Vektoren und Matrizen

Für Anwendungen in der multivariaten Statistik

U. Mortensen
FB Psychologie und Sportwissenschaften
Westfälische Wilhelms-Universität Münster

Skriptum zur Vorlesung[1]

.

Multivariate Statistik

.

Psychologisches Institut der
Johannes Gutenberg-Universität Mainz, WS 2011/12

[1] Überarbeitete Version des Skripts für das WS 2003/2004, korrigiert am 14. 11. 2011

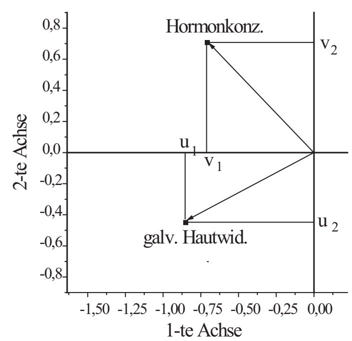
Inhaltsverzeichnis

1 Vektoren

1.1 Grundbegriffe

In einem Koordinatensystem kann ein Punkt durch seine Koordinaten charakterisiert werden. Eine gerade Linie kann durch die Koordinaten ihres Anfangs- und ihres Endpunktes spezifiziert werden. Kommt es nur auf die Länge und Richtung dieser Linie an, so genügt es, die Differenzen der Koordinaten für den Anfangs- und Endpunkt der Linie anzugeben. Denkt man sich den Endpunkt einer solchen Linie mit einer Pfeilspitze versehen und kommt es nur auf die Länge und die Orientierung des so entstandenen Pfeils an, so spricht man von einem *Vektor*. Die eben genannten Koordinatendifferenzen heißen dann die *Komponenten* des Vektors. Verschiebt man den Vektor so, daß sein Anfangspunkt mit dem Ursprung des Koordinatensystems zusammenfällt, so entsprechen die Koordinaten der Pfeilspitze gerade den Komponenten des Vektors. Abb.1 zeigt solche Vektoren; für

Abbildung 1: Vektoren



die Vektoren **u** und **v** sind die Koordinaten (u_1, u_2) und (v_1, v_2) der jeweiligen Vektorendpunkte (= Pfeilspitzen) eingezeichnet. Zusätzlich wird der Winkel α zwischen den Vektoren **u** und **v** angegeben.

Die Vektoren durch Angabe ihrer Komponenten beschrieben werden:

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \tag{1}$$

Man macht sich leicht klar, daß es bei der Charakterisierung von Vektoren auf die Reihenfolge der Komponenten ankommt: die Vektoren **u** und **v** unterscheiden sich durch die Orientierung.

Natürlich können Vektoren auch im 3-dimensionalen Raum beschrieben werden; dann muß eine jeweilige dritte Komponente u_3 bzw. v_3 hinzugefügt werden. Allgemein erweist es sich als nützlich, Vektoren gegebenenfalls auch durch mehr als drei, etwa durch n Komponenten zu definieren: man spricht dann von n-dimensionalen Vektoren. Der Punktraum, in dem die Endpunkte der Vektoren liegen, heißt dann n-dimensionaler Punktraum. Für $n > 3$ ist ein solcher Punktraum anschaulich nicht mehr vorstellbar; es handelt sich gewissermaßen bei den Begriffen des n-dimensionalen Vektors und des n-dimensionalen (Punkt-)Raumes lediglich um eine *façon de parler*. Allgemein ist ein n-dimensionaler

Vektor also durch ein n-Tupel von Zahlen $u_1, u_2, \ldots, u_n$ definiert. Die u_i, $i = 1, 2, \ldots, n$ heißen wieder die Komponenten des Vektors. Sie können insbesondere die Meßwerte von n Personen für eine bestimmte Variable repräsentieren. Man sieht daran, daß die Reihenfolge der Komponenten für die Definition des Vektors wichtig ist: vertauscht man die Komponenten u_i und u_j, $i \neq j$, so vertauscht man die Meßwerte der Personen i und j.

Schreibweisen: Ein Vektor wird immer durch eine *Spalte* von Zahlen - seiner Komponenten - definiert, wie in (1). Man spricht dementsprechend auch von einem *Spaltenvektor*. Gelegentlich ist es nützlich, die Komponenten als Zeile anzuschreiben; der Vektor ist dann *gestürzt* oder *transponiert* worden. Man schreibt dann

$$\mathbf{u}' = (u_1, u_2, \ldots, u_n). \tag{2}$$

Mit $\mathbf{u}'$ ist als der als Zeilenvektor geschriebene Vektor $\mathbf{u}$ gemeint. $\mathbf{u}''$ ist dann wieder der ursprüngliche Spaltenvektor $\mathbf{u}$. Dementsprechend kann man den Spaltenvektor $\mathbf{u}$ auch - platzsparend - in der Form $\mathbf{u} = (u_1, u_2, \ldots, u_n)'$ schreiben.

1.2 Rechenregeln

Vektoren können als neue mathematische Objekte betrachtet werden, für die gewisse Rechenregeln eingeführt werden können. Eine einzelne (relle) Zahl wird im Unterschied zum Vektor auch als *Skalar* bezeichnet.

1. **Multiplikation eines Vektors mit einem Skalar λ:**

$$\lambda\mathbf{u} = (\lambda u_1, \lambda u_2, \ldots, \lambda u_n)'. \tag{3}$$

   Es wird also jede Komponente mit λ multipliziert.

2. **Addition zweier Vektoren:**

$$\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n)'. \tag{4}$$

   Die Addition zweier Vektoren ist also durch die Summe der korrespondierenden Komponenten definiert. Allgemein hat man dann

$$\lambda\mathbf{u} + \mu\mathbf{v} = (\lambda u_1 + \mu v_1, \ldots, \lambda v_n + \mu v_n)'. \tag{5}$$

   $\lambda\mathbf{u} + \mu\mathbf{v}$ für beliebige Skalare λ und μ heißt auch *Linearkombination* der Vektoren $\mathbf{u}$ und $\mathbf{v}$.

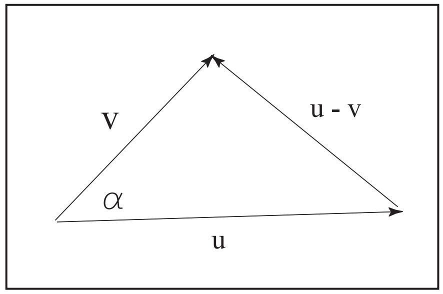
3. **Das Skalarprodukt $\mathbf{u}'\mathbf{v}$** ist durch die Summe der Produkte $u_i v_i$ definiert:

$$\mathbf{u}'\mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \sum_{i=1}^{n} u_i v_i. \tag{6}$$

   $\mathbf{u}'\mathbf{v}$ heißt Skalarprodukt, weil $\mathbf{u}'\mathbf{v}$ eben kein Vektor, sondern eine einzelne Zahl,also ein Skalar, ist. Die Schreibweise $\mathbf{u}'\mathbf{v}$ hat mit der Definition dieses Produkts zunächst nichts zu tun, erweist sich aber bei der Definition des Produktes von Matrizen (s. unten) als nützlich. Ausgeschrieben hat man jedenfalls

$$\mathbf{u}'\mathbf{v} = (u_1, u_2, \ldots, u_n) \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \sum_{i=1}^{n} u_i v_i. \tag{7}$$

Abbildung 2: Der Kosinussatz



Eine Alternative Schreibweise für das Skalarprodukt ist $\langle \mathbf{u}, \mathbf{v} \rangle$, also

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}'\mathbf{v}; \tag{8}$$

diese Schreibweise kann übersichtlicher sein.

4. Es läßt sich noch das Produkt uv' definieren; hierauf wird nach Einführung von Matrizen eingegangen.

5. **Vektorlänge und Normierung** Die Länge eines Vektors ist nach Pythagoras durch

$$|\mathbf{u}| = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2} = \sqrt{\mathbf{u}'\mathbf{u}} \tag{9}$$

gegeben; dementsprechend ist das Skalarprodukt eines Vektors mit sich selbst gleich dem Quadrat seiner Länge: $\mathbf{u}'\mathbf{u} = |\mathbf{u}|^2$. Der Vektor $\lambda\mathbf{u}$, λ ein Skalar, hat demnach die Länge $|\lambda\mathbf{u}| = \lambda|\mathbf{u}|$. Die Multiplikation eines Vektors mit einem Skalar bedeutet also einfach die Verlängerung oder Verkürzung ("Stauchung") des Vektors.

Vektoren mit der Länge 1 heißen *normiert*. Ein Vektor mit beliebiger Länge $|u|$ läßt sich stets durch Multiplikation mit einem geeigneten Skalar *normieren*: sei etwa $\lambda(1/|\mathbf{u}|)$. Dann ist

$$\lambda|\mathbf{u}| = \frac{1}{|\mathbf{u}|}|\mathbf{u}| = 1. \tag{10}$$

Ein Vektor wird also normiert, indem man jede seiner Komponenten durch die ursprüngliche Länge $|\mathbf{u}|$ teilt.

Das Skalarprodukt hängt vom Winkel α zwischen den Vektoren $\mathbf{u}$ und $\mathbf{v}$ ab. Nach dem Kosinusatz gilt

$$|\mathbf{u} - \mathbf{v}|^2 = |\mathbf{u}|^2 + |\mathbf{v}|^2 - 2|\mathbf{u}||\mathbf{v}|\cos\alpha. \tag{11}$$

Die Differenz $\mathbf{u} - \mathbf{v}$ wird dabei wie eine Linearkombination (5) mit $\lambda = 1$ und $\mu = -1$ berechnet. Nach (11) erhält man

$$(\mathbf{u} - \mathbf{v})'(\mathbf{u} - \mathbf{v}) = \mathbf{u}'\mathbf{u} + \mathbf{v}'\mathbf{v} - 2\sqrt{(\mathbf{u}'\mathbf{u})(\mathbf{v}'\mathbf{v})}\cos\alpha. \tag{12}$$

Dabei ist

$$(\mathbf{u} - \mathbf{v})'(\mathbf{u} - \mathbf{v}) = \mathbf{u}'\mathbf{u} - \mathbf{u}'\mathbf{v} - \mathbf{v}'\mathbf{u} + \mathbf{v}'\mathbf{v} = \mathbf{u}'\mathbf{u} + \mathbf{v}'\mathbf{v} - 2\mathbf{u}'\mathbf{v}, \tag{13}$$

da ja $\mathbf{u}'\mathbf{v} = \mathbf{v}'\mathbf{u}$ ist. Eingesetzt in (12) erhält man

$$-2\mathbf{u}'\mathbf{v} = -2|\mathbf{u}||\mathbf{v}|\cos\alpha,$$

denn $\sqrt{\mathbf{u}'\mathbf{u}\mathbf{v}'\mathbf{v}} = |\mathbf{u}||\mathbf{v}|$. Mithin folgt

$$\cos\alpha = \frac{\mathbf{u}'\mathbf{v}}{|\mathbf{u}||\mathbf{v}|}, \quad \text{bzw.} \quad \mathbf{u}'\mathbf{v} = |\mathbf{u}||\mathbf{v}|\cos\alpha. \tag{14}$$

Für normierte Vektoren, wenn also $|\mathbf{u}| = |\mathbf{v}| = 1$ gilt, folgt also insbesondere

$$\mathbf{u}'\mathbf{v} = \cos\alpha, \quad \text{wenn } |\mathbf{u}| = |\mathbf{v}| = 1. \tag{15}$$

Ist $\alpha = \pi/2$ bzw. $\alpha = 90^o$, so stehen die Vektoren $\mathbf{u}$ und $\mathbf{v}$ senkrecht aufeinander; man sagt, sie seien *orthogonal* (= rechtwinklig) und schreibt $\mathbf{u}\perp\mathbf{v}$.

> *Da* $\cos\pi/2 = 0$ *folgt für orthogonale Vektoren, daß ihr Skalarprodukt stets gleich Null ist.*

1.3 Linearkombinationen und Vektorräume

Linearkombinationen Es seien allgemein $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r$ r n-dimensionale Vektoren, und $\lambda_1, \lambda_2, \ldots, \lambda_r$ seien reelle Zahlen. Dann ist $\lambda_j\mathbf{u}_j$, $1 \leq j \leq r$ ebenfalls ein Vektor; er ist parallel zu $\mathbf{u}_j$, allerdings ist seine Länge um den Faktor λ_j gestreckt ($\lambda_j > 1$) oder gestaucht ($\lambda_j < 0$).

Definition 1.1 *Der Vektor*

$$\boldsymbol{v} = \lambda_1\boldsymbol{u}_1 + \lambda_2\boldsymbol{u}_2 + \cdots + \lambda_r\boldsymbol{u}_r \tag{16}$$

heißt Linearkombination *der Vektoren* $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$.

Die i-te Komponente v_i von $\mathbf{v}$ ist dann durch die Summe

$$v_i = \lambda_1 u_{i1} + \cdots + \lambda_r u_{ir} = \sum_{j=1}^{r} \lambda_j u_{ij} \tag{17}$$

gegeben.

Gegeben seien die r n-dimensionalen Vektoren $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r$. Gibt man irgendwelche λ_j, $j = 1, 2, \ldots, r$ vor, so wird ein entsprechender Vektor $\mathbf{v} = \lambda_1\mathbf{u}_1 + \lambda_2\mathbf{u}_2 + \cdots + \lambda_r\mathbf{u}_r$ resultieren. Man kann umgekehrt den Vektor $\mathbf{v}$ vorgeben und fragen, ob sich die Koeffizienten λ_j finden lassen, so daß $\mathbf{v}$ in der Form (16) darstellbar ist. Insbesondere sei

$$\mathbf{v} = \underline{0} = (0, \ldots, 0)', \tag{18}$$

d.h. die Komponenten v_i von $\mathbf{v}$ seien alle gleich Null. $\underline{0}$ heißt der Nullvektor. Es werde zunächst die folgende Definition eingeführt:

Definition 1.2 *Es seien die n-dimensionalen Vektoren* $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$ *gegeben, und es gelte*

$$\underline{0} = \lambda_1\boldsymbol{u}_1 + \lambda_2\boldsymbol{u}_2 + \cdots + \lambda_r\boldsymbol{u}_r. \tag{19}$$

Ist diese Gleichung dann und nur dann erfüllt, wenn $\lambda_1 = \cdots = \lambda_r = 0$*, so heißen die* $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$ linear unabhängig. *Gilt (19), wenn für mindestens ein* $\lambda_j \neq 0$ *gilt, so heißen die* $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$ linear abhängig.

Erläuterung: Die Vektoren $\mathbf{u}_1, \ldots, \mathbf{u}_r$ seien linear abhängig. Dann folgt aus (19), daß dann irgendeiner von ihnen als Linearkombination der übrigen dargestellt werden kann, etwa $\mathbf{u}_1$, für $\lambda_1 \neq 0$:

$$\mathbf{u}_1 = -\left(\frac{\lambda_2}{\lambda_1}\mathbf{u}_2 + \cdots + \frac{\lambda_r}{\lambda_1}\mathbf{u}_r\right), \tag{20}$$

wobei für $\mathbf{u}_1 \neq \underline{0}$ nicht alle der Koeffizienten λ_j/λ_1, $j = 2, \ldots, r$ gleich Null sind. Sind dagegen die $\mathbf{u}_1, \ldots, \mathbf{u}_r$ linear unabhängig, so gilt (20) nicht, denn die λ_j sind dann alle gleich Null.

Definition 1.3 *Die n-dimensionalen Vektoren*

$$\boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \boldsymbol{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \boldsymbol{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \tag{21}$$

heißen n-dimensionale Einheitsvektoren*; die Komponenten des j-ten Einheitsvektors sind also alle gleich Null bis auf die j-te, die gleich 1 ist.*

Folgerungen:

1. Die n Einheitsvektoren sind linear unabhängig.

   Beweis: Zu zeigen ist, daß die Darstellung

   $$\underline{0} = \lambda_1 \mathbf{e}_1 + \cdots + \lambda_n \mathbf{e}_n$$

   nur möglich ist, wenn $\lambda_1 = \cdots = \lambda_n = 0$. Tatsächlich hat man für die i-te Komponente von $\underline{0}$

   $$0 = \lambda_1 0 + \cdots + \lambda_{i-1} 0 + \lambda_i 1 + \lambda_{i+1} 0 + \cdots + \lambda_n 0 = \lambda_i 1,$$

   woraus sofort $\lambda_i = 0$ folgt. Da dies für ale i gilt, folgt, daß alle $\lambda_i = 0$, und deshalb sind Einheitsvektoren linear unabhängig. □

2. Die n Einheitsvektoren sind orthogonal.

   Beweis: Für $i \neq j$ gilt $\mathbf{e}_i{}'\mathbf{e}_j = 0$, da für das i-te Produkt $1.0 = 0$ und für das j-te Produkt $0.1 = 0$ gilt, d.h. alle Produkte sind gleich Null, und daher auch deren Summe. □

3. Die Vektoren $\mathbf{u}_1, \ldots, \mathbf{u}_r$ seien paarweise orthogonal. Dann sind sie auch linear unabhängig.

   Beweis: Es gelte

   $$\underline{0} = \lambda_1 \mathbf{u}_1 + \cdots + \lambda_r \mathbf{u}_r.$$

   Dann gilt auch

   $$\mathbf{u}_j{}'\underline{0} = \lambda_1 \mathbf{u}_j{}'\mathbf{u}_1 + \cdots + \lambda_j \mathbf{u}_j{}'\mathbf{u}_j + \cdots + \lambda_r \mathbf{u}_j{}'\mathbf{u}_r = \lambda_j \|\mathbf{u}_j\|^2 = 0,$$

   denn alle Skalarprodukte verschwinden bis auf das Skalarprodukt $\mathbf{u}_j{}'\mathbf{u}_j = \|\mathbf{u}_j\|^2$. Dann folgt aber $\lambda_j = 0$, denn $\|\mathbf{u}_j\| \neq 0$. Dies gilt analog für alle λ_j, so daß die Vektoren linear unabhängig sind. □

   Anmerkung: Die Umkehrung gilt nicht, d.h. linear unabhängige Vektoren sind nicht notwendig auch orthogonal.

4. Die Vektoren $\mathbf{u}_j$ und $\mathbf{u}_k$ haben verschiedene Orientierungen. Dann sind die beiden Vektoren linear unabhängig.

   Beweis: Es gelte $\underline{0} = \lambda_j \mathbf{u}_j + \lambda_k \mathbf{u}_k$. Angenommen, es gelte $\lambda_j \neq 0$, $\lambda_k \neq 0$. Dann folgt $\mathbf{u}_j = -(\lambda_k/\lambda_j)\mathbf{u}_k$; aber dieses Resultat bedeutet, daß $\mathbf{u}_j$ und $\mathbf{u}_k$ dieselbe Richtung haben, also entgegen der Voraussetzung, daß sie die verschiedene Richtungen haben. Also muß $\lambda_j = \lambda_k = 0$ gelten, d.h. die beiden Vektoren sind linear unabhängig. $\square$.

Vektorräume: Bevor die Bedeutung des Begriffs der Linearkombination näher erläutert wird, soll der Begriff des Vektorraumes und, damit zusammenhängend, der des Teilraumes eingeführt werden, weil dieser Begriff bei den Anwendungen eine zentrale Rolle spielt.

Definition 1.4 *Es sei V eine Menge von Vektoren. V heißt* Vektorraum, *wenn die beiden folgenden Bedingungen erfüllt sind:*

V1: Für $\boldsymbol{v} \in V$ ist auch $-\boldsymbol{v} \in V$,

V2: Es seien $\lambda \in \mathbb{R}$, $\mu \in \mathbb{R}$ Skalare. Dann gelten die folgenden Aussagen:

$$(\lambda + \mu)\boldsymbol{v} = \lambda\boldsymbol{v} + \mu\boldsymbol{v}, \quad \lambda(\boldsymbol{v} + \boldsymbol{w}) = \lambda\boldsymbol{v} + \lambda\boldsymbol{v} \tag{22}$$

$$\lambda(\mu\boldsymbol{v}) = (\lambda\mu)\boldsymbol{v} \tag{23}$$

für $\boldsymbol{v}, \boldsymbol{w} \in V$. Betrachtet man speziell n-dimensionale Vektoren, so spricht man auch vom n-dimensionalen Vektorraum und schreibt V_n.

Bemerkungen:

1. Sind $\mathbf{v}, \mathbf{w} \in \mathrm{V}_n$, so folgt, dass auch $\mathbf{u} = \lambda\mathbf{v} + \mu\mathbf{w} \in V_n$, denn V_n enthält ja alle n-dimensionalen Vektoren, und damit auch $\mathbf{u}$. Der Ausdruck $\mathbf{u} = \lambda\mathbf{v} + \mu\mathbf{w}$ kann dann als eine spezielle Darstellung des Vektors $\mathbf{u}$ aufgefasst werden.
2. Jeder Vektorraum V enthält den Nullvektor, denn nach $V2$ gilt für $\lambda \in \mathbb{R}$, dass auch $\lambda\mathbf{x} \in V$, und für $\lambda = 0$ folgt $\lambda\mathbf{x} = \underline{0} \in V$.

Definition 1.5 *Es sei $W \subset V_n$ eine Teilmenge des V_n. Dann heißt W ein* Untervektorraum *(oder einfach Teilraum), wenn die folgenden Bedingungen erfült sind:*

UV1*: $W \neq \emptyset$, $\emptyset$ die leere Menge,*

UV2*: $\boldsymbol{v}, \boldsymbol{w} \in W$ impliziert $\boldsymbol{v} + \boldsymbol{w} \in W$, d.h. W ist in bezug auf die Addition* abgeschlossen*,*

UV3*: $\boldsymbol{v} \in W$, $\lambda \in \mathbb{R}$, dann auch $\lambda\boldsymbol{v} \in W$, d.h. W ist abgeschlossen in bezug auf die Multiplikation mit Skalaren.*

Bemerkung: Die Bedingungen (UV1) und (UV2) implizieren, dass W ein Teilraum oder Untervektorraum dann ist, wenn mit $\mathbf{v}, \mathbf{w} \in W$ auch jede Linearkombination $\lambda\mathbf{v} + \mu\mathbf{w}$ in W enthalten ist.

Definition 1.6 *Es sei V ein Vektorraum und $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$ seien linear unabhängige Vektoren aus V. Wenn alle Vektoren $\boldsymbol{v} \in V$ als Linearkombination der $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$ dargestellt werden können, heißen die $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$ eine* Basis *von V.*

Satz 1.1 *Jeder Vektorraum V hat eine Basis, d.h. es existiert stets eine Menge von Vektoren $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$ derart, dass alle Vektoren aus V als Linearkombinationen der $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$ erzeugt werden können.*

Beweis: V bestehe aus mehr als nur dem Nullvektor $\underline{0}$. Im einfachsten Fall haben alle Vektoren aus V die gleiche Orientierung, bzw. die dieser Orientierung entgegengesetzte Orientierung. Sie entsprechen dann allen Vektoren auf einer Geraden. Man wählt dann einen dieser Vektoren $\mathbf{u}_0$ als Basisvektor; alle anderen Vektoren $\mathbf{u}$ entstehen dann durch geeignete Wahl eines Skalars λ, $\mathbf{u} = \lambda\mathbf{u}_0$. Hat man zwei verschiedene Orientierungen, so wählt man aus jeder Orientierung einen Vektor; die beiden Vektoren sind linear unabhängig (vergl. Folgerung 4). Die Basis besteht dann aus zwei linear unabhängigen Vektoren $\mathbf{u}_1$ und $\mathbf{u}_2$. Existiert nun ein Vektor, der sich nicht als Linearkombination dieser beiden Vektoren darstellen läßt, so nimmt man diesen Vektor zu den beiden ersten hinzu. So verfährt man weiter, bis man mit r Vektoren alle Vektoren von V erzeugen kann. Die r Vektoren bilden dann eine Basis $\mathbf{u}_1, \ldots, \mathbf{u}_r$. □

Bemerkungen:

1. Das Vorgehen im Beweis ist prinzipiell, das praktische Vorgehen zum Auffinden einer Basis wird später beschrieben.

2. Hat man eine Basis bestimmt, so wird sie nicht die einzige sein. Man beginnt ja mit irgendeinem Vektor, nimmt dann irgendeinen Vektor mit anderer Orientierung hinzu, etc. Also wird es viele verschiedene Basen eines Vektorraumes geben, aus denen jeweils alle Vektoren des Vektorraumes erzeugt werden können. Geht man von einer Basis aus und erzeugt damit alle Vektoren von V, so erzeugt man damit auch die Vektoren der anderen Basen. Dies legt nahe, daß man von einer Basis zu einer anderen "übergehen" kann.

Satz 1.2 *Gegeben seien zwei Basen $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$ und $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_s$ eines Vektorraums V. Dann gilt stets $r = s$.*

Beweis: V enthalte mehr Vektoren als nur den Nullvektor. Es seien $\mathbf{u}_1, \ldots, \mathbf{u}_r$ und $\mathbf{v}_1, \ldots, \mathbf{v}_s$ zwei Basen von V und es gelte $r < s$. Alle Vektoren aus V können aus $\mathbf{u}_1, \ldots, \mathbf{u}_r$ als Linearkombination erzeugt werden, also auch die Vektoren $\mathbf{v}_1, \ldots, \mathbf{v}_s$. Die Vektoren $\mathbf{u}_1, \ldots, \mathbf{u}_r, \mathbf{v}_1$ sind dann jedenfalls linear abhängig; man kann dann etwa $\mathbf{u}_1$ als Linearkombination der Vektoren $\mathbf{u}_2, \ldots, \mathbf{u}_r, \mathbf{v}_1$ darstellen. Weiter kann man mit den Vektoren $\mathbf{u}_2, \ldots, \mathbf{u}_r, \mathbf{v}_1, \mathbf{v}_2$ alle Vektoren von V erzeugen; dabei sind diese Vektoren linear abhängig, da ja $\mathbf{v}_2$ bereits eine Linearkombination der Vektoren $\mathbf{u}_2, \ldots, \mathbf{u}_r, \mathbf{v}_1$ ist. Also kann etwa $\mathbf{u}_2$ als Linearkombination der Vektoren $\mathbf{u}_3, \ldots, \mathbf{u}_r, \mathbf{v}_1, \mathbf{v}_2$ dargestellt werden. So verfährt man weiter, bis man schließlich $\mathbf{v}_1, \ldots, \mathbf{v}_r$ also Basis erhält, aus der alle Vektoren aus V erzeugt werden können. Da $r < s$, kann $\mathbf{v}_{r+1}$ als Linearkombination der $\mathbf{v}_1, \ldots, \mathbf{v}_r$ berechnet werden, entgegen der Annahme, daß $\mathbf{v}_1, \ldots, \mathbf{v}_s$ eine Basis ist, diese Vektoren also linear unabhängig sind. Also kann $r < s$ nicht gelten; es gilt also $r = s$. □

Satz 1.3 *Es sei V_n der Raum der n-dimensionalen Vektoren. Jede Basis von V_n enthält genau $r = n$ linear unabhängige Vektoren $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$.*

Beweis: Es existieren genau n n-dimensionale Einheits vektoren $\mathbf{e}_j$, $j = 1, \ldots, n$. Ein beliebiger Vektor $\mathbf{u} \in \mathrm{V}_n$ kann dann als Linearkombination der $\mathbf{e}_j$ dargestellt werden,

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = u_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + u_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + u_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}. \tag{24}$$

Also bilden die $\mathbf{e}_j$, $j = 1, \ldots, n$ eine Basis. Nach Satz 1.2 haben alle Basen die gleiche Anzahl von Vektoren, d.h. jede Basis des V_n enthält genau n Vektoren.

□

Satz 1.4 *Es seien $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ linear unabhängige, n-dimensionale Vektoren. Die von $bu_1, \ldots, \boldsymbol{u}_r$, $r < n$, erzeugten Vektoren bilden einen r-dimensionalen Teilraum des V_n, ebenso bilden die von den $n - r$ linear unabhängigen Vektoren $u_{r+1}, \ldots, \boldsymbol{u}_n$ erzeugten Vektoren einen $(n - r)$-dimensionalen Teilraum V_{n-r} des V_n. Es gilt $V_r \cap V_{n-r} = \underline{0}$, d.h. keiner der von den $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$ erzeugten Vektoren $\neq \underline{0}$ liegt in V_{n-r}, und keiner der von den $\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_n$ erzeugten Vektoren $\neq \underline{0}$ liegt in V_n.*

Bemerkung: Da die $\mathbf{u}_1, \ldots, \mathbf{u}_r$ n-dimensional sind, sind die Vektoren des r-dimensionalen Teilraums ebenfalls n-dimensional. Die Bedeutung dieses Satzes wird in Beispiel 1.2, Punkt 3 (Seite 11) illustriert.

Beweis: $\underline{0}$ ist stets in V_r und V_{n-r}, also folgt sicherlich $\underline{0} \in V_r \cap V_{n-r}$.

Nun sei $\mathbf{v} = \lambda_1 \mathbf{u}_1 + \cdots + \lambda_r \mathbf{u}_r \neq \underline{0}$. Angenommen, es gäbe Zahlen $\mu_{r+1}, \ldots, \mu_n$ derart, daß auch $\mathbf{v} = \mu_{r+1} \mathbf{u}_{r+1} + \cdots + \mu_n \mathbf{u}_n$. Dann folgt

$$\lambda_1 \mathbf{u}_1 + \cdots + \lambda_r \mathbf{u}_r = \mu_{r+1} \mathbf{u}_{r+1} + \cdots + \mu_n \mathbf{u}_n.$$

Es sei insbesondere $\mu_{r+1} \neq 0$. Dann folgt

$$\frac{\lambda_1}{\mu_{r+1}} \mathbf{u}_1 + \cdots + \frac{\lambda_r}{\mu_{r+1}} \mathbf{u}_r - \frac{\mu_{r+2}}{\mu_{r+2}} \mathbf{u}_{r+2} - \cdots - \frac{\mu_n}{\mu_{r+1}} \mathbf{u}_n = \mathbf{u}_{r+1},$$

d.h. der Vektor $\mathbf{u}_{r+1}$ ist eine Linearkombination der Vektoren $\mathbf{u}_1, \ldots, \mathbf{u}_r$ und $\mathbf{u}_{r+2}, \ldots, \mathbf{u}_n$. Da die diese Vektoren aber als linear unabhängig vorausgesetzt wurden, kann $\mathbf{v}$ nur der Nullvektor sein, so daß $\underline{0} \in V_r = V_{n-r}$. □

Definition 1.7 *Die Vektoren $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ seien eine Basis des V_n, so daß jeder Vektor $\boldsymbol{x} \in V_n$ als Linearkombination*

$$\boldsymbol{x} = a_1 \boldsymbol{u}_1 + a_2 \boldsymbol{u}_2 + \cdots + a_n \boldsymbol{u}_n$$

der $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ dargestellt werden kann. Dann heißen die Koeffizienten $a_1, \ldots, a_n$ die Koordinaten *von $\boldsymbol{x}$ in bezug auf die Basis $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$.*

Beispiel 1.1 Die Einheitsvektoren $\mathbf{e}_1, \ldots, \mathbf{e}_n$ bilden eine Basis des V_n. Ist $\mathbf{x} \in V_n$ ein beliebiger Vektor des V_n, so hat man in bezug auf diese Basis die Darstellung

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + x_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}, \tag{25}$$

(vergl. (24)), d.h. die Komponenten von $\mathbf{x}$ sind auch die Koordinaten von $\mathbf{x}$ in bezug auf die Basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$. □

Beispiel 1.2 Es werden die Teilräume des V_3 betrachtet. Eine mögliche Basis des V_3 sind die drei Einheitsvektoren $\mathbf{e}_1$, $\mathbf{e}_2$ und $\mathbf{e}_3$.

1. Jeder dieser Vektoren $\mathbf{e}_j$ ist auch eine Basis eines 1-dimensionalen Teilraums des V_3. Dieser Teilraum ist die Gerade, in der $\mathbf{e}_j$ liegt. Denn mit $\mathbf{e}_j$ liegt auch $\lambda\mathbf{e}_j$ auf dieser Geraden, ebenso $(\lambda + \mu)\mathbf{e}_j$, mit $\lambda, \mu \in \mathbb{R}$.

2. Es sei $\mathbf{u} = (u_1, u_2, u_3)'$ ein beliebiger Vektor des V_3. Alle Vektoren mit der gleichen oder der entgegengesetzten Richtung von $\mathbf{u}$ bilden einen 1-dimensionalen Teilraum des V_3.

3. (Illustration zu Satz 1.4) Gegeben seien 2 Vektoren $\mathbf{u}$ und $\mathbf{v}$ mit verschiedener Orientierung. Sie bilden eine Basis eines 2-dimensionalen Teilraums des V_3. Der Teilraum ist die Ebene, in der die beiden Vektoren liegen, und alle Linearkombinationen dieser beiden Vektoren liegen in dieser Ebene.

   Denn sicher sind die zwei Vektoren linear unabhängig (vergl. Folgerung 4, p. 8). Man betrachte die Ebene, die durch den Ursprung $(0, 0, 0)$ des Koordinatensystems sowie durch die Endpunkte (u_1, u_2, u_3) und (v_1, v_2, v_3) der Vektoren $\mathbf{u}$ und $\mathbf{v}$ gegeben sind. Dann existieren Zahlen a_1, a_2 und a_3 derart, dass für jeden Punkt mit den Koordinaten x_1, x_2, x_3 in dieser Ebene die Gleichung

$$a_1x_1 + a_2x_2 + a_3x_3 = k \tag{26}$$

   Ohne Beschränkung der Allgemeinheit kann $k = 0$ gesetzt werden. Insbesondere gilt dann

$$\begin{aligned} 0 &= a_1u_1 + a_2u_2 + a_3u_3 \\ 0 &= a_1v_1 + a_2v_2 + a_3v_3, \end{aligned}$$

   denn die Endpunkte von $\mathbf{u}$ und $\mathbf{v}$ liegen ja in dieser Ebene. Zu zeigen ist, daß jede Linearkombination $\mathbf{w} = (w_1, w_2, w_3)' = \lambda\mathbf{u} + \mu\mathbf{v}$ von $\mathbf{u}$ und $\mathbf{v}$ ebenfalls in dieser Ebene liegt.

   Für die i-te Komponente von $\mathbf{w}$ gilt

$$w_i = \lambda u_i + \mu v_i, \quad i = 1, 2, 3,$$

   und nach Multiplikation der Gleichung mit a_i gilt natürlich auch

$$a_iw_i = \lambda a_iu_i + \mu a_iv_i.$$

Dann folgt

$$\sum_i a_i w_i = \lambda \sum_i a_i u_i + \mu \sum_i a_i v_i.$$

Aber nach Voraussetzung gilt $\sum_i a_i u_i = \sum_i a_i v_i = 0$, so daß $\sum_i a_i w_i = 0$ folgt. Also muß $\mathbf{w}$ ebenfalls in der Ebene liegen. Dies gilt für beliebige $\lambda, \mu \in \mathbb{R}$, also liegen alle Linearkombinationen von $\mathbf{u}$ und $\mathbf{v}$ in der von ihnen definierten Ebene. □

2 Matrizen

2.1 Operationen mit Matrizen

Schreibt man n m-dimensionale (Spalten-)Vektoren nebeneinander (oder m n-dimensionale Zeilenvektoren untereinander) an, so entsteht eine *Matrix*:

$$U = \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ u_{21} & u_{22} & \dots & u_{2n} \\ & & \vdots & \\ u_{m1} & u_{m2} & \dots & u_{mn} \end{pmatrix}. \tag{27}$$

Die Komponenten sind hier zweifach indiziert worden: der zweite Index gibt die Nummer des Spaltenvektors an, der erste die Nummer des Zeilenvektors. m und n heißen die *Dimensionen* der Matrix. Man schreibt auch kurz

$$U = (u_{ij}), \quad i = 1, 2, \dots, m; \quad j = 1, 2, \dots, n. \tag{28}$$

Die u_{ij} heißen die *Elemente* der Matrix. Für Matrizen lassen sich wieder die Multiplikation mit einem Skalar, Summe und Produkt definieren:

1. **Multiplikation mit einem Skalar λ:**

   $$\lambda U = (\lambda u_{ij}), \tag{29}$$

   d.h. es wird jedes Element u_{ij} mit λ multipliziert.

2. **Summe zweier Matrizen U und V:** es gilt

   $$U + V = (u_{ij} + v_{ij}), \tag{30}$$

   d.h. die Matrizen werden addiert, indem man die zueinander korrespondierenden Elemente addiert. Die Addition zweier Matrizen setzt voraus, daß sie gleiche Dimensionen haben.

3. **Das Produkt zweier Matrizen:** das Produkt zweier Matrizen entsteht, indem man das Skalarprodukt jeden Zeilenvektors $\mathbf{u}'_i$ von U mit jedem Spaltenvektor $\mathbf{v}_j$ von V bildet:

   $$U \cdot V = (\mathbf{u}'_i \mathbf{v}_j) = \left(\sum_{k=1}^{n} u_{ik} v_{kj} \right). \tag{31}$$

   Die Möglichkeit, das Produkt $U \cdot V$ (oder einfach UV) zu bilden, setzt demnach voraus, daß die Anzahl der Spalten von U gleich der Anzahl der Zeilen von V ist.

Die Multiplikation von Matrizen ist im allgemeinen *nicht kommutativ*, d.h. es gilt im allgemeinen

$$UV \neq VU. \tag{32}$$

4. **Die gestürzte oder transponierte Matrix:** Ebenso wie sich ein Vektor stürzen läßt, kann eine Matrix gestürzt werden. Dabei werden die Spalten als Zeilen angeschrieben. Man schreibt U' für die Gestürzte Matrix U. Für das Produkt $W = UV$ zweier Matrizen U und V gilt

$$W' = (UV)' = V'U' = T, \tag{33}$$

d.h. Die Gestürzte eines Produkts ist gleich dem Produkt der Gestürzten in umgekehrter Reihenfolge. Es sei w_{ij} das Element in der i-ten Zeile und j-ten Spalte von W, und es ist $w_{ij} = \mathbf{u}_i'\mathbf{v}_j$, $\mathbf{u}_i'$ der i-te Zeilenvektor von U, $\mathbf{v}_j$ der j-te Spaltenvektor von V. Offenbar ist $w_{ij} = t_{ji} = \mathbf{v}_j'\mathbf{u}_i$, $\mathbf{v}_j'$ der j-te Zeilenvektor von V' und $\mathbf{u}_i$ der i-te Spaltenvektor von U'; damit folgt (33).

5. **Die Einheitsmatrix I:** Dies ist die Matrix

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & 1 \end{pmatrix}. \tag{34}$$

I ist eine Diagonalmatrix, in deren Diagonalzellen jeweils die 1 steht, und in den Nicht-Diagonalzellen stehen Nullen. I spielt im Matrixkalkül die Rolle der 1 bei den "normalen" reellen Zahlen.

Bisher ist nur das Skalarprodukt zweier n-dimensionaler Vektoren eingeführt worden: $\mathbf{x}'\mathbf{y} = k \in \mathbb{R}$, d.h. das Skalarprodukt ist stets gleich einer reellen Zahl. Man kann auch das Produkt $\mathbf{x}\mathbf{y}'$ betrachten: es ist gleich einer Matrix:

Definition 2.1 *Es sei $\boldsymbol{x} = (x_1, \ldots, x_m)'$ und $\boldsymbol{y} = (y_1, \ldots, y_n)'$. Dann ist das Produkt $\boldsymbol{x}\boldsymbol{y}'$ durch die Matrix*

$$\boldsymbol{x}\boldsymbol{y}' = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} (y_1, y_2, \ldots, y_n) = \begin{pmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_n \\ x_2y_1 & x_2y_2 & \cdots & x_2y_n \\ & & \vdots & \\ x_my_1 & x_my_2 & \cdots & x_my_n \end{pmatrix} \tag{35}$$

definiert.

Bemerkungen: Man vergleiche die Reihenfolge von Zeilen- und Spaltenvektor mit der beim Skalarprodukt, Gleichung (7), Seite 4. Das Produkt $\mathbf{x}\mathbf{y}'$ ergibt nicht notwendig eine symmetrische Matrix; gilt $\mathbf{x} = \mathbf{y}$, so ist $\mathbf{x}\mathbf{y}'$ symmetrisch.

□

2.2 Transformation eines Vektors

Es sei T die (m,n)-Matrix

$$T = \begin{pmatrix} t_{11} & t_{12} & \dots & t_{1n} \\ t_{21} & t_{22} & \dots & t_{2n} \\ & & \vdots & \\ t_{m1} & t_{m2} & \dots & t_{mn} \end{pmatrix}, \tag{36}$$

und $\mathbf{x} = (x_1, \dots, x_n)'$. $\mathbf{x}$ ist der Spezialfall einer einspaltigen Matrix. Das Produkt $T\mathbf{x}$ ist dann ein Vektor $\mathbf{y}$:

$$\begin{aligned} T\mathbf{x} &= \begin{pmatrix} t_{11} & t_{12} & \dots & t_{1n} \\ t_{21} & t_{22} & \dots & t_{2n} \\ & & \vdots & \\ t_{m1} & t_{m2} & \dots & t_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \\ &= \begin{pmatrix} \sum_j t_{1j}x_j \\ \sum_j t_{2j}x_j \\ \vdots \\ \sum_j t_{mj}x_j \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \mathbf{y}. \end{aligned} \tag{37}$$

Für die Komponenten des Vektors $\mathbf{y}$ gilt also

$$y_i = \sum_{j=1}^{n} t_{ij}x_j, \quad i = 1, 2, \dots, m. \tag{38}$$

Durch Multiplikation des n-dimensionalen Vektors $\mathbf{x}$ von links mit der (m,n)-Matrix T entsteht der n-dimensionale Vektor $\mathbf{y}$; man sagt auch, daß $\mathbf{x}$ durch T in den Vektor $\mathbf{y}$ transformiert wird.

Gilt $\mathbf{y} = T\mathbf{x}$, $\mathbf{x} = (x_1, \dots, x_n)'$, so bedeutet dies, dass $\mathbf{y}$ als Linearkombination der Spaltenvektoren von T dargestellt wird; sind $\mathbf{T}_1, \dots, \mathbf{T}_n$ die Spaltenvektoren von T, so gilt also

$$\mathbf{y} = T\mathbf{x} = x_1\mathbf{T}_1 + \dots + x_n\mathbf{T}_n. \tag{39}$$

Dies folgt sofort aus (37). Denn es ist

$$\mathbf{y} = \begin{pmatrix} \sum_j t_{1j}x_j \\ \sum_j t_{2j}x_j \\ \vdots \\ \sum_j t_{mj}x_j \end{pmatrix} = \begin{pmatrix} t_{11} \\ t_{21} \\ \vdots \\ t_{m1} \end{pmatrix} x_1 + \begin{pmatrix} t_{12} \\ t_{22} \\ \vdots \\ t_{m2} \end{pmatrix} x_2 + \dots + \begin{pmatrix} t_{1n} \\ t_{2n} \\ \vdots \\ t_{mn} \end{pmatrix} x_n, \tag{40}$$

und dieser Ausdruck ist äquivalent mit dem Ausdruck (39).

Jetzt sei $\mathbf{x} = (x_1, \dots, x_m)'$ ein m-dimensionaler Vektor. Dann kann T von links mit $\mathbf{x}'$ multipliziert werden; $\mathbf{x}'$ ist der Spezialfall einer einzeiligen Matrix. Man hat dann

$$\begin{aligned} \mathbf{z}' = \mathbf{x}'T &= (x_1, \dots, x_m) \begin{pmatrix} t_{11} & t_{12} & \dots & t_{1n} \\ t_{21} & t_{22} & \dots & t_{2n} \\ & & \vdots & \\ t_{m1} & t_{m2} & \dots & t_{mn} \end{pmatrix} \\ &= \left(\sum_{i=1}^{n} x_i t_{i1}, \dots, \sum_{i=1}^{n} x_i t_{in} \right) = \mathbf{z}' = (z_1, \dots, z_m). \end{aligned} \tag{41}$$

Schreibt man den Vektor $\mathbf{z}$ als Spaltenvektor an, so erhält man also

$$\mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{pmatrix} = \begin{pmatrix} x_1t_{11} + \cdots + x_mt_{m1} \\ x_1t_{12} + \cdots + x_mt_{m2} \\ \vdots \\ x_1t_{1n} + \cdots + x_mt_{mn} \end{pmatrix}$$
$$= x_1 \begin{pmatrix} t_{11} \\ t_{12} \\ \vdots \\ t_{1n} \end{pmatrix} + \cdots + x_m \begin{pmatrix} t_{m1} \\ t_{m2} \\ \vdots \\ t_{mn} \end{pmatrix} \tag{42}$$

$\mathbf{z}$ ist demnach eine Linearkombination der Zeilenvektoren von T.

Zusammenfassung

Es sei T eine $(m \times n)$-Matrix. Ist $\mathbf{x}$ ein m-dimensionaler Vektor, so ist $\mathbf{z}' = \mathbf{x}'T$ eine Linearkombination der m Zeilenvektoren von T. Ist $\mathbf{x}$ ein n-dimensionaler Vektor, so ist $\mathbf{z} = T\mathbf{x}$ eine Linearkombination der n Spaltenvektoren von T.

Beispiel 2.1 Es sei $X = (x_{ij})$ eine Matrix mit Meßwerten x_{ij}, $i = 1, 2, \ldots, m$ und $j = 1, 2, \ldots, n$. Man kann für jede Spalte von X den Mittelwert und die Standardabweichung berechnen und damit zu den standardisierten Werten $z_{ij} = (x_{ij} - \bar{x}_j)/s_j$ übergehen, es entsteht die Matrix $Z = (z_{ij})$. Man überzeugt sich leicht, daß dann

$$\frac{1}{m} Z'Z = R = (r_{jk}), \quad j = 1, 2, \ldots, n, \quad k = 1, 2, \ldots, n \tag{43}$$

die Matrix der Korrelationen zwischen den Spalten von X ist. Man kann auch die Matrix $\hat{Z} = (1/\sqrt{m})Z$ definieren; dann ist $\hat{Z}'\hat{Z} = R$, oder $\hat{Z} = (1/\sqrt{m-1})Z$. In diesem Fall berücksichtigt man, daß für kleine Werte von m die Varianz-Kovarianz-Formel

$$\mathrm{Kov}(X_j, X_k) = \frac{1}{m} \sum_i (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$$

einen Bias hat. Für die Berechnung einer Faktorenanalyse sollte allerdings möglichst die Regel $m \approx 3n$, n die Anzahl der "Tests", befolgt werden, so daß die Biaskorrektur kaum notwendig wird.

Repräsentieren diese Spalten Meßwerte in irgendwelchen "Tests" (wirkliche Tests, oder Einschätzungen auf Ratingskalen), so enthält R die Korrelationen zwischen diesen Tests. Da $r_{ij} = r_{ji}$, ist R ein Beispiel für eine *symmetrische* Matrix. □

Beispiel 2.2 Es sei U eine Matrix mit den Dimensionen $m = n$; dann heißt U *quadratisch*. Die Spaltenvektoren von U seien alle paarweise orhogonal. Dann verschwinden alle Skalarprodukte zwischen ihnen, mit Ausnahme der Skalarprodukte der Vektoren mit sich selbst. Insbesondere gilt

$$U'U = \begin{pmatrix} |\mathbf{u}_1|^2 & 0 & \ldots & 0 \\ 0 & |\mathbf{u}_2|^2 & \ldots & 0 \\ & & \vdots & \\ 0 & 0 & \ldots & |\mathbf{u}_m|^2 \end{pmatrix} = D. \tag{44}$$

Die Matrix $D = U'U$ enthält also bis auf die Diagonalzellen nur Nullen; D heißt deshalb auch *Diagonalmatrix*. Gilt weiter $|\mathbf{u}_j| = 1$, so stehen in der Diagonalen nur Einsen; die Matrix heißt deshalb *Einheitsmatrix*, sie wird mit I bezeichnet (vergl. 5, Seite 13).

Sind die Spaltenvektoren einer quadratischen Matrix orthogonal, so folgt, daß auch die Zeilenvektoren orthogonal sind. Dazu betrachte man $V = U'$. Die Spaltenvektoren von U sind die Zeilenvektoren von V. Dann ist

$$V'V = (U')'U' = U''U' = UU' = (\sum_k u_{ik}u_{kj}).$$

Aber $\sum_k u_{ik}u_{kj} = \sum_k u_{kj}u_{ik}$, so daß $U'U = UU'$ (dies gilt nur für quadratische Matrizen). UU' ist also wieder eine Diagonalmatrix, d.h. auch die Zeilenvektoren sind orthogonal. U heißt dann *orthogonal*. Haben die Zeilen- und Spaltenvektoren darüber hinaus noch die Länge 1, so heißt U *orthonormal*. □

2.3 Der Rang einer Matrix

Die folgenden Betrachtungen über den Zeilen- und Spaltenrang einer Matrix mögen auf den ersten Blick ein wenig abstrakt und, für die Zwecke der Darstellung des faktorenanalytischen Ansatzes, ein wenig sehr grundsätzlich sein. Tatsächlich aber zeigt sich, dass der Zusammenhang zwischen Zeilen- und Spaltenrang gerade dem grundlegenden Ansatz der Faktorenanalyse entspricht. Deswegen wird dieser Zusammenhang in diesem Abschnitt etwas ausführlicher besprochen.

Die m-dimensionalen Vektoren $\mathbf{u}_1, \ldots, \mathbf{u}_n$ können zu einer $m \times n$-Matrix

$$U = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n] = \begin{pmatrix} u_{11} & u_{12} & \ldots & u_{1n} \\ u_{21} & u_{22} & \ldots & u_{2n} \\ & & \vdots & \\ u_{m1} & u_{m2} & \ldots & u_{mn} \end{pmatrix} \tag{45}$$

zusammengefasst werden. Liegen die $\mathbf{u}_1, \ldots, \mathbf{u}_n$ in einem r_s-dimensionalen Teilraum, so sagt man, die Matrix U habe den *Spaltenrang* r_s. Die Vektoren $\mathbf{u}_1, \ldots, \mathbf{u}_n$ können dann als Linearkombinationen von r_s linear unabhängigen, m-dimensionalen Vektoren dargestellt werden.

Ebenso kann man der Matrix U einen *Zeilenrang* r_z zuordnen; r_z ist die Dimensionaltität des Teilraumes, in dem die n-dimensionalen Zeilenvektoren liegen, und die Zeilenvektoren können als Linearkombinationen von r_z linear unabhängigen n-dimensionalen Vektoren dargestellt werden. Es gilt nun der

Satz 2.1 *Es gilt stets*

$$r_z = r_s = r \leq \min(m, n). \tag{46}$$

Beweis: Nach Voraussetzung existieren linear unabhängige n-dimensionale Vektoren $\mathbf{b}_1, \ldots, \mathbf{b}_{r_z}$ derart, dass die n-dimensionalen Zeilenvektoren $\mathbf{u}_i$ in der Form

$$\mathbf{u}_i = (u_{i1}, \ldots, u_{ij}, \ldots, u_{in})' = q_{i1}\mathbf{b}_1 + \cdots + q_{ir_z}\mathbf{b}_{r_z}, \quad i = 1, \ldots, m \tag{47}$$

geschrieben werden können, wobei

$$u_{ij} = \sum_{k=1}^{r_z} q_{ik}b_{kj}, \tag{48}$$

und b_{kj} ist die k-te Komponente von $\mathbf{b}_j$. Die Koeffizienten $q_{ik} \in \mathbb{R}$, $k = 1, \dots, r_z$ sind die "Gewichte", mit denen die $\mathbf{b}_k$ in den jeweiligen Zeilenvektor $\mathbf{u}_i$ eingehen.

$$(u_{i1}, \dots, u_{ij}, \dots, u_{in})$$

ist die i-te Zeile von U. Die j-te Spalte von U läßt sich dann in der Form

$$\begin{aligned}
\mathbf{v}_j &= \begin{pmatrix} u_{1j} \\ \vdots \\ u_{ij} \\ \vdots \\ u_{mj} \end{pmatrix} = \begin{pmatrix} q_{11}b_{1j} + q_{12}b_{2j} + \cdots + q_{1r_z}b_{r_zj} \\ \vdots \\ q_{i1}b_{1j} + q_{i2}b_{2j} + \cdots + q_{ir_z}b_{r_zj} \\ \vdots \\ q_{m1}b_{1j} + q_{m2}b_{2j} + \cdots + q_{mr_z}b_{r_zj} \end{pmatrix} \\
&= b_{1j} \begin{pmatrix} q_{11} \\ \vdots \\ q_{i1} \\ \vdots \\ q_{m1} \end{pmatrix} + \cdots + b_{r_zj} \begin{pmatrix} q_{1r_z} \\ \vdots \\ q_{ir_z} \\ \vdots \\ q_{mr_z} \end{pmatrix} = b_{1j}\mathbf{q}_1 + \cdots + b_{r_zj}\mathbf{q}_{r_z} \qquad (49)
\end{aligned}$$

darstellen. Die Spaltenvektoren $\mathbf{v}_j$ werden also als Linearkombinationen von r Vektoren $\mathbf{q}_k$, $k = 1, \dots, r_z$ repräsentiert, deren Komponenten gerade die "Gewichte" q_{ik} der Vektoren $\mathbf{b}_k$ sind, deren Linearkombinationen die Zeilenvektoren $\mathbf{u}_i$ liefern. Die Komponenten der $\mathbf{b}_k$ sind jetzt die "Gewichte" der $\mathbf{q}_k$. Zu zeigen ist, dass die $\mathbf{q}_k$ linear unabhängig sind.

Dazu werde angenommen, dass sie es *nicht* sind; diese Annahme wird dann auf einen Widerspruch führen, aus dem die lineare Unabhängigkeit folgt. Nimmt man also an, dass die $\mathbf{q}_k$ linear abhängig sind, so folgt aus $\underline{0} = \mu_1\mathbf{q}_1 + \cdots + \mu_{r_z}\mathbf{q}_{r_z}$ dann

$$\mu_{r_z}\mathbf{q}_{r_z} = -(\mu_1\mathbf{q}_1 + \cdots + \mu_{r_z-1}\mathbf{q}_{r_z-1}),$$

und für das i-te Element von $\mathbf{q}_{r_z}$ folgt insbesondere

$$q_{ir_z} = -(\frac{\mu_1}{\mu_{r_z}}q_{i1} + \cdots + \frac{\mu_{r_z-1}}{\mu_{r_z}}q_{i,r_z-1}).$$

Aus (47) erhält man dann

$$\begin{aligned}
\mathbf{v}_i &= q_{i1}\mathbf{b}_1 + \cdots - (\frac{\mu_1}{\mu_{r_z}}q_{i1} + \cdots + \frac{\mu_{r_z-1}}{\mu_{r_z}}q_{i,r_z-1})\mathbf{b}_{r_z} \\
&= q_{i1}(\mathbf{b}_1 - \frac{\mu_1}{\mu_{r_z}}q_{i1}\mathbf{b}_{r_z}) + \cdots + q_{i,r_z-1}(\mathbf{b}_{r_z-1} - \frac{\mu_{r_z-1}}{\mu_{r_z}}q_{i,r_z-1}\mathbf{b}_{r_z}) \qquad (50)
\end{aligned}$$

Damit würde dann aber $\mathbf{v}_i$ als Linearkombination von $r_z - 1$ Vektoren

$$(\mathbf{b}_1 - \frac{\mu_1}{\mu_{r_z}}q_{i1}\mathbf{b}_{r_z}), \dots, (\mathbf{b}_{r_z-1} - \frac{\mu_{r_z-1}}{\mu_{r_z}}q_{i,r_z-1}\mathbf{b}_{r_z})$$

dargestellt werden, im Widerspruch zur Annahme, dass r_z Vektoren zur Darstellung der $\mathbf{v}_i$ notwendig sind. Also können die $\mathbf{q}_j$ nicht linear abhängig sein; sie müssen linear unabhängig sein. Damit folgt aber auch $r_z = r_s = r$. □

Bemerkung: Nach Satz 2.1 kann einer Matrix U ganz allgemein ein Rang r zugeordnet werden. Gilt $r = n$, so sind alle Spaltenvektoren linear unabhängig. Gilt aber $r < n$, so sind *nicht alle* Spaltenvektoren linear unabhängig. Dann existieren irgendwelche r linear unabhängigen, m-dimensionalen Vektoren $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_r$ derart, daß jeder der m-dimensionalen Spaltenvektoren $\mathbf{u}_j$ als Linearkombination der $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_r$ dargestellt werden kann. Die $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_r$ bilden dann eine *(Teil-)Basis* des V_m für die Spaltenvektoren von U. Gleichzeitig können die Zeilenvektoren als Linearkombinationen von r linear unabhängigen, n-dimensionalen Vektoren $\mathbf{b}_1^*, \ldots, \mathbf{v}_r^*$ dargestellt werden, die dann eine Teilbasis des V_n bilden. Die $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_r$ bzw. die $\mathbf{b}_1^*, \ldots, \mathbf{v}_r^*$ können, müssen aber nicht orthogonal sein; wichtig ist nur, dass sie linear unabhängig sind. □

Man kann die Vektoren $\mathbf{q}_1, \ldots, \mathbf{q}_r$ zu einer Matrix Q zusammenfassen:

$$Q = [\mathbf{q}_1, \ldots, \mathbf{q}_r] = \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1r} \\ q_{21} & q_{22} & \cdots & q_{2r} \\ & & \vdots & \\ q_{m1} & q_{m2} & \cdots & q_{mr} \end{pmatrix}. \tag{51}$$

Ebenso kann man die Vektoren $\mathbf{b}_1, \ldots, \mathbf{b}_r$ zu einer Matrix B zusammenfassen:

$$B = [\mathbf{b}_1, \ldots, \mathbf{b}_r] = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1r} \\ b_{21} & b_{22} & \cdots & b_{2r} \\ & & \vdots & \\ b_{n1} & b_{n2} & \cdots & b_{nr} \end{pmatrix}. \tag{52}$$

Der Vektor $\mathbf{u}_j$ ist dann in der Form

$$\mathbf{u}_j = Q\mathbf{b}_j \tag{53}$$

darstellbar. Gemäß (39) bedeutet diese Gleichung, dass $\mathbf{u}_j$ eine Linearkombination der Spaltenvektoren von Q ist.

Die folgende, zusammenfassende Aussage wird wegen der leichteren Bezugnahme als Satz formuliert:

Satz 2.2 *Hat die (m, n)-Matrix U den Rang $r \leq \min(m, n)$, so existieren stets die (m, r)-Matrix Q und die (n, r)-Matrix B, mit jeweils linear unabhängigen Spaltenvektoren, derart, dass*

$$U = QB' \tag{54}$$

gilt.

Anmerkungen:

1. Man kann das Element u_{ij}, also das Element in der i-ten Zeile und j-ten Spalte gemäß der Gleichung (31) ausdrücken:

$$u_{ij} = q_{i1}b_{j1} + q_{i2}b_{j2} + \cdots + q_{ir}b_{jr} \tag{55}$$

wobei $b_{j1}, \ldots, b_{jr}$ die j-te Zeile der Matrix B bzw. die j-te Spalte von B' ist. Repräsentieren die Zeilen von U Personen und die Spalten von U "Tests", d.h. bei allen Personen gemessene Variablen, so zeigt (55), wie sich ein individueller Meßwert, nämlich der bei der i-ten Person im j-ten "Test", zusammensetzt. Die Koeffizienten

$q_{i1}, \ldots, q_{ir}$ definieren gerade die i-te Zeile der Matrix Q und repräsentieren die i-te Person auf den r "latenten" Dimensionen. Die $b_{j1}, \ldots, b_{jr}$ repräsentieren den j-ten "Test" auf eben diesen latenten Dimensionen, und nach (55) ist der Meßwert u_{ij} gerade das Skalarprodukt, also die Korrelation, zwischen der Repräsentation der i-ten Person und des j-ten Tests. Dies ist der Ansatz der Faktorenanalyse, wobei allerdings noch ein zusätzlicher additiver "Fehler" e_{ij} angenommen wird, der in (55) noch nicht auftaucht. Weitere Implikationen der Gleichung (55) werden bei der Behandlung der Faktorenanalyse diskutiert.

   Man beachte, dass die Spaltenvektoren von Q und B nicht orthogonal sein müssen. Der Spezialfall orthogonaler Vektoren wird in Abschnitt 2.6 behandelt.

2. Nach (33) gilt $U' = (QB')' = BQ'$. Die Matrix $U'U$ enthält alle Skalarprodukte der Spaltenvektoren von U; man hat dann

$$U'U = BQ'QB'. \tag{56}$$

   Hierin ist $Q'Q$ die Matrix der Skalarprodukte für die Spaltenvektoren $\mathbf{q}_1, \ldots, \mathbf{q}_r$ von Q. So lange man nur fordert, dass die $\mathbf{q}_j$, $j = 1, 2, \ldots, r$ linear unabhängig sind, müssen die Skalarprodukte $\mathbf{q}_j'\mathbf{q}_k$, $j \neq k$, nicht gleich Null sein. Für den Fall, dass man insbesondere orthogonale Vektoren $\mathbf{q}_j$ als Teilbasis des V_m wählt, folgt aber, dass für alle $j \neq k$ die Skalarprodukte gleich Null sein müssen. $Q'Q$ ist dann eine Diagonalmatrix, in deren Diagonale die Quadrate $|\mathbf{v}_j|^2$ der Längen der $\mathbf{v}_j$ stehen. Auf diesen Spezialfall wird weiter unten genauer eingegangen.

Einige Sätze über den Rang von Matrixprodukten:

Satz 2.3 *Es sei A eine beliebige (m,k)-Matrix und B sei eine beliebige (k,n)-Matrix. Dann ist der Rang $rg(AB)$ der Matrix AB kleiner/gleich dem kleineren der Ränge $rg(A)$ oder $rg(B)$, d.h. $rg(AB) \leq \min[rg(A), rg(B)]$.*

Beweis: Die Zeilen von AB sind Linearkombinationen der Zeilen von B, so dass die Anzahl l.u. Zeilenvektoren kleiner oder höchstens gleich der Anzahl l.u. Vektoren von B ist, also muss $rg(AB) \leq rg(B)$ sein. Die Spalten von AB sind aber Linearkombinationen der Spalten von A, also muss $rg(AB) \leq rg(A)$ gelten. Beide Aussagen zusammen ergeben den Satz. □

Satz 2.4 *Es sei P eine (n,n)-Matrix mit dem Rang $rg(P) = n$, und Q sei eine (m,m)-Matrix mit dem Rang $rg(Q) = m$. Dann gilt $rg(PAQ) = rg(A)$.*

Beweis: Wegen Satz 2.3 gilt $rg(A) \geq rg(AQ) \geq rg(AQQ^{-1})$, d.h. aber $rg(AQ) = rg(A)$. Weiter gilt $rg(AQ) \geq rg(PAQ) \geq rg(P^{-1}PAQ)$, d.h. aber $rg(AQ) = rg(PAQ) = rg(A)$. □

Satz 2.5 *Es sei D eine (n,n)-Diagonalmatrix mit $r < n$ von Null verschiedenen Diagonalelementen, d.h.*

$$D = \begin{pmatrix} a_1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & a_2 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \vdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & a_r & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix}, \quad a_j \neq 0, \quad j = 1, \ldots r.$$

Dann hat D den Rang r.

Beweis: Die ersten r Spaltenvektoren $\mathbf{d}_1, \ldots, \mathbf{d}_r$ von D sind sicher linear unabhängig, da $\underline{0} = \alpha_1 \mathbf{d}_1 + \cdots + \alpha_r \mathbf{d}_r$ nur möglich ist, wenn $\alpha_1 = \cdots = \alpha_r = 0$; also ist $rg(D) \geq r$. Andererseits ist $\underline{0} = \alpha_1 \mathbf{d}_1 + \cdots + \alpha_r \mathbf{d}_r + \alpha_{r+1} \mathbf{d}_{r+1} = \alpha_1 \mathbf{d}_1 + \cdots + \alpha_r \mathbf{d}_r + \alpha_{r+1} \underline{0}$ möglich für $\alpha_{r+1} \neq 0$, so dass $r + 1$ Spaltenvektoren von D linear abhängig sind. Dann folgt $rg(D) = r$. □

2.4 Die Inverse einer Matrix

Es sei A eine beliebige quadratische Matrix, d.h. A habe so viele Zeilen wie Spalten. Insbesondere habe A n Zeilen wie Spalten, und der Rang von A sei gleich n. Dann existiert eine Matrix $\tilde{A}$ mit ebenfalls n Zeilen und Spalten derart, daß $A\tilde{A} = I$, I die Einheitsmatrix. Ist eine a eine reelle Zahl, so ist $a(1/a) = aa^{-1} = 1$. Die Rolle der 1 ist bei den Matrizen die Einheitsmatrix I, vergl. (34). Deshalb schreibt man auch A^{-1} für die inverse Matrix $\tilde{A}$, also

$$AA^{-1} = A^{-1}A = I. \tag{57}$$

> Die inverse Matrix für eine gegebene *quadratische* Matrix A existiert nur, wenn alle Zeilen- bzw. Spaltenvektoren der Matrix linear unabhängig sind!
>
> Für nichtquadratische Matrizen gibt es keine Inverse, die die Beziehung (57) erfüllt. Für solche Matrizen kann man die *generalisierte Inverse* einführen, worauf hier aber nicht eingegangen werden kann.

Satz 2.6 *Es sei A eine symmetrische $(n \times n)$-Matrix, d.h. A ist quadratisch und es gilt $A = A'$. Weiter seien die Zeilen- bzw. Spaltenvektoren von A orthogonal und auf die Länge 1 normiert (die Matrix heißt dann* orthonormal*). Dann gilt*

$$A' = A^{-1}, \tag{58}$$

d.h. die zu A inverse Matrix ist einfach durch die "gestürzte" Matrix A' gegeben, so daß

$$A'A = AA' = I. \tag{59}$$

Beweis: Es seien A_j und A_k $j \neq k$, zwei beliebige Spaltenvektoren von A. Nach Voraussetzung sind sind orthogonal, so daß $A'_j A_k = 0$. Nach Voraussetzung sind sie auch *normal*, d.h. die Vektoren haben die Länge 1. Für $j = k$ gilt also $A'_j A_j = 1$. Folglich muß $A'A$ die Einheitsmatrix sein; ein analoges Argument gilt für AA'. Aus der Definition der Inversen folgt dann, daß $A' = A^{-1}$. □

2.5 Die Eigenvektoren einer Matrix

Die Transformation eines Vektors. Es sei $\mathbf{p} = (p_1, \ldots, p_n)'$ ein n-dimensionaler Vektor, und $T = (t_{ij})$ eine $(m \times n)$-Matrix, d.h. $i = 1, 2, \ldots, m$ und $j = 1, 2, \ldots, n$. Das Skalarprodukt des i-ten Zeilenvektors $(t_{i1}, t_{i2}, \ldots, t_{in})'$ mit dem Vektor $\mathbf{p}$ sei durch

$$y_i = \sum_{j=1}^{n} t_{ij} p_j$$

gegeben. Bildet man alle Skalarprodukte für $i = 1, \ldots, m$, so erhält man einen Vektor $\mathbf{q} = (q_1, q_2, \ldots, q_m)'$. In Matrixform gilt dann

$$T\mathbf{p} = \mathbf{q}. \tag{60}$$

Man kann sagen, daß $\mathbf{p}$ durch Multiplikation mit T in den Vektor $\mathbf{q}$ übergeht; insofern ist $\mathbf{q}$ eine *Transformation* von $\mathbf{p}$. Dementsprechend heißt T auch *Transformationsmatrix.* Für $m \neq n$ geht der n-dimensionale Vektor $\mathbf{q}$ in einen m-dimensionalen Vektor $\mathbf{p}$ über. Damit das Produkt $T\mathbf{p}$ gebildet werden kann, muß nur gewährleistet sein, dass die Anzahl der Spalten von T gleich der Anzahl der Komponenten von $\mathbf{p}$ ist.

Eigenvektoren. Ist T irgendeine (m, n)-Matrix und $\mathbf{p}$ ein beliebiger n-dimensionaler Vektor, so wird der Vektor $T\mathbf{p} = \mathbf{q}$ eine andere Länge und i.a. eine andere Richtung als $\mathbf{p}$ haben; für $m \neq n$ wird er sich in der Dimensionalität von $\mathbf{p}$ unterscheiden. Ein Spezialfall wird in der folgenden Definition spezifiziert:

Definition 2.2 *Es werde die Matrix T so gewählt, daß*

$$T\boldsymbol{p} = \lambda\boldsymbol{p} \tag{61}$$

gilt, d.h. es soll $\boldsymbol{q} = \lambda\boldsymbol{p}$ gelten. Dann heißt $\boldsymbol{p}$ ein Rechtseigenvektor *von T, und λ ist der zugehörige* Eigenwert. *Gilt andererseits*

$$\boldsymbol{p}'T = \lambda\boldsymbol{p}', \tag{62}$$

so heißt $\boldsymbol{p}$ Linkseigenvektor *von P, und λ ist wieder der zugehörige Eigenwert.*

Die Gleichung (61) impliziert zunächst, dass T quadratisch, also eine (n, n)-Matrix sein muß, denn sonst kann $\mathbf{q} = \lambda\mathbf{p}$ nicht gelten. Diese Beziehung bedeutet, daß sich $\mathbf{q}$ nur durch eine Streckung oder Stauchung von $\mathbf{p}$ unterscheidet; die beiden Vektoren haben also die gleiche Orientierung. Analoge Aussagen gelten für Linkseigenvektoren.

Der Ausdruck *Eigenvektor* ergibt sich aus der gewissermaßen umgekehrten Betrachtung: für eine gegebene Matrix T sucht man diejenigen Vektoren, die der Bedingung (61) genügen. Für gegebene Matrix T wird ein *beliebiger* Vektor $\mathbf{p}$ der Bedingung (61) nicht genügen. Es müssen also spezielle Vektoren $\mathbf{p}$ sein, die für die gegebene Matrix T der Bedingung (61) erfüllen. In Abschnitt 2.7 wird gezeigt, daß die Eigenvektoren symmetrischer Matrizen die Richtungen der Hauptachsen von Ellipsoiden definieren. Dieser Sachverhalt stellt nicht nur eine geometrische Veransschaulichung des Eigenvektorbegriffs dar, sondern liefert auch eine Beziehung zur multivariaten Normalverteilung, die für Fragen der Parameterschätzung und des Tests von Hypothesen von Bedeutung ist (vergl. Abschnitt 2.10).

Es werde angenommen, dass es möglicherweise mehr als einen Eigenvektor gibt. Dazu sei $P = [\mathbf{p}_1, \ldots, \mathbf{p}_r]$ die (n, r)-Matrix der Eigenvektoren der symmetrischen (n, n)-Matrix T, und

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & \\ 0 & \lambda_2 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \tag{63}$$

sei die Diagonalmatrix der zugehörigen Eigenwerte. Man rechnet leicht nach, dass dann

$$TP = P\Lambda \tag{64}$$

gilt.

Die Eigenvektoren $\mathbf{p}_k$ können als normiert betrachtet werden, d.h. man kann festsetzen, dass sie die Länge 1 haben sollen. Tatsächlich ist die Länge eines Eigenvektors für die Beziehung (61) unerheblich. Um dies zu sehen, sei von einem Vektor $\mathbf{p}$ bereits nachgewiesen worden, dass der die Beziehung $T\mathbf{p} = \lambda\mathbf{p}$ erfüllt, so dass $\mathbf{p}$ ein Eigenvektor von T ist. Nun sei $a \in \mathbb{R}$ ein Skalar. Dann ist auch $a\mathbf{p}$ ein Eigenvektor, denn $T(a\mathbf{p}) = aT\mathbf{p} = \lambda(a\mathbf{p}) = a\lambda T$, und der Faktor a kürzt sich wieder heraus. Die Länge eines Vektors ist also in Bezug auf die Eigenschaft, Eigenvektor von T zu sein, irrelevant; in der Tat bezieht sich die Definition von Eigenvektoren (i) auf die Orientierung, (ii) auf das Verhältnis der Komponenten von $\mathbf{p} = (p_1, \ldots, p_n)'$ und $\mathbf{q} = (q_1, \ldots, q_n)' = \lambda\mathbf{p}$; es soll eben $q_k/p_k = \lambda$ für $k = 1, \ldots, n$ gelten, die tatsächliche Länge von $\mathbf{p}$ und $\mathbf{q}$ ist dafür irrelevant. Setzt man also die Länge der Eigenvektoren fest gleich 1, so ändert dies nichts an der Eigenschaft, eben Eigenvektoren von T zu sein, hat aber den Vorteil, dass eine Reihe von Rechnungen vereinfacht wird.

Einen für die multivariate Statistik wichtigen Spezialfall bilden die symmetrischen Matrizen (so sind z.B. Korrelationsmatrizen symmetrisch). Im folgenden Satz werden Eigenschaften der Eigenvektoren und Eigenwerte symmetrischer Matrizen zusammengefasst.

Satz 2.7 *Es sei T eine reelle, symmetrische $(n \times n)$-Matrix, so daß $T = T'$ gilt. Dann gilt*

1. *Die Menge der Rechtseigenvektoren ist gleich der Menge der Linkseigenvektoren,*
2. *Je zwei verschiedene Eigenvektoren $\boldsymbol{x}_j$ und $\boldsymbol{x}_k$ mit zugehörigen verschiedenen Eigenwerten λ_j und λ_k sind orthogonal zueinander, d.h. die Eigenvektoren einer symmetrischen Matrix sind linear unabhängig (vergl. Folgerung 3, Seite 7).*
3. *die Anzahl der von Null verschiedenen Eigenwerte λ_j von T gleich dem Rang r, d.h. $\lambda_j \neq 0$, $j = 1, 2, \ldots, r$,*
4. *Für alle Eigenwerte gilt $\lambda_j \geq 0$, $j = 1, 2, \ldots, r$.*

Beweis: (1) ist leicht einzusehen. Es sei nämlich $T\mathbf{x} = \lambda\mathbf{x}$, d.h. X sei ein Eigenvektor von T mit dem Eigenwert λ. Dann ist $(T\mathbf{x})' = \mathbf{x}'T' = \mathbf{x}'T = \lambda\mathbf{x}'$, da ja $T' = T$ vorausgesetzt wurde. Das heißt aber, das Rechts- und Linkseigenvektoren identisch sind.

(2): Es gelte $T\mathbf{x} = \lambda\mathbf{x}$ und $T\mathbf{y} = \mu\mathbf{y}$, $\lambda \neq \mu$. Dann gilt sicher $\mathbf{y}'T\mathbf{x} = \lambda\mathbf{y}'T$, und $\mathbf{x}'T\mathbf{y} = \mu\mathbf{x}'\mathbf{y}$. Aber $(\mathbf{y}'T\mathbf{x})' = \mathbf{x}'T'\mathbf{y} = \mathbf{x}'T\mathbf{y}$, also muß $\lambda\mathbf{y}'\mathbf{x} = \mu\mathbf{x}'\mathbf{y}$ folgen. Das heißt aber $\lambda\mathbf{y}'\mathbf{x} - \mu\mathbf{x}'\mathbf{y} = (\lambda - \mu)\mathbf{y}'\mathbf{x} = 0$, denn es ist ja $\mathbf{y}'\mathbf{x} = \mathbf{x}'\mathbf{y}$. Da aber $\lambda - \mu \neq 0$ vorausgesetzt wurde, muß jetzt $\mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x} = 0$ folgen, und das bedeutet die Orthogonalität der Eigenvektoren $\mathbf{x}$ und $\mathbf{y}$.

(3): Es sei P die Matrix der Eigenvektoren von T; sie können ohne Beschränkung der Allgemeinheit als normiert vorausgesetzt werden. Aus (64), d.h. aus $TP = P\Lambda$, folgt demnach durch Multiplikation von links mit P'

$$P'TP = \Lambda, \tag{65}$$

da ja $P'P = PP' = I$. Aus (65) und dem Satz 2.4 folgt sofort $rg(P'TP) = rg(T) = rg(\Lambda)$, und da Λ eine Diagonalmatrix ist, bedeutet dies nach Satz 2.5, dass genau r der Eigenwerte ungleich Null sind.

(4) Es sei $G = \Lambda^{1/2}P'$ mit $\Lambda^{1/2} = \text{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})'$; dann gilt sicherlich $\Lambda = \Lambda^{1/2}\Lambda^{1/2}$. Aus $T = P\Lambda P'$ folgt dann $T = G'G$ und $\mathbf{x}'T\mathbf{x} = \mathbf{x}'G'G\mathbf{x}$, wobei der Vektor $\mathbf{x}$ ein beliebiger reeller Vektor sein darf. Setzt man $\mathbf{y} = G\mathbf{x}$, so sieht man, dass $\mathbf{y}'\mathbf{y} = \sum_i y_i^2 = \mathbf{x}'G'G\mathbf{x} = \mathbf{x}'T\mathbf{x} \geq 0$. Dann gilt aber auch $\mathbf{x}'T\mathbf{x} = \mathbf{x}'P\Lambda P'\mathbf{x}$. Setzt man nun $\mathbf{x} = P'\mathbf{z}$, so hat man wegen (65)

$$\mathbf{x}'T\mathbf{x} = \mathbf{z}'PTP'\mathbf{z} = \mathbf{z}'\Lambda\mathbf{z} = \lambda_1 z_1^2 + \cdots + \lambda_n z_n^2 \geq 0.$$

Diese Beziehung muß *für alle* $\mathbf{x}$ und damit $\mathbf{z}$ gelten, und die λ_j sind von der speziellen Wahl von $\mathbf{z}$ unabhängig. Also muß diese Bedingung auch für den Einheitsvektor $\mathbf{z} = \mathbf{e}_1 = (1, 0, \ldots, 0)'$ gelten. Dann folgt aber

$$\mathbf{x}'T\mathbf{x} = \lambda_1 \geq 0.$$

Ebenso kann man $\mathbf{z} = \mathbf{e}_2 = (0, 1, 0, \ldots, 0)'$ setzen; dann folgt

$$\mathbf{x}'T\mathbf{x} = \lambda_2 \geq 0,$$

etc. Dies bedeutet aber, dass alle Eigenwerte größer oder gleich Null, d.h. nicht negativ sind. $\square$

Anmerkungen:

1. Nach (64) gilt $TP = P\Lambda$. Da die Eigenvektoren in P orthonormal sind, folgt durch Multiplikation von rechts mit P' die Beziehung $T = P\Lambda P'$; zusammenfassend gilt also

$$P'TP = \Lambda \tag{66}$$
$$T = P\Lambda P'. \tag{67}$$

   Eine beliebige symmetrische Matrix T läßt sich also als ein Produkt der Matrix P der Eigenvektoren und der Diagonalmatrix Λ der Eigenwerte darstellen.

2. Die Matrix T sei symmetrisch. Die Größe $\mathbf{x}'T\mathbf{x}$ heißt *quadratische Form*; in Abschnitt 2.7 wird deutlich, dass quadratische Formen Elipsen definieren.

3. Gilt für eine symmetrische T, dass $\mathbf{x}'T\mathbf{x} \geq 0$ für alle reellen Vektoren $\mathbf{x}$, so heißt T *positiv-semidefinit*. Nach Satz 2.7 sind reelle symmetrische Matrizen offenbar positiv-semidefinit.

Anwendung: Nach (33) gilt $U' = (QB')' = BQ'$, so dass nach (56)

$$U'U = BQ'QB'.$$

gilt. Nimmt man an, dass die linear unabhängigen Spaltenvektoren von Q insbeondere orthogonal sind, folgt, dass $Q'Q = \Lambda$ der Form (63) ist, so dass

$$U'U = B\Lambda B' \tag{68}$$

folgt. Der Vergleich mit (67) zeigt dann, dass B die Matrix der Eigenvektoren von $U'U$ sein muß! Nach (31) soll aber $U = QB'$ gelten. Ist B die Matrix der Eigenvektoren von $U'U$, so folgt, dass $C = B'B$ eine Diagonalmatrix sein muss, da diese Vektoren orthogonal sind. Durch Multiplikation von U von rechts mit U' erhält man dann $UU' = QB'BQ' = QCQ'$. Damit ist die rechte Seite dieser Gleichung wieder von der Form (67), so dass man folgern kann, dass die Vektoren in Q ebenfalls Eigenvektoren sind, und zwar die der symmetrischen Matrix UU'. Diese Sachverhalte werden in Abschnitt 2.6 noch einmal in etwas anderer Form diskutiert. □

2.6 Die Singularwertzerlegung

Eine $(m \times n)$-Matrix U kann als Zusammenfassung von n m-dimensionalen Spaltenvektoren oder von m n-dimensionalen Zeilenvektoren aufgefaßt werden. Der Rang von U sei $r \leq \min(m, n)$, d.h. die Spalten- oder die Zeilenvektoren können als Linearkombinationen von jeweils r linear unabhängigen Vektoren dargestellt werden. Für die psychologische Praxis bedeutet dies z.B., daß die Messungen von n Tests bei m Personen letzlich durch r "latente" Variable (dies sind die linear unabhängigen Vektoren) dargestellt werden, die eine nicht-redundante Beschreibung der Messergebnisse erlauben. Der folgende Satz ist die Grundlage für das Auffinden der latenten Variablen.

Satz 2.8 *Es sei* $\mathbf{u}$ *eine beliebige* $(m \times n)$*-Matrix mit dem Rang* r*. Dann gilt*

$$U = Q\Lambda^{1/2}P', \tag{69}$$

wobei $Q = [Q_1, Q_2, \ldots, Q_r]$ *eine Matrix ist, deren Spaltenvektoren die* r *orthonormalen Eigenvektoren von* UU' *sind,* $P = [P_1, \ldots, P_r]$ *ist eine Matrix mit den* r *orthonormalen Eigenvektoren von* $U'U$ *und*

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_2 & \cdots & 0 & 0 \\ 0 & 0 & \vdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & \lambda_r \end{pmatrix} \tag{70}$$

ist die Diagonalmatrix der r *von Null verschiedenen Eigenwerte von* UU' *bzw.* $U'U$*, und*

$$\Lambda^{1/2} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 & 0 \\ 0 & 0 & \vdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & \sqrt{\lambda_r} \end{pmatrix}. \tag{71}$$

Beweis: U sei aus den Spaltenvektoren $\mathbf{u}_j$, $1 \leq j \leq n$ gebildet worden, d.h. $U = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$. Es existieren stets $r \leq \min(m, n)$ linear unabhängige, insbesondere orthogonale Vektoren $\mathbf{L}_k$, $k = 1, \ldots, r$, aus denen die $\mathbf{u}_j$ als Linearkombinationen gebildet werden können, d.h.

$$\mathbf{u}_j = p_{j1}\mathbf{L}_1 + \cdots + p_{jr}\mathbf{L}_r. \tag{72}$$

Die $\mathbf{L}_k$ können zu einer Matrix $L = [\mathbf{L}_1, \ldots, \mathbf{L}_r]$ zusammengefaßt werden, und die *Zeilenvektoren* $(\mathbf{p}_{j1}, \mathbf{p}_{j2}, \ldots, \mathbf{p}_{jr})$ zu einer Matrix P, so daß (72) in der Form

$$U = LP' \tag{73}$$

geschrieben werden kann. Die Längen der Vektoren $\mathbf{L}_k$ werden nicht spezifiert, Spaltenvektoren von P können als normiert (d.h. sie haben die Länge 1) vorausgesetzt werden. Denn angenommen, die Spaltenvektoren $\mathbf{P}_k$ von P hätten die Längen $\tilde{\lambda}_k^{1/2} \neq 1$. Dann enthält $B = P\tilde{\Lambda}^{-1/2}$, $\tilde{\Lambda} = \text{diag}(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_k)$, normierte Spaltenvektoren, und (73) kann in der Form $U = L\tilde{\Lambda}^{-1/2}B' = \tilde{L}B'$ geschrieben werden. Die Spaltenvektoren von $\tilde{L}$ sind wieder orthogonal. Umbenennung von $\tilde{L}$ in L und B in P bedeutet dann, daß P in (69) als normiert vorausgesetzt werden kann.

P muß orthogonal sein. Denn wegen der vorausgesetzten Orthogonalität von L muß $L'L = \text{diag}(\lambda_1, \ldots, \lambda_k)$ eine Diagonalmatrix sein, und $\lambda_k = |\mathbf{L}_k|^2$ das Quadrat der Länge des Spaltenvektors $\mathbf{L}_k$ von L. Also hat man

$$U'U = PL'LP' = P\Lambda P'. \tag{74}$$

Andererseits bedeutet die rechte Seite von (74), daß P die Matrix mit den Eigenvektoren von $U'U$ sein muß, und die sind wegen der Symmetrie von $U'U$ orthogonal, und Λ enthält die zugehörigen Eigenwerte (d.h. die Längen der $\mathbf{L}_k$ sind die Eigenwerte von $U'U$!). Dann enthält $Q = L\Lambda^{-1/2}$ die auf die Länge 1 normierten Spaltenvektoren von L, d.h. L kann in der Form $L = Q\Lambda^{1/2}$ geschrieben werden. Eingesetzt in (73) erhält man

$$U = Q\Lambda^{1/2}P',$$

also (69). Dann aber erhält man $UU' = Q\Lambda^{1/2}P'P\Lambda^{1/2}Q' = Q\Lambda Q'$, denn $P'P = I$, da P orthonormal ist. Dann aber muß Q die Matrix der normierten Eigenvektoren von UU' sein. $\square$

Anmerkung: In (31) wurde die Zerlegung $U = QB'$ angegeben; die Spaltenvektoren von Q und B sind linear unabhängig. Sie können, müssen aber nicht orthogonal sein. Die Singularwertzerlegung (69) entspricht dieser Gleichung, wenn man $B = P\Lambda^{1/2}$ setzt. (69) ist ein Spezialfall von (31), weil in (69) die Vektoren in Q und B orthogonal sein sollen. (69) hat den Vorteil, dass die Vektoren in Q und P als Eigenvektoren konkret berechnet werden können; der Übergang zu obliquen latenten Vektoren ist dann immer noch möglich.

2.7 Ellipsen und Eigenvektoren symmetrischer Matizen

In diesem Abschnitt soll eine Veranschaulichung des Eigenvektor- und Eigenwertbegriffs durch die Bezug auf die Darstellung von Ellipsen geliefert werden.

Es sei $\mathbf{u} = (u_1, u_2)'$ und M sei eine symmetrische Matrix, d.h.

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

Die in der folgenden Gleichung eingeführte quadratische Form definiert eine Ellipse:

$$\mathbf{u}'M\mathbf{u} = (u_1, u_2) \begin{pmatrix} a & c \\ c & b \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = k, \tag{75}$$

denn

$$\mathbf{u}'M\mathbf{u} = au_1^2 + bu_2^2 + 2cu_1u_2 = k, \tag{76}$$

und dies ist die Gleichung einer Ellipse, deren Hauptachsen nicht parallel zu den Koordinatenachsen liegen, falls $c \neq 0$. u_1 und u_2 treten hier als Variable auf: für einen bestimmten Wert für u_1 liefert (76) den zugehörigen Wert u_2, falls u_1 im erlaubten Wertebereich liegt. Da die Punkte mit den Koordinaten (u_1, u_2) auf einer Ellipse liegen sollen, kann man keinen Wert für u_2 finden, wenn Punkte mit der Koordinate u_1 gar nicht auf der Ellipse liegen. $\mathbf{u} = (u_1, u_2)'$ steht damit für eine Menge von Vektoren, nämlich die Menge der Vektoren, deren Endpunkte auf einer Ellipse liegen.

Setzt man in (76) den Wert von c gleich Null, so beschreibt die Gleichung eine achsenparallele Ellipse; man kann also

$$\mathbf{u}'M\mathbf{u} = au_1^2 + bu_2^2 = k$$

schreiben, wobei jetzt aber M eine Diagonalmatrix ist, denn es gilt ja nun $c = 0$. Die Menge der Punkte (u_1, u_2) unterscheidet sich von der Menge der Punkte (u_1, u_2), die durch die Gleichung (76)für den Fall $c \neq 0$ definiert wird. Deswegen ist es sinnvoll, die Punkte für die achsenparallele Ellipse in (x_1, x_2) umzubenennen, d.h. die Vektoren Vektoren $\mathbf{x} = (x_1, x_2)'$ zu betrachten. Eine achsenparallele Ellipse hat dann die Gleichung

$$\mathbf{x}'N\mathbf{x} = (x_1, x_2)\begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = a_1x_1^2 + a_2x_2^2 = k, \tag{77}$$

wobei N eine Diagonalmatrix $N = \text{diag}(a_1, a_2)$ ist.

Es sei $\mathcal{E}_\alpha$ eine nicht achsenparallele Ellipse; $\alpha \neq 0$ ist der Winkel, den die erste Hauptachse der Ellipse mit der x-Achse des Koordinatensystems bildet. $\mathcal{E}_\alpha$ ist dann durch (76) mit $c \neq 0$ definiert. $\mathcal{E}_\alpha$ soll so rotiert werden, daß ihre Achsen parallel zu den Koordinatenachsen liegen, d.h. die Ellipse soll in eine Ellipse $\mathcal{E}_0$ überführt werden, die durch eine Gleichung der Form (77) definiert ist, für die $\alpha = 0$ gilt. Dieser Rotation entspricht eine Rotation der Koordinaten derart, dass die neuen Koordinaten die Richtung der Hauptachsen der Ellipse haben.

Der Rotation der Ellipse entspricht eine Transformation der Vektoren $\mathbf{u}$ für die Ellipse (76), bzw. der Vektoren $\mathbf{x}$ für die Ellipse (77), je nachdem, von welcher Ellipse man ausgeht. Denn rotiert man die Ellipse (76), so geht ein Vektor $\mathbf{u}$ in einen dazu korrespondierenden Vektor $\mathbf{x}$ über. Einen Vektor transformiert man aber, indem man ihn mit einer geeigneten Matrix T mulitpliziert, d.h. man sucht eine Transformationsmatrix $T = T_0$ derart, dass z.B.

$$T_0\mathbf{u} = \mathbf{x} \tag{78}$$

gilt für alle Vektoren $\mathbf{u}$, die der Gleichung (76) genügen, und alle Vektoren $\mathbf{x}$, die der Gleichung (77) genügen. Umgekehrt kann man eine Transformationsmatrix $T = T_\alpha$ suchen, die jeden Vektor $\mathbf{x}$ in den zugehörigen Vektor $\mathbf{u}$ überführt, so dass

$$T_\alpha\mathbf{x} = \mathbf{u}. \tag{79}$$

Zur Vereinfachung werde jetzt $T_\alpha = T$ gesetzt, so dass statt $T_\alpha\mathbf{x} = \mathbf{u}$ einfach $T\mathbf{x} = \mathbf{u}$ geschrieben werden kann. Man kann dann $T\mathbf{x}$ für $\mathbf{u}$ in (75) einsetzen und erhält

$$(T\mathbf{x})'M(T\mathbf{x}) = \mathbf{x}'T'MT\mathbf{x} = k.$$

Nun genügen aber die Vektoren $\mathbf{x}$ der Gleichung (77), so dass

$$T'MT = N$$

gelten muß. Diese Gleichung ist aber von der Form (65), mit $N = \Lambda$, $T = P$. Also sind die Spaltenvektoren $\mathbf{T}_1, \mathbf{T}_2$ in T gerade die Eigenvektoren der symmetrischen Matrix M, und die Diagonalelemente a_1 und a_2 sind gerade die Eigenwerte λ_1 und λ_2 von M. Die Länge x_1^{max} der ersten Halbachse erhält man, indem man $x_2 = 0$ setzt, so dass

$$x_1^{max} = \sqrt{k/a_1}. \tag{80}$$

Für die Länge der zweiten Halbachse erhält man analog

$$x_2^{max} = \sqrt{k/a_2}. \tag{81}$$

Da T die Matrix der normierten Eigenvektoren von M ist, folgt, dass die Spaltenvektoren von T orthonormal sind, so dass $T'T = I$, I die Einheitsmatrix. Nun gilt einerseits $T\mathbf{x} = T_\alpha \mathbf{x} = \mathbf{u}$, und nach (78) gilt andererseits $T_0\mathbf{u} = \mathbf{x}$. Hier kann man für $\mathbf{u}$ den Ausdruck $T\mathbf{x}$ einsetzen, so dass $T_0T\mathbf{x} = \mathbf{x}$ folgt. Das heißt aber, dass $T_0T = I$ die Einheitsmatrix ist, so dass $T_0 = T^{-1}$ folgt. Die Transformationsmatrizen T_0 und $T_\alpha = T$ stehen also in einer inversen Beziehung zueinander.

Eigenvektoren und Hauptachsentransformation: Die durch $\mathbf{x}'N\mathbf{x} = k$ beschriebene, achsenparallele Ellipse wird durch eine Transformation T von $\mathbf{x}$ in $\mathbf{u}$ in die nicht achsenparallele Ellipse $\mathbf{u}'M\mathbf{u} = k$ überführt, und die Spaltenvektoren von T sind die Eigenvektoren von M, d.h. sie genügen den Gleichungen

$$M\mathbf{t}_i = \lambda_i \mathbf{t}_i.$$

Die Eigenvektoren haben die Richtung der Hauptachsen der durch M charakterisierten Ellipse. Dazu betrachte man den Punkt $(x_{01}, 0)$ der achsenparallelen Ellipse. Dieser Punkt liegt auf der x-Achse und ist ein Scheitelpunkt. Durch T wird der dazu korrespondierende Vektor $\mathbf{x}_0$ in den Punkt $T\mathbf{x}_0 = \mathbf{u}_0 = (u_{01}, u_{02})'$ überführt. Der Punkt (u_{01}, u_{02}) ist wieder ein Scheitelpunkt: er liegt auf der rotierten Hauptachse. Allgemein ist $\mathbf{u} = T\mathbf{x}$ eine Linearkombination der Spalten von T,

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = x_1 \begin{pmatrix} t_{11} \\ t_{21} \end{pmatrix} + x_2 \begin{pmatrix} t_{21} \\ t_{22} \end{pmatrix} = x_1\mathbf{t}_1 + x_2\mathbf{t}_2.$$

Für $\mathbf{u}_0$ und $\mathbf{x}_0$ erhält man insbesondere $\mathbf{u}_0 = x_{01}\mathbf{t}_1 + x_{02}\mathbf{t}_2 = x_{01}\mathbf{t}_1$, da ja $x_{02} = 0$, d.h.

$$\mathbf{u}_0 = \begin{pmatrix} u_{01} \\ u_{02} \end{pmatrix} = x_{01} \begin{pmatrix} t_{11} \\ t_{21} \end{pmatrix} = \begin{pmatrix} x_{01}t_{11} \\ x_{01}t_{21} \end{pmatrix}.$$

Der Vektor $\mathbf{u}_0$ bildet mit der x-Achse den Winkel α, so dass $\tan\alpha = u_{02}/u_{01}$. Dies bedeutet

$$\tan\alpha = \frac{u_{02}}{u_{01}} = \frac{x_{01}t_{21}}{x_{01}t_{11}} = \frac{t_{21}}{t_{11}}.$$

Damit gibt aber $\tan\alpha$ bzw α auch die Orientierung des Eigenvektors $\mathbf{t}_1$ an. Die Eigenvektoren haben also die Orientierung der Hauptachsen der rotierten Ellipse.

Generell haben die Spaltenvektoren $\mathbf{t}_j$ von $T = [\mathbf{t}_1, \ldots, \mathbf{t}_n]$ die Orientierungen der Hauptachsen der rotierten Ellipse. Andererseits ist T eine Transformationsmatrix. Man sagt deshalb, T definiere eine *Hauptachsentransformation*. Die Hauptachsentransformation ist äquivalent der Transformation der Matrix M in die Diagonalmatrix $N = T'MT$, so daß man von $T'MT$ auch als einer Repräsentation einer Hauptachsentransformation sprechen kann. Für die Längen der Hauptachsen erhält man also

$$x_1^{max} = \sqrt{k/\lambda_1}, \quad x_2^{max} = \sqrt{k/\lambda_2}. \tag{82}$$

Diese Werte sind natürlich mit denen in (80) und (81) identisch.

Der Vergleich mit (75) und (76) zeigt, dass eine quadratische Form für $n = 2$ gerade eine Ellipse definiert. Man sieht dann leicht, dass für $n = 3$ ein Ellipsoid entsteht, und für $n > 3$ eben ein n-dimensionales Ellipsoid. Die quadratische Form für beliebiges n und nicht notwendig diagonale, symmetrische Matrix M hat die Form

$$\mathbf{x}'M\mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}x_ix_j = k \in \mathbb{R}, \tag{83}$$

m_{ij} das Element in der i-ten und j-ten Spalte von M.

2.8 Lineare Gleichungssysteme

Es sei A eine $(m \times n)$-Matrix mit den Elementen a_{ik}, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, $\mathbf{x}$ sei ein n-dimensionaler, $\mathbf{y}$ sei ein m-dimensionaler Vektor. Die Gleichung

$$A\mathbf{x} = \mathbf{y} \tag{84}$$

repräsentiert dann ein lineares Gleichungssystem; ausgeschrieben hat es die Form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= y_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= y_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= y_m. \end{aligned} \tag{85}$$

Tatsächlich ist das Gleichungssystem der Vektorgleichung

$$x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \tag{86}$$

äquivalent, d.h. $\mathbf{y}$ ist eine Linearkombination der Spaltenvektoren von A. Für den Fall $\mathbf{y} = \underline{0}$ heißt das Gleichungssystem *homogen*, andernfalls heißt es *inhomogen*. Es sei $\mathbf{a}_1$ der erste Spaltenvektor von A, $\mathbf{a}_2$ der zweite, etc. Dann kann (86) in der Form

$$x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{y} \tag{87}$$

geschrieben werden.

Lösungen des homogenen Systems: Es werde zunächst der Spezialfall $\mathbf{y} = \underline{0}$, also das homogene System

$$x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \underline{0} \tag{88}$$

betrachtet. Dann ist $\mathbf{x} = \underline{0}$ offenbar stets eine Lösung; dies ist die *triviale Lösung*. Ist der Rang von A gleich n, so sind die $\mathbf{a}_j$, $j = 1, 2, \ldots, n$ linear unabhängig, und $\mathbf{x} = \underline{0}$ ist die einzige Lösung.

Nun sei $r < \min(m, n)$, und $\mathbf{x}_1$ und $\mathbf{x}_2$ seien zwei Lösungen, so dass $A\mathbf{x}_1 = \underline{0}$ und $A\mathbf{x}_2 = \underline{0}$. Dann sind auch $\lambda_1\mathbf{x}_1$ und $\lambda_2\mathbf{x}_2$ Lösungen, wenn λ_1 und λ_2 Skalare sind, denn $A\lambda_1\mathbf{x}_1 = \lambda_1\underline{0} = \underline{0}$ und $A\lambda_2\mathbf{x}_2 = \lambda_2\underline{0} = \underline{0}$; Addition der beiden Gleichungen liefert

$$A(\lambda_1\mathbf{x}_1 + \lambda_2\mathbf{x}_2) = \underline{0},$$

also ist $\lambda_1\mathbf{x}_1 + \lambda_2\mathbf{x}_2$ ebenfalls eine Lösung.

Ohne Einschränkung der Allgemeinheit kann angenommen werden, dass die ersten r Spaltenvektoren linear unabhängig sind. Dann können die restlichen $n-r$ Spaltenvektoren als Linearkombinationen dieser r Vektoren dargestellt werden, d.h. es gibt Koeffizienten $-\lambda_{1k}, \ldots, -\lambda_{r,k}$, $k = 1, 2, \ldots, n-r$ derart, dass

$$\mathbf{a}_{r+k} = -\lambda_{1k}\mathbf{a}_1 - \cdots - \lambda_{r,k}\mathbf{a}_r, \quad k = 1, \ldots, r,$$

bzw.

$$\lambda_{1k}\mathbf{a}_1 + \cdots + \lambda_{r,k}\mathbf{a}_r + \mathbf{a}_{r+k} = \underline{0}, \quad k = 1, \ldots, n-r. \tag{89}$$

ist. Dies bedeutet aber, dass jeder Vektor

$$\mathbf{x}_k = (\lambda_{1k}, \ldots, \lambda_{r,k}, 0, \ldots, 0, 1, 0, \ldots, 0)'$$

ein Lösungsvektor für das homogene System ist, wobei die 1 an der k-ten Stelle steht, $k = 1, \ldots, n-r$. Diese Vektoren sind aber linear unabhängig[2]. Da jede Linearkombination von Lösungsvektoren wieder eine Lösung des homogenen Systems darstellt, gilt

Satz 2.9 *Die Menge der Lösungsvektoren des homogene durch die Elemente* $(n-r)$-*dimensionalen Teilraums des* V_n *gegeben.*

Lösungen des inhomogenen Systems: Man muß zunächst fragen, unter welchen Bedingungen überhaupt eine Lösung, d.h. ein Vektor $\mathbf{x}$, existiert, der der Gleichung $A\mathbf{x} = \mathbf{y}$, also (86) bzw. (87), genügt. In diesen Gleichungen wird der Vektor $\mathbf{y}$ als Linearkombination der Spaltenvektoren $\mathbf{a}_1, \ldots, \mathbf{a}_n$ dargestellt. Die notwendige Bedingung für die Existenz von $\mathbf{x}$ ist daher, dass $\mathbf{y}$ tatsächlich auch eine Linearkombination der $\mathbf{a}_1, \ldots, \mathbf{a}_n$ ist. Ist $\mathbf{y}$ keine Linearkombination der Spaltenvektoren $\mathbf{a}_j$, $j = 1, \ldots, n$, so existiert auch der Vektor $\mathbf{x}$ nicht.

Der Rang von A sei $r \leq \min(m, n)$. Ist $\mathbf{y}$ eine Linearkombination der $\mathbf{a}_j$, so liegen die $\mathbf{a}_j$ und $\mathbf{y}$ im gleichen Teilraum des m-dimensionalen Vektorraums (die $\mathbf{a}_j$ und $\mathbf{y}$ sind ja m-dimensionale Vektoren). Der Rang von A ist die Anzahl linear unabhängiger Vektoren von A. Es sei B die Matrix, die aus A entsteht, wenn als $n+1$-te Spalte der Vektor $\mathbf{y}$ hinzugefügt wird. Der vorangegangenen Argumentation entspricht der

Satz 2.10 *Das Gleichungssystem (84) hat mindestens eine Lösung, wenn die Matrix* A *und die Matrix* $B = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n, \boldsymbol{y}]$ *den gleichen Rang* r *haben, andernfalls existiert keine Lösung.*

[2] Man überzeuge sich davon, indem man die Vektorgleichung $\mu_1\mathbf{x}_1 + \cdots + \mu_k\mathbf{x}_k = \underline{0}$ betrachtet. Diese Gleichung kann nur für $\mu_1 = \cdots = \mu_k = 0$ gelten!

Beispiel 2.3 Das Gleichungssystem

$$\begin{aligned} x_1 3 - x_2 6 + x_3 3 &= 0 \\ x_1 2 - x_2 4 + x_3 2 &= 2 \end{aligned}$$

hat keine Lösung. Um dies zu sehen, multipliziere man die erste Gleichung mit 2, die zweite mit 3; es entsteht das Gleichungssystem

$$\begin{aligned} x_1 6 - x_2 12 + x_3 6 &= 0 \\ x_1 6 - x_2 12 + x_3 6 &= 6 \end{aligned}$$

Subtrahiert man die zweite von der ersten Gleichung, so erhält man die Gleichung $0 = -6$. Dieser Widerspruch impliziert, dass die Gleichungen keine Lösung $\mathbf{x}$ haben. Der Vektor $\mathbf{y} = (0, 2)$ ist demnach keine Linearkombination der Vektoren $\mathbf{a}_1 = (3, 2)'$, $\mathbf{a}_2 = (-6, -4)'$ und $\mathbf{a}_3 = (3, 2)'$. Nun ist $\mathbf{a}_1 = \mathbf{a}_3$, und $\mathbf{y}$ ist dann keine Linearkombination der $\mathbf{a}_j$, wenn $\mathbf{y}$ nicht im gleichen Teilraum wie die beiden Vektoren $\mathbf{a}_1$ und $\mathbf{a}_2$ liegt. Da der (Spalten-)Vektorraum 2-dimensional ist, kann dies nur bedeuten, dass $\mathbf{a}_1$ und $\mathbf{a}_2$ in einem 1-dimensionalen Teilraum liegen, also die gleiche Orientierung haben. Tatsächlich bilden $\mathbf{a}_1$ und $\mathbf{a}_2$ den gleichen Winkel mit der x-Achse: $\tan\alpha_1 = 2/3$ und $\tan\alpha_2 = 4/6 = 2/3$. Linearkombinationen dieser beiden Vektoren können nicht aus diesem Teilraum hinausführen und deshalb nie den Vektor $\mathbf{y}$ erzeugen. □

Generell ist es so, dass das Hinzufügen von $\mathbf{y}$ zu den Spalten von A den Rang der entstehenden Matrix nicht verringern kann, denn die Anzahl linear unabhängiger Vektoren wird dadurch ja nicht reduziert. Der Rang kann also nur gleich bleiben oder größer werden. Er bleibt gleich, wenn $\mathbf{y}$ eine Linearkombination der $\mathbf{a}_j$ ist, und erhöht sich um 1, wenn $\mathbf{y}$ keine Linearkombination der $\mathbf{a}_j$ ist; in diesem Fall gibt es keine Lösung.

Satz 2.11 *Das inhomogene Gleichungssystem (84), d.h. $A\boldsymbol{x} = \boldsymbol{y}$, sei lösbar, und es sei $\boldsymbol{x}_0$ ein bestimmter, fester Lösungsvektor, so daß $A\boldsymbol{x}_0 = \boldsymbol{y}$. Das zugehörige homogene System $A\boldsymbol{x} = \underline{0}$ habe die Lösungen V_{n-r}, d.h. alle Vektoren aus dem $(n-r)$-dimensionalen Vektorraum. Die Menge der Lösungen für das inhomogene System ist dann durch*

$$\boldsymbol{x}_0 + V_{n-r} = \{\boldsymbol{x}_0 + \boldsymbol{x} | \boldsymbol{x} \in V_{n-r}\} = \mathbf{L} \tag{90}$$

gegeben.

Beweis: Es wird zuerst gezeigt, dass jeder Vektor aus $\mathbf{L}$ tatsächlich eine Lösung von $A\mathbf{x} = \mathbf{y}$ ist. Sei also $\mathbf{x} = \mathbf{x}_0 + \mathbf{x}_1$, $\mathbf{x}_1 \in \mathbf{L}$. Dabei ist $\mathbf{x}_1$ definitionsgemäß eine Lösung des homogenen Systems. Dementsprechend ist dann

$$A\mathbf{x} = A(\mathbf{x}_0 + \mathbf{x}_1) = A\mathbf{x}_0 = \mathbf{y}$$

da ja voraussetzungsgemäß $A\mathbf{x}_1 = \underline{0}$. Ebenso wurde vorausgesetzt, dass $\mathbf{x}_0$ eine Lösung ist. Damit ist nachgewiesen, dass $\mathbf{x}_0 + \mathbf{x}_1$ tatsächlich ebenfalls eine Lösung ist.

Nun sei $\mathbf{x}$ ein Lösungsvektor. Dann muß gezeigt werden, dass $\mathbf{x} \in \mathbf{L}$ ist. Gilt dies, so muß auch $\mathbf{x} - \mathbf{x}_1$ ein Lösungsvektor sein, also

$$A(\mathbf{x} - \mathbf{x}_1) = A\mathbf{x} - A\mathbf{x}_1 = A\mathbf{x},$$

da ja $A\mathbf{x}_1 = \underline{0}$. Aber $\mathbf{x}$ war als Lösungsvektor vorausgesetzt worden, so dass $A\mathbf{x} = \mathbf{y}$ gelten muß. Damit ist der Satz bewiesen. □

Der Satz impliziert, dass die Lösung nicht eindeutig ist, wenn es eine von $\underline{0}$ verschiedene Lösung des zugehörigen homogenen Systems gibt, und dies ist der Fall, wenn $r < \min(m, n)$ ist. Hat aber A den vollen Rang, so ist $\underline{0}$ die einzige Lösung des homogenen Systems und das inhomogene System hat nur eine einzige Lösung.

Spezialfall: Ein wichtiger Spezialfall ergibt sich, wenn A quadratisch ist, d.h. wenn $m = n$ und der Rang von A gleich n ist. Dann existiert die Inverse A^{-1} von A und aus $A\mathbf{x} = \mathbf{y}$ folgt dann

$$\mathbf{x} = A^{-1}\mathbf{y}; \tag{91}$$

man muß also die Inverse von A bestimmen, um zur Lösung zu kommen. Ist der Rang von $A = A_{(n,n)}$ gleich n, so *muß* $\mathbf{y}$ eine Linearkombination der Spaltenvektoren von A sein. Denn bei diesen handelt es sich um n linear unabhängige n-dimensionale Vektoren, und jeder weitere n-dimensionale Vektor, also eben auch $\mathbf{y}$, kann als Linearkombination dieser Spaltenvektoren dargestellt werden.

2.9 Zufällige Vektoren und Matrizen

Bestimmt man die Werte X_j als Messungen von n Variablen V_j, $j = 1, \ldots, n$, so wird man die X_j als zufällige Veränderliche betrachten. Man kann dann $\mathbf{x} = (X_1, \ldots, X_n)'$ als *zufälligen Vektor* oder auch als *Zufallsvektor* definieren. Dabei geht man i.a. davon aus, dass die Komponenten X_j jeweils einen Erwartungswert $\mu_j = E(X_j)$ und eine Varianz σ_j^2 haben. Faßt man die μ_j zu einem Vektor zusammen, so erhält man

$$E(\mathbf{x}) = \vec{\mu} = (\mu_1, \ldots, \mu_n)'. \tag{92}$$

$\mathbf{x} - E(\mathbf{x}) = \mathbf{x} - \vec{\mu}$ ist der Vektor der Abweichungen $X_j - \mu_j$, $j = 1, \ldots, n$. Die Varianzen und Kovarianzen der Komponenten von $\mathbf{x}$ können dann in der durch (35) definierten Matrix zusammengefasst werden:

$$\begin{aligned} \mathbf{\Sigma} &= \mathrm{Kov}(\mathbf{x}) = E(\mathbf{x} - \vec{\mu})(\mathbf{x} - \vec{\mu})' \qquad (93) \\ &= \begin{pmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) & \cdots & (X_1 - \mu_1)(X_n - \mu_n) \\ X_2 - \mu_2)(X_1 - \mu_2) & (X_2 - \mu_2)^2 & \cdots & (X_2 - \mu_2)(X_n - \mu_n) \\ & & \vdots & \\ (X_n - \mu_n)(X_1 - \mu_1) & (X_n - \mu_n)(X_2 - \mu_2) & \cdots & (X_n - \mu_n)^2 \end{pmatrix} \end{aligned}$$

Schreibt man σ_{ij} für $(X_1 - \mu_i)(X_2 - \mu_j)$, so erhält man die kürzere Form

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ & & \vdots & \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{pmatrix}. \tag{94}$$

Sicherlich ist $\sigma_{ij} = \sigma_{ji}$, so dass Σ eine symmetrische Matrix ist. Gelegentlich wird auch σ_{ii} statt σ_i^2 geschrieben.

Der Begriff des zufälligen Vektors kann zu dem der zufälligen Matrix verallgemeinert werden. $X = (x_{ij})$ heißt *zufällige Matrix* oder *Zufallsmatrix*, wenn die Elemente x_{ij} zufällige Veränderliche sind. Ist $\mu_{ij} = E(x_{ij})$, so ist $\mu = (\mu_{ij})$ die Matrix der Erwartungswerte.

Bekanntlich gilt

$$E(aX_j) = aE(X_j) = a\mu_j, \tag{95}$$

und

$$\mathrm{Var}(aX_j) = E(aX_j - a\mu_j)^2 = a^2\mathrm{Var}(X_j) = a^2\sigma_j^2. \tag{96}$$

Für eine Linearkombination zweier zufällige Vektoren erhält man dann

$$E(aX_j + bX_k) = a\mu_j + b\mu_k, \tag{97}$$

und für die entsprechende Varianz

$$\mathrm{Var}(aX_j + bX_k) = a^2\mathrm{Var}(X_j) + b^2\mathrm{Var}(X_k) + ab\mathrm{Kov}(X_jX_k). \tag{98}$$

Setzt man $\mathbf{c} = (a,b)'$, $\mathbf{x} = (X_j, X_k)'$, so erhält man (97) in der Form

$$E(aX_j + bX_k) = (a,b)\begin{pmatrix} X_j \\ X_k \end{pmatrix} = \mathbf{c}'\mathbf{x}. \tag{99}$$

Für den Fall $n = 2$ liefert (94) die Varianz-Kovarianz-Matrix $\mathbf{\Sigma}$ für $\mathbf{x}$, und (98) kann in der Form

$$\mathrm{Var}(aX_j + bX_k) = \mathrm{Var}(\mathbf{cx}) = \mathbf{c}'\mathbf{\Sigma}\mathbf{c} \tag{100}$$

dargestellt werden. Diese Form ist insbesondere dann nützlich, wenn $n > 2$.

2.10 Die multivariate Normalverteilung

Es sei $\mathbf{x} = (x_1, \ldots, x_n)'$ ein Vektor, dessen Komponenten Messungen irgendwelcher Variablen sind. Da Messungen i.a. fehlerbehaftet sind, kann man die x_i als zufällige Veränderliche betrachten; dementsprechend heißt $\mathbf{x}$ auch *zufälliger Vektor*.

Es sei

$$E(\mathbf{x}) = \vec{\mu} = (E(x_1), \ldots, E(x_n))' = (\mu_1, \ldots, \mu'_) \tag{101}$$

der Vektor der Erwartungswerte von $\mathbf{x}$. Weiter sei $\mathbf{y} = \mathbf{x} - E(\mathbf{x})$. Dann enthält die Matrix $C = \mathbf{yy}'$ gerade alle Varianzen (in den Diagonalzellen) der x_i und die Kovarianzen zwischen den x_i und x_j. Es sei

$$g(\mathbf{x}) = \exp(-\frac{1}{2}(\mathbf{x} - \vec{\mu})'C^{-1}(\mathbf{x} - \vec{\mu})). \tag{102}$$

g ist eine Funktion in den n Variablen $x_1, \ldots, x_n$), und C^{-1} ist die zu C inverse Matrix. Für alle Werte von x_i, $i = 1, 2, \ldots, n$ ist $g \geq 0$, und das Integral über alle möglichen $\mathbf{x}$ ist gleich $(2\pi)^{n/2}\sqrt{|C|}$, wobei $|C|$ die Determinante von C ist. Von Determinanten wird in der Veranstaltung nicht explizit Gebrauch gemacht, so dass sie hier nicht weiter behandelt werden, es genügt, zu wissen, dass sie gleich einer reellen Zahl sind, d.h. $|C| \in \mathbb{R}$. Dann kann man die n-dimensionale Normalverteilung definieren:

Definition 2.3 *Es sei $\boldsymbol{x}$ ein zufälliger Vektor mit dem Vektor der Erwartungswerte $\vec{\mu}$ und der Varianz-Kovarianz-Matrix C. Dann ist*

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{|C|}} \exp(-\frac{1}{2}(\boldsymbol{x} - \vec{\mu})'C^{-1}(\boldsymbol{x} - \vec{\mu})) \tag{103}$$

die Dichtefunktion der n-dimensionalen Normalverteilung.

Der Faktor $1/(2\pi)^{n/2}\sqrt{|C|}$ bewirkt, dass das Integral von f über alle $\mathbf{x}$ gleich 1 ist; eine notwendige Bedingung dafür, dass f eine Wahrscheinlichkeitsdichte ist.

Der Fall $n = 2$: Der 2-dimensionale Fall kann in etwas expliziterer Form angegeben werden:

$$f(x_1, x_2) = k \exp\left(-\frac{1}{2(1-r^2)}\phi(x_1, x_2)\right) \tag{104}$$

mit

$$\phi(x_1, x_2) = \left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} - \frac{2r(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right) \tag{105}$$

und

$$k = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}; \tag{106}$$

k ist hier der Normierungsfaktor, und r ist die Korrelation zwischen x_1 und x_2. Offenbar ist $f(x_1, x_2) = c_0$ eine Konstante, wenn

$$\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} - \frac{2r(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right) = k_0, \tag{107}$$

k_0 eine Konstante, d.h. der geometrische Ort von Punkten (d.h. Messungen) (x_1, x_2) mit gleicher Dichte ist für $r \neq 0$ jeweils eine Ellipse; für $r = 0$ und $\sigma_1 = \sigma_2$ ein Kreis.

Es sei $r = 0$, $\sigma_1 \neq \sigma_2$; der geometrische Ort von Punkten gleicher Dichte ist dann eine Ellipse. Der Vergleich mit (80) zeigt, dass nun $a_1 = \lambda_1 = 1/\sigma_1$ und $\lambda_2 = 1/\sigma_2$, und nach (80) ist

$$x_i^{max} = \sqrt{k/\lambda_i} = \sigma_i\sqrt{k}, \quad i = 1, 2 \tag{108}$$

Andererseits ist nach (103)

$$(\mathbf{x} - \vec{\mu})' C^{-1} (\mathbf{x} - \vec{\mu}) = k,$$

d.h. die Matrix, die die Elipse definiert, ist C^{-1}, und für den Fall $r = 0$ hat man

$$C = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}.$$

Diese Matrix hat die Eigenwerte

$$\lambda_1 = \sigma_1^2, \quad \lambda_2 = \sigma_2^2. \tag{109}$$

Dann ist

$$C^{-1} = \begin{pmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{pmatrix}$$

Für die Eigenwerte dieser Matrix findet man

$$\mu_1 = 1/\sigma_1^2, \quad \mu_2 = 1/\sigma_2^2.$$

(80) liefert nun wieder das Resultat (108), d.h.

$$x_i^{max} = \sqrt{k/\mu_1} = \sqrt{k\lambda_i} = \sigma_i\sqrt{k}, \quad i = 1, 2$$

wobei die λ_i die Eigenwerte der Matrix C sind; die Länge Halbachsen und damit der Achsen ist damit proportional zu den Eigenwerten der Varianz-Kovarianz-Matrix für den

Fall $r = 0$, bei die Eigen werte gerade gleich den Varianzen sind. Da eine Rotation der Ellipse die Länge der Achsen nicht verändert gilt auch für den Fall $r \neq 0$, dass die Länge Achsen proportional zu den Varianzen ist. Die Eigenwerte hängen in diesem Fall aber noch von r ab:

$$\lambda_i \quad = \quad \frac{\sigma_1^2 + \sigma_2^2 \pm \sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4r^2\sigma_1^2\sigma_2^2}}{2(r^2\sigma_1^2\sigma_2^2 - \sigma_1^2\sigma_2^2)}, \quad i = 1, 2 \tag{110}$$

Anmerkung: Die Inverse der Korrelationsmatrix ist

$$R^{-1} = \begin{pmatrix} \frac{s_{22}}{s_{11}s_{22} - s_{12}^2)}, & -\frac{s_{12}}{s_{11}s_{22} - s_{12}^2} \\ -\frac{s_{12}}{s_{11}s_{22} - s_{12}^2}, & \frac{s_{11}}{s_{11}s_{22} - s_{12}^2} \end{pmatrix}. \tag{111}$$

2.11 PCAs von nichtstandardisierten Werten

Da die Zerlegung $X = Q\Lambda^{1/2}P'$ für jede reelle Matrix X gilt, kann man auch nichtstandardisierte Messwerte einer PCA (Principal Component Analysis = Hauptachsentransformation) unterziehen. Das Problem ist, dass - abgesehen von Fragen der Vergleichbarkeit der Maßeinheiten für die verschiedenen Variablen - die erste Achse dann i.a. nur die Mittelwerte der Spalten von X widerspiegelt, also wenig über die Kovarianzstruktur der Variablen aussagt. Dies ist ein weiterer Grund, von den standardisierten Variablen auszugehen.

Tatasächlich gilt aber, dass zwar die direkte Verrechnung von Messwerten möglich ist, aber zu irrelevanten Resultaten führen kann. Es werde das folgende Beispiel betrachtet:

Es sei

$$X = \begin{pmatrix} 20 & 24 & 36 \\ 40 & 16 & 42 \\ 30 & 26 & 28 \\ 40 & 20 & 16 \end{pmatrix}$$

Dann ist

$$X'X = \begin{pmatrix} 4776 & 2844 & 4096 \\ 2844 & 1908 & 2584 \\ 4098 & 2584 & 4100 \end{pmatrix}$$

Diese Matrix enthält also nur die Skalarprodukte der Spaltenvektoren von X, nicht die Kovarianzen oder gar Korrelationen. Die Eigenwerte Λ^* und die Eigenvektoren P^* von $X'X$ sind

$$\Lambda^* = \begin{pmatrix} 10306.05 & 0 & 0 \\ 0 & 331.09 & 0 \\ 0 & 0 & 146.86 \end{pmatrix}, \quad P^* = \begin{pmatrix} .670 & .632 & .392 \\ .416 & .116 & -.902 \\ .615 & -.766 & .185 \end{pmatrix}$$

Es sei nun $\bar{X}$ eine Matrix, deren Spalten den jeweiligen Mittelwert für die entsprechende Spalte von X enthalte:

$$\bar{X} = \begin{pmatrix} 34 & 21.5 & 30.5 \\ 34 & 21.5 & 30.5 \\ 34 & 21.5 & 30.5 \\ 34 & 21.5 & 30.5 \end{pmatrix},$$

und es sei

$$Y = X - \bar{X} = \begin{pmatrix} -8 & 2.5 & 5.5 \\ 6 & -5.5 & 11.5 \\ -4 & 4.5 & -2.5 \\ 6 & -1.5 & -14.5 \end{pmatrix}, \quad Y'Y = \begin{pmatrix} 152 & -80 & -52 \\ -80 & 59 & -39 \\ -52 & 39 & 379 \end{pmatrix}.$$

Die Eigenwerte und Eigenvektoren von $Y'Y$ sind

$$\Lambda = \begin{pmatrix} 391.81 & 0 & \\ 0 & 194.81 & 0 \\ 0 & 0 & 3.38 \end{pmatrix}, \quad P = \begin{pmatrix} .189 & .839 & .510 \\ .069 & -.530 & .845 \\ -.979 & -125 & .158 \end{pmatrix}.$$

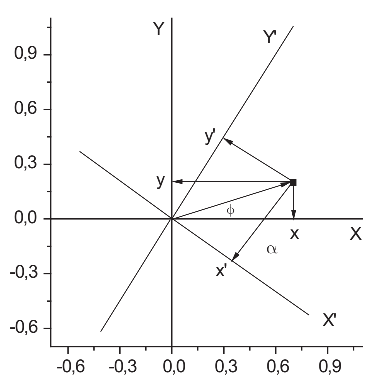
Man kann nun sowohl X wie auch Y auf Hauptachsen transformieren; dazu bildet man die Matrizen $L^* = XP^*$ und $L = YP$. Man erhält insbesondere für L^* und L

$$L^* = \begin{pmatrix} -6.73 & -7.35 & -1.10 \\ -10.51 & 9.38 & .235 \\ 2.00 & -6.05 & 1.37 \\ 15.23 & 4.02 & -.50 \end{pmatrix}, \quad L = \begin{pmatrix} -1.17 & -0.64 & .21 \\ 1.51 & -.99 & -.07 \\ -1.13 & .27 & .23 \\ .78 & 1.37 & .10 \end{pmatrix}.$$

3 Rotationen von Vektoren und Koordinaten

Gegeben sei ein Koordinatensystem (X, Y), in dem der Punkt P die Koordinaten (x, y) habe, vergl. Abb. 3. Das Koordinatensystem (X', Y') unterscheide sich vom System

Abbildung 3: Koordinatentransformation



(X, Y) durch eine Rotation um den Winkel α. Der Vektor $\mathbf{u}$, dessen Anfangspunkt um Ursprung der beiden Koordinatensysteme habe und dessen Endpunkt P sei, bilde mit der X-Achse die Winkel ϕ. Der Winkel, den der Vektor $\mathbf{u}$ mit der X'-Achse bildet, hat den den Wert $\alpha + \phi$. Es gelten sicherlich einerseits die Beziehungen

$$\sin\phi = y/r, \quad r = \|\mathbf{u}\| \tag{112}$$

$$\cos\phi = x/r, \tag{113}$$

und andererseits

$$\begin{aligned} \cos(\alpha + \phi) &= x'/r && (114) \\ \sin(\alpha + \phi) &= y'/r && (115) \end{aligned}$$

Weiter gilt

$$\begin{aligned} \cos(\alpha + \phi) &= \cos\alpha\cos\phi - \sin\alpha\sin\phi && (116) \\ \sin(\alpha + \phi) &= \sin\alpha\cos\phi + \cos\alpha\sin\phi, && (117) \end{aligned}$$

Setzt man die Ausdrücke auf den rechten Seiten dieser Gleichungen für die linken Seiten der Gleichungen (115) und (114) ein, so erhält man

$$\begin{aligned} \cos\alpha\cos\phi + \sin\alpha\sin\phi &= x'/r, && (118) \\ \sin\alpha\cos\phi - \cos\alpha\sin\phi &= y'/r. && (119) \end{aligned}$$

Substituiert man hier für $\sin\phi$ und $\cos\phi$ die Ausdrücke aus den Gleichungen (112) und (113), so erhält man eine Beziehung zwischen den Koordinaten x und x' sowie y und y':

$$\begin{aligned} x\cos\alpha - y\sin\alpha &= x' && (120) \\ x\sin\alpha + y\cos\alpha &= y', && (121) \end{aligned}$$

wobei berücksichtigt wurde, dass r sich herauskürzt; der Winkel ϕ tritt hier nicht mehr auf. Es sei T die Matrix

$$T = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}, \tag{122}$$

so hat man statt der Gleichungen (120) und (121) die Gleichung

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x' \\ y' \end{pmatrix}. \tag{123}$$

T ist eine Transformationsmatrix, die das Koordinatensystem (X, Y) rotiert, wobei der Vektor $\mathbf{u}$ die neuen Koordinaten (x', y') bekommt, oder, äquivalent dazu, die den Vektor $\mathbf{u} = (x, y)'$ im Koordinatensystem (X, Y) in den Vektor $\mathbf{v} = (x', y')'$ rotiert. T ist eine orthonormale Matrix:

$$\begin{aligned} T'T &= \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} && (124) \\ &= \begin{pmatrix} \cos^2\alpha + \sin^2\alpha & -\cos\alpha\sin\alpha + \cos\alpha\sin\alpha \\ -\sin\alpha\cos\alpha + \sin\alpha\cos\alpha & \sin^2\alpha + \cos^2\alpha \end{pmatrix} && (125) \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I. && (126) \end{aligned}$$

Auf analoge Weise folgt $TT' = I$.

In Abschnitt 2.7 wurde gezeigt, dass die Eigenvektoren T einer symmetrischen Matrix M, die eine Ellipse $\mathbf{x}'M\mathbf{x} = k \in \mathbb{R}$ definiert, gerade die Rotation der Ellipse auf eine achsenparallele Form implizieren. Rotationen werden aber durch Matrizen der Form (122) definiert. Dies bedeutet, dass die Komponenten der Eigenvektoren $\mathbf{t}_1$ und $\mathbf{t}_2$ von T gerade durch

$$\mathbf{t}_1 = \begin{pmatrix} \cos\alpha \\ \sin\alpha \end{pmatrix}, \quad \mathbf{t}_2 = \begin{pmatrix} -\sin\alpha \\ \cos\alpha \end{pmatrix} \tag{127}$$

gegeben sind; α ist dann der Winkel, den die erste Hauptachse der Ellipse mit der x-Achse bildet.

Um eine Verallgemeinerung für n-dimensionale Vektoren vornehmen zu können, ist es nützlich, die Betrachtung in einer etwas anderen Weise durchzuführen. Nach (14) gilt ja für irgend zwei Vektoren $\mathbf{u}$ und $\mathbf{v}$ die Beziehung

$$\cos\alpha = \frac{\mathbf{u}'\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}. \tag{128}$$

Das Koordinatensystem in Abb. 3 kann als durch die beiden Einheitsvektoren $\mathbf{e}_1 = (1,0)'$ und $\mathbf{e}_2 = (0,1)'$ gegeben gedacht werden. Der Vektor $\mathbf{z} = (x,y)'$ ist dann durch die Linearkombination

$$\mathbf{z} = x\begin{pmatrix} 1 \\ 0 \end{pmatrix} + y\begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{129}$$

gegeben.Dann gilt $\mathbf{z}'\mathbf{e}_1 = x\cdot 1 + y\cdot 0 = x$ und $\mathbf{z}'\mathbf{e}_2 = x\cdot 0 + y\cdot 1 = y$, und nach (128)

$$\cos\alpha = \mathbf{z}'\mathbf{e}_1 r = \frac{x}{r}, \quad \cos(\pi/2-\alpha) = \frac{\mathbf{z}'\mathbf{e}_2}{r} = \frac{y}{r} \tag{130}$$

wobei $r = \|\mathbf{z}\|$ und $\|\mathbf{e}_1\| = \|\mathbf{e}_2\| = 1$. $\pi/2-\alpha$ ist der Winkel, den der Vektor $\mathbf{z}$ mit der durch $\mathbf{e}_2$ definierten Koordinatenachse bildet $\pi/2$ entspricht 90^0, denn ein rechter Winkel ist ein Viertel von $360^0 \doteqdot 2\pi$. Die Größen x/r und y/r heißen auch *Richtungskosinus*. Wendet man (116) auf $\cos(\pi/2-\alpha)$ an, so findet man $\cos(\pi/2-\alpha) = -\sin\alpha$. Dementsprechend kann man sagen, dass die Komponenten der Matrix T die Richtungskosinus des Vektors $\mathbf{z}$ enhalten. Eines symmetrische (2×2)-Matrix definiert eine Ellipse, und die Eigenvektoren T von M definieren die Transformationsmatrix, die die Ellipse achsenparallel rotiert. Die Komponenten der Eigenvektoren enthalten also gerade die Richtungskosinus der Hauptachsen der Ellipse.

Ist M eine $n\times n$-Matrix, so definiert sie ein n-dimensionales Ellipsoid. Ein Vektor, dessen Anfangspunkt im Ursprung des Koordinatensystems liegt und dessen Endpunkt auf der Oberfläche des Ellipsoids liegt, bildet zur k-ten Achse einen Winkel α_k; die Komponenten der Eigenvektoren von M sind wieder die Richtungskosinus der α_k.