk} der Winkel zwischen dem Vektor $\mathbf{w}_j$ und $\mathbf{q}_k$ (k-te latente Dimension). Nach (3.171) gilt $\|\mathbf{w}_j\| = 1$ für alle j, und $\|\mathbf{q}_k\| = 1$, da $\mathbf{q}_k$ der auf die Länge 1 normierte k-te Eigenvektor von WW' ist. Also kann kurz

$$a_{jk} = \cos\phi_{jk} \tag{3.178}$$

geschrieben werden. Die Ladung a_{jk} der j-ten Variablen ξ_j auf der k-ten latenten Dimension/Variablen kann also als Korrelation zwischen der gemessenen Variablen ξ_j und der k-ten latenten Variablen interpretiert werden.

Nach Definition von A ist ein Spaltenvektor $\mathbf{a}_k$ von A durch

$$\mathbf{a}_k = \mathbf{v}_k\sqrt{\lambda_k}, \quad k = 1, \ldots, n \tag{3.179}$$

definiert[71]. Es gilt

$$\mathbf{a}_k'\mathbf{a}_{k'} = \left\{ \begin{array}{ll} \|\mathbf{a}_k\|^2 = \lambda_k, & k = k' \\ \sqrt{\lambda_k\lambda_{k'}}\mathbf{v}_k'\mathbf{v}_{k'} = 0, & k \neq k' \end{array} \right. ; \tag{3.180}$$

wegen $\mathbf{v}_k'\mathbf{v}_{k'} = 0$ für $k \neq k'$.

Die Zeilenvektoren $\tilde{\mathbf{a}}_j$ von A repräsentieren die Variablen ξ_j. Aus $W = QA'$ folgt $A' = Q'W$, so dass

$$\tilde{\mathbf{a}}_j = Q'\mathbf{w}_j, \tag{3.181}$$

und

$$\tilde{\mathbf{a}}_j'\tilde{\mathbf{a}}_{j'} = \mathbf{w}_j'QQ'\mathbf{w}_{j'} = \mathbf{w}_j'\mathbf{w}_{j'} = \|\mathbf{w}_j\|\|\mathbf{w}_{j'}\|\cos\theta_{jj'} = \cos\theta_{jj'} = r_{jj'}, \tag{3.182}$$

wegen $\|\mathbf{w}_j\|^2 = 1$ (s. auch (3.171)), $\theta_{jj'}$ der Winkel zwischen den Vektoren $\tilde{\mathbf{a}}_j$ und $\tilde{\mathbf{a}}_{j'}$. Für $j = j'$ ist $\theta_{jj'} = 0$, also $\cos\theta_{jj} = 1$ und es folgt insbesondere

$$\tilde{\mathbf{a}}_j'\tilde{\mathbf{a}}_j = \|\mathbf{w}_j\|^2 = r_{jj} = 1. \tag{3.183}$$

Dies bedeutet, dass die Variablen alle durch Vektoren der Länge 1 abgebildet werden, also durch Vektoren, deren Endpunkte auf einer r-dimensionalen Hyperkugel liegen, wenn r der Rang der Datenmatrix Z ist. Ist insbesondere $r = 2$, so liegen die Endpunkte auf einem Kreis.

Anmerkung: Die Aussage (3.183) ist eine Implikation der Spalten-, also der Variablenstandardisierung der Datenmatrix X. Betrachtet man die Kovarianzen

[71] $A = V\Lambda^{1/2}$ bedeutet, dass die Spaltenvektoren $\mathbf{a}_k$ von A Linearkombinationen der Spalten von V sind: $\mathbf{a}_k = V\Lambda_k^{1/2}$, wobei Λ_k der k-te Spaltenvektor von $\Lambda^{1/2}$ ist. Die Komponenten dieses Vektors sind gleich Null bis auf die k-te, die gleich $\sqrt{\lambda_k}$ ist.

statt der Korrelationen zwischen den Variablen ξ_j, so muß (3.183) nicht gelten! $\square$

Zusammengefasst hat man die Beziehungen

$$
\begin{aligned}
AA' &= V\Lambda^{1/2}\Lambda^{1/2}V' = V\Lambda V' = R, && (3.184)\\
A'A &= \Lambda^{1/2}V'V\Lambda^{1/2} = \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n) && (3.185)
\end{aligned}
$$

Die Elemente von AA' – die Korrelationen $r_{jj'}$ zwischen Variablen, s. (3.182), – sind Skalarprodukte der Zeilenvektoren $\tilde{\mathbf{y}}_j$ von A. Die Elemente von $A'A$ sind die Korrelationen bzw. Skalaarprodukte der Spaltenvektoren von A, s. (3.180).

Die Abbildung 20 illustriert den Sachverhalt (3.183):

> Die Begriffe Mann, Intelligenz, Vater etc waren die Variablen, die Fälle waren Adjektive, die Messungen waren Einschätzungen auf einer Rating-Skala, wie gut das jeweilige Adjektiv auf den Begriff zutrifft, – auf diese Weise werden Stereotype erfasst. Die Punkte, die die Begriffe repräsentieren, liegen nahe beim Einheitskreis, so dass die Stereotype gut durch eine 2-dimensionales System von latenten Variablen beschrieben werden kann. D_1 repräsentiert das "weibliche Prinzip", D_2 das "männliche Prinzip", die einzelnen Sterotype setzen sich anteilig aus diesen Dimensionen zusammen; die Anteile sind durch die Koordinaten der Begriffe auf den beiden Dimensionen gegeben. Entgegen geisteswissenschaftlichen Vorstellungen (z.B. Wellek, 1966)[72] sind die "Prinzipien" nicht entgegengesetzte Pole einer 1-dimensionalen Skala, wie sie durch die Gerade repräsentiert wird, die die Punkte "Mann" und "Frau" verbindet; die Projektionen der Punkte auf diese Gerade entsprechen aber der hermeneutisch aus Assoziationen zu den Begriffen 'Mann', 'Frau' etc. herausdestillierten Polarität von "männlich" und "weiblich", also als Gegensätze auf einer einzelnen Dimension. Tatsächlich sind die beiden Prinzipien voneinander unabhängig. In einer Person, gleich ob weiblich oder männlich, können also beide Prinzipien gleichermaßen vorhanden oder nicht vorhanden sein.

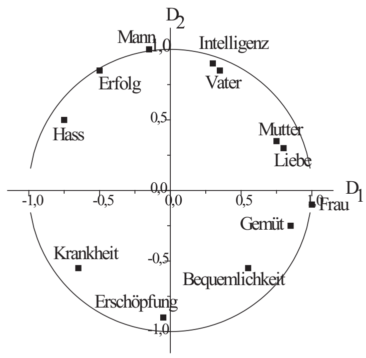
Während λ_k die Varianz der Projektionen der Fälle auf die k-te Hauptachse ist, ist $\|\mathbf{a}_k\|^2 = \lambda_k$ in (3.180) keine Varianz im strengen Sinn, weil die Komponenten von $\mathbf{a}_k$ nicht notwendig einen Mittelwert gleich Null haben, aber $\|\mathbf{a}_k\|^2$ als Quadrat der Länge von $\mathbf{a}_k$ kann gleichwohl als ein Maß für die Variation der Komponenten von $\mathbf{a}_k$ gesehen werden, – dass sie gerade gleich λ_k ist, also ein Maß für das Ausmaß, in dem das durch die k-te Hauptachse repräsentierte latente Merkmal zwischen den Variablen differenziert.

Die Betrachtungen zu den F-Scores, also den Komponenten u_{ik} der Vektoren $\mathbf{u}_k$ sind im Prinzip analog, wobei aber die Art der Standardisierung der Messungen (Zeilen- oder Spaltenstandardisierung) berücksichtigt werden muß. Dass $U'U = \Lambda$ eine Diagonalmatrix gilt ist bereits durch die Annahme **A2** festgelegt worden. Andererseits hat man wegen $WV = U$

$$UU' = WVV'W' = WW', \tag{3.186}$$

[72] Wellek, A.: Die Polarität im Aufbau des Charakters – System der konkreten Charakterkunde. Bern-München 1966

Abbildung 20: Begriffliche Stereotypen in den 50-er Jahren nach P. R. Hofstätter.



da ja $V'V = I$ auch $VV' = I$ impliziert. Dementsprechend hat man

$$\tilde{\mathbf{w}}_i'\tilde{\mathbf{w}}_{i'} = \tilde{\mathbf{u}}_i'\tilde{\mathbf{u}}_{i'} = \|\tilde{\mathbf{u}}_i\|\|\tilde{\mathbf{u}}_{i'}\|\cos\varphi_{ii'} \tag{3.187}$$

und $\varphi_{ii'}$ der Winkel zwischen den Vektoren, die die Fälle i und i' repräsentieren.

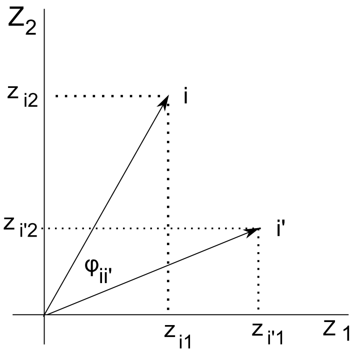
$$\tilde{\mathbf{w}}_i'\tilde{\mathbf{w}}_{i'} = \frac{1}{m}\sum_{j=1}^{n}\frac{(x_{ij}-\bar{x}_j)}{s_j}\frac{(x_{i'j}-\bar{x}_j)}{s_j} \tag{3.188}$$

$$\mathbf{w}_j'\mathbf{w}_k = \frac{1}{m}\sum_{i=1}^{m}\frac{(x_{ij}-\bar{x}_j)}{s_j}\frac{(x_{ik}-\bar{x}_k)}{s_k} \tag{3.189}$$

Die Skalarprodukte $\tilde{\mathbf{w}}_i'\tilde{\mathbf{w}}_{i'}$ sind keine Produkt-Moment-Korrelationen, – damit sie das sind, müsste eine Zeilenstandardisierung vorgenommen werden. Die $\tilde{\mathbf{w}}_i'\tilde{\mathbf{w}}_{i'}$ sind aber Ähnlichkeitsmaße für die Fälle i und i' (vergl. (3.187)). (3.189) ist natürlich eine Produkt-Moment-Korrelation. Da jede Achse ein Merkmal repräsentiert behält der Begriff der Orientierung (1-dimensionaler Teilraum) im Raum der Fälle seine Bdeutung als Repräsentation einer bestimmten Mischung von Merkmalen.

Abschätzung des Rangs $r < n$: Die Verteilung der Werte der Eigenwerte liefert Aufschluß über mögliche Werte des Ranges r von Z. Ein spezieller Fall gibt einen ersten Anhaltspunkt. Es gelte $\mathbf{z}_j'\mathbf{z}_{j'} = 0$ für alle $j \neq j'$, und $\mathbf{z}_j'\mathbf{z}_{j'}/m = 1$ für $j = j'$. Dann ist $\frac{1}{m}Z'Z = R = I$ die Identitätsmatrix. Nun ist $R = V\Lambda V'$, also $V\Lambda V' = I$, und so folgt $V'V\Lambda V'V = \Lambda = I$, d.h. $\lambda_j = 1$ für alle Variable j. Das heißt, es werden $r = n$ linear unabhängige Variable benötigt, um die Matrix Z darzustellen, und jede dieser Variablen trägt mit gleichem Anteil an der "Vorhersge" der $\mathbf{z}_j$ bei. Ungleiche Eigenwerte legen also nahe, dass entweder $r < n$ und/oder die latenten Variablen gehen mit ungleichem Gewicht in die Vorhersage ein. In der Praxis wird dieser Befund die Regel sein, da schon die stets vorhandenen Messfehler Korrelationen $\mathbf{z}_j'\mathbf{z}_{j'} \neq 0$ implizieren.

Abbildung 21: Zum Skalarprodukt zweier Fälle; die Komponenten sind $z_{ij} = (x_{ij} - \bar{x}_j)/s_j$ etc sind Mittelwerte und Standardabweichungen für die j-te Variable, vergl. (3.188).



Die Eigenwerte sind Varianzen: $\lambda_k = \|\mathbf{u}_k\|^2$, und $\sum_{k=1}^n \lambda_k$ ist dann die Gesamtvarianz der Daten. Der Quotient

$$\pi_k = \frac{\lambda_k}{\sum_{k=1}^n \lambda_k} \tag{3.190}$$

ist dann der Anteil der Gesamtvarianz, der durch die k-te Hauptachse erklärt wird. Es sei

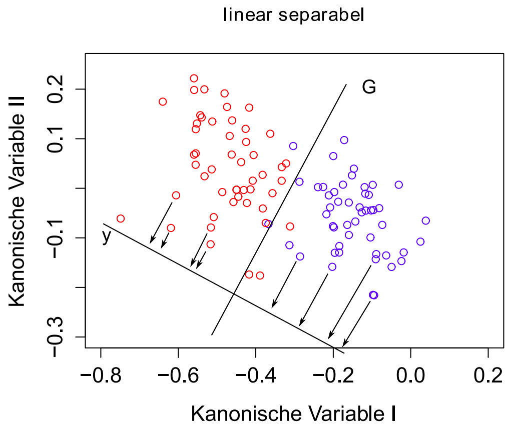
$$P_r = \frac{\sum_{k=1}^r \lambda_k}{\sum_{k=1}^n \lambda_k}, \quad r < n \tag{3.191}$$

Ein bestimmter Wert von P_r kann als Kriterium verwendet werden, um den Wert von r zu bestimmen, s. a. den Satz 3.23 von Eckart & Young uf Seite 132. Weitere Details und Beispiele findet man im Skriptum zur PCA, 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 22. Gegeben ist eine $(m \times n)$ Matrix X, deren 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 $\mathbf{u}$.

Abbildung 22: Klassifikation nach Fisher (1936) (I): Ω_1 blau, Ω_2 rot, eine mögliche Trennlinie G, eine Projektionsgerade Y



Da die Komponenten von $\mathbf{y}$ Koordinaten auf einer Koordinatenachse darstellen, kann man sie zweifach indizieren: die j-te Komponente von $\mathbf{y}$ sei sei die Koordinate y_{ik} des i-ten Falls in der k-ten Kategorie. $\bar{y}_k$ sei der Mittelwert der Koordinaten der k-ten Gruppe. $\mathbf{u}$ soll so bestimmt werden, dass die Varianz der $\bar{y}_k$, $k = 1, \ldots, K$ (es gebe K Gruppen oder Kategorien) maximal wird relativ zur durchschnittlichen Varianz innerhalb der Gruppen. Die Varianz der $\mathbf{a}y_k$ wird durch eine Quadratsumme QS_{zw} (zw für "zwischen") definiert, und die durchschnittliche Varianz innerhalb der Gruppen wird durch eine Quadratsumme QS_{inn} (inn für "innerhalb") definiert. Es soll

$$\lambda = \frac{QS_{zw}}{QS_{inn}} \tag{3.192}$$

maximiert werden; λ ist als *Diskriminanzkoeffizient* bekannt. Die Varianz aller Komponenten von $\mathbf{y}$ wird durch eine Quadratsumme QS_{tot} (tot für "total") definiert, und wie aus der Varianzanalyse bekannt gilt

$$QS_{ges} = QS_{inn} + QS_{zw}. \tag{3.193}$$

Da $\mathbf{u}$ bestimmt werden muss, muss der Quotient (3.192) aus Funktion des unbekannten Vektors $\mathbf{u}$ angeschrieben werden. Wegen $\mathbf{y} = X\mathbf{u}$ muss also $\mathbf{y}$ durch $X\mathbf{u}$ ersetzt werden.

Zunächste eine kleine Vorbetrachtung. Es seien $\mathbf{u} = (u_1, \ldots, u_n)'$ und $\mathbf{x} = (x_1, \ldots, x_n)'$ zwei n-dimensionale Vektoren. Für das Quadrat des Skalarprodukts $(\mathbf{x}'\mathbf{u})^2$ gilt

$$(\mathbf{x}'\mathbf{u})^2 = (x_1u_1 + \cdots x_nu_n)^2 = (x_1u_1)^2 + \cdots + (x_nu_n)^2 + \sum_{i \neq j} x_ix_ju_iu_j. \tag{3.194}$$

Der Ausdruck rechts erinnert an eine quadratische Form (vergl. (3.36)), Seite 91). In der Tat läßt sich das Produkt $u_i u_j$ in der Summe rechts als das (i,j)-te Element des dyadischen Produkts $\mathbf{uu}'$ interpretieren:

$$\mathbf{uu}' = \begin{pmatrix} u_1^2 & u_1u_2 & \cdots & u_1u_n \\ u_2u_1 & u_2^2 & \cdots & u_2u_n \\ \vdots & \vdots & \ddots & \vdots \\ u_nu_2 & u_nu_2 & \cdots & u_n^2 \end{pmatrix}$$

Man rechnet leicht nach, dass nun

$$(\mathbf{x}'\mathbf{u})^2 = \mathbf{x}'\mathbf{uu}'\mathbf{x} \tag{3.195}$$

gilt.

Es sei X eine (m,n)-Matrix; die m Fälle seien in K Gruppen mit den Umfängen $n_1, n_2, \ldots, n_K$ aufgeteilt, $m = \sum_{k=1}^K n_k$. Für jeden Fall werden Messwerte bei insgesamt n Variablen bestimmt. Die Matrix X läßt sich dann wie in (3.196) anschreiben

$$\mathbf{y} = \left(\begin{array}{c} Y_{11} \\ Y_{21} \\ \vdots \\ Y_{n_11} \\ \hline Y_{12} \\ Y_{22} \\ \vdots \\ Yy_{n_22} \\ \hline \vdots \\ \hline Y_{1K} \\ Y_{2K} \\ \vdots \\ Y_{n_KK} \end{array}\right), \quad X = \left(\begin{array}{cccc} X_{111} & X_{112} & \cdots & X_{11p} \\ X_{211} & X_{212} & \cdots & X_{21p} \\ \vdots & \vdots & \cdots & \vdots \\ X_{n_111} & X_{n_112} & \cdots & X_{n_11p} \\ \hline X_{121} & X_{122} & \cdots & X_{12p} \\ X_{221} & X_{222} & \cdots & X_{22p} \\ \vdots & \vdots & \cdots & \vdots \\ X_{n_221} & X_{n_222} & \cdots & X_{n_22p} \\ \hline \vdots & \vdots & \cdots & \vdots \\ \hline X_{1K1} & X_{1K2} & \cdots & X_{1Kp} \\ X_{2K1} & X_{2K2} & \cdots & X_{2Kp} \\ \vdots & \vdots & \cdots & \vdots \\ X_{n_KK1} & X_{n_KK2} & \cdots & X_{n_KKp} \end{array}\right), \tag{3.196}$$

Die Indizierung der Elemente von X bezieht sich (i) auf einen Fall in einer gegebenen Gruppe, (ii) auf die gegebene Gruppe, und (iii) auf eine Variable. Es soll eine Linearkombination $\mathbf{y}$ der Spaltenvektoren von X bestimmt werden, die eine möglichst gute Separierung der Gruppen gestattet. Dies bedeutet, dass ein Vektor $\mathbf{u}$ gesucht wird derart, dass

$$\mathbf{y} = X\mathbf{u}. \tag{3.197}$$

Man muss natürlich spezifizieren, was mit "möglichst gute Separierung der Gruppen" gemeint ist. Dem Ansatz (3.197) zufolge wird für jeden Fall eine Komponente von $\mathbf{y}$ bestimmt; die Indizierung der Komponenten von $\mathbf{y}$ werde dabei wie in (3.196) angegeben vorgenommen, so dass gruppenspezifische Mittelwerte $\bar{y}_k$ der Komponenten von $\mathbf{y}$ bestimmt werden können:

$$\bar{y}_k = \frac{1}{n_k}\sum_{i=1}^{n_k} Y_{ik}, \quad k = 1, \ldots, K \tag{3.198}$$

Die "möglichst gute Separierung der Gruppen" soll nun bedeuten, dass die $\bar{y}_k$ sich maximal voneinander unterscheiden sollen. Dies bedeutet, dass die Varianz der $\bar{y}_k$ so groß wie möglich sein soll, was natürlich die Einführung geeigneter Nebenbedingungen erfordert (die maximale Varianz wäre ohne Nebenbedingungen unendlich), auf die später eingangen wird. Wie aus der Varianzanalyse bekannt läßt sich nun die Gesamtvarianz, bzw. die ihr entsprechende Quadratsumme QS_{ges}, in eine Quadratsumme QS_{inn} und eine Quadratsumme QS_{zw} "zwischen" den Gruppen aufteilen; diese entspricht der Varianz der Mittelwerte der Gruppen. Man hat

$$QS_{ges} = \sum_{k=1}^{K}\sum_{i=1}^{n_k}(Y_{ik} - \bar{y})^2 \tag{3.199}$$

$$QS_{inn} = \sum_{k=1}^{K}\sum_{i=1}^{n_k}(Y_{ik} - \bar{y}_k)^2, \quad QS_{zw} = \sum_{k=1}^{K} n_k(\bar{y}_k - \bar{y})^2 \tag{3.200}$$

und es gilt (3.193). 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 (3.198) 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. (3.194))

$$QS_{inn} = \sum_{k=1}^{K}\sum_{i=1}^{n_k}\mathbf{u}'\tilde{\mathbf{x}}_{ik}\tilde{\mathbf{x}}'_{ik}\mathbf{u} = \mathbf{u}'\underbrace{\left(\sum_{k=1}^{K}\sum_{i=1}^{n_k}\tilde{\mathbf{x}}_{ik}\tilde{\mathbf{x}}'_{ik}\right)}_{W}\mathbf{u}. \tag{3.201}$$

$$QS_{inn} = \mathbf{u}'W\mathbf{u}. \tag{3.202}$$

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.

$$QS_{zw} = \mathbf{u}'B\mathbf{u} \tag{3.203}$$

Die Maximierung von QS_{zw}, d.h. der Maximierung der Gruppenmittelwerte $\bar{y}_k$, kann nun als Maximierung des Quotienten

$$\lambda(\mathbf{u}) := \frac{QS_{zw}}{QS_{inn}} = \frac{\mathbf{u}'B\mathbf{u}}{\mathbf{u}'W\mathbf{u}} \tag{3.204}$$

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 (3.104), 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. (3.38), Seite 93) 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

$$\lambda = \frac{\mathbf{v}'A\mathbf{v}}{\mathbf{v}'\mathbf{v}}, \tag{3.205}$$

d.h. λ entspricht einem Rayleigh-Quotienten. Nach dem Satz von Courant-Fisher wird λ maximal, wenn $\mathbf{v} = \mathbf{t}$, $\mathbf{t}$ der Eigenvektor von A, der zum maximalen Eigenwert λ_1 von A korrespondiert, d.h. es muss $A\mathbf{t} = \lambda_{\max}\mathbf{t}$ und damit

$$W^{-1/2}BW^{-1/2}\mathbf{t} = \lambda_{\max}\mathbf{t} \tag{3.206}$$

gelten. Multiplikation von links mit $W^{-1/2}$ liefert dann

$$W^{-1}BW^{-1/2}\mathbf{t} = \lambda_{\max}W^{-1/2}\mathbf{t}.$$

$\mathbf{t}$ ist ein spezieller Vektor für $\mathbf{v} = W^{1/2}\mathbf{u}$; setzt man $\mathbf{t} = W^{1/2}\mathbf{u}_t$, so folgt $\mathbf{t}W^{-1/2} = \mathbf{u}_t$ und man erhält

$$W^{-1}B\mathbf{u}_t = \lambda_{\max}\mathbf{u}_t. \tag{3.207}$$

$\mathbf{u}_t$ ist also ein Eigenvektor von $W^{-1}B$; $\lambda_{\max}$, der maximale Wert von λ, ist der zugehörige Eigenwert. $\mathbf{y}_t = X\mathbf{u}_t$ ist der Vektor, dessen Komponenten die Projektionen der Fälle auf eine Diskriminanzdimension sind, auf der die Gruppenmittelmittelwerte maximal separiert sind.

Die Gleichung (3.207) charakterisiert das generalisierte Eigenwertproblem, vergl. Gleichung (3.103), Seite 113, – man muss Gleichung (3.207) nur von links mit W multipliplizieren, und (3.205) ist ein generalisierter Rayleigh-Quotient, vergl. Gleichung (3.104), ebenfalls Seite 113. Existiert für (3.207) 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 3.25 *Sind $\boldsymbol{y}_j = X\boldsymbol{u}_j$ und $\boldsymbol{y}_k = X\boldsymbol{u}_k$ zwei verschiedene Vektoren,so gilt*

$$\boldsymbol{y}_j'\boldsymbol{y}_k = 0, \quad j \neq k \tag{3.208}$$

Beweis: Es ist

$$\begin{aligned}\mathbf{y}_j'\mathbf{y}_k = \mathbf{u}_j'X'X\mathbf{u}_k &= \mathbf{v}_jW^{-1/2}(X'X)W^{-1/2}\mathbf{v}_k \\ &= \mathbf{v}_jW^{-1/2}WW^{-1/2]}\mathbf{v}_k = \mathbf{v}_j'\mathbf{v}_k = 0\end{aligned}$$

denn $\mathbf{v}_j$ und $\mathbf{v}_k$ sind Eigenvektoren der symmetrischen Matrix $W^{-1/2}BW^{-1/2}$ und deswegen orthonormal, und W ist eine Schätzung für $X'X$, Gleichung (3.201).

□

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 42 betrachtet worden. Demnach ist eine Ebene durch

$$\mathbb{E} = \{\mathbf{x} | \, \mathbf{x} = \mathbf{u} + s\mathbf{b}_1 + t\mathbf{b}_2, \, s, t \in \mathbb{R}\} \tag{3.209}$$

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 (3.210) definiert wird. $B = [\mathbf{b}_1, \mathbf{b}_2]$ ist eine Basis für diesen 2-dimensionalen Teilraum des $\mathbb{R}^3$ (vergl. (1.116), Seite 52). Die allgemeine Darstellung der Ebene ist

$$\mathbb{E}_2^3 = \{\mathbf{p} < 0, \mathbf{p} = \mathbf{p}_0 + c_1\mathbf{b}_1 + c_2\mathbf{b}_2, \quad \mathbf{b}_1, \mathbf{b}_2 \text{ nicht parallel}, c_1, c_2 \in \mathbb{R}\} \tag{3.210}$$

(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 42, 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

$$c_1 = \mathbf{x}'\mathbf{b}_1 \tag{3.211}$$

$$c_2 = \mathbf{x}'\mathbf{b}_2 \tag{3.212}$$

und man erhält

$$\mathbf{y}_0 = (\mathbf{x}'\mathbf{b}_1)\mathbf{b}_1 + (\mathbf{x}'\mathbf{b}_2)\mathbf{b}_2, \tag{3.213}$$

also eine orthonormale Basisentwicklung von $\mathbf{y}_0$, s. Gleichung (1.117), Seite 52.

3.7.2 Der allgemeine Fall: Projektionsmatrizen

Es werde allgemein der $\mathbb{R}^n$ und eine k-dimensionale Ebene E_n^k betrachtet, wobei

$$E_n^k = \left\{ \mathbf{y} |\, \mathbf{y} = \sum_{j=1}^{k} c_j\mathbf{b}_j,\ c_j \in \mathbb{R},\ j = 1, \ldots, k < n \right\} \tag{3.214}$$

sei und die $\mathbf{b}_1, \ldots, \mathbf{b}_k$ linear unabhängige Vektoren aus $\mathbb{R}^n$ sind, d.h. die $\mathbf{b}_j$, $1 \le j \le k$ bilden eine Teilbasis des $\mathbb{R}^n$. Setzt man $B = [\mathbf{b}_1, \ldots, \mathbf{b}_k]$ und $\mathbf{c} = (c_1, \ldots, c_k)'$, so hat man

$$\mathbf{y} = B\mathbf{c}. \tag{3.215}$$

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 10, Seite 29. Weiter ist

$$(\mathbf{y}_0 - \mathbf{x})'\mathbf{b}_j = 0, \quad j = 1, \ldots, k \tag{3.216}$$

d.h.

$$\mathbf{y}_0'\mathbf{b}_j = \mathbf{x}'\mathbf{b}_j, \text{ oder } \mathbf{y}_0'B = \mathbf{x}'B \tag{3.217}$$

d.h. $B'\mathbf{y}_0 = B'\mathbf{x}$, d.h. $B'B\mathbf{c} = B'\mathbf{x}$ nach (3.215), so dass

$$\mathbf{c} = (B'B)^{-1}B'\mathbf{c}$$

folgt und die Multiplikation von links mit B liefert, wieder wegen (3.215),

$$\mathbf{y}_0 = \underbrace{B(B'B^{-1})B'}_{P}\mathbf{x} = P\mathbf{x} \tag{3.218}$$

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 3.26 *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 **??**, Seite **??**), hat P den Rang $\text{rg}(P) = m - 1$.

□

3.7.3 Beispiel: das Allgemeine Lineare Modell

Das Allgemeine Lineare Modell (ALM) ist durch

$$\mathbf{y} = X\mathbf{b} + \mathbf{e} \tag{3.219}$$

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: (3.219) liefert $\mathbf{y} - X\mathbf{b} = \mathbf{e}$ und man 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 6.9.3, Seite 194 (Anhang))

$$\hat{\mathbf{b}} = (X'X)^{-1}X'\mathbf{y}, \tag{3.220}$$

so dass

$$\mathbf{y} = X\hat{\mathbf{b}} + \hat{\mathbf{e}} = \hat{\mathbf{y}} + \hat{\mathbf{e}} \tag{3.221}$$

mit

$$\hat{\mathbf{y}} = X\hat{\mathbf{b}} = \underbrace{X(X'X)^{-1}X'}_{P}\mathbf{y} = P\mathbf{y}. \tag{3.222}$$

P erfüllt offenbar die Bedingungen 1. und 2. für eine Projektionsmatrix. Man hat demnach

$$\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{e}} = P\mathbf{y} + \hat{\mathbf{e}}, \tag{3.223}$$

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 orthgonalen Komplements von $\mathcal{L}(X)$ (s. Definition 1.17, Seite 56). Wegen $\mathbf{y} - P\mathbf{y} = \hat{\mathbf{e}}$ folgt

$$(I - P)\mathbf{y} = \hat{\mathbf{e}}. \tag{3.224}$$

Zusammenfassend hat man mit (3.222)

$$\hat{\mathbf{y}} = P\mathbf{y} \tag{3.225}$$
$$\hat{\mathbf{e}} = (I - P)\mathbf{y}, \tag{3.226}$$

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:

$$\begin{aligned}\hat{\mathbf{y}}'\hat{\mathbf{e}} &= (P\mathbf{y})'(\mathbf{y} - \hat{\mathbf{y}}) = \mathbf{y}'P'\mathbf{y} - \mathbf{y}'P'\hat{\mathbf{y}} \\ &= \hat{\mathbf{y}}'P\mathbf{y} - \mathbf{y}'PP\mathbf{y} = \hat{\mathbf{y}}'P\mathbf{y} - \hat{\mathbf{y}}'P\mathbf{y} = 0 \qquad (3.227)\end{aligned}$$

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 (3.227) gilt dementsprechend der Satz des Pythagoras

$$\|\mathbf{y}\|^2 = \|\hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{e}}\|^2. \tag{3.228}$$

Die Einflußmatrix Nach (3.222) 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 (3.222) 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:

$$\hat{y}_i = p_{i1}y_1 + p_{i2}y_2 + \cdots + p_{im}y_m. \tag{3.229}$$

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. 14, S. 84). 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

$$p_{ii} = \sum_{j=1}^{m} p_{ij}^2 = p_{ii}^2 + \sum_{i\neq j} p_{ij}^2 \tag{3.230}$$

folgt. Man sieht leicht, dass diese Beziehung nicht gelten kann, wenn die $p_{ij} > 1$ sein können, so dass folgt, dass

$$0 \leq p_{ij} \leq 1. \tag{3.231}$$

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

$$\sum_{i=1}^{m} p_{ii} = n \tag{3.232}$$

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 (3.230) 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

$$\ell_{ik} = \tilde{\mathbf{x}}_i'\mathbf{t}_k. \tag{3.233}$$

ℓ_{ik} ergibt sich als Projektion des Vektors $\tilde{\mathbf{x}}_i$ auf eine Gerade durch den Nullpunkt des Koordinatensystems mit em Orientíerungsvektor $\mathbf{t}_k$, der als Eigenvektor von $X'X$ die Länge 1 hst, $\|\mathbf{t}_k\| = 1$. Tatsächlich ist nach (1.72), Seite 30,

$$\tilde{\mathbf{x}}_{i,\mathbf{t}_k} = \frac{\tilde{\mathbf{x}}_i'\mathbf{t}_k}{\mathbf{t}_k'\mathbf{t}_k}\mathbf{t}_k,$$

und

$$\|\tilde{\mathbf{x}}_{i,\mathbf{t}_k}\| = |\tilde{\mathbf{x}}_i'\mathbf{t}_k|\|\mathbf{t}_k\| = |\tilde{\mathbf{x}}_i'\mathbf{t}_k|,$$

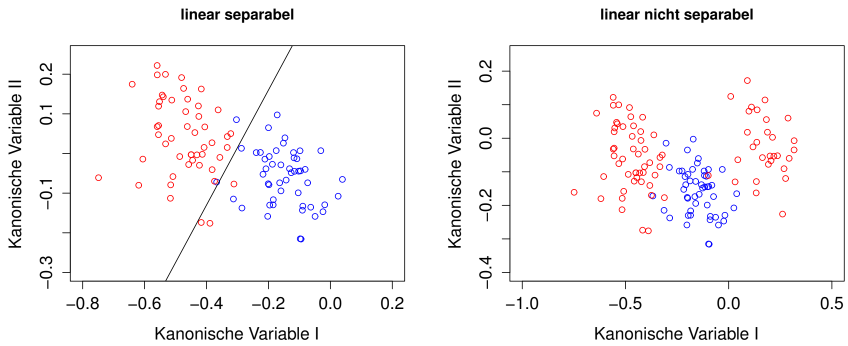
und dies ist der Betrag des Skalarprodukts ℓ_{ik} in (3.233) (vergl Gleichung (1.73), Seite 30).

4 Nichtlineare latente Strukturen

4.1 Nichtlineare Funktionen und Kernfunktionen

In den klassischen Verfahren wie etwa PCA, Diskriminzanalyse, etc wird davon ausgegangen, dass gemessenen Variaben ebenso wie die Fälle als Linearkombinationen "latenter" Vektoren darstellen lassen, wobei die latenten Vektoren sich ebenfalls als Linearkombinationen gemessener Größen darstellen lassen. So kann man bei einer (m,n)-Datenmatrix X wegen der allgegenwärtigen Messfehler davon ausgehen, dass für den Fall $m \geq n$ X den Rang n hat und man von der algebraisch stets gültigen Beziehung $X = UV'$ ausgehen kann, wobei U eine (m,n)-Matrix und V eine (n,n)-Matrix ist; beide sind ebenfalls vom Rang n, so dass die Inverse Matrix V^{-1} existiert. Dann ist $U = XV^{-1}$ und die Spaltenvektoren von U ergeben sich als Linearkombinationen der Spaltenvektoren von X, sind aber nicht notwendig orthogonal, wenn V keine Rotationsmatrix ist. Das ist mathematisch elegant, unterstellt aber eine lineare Beziehung zwischen latenten und gemessenen Variablen. So kann man bei Konfigurationen der Fälle wie in Abbildung 23 eine Gerade – also eine lineare bzw. affine Funktion – finden, die zwei Subpopulationen optimal voneinander trennt, wobei 'optimal' heißt, dass die Wahrscheinlichkeit von Fehlklassifikationen minimiert wird. Im zweiten Beispiel sind die zwei Subpopulationen nicht linear trennbar. Es werden Klassifikationsaufgaben betrachtet.

Abbildung 23: Linear trennbare und linear nicht trennbare Konfigurationen



Zur Illustration werde angenommen, dass eine Klassifikation anhand von nur zwei Variablen vorgenommmen werden werden soll, der "Input" demnach durch einen Vektor $\mathbf{x} = (x_1, x_2)'$ gegeben ist. Die Frage ist, ob eine Trennung von zwei Subpopulationen wie in Abb. 23, linear separabel, möglich ist, oder ob eine Situation wie in Abb. Abb. 23, nicht linear separabel, angenommen werden muß. Auch in diesem Fall lassen sich die beiden Populationen trennen, aber nicht einen Teilraum des $\mathbb{R}^2$, also eine Gerade sucht, sondern indem man umgekehrt den $\mathbb{R}^2$ zu einem dreidimensionalen Raum erweitert und die Klasse der blauen Fälle bezüglich der dritten Dimension von der Klasse der roten Fälle separiert. Die roten Fälle einerseits und die blauen andererseits können durch eine Ebene, also eine lineare Funktion, separiert werden.

Es werde zunächst die Möglichkeit einer linearen Klassifikation betrachtet. Für eine gegebene Menge Ω von "Objekten" (die auch Personen sein können) liegt für jedes Objekt $\omega \in \Omega$ ein Vektor $\mathbf{x}(\omega) = (x_1, \ldots, x_n)'$ vor; die Komponenten x_j sind natürlich für jedes ω charakteristisch, aber zur Vereinfachung der Notation wird die Schreibweise $x_j(\omega)$ aufgegeben. Ω sei die Vereinigung zweier durchschnittsfremder Klassen Ω_1 und Ω_2. Gesucht wird ein Vektor $\mathbf{w} = (w_1, \ldots, w_n)'$ derart, dass für jeden Vektor $\mathbf{x}$ die *lineare Diskriminanzfunktion*

$$g(\mathbf{x}) = w_1 x_1 + \cdots + w_n x_n = \mathbf{w}'\mathbf{x} = \langle \mathbf{w}, \mathbf{x} \rangle \tag{4.1}$$

berechnet wird, so dass mit minimal möglichen Fehler gemäß

$$g(\mathbf{x}) = \begin{cases} g(\mathbf{x}) \in K_1, & \omega \in \Omega_1 \\ g(\mathbf{x}) \in K_2, & \omega \in \Omega_2 \end{cases} \tag{4.2}$$

entschieden werden kann (das Schema kann auf mehr als zwei Alternativen verallgemeinert werden. Jedenfalls ist $g(\mathbf{x})$ nach (4.1) ein Skalarprodukt.

Um eventuelle Nichtlinearitäten modellieren zu können werden *feature mappings* – also Merkmalsabbildungen – eingeführt. Diese Abbildungen sind nichtlineare Funktionen der Komponenten x_j von $\mathbf{x}$. Ist z.B. $\mathbf{x} \in \mathbb{R}^n$ ein Vektor, dessen

Komponenten x_j Messungen von Variablen sind, die für die Klassifikation von Fällen verwendet werden sollen. So sei eine Menge Ω von Objekten gegeben, ein Objekt $\omega \in \Omega$ wird durch einen Vektor $\mathbf{x}(\omega) = (x_1, \ldots, x_n)'$ beschrieben. Für $\omega \in \Omega_1$ gehört ω zur Klasse C_1, für $\omega \in \Omega_2$ gehört ω zur Klasse C_2.

Beispiel 4.1 So sei $n = 2$, also $\mathbf{x} = (x_1, x_2)'$ und es sei

$$\phi_1 = x_1^2, \quad \phi_2 = x_2^2, \quad \phi_3 = \sqrt{2}x_1 x_2.$$

Man erhält dementsprechend einen 3-dimensionalen Vektor

$$\phi(\mathbf{x}) = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{pmatrix} = \begin{pmatrix} x_1^2 \\ x_2 \\ \sqrt{2}x_1 x_2 \end{pmatrix} \tag{4.3}$$

und es kann eine Diskriminanzfunktion analog zu (4.1) erklärt werden:

$$\mathbf{y}(\mathbf{x}) = b_{11}x_1^2 + b_{22}x_2^2 + b_{12}\sqrt{2}x_1 x_2 = \langle \phi, \mathbf{b} \rangle, \quad \mathbf{b} \in \mathbb{R}^3 \tag{4.4}$$

Statt wie in der klassischen linearen Analyse die Dimensionalität der Beschreibung der Objekte zu reduzieren wird sie nun erhöht, – hier von zwei auf drei Dimensionen, um dann auf die Vektoren ϕ wieder Begriffe aus der linearen Analyse anzuwenden.

Beispiel 4.2 Wie man für zwei Vektoren $\mathbf{x}$ und $\mathbf{z}$ das Skalarprodukt $\mathbf{x}'\mathbf{z} = \sum_i x_i z_i$ berechnen kann, so kann auch ein Skalarprodukt für $\phi(\mathbf{x})$ und $\phi(\mathbf{z})$ berechnet werden:

$$\begin{aligned} \phi(\mathbf{x})'\phi(\mathbf{z}) = \langle \phi(\mathbf{x}), \phi(\mathbf{z}) \rangle &= \phi_{x1}\phi_{z1} + \cdots + \phi_{x3}\phi_{z3} \\ &= x_1^2 z_2^2 + x_2^2 z_2^2 + 2x_1 z_1 x_2 z_2 \qquad (4.5) \\ &= (x_1 z_1 + x_2 z_z)^2 = \langle \mathbf{x}, \mathbf{z} \rangle^2 \qquad (4.6) \end{aligned}$$

Dieses Beispiel läßt sich auf den Fall allgemein n-dimensionaler Vektore $\mathbf{x}, \mathbf{z}$ verallgemeinern:

Es seien $\mathbf{x}, \mathbf{z} \in \mathbb{R}^n$ und ϕ sei durch

$$\phi(\mathbf{x}) = \{x_i x_j\}_{i,j=1}^n \in V = \mathbb{R}^{n^2}$$

definiert, d.h. durch die Folge

$$\{x_i x_j\}_{i,j=1}^n = \{x_1 x_1, x_2 x_1, \ldots, x_n x_n\},$$

also durch die Folge aller möglichen Paare der Komponenten von $\mathbf{x} = (x_1, \ldots, x_n)'$. Es sei $\mathbf{z}$ ebenfalls ein Vektor aus $\mathbb{R}^n$, und $\phi(\mathbf{z})$ sei wie $\phi(\mathbf{x})$ definiert. Für das Skalarprodukt $\langle \phi(\mathbf{x}), \phi(\mathbf{z}) \rangle$ erhält man

$$\begin{aligned} \langle \phi(\mathbf{x}), \phi(\mathbf{z}) \rangle &= \sum_{i=1}^n \sum_{j=1}^n x_i x_j z_i z_j \\ &= \sum_{i=1}^n x_i z_i \sum_{j=1}^n x_j z_j = \langle \mathbf{x}, \mathbf{z} \rangle^2, \qquad (4.7) \end{aligned}$$

analog zu (4.5). □

Der Kernel-Trick Das Bemerkenswerte an (4.6) bzw. (4.7) ist nun, dass das Skalarprodukt $\langle\phi(\mathbf{x}), \phi(\mathbf{z})\rangle$ eine nichtlineare Funktion des Skalarprodukts $\langle\mathbf{x}, \mathbf{z}\rangle = \mathbf{x}'\mathbf{z}$ von $\mathbf{x}$ und $\mathbf{z}$ ist, im Fall (4.7) des Quadrats $(\mathbf{x}'\mathbf{z})^2$. Dieses Resultat bedeutet, dass man das Skalarprodukt $\phi(\mathbf{x})'\phi(\mathbf{z})$ gar nicht gemäß (4.5) auszurechnen braucht, es genügt, eine nichtlineare Funktion des Skalarprodukts $\langle\mathbf{x}, \mathbf{z}\rangle$ der Ausgangsvektoren $\mathbf{x}$ und $\mathbf{z}$ zu berechnen. Man führt für diese Funktion eine eigene Bezeichnung ein[73]:

$$\kappa(\mathbf{x}, \mathbf{z}) := \langle\phi(\mathbf{x}), \phi(\mathbf{z})\rangle \tag{4.8}$$

$\kappa(\mathbf{x}, \mathbf{z})$ ist ein Skalarprodukt, daas von $\mathbf{x}$ und $\mathbf{z}$ über die nichtlinearen Funktionen $\phi(\mathbf{x})$ und $\phi(\mathbf{z})$ abhängt. Im Falle (4.7) ist insbesondere $\kappa(\mathbf{x}, \mathbf{z}) = \langle\mathbf{x}, \mathbf{z}\rangle^2$. κ steht zwar für das Skalarprodukt $\langle\phi(\mathbf{x}), \phi(\mathbf{z})\rangle$ und wird hier als Funktion der gegebenen Vektoren $\mathbf{x}$ und $\mathbf{z}$ interpretiert. Die Definition von κ kann auf den Fall, dass $\mathbf{x}$ nicht mehr notwendig ein Vektor aus $\mathbb{R}^n$ ist, verallgemeinert werden (Schölkopf et al. (2002), p. 32):

Definition 4.1 *Es sei $\mathcal{X}$ eine nichtleere Menge, und κ sei eine Funktion auf $\mathcal{X} \times \mathcal{X}$, so dass für alle $m \in \mathbb{N}$ alle $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ eine positiv-definite Gram-Matrix mit den Elementen $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)$ erzeugt wird. Dann heißt κ* positiv-definiter Kern (pd-Kern).

Beispiele für Kern-Funktionen: $\kappa(\mathbf{x}, \mathbf{z}) = \langle\mathbf{x}, \mathbf{z}\rangle^p$, $p \in \mathbb{N}$, oder $\kappa(\mathbf{x}, \mathbf{z}) = \exp(-\|\mathbf{x} - \mathbf{z}\|)$.

Kern-Trick (kernel trick): (Schölkopf et al., p. 34, Bishop (2006), p. 292), auch *Kern-Substitution*) Wie insbesondere in Beispiel 4.2 illustriert wurde kann eine nichtlineare Transformation der ursprünglichen Variablen (Komponenten von) $\mathbf{x}$, $\phi(\mathbf{z})$ vorgenommen werden, was mit einer Erhöhung der Dimensionalität einhergeht, und für die neuen Dimensionen ein Skalarprodukt $\langle\phi(\mathbf{x}), \phi(\mathbf{z})\rangle$ definiert werden, wobei sich zeigt, dass dieses Skalarprodukt gleich einer *nichtlinearen Transformation des Skalarprodukts* $\langle\mathbf{x}, \mathbf{z}\rangle$ ist. Dies ist der *Kern-Trick* Dieser "Trick" erlaubt es, lineare Methoden, z.B. PCAs, oder Klassifikationen, auch auf zunächst nicht linear-trennbare Daten anzuwenden. Dazu werden die Daten in einen höherdimensionalen Raum transformiert, in dem eine lineare Trennung von Klassen erwartet werden kann. □

Bis hier sind nur endlich-dimensionle Vektoren betrachtet worden, d.h. n-Tupel von reellen Zahlen mit $n < \infty$. Sind $\mathbf{x}_1, \ldots, \mathbf{x}_m$ Elemente aus $\mathbb{R}^n$, so können sie stets als Linearkombinationen von $r \leq \min(m, n)$ lineare unabhängigen Vek-

[73] κ ist der griechische Buchstabe kappa. Der Mathematiker David Hilbert diskutierte Integrale der Form

$$(T_k f)(x) = \int_{\mathcal{X}} \kappa(x, x') f(x') dx',$$

wobei er die Funktion $\kappa(x, x')$ als den 'Kern' des Integrals bezeichnete. Mit ":=" soll in der folgenden Gleichung angezeigt werden, dass κ durch die rechte Seite der Gleichung definiert wird.

toren $\mathbf{u}_1, \ldots, \mathbf{u}_r$ dargestellt werden, und die Skalarprodukte

$$\mathbf{x}'_j \mathbf{x}_{j'} = \sum_{i=1}^{n} x_{ij} x_{ij'}$$

als Summen von endlich vielen reellen Zahlen existieren stets.

Hilbert-Räume Man kann aber auch über einem Intervall $[a, b]$ definierte Funktionen als Vektoren interpretieren. Die Begriffe Linearkombination und Skalarproduktion übertragen sich auf diesen Fall, allerdings ergeben sich bestimmte Unterschiede. Ist f eine auf $[a, b]$ definierte Funktion, so ist die Menge der Werte $f(x)$ für $x \in [a, b]$, i.A. nicht mehr endlichm sondern überabzählbar, und die Anzahl der linear unbhängigen Basisfunktionen ist nicht mehr notwendig endlich. Das Skalarprodukt zweier Funktionen f und g ist durch

$$\langle f, g \rangle = \int_a^b f(x)g(x)dx \tag{4.9}$$

gegeben, – falls das Integral existiert. Dies gilt insbesondere für den Fall $f = g$:

$$\langle f, f \rangle = \|f\|^2 = \int_a^b |f(x|^2 dx \tag{4.10}$$

Funktionen, für die dieses Integral existiert, heißen *quadratintegrierbar*; die Menge dieser Funktionen wird mit ℓ^2 bezeichnet. jjjjj

XXXX

Hier ausführlicher werden, Courant-Hilbert wegen derkonvergenzprobleme zitieren!!!

Für endlich-dimensionale Vektorräume – also Mengen von Vektoren $\mathbf{x} \in \mathbb{R}^n$ – ist die Definition von Skalarprodukten im Allgemeinen kein Problem. Anders ist es, wenn etwa die Vektoren durch Funktionen definiert sind. Man hat es dann mit unendlichdimensionalen Räumen zu tun; die Skalarprodukte sind dann durch Integrale gegeben, die aber für die betrachteten Funktionen nicht existieren müssen. Man muß sich also auf Funktionen beschränken, für die die Skalarprodukte definiert sind.

4.2 Kernreproduzierende Hilbert-Räume (RKHS)

Das Ziel ist, Skalarprodukte $\langle \phi(\mathbf{x}), \phi(\mathbf{z}) \rangle$ durch Kernfunktionen $\kappa(\mathbf{x}, \mathbf{z})$ darzustellen. Die Menge der Funktionen $\langle \cdot, \cdot \rangle$ bildet einen Vektorraum, und wie bei Vektorräumen können Funktionen punktweise als Linearkombinationen von Funktionen dargestellt werden. In den bisher behandelten Vektorräumen bestimmt die Anzahl der Komponenten der Vektoren die Dimensionalität des Vektorraums. Für die punktweise Darstellung von Funktionen als Linearkombinationen bedeutet dies, dass der in Frage kommende Vektorraum unendlich dimensional sein kann. Da Skalarprodukte der Form $\langle \phi(\mathbf{x}), \phi(\mathbf{z}) \rangle$ dargestellt werden sollen, muß für den Vektorraum ein Skalarprodukt erklärt sein, – während für die endlichdimensionalen Vektorräume die Definition eines Skalarprodukts kein Problem ist, muß ein Skalarprodukt für Funktionen nicht notwendig existieren, die Forderung nach der Existenz eines solchen Produkts stellt also eine Einschränkung für die Wahl der betrachteten Funktionen dar. Dem eben Gesagten entsprechend muß der Raum der Funktionen ein HilbertRaum sein:

Definition 4.2 *Gegeben sei ein Vektorraum $\mathcal{H}$ über der Menge $\mathbb{R}$ der reellen oder der komplexen Zahlen $\mathcal{C}$, in dem ein Skalarprodukt $\langle\cdot,\cdot\rangle$ definiert, durch das eine Norm $\langle \boldsymbol{x}, \boldsymbol{x}\rangle^{1/2}$ induziert wird, und $\mathcal{H}$ ist vollständig*[74]*. Dann ist $\mathcal{H}$ ein* Hilbert-Raum[75].

Es sei κ eine reellwertige, positiv definite Kernfunktion, und $\mathcal{X}$ sei eine nichtleere Menge; weiter sei

$$\mathbb{R}^{\mathcal{X}} := \{f : \mathcal{X} \to \mathbb{R}\};$$

$\mathbb{R}^{\mathcal{X}}$ ist die Menge der Funktionen, die $\mathcal{X}$ in $\mathbb{R}$ abbilden:

$$\begin{aligned} \phi : \mathcal{X} &\to \mathbb{R}^{\mathcal{X}}, \\ \mathbf{x} &\mapsto \kappa(\cdot, \mathbf{x}) \end{aligned}$$

d.h. $\phi(x)(\cdot) = \kappa(\cdot, x)$ ($\cdot$ bezeichnet einen Platzhalter). $\phi(\mathbf{x})$ bezeichnet die Funktion, die κ den Wert $\kappa(\mathbf{z}, \mathbf{x})$ für $\mathbf{z} \in \mathcal{X}$ zuordnet (Schölkopf et al., p. 32). Es wird angenommen, dass der korrepondierende Vektorraum ein Hilbert-Raum ist.

Es werden insbesondere *reproduzierende Kernfunktionen* betrachtet

$$f(\cdot) = \sum_{i=1}^{m} \alpha_i \kappa(\cdot, \mathbf{x}_i), \quad m \in \mathbb{N},\, \alpha_i \in \mathbb{R},\, (\mathbf{x}_1, \ldots, \mathbf{x}_m)' \in \mathcal{X} \tag{4.11}$$

$$g(\cdot) = \sum_{j=1}^{m'} \beta_j \kappa(\cdot, \mathbf{z}_j) \tag{4.12}$$

spezifiziert. Das gesuchte Skalarprodukt kann wie folgt erklärt werden:

Definition 4.3 *Das Skalarprodukt $\langle f, g\rangle$ sei durch*

$$\langle f, g\rangle := \sum_{i=1}^{m} \sum_{j=1}^{m'} \alpha_i \beta_j \kappa(x_i, z_j) \tag{4.13}$$

[74]Die Norm ordnet jedem Vektor in $\mathcal{H}$ eine "Länge" oder "Größe" zu, und Vollständigkeit von $\mathcal{H}$ bedeutet, dass jede Cauchy-Folge in $\mathcal{H}$ gegen einen Vektor in $\mathcal{H}$ konvergiert. Eine Folge $\{a_n\}_{n\in\mathbb{N}}$ ist eine Cauchy-Folge, wenn die Differenzen zwischen Gliedern a_n, a_m mit wachsendem n und m gegen Null gehen. Vollständigkeit bedeutet, dass es keine "Löcher" in $\mathcal{H}$ gibt, dh dass es für gewünschte Vektoren (Funktionen) stets eine beliebig genaue Approximation gibt.

[75]Nach David Hilbert (1862 – 1943), deutscher Mathematiker

gegeben[76].

Natürlich kann man die Reihenfolge der Summation vertauschen:

$$\sum_{i=1}^{m}\sum_{j=1}^{m'}\alpha_i\beta_j\kappa(x_i,z)=\sum_{j=1}^{m'}\sum_{i=1}^{m}\alpha_i\beta_j\kappa(x_i,z)=\sum_{j=1}^{m'}\beta_j\underbrace{\sum_{i=1}^{m}\alpha_i\kappa(x_i,z)}_{f(\mathbf{z}_j)}=\sum_{j=1}^{m'}\beta_j f(\mathbf{z}_j)$$

d.h. es gilt

$$\langle f,g\rangle=\sum_{j=1}^{m'}\beta_j f(\mathbf{z}_j). \tag{4.14}$$

In analoger Weise folgt

$$\langle f,g\rangle=\sum_{i=1}^{m}\alpha_i g(\mathbf{x}_i) \tag{4.15}$$

Damit ist gezeigt, dass $\langle\cdot,\cdot\rangle$ bilinear ist. Auch zeigt sich, dass $\langle f,g\rangle=\langle g,f\rangle$ gilt, d.h. $\langle\cdot,\cdot\rangle$ ist symmetrisch. Weiter findet man

$$\langle f,f\rangle=\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j\kappa(\mathbf{x}_i,\mathbf{x}_j)\geq 0, \tag{4.16}$$

d.h. $\langle\cdot,\cdot\rangle$ ist positiv semi-definit. Es sei insbesondere $f=\kappa(\cdot,\mathbf{x})$, $g=\kappa(\cdot,\mathbf{x}')$. Dann folgt aus (4.13)

$$\langle f,g\rangle=\kappa(\mathbf{x},\mathbf{x}'),$$

wenn man für ein $\alpha_i=1$, und für ein $\beta_j=1$ setzt und die restlichen $\alpha-$ und $\beta-$Werte gleich Null setzt, so folgt die *representer evaluation*

$$\langle\kappa(\cdot,\mathbf{x}),f\rangle=f(x), \tag{4.17}$$

und damit

$$\langle\kappa(\cdot,\mathbf{x}),\kappa(\cdot,\mathbf{x}')\rangle=\kappa(\mathbf{x},\mathbf{x}') \tag{4.18}$$

(explizit machen, siehe Bleistifteintrag Schölkopf et al. p. 33). Deshalb heißen die κ-Kerne *reproduzierende Kerne*. Aus der Cauchy-Schwarz-Ungleichung (vergl. Satz 1.2, Seite 30)

$$|\kappa(x_1,x_2)|^2\leq\kappa(x_1,x_1)\kappa(x_2,x_2) \tag{4.19}$$

(Schölkopf et al., p. 21) folgt dann

$$|f(x)|^2=|\langle\kappa(\cdot,\mathbf{x}),f\rangle|^2\leq\kappa(\mathbf{x},\mathbf{x}')\cdot\langle f,f\rangle. \tag{4.20}$$

[76] Schhölkopf et al., p.33, führen diese Beziehung als Definition ein, indem sie das :=-Zeichen verwenden. Weiter unten führen sie den Ausdruck $\langle\kappa(\cdot x),\kappa(\cdot x')\rangle=\kappa(x,x)$ ein (als *representer evaluation*), ohne daran zu erinnern, dass diese Beziehung aus der *Definition* (4.13) folgt, was zu Verständnisschwierigkeiten führen kann, s. Stackexchange -Eintrag zu *representer-of-evaluation.* deswegen wird die Definition $\langle f,g\rangle$ hier durch eine explizite Definition gekennzeichnet.

Hieraus folgt $\langle f, f\rangle = 0$ genau dann, wenn $f = 0$, womit gezeigt ist, dass $\langle \cdot, \cdot\rangle$ ein inneres, d.h. ein Skalarprodukt ist, und die Eigenschaft, ein reproduzierender Kern zu sein, bedeutet, dass tatsächlich

$$\langle \phi(\mathbf{x}), \phi(\mathbf{x}')\rangle = \kappa(\mathbf{x}, \mathbf{x}') \tag{4.21}$$

gilt. Diese Bezeichung besagt nun, dass Kerne durch Skalarprodukte der ϕ-Funktionen definiert werden können; dies ist der Kern-Trick, s. oben.

Das Representer-Theorem Intuitive Formulierung $J(\mathbf{w}) = \min$ (Bishop. p. 293) für bestimmten Vektor $\mathbf{w}$ minimal, Dann exististieren für die $\phi(x_i)$ reproduzierende Kernfunktionen, durch die $J(\mathbf{w})$ ausgedrückt werden kann.

Betrachtet werde eine Diskriminanz- oder Regressionsfunktion entsprechend (4.24), Seite 158,

$$\mathbf{y}(\mathbf{x}) = w_1\phi(\mathbf{x}_1) + \cdots, +w_n\phi(\mathbf{x}_n) + \mathbf{e}$$

Nee: $\mathbf{w} \in \mathbb{R}^m$,, $\phi(\mathbf{x}_j) \in \mathbb{R}^m$, $\mathbf{w}'\phi(\mathbf{x}_i)$, $j = 1, \ldots, N$

Ab hier Shawe-Taylor:

Definition 4.4 *Das Skalarprodukt*

$$\kappa(\boldsymbol{x}, \boldsymbol{z}) = \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{z})\rangle \tag{4.22}$$

heißt Kernfunktion *(engl.: kernel function) oder* Kern*; ein* Kern *ist eine Funktion, wobei ϕ durch*

$$\phi:\ \boldsymbol{x} \in \mathbb{R}^n \longmapsto \phi(\boldsymbol{x}) \in \mathcal{F} \subseteq \mathbb{R}^N, \tag{4.23}$$

definiert ist. $\mathcal{F}$ ist der Merkmalsraum *(*Feature Space*). Im Allgemeinen ist $N \geq n$.*

Kerne können als generalisierte Skalarprodukte betrachtet werden. Viele Aussagen über Skalarprodukte von Vektoren aus $\mathbb{R}^n$ übertragen sich auf Kerne.

Gleichung (4.3) ist ein Beispiel für das Bild eines Vektors $\mathbf{x} = (x_1, x_2)'$ in einem 3-dimensionalen Merkmalsraum ($N = 3$), der die Dimensionen x_1^2, x_2^2 und $\sqrt{2}x_1x_2$ hat. Bei der Klassifikation eines Objekts werden also nicht, wie im linearen Fall, Werte der die Variablen x_1 und x_2 direkt "verrechnet", sondern die "Features" x_1^2, x_2^2 und $\sqrt{2}x_1x_2$. Um Klassifikationen in $\mathcal{F}$ vornehmen zu können erhält man für

$$\phi(\mathbf{x}) = (\phi(x_1), \phi(x_2), \ldots, \phi(x_m), y_m)'$$

die Diskriminanzfunktion

$$g(\mathbf{x}) = b_1\phi(x_1) + \cdots + b_m\phi(x_m) = \langle \mathbf{b}, \phi(\mathbf{x})\rangle \tag{4.24}$$

d.h. die Klassifikations- bzw. Diskriminanzfunktion ist wieder eine lineare Funktion, die aber nicht direkt auf den Vektoren $\mathbf{x}$, sondern auf den Vektoren ϕ. kkkk

Der Vektor $\mathbf{b} = (b_1, \ldots, b_N)'$ muß geschätzt werden. Deswegen wird wie bei der linearen Regression wie in Abschnitt 3.4.2 eine Verlust-Funktion, die für geeignet gewählten Vektor $\mathbf{b}$ minimalisiert wird: mit $\mathbf{y} = g(\mathbf{x}) + \mathbf{e}$, $\mathbf{e}$ der übliche "Fehlervektor", hat man die Verlustfunktion

$$\ell(\mathbf{x}, \mathbf{y}) = |\mathbf{y} - g(\mathbf{x})| = |\mathbf{y} - \langle \mathbf{b}, \phi(\mathbf{x}) \rangle| \tag{4.25}$$

Legt man hierfür den Ridge-Regressions-Ansatz zugrunde (Abschnitt 3.4.2, Seite 120), so wird man für $g(\mathbf{x})$ auf

$$g(\mathbf{x})) = \mathbf{y}'(G + \lambda I_m)^{-1}\mathbf{k} \tag{4.26}$$

geführt, wobei nun G eine Gram-Matrix mit den Elementen

$$G_{ij} = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle, \quad i, j = 1, \ldots, m \tag{4.27}$$

ist, d.h. eine Reihe von X sind die Feature-Vektoren $\phi(\mathbf{x}_1)', \ldots, \phi(\mathbf{x}_m)'$. $\mathbf{k}$ ist ein Vektor mit den Elementen

$$k_i = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}) \rangle. \tag{4.28}$$

(vergl. (3.133), Seite 122).

Wenn von einem Kern $\kappa(\mathbf{x}, \mathbf{z})$ die Rede ist, so wird oft auch von der *Kern-Funktion* gesprochen (z.B. Shawe-Taylor & Christianini (2004), p. 34)); gemeint ist das Skalarprodukt $\langle \phi(\mathbf{x}), \phi(\mathbf{z}) \rangle$ als Funktion von $\mathbf{x}$ bzw. $\mathbf{z}$; der Ausdruck soll ausdrücken, dass κ eben von $\mathbf{x}$ und $\mathbf{z}$ abhängt. In diesem Sinne wird der Ausdruck auch im Folgenden verwendet. Es soll auch die Möglichkeit unendlich-dimensionaler Vektoren in Betracht gezogen werden. Damit der Begriff des Kerns dann noch Sinn macht, müssen gewisse Voraussetzungen insbesondere für den Feature-Raum $\mathcal{F}$ erfüllt sein. Diese sind

1. Für jedes Paar $\phi(\mathbf{x}), \phi(\mathbf{z}) \in \mathcal{F}$ existiert ein Skalarprodukt, das *strikt* ist, d.h. $\langle \phi(\mathbf{x}), \phi(\mathbf{x}) \rangle = 0$ dann und nur dann, wenn $\phi(\mathbf{x}) = \vec{0}$,

2. Die Elemente von $\mathcal{F}$ sind *vollständig*, d.h. für jede Folge[77] von Elementen $\{\phi_n\}_{n \geq 1}$ gilt, dass sie gegen ein Element ϕ konvergiert, wenn

   $$\sup_{m>n} \|\phi_n - \phi_m\| \to 0$$

   für $n \to \infty$ gilt[78], und

3. Die Elemente von $\mathcal{F}$ sind *separierbar*, wenn es eine abzählbare Folge $\phi_1, \phi_2, \ldots$ gibt derart, dass für alle $\phi \in \mathcal{F}$ und $\epsilon > 0$ die Bedingung $\|\phi_i - \phi\| < \epsilon$ erfüllt ist

[77] $\{\phi_n\}_{n\geq 1}$ steht für $\phi_1, \phi_2, \phi_3, \ldots, \phi_n, \ldots$. Der Ausdruck sup steht für Supremum, das ist das kleinste Element einer Menge, das größer oder gleich als alle Elemente der Menge ist.

[78] Dies sind die Cauchy-Folgen, nach Augustin-Louis Cauchy (1789 - 1857), französischer Mathematiker.

Sind diese drei Bedingungen für eine Menge $\mathcal{F}$ erfüllt, so ist $\mathcal{F}$ ein *Hilbert-Raum*[79]. Ein Hilbert-Raum ist ein Vektorraum, für den ein Skalarprodukt definiert ist und der vollständig im Sinne von Punkt 2 ist; die in Abschhnitt 1.3 betrachteten Vektorräume sind Hilbert-Räume; die Definition von Hilbert-Räumen wird wichtig, wenn, wie im hier gegebenene Zusammenhang, Funktionen als Vektoren betrachtet werden.

Beispiel 4.3 Es sei X die Menge aller abzählbaren Folgen $\mathbf{x}$ von reellen Zahlen, $\mathbf{x} = (x_1, x_2, \ldots, x_n, \ldots)$, für die

$$\sum_{i=1}^{\infty} x_i^2 < \infty$$

gilt. Ist $\mathbf{y} = (y_1, y_2, \ldots, y_n, \ldots) \in X$, und gilt ebenfalls

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{\infty} x_i y_i < \infty,$$

so bilden die Folgen einen Hilbert-Raum. $\square$

Gegeben sei eine Funktion $\kappa : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$. Die Kernfunktion ist. Es gilt der

Satz 4.1 *Damit κ eine Kernfunktion ist, muß (i) $\kappa(\boldsymbol{x}, \boldsymbol{z}) = \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{z}) \rangle$ gelten, ϕ eine Feature-Abbildung, und (ii) sie muß positive semi-definit sein.*

Beweis:[80] Nach Definition 3.2, Seite 90, ist eine symmetrische (n, n)-Matrix M positiv-semidefinit,wenn für einen beliebigen Vektor $\mathbf{x} \in R^n$ die Bedingung $\mathbf{x}' M \mathbf{x} \geq 0$ erfüllt ist. Es sei $G = (G_{ij})$ eine $(\ell \times \ell)$-Matrix von Kernen, d.h. es gelte

$$G_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle, \quad i, j = 1, \ldots, \ell$$

Für einen beliebigen Vektor $\mathbf{v} \in \mathbb{R}^\ell$ gilt dann (vergl. Abschnitt 3.2.2, Abschnitt 'Quadratische Formen und Ellipsoide')

$$\begin{aligned} \mathbf{v}' G \mathbf{v} &= \sum_{i,j} v_i v_j G_{ij} \\ &= \left\langle \sum_{i=1}^{\ell} v_i \phi(\mathbf{x}_i), \sum_{j=1}^{\ell} v_j \phi(\mathbf{x}_j) \right\rangle \\ &= \left\| \sum_{i=1}^{\ell} v_i \phi(\mathbf{x}_i) \right\|^2 \geq 0 \end{aligned}$$

$\square$

[79] David Hilbert (1862 – 1943), deutscher Mathematiker

[80] Propos. 3.7, Shawe-Taylor et al., p. 57.

Dies ist nicht der Beweis aber man findet den Text auf p. 61 Shawe-Taylor. Es sei $\mathcal{F}$ der Feature-Raum: Diese Funktion wird bei Schölkopf-Smola auf Seite 32 gegebenen allderdings unter dem Label 'The reproducing kernel-map'. Also:

Schölkopf-Formulierung übernehmen!!! Mit RKHS-Definition, Entspricht (4.29)

$$\mathcal{F} = \left\{ \sum_{i=1}^{m} \alpha_i \kappa(x_i, \cdot) | x_i \in X, \alpha_i \in \mathbb{R} \right\} \tag{4.29}$$

Die Elemente von $\mathcal{F}$ bilden eine abgeschlossene Menge bezüglich der Multiplikation mit einem Skalar und bezüglich der Addition:

$$f, g \in \mathcal{F} \rightarrow (f+g)(x) = f(x) + g(x),$$

wovon man sich durch Einsetzen in $\sum_i \alpha_i \kappa(x_i, \cdot)$ überzeugt. Weiter kann ein Skalarprodukt definiert werden. Die Funktionen f und g seien durch

$$f(\mathbf{x}) = \sum_{i=1}^{m} \alpha_i \kappa(x_i, x), \quad g(\mathbf{x}) = \sum_{i=1}^{m} \beta_i \kappa(z_i, x).$$

definiert. Für das Skalarprodukt von $f, g \in \mathcal{F}$ erhält man

$$\langle f, g \rangle = \sum_i \sum_j \alpha_i \beta_j \kappa(x_i, z_j) = \sum_i \alpha_i g(x_i) + \sum_j \beta_j f(z_j). \tag{4.30}$$

Damit ist $\langle f, g \rangle$ eine reellwertige, symmetrische Funktion, die linear in f und g (d.h. sie ist *bilinear*), und es gilt

$$\langle f, f \rangle \geq 0, \quad \text{für alle}\, f \in \mathcal{F}$$

denn

$$\langle f, f \rangle = \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j \kappa(x_i, x_j) = \boldsymbol{\alpha}' K \boldsymbol{\alpha}$$

$\boldsymbol{\alpha} = (\alpha_1, \dots \alpha_m)' \geq 0$ (vergl. Gleichung (3.34), Seite 90).

Reproduzierende Eigenschaft eines Kerns: Es werde $g = \kappa(\mathbf{x}, \cdot)$ in Gleichung (4.30) gesetzt. Dann folgt

$$\langle f, \kappa(\mathbf{x}, \cdot) \rangle = \sum_{i=1}^{m} \alpha_i \kappa(x_i, x) = f(x), \tag{4.31}$$

das Skalarprodukt von f mit $\kappa(\mathbf{x}, \cdot)$ reproduziert den Wert von f für das Argument $\mathbf{x}$.

RKHS Wenn eine Funktion κ positiv semi-definit ist heißt der korrespondierende $\mathcal{F}_\kappa$ *Reproducing Kernel Hilbert Space* (RKHS). Wie gezeigt wurde, kann jeder Kern dazu verwendet werden, um einen Hilbert-Raum zu erzeugen. Dazu muß man nur zeigen, dass eine symmetrische Funktion $\kappa(\cdot,\cdot)$ die reproduzierende Eigenschaft in einem Hilbert-Raum $\mathcal{F}$ hat,

$$\langle\kappa(,\cdot\rangle, f(\cdot)\rangle_{\mathcal{F}} = f(x), \quad f \in \mathcal{F},$$

dann ist κ positiv-semidefinit:

$$\begin{aligned}
\sum_{i,j}^{m} \alpha_i\alpha_j\kappa(x_i,x_j) &= \sum_{i,j}^{m} \alpha_i\alpha_j\langle\kappa(x_i,\cdot)\kappa(x_j,\cdot)\rangle \\
&= \left\langle \sum_{i=1}^{m} \alpha_i\kappa(x_i,\cdot), \sum_{j=1}^{m} \alpha_j(\kappa(x_j,\cdot) \right\rangle \\
&= \left\| \sum_{i=1}^{m} \alpha_i\kappa(x_i,\cdot) \right\|_{\mathcal{F}}^2 \geq 0
\end{aligned}$$

Zur Vorbereitung desfolgenden Theorem sollen zwei Begriffe aus der Analysis eingeführt werden:

Definition 4.5 *Gegeben sei eine Teilmenge U es $\mathbb{R}^n$. $U \subset \mathbb{R}^n$.*
1. U heißt abgeschlossen*, wenn für jede Folge $\{a_n\}_{n\in\mathbb{N}}$, die gegen einen Punkt a konvergiert $a \in U$ gilt.*
2. U ist beschränkt*, wenn sie in einer Kugel $K \subset \mathbb{R}^n$ mit endlichem Radius liegt.*
3. U ist kompakt*, wenn U beschränkt und abgeschlossen ist.*[81]

Definition 4.6 *Eine Funktion f heißt* quadratisch integrierbar *über einer Menge X, wenn*

$$\int_X |f(x)|^2 dx < \infty \tag{4.32}$$

gilt (Schreibweise: $f \in L^2(X)$).

Die Eigenschaft (4.32) ist wichtig, wenn es darum geht, ob eine Funktion f Element eines Hilbert-Raums ist, denn dann kann eine Norm $\langle f, f\rangle$ bestimmt werden. Sind f und g Elemente eine Hilbert-Raums, so existiert das Skalarprodukt $\langle f, g\rangle$ und der Winkel θ zwischen f und g ist definiert (vergl. (1.45), Seite 20), damit ist auch die Rede von *orthogonalen Funktionen* erklärt.

Definition 4.7 *Es sei $\kappa(\cdot,\cdot)$ ein symmetrischer, stetiger und positiv-semideiniter Kern. Ein* linearer Operator *ist definiert durch*

$$[T_{\kappa_f}](s) = \int_\Omega \kappa(s,t)f(t)dt, \tag{4.33}$$

[81] Man findet äquivalente Definitionen von Kompaktheit, z.B. U ist kompakt, wenn jede Folge von Punkten in U eine Teilfolge enthält, die gegen einen Punkt in U konvergiert, vergl. Heuser (2001), p. 225, und Werner (2011), p. 523.

d.h. das Integral auf der rechten Seite definiert eine Funktion $g(x)$, die einerseits von κ abhängt und andererseits von der Funktion f.

f wird hier als Vektor aufgefasst und κ entspricht einer Matrix, und das Produkt einer Matrix mit einem Vektor kann ja als lineare Transformation des Vektors f gesehen werden, mit κ als Transformationsmatrix. $g(s) = [T_{\kappa_f}](s)$ entsteht dementsprechend durch eine lineare Transformation, und das Integral entspricht einer Summation, – also einer linearen Transformation[82] (4.33) bezieht sich explizit auf Funktionen als Vektoren.

Ein spezieller linearer Operator ist

Definition 4.8 *Es gelte*

$$\lambda_i e_i(t) = [T_\kappa e_i](t) = \int_\Omega \kappa(t,s) e_i(s) ds, \quad \lambda_i \neq 0 \tag{4.34}$$

Dann ist e_i eine Eigenfunktion von κ und λ_i der zugehörige Eigenwert.

Das folgende Theorem ergibt sich aus den vorangegangenen Betrachtungen:

Satz 4.2 *(***Mercers Theorem***): Es sei $\mathcal{X}$ ein Hilbertraum und κ sei eine stetige, symmetrische Funktion und für alle Funktionen $f \in L^2(\mathcal{X})$ gelte*

$$\int_{\mathcal{X}\times\mathcal{X}} \kappa(x,z) f(x) f(z) dx dz \geq 0. \tag{4.35}$$

Dann kann κ in eine gleichförmig konvergente[83] *Reihe auf $\mathcal{X} \times \mathcal{X}$ in Termen von Funktionen ϕ_i entwickelt werden:*

$$\kappa(x,z) = \sum_{i=1}^{\infty} \phi_i(x)\phi_i(z), \tag{4.36}$$

wobei die ϕ_i paarweise orthogonal sind, d.h. es gilt für zwei Funktionen ϕ_i und ϕ_j

$$\langle \phi_i, \phi_{i'} \rangle = \begin{cases} 1, & i = i' \\ 0, & i \neq i' \end{cases} .$$

Beweis: Hier nur Lit'angabe machen, weil der Beweis nur nachvollziehbar wird, wenn eine Reihe von Begriffen aus der Funktionalanalysis une Maßtheorie verstehbar wird. □

[82] Als allgemeine Definition eines linearen Operators findet man auch die Gleichung $L(ax + bx) = aL(x) + bL(y)$.

[83] Eine Folge $\{f_n\}_{n>0}$ konvergiert *gleichmäßig konvergent* gegen eine Funktion f, wenn $|f_n(x) - f(x)| < \varepsilon > 0$ für $n \to \infty$ und falls x aus dem Definitionsbereich von f gilt.

5 Funktionenräume und PCA

5.1 Einführung

Vielfach ist die Untersuchung zeitlicher oder räumlicher Verläufe von Interesse; hier werden nur zeitliche Verläufe betrachtet, viele der zur Analyse dieser Verläufe eingeführten Begriffsbildungen übertragen sich auf räumliche Verläufe. Man betrachtet *Funktionenräume* $\mathcal{F}$, d.h. Mengen von Funktionen, die über einem bestimmten Bereich D definiert sind. Linearkombinationen von Funktionen lassen sich analog zu den bisher betrachteten Linearkombinationen definieren: sind f, g Elemente eines Funktionenraums $\mathcal{F}$, so soll auch $\lambda f + \mu g \in \mathcal{F}$ gelten; ist außerdem für alle $f, g \in \mathcal{F}$ außerdem das Skalarprodukt zweier Funktionen erklärt[84]

$$\langle f, g\rangle = \int_D f(t)g(t)dt, \quad f, g \in \mathcal{F} \tag{5.1}$$

(d.h. existiert das Integral), so hat man über

$$\|f\|^2 = \int_D |f(t)|^2 dt < \infty \tag{5.2}$$

auch eine Norm $\|f\|$ erklärt und $\mathcal{F}$ ist ein Vektorraum. Es sei $\{\phi_1, \phi_2, \ldots\}$ eine Menge von Funktionen aus $\mathcal{F}$ derart, dass

$$f(t) = \sum_{i=1}^{\infty} a_i \phi_i(t) \tag{5.3}$$

gilt, wobei die a_i für f spezifische Koeffizienten sind. Die ϕ_i heißen dann *Basisfunktionen*. Werden zur Darstellung von f unendlich viele Basisfunktionen benötig, so heißt der Vektorraum *unendlich-dimensional*. Wie Basisvektoren sind Basisfunktionen linear unabhängig. Gilt für die Basisfunktionen ϕ_i, $i = 1, 2, 3, \ldots$

$$\langle \phi_i, \phi_j\rangle = \int_D \phi_i(t)\phi_j(t)dt = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \tag{5.4}$$

so heißen die ϕ_i *orthogonal*; δ_{ij} heißt *Kronecker-Delta*.

Beispiel 5.1 Fourier-Reihen[85] Eine bekannte Reihenentwicklung einer Funktion $f(t)$ ist die Fourier-Reihe

$$f(t) = a_0 + \sum_{n=1}^{\infty}(a_n \cos(nt) + b_n \sin(nt)), \tag{5.5}$$

[84] Die 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.

[85] Joseph Fourier, (1768 – 1830), französischer Mathematiker und Physiker

wobei die Koeffizienten a_n und b_n durch

$$a_0 = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(t)dt \tag{5.6}$$

$$a_n = \frac{1}{\pi}\int_{-\infty}^{\infty} f(t)\cos(nt)dt \tag{5.7}$$

$$b_n = \frac{1}{\pi}\int_{-\infty}^{\infty} f(t)\sin(nt)dt \tag{5.8}$$

gegeben sind. Die cos- und sin-Funktionen sind orthogonal, denn wie gezeigt werden kann gilt

$$\int_{-T/2}^{T/2}\cos(m\omega_0 t)\cos(n\omega_0 t)dt = \begin{cases} 0, & m \neq n \\ T/2, & m = n \neq 0 \end{cases} \tag{5.9}$$

$$\int_{-T/2}^{T/2}\sin(m\omega_0 t)\sin(n\omega_0 t)dt = \begin{cases} 0. & m \neq n \\ T/2, & m = n \neq 0 \end{cases} \tag{5.10}$$

$$\int_{-T/2}^{T/2}\sin(m\omega_0 t)\cos(n\omega_0 t)dt = 0, \quad \text{für alle } m \text{ und } n \tag{5.11}$$

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 24 und die folgende Erläuterung). □

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 $\exp(-(t-t_n)^2$, $n = 1, 2, \ldots$ verwendet werden (Gishtasby & O'Neill (1994), Calcaterra (2008)), oder Gabor-Funktionen:

$$f_n(t) = \exp\left(-\frac{(x-x_n)^2}{2\sigma^2}\right)\phi(k\omega_o t), \tag{5.12}$$

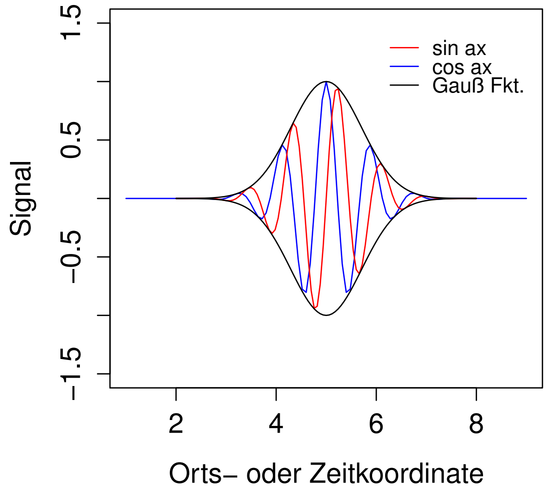
wobei ϕ eine Sinus- oder Kosinus-Funktion ist (Gabor (1948)), vergl. Abbildung 24. Weitere Basisfunktionen sind Polynome (Hermite- oder Laguerre-Polynome); die Diskussion der verschiedenen Möglichkeiten geht über den Rahmen dieses Skripts hinaus.

Unter Umständen lassen sich die Basisfunktionen auch empirisch bestimmen. Dies geschieht bei der Karhunen-Loéve-Analyse, auf die im Folgenden eingegangen werden soll.

5.2 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

Abbildung 24: Gabor-Funktion



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 repäsentieren.

Was jeweils gemeint ist wird durch den jeweiligen Zusammenhang bestimmt. Ein stochastischer Prozess ist dann eine Familie $X_t = \{X(t, \omega)\}_{t \in D}$ von zufälligen

Funktionen (Pfaden, Trajektorien) über dem Zeitbereich D. Um die Notation zu vereinfachen wird im Folgenden einfach $X(t)$ statt $X(t,\omega)$ geschrieben, und der Prozess wird kurz mit $X_t = \{X(t)\}_{t\in D}$ bezeichnet.

Erwartungswert und Varianz eines stochastischen Prozesses X_t sind durch

$$\mathbb{E}(X_t) = \int_\Omega X(t,\omega)dP(\omega) = m(t), \quad Var(X_t) = \mathbb{E}[(X_t - \mathbb{E}(X_t))^2] \tag{5.13}$$

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

$$X_t = \{X_t\}_{t\in D}, \quad X_t = Y_t - m(t) \tag{5.14}$$

zentrierter stochastischer Prozess.

Definition 5.1 *Der stochastische Prozess heißt* stetig im quadratischen Mittel*, wenn*

$$\mathbb{E}[(X_{t+\varepsilon} - X_t)^2] = 0. \tag{5.15}$$

Es sei X_t ein zentrierter stochastischer Prozess; dann ist

$$R_X(s,t) = \mathbb{E}[X(s)X(t)], \quad s,t \in D \tag{5.16}$$

die *Autokorrelationsfunktion* von X_t.

Satz 5.1 *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: z.B. Wong (1971). □

Der folgende Satz geht auf Loève (1945) und Karhunen (1947) zurück.

Satz 5.2 *Es sei $X_t = \{X(t) | t \in [a,b]\}$ mit $\mathbb{E}(X_t) = 0$ für alle $t \in [a,b]$ und stetiger Kovarianzfunktion $C(s,t)$, wobei die Funktionen $X(t)$ quadratintegrierbar seien. Dann gilt*

$$\hat{X}(t) = \sum_{i=1}^n a_i \phi_i(t), \tag{5.17}$$

und die

$$a_i = \int_a^b X(t)\phi_i(t) \tag{5.18}$$

sind zufällige Veränderliche[86] *mit*

$$\mathbb{E}(a_i) = 0, \quad \mathbb{E}(a_i a_j) = \delta_{ij}\lambda_i \tag{5.19}$$

[86] a_i ist zufällig weil $X(t)$ eine zufällige Funktion aus der Menge der zufälligen Funktionen ist, die den stochastischen Prozess X_t definieren.

mit

$$\delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} \qquad \textit{(Kronecker-Delta)}$$

sind; die ϕ_i sind die orthonormalen Eigenfunktionen von $C(s,t)$, mit den λ_i als dazu korrespondierenden Eigenwerten. Weiter gilt

$$\lim_{n\to\infty} \hat{X}(t) = X(t). \tag{5.20}$$

Beweis: Der Ansatz (5.17) gilt für eine willkürliche Funktion $f(t)$, also insbesondere für eine zufällige Funktion $X(t)$ aus der Menge X_t; da $X(t)$ zufällig ist, ist a_i eine zufällige Veränderliche. Es gilt

$$\mathbb{E}(a_i) = \mathbb{E}\left[\int_a^b X_t\phi_i(t)\right] = \int_a^b \mathbb{E}[X_t]\phi_i(t) = 0, \tag{5.21}$$

da ja $\mathbb{E}[X_t] = 0$. Nach Mercers Theorem gilt $C(s,t) = \sum_{i=1}^\infty \lambda_i\phi_i(s)\phi_i(t)$ und ϕ_i und ϕ_j sind orthonormal, $i \neq j$. Dann folgt

$$\begin{aligned}
\mathbb{E}[a_i a_j] &= \mathbb{E}\left[\int_a^b\int_a^b X_s X_t \phi_i(s)\phi_j(t)dsdt\right] \\
&= \int_a^b\int_a^b \mathbb{E}[X_s X_t]\phi_i(s)\phi_j(t)dsdt \\
&= \int_a^b\int_a^b k(s,t)\phi_i(s)\phi_j(t)dsdt \\
&= \int_a^b \phi_i(s)\left(\int_a^b k(s,t)\phi_j(t)dt\right)ds \\
&= \lambda_i \int_a^b \phi_i(s)\phi_j(s)ds \\
&= \delta_{ij}\lambda_i,
\end{aligned} \tag{5.22}$$

δ_{ij} das Kronecker-Delta. Es sei

$$\varepsilon_n(t) = \mathbb{E}\left[\left(X(t) - \sum_{i=1}^n a_i\phi_i(t)\right)^2\right]. \tag{5.23}$$

Dann folgt

$$\varepsilon_n(t) = \mathbb{E}[X^2(t)] - 2\mathbb{E}\left[X(t)\sum_{i=1}^n a_i\phi_i(t)\right] + \mathbb{E}\left[a_i a_j \phi_i(t)\phi_j(t)\right] \tag{5.24}$$

Es ist $\mathbb{E}(X^2(t)] = C(t,t)$, und

$$\begin{aligned}
\mathbb{E}\left[X(t)\sum_{i=1}^{n} a_i\phi_i(t)\right] &= \mathbb{E}\left[X(t)\sum_{i=1}^{n}\int_D X(s)\phi_i(s)ds\phi_i(t)\right] \\
&= \sum_{i=1}^{n}\left(\int_D \mathbb{E}[X(s)X(t)]\phi_i(s)ds\right)\phi_i(t) \\
&= \sum_{i=1}^{n}\left(\int_D C(s,t)\phi_i(s)ds\right)\phi_i(t) \\
&= \sum_{i=1}^{n}\lambda_i\phi_i^2(t),
\end{aligned}$$

und schließlich folgt analog

$$\mathbb{E}\left[\sum_{i=1}^{n} a_i a_j\phi_i(t)\phi_j(t)\right] = \sum_{i=1}^{n}\lambda_i\phi_i^2(t).$$

Also folgt

$$\varepsilon_n(t) = C(t,t) - \sum_{i=1}^{n}\lambda_i\phi_i(t)\phi_i(t),$$

und wegen Mercers Theorem folgt

$$\lim_{n\to\infty}\varepsilon_n(t) = 0$$

d.h. $\hat{X}(t) \to X(t)$. □

$X(t)$ **als Zeitreihe:** Erhebt man $X(t)$ als kontinuierliche Funktion über $D = [a,b]$, so muss man die Integralgleichung

$$\int_D k(s,t)\phi_i(t)dt = \lambda_i\phi_i(s)$$

für $i = 1, 2, \ldots, k$ lösen, d.h. die Eigenfunktionen $\phi_i(t)$ für alle $t \in [a,b]$ bestimmen. Das ist im Allgemeinen sehr aufwendig. Einfacher wird die Aufgabe, wenn $X(t)$ für diskrete Werte $t_1, \ldots, t_N$ bestimmt wird. Man erhält dann den Vektor

$$X = (X_1, \ldots, X_N),'$$

für den die (m,n)-Matrix C der Autokorrelationen berechnet werden kann. Man erhält man dann die Gleichung

$$C\vec{\phi}_j = \lambda_j\vec{\phi}_j, \quad j = 1, \ldots, N \tag{5.25}$$

$\vec{\phi}_j$ ist jetzt ein N-dimensionaler (Eigen-)Vektor mit λ_j als zugehörigem Eigenwert. $\vec{\phi}_j$ repräsentiert die entsprechende Eigenfunktion an den Stellen $t_1, \ldots, t_N$, und natürlich sind die Eigenvektoren $\vec{\phi}_j$ paarweise orthonormal. Gesucht ist die Entwicklung für $X = (X(t_1), X(t_2) \ldots, X(t_N))'$ als Reihe:

$$X = \sum_{j=1}^{N} a_j \phi_j(t),$$

mit $\phi_j = (\phi_j(t_1), \ldots, \phi(t_N))'$. Die Koeffizienten a_j erhält man durch orthogonale Basisentwicklung:

$$\phi_j' X = a_j, \quad j = 1, \ldots, N \tag{5.26}$$

Natürlich ist es von Interesse, weniger als N Eigenfunktionen zu wählen, – gesucht ist ja die sparsamste Repräsentation der Daten. Also betrachtet man

$$\hat{X} = \sum_{j=1}^{r} a_j \phi_j(t), \quad r < N \tag{5.27}$$

Die Güte der Abschätzung $\hat{X}$ läßt sich durch den Quotienten

$$\frac{\sum_{j=1}^{r} \lambda_j}{\sum_{j=1}^{N} \lambda_j} \geq \alpha \tag{5.28}$$

bewerten, wobei α ein festgesetzter Anteil an erklärter Varianz ist.

6 Anhang

6.1 Punkträume

Mit dem Symbol $\mathbb{R}$ wird die Menge der reellen Zahlen bezeichnet[87]. 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

$$\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n)|x_1, x_2, \ldots, x_n \in \mathbb{R}\}, \quad n \in \mathbb{N} \tag{6.1}$$

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

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

Anmerkung: Gelegentlich findet man für einen Punkt x auch die Schreibweise $x = (x_1|x_2|\cdots|x_n)$, um ihn vom Vektor $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$ 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 $x = (x_1, \ldots, x_n)$ und $y = (y_1, \ldots, y_n)$ irgendzwei Punkte, so soll
1. $ax = (ax_1, ax_2, \ldots, ax_n)$, für $a \in \mathbb{R}$, und
2. $x + y = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$
gelten. $x + y$ definiert also ebenfalls einen Punkt. Es gilt der folgende

Satz 6.1 *Es seien x, y, z Punkte im $\mathbb{R}^n$. Dann gelten die Aussagen*
1. $x + y = y + x$, (Kommutativität)
2. $(x + y) + z = x + (y + z)$, (Assoziativität der Summation)
3. $x + 0 = x$; 0 ist der neutrale Punkt *(der Ursprung des Koordinatensystems)*
4. Für jeden Punkt x existiert ein Punkt $-x$ derart, dass $x + (-x) = 0$,
5. $1x = x$,
6. $(ab)x = a(bx)$, für $a, b \in \mathbb{R}$,
7. $(a + b)x = ax + bx$, für $a, b \in \mathbb{R}$,

Beweis: Es genügt ein einfaches Nachrechnen. □

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

Definition 6.1 *Die Distanz $d(x, y)$ ist eine Funktion der Differenzen $x_i - y_i$, $i = 1, \ldots, n$ der Koordinaten x_i, y_i der Punkte x und y; genügen die Distanzen den folgenden Axiomen*
1. $d(x, y) \geq 0$,
2. $d(x, y) = d(y, x)$ (Reflexivität)
3. $d(x, z) \leq d(x, y) + d(y, z)$ (Dreiecksungleichung),
so definiert d eine Metrik *des Raumes, in dem die Punkte x und y liegen.*

Die Distanzen zwischen den Punkten eines Raumes können auf verschiedene Weisen definiert sein, und dementpsrechend ist ein Raum durch eine bestimmte Metrik definiert, die die geometrische Struktur des Raumes kennzeichnet. So ist nach Euklid ist die kürzeste Verbindung zwischen zwei Punkten eine Gerade. Für irgendzwei Punkte x und y gilt dann

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} = \left[\sum_{i=1}^{n}(x_i - y_i)^2\right]^{\frac{1}{2}}. \tag{6.2}$$

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 (6.2) ist dann der Satz des Pythagoras für den n-dimensionalen Raum. Ist die Metrik eines Raums durch (6.2) definiert, so heißt der Raum *euklidisch*. Die euklidische Metrik ist nicht die einzig mögliche Metrik, wie weiter unten illustriert wird.

Die Axiome in Definition 6.1 definieren die allgemeinen Eigenschaften einer Metrik. Die Forderung, dass eine Distanz $d(x, y)$ nicht negativ sein darf, entspricht dem umgangssprachlichen Begriff von "Distanz": die Länge eines Weges von Ort A zu Ort B kann nicht negativ sein. Die Forderung der Reflexivität $d(x, y) = d(y, x)$ stellt eine Einschränkung dar: will man in einer Stadt mit dem Auto von der Adresse A_1 zu Adresse A_2 fahren, so kann wegen eines Einbahnstrassensystems die Distanz $d(A_1, A_2)$ größer als die Distanz $d(A_2, A_1)$ sein. Die Dreiecksungleichung ist wiederum intuitiv einleuchtend. Sind A_1, A_2 und A_3 drei Adressen in einer Stadt, so ist es möglich, dass es zwischen A_1 und A_3 keinen direkten Weg gibt und man immer über die Adresse A_2 fahren muss; dann gilt eben $d(A_1, A_3) = d(A_1, A_2) + d(A_2, A_3)$. Es wird aber nie $d(A_1, A_3) > d(A_1, A_2) + d(A_2, A_3)$ gelten, d.h. es muss $d(A_1, A_3) \leq d(A_1, A_2) + d(A_2, A_3)$ gelten.

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

$$d(x, y) = \left[\sum_{i=1}^{n} |x_i - y_i|^p \right]^{\frac{1}{p}}, \quad 0 < p \in \mathbb{R} \tag{6.3}$$

definiert ist. Für $p = 2$ ergibt sich die euklidische Metrik (6.2). 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 2). Die Orientierung hängt von gewählten Koordinatensystem ab. Insbesondere lassen sich Abhängigkeiten zwischen Variablen leicht mittels des Vektorbegriffs darstellen und latente, also nicht direkt gemessene Variablen, zur Erklärung von Korrelationen zwischen Variablen bestimmen. Analog zum Punktraum kann dann der Begriff des Vektorraums eingeführt werden, der isomorph zum jeweiligen Punktraum ist. Je nach Perspektive macht man in der multivariaten Analyse sowohl vom Begriff des Punkt- wie des Vektorraums Gebrauch. In den folgenden Abschnitten wird der Begriff des Vektors und der des Vektorraums eingeführt und einige Resultate aus der Vektor- und Matrixrechung vorgestellt, soweit sie für die üblichen multivariaten Verfahren notwendig sind. Neuere Verfahren wie zum Beispiel die Klassifikation von Mustern oder Objekten anhand von "Support Vector Machines" erfordern mathematische Grundlagen, die in einem gesonderten Skript vorgestellt werden.

6.2 Allgemeine Vektorräume

vvvv

6.2.1 Definition eines Vektorraums

Die Begriffe des Vektorraums und des Teilvektorraums spielen u.a. bei der Bestimmung von "latenten", also nicht direkt gemessenen Variablen eine wichtige Rolle. In der multivariaten Statistik spielen n-dimensionale Vektoren mit $n < \infty$ eine zentrale Rolle, bei neueren Verfahren jedoch von wird auch von allgemeineren Definitionen des Vektorbegriffs Gebrauch gemacht. Der Anschaulichkeit halber wir zunächst eine auf n-dimensionale Vektoren beschränkte Definition eines Vektorraums gegeben. allgemallgem

Definition 6.2 *Ein* Vektorraum über $\mathbb{R}$ *ist eine nichtleere Menge* V*, mit zwei Operationen für beliebige* $\boldsymbol{x}, \boldsymbol{y} \in V$*,* $a \in \mathbb{R}$*:*
(1) $V \times V \to V$*,*
. $(\boldsymbol{x}, \boldsymbol{y}) \mapsto \boldsymbol{x} + \boldsymbol{y}$*, (Summenbildung), und*
(2) $\mathbb{R} \times V \to V$*,*
. $(a, \boldsymbol{x}) \mapsto a\boldsymbol{x}$*, (Multiplikation mit einem Skalar)*

Für alle $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$ *sollen die Operationen die folgenden Eigenschaften gelten (Tabelle 2)*[88]

Tabelle 2: Eigenschaften der Operationen in einem Vektorraum

(1)	Reihenfolge	$(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$
(2)	Kommutivität	$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}, \quad \mathbf{x}, \mathbf{y} \in V,$
(3)	Nullvektor $\vec{0}$	$\mathbf{x} + \vec{0} = \mathbf{x}$
(4)	inverses Element $-\mathbf{x}$	$\mathbf{x} + (-\mathbf{x}) = 0$
(5)	Assoziativität	$(ab)\mathbf{x} = a(b\mathbf{x}),\ a, b \in \mathbb{R},\ \mathbf{x} \in V$
(6)	Einselement	$1 \cdot \mathbf{x} = \mathbf{x},\ \mathbf{x} \in V$
(7)	Distribtutivität I	$a(\mathbf{x} + \mathbf{y}) = a\mathbf{x} + a\mathbf{y},\ a \in \mathbb{R},\ \mathbf{x}, \mathbf{y} \in V$
(8)	Distributivität II	$(a + b)\mathbf{x} = a\mathbf{x} + b\mathbf{x}, \quad a, b \in \mathbb{R},\ \mathbf{x} \in V$

Ist $U \subseteq V$ *eine Teilmenge von* V*, und gelten für die Elemente von* U *wieder die Aussagen (1) und (2) für Operationen mit den Eigenschaften (1) bis (8), so ist* U *ein Teilvektorraum von* V*. Die leere Menge* $\emptyset$ *ist stets Element des Vektorraums;*$\{\emptyset\}$ *und der Vektorraum* V *selbst heißen* triviale Teilräume *(Vektorräume, die aus nur einem Element bestehen).*

Der Grund für diese ausführlichere Definition 6.2 ist, dass eben nicht nur n-Tupel aus $\mathbb{R}^n$ als Vektoren betrachtet werden können. So läßt sich zeigen, dass die Menge der (m, n)-Matrizen ein Vektorraum ist, oder dass die Menge der stetigen Funktionen über einem Intervall $[a, b]$ ein Vektorraum ist. Die Definition 6.2 ist noch nicht die allgemeinste Definition eines Vektorraumes, weil Vektoren noch "über der Menge $\mathbb{R}$" angenommen werden. Die allgemeinste Definition erhält man, wenn man $\mathbb{R}$ durch eine als *Körper* definierte Menge K ersetzt; die Menge $\mathbb{R}$ ist der Speziallfall eines Körpers[89]. Eine erste Verallgemeinerung ergibt sich,

[88] Jänich (2002), p. 23

[89] Zunächst wird der Begriff des *Körpers* definiert: Ein Körper ist eine Menge K, für die zwei Arten der Verknüpfung von Elementen aus K erklärt sind: (i) die Addition ("+"), und (ii) die Multiplikation ("·"). Die Addition ist *assoziativ*, d.h. für drei Elemente $a, b, c \in K$ gilt $(a + b) + c = a + (b + c)$, (iv) sie ist *kommutativ*, d.h. für alle $a, b \in K$ gilt $a + b = b + a$, (v) es existiert ein *neutrales Element 0.* d.h. $a + 0 = a$, und (vi) zu jedem Elemente $a \in K$ existiert ein inverses Element $-a$ bezüglich der Addition, so dass $a + (-a) = 0$. Für die Multiplikation gilt (i) für die Elemente $a, b, c \in K$ gilt $(a \cdot b) \cdot c = a \cdot (b \cdot c)$, (ii) sie ist*kommutativ* in Bezug auf "·", d.h. für $a, b \in K$ gilt $a \cdot b = b \cdot a$, (iii) es existiert ein neutrales Element 1, so dass $a \cdot 1 = a$, (iv) für jedes Element $0 \neq a \in K$ existiert ein inverses Element a^{-1} derrart, dass $a^{-1} \cdot a = 1$, und (v) die Multiplikation ist distributiv bezüglich der Addition, d.h. für $a, b, c \in K$ gilt $a \cdot (b + c) = a \cdot b + a \cdot c$).

wenn statt der Menge $\mathbb{R}$ die Menge $\mathbb{C}$ der komplexen Zahlen $z = x + iy$, $x, y \in \mathbb{R}$ und $i = \sqrt{-1}$ zugrunde gelegt wird.

6.3 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 $\lambda_1, \ldots, \lambda_k$, so wird man die Dimensionen $L_{k+1}, \ldots, \lambda_n$ als nicht mehr relevante Dimensionen ansehen. Als Eigenwerte von $X'X$ werden die λ_k von den Maßeinheiten der Variablen $V_1, \ldots, V_n$ 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 Skla transformieren, wenn etwa an einigen Orten in Fahrenheit, an anderen Orten in Celsius gemessen wurde. Anders ist es, wenn die Variablen verschiedene Merkmale wir Körperlänge und Körpergewicht repräsentieren. Für eine gegebene Längenskala gibt es keine Gewichtsskala, die der Längenskala auf "natürliche" Weise entpräche: 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: $x - \mathbf{x}$ und s_x haben stets dieselbe Maßeinheit, die sich im Audruck für z herauskürzt, d.h. z ist an keine Maßeinheit gekkoppelt. 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

$$P(|x - \mu| \geq k) \leq \frac{\sigma^2}{k^2}, \quad k > 0 \tag{6.4}$$

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

$$P(|x - \mu| > z\sigma) \leq \frac{\sigma^2}{z^2\sigma^2} = \frac{1}{z^2} \tag{6.5}$$

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 gesammte 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 25 illustriert den Effekt verschiedener Maßeinheiten auf die Form der Punktekonfiguration. Betrachtet werden Konfigurationen von Fällen, die als Stichproben einer 2-dimensionalen Gaußverteilung mit den Erwartungswerken $\mu_1 = \mu_2 = 0$ und den Varianzen $\sigma_1^2 = \sigma_2^2 = 1$ gewonnen wurden (Package mvrnorm[90]. Der Effekt unterschiedlicher Skalan wurde erzeugt durch Multiplikation der x- bzw. y-Werte mit einem Faktor fx bzw. fy. Die Tabelle 3 enthält die Eigenwerte für die verschiedenen Konfigurationen.

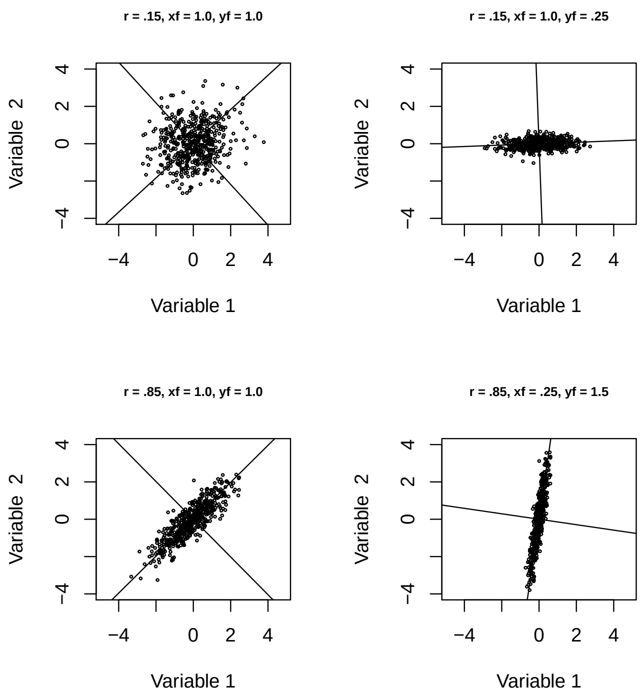
Tabelle 3: Eigenwerte für verschiedene fx, fy-Faktoren; λ_1, λ_2 Eigenwerte der Kovarianzmatrix

$r = .15$	$r = .15$	$r = .85$	$r = .85$
$fx = fy = 1$	$fx = 1,\ fy = .25$	$xf = fy = 1$	$fy = .25,\ yf = 1.5$
$\lambda_1 = 1.140$	$\lambda_1 = .980$	$\lambda_1 = 1.905$	$\lambda_1 = 2.330$
$\lambda_2 = .861$	$\lambda_2 = .063$	$\lambda_2 = .142$	$\lambda_2 = .016$
$\lambda_1/\lambda_2 = 1.323$	$\lambda_1/\lambda_2 = 15.660$	$\lambda_1/\lambda_2 = 13.443$	$\lambda_1/\lambda_2 = 142.998$

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. Am Ort A werde die Temperatur auf einer Celsius-Skala, an Ort B auf einer Fahrenheit-Skala gemessen. Die beiden Skalen sind durch eine affine Beziehung $F = aC + b$ aufeinander bezogen. Bekanntlich gilt dann $Var(F) = a^2 Var(C)$, mit $a = 1.8$, $b = 32$, d.h. $Var(F) = 3.24 Var(C)$, oder $Var(C) = .31 Var(F)$

[90] R 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 25: Effekt verschiedener Skalen



6.4 Zur geometrischen Definition des Skalarprodukts

In Gleichung (1.49) 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

$$\mathbf{x}(\mathbf{y} + \mathbf{z}) = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z} \tag{6.6}$$

gelten, wobei $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ sind, vergl. Abbildung 26. Nach (1.73)), Seite 30 gilt

$$\|\mathbf{x}_{\mathbf{y}}\| = |\mathbf{x}'\mathbf{y}|\|\mathbf{y}\| = |\mathbf{x}'\mathbf{y}|,$$

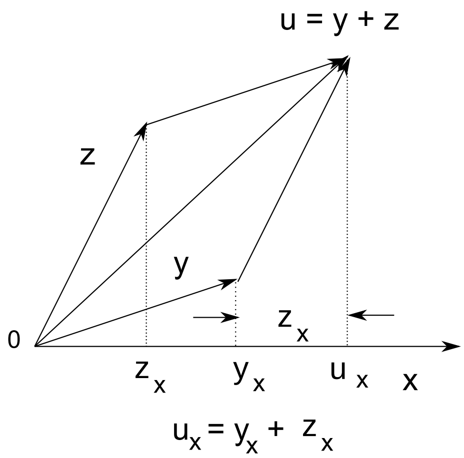
d.h. $\mathbf{x}_{\mathbf{y}}$ ist gleich dem Skalarprodukt $\mathbf{x}_{\mathbf{y}}$. Insbesondere werde die Projektion des Vektors $\mathbf{y} + \mathbf{z}$ auf den Vektor $\mathbf{x}$ (Abb. 26) betrachtet. Wie die Abbildung zeigt, gilt

$$(\mathbf{y} + \mathbf{z})_x = \mathbf{x}_x + \mathbf{z}_x, \tag{6.7}$$

d.h. aber

$$\mathbf{x}(\mathbf{y} + \mathbf{z}) = \mathbf{x}(\mathbf{y} + \mathbf{z})_x = \mathbf{x}\mathbf{y}_x + \mathbf{x}\mathbf{z}_x = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z},$$

Abbildung 26: Distributivgesetz für die geometrische Definition des Skalarprodukts



d.h. das Distributivgesetz gilt auch für die geometrische Definition des Skalarprodukts.

6.5 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 6.2 *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 6.3 *Die*[91] *(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.

[91] Im Folgenden wird die in https://de.wikipedia.org/wiki/Elementarmatrix vorgestellte Notation übernommen.

Definition 6.4 *Es sei I_n die (n,n)-Einheitsmatrix. Eine* Elementarmatrix *entsteht, wenn auf I_n eine der möglichen elementaren Umformungen angewendet wird; dabei entstehen drei Typen von Elementarmatrizen:*
1. $I_n \to Q_i^j(\lambda)$: das (i,j)-te Element von I_n wird durch $\lambda \in \mathbb{R}$ ersetzt, d.h.[92]

$$Q_i^j(\lambda) = I_n + \lambda E_{ij} \tag{6.8}$$

2. $I_n \to P_i^j$: die i-te Zeile von I_n wir mit der j-ten Zeile vertauscht, d.h.

$$P_i^j = I_n - E_{ii} - E_{jj} + E_{ij} + E_{ji}, \quad i \neq j. \tag{6.9}$$

Die Subtraktion von E_{ii} und E_{jj} bewirkt die Ersetzung der Einsen an der i-ten und j-ten Diagonalzelle von I_n, und die Addition von E_{ij} und E_{ji} bewirkt die Ersetzung der Null durch eine Eins an der (i,j)-ten und der (j,i)-ten Position von I_n.
3. $I_n \to S_i(\lambda)$: die Eins in der i-ten Diagonalzelle von I_n wird durch $1 \neq \lambda \in \mathbb{R}$ ersetzt:

$$S_i(\lambda) = I_n + \lambda E_{ii} - E_{ii} = I_n + (\lambda - 1)E_{ii}. \tag{6.10}$$

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} entsehe aus A durch Addition des λ-fachen der j-ten Zeile zur i-ten Zeile. Dann gilt

$$A_{III} = Q_j^i(\lambda)A.$$

3. A_{IV} entstehe aus A durch Vertauschen der i-ten mit der j-ten Zeile. Dann gilt

$$A_{IV} = P_i^j A.$$

Will man Spalten- statt Zeilenumformungen vornehmen, muss A statt von links mit einer Elementarmatrix von rechts multipliziert werden.

Die Elementarmatrizen sind invertierbar und ihre Inversen sind wieder Elementarmatrizen, d.h. es gelten die Gleichungen

$$(S_i(\lambda))^{-1} = S_i(\frac{1}{\lambda}), \quad (Q_i^j)^{-1} = Q_i^j(-1) \tag{6.11}$$

$$(Q_i^j(\lambda))^{-1} = Q_i^j(-\lambda), \quad (Pi^j)^{-1} = P_i^j \tag{6.12}$$

[92]Die Bezeichnung der Elementarmatrizen entspricht der in Fischer (1997), p.155 gewählten.

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:

$$B = \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & b_{nn} \end{pmatrix} \tag{6.13}$$

Da die elementaren Umformungen durch Elementarmatrizen repräsentiert werden, heißt dies, dass Elementarmatrizen $B_1, \ldots, B_r$ existieren derart, dass

$$B = B_r B_{r-1} \cdots B_1 A. \tag{6.14}$$

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.

6.5.1 Lineare Gleichungen und Gauß-Algorithmus

Gegeben sei ein lineares Gleichungssystem $A\mathbf{x} = \mathbf{y}$; gesucht ist eine möglichst systematische Art und Weise, einen Lösungsvektor $\mathbf{x}$ zu bestimmen. In diesem Abschnitt soll das gaußsche Eliminationsverfahren (auch Gauß-Algorithmus vorgestellt werden. Die Idee des Verfahrens ist, die Koeffizientenmatirx A in eine Dreiecksmatrix zu transformieren, anhand der serh schnell eine Lösung gefunden werden kann, sofern eine Lösung existiert. Zur Illustration werde ein System mit drei Unbekannten betrachtet:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= y_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= y_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= y_3 \end{aligned} \tag{6.15}$$

Bei der Bestimmung der Dreiecksmatrix mit den Elementen $\tilde{a}_{ij}$ wird auch der Vektor $\mathbf{y}$ in einen Vektor $\tilde{\mathbf{y}}$ transformiert:

$$\begin{aligned} \tilde{a}_{11}x_1 + \tilde{a}_{12}x_2 + \tilde{a}_{13}x_3 &= \tilde{y}_1 \\ \tilde{a}_{22}x_2 + \tilde{a}_{23}x_3 &= \tilde{y}_2 \\ \tilde{a}_{33}x_3 &= \tilde{y}_3 \end{aligned} \tag{6.16}$$

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 (6.16) ist eine Dreiecksmatrix

$$\tilde{A} = \begin{pmatrix} \tilde{a}_{11} & \tilde{a}_{12} & \cdots & \tilde{a}_{1n} \\ 0 & \tilde{a}_{22} & \cdots & \tilde{a}_{2n} \\ & & \ddots & \\ 0 & 0 & \cdots & \tilde{a}_{nn} \end{pmatrix}. \tag{6.17}$$

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

$$|\tilde{A}| = \begin{vmatrix} \tilde{a}_{11} - \lambda & \tilde{a}_{12} & \cdots & \tilde{a}_{1n} \\ 0 & \tilde{a}_{22} - \lambda & \cdots & \tilde{a}_{2n} \\ & & \ddots & \\ 0 & 0 & \cdots & \tilde{a}_{nn} - \lambda \end{vmatrix} = (\tilde{a}s_{11} - \lambda)(\tilde{a}_{22} - \lambda) \cdots (\tilde{a}_{nn} - \lambda) = 0. \tag{6.18}$$

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

$$\left(\begin{array}{rrr|r} 2 & -1 & 3 & 1 \\ 3 & 1 & -2 & 0 \\ 1 & 1 & 1 & 3 \end{array}\right)$$

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

$$\left(\begin{array}{rrr|r} 2 & -1 & 3 & 1 \\ 0 & 2 & 5 & 9 \\ 0 & 3 & 1 & 5 \end{array}\right)$$

Ersetzt man die Gleichung (6.19) durch 3(6.19) - 2(6.19), 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

$$\left(\begin{array}{rrr|r} 2 & -1 & 3 & 1 \\ 0 & 2 & 5 & 9 \\ 0 & 0 & 17 & 17 \end{array}\right) \tag{6.19}$$

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:

$$\begin{pmatrix} 2 & -1 & 3 \\ 3 & 1 & -2 \\ 1 & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 2 & -1 & 3 \\ 0 & 2 & 5 \\ 0 & 0 & 17 \end{pmatrix}$$

und sieht die Transformation von A in eine Dreiecksmatrix direkt. Die hier vorgeführte Transformation hat einen gewissen ad hoc-Charakter und mag insofern nicht zufriedenstellend erscheinen, und natürlich existieren kanonische Darstellungen des Algorithmus (Golub & van Loan (2013), Kapitel 3), auf die hier aber nicht eingegangen werden soll bzw. kann, da diese Darstellungen auf der einen Seite recht lang sind und auf der anderen nicht benötigt werden, wenn man nur an der Anwendung interessiert ist, was im Allgemeinen der Fall sein wird; Statistikpakete enthalten entsprechende Module. Hier soll nur festgestellt werden, dass A und $\tilde{A}$ natürlich denselben Rang haben.

Im folgenden Abschnitt wird anhand der SVD von A der Teilraum $\mathbf{L}_{n-r}$ näher bestimmt.

6.5.2 Herleitung einer Rotationsmatrix

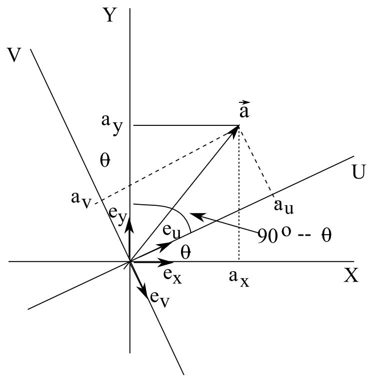
Es sei (X, Y) ein Koordinatensystem für die Fälle, wobei X und Y gemessene Variablen seien. Sei

$$\mathbf{a} = \begin{pmatrix} a_x \\ a_y \end{pmatrix} = \begin{pmatrix} a_u \\ a_v \end{pmatrix},$$

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 $(X.Y)$ 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

Abbildung 27: Vektorrotation, $\measuredangle(X,U) = \theta$



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

$$\begin{aligned} \mathbf{a} &= a_x\mathbf{e}_x + a_y\mathbf{e}_y, \quad (X,Y) - \text{System} && (6.20) \\ &= a_u\mathbf{e}_u + a_v\mathbf{e}_v, \quad (U,V) - \text{System} && (6.21) \end{aligned}$$

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 Skalarprodukte[93]

$$\begin{aligned} \mathbf{e}_x'\mathbf{e}_u &= (1,0)\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = u_1 = \cos\theta && (6.22) \\ \mathbf{e}_y'\mathbf{e}_u &= (0,1)\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = u_2 = \cos(90^0 - \theta) = \sin\theta && (6.23) \\ \mathbf{e}_x'\mathbf{e}_v &= (1,0)\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = v_1 = \cos(90^0 + \theta) = -\sin\theta && (6.24) \\ \mathbf{e}_y'\mathbf{e}_v &= (0,1)\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = v_2 = \cos\theta && (6.25) \end{aligned}$$

In (6.21) eingesetzt ergibt sich das Gleichungssystem in den zwei Unbekannten a_u und a_v

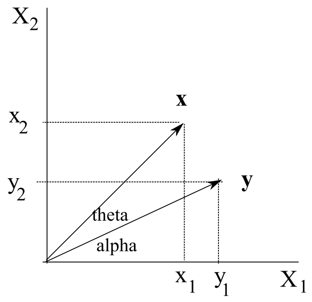
$$\begin{aligned} a_x &= a_u\cos\theta - a_v\sin\theta && (6.26) \\ a_y &= a_u\sin\theta + a_v\cos\theta && (6.27) \end{aligned}$$

In Matrixform hat das System die Form

$$\mathbf{a} = \underbrace{\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}}_{T'} \begin{pmatrix} a_u \\ a_v \end{pmatrix} = T'\begin{pmatrix} a_u \\ a_v \end{pmatrix} \qquad (6.28)$$

[93] Es wird von der bekannten Identität $\cos(\alpha\pm\beta) = \cos\alpha\cos\beta \mp \sin\alpha\sin\beta$ Gebrauch gemacht; 90^o entspricht $\pi/2$ im Radians-System.

Abbildung 28: Rotation eines Vektors (Alibi-Transformation)



T' ist orthonormal, wie man durch Nachrechnen bestätigt, so dass man die Lösung

$$T\mathbf{a} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} a_x \\ a_y \end{pmatrix} = \begin{pmatrix} a_u \\ a_v \end{pmatrix} \tag{6.29}$$

erhält.

Es 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. 28)

$$x_1 = \frac{\cos(\alpha+\theta)}{r}, \quad x_2 = \frac{\sin(\alpha+\theta)}{r}, \quad y_1 = \frac{\cos\alpha}{r,}, \quad y_2 = \frac{\sin\alpha}{r}$$

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.

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \tag{6.30}$$

d.h.

$$\mathbf{x} = T'\mathbf{y} \tag{6.31}$$

und T wurde in (6.29) definiert.

6.5.3 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-Zerlegung[94]. 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 6.3 *Es sei A eine symmetrische, positiv-definite Matrix. Dann existiert eine untere Dreiecksmatrix L derart, dass*

$$A = LL'. \tag{6.32}$$

Beweis: Statt eines allgemeinen Beweises wird der Satz anhand einer (3×3)-Matrix illustriert; das angewendetete Prinzip übertzrägt 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

$$A = LL' = \begin{pmatrix} L_{11}^2 & L_{11}L_{21} & L_{11}L_{31} \\ L_{21}L_{11} & L_{21}^2 + L_{22}^2 & L_{21}L_{31} + L_{22}L_{22}L_{32} \\ L_{31}L_{11} & L_{31}L_{21} + L_{32}L_{22} & L_{31}^2 + L_{32}^2 + L_{33}^2 \end{pmatrix}.$$

Das Element a_{ij} von A ist dann gleich dem Element in der i-ten Zeile und j-ten Spalte der Matrix auf der rechten Seite, also

$$a_{11} = L_{11}^2, \quad a_{22} = L_{21}^2 + L_{22}^2, \quad a_{33} = L_{31}^2 + L_{32}^2 + L_{33}^2,$$

und

$$a_{21} = L_{21}L_{11}, \quad a_{23} = L_{21}L_{31} + L_{22}L_{22}L_{32}, \; a_{31} = L_{31}L_{11}$$

und wegen der vorausgesetzten Symmetrie von A gilt $a_{ij} = a_{ji}$, so dass nicht mehr Elemente bestimmt werden müssen. Wegen der vorausgesetzten Positiv-Definitheit gilt $a_{11} > 0$, also folgt $L_{11} = \sqrt{a_{11}}$. Dann folgt $L_{21} = a_{21}/L_{11} = a_{21}/\sqrt{a_{11}}$ und $L_{31} = a_{31}/\sqrt{a_{11}}$, etc. Auf diese Weise lassen sich die L_{ij} aus den

[94] André-Louis Cholesky (1875 – 1918), französischer Mathematiker

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:

$$A = LDL', \tag{6.33}$$

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

$$A = GG'. \tag{6.34}$$

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 zulösen. Man findet für die Elemente g_{ik} von G

$$g_{ik} = \begin{cases} 0, & i < k \\ \sqrt{a_{ik} - \sum_{j=1}^{k-1} g_{kj}^2}, & i = k \\ \frac{1}{a_{ik}}\left(a_{ik} - \sum_{j=1}^{k-1} g_{ij}g_{kj}\right), & i > k \end{cases} \tag{6.35}$$

Eine ausführliche Darstellung von Lösungen von linearen Gleichungssytemen durch Rückführung auf Dreiecksmatrizen wird in Golub & vanLoan (2013), Kap. 3 und 4, gegeben.

6.6 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. $P = P(t)$.

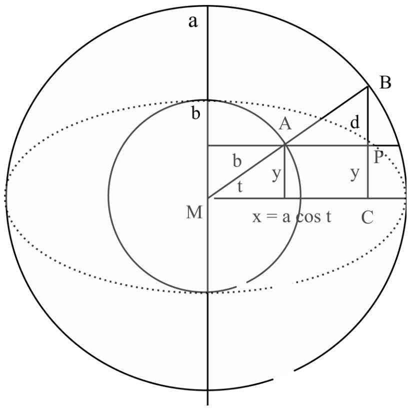
Behauptung: Die Punkte $P(t)$, $0 \le t \le 2\pi$, liegen auf einer Ellipse $\mathcal{E}$ der Form

$$\mathcal{E}\colon\ \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad x = x(t),\ y = y(t) \tag{6.36}$$

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

$$\frac{y}{b} = \frac{\sqrt{a^2 - x^2}}{a},$$

Abbildung 29: Kreise mit den Radien a bzw. $b < a$ und zugehörige Ellipse (Menge der Punkte P) mit den Halbachsen a und b



so dass

$$\frac{y^2}{b^2} = \frac{a^2 - x^2}{a^2} = 1 - \frac{x^2}{a^2},$$

woraus sofort (6.36) folgt. □

Parameterform: Bei dieser Darstellung wird die Ellipse (6.36) als Menge der Punkte $P = P(t)$ für $0 \leq t \leq 2\pi$ definiert. Aus der Abbildung 29 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 $\mathcal{E}$

$$\mathcal{E} = \{\mathbf{p}(t) | \mathbf{p}(t) = (x(t), y(t))' = (a \cos t, b \sin t)\}. \tag{6.37}$$

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

Die Gleichung (6.36) 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 und $x = 0$). Gesucht sind Werte für a und b derart, dass die erste Hauptachse der Ellipse der maximalen Ausdehnung der Punktekonfiguration entspricht; die k-te Spalte $\mathbf{L}_k$ hat die Länge $\sqrt{\lambda_k}$). Diese Forderung führt zu

$$a = \sqrt{\lambda_1}, \quad b = \sqrt{\lambda_2}. \tag{6.38}$$

In Bezug auf (6.37) 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 (6.36) äquivalent zu

$$\mathbf{y}' \Lambda^{-1} \mathbf{y} = 1. \tag{6.39}$$

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

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

6.7 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 $\Lambda = \text{diag}(\lambda_1, \ldots, \mathbf{v}_n)$ 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. $\mathcal{L}(V) = \mathcal{L}(\mathbf{v}_1, \ldots, \mathbf{v}_n) = \mathbb{R}^n$, $\mathcal{L}$ die lineare Hülle der Eigenvektoren. Sollte es eine zweite Matrix W von Eigenvektoren $\mathbf{w}_1, \ldots, \mathbf{w}_n$ geben, so können die $\mathbf{w}_k$ als Linearkombinationen der $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 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 $T_a = \text{diag}(a, \ldots, a)$, 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 6.4 *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, $\lambda_1 = \cdots = \lambda_n = \lambda$ 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

$$MV = V\Lambda_v \tag{6.40}$$

$$MW = W\Lambda_w \tag{6.41}$$

gelten[95]. Aus (6.40) folgt $M = V\Lambda_V V'$, und die Multiplikation von (6.41) von links mit V' liefert

$$V'MW = V'V\Lambda_v V'W = V'W\Lambda_w$$

d.h.

$$\Lambda_v V'W = V'W\Lambda_w. \tag{6.42}$$

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 $(a_{i1}, \ldots, a_{in})$ der i-te Zeilenvektor von A; (6.42) bedeutet dann

$$(\lambda_i a_{i1}, \lambda_i a_{i2} \ldots, \lambda_i a_{in}) = (\mu_1 a_{i1}, \mu_2 a_{i2}, \ldots, \mu_n a_{in}).$$

Zwei Vektoren sind identisch, wenn die korrespondierenden Komponenten identisch sind. Betrachtet man also die Gleichungen $\lambda_i a_{i1} = \mu_1 a_{i1}$, $\lambda_i a_{i2} = \mu_2 a_{i2}$, $\ldots$ $\lambda_i a_{in} = \mu_n a_{in}$, so sieht man, dass sich die Komponenten $a_{i1}, a_{i2}, \ldots$ herauskürzen, – insgesamt (also für $i = 1, \ldots, m$) hat man man die Gleichungen

$$\begin{array}{cccc}
\lambda_1 = \mu_1, & \lambda_1 = \mu_2, & \cdots & \lambda_1 = \mu_n \\
\lambda_2 = \mu_1, & \lambda_2 = \mu_2, & \cdots & \lambda_2 = \mu_n \\
\vdots & \vdots & \cdots & \vdots \\
\lambda_n = \mu_1, & \lambda_n = \mu_2, & \cdots & \lambda_n = \mu_n
\end{array}$$

und damit

$$\lambda_1 = \cdots = \lambda_n = \lambda, \quad \mu_1 = \ldots = \mu_n = \mu, \quad \lambda = \mu, \tag{6.43}$$

$V \neq W = VT$, $T \neq T_a$ 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: $\neg(V \neq W) \equiv (V = W)$[96]. Dies aber heißt, dass Eigenvektoren von M bis auf eine Längentransformation eindeutig bestimmt sind.

(ii) Identische Eigenwerte Nun werde der Fall $\lambda_1 = \cdots = \lambda_n = \lambda$ betrachtet. Es sei V eine orthonormale Matrix derart, dass $MV = V\Lambda$, $\Lambda = \text{diag}(\lambda, \ldots, \lambda)$. 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

$$\mathbf{w}_k = V\mathbf{t}_k = t_{1k}\mathbf{v}_1 + t_{2k}\mathbf{v}_2 + \cdots + t_{nk}\mathbf{v}_n,$$

[95] Diese Bedingung schließt auch die triviale Transformation $T = I$, I die Identitätsmatrix (d.h. $a = 1$) aus!

[96] $\neg$ ist das Negationszeichen $\neg(V \neq W)$ heißt, die Aussage $V \neq W$ gilt nicht.

$t_{1k}, \ldots, t_{nk}$ die Komponenten des k-ten Spaltenvektors $\mathbf{t}_k$ von T und $\mathbf{v}_1, \ldots, \mathbf{v}_n$ die Spaltenvektoren von V. Zu zeigen ist, dass $\mathbf{w}_k$ ein Eigenvektor von M ist, dass also $M\mathbf{w}_k = \lambda\mathbf{w}_k$ gilt. Es ist

$$\begin{aligned} M\mathbf{w}_k = MT\mathbf{v}_k &= M(t_{1k}\mathbf{v}_1 + \cdots + t_{nk}\mathbf{v}_n) \\ &= t_{1k}M\mathbf{v}_1 + \cdots + t_{nk}M\mathbf{v}_n \\ &= t_{1k}\lambda\mathbf{v}_1 + \cdots + t_{nk}\lambda\mathbf{v}_n \\ &= \lambda(t_{1k}\mathbf{v}_1 + \cdots + t_{nk}\mathbf{v}_n) \\ &= \lambda\mathbf{w}_k \end{aligned} \tag{6.44}$$

Damit ist gezeigt, dass die Linearkombination der Eigenvektoren $\mathbf{v}_1, \ldots, \mathbf{v}_n$ ebenfalls ein Eigenvektor ist, der zum Eigenwert λ korrespondiert. Da T beliebig gewählt werden kann folgt, dass es beliebig viele Eigenvektormatrizen außer V gibt. □

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 (6.44) 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.

□

Auf den Fall identischer Eigenwerte wird in Abschnitt 3.3.2 zurückgekommen.

6.8 Alternative Herleitung der Singularwertzerlegung

Die folgende Herleitung entspricht der von Shawe-Taylor & Christianini (2004), p.141.

Es sei also X eine (m, n)-Matrix vom Rang $\mathrm{rg}(X) \leq \min(m, n)$, und

$$X'X = V\Lambda_n V', \quad XX' = Q\Lambda_m Q'. \tag{6.45}$$

V die (n, n)-Matrix der Eigenvektoren von $X'X$, Λ_n die (n, n)-Diagonalmatrix der Eigenwerte von $X'X$. Ist $r < n$, so sind $n - r$ Eigenwerte gleich Null, d.h. Λ_n $n - r$ Zeilenvektoren und ebensoviele Spaltenvektoren, deren Komponenten alle gleich Null sind. Q ist die (m, m)-Matrix der Eigenvektoren von XX', und Λ_m ist eine (m, m)-Diagonalmatrix mit den Eigenwerten von XX'. Da $\mathrm{rg}(X) = \mathrm{rg}(X'X) = \mathrm{rg}(XX') = r \leq \min(m, n)$ haben Λ_m und Λ_n gleich viele von Null verschiedene Eigenwerte.

Die Matrix $X'X$ werde von rechts mit $X'\mathbf{q}$ multipliziert, wobei $\mathbf{q}$ ein Eigenvektor von XX' mit zugehörigem Eigenwert λ ist; dann ist

$$X'XX'\mathbf{q} = X'(\lambda\mathbf{q}) = \lambda(X'\mathbf{q}),$$

denn $XX'\mathbf{q} = \lambda\mathbf{q}$ nach Voraussetzung. Das heißt aber $(X'X)(X'\mathbf{q}) = \lambda(X'\mathbf{q})$, also ist $X'\mathbf{q}$ ein Eigenvektor von $X'X$, der allerdings nicht notwendig die Länge 1 hat. $X'\mathbf{q}$ wird normalisiert, indem man $X'\mathbf{q}$ mit $1/\|X'\mathbf{q}\|$ multipliziert, wobei

$$\|X'\mathbf{q}\|^2 = \mathbf{q}'XX'\mathbf{q}' = \lambda,$$

(aus $XX'\mathbf{q} = \lambda\mathbf{q}$ folgt $\mathbf{q}'XX'\mathbf{q} = \lambda\mathbf{q}'\mathbf{q} = \lambda$), d.h.

$$\frac{1}{\|X'\mathbf{q}\|} = \lambda^{-1/2}.$$

λ ist der zu $\mathbf{q}$ gehörige Eigenwert, wie man (6.45) entnehmen kann. Da $X'\mathbf{q}$ ein Eigenvektor von $X'X$ ist, existiert also ein zu $\mathbf{q}$ korrespondierender Eigenvektor $\mathbf{v}$ von $X'X$, für den

$$\mathbf{v} = \lambda^{-1/2}X'\mathbf{q} = X'(\lambda^{-1/2}\mathbf{q}) \tag{6.46}$$

gilt. Es sei $\mathbf{v}$ ein Eigenvektor von $X'X$; in analoger Weise findet man, wenn man XX' von rechts mit $X\mathbf{v}$ multipliziert, $XX'(X\mathbf{v}) = \lambda X\mathbf{v}$, d.h. $X\mathbf{v}$ ist ein nichtnormierter Eigenvektor von XX', es existiert also ein zu $X\mathbf{v}$ korrespondierender Eigenvektor $\mathbf{q}$ von XX' derart, dass

$$\mathbf{q} = \lambda^{-1/2}X\mathbf{v} = X(\lambda^{-1/2}\mathbf{v}) \tag{6.47}$$

wobei jetzt $\lambda^{-1/2} = 1/\|X\mathbf{v}\|$. Es folgt, dass λ ein sowohl zu $\mathbf{q}$ wie auch zu $\mathbf{v}$ ist, d.h. die Matrizzen $X'X$ und XX' haben identische von Null verschiedene Eigenwerte. Hat X den Rang $r \leq \min(m,n)$, so sind genau r Eigenwerte von Null verschieden, Fasst man die $\mathbf{v}_k$ in (6.46) zu einer (m,r)-Matrix V_r zusammen, die $\mathbf{q}_k$ zu einer (m,r)-Matrix Q_r und die λ_k zu einer (r,r)-Diagonalmatrix Λ_r, so ergeben sich die Beziehungen

$$V_r = X'Q_r\Lambda_r^{-1/2}, \tag{6.48}$$
$$Q_r = XV_r\Lambda_r^{-1/2} \tag{6.49}$$

ergeben (eine Längenskalierung bedeutet die Postmultiplikation mit einer Diagonalmmatrix, vergl. Gleichung 2.19, Seite 65). Löst man die Gleichungen nach X auf, so ergibt sich

$$X = Q_r\Lambda_r^{1/2}V_r' \tag{6.50}$$

Man kann zur kompletten (m,m)-Matrix Q übergehen, indem man Q_r restlichen $m-r$ Vektoren der Basis $\underline{1},\ldots,\mathbf{q}_m$ zufügt, ebenso kann man zur kompletten Basis $\mathbf{v}_1,\ldots,\mathbf{v}_n$ übergehen, indem man die restlichhen $n-r$ der Basis zu V_r hinzufügt. Dann muß Λ_r um $m-r$ Zeilen und $n-r$ Spalten mit Nullen erweitert werden, so dass eine (m,n)-Matrix Λ entsteht:

$$X = Q\Lambda^{1/2}V', \quad \Lambda^{1/2} = \begin{pmatrix} \Lambda_r^{1/2} & 0_{12} \\ 0_{21} & 0_{22} \end{pmatrix}, \tag{6.51}$$

wobei 0_{12} eine $(r, n-r)$-Matrix ist, 0_{21} eine $(m-r, r)$-Matrix und 0_{22} eine $(m-r, n-r)$-Matrix ist; die Elemente von 0_{12}, 0_{21} und 0_{22} sind alle gleich Null. Setzt man $\Sigma = \Lambda^{1/2}$, so erhält man die Darstellung der *Singularwertzerlegung* mit der üblichen Abkürzung SVD (Singular Value Decomposition); im Deutschen ist auch der Ausdruck *Grundstruktur* gebräuchlich):

$$X = Q\Sigma V'. \tag{6.52}$$

Anmerkungen:
1. Die Diagonalmatrix Σ enthält die Wurzeln aus den Eigenwerten; sie heißen auch *Singularwerte*; $\sigma_j = \sqrt{\lambda_j}$, $j = 1, \ldots, r$. Die Spaltenvektoren von Q heißen *Linkseigenvektoren*, die von V heißen *Rechtseigenvektoren* von X.
2. Die SVD ist nicht an eine Spaltenzentrierung oder Standardisierung der Matrix X gebunden; eine SVD kann für eine beliebige Matrix X bestimmt werden. □

6.9 Die Differentiation von Vektoren

6.9.1 Die allgemeine Differentiationsformel

Ein Vektor ist durch seine Komponenten festgelegt. Man kann dann fragen, wie sich der Vektor verändert, wenn man seine Komponenten verändert. Solche Veränderungen lassen sich oft durch einen Differentialquotienten beschreiben. So sei etwa $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$. Dabei wird angenommen, dass keine Komponente von der anderen abhängt. Man definiert nun den Differentialquotienten von $\mathbf{x}$ in Bezug auf die j-te Komponente x_j durch

$$\frac{d\mathbf{x}}{dx_j} = \begin{pmatrix} dx_1/dx_j \\ \vdots \\ dx_j/dx_j \\ \vdots \\ dx_n/dx_j \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{e}_j. \tag{6.53}$$

Dann hat man

$$\frac{d\mathbf{x}}{d\mathbf{x}} = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n] = I, \tag{6.54}$$

I die (n, n)-Einheitsmatrix.

Der Fall (6.53) tritt u.a. dann auf, wenn eine Größe in Abhängigkeit von einem Vektor $\mathbf{b} = (b_1, \ldots, b_p)'$ von Parametern maximiert oder minimiert werden soll.

Es sei $\mathbf{y} = A\mathbf{x}$; für eine gegebene Matrix A hängt $\mathbf{y}$ von $\mathbf{x}$ ab, d.h. $\mathbf{y}$ ist eine Funktion von $\mathbf{x}$, so dass man $\mathbf{y} = \mathbf{y}(\mathbf{x})$ schreiben kann. Weiter ist

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

so dass

$$\frac{\partial \mathbf{y}}{\partial x_j} = \frac{\partial A\mathbf{x}}{\partial x_j} = \mathbf{a}_j, \quad j = 1, \ldots, n$$

Dementsprechend erhält man

$$\frac{d\mathbf{y}}{d\mathbf{x}} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n] = A. \tag{6.55}$$

Beispiel 6.1 Alternativer Beweis (I) für $T'T = I \Rightarrow TT' = I$, T eine (n,n)-Matrix: Für einen beliebigen Vektor $\mathbf{y} \in \mathbb{R}^n$ gelte $T\mathbf{y} = \mathbf{x}$, $\mathbf{x} \in \mathbb{R}^n$. Dann

$$\begin{aligned} T\mathbf{y} &= \mathbf{x} \Rightarrow T'T\mathbf{y} = T'\mathbf{x} \\ \text{d.h.}\, \mathbf{y} &= T'\mathbf{x} \\ \Rightarrow T\mathbf{y} &= TT'\mathbf{x} = \mathbf{x}, \quad \text{für alle } \mathbf{x} \end{aligned}$$

Die letzte Zeile ist mit der Folgerung $TT' = I$ kompatibel. Andererseits verweist die Gleichung $TT'\mathbf{x} = \mathbf{x}$ auf die Möglichkeit, dass $\mathbf{x}$ ein Eigenvektor von TT' ist mit zugehörigem Eigenwert $\lambda = 1$ und $TT' \neq I$. Allerdings soll die Annahme $T\mathbf{y} = \mathbf{x}$ für einen beliebigen Vektor $\mathbf{y}$ und zugehörigem Vektor $\mathbf{x}$ gelten. Dann kann man die Ableitung $dTT'/d\mathbf{x}$ betrachten, für die sich allerdings

$$\frac{dTT'\mathbf{x}}{d\mathbf{x}} = TT' = \frac{d\mathbf{x}}{d\mathbf{x}} = I$$

nach (6.54) und (6.55) ergibt. □

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.

6.9.2 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)

$$\frac{\partial Q}{\partial x_j} = \mathbf{e}_j'C\mathbf{x} + \mathbf{x}'C\mathbf{e}_j \tag{6.56}$$

$\mathbf{e}_j$ 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

$$\frac{\partial Q}{\partial x_j} = 2\mathbf{e}_j'C\mathbf{x}. \tag{6.57}$$

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

$$\frac{\partial Q}{\partial \mathbf{x}} = 2IC\mathbf{x} = 2C\mathbf{x}. \tag{6.58}$$

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 (6.58)

$$\frac{\partial \mathbf{x}'\mathbf{x}}{\partial \mathbf{x}} = 2I\mathbf{x} = 2\mathbf{x}. \tag{6.59}$$

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 6.9.4 nachgelesen werden.

Zu maximieren sei

$$Q(\mathbf{x}) = \mathbf{x}'A\mathbf{x} + \lambda(\mathbf{x}'\mathbf{x} - a), \quad \mathbf{x}'\mathbf{x} - a = 0, \quad A' = A \tag{6.60}$$

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

$$A\mathbf{x}_{\max} = \lambda\mathbf{x}_{\max}, \tag{6.61}$$

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

6.9.3 Die Kleinste-Quadrate-Schätzung für das Lineare Modell

Es sei X eine (m, n)-Matrix mitvollem Rang $r = n \leq m$, $\mathbf{b} \in \mathbb{R}^n$ sei ein unbekannter Vektor von (Regressions-)Parametern und $\mathbf{y}, \mathbf{e} \in \mathbb{R}^m$ seinen Vektoren; es gelte

$$\mathbf{y} = X\mathbf{b} + \mathbf{e} \tag{6.62}$$

$\mathbf{e}$ ist ein Fehlervektor, und $\mathbf{b}$ soll so bestimmt werden, dass der Fehler minimalisiert wird. Dies soll heißen, dass $\mathbf{e}'\mathbf{e} = \|\mathbf{e}\|^2$ minimal werden soll; $\mathbf{e}$ soll so kurz wie möglich werden. Es ist

$$\mathbf{e}'\mathbf{e} = (\mathbf{y} - X\mathbf{b})'(\mathbf{y} - X\mathbf{b}) = \mathbf{y}'\mathbf{y} - \mathbf{y}'X\mathbf{b} - \mathbf{b}'X'\mathbf{y} + \mathbf{b}'X'X\mathbf{b} \tag{6.63}$$

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 (6.55) ist $d(X\mathbf{b})/d\mathbf{b} = X$, und nach (6.58) ist $d(\mathbf{b}'X'X\mathbf{b})/d\mathbf{b} = 2X'X\mathbf{b}$. $\mathbf{e}'\mathbf{e}$ wird minimal für $\mathbf{b} = \hat{\mathbf{b}}$ derart, dass

$$\left.\frac{d(\mathbf{e}'\mathbf{e})}{d\mathbf{b}}\right|_{\mathbf{b}=\hat{\mathbf{b}}} = 2(X'\mathbf{y} - (X'X)\mathbf{b}) = 0 \tag{6.64}$$

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

$$\hat{\mathbf{b}} = (X'X)^{-1}X'\mathbf{y}, \tag{6.65}$$

Einige Implikationen: Es sei $X = Q\Sigma V'$, $\Sigma = \Lambda^{1/2}$, die SVD von X. Dann ist $X'X = V\Lambda V'$ und

$$(X'X)^{-1} = (V\Lambda V')^{-1} = (V')^{-1}\Lambda^{-1}V^{-1} = V\Lambda^{-1}V', \tag{6.66}$$

denn $V^{-1} = V'$ wegen der Orthonormalität von V. (6.65) impliziert dann

$$\hat{\mathbf{b}} = V\Lambda^{-1}V'V\Lambda^{1/2}Q'\mathbf{y} = V\Lambda^{-1/2}Q'\mathbf{y}. \tag{6.67}$$

Aber es ist $\mathbf{y} = X\mathbf{b} + \mathbf{e} = Q\Lambda^{1/2}Q'\mathbf{b} + \mathbf{e}$, so dass aus (6.67)

$$\hat{\mathbf{b}} = V\Lambda^{-1/2}Q'(Q\Lambda^{1/2}Q'\mathbf{b} + \mathbf{e}) = \mathbf{b} + V\Lambda^{-1/2}Q'\mathbf{e} \tag{6.68}$$

folgt, und schreibt man $V\Lambda^{-1/2}Q'$, indem man das dyadische Produkt anwendet, so ergibt sich (vergl. (3.31), Seite 89, wo der Fall $X = Q\Lambda^{1/2}V'$ betrachtet wird)

$$\hat{\mathbf{b}} = \mathbf{b} + \left(\sum_{k=1}^{n} \frac{\mathbf{v}_k\mathbf{q}_k'}{\sqrt{\lambda_k}}\right)\mathbf{e} \tag{6.69}$$

Der wahre Parametervektor $\mathbf{b}$ und die Kleinste-Quadrate-Schätzung $\hat{\mathbf{b}}$ unterschieden sich also um den Vektor $V\Lambda^{-1/2}Q'\mathbf{e} = (\sum_{k=1}^{n} \mathbf{v}_k\mathbf{q}_k'/\sqrt{\lambda_k})\mathbf{e}$. 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 $(v_{1j}, \ldots, v_{kj}, \ldots, v_{nj})'$ gegeben und man hat

$$\mathbf{c}_j = v_{1j}\lambda_1\mathbf{v}_1 + v_{2j}\lambda_2\mathbf{v}_2 + \cdots + v_{nj}\lambda_n\mathbf{v}_n. \tag{6.70}$$

Sind nur die ersten Eigenwerte $\lambda_1, \lambda_2, \ldots$ "groß" und sind die restlichen "klein", so werden die Kovarianzen bzw. Korrelationen in den Spalten von C nur durch die ersten Eigenvektoren "erklärt", die restlichen spielen eine geringere Rolle, d.h. die Messwerte werden durch wenige latente Variablen (repräsentiert durch die Eigenvektoren von C) bestimmt, – was hohe Korrelationen (Absolutbetrag) zwischen den Variablen bedeutet. In der multiplen Regression sind die Variablen die Prädiktoren für $\mathbf{y}$. Korrelierte Prädiktoren bedeuten also die Existenz kleiner Eigenwerte und damit große Differenzen zwischen $\mathbf{b}$ und der Schätzung $\hat{\mathbf{b}}$, – und damit ungenaue Voraussagen für $\mathbf{y}$. Die Prädiktoren sollten daher möglichst unkorreliert sein. Eine Möglichkeit, den Effekt von Korrelationen zwischen den Prädiktoren zu reduzieren, besteht wieder darin, für X die SVD $Q\Sigma V'$ einzusetzen:

$$\mathbf{y} = Q\Sigma V'\mathbf{b} + \mathbf{e} = Q\Sigma(V'\mathbf{b}) + \mathbf{e} = L\boldsymbol{\beta} + \mathbf{e}, \quad L = Q\Sigma \tag{6.71}$$

mit L als neuer Prädiktormatrix und $\boldsymbol{\beta} = V'\mathbf{b}$ als neuem Parametervektor; die Spaltenvektoren von L sind orthgonal und also unkorreliert. Im Skriptum über Regressionsverfahren wird dieser Ansatz ausführlicher diskutiert.

6.9.4 Extrema unter Nebenbedingungen

Es sei $f(x_1, \ldots, x_n)$ eine Funktion der Variablen $x_1, \ldots, x_n$. Gesucht sind diejenigen Werte x_{0j} von x_j, $j = 1, \ldots, n$, für die f ein Maximum annimmt, wobei aber die Nebenbedingung $g(x_1, \ldots, x_n) = k$, k ein Konstante berücksichtigt werden soll, d.h. der Vektor $\mathbf{x}_0 = (x_{01}, \ldots, x_{0n})'$ soll so bestimmt werden, dass auch $g(x_{01}, \ldots, sx_{0n}) = k$ erfüllt ist. Man kann i.A. g so definieren, dass $k = 0$ gesetzt werden kann.

Der Einfachheit halber werden die Überlegungen zur Maximierung unter Nebenbedingungen für den Fall $n = 2$ durchgeführt; das Resultat überträgt sich unmittelbar auf den Fall $n > 2$. Dazu wird $x = x_1$, $y = x_2$ gesetzt. Es soll also $f(x, y)$ unter der Nebenbedingung $g(x, y) = 0$ maximiert werden (oder allgemein ein Extremwert bestimmt werden).

$g(x, y) = 0$ bedeutet, dass es eine Funktion $y = g(x)$ gibt, so dass $f(x, y) = f(x, g(x))$ und $g(x, g(x)) = 0$ geschrieben werden kann. Geometrisch beschreibt $f(x, y)$ eine Fläche im 3-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

$$f_x + f_y g' = 0 \tag{6.72}$$

$$g_x + g_y g' = 0 \tag{6.73}$$

Die bisher hergeleiteten Ableitungen enthalten noch die Ableitung g' von g. Um das Extremum zu bestimmen, eliminiert man g' am besten, da die Bestimmung von g' kompliziert sein kann. Man hat nun $g' = -f_x/f_y = -g_x/g_y$; diese Beziehung bedeutet, dass die *Gradientenvektoren* $(f_x, f_y)'$ und $(g_x, g_y)'$ dieselbe Orientierung haben, d.h. sie unterscheiden sich allenfalls in ihrer Länge, so dass man

$$\begin{pmatrix} f_x \\ f_y \end{pmatrix} = \lambda \begin{pmatrix} g_x \\ g_y \end{pmatrix} \tag{6.74}$$

schreiben kann. $\lambda \in \mathbb{R}$ ist ein neuer, freier Parameter, der sogenante *Lagrange-Faktor* oder *Lagrange-Multiplikator*. Er drückt einfach aus, dass man nur etwas über die Orientierung, nicht aber über die Länge der Gradientenvektoren am Ort des Maximums weiß. Die Vektorgleichung (6.74) zusammen mit der Bedingung $g(x, y) = 0$ führt sofort auf ein System von drei Gleichungen mit den drei Unbekannten x, y und λ:

$$f_x - \lambda g_x = 0 \tag{6.75}$$

$$f_y - \lambda g_y = 0 \tag{6.76}$$

$$g(x, y) = 0 \tag{6.77}$$

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

$$L(x, y, \lambda) = f(x, y) + \lambda g(x, y), \quad g(x, y) = 0. \tag{6.78}$$

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 (6.75), (6.76) und (6.77) heißen zusammen die *Lagrangesche Multiplikatorenregel*, nach dem Mathematiker und Astronomen Jean-Louis Lagrange (1736 – 1813), der diese Regel 1788 herleitete.

Beispiel 6.2 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

$$f_x = -2x, \; f_y = -\frac{2}{3}y, \; g_x = 1, \; g_y = 1,$$

und man erhält das Gleichungssystem

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

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

Beispiel 6.3 (Satz von Courant-Fisher). Es sei A eine symmetrische, positiv-definite $n \times n$-Matrix mit den Eigenwerten $\lambda_1 \geq \lambda_2 \leq \cdots \geq \lambda_n$. Dann gilt

$$\max_{\mathbf{x} \neq \vec{0}} \frac{\mathbf{x}' A \mathbf{x}}{\mathbf{x}'\mathbf{x}} = \max_j \lambda_j = \lambda_1, \tag{6.79}$$

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

$$\min_{\mathbf{x} \neq \vec{0}} \frac{\mathbf{x}' A \mathbf{x}}{\mathbf{x}'\mathbf{x}} = \min_j \lambda_j = \lambda_n \tag{6.80}$$

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. □

6.10 Der Rayleigh-Koeffizient und der Satz von Courant-Fischer

Satz 6.5 *(Satz von Courant-Fischer) Es sei M eine symmetrische, positiv definite Matrix mit $M = T\Lambda T'$, T die Matrix der Eigenvektoren und*

$$\Lambda = diag(\lambda_1, \ldots, \lambda_n), \quad \lambda_1 \geq \cdots \geq \lambda_n$$

die Diagonalmatrix der zugehörigen Eigenwerte von M. Dann ist

$$\max_{\boldsymbol{x} \neq \vec{0}} R(\boldsymbol{x}) = \max_{\boldsymbol{x} \neq \vec{0}} \frac{\boldsymbol{x}' M \boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}} = \max_j \lambda_j = \lambda_1, \tag{6.81}$$

und der Vektor $\boldsymbol{x}$, für den das Maximum angenommen wird, ist der zu λ_1 korrespondierende Eigenvektor $\boldsymbol{t}_1$, so dass $\max_{\boldsymbol{x} \neq \vec{0}} R(\boldsymbol{x}) = \boldsymbol{t}_1' M \boldsymbol{t}_1 = \lambda_1$. Weiter gilt

$$\max_{\boldsymbol{x} \perp \boldsymbol{t}_1, \ldots, \boldsymbol{t}_k} \frac{\boldsymbol{x}' M \boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}} = \lambda_{k+1}, \quad k < n, \tag{6.82}$$

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

$$\min_{\boldsymbol{x} \neq \vec{0}} R(\boldsymbol{x}) = \frac{\boldsymbol{x}' M \boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}} = \min_j \lambda_j, \tag{6.83}$$

mit dem zugehörigen Eigenvektor $\boldsymbol{t}_{\min}$, d.h. $\min_{\boldsymbol{x} \neq \vec{0}} R(\boldsymbol{x}) = \boldsymbol{t}_{\min} M \boldsymbol{t}_{\min} = \lambda_{\min}$.

Beweis: Sei $T = [\mathbf{t}_1, \ldots, \mathbf{t}_n]$ die Matrix der Eigenvektoren von M mit den zugehörigen Eigenwerten $\lambda_1 \geq \cdots, \geq \lambda_n$, so dass $MT = T\Lambda$, $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$. Die Eigenvektoren $\mathbf{t}_j$, $j = 1, \ldots, n$ bilden eine orthonormale Basis des $V_n = \mathcal{L}(\mathbf{t}_1, \ldots, \mathbf{t}_n)$. Es sei nun $\mathbf{x}$ ein beliebiger n-dimensionaler Vektor. Er kann stets als Linearkombination einer Basis dargestellt werden, insbesondere können die Spaltenvektoren $\mathbf{t}_j$ von T gewählt werden[97]. Es gilt also insbesondere

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

für für geeignet gewählte Koeffizienten $c_1, \ldots, c_n$. Dann ist

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

Alle Kreuzprodukte $c_jc_k\mathbf{t}_j'\mathbf{t}_k$ verschwinden, weil $\mathbf{t}_j'\mathbf{t}_k = 0$ für alle $j \neq k$, so dass

$$\frac{\mathbf{x}'M\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \frac{\sum_{j=1}^n c_j^2\lambda_j}{\sum_{j=1}^n c_j^2}$$

und $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Ersetzt man also auf der rechten Seite alle $\lambda_j \neq \lambda_1$ durch λ_1, so folgt

$$\frac{\mathbf{x}'M\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \frac{\sum_{j=1}^n c_j^2\lambda_j}{\sum_{j=1}^n c_j^2} \leq \frac{\lambda_1\sum_{j=1}^n c_j^2}{\sum_{j=1}^n c_j^2} = \lambda_1,$$

so dass $\max_{\mathbf{x}} R(\mathbf{x}) = \lambda_1$, λ_1 der größte Eigenwert. Jetzt muss noch gezeigt werden, für welchen Vektor $\mathbf{x}$ das Maximum angenommen wird. Es sei $\mathbf{t}_1$ der zu $\lambda_{\max} = \lambda_1$ korrespondierende Eigenvektor. Für $\mathbf{x} = \mathbf{t}_1$ folgt $\mathbf{t}_1'M\mathbf{t}_1 = \lambda_1$ und wegen $\mathbf{t}_1'\mathbf{t}_1 = 1$ sieht man, dass $R(\mathbf{x}) = \max$ für $\mathbf{x} = \mathbf{t}_1$.

Die Aussage (6.83) ergibt sich aus (6.82). Es sei wieder $T = [\mathbf{t}_1, \cdots, \mathbf{t}_n]$ die Matrix der Eigenvektoren von M. Dann existieren relle Zahlen $y_1, \ldots, y_n$ derart, dass ein Vektor $\mathbf{x}$ in der Form $\mathbf{x} = T\mathbf{y}$ dargestellt werden kann, wobei die Komponenten von $\mathbf{y}$ durch die y_j gegeben sind. Nun soll speziell $\mathbf{x} \perp \mathbf{t}_1, \ldots, \mathbf{t}_k$ gelten. Dann folgt aber

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

denn $\mathbf{t}_k'\mathbf{t}_k = 1$, so dass $y_k\mathbf{t}_k'\mathbf{t}_k = y_k$. Die Forderung der Orthogonalität von $\mathbf{x}$ zu den ersten k Eigenvektoren impliziert also $y_1 = \cdots = y_k = 0$. Dann folgt aus (6.81)

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

und analog zur Argumentation im Beweis zu Satz 6.5, Seite 198, folgt (6.82). □

Anmerkungen:

[97] s. die Anmerkung 1. auf Seite 199

1. Der Beweis geht davon aus, dass die Eigenvektoren $\mathbf{t}_1, \ldots, \mathbf{t}_n$ als Basis des V_n gewählt werden können. Diese Wahl ist keine Einschränkung der Allgemeinheit, da man von einer beliebigen Basis V stets zu einer orthonormalen Basis übergehen kann, vergl. die Singularwertzerlegung (Abschnitt **??**), angewendet auf V. Deswegen bedeutet dieser Übergang auch nur, einen Umweg zu gehen: man kann gleich mit den Eigenvektoren als Basis beginnen.

2. Für $k = 0$ betrachtet man $\max_{\mathbf{x} \neq \vec{0}} \frac{\mathbf{x}'M\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_1$, und für $k = n-1$ erhält man
$$\max_{\mathbf{x} \perp \mathbf{t}_1, \ldots, \mathbf{t}_{n-1}} \frac{\mathbf{x}'M\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_n,$$
d.h.
$$\min_{\mathbf{x}} R(\mathbf{x}) = \lambda_n \tag{6.84}$$
so dass
$$\lambda_n \leq R(\mathbf{x}) \leq \lambda_1. \tag{6.85}$$

3. Es sei M eine beliebige symmetrische (m, n)-Matrix und $\mathbf{x} \in \mathbb{R}^n$. Weiter sei $\lambda = \mathbf{x}'M\mathbf{x}/\mathbf{x}'\mathbf{x}$. Die Werte λ heißen *maximal*, wenn sie den Rayleigh-Quotienten $\mathbf{x}'M\mathbf{x}/\mathbf{x}'\mathbf{x}$ maximieren.

4. Es sei T die Matrix der Eigenvektoren von $M = X'X$; für irgendzwei Eigenvektoren $\mathbf{t}_j$ und $\mathbf{t}_k$, $j \neq k$, gilt also $\mathbf{t}_j'\mathbf{t}_k = 0$. Es sei $L = XT$, so dass $\mathbf{L}_j = X\mathbf{t}_j$, $j = 1, \ldots, n$. Dann gilt
$$\mathbf{L}_j'\mathbf{L}_k = \mathbf{t}_j'X'X\mathbf{t}_k = \mathbf{t}_j'M\mathbf{t}_k = 0, \tag{6.86}$$
d.h. die Matrix L ist orthogonal. Denn $M = T\Lambda T'$, Λ die Diagonalmatrix der Eigenwerte on M. Dann hat man
$$\mathbf{t}_j'M\mathbf{t}_k = \mathbf{t}_j'T\Lambda T'\mathbf{t}_k = \mathbf{e}_j'\Lambda\mathbf{e}_k = \lambda_j\mathbf{e}_j'\mathbf{e}_k = 0, \quad j \neq k.$$

$\square$

6.11 Alternativer Beweis von Satz 3.17

Es sei $\mathbf{a}_{zi}$ der i-te Zeilenvektor von A, $i = 1, \ldots, m$. Es sei $A\mathbf{x} = \vec{0}$; dies bedeutet, dass die Skalarprodukte $\mathbf{a}_{ui}'\mathbf{x} = 0$ für alle i, d.h. $\mathbf{x}$ ist orthogonal zu allen Zeilenvektoren von A. Für die $\mathbf{x}$ mit $A\mathbf{x} = \mathbf{y} \neq \vec{0}$ gilt diese Aussage nicht, d.h. $\mathbb{R}^n$ wird in zwei Teilmengen U und V zerlegt:
$$U = \{\mathbf{x} \in \mathbb{R}^n | A\mathbf{x} = \vec{0}\} = \text{kern}(A), \quad V = \{\mathbf{x} \in \mathbb{R} | A\mathbf{x} = \mathbf{y} \neq \vec{0}\}.$$
$U = \text{kern}(A)$ ist ein Teilraum des $\mathbb{R}^n$. V ist ebenfalls ein Teilraum des $\mathbb{R}^n$, denn es sei $A\mathbf{x}_1 = \mathbf{y}_1$, $A\mathbf{x}_2 = \mathbf{y}_2$, $\mathbf{x}_1, \mathbf{x}_2 \notin \text{kern}(A)$. Dann ist, für $\lambda, \mu \in \mathbb{R}$ beliebig, $A\lambda\mathbf{x}_1 = \lambda\mathbf{y}_1$, $A\mu\mathbf{x}_2 = \mu\mathbf{y}_2$ und $A(\lambda\mathbf{x}_1 + \mu\mathbf{x}_2) = \lambda\mathbf{y}_1 + \mu\mathbf{y}_2 = \mathbf{y} \in \mathcal{L}(A)$, mithin

ist $\mathbf{x} = \lambda\mathbf{x}_1 + \mu\mathbf{x}_2 \in V$. Offenbar ist $U \cap V = \emptyset$, denn ein Vektor $\mathbf{x} \in \mathbb{R}^n$ kann nicht zugleich in U und in V sein. Also folgt $U + V = \mathbb{R}$ und es folgt

$$\dim(U + V) = \dim(U) + \dim(V) = \dim \mathbb{R}^n = n. \tag{6.87}$$

Zur Bestimmung von $\dim(U)$ werde

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

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

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

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

$$\mathbf{a}_{r+k} = \lambda_{k1}\mathbf{a}_1 + \cdots + \lambda_{kr}\mathbf{a}_r = \sum_{j=1}^{r} \lambda_{kj}\mathbf{a}_j \tag{6.88}$$

gelten muss, so hat man

$$\sum_{j=1}^{r} x_j\mathbf{a}_j = -\sum_{k=1}^{n-r} x_{r+k} \sum_{j=1}^{r} \lambda_{kj}\mathbf{a}_j. \tag{6.89}$$

Über den Vektor $\mathbf{x}$ ist bisher keine weitere Annahme gemacht worden außer, dass $A\mathbf{x} = \vec{0}$ gelten soll. Man kann also insbesondere

$$\mathbf{x} = \mathbf{x}_{r+k} = (x_1, \ldots, x_r, 0, \ldots, 0, 1, 0, \ldots, 0)', \quad k = 1, \ldots, n-r$$

setzen, wobei die 1 an der $(r+k)$-ten Stelle stehen soll, also

$$\begin{aligned}
\mathbf{x}_{r+1} &= (x_1, \ldots, x_r, 1, 0, \ldots, 0)', \\
\mathbf{x}_{r+2} &= (x_1, \ldots, x_r, 0, 1, 0, \ldots, 0)', \\
\mathbf{x}_{r+3} &= (x_1, \ldots, x_r, 0, 0, 1, 0, \ldots, 0)' \\
&\vdots
\end{aligned}$$

Die Gleichung (6.89) nimmt dann die Form

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

so dass

$$\sum_{j=1}^{r} \lambda_{kj}\mathbf{a}_j + \sum_{j=1}^{r} x_j\mathbf{a}_j = \sum_{j=1}^{n} (x_j + \lambda_j)\mathbf{a}_j = 0.$$

Da die $\mathbf{a}_j$ linear unabhängig sind folgt $x_j + \lambda_j = 0$ oder $-\lambda_j = x_j$ für alle j. es ist also

$$\mathbf{x}_k = (-\lambda_{k1}, \ldots, -\lambda_{kr}, 0, \ldots, 0, 1, 0, \ldots, 0)', \quad k = 1, \ldots, n-r \tag{6.90}$$

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

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

wegen (6.88).

Die $\mathbf{x}_k$ sind linear unabhängig, wie man sofort sieht, denn

$$\mu_1\mathbf{x}_1 + \cdots + \mu_k\mathbf{x}_k = 0$$

impliziert $\mu_i = 0$ für $i = 1, \ldots, n-r$. Sie bilden damit eine Basis für einen $(n-r)$-dimensionalen Teilraum. Damit ist gezeigt, dass kern(A) mindestens $(n-r)$-dimensional ist. Die Frage ist, ob die Dimensionalität von kern(A) nicht größer ist. Das ist aber nicht möglich, da ja bereits rg$(A) = r$ angebommen wurde, der Rang von kern(A) kann also nicht größer als $n - r$ sein. Wegen (6.87) (= Dimensionssatz) folgt weiter, dass rg$(V) = \mathcal{L}(A) = r$ ist. $\square$

Für $r = n$ folgt demnach rg[kern(A)] $= 0$, d.h. in diesem Fall hat $A\mathbf{x} = \vec{0}$ nur eine Lösung: $\mathbf{x} = \vec{0}$. Dies ist evident, denn in diesem Fall sind die $\mathbf{a}_1, \ldots, \mathbf{a}_n$ linear unabhängig und $\sum_j x_j\mathbf{a}_j = \vec{0}$ nur dann, wenn $x_1 = \cdots = x_n = 0$. $\square$

6.12 Vektortransformationen und Abbildungen

Dieser Abschnitt enthält einige grundsätzliche Betrachtungen über Produkte von Matrizen und Vektoren, die einerseits das Verständnis der Vektor- und Matrixrechung vertiefen, andererseits für das Verständnis der unmittelbaren Anwendung der Matrixrechung auf Fragen der multivariaten Statistik nicht unbedingt notwendig sind und deshalb übersprungen werden können.

Die Matrix als Abbildung 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 ein* Element aus $\mathcal{N}$ zu:

$$f : \mathcal{M} \to \mathcal{N}, \quad x \mapsto y = f(x), \quad x \in \mathcal{M}, y \in \mathcal{N} \tag{6.91}$$

Man schreibt gelegentlich auch

$$f(\mathcal{M}) = \mathcal{N}. \tag{6.92}$$

$f(\mathcal{M})$ heißt das *Bild von* $\mathcal{M}$ in $\mathcal{N}$, und $\mathcal{M}$ ist das *Urbild* von $f(\mathcal{M})$. Man schreibt auch $\mathrm{Im} f = \mathcal{N}$[98].

Eine spezielle Abbildung ist die *Identität* oder identische Abbildung

$$\mathrm{id}(\mathcal{M}) = \mathcal{M}. \tag{6.93}$$

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

$$I_n \mathbf{x} = \mathbf{x}, \quad \mathbf{x} \in \mathbb{R}^n. \tag{6.94}$$

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

$$f(\mathcal{M}) = \mathcal{N}, \quad f^{-1} f(\mathcal{M}) = \mathcal{M} = f^{-1}(\mathcal{N}). \tag{6.95}$$

Wenn f durch eine Matrix M definiert ist, so bedeutet die Existenz der inversen Abbildung f^{-1} die Existenz einer inversen Matrix M^{-1}. Es wird deutlich werden, dass inverse Matrizen M^{-1} für eine Matrix M nur für spezielle Matrizen existieren.

Die Forderung, dass einem Element $x \in \mathcal{M}$ nur *ein* Element $y \in \mathcal{N}$ zugeordnet wird schließt nicht aus, das verschiedenen Elementen $x, x' \in \mathcal{M}$ der gleiche Wert $y \in \mathcal{N}$ zugeordnet werden kann. In diesem Fall kann von einem Element $y \in \mathcal{N}$ nicht eindeutig auf das Element $x \in \mathcal{M}$ mit $f(x) = y$ zurückgeschlossen werden.

Mit der Schreibweise $f(\mathcal{M})$ ist nicht ein einzelnes Element gemeint, sondern die Menge der Werte, die man erhält, wenn man f für alle Werte aus X bestimmt, also

$$f(\mathcal{M}) = \{f(x), x \in \mathcal{M}\}. \tag{6.96}$$

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

Definition 6.5 *Es sei $f : \mathcal{M} \to \mathcal{N}$. Dann ist f*

1. injektiv*, wenn aus $x, x' \in \mathcal{M}$ und $f(x) = f(x')$ folgt, dass $x = x'$ (und damit $f(x) \neq f(x') \Rightarrow x \neq x'$). Es kann $f(\mathcal{M}) \subset \mathcal{N}$ gelten, d.h. $f(\mathcal{M})$ kann eine echte Teilmenge von $\mathcal{N}$ sein.*
2. surjektiv*, wenn zu jedem $y \in \mathcal{N}$ ein $x \in \mathcal{M}$ existiert derart, dass $y = f(x)$. Es gilt $f(\mathcal{M}) = \mathcal{N}$.*
3. bijektiv*, wenn f sowohl injektiv als auch surjektiv ist. Es gilt $f(\mathcal{M}) = \mathcal{N}$.*
4. Die Menge $kern(f) = \{x \in \mathcal{M} | f(x) = 0\}$ heißt Kern *der Abbildung f; man schreibt für den Kern auch $kern(f) = f^{-1}(\vec{0})$.*
5. Es sei $f(\mathcal{M}) = \mathcal{N}$, d.h. $f(\boldsymbol{x}) = \boldsymbol{y}$ für $\boldsymbol{x} \in \mathcal{M}$, $\boldsymbol{y} \in \mathcal{N}$. Dann heißt $f^{-1}(\boldsymbol{y}) = \{\boldsymbol{x} \in \mathcal{M} | f(\boldsymbol{x}) = \boldsymbol{y}\}$ die Faser *über $\boldsymbol{y} \in \mathcal{N}$.*

Anmerkung: Die Schreibweise $f^{-1}(f)$ für den Kern einer Abbildung f ergibt sich aus der in 4. gegebenen Definition: ist $\mathbf{x} \in \mathrm{kern} f$, so gilt $f(\mathbf{x}) = \vec{0}$. Aus der

[98] Im 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 Defintion der Faser offenbar voraus, dass die Inverse existiert. $\square$

Beispiele: $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto ax + b$ für $a, b \in \mathbb{R}$ fest gewählte Konstante. $\mathbb{R}$ bezeichnet die Menge der reellen Zahlen. f ist sicher injektiv, denn $f(x) = f(x')$ impliziert $ax + b = ax' + b$ und damit $x = x'$, wie man leicht nachrechnet. f ist auch surjektiv, denn für $y = ax + b$ existiert genau ein $x = (y - b)/a$ derart, dass $y = f(x)$. Da f sowohl injektiv wie surjektiv ist, ist f auch bijektiv.

Mit $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ wird die Menge der Paare (x, y), $x, y, \in \mathbb{R}$, bezeichnet, allgemein mit

$$\mathbb{R}^m = \underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}}_{m-\text{mal}}$$

Die Menge der m-tupel $(x_1, x_2, \ldots, x_m)$, $x_j \in \mathbb{R}$, d.h. der m-dimensionalen Vektoren. $\mathbb{R}^n$ ist dann die Menge der n-dimensionalen Vektoren, etc. Mit $\mathbb{R}^{m,n}$ wird die Menge der (m, n)-Matrizen bezeichnet. Alle diese Definitionen übertragen sich auf $\mathcal{C}$, die Menge der komplexen Zahlen $x + iy$, $x, y \in \mathbb{R}$ und $i = \sqrt{-1}$.

Die Schreibweise $M \in \mathbb{R}^{m,n}$ bedeutet, dass M eine (m, n)-Matrix ist. Die Schreibweise $f\colon V_m \to V_n$ bedeutet dann, dass f eine Abbildung der m-dimensionalen Vektoren in die Menge der n-dimensionalen Vektoren ist. Wenn $V_m = \mathbb{R}^m$, $V_n = \mathbb{R}^n$ kann man auch $f\colon \mathbb{R}^m \to \mathbb{R}^n$ schreiben.

Da $M\mathbf{x} = \mathbf{y}$ mit $\mathbf{x} \in V_m$, $\mathbf{y} \in V_n$, folgt, dass f durch eine Matrix $M \in \mathbb{R}^{m,n}$ definiert ist.

Definition 6.6 *Es seien V und W Vektorräume und $f\colon V \to W$ sei eine Abbildung von V ind W.*
1. f heißt linear *bzw.* homomorph*, wenn*

$$f(\lambda \boldsymbol{v} + \mu \boldsymbol{w}) = \lambda f(\boldsymbol{v}) + \mu f(\boldsymbol{w}) \tag{6.97}$$

für $\lambda, \mu \in \mathbb{R}$ und für alle $\boldsymbol{v} \in V$ und $\boldsymbol{w} \in W$.
2. f heißt isomorph*, wenn f bijektiv ist; man sagt auch, f definiere einen* Homomorphismus *bzw.* Isomorphismus *wenn f bijektiv ist.*
3. f definiert einen Endomorphismus*, wenn $V = W$, und*
4. einen Automorphismus*, wenn f bijektiv ist und außerdem $V = W$ gilt.*

Es sei $M \in \mathbb{R}^{m,n}$; M definiert eine lineare, also homomorphe Abbildung, denn $M\mathbf{x} = \mathbf{y}$ erfüllt die Bedingungen einer linearen Abbildung. für $m \neq n$ ist f offenbar weder ein Endomorphismus noch ein Automorphismus.

Dem Begriff des Kerns in der allgemeinen Definition 6.5 von Abbildungen entspricht für $f \in \mathbb{R}^{m,n}$ der Nullvektor $\vec{0}$.

Satz 6.6 *Es sei $f(\mathcal{M}) = \mathcal{N}$. Dann gilt*
1. f ist surjektiv genau dann, wenn $\operatorname{Im} f = \mathcal{N}$
2. f ist injektiv genau dann, wenn $kern f = \{\vec{0}\}$.

3. f sei injektiv und die Vektoren $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathcal{M}$ seien linear unabhängig. Dann sind auch die Bilder $f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_n)$ linear unabhängig.

Anmerkung: Die Schreibweise kern$f = \{\vec{0}\}$ bedeutet, dass kern$f$ nur das eine Element $\vec{0}$ enthält. □

Beweis: $\Rightarrow$ für 1. und 2. folgt sofort aus der Definition von injektiv und surjektiv. Um $\Leftarrow$ zu sehen, betrachte man zwei Vektoren $\mathbf{u}, \mathbf{v} \in \mathcal{M}$ mit $\mathbf{u} \neq \mathbf{v}$, aber $f(\mathbf{u}) = f(\mathbf{v})$. Wegen der Linearität von f folgt dann

$$f(\mathbf{v}) - f(\mathbf{u}) = f(\mathbf{v} - \mathbf{u}) = \vec{0},$$

d.h. es gilt $\mathbf{v} - \mathbf{u} \in \text{kern}f$.

Um 3. einzusehen sei angenommen, dass

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

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

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

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

Definition 6.7 *Es sei f eine Abbildung $f\colon \mathbb{R}^m \to \mathbb{R}^n$; dann heißt die Dimensionalität des Bildes* Imf *der* Rang; *man schreibt rg$(f) = \dim \in f$.*

f sei durch eine Matrix $A \in \mathbb{R}^{m,n}$ definiert, so dass $A : \mathbb{R}^m \to \mathbb{R}^n$ und $\mathbf{x} \mapsto \mathbf{y} = A\mathbf{x}$. Dann ist $f = A$. $\mathbf{e}_1, \ldots, \mathbf{e}_m$ ist die kanonische Basis von $\mathbb{R}^m$, d.h. die n m-dimensionalen Spaltenvektoren von A können als Linearkombinationen $A\mathbf{e}_1, A\mathbf{e}_2, \ldots, A\mathbf{e}_n$ geschrieben werden. Dann ist das Bild der durch A definierten Abbildung die lineare Hülle

$$\text{Im}A = \mathcal{L}(A\mathbf{e}_1, A\mathbf{e}_2, \ldots, A\mathbf{e}_n).$$

Demnach wird ImA auch der *Spaltenraum* von A bezeichnet. Der Begriff des Ranges einer Matrix A wird in Abschnitt 2.3 noch ausführlich diskutiert.

Beispiel 6.4 Es sei $f : \mathbb{R}^3 \to \mathbb{R}^4$; für $\mathbf{x} \in \mathbb{R}^3$ und $\mathbf{y} \in \mathbb{R}^4$ soll also $f(\mathbf{x}) = \mathbf{y}$ gelten; insbesondere sei f durch

$$f(\mathbf{x}) = \begin{pmatrix} x_1 + 2x_2 \\ 0 \\ \frac{1}{2}x_1 + x_2 \\ 0 \end{pmatrix}$$

definiert. Gesucht ist die zu f gehörige Matrix $M = A$ sowie der Kern von f.

Der Kern von f ist diejenige Menge von Vektoren $\mathbf{x}$, für die $f(\mathbf{x}) = \vec{0}$. Für dies Komponenten dieser Vektoren $\mathbf{x}$ muss also gelten

$$\begin{aligned} x_1 + 2x_2 &= 0 \\ \frac{1}{2}x_1 + x_2 &= 0, \end{aligned}$$

d.h. $x_1 = -2x_2$. Der Kern ist dann

$$kern(f) = \{(x_1, x_2, x_3)' | x_1 = -2x_2\} = \{(-2x_2, x_2, x_3)'\}.$$

Es gilt

$$A\mathbf{x} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \mathbf{y}.$$

Es gibt also 12 Elemente a_{ij}, die zu bestimmen sind, wobei allerdings nur bestimmte Relationen zwischen den Komponenten gegeben sind, die aus dem Spezialfall $A\mathbf{x} = \vec{0}$ folgen. Wie die Diskussion linearer Gleichungssysteme zeigen wird, läßt sich aus diesen Bedingungen keine eindeutige Lösung für die a_{ij} ableiten.

Andererseits ist das Bild von f eine Linearkombination der Spalten von A, und damit folgt

$$\begin{pmatrix} x_1 + 2x_2 \\ 0 \\ \frac{1}{2}x_1 + x_2 \\ 0 \end{pmatrix} = \frac{x_1}{2} \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix} = (\frac{x_1}{2} + x_2) \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix},$$

d.h. die Spalten von A haben die Form $(2c, 0, c, 0)'$ mit $c \in \mathbb{R}$ (Barrantes Campos (2012), p. 231). $\square$

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

Beispiel 6.5 Es sei

$$T = \begin{pmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}. \tag{6.98}$$

T definiert eine Abbildung $\mathbb{R}^2 \to \mathbb{R}^2$:

$$T\mathbf{x} = \begin{pmatrix} x_1\cos\phi + x_2\sin\phi \\ x_1\sin\phi - x_2\cos\phi \end{pmatrix} = x_1\begin{pmatrix} \cos\phi \\ \sin\phi \end{pmatrix} + x_2\begin{pmatrix} \sin\phi \\ -\cos\phi \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

T definiert die Rotation eines Vektors $\mathbf{x}$ um einen Winkel ϕ. Dadurch werden die Elemente von $\mathbb{R}^2$ auf Elemente von $\mathbb{R}^2$ abgebildet, – $\mathbf{y}$ ist ja wieder ein Element von $\mathbb{R}^2$. Die Abbildung ist sicher injektiv und surjektiv, also bijektiv und damit umkehrbar, d.h. man kann einen Vektor $\mathbf{y} \in \mathbb{R}^2$ wählen und in "zurückdrehen", so dass man wieder bei $\mathbf{x}$ landet. Die Abbildung bzw. Matrix, die diese inverse Rotation bewirkt, wird mit T^{-1} bezeichnet.

□

6.13 Determinanten

Der Begriff der Determinante ist bereits auf Seite 77 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; allgmeiner wurden sie dann von Gottfried Wilhelm Leibniz (1646 – 1716) im Zusammenhang mit Gleichungssytemen mit $n > 2$ Unbekannten untersucht. Es war allerdings der Genfer Mathematiker Gabriel Cramer (1704 – 1752), der die allgemeine Verwendung von Determinanten zur Lösung linearer Gleichungssysteme formulierte (Cramersche Regel), und die eigentliche Entwicklung der Determinantentheorie fand erst im 19-ten Jahrhundert statt. Die heutige axiomatische Definition des Begriffs der Determinante wurde von Karl Theodor Wilhelm Weierstraß (1815 – 1897) eingeführt.

Definition 6.8 *Es sei* $A = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$ *eine quadratische Matrix, d.h. die Spaltenvektoren* $\boldsymbol{a}_j$ *von* A *seien* n*-dimensional. Die* Determinante $|A|$ *bzw.* $\det(A)$ *von* A *ist eine Zahl (im Folgenden werden nur Matrizen mit reellen Elementen betrachtet, so dass die Determinante eine reelle Zahl ist), wobei* $|A|$ *die folgenden Eigenschaften hat:*
1. $|A|$ *ist* multilinear, *d.h. es gilt mit* $j \neq k$
(i) $\det(\boldsymbol{a}_1, \ldots, \boldsymbol{u} + \boldsymbol{v}, \ldots, \boldsymbol{a}_n) = \det(\boldsymbol{a}_1, \ldots, \boldsymbol{u}, \ldots, \boldsymbol{a}_n) + \det(\boldsymbol{a}_1, \ldots, \boldsymbol{v}, \ldots, \boldsymbol{a}_n)$,

(ii) $\det(\boldsymbol{a}_1, \ldots, \lambda\boldsymbol{a}_j, \ldots, \boldsymbol{a}_n) = \lambda\det(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_j, \ldots, \boldsymbol{a}_n)$
2. $|A|$ *ist* alternierend, *d.h. gilt* $\boldsymbol{a}_j = \boldsymbol{a}_k$ *für irgendzwei Spaltenvektoren, so gilt* $|A| = 0$.
3. $\det(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n) = 1$.
Dann heißt $\det(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$ *die* Determinante *der Matrix* $A = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$.

Betrachtet man statt der Spaltenvektoren die Zeilenvektoren einer Matrix A, so ändert dies nicht den Wert der Determinante von A.

Permutationen: Eine Permutation der Zahlen $1, 2, \ldots, n$ ist eine mögliche Anordnung dieser Zahlen: für die Zahlen $\{1, 2, 3\}$ ist $(3, 1, 2)$ eine mögliche Permutation. Für n Zahlen $1, 2, \ldots, n$ enthält die Menge $\mathcal{P}$ der möglichen Permutationen $n! = 1 \cdot 2 \cdot \cdots \cdot n$ Elemente. Es sei $\tau = (i_1, i_2, \ldots, i_n)$ ein Element von $\mathcal{P}$, wobei die i_k, $k = 1, \ldots, k$ Elemente aus $\{1, 2, \ldots, n\}$ sind. $\tau(i)$ und $\tau(k)$ definierten eine *Transposition*, wenn $\tau(i) = k$ und $\tau(k) = i$. Jedem $\tau \in \mathcal{P}$ kannn ein Vorzeichen zugeordnet werden: unterscheiden sich $\tau(i)$ und $\tau(j)$ durch nur eine Transposition, so ist $\tau(j) = -1\tau(i)$. Benötigt man m Transpositionen, um $\tau(i)$ in die Folge $\{1, 2, \ldots, n\}$ zu transponieren, so ist das Vorzeichen von $\tau(i)$ durch $\text{sgn}(\tau) = (-1)^m$ gegeben. Für jede Permutation τ existiert eine Inverse τ^{-1} derart, dass $\tau^{-1}\tau = \{1, 2, \ldots, n\}$; τ^{-1} ist die auf τ angewandte Permutation, die die Permutaiton τ wieder in die "natürliche" Anordnung $1, 2, \ldots, n$ überführt. Die Determinante $|A|$ einer (n, n)-Matrix $A = (a_{ij})$ ist dann durch

$$|A| = \sum_{\tau \in \mathcal{P}} \text{sgn}(\tau) a_{\tau(1),1} a_{\tau(2),2} \cdots a_{\tau(n),n} \tag{6.99}$$

gegeben (s. a. Leibnizregel und Laplacescher Entwicklungssatz weiter unten).

Folgerungen:

1. *Die Matrix B gehe aus der Matrix A durch Vertauschung zweier Spalten hervor. Dann gilt $|B| = -|A|$.*

   Die Aussage ergibt sich aus der sukzessiven Anwendung der definierenden Eigenschaften der Determinante, insbesondere
   a. Addition des Vektors an j-ter Stelle zum Vektor an der k-ten Stelle,
   b. Subtraktion des neuen Vektors an der k-ten Stelle vom Vektor an der j-ten Stelle,
   c. Addition des Vektors an der j-ten Stelle zum Vektor an der k-ten Stelle,
   d. Ersetzung des Vektors an der j-ten Stelle durch sein Negatives.

2. *Es sei $A = [\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n]$. Dann gilt*

   $$|[\lambda_1\mathbf{a}_1, \lambda_2\mathbf{a}_2, \ldots, \lambda_n\mathbf{a}_n]| = \lambda_1\lambda_2\cdots\lambda_n|A|. \tag{6.100}$$

   Für $\lambda_1, \cdots = \lambda_n = \lambda$ ist dann

   $$|[\lambda\mathbf{a}_1, \lambda\mathbf{a}_2, \ldots, \lambda\mathbf{a}_n]| = |\lambda A| = \lambda^n|A| \tag{6.101}$$

   Die Aussagen folgen sofort aus Definition 6.8, 1. (ii).

3. *Die Matrix B entstehe aus A, indem der Vektor $\boldsymbol{a}_j$ durch $\boldsymbol{a}_j + \lambda \boldsymbol{a}_k$ ersetzt wird, wobei $\boldsymbol{a}_k$ ebenfalls ein Vektor von A ist und $\lambda \in \mathbb{R}$. Dann gilt $|A| = |B|$.*

   Denn

   $$\begin{aligned} \det(\mathbf{a}_1, \cdots, \mathbf{a}_j + \lambda \mathbf{a}_k, \ldots, \mathbf{a}_k, \ldots, \mathbf{a}_n) &= |A| + \lambda \det(\mathbf{a}_1, \ldots, \mathbf{a}_k, \ldots, \mathbf{a}_k, \ldots, \mathbf{a}_n) \\ &= |A|, \end{aligned}$$

   denn $\det(\mathbf{a}_1, \ldots, \mathbf{a}_k, \ldots, \mathbf{a}_k, \ldots, \mathbf{a}_n) = 0$ wegen Folgerung 1

4. *$|A| = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$ verändert ihren Wert nicht, wenn einer der Vektoren $\boldsymbol{a}_j$ durch eine Linearkombination der übrigen ersetzt wird.*

   Diese Folgerung ist eine 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 gilt $|B| = \lambda |A| = 0 \cdot |A| = 0$.

6. Es sei $A = \text{diag}(d_1, d_2, \ldots, d_n)$ eine Diagonalmatrix, d.h. es sei $a_{ij} = 0$ für alle $i \neq j$ und $a_{ii} = d_i$, $i = 1, \ldots, n$. Dann gilt

   $$|A| = \prod_{i=1}^{n} d_i. \tag{6.102}$$

   Denn nach (6.99) 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 $A = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$ und die $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ seien linear abhängig. Dann gilt $|A| = 0$.*

   Es sei $\mathbf{a}_j$ eine Linearkombination der übrigen $\mathbf{a}_k$, 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 $C = [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n]$ und die $\boldsymbol{c}_j$ seien linear abhängig, so dass nach Folgerung 7 $|C| = 0$ folgt. Es sei $\mathcal{L}_C = \mathcal{L}(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n)$ die lineare Hülle der $\boldsymbol{c}_j$. Die Vektoren $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ seien Elemente aus $\mathcal{L}_C$ und es sei $A = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$. Dann gilt $|A| = 0$.*

    Die $\mathbf{a}_j$ sind Linearkombinationen der $\mathbf{c}_j$, und da die $\mathbf{c}_j$ linear abhängig sind, sind auch die $\mathbf{a}_j$ linear abhängig, so dass nach Folgerung 7 $|A| = 0$ sein muss.

11. *Es sei A eine (n, n)-Matrix und A' sei die Transponierte von A. Dann gilt*

$$|A'| = |A|. \tag{6.103}$$

    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*

$$|AB| = |A||B|. \tag{6.104}$$

    Es sei $C = AB$; der j-te Spaltenvektor $\mathbf{c}_j$ von C ist dann durch $\mathbf{c}_j = A\mathbf{b}_j$ gegeben, $\mathbf{b}_j$ der j-te Spaltenvektor von B, $j = 1, \ldots, n$. Dann ist

$$|C| = |[\sum_{j=1}^n b_{j1}\mathbf{a}_j, \sum_{j=1}^n b_{j2}\mathbf{a}_j, \ldots, \sum_{j=1}^n b_{jn}\mathbf{a}_j]|.$$

Die Aussage (6.104) folgt dann aus der wiederholten Anwendung der Eigenschaft 1. (i) der Definition 6.8 und der Folgerung 6.100.

Beispiel 6.6 Alternativer Beweis (II) von $TT' = I$, Satz 3.1, Seite 81: Nach (6.104) gilt $|AB| = |A||B|$, und nach (6.119) 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*

$$|A^{-1}| = \frac{1}{|A|}. \tag{6.105}$$

Da A vollen Rang hat existiert die Inverse A^{-1}, so dass $A^{-1}A = I$, I die Einheitsmatrix. (6.104) impliziert dann $|A^{-1}A| = |A^{-1}||A| = |I| = 1$ (s. Definition 6.8, 3.). (6.105) folgt unmittelbar.

14. **Cramersche Regel**[99] 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*,

$$x_j = \frac{|A_j(\mathbf{b})|}{|A|}. \tag{6.106}$$

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

$$\begin{vmatrix} x_1 & 0 \\ x_2 & 1 \end{vmatrix} = x_1 \cdot 0 - 0 \cdot x_2 = x_1, \qquad \begin{vmatrix} 1 & x_2 \\ 0 & x_2 \end{vmatrix} = 1 \cdot x_2 - 0 \cdot x_1 = x_2.$$

Den Fall $n = 3$ illustriert man anhand der Sarrusschen Regel (s. unten), und für $n > 3$ kann man den Laplaceschen Entwicklungssatz (s. unten) heranziehen.

Weiter ist

$$A_j(\mathbf{b}) = [\mathbf{a}_1, \ldots, \mathbf{a}_{j-1}, \mathbf{b}, \mathbf{a}_{j+1}, \ldots, \mathbf{a}_n],$$

und weiter ist $A^{-1}\mathbf{a}_k = \mathbf{e}_k$ der k-te Einheitsvektor (wegen $A^{-1}A = I$). Dann folgt

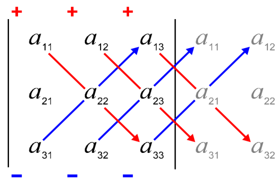
$$A^{-1}A_j(\mathbf{b}) = [\mathbf{e}_1, \ldots, \mathbf{e}_{j-1}, A^{-1}\mathbf{b}, \mathbf{e}_{j+1}, \ldots, \mathbf{e}_n] = I_j(\mathbf{b}) = x_j,$$

und

$$|A^{-1}A_j(\mathbf{b})| = |A^{-1}||A_j(\mathbf{b})| = x_j,$$

wegen (6.104), und wegen (6.105) folgt (6.106).

Abbildung 30: Zur Regel von Sarrus



15. **Regel von Sarrus** Es sei also A eine $(3,3)$-Matrix. Sarrus[100] 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 30. Dann verbindet man die Elemente a_{12}, a_{23} und a_{31}: sie bilden das zweite Produkt, und schließlich noch die Elemente a_{13}, a_{21} und a_{32}, sie bilden das dritte Produkt. Die ersten drei Produkte werden addiert. Dann folgen drei Produkte mit negativem Vorzeichen: man verbindet die Elemente von a_{13} bis a_{32}, deren Produkt das erste mit negativem Vorzeichen ist, dann folgt das Produkt $a_{11}a_{23}a_{12}$ und schließlich das Produkt $a_{12}a_{21}a_{33}$, vergl. Abbildung[101] 30. Für

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

hat man also

$$|A| = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{12} - a_{12}a_{21}a_{33}. \tag{6.107}$$

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 (6.107) gegebene Ausdruck für* $\det(A)$ *ist durch eine Summe von Produkten aus jeweils n Elementen von A gegeben; der allgemeine Ausdruck ist nach (6.99)*

$$|A| = \sum_{\tau \in \mathcal{P}} sgn(\tau) a_{\tau(1),1} a_{\tau(2),2} \cdots a_{\tau(n),n}$$

τ ist eine der $n!$ Anordnungen (Permutationen) von n Elementen aus A, $\tau(k) = j$ ist eine der Zahlen $j \in \{1, 2, \ldots, n\}$. Die Anordnung der Elemente

[99] Gabriel Cramer (1704 – 1752), Genfer Mathematiker
[100] Pierre Frédéric Sarrus (1798 – 1861), französischer Mathematiker
[101] kopiert aus dem Wikipedia-Eintrag 'Regel von Sarrus'

innerhalb eines Produktes in (6.107) ist eine Konsequenz der Sarusschen Regel; da es auf die Anordnung der Elemente innerhalb eines Produktes nicht ankommt, kann diese so anordnen, dass die Komponente von $\boldsymbol{a}_1$ an erster, die von $\boldsymbol{a}_2$ an zweiter und die von $\boldsymbol{a}_3$ an dritter Stelle steht:

$$\begin{aligned} |A| &= a_{11}a_{22}a_{33} + a_{31}a_{12}a_{23} + a_{21}a_{32}a_{13} \\ &\quad -a_{31}a_{22}a_{13} - a_{11}a_{12}a_{23} - a_{21}a_{12}a_{33} \end{aligned} \tag{6.108}$$

Diese Anordnung entpricht der "kanonischen" Anordnung in (6.99), wie sie oben noch einmal angegeben wurde: die zweiten Indices zeigen immer die Reihenfolge 1, 2, 3.

Inspiziert man die Produkte, so fällt auf, dass die jeweils erste Zahl in einem Produkt, etwa a_{11}, die folgenden Zahlen so bestimmt, dass sie aus einer Restmatrix A_{11} kommen, die entsteht, wenn man die dem ersten Index entsprechende Zeile und die dem zweiten Indexes entpsrechhende Spalte aus A streicht. Für a_{11} erhält man die Restmatrix

$$A_{11} = \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix}, \quad |A_{11}| = a_{22}a_{33} - a_{23}a_{32} \tag{6.109}$$

Man sieht in (6.108), 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 (6.108) läßt sich also zu

$$a_{11}a_{22}a_{33} - a_{11}a_{12}a_{23} = a_{11}|A_{11}| \tag{6.110}$$

zusammenfassen. Die zweite Komponente a_{21} von $\mathbf{a}_1$ taucht in (6.108) ebenfalls in zwei Summanden (Produkten) auf, und zwar als Differenz

$$a_{21}a_{32}a_{13} - a_{21}a_{12}a_{33} = -a_{21}|A_{21}|, \tag{6.111}$$

wobei jetzt

$$A_{21} = \begin{pmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{pmatrix}, \quad |A_{21}| = a_{12}a_{33} - a_{13}a_{32} \tag{6.112}$$

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 (6.108) für $|A|$. Für diese Differenz wird die dritte Zeile und wieder die erste Spalte aus A gestrichen; man erhält

$$A_{31} = \begin{pmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{pmatrix}, \quad |A_{31}| = a_{12}a_{23} - a_{13}a_{23}, \tag{6.113}$$

so dass

$$a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13} = a_{31}|A_{31}|. \tag{6.114}$$

Hier ist das Vorzeichen wieder positiv, weit 1 + 3 eine gerade Zahl ist. Die Determinante $|A|$ ist die Summe der Differenzen (6.110), (6.111) und (6.114):

$$\begin{aligned} |A| &= a_{11}|A_{11}| - a_{21}|A_{21}| + a_{31}|A_{31}| \\ &= (-1)^{1+1}a_{11}|A_{11}| + (-1)^{2+1}a_{21}|A_{21}| + (-1)^{3+1}a_{31}|A_{31}| \\ &= \sum_{k=1}^{3}(-1)^{k+1}a_{k1}|A_{k1}| \end{aligned} \tag{6.115}$$

(6.115) ist die *Entwicklung von $|A|$ nach der ersten Spalte von A*, eine andere Bezeichnung ist *Laplacescher Entwicklungssatz*[102]. 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
$$y_{ij} = \frac{|A_j(\mathbf{e}_i)|}{|A|} \tag{6.116}$$
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
$$|A_j(\mathbf{e}_i)| = (-1)^{i+j}|A_{ij}| := \alpha_{ij}. \tag{6.117}$$
$|A_{ij}|$ heißt der *Kofaktor* oder *Minor* des Elements a_{ij} von A. Die Matrix $\mathrm{Adj}(A)$ der α_{ij}-Werte heißt *Adjunkte* der Matrix A, und (6.116) impliziert dann
$$A^{-1} = \frac{1}{|A|}\mathrm{Adj}(A). \tag{6.118}$$
Zu zeigen ist (6.117). $A_j(\mathbf{e}_i)$ entsteht, wenn die j-te Spalte von A durch en Einheitsvektor $\mathbf{e}_i$ ersetzt wird. Man kann die Determinante $|A_j(\mathbf{e}_i)|$ bestimmen, indem man sie nach der j-ten Spalte entwickelt:
$$|A_j(\mathbf{e}_i)| = \sum_{k=1}^{n}(-1)^{k+j}e_{ij}|A_{ij}|.$$
Aber $e_{ij} = 0$ für alle $i \neq k$, so dass nur ein Summand – der für $k = i$ – übrig bleibt, also hat man (6.117).

[102]Pierre Simon Laplace (1749 – 1827), französischer Mathematiker

18. *Es sei A eine orthonormale (n, n)-Matrix mit vollem Rang. Dann gilt*

$$|A| = 1. \tag{6.119}$$

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 $A = diag(a_{11}, a_{22}, \ldots, a_{nn})$ eine Diagonalmatrix. Dann gilt*

$$|A| = |\text{diag}(a_{11}, a_{22}, \ldots, a_{nn})| = \prod_{i=1}^{n} a_{ii}, \tag{6.120}$$

A kann in der Form $A = [a_{11}\mathbf{e}_1, a_{22}\mathbf{e}_2, \ldots, a_{nn}\mathbf{e}_n]$ geschrieben werden. Nach Folgerung 2, Gleichung (6.100) 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

$$|X^{-1}MX| = |N|. \tag{6.121}$$

Insbesondere gilt

$$|N| = \prod_{i=1}^{n} \lambda_i, \tag{6.122}$$

wobei $\lambda_1, \ldots, \lambda_n$ die (positiven) Eigenwerte von M sind.

Nach dem Produktsatz und nach Folgerung 14, Gleichung (6.105) gilt

$$|X^{-1}MX = |X^{-1}M||X| = |X^{-1}||M||X| = \frac{1}{|X|}|M||X| = |M| = |N|.$$

Insbesondere sei $X = P$, P die Matrix der (orthognormalen) Eigenvektoren von M. Dann gilt $P'MP = \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$, λ_j, $j = 1, \ldots, n$ die zu den Eigenvektoren korrespondierenden Eigenwerte von M. Man hat dann $MP = P\Lambda$ und der Produktsatz liefert

$$|PM| = |P||M| = |P||\Lambda|.$$

Da P aber orthonormal ist folgt $|P| = 1$, so dass $|M| = |\Lambda| = \prod_{i=1}^{n} \lambda_i$ nach Folgerung 6.128, Gleichnung (6.120).

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

$$|A - \lambda I| = \lambda^n - a_1\lambda^{n-1} + a_2\lambda^{n-2} + \cdots + (-1)^n a_n = 0 \tag{6.123}$$

heißt *charakteristische Gleichung* von A, das Polynom auf der rechten Seite heißt *charakteristisches Polynom* von A. Dabei gilt

$$a_1 = \sum_{i=1}^{n} a_{ii}, \quad a_n = |A|, \tag{6.124}$$

d.h. a_1 ist die Spur von A. Die Nullstellen des Polynoms sind gleich den Eigenwerten von A.

Das charakteristische Polynom kann z.B. durch die Anwendung des Laplaceschen Entwicklungssatzes auf $|A - \lambda I|$ hergeleitet werden; die Koeffizienten a_i ergeben sich aus den Elementen der Matrix $A - \lambda I$. Die Herleitung soll hier übergangen werden. Die Bestimmung der Eigenwerte und der Eigenvektoren wird im Allgemeinen von Programmen übernommen (Golub & van Loan (2013)).

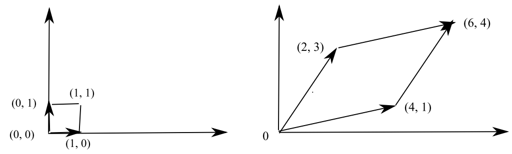
6.13.1 Transformationen und Volumen

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 31. Das Quadrat ist durch die Einheitsvektoren $\mathbf{e}_1 = (1,0)'$ und $\mathbf{e}_2 = (1,0)'$ definiert. Die Matrix A ist durch

$$A\mathbf{x} = \mathbf{y}, \quad A = [\mathbf{a}_1, \mathbf{a}_2] = \begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix} \tag{6.125}$$

Abbildung 31: 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

$$A\mathbf{e}_1 = \mathbf{a}_1 = \begin{pmatrix} 4 \\ 1 \end{pmatrix}, \quad A\mathbf{e}_2 = \mathbf{a}_2 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \tag{6.126}$$

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.

Der Flächeninhalt des Parallelogramms ergibt sich aus der bekannten Formel $F = hg$, wobei h die Höhe des Paralelogramms und g die Länge des Vektors $\mathbf{a}_2 = (2,3)'$ ist. h ist die Länge des Vektors $\mathbf{z}$ in Abbildung 26, Seite 178, und nach Gleichung (1.73), Seite 30, läßt sich $h = \|\mathbf{z}\|$ ausrechnen. Um aber die Beziehung zwischen dem Flächeninhalt des Parallelogramms und der Determinnante zu illustrieren, soll der Flächeninhalt auf andere Weise bestimmt werden. Um die dazu benötigte Grafik (Abb. 32) nicht zu unübersichtlich werden zu lassen sollen Doppelindices bei Vektoromponenten vermieden werden, weshalb die Spalten der Matrix A umbenannt werden: es gelte jetzt $A = [\mathbf{u}, \mathbf{v}]$, mit $\mathbf{u} = (u_1, u_2)'$, $\mathbf{v} = (v_1, v_2)'$.

Satz 6.7 *Es ist det$(A) = u_1 v_2 - u_2 v_1$. Der Flächeninhalt F_p des durch die Vektoren $\boldsymbol{u}$ und $\boldsymbol{v}$ definierten Parallelogramms (s. Abb. 32) ist durch*

$$F_p = u_1 v_2 - u_2 v_1 \tag{6.127}$$

gegeben.

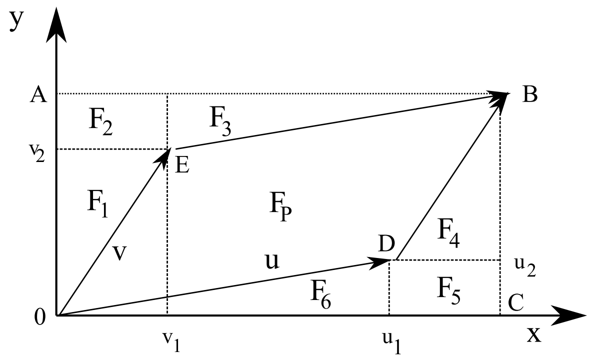
Beweis: Offenbar ist F_P gleich der Fläche $F_{tot} = (u_1 + v_1)(u_2 + v_2)$ des Rechtecks 0ABC minus der Summe der Flächen F_1 bis F_6. Es ist

$$F_1 = F_4 = \frac{1}{2} v_1 v_2 \tag{6.128}$$

$$F_2 = F_5 = v_1 u_2 \tag{6.129}$$

$$F_3 = F_6 = \frac{1}{2} u_1 u_2 \tag{6.130}$$

Abbildung 32: Durch die Vektoren $\mathbf{a}_1 = (u_1, u_2)'$, und $\mathbf{a}_2 = (v_1, v_2)'$ definiertes Parallelogramm



Einsetzen der Ausdrücke für F_1 bis F_6 liefert sofort

$$F_p = F_{tot} - (F_1 + \cdots + F_6) = |u_1 v_2 - v_1 u_2| = |det(A)| \tag{6.131}$$

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

$$|B| = -(u_1 v_2 - v_1 u_2) = -det(A)$$

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 31 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

$$A a\mathbf{e}_1 = a\mathbf{a}_1 = \begin{pmatrix} 4a \\ 1a \end{pmatrix}, \quad A a\mathbf{e}_2 = a\mathbf{a}_2 = \begin{pmatrix} 2a \\ 3a \end{pmatrix}, \tag{6.132}$$

abgebildet werden; es entsteht ein Parallelogramm, dessen Seitenlänge und Diagonale um den Faktor a verändert werden. Der Flächeninhalt des Parallelogramms beträgt nun nach (6.131) (die Komponenten der Vektoren $\mathbf{u}$ und $\mathbf{v}$ werden alle mit dem Faktor a multipliziert)

$$F_p = F_{\mathcal{G}^*} = a^2|u_1v_2 - v_1u_2| = F_G|A|. \tag{6.133}$$

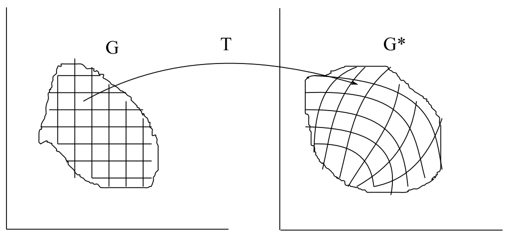
Für die Determinante $\det(A) = |A|$ folgt daraus sofort

$$det(A)| = \frac{F_{\mathcal{G}^*}}{F_G}, \tag{6.134}$$

d.h. im allgemeinen Fall ist $|A|$ gleich dem *Verhältnis* der Flächeninhalte bzw. im mehr als 2-dimensionalen Fall der Volumina der Gebiete $\mathcal{G}^*$ und F_G. Die Redeweise, dass $|A|$ gleich dem Volumen eines Parallelepipeds ist, impliziert den Fall $F_G = 1$.

Das Gebiet G ist aber nicht notwendig ein Quadrat, sondern kann irgendein zusammenhängedes Gebiet sein (Abbildung 33). T in Abb. 33 kann irgendeine Transformation sein, die lineare Transformation, wie sie durch eine (n, n)-Matrix

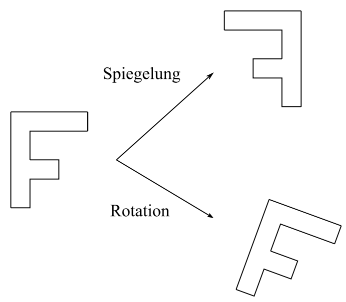
Abbildung 33: Allgemeine Gebietstransformation



gegeben ist, ist ein Spezialfall. Das Volumen des Gebiets G kann abgechätzt werden, indem man G in Quadrate unterteilt und die Summe der Fläche dieser Quadrate bildet. Die Approximation ist um so genauer, die kleiner die Quadrate und je größer ihre Anzahl. Im Grenzfall geht man zu infinitesimalen Flächen über und bestimmt die Fläche (das Volumen) von G durch ein Gebietsintegral. Analog dazu verfährt man mit der Fläche bzw. dem Volumen von $\mathcal{G}^* = T(G)$. Die zu den Quadraten in G korrespondierenden Flächen in $\mathcal{G}^*$ können im Limit durch infinitesímale Quadrate angenähert werden, die eine lineare Transformation A der Quadrate in G sind, d.h. A ist eine Matrix, und man findet analog zu (6.133)

$$\text{vol}(G) = det(A)|\text{vol}(\mathcal{G}^*) \tag{6.135}$$

Abbildung 34: Transformation: Spiegelung (Änderung der Orientierung) und Rotation



oder

$$det(A)| = \frac{\text{vol}(G)}{\text{vol}(\mathcal{G}^*)}. \tag{6.136}$$

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.

6.13.2 Das Vektorprodukt, Transformationen und Orientierung

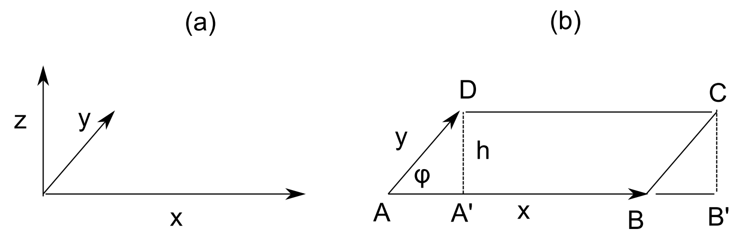
Zum Abschluß soll noch kurz auf die Bedeutung des Vorzeichens einer Determinante eingegangen werden. Dazu werden die beiden Matrizen[103]

$$A = \begin{pmatrix} -1 & -\frac{1}{4} \\ 1 & -\frac{1}{2} \end{pmatrix}, \quad B = \begin{pmatrix} -\frac{1}{4} & 1 \\ 1 & \frac{1}{2} \end{pmatrix}. \tag{6.137}$$

Man rechnet leicht nach, dass $|A| = 3/4$ und $|B| = -9/8$. Die Determinanten sind ungleich Null, weshalb die Spaltenvektoren $\mathbf{a}_1$ und $\mathbf{a}_2$ von A linear unabhängig sind, ebenso die Spaltenvektoren $\mathbf{b}_1$ und $\mathbf{b}_2$ von B, d.h. sowohl $\{\mathbf{a}_1, \mathbf{a}_2\}$ wie $\{\mathbf{b}_1, \mathbf{b}_2\}$ sind Basen des $\mathbb{R}^2$, also kann man mit beiden Basen alle Vektoren des $\mathbb{R}^2$ erzeugen. Wie gezeigt wurde, kann man $|A|$ als Volumen bzw. als Quotien von Volumina deuten, aber $|B|$ nicht, da Volumina nicht negativ sein können. Allgemein gilt für Matrizen, dass sie Vektoren transformieren, d.h. wenn $A\mathbf{x} = \mathbf{y}$ gilt, so transformiert A den Vektor $\mathbf{x}$ in den Vektor $\mathbf{y}$, und natürlich gilt die analoge Aussage für B. Die Transformation bewirkt allgemein eine Rotation und eine Veränderung der Länge von $\mathbf{x}$, oder eines von beiden (und in speziellen Fällen gar keine Veränderung). Nun können die Vektoren $\mathbf{x}_1, \ldots, \mathbf{x}_p$ eine Figur abbilden, eben z.B. ein F. Die Matrix A rotiert alle $\mathbf{x}_j$ und verändert die Längen, und die Figur, hier das F, wird als Ganzes "nur" rotiert. Wendet man dagegen

[103]Fischer (1997), p. 203)

Abbildung 35: Durch $\mathbf{x}$ und $\mathbf{y}$ definiertes Vektorprodukt (a), korrespondierendes Parallelogramm (b)



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 34). 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.

Das Vektorprodukt Der Begriff der Orientierung kann mithilfe des *Vektorprodukts* illustriert werden, das nur 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 spannen eine Ebene im $\mathbb{R}^3$ auf), und $\mathbf{z}$ stehe senkrecht auf dieser Ebene (Abb. 35, (a)). Man stelle sich vor, der Vektor $\mathbf{x}$ werde in dieser Ebene so rotiert, dass er dieselbe Orientierung wie $\mathbf{y}$ hat, und dass der Vektor $\mathbf{z}$ sich dabei sozusagen um seine Achse rotiert. Weiter sei $\mathbf{n}$ ein auf $\mathbf{x}$ und $\mathbf{y}$ senkrecht stehender (Normalen-)Vektor der Länge 1; $\mathbf{n}$ und $\mathbf{z}$ haben dieselbe Orientierung im $\mathbb{R}^3$.

Definition 6.9 *Das Vektorprodukt* $\boldsymbol{x} \times \boldsymbol{y}$ *ist durch die Beziehung*

$$\boldsymbol{z} = \boldsymbol{x} \times \boldsymbol{y} = (\|\boldsymbol{x}\| \|by\| \sin\varphi)\boldsymbol{n} \tag{6.138}$$

definiert.

Offenbar ist die Länge von $\mathbf{z}$ durch

$$\|\mathbf{z}\| = \|\mathbf{x}\| \|by\| \sin\varphi \tag{6.139}$$

gegeben und gleich der Fläche F des durch $\mathbf{x}$ und $\mathbf{y}$ definierten Paralellogramms ABCD, s. Abb. 35, (b), denn $\sin\varphi = h/\|\mathbf{y}\|$, so dass $h = \|\mathbf{y}\| \sin\varphi$ und $F = h\|\mathbf{x}\|$. Die durch $\mathbf{z}$ gehende, senkrecht auf der drch $\mathbf{x}$ und $\mathbf{y}$ definierten Ebene im $\mathbb{R}^3$ ist gewissermaßen eine Rotationsachse, wenn $\mathbf{x}$ in die Richtung von $\mathbf{y}$ rotiert wird, weshalb $\mathbf{z}$ auch *Axialvektor* genannt wird.

Ohne Beschränkung der Allgemeinheit kann man annehmen, dass diese Ebene durch die kanonischen Einheitsvektoren $\mathbf{e}_1$ und $\mathbf{e}_2$ definiert wird; dann steht $\mathbf{e}_3$

senkrecht auf dieser Ebene, hat also dieselbe Orientierung wie $\mathbf{z}$. Das Vektorprodukt ist dann definiert durch

$$\mathbf{x} \times \mathbf{y} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} \tag{6.140}$$

Entwickelt man diese Determinante nach der ersten Zeile, so erhält man

$$\begin{aligned} \mathbf{z} = \mathbf{x} \times \mathbf{y} &= \mathbf{e}_1 \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} - \mathbf{e}_2 \begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix} + \mathbf{e}_3 \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \\ &= (x_2 y_3 - x_3 y_2)\mathbf{e}_1 + (x_3 y_1 - x_1 y_3)\mathbf{e}_2 + (x_1 y_2 - x_2 y_1)\mathbf{e}_3 \\ &= \begin{pmatrix} x_2 y_3 - x_3 y_2 \\ x_3 y_1 - x_1 y_3 \\ x_1 y_2 - x_2 y_1 \end{pmatrix} \end{aligned} \tag{6.141}$$

$\mathbf{x} \times \mathbf{y}$ ist ein Vektor, der orthogonal zu den Vektoren $\mathbf{x}$ und $\mathbf{y}$ ist; so ist

$$\mathbf{x}'\mathbf{z} = \mathbf{x}'(\mathbf{x} \times \mathbf{y}) = x_1 x_2 y_3 - x_1 x_3 y_2 + x_2 x_3 y_1 - x_1 x_2 y_3 + x_3 x_1 y_2 - x_3 x_2 y_1 = 0,$$

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

$$\begin{aligned} \mathbf{a} \times \mathbf{b} &= -(\mathbf{a} \times \mathbf{b}) &(6.142)\\ \mathbf{a} \times (\mathbf{b} + \mathbf{c}) &= \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c} &(6.143)\\ \mathbf{a} \times (\mathbf{b} \times \mathbf{c}) &\neq (\mathbf{a} \times \mathbf{b}) \times \mathbf{x} &(6.144)\\ p(\mathbf{a} \times \mathbf{b}) &= (p\mathbf{a}) \times \mathbf{b} = \mathbf{a} \times (p\mathbf{b}). &(6.145) \end{aligned}$$

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 Schraubenbewegung; $-\mathbf{z}$ entpricht dann der Orientierung der Drehung beim Herausschrauben der Schraube.

Um den Begriff der Orientierung formal fassen zu können, wird der Begriff der *geordneten Basis* eingeführt; dies ist eine Basis $\mathbf{b}_1, \ldots, \mathbf{b}_n$, bei der es auf die Anordnung der $\mathbf{b}_j$ ankommt. Nun seien $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ und $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ zwei geordnete Basen eine n-dimensionalen Vektorraumes V. Dann existiert stets eine Matrix A, die die Basis $\mathcal{B}$ in die Basis $\mathcal{C}$ überführt[104], so dass mit $B = [\mathbf{b}_1, \ldots, \mathbf{b}_n]$ und $C = [\mathbf{c}_1, \ldots, \mathbf{c}_n]$ die Beziehung $C = AB$ gilt. Dann heißen die

[104]Warum?

Basen $\mathcal{B}$ und $\mathcal{C}$ *gleichorientiert*, wenn $|A| > 0$ ist. Gilt diese Bedingung *nicht*, so heißen sie *verschieden orientiert*; dann gilt $|A| < 0$.

Vertauscht man in einer Matrix zwei Vektoren miteinander, so ändert sich das Vorzeichen der Determinante. Hat man also $\mathcal{B}_1 = \{\mathbf{b}_1, \ldots, \mathbf{b}_j, \mathbf{b}_{j+1}, \ldots, \mathbf{b}_n\}$ und $\mathcal{B}_2 = \{\mathbf{b}_1, \ldots, \mathbf{b}_{j+1}, \mathbf{b}_j, \ldots, \mathbf{b}_n\}$, so sind $\mathcal{L}(\mathcal{B}_1$ und $\mathcal{L}(\mathcal{B}_2)$ verschieden orientierte Vektorräume. insbesondere heißt eine geordnete Basis *positiv orientiert*, wenn sie dieselbe Orientierung wie die kanonische Basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ hat; andernfalls heißt sie *negativ orientiert*.

6.13.3 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 6.10 *Eine Funktion $f(x)$ mit einer Variablen heißt differenzierbar in x, wenn eine Zahl $f'(\xi)$ existiert derart, dass*

$$f(x+\xi) = f(x) + f'(x)\xi + o(\xi), \quad \lim_{\xi \to 0} \frac{o(\xi)}{\|\xi\|} = 0. \tag{6.146}$$

gilt. $o(\xi)$ ist eine Funktion von ξ, die schneller gegen Null geht als ξ. Dabei ist

$$f'(x) = \lim_{\xi \to 0} \frac{f(x+\xi) - f(x)}{\xi} \tag{6.147}$$

Der mehrdimensionale Fall ist analog definiert:

Definition 6.11 *Es sei $U \subset \mathbb{R}^n$ eine offene Menge und es sei $\boldsymbol{f}$ eine Abbildung $\boldsymbol{f}$: $U \to \mathbb{R}^m$. in $\boldsymbol{x} \in U$* total differenzierbar *(auch einfach* differenzierbar*), wenn eine weitere lineare Abbildung A: $\mathbb{R}^n \to \mathbb{R}^m$ gibt derart, dass in einer Umgebung von $\boldsymbol{x}$*

$$\boldsymbol{f}(\boldsymbol{x}+\boldsymbol{\xi}) = \boldsymbol{f}(\boldsymbol{x}) + A\boldsymbol{\xi} + o(\|\boldsymbol{\xi}\|) \tag{6.148}$$

gilt.

Hier ist $A = (a_{ij})$ eine (m,n)-Matrix und $\boldsymbol{\xi}$ ein n-dimensionaler Vektor; weiter ist $\mathbf{f} = (f_1, \ldots, f_n)'$ ein n-dimensionaler Vektor. Offenbar ist differenzierbar, wenn die Komponenten f_i differenzierbar sind.

Satz 6.8 *Es seien $\boldsymbol{f}$ und A wie in Definition 6.11 definiert. Dann gilt (i) $\boldsymbol{f}$ ist in $\boldsymbol{x}$ stetig, (ii) die Komponenten f_i von $\boldsymbol{f}$ sind in $\boldsymbol{x}$ partiell differenzierbar, d.h. es gilt*

$$\frac{\partial f_i}{\partial x_j} = a_{ij}. \tag{6.149}$$

Beweis: (vergl. Forster (2), p. 78) Es ist

$$\lim_{\boldsymbol{\xi}\to\vec{0}} A\boldsymbol{\xi} = \mathbf{f}(\boldsymbol{\xi}),$$

und dies heißt, dass $\mathbf{f}$ in $\mathbf{x}$ stetig ist. Weiter gilt

$$f_i(\mathbf{x}+\boldsymbol{\xi}) = f_i(\mathbf{x}) + \sum_{i=1}^{n} a_{ij}\xi_j + o(\|x_i\|), \quad i = 1, \dots n$$

d.h.

$$f_i(\boldsymbol{\xi} + h\mathbf{e}_j) = f_i(\mathbf{x}) + ha_{ij} + o(\|h\mathbf{e}_j\|),$$

so dass

$$\frac{\partial f_i}{\partial x_j} = \lim_{h\to 0} \frac{f_i(\mathbf{x}+h\mathbf{e}_j) - f_i(\mathbf{x})}{h} = a_{ij} + o(\|h\mathbf{e}_j\|) = a_{ij}.$$

□

Die Implikation des Satzes ist, dass die Matrix A durch die Abbildung $\mathbf{f}$ eindeutig bestimmt ist.

Definition 6.12 *Die Matrix A heißt das* Differential *oder* Funktionalmatrix *oder* Jacobi-Matrix[105] *der Abbildung* $\boldsymbol{f}$ *im Punkt* $\boldsymbol{x}$.

Im mehrdimensionalen Fall wird also der Differentialquotient durch eine Matrix von partiellen Differentialquotienten ersetzt. Man hat die Schreibweisen

$$(D\mathbf{f})(\mathbf{x}) := J_{\mathbf{f}} := \frac{\partial(f_1, \dots, f_n)}{\partial(x_1, \dots, x_n)}(\mathbf{x}) := \left(\frac{\partial f_i}{\partial x_j}\right), \tag{6.150}$$

wobei ":=" bedeutet, dass der Ausdruck links von diesem Zeichen durch den rechts davon stehenden definiert wird. In ausgeschriebener Form hat die Funktional- oder Jacobi-Matrix die Gestalt

$$J_{\mathbf{f}} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix} \tag{6.151}$$

Die Differenzierbarkeit von bedeutet nach (6.148), 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 6.9 *Es seine* $U \subset \mathbb{R}^n$ *und* $V \subset \mathbb{R}^m$ *offene Mengen und* $\boldsymbol{f}\colon U \to \mathbb{R}^m$, $\boldsymbol{g}\colon V \to \mathbb{R}^k$ *offene Mengen und es gelte* $f\colon U \to \mathbb{R}^m$, $g : V \to \mathbb{R}^k$, $(\mathbf{U}) \subset \mathbf{V}$. $\boldsymbol{f}$ *sei in* $\boldsymbol{x}$ *und* $\boldsymbol{g}$ *sei in* $\boldsymbol{y} := \boldsymbol{f}(\boldsymbol{x})$ *differenzierbar. Es sei* $\boldsymbol{g} \circ \boldsymbol{f}\colon U \to \mathbb{R}^k$ *die aus* $\boldsymbol{g}$ *und* $\boldsymbol{f}$ *zusammengesetzte Abbildung; sie sei in* $\boldsymbol{x}$ *differenzierbar und es gilt*

$$D(\boldsymbol{g} \circ \boldsymbol{f})(\boldsymbol{x}) = (D\boldsymbol{g})\boldsymbol{f}(\boldsymbol{x}) \cdot D\boldsymbol{f}(\boldsymbol{x}). \tag{6.152}$$

[105] Carl 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 $F(x) = \int f(x)dx$[106]. So kann F als Verkettung $Fo\varphi = F(\varphi(u))$ gesehen werden. Man kann dann

$$F(\varphi(u)) = G(u). \tag{6.153}$$

iund die Kettenregel führt auf

$$\frac{dF(\varphi(u))}{du} = \frac{dF(\varphi)}{d\varphi}\frac{d\varphi}{du} = f(\varphi)\varphi' = g(u). \tag{6.154}$$

und

$$\begin{aligned} \int_a^b f(\varphi(u))\varphi'(u)du &= F(\varphi(b)) - F(\varphi(a)) \\ &= \int_{\varphi(a)}^{\varphi(b)} f(x)dx, \quad x = \varphi(u) \end{aligned} \tag{6.155}$$

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 (6.155) von rechts nach links: gegeben ist eine zu integrierende Funktion $f(x)$, Die Aufgabe, das Integral zu bestimmen, kann u. U. vereinfacht werden, indem man für x die Funktion $x = \varphi(u)$ und damit die neue Integrationsvariable u einführt. Man bestimmt also

$$G(u) = \int f(\varphi(u))\varphi'(u)du$$

und ersetzt dann u wieder durch x. Dazu muss φ umkehrbar sein, d.h. es muss die inverse Funktion $\psi(x) = u = \varphi^{-1}(x)$ existieren mit $\psi'(x) \neq 0$ für alle $x \in I$. Als *Grundformel* (Courant I, p. 184) erhält man dann

$$\int f(x)dx = \int f(\varphi(u))\varphi'(u)du, \quad u = \psi(x) \tag{6.156}$$

Man bestimmt also das Integral auf der linken Seite der Gleichung, indem man das Integral auf der rechten Seite findet und anschließend u über die Abbildung $u = \psi(x)$ wieder einführt. Es sei angemerkt, dass

$$\varphi' = \frac{1}{\psi'} \tag{6.157}$$

[106]Diese Schreibweise mag etwas lax sein, weil die Laufvariable x nicht vom Argument x in $F(x)$ unterschieden wird, ist aber als möglichst einfache, intuitive Schreibweise zu lesen (vergl. Courant I (1961), p. 182)

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 (6.155) 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 (6.155) erhält man die Beziehung

$$dx = \varphi'(u)du \tag{6.158}$$

Die Transformation impliziert i. A., dass x und u und damit auch dx und du auf verschiedenen Skalen laufen[107] 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 (6.156) 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 6.7 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

$$\frac{du}{dx} = \frac{d\psi(x)}{dx} = \psi'(x) = 2, \quad \text{d.h. } du = 2dx, \text{bzw. } dx = \frac{1}{2}du$$

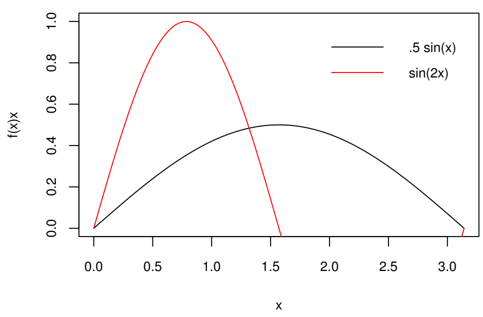
einer Veränderung auf der x-Skala entspricht eine doppelt so große Veränderung auf der u-Skala, und einer Veränderung auf der u-Skala entspricht die Hälfte dieser Verändrung auf der x-Skala. Diese Beziehungen übertragen sich auf die infinitesimalen Größen dx und du. Sind a und b die Grenzen des Integrals auf der u-Skala, so erhält man wegen $u/2 = x$ die Grenzen $a/2$ und $b/2$ auf der x-Skala; speziell für $a = 0$ und $b = \pi$ erhält man also

$$\begin{aligned} \int_0^{\pi/2} \sin(2x)dx &= \int_0^{\pi} \sin(u)\frac{1}{2}du = \frac{1}{2}\int_0^{\pi} \sin(u)du \\ &= -\frac{1}{2}\cos(\pi) + \frac{1}{2}\cos(0) = 1. \end{aligned} \tag{6.159}$$

Abbildung 36 illustriert das Prinzip: Die Funktion $\sin(2x)$ "überdeckt" nur die

[107] Ein einfacher Fall sind die Celsius- und Fahrenheitskalen für die Temperatur, die durch eine lineare Transformation aufeinander bezogen sind: $F = aC + b$ mit $a = 1.8$, $b = 32$. Der Unterschied von einem Grad Celsius entspricht einem Unterschied von 1.8 Grad F. Differenziert man, so erhält man $dF/dC = a$, d.h. $dF = a \cdot dC$. d.h. der Faktor a bezieht auch die Differentiale aufeinander. Da dx und dy infinitesimale *Differenzen* sind, verschwindet die additive Konstante b.

Abbildung 36: Skalen und Fläche unter Funktionen



Hälfte des Intervalls, den die Funktion $.5\sin(u)$ überdeckt, und damit die Flächen unter den Funktionen – also die Integrale über die entsprechenden Intervalle – gleich groß sind, müssen die Funktionswerte von $\sin(u)$ in diesem Fall mit dem Faktor $1/2$ multipliziert werden. Eine intuitive Plausibiltätsbetrachtung mag hier hilfreich sein: Geht man von der Riemannschen Definition des Integrals als Grenzwert einer Summe von Rechtecken aus, so gilt speziell für eine Zerlegung des Integrationsintervalls $[a, b]$ in gleichgroße Intervalle $\Delta x = (b-a)/n$

$$\int f(x)dx = \lim_{\Delta x_i \to 0} \lim_{n\to\infty} \sum_{i=1}^{n} f(t_i)\Delta x, \quad t_i \in [x_i - x_{i-1}) \tag{6.160}$$

d.h. für $\Delta x \to dx$ für $n \to \infty$. Gilt nun $f(t_i)\Delta x = f(u_i)\varphi'\Delta x$, so erhält man im Limit für die Rechtecke $du = \varphi' dx$, und φ' kann als Dehnung oder Verkürzung von Δx oder als Streckung oder Stauchung von $f(t_i)$ interpretiert werden, so dass $f(u)\varphi' du = f(x)dx$ gilt. So haben die Rechtecke von $\sin(u)$ nur die Breite $du = dx/2$, haben also nur den halben Flächeninhalt. Da man den Faktor $1/2$ auch auf die Höhe der Rechtecke beziehen kann, kann man ebensogut sagen, dass die Rechtecke für $\sin(u)$ nur die halbe Höhe der Rechtecke für die Funktion $\sin(2x)$ haben. □

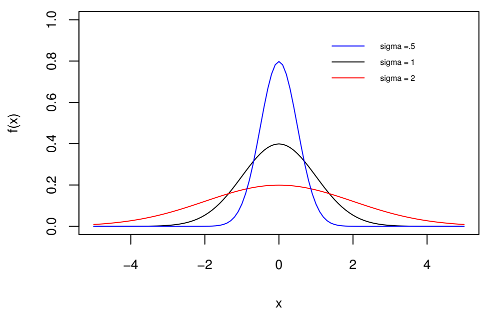
Beispiel 6.8 Es sei Z eine $N(0, 1)$-verteilte zufällige Veränderliche, d.h. Z sei standardnormalverteilt, so dass $\mathbb{E}(Z) = 0$ und $Var(Z) = 1$. Es sei nun $X = \varphi(Z) = \sigma Z + \mu$. Die Verteilung von X wird sich also über die Verteilung von Z ergeben, denn $\varphi^{-1}(X) = (X - \mu)/\sigma = Z$. 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 37: 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 $\Delta Z = \sigma \Delta Z$, d.h. eine Veränderung von Z um ΔZ bedeutet eine σ-fache Veränderung Δx von X. Setzt man $(x - \mu)/\sigma$ für z in die Dichtefunktion $f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}z^2)$ für Z ein und multipliziert, mit $1/\sigma$, so erhält man die Dichtefunktion für X:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \tag{6.161}$$

Allgemein gelte nun $X = \varphi(Z)$, wobei φ^{-1} für alle X existiere und von Null verschieden sei. Dann ergibt sich

$$P(X \leq x) = P\varphi(Z) \leq x) = P(Z \leq \varphi^{-1}(x)),$$

so dass

$$f(x) = f_Z(\varphi^{-1}(z))\frac{d\varphi^{-1}(x)}{dx} = f_Z(\psi(x))\frac{d\psi(x)}{dx} \tag{6.162}$$

f_Z ist hier wieder in der in (6.156) angegebenen allgemeinen Form. Während eine lineare Transformation von Z wieder auf die Gaußsche Dichte (6.161) führt, muss dies bei einer nichtlinearen Transformation φ nicht mehr der Fall sein. □

Der mehrdimensionale Fall: Es seien $\mathbf{x} = (x_1, \ldots, x_m)'$ und $\mathbf{y} = (y_1, \ldots, y_n)'$ zufällige Vektoren, d.h. die x_i und y_i seien zufällige Veränderliche, und es gelte $\mathbf{y} = A\mathbf{x}$, A eine (n,n)-Matrix mit vollem Rang, so dass die Inverse A^{-1} existiert. Gesucht ist die Verteilung von $\mathbf{y}$, wenn die von $\mathbf{x}$ gegeben ist. Der Übersichlichkeit halber wird das vorgehen am Fall $n = 2$ illustriert.

Gegeben sei ein Doppelintegral einer Funktion mit zwei unabhängigen Veränderlichen; eine Transformation der beiden Variablen kann oft die Aufgabe, das Integral zu bestimmen, erleichtern. Dazu werden die Funktionen

$$x = \varphi(u, v), \quad y = \psi(u, v) \tag{6.163}$$

eingeführt. Die Jacobi-Matrix ist

$$J = \frac{\partial(x, y)}{\partial(u, v)} = \begin{pmatrix} \frac{\partial\varphi}{\partial u} & \frac{\varphi}{\partial v} \\ \\ \frac{\partial\psi}{\partial u} & \frac{\partial\psi}{\partial v} \end{pmatrix} = \begin{pmatrix} \varphi_u & \varphi_v \\ \psi_u & \psi_v \end{pmatrix}. \tag{6.164}$$

Dann gilt der

Satz 6.10 Transformationssatz

$$\iint_G f(x, y)dxdy = \iint_{G*} f(\varphi(u, v), \psi(u, v))|det(J_{\varphi\psi})|dudv \tag{6.165}$$

Beweis: Statt eines ausführlichen Beweises wird hier eine eher intuitive Betrachtung vorgestellt (Courant II, p. 231), einen vollständigen Beweis findet man zB in Courant II, pp. 222 – 225. Man denke sich das Gebiet $\mathcal{G}^*$ in Rechtecke aufgeteilt: Geraden u_i parallel zur v-Achse und Geraden v_j parallel zur u-Achse. Die Differenz zwischen u_i und u_{i+1} betrage $u_{i+1} - u_i = \Delta u$, die zwischen v_{j+1} und v_j sei Δv für alle i und j. Das i-te Rechteck G_i hat dann die Fläche $\Delta u \Delta v$. in der xy-Ebene entspricht g_i ein Teilgebiet G_i^*, definiert durch die Abbildungen $\varphi(u_i, v_j)$, $\psi(u_i, v_j)$ etc. Im allgemeinen Fall sind die Gebiete G_i^+ nicht mehr notwendig Rechtecke, denn die Verbindungslinien zwischen den Eckpunkten können nun krummlinig sein. Sind φ und ψ allerdings linear, so sind die G_i^* Parallelogramme. Man betrachte nun die Determinante

$$D := \begin{vmatrix} \varphi(u_i + h, v_j) - \varphi(u_i, v_j) & \varphi(u_i, v_j + k) - \varphi(u_i, v_j) \\ \psi(u_i + h, v_j) - \psi(u_i, v_j) & \psi(u_i, v_j + k) - \psi(u_i, v_j) \end{vmatrix}$$

Für hinreichend kleine Werte von h und k kann man dafür

$$\begin{vmatrix} \varphi_u(u_i, v_j) & \varphi_v(u_i, v_j) \\ \psi_u(u_i, v_j) & \psi_v(u_i, v_j) \end{vmatrix} hk \approx hkD \tag{6.166}$$

schreiben; hkD approximiert die Fläche von G_i^*. $f(\varphi(u_i, v_j), \psi(u, v))Dkh$ definiert dann ein Volumenelement, und die Summe über alle diese Elemente approximiert dann das Integral und man hat

$$\lim_{h\to 0, k\to 0} \sum_{i,j} f(\varphi(u_i, v_j), \psi(u_i, v_j))Dkh = \iint_{\mathcal{G}^*} f(\varphi(u, v), \psi(u, v))Ddudv.$$

□

Um das folgende Beispiel vorzubereiten, wird das folgende Lemma bewiesen:

Lemma: A habe vollen Rang. Dann gilt

$$(A^{-1})'A^{-1} = (AA')^{-1}. \tag{6.167}$$

Beweis: Es ist[108] $(AA')^{-1} = (A')^{-1}A^{-1}$. Multiplikation von links mit A' liefert $A'(AA')^{-1} = A^{-1}$, und die Transponierung der Gleichung führt auf $((AA')^{-1})'A = (A^{-1})'$. Multiplikation von rechts mit A^{-1} liefert schließlich $(AA')^{-1} = (A^{-1})'A^{-1}$, und das war zu zeigen. □

Beispiel 6.9 Die multivariate Normalverteilung Es sei $\mathbf{z} = (z_1, \ldots, z_n)'$ ein zufälliger Vektor, dessen Komponenten stochastisch unabhängig $N(0,1)$-verteilt seien, d.h. es soll

$$f(\mathbf{z}) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\mathbf{z}'\mathbf{z}\right). \tag{6.168}$$

gelten. Weiter sei $\mathbf{x} = A\mathbf{z} + \boldsymbol{\mu}$, d.h. $\mathbf{x}$ und $\boldsymbol{\mu}$ seien ebenfalls n-dimensionale Vektoren und A sei eine (n,n)-Matrix mit vollem Rang. Dann ist die Verteilung von $\mathbf{x}$ durch

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right). \tag{6.169}$$

gegeben.

Beweis: Es ist $\mathbf{x} - \boldsymbol{\mu} = A\mathbf{z}$, so dass

$$(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\mathbf{u})' = A\mathbf{z}\mathbf{z}'A'.$$

Dann ist $\mathbb{E}[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})'] = A\mathbb{E}(\mathbf{z}\mathbf{z}')A'$, und

$$\mathbb{E}(\mathbf{z}\mathbf{z}') = (\mathbb{E}(z_i z_j)), \quad \mathbb{E}(z_i z_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases},$$

wegen der vorausgesetzten Unabhängigkeit der Komponenten von $\mathbf{z}$, d.h. $\mathbb{E}(\mathbf{z}\mathbf{z}') = I$ die Einheitsmatrix, so dass

$$\boldsymbol{\Sigma} = \mathbb{E}[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})'] = AA', \tag{6.170}$$

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 (6.168) 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 (6.167) und nach dem Produktsatz für Determinanten gilt mit $\boldsymbol{\Sigma} = AA'$

$$|\Sigma| = |AA'| = |A||A'| = |A|^2$$

[108]Wegen $(AB)^{-1} = B^{-1}A^{-1}$

so dass $|A| = |\mathbf{\Sigma}|^{1/2}$, also erhält man für die Dichte von $\mathbf{x}$ die Dichte (6.169). $\square$

Kommentar: $|\Sigma|^{1/2}$ ist gleich der Funktionaldeterminante, die dem φ' im 1-dimensionalen Fall entspricht. $\square$

nebeneinander, so entsteht die Matrix

Literatur

[1] Anderson, T.W.:An Introduction to Multivariate Statistial Analysis. John Wiley & Sons, New York 1958

[2] Barrantes Campos, H.: Elementos de álgebra lineal. *Editorio Unniversidad a Distancia*, San José, 2012

[3] Bishop, C.M.: Pattern Recognition and Machine Learning. Springer 2006

[4] Bortz, J.: Statistik für Sozialwissenschaftler, Springer Verlag, Berlin 1999

[5] Cadima, J. Jolliffe, I. (2009) On relationships between uncentred and column-centred Principal Component Analysis. *PAk. J. Statist.*, 25(4), 473-503

[6] 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

[7] Cangelosi, R., Goriely, A. (2007) Component retention in principal component analysis with application to cDNA microarray data, *Bioloigy Direct*, 2(2), doi:10.1186/1745-6150-2-2

[8] Cattell, R.B. The data box: its orderung of total resources in terms of possible relational systems. In: Cattell, R.B. (ed): Handbook of multivariate experimental psychology. Chicago 1966

[9] Chu, C., Ni, Y., Tan, G., Saunders, C. J., & Ashburner, J. (2011). Kernel regression for fMRI pattern prediction. NeuroImage, 56(2), 662-673.

[10] Courant, R.: Vorlesungen über Differential- und Integralrechung, Erster Band: Funktionen mehrerer Veränderlicher. Springer Verlag Berlin 1961

[11] Courant, R.: Vorlesungen über Differential- und Integralrechung, Zweiter Band: Funktionen einer Veränderlichen. Springer Verlag Berlin 1963

[12] Courant, R. (1929) Über die Eigenwerte bei den Differentialgleichungen der mathematischen Physik. *Mathematische Zeitschrift* Band 7, Nr. 1–4, 1–57.

[13] Courant, R., Hilbert, D.: Methoden der Mathematischen Physik I, Springer-Verlag Berlin, eidelberg, New York 1968

[14] Dorier, J.L. (1995) A general outline of the genesis of vector space theory. *Historia Mathematica* 22, 227–261

[15] Eckart, C., Young, G. (1936), The approximation of one matrix by another of lower rank. *Psychometrika*, 1 (3): 211–8. doi:10.1007/BF02288367.

[16] Fischer, E. (1905) Über quadratische Formen mit reellen Koeffizienten. *Monatshefte für Mathematik und Physik*, Band 16, 234–249.

[17] Fischer, G.: Lineare Algebra. Braunschweig Wiesbaden 1997

[18] Forster, G.: Analysis 2 – Differentialrechnung im $\mathbb{R}^n$, gewöhnliche Differentialgelichungen. Springer- Sprektum, Wiesbaden 2017

[19] Forster, G.: Analysis 3 – Maß- und Integrationstheorie, Integralsätze im $\mathbb{R}^n$ und Anwendungen. Springer- Sprektum, Wiesbaden 2017

[20] Gabor, D. (1948) A new microscopic principle. *Nature*, 161, 777–778

[21] Gabriel, K.R. (1971) The biplot display of matrices with application to principal component analyis. *Biometrica* 58 (3), 453 – 467

[22] Golub, G.H., van Loan, C. F.: Matrix computations. Baltimore 2013

[23] Gumbel, E.J. (1961) Bivariate Logistic Distributions, *Journal of the American Statistical Association*, 56, No. 294. (Jun., 1961), 335-349

[24] Goshtaby, A., O'Neill, W. (1994) Curve fitting by a sum of Gaussians. *CV-GIP: Graphical Models and Image Processing*, 56(4). 281–288

[25] Graybill, F.: Matrices with Applications in Statistics. Belmont 1969

[26] Heuser, H.: Lehrbuch der Analysis, Teil 1. Stuttgart, 2001

[27] Hoaglin, D,C, Welsch, R.E. (1978) The hat matrix in Regression and ANOVA. *The American Statistician*, 32(1), 17 – 22

[28] Honeine, P. (2014) An eigenanalysis of data centering in machine learning. *Statistics Machine Learning (stat.ML)* arXiv:1407.2904fl, 10 Jul 2014.

[29] 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)

[30] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology,* 24(6), 417.

[31] Karhunen, K. (1947), Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys.*, 37, 1–79.

[32] Koecher, M.: Lineare Algebra und analytische Geometrie. Springer-Verlag, Berlin, Heidelberg, New York 1997

[33] Lange, M. (2009) A Tale of two Vectors. *Dialectica*, 63(4), 397–431

[34] Loève, M. (1978), Probability theory. Vol. II, 4th ed. Graduate Texts in Mathematics, 46. Springer-Verlag.

[35] Lorenz, F.: Lineare Algebra I, II. Mannheim, 1988

[36] Mardia, K.V., Kent, J. T., Bibby, J.M.: Multivariate Analysis. Academic Press, London, New York, Toronto 1979

[37] Margalit, D. & Rabinoff, J.: Interactive Linear Algebra, June 2019 https://textbooks.math.gatech.edu/ila/

[38] Papoulis, A.: Probability, random variables, and stochastic processes. Tokyo 1965

[39] Ni, Y., Chu, C., Saunders, C. J., & Ashburner, J. (2008, June). Kernel methods for fmri pattern prediction. In 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence) (pp. 692-697). IEEE

[40] 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.

[41] Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science,* 2(11), 559-572.

[42] Searle, S. R. (1999). The infusion of matrices into statistics.

[43] Seber, G.A.F.: Linear regression analysis. John Wiley & Sons, New York, London, Sydney Toronto 1977

[44] Shaw-Taylor, J. Christianini, N.: Kernel Methods for Pattern Analysis. Cambridge University Press, Cambridge 2004

[45] Stewart, G. W.: Introduction to Matrix Computations. Academic Press, New York 1973

[46] Stewart, G. W. (1993) On the early history of the singular value decomposition. *SIAM Review*, 35 (4), 551 – 566

[47] Wall, M. E., Rechtsteiner, A., & Rocha, L. M. (2003). Singular value decomposition and principal component analysis. In: A practical approach to microarray data analysis (pp. 91-109). Boston, MA: Springer US.

[48] Werner, D.: Funktionalanalysis. Berlin, 2011hhhh

[49] Wong, E.: Stochastic processes in information and dynamical systems. New York 1971

[50] Wu, A., Nastase, S. A., Baldassano, C. A., Turk-Browne, N. B., Norman, K. A., Engelhardt, B. E., & Pillow, J. W. (2021). Brain kernel: a new spatial covariance function for fMRI data. NeuroImage, 245, 118580.

Index

$\mathbf{c}_x$ Vektor der Richtungskosinus, 26
ℓ^2, 155

Produkt
  dyadisches und Kovarianz, 102

Abbildung, 15, 202
  bijektiv, 203
  Bild von, 203
  definiert durch Matrix, 202
  homomorphe, 204
  identische, 203
  injektiv, 203
  inverse, 203
  isomorphe, 204
  Kern einer, 203
  lineare, 204
  Rang der Abbildung, 205
  surjektiv, 203
abhängig
  linear, 32
absolute Homogenität, 40
abzählbare, 159
achsenparallel, s.a. orientiert, 91
Adjunkte
  einer quadratischen Matrix, 214
Algorithmus, Gauß-, 180
assoziativ, 65
Assoziativität, 64
Ausreißer, 150
Austauschsatz von Steinitz, 48
Autokorrelation, 167
Automophismus, 204
Axialvektor, 221

Basis
  eines Vektorraums, 45
  geordnete, 222
  gleichorientierte, 223
  kanonische, 45
  orthonormale, 52, 53
Basisentwicklung eines Vektors
  orthonormale, 55
Basisergänzungssatz, 50
Basisfunktion, 164
Basisfunktionen
  orthogonal, 164
Basiswechsel, 53
Beweis
  Existenz der Basis eines Vektorraums, 51
bilinear, 161

Cartesisches Produkt, 9
charakteristische Gleichung, 216
charakteristisches Polynom, 216
Cholesky-Zerlegung, 185
Cramersche Regel, 119
cramersche Regel, 77

Datenkompression, 6
Datenreduktion, 6
Definitheit, 40
Determinante, 77, 207
Diagonalisierung, 85
Diagonalmatrix, 61
  Rang, 74
Differential, 224
differenzierbar, total, 223
Dimension (s.a. Rang), 59
Dimensionalitätsatz, 57
Diskriminanzkoeffizient, 142
Diskriminnanzfunktion, 153
  lineare, 152
Distanz
  Dreiecksungleichung, 171
  euklidische, 40
  Minkowski, 41
  Reflexivität, 171
Distributivität
  Skalarprodukt, 18
Dreiecksmatrix, 180
Dreiecksungleichung, 20, 40
dyadisches Produkt, 73
  Rang, 70

Ebene
  k-dimensionale, 51
  in einem Vektorraum, 42, 52
Eckart & Young
  Satz von, 132
Eigenfunktion, 163
Eigenraum, 109
Eigenstruktur
  einer $(n \times n)$-Matrix, 85
Eigenvektor, 63
  Links-, 106
  Rechts-, 106
Eigenwert, 63, 85
  komplexer, 108
  Nullstellen eines Polynoms, 108
Eindeutigkeit
  von Linearkombinationen, 46
Einflußmatrix, 150
eingefüht, 159
Einheitsvektor
  kanonischer, 12
Einsvektor, 12
Elementarmatrix, 179
Eliminationsverfahren, gaußsches, 180
Ellipsoid, 91
  Punktekonfiguration, 96
Endomorphismus, 204, 216
Erwartungswert, 102
Erzeugendensystem, 44

Faser, 203
Featur Space, 158
feature mappings, 152
Funktion
  homogene, 15
  lineare, 14
Funktionalmatrix, 224
Funktionenraum, 164

Gauss-Dichte, 31
Gaußverteilung
  n-dimensionale, 103
Gebietsintegral, 219
Gerade
  Ortsvektor, 41
  Parametergleichung, 41
  Richtungsvektor, 41
Gesamtvarianz
  Daten, 98
Gleichung, charakteristische, 105
Gleichungssystem
  homogen/inhomogen, 116
Gradientenvektor, 197
Gram-Matrix, 72
Gram-Schmidt-Verfahren, 55
Gramsche Matrizen, 72
Grundformel einf. einer neuen Variablen, 225
Grundstruktur, 88, 192

Hauptachsentransformation, 84, 95
Hauptraum, 110
Hauptvektor, 110
Hebelwirkung, 150
Hilbert-Raum, 156, 160
  kernreduzierender, 156
Homogenität
  Skalarprodukt, 18
Homomorphismus, 204
Hyperebene, 51

idempotent, 123, 148
Identität, 203
influence matrix, 150
integrierbar
  quadratisch, 162
Inverse, 76, 195
  generalisierte, 126
  Moore-Penrose, 126
  Pseudo, 126
Isomorphismus, 204

Jacobi-Matrix, 224

Karhunen-Loéve-Analyse, 165
Kern, 70, 116
  einer Matrix, 109
  positiv definiter, 154
  reproduzierender, 157
Kern-Funktion, 159
Kern-Trick, 154

Kernel-Trick, 154
Kernfunktionen
  reproduzierende, 156
Kernfuntkion, 158
Kettenregel, 224
Kodimension eines Teilraums, 59
Kofaktor, 214
Kofaktoren, 214
kollinear, 36
kommutativ, 65
Komplement
  orthogonales, 56, 86
Komponenten in Bezug auf eine Basis, 45
Kontinuum, 170
Koordinatenform
  Ebene, 43
  Gerade, 42
Korrelation
  kanonische, 113
  Produkt-Moment-, 25
  Stichproben-, 25
Kosinussatz, 20
Kovarianz
  Stichpoben-, 25
Kovarianzen
  von Linearkombinationen unabhängiger Variablen, 102
Kovarianzmatrix, 73
Kreuzprodukt, 72
Kreuzproduktmatrix
  Rang von, 72
Kronecker-Delta, 19, 164
Körper, 174

Ladung, 55
Lagrange
  -Faktor, 197
  -Funktion, 197
  -Multplikator, 197
  sche Multiplikatorenregel, 197
LDL-Zerlegung, 186
Leibniz-Formel, 210
leverage, 150
linear abhängig, 32
lineare Hülle, 44
lineare Hülle (Gleich'syst.), 116
Linearkombination, 14
  als Abbildung, 15
  von Funktionen, 164
Linkseigenvektoren, 87, 192
Linksinverse, 75
Loss function, 121
Längenskalierung, 65
Lösung
  duale, 123
  primäre, 123

Matrix
  Wurzel einer symmetrischen, 100
  als Abbildung, 202
  assoziierte, 109
  charakterische Gleichung, 105
  Diagonal-, 61
  Dreiecks, 180
  elliptisch, 90
  gestürzte, transponierte, 60
  hermitesch, 109
  hyperbolisch, 90
  imaginär, 109
  inverse, 76
  inverse (2-dim), 77
  konjugierte, 109
  negativ semidefinit, 90
  positiv semidefinit, 90
  Präzisions, 78
  schief-symmetrisch, 109
  spaltenstandardisiert, 73
  spaltenzentriert, 73
  symmetrische, 61
  Transformations-, 189
  zufällige, 101
  ähnliche, 215
Matrixnorm, 129
Matrizen
  ähnliche, 110
Mercers Theorem, 168, 169
Merkmalsraum, 158
Metrik, 171
  City-Block, 172

euklidisch, 172
euklidische, 40, 172
Manhattan, 172
Minkowski, 41, 172
Mini-max-Theorem
von Courant-Fischer, 93
Minor, 214
Mittelwertfunktion, 167

Norm, 19
p-, 129
-Matrix, 129
allgemein, 40
definierende Eigenschaften, 40
euklidische, 19, 128
Frobenius-, 129
Hilbert-Schmidt-, 129
induziert, 19
induzierte, 129
Maximum, 128
Maximums, 19
Schur-, 129
Normalenvektor, 37, 52, 56
Ebene, 42
Gerade, 41
Normalverteilung
multivariate, 230
Nullität, 70
Nullmatrix, 69
Nullraum, 70
Nullvektor, 12

Operator
linearer, 102
Opertor
linearer, 162
orientiert
negativ, 223
positiv, 223
Orientierung
eines Vektorraums, 220
Orientierungsvektor, 26
orthogonal, 22
orthogonales Komplement, 56
orthonormal, 22
Orthonormalbasis (ONB), 52
orthonormale Basisentwicklung, 54
Ortsvektor, 41

Parameterform
Ebene, 42
Gerade, 187
penalty, 121
Permutation, 208
Pfad, 166
Polynom, charakteristisches, 105
positiv semidefinit, 98
Principal Component Anylsyis, 84
Produkt
dyadisches, 16, 131
dyadisches und SVD, 130
inneres, 16
tensorielles, 16
Produktregel
stoch.unabh. Variablen, 35
Produktsatz, 210
Projektion
Vektor auf Gerade, 29
Vektor auf Vektor, 29
Projektionsmatrix, 148
Präzisionsmatrix, 78
Pseudoinverse, 119
Punkt: Schreibweise, 171
Punktekonfigurationen K_x, K_y, 96
Pythagoras
verallgem. Satz von, 19

quadratintegrierbar, 155
quadratische Form, 90

Rang
Diagonalmatrix, 74
dyadisches Produkt, 70
einer Matrix, 67
einer Vektormenge, 59
Kreuzproduktmatrizen, 72
voller, 59
Rangsatz, 67
Rayleigh-Quotient, 91, 93
generalisierter, 113
Rechtseigenvektoren, 87, 192

Rechtsinverse, 75
Regularisierung, 121
Regularisierungsterm, 121
representer evaluation, 157
Richtungskosinus, 25
Richtungsvektor, 26, 41
Richtungsvektoren
  Ebene, 42
Ridge-Regression, 121
Ridge-Regression-Schätzung, 122
RKHS, 161
Rotation, 53
Rotationsmatrix, 182

Sarrussche Regel, 212
Satz
  des Thales, 23
  Eckart-Young, 132
  von Courant-Fischer, 93, 198
  von Schmidt-Mirsky, 132, 133
  von Steinitz, 47
schlecht formulierte Probleme, 120
Schur-Norm, 129
Schwarzsche Ungleichung, 30
separierbar, 159
Singularwert, 87, 192
Singularwertzerlegung, 87, 192
Skalar, 12, 170
Skalarprodukt, 16
  algebraisch, 22
  als Koordinate, 30
  alternative Definition, 17
  geometrisch, 22
Skalarproduktraum, 39
Skalierung
  eines Vektors, 20, 65
Spaltenraum, 66, 205
Spaltenvektor, 26
Spaltenvektoren, 60
span, 44
Spur, 97, 130
Standardmatrix, 178
Steinitzsches Lemma, 47
stetig (im quadrat. Mittel), 167
stochastischer Prozess
  zentrierter, 167
Strafterm, 121
Stütz-/Ortsvektor
  Ebene, 42
Subaddditivität, 40

Teilraum
  r-dimensionaler, 50
  Dimension (Rang) des, 59
  eindimensionaler, 36
  invarianter, 110
  orthogonales Komplement, 56
  orthonormale Basis, 55
  trivialer, 38, 174
  zweidimensionaler, 37
Teilräume
  Dimensionsformel, 59
  Durchschnitt, 57
  Kombination von, 57
  orthogonale, 55
  orthogonale Summe, 57
  Summe zweier, 58
  Vereinigung, 57
Teilvektorraum, 38, 44
Trajektorie, 166
Transformation
  eines Spaltenvektors, 63
  eines Vektors, 15
Transformationssatz, 229
Transposition, 208
triviale Lösung, 32
Tychonow-Matrix, 121
Tychonow-Regularisierung, 121
Typ, 28
Typus, 11

Umformungen, elementare, 181
Unabhägigkeit
  lineare und Korrelation, 34
unabhängig
  linear, 32
unendlich-dimensional, 164
Unterdvektorraum, 38
Unterraum, 59

Varianz

Parameterschätzungen, 100
Vektor
n-dimensionaler, 33
-transformation, 15
Axial, 221
charakteristischer, 85
Eins-, 12
gerichtete Größe, 10
Komponenten, 10
Länge, 10
normiert, 20
Null-, 12
zentrierter, 24
zufälliger, 16, 101
Vektoren
charakeristische, 63
Vektornorm, 128, 129
Vektorprodukt, 221
Vektorraum, 7, 36
n-dimensionaler, 50
euklidischer, 39, 40
normierter, 129
Verlustfunktion, 121

Zeilenraum, 66
Zeilenvektor, 26
Zeilenvektoren, 60
Zentrierungsmatrix, 123
Zufallsmatrix, 101
Zufallsvektor, 101

Übergangsmatrix, 80
überabzählbar, 44, 170