Lineare Algebra in Hinblick auf die Hauptachsentransformation und Faktorenanalysen

Paul Ruppen

31. März 2020

Dieses Dokument wurde mit LaTeX gesetzt.

Inhaltsverzeichnis

Kapitel 1

Lineare Funktionen, Matrizen und lineare Gleichungssysteme

Beim Berechnen von Polynomen durch Punkte mussten Gleichungssysteme ersten Grades aufgelöst werden, in denen mehrere Variablen vorkamen. In diesem Kapitel geht es darum, hierfür effiziente Lösungsmöglichkeiten bereitzustellen. Dazu sind ein paar Vorarbeiten nötig, die eine eigenständige Bedeutung haben - oft auch für die unmittelbare ökonomische Anwendung.

1.1 Lineare Abbildungen

Den Ausdruck

$$5x_1 + 3x_2 + 4x_3 + 17x_4,$$

kann man verwenden, um einem 4-Tupel von Zahlen genau eine Zahl zuzuordnen. Setzen wir nämlich z.B.

$$(x_1, x_2, x_3, x_4) = (1.2, 3, 4.5, 0.5)$$

so gilt

$$\begin{aligned} 5x_1 + 3x_2 + 4x_3 + 17x_4 &= 5 \cdot 1.2 + 3 \cdot 3 + 4 \cdot 4.5 + 17 \cdot 0.5 \\ &= 41.5 \end{aligned}$$

Der obige Ausdruck ordnet damit dem $4-Tupel$ $(1.2, 3, 4.5, 0.5)$ die Zahl 41.5 zu. Anderen $4-$Tupeln wird ebenfalls derart eine Zahl zugeordnet. Damit wird durch den Ausdruck eine Abbildung von $\mathbb{R}^4$ nach $\mathbb{R}$ definiert:

$$\begin{aligned} f &: \mathbb{R}^4 \to \mathbb{R} \\ f((x_1, x_2, x_3, x_4)) &:= 5x_1 + 3x_2 + 4x_3 + 17x_4 \end{aligned} \tag{1.1.1}$$

Wir nennen $5, 3, 4$ und 17 die Koeffizienten des Ausdrucks $5x_1 + 3x_2 + 4x_3 + 17x_4$. Wir beziehen uns auf den Koeffizienten des Faktors x_i mit a_i.

Beispiel 1.1.1. *Funktionen der Art (1.1.1) finden ökonomische Interpretationen. So braucht man z.B. fürs Brotbacken Mehl. Ebenso fürs Backen von Wegglis und Nussgipfeln. Wir nehmen an, man brauche für 1 Kg Brot 800 g Mehl, für ein Weggli 80 g und für einen Nussgipfel 70 g. Wir nehmen nun an, wir möchten 70 Brote, 120 Wegglis und 17 Nussgipfel backen und möchten dazu die benötigte Menge Mehl berechnen. Wir multiplizieren die gewünschte Anzahl Brote, Anzahl Wegglis und Anzahl Nussgipfel mit der jeweils pro Stück nötigen Menge und addieren diese Produkte, d.h. wir berechnen*

$$f((70, 120, 17)) = 800 \cdot 70 + 80 \cdot 120 + 70 \cdot 17 = 66790.$$

Die Funktion $f((x_1, x_2, x_3)) = 800x_1+80x_2+70x_3$ *ordnet die benötigte Menge Mehl jeder gewünschten Outputkombination* (x_1, x_2, x_3) *zu.* ◊

Bemerkung 1.1.2. *Auf die inneren Klammern bei* $f((x_1, x_2, x_3, x_4))$ *wird künftig verzichtet, so dass nur mehr* $f(x_1, x_2, x_3, x_4)$ *geschrieben wird.* ◊

Bemerkung 1.1.3. *Allgemein kann man eine Funktion der Art (1.1.1) ausdrücken durch*

$$f : \mathbb{R}^n \to \mathbb{R}$$
$$f(\mathbf{x}) = f(x_1, ..., x_n) := \sum_{i=1}^{n} a_i x_i$$

wobei $a_i \in \mathbb{R}$ *die Koeffizienten sind und* $\mathbf{x} := (x_1, ..., x_n)$. ◊

Betrachten wir m Ausdrücke der Art

$$\sum_{i=1}^{n} a_i x_i$$

zusammen - wir nennen solche Ansammlungen von Ausdrücken künftig „System von Ausdrücken“, so können wir Abbildungen von $\mathbb{R}^n$ nach $\mathbb{R}^m$ definieren. So definieren die Ausdrücke

$$1.2x_1 + 4x_2 + 8x_3 + 12x_4$$
$$3x_1 + 12x_2 - x_3 + 7x_4$$
$$15x_1 + 200x_2 + 4.3x_3 + 2x_4$$

eine Funktion von $\mathbb{R}^4$ nach $\mathbb{R}^3$, wenn wir die Tripel betrachten, die durch diese drei Ausdrücke in ihrer Reihenfolge beim Einsetzen von Zahlen für x_1, x_2, x_3 und x_4 bestimmt sind. Diese Funktion würde z.B. dem $4-Tupel$ (1.2, 3, 4.5, 0.5) das $Tripel$ (55.44, 38.6, 638.35) zuordnen. Schreiben wir $n-$Tupel statt auf einer Zeile als Spalte hin - dies machen wir aus praktischen Gründen - so gilt damit:

$$f : \mathbb{R}^4 \to \mathbb{R}^3 \tag{1.1.2}$$
$$f\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} := \begin{pmatrix} 1.2x_1 + 4x_2 + 8x_3 + 12x_4 \\ 3x_1 + 12x_2 - x_3 + 7x_4 \\ 15x_1 + 200x_2 + 4.3x_3 + 2x_4 \end{pmatrix}$$

Bemerkung 1.1.4. *Es ist üblich,* $n-Tupel$ *in diesem Zusammenhang als „Vektoren“ zu bezeichnen. Wir werden uns hier an diese Sprachregelung halten. Es ist zu beachten, dass ein Vektor hier genau dasselbe wie ein* $n-Tupel$ *ist, unabhängig davon, wie es auf dem Papier dargestellt wird. Der Vektorbegriff der linearen Algebra ist allerdings weiter als der des* $n-Tupels$. *Ein Vektor ist Element eines sogenannten Vektorraumes - eine mathematische Struktur, das wir aus Zeitgründen nicht einführen können. Je nach betrachtetem Vektorraum sind Vektoren Funktionen,* $n-Tupel$ *oder andere mathematische Objekte.* ◊

Bemerkung 1.1.5. *Allgemein kann man eine Funktion der Art* (1.1.2) *ausdrücken durch*

$$f : \mathbb{R}^n \to \mathbb{R}^m$$
$$f\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} := \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + ... + a_{1n-1}x_{n-1} + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn-1}x_{n-1} + a_{mn}x_n \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{n} a_{1j}x_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj}x_j \end{pmatrix}$$

Wir brauchen nun für die Koeffizienten zwei Indizes, da diese in jeder Zeile verschieden sein können. Der erste Index drückt jeweils die Zeile i aus, in der der Koeffizient a_{ij} auftaucht. Der zweite Index drückt jeweils aus, mit welchem x_j der Koeffizient a_{ij} zu multiplizieren ist. ◊

Beispiel 1.1.6. *Auch Abbildungen $f : \mathbb{R}^n \to \mathbb{R}^m$ können ökonomisch interpretiert werden. Fürs Brotbacken braucht man ja nicht nur Mehl, sondern auch Salz, Wasser und eventuell weitere Zutaten wie Butter, Nüsse, etc. So brauche man - nehmen wir an - für ein normales Brot 800 g Mehl, 300 g Wasser, 25 g Hefe und 30 g Salz. Für ein Weggli brauche man 80 g Mehl, 30 g Wasser, 3 g Hefe, 3 g Salz und 20 g Butter. Für einen Nussgipfel brauche man 70 g Mehl, 3 g Salz, 3 g Hefe, 30 g Wasser, 10 g Butter und 30 g Nüsse (frei erfundene Zahlen; zum Nachbacken verwende man ein Kochbuch!). Wir möchten wissen, wieviel von diesen Zutaten wir brauchen, wenn wir 50 Brote, 27 Wegglis und 30 Nussgipfel backen wollen. Wir können das folgende System von Ausdrücken aufstellen:*

$$\begin{aligned} Mehl &: 800x_1 + 80x_2 + 70x_3 \\ Wasser &: 300x_1 + 30x_2 + 30x_3 \\ Hefe &: 25x_1 + 3x_2 + 3x_3 \\ Salz &: 30x_1 + 3x_2 + 3x_3 \\ Butter &: 20x_2 + 10x_3 \\ Nüsse &: 30x_3 \end{aligned}$$

Durch dieses System ist die Funktion

$$f : \mathbb{R}^3 \to \mathbb{R}^6$$
$$f\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 800x_1 + 80x_2 + 70x_3 \\ 300x_1 + 30x_2 + 30x_3 \\ 25x_1 + 3x_2 + 3x_3 \\ 30x_1 + 3x_2 + 3x_3 \\ 20x_2 + 10x_3 \\ 30x_3 \end{pmatrix}$$

festgelegt. Jeder Mengenkombination an gewünschtem Output wird die nötige Inputmengenkombination der Rohstoffe zugordnet. Wollen wir z.B. die Einkaufsmengen für den angestrebten Output-Vektor $(50, 27, 30)$ berechnen, so berechnen wir $f(x_1, x_2, x_3)$ für $(x_1, x_2, x_3) = (50, 27, 30)$. Man erhält:

$$f\begin{pmatrix} 50 \\ 27 \\ 30 \end{pmatrix} = \begin{pmatrix} 800 \cdot 50 + 80 \cdot 27 + 70 \cdot 30 \\ 300 \cdot 50 + 30 \cdot 27 + 30 \cdot 30 \\ 25 \cdot 50 + 3 \cdot 27 + 3 \cdot 30 \\ 30 \cdot 50 + 3 \cdot 27 + 3 \cdot 30 \\ 20 \cdot 27 + 10 \cdot 30 \\ 30 \cdot 30 \end{pmatrix} = \begin{pmatrix} 44260 \\ 16710 \\ 1421 \\ 1671 \\ 840 \\ 900 \end{pmatrix}$$

◊

Als nächstes definieren wir ein paar Operationen auf $n - Tupel$.

Definition 1.1.7. *(Multiplikation eines Vektors mit einer reellen Zahl) Sei* $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ *und* $a \in \mathbb{R}$, *dann ist*

$$a\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} ax_1 \\ \vdots \\ ax_n \end{pmatrix}$$

Beispiel 1.1.8. *a)*

$$5\begin{pmatrix} 3 \\ 2 \\ 8 \end{pmatrix} = \begin{pmatrix} 5 \cdot 3 \\ 5 \cdot 2 \\ 5 \cdot 8 \end{pmatrix} = \begin{pmatrix} 15 \\ 10 \\ 40 \end{pmatrix}$$

b) ökonomisches Beispiel: Eine Unternehmung produziere von 4 verschiedene Produkte und pro Monat folgende Mengen

$$\begin{pmatrix} 40 \\ 30 \\ 50 \\ 100 \end{pmatrix}$$

Dann produziert Sie in 12 Monaten:

$$12\begin{pmatrix} 40 \\ 30 \\ 50 \\ 100 \end{pmatrix} = \begin{pmatrix} 12 \cdot 40 \\ 12 \cdot 30 \\ 12 \cdot 50 \\ 12 \cdot 100 \end{pmatrix} = \begin{pmatrix} 480 \\ 360 \\ 600 \\ 1200 \end{pmatrix}$$

◊

◊

Definition 1.1.9. *: Sei* $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n$*, dann ist*

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

◊

Beispiel 1.1.10. *a)*

$$\begin{pmatrix} 2 \\ 1 \\ 5 \\ 6 \end{pmatrix} + \begin{pmatrix} 4 \\ 16 \\ 3 \\ 1.5 \end{pmatrix} = \begin{pmatrix} 6 \\ 17 \\ 8 \\ 7.5 \end{pmatrix}$$

b) ökonomisches Beispiel: Eine Unternehmung habe zwei Standorte, an denen dieselben Produkte erzeugt werden. Die produzierte Menge pro Standort sei durch die folgenden zwei Vektoren gegeben:

$$\begin{pmatrix} 30 \\ 40 \\ 500 \\ 80 \end{pmatrix}; \begin{pmatrix} 5 \\ 80 \\ 100 \\ 250 \end{pmatrix}$$

Dann ist die Gesamtproduktion des Betriebes gegeben durch:

$$\begin{pmatrix} 30 \\ 40 \\ 500 \\ 80 \end{pmatrix} + \begin{pmatrix} 5 \\ 80 \\ 100 \\ 250 \end{pmatrix} = \begin{pmatrix} 35 \\ 120 \\ 600 \\ 330 \end{pmatrix}$$

◊

Definition 1.1.11. *Eine Abbildung $f : \mathbb{R}^n \to \mathbb{R}^m$ für $n, m \in \mathbb{N}^*$ heisst „linear“ genau dann, wenn*

$$f(ax) = af(x)$$
$$f(x+y) = f(x) + f(y)$$

*für $a \in \mathbb{R}$ (s. auch Seite **??**)*

$\Diamond$

Praktisch bedeutet die Linearität folgendes. Kennen wir $f(x)$, und möchten wir wissen, wieviel $f(ax)$ ist, genügt es, $f(x)$ mit a zu multiplizieren. Kennen wir $f(x)$ und $f(y)$, und möchten wir $f(x+y)$ kennen, so genügt es, $f(x)$ und $f(y)$ zu addieren.

Satz 1.1.12. *Funktionen*

$$f : \mathbb{R}^n \to \mathbb{R}^m$$
$$f\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} := \begin{pmatrix} \sum_{j=1}^{n} a_{1j}x_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj}x_j \end{pmatrix}$$

mit $a_{ij} \in \mathbb{R}$ sind linear.

Beweis. a) Sei $a_{ij}, a \in \mathbb{R}$ und $\mathbf{x} := \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, dann gilt mit den obigen Operationen:

$$f(a\mathbf{x}) = f\left(a\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right) = f\left(\begin{pmatrix} ax_1 \\ \vdots \\ ax_n \end{pmatrix}\right) = \begin{pmatrix} \sum_{j=1}^{n} a_{1j}ax_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj}ax_j \end{pmatrix}$$
$$= \begin{pmatrix} a\sum_{j=1}^{n} a_{1j}x_j \\ \vdots \\ a\sum_{j=1}^{n} a_{mj}x_j \end{pmatrix} = a\begin{pmatrix} \sum_{j=1}^{n} a_{1j}x_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj}x_j \end{pmatrix} = af\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = af(\mathbf{x})$$

b) Sei $x := \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ und $y := \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ dann gilt mit den obigen Operationen:

$$f(\mathbf{x}+\mathbf{y}) = f\left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}\right) = f\left(\begin{pmatrix} x_1+y_1 \\ \vdots \\ x_n+y_n \end{pmatrix}\right) = \begin{pmatrix} \sum_{j=1}^{n} a_{1j}(x_j+y_j) \\ \vdots \\ \sum_{j=1}^{n} a_{mj}(x_j+y_j) \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{j=1}^{n} a_{1j}x_j + \sum_{j=1}^{n} a_{1j}y_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj}x_j + \sum_{j=1}^{n} a_{mj}y_j \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{n} a_{1j}x_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj}x_j \end{pmatrix} + \begin{pmatrix} \sum_{j=1}^{n} a_{1j}y_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj}y_j \end{pmatrix}$$

$$= f\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + f\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = f(\mathbf{x}) + f(\mathbf{y})$$

□

Als nächstes führen wir eine übersichtlichere Schreibweise für lineare Abbildungen ein. Wir ordnen die Koeffizienten a_{ij} als Tabelle an, die wir mit einer Klammer umschliessen:

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}$$

So würden die Koeffizienten des Systems

$$\begin{aligned} &1.2x_1 + 4x_2 + 8x_3 + 12x_4 \\ &3x_1 + 12x_2 - x_3 + 7x_4 \\ &15x_1 + 200x_2 + 4.3x_3 + 2x_4 \end{aligned}$$

dargestellt durch

$$\begin{pmatrix} 1.2 & 4 & 8 & 12 \\ 3 & 12 & -1 & 7 \\ 15 & 200 & 4.3 & 2 \end{pmatrix} \tag{1.1.3}$$

Solche Tabellen werden wir künftig „Matrizen“ (Einzahl „Matrix“) nennen. Eine Matrix, die m Zeilen und n Spalten aufweist, wird $(m,n) - Matrix$ genannt (Sprich: m mal n Matrix).

Bemerkung 1.1.13. *Matrizen werden in der Mathematik nicht nur als tabellarische Darstellung von Zahlen auf Papier betrachtet sondern als mathematische Objekte (wie Zahlen, Mengen, Funktionen, etc.). Eine $(m,n) - Matrix$ wird definiert als Funktion $N_m^* \times N_n^* \to \mathbb{R}$. So wäre der Graph der obigen Matrix* (1.1.3)

$$\begin{aligned} &\{(1,1,1.2),(1,2,4),(1,3,8),(1,4,12), \\ &(2,1,3),(2,2,12),(2,3,-1),(2,4,7), \\ &(3,1,15),(3,2,200),(3,3,4.3),(3,4,2)\}. \end{aligned}$$

Diese Darstellung ist nicht üblich, ist aber manchmal hilfreich fürs Verständnis. ◊

Wir verwenden als Variablen für Matrizen die fetten Grossbuchstaben $\mathbf{A}, \mathbf{B}, \mathbf{C}, \ldots$. Da der Zusammenhang klar macht, was gemeint ist, kann man handschriftlich normale Grossbuchstaben verwenden. Die Zahlen in einer Matrix $\mathbf{A}$ nennen wir „Komponenten der Matrix“. Komponenten der Matrix $\mathbf{A}$ werden allgemein bezeichnet durch a_{ij}, wobei i die Zeile der Matrix angibt und j die Spalte der Matrix, in der sich die Komponente befindet. So ist in der obigen Matrix (1.1.3) $a_{23} = -1; a_{31} = 15; a_{13} = 8$.
Wir definieren die Multiplikation einer Matrix mit einem Vektoren wie folgt:

$$\begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} =: \begin{pmatrix} \sum_{j=1}^{n} a_{1j}x_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj}x_j \end{pmatrix}$$

An einem Beispiel:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 5 \\ 4 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \cdot x_1 + 2x_2 \\ 3 \cdot x_1 + 5x_2 \\ 4 \cdot x_1 + 6x_2 \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{2} a_{1j}x_j \\ \sum_{j=1}^{2} a_{2j}x_j \\ \sum_{j=1}^{2} a_{3j}x_j \end{pmatrix}$$

Man sieht, dass das Resultat der Matrixmultiplikation mit der rechten Seite von

$$f \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} := \begin{pmatrix} \sum_{j=1}^{n} a_{1j}x_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj}x_j \end{pmatrix}$$

identisch ist (s. Bemerkung 1.1.5). Wir können lineare Abbildungen also künftig mit

$$f : \mathbb{R}^n \to \mathbb{R}^m$$
$$f \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} := A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

definieren. Kürzen wir $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ durch $\mathbf{x}$ ab, so entsteht die handliche Darstellung:

$$f : \mathbb{R}^n \to \mathbb{R}^m$$
$$f(\mathbf{x}) := \mathbf{A}\mathbf{x}$$

mit einer (m, n) −Matrix $\mathbf{A}$.

Beispiel 1.1.14. *Für die obige Zusammenstellung*

$$\begin{aligned} Mehl &: 800x_1 + 80x_2 + 70x_3 \\ Wasser &: 300x_1 + 30x_2 + 30x_3 \\ Hefe &: 25x_1 + 3x_2 + 3x_3 \\ Salz &: 30x_1 + 3x_2 + 3x_3 \\ Butter &: 20x_2 + 10x_3 \\ Nüsse &: 30x_3 \end{aligned}$$

erhält man die Matrix

$$\begin{pmatrix} 800 & 80 & 70 \\ 300 & 30 & 30 \\ 25 & 3 & 3 \\ 30 & 3 & 3 \\ 0 & 20 & 10 \\ 0 & 0 & 30 \end{pmatrix}$$

Diese nennen wir „Produktionsmatrix". Will man die nötige Menge an Rohstoffen berechnen, so multipliziert man die Produktionsmatrix mit dem Vektor der gewünschten Output-Mengen. So erhält man z.B. für $\begin{pmatrix} 50 \\ 27 \\ 30 \end{pmatrix}$*:*

$$\begin{pmatrix} 800 & 80 & 70 \\ 300 & 30 & 30 \\ 25 & 3 & 3 \\ 30 & 3 & 3 \\ 0 & 20 & 10 \\ 0 & 0 & 30 \end{pmatrix} \begin{pmatrix} 50 \\ 27 \\ 30 \end{pmatrix} = \begin{pmatrix} 44\,260 \\ 16\,710 \\ 1421 \\ 1671 \\ 840 \\ 900 \end{pmatrix}$$

Der letzte Vektor ist der Input-Vektor.

◊

Beispiel 1.1.15. *Für das Brotbackbeispiel 1.1.1 mit der Funktion*

$$f : \mathbb{R}^3 \to \mathbb{R}$$
$$f((x_1, x_2, x_3)) = 800x_1 + 80x_2 + 70x_3$$

erhält man die einzeilige Matrix

$$\begin{pmatrix} 800 & 80 & 70 \end{pmatrix},$$

womit

$$f(70, 120, 17) = \begin{pmatrix} 800 & 80 & 70 \end{pmatrix} \begin{pmatrix} 70 \\ 120 \\ 17 \end{pmatrix} = 66\,790$$

◊

Bemerkung 1.1.16.

- *Im vorangegangenen Beispiel 1.1.15 zeigt sich, dass einer linearen Abbildung $f : \mathbb{R}^n \to \mathbb{R}$ eine einzeilige Matrix mit n Spalten entspricht. Diese unterscheidet sich äusserlich von einem $n-Tupel$ dadurch, dass die Zahlen nicht durch Kommas abgetrennt werden. Eine einzeilige Matrix ist gemäss Bemerkung 1.1.13 eine Funktion $A : \{1\} \times \{1, ..., n\} \to \mathbb{R}$. Damit ist eine einzeilige Matrix mit n Spalten von einem $n-$Tupel verschieden. Fürs obige Beispiel würden wir als Graphen der Matrix erhalten:*

$$\{(1, 1, 800), (1, 2, 80), (1, 3, 70)\} \neq (800, 80, 70).$$

- *Es gibt auch einspaltige Matrizen. Diese sind zu verstehen als Minifunktionen*

$$\{1, ..., n\} \times \{1\} \to \mathbb{R}$$

So gilt z.B. für die folgende einspaltige Matrix

$$\begin{pmatrix} 660 \\ 626 \\ 40 \\ 22 \end{pmatrix}$$

der Graph

$$\{(1,1,660),(2,1,626),(3,1,40),(4,1,22)\}$$

Dieser ist vom Graphen

$$\{(1,1,660),(1,2,626),(1,3,40),(1,4,22)\}$$

der entsprechenden einzeiligen Matrix

$$\begin{pmatrix} 660 & 626 & 40 & 22 \end{pmatrix}$$

verschieden.
Die einspaltige Matrix ist aber auch vom 4-Tupel $(660,626,40,22)$ *verschieden. Als Spalte geschrieben ist der Vektor von der einspaltigen Matrix mit der bisherigen Notation nicht zu unterscheiden. Allerdings werden wir keine Schreibweise einführen, um Vektoren von einspaltigen Matrizen zu unterscheiden, da Vektoren und einspaltige Matrizen bezüglich der mathematischen Operationen der Addition und der Multiplikation mit einer Konstanten genau gleich behandelt werden. Die Multiplikation einer Matrix mit einem Vektor und einer Matrix mit einer einspaltigen Matrix unterscheidet sich äusserlich ebenfalls nicht. Entsprechend lohnt sich eine Unterscheidung der beiden Objektarten nicht. Da Vektoren und einspaltige Matrizen aber verschieden sind, darf man sie nicht gleichsetzen, selbst wenn man sie gleich behandelt!*

- *Die Multiplikation einer Matrix mit einem Vektor mit n Komponenten ist nur definiert, falls die Matrix n Spalten hat.*
- *Die Multiplikation einer* $(m,n)-$*Matrix und eines Vektors mit n Komponenten ergibt einen Vektor mit m Komponenten.*

$\Diamond$

Mit Excel gibt man zwecks Multiplikation der Matrix mit einem Vektor die Zahlen der Matrix in eine Tabelle ein (z.B. eine (m,n)-Matrix in $A1 : Xm$ (X sei der n-te Buchstabe des Alphabets). Dann gibt man den Vektor als Spalte in die Tabelle ein (z.B. $A\,(m+1) : A(m+1+n)$). Dann wird eine von Zahlen freie Spalte mit m Zellen beleuchtet und in die erste „$= mmult(A1 : Xm; A\,(m+1) : A(m+1+n))$“ geschrieben. Es wird mit „ctrl+shift+enter“ bestätigt.
In R wird eine Matrix eingegeben und unter A abgespeichert durch

$$A = matrix(c(a_{11},a_{12},a_{13},...,a_{1n},a_{21},...,a_{mn}), byrow = T, ncol = n)$$

(`byrow=T` gibt an, dass die Komponenten zeilenweise im Vektor $c(a_{11},a_{12},a_{13},...,a_{1n},a_{21},...,a_{mn})$ aufgeschrieben sind, fehlt die Angabe geht R davon aus, dass die Komponenten spaltenweise eingegeben sind; `ncol` gibt die Anzahl der Spalten der Matrix an). So wird für die Matrix

$$\mathbf{A} = \begin{pmatrix} 4 & 3 & 1 \\ 5 & 8 & 11 \\ 12 & 12 & 10 \end{pmatrix}$$

geschrieben: $A = matrix(c(4,3,1,5,8,11,12,12,10), byrow = T, ncol = 3)$.
Die Multiplikation mit einem Vektor $\mathbf{x}$ erfolgt durch
$A\%*\%c(x_1,x_2,...x_n)$ oder z.B. durch $A\%*\%x$, wenn man vorgängig $x = c(x_1,...,x_n)$ definiert.

Definition 1.1.17. *Vektoren, die bis auf eine Zelle i nur Nullen aufweisen und die in der* $i-$*ten Zelle eine* 1 *aufweisen, werden Einheitsvektoren genannt. Wir kürzen sie ab durch* $\mathbf{e}_i$*. Wenn die Anzahl n der Komponenten des Vektors erwähnt werden soll, schreibt man* $\mathbf{e}_i^n$ $\Diamond$

Beispiel 1.1.18.

$$\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \textit{ ist der Vektor } \mathbf{e}_2^3$$

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \textit{ ist der Vektor } \mathbf{e}_1^4$$

$$\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \textit{ ist der Vektor } \mathbf{e}_3^4$$

$\Diamond$

Satz 1.1.19. *Multipliziert man eine $(m,n)-$Matrix mit dem Vektor e_i^n, so erhält man die $i-$te Spalte der Matrix als Resultat.*

Beispiel 1.1.20.

$$\begin{pmatrix} 1.2 & 4 & 8 & 12 \\ 3 & 12 & -1 & 7 \\ 15 & 200 & 4.3 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 12 \\ 200 \end{pmatrix}$$

$\Diamond$

Übungen

1. Sei

$$f : \mathbb{R}^5 \to \mathbb{R}$$
$$f(\mathbf{x}) = 5x_1 + 3x_2 - 2x_3 + 14x_4 + 1.5x_5$$

   Berechnen Sie $f(7, 8.9, 10, 11, 12)$ mit Hilfe der Gleichung und mit Hilfe der Matrixmultiplikation.

2. Sei

$$f : \mathbb{R}^2 \to \mathbb{R}^3$$
$$f\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5x_1 + 3.3x_2 \\ 4x_1 - 2x_2 \\ 15x_1 + \sqrt{5}x_2 \end{pmatrix}$$

   Berechnen Sie $f(7, 8.9)$ a) durch Einsetzung in die obige Gleichung.
   b) Mit Hilfe der Matrixmultiplikation (von Hand und mit Excel).

3. Eine Baufertigelementefirma produziert Zementplatten, Zementdeckel und Zementziegel. Für eine Platte braucht es 500 kg Zement, für einen Deckel 30 kg und für einen Ziegel 60 kg.
   a) Geben Sie die Produktionsmatrix an.
   b) Berechnen Sie die Menge Zement, die es für die Produktion von 60 Platten, 100 Deckeln und 40 Ziegeln braucht.

4. Eine Baufertigelementefirma produziert Zementplatten, Zementdeckel und Zementziegel. Für eine Platte braucht es 500 kg Zement, 200 kg Sand und 50 kg Stahl, für einen Deckel 30 kg Zement, 15 kg Sand und 3 kg Stahl und für einen Ziegel 60 kg Zement und 30 kg Sand.
   a) Geben Sie die Produktionsmatrix an.
   b) Berechnen Sie den Zement, den Sand und den Stahl, den es für die Produktion von 60 Platten, 100 Deckeln und 40 Ziegeln braucht.

5. Bei der Produktion von Kunststein brauche man pro 100 kg 75 kg Zement, 20 kg Marmorkies und 5 kg Spezialleim.
   a) Geben Sie die Funktion an, welche einem gewünschten Output den erforderlichen Input zuordnet. Berechnen sie den nötigen Input für einen Output von 1780.
   b) Geben Sie die Funktion mit Hilfe der Matrixmultiplikation an.

Lösungen

1. Wir erhalten

$$\begin{aligned} f(7, 8.9, 10, 11, 12) &= 5x_1 + 3x_2 - 2x_3 + 14x_4 + 1.5x_5 \\ &= 5 \cdot 7 + 3 \cdot 8.9 - 2 \cdot 10 + 14 \cdot 11 + 1.5 \cdot 12 \\ &= 213.\,7 \end{aligned}$$

   oder mit Matrixmultiplikation

$$\begin{pmatrix} 5 & 3 & -2 & 14 & 1.5 \end{pmatrix} \begin{pmatrix} 7 \\ 8.9 \\ 10 \\ 11 \\ 12 \end{pmatrix} = 213.7$$

2.

$$f\begin{pmatrix} 7 \\ 8.9 \end{pmatrix} = \begin{pmatrix} 5 \cdot 7 + 3.3 \cdot 8.9 \\ 4 \cdot 7 - 2 \cdot 8.9 \\ 15 \cdot 7 + \sqrt{5} \cdot 8.9 \end{pmatrix} = \begin{pmatrix} 64.37 \\ 10.2 \\ 124.9 \end{pmatrix}$$

   Mit Matrixmultiplikation:

$$\begin{pmatrix} 5 & 3.3 \\ 4 & -2 \\ 15 & \sqrt{5} \end{pmatrix} \begin{pmatrix} 7 \\ 8.9 \end{pmatrix} = \begin{pmatrix} 64.\,37 \\ 10.\,2 \\ 8.\,9\sqrt{5} + 105 \end{pmatrix}$$

   $8.9\sqrt{5} + 105 = 124.\,9.$

3. a) $\begin{pmatrix} 500 & 30 & 60 \end{pmatrix}$ b) $f(60, 100, 40) = \begin{pmatrix} 500 & 30 & 60 \end{pmatrix} \begin{pmatrix} 60 \\ 100 \\ 40 \end{pmatrix} = 35400$

4. a)

$$\begin{pmatrix} 500 & 30 & 60 \\ 200 & 15 & 30 \\ 50 & 3 & 0 \end{pmatrix}$$

   b) $f(60, 100, 40) = \begin{pmatrix} 500 & 30 & 60 \\ 200 & 15 & 30 \\ 50 & 3 & 0 \end{pmatrix} \begin{pmatrix} 60 \\ 100 \\ 40 \end{pmatrix} = \begin{pmatrix} 35\,400 \\ 14\,700 \\ 3300 \end{pmatrix}$

5. a)

$$f : \mathbb{R} \to \mathbb{R}^3$$
$$f(x) = \begin{pmatrix} 75x \\ 20x \\ 5x \end{pmatrix}$$
$$f(1780) = \begin{pmatrix} 75 \cdot 1780 \\ 20 \cdot 1780 \\ 5 \cdot 1780 \end{pmatrix} = \begin{pmatrix} 133\,500 \\ 35\,600 \\ 8900 \end{pmatrix}$$

b) Der Vektor, mit dem die Matrix multipliziert wird, besteht nur aus einer Zahl. Die Multiplikation kann man sich als Multiplikation der Matrix mit dem einstelligen Vektor $\mathbf{x}$ vorstellen:

$$f : \mathbb{R}^3 \to \mathbb{R}$$
$$f(x) = \begin{pmatrix} 75 \\ 20 \\ 5 \end{pmatrix} x$$
$$f\,(1780) = \begin{pmatrix} 75 \\ 20 \\ 5 \end{pmatrix} 1780 = \begin{pmatrix} 133\,500 \\ 35\,600 \\ 8900 \end{pmatrix}$$

1.2 Operationen auf lineare Abbildungen und Matrizen

Mit der Definition **??** bezüglich der (punktweisen) Addition von Funktionen ergibt sich auch eine punktweise Addition von linearen Funktionen f und g, sofern diese denselben Definitions- und Wertebereich aufweisen. Die Summe linearer Funktionen ist wiederum eine lineare Funktion. Es stellt sich die Frage, ob sich die Addition von f und g mit Hilfe der zugehörigen Matrizen darstellen lässt. In der Tat können wir zwei $(m, n) - Matrizen$ $\mathbf{A}$ und $\mathbf{B}$, welche zwei lineare Funktionen f und g definieren, gemäss Definition **??** punktweise addieren und wir erhalten eine Matrix, die der Addition $f + g$ entspricht. Wir halten dies noch formeller und ohne Beweis fest:

Satz 1.2.1. *Seien* $\mathbf{A}$ *und* $\mathbf{B}$ *zwei* $(m, n) - Matrizen$*, welche die linearen Abbildungen* f *und* g *definieren. Dann gilt:*

$$f(\mathbf{x}) + g(\mathbf{x}) = (\mathbf{A} + \mathbf{B})\,\mathbf{x}$$

wobei $\mathbf{A} + \mathbf{B}$ *definiert ist durch* $a_{ij} + b_{ij}$ *für* $1 \leq i \leq m; 1 \leq j \leq n$.

Beispiel 1.2.2. *Sei*

$$\mathbf{A} = \begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix}$$
$$\mathbf{B} = \begin{pmatrix} 11 & 1 \\ 3 & 5 \end{pmatrix}$$

dann ist

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix} + \begin{pmatrix} 11 & 1 \\ 3 & 5 \end{pmatrix} = \begin{pmatrix} 16 & 7 \\ 11 & 14 \end{pmatrix}$$

Sei nun der Vektor $(20, 22)$ *gegeben. Dann gilt*

$$f\begin{pmatrix} 20 \\ 22 \end{pmatrix} = \begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix}\begin{pmatrix} 20 \\ 22 \end{pmatrix} = \begin{pmatrix} 232 \\ 358 \end{pmatrix}$$
$$g\begin{pmatrix} 20 \\ 22 \end{pmatrix} = \begin{pmatrix} 11 & 1 \\ 3 & 5 \end{pmatrix}\begin{pmatrix} 20 \\ 22 \end{pmatrix} = \begin{pmatrix} 242 \\ 170 \end{pmatrix}$$
$$f\begin{pmatrix} 20 \\ 22 \end{pmatrix} + g\begin{pmatrix} 20 \\ 22 \end{pmatrix} = \begin{pmatrix} 474 \\ 528 \end{pmatrix}$$
$$(\mathbf{A}+\mathbf{B})\begin{pmatrix} 20 \\ 22 \end{pmatrix} = \begin{pmatrix} 16 & 7 \\ 11 & 14 \end{pmatrix}\begin{pmatrix} 20 \\ 22 \end{pmatrix} = \begin{pmatrix} 474 \\ 528 \end{pmatrix}$$

◊

Die punktweise Multiplikation gemäss Definition **??** führt ausser im Fall der Multiplikation mit einer Konstanten nicht auf lineare Funktionen und ist deshalb nicht im Rahmen linearer Funktionen zu behandeln. Im Falle der Multiplikation mit einer Konstanten gilt hingegen wie bei der Addition, dass wiederum eine lineare Funktion entsteht. Die punktweise Multiplikation der $(m, n)-$Matrix $\mathbf{A}$ mit einer reellen Konstanten b entspricht dabei der punktweisen Multiplikation der durch $\mathbf{A}$ bestimmten Funktion f mit einer Konstanten b, wobei $b\mathbf{A}$ gemäss Definition **??** festgelegt ist durch ba_{ij} für alle $1 \leq i \leq m, 1 \leq j \leq n$. Es gilt also, wie wir etwas formeller und ebenfalls ohne Beweis festhalten:

Satz 1.2.3. *Für eine Matrix* $\mathbf{A}$*, welche die lineare Funktion* f *festlegt, und eine reelle Zahl* b *gilt:*

$$bf(\mathbf{x}) = (b\mathbf{A})\,\mathbf{x}$$

wobei $b\mathbf{A}$ *durch* ba_{ij} *für alle* $1 \leq i \leq m, 1 \leq j \leq n$ *festgelegt ist.*

Beispiel 1.2.4. *Für* $b = 4$,

$$\mathbf{A} = \begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix}$$

und den Vektor $(20, 22)$ *gilt:*

$$4f\begin{pmatrix} 20 \\ 22 \end{pmatrix} = 4\left(\begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix}\begin{pmatrix} 20 \\ 22 \end{pmatrix}\right) = 4\begin{pmatrix} 232 \\ 358 \end{pmatrix} = \begin{pmatrix} 928 \\ 1432 \end{pmatrix}$$
$$\left(4\begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix}\right)\begin{pmatrix} 20 \\ 22 \end{pmatrix} = \begin{pmatrix} 20 & 24 \\ 32 & 36 \end{pmatrix}\begin{pmatrix} 20 \\ 22 \end{pmatrix} = \begin{pmatrix} 928 \\ 1432 \end{pmatrix}$$

◊

Definition 1.2.5. *Eine Matrix, die als Komponenten nur* 0 *enthält, heisst Nullmatrix. Sie wird durch* $\mathbf{O}$ *abgekürzt. Wird die Nullmatrix mit einer* $(m, n) -$ *Matrix addiert, gehen wir gewöhnlich stillschweigend davon aus, dass* $\mathbf{O}$ *eine* $(m, n) -$ *Matrix ist.* ◊

Wir halten in Bezug auf die zwei eingeführten Operationen noch ein paar offensichtliche Regeln fest:

Satz 1.2.6. *Für* $k, r \in \mathbb{R}$ *und* $(m, n) - Matrizen$ $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{O}$ *gilt:*

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$$
$$\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$$
$$k(r\mathbf{A}) = (kr)\mathbf{A}$$
$$k(\mathbf{A} + \mathbf{B}) = k\mathbf{A} + k\mathbf{B}$$
$$(k + r)\mathbf{A} = k\mathbf{A} + r\mathbf{A}$$
$$\mathbf{A} + \mathbf{O} = \mathbf{A} = \mathbf{O} + \mathbf{A}$$

Mit Excel: In einer leeren Zelle die Zellen a_{11} und b_{11} addieren. Über die entsprechend Anzahl Spalten und Zeilen ziehen.
Mit R: A+B.
Neben der eingeführten Addition linearer Funktionen und der Multiplikation mit einer reellen Zahl, spielt die Berechnung der Verkettungen von linearen Funktionen eine wichtige Rolle. In Bezug auf lineare ökonomische Funktionen wird die Verkettung etwa verwendet, wenn eine Firma zuerst Zwischenprodukte herstellt und diese dann für eine Endproduktion von Gütern verwendet. Hat man eine Produktionsmatrix, welche dem Vektor der Zwischenprodukte einen spezifischen Rohstoff-Input-Vektor zuordnet - wir nennen diese Produktionsmatrix künftig „Rohstoffmatrix“ - und eine Produktionsmatrix, welche einem Vektor von Endprodukten den Inputvektor der Zwischenprodukte zuordnet - wir nennen diese Matrix künftig kurz „Produktionsmatrix“- so möchten wir für einen Vektor von Endprodukten unmittelbar den Rohstoffinputvektor berechnen können. Wenn $\mathbf{B}$ die Rohstoffmatrix ist und $\mathbf{A}$ die Produktionsmatrix, so möchten wir

$$h(\mathbf{x}) = \mathbf{B}(\mathbf{A}\mathbf{x})$$

berechnen. Es stellt sich die Frage, ob wir eine Matrix berechnen können, so dass

$$h(\mathbf{x}) = \mathbf{C}\mathbf{x}$$

In der Tat gilt:

Satz 1.2.7. *Sei* $\mathbf{B}$ *eine* $(n,k)-$*Matrix und* $\mathbf{A}$ *eine* $(k,m)-$*Matrix so dass*

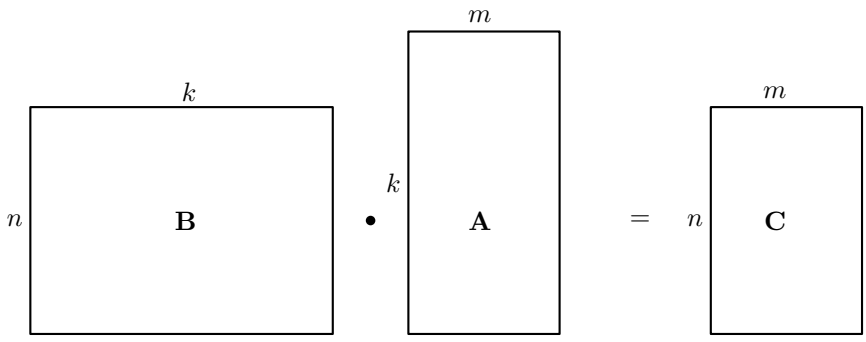
$$h(\mathbf{x}) = \mathbf{B}(\mathbf{A}\mathbf{x}).$$

Dann gibt es eine $(n,m)-$*Matrix* $\mathbf{C}$*, so dass*

$$h(\mathbf{x}) = \mathbf{C}\mathbf{x}.$$

Diese Matrix ist bestimmt durch

$$c_{ij} = \sum_{l=1}^{k} b_{il} a_{lj}$$



Beweis. Sei $\mathbf{B}$ eine (n,k)-Matrix und $\mathbf{A}$ eine (k,m)-Matrix und und $\mathbf{x} \in \mathbb{R}^m$. Damit gilt

$$\mathbf{A}x = \begin{pmatrix} \sum_{j=1}^{m} a_{1j} x_j \\ \vdots \\ \sum_{j=1}^{m} a_{1j} x_j \\ \vdots \\ \sum_{j=1}^{m} a_{kj} x_j \end{pmatrix}$$

und

$$\mathbf{B}(\mathbf{A}x)) = \begin{pmatrix} \sum_{l=1}^{k} b_{1l} \sum_{j=1}^{m} a_{lj} x_j \\ \vdots \\ \sum_{l=1}^{k} b_{il} \sum_{j=1}^{m} a_{lj} x_j \\ \vdots \\ \sum_{n=1}^{k} b_{nl} \sum_{j=1}^{m} a_{lj} x_j \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{m} \left(\sum_{l=1}^{k} b_{1l} a_{lj} \right) x_j \\ \vdots \\ \sum_{j=1}^{m} \left(\sum_{l=1}^{k} b_{il} a_{lj} \right) x_j \\ \vdots \\ \sum_{j=1}^{m} \left(\sum_{n=1}^{k} b_{nl} a_{lj} \right) x_j \end{pmatrix}$$

d.h. es gilt

$$c_{ij} = \sum_{l=1}^{k} b_{il} a_{lj}$$

□

Zusätzlich zum Beweis, ein illustratives Beispiel:

Beispiel 1.2.8. *Sei*

$$f \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} =: \mathbf{Ax}$$

und

$$g \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} =: \mathbf{Bx}.$$

Dann ist

$$\begin{aligned} g(f(\mathbf{x})) = \mathbf{B}(\mathbf{A}x) &= \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \left(\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right) \\ &= \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} x_1 a_{11} + x_2 a_{12} \\ x_1 a_{21} + x_2 a_{22} \end{pmatrix} \\ &= \begin{pmatrix} b_{11}(x_1 a_{11} + x_2 a_{12}) + b_{12}(x_1 a_{21} + x_2 a_{22}) \\ b_{21}(x_1 a_{11} + x_2 a_{12}) + b_{22}(x_1 a_{21} + x_2 a_{22}) \end{pmatrix} \\ &= \begin{pmatrix} b_{11} x_1 a_{11} + b_{11} x_2 a_{12} + b_{12} x_1 a_{21} + b_{12} x_2 a_{22} \\ b_{21} x_1 a_{11} + b_{21} x_2 a_{12} + b_{22} x_1 a_{21} + b_{22} x_2 a_{22} \end{pmatrix} \\ &= \begin{pmatrix} b_{11} x_1 a_{11} + b_{11} x_2 a_{12} + b_{12} x_1 a_{21} + b_{12} x_2 a_{22} \\ b_{21} x_1 a_{11} + b_{21} x_2 a_{12} + b_{22} x_1 a_{21} + b_{22} x_2 a_{22} \end{pmatrix} \\ &= \begin{pmatrix} (b_{11} a_{11} + b_{12} a_{21}) x_1 + (b_{11} a_{12} + b_{12} a_{22}) x_2 \\ (b_{21} a_{11} + b_{22} a_{21}) x_1 + (b_{21} a_{12} + b_{22} a_{22}) x_2 \end{pmatrix} \\ &= \begin{pmatrix} b_{11} a_{11} + b_{12} a_{21} & b_{11} a_{12} + b_{12} a_{22} \\ b_{21} a_{11} + b_{22} a_{21} & b_{21} a_{12} + b_{22} a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= \begin{pmatrix} \sum_{l=1}^{2} b_{1l} a_{l1} & \sum_{l=1}^{2} b_{1l} a_{l2} \\ \sum_{l=1}^{2} b_{2l} a_{l1} & \sum_{l=1}^{2} b_{2l} a_{l2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \end{aligned}$$

Es gibt also eine Matrix

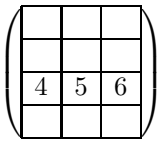
$$\mathbf{C} := \begin{pmatrix} \sum_{k=1}^{2} b_{1k} a_{k1} & \sum_{k=1}^{2} b_{1k} a_{k2} \\ \sum_{k=1}^{2} b_{2k} a_{k1} & \sum_{k=1}^{2} b_{2k} a_{k2} \end{pmatrix},$$

so dass $\mathbf{Cx} = \mathbf{A}(\mathbf{Bx})$*, wobei* $c_{ij} = \sum_{l=1}^{2} b_{il} a_{lj}$*.* ◊

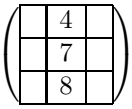
Bemerkung 1.2.9. *Man muss für* c_{ij} *berechnen:*

$$a_{i1} b_{1j} + a_{i2} b_{2j} + a_{i3} b_{3j} + ... + a_{im-1} b_{m-1j} + a_{im} b_{mj}$$

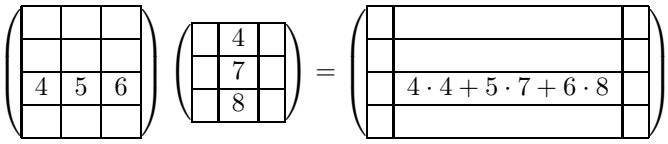
Hierzu noch ein konkretes Beispiel: Sei eine $(4,3)-$*Matrix* $\mathbf{A}$



und eine $(3,3)-$*Matrix* $\mathbf{B}$



Dann ist $c_{3,2} = \sum_{l=1}^{3} b_{3l}a_{l2} = 4 \cdot 4 + 5 \cdot 7 + 6 \cdot 8$ *oder in Matrixdarstellung*



◊

Beispiel 1.2.10. *Sei*

$$\mathbf{A} = \begin{pmatrix} 1 & 3 & 2 \\ 4 & 8 & 11 \end{pmatrix}$$
$$\mathbf{B} = \begin{pmatrix} 14 & 8 \\ 1 & 4 \\ 5 & 7 \end{pmatrix}$$

dann ist

$$\mathbf{BA} = \begin{pmatrix} 14 & 8 \\ 1 & 4 \\ 5 & 7 \end{pmatrix} \begin{pmatrix} 1 & 3 & 2 \\ 4 & 8 & 11 \end{pmatrix} = \begin{pmatrix} 46 & 106 & 116 \\ 17 & 35 & 46 \\ 33 & 71 & 87 \end{pmatrix}$$

Wir können am Beispiel und für $\mathbf{x} = (14, 20, 5)$ *nachprüfen, dass*

$$\mathbf{B}\,(\mathbf{Ax}) = \mathbf{Cx}$$

denn

$$\mathbf{A} \begin{pmatrix} 14 \\ 20 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 & 3 & 2 \\ 4 & 8 & 11 \end{pmatrix} \begin{pmatrix} 14 \\ 20 \\ 5 \end{pmatrix} = \begin{pmatrix} 84 \\ 271 \end{pmatrix}$$
$$\mathbf{B} \begin{pmatrix} 84 \\ 271 \end{pmatrix} = \begin{pmatrix} 14 & 8 \\ 1 & 4 \\ 5 & 7 \end{pmatrix} \begin{pmatrix} 84 \\ 271 \end{pmatrix} = \begin{pmatrix} 3344 \\ 1168 \\ 2317 \end{pmatrix}$$
$$\mathbf{C} \begin{pmatrix} 14 \\ 20 \\ 5 \end{pmatrix} = \begin{pmatrix} 46 & 106 & 116 \\ 17 & 35 & 46 \\ 33 & 71 & 87 \end{pmatrix} \begin{pmatrix} 14 \\ 20 \\ 5 \end{pmatrix} = \begin{pmatrix} 3344 \\ 1168 \\ 2317 \end{pmatrix}$$

◊

Definition 1.2.11. *Für die $(k,n)-$Matrix $\mathbf{C}$, so dass*

$$\mathbf{Cx} = \mathbf{B}(\mathbf{Ax})$$

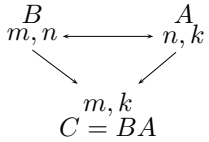
mit einer $(k,m)-$Matrix $\mathbf{B}$ und einer $(m,n)-$Matrix $\mathbf{A}$, schreiben wir

$$\mathbf{C} = \mathbf{BA}.$$

Wir nennen $\mathbf{C}$ das Produkt der Matrizen $\mathbf{A}$ und $\mathbf{B}$ (in dieser Reihenfolge). Wir sagen, dass $\mathbf{C}$ mittels Matrixmultiplikation von $\mathbf{A}$ und $\mathbf{B}$ resultiert. ◊

Bemerkung 1.2.12.

- *Die Matrixmultiplikation hat mit der punktweisen Multiplikation von Funktionen nichts zu tun!*
- *Die Matrixmultiplikation hat mit der Multiplikation, die auf reelle Zahlen definiert ist, wenig zu tun. Man könnte die Matrixmultiplikation auch „Matrixverkettung“ oder sonst wie nennen. Betrachtet man aber die Eigenschaften der Matrixmultiplikation im Satz, so ergeben sich gewisse Strukturähnlichkeiten mit der Multiplikation in den reellen Zahlen (Geltung der Assoziativ- und Distributivgesetze; Kommutativgesetz gilt aber nicht!). Es handelt sich um die Eigenschaften, die auch für die Verkettung gelten.*
- *Die Matrixmultiplikation $\mathbf{BA}$ ist nur definiert, wenn die Anzahl der Spalten von $\mathbf{B}$ mit der Anzahl der Zeilen von $\mathbf{A}$ übereinstimmt. Das Produkt einer $(m,n)-$Matrix und einer $(n,k)-$Matrix ist eine $(m,k)-$Matrix*



- *Wenn $\mathbf{BA}$ definiert ist, ist somit $\mathbf{AB}$ nur definiert, wenn $\mathbf{A}$ eine $(m,n)-$Matrix und $\mathbf{B}$ eine $(n,m)-$Matrix ist.*
- *Sind sowohl $\mathbf{AB}$ als auch $\mathbf{BA}$ definiert, so ist im allgemeinen $\mathbf{AB} \neq \mathbf{BA}$, wie folgendes Beispiel zeigt:*

$$\begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix}\begin{pmatrix} 11 & 1 \\ 3 & 5 \end{pmatrix} = \begin{pmatrix} 73 & 35 \\ 115 & 53 \end{pmatrix}$$
$$\begin{pmatrix} 11 & 1 \\ 3 & 5 \end{pmatrix}\begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix} = \begin{pmatrix} 63 & 75 \\ 55 & 63 \end{pmatrix}$$

- *Die Matrix $\mathbf{BA}$ hat so viele Zeilen wie $\mathbf{B}$ Zeilen hat und so viele Spalten wie $\mathbf{A}$ Spalten hat.*

◊

Definition 1.2.13. *Eine $(n,n)-$Matrix $\mathbf{A}$ wird „quadratisch“ genannt. Wir verwenden dafür oft das Symbol $\mathbf{A}_n$ oder $\mathbf{A}_{nn}$*
Eine quadratische Matrix mit $a_{ij} = 0$ für $i \neq j$ wird „Diagonalmatrix“ genannt.
Eine $(n,n)-$Diagonalmatrix mit $a_{ii} = 1$ für alle $1 \leq i \leq n$, wird „Einheitsmatrix“ genannt. Wir kürzen sie ab durch $\mathbf{E}$ (oder durch $\mathbf{E}_n$). ◊

Beispiel 1.2.14. *Beispiel für eine quadratische Matrix $\mathbf{A}_3$*

$$\begin{pmatrix} 5 & 3 & 4 \\ 2 & 1.5 & 6 \\ 7 & 7.5 & 8 \end{pmatrix}$$

Beispiel für eine Diagonalmatrix

$$\begin{pmatrix} 5 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 8 \end{pmatrix}$$

Beispiel für eine Einheitsmatrix:

$$\mathbf{E}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

◊

Definition 1.2.15. *Sei* $\mathbf{a}$ *ein Vektor. Dann ist* $\mathbf{diag}\,(\mathbf{a})$ *die Diagonalmatrix* $\mathbf{A}$ *mit* $a_{ii} := a_i$ *und* $a_{ij} = 0$ *für* $i \neq j$. ◊

Beispiel 1.2.16. *Mit* $\mathbf{a} = (1, 4, 5, 2, 3)$ *ist*

$$\mathbf{diag}\,(\mathbf{a}) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix}$$

◊

Machen Sie sich an Beispielen klar, wie sich die Multiplikation einer Diagonalmatrix mit einem Vektor (mit einer Matrix) auswirkt.
Multiplizieren wir eine quadratische $(n, n)-$Matrix $\mathbf{A}$ mit $\mathbf{E}_n$ schreiben wir einfach $\mathbf{AE}$. $\mathbf{E}$ ist immer passend zu $\mathbf{A}$ zu wählen, muss also gleichviele Spalten und Zeilen wie $\mathbf{A}$ haben. Ohne Beweis halten wir fest:

Satz 1.2.17.

$$\mathbf{AE} = \mathbf{A} = \mathbf{EA}$$

Beispiel 1.2.18. *Sei*

$$\mathbf{A} = \begin{pmatrix} 5 & 3 & 4 \\ 2 & 1.5 & 6 \\ 7 & 7.5 & 8 \end{pmatrix}.$$

Dann ist

$$\mathbf{AE} = \begin{pmatrix} 5 & 3 & 4 \\ 2 & 1.5 & 6 \\ 7 & 7.5 & 8 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 3 & 4 \\ 2 & 1.5 & 6 \\ 7 & 7.5 & 8 \end{pmatrix}$$

$$\mathbf{EA} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 5 & 3 & 4 \\ 2 & 1.5 & 6 \\ 7 & 7.5 & 8 \end{pmatrix} = \begin{pmatrix} 5 & 3 & 4 \\ 2 & 1.5 & 6 \\ 7 & 7.5 & 8 \end{pmatrix}$$

◊

Bemerkung 1.2.19.

- *Die Eigenschaft* $\mathbf{AE} = \mathbf{A}$ *macht klar, wieso* $\mathbf{E}$ *„Einheitsmatrix“ genannt wird. Die Einheitsmatrix ist das neutrale Element bezüglich der Matrixmultiplikation in der Menge der quadratischen Matrizen.*
- $\mathbf{AE}$ *entspricht der Verkettung der durch* $\mathbf{A}$ *definierten linearen Funktion und der Funktion, welche einem Vektor diesen selber zuordnet.*

- *Durch die Multiplikation von* $\mathbf{E}$ *und einem Vektor* $\mathbf{x}$ *mit passender Anzahl Komponenten wird dem Vektor* $\mathbf{x}$ *dieser selber zugeordnet, d.h.* $f(\mathbf{x}) = \mathbf{E}\mathbf{x} = \mathbf{x}$. *Rechnen Sie das an einem Beispiel selber nach. Entsprechend kann man umgekehrt im Rechengang* $\mathbf{x}$ *durch* $\mathbf{E}\mathbf{x}$ *ersetzen.*

$\Diamond$

Bemerkung 1.2.20. *Durch Diagonalmatrizen, die auf der Diagonalen nur* 1 *oder* 0 *enthalten, erfolgen senkrechte Projektionen auf Teile des Koordinatennetzes. Es handelt sich um eine Abbildung, welche der* $i-$*ten Koordinate eines Punktes diese zuordnet, wenn* $a_{ii} = 1$ *und welche der* i*-te Koordinate* 0 *zuordnet, wenn* $a_{ii} = 0$. *So wird z.B. der Punkt* $(5, 3, 6)$ *im dreidimensionalen Raum* $\mathbb{R}^3$ *durch die Matrix*

$$A := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

in den Punkt $(5, 3, 0)$ *im zweidimensionalen Raum* $\mathbb{R}^2$ *abgebildet. Die Gerade* $\{z \mid z = (x_1, x_2, x_3)\,\lambda + (y_1, y_2, y_3)\,, \lambda \in \mathbb{R}\ \}$ *im* $\mathbb{R}^3$ *wird durch* A *in die Gerade* $\{z' \mid z' = (x_1, x_2, 0)\,\lambda + (y_1, y_2, 0)\,, \lambda \in \mathbb{R}\}$ *im* $\mathbb{R}^2$ *abgebildet. Wir betrachten als Beispiel die Gerade, die durch die Punkte* $(0, 2, 3)$ *und* $(4, 0, 2)$ *geht. Deren Projektion auf die* (x_1, x_2)*-Ebene ist die Gerade durch die Punkte* $(0, 2, 0)$ *und* $(4, 0, 0)$ *(s. Abbildung 1.2.1).*

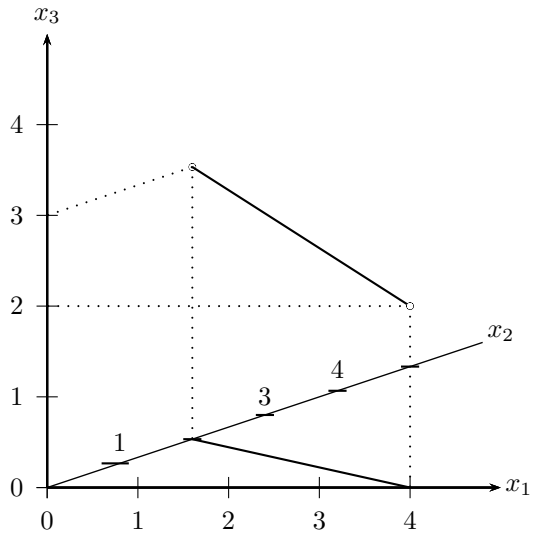


Abbildung 1.2.1: Projektion der Gerade durch die Punkte $(0, 2, 3)$ und $(4, 0, 2)$ auf die $x_1 - x_2$-Ebene

$\Diamond$

Bezüglich der Matrixmultiplikation gelten folgende Gesetze, die wir ohne Beweis festhalten, die Sie aber an einfachen Beispielen exemplarisch durchrechnen sollten:

Satz 1.2.21. *Für* $a \in \mathbb{R}$ *und für geeignete Spalten- und Zeilenanzahlen aufweisende Matrizen* $\mathbf{A}$,$\mathbf{B}$ *und* $\mathbf{C}$ *gilt:*

$$\begin{aligned} (\mathbf{AB})\mathbf{C} &= \mathbf{A}(\mathbf{BC}) \\ a(\mathbf{AB}) &= (a\mathbf{A})\,\mathbf{B} = \mathbf{A}(a\mathbf{B}) \\ \mathbf{A}(\mathbf{B} + \mathbf{C}) &= \mathbf{AB} + \mathbf{AC} \\ (\mathbf{A} + \mathbf{B})\mathbf{C} &= \mathbf{AC} + \mathbf{BC} \end{aligned}$$

Da das Kommutativgesetz für die Matrixmultiplikation nicht gilt, muss die in den Gesetzen festgehaltene Reihenfolge der Matrizen immer strikt eingehalten werden.
Als nächstes folgt ein ökonomisches Beispiel für die Matrixmultiplikation:

Beispiel 1.2.22. *Eine Unternehmung produziere die Zwischenprodukte* 1, 2 *und* 3, *wobei für die Produktion der Zwischenprodukte die Rohstoffe* 1, 2, 3 *und* 4 *verwendet werden. Für eine Einheit des Zwischenproduktes 1 werden* 40, 20, 4 *und* 6 *Einheiten der Rohstoffe* 1, 2, 3 *und* 4 *verwendet.*
Für eine Einheit des Zwischenproduktes 2 *werden* 20, 30, 0 *und* 6 *Einheiten der Rohstoffe* 1, 2, 3 *und* 4 *verwendet.*
Für eine Einheit des Zwischenproduktes 3 *werden* 0, 3, 0 *und* 7 *Einheiten der Rohstoffe* 1, 2, 3 *und* 4 *verwendet.*
Für die 5 *Endprodukte werden pro Einheit Endprodukt jeweils die folgenden Zwischenprodukte verwendet (erstes, zweites und drittes Zwischenprodukt in dieser Reihenfolge):*
$1 : 5, 6, 3$
$2 : 0, 7, 1$
$3 : 4, 8, 3$
$4 : 0, 1, 2$
$5 : 10, 13, 12$
Zu berechnen ist der benötigte Rohstoffvektor für einen Endproduktvektor von $(50, 200, 30, 150, 40)$.
Rohstoffmatrix:

$$\begin{pmatrix} 40 & 20 & 0 \\ 20 & 30 & 3 \\ 4 & 0 & 0 \\ 6 & 6 & 7 \end{pmatrix}$$

Produktionsmatrix:

$$\begin{pmatrix} 5 & 0 & 4 & 0 & 10 \\ 6 & 7 & 8 & 1 & 13 \\ 3 & 1 & 3 & 2 & 12 \end{pmatrix}$$

Das Produkt der beiden Matrizen ist:

$$\begin{pmatrix} 40 & 20 & 0 \\ 20 & 30 & 3 \\ 4 & 0 & 0 \\ 6 & 6 & 7 \end{pmatrix} \begin{pmatrix} 5 & 0 & 4 & 0 & 10 \\ 6 & 7 & 8 & 1 & 13 \\ 3 & 1 & 3 & 2 & 12 \end{pmatrix} = \begin{pmatrix} 320 & 140 & 320 & 20 & 660 \\ 289 & 213 & 329 & 36 & 626 \\ 20 & 0 & 16 & 0 & 40 \\ 87 & 49 & 93 & 20 & 222 \end{pmatrix}$$

Der Vektor der benötigten Rohstoffe ist: $\begin{pmatrix} 320 & 140 & 320 & 20 & 660 \\ 289 & 213 & 329 & 36 & 626 \\ 20 & 0 & 16 & 0 & 40 \\ 87 & 49 & 93 & 20 & 222 \end{pmatrix} \begin{pmatrix} 50 \\ 200 \\ 30 \\ 150 \\ 40 \end{pmatrix} = \begin{pmatrix} 83000 \\ 97360 \\ 3080 \\ 28820 \end{pmatrix}$

$\Diamond$

Multiplizieren wir eine $(n, 1)-$Matrix mit einer $(1, m)-$Matrix, erhält man eine $(n, m)-$Matrix.

Beispiel 1.2.23.

$$\begin{pmatrix} 4 \\ 16 \\ 17 \\ 28 \end{pmatrix} \begin{pmatrix} 5 & 6 & 7 & 20 \end{pmatrix} = \begin{pmatrix} 20 & 24 & 28 & 80 \\ 80 & 96 & 112 & 320 \\ 85 & 102 & 119 & 340 \\ 140 & 168 & 196 & 560 \end{pmatrix}$$

$\Diamond$

Multiplizieren wir eine $(1, n)-$Matrix mit einem $n-$Vektor oder einer $(n, 1)-$Matrix, erhält man eine $(1, 1)-$Matrix (= reelle Zahl).

Beispiel 1.2.24.

$$\begin{pmatrix} 5 & 6 & 7 & 20 \end{pmatrix} \begin{pmatrix} 4 \\ 16 \\ 17 \\ 28 \end{pmatrix} = 795$$

$\Diamond$

Die Matrixmultiplikation von Hand ist mühsam und fehleranfällig. Mit Excel ist diese hingegen einfach zu berechnen. Sie erfolgt mit den gleichen Befehlen wie die Multiplikation einer Matrix mit einem Vektor (=mmult(;)). Statt des Vektors gibt man die zweite Matrix ein. Statt einer Spalte wird ein Tabellenabschnitt mit den Spalten- und Zeilenzahlen von **C** beleuchtet. Entsprechend muss man wissen, wie viele Spalten und Zeilen die Matrix **C** hat, wenn man die Spalten- und Zeilenanzahlen der Matrizen **A** und **B** kennt. Es wird mit „ctrl+shift+enter“ bestätigt. Für $diag(\mathbf{a})$ gibt es keinen speziellen Befehl.
Mit R: $A\%*\%B$. Für (n,n)-Einheitsmatrix $diag(n)$. Für Diagonalmatrix mit der Diagonale $\mathbf{a} = c(a_1,...,a_n)$: $diag(\mathbf{a})$.
Die Zeilensumme einer (n×m)-Matrix **A** können wir durch die Multiplikation

$$\mathbf{A1}$$

erreichen, wobei $\mathbf{1} := (1,...,1) \in \{1\}^m$ ist. So ist

$$\begin{pmatrix} 5 & 3 & 6 \\ 8 & 2 & 3 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 14 \\ 13 \\ 6 \\ 15 \end{pmatrix}$$

Die Spaltensumme einer (n×m)-Matrix **A** können wir durch die die Multiplikation

$$\mathbf{1}^T\mathbf{A}$$

erreichen, wobei $\mathbf{1}^T$ die (1×n)-Zeilenmatrix ist, die nur Einen enthält. So ist

$$\begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 5 & 3 & 6 \\ 8 & 2 & 3 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 18 & 12 & 18 \end{pmatrix}.$$

Mit R für $(1,1,1,1,1)$: $rep(1,5)$, für $\mathbf{1}^T = \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix} : t(rep(1,4))$

Übungen

1. Seien f und g durch die Matrizen

$$A = \begin{pmatrix} 5 & 3 & 6 \\ 8 & 2 & 3 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}; B = \begin{pmatrix} 1 & 13 & 0 \\ 2 & 12 & 3 \\ 3 & 12 & 0 \\ 4 & 15 & 5 \end{pmatrix}$$

   gegeben. Berechnen Sie $(f+g)(5,2,3)$.

2. Sei f durch die Matrix

$$A = \begin{pmatrix} 5 & 3 & 6 \\ 8 & 2 & 3 \end{pmatrix}$$

   gegeben. Berechnen Sie $5f(5,2,3)$.

3. Berechnen Sie folgende Matrixmultiplikationen (von Hand und mit Excel):

$$\begin{pmatrix} 5 & 3 & 6 \\ 8 & 2 & 3 \end{pmatrix} \begin{pmatrix} 4 & 2 \\ 7 & 1 \\ 4 & 3 \end{pmatrix}$$

$$\begin{pmatrix} 4 & 2 \\ 7 & 1 \\ 4 & 3 \end{pmatrix} \begin{pmatrix} 5 & 3 & 6 \\ 8 & 2 & 3 \end{pmatrix}$$

$$\begin{pmatrix} 4 & 2 & 13 \\ 7 & 1 & 4 \\ 4 & 3 & 6 \end{pmatrix} \begin{pmatrix} 4 & 2 & 13 \\ 7 & 1 & 4 \\ 4 & 3 & 6 \end{pmatrix}$$

$$\begin{pmatrix} 7 & 1 & 4 \end{pmatrix} \begin{pmatrix} 4 & 2 & 13 \\ 7 & 1 & 4 \\ 4 & 3 & 6 \end{pmatrix}$$

$$\begin{pmatrix} 5 \\ 8 \end{pmatrix} \begin{pmatrix} 7 & 1 & 4 \end{pmatrix}$$

4. Seien f und g durch die Matrizen

$$A = \begin{pmatrix} 5 & 3 & 6 \\ 8 & 2 & 3 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}; B = \begin{pmatrix} 1 & 13 & 0 & 14 \\ 2 & 12 & 3 & 0 \\ 3 & 12 & 0 & 8 \end{pmatrix}$$

gegeben. Berechnen Sie $f(g(5, 2, 3, 4))$.

5. Seien die folgende Rohstoff- und Produktionsmatrix gegeben (Zwischenprodukte (ZP); Endprodukt (EP)):

	$ZP\ 1$	$ZP\ 2$	$ZP\ 3$	$ZP\ 4$
$Rohstoff\ 1$	5	3	6	4
$Rohstoff\ 2$	8	2	3	3
$Rohstoff\ 3$	1	2	0	3
$Rohstoff\ 4$	4	5	6	1

	$EP\ 1$	$EP\ 2$	$EP\ 3$
$ZP\ 1$	16	12	25
$ZP\ 2$	228	12	31
$ZP\ 3$	2.5	0	3
$ZP\ 4$	4	18	0.5

a) Berechnen Sie für den Outputvektor der Endprodukte $(20, 400, 50)$ den Inputvektor der Rohstoffe.
b) Wenn die Rohstoffpreise pro Einheit durch die folgende Preismatrix gegeben ist (für RS 1 ... RS 4 in dieser Reihenfolge)

$$\begin{pmatrix} 30 & 3 & 40 & 50 \end{pmatrix},$$

wieviel betragen dann die Warenkosten für den Outputvektor $(20, 400, 50)$.

Lösungen

1.

$$
\begin{aligned}
(f+g)(5,2,3) = (A+B)\begin{pmatrix} 5 \\ 2 \\ 3 \end{pmatrix} = \\
&= \left(\begin{pmatrix} 5 & 3 & 6 \\ 8 & 2 & 3 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \begin{pmatrix} 1 & 13 & 0 \\ 2 & 12 & 3 \\ 3 & 12 & 0 \\ 4 & 15 & 5 \end{pmatrix}\right)\begin{pmatrix} 5 \\ 2 \\ 3 \end{pmatrix} \\
&= \begin{pmatrix} 6 & 16 & 6 \\ 10 & 14 & 6 \\ 4 & 14 & 3 \\ 8 & 20 & 11 \end{pmatrix}\begin{pmatrix} 5 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 80 \\ 96 \\ 57 \\ 113 \end{pmatrix}
\end{aligned}
$$

2.

$$
\begin{aligned}
5f(5,2,3) = (5A)\begin{pmatrix} 5 \\ 2 \\ 3 \end{pmatrix} \\
&= \left(5\begin{pmatrix} 5 & 3 & 6 \\ 8 & 2 & 3 \end{pmatrix}\right)\begin{pmatrix} 5 \\ 2 \\ 3 \end{pmatrix} \\
&= \begin{pmatrix} 25 & 15 & 30 \\ 40 & 10 & 15 \end{pmatrix}\begin{pmatrix} 5 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 245 \\ 265 \end{pmatrix}
\end{aligned}
$$

3.

$$
\begin{pmatrix} 5 & 3 & 6 \\ 8 & 2 & 3 \end{pmatrix}\begin{pmatrix} 4 & 2 \\ 7 & 1 \\ 4 & 3 \end{pmatrix} = \begin{pmatrix} 65 & 31 \\ 58 & 27 \end{pmatrix}
$$

$$
\begin{pmatrix} 4 & 2 \\ 7 & 1 \\ 4 & 3 \end{pmatrix}\begin{pmatrix} 5 & 3 & 6 \\ 8 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 36 & 16 & 30 \\ 43 & 23 & 45 \\ 44 & 18 & 33 \end{pmatrix}
$$

$$
\begin{pmatrix} 4 & 2 & 13 \\ 7 & 1 & 4 \\ 4 & 3 & 6 \end{pmatrix}\begin{pmatrix} 4 & 2 & 13 \\ 7 & 1 & 4 \\ 4 & 3 & 6 \end{pmatrix} = \begin{pmatrix} 82 & 49 & 138 \\ 51 & 27 & 119 \\ 61 & 29 & 100 \end{pmatrix}
$$

$$
\begin{pmatrix} 7 & 1 & 4 \end{pmatrix}\begin{pmatrix} 4 & 2 & 13 \\ 7 & 1 & 4 \\ 4 & 3 & 6 \end{pmatrix} = \begin{pmatrix} 51 & 27 & 119 \end{pmatrix}
$$

$$
\begin{pmatrix} 5 \\ 8 \end{pmatrix}\begin{pmatrix} 7 & 1 & 4 \end{pmatrix} = \begin{pmatrix} 35 & 5 & 20 \\ 56 & 8 & 32 \end{pmatrix}
$$

4.

$$f(g(5,2,3,4)) = (AB)\begin{pmatrix} 5 \\ 2 \\ 3 \\ 4 \end{pmatrix} =$$

$$= \left(\begin{pmatrix} 5 & 3 & 6 \\ 8 & 2 & 3 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}\begin{pmatrix} 1 & 13 & 0 & 14 \\ 2 & 12 & 3 & 0 \\ 3 & 12 & 0 & 8 \end{pmatrix}\right)\begin{pmatrix} 5 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

$$= \begin{pmatrix} 29 & 173 & 9 & 118 \\ 21 & 164 & 6 & 136 \\ 14 & 73 & 6 & 38 \\ 32 & 184 & 15 & 104 \end{pmatrix}\begin{pmatrix} 5 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 990 \\ 995 \\ 386 \\ 989 \end{pmatrix}$$

5. a)

$$\begin{pmatrix} 5 & 3 & 6 & 4 \\ 8 & 2 & 3 & 3 \\ 1 & 2 & 0 & 3 \\ 4 & 5 & 6 & 1 \end{pmatrix}\begin{pmatrix} 16 & 12 & 25 \\ 228 & 12 & 31 \\ 2.5 & 0 & 3 \\ 4 & 18 & 0.5 \end{pmatrix} = \begin{pmatrix} 795 & 168 & 238 \\ 603.5 & 174 & 272.5 \\ 484 & 90 & 88.5 \\ 1223 & 126 & 273.5 \end{pmatrix}$$

$$\begin{pmatrix} 795 & 168 & 238 \\ 603.5 & 174 & 272.5 \\ 484 & 90 & 88.5 \\ 1223 & 126 & 273.5 \end{pmatrix}\begin{pmatrix} 20 \\ 400 \\ 50 \end{pmatrix} = \begin{pmatrix} 95000 \\ 95295 \\ 50105 \\ 88535 \end{pmatrix}$$

b)

$$\begin{pmatrix} 30 & 3 & 40 & 50 \end{pmatrix}\begin{pmatrix} 95000 \\ 95295 \\ 50105 \\ 88535 \end{pmatrix} = 9566\,835$$

1.3 Umkehrfunktionen linearer Abbildungen und Matrizen

Wir haben oben mit Hilfe der Multiplikation einer Produktionsmatrix $\mathbf{A}$ mit dem Outputvektor $\mathbf{x}$ den Inputvektor $\mathbf{y}$ berechnet:

$$f(\mathbf{x}) = \mathbf{A}\mathbf{x} = \mathbf{y}$$

Manchmal ist es nützlich, für einen gegebenen Inputvektor $\mathbf{y}$ den Outputvektor $\mathbf{x}$ zu berechnen, wobei es nicht unbedingt eine sinnvolle Lösung gibt - es ist nicht sicher, dass es für einen gegebenen Inputvektor eine Produktion gibt, die fertige Endprodukte liefert und alle Rohstoffe aufbraucht. Rechnerisch entspricht das Problem der Berechnung der Umkehrfunktion von f :

$$f^{-1}(\mathbf{y}) = \mathbf{x}.$$

Die Umkehrfunktion ist nur definiert, wenn die lineare Funktion f bijektiv ist. Ohne Beweis halten wir fest: Lineare Funktionen sind nur bijektiv, wenn sie durch eine quadratische Matrix bestimmt sind. Allerdings legen nicht alle quadratischen Matrizen bijektive Funktionen fest. Welche Teilmenge der quadratischen Matrizen bijektive Funktionen entsprechen, wird später zu diskutieren sein.
Auf Grund des Zusammenhangs von linearen Funktionen und Matrizen, suchen wir für eine bijektive lineare Funktion f, die durch die Matrix $\mathbf{A}$ bestimmt ist, eine Matrix $\mathbf{B}$, welche die Umkehrfunktion von f bestimmt. Für $\mathbf{B}$ gilt also:

$$f^{-1}(\mathbf{y}) = \mathbf{B}\mathbf{y} = \mathbf{x}$$

Rein formal ist das Problem einfach zu lösen. Gibt es eine Matrix $\mathbf{B}$, für die

$$\mathbf{BA} = \mathbf{E},$$

gilt:

$$\mathbf{BAx} = \mathbf{Ex} = \mathbf{x}$$

und damit durch Multiplikation beider Seiten der Gleichung $\mathbf{Ax} = \mathbf{y}$ von links her mit $\mathbf{B}$

$$\begin{aligned}\mathbf{BAx} &= \mathbf{By} \Longrightarrow \\ \mathbf{x} &= \mathbf{By}\end{aligned}$$

Wir halten diese Überlegungen noch formaler fest:

Satz 1.3.1. *Sei die lineare Abbildung f bijektiv und $\mathbf{A}$ die Matrix mit $f(x) = \mathbf{Ax}$. Dann gibt es eine Matrix $\mathbf{B}$, welche die Umkehrfunktion f^{-1} von f definiert, d.h. es gilt $f^{-1}(\mathbf{x}) = \mathbf{Bx}$.*

Analysieren wir die Bedingung $\mathbf{BA} = \mathbf{E}$, muss gelten:

$$\sum_{l=1}^{n} b_{il} a_{lj} = 0$$

für $i \neq j$ sowie $1 \leq i, j \leq n$ und

$$\sum_{l=1}^{n} b_{il} a_{li} = 1$$

für $1 \leq i \leq n$, wobei b_{il} unbekannt ist. Dies führt auf ein Gleichungssystem von n^2 Gleichungen und n^2 Unbekannten. Dieses Gleichungssystem besteht aus n Untergleichungssystemen mit je n Unbekannten, die von Untergleichungssystem zu Untergleichungssystem verschieden sind.

Definition 1.3.2. *Eine Matrix $\mathbf{B}$ mit $\mathbf{BA} = \mathbf{E}$ wird „inverse Matrix von $\mathbf{A}$“ oder „die Inverse von $\mathbf{A}$“ oder „Umkehrmatrix von $\mathbf{A}$“ genannt. Wir kürzen sie ab durch $\mathbf{A}^{-1}$.* ◊

Beispiel 1.3.3. *Zu bestimmen sei die Umkehrmatrix von*

$$\begin{pmatrix} 4 & 5 \\ 3 & 2 \end{pmatrix}.$$

Für die Inverse muss gelten:

$$\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} 4 & 5 \\ 3 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Durch Ausmultiplizieren von

$$\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} 4 & 5 \\ 3 & 2 \end{pmatrix}$$

erhält man

$$\begin{pmatrix} 4b_{11} + 3b_{12} & 5b_{11} + 2b_{12} \\ 4b_{21} + 3b_{22} & 5b_{21} + 2b_{22} \end{pmatrix}$$

und damit

$$\begin{pmatrix} 4b_{11} + 3b_{12} & 5b_{11} + 2b_{12} \\ 4b_{21} + 3b_{22} & 5b_{21} + 2b_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Dies führt auf folgendes Gleichungssystem - das eigentlich aus zwei Gleichungssystemen mit je zwei Unbekannten besteht:

$$\begin{aligned} 4b_{11} + 3b_{12} &= 1 \\ 5b_{11} + 2b_{12} &= 0 \end{aligned}$$

$$4b_{21} + 3b_{22} = 0$$
$$5b_{21} + 2b_{22} = 1$$

Dessen Auflösung führt zu $b_{11} = -\frac{2}{7}, b_{12} = \frac{5}{7}, b_{21} = \frac{3}{7}, b_{22} = -\frac{4}{7}$
Damit ist die Inverse $\mathbf{A}^{-1}$*von* $\mathbf{A}$:

$$\begin{pmatrix} -\frac{2}{7} & \frac{5}{7} \\ \frac{3}{7} & -\frac{4}{7} \end{pmatrix}$$

Zur Kontrolle multiplizieren wir $\mathbf{A}^{-1}$ *und* $\mathbf{A}$:

$$\begin{pmatrix} -\frac{2}{7} & \frac{5}{7} \\ \frac{3}{7} & -\frac{4}{7} \end{pmatrix} \begin{pmatrix} 4 & 5 \\ 3 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

◊

Beispiel 1.3.4. *Zu bestimmen sei die Umkehrmatrix von*

$$\begin{pmatrix} 2 & 5 & 3 \\ 2 & 1 & 4 \\ 3 & 16 & 7 \end{pmatrix}.$$

Für die Inverse muss gelten:

$$\begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \begin{pmatrix} 2 & 5 & 3 \\ 2 & 1 & 4 \\ 3 & 16 & 7 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Durch Ausmultiplizieren von

$$\begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \begin{pmatrix} 2 & 5 & 3 \\ 2 & 1 & 4 \\ 3 & 16 & 7 \end{pmatrix}$$

erhält man

$$\begin{pmatrix} 2b_{11} + 2b_{12} + 3b_{13} & 5b_{11} + b_{12} + 16b_{13} & 3b_{11} + 4b_{12} + 7b_{13} \\ 2b_{21} + 2b_{22} + 3b_{23} & 5b_{21} + b_{22} + 16b_{23} & 3b_{21} + 4b_{22} + 7b_{23} \\ 2b_{31} + 2b_{32} + 3b_{33} & 5b_{31} + b_{32} + 16b_{33} & 3b_{31} + 4b_{32} + 7b_{33} \end{pmatrix}$$

und damit

$$\begin{pmatrix} 2b_{11} + 2b_{12} + 3b_{13} & 5b_{11} + b_{12} + 16b_{13} & 3b_{11} + 4b_{12} + 7b_{13} \\ 2b_{21} + 2b_{22} + 3b_{23} & 5b_{21} + b_{22} + 16b_{23} & 3b_{21} + 4b_{22} + 7b_{23} \\ 2b_{31} + 2b_{32} + 3b_{33} & 5b_{31} + b_{32} + 16b_{33} & 3b_{31} + 4b_{32} + 7b_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Dies führt auf das Gleichungssystem - das eigentlich aus drei Gleichungssystemen mit je drei Unbekannten besteht:

$$2b_{11} + 2b_{12} + 3b_{13} = 1$$
$$5b_{11} + b_{12} + 16b_{13} = 0$$
$$3b_{11} + 4b_{12} + 7b_{13} = 0$$

$$2b_{21} + 2b_{22} + 3b_{23} = 0$$
$$5b_{21} + b_{22} + 16b_{23} = 1$$
$$3b_{21} + 4b_{22} + 7b_{23} = 0$$

$$2b_{31} + 2b_{32} + 3b_{33} = 0$$
$$5b_{31} + b_{32} + 16b_{33} = 0$$
$$3b_{31} + 4b_{32} + 7b_{33} = 1$$

Dessen Auflösung führt zu $b_{11} = \frac{57}{37}, b_{12} = -\frac{13}{37}, b_{21} = \frac{2}{37}, b_{13} = -\frac{17}{37}$,
$b_{22} = -\frac{5}{37}, b_{31} = -\frac{29}{37}, b_{23} = \frac{2}{37}, b_{32} = \frac{17}{37}, b_{33} = \frac{8}{37}$
Damit ist die Inverse $\mathbf{A}^{-1}$ *von* $\mathbf{A}$:

$$\begin{pmatrix} \frac{57}{37} & -\frac{13}{37} & -\frac{17}{37} \\ \frac{2}{37} & -\frac{5}{37} & \frac{2}{37} \\ -\frac{29}{37} & \frac{17}{37} & \frac{8}{37} \end{pmatrix}$$

Zur Kontrolle multiplizieren wir $\mathbf{A}^{-1}$ *und* $\mathbf{A}$:

$$\begin{pmatrix} \frac{57}{37} & -\frac{13}{37} & -\frac{17}{37} \\ \frac{2}{37} & -\frac{5}{37} & \frac{2}{37} \\ -\frac{29}{37} & \frac{17}{37} & \frac{8}{37} \end{pmatrix} \begin{pmatrix} 2 & 5 & 3 \\ 2 & 1 & 4 \\ 3 & 16 & 7 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

◊

Man sieht, dass die Berechnung der Inversen ziemlich mühsam ist. Mit Excel erfolgt sie wie folgt: Man gibt die zu invertierende Matrix ein. Dann beleuchtet man einen Tabellenabschnitt mit gleichviel Spalten und Zeilen wie die zu invertierende Matrix. Dann wird in die erste Zelle der beleuchteten Teilmatrix „minv()“ eingegeben und mit „ctrl+shift+enter“ bestätigt, wobei wir im Beispiel 1.3.4 mit Excel (gerundete) Dezimalbrüche erhalten. Man erhält

$$\mathbf{A}^{-1} := \begin{pmatrix} 1.540540541 & -0.351351351 & -0.459459459 \\ 0.054054054 & -0.135135135 & 0.054054054 \\ -0.783783784 & 0.459459459 & 0.216216216 \end{pmatrix}$$

$$\mathbf{A}^{-1} \begin{pmatrix} 2 & 5 & 3 \\ 2 & 1 & 4 \\ 3 & 16 & 7 \end{pmatrix} = \begin{pmatrix} 1 & 1 \times 10^{-8} & 6 \times 10^{-9} \\ -1.3553 \times 10^{-20} & 1 & -5.421 \times 10^{-20} \\ -2 \times 10^{-9} & -5 \times 10^{-9} & 1 \end{pmatrix}$$

.

Wegen Rundungsfehlern führt die Multiplikation $\mathbf{A}^{-1}\mathbf{A}$ nicht genau auf die Einheitsmatrix.
Mit R: solve(A)

Satz 1.3.1. *Die Umkehrmatrix* $\mathbf{B}$ *einer Diagonalmatrix* $\mathbf{diag}(\mathbf{a})$ *mit* $a = (a_1, ...a_n)$ *ist* $\mathbf{diag}\left((\frac{1}{a_1}, ..., \frac{1}{a_n})\right)$.

Beweis. Es muss gelten

$$\mathbf{Bdiag}(\mathbf{a}) = \mathbf{E}$$

und man erhält das Gleichungssystem:

$$b_{11}a_1 = 1$$
$$b_{22}a_2 = 1$$
$$\vdots$$
$$b_{nn}a_n = 1$$

Entsprechend gilt $b_{11} = \frac{1}{a_1}; b_{22} = \frac{1}{a_2}; ...; b_{nn} = \frac{1}{a_n}$. □

Beispiel 1.3.5. $\mathbf{diag}((5,3,2)) = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix}$.

$$\begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix} = \begin{pmatrix} 5b_{11} & 5b_{12} & 5b_{13} \\ 3b_{21} & 3b_{22} & 3b_{23} \\ 2b_{31} & 2b_{32} & 2b_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Für $5b_{12} = 0$ z.B. ist $b_{12} = 0$. Dies gilt für alle Koeffizienten b_{ij} mit $i \neq j$. Für $5b_{11} = 1$ erhält man $b_{11} = \frac{1}{5}$. Es gilt also $b_{11} = \frac{1}{5}, b_{22} = \frac{1}{3}$ und $b_{33} = \frac{1}{2}$. Die Inverse von $\mathbf{diag}\,((5,3,2))$ *ist damit*

$$\begin{pmatrix} \frac{1}{5} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix}$$

und es gilt

$$\begin{pmatrix} \frac{1}{5} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$\Diamond$

Beispiel 1.3.6. *Input-Output-Analyse (Wassily Leontief, Deutscher Ökonom, Nobelpreisträger, 1905-1999): Volkswirtschaften, Regionalwirtschaften oder grössere Unternehmen können als Netz von Sektoren oder Teilunternehmungen betrachtet werden, die*

- *sich gegenseitig beliefern,*
- *Güter von Ausserhalb des Verflechtungssystems wie Arbeit, Rohstoffe und Zwischenprodukte brauchen, um produzieren zu können und die*
- *Produkte für Wirtschaftssubjekte ausserhalb des Verflechtungsnetzes bereitstellen.*

Wir nennen Inputs, die von Aussen in ein solches Verflechtungssystem gelangen, „exogene Inputs“. Inputs in das Verflechtungssystem, die innerhalb dieses produziert werden, nennen wir „endogene“ Inputs. Produkte und Dienstleistungen, die vom System produziert werden und die ausserhalb des Verflechtungssystems verwendet oder konsumiert werden, nennen wir Outputs (s. Abbildung 1.3.2). Die Summe der Outputs und der endogenen Inputs ist die Gesamtproduktion des Verflechtungssystems („exo“ griechisch für „ausserhalb“, „endo“ griechisch für „innerhalb“; „gen“ griechisch für „produziert“)
Ziel der Input-Output-Analyse ist

- *die Berechnung der Gesamtproduktion pro Sektor bei gegebenem Output-Vektor und gegebenem Verflechtungssystem,*
- *Berechnung der Liefermengen der Sektoren an die Sektoren bei gegebenem Output-Vektor.*
- *die Berechnung des Output-Vektors bei gegebener Gesamtproduktion,*
- *die Bestimmung des Rohstoffvektors bei gegebenem Output-Vektor und gegebener Rohstoffmatrix.*
- *zu gegebenen Rohstoffvektor Bestimmung des Output-Vektors (falls definiert).*

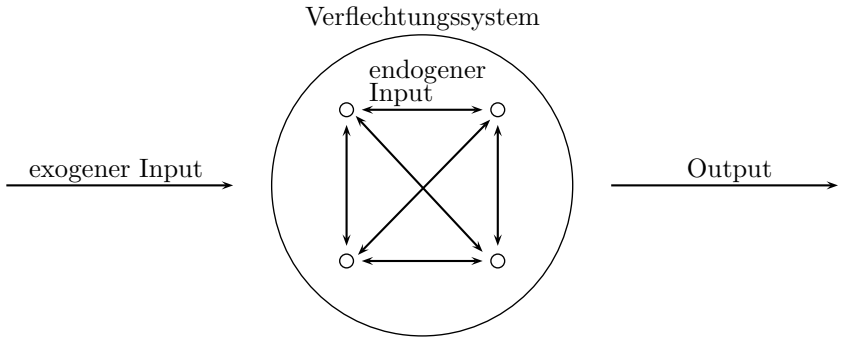


Abbildung 1.3.2: Graphische Erläuterung einiger Begriffe der Input-Output-Analyse

Die obige Situation kann mathematisch durch die folgenden Objekte dargestellt werden:

- *Das Geflecht der Lieferbeziehungen im System durch eine Matrix* $\mathbf{X}_{nn}$*:* x_{ij} *gibt die Güter und Dienstleistungen an, die aus dem Sektor* i *an den Sektor* j *geliefert werden.*
- *Die Outputs können wir durch einen Vektor* $\mathbf{y}_n$ *ausdrücken: für jeden Sektor* i *gibt er den Output* y_i *an, der im Spezialfall auch* 0 *sein kann.*
- $\mathbf{x}$ *sei der Vektor der Gesamtproduktion der* n *Sektoren, wobei offensichtlich gilt:*

$$x_i = \sum_{j=1}^{n} x_{ij} + y_i$$

  oder in Matrixschreibweise

$$\mathbf{x} = \mathbf{X1} + \mathbf{y}$$

  mit $\mathbf{1} := (1, 1, ..., 1, 1) \in \{1\}^n$.

- *Den Zusammenhang von exogenem Input und Gesamtproduktion* $\mathbf{x}$ *des Verflechungssystems können wir durch eine Rohstoffmatrix oder Produktionsmatrix* $\mathbf{R}_{mn}$ *mit den Komponenten* r_{ij} *darstellen, welche angeben, wieviel externen Input oder Rohstoff* i *man für die Produktion einer Einheit des Sektors* j *braucht.*

Wir können $\sum_{j=1}^{n} x_{ij}$ *mit Hilfe von* $\mathbf{x}$ *darstellen, indem wir geeignete Koeffizienten* a_{ij} *wählen. Mit*

$$a_{ij} := \frac{x_{ij}}{x_j}$$

gilt nämlich

$$\sum_{j=1}^{n} a_{ij} x_j = \sum_{j=1}^{n} \frac{x_{ij}}{x_j} x_j = \sum_{j=1}^{n} x_{ij}.$$

Damit erhält man:

$$x_i = \sum_{j=1}^{n} x_{ij} + y_i = \sum_{j=1}^{n} a_{ij} x_j + y_i$$

oder in Matrixschreibweise mit $\mathbf{A} := (a_{ij})$

$$\begin{aligned} \mathbf{x} &= \mathbf{X1} + \mathbf{y} \\ &= \mathbf{Ax} + \mathbf{y}. \end{aligned} \tag{1.3.4}$$

Damit gilt auch

$$\mathbf{X1} = \mathbf{Ax}.$$

Die Matrix $\mathbf{A}$ *wird Technologie- oder Inputmatrix genannt wird. Sie ist von der produzierten Menge* $\mathbf{x}$ *unabhängig, da wir die vom Teilbetrieb* j *erhaltenen endogenen Inputs durch die von* j *produzierte Gesamtmenge dividieren. Sie gibt damit den Stand der Technologie bezüglich des nötigen endogenen Inputs pro Einheit Gesamtproduktion wieder. Rechnerisch erhalten wir* $\mathbf{A}$*, indem wir die Spalte* j *von* $\mathbf{X}$ *mit* $\frac{1}{x_j}$ *multiplizieren. Mit Matrixmultiplikation erhält man* $\mathbf{A}$ *aus* $\mathbf{X}$ *durch*

$$\mathbf{A} = \mathbf{X} \cdot (\mathbf{diag}(\mathbf{x}))^{-1}. \tag{1.3.5}$$

Dies sieht man, wenn man ein einfaches Beispiel betrachtet:

$$\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} \begin{pmatrix} \frac{1}{x_1} & 0 \\ 0 & \frac{1}{x_2} \end{pmatrix} = \begin{pmatrix} \frac{1}{x_1}x_{11} & \frac{1}{x_2}x_{12} \\ \frac{1}{x_1}x_{21} & \frac{1}{x_2}x_{22} \end{pmatrix}$$

Die Technologiematrix liefert die Möglichkeit, für eine beliebige Gesamtproduktion $\mathbf{x}$ *die von* i *an* j *zu liefernden Mengen zu berechnen, d.h.* $\mathbf{X}$ *zu berechnen. Man erhält* $\mathbf{X}$ *aus* $\mathbf{A}$ *für beliebige* $\mathbf{x}$ *durch Multiplikation der* j*-ten Spalte mit* x_j *oder durch Matrixmultiplikation mittels*

$$\mathbf{X} = \mathbf{A} \cdot \mathbf{diag}(\mathbf{x}).$$

$\mathbf{X} = \mathbf{A} \cdot \mathbf{diag}(\mathbf{x})$ *erhält man aus (1.3.5), indem man beide Seiten von rechts her mit* $\mathbf{diag}(\mathbf{x})$ *multipliziert.*
Gemäss (1.3.4) gilt:

$$\begin{aligned} \mathbf{y} &= \mathbf{x} - \mathbf{A}\mathbf{x} \\ &= \mathbf{E}\mathbf{x} - \mathbf{A}\mathbf{x} \\ &= (\mathbf{E} - \mathbf{A})\,\mathbf{x} \end{aligned}$$

und

$$\begin{aligned} (\mathbf{E} - \mathbf{A})^{-1}\,\mathbf{y} &= \mathbf{E}\mathbf{x} \\ &= \mathbf{x} \end{aligned}$$

sofern die Inverse von $(\mathbf{E} - \mathbf{A})$ *existiert.*

Geht man von einer bestehenden Produktion mit entsprechender Matrix $\mathbf{X}$ *und Outputvektor* $\mathbf{y}$ *aus, besteht der Nutzen der Berechnung der Technologiematrix darin, dass man damit für alternative Outputs* $\mathbf{y}'$ *die benötigte Gesamtleistung der Sektoren sowie die benötigten Lieferungen der Sektoren an die anderen Sektoren berechnen kann.*
— Der erste Schritt besteht in der Berechnung der Technologiematrix: $\mathbf{A} = \mathbf{X} \cdot (\mathbf{diag}(\mathbf{x}))^{-1}$
— Für gegebenen alternativen Output $\mathbf{y}'$ *berechnet man dann die benötigte, von den Sektoren zu erbringende Gesamtproduktion:* $\mathbf{x}' = (\mathbf{E} - \mathbf{A})^{-1}\,\mathbf{y}'$
— Man berechnet die Lieferungen der Sektoren an die anderen Sektoren: $\mathbf{X}' = \mathbf{A} \cdot \mathbf{diag}(\mathbf{x}')$

Bemerkung*: (1) Liegt noch eine Rohstoffmatrix vor, so gilt für den Vektor* $\mathbf{r}$ *der benötigten Rohstoffe:*

$$\mathbf{r} = \mathbf{R}\mathbf{x}$$

und mit $\mathbf{x} = (\mathbf{E} - \mathbf{A})^{-1}\,\mathbf{y}$ *erhält man für die benötigten Rohstoffe:*

$$\mathbf{r} = \mathbf{R}\,(\mathbf{E} - \mathbf{A})^{-1}\mathbf{y}$$

(2) Umgekehrt erhält man unter der Voraussetzung der Invertierbarkeit der Rohstoffmatrix aus $\mathbf{r} = \mathbf{R}\left(\mathbf{E} - \mathbf{A}\right)^{-1}\mathbf{y}$*:*

$$\begin{aligned}
\mathbf{R}^{-1}\mathbf{r} &= \mathbf{R}^{-1}\mathbf{R}\left(\mathbf{E} - \mathbf{A}\right)^{-1}\mathbf{y} \\
\mathbf{R}^{-1}\mathbf{r} &= \left(\mathbf{E} - \mathbf{A}\right)^{-1}\mathbf{y} \\
\left(\mathbf{E} - \mathbf{A}\right)\mathbf{R}^{-1}\mathbf{r} &= \left(\mathbf{E} - \mathbf{A}\right)\left(\mathbf{E} - \mathbf{A}\right)^{-1}\mathbf{y} \\
\left(\mathbf{E} - \mathbf{A}\right)\mathbf{R}^{-1}\mathbf{r} &= \mathbf{y}
\end{aligned}$$

Man kann also für gegebenen Rohstoffvektor den Outputvektor berechnen, wobei diese eher selten vernünftige Resultate ergeben wird. Es ist oft nicht möglich, für gegebenen Rohstoffvektor einen Output zu produzieren, so dass alle Rohstoffe aufgebraucht werden. ◊

Übungen

1. Berechnen Sie die Umkehrmatrix von

$$\begin{pmatrix} 5 & 3 \\ 1 & 4 \end{pmatrix}$$

   von Hand und mit Excel. Kontrollieren Sie, ob $\mathbf{A}\mathbf{A}^{-1} = \mathbf{E}$.

2. Berechnen Sie, die Umkehrmatrix von $\begin{pmatrix} 4 & 15 & 2 \\ 1 & 2 & 0 \\ 3 & 7 & 4 \end{pmatrix}$ von Hand und mit Excel. Kontrollieren Sie, ob $\mathbf{A}\mathbf{A}^{-1} = \mathbf{E}$.

3. Sei die folgende Produktionsmatrix $\mathbf{A}$ gegeben, welche den nötigen Input $\mathbf{y}$ als Funktion $\mathbf{y} = f(\mathbf{x}) = \mathbf{A}\mathbf{x}$ des Outputs $\mathbf{x}$ ausdrückt. Suchen Sie die Matrix, welche den Output als Funktion des Inputs ausdrückt.

$$\mathbf{A} = \begin{pmatrix} 15 & 2 & 35 \\ 28 & 11 & 12 \\ 1.5 & 5 & 0 \end{pmatrix}$$

   Berechnen Sie den Output für den Inputvektor $(15, 20, 2)$

4. In einer Volkswirtschaft werden die drei Sektoren 1, 2 und 3 unterschieden: Es ist die folgende Verflechtungsmatrix gegeben, welche die Flüsse der in den Sektoren produzierten Produkte und Dienstleistungen (in Geldeinheiten) angibt:

$$\mathbf{X} = \left(\begin{array}{c|ccc} & S_1 & S_2 & S_3 \\ \hline S_1 & 500 & 200 & 0 \\ S_2 & 1000 & 7000 & 15000 \\ S_3 & 600 & 700 & 5000 \end{array}\right)$$

   Die drei Sektoren produzieren auch für den Konsum und den Export, wobei der Outputvektor der drei Sektoren gegeben ist durch

$$\mathbf{y} = \begin{pmatrix} 500 \\ 2000 \\ 3000 \end{pmatrix}$$

Zudem sei die folgende Rohstoffmatrix gegeben, welche die benötigen Rohstoffeinheiten angibt, die für die Produktion einer Einheit des Sektors S_i benötigt werden.

$$\mathbf{R} = \left(\begin{array}{c|ccc} & S_1 & S_2 & S_3 \\ \hline R_1 & 3 & 1 & 0 \\ R_2 & 5 & 0.3 & 5 \\ R_3 & 4 & 2 & 3 \end{array}\right)$$

a) Berechnen Sie die Gesamtproduktion $\mathbf{x}$
b) Berechnen Sie die Technologiematrix $\mathbf{A}$
c) Berechnen Sie für den Outputvektor $\begin{pmatrix} 2650 \\ 10600 \\ 15900 \end{pmatrix}$ die Gesamtproduktion $\mathbf{x}'$.
d) Berechnen Sie für den Outputvektor von c) die Verflechtungsmatrix $\mathbf{X}'$ (Lieferungen der Sektoren an die Sektoren)
e) Berechnen Sie für den Outputvektor von c) den Rohstoffvektor.
f) Berechnen Sie für den Outputvektor $(3500, 8000, 12000)$ den Rohstoffvektor.
g) Berechnen Sie den Outputvektor, wenn für die Sektoren folgender Rohstoffvektor vorliegen würde $(500, 600, 1200)$:
h) Output für den Rohstoffvektor $(700, 3000, 4120)$

5. Gegeben sei die folgende Matrix, welche die Leistungen angibt, welche Teilbetriebe für andere Teilbetriebe einer Firma beim spezifischen, gegebenen Output erbringen:

Teilbetriebe	S1	S2	S3	S4	S5	Output	
S1	40	30	50	20	60	500	
S2	30	40	40	0	30	450	
S3	50	60	70	90	10	400	
S4	40	110	120	25	3	60	
S5	4	5	13	43	15	700	

Tabelle 1.3.1: Daten der Aufgabe 3

a) Berechnen Sie die Technologiematrix.
b) Berechnen Sie die Leistungen, welche die Teilbetriebe für die Teilbetriebe erbringen müssen, um den Output (59,65, 73, 200, 30) zu produzieren.

Lösungen

1.
$$\begin{pmatrix} \frac{4}{17} & -\frac{3}{17} \\ -\frac{1}{17} & \frac{5}{17} \end{pmatrix} \begin{pmatrix} 5 & 3 \\ 1 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

2.
$$\begin{pmatrix} -\frac{4}{13} & \frac{23}{13} & \frac{2}{13} \\ \frac{2}{13} & -\frac{5}{13} & -\frac{1}{13} \\ -\frac{1}{26} & -\frac{17}{26} & \frac{7}{26} \end{pmatrix} \begin{pmatrix} 4 & 15 & 2 \\ 1 & 2 & 0 \\ 3 & 7 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

3.
$$\begin{pmatrix} -1.734\,9 \times 10^{-2} & 5.060\,0 \times 10^{-2} & -0.104\,38 \\ 5.204\,6 \times 10^{-3} & -1.518\,0 \times 10^{-2} & 0.231\,31 \\ 3.570\,9 \times 10^{-2} & -2.081\,8 \times 10^{-2} & 3.151\,7 \times 10^{-2} \end{pmatrix} \begin{pmatrix} 15 \\ 20 \\ 2 \end{pmatrix} = \begin{pmatrix} 0.543\,01 \\ 0.237\,09 \\ 0.182\,31 \end{pmatrix}$$

Da in der Matrix negative Zahlen auftauchen, ist es möglich, einen Outputvektor mit negativen Zahlen zu erhalten, was dann wenig Sinn macht!

4. a) Gesamtproduktion $\mathbf{x}$

$$\mathbf{x} = \mathbf{X1} + \mathbf{y} = \begin{pmatrix} 500+200+0 \\ 1000+7000+15000 \\ 600+700+5000 \end{pmatrix} + \begin{pmatrix} 500 \\ 2000 \\ 3000 \end{pmatrix} = \begin{pmatrix} 1200 \\ 25\,000 \\ 9300 \end{pmatrix}$$

Mit R:
X=matrix(c(500,200,0,1000,7000,15000,600,700,5000),ncol=3,byrow=T)
y=c(500,2000,3000)
R=matrix(c(3,1,0,5,0.3,5,4,2,3),ncol=3,byrow=T)
x=rowSums(X)+y

b) Technologiematrix:

$$\mathbf{A} = \mathbf{X}(\mathbf{diag}(\mathbf{x})^{-1}) = \begin{pmatrix} 500 & 200 & 0 \\ 1000 & 7000 & 15000 \\ 600 & 700 & 5000 \end{pmatrix} \begin{pmatrix} \frac{1}{1200} & 0 & 0 \\ 0 & \frac{1}{25000} & 0 \\ 0 & 0 & \frac{1}{9300} \end{pmatrix} = \begin{pmatrix} \frac{5}{12} & \frac{1}{125} & 0 \\ \frac{5}{6} & \frac{7}{25} & \frac{50}{31} \\ \frac{1}{2} & \frac{7}{250} & \frac{50}{93} \end{pmatrix}$$

Mit R:
A=X% ∗ %solve(diag(x))

c) Gesamtproduktion $\mathbf{x}'$ für den Outputvektor $\mathbf{y}' = \begin{pmatrix} 2650 \\ 10600 \\ 15900 \end{pmatrix}$

$$\mathbf{x}' = (\mathbf{E}-\mathbf{A})^{-1}\mathbf{y}' = \left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} \frac{5}{12} & \frac{1}{125} & 0 \\ \frac{5}{6} & \frac{7}{25} & \frac{50}{31} \\ \frac{1}{2} & \frac{7}{250} & \frac{50}{93} \end{pmatrix} \right)^{-1} \begin{pmatrix} 2650 \\ 10600 \\ 15900 \end{pmatrix} = \begin{pmatrix} 6360 \\ 132\,500 \\ 49\,290 \end{pmatrix}$$

mit R:
yp=c(2650,10600,15900)
xp=solve(diag(3)-A)%*%yp

d) neue Verflechtungsmatrix $\mathbf{X}'$ für den Outputvektor von c)

$$\mathbf{X}' = \mathbf{A} \cdot \mathbf{diag}\,(\mathbf{x}') = \begin{pmatrix} \frac{5}{12} & \frac{1}{125} & 0 \\ \frac{5}{6} & \frac{7}{25} & \frac{50}{31} \\ \frac{1}{2} & \frac{7}{250} & \frac{50}{93} \end{pmatrix} \begin{pmatrix} 6360 & 0 & 0 \\ 0 & 132500 & 0 \\ 0 & 0 & 49290 \end{pmatrix} = \begin{pmatrix} 2650 & 1060 & 0 \\ 5300 & 37\,100 & 79\,500 \\ 3180 & 3710 & 26\,500 \end{pmatrix}$$

mit R:
Xp=A%*%diag(as.vector(xp))

e) Rohstoffvektor für den Outputvektor von c)

$$\mathbf{r}' = \mathbf{R}\mathbf{x}' = \begin{pmatrix} 3 & 1 & 0 \\ 5 & 0.3 & 5 \\ 4 & 2 & 3 \end{pmatrix} \begin{pmatrix} 6360 \\ 132\,500 \\ 49\,290 \end{pmatrix} = \begin{pmatrix} 151580 \\ 318000 \\ 438310 \end{pmatrix}$$

Mit R: R%*%xp

f) Rohstoffvektor für den Outputvektor $(3500, 8000, 12000)$:

$$\mathbf{r}'' = \mathbf{R}\,(\mathbf{E}-\mathbf{A})^{-1}\mathbf{y}'' = \begin{pmatrix} 3 & 1 & 0 \\ 5 & 0.3 & 5 \\ 4 & 2 & 3 \end{pmatrix} \left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} \frac{5}{12} & \frac{1}{125} & 0 \\ \frac{5}{6} & \frac{7}{25} & \frac{50}{31} \\ \frac{1}{2} & \frac{7}{250} & \frac{50}{93} \end{pmatrix} \right)^{-1} \begin{pmatrix} 3500 \\ 8000 \\ 12000 \end{pmatrix}$$

$$= \begin{pmatrix} 133870.5 \\ 275178.9 \\ 375184.3 \end{pmatrix}$$

Mit R: ys=c(3500,8000,12000)
R%*%solve(diag(3)-A)%*%ys

g) Outputvektor für Rohstoffvektor $(500, 600, 1200)$:

$$\mathbf{y}=(\mathbf{E}-\mathbf{A})\,\mathbf{R}^{-1}\mathbf{r}=\left(\left(\begin{array}{ccc}1&0&0\\0&1&0\\0&0&1\end{array}\right)-\left(\begin{array}{ccc}\frac{5}{12}&\frac{1}{125}&0\\\frac{5}{6}&\frac{7}{25}&\frac{50}{31}\\\frac{1}{2}&\frac{7}{250}&\frac{50}{93}\end{array}\right)\right)\left(\begin{array}{ccc}3&1&0\\5&0.3&5\\4&2&3\end{array}\right)^{-1}\left(\begin{array}{c}500\\600\\1200\end{array}\right)$$
$$=\left(\begin{array}{c}5.532\,1\\188.62\\15.133\end{array}\right)$$

Mit R: rt=c(500,600,1200)
(diag(3)-A)%*%solve(R)%*%rt

h) Output für den Rohstoffvektor $(700, 3000, 4120)$

$$\left(\left(\begin{array}{ccc}1&0&0\\0&1&0\\0&0&1\end{array}\right)-\left(\begin{array}{ccc}\frac{5}{12}&\frac{1}{125}&0\\\frac{5}{6}&\frac{7}{25}&\frac{50}{31}\\\frac{1}{2}&\frac{7}{250}&\frac{50}{93}\end{array}\right)\right)\left(\begin{array}{ccc}3&1&0\\5&0.3&5\\4&2&3\end{array}\right)^{-1}\left(\begin{array}{c}700\\3000\\4120\end{array}\right)=\left(\begin{array}{c}-148.04\\-4.1598\\424.88\end{array}\right)$$

Ein negativer Output ist nicht möglich ist! Es ist nicht möglich, so zu produzieren, dass alle Rohstoffe aufgebraucht werden.
Mit R: r4=c(700,3000,4120)
(diag(3)-A)%*%solve(R)%*%r4

5. Mit R: X=matrix(c(40,30,50,20,60,30,40,40,0,30,50,60,70,90,10, 40,110,
120,25,3,4,5,13,43,15),ncol=5,byrow=T)
y=c(500,450,400,60,700)
x=X%*%rep(1,5)+y
A= X%*%solve(diag(as.vector(x)))
für a) A
für b) y2=c(59,65,73,200,30)
A%*%diag(y2)

1.4 Lineare Gleichungssysteme

Ein lineares Gleichungssystem hat die Form

$$\left\{\begin{array}{c}a_{11}x_1+a_{12}x_2+...+a_{1n-1}x_{n-1}+a_{1n}x_n=b_1\\\vdots\\a_{m1}x_1+a_{m2}x_2+...+a_{mn-1}x_{n-1}+a_{mn}x_n=b_m\end{array}\right.$$

Fassen wir die Variablen x_i, die Konstanten b_i in den Vektoren

$$\mathbf{x}:=(x_1,...,x_n)$$
$$\mathbf{b}:=(b_1,...,b_m)$$

und die Koeffizienten in der Matrix $\mathbf{A}$ zusammen, so kann man das Gleichungssystem kompakt wie folgt ausdrücken:

$$\mathbf{Ax}=\mathbf{b}$$

Die Lösungsmenge des Gleichungssystems ist dann die Menge aller Vektoren $\mathbf{x}$, so dass die Aussage „$\mathbf{Ax}=\mathbf{b}$“ wahr wird. Im Falle eines Gleichungssystems mit $n=m$ ist die Anzahl der Gleichungen mit der Anzahl der Unbekannten identisch und wir erhalten eine quadratische Matrix. Ist

diese invertierbar, haben wir das Problem des Auffindens der Lösungsmenge oben bereits gelöst: wir suchen das Argument einer bijektiven Abbildung, dessen Funktionswert unter der Abbildung bekannt ist, d.h. wir berechnen die Umkehrmatrix und multiplizieren diese mit **b:**

$$\mathbf{A}^{-1}\mathbf{A}\mathbf{x} = \mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

Wir erhalten das einzige Element $\mathbf{x}$ der Lösungsmenge. Allerdings ist diese Berechnungsmethode nur schneller als die Anwendung der vermutlich bekannten Methoden der Auflösung linearer Gleichungssysteme (Substitution und Gleichsetzen, Addition), wenn man mit Computern arbeitet. Die Berechnung einer (n, n)-Umkehrmatrix führt ja auf jeweils n Gleichungssysteme, welche jedes n Unbekannte hat - jedes gleich viele wie das ursprünglich zu lösende Gleichungssystem. Zudem haben wir mit dem Verfahren das Problem des Auffindens der Lösungsmenge nur für invertierbare Matrizen von Gleichungssystemen gelöst, die gleichviele Unbekannte wie Gleichungen aufweisen. Deshalb führen wir noch ein allgemeineres Lösungsverfahren für lineare Gleichungssysteme ein.
Offensichtlich gelten folgende Regeln für die Umwandlung von Gleichungssystemen, ohne das die Lösungsmenge verändert wird:

- Wird eine Gleichung eines Gleichungssystems durchgängig mit einer von Null verschiedenen reellen Konstanten multipliziert, verändert sich die Lösungsmenge nicht. So ist für

$$\begin{aligned} 2x_1 + 3x_2 &= 5 \\ 3x_1 + 15x_2 &= 7 \end{aligned} \tag{1.4.6}$$

  $L = \left\{(\frac{18}{7}, -\frac{1}{21})\right\}$, ebenso für das Gleichungssystems, in dem wir die ersten Zeile durchgehend mit 5 multipliziert haben:

$$\begin{aligned} 10x_1 + 15x_2 &= 25 \\ 3x_1 + 15x_2 &= 7 \end{aligned}$$

- Wird die linke und die rechte Seite einer der Gleichungen zur linken respektive rechten Seite einer anderen Gleichung hinzugezählt - wir sagen künftig kurz, dass „eine Gleichung zur anderen addiert wurde“ - verändert sich die Lösungsmenge nicht. So ist für

$$\begin{aligned} 2x_1 + 3x_2 &= 5 \\ 5x_1 + 18x_2 &= 12 \end{aligned}$$

  die Lösungsmenge wiederum $L = \left\{(\frac{18}{7}, -\frac{1}{21})\right\}$, wobei das Gleichungssystem aus dem System (1.4.6) durch die Addition der ersten zur zweiten Gleichung entstand, denn $2x_1+3x_2+3x_1+15x_2 = 5x_1 + 18x_2$ und $5 + 7 = 12$.
  Die erste und die zweite Regel zusammen ergeben die Möglichkeit, die addierte Gleichung z vorgängig mit einer von Null verschiedenen reellen Konstanten a zu multiplizieren, wobei im Gleichungssystem die Zeile z unverändert bleibt - man kann ja eine Zeile z mit einer Konstanten multiplizieren, diese dann zu einer anderen Zeile addieren und dann z wieder mit der Inversen $\frac{1}{a}$ der Konstanten multiplizieren. So kann man z.B. in (1.4.6) die erste Zeile mit 5 multiplizieren und diese zur zweiten dazuzählen, wobei die erste Zeile ins neue System unverändert übernommen wird, und man erhält:

$$\begin{aligned} 2x_1 + 3x_2 &= 5 \\ 13x_1 + 30x_2 &= 32 \end{aligned}$$

- Vertauschen wir zwei Zeilen eines Gleichungssystems, so verändert sich die Lösungsmenge nicht.

Jedem Gleichungssystem entspricht eine Matrix $(\mathbf{A}, \mathbf{b})$, wenn wir die Matrix $\mathbf{A}$ der Koeffizienten bilden und zu dieser die Spalte der Konstanten $\mathbf{b}$ hinzufügen. Die Matrix $(\mathbf{A}, \mathbf{b})$ nennen wir die „erweiterte Matrix“. Im Beispiel des Gleichungssystems (1.4.6) erhalten wir:

$$\left(\begin{array}{cc|c} 2 & 3 & 5 \\ 3 & 15 & 7 \end{array}\right)$$

Die Abtrennung des Vektors $\mathbf{b}$ von der Matrix $\mathbf{A}$ durch einen Trennstrich ist fakultativ. Gleichungssystem und erweiterte Matrix entsprechen sich eineindeutig. Offensichtlich entspricht die Multiplikation einer Zeile der erweiterten Matrix mit einer Konstanten ($\neq 0$) der obigen Operation der Multiplikation einer Zeile des Gleichungssystems mit einer Konstanten ($\neq 0$). Multiplizieren wir also die Zeile einer erweiterten Matrix mit einer von Null verschiedenen reellen Konstanten, verändert sich die Lösungsmenge des der Matrix entsprechenden Gleichungssystems nicht. Dies gilt ebenso für die Addition einer - eventuell mit einer von Null verschiedenen reellen Konstanten multiplizierten - Zeile zu einer anderen Zeile. Zudem können Zeilen der Matrix ausgetauscht werden, ohne dass sich die Lösungsmenge des entsprechenden Gleichungssystems ändert. Mit Hilfe dieser Operationen versuchen wir die erweiterte Matrix so umzuwandeln, das wir die Lösungsmenge unmittelbar aus der Matrix ablesen können. Wir führen dies nun ohne lange Erläuterungen an ein paar Beispielen vor, um die verschiedenen dabei möglichen Ergebnisse aufzuzeigen. Das Verfahren wird „Gausssches Eliminationsverfahren“ genannt (Johann Carl Friedrich Gauß, 1777 - 1855, deutscher Mathematiker).

Beispiel 1.4.1. *Wir lösen mit dem Eliminationsverfahren das Gleichungssystem (1.4.6). Wir multiplizieren die erste Zeile mit $\frac{1}{2}$ und erhalten:*

$$\left(\begin{array}{cc|c} 1 & \frac{3}{2} & \frac{5}{2} \\ 3 & 15 & 7 \end{array}\right)$$

Als nächstes addieren wir die erste mit -3 multiplizierte Zeile zur zweiten Zeile und erhalten (rechnen Sie das nach!):

$$\left(\begin{array}{cc|c} 1 & \frac{3}{2} & \frac{5}{2} \\ 0 & \frac{21}{2} & -\frac{1}{2} \end{array}\right)$$

Nun multiplizieren wir die zweite Zeile mit $\frac{2}{21}$ und erhalten:

$$\left(\begin{array}{cc|c} 1 & \frac{3}{2} & \frac{5}{2} \\ 0 & 1 & -\frac{1}{21} \end{array}\right) \tag{1.4.7}$$

Wir addieren die mit $-\frac{3}{2}$ multiplizierte zweite Zeile der Matrix (1.4.7) zur ersten und erhalten:

$$\left(\begin{array}{cc|c} 1 & 0 & \frac{18}{7} \\ 0 & 1 & -\frac{1}{21} \end{array}\right)$$

Diese Matrix entspricht dem Gleichungssytem

$$\begin{aligned} x_1 &= \frac{18}{7} \\ x_2 &= -\frac{1}{21}, \end{aligned}$$

Die Lösungsmenge $\left\{\left(\frac{18}{7}, -\frac{1}{21}\right)\right\}$ kann direkt abgelesen werden. ◊

Beispiel 1.4.2. *Gegeben sei das Gleichungssystem*

$$\begin{aligned} 15x_1 + 2x_2 + 35x_3 &= 5 \\ 28x_1 + 11x_2 + 12x_3 &= 12 \\ 1.5x_1 + 5x_2 &= 14 \end{aligned}$$

Künftig verwenden wir folgende Abkürzungen: Steht zwischen zwei Matrizen aZ_i, so wird die Zeile i der vorausgehenden Matrix mit a multipliziert und man erhält die nachfolgende Matrix. Steht $Z_i + (aZ_j)$, so wird zur $i-$ten Zeile der vorausgehenden Matrix das $a-$fache der $j-$ten Zeile hinzugezählt, um die nachfolgende Matrix zu erhalten. Steht $Z_i \leftrightarrow Z_j$ so werden die Zeilen i und j der vorausgehenden Matrix vertauscht und man erhält die nachfolgende Matrix. Wir gehen von der erweiterten Matrix aus:

$$\left(\begin{array}{ccc|c} 15 & 2 & 35 & 5 \\ 28 & 11 & 12 & 12 \\ 1.5 & 5 & 0 & 14 \end{array}\right) \xrightarrow{\frac{1}{15}Z_1} \left(\begin{array}{ccc|c} 1 & \frac{2}{15} & \frac{7}{3} & \frac{1}{3} \\ 28 & 11 & 12 & 12 \\ 1.5 & 5 & 0 & 14 \end{array}\right) \xrightarrow{Z_2 + (-28Z_1)}$$

$$\left(\begin{array}{ccc|c} 1 & \frac{2}{15} & \frac{7}{3} & \frac{1}{3} \\ 0 & \frac{109}{15} & -\frac{160}{3} & \frac{8}{3} \\ 1.5 & 5 & 0 & 14 \end{array}\right) \xrightarrow{Z_3 + (-1.5Z_1)} \left(\begin{array}{ccc|c} 1 & \frac{2}{15} & \frac{7}{3} & \frac{1}{3} \\ 0 & \frac{109}{15} & -\frac{160}{3} & \frac{8}{3} \\ 0 & 4.8 & -3.5 & 13.5 \end{array}\right)$$

$$\xrightarrow{\frac{15}{109}Z_2} \left(\begin{array}{ccc|c} 1 & \frac{2}{15} & \frac{7}{3} & \frac{1}{3} \\ 0 & 1 & -\frac{800}{109} & \frac{40}{109} \\ 0 & 4.8 & -3.5 & 13.5 \end{array}\right) \xrightarrow{Z_1 + \left(-\frac{2}{15}Z_2\right)} \left(\begin{array}{ccc|c} 1 & 0 & \frac{361}{109} & \frac{31}{109} \\ 0 & 1 & -\frac{800}{109} & \frac{40}{109} \\ 0 & 4.8 & -3.5 & 13.5 \end{array}\right)$$

$$\xrightarrow{Z_3 + (-4.8Z_2)} \left(\begin{array}{ccc|c} 1 & 0 & \frac{361}{109} & \frac{31}{109} \\ 0 & 1 & -\frac{800}{109} & \frac{40}{109} \\ 0 & 0 & 31.729 & 11.739 \end{array}\right) \xrightarrow{\frac{1}{31.729}Z_3} \left(\begin{array}{ccc|c} 1 & 0 & \frac{361}{109} & \frac{31}{109} \\ 0 & 1 & -\frac{800}{109} & \frac{40}{109} \\ 0 & 0 & 1 & 0.369\,98 \end{array}\right)$$

$$\xrightarrow{Z_1 + \left(-\frac{361}{109}Z_3\right)} \left(\begin{array}{ccc|c} 1 & 0 & 0 & -0.940\,94 \\ 0 & 1 & -\frac{800}{109} & \frac{40}{109} \\ 0 & 0 & 1 & 0.369\,98 \end{array}\right) \xrightarrow{Z_2 + \left(\frac{800}{109}Z_3\right)} \left(\begin{array}{ccc|c} 1 & 0 & 0 & -0.940\,94 \\ 0 & 1 & 0 & 3.\,082\,4 \\ 0 & 0 & 1 & 0.369\,98 \end{array}\right)$$

Damit erhält man die Lösungsmenge $\{(-0.940\,94, 3.0824, 0.36998)\}$ ◊

Die Transformationen kann man auch mittels Matrixmultiplikation erreichen - womit man diese mit Excel oder R durchführen kann, um Fehler zu vermeiden. Wir führen die folgenden drei Elementarmatrizen ein und erläutern deren Effekte jeweils mit einem Beispiel.
Typ I: Elementarmatrizen $\mathbf{E}_{\lambda i}$ vom Typ I erhält man durch Ersetzen der Komponente e_{ii} in der Einheitsmatrix $\mathbf{E}$ durch eine Konstante λ. $\mathbf{E}_{\lambda i}\mathbf{A}$ bewirkt die Multiplikation der Zeile i der Matrix $\mathbf{A}$ mit der Konstanten λ (im Beispiel wird die zweite Zeile mit 5 multipliziert).

$$\left(\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right) \left(\begin{array}{ccc|c} 2 & 6 & 8 & 5 \\ 3 & 9 & 12 & 12 \\ 5 & 2 & 2 & 14 \\ 4 & 12 & 11 & 14 \end{array}\right) = \left(\begin{array}{cccc} 2 & 6 & 8 & 5 \\ 15 & 45 & 60 & 60 \\ 5 & 2 & 2 & 14 \\ 4 & 12 & 11 & 14 \end{array}\right)$$

Typ II: Elementarmatrizen $\mathbf{E}_{i+\lambda j}$ vom Typ II erhält man durch Ersetzen der Komponente e_{ij} in der Einheitsmatrix $\mathbf{E}$ durch eine Konstante λ. $\mathbf{E}_{i+\lambda j}\mathbf{A}$ bewirkt die Addition einer mit λ multiplizierten Zeile j der Matrix $\mathbf{A}$ zur Zeile i $(i \neq j)$ der Matrix $\mathbf{A}$ (im Beispiel wird die 4. Zeile mit 3 multipliziert und zur 2. Zeile hinzugezählt):

$$\left(\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right) \left(\begin{array}{ccc|c} 2 & 6 & 8 & 5 \\ 3 & 9 & 12 & 12 \\ 5 & 2 & 2 & 14 \\ 4 & 12 & 11 & 14 \end{array}\right) = \left(\begin{array}{cccc} 2 & 6 & 8 & 5 \\ 15 & 45 & 45 & 54 \\ 5 & 2 & 2 & 14 \\ 4 & 12 & 11 & 14 \end{array}\right)$$

Typ III: Elementarmatrizen $\mathbf{E}_{ij}$ vom Typ III erhält man aus der Einheitsmatrix $\mathbf{E}$ durch Nullsetzen der Diagonalkomponenten e_{ii} und e_{jj} und durch Ersetzung der Komponenten e_{ij} und e_{ji} durch 1. $\mathbf{E}_{ij}\mathbf{A}$ vertauscht die Zeilen i und j von $\mathbf{A}$ (im folgenden Beispiel werden die zweite und die vierte Zeile vertauscht - Vertauschungsmatrizen sind immer symmetrisch). Diese Matrizen verwendet man etwa, wenn auf der Diagonalen eine Null auftaucht und man die entsprechende

Zeile durch eine tiefere Zeile ersetzen möchte, deren unter der Null liegende Komponente von Null verschieden ist.

$$\left(\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{array}\right)\left(\begin{array}{ccc|c} 2 & 6 & 8 & 5 \\ 3 & 9 & 12 & 12 \\ 5 & 2 & 2 & 14 \\ 4 & 12 & 11 & 14 \end{array}\right) = \left(\begin{array}{cccc} 2 & 6 & 8 & 5 \\ 4 & 12 & 11 & 14 \\ 5 & 2 & 2 & 14 \\ 3 & 9 & 12 & 12 \end{array}\right)$$

Wie das folgenden Beispiel zeigt, kann man die Operationen des Typs II gleichzeitig auf mehrere Zeilen anwenden, um ein Gleichungssystem effizient zu lösen.

Beispiel 1.4.3.

$$\left(\begin{array}{ccc} \frac{1}{15} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right)\left(\begin{array}{ccc|c} 15 & 2 & 35 & 5 \\ 28 & 11 & 12 & 12 \\ 1.5 & 5 & 0 & 14 \end{array}\right) = \left(\begin{array}{cccc} 1 & 0.133\,33 & 2.333\,3 & 0.333\,33 \\ 28 & 11 & 12 & 12 \\ 1.5 & 5 & 0 & 14 \end{array}\right)$$

$$\left(\begin{array}{ccc} 1 & 0 & 0 \\ -28 & 1 & 0 \\ -1.5 & 0 & 1 \end{array}\right)\left(\begin{array}{cccc} 1 & 0.133 & 2.33 & 0.333 \\ 28 & 11 & 12 & 12 \\ 1.5 & 5 & 0 & 14 \end{array}\right) = \left(\begin{array}{cccc} 1 & 0.133 & 2.333 & 0.333 \\ 0 & 7.266 & -53.332 & 2.666 \\ 0 & 4.8 & -3.500 & 13.5 \end{array}\right)$$

$$\left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & \frac{1}{7.2668} & 0 \\ 0 & 0 & 1 \end{array}\right)\left(\begin{array}{cccc} 1 & 0.133 & 2.33 & 0.333 \\ 0 & 7.266 & -53.33 & 2.66 \\ 0 & 4.8 & -3.50 & 13.5 \end{array}\right) = \left(\begin{array}{cccc} 1 & 0.133 & 2.33 & 0.33 \\ 0 & 1 & -7.339 & 0.366 \\ 0 & 4.8 & -3.5 & 13.5 \end{array}\right)$$

$$\left(\begin{array}{ccc} 1 & -0.133 & 0 \\ 0 & 1 & 0 \\ 0 & -4.8 & 1 \end{array}\right)\left(\begin{array}{cccc} 1 & 0.133 & 2.33 & 0.333 \\ 0 & 1 & -7.33 & 0.366 \\ 0 & 4.8 & -3.5 & 13.5 \end{array}\right) = \left(\begin{array}{cccc} 1.0 & 0 & 3.31 & 0.284 \\ 0 & 1.0 & -7.33 & 0.366\,98 \\ 0 & 0 & 31.72 & 11.738 \end{array}\right)$$

$$\left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{31.728} \end{array}\right)\left(\begin{array}{cccc} 1.0 & 0 & 3.31 & 0.284 \\ 0 & 1.0 & -7.33 & 0.366\,98 \\ 0 & 0 & 31.72 & 11.738 \end{array}\right) = \left(\begin{array}{cccc} 1.0 & 0 & 3.31 & 0.284 \\ 0 & 1.0 & -7.33 & 0.366\,98 \\ 0 & 0 & 1 & 11.738 \end{array}\right)$$

$$\left(\begin{array}{ccc} 1 & 0 & -3.311\,8 \\ 0 & 1 & 7.339\,1 \\ 0 & 0 & 1 \end{array}\right)\left(\begin{array}{cccc} 1.0 & 0 & 3.311\,8 & 0.284\,4 \\ 0 & 1.0 & -7.339\,1 & 0.366\,98 \\ 0 & 0 & 1.0 & 0.369\,96 \end{array}\right) = \left(\begin{array}{cccc} 1 & 0 & 0 & -0.940\,83 \\ 0 & 1 & 0 & 3.082\,2 \\ 0 & 0 & 1 & 0.369\,96 \end{array}\right)$$

Vollziehen Sie das Beispiel und die folgenden mittels Multiplikation mit Elementarmatrizen mit Excel nach. ◊

Beispiel 1.4.4.

$$\left(\begin{array}{ccc|c} 5 & 12 & 3 & 9 \\ 28 & 11 & 12 & 12 \\ 5.6 & 2.2 & 2.4 & 14 \end{array}\right) \xrightarrow{\frac{1}{5}Z_1} \left(\begin{array}{ccc|c} 1 & \frac{12}{5} & \frac{3}{5} & \frac{9}{5} \\ 28 & 11 & 12 & 12 \\ 5.6 & 2.2 & 2.4 & 14 \end{array}\right) \xrightarrow{Z_2+(-28Z_1)}$$

$$\left(\begin{array}{ccc|c} 1 & \frac{12}{5} & \frac{3}{5} & \frac{9}{5} \\ 0 & -\frac{281}{5} & -\frac{24}{5} & -\frac{192}{5} \\ 5.6 & 2.2 & 2.4 & 14 \end{array}\right) \xrightarrow{Z_3+(-5.6Z_1)} \left(\begin{array}{ccc|c} 1 & \frac{12}{5} & \frac{3}{5} & \frac{9}{5} \\ 0 & -\frac{281}{5} & -\frac{24}{5} & -\frac{192}{5} \\ 0 & -11.24 & -0.96 & 3.92 \end{array}\right)$$

$$\xrightarrow{-\frac{5}{281}Z_2} \left(\begin{array}{ccc|c} 1 & \frac{12}{5} & \frac{3}{5} & \frac{9}{5} \\ 0 & 1 & \frac{24}{281} & \frac{192}{281} \\ 0 & -11.24 & -0.96 & 3.92 \end{array}\right) \xrightarrow{Z_1+\left(-\frac{12}{5}Z_2\right)} \left(\begin{array}{ccc|c} 1 & 0 & \frac{111}{281} & \frac{45}{281} \\ 0 & 1 & \frac{24}{281} & \frac{192}{281} \\ 0 & -11.24 & -0.96 & 3.92 \end{array}\right)$$

$$\xrightarrow{Z_3+(11.24Z_2)} \left(\begin{array}{ccc|c} 1 & 0 & \frac{111}{281} & \frac{45}{281} \\ 0 & 1 & \frac{24}{281} & \frac{192}{281} \\ 0 & 0 & 0 & 11.6 \end{array}\right)$$

Die letzte Zeile der Matrix entspricht im zur Matrix gehörenden Gleichungssystem der Gleichung

$$0x_1 + 0x_2 + 0x_3 = 11.6 \Longrightarrow$$
$$0 = 11.6$$

Man erhält einen Widerspruch. Das Gleichungssystem ist nicht lösbar. ◊

Beispiel 1.4.5.

$$\left(\begin{array}{ccc|c} 2 & 6 & 8 & 5 \\ 3 & 9 & 12 & 12 \\ 5 & 2 & 2 & 14 \end{array}\right) \underrightarrow{\frac{1}{2}Z_1} \left(\begin{array}{ccc|c} 1 & 3 & 4 & \frac{5}{2} \\ 3 & 9 & 12 & 12 \\ 5 & 2 & 2 & 14 \end{array}\right) \underrightarrow{Z_2 + -3Z_1}$$

$$\left(\begin{array}{ccc|c} 1 & 3 & 4 & \frac{5}{2} \\ 0 & 0 & 0 & \frac{9}{2} \\ 5 & 2 & 2 & 14 \end{array}\right) \underrightarrow{Z_3 + -5Z_1} \left(\begin{array}{ccc|c} 1 & 3 & 4 & \frac{5}{2} \\ 0 & 0 & 0 & \frac{9}{2} \\ 0 & -13 & -18 & 1.5 \end{array}\right)$$

Da die Zeile 2 an zweiter Stelle eine 0 enthält, vertauschen wir Zeilen so, dass dort eine von 0 verschiedene Zahl steht. Die Zeile 1 fällt dabei als Kandidatin ausser Betracht:

$$\underrightarrow{Z_3 \leftrightarrow Z_2} \left(\begin{array}{ccc|c} 1 & 3 & 4 & \frac{5}{2} \\ 0 & -13 & -18 & 1.5 \\ 0 & 0 & 0 & \frac{9}{2} \end{array}\right) \underrightarrow{-\frac{1}{13}Z_2} \left(\begin{array}{ccc|c} 1 & 3 & 4 & \frac{5}{2} \\ 0 & 1 & \frac{18}{13} & -0.115\,38 \\ 0 & 0 & 0 & \frac{9}{2} \end{array}\right)$$

$$\underrightarrow{Z_1 + (-3Z_2)} \left(\begin{array}{ccc|c} 1 & 0 & -\frac{2}{13} & 2.846\,1 \\ 0 & 1 & \frac{18}{13} & -0.115\,38 \\ 0 & 0 & 0 & \frac{9}{2} \end{array}\right)$$

Auch dieses System ist nicht lösbar. Das hätte man schon oben sehen können, wo die zweite Zeile zuerst drei Nullen aufweist und dann eine von Null verschiedene Zahl b_2 auftaucht. Allerdings wollten wir ein mechanisches Verfahren vorführen, das sich noch als nützlich erweisen wird. ◊

Beispiel 1.4.6. *Gegeben sei das Gleichungssystem*

$$5x_1 + 3x_2 + 4x_3 = 15$$
$$2x_1 + 8x_2 + 7x_3 = 14$$

Die entsprechende erweiterte Matrix ist

$$\left(\begin{array}{ccc|c} 5 & 3 & 4 & 15 \\ 2 & 8 & 7 & 14 \end{array}\right)$$

Wir wandeln diese um:

$$\left(\begin{array}{ccc|c} 5 & 3 & 4 & 15 \\ 2 & 8 & 7 & 14 \end{array}\right) \underrightarrow{\frac{1}{5}Z_1} \left(\begin{array}{ccc|c} 1 & \frac{3}{5} & \frac{4}{5} & 3 \\ 2 & 8 & 7 & 14 \end{array}\right) \underrightarrow{Z_2 + -2Z_1}$$

$$\left(\begin{array}{ccc|c} 1 & \frac{3}{5} & \frac{4}{5} & 3 \\ 0 & \frac{34}{5} & \frac{27}{5} & 8 \end{array}\right) \underrightarrow{\frac{5}{34}Z_2} \left(\begin{array}{ccc|c} 1 & \frac{3}{5} & \frac{4}{5} & 3 \\ 0 & 1 & \frac{27}{34} & \frac{20}{17} \end{array}\right) \underrightarrow{Z_1 + -\frac{3}{5}Z_2}$$

$$\left(\begin{array}{ccc|c} 1 & 0 & \frac{11}{34} & \frac{39}{17} \\ 0 & 1 & \frac{27}{34} & \frac{20}{17} \end{array}\right)$$

Das Verfahren stoppt. Es ergibt sich weder ein Widerspruch noch eine eindeutige Lösung. In diesem Fall gibt es unendlich viele Lösungen. Fürs obige Beispiel erhält man das folgende der Matrix entsprechende Gleichungssystem

$$x_1 + \frac{11}{34}x_3 = \frac{39}{17}$$
$$x_2 + \frac{27}{34}x_3 = \frac{20}{17}$$

Eine der unendlich vielen konkreten Lösungen erhält man wie folgt. Wir setzen für x_i mit $i = 1$ anfangend solange Zahlen ein, bis ein eindeutig lösbares Gleichungssystem entsteht. Dieses lösen wir dann auf. Die willkürlich gewählten Werte zusammen mit den berechneten ergeben dann eine Lösung. Setzen wir z.B. $x_1 = 1$, gilt

$$1 + \frac{11}{34}x_3 = \frac{39}{17}$$

und damit $x_3 = 4$. Wir können nun x_2 ausrechnen:

$$x_2 + \frac{27}{34} \cdot 4 = \frac{20}{17}$$

Man erhält: $x_2 = -2$. Ein Element der Lösungsmenge ist damit $(1, -2, 4)$. Wir kontrollieren:

$$\begin{pmatrix} 5 & 3 & 4 \\ 2 & 8 & 7 \end{pmatrix} \begin{pmatrix} 1 \\ -2 \\ 4 \end{pmatrix} = \begin{pmatrix} 15 \\ 14 \end{pmatrix}$$

Allgemeiner können wir die Gleichungen

$$x_1 + \frac{11}{34}x_3 = \frac{39}{17}$$
$$x_2 + \frac{27}{34}x_3 = \frac{20}{17}$$

nach zwei der Variablen auflösen. Dabei wählt man mit Vorteil die Variablen, die bereits isoliert (konstantenfrei) sind, wodurch sich das oben entwickelte Verfahren der Zeilennormalform als nützlich erweist:

$$x_1 = \frac{39}{17} - \frac{11}{34}x_3$$
$$x_2 = \frac{20}{17} - \frac{27}{34}x_3$$

Damit ist der Lösungsvektor allgemein ausgedrückt:

$$\begin{pmatrix} \frac{39}{17} - \frac{11}{34}x_3 \\ \frac{20}{17} - \frac{27}{34}x_3 \\ x_3 \end{pmatrix}$$

und die Lösungsmenge

$$\left\{ \mathbf{x} : \mathbf{x} = \begin{pmatrix} \frac{39}{17} - \frac{11}{34}x_3 \\ \frac{20}{17} - \frac{27}{34}x_3 \\ x_3 \end{pmatrix} \text{ mit } x_3 \in \mathbb{R} \right\}$$

Man könnte auch nach anderen Variablen auflösen - die Lösungsmenge verändert sich nicht. $\Diamond$

Beispiel 1.4.7. *Gegeben sei das Gleichungssystem*

$$\begin{pmatrix} 5 & 3 & 2 & 4 \\ 8 & 2 & 1 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 20 \\ 44 \end{pmatrix}.$$

Die entsprechende erweiterte Matrix ist

$$\begin{pmatrix} 5 & 3 & 2 & 4 & 20 \\ 8 & 2 & 1 & 4 & 44 \end{pmatrix}.$$

Wir erhalten:

$$\begin{pmatrix} 5 & 3 & 2 & 4 & 20 \\ 8 & 2 & 1 & 4 & 44 \end{pmatrix} \underrightarrow{\frac{1}{5}Z_1} \begin{pmatrix} 1 & \frac{3}{5} & \frac{2}{5} & \frac{4}{5} & 4 \\ 8 & 2 & 1 & 4 & 44 \end{pmatrix} \underrightarrow{Z_2 + -8Z_1}$$

$$\begin{pmatrix} 1 & \frac{3}{5} & \frac{2}{5} & \frac{4}{5} & 4 \\ 0 & \frac{-14}{5} & -\frac{11}{5} & \frac{-12}{5} & 12 \end{pmatrix} \underrightarrow{\frac{-5}{14}Z_1} \begin{pmatrix} 1 & \frac{3}{5} & \frac{2}{5} & \frac{4}{5} & 4 \\ 0 & 1 & \frac{11}{14} & \frac{6}{7} & -\frac{30}{7} \end{pmatrix} \underrightarrow{Z_1 + -\frac{3}{5}Z_2}$$

$$\begin{pmatrix} 1 & 0 & -\frac{1}{14} & \frac{2}{7} & \frac{46}{7} \\ 0 & 1 & \frac{11}{14} & \frac{6}{7} & -\frac{30}{7} \end{pmatrix}$$

Das entspricht dem Gleichungssystem

$$x_1 - \frac{1}{14}x_3 + \frac{2}{7}x_4 = \frac{46}{7}$$
$$x_2 + \frac{11}{14}x_3 + \frac{6}{7}x_4 = -\frac{30}{7}$$

Wir können jeweils zwei Variablen willkürlich festlegen und erhalten ein lösbares Gleichungssystem. Allgemein erhält man also als Lösungsmenge mit

$$x_1 = \frac{46}{7} + \frac{1}{14}x_3 - \frac{2}{7}x_4$$
$$x_2 = -\frac{30}{7} - \frac{11}{14}x_3 - \frac{6}{7}x_4$$

$$L = \left\{ x \in \mathbb{R}^4 | \begin{pmatrix} \frac{46}{7} + \frac{1}{14}x_3 - \frac{2}{7}x_4 \\ -\frac{30}{7} - \frac{11}{14}x_3 - \frac{6}{7}x_4 \\ x_3 \\ x_4 \end{pmatrix} \right\}$$

So gilt etwa mit $x_3 = 8$ und $x_4 = 5$:

$$x_1 = \frac{46}{7} + \frac{1}{14} \cdot 8 - \frac{2}{7} \cdot 5 = \frac{40}{7}$$
$$x_2 = -\frac{30}{7} - \frac{11}{14} \cdot 8 - \frac{6}{7} \cdot 5 = -\frac{104}{7}$$

Und damit

$$\begin{pmatrix} 5 & 3 & 2 & 4 \\ 8 & 2 & 1 & 4 \end{pmatrix} \begin{pmatrix} \frac{40}{7} \\ -\frac{104}{7} \\ 8 \\ 5 \end{pmatrix} = \begin{pmatrix} 20 \\ 44 \end{pmatrix}$$

◊

Das Gausssche Eliminationsverfahren sollte durch die obigen Beispiele 1.4.1 ff. deutlich geworden sein.

Definition 1.4.8. *Eine Matrix* **B**, *die aus* **A** *mit dem Gaussschen Eliminationsverfahren gewonnen wurde, so*

- *dass* **B** *pro Zeile entweder mit einer* 1 *anfängt oder nur Nullen aufweist und*
- *in Zeilen, die mit 1 anfangen spaltenweise aus Einheitsvektoren besteht,*

wird „Matrix in Zeilennormalform“ von **A** *genannt.*

◊

Wir können die bisher gewonnen Erkenntnisse - ohne Beweis - wie folgt zusammenfassen:

Satz 1.4.9. *Sei* $(\mathbf{A}', \mathbf{b}')$ *die Zeilennormalform der erweiterten Matrix* $(\mathbf{A}, \mathbf{b})$, *wobei* $\mathbf{A}$ *eine* $(m, n)-$*Matrix sei:*

- *Besteht* $\mathbf{A}'$ *nur mehr aus paarweise verschiedenen Einheitsvektoren (*$\mathbf{A}'$ *ist also quadratisch), so weist das der Matrix entsprechende Gleichungssystem genau eine Lösung auf.*
- *Gibt es eine Zeile* i *von* $\mathbf{A}'$, *so dass* $a_{ij} = 0$ *für* $1 \leq j \leq n$ *und ist* $b_i \neq 0$, *dann hat das der Matrix* $(\mathbf{A}, \mathbf{b})$ *entsprechende Gleichungssystem keine Lösung.*
- *Tritt keiner der beiden obigen Fälle ein, gibt es unendlich viele Lösungen.*

Bemerkung 1.4.10. *Man kann beweisen, dass die Bedingung „*$\mathbf{A}'$ *besteht nur mehr aus paarweise verschiedenen Einheitsvektoren“ des ersten Teilsatzes von 1.4.9 mit der Invertierbarkeit von quadratischen Matrizen äquivalent ist. Damit haben wir ein Verfahren zu Hand, um die Invertierbarkeit zu prüfen. Bevor man Gleichungssysteme mit der Inversen löst, sollte man die Invertierbarkeit nämlich abklären. Beim obigen Nachweis der Invertierbarkeit haben wir das Gleichungssystem aber bereits gelöst. Allerdings ist es manchmal nützlich, für oft wiederholte Rechnungen die Inverse zu haben. Entsprechend kann die Verwendung von Inversen für die Berechnung von Gleichungssystemen trotzdem nützlich sein, solange die Matrix der Koeffizienten unverändert bleibt.*

Wollen wir für die einmalige Lösung einer vermutlich eindeutig lösbaren Gleichung das Lösungsverfahren mit der Inversen (etwa mit Excel) verwenden, gibt es weitere Verfahren, um die Invertierbarkeit quadratischer Matrizen zu überprüfen. Diese sind allerdings von Hand mindestens so rechenintensiv wie das eben eingeführte. Mit Excel kann man die Invertierbarkeit mit einem der alternativen Verfahren leicht überprüfen. Man gibt für eine quadratische Matrix ein: „mdet()“ - $\det \mathbf{A}$ *ist eine Funktion, die Matrizen reelle Zahlen zuordnet und die genau dann von Null verschiedene Werte annimmt, wenn die Matrix invertierbar ist. Die zugeordneten Zahlen werden „Determinanten“ genannt. Allerdings ist zu beachten, dass bei Resultaten sehr nahe bei 0, z.B.* $3.4639E-14$, *die Matrix vermutlich nicht invertierbar ist. Es ist anzunehmen, dass eine solche Abweichung von* 0 *durch Rundungsfehler entsteht. Will man dies ausschliessen, hilft nur eine Rechnung, die auf Rundung verzichtet. Dazu ist das obige Verfahren - wenigstens bei kleinen Matrizen - recht geeignet, da man bei einem Ausgangsgleichungssystem, das nur rationale Zahlen aufweist, durchgehend mit Brüchen rechnen kann.*
◊

Definition 1.4.11. *Weist die Zeilennormalform einer quadratischen* $n \times n$*-Matrix in jeder Zeile von links her eine* 1 *auf, so sagt man, dass die ursprüngliche Matrix den Rang* n *hat. Der Rang einer Matrix besteht allgemeiner aus der Anzahl der mit* 1 *anfangenden Zeilen der Zeilennormalform der Matrix.* ◊

Quadratische (n, n)-Matrizen $\mathbf{M}$ mit dem Rang n legen entsprechend bijektive Abbildungen $\mathbb{R}^n \longrightarrow \mathbb{R}^n$ fest, die sie unter anderem mittels $\mathbf{Mx}$ für $\mathbf{x} \in \mathbb{R}^n$ definieren. $\mathbf{M}^{-1}$ legt die entsprechende Umkehrfunktion fest.
Ohne deren Bedeutung für lineare Abbildungen zu erläutern, führen wir noch die Transposition von Matrizen ein, da sich diese oft als nützlich erweist.

Definition 1.4.12. *Seien* (i, j, x) *die Elemente des Graphen einer Matrix* $\mathbf{A}$ *(s. Bemerkung 1.1.13), dann enthält der Graph der Transponierten der Matrix* $\mathbf{A}$ *die Elemente* (j, i, x). *Für die Transponierte von* $\mathbf{A}$ *schreibt man* $\mathbf{A}^T$. ◊

In der Literatur findet man statt $\mathbf{A}^T$ unter anderem auch $\mathbf{A}'$, $\mathbf{A}^t$ oder ${}^t\mathbf{A}$. Aus einer $(m, n)-$Matrix wird eine (n, m) $-$Matrix. Die Komponente der i-ten Zeile und j-ten Spalte wird in die j-te Zeile und die i-te Spalte gesetzt. Bei einer (n, n) $-$Matrix erhält man die Transponierte, indem man die Matrix an der Hauptdiagonalen spiegelt. Wenn die (m, n)-Matrix $\mathbf{A}$ eine Abbildung von $\mathbb{R}^n$ nach $\mathbb{R}^m$, so definiert die transponierte Matrix unter anderem eine Abbildung von $\mathbb{R}^m$ nach

$\mathbb{R}^n$, wobei die entsprechende Abbildung i.A. keineswegs die Umkehrfunktion ist. Diese Abbildung ist durch die Matrix $\mathbf{A}$ aber eindeutig bestimmt.
Mit Excel erhält man die Transponierte einer $(m,n)-$Matrix, indem man $n \times m$ freie Zellen beleuchtet und dann „=mtrans(Referenz auf zu transponierende Matrix)“ fortfährt oder durch Kopieren der Matrix, Anwählen einer Zelle mit genügend freiem Raum rechts sowie unten, dann Inhalte einfügen und transponieren.
Mit R: $t(\mathbf{A})$

Beispiel 1.4.13.

$$\begin{pmatrix} 5 & 3 & 4 \\ 15 & 1 & 2 \\ 16 & 6 & 18 \end{pmatrix}^T = \begin{pmatrix} 5 & 15 & 16 \\ 3 & 1 & 6 \\ 4 & 2 & 18 \end{pmatrix}$$

$\Diamond$

Beispiel 1.4.14.

$$\begin{pmatrix} 5 & 3 & 4 \\ 15 & 1 & 2 \\ 16 & 6 & 18 \\ 14 & 22 & 100 \end{pmatrix}^T = \begin{pmatrix} 5 & 15 & 16 & 14 \\ 3 & 1 & 6 & 22 \\ 4 & 2 & 18 & 100 \end{pmatrix}$$

$\Diamond$

Es folgen ein paar nützlich Rechenregeln für Matrizen.
Offensichtlich gilt:

Satz 1.4.15. $(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$

Satz 1.4.16. $\left(\mathbf{A}^T\right)^T = \mathbf{A}$

Weniger offensichtlich sind:

Satz 1.4.17. $\mathbf{A}^T\mathbf{B}^T = (\mathbf{BA})^T$

Beweis. Für $\mathbf{C} = \mathbf{BA}$ gilt: $c_{ij} = \sum_{l=1}^{k} b_{il}a_{lj}$. Ist a_{lj} die Eintrag in die entsprechende Zelle von $\mathbf{A}$, ist a'_{jl} der Eintrag in der j-ten Zeile und der l-ten Spalte von $\mathbf{A}^T$, wobei gilt $a_{lj} = a'_{jl}$. Analog für $\mathbf{B}$ und $\mathbf{C}$. Damit gilt:

$$c'_{ji} = c_{ij} = \sum_{l=1}^{k} b_{il}a_{lj} = \sum_{l=1}^{k} a_{lj}b_{il} = \sum_{l=1}^{k} a'_{jl}b'_{li}$$

Also gilt der Satz. □

Satz 1.4.18. $\left(\mathbf{A}^T\right)^{-1} = \left(\mathbf{A}^{-1}\right)^T$

Beweis. Es gilt $\mathbf{E} = \mathbf{E}^T$ und damit $\mathbf{E} = \mathbf{A}\mathbf{A}^{-1} = \left(\mathbf{A}\mathbf{A}^{-1}\right)^T = \left(\mathbf{A}^{-1}\right)^T \mathbf{A}^T$. Durch Multiplikation durch $\left(\mathbf{A}^T\right)^{-1}$ von rechts her, erhält man $\left(\mathbf{A}^T\right)^{-1} = \left(\mathbf{A}^{-1}\right)^T$ □

Satz 1.4.19. $(\mathbf{BA})^{-1} = \mathbf{A}^{-1}\mathbf{B}^{-1}$

Beweis. $(\mathbf{BA})(\mathbf{BA})^{-1} = \mathbf{E} \iff \mathbf{A}(\mathbf{BA})^{-1} = \mathbf{B}^{-1} \iff (\mathbf{BA})^{-1} = \mathbf{A}^{-1}\mathbf{B}^{-1}$ (durch jeweilige Multiplikation von links her). □

Satz 1.4.20. *Für* $\mathbf{b}, \mathbf{a} \in \mathbb{R}^n$ *gilt*

$$\mathbf{diag}(\mathbf{a})\,\mathbf{b} = \begin{pmatrix} a_1b_1 \\ \vdots \\ a_nb_n \end{pmatrix}$$

Mit $\mathbf{diag}\,(\mathbf{x})\,\mathbf{p}$ kann man z.B. einen Kostenvektor $\mathbf{k}$ berechnen, wenn der Vektor der Gütermenge (Leistungen) $\mathbf{x}$ und der Vektor der Preise $\mathbf{p}$ dieser Güter bekannt ist. Es gilt nämlich $\mathbf{k} = \mathbf{diag}\,(\mathbf{x})\,\mathbf{p}$, wie etwa das folgende Beispiel zeigt: $\mathbf{p} = (p_1, p_2, p_3)\,; \mathbf{x} = (300, 400, 50)$

$$\mathbf{diag}\,(\mathbf{x})\,\mathbf{p} = \begin{pmatrix} 300 & 0 & 0 \\ 0 & 400 & 0 \\ 0 & 0 & 50 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = \begin{pmatrix} 300p_1 \\ 400p_2 \\ 50p_3 \end{pmatrix}$$

Beispiel 1.4.21. *Innerbetriebliche Leistungsverrechnung. In Betrieben gibt es oft Unternehmensbereiche, welche für die übrigen Betriebseinheiten und sich selbst Leistungen erbringen. Will man überprüfen, ob z.B. diese Leistungen im Betrieb billiger erbracht werden als von auswärtigen Produzenten, muss man die innerbetrieblichen Verrechnungskosten berechnen. Zudem braucht man bei der Produktion verschiedener Güter und Dienstleistungen für die Preiskalkulation die genauen internen Preise. Dazu kann man nicht einfach die sogenannten Primärkosten (Löhne, Material - also Input von ausserhalb des Verflechtungssystems, Abschreibungen, etc.) durch die an die Hauptproduktion gelieferte Anzahl Produkte dividieren, denn in die Zwischenprodukte sind unterschiedlich viele Leistungen der verschiedenen Betriebseinheiten geflossen - deren Kosten „Sekundärkosten" genannt werden. Um die effektiven Kosten zu berechnen müsste man entsprechend bereits die Kosten der Leistungen, welche die Unterbetriebseinheiten anderen Unterbetriebseinheiten gegenüber erbringen, kennen. Das Problem führt auf ein Gleichungssystem, wie wir am folgenden Beispiel zeigen wollen:*

Hilfsbetriebe	H	F	P	D	Leistungen an Hauptbetrieb ($= \mathbf{l}_{HB}$)	Gesamtleistung der Hilfsbetriebe ($= \mathbf{l}_g$)	Primärkosten GE
Heizung (H)	0	500	0	600	1000	2100	5123
Finanzen (F)	100	1000	100	5000	20000	26200	4006
Putzdienst (P)	1000	1200	0	2000	8000	12200	8015
Druckerei (D)	0	2000	0	500	12000	14500	12212
Summe							29356

Der Hilfsbetrieb Heizung liefert z.B. an die Druckerei 600 Einheiten Heizdienstleistungen (z.B. in kwh). Die Druckerei liefert an den Hauptbetrieb 12000 Einheiten Druckdienstleistungen. Offenbar muss für die Produktion eines Hilfsbetriebs gelten:

$$\textit{Gesamtleistung mal effektive Preise } (\mathbf{k}_g) = \textit{Primärkosten } (\mathbf{k}_p) + \textit{Sekundärkosten } (\mathbf{k}_s) \tag{1.4.8}$$

$$\mathbf{k}_g = \mathbf{k}_p + \mathbf{k}_s \tag{1.4.9}$$

wobei die sekundären Kosten k_{si} die von den Hilfsbetrieben i empfangenen Leistungen sind, die mit dem (unbekannten) Verrechnungspreis p_i dieser Leistungen multipliziert werden.
Die Gesamtleistung mal effektive Preise pro Hilfsbetreib k_{gi} sind die Gesamtleistung der Hilfsbetriebe i, die mit dem (unbekannten) Verrechnungspreis p_i dieser Leistungen multipliziert werden.

Der Zusammenhang (1.4.8) führt auf das folgende Gleichungssystem mit den gesuchten Verrechnungspreisen p_i der Leistungen von Hilfsbetrieb i pro Zeile:

$$\begin{aligned} k_p + (k_{s1} + k_{s2} + k_{s3} + k_{s4}) &= k_g \\ 5123 + 0p_1 + 100p_2 + 1000p_3 + 0p_4 &= 2100p_1 \\ 4006 + 500p_1 + 1000p_2 + 1200p_3 + 2000p_4 &= 26200p_2 \\ 8015 + 0p_1 + 100p_2 + 0p_3 + 0p_4 &= 12200p_3 \\ 12212 + 600p_1 + 5000p_2 + 2000p_3 + 500p_4 &= 14500p_4 \end{aligned}$$

Dies kann umgeformt werden in:

$$\begin{aligned} 2100p_1 - 100p_2 - 1000p_3 &= 5123 \\ -500p_1 + 25200p_2 - 1200p_3 - 2000p_4 &= 4006 \\ -100p_2 + 12200p_3 &= 8015 \\ -600p_1 - 5000p_2 - 2000p_3 + 14000p_4 &= 12212 \end{aligned}$$

Mit Excel gelangen aus der ursprünglichen Tabelle der Leistungen der Hilfsbetriebe an die Hilfsbetriebe wie folgt möglichst effizient zur Matrixdarstellung dieses Gleichungssystems: wir berechnen die Transponierte der Matrix. Wir erhalten damit

$$\mathbf{L}^T := \begin{pmatrix} 0 & 500 & 0 & 600 \\ 100 & 1000 & 100 & 5000 \\ 1000 & 1200 & 0 & 2000 \\ 0 & 2000 & 0 & 500 \end{pmatrix}^T = \begin{pmatrix} 0 & 100 & 1000 & 0 \\ 500 & 1000 & 1200 & 2000 \\ 0 & 100 & 0 & 0 \\ 600 & 5000 & 2000 & 500 \end{pmatrix}$$

Die linke Seite des obigen Gleichungssystems in Matrixform erhalten wir dann mit

$$\mathbf{l}_g = \mathbf{L1} + \mathbf{l}_{HB} = \begin{pmatrix} 0 & 500 & 0 & 600 \\ 100 & 1000 & 100 & 5000 \\ 1000 & 1200 & 0 & 2000 \\ 0 & 2000 & 0 & 500 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1000 \\ 20000 \\ 8000 \\ 12000 \end{pmatrix} = \begin{pmatrix} 2100 \\ 26\,200 \\ 12\,200 \\ 14\,500 \end{pmatrix}$$

und

$$\begin{aligned} diag\,(\mathbf{l}_g) - \mathbf{L}^T &= \begin{pmatrix} 2100 & 0 & 0 & 0 \\ 0 & 26200 & 0 & 0 \\ 0 & 0 & 12200 & 0 \\ 0 & 0 & 0 & 14500 \end{pmatrix} - \begin{pmatrix} 0 & 100 & 1000 & 0 \\ 500 & 1000 & 1200 & 2000 \\ 0 & 100 & 0 & 0 \\ 600 & 5000 & 2000 & 500 \end{pmatrix} \\ &= \begin{pmatrix} 2100 & -100 & -1000 & 0 \\ -500 & 25\,200 & -1200 & -2000 \\ 0 & -100 & 12\,200 & 0 \\ -600 & -5000 & -2000 & 14\,000 \end{pmatrix} =: \mathbf{A} \end{aligned}$$

Mit R:
L=matrix(c(0,100,1000,0,500,1000,1200,2000,0,100,0,0,600,5000,2000,500),ncol=4)
lHB=c(1000,20000,8000,12000)
lg=rowSums(L)+lHB #oder L%%c(1,1,1,1)+lHB oder L%*%rep(1,4)+lHB*
A=diag(lg)-t(L)

Mit Excel kontrollieren wir, ob $\mathbf{A}$ *invertierbar ist:* $\det(\mathbf{A}) = 8.76734E + 15 > 0$. *Die Matrix ist invertierbar. Wir berechnen die Inverse:*
Mit R:
det(A)

Wir berechnen die Inverse:

$$\mathbf{A}^{-1} = \begin{pmatrix} 0.000476778 & 3.93369E-05 & 2.10782E-06 & 3.01117E-07 \\ 1.14105E-05 & 4.09109E-05 & 5.91741E-06 & 5.84442E-06 \\ 9.35289E-08 & 3.35335E-07 & 8.20157E-05 & 4.7905E-08 \\ 2.45219E-05 & 1.47493E-05 & 1.55158E-05 & 7.35356E-05 \end{pmatrix}$$

Mit $\mathbf{A}^{-1}\mathbf{k}_p = \mathbf{p}$ *mit* $\mathbf{k}_p$ *für die Primärkosten erhält man den Vektor* $\mathbf{p}$ *der Verrechnungspreise*

$$\mathbf{A}^{-1}\mathbf{k}_p = \mathbf{A}^{-1}\begin{pmatrix} 5123 \\ 4006 \\ 8015 \\ 12212 \end{pmatrix} = \begin{pmatrix} 2.769941912 \\ 0.341145286 \\ 0.659763486 \\ 1.207087039 \end{pmatrix} = \mathbf{p}$$

Mit R:
kp=c(5123,4006,8015,12212)
solve(A)%%kp*

Insgesamt geht man also wie folgt vor: Sei $\mathbf{L}$ *die Leistungsmatrix (Leistungen der Teilbetriebe an die Teilbetriebe),* $\mathbf{l}_g = \mathbf{L1} + \mathbf{l}_{HB}$ *der Vektor der Gesamtleistungen und* $\mathbf{k}_p$ *der Vektor der Primärkosten. Gesucht sei der Vektor der effektiven Preise* $\mathbf{p}$*: Man rechnet:*

$$\mathbf{p} = \left(\mathbf{diag}\left(\mathbf{l}_g\right) - \mathbf{L}^T\right)^{-1}\mathbf{k}_p$$

Mit R alles in einem Zug:
L=matrix(c(0,100,1000,0,500,1000,1200,2000,0,100,0,0,600,5000,2000,500),ncol=4)
lHB=c(1000,20000,8000,12000)
lg=rowSums(L)+lHB
solve(diag(lg)-t(L))%%kp*

Bemerkungen*: (1) Summiert man die Kosten für die Lieferungen an den Hauptbetrieb pro Teilbetrieb, so erhält man:*

$$\sum_{i=1}^{4}\mathbf{k}_{g_i} = \mathbf{l}_{HB}^T\mathbf{p} = \begin{pmatrix} 1000 & 20000 & 8000 & 12000 \end{pmatrix}\begin{pmatrix} 2.769941912 \\ 0.341145286 \\ 0.659763486 \\ 1.207087039 \end{pmatrix} = 29356.$$

Dies ist identisch mit der Summe der Primärkosten, d.h. $\left(\mathbf{l}_{HB}^T\mathbf{p}\right)\mathbf{1} = \mathbf{k}_p^T\mathbf{1} = \sum_{i=1}^{4}\mathbf{k}_{p_i}$*. Mit den korrekten Preisen für die Lieferung an den Hauptbetrieb deckt man also genau die Primärkosten.*

(2) Vergleicht man den Vektor der Verrechnungspreise mit den falschen Preisen, die aus der Division der Primärkosten durch die Leistung der Teilbetriebe an den Hauptbetrieb entstehen:

$$\left(\mathbf{diag}\left(\mathbf{l}_{HB}\right)\right)^{-1}\mathbf{k}_p = \begin{pmatrix} \frac{5123}{1000} \\ \frac{4006}{20000} \\ \frac{8015}{8000} \\ \frac{12212}{12000} \end{pmatrix} = \begin{pmatrix} 5.123 \\ 0.2003 \\ 1.0019 \\ 1.0177 \end{pmatrix}$$

sieht man, dass für manche Leistungen zuviel, für manche zu wenig verrechnet würde, was die Preiskalkulation der Endprodukte verfälschen würde, da in deren Produktion im Allgemeinen unterschiedlich viele Zwischenleistungen fliessen!

(3) Die Sekundärkosten $\mathbf{k}_s$ *kann man berechnen durch* $\mathbf{L}^T\mathbf{p}$ *Im Beispiel:*

$$\mathbf{k}_s = \mathbf{L}^T\mathbf{p} = \begin{pmatrix} 0 & 500 & 0 & 600 \\ 100 & 1000 & 100 & 5000 \\ 1000 & 1200 & 0 & 2000 \\ 0 & 2000 & 0 & 500 \end{pmatrix}^T\begin{pmatrix} 2.769941912 \\ 0.341145286 \\ 0.659763486 \\ 1.207087039 \end{pmatrix} = \begin{pmatrix} 693.88 \\ 4932.0 \\ 34.115 \\ 5290.8 \end{pmatrix}$$

◊

Wir behandeln nun die Grundideen des Beispiels 1.4.21 etwas allgemeiner.

Definition 1.4.22. *1.* $\mathbf{L} : (n,n) -$*Matrix der Leistungen der Teilbetriebe i an Teilbetriebe j - pro Zeile i stehen die Leistungen des Teilbetriebes i an die Teilbetriebe* $1,...,n$*.*

2. $\mathbf{l}_g, \mathbf{l}_h, \mathbf{l}_t \in \mathbb{R}^n : \mathbf{l}_g$ *Vektor der Gesamtleistungen der Teilbetriebe i an an die anderen Teilbetriebe und an den Hauptbetrieb,* $\mathbf{l}_h$ *Vektor der Leistungen der Teilbetriebe an den Hauptbetrieb;* $\mathbf{l}_t$ *Vektor der Gesamtleistungen der Teilbetriebe an die anderen Teilbetriebe, d.h.* $\mathbf{l}_t = \mathbf{L1}$*).*

3. $\mathbf{p} \in \mathbb{R}^n$*: Vektor der Verrechnungspreise (= effektive Preise) der von den Teilbetrieben i produzierten Güter.*

4. $\mathbf{k}_g, \mathbf{k}_s, \mathbf{k}_p \in \mathbb{R}^n : \mathbf{k}_g$ *Vektor der Gesamtkosten pro Teilbetrieb;* $\mathbf{k}_s$ *Vektor der Sekundärkosten (Kosten, die für die Teilbetriebe anfallen);* $\mathbf{k}_p$ *Vektor der Primärkosten pro Teilbetrieb).*

◊

Es gilt:

Satz 1.4.23. *Mit den Bezeichnungen der Definition 1.4.22 und mit*

$$\begin{aligned}\mathbf{k}_g &:= \mathbf{k}_s + \mathbf{k}_p\\ \mathbf{k}_g &:= \mathbf{diag}\,(\mathbf{l}_g)\,\mathbf{p}\\ \mathbf{k}_s &:= \left(\mathbf{L}^T\right)\mathbf{p}\end{aligned}$$

gilt:

$$\mathbf{p} = \left[\mathbf{diag}\,(\mathbf{l}_g) - \left(\mathbf{L}^T\right)\right]^{-1}\mathbf{k}_p$$

Beweis. Mit den Voraussetzungen gilt

$$\mathbf{diag}\,(\mathbf{l}_g)\,\mathbf{p} = \left(\mathbf{L}^T\right)\mathbf{p} + \mathbf{k}_p$$

sowie

$$\begin{aligned}-\mathbf{k}_p &= \left(\mathbf{L}^T\right)\mathbf{p} - \mathbf{diag}\,(\mathbf{l}_g)\,\mathbf{p}\\ &= \left[\left(\mathbf{L}^T\right) - \mathbf{diag}\,(\mathbf{l}_g)\right]\mathbf{p}\end{aligned}$$

oder auch

$$\left[\mathbf{diag}\,(\mathbf{l}_g) - \left(\mathbf{L}^T\right)\right]\mathbf{p} = \mathbf{k}_p \tag{1.4.10}$$

Dieses Gleichungssystem löst man auf durch

$$\mathbf{p} = \left[\mathbf{diag}\,(\mathbf{l}_g) - \left(\mathbf{L}^T\right)\right]^{-1}\mathbf{k}_p$$

□

Damit ist das allgemeine Lösungsverfahren für die Berechnung der effektiven Kosten beschrieben. Liefert ein Teilbetrieb verschiedene Produkte an andere Teilbetriebe, so behandelt man jedes der Produkte als von einer eigenen Teilunternehmung hergestellt.
Im Beispiel galt: Die Summe der Leistungen an Hauptbetrieb multipliziert mit ihren Preisen $\left(\mathbf{l}_h^T\mathbf{p}\right)$ ergibt die Summe der Primärkosten ($\sum_{i=1}^{n} k_{p_i} = \mathbf{1}^T\mathbf{k}_p$ mit $\mathbf{k}_p = (k_{p_1},...,k_{p_n})$). Allgemein gilt:

Satz 1.4.24. *Wenn die Voraussetzungen des Satzes 1.4.23 gelten, gilt mit den Bezeichnungen der Definition 1.4.22:*

$$\mathbf{l}_h^T\mathbf{p} = \mathbf{1}^T\mathbf{k}_p$$

Beweis. Wir nehmen an, es gilt:

$$\mathbf{l}_h^T \mathbf{p} = \mathbf{1}^T \mathbf{k}_p \tag{1.4.11}$$

Mit obiger Gleichung (1.4.10) gilt dies genau dann, wenn

$$\mathbf{l}_h^T \mathbf{p} = \mathbf{1}^T \left[\mathbf{diag}\left(\mathbf{l}_g\right) - \left(\mathbf{L}^T\right) \right] \mathbf{p}$$

genau dann, wenn

$$\begin{aligned} \mathbf{l}_h^T &= \mathbf{1}^T \left[\mathbf{diag}\left(\mathbf{l}_g\right) - \left(\mathbf{L}^T\right) \right] \\ &= \mathbf{1}^T \mathbf{diag}\left(\mathbf{l}_g\right) - \mathbf{1}^T \left(\mathbf{L}^T\right). \end{aligned} \tag{1.4.12}$$

Nun ist $\mathbf{1}^T \left(\mathbf{L}^T\right) = (\mathbf{L1})^T$, wobei $\mathbf{L1} = \mathbf{l}_t$ ist. Zudem ist $\mathbf{1}^T \mathbf{diag}\left(\mathbf{l}_g\right)$ die als Zeilenmatrix angeordneten Gesamtleistungen der Teilbereiche i, d.h. $\mathbf{1}^T \mathbf{diag}\left(\mathbf{l}_g\right) = \mathbf{l}_g^T$. Nun gilt $\mathbf{l}_g = \mathbf{l}_h + \mathbf{l}_t$ und damit $\mathbf{l}_h = \mathbf{l}_h - \mathbf{l}_t$. Zudem gilt $\mathbf{l}_h = \mathbf{l}_h - \mathbf{l}_t$ genau dann, wenn $\mathbf{l}_h^T = \left(\mathbf{l}_h - \mathbf{l}_t\right)^T = \mathbf{l}_h^T - \mathbf{l}_t^T$. Damit gilt (1.4.12) und entsprechend (1.4.11). □

1.4.1 Übungen

1. Lösen Sie die folgenden Gleichungssysteme von Hand (Gausssche Eliminationsmethode) und mit Excel
   a)

$$\begin{aligned} 2x_1 + 3x_2 &= 5 \\ 14x_1 - 2x_2 &= 18 \end{aligned}$$

   b)

$$\begin{aligned} 2x_1 + 3x_2 + 4x_3 &= 5 \\ 14x_1 - 2x_2 + 2x_3 &= 18 \\ 3x_1 + 12x_2 + 3x_3 &= -2 \end{aligned}$$

2. Lösen Sie das folgende Gleichungssystem von Hand (Gausssche Eliminationsmethode) und mit Excel

$$\begin{aligned} -3x_1 + 0.3x_2 - 1.2x_3 &= -5.4 \\ 50x_1 - 5x_2 + 20x_3 &= 90 \\ 10x_1 - x_2 + 4x_3 &= 18 \end{aligned}$$

3. Lösen Sie das folgende Gleichungssystem von Hand (Gausssche Eliminationsmethode) und mit Excel

$$\begin{aligned} 5x_1 + 3x_2 + 4x_3 + 15x_4 &= 16 \\ x_1 + 0.6x_2 + 0.8x_3 + 3x_4 &= 6 \\ 20x_1 + x_2 + 3x_3 + 6x_4 &= 5 \end{aligned}$$

4. Lösen Sie das Gleichungssystem der Übung **??** auf der Seite **??** mit Excel.

5. In einer Firma gebe es 5 Unterbetriebe UB_i, die für diese selber und an die Endproduktion (EP) folgende Leistungen erbringen (UB_2 erbringt z.B. 600 Einheiten Leistung an UB_5).

	UB_1	UB_2	UB_3	UB_4	UB_5	EP	Primärkosten
UB_1	50	200	300	100	0	20000	5000
UB_2	150	350	0	700	600	15000	4500
UB_3	300	7500	800	0	300	22000	12000
UB_4	1300	4900	80	700	140	10000	2000
UB_5	400	550	7630	380	50	12580	4000

Berechnen Sie
a) die Gesamtleistung der Unterbetriebe
b) die korrekten internen Verrechnungskosten.

1.4.2 Lösungen

1. a) Wir erhalten die erweiterte Matrix

$$\begin{pmatrix} 2 & 3 & 5 \\ 14 & -2 & 18 \end{pmatrix}$$

und daraus mit Zeilenumformungen:

$$\underset{\longrightarrow}{\frac{1}{2}z_1}\begin{pmatrix} 1 & 1.5 & 2.5 \\ 14 & -2 & 18 \end{pmatrix} \xrightarrow{z_2+(-14z_1)} \begin{pmatrix} 1 & 1.5 & 2.5 \\ 0 & -23 & -17 \end{pmatrix}$$

$$\underset{\longrightarrow}{-\frac{1}{23}z_2}\begin{pmatrix} 1 & 1.5 & 2.5 \\ 0 & 1 & \frac{17}{23} \end{pmatrix} \xrightarrow{z_1+(-1.5z_2)} \begin{pmatrix} 1 & 0 & 1.3913 \\ 0 & 1 & \frac{17}{23} \end{pmatrix}$$

Damit ist $x_1 = 1.3913$, $x_2 = \frac{17}{23}$ und die Lösungsmenge: $\left\{\left(1.3913, \frac{17}{23}\right)\right\}$ wobei $\frac{17}{23} = 0.739\,13$
Mit Excel überprüfen wir, ob die Matrix

$$\mathbf{A} := \begin{pmatrix} 2 & 3 \\ 14 & -2 \end{pmatrix}$$

invertierbar ist. $\det(\mathbf{A}) = -46 \neq 0$ und berechnen dann die inverse Matrix:

$$\begin{pmatrix} \frac{1}{23} & \frac{3}{46} \\ \frac{7}{23} & -\frac{1}{23} \end{pmatrix} = \begin{pmatrix} 4.347\,8 \times 10^{-2} & 6.521\,7 \times 10^{-2} \\ 0.304\,35 & -4.347\,8 \times 10^{-2} \end{pmatrix}$$

Die Lösung ist dann:

$$\begin{pmatrix} \frac{1}{23} & \frac{3}{46} \\ \frac{7}{23} & -\frac{1}{23} \end{pmatrix}\begin{pmatrix} 5 \\ 18 \end{pmatrix} = \begin{pmatrix} \frac{32}{23} \\ \frac{17}{23} \end{pmatrix} = \begin{pmatrix} 1.391\,3 \\ 0.739\,13 \end{pmatrix}$$

Mittels Multiplikation mit Elementarmatrizen erhalten wir alternativ:

$$\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 2 & 3 & 5 \\ 14 & -2 & 18 \end{pmatrix} = \begin{pmatrix} 1 & \frac{3}{2} & \frac{5}{2} \\ 14 & -2 & 18 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 \\ -14 & 1 \end{pmatrix}\begin{pmatrix} 1 & \frac{3}{2} & \frac{5}{2} \\ 14 & -2 & 18 \end{pmatrix} = \begin{pmatrix} 1 & \frac{3}{2} & \frac{5}{2} \\ 0 & -23 & -17 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 \\ 0 & -\frac{1}{23} \end{pmatrix}\begin{pmatrix} 1 & \frac{3}{2} & \frac{5}{2} \\ 0 & -23 & -17 \end{pmatrix} = \begin{pmatrix} 1 & \frac{3}{2} & \frac{5}{2} \\ 0 & 1 & \frac{17}{23} \end{pmatrix}$$

$$\begin{pmatrix} 1 & -\frac{3}{2} \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & \frac{3}{2} & \frac{5}{2} \\ 0 & 1 & \frac{17}{23} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \frac{32}{23} \\ 0 & 1 & \frac{17}{23} \end{pmatrix}$$

$x_1 = \frac{32}{23} = 3.608\,7; \quad x_2 = \frac{17}{23} = 0.739\,13$
Multipliziert man die Elementarmatrizen, erhält man übrigens die Inverse:

$$\begin{pmatrix} 1 & -\frac{3}{2} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -\frac{1}{23} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -14 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{23} & \frac{3}{46} \\ \frac{7}{23} & -\frac{1}{23} \end{pmatrix}$$

Die Multiplikation der Matrix von links her mit Elementarmatrizen ist also schliesslich dasselbe wie die entsprechende Multiplikation mit der Umkehrmatrix.

b) Wir erhalten die erweiterte Matrix

$$\begin{pmatrix} 2 & 3 & 4 & 5 \\ 14 & -2 & 2 & 18 \\ 3 & 12 & 3 & -2 \end{pmatrix}$$

und mit den bekannten Zeilenumwandlungen:

$$\xrightarrow{\frac{1}{2}z_1} \begin{pmatrix} 1 & \frac{3}{2} & 2 & \frac{5}{2} \\ 14 & -2 & 2 & 18 \\ 3 & 12 & 3 & -2 \end{pmatrix} \xrightarrow{z_2 + (-14z_1)} \begin{pmatrix} 1 & \frac{3}{2} & 2 & \frac{5}{2} \\ 0 & -23 & -26 & -17 \\ 3 & 12 & 3 & -2 \end{pmatrix} \xrightarrow{z_3 + (-3z_1)}$$

$$\begin{pmatrix} 1 & \frac{3}{2} & 2 & \frac{5}{2} \\ 0 & -23 & -26 & -17 \\ 0 & \frac{15}{2} & -3 & -\frac{19}{2} \end{pmatrix} \xrightarrow{-\frac{1}{23}z_2} \begin{pmatrix} 1 & \frac{3}{2} & 2 & \frac{5}{2} \\ 0 & 1 & \frac{26}{23} & \frac{17}{23} \\ 0 & \frac{15}{2} & -3 & -\frac{19}{2} \end{pmatrix} \xrightarrow{z_1 + \left(-\frac{3}{2}z_2\right)}$$

$$\begin{pmatrix} 1 & 0 & \frac{7}{23} & \frac{32}{23} \\ 0 & 1 & \frac{26}{23} & \frac{17}{23} \\ 0 & \frac{15}{2} & -3 & -\frac{19}{2} \end{pmatrix} \xrightarrow{z_3 + \left(-\frac{15}{2}z_2\right)} \begin{pmatrix} 1 & 0 & \frac{7}{23} & \frac{32}{23} \\ 0 & 1 & \frac{26}{23} & \frac{17}{23} \\ 0 & 0 & -\frac{264}{23} & -\frac{346}{23} \end{pmatrix} \xrightarrow{-\frac{23}{264}z_3}$$

$$\begin{pmatrix} 1 & 0 & \frac{7}{23} & \frac{32}{23} \\ 0 & 1 & \frac{26}{23} & \frac{17}{23} \\ 0 & 0 & 1 & \frac{173}{132} \end{pmatrix} \xrightarrow{z_1 + \left(-\frac{7}{23}z_3\right)} \begin{pmatrix} 1 & 0 & 0 & \frac{131}{132} \\ 0 & 1 & \frac{26}{23} & \frac{17}{23} \\ 0 & 0 & 1 & \frac{173}{132} \end{pmatrix} \xrightarrow{z_2 + \left(\frac{26}{23}z_3\right)}$$

$$\begin{pmatrix} 1 & 0 & 0 & \frac{131}{132} \\ 0 & 1 & 0 & -\frac{49}{66} \\ 0 & 0 & 1 & \frac{173}{132} \end{pmatrix}$$

Damit ist $x_1 = \frac{131}{132} = 0.992\,42; x_2 = -\frac{49}{66} = -0.742\,42, x_3 = \frac{173}{132} = 1.310\,6$ und die Lösungsmenge

$$\left\{\left(\frac{131}{132}, -\frac{49}{66}, \frac{173}{132}\right)\right\}$$

Mit Excel prüfen wir, ob die Matrix

$$\mathbf{A} := \begin{pmatrix} 2 & 3 & 4 \\ 14 & -2 & 2 \\ 3 & 12 & 3 \end{pmatrix}$$

invertierbar ist: $\det(\mathbf{A}) = 528 \neq 0$. Damit ist die Matrix invertierbar. Wir berechnen die Inverse:

$$\mathbf{A}^{-1} = \begin{pmatrix} -0.056818182 & 0.073863636 & 0.026515152 \\ -0.068181818 & -0.011363636 & 0.098484848 \\ 0.329545455 & -0.028409091 & -0.087121212 \end{pmatrix}$$

Damit ist dann mit

$$\mathbf{b} := \begin{pmatrix} 5 \\ 18 \\ -2 \end{pmatrix}$$

$$\mathbf{A}^{-1}\mathbf{b} = \begin{pmatrix} 0.992\,42 \\ -0.742\,42 \\ 1.310\,6 \end{pmatrix}$$

Auch hier kann man mit den Elementarmatrizen arbeiten. Tun Sie das mit Excel!

2. Man erhält die erweiterte Matrix

$$\begin{pmatrix} -3 & 0.3 & -1.2 & -5.4 \\ 50 & -5 & 20 & 90 \\ 10 & -1 & 4 & 18 \end{pmatrix}$$

und mit den bekannten Zeilenumwandlungen:

$$\xrightarrow{-\frac{1}{3}z_1} \begin{pmatrix} 1 & -0.1 & 0.4 & 1.8 \\ 50 & -5 & 20 & 90 \\ 10 & -1 & 4 & 18 \end{pmatrix} \xrightarrow{z_2+(-50z_1)} \begin{pmatrix} 1 & -0.1 & 0.4 & 1.8 \\ 0 & 0 & 0 & 0 \\ 10 & -1 & 4 & 18 \end{pmatrix}$$

$$\xrightarrow{z_3+(-10z_1)} \begin{pmatrix} 1 & -0.1 & 0.4 & 1.8 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Damit gibt es unendlich viele Lösungen. Die Lösung kann man allgemein ausdrücken durch

$$\begin{pmatrix} 1.8 + 0.1x_2 - 0.4x_3 \\ x_2 \\ x_3 \end{pmatrix}$$

Die Lösungsmenge ist damit:

$$\left\{ \mathbf{x} \mid \mathbf{x} = \begin{pmatrix} 1.8 + 0.1x_2 - 0.4x_3 \\ x_2 \\ x_3 \end{pmatrix} \text{ mit } x_2, x_3 \in \mathbb{R} \right\}$$

Eine konkrete Lösung wäre z.B. mit $x_2 = 3$ und $x_3 = 4$

$$\begin{pmatrix} 0.5 \\ 3 \\ 4 \end{pmatrix}$$

Wir kontrollieren:

$$\begin{pmatrix} -3 & 0.3 & -1.2 \\ 50 & -5 & 20 \\ 10 & -1 & 4 \end{pmatrix} \begin{pmatrix} 0.5 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} -5.4 \\ 90 \\ 18 \end{pmatrix}$$

Mit Excel überprüfen wir die Invertierbarkeit: $\det A = 0$. Damit ist die Matrix nicht invertierbar. Wir müssen die Lösungsmenge mit dem Eliminationsverfahren suchen - am besten mit Excel und mit Hilfe von Elementarmatrizen.

3. Wir erhalten die erweiterte Matrix

$$\begin{pmatrix} 5 & 3 & 4 & 15 & 16 \\ 1 & 0.6 & 0.8 & 3 & 6 \\ 20 & 1 & 3 & 6 & 5 \end{pmatrix}$$

und mit den bekannten Zeilenumwandlungen:

$$\xrightarrow{\frac{1}{5}z_1} \begin{pmatrix} 1 & \frac{3}{5} & \frac{4}{5} & 3 & \frac{16}{5} \\ 1 & 0.6 & 0.8 & 3 & 6 \\ 20 & 1 & 3 & 6 & 5 \end{pmatrix} \xrightarrow{z_2+(-z_1)} \begin{pmatrix} 1 & \frac{3}{5} & \frac{4}{5} & 3 & \frac{16}{5} \\ 0 & 0 & 0 & 0 & \frac{14}{5} \\ 20 & 1 & 3 & 6 & 5 \end{pmatrix}$$

Damit ist das Gleichungssystem nicht lösbar. Die Zeilentransformationen können abgebrochen werden, bevor die Zeilennormalform erreicht ist.
Mit der Inverse kommt man so oder so nicht weiter, da nicht eine quadratische Matrix vorliegt. Kontrollieren Sie die Berechnung mit der Multiplikation von Elementarmatrizen.

4. Man erhält in Matrixdarstellung

$$\mathbf{Aa} = \begin{pmatrix} 5^5 & 5^4 & 5^3 & 5^2 & 5 & 1 \\ 6^5 & 6^4 & 6^3 & 6^2 & 6 & 1 \\ 7^5 & 7^4 & 7^3 & 7^2 & 7 & 1 \\ 8^5 & 8^4 & 8^3 & 8^2 & 8 & 1 \\ 9^5 & 9^4 & 9^3 & 9^2 & 9 & 1 \\ 10^5 & 10^4 & 10^3 & 10^2 & 10 & 1 \end{pmatrix} \begin{pmatrix} a_5 \\ a_4 \\ a_3 \\ a_2 \\ a_1 \\ a_0 \end{pmatrix} = \begin{pmatrix} 10 \\ 12 \\ 5 \\ 3 \\ 20 \\ 50 \end{pmatrix} =: \mathbf{b}$$

Invertierbarkeit? $\det A = -34\,560 \neq 0$. Wir berechnen die Inverse:

$$\begin{pmatrix} -8.333\,3E-3 & 4.166\,7E-2 & -8.333\,3E-2 & 8.333\,3E-2 & -4.166\,7E-2 & 8.333\,3E-3 \\ 0.333\,33 & -1.625 & 3.166\,7 & -3.083\,3 & 1.5 & -0.291\,67 \\ -5.291\,7 & 25.042 & -47.417 & 44.917 & -21.292 & 4.041\,7 \\ 41.667 & -190.38 & 349.33 & -321.92 & 149.0 & -27.708 \\ -162.7 & 712.92 & -1265.0 & 1135.0 & -514.17 & 93.95 \\ 252.0 & -1050.0 & 1800.0 & -1575.0 & 700.0 & -126.0 \end{pmatrix}$$

und damit $\mathbf{A}^{-1}\mathbf{b}$

$$\begin{pmatrix} -0.166\,67 \\ 5.833\,4 \\ -78.502 \\ 507.6 \\ -1577.9 \\ 1895.0 \end{pmatrix}$$

5. a) Man muss die Leistungen zeilenweise zusammenzählen und erhält:

$$\begin{pmatrix} 20650 \\ 16800 \\ 30900 \\ 17120 \\ 21590 \end{pmatrix}$$

b) Wir erhalten das Gleichungssystem:

$$\begin{aligned} 5000 + 50p_1 + 150p_2 + 300p_3 + 1300p_4 + 400p_5 &= 20650p_1 \\ 4500 + 200p_1 + 350p_2 + 7500p_3 + 4900p_4 + 550p_5 &= 16800p_2 \\ 12000 + 300p_1 + 800p_3 + 80p_4 + 7630p_5 &= 30900p_3 \\ 2000 + 100p_1 + 700p_2 + 700p_4 + 380p_5 &= 17120p_4 \\ 4000 + 600p_2 + 300p_3 + 140p_4 + 50p_5 &= 21590p_5 \end{aligned}$$

oder in Standardform:

$$\begin{aligned} 20600p_1 - 150p_2 - 300p_3 - 1300p_4 - 400p_5 &= 5000 \\ -200p_1 + 16450p_2 - 7500p_3 - 4900p_4 - 550p_5 &= 4500 \\ -300p_1 + 30100p_3 - 80p_4 - 7630p_5 &= 12000 \\ -100p_1 - 700p_2 + 16420p_4 - 380p_5 &= 2000 \\ -600p_2 - 300p_3 - 140p_4 + 21540p_5 &= 4000 \end{aligned}$$

Wir kontrollieren die Invertierbarkeit von

$$\mathbf{A} := \begin{pmatrix} 20600 & -150 & -300 & -1300 & -400 \\ -200 & 16450 & -7500 & -4900 & -550 \\ -300 & 0 & 30100 & -80 & -7630 \\ -100 & -700 & 0 & 16420 & -380 \\ 0 & -600 & -300 & -140 & 21540 \end{pmatrix}$$

$\det \mathbf{A} = 3530065\,867850400\,000\,000 \neq 0$. Wir können also die Inverse berechnen:

$$\begin{pmatrix} 4.85794E-05 & 6.60906E-07 & 1.22466E-06 & 6.60272E-07 & 4.05682E-06 \\ 9.16955E-07 & 6.18637E-05 & 1.54976E-05 & 1.86725E-05 & 7.41569E-06 \\ 4.93844E-07 & 4.56555E-07 & 3.34597E-05 & 4.3966E-07 & 1.18808E-05 \\ 3.35747E-07 & 2.68176E-06 & 6.8558E-07 & 6.17435E-05 & 1.40681E-06 \\ 3.46022E-08 & 1.74701E-06 & 9.02156E-07 & 9.27553E-07 & 4.68064E-05 \end{pmatrix}$$

und damit

$$\mathbf{k}_p := \begin{pmatrix} 5000 \\ 4500 \\ 12000 \\ 2000 \\ 4000 \end{pmatrix}$$

$$\mathbf{A}^{-1}\mathbf{k}_p = \begin{pmatrix} 0.266811463 \\ 0.535950179 \\ 0.454442606 \\ 0.151087939 \\ 0.207941281 \end{pmatrix} = \mathbf{p}$$

Wir kontrollieren, ob die an die Endproduktion gelieferten Leistungen zu diesen Preisen die Summe der Primärkosten ergeben:

$$\mathbf{l}_{HB}^T p = \begin{pmatrix} 20000 & 15000 & 22000 & 10000 & 12580 \end{pmatrix} \begin{pmatrix} 0.266811463 \\ 0.535950179 \\ 0.454442606 \\ 0.151087939 \\ 0.207941281 \end{pmatrix} = 27500$$

Summe der Primärkosten: $5000 + 4500 + 12000 + 2000 + 4000 = 27\,500$
Mit R:

```
L=matrix(c(50,200,300,100,0,150,350,0,700,600,300,7500,800,0,300,
1300,4900,80,700,140,400,550,7630,380,50),ncol=5,byrow=T)
lHB=c(20000,15000,22000,10000,12580) #Lieferungen an Hauptbetrieb
kp=c(5000,4500,12000,2000,4000) #Primaerkosten
lt=L%*%rep(1,5) # (Summe der Leistungen an Teilbetriebe pro Teilbetrieb)
lg=lt+lHB #(Gesamtleistung pro Teilbetrieb)
A=diag(as.vector(lg))-t(L) #diag() funktioniert korrekt nur für R-Vektoren, nicht für R-Matrizen
det(A) # (Ueberprüfung auf Invertierbarkeit)
p=solve(A)%*%kp # (effektive Preise)
p
diag(as.vector(lHB))%*%p # (Gesamtkosten pro Teilbetrieb)
t(L)%*%p # (Sekundaerkosten pro Teilbetrieb)
```

1.5 Lernziele

- Die allgemeine Definition linearer Abbildungen hinschreiben können. Wissen, wie man mit Systemen von Ausdrücken lineare Abbildungen definieren kann. Werte von linearen Abbildung, die mit Systemen von Ausdrücken definiert sind, berechnen können. Die ökonomischen Beispiele (Produktionsmatrix) kennen und entsprechende Berechnungen vornehmen können.
- Den Ausdruck „Vektor“ kennen und den Zusammenhang mit n-Tupeln erklären können. Die Multiplikation einer Konstanten mit einem Vektor und die Addition zweier Vektoren durchführen können. Den Begriff des Einheitsvektors kennen.
- Den Zusammenhang von Matrizen und linearen Abbildungen kennen. Lineare Abbildungen mit Hilfe von Matrizen definieren können. Die Multiplikation einer Matrix mit einem Vektor durchführen können und erklären können, was diese repräsentiert. Die Addition von Matrizen und die Multiplikation von Matrizen mit Konstanten durchführen können. Erklären können, was diese Operationen bezüglich der entsprechenden linearen Abbildungen bedeuten. Die Begriffe der Einheitsmatrix und der Nullmatrix kennen und verstehen.
- Den Begriff der „Verkettung“ von Funktionen erläutern können. Erklären können, was die Verkettung linearer Funktionen mit der sogenannten Matrixmultiplikation zu tun hat. Die Multiplikation von Matrizen durchführen können. Die entsprechende ökonomische Anwendung kennen und berechnen können (Produktionsmatrix bezüglich Zwischenprodukten; Produktionsmatrix bezüglich Endprodukten; Berechnung des Rohstoffvektors unter Angabe eines Outputvektors).
- Den Begriff der Umkehrfunktion auf lineare Abbildungen anwenden können und erläutern können, was dieser mit der inversen Matrix zu tun hat. Die Inverse von Matrizen berechnen können. Die ökonomische Anwendung (Input-Output-Analyse) verstehen und entsprechenden Berechnungen durchführen können.
- Die Definition der linearen Gleichungssysteme kennen. Lineare Gleichungssysteme mit Hilfe von Matrizen und Vektoren darstellen können. Die allgemeine Lösung von linearen Gleichungssystemen mit Hilfe der Inversen einer invertierbaren Matrix vorführen können. Den Gaussschen Algorithmus anwenden können. Entscheiden können, ob ein Gleichungssystem genau eine Lösung, keine Lösung oder unendlich viele Lösungen hat. Im Falle genau einer Lösung dies berechnen können. Im Falle unendlich vieler Lösungen die Lösungsmenge angeben können. Den Begriff der Zeilennormalform kennen und solche für Matrizen berechnen können. Die eingeführten Methoden der innerbetrieblichen Leistungsverrechnung verstehen und berechnen können.
- Die Begriffe „Transponierte Matrix“, „Diagonalmatrix“ und „Elementarmatrix“ kennen. Matrizen transponieren können. Die Umkehrmatrix einer Diagonalmatrix angeben können. Die Elementarmatrizen verwenden können, um Gleichungssysteme aufzulösen.

Kapitel 2

Determinanten

I:\dateien\Mathematik\111Math\Lineare_Algebra\Determinanten.tex
Determinanten spielen in der linearen Algebra eine wichtige Rolle. Es handelt sich um die Werte einer Abbildung, die quadratischen Matrizen reelle Zahlen zuordnet. Determinanten sind genau dann von 0 verschieden, wenn die quadratischen Matrizen den vollen Rang haben und damit injektive Abbildungen beschreiben.

Definition 2.0.1. *Sei M die Menge der quadratischen Matrizen.*

$$\det : M \to \mathbb{R}$$

heisst Determinantenfunktion (und die Funktionswerte Determinanten) genau dann, wenn:

1. $\det \mathbf{C} = \det \mathbf{A} + \det \mathbf{B}$ *für eine Matrix* $\mathbf{A}$ *mit der i-ten Zeile* $a_i = (a_{i1}, ..., a_{in})$ *und einer Matrix* $\mathbf{B}$ *mit der i-ten Zeile* $b_i = (b_{i1}, ..., b_{in})$*, wobei* $\mathbf{A}$ *und* $\mathbf{B}$ *bis auf die i-te Zeile identisch sind und* $\mathbf{C}$ *die Matrix ist, die mit* $\mathbf{A}$ *bis auf die i-te Zeile identisch ist, und für die i-Zeile von* $\mathbf{C}$ *gilt:* $c_i = (a_{i1} + b_{i1}, ..., a_{in} + b_{in})$.
2. $\det \mathbf{B} = \lambda \det \mathbf{A}$*, wobei die Matriz* $\mathbf{B}$ *aus* $\mathbf{A}$ *entsteht, indem eine Zeile* a_i *von* $\mathbf{A}$ *mit einer Konstanten* $\lambda \neq 0$ *multipliziert (d.h.* $b_i := \lambda a_i = (\lambda a_{i1}, ..., \lambda a_{in})$*).*
3. $\det \mathbf{A} = 0$*, wenn* $\mathbf{A}$ *zwei identische Zeilen aufweist.*
4. $\det \mathbf{E} = 1$.

$\Diamond$

Es wäre noch zu beweisen, dass eine Abbildung mit diesen Eigenschaften existiert, worauf hier verzichtet wird. Mit Hilfe der Eigenschaften der Definition kann man weitere Eigenschaften herleiten:

Satz 2.0.2. $\det(\lambda \mathbf{A}) = \lambda^n \det \mathbf{A}$

Beweis. Folgt unmittelbar aus Punkt 2 der Definition, da $\lambda\mathbf{A}$ jede Zeile von $\mathbf{A}$ mit λ multipliziert. □

Satz 2.0.3. *Besteht eine Zeile von* $\mathbf{A}$ *aus Nullen, so ist* $\det \mathbf{A} = 0$

Beweis. Seien zwei Matrizen $\mathbf{A}$ und $\mathbf{B}$ gegeben, die bis auf die i-te Zeile identisch sind, wobei die i-Zeile von $\mathbf{B}$ aus Nullen besteht. Bilden wir aus $\mathbf{A}$ und $\mathbf{B}$ eine Matrix $\mathbf{C}$ gemäss Punkt 1 der Definition der Determinantenabbildung, so gilt $\mathbf{C} = \mathbf{A}$ und wegen $\det \mathbf{C} = \det \mathbf{A} + \det \mathbf{B}$ gilt: $\det \mathbf{C} = \det \mathbf{C} + \det \mathbf{B}$. Damit ist $\det \mathbf{B} = 0$. □

Satz 2.0.4. *Vertauscht man zwei Zeilen einer Matrix* $\mathbf{A}$*, so dass eine Matrix* $\mathbf{B}$ *entsteht, gilt* $\det \mathbf{B} = -\det \mathbf{A}$

Beweis. Sei $\begin{pmatrix} a_i \\ a_j \end{pmatrix}$ eine Darstellung der Matrix $\mathbf{A}$, mit den zu vertauschenden Zeilen a_i und a_j (die übrigen Zeilen werden nicht angegeben, das sie hier keine Rolle spielen). Dann ist $\begin{pmatrix} a_j \\ a_i \end{pmatrix}$ eine entsprechende Darstellung von $\mathbf{B}$. Damit gilt

$$\det \mathbf{A} + \det \mathbf{B} = \det \begin{pmatrix} a_i \\ a_j \end{pmatrix} + \det \begin{pmatrix} a_j \\ a_i \end{pmatrix}$$

Wegen Punkt 3 der Determinanten-Definition gilt

$$\det \begin{pmatrix} a_i \\ a_i \end{pmatrix} = 0 = \det \begin{pmatrix} a_j \\ a_j \end{pmatrix}$$

und damit mit Punkt 1 und im letzten Schritt mit Punkt 3 der Determinanten-Definition

$$\begin{aligned} \det \mathbf{A} + \det \mathbf{B} &= \det \begin{pmatrix} a_i \\ a_i \end{pmatrix} + \det \begin{pmatrix} a_i \\ a_j \end{pmatrix} + \det \begin{pmatrix} a_j \\ a_i \end{pmatrix} + \det \begin{pmatrix} a_j \\ a_j \end{pmatrix} \\ &= \det \begin{pmatrix} a_i \\ a_i + a_j \end{pmatrix} + \det \begin{pmatrix} a_j \\ a_j + a_i \end{pmatrix} \\ &= \det \begin{pmatrix} a_i + a_j \\ a_i + a_j \end{pmatrix} = 0 \end{aligned}$$

Entsprechend ist

$$\det \mathbf{B} = -\det \mathbf{A}$$

□

Satz 2.0.5. *Entsteht die Matrix* $\mathbf{B}$ *aus der Matrix* $\mathbf{A}$*, indem man zur Zeile* a_i *die mit* λ *multiplizierte Zeile* a_j *dazuzählt* ($i \neq j$, $\lambda \neq 0$)*, d.h.* $b_i = a_i + \lambda a_j$*, so gilt*

$$\det \mathbf{B} = \det \mathbf{A}$$

Beweis.

$$\det \mathbf{B} = \det \begin{pmatrix} a_i + \lambda \alpha_j \\ a_j \end{pmatrix} = \det \mathbf{A} + \lambda \det \begin{pmatrix} a_j \\ a_j \end{pmatrix} = \det \mathbf{A} + 0 = \det \mathbf{A}$$

□

Damit kennen wir die Auswirkungen der Regeln, die wir beim Gausschen Eliminationsverfahren verwenden, auf die Determinanten. Um Determinanten berechnen zu können, fehlt uns noch der folgende Satz:

Satz 2.0.6. *Ist* $\mathbf{A}$ *eine obere (untere) Dreiecksmatrix (alle Zellen oberhalb (unterhalb) der Hauptdiagonale weisten die Komponenten 0 auf), so gilt*

$$\det \mathbf{A} = \prod_{i=1}^{n} a_{ii}$$

Die Determinante ist also das Produkt der Einträge auf der Hauptdiagonalen.

Beweis. Sind alle Einträge auf der Diagonalmatrix von 0 verschieden, kann man mit Hilfe von Satz 2.0.5 die Dreiecksmatrix in eine Diagonalmatrix $diag\,(a_{11}, ..., a_{nn})$ umwandeln. Für diese erhält man dann durch das zeilenweise „Ausklammern“ der Diagonaleinträge gemäss Punkt 2 der Definition der Determinantenfunktion

$$\det\,(diag\,(a_{11}, ..., a_{nn})) = \prod_{i=1}^{n} a_{ii} \cdot \det \mathbf{E} = \prod_{i=1}^{n} a_{ii}$$

Sind die Einträge auf der Diagonalmatrix nicht von 0 verschieden, so wählt man das grösste i, so dass $a_{i+1,i+1}, ..., a_{nn}$ von 0 verschieden sind. Mit Hilfe von Satz 2.0.5 werden die Zellen der i-ten Zeile rechts von $a_{ii} = 0$ in Nullen verwandelt. Die i-te Zeile besteht damit aus Nullen und mit Satz 2.0.3 gilt

$$\det \mathbf{A} = 0 = \prod_{i=1}^{n} a_{ii}$$

Der Beweis für die untere Dreiecksmatrix verläuft analog. □

Mit Hilfe dieser Sätze kann man nun Determinanten berechnen, indem man das Gausssche Eliminationsverfahren anwendet und bezüglich der Veränderungen der Determinante Protokoll führt, wobei wir die Einträge auf der Hauptdiagonale nicht auf 1 setzen, um weniger Protokoll führen zu müssen.

Beispiel 2.0.7. *Sei*

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

mit $a \neq 0$. Damit gilt durch Addition der ersten mit $-\frac{c}{a}$ multiplizierten Zeile zur zweiten Zeile:

$$\begin{pmatrix} a & b \\ 0 & d - \frac{b}{a}c \end{pmatrix}$$

und gemäss Satz 2.0.6

$$\begin{aligned} \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} &= \det \begin{pmatrix} a & b \\ 0 & d - \frac{b}{a}c \end{pmatrix} \\ &= a\left(d - \frac{b}{a}c\right) \\ &= ad - a\frac{b}{a}c \\ &= ad - bc \end{aligned}$$

Ist $a = 0$ und $c \neq 0$, so vertauscht man die zwei Zeilen und erhält

$$\begin{aligned} \det \begin{pmatrix} 0 & b \\ c & d \end{pmatrix} &= -\det \begin{pmatrix} c & d \\ 0 & b \end{pmatrix} \\ &= -cd \end{aligned}$$

womit auch hier

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ab - cd$$

gilt. Gilt $a = 0 = c$, gilt gemäss Dreiecksregel:

$$\det \begin{pmatrix} 0 & b \\ 0 & d \end{pmatrix} = 0 \cdot d = 0 = ab - cd$$

Für 2×2-Matrizen gibt es also eine sehr einfach Regel zur Berechnung der Determinante. Damit gilt z.B.

$$\det \begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix} = 1 \cdot 3 - 2 \cdot 4 = -5$$

◊

Beispiel 2.0.8. *Für 3×3-Matrizen erhält man eine etwas kompliziertere, aber immer noch leicht anwendbare Regel. Für $a\neq 0\neq e-\frac{b}{a}d$ gilt:*

$$\begin{pmatrix} a & b & c\\ d & e & f\\ g & h & k\end{pmatrix} \xrightarrow{a_2=a_2+a_1\cdot -\frac{d}{a}} \begin{pmatrix} a & b & c\\ 0 & e-\frac{d}{a}b & f-\frac{d}{a}c\\ g & h & k\end{pmatrix}$$
$$\xrightarrow{a_3+a_1\cdot -\frac{g}{a}} \begin{pmatrix} a & b & c\\ 0 & e-\frac{b}{a}d & f-\frac{c}{a}d\\ 0 & h-\frac{g}{a}b & k-\frac{g}{a}c\end{pmatrix}$$
$$\xrightarrow{a_3+a_2\cdot -\frac{\left(h-\frac{b}{a}g\right)}{e-\frac{b}{a}d}} \begin{pmatrix} a & b & c\\ 0 & e-\frac{b}{a}d & f-\frac{c}{a}d\\ 0 & 0 & k-\frac{1}{a}cg-\frac{\left(h-\frac{b}{a}g\right)}{e-\frac{b}{a}d}\left(f-\frac{c}{a}d\right)\end{pmatrix}$$

Damit gilt

$$\begin{aligned}\det\begin{pmatrix} a & b & c\\ d & e & f\\ g & h & k\end{pmatrix} &= a\left(e-\frac{b}{a}d\right)\left(k-\frac{1}{a}cg-\frac{\left(h-\frac{b}{a}g\right)}{e-\frac{b}{a}d}\left(f-\frac{c}{a}d\right)\right)\\ &= (ae-bd)\left(\frac{ak-cg}{a}-\frac{\frac{ha-bg}{a}}{\frac{ea-bd}{a}}\left(\frac{af-cd}{a}\right)\right)\\ &= \frac{1}{a}(ae-bd)\left(ak-cg-\frac{ha-bg}{ae-bd}(af-cd)\right)\\ &= \frac{1}{a}(ae-bd)\left(\frac{(ak-cg)(ae-bd)-(ha-bg)(af-cd)}{ae-bd}\right)\\ &= \frac{1}{a}\left((ak-cg)(ae-bd)-(ha-bg)(af-cd)\right)\\ &= \frac{1}{a}\left(a^2ke-a^2fh-acge+abfg+acdh-abdk\right)\\ &= ake-cge-afh+bfg+cdh-bdk\end{aligned}$$

Man bildet die Summe der Produkte der Komponenten in Hauptdiagonalrichtung (aek, bfg, cdh) und zählt von diesen die Produkte in Nebendiagonalrichtung ab (ceg, bdk, hfa). Mit der Regel gilt z.B.

$$\begin{aligned}\det\begin{pmatrix} 4 & 5 & 3\\ 2 & 1 & 8\\ 3 & 2 & 7\end{pmatrix} &= 4\cdot 1\cdot 7+5\cdot 8\cdot 3+3\cdot 2\cdot 2-(3\cdot 1\cdot 3+2\cdot 5\cdot 7+4\cdot 8\cdot 2)\\ &= 17\end{aligned}$$

◊

Beispiel 2.0.9. *Man berechne die Determinante von*

$$\begin{pmatrix} 4 & 6 & 7 & 5\\ 3 & 7 & 8 & 12\\ 2 & 9 & 2 & 6\\ 1 & 12 & 1 & 7\end{pmatrix}$$

Für $n\times n$-Matrizen mit $n>3$, gibt es nicht mehr einfache Formeln für die Berechnung: Man erhält mittels Multiplikation von etwas angepassten Elementarmatrizen (um nicht beim Setzen auf

Einsen auf der Hauptdiagonale Protokoll führen zu müssen).

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ -\frac{1}{4}3 & 1 & 0 & 0 \\ -\frac{1}{4}2 & 0 & 1 & 0 \\ -\frac{1}{4}1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 4 & 6 & 7 & 5 \\ 3 & 7 & 8 & 12 \\ 2 & 9 & 2 & 6 \\ 1 & 12 & 1 & 7 \end{pmatrix} = \begin{pmatrix} 4 & 6 & 7 & 5 \\ 0 & 2.5 & 2.75 & 8.25 \\ 0 & 6 & -1.5 & 3.5 \\ 0 & 10.5 & -0.75 & 5.75 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -\frac{1}{2.5}6 & 1 & 0 \\ 0 & -\frac{1}{2.5}10.5 & 0 & 1 \end{pmatrix} \begin{pmatrix} 4 & 6 & 7 & 5 \\ 0 & 2.5 & 2.75 & 8.25 \\ 0 & 6 & -1.5 & 3.5 \\ 0 & 10.5 & -0.75 & 5.75 \end{pmatrix} = \begin{pmatrix} 4 & 6 & 7 & 5 \\ 0 & 2.5 & 2.75 & 8.25 \\ 0 & 0 & -8.1 & -16.3 \\ 0 & 0 & -12.3 & -28.9 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{-8.1}(-12.3) & 1 \end{pmatrix} \begin{pmatrix} 4 & 6 & 7 & 5 \\ 0 & 2.5 & 2.75 & 8.25 \\ 0 & 0 & -8.1 & -16.3 \\ 0 & 0 & -12.3 & -28.9 \end{pmatrix} = \begin{pmatrix} 4 & 6 & 7 & 5 \\ 0 & 2.5 & 2.75 & 8.25 \\ 0 & 0 & -8.1 & -16.3 \\ 0 & 0 & 0 & -4.1481 \end{pmatrix}$$

Damit ist die Determinante:

$$4 \cdot 2.5 \cdot (-8.1) \cdot (-4.1481) = 336$$

$\Diamond$

Die Determinanten von Elementarmatrizen sind einfach zu bestimmen. Elementarmatrizen des Typs I (Ersetzen einer Eins der Einheitsmatrix durch eine von 0 verschiedene reelle Zahl λ) weisen gemäss Punkt 2 der Definition 2.0.1 die Derminante λ auf. Elementarmatrizen des Typs II sind Dreiecksmatrizen mit Einsen auf der Diagonalen. Sie weisen damit die Determinante 1 auf. Elementarmatrizen des Typs III (Austauschen von Zeilen der Einheitsmatrix) weisen die Determinante -1 auf.

Satz 2.0.10. *Sei* $\mathbf{A}$ *eine* $n \times n$*-Matrix. Dann gilt:* $\det A = 0$ *genau dann, wenn der Rang von* $\mathbf{A} < n$.

Beweis. Gemäss Definition des Rangs n einer $n \times n$ Matrix, hat eine Matrix den Rang n, wenn sie in Zeilennormalform gebracht werden kann, so dass alle Zeilen mit einer 1 anfangen. Genau in diesem Fall gilt aber auch $\det \mathbf{A} \neq 0$. Fangen nicht alle Zeilen mit einer 1 an, bestehen entsprechende Zeilen aus Nullen. Der Rang der Matrix ist kleiner als n und $\det \mathbf{A} = 0$. □

Satz 2.0.11. $\det\left(\mathbf{A}^{-1}\right) = \frac{1}{\det \mathbf{A}}$ *(*$\det \mathbf{A} \neq 0$*, da* $\mathbf{A}^{-1}$ *definiert ist).*

Beweis. Es gilt $\det\left(\mathbf{A}\mathbf{A}^{-1}\right) = \det \mathbf{A} \cdot \det \mathbf{A}^{-1}$ und $\det\left(\mathbf{A}\mathbf{A}^{-1}\right) = \det \mathbf{E} = 1$. □

Inversen von Elementarmatrizen sind wiederum Elementarmatrizen und zwar des gleichen Typs. Die Determinante der Inversen einer Elementarmatrix des Typs I mit der Determinante λ ist gemäss vorhergehendem Satz $\frac{1}{\lambda}$. Die Determinante der Inversen einer Elementarmatrix des Typs II ist gemäss vorangehendem Satz 1. Die Inverse einer Elementarmatrix des Typs III ist mit ihr identisch (Orthogonalmatrix, s. später) - deren Determinante beträgt also -1.
Jede quadratische Matrix mit vollem Rang kann als Produkt von Elementarmatrizen de Typs I und II dargestellt werden. Durch Multiplikation einer Matrix $\mathbf{A}$ von links her mit Elementarmatrizen $\mathbf{E}_i$ entsteht die Einheitsmatrix, d.h. es gilt $\mathbf{E}_n \cdot \ldots \cdot \mathbf{E}_1 \cdot \mathbf{A} = \mathbf{E}$. Multipliziert man diese Gleichung von links her mit den Inversen der Einheitsmatrizen, erhält man $\mathbf{A} = \mathbf{E}_1^{-1} \cdot \ldots \cdot \mathbf{E}_n^{-1} \cdot \mathbf{E}$. Multipliziert man eine Matrix $\mathbf{A}$ von links her mit einer Elementarmatrix des Typs I mit der Determinante λ, so gilt mit obigen Regeln $\det(\mathbf{E}_i\mathbf{A}) = \det(\mathbf{E}_i) \cdot \det \mathbf{A} = \lambda \det \mathbf{A}$, es wird ja eine Zeile mit λ multipliziert. Multipliziert man eine Matrix $\mathbf{A}$ von links her mit einer Elementarmatrix $\mathbf{E}_i$ des Typs II, so gilt gemäss obigen Regeln $\det(\mathbf{E}_i\mathbf{A}) = \det(\mathbf{E}_i) \det \mathbf{A} = \det \mathbf{A}$, es wird ja eine Zeile zu einer anderen dazugezählt. Damit gilt:

$$\det(\mathbf{E}_n \cdot \ldots \cdot \mathbf{E}_1 \cdot \mathbf{A}) = \det \mathbf{E}_n \cdot \ldots \cdot \det \mathbf{E}_1 \cdot \det \mathbf{A} = \det \mathbf{E} = 1$$

und

$$\det \mathbf{A} = \det \mathbf{E}_1^{-1} \cdot \ldots \cdot \det \mathbf{E}_n^{-1} \cdot \det \mathbf{E}$$

Satz 2.0.12. $\det(\mathbf{AB}) = \det \mathbf{A} \cdot \det \mathbf{B}$ *für* $n \times n$*-Matrix* $\mathbf{A}$ *und* $\mathbf{B}$ *mit vollem Rang* n*.*

Beweis. Da jede quadratische Matrix mit vollem Rang als Produkt von Elementarmatrizen des Typs I und II ausgedrückt werden kann, kann ihr Produkt ebenfalls als solches Produkt dargestellt werden. Mit

$$\mathbf{A} = \mathbf{E}_a^{-1} \cdot \ldots \cdot \mathbf{E}_a^{-1} \cdot \mathbf{E}$$
$$\mathbf{B} = \mathbf{E}_b^{-1} \cdot \ldots \cdot \mathbf{E}_b^{-1} \cdot \mathbf{E}$$

gilt

$$\mathbf{BA} = \mathbf{E}_b^{-1} \cdot \ldots \cdot \mathbf{E}_b^{-1} \cdot \mathbf{E}_a^{-1} \cdot \ldots \cdot \mathbf{E}_a^{-1} \cdot \mathbf{E}$$

und damit

$$\begin{aligned} \det(\mathbf{BA}) &= \det\left(\mathbf{E}_b^{-1} \cdot \ldots \cdot \mathbf{E}_b^{-1} \cdot \mathbf{E}_a^{-1} \cdot \ldots \cdot \mathbf{E}_a^{-1} \cdot \mathbf{E}\right) \\ &= \det\left(\mathbf{E}_b^{-1} \cdot \ldots \cdot \mathbf{E}_b^{-1}\right) \det\left(\mathbf{E}_a^{-1} \cdot \ldots \cdot \mathbf{E}_a^{-1}\right) \\ &= \det(\mathbf{B}) \det(\mathbf{A}) \end{aligned}$$

□

Der Satz gilt auch für quadratische Matrizen mit $\det(\mathbf{A}) = 0$ oder $\det(\mathbf{B}) = 0$. Die Multiplikation einer Matrix mit einer Matrix $\mathbf{A}$ mit $\det(\mathbf{A}) = 0$ erzeugt eine Matrix, die nicht vollen Rang aufweist - verkettet man eine nicht injektive Funktion mit einer injektiven oder nicht injektiven Funktion, entsteht eine nicht-injektive Funktion. Entsprechend gilt auch hier $\det(\mathbf{AB}) = \det \mathbf{A} \cdot \det \mathbf{B}$.

Beispiel 2.0.13. *Wir betrachten die Darstellung der Matrix*

$$\begin{pmatrix} 5 & 3 \\ 2 & 4 \end{pmatrix}$$

mittels Elementarmatrizen. Mit Hilfe von Elementarmatrizen erhält man (Vorgehensweise: von rechts nach links werden die entsprechenden Matrizen vorgeschaltet. Statt die jeweiligen Operationen durchzuführen und dann die nächste Elementarmatrix vorne dranzusetzen, lässt man die Faktoren jeweils stehen):

$$\begin{pmatrix} 1 & -\frac{3}{5} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \frac{5}{14} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{5} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 5 & 3 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Entsprechend ergibt (Wenn $\mathbf{ABC} = \mathbf{D}$*, dann ist* $\mathbf{BC} = \mathbf{A}^{-1}\mathbf{D}$ *und* $\mathbf{C} = \mathbf{B}^{-1}\mathbf{A}^{-1}\mathbf{D}$*.*

$$\begin{aligned} &\begin{pmatrix} \frac{1}{5} & 0 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ 0 & \frac{5}{14} \end{pmatrix}^{-1} \begin{pmatrix} 1 & -\frac{3}{5} \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 5 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \frac{14}{5} \end{pmatrix} \begin{pmatrix} 1 & \frac{3}{5} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 5 & 3 \\ 2 & 4 \end{pmatrix} \end{aligned}$$

Die Determinante der ersten Matrix ist 5*, die der zweiten ist* 1 *(Diagonalmatrix), die der dritten ist* $\frac{14}{5}$*, die der vierten ist* 1*. Das ergibt das Produkt* 14*.* ◊

Satz 2.0.14. $\det \mathbf{A}^T = \det \mathbf{A}$

Beweis. Die Transponierten der Elementarmatrizen des Typs I und II sind Elementarmatrizen des jeweils gleichen Typs (Typ I weist als Determinante λ auf, Typ II die Zahl 1). Durchs Transponieren verändert sich die Determinante nicht (die Diagonale der Dreiecksmatrizen bleibt ja unverändert). Zudem gilt für Matrizen: Wenn

$$\mathbf{A} = \mathbf{BCDF}$$

dann ist

$$\mathbf{A}^T = \mathbf{F}^T\mathbf{D}^T\mathbf{C}^T\mathbf{B}^T$$

Entsprechend ist die Transponierte einer Matrix das Produkt der transponierten Elementarmatrizen, welche untransponiert $\mathbf{A}$ bilden. □

Bemerkung 2.0.15. *Manchmal wird statt* $\det \mathbf{A}$ *die Notation* $|\mathbf{A}|$ *verwendet, auch in diesem Skript.* ◊

Kapitel 3

Vektorräume

I:\dateien\Mathematik\111Math\Lineare_Algebra\002vektorraeueme.txt

3.1 Graphische Bedeutung der Operationen auf n-Tupel

n-Tupel werden - mittels Bijektion von $\mathbb{R}^n$ in den $n-$dimensionalen Raum - graphisch als Punkte gedeutet und werden auch „Vektoren“ genannt. Wir schreiben sie als $(x_1, .., x_n)$ oder $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$.
Oben wurden für $a \in \mathbb{R}$ und $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ die Operationen

$$a\mathbf{x} = a \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} ax_1 \\ \vdots \\ ax_n \end{pmatrix}$$

und

$$\mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_1 \end{pmatrix}$$

eingeführt. Es lohnt sich, die graphische Auswirkungen dieser Operationen zu betrachten.
$a\mathbf{x}$ ist ein Punkt auf der Geraden, die durch den Punkt $\mathbf{x}$ und den $\mathbf{0}$-Punkt des Koordinatensystems verläuft.

Beispiel 3.1.1. *(zweidimensionaler Raum) Sei* $\mathbf{x} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$. *Damit ist*

$$2\mathbf{x} = 2 \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \\ 6 \end{pmatrix}$$

auf der Geraden

$$g := \{\mathbf{z} | \mathbf{z} = a\mathbf{x}, a \in \mathbb{R}\}$$

(s. Abbildung 3.1.1) .

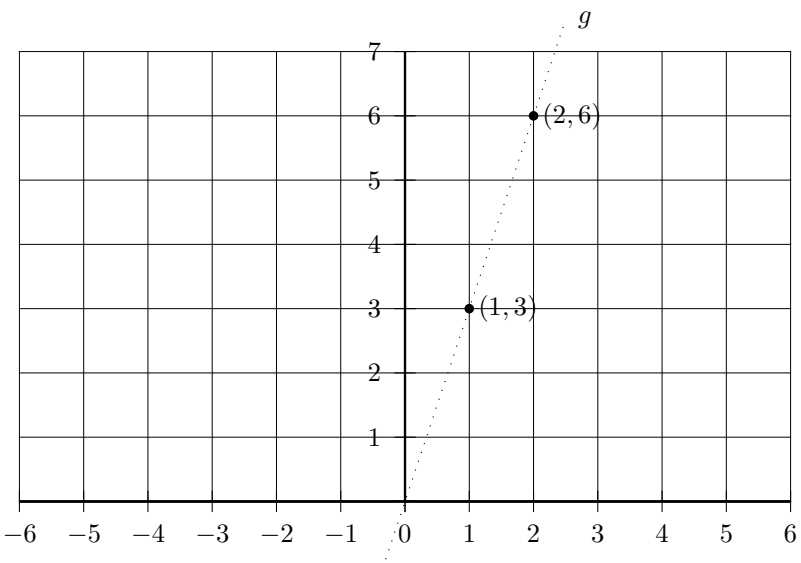


Abbildung 3.1.1: Punkt $\mathbf{x} = (1, 3)$ und $2\mathbf{x} = (2, 6)$. Menge g der Punkte, die durch Multiplikation von $\mathbf{x}$ mit einer reellen Zahl entsteht.

$\Diamond$

Die graphische Auswirkung der Addition von zwei Vektoren kann an der folgenden Graphik 3.1.2 nachvollzogen werden, mit

$$\mathbf{x} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$
$$\mathbf{y} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$
$$\mathbf{x} + \mathbf{y} = \begin{pmatrix} 1 \\ 3 \end{pmatrix} + \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

ergibt sich graphisch (s. Abbildung 3.1.2)

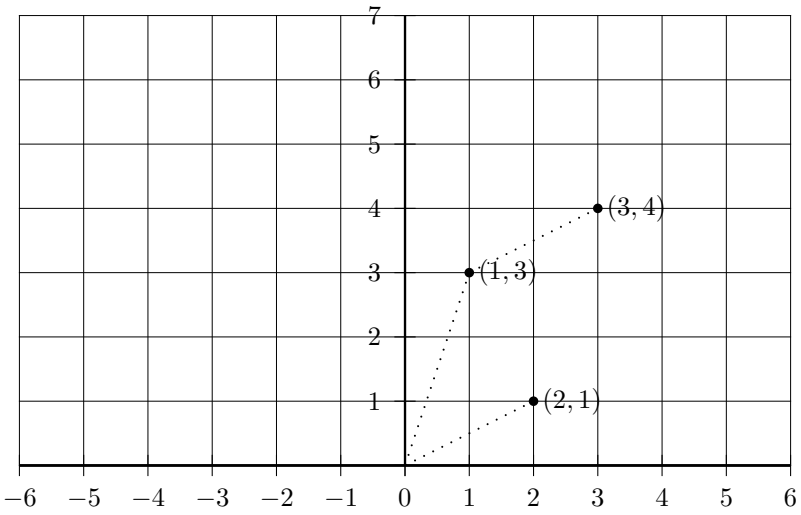


Abbildung 3.1.2: Addition der Punkte $\mathbf{x} = (1, 3)$ und $\mathbf{y} = (2, 1)$. Um graphisch die Summe zu erhalten, verschiebt man die Strecke zwischen $\mathbf{0}$ und $(2, 1)$ parallel nach oben, so dass der $\mathbf{0}$-Punkt auf $(1, 3)$ verschoben wird. Die Summe ist der Punkt, in den $(2, 1)$ verschoben wird.

3.2 $\mathbb{R}^n$ als Vektorraum

Wir können die Menge der Punkte betrachten, die man durch die Multiplikation von $\mathbf{x}$ durch eine reelle Zahl a erhält. Es entsteht für jedes $\mathbf{x} \in \mathbb{R}^n$ eine Gerade g :

$$g = \{\mathbf{z} | \mathbf{z} = a\mathbf{x}, a \in \mathbb{R}\}.$$

$\mathbf{x}$ ist offensichtlich Element von g - mit $\mathbf{x} = \mathbf{z} = 1 \cdot \mathbf{x}$. Ebenso ist der Nullvektor $\mathbf{0} \in g$ mit $\mathbf{0} = \mathbf{z} = 0 \cdot \mathbf{x}$. Die Gerade g bildet einen sogenannten Vektorraum.

Definition 3.2.1. *V ist ein Vektorraum, wenn für $a \in \mathbb{R}$ und $\mathbf{x}, \mathbf{y} \in V \subset \mathbb{R}^n$ gilt:*

$$a\mathbf{x} = a \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} ax_1 \\ \vdots \\ ax_n \end{pmatrix} \in V \tag{3.2.1}$$

und

$$\mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_1 \end{pmatrix} \in V \tag{3.2.2}$$

◊

Ein Vektorraum enthält immer den Nullpunkt des Koordinatensystems (gemäss (3.2.1), für $a = 0$).

Bemerkung 3.2.2. *Die gelieferte Vektorraumdefinition ist nicht allgemein - deshalb steht statt „wenn“ nicht „genau dann, wenn“. Es gibt Vektorräume, die statt Mengen von n-Tupeln z.B. spezifische Klassen von Funktionen oder Matrizen enthalten. Statt $a \in \mathbb{R}$, kann a Element anderer Mengen sein, die spezifische Bedingungen erfüllen (die mit Addition und der Multiplikation einen sogenannten „Körper“ K bilden). Im letzten Kapitel wird ein Beispiel geliefert mit spezifischen Funktionen als Vektoren.* ◊

Satz 3.2.3. *Für jedes* $\mathbf{x} \in \mathbb{R}^n$ *bildet die Menge*

$$g = \{z | z = a\mathbf{x}, a \in \mathbb{R}\},$$

$n \geq 2$, *einen Vektorraum.*

Beweis. (i) Zu zeigen ist: wenn $\mathbf{y} \in g$, dann ist $b\mathbf{y} \in g$ mit $b \in \mathbb{R}$. Sei also $\mathbf{y} \in g$. Dann gibt es gemäss Definition von g ein $a \in \mathbb{R}$, so dass $\mathbf{y} = a\mathbf{x}$. Damit ist $b\mathbf{y} = ba\mathbf{x}$ und da ba eine reelle Zahl ist, $b\mathbf{y} = ba\mathbf{x} \in g$.
(ii) Zu zeigen ist: wenn $\mathbf{y}, \mathbf{z} \in g$, dann ist $\mathbf{y} + \mathbf{z} \in g$. Seien also $\mathbf{y}, \mathbf{z} \in g$. Damit gibt es gemäss Definition von g reelle Zahlen a und b, so dass $\mathbf{y} = a\mathbf{x}$ und $\mathbf{z} = b\mathbf{x}$. Es gilt damit $\mathbf{y} + \mathbf{z} = a\mathbf{x} + b\mathbf{x} = (a + b)\,\mathbf{x}$. Da $a + b$ eine reelle Zahl ist, ist $(a + b)\,\mathbf{x}$ gemäss Definition von g Element von g. Damit ist $\mathbf{y} + \mathbf{z} \in g$. □

Siehe Abbildung 3.1.1 für ein Beispiel für „Wenn $\mathbf{x} \in g$, dann $a\mathbf{x} \in g$“ im $\mathbb{R}^2$. Für „Wenn $\mathbf{x}, \mathbf{y} \in g$, dann $\mathbf{x} + \mathbf{y} \in g$“ betrachte man folgendes

Beispiel 3.2.4.

$$\mathbf{x} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$
$$\mathbf{y} = \begin{pmatrix} 1.7 \\ 5.1 \end{pmatrix}$$

Offensichtlich ist

$$\begin{pmatrix} 1.7 \\ 5.1 \end{pmatrix} \in g := \left\{ \mathbf{z} | \mathbf{z} = a \begin{pmatrix} 1 \\ 3 \end{pmatrix}, a \in \mathbb{R} \right\},$$

denn mit $a = 1.7$ *gilt*

$$1.7 \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 1.7 \\ 5.1 \end{pmatrix}.$$

Es gilt:

$$\mathbf{x} + \mathbf{y} = \begin{pmatrix} 1 \\ 3 \end{pmatrix} + \begin{pmatrix} 1.7 \\ 5.1 \end{pmatrix} = \begin{pmatrix} 2.7 \\ 8.1 \end{pmatrix} \in g,$$

denn

$$2.7 \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 2.7 \\ 8.1 \end{pmatrix}$$

(s. Abbildung 3.2.3).

◊

Definition 3.2.5. *Ein Vektor* $\mathbf{x}$ *ist von den Vektoren* $\mathbf{y}_1, .., \mathbf{y}_n$, $n \geq 1$, *linear abhängig genau dann, wenn es reelle Zahlen* a_i *gibt, so dass*

$$\mathbf{x} = \sum_{i=1}^{n} a_i \mathbf{y}_i$$

Wir sagen dann, $\mathbf{x}$ *lasse sich als „Linearkombination von* $\mathbf{y}_1, ..., \mathbf{y}_n$ *“ darstellen. Lässt sich* $\mathbf{x}$ *nicht so darstellen, wird* $\mathbf{x}$ *„linear unabhängig von* $\mathbf{y}_1, ..., \mathbf{y}_n$ *“ genannt.* ◊

Bemerkung 3.2.6. *(1) Gemäss Definition ist im Falle* $n = 1$ $\mathbf{x}$ *von* $\mathbf{y}$ *linear abhängig genau dann, wenn es ein eine reelle Zahl a gibt, so dass* $\mathbf{x} = a\mathbf{y}$.
(2) Ist $\mathbf{x} \neq \mathbf{0}$ *und* $\mathbf{x} = \sum_{i=1}^{n} a_i \mathbf{y}_i$ *ist* $a_i \neq 0$ *für mindestens ein* a_i.
(3) $\mathbf{0}$ *ist linear abhängig, da* $\mathbf{0} = \sum_{i=1}^{n} a_i \mathbf{y}_i$, *wenn* $a_i = 0$ *für alle* $i \in \mathbb{N}_n^*$. ◊

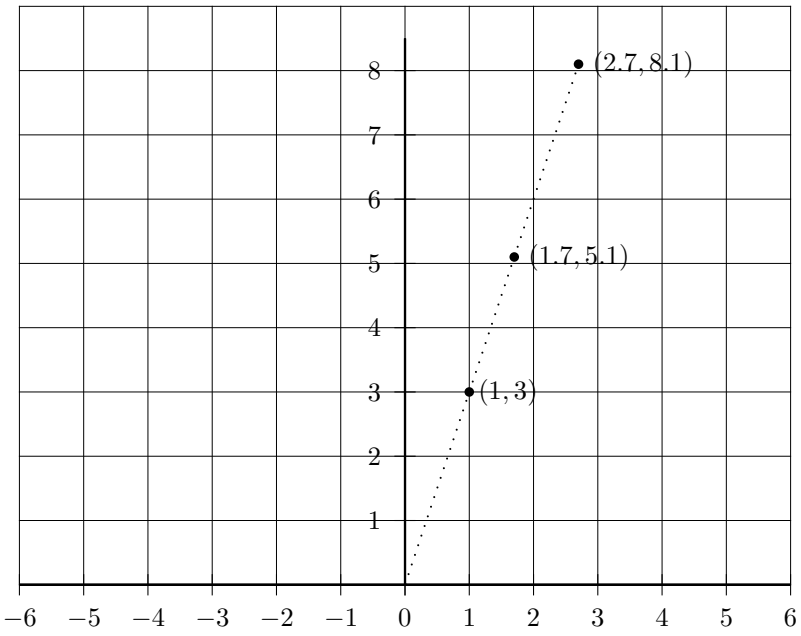


Abbildung 3.2.3: Addition zweier Punkte desselben Vektorraums am Beispiel einer Geraden im zweidimensionalen Raum, die durch den **0**-Punkt verläuft.

Beispiel 3.2.7. *Im obigen Beispiel 3.2.4 ist jeder Vektor aus*

$$g := \left\{ \mathbf{z} | \mathbf{z} = a \begin{pmatrix} 1 \\ 3 \end{pmatrix}, a \in \mathbb{R} \right\}$$

linear abhängig von $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$*, da sich jeder Vektor aus g als* $a \begin{pmatrix} 1 \\ 3 \end{pmatrix}$ *darstellen lässt.* ◊

Definition 3.2.8. *Lässt sich kein Vektor* $\mathbf{x}_i$ *aus* $(\mathbf{x}_1, ..., \mathbf{x}_n)$ *als Linearkombination von Vektoren* $\mathbf{x}_j \neq \mathbf{x}_i$ *aus* $(\mathbf{x}_1, ..., \mathbf{x}_n)$ *darstellen, nennen wir die Vektoren* $(\mathbf{x}_1, ..., \mathbf{x}_n)$ *linear unabhängig, sonst linear abhängig.* ◊

Der folgende Satz erweist sich für die Überprüfung auf lineare Unabhängigkeit oft als nützlich:

Satz 3.2.9. *Die von* **0** *verschiedenen Vektoren* $(\mathbf{x}_1, ..., \mathbf{x}_n)$ *sind linear unabhängig genau dann, wenn - wenn* $\sum_{i=1}^{n} a_i \mathbf{x}_i = \mathbf{0}$*, dann* $a_i = 0$ *für alle* $i \in \mathbb{N}_n^*$*.*

Beweis. Der Satz ist äquvalent mit: Die von **0** verschiedenen Vektoren $(\mathbf{x}_1, ..., \mathbf{x}_n)$ sind linear abhängig genau dann, wenn $\sum_{i=1}^{n} a_i \mathbf{x}_i = \mathbf{0}$ ohne dass $a_i = 0$, für alle $i \in \mathbb{N}_n^*$. Wir beweisen dieses Satz:
(i) Seien die Vektoren $(\mathbf{x}_1, ..., \mathbf{x}_n)$ linear abhängig. Dann gibt es ein $\mathbf{x}_i$ aus $(\mathbf{x}_1, ..., \mathbf{x}_n)$ mit $\mathbf{x}_j \neq \mathbf{x}_i$, so dass $\mathbf{x}_i = \sum_{j \in J}^{n} a_j \mathbf{x}_j$ mit $J \subset \mathbb{N}_n^* \backslash \{i\}$. Da $\mathbf{x}_i \neq \mathbf{0}$, gibt es mindestens ein a_j in $\sum_{j \in J}^{n} a_j \mathbf{x}_j$, so dass $a_j \neq 0$. Damit gilt $\sum_{j \in J}^{n} a_j \mathbf{x}_j - \mathbf{x}_i = \mathbf{0}$, ohne dass alle $a_j = 0$.
(ii) Sei $\sum_{i=1}^{n} a_i \mathbf{x}_i = \mathbf{0}$, ohne dass $a_i = 0$ für alle $i \in \mathbb{N}_n^*$. Wir wählen eines der $a_i \neq 0$. Damit gilt $\sum_{j=1;j \neq i}^{n} a_j \mathbf{x}_j + a_i \mathbf{x}_i = \mathbf{0}$, woraus folgt $\sum_{j=1;j \neq i}^{n} a_j \mathbf{x}_j = -a_i \mathbf{x}_i$ und $\sum_{j=1;j \neq i} -\frac{a_j}{a_i} \mathbf{x}_j = \mathbf{x}_i$, womit $\mathbf{x}_i$ linear abhängig ist von $(\mathbf{x}_1, ..., \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \mathbf{x}_n)$. □

Definition 3.2.10. *Gilt für einen Vektorraum V, dass es linear unabhängige Vektoren* $(\mathbf{v}_1, ..., \mathbf{v}_n)$

mit $\mathbf{v}_i \in V$ gibt, so dass für alle $\mathbf{x} \in V$ reelle Zahlen a_i gibt, derart dass

$$\mathbf{x} = \sum_{i=1}^{n} a_i \mathbf{v}_i,$$

nennt man $(\mathbf{v}_1, ..., \mathbf{v}_n)$ eine Basis von V. Man sagt, dass eine Basis einen Vektorraum „aufspanne“. ◊

Definition 3.2.11. *Die Anzahl Vektoren einer Basis $(v_1, ..., v_n)$, die einen Vektorraum V aufspannt, wird Dimension von V genannt. Wir schreiben auch* $\dim(V) = n$. ◊

Definition 3.2.12. *Ein n-Tupel von Vektoren $(v_1, ..., v_n)$ nennen wir Erzeugendensystem des Vektorraums V, der durch diese aufgespannt wird. Wir schreiben auch $V = \mathtt{span}(v_1, ..., v_n)$ oder $V = \mathtt{span}\{v_1, ..., v_n\}$.* ◊

Eine Basis eines Vektorraums V ist ein Erzeugendensystem dieses Vektorraums. Während eine Basis keine redundanten Vektoren enthält, können andere Erzeugendensysteme fürs Aufspannen des Vektorraums überflüssige Vektoren enthalten.

Beispiel 3.2.13. *In*

$$g = \{z | z = a\mathbf{x}, a \in \mathbb{R}\}$$

nennen wir $\mathbf{x} \in \mathbb{R}^n$ Basis des Vektorraums g. Im konkreten Fall

$$g := \left\{ \mathbf{z} | \mathbf{z} = a \begin{pmatrix} 1 \\ 3 \end{pmatrix}, a \in \mathbb{R} \right\}$$

wird $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$ die Basis des Vektorraums g genannt. Jeder andere Vektor von g - ausser dem Nullvektor - kann als Basis von g verwendet werden. ◊

Es gibt in jedem $n-$dimensionalen Raum, $n \geq 2$, überabzählbar unendlich viele Vektorräume der Form $g = \{\mathbf{z} | \mathbf{z} = a\mathbf{x}, a \in \mathbb{R}\}$ mit $\mathbf{x} \in \mathbb{R}^n$, da wir mit überabzählbar vielen reellen Zahlen mindestens ebenso viele Basen konstruieren können. Da ein Vektor als Basis genügt, spricht man von einem eindimensionalen Vektorraum.

Beispiel 3.2.14. *Der Vektor $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ liegt nicht im Vektorraum*

$$g = \left\{ z | z = a \begin{pmatrix} 3 \\ 1 \end{pmatrix}, a \in \mathbb{R} \right\},$$

denn man findet keine reelle Zahl a, so dass $\begin{pmatrix} 1 \\ 2 \end{pmatrix} = a \begin{pmatrix} 3 \\ 1 \end{pmatrix}$. Diese Gleichung führt nämlich auf einen Widerspruch, da dann gilt:

$$a = \frac{1}{3}$$
$$a = \frac{1}{2}$$

und damit

$$\frac{1}{3} = \frac{1}{2}.$$

In der Mathematik gilt jedoch

$$\frac{1}{3} \neq \frac{1}{2}.$$

Entsprechend ist $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ *von* $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$ *linear unabhängig. Bilden wir die Addition*

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \end{pmatrix},$$

entsteht entsprechend ein Vektor, der nicht auf g liegt (s. Abbildung 3.2.4).

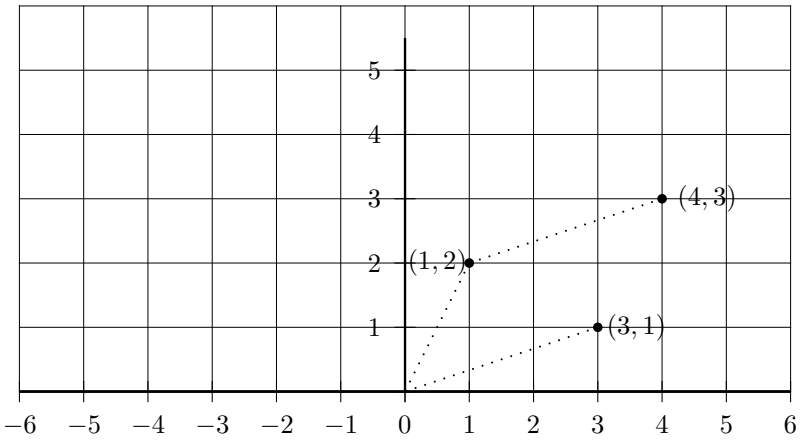


Abbildung 3.2.4: Addition zweier linear unabhängiger Punkte.

Durch

$$a\begin{pmatrix} 1 \\ 2 \end{pmatrix} + b\begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

können aber alle Punkte des $\mathbb{R}^2$ *ausgedrückt werden, d.h. alle Punkte des* $\mathbb{R}^2$ *sind von* $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ *und* $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$ *linear abhängig. Will man z.B. a und b für den Punkt*

$$\begin{pmatrix} 1.5 \\ 3.2 \end{pmatrix}$$

bestimmen, wird man auf ein Gleichungssystem mit zwei Gleichungen und zwei Unbekannten geführt:

$$1.5 = 1a + 3b$$
$$3.2 = 2a + 1b$$

Es ergibt sich $a = 1.62, b = -0.04$ *und es gilt*

$$1.62\begin{pmatrix} 1 \\ 2 \end{pmatrix} - 0.04\begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 1.5 \\ 3.2 \end{pmatrix}.$$

Für beliebige $\begin{pmatrix} x \\ y \end{pmatrix}$ *lässt sich die Gleichung*

$$a\begin{pmatrix} 1 \\ 2 \end{pmatrix} - b\begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}$$

allgemein lösen, womit nachgewiesen ist, dass sich alle $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$ *so darstellen lassen. Dies lässt vermuten, dass* $\mathbb{R}^2$ *eine Vektorraum ist, den man mit der Basis* $\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \end{pmatrix}\right)$ *aufspannen kann.* ◊

Satz 3.2.15. $\mathbb{R}^n$ *ist ein Vektorraum. für* $n \geq 1$.

Beweis. (i) Sei $x \in \mathbb{R}^n$. Dann ist $ax = a \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} ax_1 \\ \vdots \\ ax_n \end{pmatrix} \in \mathbb{R}^n$.

(ii) Seien $x, y \in \mathbb{R}^n$. Dann ist $x + y = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_2 \end{pmatrix} \in \mathbb{R}^n$. □

Damit ist auch $\mathbb{R}^2$ ein Vektorraum. Da er durch zwei linear unabhängige Vektoren des $\mathbb{R}^2$ aufgespannt wird, sprechen wir von einem zweidimensionalen Vektorraum. Da es in $\mathbb{R}^2$ überabzählbar unendlich viele Möglichkeiten gibt, zwei linear unabhängige Vektoren zu finden, gibt es überabzählbar unendlich viele Basen von $\mathbb{R}^2$. Die Standardbasis des $\mathbb{R}^2$ sind die Vektoren $(1, 0)$ und $(0, 1)$. Sie sind linear unabhängig, da es keine relle Zahl a gibt, so dass

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} = a \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Alle Elemente von $\mathbb{R}^2$ lassen sich durch

$$a \begin{pmatrix} 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

ausdrücken. So lässt sich etwa $(1.5, 3.2)$ darstellen als:

$$1.5 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 3.2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1.5 \\ 3.2 \end{pmatrix}.$$

Die eindimensionalen Vektorräume für $\mathbf{x} \in \mathbb{R}^2$ mit $g := \{\mathbf{z} | \mathbf{z} = a\mathbf{x}, a \in \mathbb{R}\}$ sind offenbar Teilmengen des Vektorraums $\mathbb{R}^2$. Wir sagen, dass die g Untervektorräume von $\mathbb{R}^2$ sind. Die Vereinigungsmenge aller Untervektoräume g von $\mathbb{R}^2$ fällt mit $\mathbb{R}^2$ zusammen. Die Vereinigungsmenge zweier eindimensionaler Untervektorräume g und f, $g \neq f$, des $\mathbb{R}^2$ ist kein Vektorraum, da nicht alle Linearkombinationen von $\mathbf{x} \in f$ und $\mathbf{y} \in g$ zu $f \cup g$ gehören. Demgegenüber ist die Schnittmenge von Vektorräumen immer ein Vektorraum - die Menge, die nur den Nullpunkt enthält wird als Vektorraum betrachtet.

Als nächstes betrachten wir den $\mathbb{R}^3$, der gemäss Satz 3.2.15 ein Vektorraum ist. Der $\mathbb{R}^3$ wird durch eine Basis aus drei linear unabhängige Vektoren aufgespannt. Offenbar genügt ein Vektor nicht, um alle Vektoren aus $\mathbb{R}^3$ als Linearkombinationen auszudrücken. So gibt es z.B. kein von 0 verschiedenes $a \in \mathbb{R}$, so dass

$$a \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \\ 1 \end{pmatrix}.$$

Wir erhalten drei Gleichungen mit einer Unbekannten.

$$\begin{aligned} a &= 3 \\ a &= 4 \\ 2a &= 1 \end{aligned}$$

Das Gleichungssystem führt auf $a = 3$ und $a = 4$, d.h. $3 = 4$, was im Widerspruch zu $3 \neq 4$ steht. $(3, 4, 1)$ ist von $(1, 1, 2)$ linear unabhängig und liegt nicht in dem von $(1, 1, 2)$ aufgespannten Vektorraum.

Ebenso wenig genügen zwei linear unabhängige Vektoren, um den $\mathbb{R}^3$ aufzuspannen, d.h. um alle Element von $\mathbb{R}^3$ als Linearkombinationen auszudrücken. So gibt es z.B. keine reelle von 0 verschiedene a und b so dass

$$a \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + b \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ 4 \\ 6 \end{pmatrix}.$$

Wir erhalten drei Gleichungen mit zwei Unbekannten.

$$\begin{aligned} a + 3b &= 5 \\ 2a + 2b &= 4 \\ 3a + b &= 6 \end{aligned}$$

Die Lösung führt auf einen Widerspruch, denn

$$\begin{aligned} a + 3b &= 5 \\ 2a + 2b &= 4 \end{aligned}$$

führt auf die Lösung $a = \frac{1}{2}, b = \frac{3}{2}$. Eingesetzt in die dritte Gleichung ergibt sich

$$3\frac{1}{2} + \frac{3}{2} = 6,$$

wobei $3\frac{1}{2} + \frac{3}{2} = 5$. Auch hier gilt wiederum, dass $(5, 4, 6)$ linear unabhängig von $(1, 2, 3)$ und $(3, 2, 1)$ ist.
Um den $\mathbb{R}^3$ aufzuspannen brauchen wir entsprechend eine Basis aus drei linear unabhängigen Vektoren. So ist z.B.

$$\left(\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 5 \\ 4 \\ 6 \end{pmatrix} \right)$$

eine Basis des $\mathbb{R}^3$. Jeder Vektor des $\mathbb{R}^3$ kann durch eine Linearkombination dieser Vektoren ausgedrückt werden. Die Berechnung der Koeffizienten erfolgt über ein Gleichungssystem mit drei Unbekannten. So führt z.B.

$$a \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + b \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} + c \begin{pmatrix} 5 \\ 4 \\ 6 \end{pmatrix} = \begin{pmatrix} 4 \\ 7 \\ 1 \end{pmatrix}$$

auf

$$\begin{aligned} a + 3b + 5c &= 4 \\ 2a + 2b + 4c &= 7 \\ 3a + b + 6c &= 1 \end{aligned}$$

mit der Lösung $a = \frac{19}{4}, b = \frac{19}{4}, c = -3$. Allgemein kann man zeigen, dass diese drei Vektoren genügen, um alle Vektoren aus $\mathbb{R}^3$ als Linearkombinationen darzustellen, indem man die folgende Gleichung allgemein löst:

$$a \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + b \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} + c \begin{pmatrix} 5 \\ 4 \\ 6 \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

Da wir drei Vektoren für eine Basis des $\mathbb{R}^3$ brauchen, sprechen wir von einem dreidimensionalen Vektorraum. Es gibt überabzählbar unendliche viele Basen des $\mathbb{R}^3$. Die Standardbasis ist

$$\left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right)$$

wodurch sich alle Vektoren des $\mathbb{R}^3$ als Linearkombinationen dieser Vektoren darstellen lassen.
Die eindimensionalen Untervektorräume des $\mathbb{R}^3$ für jedes $\mathbf{x} \in \mathbb{R}^3$ sind die Geraden

$$g = \{\mathbf{z} | \mathbf{z} = a\mathbf{x}, a \in \mathbb{R}\}.$$

Es handelt sich um die Untervektorräume des $\mathbb{R}^3$, die von einem Vektor aufgespannt werden - s. Abbildung 3.2.5 für

$$g = \left\{ \mathbf{z} | \mathbf{z} = a \begin{pmatrix} 1 \\ 3 \\ 4 \end{pmatrix}, a \in \mathbb{R} \right\}.$$

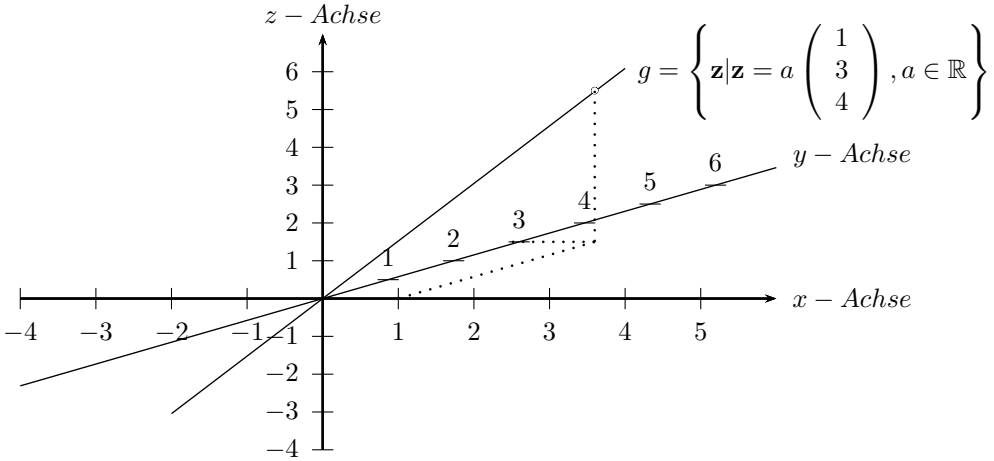


Abbildung 3.2.5: Beispiel für einen eindimensionalen Untervektorraum des dreidimensionalen Vektorraums, der durch die Basis $(1, 3, 4)$ aufgespannt wird.

Zudem können wir zweidimensionale Untervektorräume des $\mathbb{R}^3$ betrachten. Mit zwei linear unabhängigen Vektoren $\mathbf{x}, \mathbf{y}$ des $\mathbb{R}^3$ wird jeweils ein zweidimensionaler Untervektorraum des $\mathbb{R}^3$ aufgespannt. Es handelt sich um Ebenen E im dreidimensionalen Raum mit $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$; $\mathbf{x}, \mathbf{y}$ linear unabhängig,

$$E := \{\mathbf{z} | \mathbf{z} = a\mathbf{x} + b\mathbf{y}\},$$

die durch den $\mathbf{0}$-Punkt des Koordinatensystems verlaufen - s. Abbildung 3.2.6 mit den linear unabhängigen Vektoren

$$\begin{pmatrix} 4 \\ 5 \\ 220 \end{pmatrix}, \begin{pmatrix} 7 \\ 6 \\ 330 \end{pmatrix}$$

- der Ansatz $a\,(4, 5, 220) = (7, 6, 330)$ führt auf $a = \frac{7}{4} = \frac{6}{5}$ und damit zum Widerspruch mit $\frac{7}{4} \neq \frac{6}{5}$.
Im $\mathbb{R}^3$ gibt es überabzählbar unendliche viele eindimensionale wie zweidimensionale Unterräume. $\mathbb{R}^2$ ist einer dieser zweidimensionalen Untervektorräume von $\mathbb{R}^3$, wenn wir $\mathbb{R}^2$ im $\mathbb{R}^3$ in $\mathbb{R}^2 \times \{0\}$ übergehen lassen.
Die obigen Beispiele führen zur Vermutung, dass der Vektorraum $\mathbb{R}^n$ durch genau n unabhängige Vektoren aus $\mathbb{R}^n$ aufgespannt wird.

Satz 3.2.16. *Sei $(\mathbf{v}_1, ..., \mathbf{v}_n)$ ein n-Tupel von Vektoren $\mathbf{v}_i$ des Vektorraums V, dann sind die folgenden Aussagen äquivalent:*
(1) $(\mathbf{v}_1, ..., \mathbf{v}_n)$ ist eine Basis von V.
(2) $(\mathbf{v}_1, ..., \mathbf{v}_n)$ ist ein Erzeugendensystem von V und die linear unabhängigen Vektoren $(\mathbf{v}_1, ..., \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \mathbf{v}_n)$ bilden kein Erzeugendensystem von V.
(3) Zu jedem $\mathbf{v} \in V$ gibt es eindeutig bestimmte a_i, so dass $\mathbf{v} = \sum_{i=1}^{n} a_i \mathbf{v}_i$ mit linear unabängigen $(\mathbf{v}_1, ..., \mathbf{v}_n)$.
(4) $(\mathbf{v}_1, ..., \mathbf{v}_n, \mathbf{v}_{n+1})$ mit linear unabhängigen $(\mathbf{v}_1, ..., \mathbf{v}_n)$ ist ein Erzeugendensystem von V, ohne eine Basis von V zu sein.

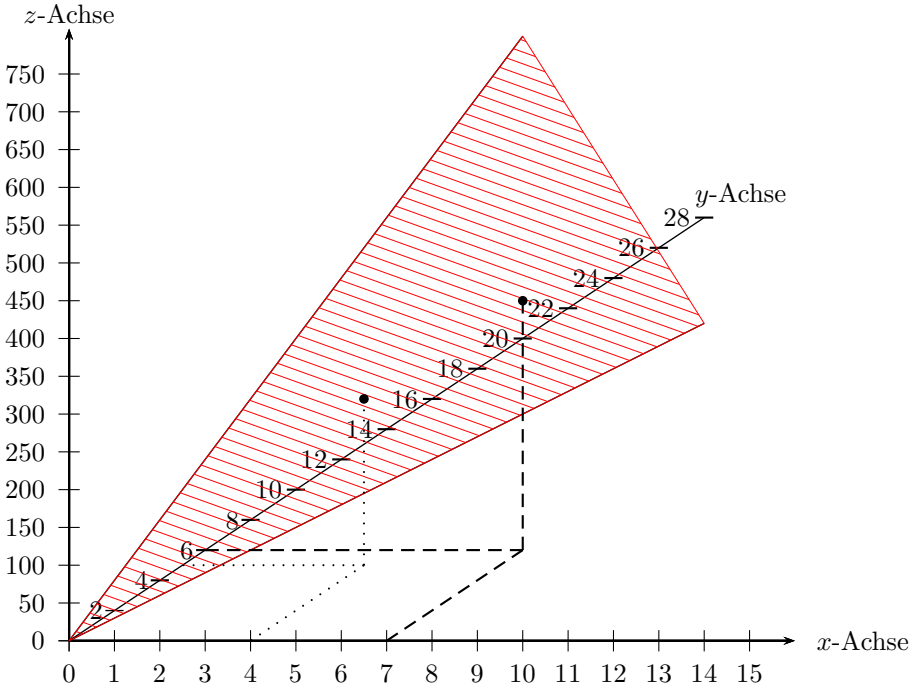


Abbildung 3.2.6: Beispiel für den zweidimensionalen Untervektorraum des $\mathbb{R}^3$, der durch $(4, 5, 220)$ und $(7, 6, 330)$ aufgespannt wird. In der Graphik wird nur der Auschnitt des Untervektorraums gezeigt, der Tripel mit positiven Komponenten enthält.

Beweis. (1) $\Longrightarrow$ (2) : Sei $(\mathbf{v}_1, ..., \mathbf{v}_n)$ eine Basis von V. Damit ist $(\mathbf{v}_1, ..., \mathbf{v}_n)$ ein Erzeugendensystem von V. Zudem lässt sich $\mathbf{v}_k$ nicht als Linearkombination von $(\mathbf{v}_1, ..., \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \mathbf{v}_n)$ darstellen, da gemäss Definition der Basis $\mathbf{v}_k$ von $(\mathbf{v}_1, ..., \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \mathbf{v}_n)$ unabhängig ist. Es gäbe sonst ein $\mathbf{v} \in \mathbb{R}^n$, das nicht Linearkombination von $(\mathbf{v}_1, ..., \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \mathbf{v}_n)$ ist. Also sind die linear unabhängigen Vektoren $(\mathbf{v}_1, ..., \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \mathbf{v}_n)$ keine Erzeugendensystem von V.
(2) $\Longrightarrow$ (3) : Sei $(\mathbf{v}_1, ..., \mathbf{v}_n)$ ein Erzeugendensystem von V, ohne dass die linear unabhängigen Vektoren $(\mathbf{v}_1, ..., \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \mathbf{v}_n)$ eine Erzeugendensystem von V bilden. Entsprechend muss $\mathbf{v}_k$ von $(\mathbf{v}_1, ..., \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \mathbf{v}_n)$ unabhängig sein, damit $(\mathbf{v}_1, ..., \mathbf{v}_n)$ V aufspannen kann, und $(\mathbf{v}_1, ..., \mathbf{v}_n)$ bildet eine Basis von V. Gibt es nicht eine eindeutige Darstellung von $\mathbf{v}$, so gibt es für ein $\mathbf{v} \in \mathbb{R}^n$ Darstellungen $\mathbf{v} = \sum_{i=1}^n a_i \mathbf{v}_i$ und $\mathbf{v} = \sum_{i=1}^n b_i \mathbf{v}_i$ so dass $a_i \neq b_i$ für mindestens ein i. Durch Subtraktion erhält man $\mathbf{0} = \sum_{i=1}^n a_i \mathbf{v}_i - \sum_{i=1}^n b_i \mathbf{v}_i = \sum_{i=1}^n \mathbf{v}_i (a_i - b_i)$ und durch Division mit $b_j - a_j$ für ein j mit $a_j \neq b_j$ erhält man: $\sum_{i=1}^n \mathbf{v}_i \frac{(a_i - b_i)}{a_j - b_j} = \sum_{i=1; i \neq j}^n \mathbf{v}_i \frac{(a_i - b_i)}{a_j - b_j} + \mathbf{v}_j = 0$ und damit $\sum_{i=1; i \neq j}^n \mathbf{v}_i \frac{(a_i - b_i)}{a_j - b_j} = -\mathbf{v}_j$, d.h. v_j lässt sich als Linearkombination der übrigen Vektoren von $(\mathbf{v}_1, ..., \mathbf{v}_n)$ darstellen, so dass $(\mathbf{v}_1, ..., \mathbf{v}_n)$ nicht aus linear unabhängigen Vektoren bestünde, also keine Basis wäre. Unter der Voraussetzung gibt es also eine eindeutige Darstellung.
(3) $\Longrightarrow$ (4) Es gebe zu jedem $\mathbf{v} \in V$ eindeutig bestimmte a_i, so dass $\mathbf{v} = \sum_{i=1}^n a_i \mathbf{v}_i$ mit linear unabängigen $(\mathbf{v}_1, ..., \mathbf{v}_n)$. Damit ist $(\mathbf{v}_1, ..., \mathbf{v}_n)$ ein Erzeugendensystem von V
(4) $\Longrightarrow$ (1) Sei $(\mathbf{v}_1, ..., \mathbf{v}_n, \mathbf{v}_{n+1})$ mit linear unabhängigen $(\mathbf{v}_1, ..., \mathbf{v}_n)$ ein Erzeugendensystem von V, ohne eine Basis von V zu sein. Da $(\mathbf{v}_1, ..., \mathbf{v}_n, \mathbf{v}_{n+1})$ eine Erzeugendensystem von V ist, ohne eine Basis von V zu sein und die $(\mathbf{v}_1, ..., \mathbf{v}_n)$ unabhängig sind, lässt sich $\mathbf{v}_{n+1}$ als Linearkombination der $(\mathbf{v}_1, ..., \mathbf{v}_n)$ darstellen und $(\mathbf{v}_1, ..., \mathbf{v}_n)$ ist eine Basis von V, da $(\mathbf{v}_1, ..., \mathbf{v}_n)$ als Vektor von linear unabhängigen Vektoren keine überflüssigen Vektoren enthält. □

$\mathbb{R}^n$ enthält überabzählbar unendlich viele i dimensionale $(0 < i < n)$ Untervektorräume.
Stellen wir einen Vektor $\mathbf{x} \in V = \mathbb{R}^n$ als Linearkombination einer Basis $(\mathbf{v}_1, ..., \mathbf{v}_n)$ von V dar,

rechnen wir

$$\sum_{i=1}^{n} a_i \mathbf{v}_i = \mathbf{x}.$$

Dies kann man auch mit Hilfe der Matrixmultiplikation darstellen, indem man die Basisvektoren als Spaltenvektoren einer Matrix $\mathbf{V}$ auffasst und die Koeffizienten a_i zum Vektor $\mathbf{a}$ zusammenfasst. Damit gilt:

$$\mathbf{x} = \mathbf{V}\mathbf{a} = \sum_{i=1}^{n} a_i \mathbf{v}_i.$$

Die Matrix $\mathbf{V}$ repräsentiert eine lineare Abbbildung von $\mathbb{R}^n$ nach $\mathbb{R}^n$, so dass jedem Vektor $\mathbf{a} \in \mathbb{R}^n$ ein Vektor $\mathbf{x}$ zugeordnet wird. Umgekehrt erhält man durch $\mathbf{V}^{-1}\mathbf{x}$ die Koeffizienten $\mathbf{a}$, so dass $\mathbf{v}$ eine Linearkombination $\sum_{i=1}^{n} a_i \mathbf{v}_i$ ist. a_i sind die Koordinaten von $\mathbf{x}$ bezüglich der Basis $(\mathbf{v}_1, ..., \mathbf{v}_n)$.

Definition 3.2.17. *Sei $V := (\mathbf{v}_1, ..., \mathbf{v}_n)$ eine Basis. Dann ist dim$(V) = n$.* ◊

Sei $\mathbf{A}$ eine Matrix, die aus spaltenweise aus einer Basis von V besteht. Dann ist $\mathbf{rgA}$ =dim(V) (Der Rang von $\mathbf{A}$ ist mit der Dimension von V identisch).

3.3 Affine Unterräume

Betrachtet man

$$\left\{\mathbf{x} | \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \lambda \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right\}$$

mit $x_i, y_i, \lambda \in \mathbb{R}$ so erhält man eine Gerade durch den Punkt $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ mit der Steigung $\frac{y_2 - x_2}{y_1 - x_2}$. Sind $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ und $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ linear unabhängig, so handelt sich um eine Parallele zur Geraden, die durch

$$\left\{\mathbf{x} | \mathbf{x} = \lambda \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right\}$$

gegeben ist. Diese dazu parallele Gerade entspricht einer affinen Funktion. Wir nennen die Menge der Punkte einer solchen Geraden „affine Unterräume“, wobei affine Unterräume nicht Vektorräume sind. Analoge Überlegungen führen zu affinen Unterräumen des $\mathbb{R}^n$.

Beispiel 3.3.1.

$$\left\{\mathbf{x} | \mathbf{x} = \begin{pmatrix} 1 \\ 3 \end{pmatrix} + \lambda \begin{pmatrix} 2 \\ 1 \end{pmatrix}\right\}$$

(s. Abbildung 3.3.7)

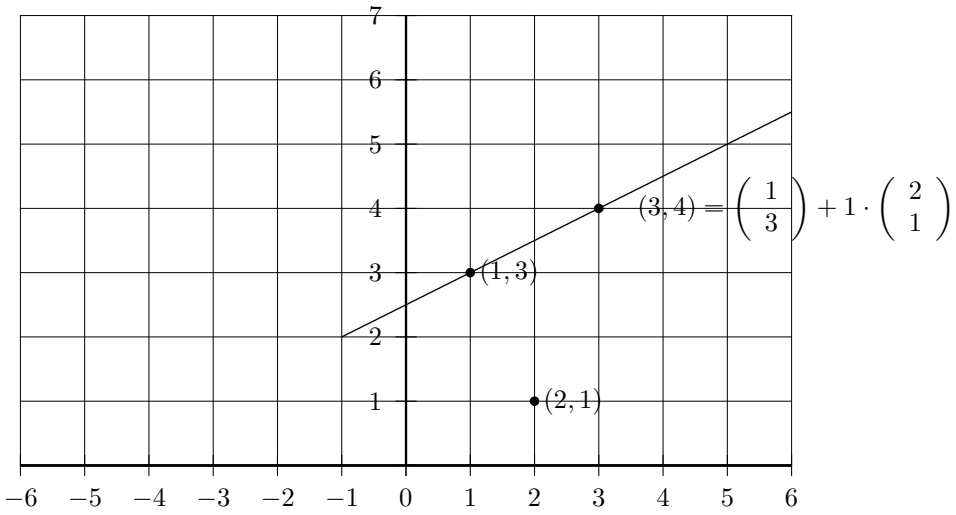


Abbildung 3.3.7: Graphische Darstellung des Beispiels 3.3.1

◊

Ein affiner Unterraum durch zwei Punkte $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ und $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ des $\mathbb{R}^2$ ist gegeben durch

$$\left\{\mathbf{x}|\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \lambda\left(\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right)\right\}$$

Beispiel 3.3.2.

$$\left\{\mathbf{x}|\mathbf{x} = \begin{pmatrix} 1 \\ 3 \end{pmatrix} + \lambda\left(\begin{pmatrix} 3 \\ 4 \end{pmatrix} - \begin{pmatrix} 1 \\ 3 \end{pmatrix}\right)\right\} = \left\{x|x = \begin{pmatrix} 1 \\ 3 \end{pmatrix} + \lambda\begin{pmatrix} 2 \\ 1 \end{pmatrix}\right\}$$

◊

Zwischen den zwei Punkten $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ und $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ des $\mathbb{R}^2$ ist die Strecke, die auf der entsprechenden Gerade liegt, gegeben durch

$$c\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + (1-c)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = c\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + (1-c)\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

mit $c \in [0, 1]$, denn es gilt

$$\begin{aligned} c\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + (1-c)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= c\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - c\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + c\left(\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right). \end{aligned}$$

Kapitel 4

Definition von linearen Abbildungen mittels Vektorraum-Basen

I:\dateien\Mathematik\111Math\Lineare_Algebra\003Uebersicht_Basistausch.tex. Mit Hilfe von Matrizen können lineare Abbildungen definiert werden. Diese können aber auch mit Hilfe von Vektoren oder Basen festgelegt werden, wie in der Folge zu zeigen sein wird. Dabei taucht ein neuer Aspekt von Matrizen und deren Zusammenhang mit der Definition von linearen Abbildungen auf. In der Tat ist die Definition von $f(\mathbf{x})$ mittels $\mathbf{Ax}$ ein Spezialfall einer allgemeineren Definition von Abbildungen mittels Basen, wobei Matrizen als Hilfsmittel wieder auftauchen. In diesem Zusammenhang stellt sich auch die Frage, wie die Festlegung einer Abbildung mittels einer Basis in die Festlegung derselben Abbildung mittels einer anderen Basis umgewandelt werden kann (Basistausch).

4.1 Koordinaten- und Transformationsfunktionen

In der Folge verwenden wir die folgenden Abkürzungen: Sei $(\mathbf{v}_1, ..., \mathbf{v}_n)$ eine Basis $\mathcal{A}$ von V und $\mathbf{A}$ die Matrix, so dass $\mathbf{a}_j = \mathbf{v}_j$ (Spaltenvektoren); Sei $(\mathbf{w}_1, ..., \mathbf{w}_n)$ eine Basis $\mathcal{B}$ von V und $\mathbf{B}$ die Matrix, so dass $\mathbf{b}_j = \mathbf{w}_j$ ($\mathbf{a}_j$ und $\mathbf{b}_j$ sind die Spalten von $\mathbf{A}$ beziehungsweise $\mathbf{B}$). Damit kann man die folgenden Sätze festhalten:

Definition 4.1.1. *Sei K^n ein Körper (hier $K^n = \mathbb{R}^n$): , $\mathbf{e}_j^n$ die Einheitsvektoren von K^n und V ein Vektorraum mit der Basis $\mathcal{A} := (\mathbf{v}_1, ..., \mathbf{v}_n)$. Dann definiert man die lineare Abbildung:*

$$\Phi_{\mathcal{A}} : K^n \to V$$
$$\Phi_{\mathcal{A}}(\mathbf{e}_j^n) = \mathbf{v}_j$$

Die Hochindizes bei den Einheitsvektoren lassen wir weg, wenn keine Missverständnisse möglich sind. ◊

Satz 4.1.2. *$\Phi_{\mathcal{A}}$ ist eine Bijektion (in der linearen Algebra „Isomorphismus“ genannt).*

Beweis. Sei $\mathbf{x}_1 \neq \mathbf{x}_2, \mathbf{x}_1, \mathbf{x}_2 \in K^n$ und $\Phi_{\mathcal{A}}(\mathbf{x}_1) = \Phi_{\mathcal{A}}(\mathbf{x}_2)$. Damit gilt: $\Phi_{\mathcal{A}}(\mathbf{x}_1) = \Phi_{\mathcal{A}}\left(\sum_{i=1}^n x_{i1}\mathbf{e}_i\right) = \sum_{i=1}^n x_{i1}\Phi_{\mathcal{A}}(\mathbf{e}_i) = \sum_{i=1}^n x_{i1}\mathbf{v}_i$ und analog $\Phi_{\mathcal{A}}(\mathbf{x}_2) = \sum_{i=1}^n x_{i2}\mathbf{v}_i$ sowie
$\sum_{i=1}^n x_{i2}\mathbf{v}_i - \sum_{i=1}^n x_{i1}\mathbf{v}_i = \sum_{i=1}^n (x_{i2} - x_{i1})\mathbf{v}_i = 0$. Da die $(\mathbf{v}_1, ..., \mathbf{v}_n)$ unabhängig sind, gilt dies nur, wenn sind $x_{i2} - x_{i1} = 0$ für alle i, dies im Widerspruch zur Voraussetzung $\mathbf{x}_1 \neq \mathbf{x}_2$. Zudem gilt, dass $\Phi_{\mathcal{B}}$ surjektiv ist: für alle $\mathbf{v} \in V$ gibt es jeweils ein $\mathbf{x} \in K^n$, so dass $\mathbf{v} = \sum_{i=1}^n x_i\mathbf{v}_i$, d.h. $\Phi_{\mathcal{A}}(\mathbf{x}) = \mathbf{v}$ (V ist ja ein Vektorraum, der von $(\mathbf{v}_1, ..., \mathbf{v}_n)$ aufgespannt wird.. □

Definition 4.1.3. *Sei* $(\mathbf{v}_1, ..., \mathbf{v}_n)$ *eine Basis von* V *und* $\mathbf{v} \in V$*. Dann nennt man die* x_i *in* $\mathbf{v} = \sum_{i=1}^{n} x_i \mathbf{v}_i$ *die Koordinaten von* $\mathbf{v}$ *bezüglich der Basis* $(\mathbf{v}_1, ..., \mathbf{v}_n)$*.* ◊

Bemerkung 4.1.4. $\Phi_{\mathcal{B}}$ *wird „Koordinatenfunktion" genannt. Sie ordnet jedem* $\mathbf{x}$ *einen Vektor* $\mathbf{v}$ *zu, wobei* $\mathbf{x}$ *die Koordinaten des Vektors* $\mathbf{v}$ *bezüglich der Basis* $(\mathbf{v}_1, ..., \mathbf{v}_n)$ *sind. Es gilt für* $\mathbf{v} := \sum_{i=1}^{n} x_i \mathbf{v}_i$ *nämlich*

$$\Phi_{\mathcal{A}}(\mathbf{x}) = \Phi_{\mathcal{A}}(\sum_{i=1}^{n} x_i \mathbf{e}_i) = \sum_{i=1}^{n} x_i \Phi_{\mathcal{A}}(\mathbf{e}_i) = \sum_{i=1}^{n} x_i \mathbf{v}_i = \mathbf{v},$$

was den Namen „Koordinatenfunktion" erklärt. Umgekehrt werden durch $\Phi_{\mathcal{A}}^{-1}$ *jedem Vektor* $\mathbf{v}$ *seine Koordinaten* $\mathbf{x}$ *bezüglich der Basis* $(\mathbf{v}_1, ..., \mathbf{v}_n)$ *zugeordnet.* ◊

Satz 4.1.5. *Sei* $\mathbf{A}$ *die Matrix mit den Basisvektoren von* $\mathcal{A} = (\mathbf{v}_1, ..., \mathbf{v}_n)$ *als Spalten. Dann gilt* $\Phi_{\mathcal{A}}(\mathbf{x}) = \mathbf{A}\mathbf{x} = \mathbf{v}$ *und* $\mathbf{A}^{-1}\mathbf{v} = \mathbf{x}$

Beweis.

$$\begin{aligned}
\Phi_{\mathcal{A}}(\mathbf{x}) &= \Phi_{\mathcal{A}}(\sum_{i=1}^{n} x_i \mathbf{e}_i) \\
&= \sum_{i=1}^{n} x_i \Phi_{\mathcal{A}}(\mathbf{e}_i) \\
&= \sum_{i=1}^{n} x_i \mathbf{v}_i \\
&= \mathbf{A}\mathbf{x}
\end{aligned}$$

□

Bemerkung 4.1.6. *Es gilt*

$$\mathbf{A}\mathbf{x} = \begin{pmatrix} \sum_{j=1}^{n} a_{1j} x_j \\ \vdots \\ \sum_{j=1}^{n} a_{nj} x_j \end{pmatrix} = \sum_{j=1}^{n} \begin{pmatrix} a_{1j} x_j \\ \vdots \\ a_{nj} x_j \end{pmatrix} = \sum_{j=1}^{n} x_j \begin{pmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{pmatrix} = \sum_{j=1}^{n} x_j \mathbf{a}_j.$$

Die Multiplikation einer Matrix $\mathbf{A}$ *mit einem Vektor* $\mathbf{x}$ *ist also dasselbe wie die Multiplikation der* j*-ten Spalte von* $\mathbf{A}$ *mit der* j*-ten Komponente von* $\mathbf{x}$ *und der Summe dieser Vektoren. Konkretes Beispiel:*

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{pmatrix} = x_1 \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix}$$

Analog kann man für Matrizen überlegen:

$$\begin{aligned}
\mathbf{AB} &= \begin{pmatrix} \sum_{j=1}^{n} a_{1j} b_{j1} & \cdots & \sum_{j=1}^{n} a_{1j} b_{jn} \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^{n} a_{nj} b_{j1} & \cdots & \sum_{j=1}^{n} a_{nj} b_{jn} \end{pmatrix} = \left(\sum_{j=1}^{n} b_{j1} \begin{pmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{pmatrix}, \cdots, \sum_{j=1}^{n} b_{jn} \begin{pmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{pmatrix} \right) \\
&= \left(\sum_{j=1}^{n} b_{j1} \mathbf{a}_j, ..., \sum_{j=1}^{n} b_{jn} \mathbf{a}_j \right)
\end{aligned}$$

Konkretes Beispiel:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11}+a_{12}b_{21} & a_{11}b_{12}+a_{12}b_{22} \\ a_{21}b_{11}+a_{22}b_{21} & a_{21}b_{12}+a_{22}b_{22} \end{pmatrix}$$
$$= \left(b_{11}\begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} + b_{21}\begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix}, b_{12}\begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} + b_{22}\begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix}\right)$$

◊

Beispiel 4.1.7. *Sei*

$$\mathcal{A} := \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 \\ 3 \end{pmatrix}\right); \qquad \mathbf{A} := \begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix}$$
$$\Phi_{\mathcal{A}}\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) := \begin{pmatrix} 1 \\ 2 \end{pmatrix}; \quad \Phi_{\mathcal{A}}\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) := \begin{pmatrix} 4 \\ 3 \end{pmatrix}$$

Damit ist z.B.

$$\Phi_{\mathcal{A}}\left(\begin{pmatrix} 8 \\ 9 \end{pmatrix}\right) = \Phi_{\mathcal{A}}\left(8\begin{pmatrix} 1 \\ 0 \end{pmatrix} + 9\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) = 8\left(\Phi_{\mathcal{A}}\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) + 9\Phi_{\mathcal{A}}\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right)$$
$$= 8\begin{pmatrix} 1 \\ 2 \end{pmatrix} + 9\begin{pmatrix} 4 \\ 3 \end{pmatrix} = \begin{pmatrix} 44 \\ 43 \end{pmatrix}$$

Da $\Phi_{\mathcal{A}}$ bijektiv ist, existiert $\Phi_{\mathcal{A}}^{-1}$ und es gilt $\Phi_{\mathcal{A}}^{-1}\left(\begin{pmatrix} 44 \\ 43 \end{pmatrix}\right) = \begin{pmatrix} 8 \\ 9 \end{pmatrix}$, wobei 8 und 9 die Koordinaten von $\begin{pmatrix} 44 \\ 43 \end{pmatrix}$ bezüglich der Basis $\mathcal{A}$ sind.
Bezüglich Matrixdarstellung:

$$\mathbf{Ax} = \begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix}\begin{pmatrix} 8 \\ 9 \end{pmatrix} = \begin{pmatrix} 44 \\ 43 \end{pmatrix}$$

und

$$\mathbf{A}^{-1}\mathbf{v} = \begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix}^{-1}\begin{pmatrix} 44 \\ 43 \end{pmatrix} = \begin{pmatrix} 8 \\ 9 \end{pmatrix}$$

◊

Bemerkung 4.1.8. *Sei $\mathbf{A}$ die Matrix mit den Basisvektoren von $\mathcal{A}$ als Spalten und $\mathcal{A}$ bestehe aus den aufsteigend angeordneten Standardbasisvektoren $\mathbf{e}_i^n$. Dann gilt mit Satz 4.1.5 unmittelbar $\mathbf{A} = \mathbf{E}$ und damit $\Phi_{\mathcal{E}}(\mathbf{x}) = \mathbf{Ex} = \mathbf{x}$ und $\mathbf{E}^{-1}\mathbf{x} = \mathbf{x}$. Φ ist in diesem Fall die Identitätsabbildung id_V - definiert mittels $id_v(x) = x$ für alle $x \in span(\mathcal{A})$.* ◊

Definition 4.1.9. *(Identitäsabbildung)*

$$id_V : V \to V$$
$$id_V(x) := x$$

◊

Definition 4.1.10. *Seien $\mathcal{A}$ und $\mathcal{B}$ Basen des Vektorraums V und $\Phi_{\mathcal{A}}$ und $\Phi_{\mathcal{B}}$ die entsprechenden Koordinatenfunktionen. Dann definieren wir:*

$$T_{\mathcal{B}}^{\mathcal{A}} := \Phi_{\mathcal{B}}^{-1} \circ \Phi_{\mathcal{A}}$$

◊

Wir nennen $T_{\mathcal{B}}^{\mathcal{A}}$ die (Koordinaten-)Transformationsfunktion. Sie ist ein Isomorphismus und ordnet den Koordinaten von $\mathbf{v}$ bezüglich $\mathcal{A}$ die Koordinaten von $\mathbf{v}$ bezüglich $\mathcal{B}$ zu.

$$\begin{array}{ccc} & K^n & \\ & & \searrow^{\Phi_{\mathcal{A}}} \\ \Phi_{\mathcal{B}}^{-1} \circ \Phi_{\mathcal{A}} & \downarrow & V \\ & & \nearrow_{\Phi_{\mathcal{B}}} \\ & K^n & \end{array}$$

Es gilt nämlich mit $\Phi_{\mathcal{A}}(\mathbf{x}) = \mathbf{v}$, $\Phi_{\mathcal{B}}(\mathbf{y}) = \mathbf{v}$, $\mathcal{B} = (\mathbf{w}_1, ..., \mathbf{w}_n), \mathcal{A} = (\mathbf{v}_1, ..., \mathbf{v}_n)$

$$\Phi_{\mathcal{B}}^{-1} \circ \Phi_{\mathcal{A}}(\mathbf{x}) = \Phi_{\mathcal{B}}^{-1}(\mathbf{v}) = \Phi_{\mathcal{B}}^{-1}\left(\sum_{i=1}^{n} y_i \mathbf{w}_i\right) = \sum_{i=1}^{n} y_i \Phi_{\mathcal{B}}^{-1}(\mathbf{w}_i) = \sum_{i=1}^{n} y_i \mathbf{e}_i = \mathbf{y}.$$

Satz 4.1.11. *Sei $(\mathbf{v}_1, ..., \mathbf{v}_n)$ eine Basis $\mathcal{A}$ von V und $\mathbf{A}$ die Matrix, so dass $\mathbf{a}_j = \mathbf{v}_j$. Sei $(\mathbf{w}_1, ..., \mathbf{w}_n)$ eine Basis $\mathcal{B}$ von V und $\mathbf{B}$ die Matrix, so dass $\mathbf{b}_j = \mathbf{w}_j$. Dann gilt:*

$$T_{\mathcal{B}}^{\mathcal{A}}(\mathbf{x}) = \mathbf{B}^{-1}\mathbf{A}\mathbf{x}$$

Beweis.

$$\begin{aligned} T_{\mathcal{B}}^{\mathcal{A}}(\mathbf{x}) &= \Phi_{\mathcal{B}}^{-1} \circ \Phi_{\mathcal{A}}(\mathbf{x}) = \\ &= \mathbf{B}^{-1}\mathbf{A}\mathbf{x} \end{aligned}$$

□

Beispiel 4.1.12. *Sei*

$$\begin{aligned} \mathcal{A} &:= \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 \\ 3 \end{pmatrix}\right), \mathbf{A} = \begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix} \\ \mathcal{B} &:= \left(\begin{pmatrix} 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 5 \\ 8 \end{pmatrix}\right), \mathbf{B} = \begin{pmatrix} 3 & 5 \\ 1 & 8 \end{pmatrix} \\ \mathbf{x} &= \begin{pmatrix} 18 \\ 20 \end{pmatrix} \end{aligned}$$

Dann ist

$$\begin{aligned} \Phi_{\mathcal{A}}\begin{pmatrix} 18 \\ 20 \end{pmatrix} &= \Phi_{\mathcal{A}}\left(18\begin{pmatrix} 1 \\ 0 \end{pmatrix} + 20\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) \\ &= 18\Phi_{\mathcal{A}}\begin{pmatrix} 1 \\ 0 \end{pmatrix} + 20\Phi_{\mathcal{A}}\begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &= 18\begin{pmatrix} 1 \\ 2 \end{pmatrix} + 20\begin{pmatrix} 4 \\ 3 \end{pmatrix} = \begin{pmatrix} 98 \\ 96 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix}\begin{pmatrix} 18 \\ 20 \end{pmatrix} = \mathbf{A}\mathbf{x} \end{aligned}$$

Zur Berechnung von $\Phi_{\mathcal{B}}^{-1}\left(\begin{pmatrix} 98 \\ 96 \end{pmatrix}\right)$ drückt man $\begin{pmatrix} 98 \\ 96 \end{pmatrix}$ durch die Basis $\mathcal{B}$ aus:

$$\begin{pmatrix} 98 \\ 96 \end{pmatrix} = a\begin{pmatrix} 3 \\ 1 \end{pmatrix} + b\begin{pmatrix} 5 \\ 8 \end{pmatrix}$$

Dies führt auf das Gleichungssystem:

$$98 = 3a + 5b$$
$$96 = 1a + 8b$$

$a = 16, b = 10$ *oder mittels:*

$$\begin{pmatrix} 3 & 5 \\ 1 & 8 \end{pmatrix}^{-1} \begin{pmatrix} 98 \\ 96 \end{pmatrix} = \begin{pmatrix} 16 \\ 10 \end{pmatrix} = \mathbf{y}$$

Damit ist

$$\Phi_{\mathcal{B}}^{-1}\left(\begin{pmatrix} 98 \\ 96 \end{pmatrix}\right) = \begin{pmatrix} 16 \\ 10 \end{pmatrix}$$

oder mittels Matrizen:

$$\begin{pmatrix} 3 & 5 \\ 1 & 8 \end{pmatrix}^{-1} \begin{pmatrix} 98 \\ 96 \end{pmatrix} = \begin{pmatrix} 16 \\ 10 \end{pmatrix} = \mathbf{y}$$

Also ist

$$\Phi_{\mathcal{B}}^{-1} \circ \Phi_{\mathcal{A}} \begin{pmatrix} 18 \\ 20 \end{pmatrix} = \mathbf{B}^{-1}\mathbf{A} \begin{pmatrix} 18 \\ 20 \end{pmatrix} = \begin{pmatrix} 16 \\ 10 \end{pmatrix} = \mathbf{y}.$$

◊

Bemerkung 4.1.13. *Statt $\Phi_{\mathcal{A}}$ wird in vielen Einführungen in die LA mit der Umkehrfunktion $\varphi_{\mathcal{A}} := \Phi_{\mathcal{A}}^{-1}$, die jedem Basisvektor von V einen Einheitsvektor von K^n zuordnet, gearbeitet.* ◊

Bemerkung 4.1.14. *Durch $\varphi_{\mathcal{A}}^{-1} \circ \varphi_{\mathcal{B}} = \Phi_{\mathcal{A}} \circ \Phi_{\mathcal{B}}^{-1}$ werden den Basisvektoren von $\mathcal{B}$ die Basisvektoren von $\mathcal{A}$ zugeordnet und durch $\varphi_{\mathcal{B}}^{-1} \circ \varphi_{\mathcal{A}} = \Phi_{\mathcal{B}} \circ \Phi_{\mathcal{A}}^{-1}$ werden den Basisvektoren von $\mathcal{A}$ die Basisvektoren von $\mathcal{B}$ zugeordnet.*

$$\begin{array}{ccc} & V & \\ & & \searrow^{\Phi_{\mathcal{A}}^{-1}} \\ \Phi_{\mathcal{B}} \circ \Phi_{\mathcal{A}}^{-1} & \downarrow & K^n \\ & & \nearrow_{\Phi_{\mathcal{B}}^{-1}} \\ & V & \end{array}$$

◊

Bemerkung 4.1.15. *Sind A und B die Einheitsmatrizen und damit $\Phi_{\mathcal{A}}$ und $\Phi_{\mathcal{B}}$ die Identitätsabbildungen id_V, gilt offensichtlich $\Phi_{\mathcal{B}}^{-1} \circ \Phi_{\mathcal{A}}(\mathbf{x}) = \mathbf{E}^{-1}\mathbf{E}\mathbf{x} = \mathbf{E}\mathbf{x} = \mathbf{x}$. Die Transformationsfunktion ist ind diesem Fall also ebenfalls die Identitätsabbildung id_V.* ◊

4.2 Definition von linearen Abbildungen mit Vektoren und Basen

Lineare Abbildungen von V nach W können mit Basen aus V und Vektoren aus W definiert werden. Zuerst werden einige Ergebnisse und Definitionen festgehalten, die in diesem Zusammenhang nützlich sind.

Satz 4.2.1. *Sei $(\mathbf{v}_1, ..., \mathbf{v}_n)$ eine Basis $\mathcal{A}$ von V, V ein Vektorraum, und $\mathbf{w}_i$ n Vektoren aus W.*

1. *Wird durch $f(v_i) = w_i$ eine Abbildung von V nach W festgelegt, so dass $f(\mathbf{v}) = \sum_{i=1}^{n} a_i \mathbf{w}_i$ für $\mathbf{v} \in V$, so ist f linear und eindeutig bestimmt.*

2. $\mathbf{w}_1, ..., \mathbf{w}_n$ *sind linear unabhängig genau dann, wenn* f *injektiv ist.*

3. $\mathtt{span}(\mathbf{w}_1, ..., \mathbf{w}_n) = W$ *genau dann, wenn* f *surjektiv ist.*

4. $(\mathbf{w}_1, ..., \mathbf{w}_n)$ *ist eine Basis von* W *genau dann, wenn* f *bijektiv ist.*

Beweis. 1. $K^n \to V$ die Koordinatenfunktion und $g : K^n \to W$ mit $g(\mathbf{x}) = \sum_{i=1}^n x_i \mathbf{w}_i$ (die $\mathbf{w}_i$ sind nicht unbedingt eine Basis von W). Für die Funktion $f := g \circ \Phi_{\mathcal{A}}^{-1}$ gilt damit $f : V \to W$ und

$$\begin{aligned} f(\mathbf{v}) = g \circ \Phi_{\mathcal{A}}^{-1}(\mathbf{v}) = g \circ \Phi_{\mathcal{A}}^{-1}\left(\sum_{i=1}^n x_i \mathbf{v}_i\right) = g\left(\sum_{i=1}^n x_i \Phi_{\mathcal{A}}^{-1}(\mathbf{v}_i)\right) \\ = g\left(\sum_{i=1}^n x_i \mathbf{e}_i\right) = g(\mathbf{x}) = \sum_{i=1}^n x_i \mathbf{w}_i. \end{aligned}$$

Zudem ist f linear, denn
(1)

$$f(a\mathbf{v}) = g \circ \Phi_{\mathcal{A}}^{-1}(a\mathbf{v}) = g(a\mathbf{x}) = \sum_{i=1}^n a x_i \mathbf{w}_i = a \sum_{i=1}^n x_i \mathbf{w}_i = a g(\mathbf{x}) = a\left(g \circ \Phi_{\mathcal{A}}^{-1}\right)(\mathbf{v}) = a f(\mathbf{v})$$

und
(2)

$$\begin{aligned} f(\mathbf{v}_1 + \mathbf{v}_2)\, g \circ \Phi_{\mathcal{A}}^{-1}(\mathbf{v}_1 + \mathbf{v}_2) &= g \circ \Phi_{\mathcal{A}}^{-1}\left(\sum_{i=1}^n x_{i1} \mathbf{v}_i + \sum_{i=1}^n x_{i2} \mathbf{v}_i\right) \\ &= g\left(\sum_{i=1}^n \left(x_{i1} \Phi_{\mathcal{A}}^{-1}(\mathbf{v}_i) + x_{i2} \Phi_{\mathcal{A}}^{-1}(\mathbf{v}_i)\right)\right) = g\left(\sum_{i=1}^n (x_{i1} \mathbf{e}_i + x_{i2} \mathbf{e}_i)\right) \\ &= g\left(\sum_{i=1}^n (x_{i1} + x_{i2}) \mathbf{e}_i\right) = g(\mathbf{x}_1 + \mathbf{x}_2) = \sum_{i=1}^n (x_{i1} + x_{i2}) \mathbf{w}_i \\ &= \sum_{i=1}^n x_{i1} \mathbf{w}_i + \sum_{i=1}^n x_{i2} \mathbf{w}_i = g(\mathbf{x}_1) + g(\mathbf{x}_2) = g \circ \Phi_{\mathcal{A}}^{-1}(\mathbf{v}_1) + g \circ \Phi_{\mathcal{A}}^{-1}(\mathbf{v}_2) \\ &= f(\mathbf{v}_1) + f(\mathbf{v}_2) \end{aligned}$$

Eindeutigkeit von f: Es ist $f(\mathbf{v}_j) = g \circ \Phi_{\mathcal{A}}^{-1}(\mathbf{v}_j) = g(\mathbf{e}_j) = \sum_{i=1}^n \delta_{ij} \mathbf{w}_j = \mathbf{w}_j$ für alle j. Damit gilt:

$$f \circ \Phi_{\mathcal{A}}(\mathbf{x}) = f\left(\sum_{i=1}^n x_i \mathbf{v}_i\right) = \sum_{i=1}^n x_i f(\mathbf{v}_i) = \sum_{i=1}^n x_i \mathbf{w}_i$$

für beliebige $\mathbf{x} \in K^n$. Also ist $g = f \circ \Phi_{\mathcal{A}}$ und damit $f = g \circ \Phi_{\mathcal{A}}^{-1}$.

2. Da $\Phi_{\mathcal{A}}$ injektiv ist , ist $g = f \circ \Phi_{\mathcal{A}}$ genau dann injektiv, wenn f injektiv ist. Nun sind die $(\mathbf{w}_1, ..., \mathbf{w}_n)$ genau dann unabhängig, wenn $\sum_{i=1}^n x_{i1} \mathbf{w}_i \neq \sum_{i=1}^n x_{i2} \mathbf{w}_i$ für $\mathbf{x}_1 \neq \mathbf{x}_2$ (s. Satz 3.2.16, S. 72). Damit gilt $g(\mathbf{x}_1) = \sum_{i=1}^n x_{i1} \mathbf{w}_i \neq g(\mathbf{x}_2) = \sum_{i=1}^n x_{i2} \mathbf{w}_i$ für $\mathbf{x}_1 \neq \mathbf{x}_2$. Da g injektiv ist, ist f injektiv.

3. Da $\Phi_{\mathcal{A}}$ surjektiv ist , ist $g = f \circ \Phi_{\mathcal{A}}$ genau dann surjektiv, wenn f surjektiv ist. Die Surjektivität von g bedeutet, dass $g(K^n) = \mathtt{span}(\mathbf{w}_1, ..., \mathbf{w}_n)$.

4. Wenn f bijektiv ist, werden den Basisvektoren $\mathbf{v}_i$ unterschiedliche, linear unabhängige Vektoren zugeordnet. Diese stellen eine Basis von W dar. Ist umgekehrt $(\mathbf{w}_1, ..., \mathbf{w}_n)$ eine Basis von W, sind die $\mathbf{w}_i$ unabhängig und f ist injektiv. Da zudem $\mathtt{span}(\mathbf{w}_1, ..., \mathbf{w}_n) = W$, ist f auch surjektiv.

□

Definition 4.2.2. *Sei f eine lineare Funktion von V nach W.*
$\mathit{Kern}(f) := \{\mathbf{v} \in V | f(\mathbf{v}) = \mathbf{0}\}$
$\mathit{Bild}(f) := \{\mathbf{w} \in W | f(\mathbf{v}) = \mathbf{w}\}$ ◊

Satz 4.2.3. *Sei f eine lineare Funktion von V nach W*

1. *Für einen beliebigen Untervektorraum W' von W is $f^{-1}(W')$ ein Untervektorraum von V.*
2. *$\mathit{Kern}(f)$ ist ein Untervektorraum von V.*
3. *f ist injektiv genau dann, wenn $\mathit{Kern}(f) = \{\mathbf{0}\}$*
4. *Ist f injektiv und sind die $\mathbf{v}_i, ..., \mathbf{v}_k \in V$ linear unabhängig, dann sind auch die $f(\mathbf{v}_1), ..., f(\mathbf{v}_k)$ linear unabhängig. Für injektives f ist damit $\dim(V) = \dim(\mathtt{Bild}(f))$.*
5. *Für einen beliebigen Untervektorraum V' von V ist $f(V')$ in Untervektorraum von W mit $\dim(f(V')) \leq \dim(V')$.*
6. *Ist $V = \mathit{Kern}(f) + V'$ mit einem Untervektorraum V' von V, so dass $\mathit{Kern}(f) \cap V' = \{\mathbf{0}\}$, dann ist $f : V' \to W$ injektiv. ($V' + V'' := \{\mathbf{v} : \mathbf{v} = \mathbf{v}' + \mathbf{v}'', \mathbf{v}' \in V', \mathbf{v}'' \in V''\}$, $V' + V''$ ist also die Menge aller Linearkombinationen von Vektoren aus V' und V'')*
7. *Für Vektorräume V mit endlicher Basis (d.h. $\dim(V) < \infty$) gilt: $\dim(V) = \dim(\mathtt{Kern}(f)) + \dim(\mathtt{Bild}(f))$.*
8. *Für Vektorräume V mit endlicher Basis gilt: wenn f injektiv ist, ist $\dim(W) \geq \dim(V)$.*
9. *Für Vektorräume V mit endlicher Basis gilt: wenn f surjektiv ist, ist $\dim(V) \geq \dim(W)$.*
10. *Für Vektorräume V mit endlicher Basis: Ist $\dim(V) = \dim(W)$, dann ist f injektiv, genau dann wenn f surjektiv ist.*

Beweis. 1. Sei $\mathbf{v}_1, \mathbf{v}_2 \in f^{-1}(W')$, d.h. $f(\mathbf{v}_1), f(\mathbf{v}_2) \in W'$. Da W' ein Vektorraum ist und f linear ist, ist auch $af(\mathbf{v}_1) + bf(\mathbf{v}_2) = f(a\mathbf{v}_1 + b\mathbf{v}_2) \in W'$, d.h. $a\mathbf{v}_1 + b\mathbf{v}_2 \in f^{-1}(W')$.

2. Seien $\mathbf{v}_1, \mathbf{v}_2 \in \mathtt{Kern}(f)$, d.h. $f(\mathbf{v}_1) = f(\mathbf{v}_2) = \mathbf{0}$. Damit gilt, da f linear ist, $f(a\mathbf{v}_1 + b\mathbf{v}_2) = af(\mathbf{v}_1) + bf(\mathbf{v}_2) = \mathbf{0}$, d.h. $a\mathbf{v}_1 + b\mathbf{v}_2 \in \mathtt{Kern}(f)$.

3. Sei $\mathtt{Kern}(f) \neq \{\mathbf{0}\}$, d.h. es gibt $\mathbf{v}_1, \mathbf{v}_2 \in V$, $\mathbf{v}_1 \neq \mathbf{v}_2$, so dass $f(\mathbf{v}_1) = f(\mathbf{v}_2) = \mathbf{0}$. Entsprechend ist f nicht injektiv. Sei f nicht injektiv. Dann gibt es $\mathbf{v}_1, \mathbf{v}_2 \in V$, $\mathbf{v}_1 \neq \mathbf{v}_2$, so dass $f(\mathbf{v}_1) = f(\mathbf{v}_2)$. Entsprechend ist $f(\mathbf{v}_1) - f(\mathbf{v}_2) = f(\mathbf{v}_1 - \mathbf{v}_2) = \mathbf{0}$ und damit $\mathbf{v}_1 - \mathbf{v}_2 \in \mathtt{Kern}(f)$, wobei $\mathbf{v}_1 - \mathbf{v}_2 \neq \mathbf{0}$.

4. Sei f injektiv und seien die $\mathbf{v}_i, ..., \mathbf{v}_m \in V$ linear unabhängig. Ist z.B. $f(\mathbf{v}_i)$ eine Linearkombination der $f(\mathbf{v}_j), i \neq j$, d.h. $f(\mathbf{v}_i) = \sum_{j=1;j\neq i}^{n} a_j f(\mathbf{v}_j)$, gilt $f(\mathbf{v}_i) - \sum_{j=1;j\neq i}^{n} a_j f(\mathbf{v}_j) = \mathbf{0}$, d.h. $\sum_{j=1}^{n} a_j f(\mathbf{v}_j) = \mathbf{0}$ mit $a_i = -1$. Damit gilt $\sum_{j=1}^{n} a_j f(\mathbf{v}_j) = f\left(\sum_{j=1}^{n} a_j \mathbf{v}_j\right) = \mathbf{0}$. Da die $\mathbf{v}_j$ eine Basis bilden und $a_i = -1$, ist $\sum_{j=1}^{n} a_j \mathbf{v}_j \neq \mathbf{0}$. Damit ist $\mathtt{Kern}(f) \neq \{\mathbf{0}\}$ und f ist nicht injektiv.

5. Sei $(\mathbf{v}_1, .., \mathbf{v}_k)$ eine Basis von V'. Dann sind die $f(\mathbf{v}_i)$ linear unabhängig. Ist nun $\dim(f(V')) > \dim(V')$, gibt es ein $\mathbf{w} \in f(V')$, so dass sich $\mathbf{w}$ nicht als Linearkombinantion von $(f(\mathbf{v}_1), ..., f(\mathbf{v}_k))$ darstellen lässt. Da $\mathbf{w} \in f(V')$, gibt es ein $\mathbf{v} \in V'$, mit $f(\mathbf{v}) = \mathbf{w}$, wobei $f(\mathbf{v}) = f(\sum_{i=1}^{p} a_i \mathbf{v}_i) = \sum_{i=1}^{p} a_i f(\mathbf{v}_i) = \sum_{i=1}^{p} a_i f(\mathbf{v}_i) = \mathbf{w}$. Damit lässt sich $\mathbf{w}$ als Linearkombination der $f(\mathbf{v}_i)$ darstellen.

6. Wäre f nicht injektiv, gäbe es $\mathbf{v}_1, \mathbf{v}_2 \in V'$, $\mathbf{v}_1 \neq \mathbf{v}_2$, so dass $f(\mathbf{v}_1) = f(\mathbf{v}_2)$. Da $\mathbf{v}_1 - \mathbf{v}_2 \in V'$, wäre $f(\mathbf{v}_1 - \mathbf{v}_2) = \mathbf{0}$, ohne dass $\mathbf{v}_1 = \mathbf{v}_2$, d.h. $\mathbf{0} \neq \mathbf{v}_1 - \mathbf{v}_2 \in \mathtt{Kern}(f)$ und damit $\mathtt{Kern}(f) \cap V' \neq \{\mathbf{0}\}$.

7. Ist $\dim(\mathtt{Kern}(f)) = m, 0 \leq n$, dann ist $\dim(V') = n - m$ für $V' + \mathtt{Kern}(f) = V$ und $V' \cap \mathtt{Kern}(f) = \{\mathbf{0}\}$. Da f auf V' bijektiv ist, ist $\dim(\mathtt{Bild}(f)) = n - m$.

8. Da f injektiv ist, wie bewiesen, $\dim(f(V)) = \dim(V)$. Da $\mathtt{Bild}(f)$ ein Untervektorraum von W ist, muss W mindestens die Dimension von $\mathtt{Bild}(f)$ aufweisen.

9. Sei $\dim(V) < \dim(W)$. Für $\mathbf{v} \in V$ gilt damit: $f(\mathbf{v}) = f(\sum_{i=1}^{m} x_i \mathbf{v}_i) = \sum_{i=1}^{m} x_i f(\mathbf{v}_i) = \sum_{i=1}^{m} x_i \mathbf{w}_i$ mit Vektoren $\mathbf{w}_i$ aus W. Damit sind die Funktionswerte von f Linearkombinationen von $1 \leq m < n$ Vektoren aus W, womit man nicht jedes $\mathbf{w} \in W$ darstellen kann. Somit wäre f nicht surjektiv.

10. Sei $\dim(V) = \dim(W)$ und f injektiv. Damit ist $\dim(f(V)) = \dim(V)$ und damit f surjektiv.
Ist umgekehrt f surjektiv. Da $\dim(V) = \dim(W)$ ist $\dim(\mathtt{Kern}(f)) = 0$ genau dann, wenn $\dim(\mathtt{Bild}(f)) = \dim(V) = \dim(W)$. Da damit $\dim(\mathtt{Kern}(f)) = 0$ ist, ist f injektiv.

□

Statt mit beliebigen Vektoren aus W, können wir eine Abbildung von V nach W auch mit Basen aus W definieren. In diesem Zusammenhang liefern wir auch eine allgemeinere Definition von Matrizen:

Definition 4.2.4. *Sei $(\mathbf{v}_1, ..., \mathbf{v}_n)$ eine Basis $\mathcal{A}$ von V.*
Sei $(\mathbf{w}_1, ..., \mathbf{w}_m)$ eine Basis $\mathcal{B}$ von W. Sei $f : V \to W$ ein Vektorraumhomomorphismus (so werden lineare Abbildungen in der linearen Algebra genannt).
$M_{\mathcal{B}}^{\mathcal{A}}(f) = \mathbf{C}$ *genau dann, wenn* $f(\mathbf{v}_j) = \sum_{i=1}^{m} c_{ij} \mathbf{w}_i$ *mit* $\mathbf{C} = (c_{ij})_{mn}$. ◊

Satz 4.2.5. *Ist $M_{\mathcal{B}}^{\mathcal{A}}(f) = \mathbf{C}$, dann ist $f(\mathbf{v}) = \mathbf{B}\mathbf{C}\mathbf{A}^{-1}\mathbf{v}$ mit den Matrizen $\mathbf{B} := (\mathbf{w}_1, ..., \mathbf{w}_m)$, $\mathbf{A} := (\mathbf{v}_1, ..., \mathbf{v}_n)$*

Beweis. Es sei $\mathbf{v} = \sum_{j=1}^{n} x_j \mathbf{v}_j$. Damit ist $f(\mathbf{v}) = f\left(\sum_{j=1}^{n} x_j \mathbf{v}_j\right) = \sum_{j=1}^{n} x_j f(\mathbf{v}_j) = \sum_{j=1}^{n} x_j \sum_{i=1}^{m} c_{ij} \mathbf{w}_i = \mathbf{B}\mathbf{C}\mathbf{A}^{-1}\mathbf{v}$, denn $\sum_{i=1}^{m} c_{ij} \mathbf{w}_i = \mathbf{B}\mathbf{c}_j$ (s. Bemerkung 4.1.6) und $\sum_{j=1}^{n} x_j (\mathbf{B}\mathbf{c}_j) = \mathbf{B}\mathbf{C}\mathbf{x}$ ($\mathbf{B}\mathbf{c}_j$ ist ein Vektor, der letzte Schritt gilt also wiederum mit Bemerkung 4.1.6, denn $(\mathbf{B}\mathbf{c}_1, ..., \mathbf{B}\mathbf{c}_n) = \mathbf{B}\mathbf{C}$). Mit $\Phi^{-1}(\mathbf{v}) = \mathbf{x} = \mathbf{A}^{-1}\mathbf{v}$ erhält man $f(\mathbf{v}) = \mathbf{B}\mathbf{C}\mathbf{A}^{-1}\mathbf{v}$. □

$\mathbf{B}\mathbf{C}$ bildet die Basen $\mathbf{v}_i$ von V in $f(\mathbf{v}_i)$ ab. Für die x_i in $f(\mathbf{v}) = \sum_{j=1}^{n} x_j f(\mathbf{v}_j)$ gilt zudem $\mathbf{A}^{-1}\mathbf{v}$.

Beispiel 4.2.6. $\mathbf{v}_1 := \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ *und* $\mathbf{v}_2 := \begin{pmatrix} 4 \\ 3 \end{pmatrix}$ *stellen eine Basis des* $\mathbb{R}^2$ *dar. Ebenso* $\mathbf{w}_1 :=$

$\begin{pmatrix} 5 \\ 8 \end{pmatrix}$ *und* $\mathbf{w}_2 := \begin{pmatrix} 7 \\ 12 \end{pmatrix}$. *Zudem sei* f *festgelegt durch*

$$f\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}\right) = \begin{pmatrix} 17 \\ 13 \end{pmatrix}$$
$$f\left(\begin{pmatrix} 4 \\ 3 \end{pmatrix}\right) = \begin{pmatrix} 6 \\ 9 \end{pmatrix}$$

Damit erhält man die Matrix

$$\mathbf{C} = \begin{pmatrix} \frac{113}{4} & \frac{9}{4} \\ -\frac{71}{4} & -\frac{3}{4} \end{pmatrix},$$

denn es gilt gemäss Definition von $\mathbf{C}$*:*

$$\begin{pmatrix} 17 \\ 13 \end{pmatrix} = c_{11}\begin{pmatrix} 5 \\ 8 \end{pmatrix} + c_{21}\begin{pmatrix} 7 \\ 12 \end{pmatrix}$$

und damit

$$17 = 5c_{11} + 7c_{21}$$
$$13 = 8c_{11} + 12c_{21}$$

womit man die Koeffizienten $c_{11} = \frac{113}{4}, c_{21} = -\frac{71}{4}$ *erhält sowie*

$$\begin{pmatrix} 6 \\ 9 \end{pmatrix} = c_{12}\begin{pmatrix} 5 \\ 8 \end{pmatrix} + c_{22}\begin{pmatrix} 7 \\ 12 \end{pmatrix}$$

$$6 = 5c_{12} + 7c_{22}$$
$$9 = 8c_{12} + 12c_{22}$$

und damit die Koeffizienten $c_{12} = \frac{9}{4}, c_{22} = -\frac{3}{4}$.
Es gilt mit $\mathbf{B} := (\mathbf{b}_1, \mathbf{b}_2)$*:*

$$\mathbf{BC} = (f(\mathbf{v}_1), f(\mathbf{v}_2))$$

denn $\mathbf{BC} = (c_{11}\mathbf{b}_1 + c_{21}\mathbf{b}_1, c_{21}\mathbf{b}_2 + c_{22}\mathbf{b}_2) = (f(\mathbf{v}_1), f(\mathbf{v}_2))$. *Am Beispiel:*

$$\mathbf{BC} = \begin{pmatrix} 5 & 7 \\ 8 & 12 \end{pmatrix}\begin{pmatrix} \frac{113}{4} & \frac{9}{4} \\ -\frac{71}{4} & -\frac{3}{4} \end{pmatrix} = \begin{pmatrix} 17 & 6 \\ 13 & 9 \end{pmatrix} = (f(v_1), f(v_2))$$

d.h. multipliziert man die Matrix mit der zweiten Basis als Spalten mit der Matrix $\mathbf{C}$ *der Koeffizienten, erhält man die Matrix, welche spaltenweise die Funktionswerte - der Basisvektoren der ersten Basis - der Abbildung* f *enthält.*

Wie erhält man nur mit Hilfe von $\mathbf{C}$ *die Funktionswerte beliebiger Vektoren, z.B. von* $\mathbf{v} = \begin{pmatrix} 6 \\ 8 \end{pmatrix}$*?*

Man kann $\begin{pmatrix} 6 \\ 8 \end{pmatrix}$ *mit Hilfe der Basisvektoren* $(\mathbf{v}_1, \mathbf{v}_2)$ *darstellen:*

$$\begin{pmatrix} 6 \\ 8 \end{pmatrix} = a\begin{pmatrix} 1 \\ 2 \end{pmatrix} + b\begin{pmatrix} 4 \\ 3 \end{pmatrix}$$

und damit

$$6 = a + 4b$$
$$8 = 2a + 3b$$

$a = \frac{14}{5}, b = \frac{4}{5}$, *wobei* a *und* b *die Koordinaten von* $\begin{pmatrix} 6 \\ 8 \end{pmatrix}$ *bezüglich* $(\mathbf{v}_1, \mathbf{v}_2)$ *sind. Da* $\mathbf{BC} = (f(\mathbf{v}_1), f(\mathbf{v}_2))$ *ist* $(f(a\mathbf{v}_1), f(b\mathbf{v}_2)) = (af(\mathbf{v}_1), bf(\mathbf{v}_2)) = \mathbf{BC}\begin{pmatrix} a \\ b \end{pmatrix} = \mathbf{BCA}^{-1}\mathbf{v}$, *d.h.*

$$f\left(\begin{pmatrix} 6 \\ 8 \end{pmatrix}\right) = \begin{pmatrix} 5 & 7 \\ 8 & 12 \end{pmatrix}\begin{pmatrix} \frac{113}{4} & \frac{9}{4} \\ -\frac{71}{4} & -\frac{3}{4} \end{pmatrix}\begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix}^{-1}\begin{pmatrix} 6 \\ 8 \end{pmatrix} = \begin{pmatrix} \frac{262}{5} \\ \frac{218}{5} \end{pmatrix} = \begin{pmatrix} 52.4 \\ 43.6 \end{pmatrix}$$

◊

Definition 4.2.7. *Sei* $(\mathbf{v}_1, ..., \mathbf{v}_n)$ *eine Basis* $\mathcal{A}$ *von* V.
Sei $f : V \to V$ *ein Endomorphismus (so werden Abbildung in der linearen Algebra genannt, bei denen der Definitions- und der Wertebereich identisch sind).*
$M_{\mathcal{A}}(f) = \mathbf{C}$ *genau dann, wenn* $f(\mathbf{v}_j) = \sum_{i=1}^{n} c_{ij}\mathbf{v}_i$ *mit* $\mathbf{C} = (c_{ij})_n$. ◊

Definition 4.2.8. *Bei den Basen* $\mathcal{E}_n$=$(\mathbf{e}_1, ..., \mathbf{e}_n)$ *und* $\mathcal{E}_m$=$(\mathbf{e}_1, ..., \mathbf{e}_m)$ *gilt definitorisch:*

$$M(f) := M_{\mathcal{E}_m}^{\mathcal{E}_n}(f)$$

und im Falle eines Endomorphismus

$$M(f) := M_{\mathcal{E}_n}(f)$$

◊

Satz 4.2.9. *Ist* $\mathcal{A}$ *die Standardbasis* $(\mathbf{e}_1, ..., \mathbf{e}_n)$ *von* $\mathbb{R}^n$ f *ein Endomorphismus mit* $M(f) = \mathbf{C}$, *dann ist* $f(x) = \mathbf{C}x$

Beweis. Der Beweis folgt direkt aus $f(\mathbf{x}) = \mathbf{BCA}^{-1}(\mathbf{x}) = \mathbf{ECE}^{-1}\mathbf{x} = \mathbf{Cx}$
Alternativer Beweis: Sei $\mathbf{x} = (x_1, ..., x_n)$

$$\begin{aligned} f(\mathbf{x}) &= f\left(\sum_{j=1}^{n} x_j \mathbf{e}_j^n\right) \\ &\sum_{j=1}^{n} x_j f(\mathbf{e}_j^n) \\ &= \sum_{j=1}^{n} x_j \sum_{i=1}^{m} c_{ij}\mathbf{e}_i^m \\ &= \sum_{i=1}^{m}\left(\sum_{j=1}^{n} c_{ij}x_j\right)\mathbf{e}_i^m \\ &= \mathbf{Cx}. \end{aligned}$$

□

Beispiel 4.2.10. *Seien die Standardbasen des* $\mathbb{R}^2$ *gegeben und* f *definiert durch*

$$f\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 4 \\ 8 \end{pmatrix}, f\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 12 \\ 3 \end{pmatrix}$$

Damit gilt

$$\begin{pmatrix} 4 \\ 8 \end{pmatrix} = 4\begin{pmatrix} 1 \\ 0 \end{pmatrix} + 8\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$
$$\begin{pmatrix} 12 \\ 3 \end{pmatrix} = 12\begin{pmatrix} 1 \\ 0 \end{pmatrix} + 3\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

d.h.

$$M(f) = M_{\mathcal{E}_2}^{\mathcal{E}_2}(f) = \begin{pmatrix} 4 & 12 \\ 8 & 3 \end{pmatrix}$$

Damit ist z.B

$$\begin{aligned} f\left(\begin{pmatrix} 8 \\ 9 \end{pmatrix}\right) &= f\left(8\begin{pmatrix} 1 \\ 0 \end{pmatrix} + 9\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) \\ &= 8f\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) + 9f\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) \\ &= 8\begin{pmatrix} 4 \\ 8 \end{pmatrix} + 9\begin{pmatrix} 12 \\ 3 \end{pmatrix} \\ &= \begin{pmatrix} 140 \\ 91 \end{pmatrix} \end{aligned}$$

sowie

$$\begin{pmatrix} 4 & 12 \\ 8 & 3 \end{pmatrix}\begin{pmatrix} 8 \\ 9 \end{pmatrix} = \begin{pmatrix} 140 \\ 91 \end{pmatrix}$$

d.h. $f(\mathbf{x}) = \mathbf{C}\mathbf{x}$. ◊

Satz 4.2.11. $M_{\mathcal{B}}^{\mathcal{A}}(f) = M\left(\Phi_{\mathcal{B}}^{-1} \circ f \circ \Phi_{\mathcal{A}}\right)$

Beweis. Sei $M_{\mathcal{B}}^{\mathcal{A}}(f) = (c_{ij})_{mn}$. Dann gilt:

$$\begin{aligned} \Phi_{\mathcal{B}}^{-1} \circ f \circ \Phi_{\mathcal{A}}\left(\mathbf{e}_j^n\right) &= \Phi_{\mathcal{B}}^{-1} \circ f(\mathbf{v}_j) \\ &= \Phi_{\mathcal{B}}^{-1}\left(\sum_{i=1}^{m} c_{ij}\mathbf{w}_i\right) \\ &= \sum_{i=1}^{m} c_{ij}\Phi_{\mathcal{B}}^{-1}(\mathbf{w}_i) \\ &= \sum_{i=1}^{m} c_{ij}\mathbf{e}_i^m \end{aligned}$$

□

Bemerkung 4.2.12. *$(c_{ij})_{mn}$ beschreibt also den Homomorphismus $\Phi_{\mathcal{B}}^{-1} \circ f \circ \Phi_{\mathcal{A}}$ bezüglich der Basen $(\mathbf{e}_1, ..., \mathbf{e}_n)$ und $(\mathbf{e}_1, ..., \mathbf{e}_m)$.* ◊

Satz 4.2.13. $M_{\mathcal{A}}(f) = M\left(\Phi_{\mathcal{A}}^{-1} \circ f \circ \Phi_{\mathcal{A}}\right)$

Beweis. Sei $M_{\mathcal{A}}(f) = (c_{ij})_{nn}$. Dann gilt:

$$\begin{aligned}\Phi_{\mathcal{A}}^{-1} \circ f \circ \Phi_{\mathcal{A}} \left(\mathbf{e}_j^n\right) &= \Phi_{\mathcal{A}}^{-1} \circ f \left(\mathbf{v}_j\right) \\ &= \Phi_{\mathcal{A}}^{-1} \left(\sum_{i=1}^{n} c_{ij} \mathbf{v}_i\right) \\ &= \sum_{i=1}^{n} c_{ij} \Phi_{\mathcal{A}}^{-1} \left(\mathbf{v}_i\right) \\ &= \sum_{i=1}^{m} c_{ij} \mathbf{e}_i^n\end{aligned}$$

□

Satz 4.2.14. *$M(g \circ f) := M\left(g\left(f\right)\right) = M(g) \cdot M(f)$ für $f : V \to U$ und $g : U \to W$ (Homomorphismen); $\dim\left(V\right) = n; \dim\left(U\right) = l; \dim\left(W\right) = m$.*

Beweis. Sei $M(g) = \left(b_{ik}\right)_{ml}$ und $M(f) = \left(c_{kj}\right)_{ln}$.

$$\begin{aligned}g \circ f(\mathbf{e}_j^n) &= g \left(\sum_{k=1}^{l} c_{kj} \mathbf{e}_k^l\right) \\ &= \sum_{k=1}^{l} c_{kj} g \left(\mathbf{e}_k^l\right) \\ &= \sum_{k=1}^{l} c_{kj} \sum_{i=1}^{m} b_{ik} \mathbf{e}_i^m \\ &= \sum_{i=1}^{m} \left(\sum_{k=1}^{l} c_{ik} a_{kj}\right) \mathbf{e}_i^m\end{aligned}$$

□

Satz 4.2.15. *Für $M_{\mathcal{C}}^{\mathcal{B}}(g)$ und $M_{\mathcal{B}}^{\mathcal{A}}(f)$ gilt: $M_{\mathcal{C}}^{\mathcal{A}}(g \circ f) = M_{\mathcal{C}}^{\mathcal{B}}(g) \cdot M_{\mathcal{B}}^{\mathcal{A}}(f)$ $\quad (M_{\mathcal{A}}(g \circ f) = M_{\mathcal{A}}(g) \cdot M_{\mathcal{A}}(f))$*

Beweis.

$$\begin{aligned}M_{\mathcal{C}}^{\mathcal{A}}(g \circ f) &= M\left(\Phi_{\mathcal{C}}^{-1} \circ g \circ f \circ \Phi_{\mathcal{A}}\right) \\ &= M\left(\Phi_{\mathcal{C}}^{-1} \circ g \circ \Phi_{\mathcal{B}} \circ \Phi_{\mathcal{B}}^{-1} \circ f \circ \Phi_{\mathcal{A}}\right) \\ &= M\left(\Phi_{\mathcal{C}}^{-1} \circ g \circ \Phi_{\mathcal{B}}\right) \cdot M\left(\Phi_{\mathcal{B}}^{-1} \circ f \circ \Phi_{\mathcal{A}}\right) \\ &= M_{\mathcal{C}}^{\mathcal{B}}(g) \cdot M_{\mathcal{B}}^{\mathcal{A}}(f)\end{aligned}$$

□

Satz 4.2.16. *$M\left(\Phi_{\mathcal{A}}\right) = \mathbf{A}$ mit $\mathbf{A} = (\mathbf{v}_1, ..., \mathbf{v}_n)$ (Matrix aus Basisvektoren von $\mathcal{A}$ zusammengesetzt)*

Beweis.

$$\begin{aligned}\Phi_{\mathcal{A}}(\mathbf{e}_j) &= \mathbf{v}_j \\ &= \sum_{i=1}^{n} v_{ij} \mathbf{e}_i^n\end{aligned}$$

□

Bemerkung 4.2.17. *Den Satz 4.2.5 kann man mit Hilfe dieser Ergebnisse auch wie folgt beweisen: Wegen* $M_{\mathcal{B}}^{\mathcal{A}}(f) = M\left(\Phi_{\mathcal{B}}^{-1} \circ f \circ \Phi_{\mathcal{A}}\right) = M\left(\Phi_{\mathcal{B}}^{-1}\right) \cdot M(f) \cdot M(\Phi_{\mathcal{A}})$ *und damit wegen der Injektivität von* $\Phi_{\mathcal{B}}$ *und* $\Phi_{\mathcal{A}}$ *gilt*

$$M(f) = M\left(\Phi_{\mathcal{B}}^{-1}\right)^{-1} M_{\mathcal{B}}^{\mathcal{A}}(f) M(\Phi_{\mathcal{A}})^{-1} = M(\Phi_{\mathcal{B}}) M_{\mathcal{B}}^{\mathcal{A}}(f) M(\Phi_{\mathcal{A}})^{-1}$$

und damit wegen $M(\Phi_{\mathcal{B}}) = \mathbf{B}$ *und* $M(\Phi_{\mathcal{A}})^{-1} = \mathbf{A}^{-1} : f(\mathbf{v}) = \mathbf{B}\mathbf{C}\mathbf{A}^{-1}\mathbf{v}$ *für* $\mathbf{C} := M_{\mathcal{B}}^{\mathcal{A}}(f)$. ◊

Satz 4.2.18. $M_{\mathcal{A}}^{\mathcal{E}}(\Phi_{\mathcal{A}}) = \mathbf{E}$

Beweis. $M_{\mathcal{A}}^{\mathcal{E}}(\Phi_{\mathcal{A}}) = M\left(\Phi_{\mathcal{A}}^{-1} \circ \Phi_{\mathcal{A}} \circ \Phi_{\mathcal{E}}\right) = M(\Phi_{\mathcal{E}}) = \mathbf{E}$ □

Satz 4.2.19. *Sei* $(\mathbf{v}_1, ..., \mathbf{v}_n)$ *eine Basis* $\mathcal{A}$ *von* V *und* $\mathbf{A}$ *die Matrix, so dass* $\mathbf{a}_j = \mathbf{v}_j$.
Sei $(\mathbf{w}_1, ..., \mathbf{w}_n)$ *eine Basis* $\mathcal{B}$ *von* V *und* $\mathbf{B}$ *die Matrix, so dass* $\mathbf{b}_j = \mathbf{w}_j$.
Dann gilt:

$$M_{\mathcal{B}}^{\mathcal{A}}(id_V) = \mathbf{B}^{-1}\mathbf{A} = M\left(T_{\mathcal{B}}^{\mathcal{A}}\right)$$

Beweis.

$$\begin{aligned} M_{\mathcal{B}}^{\mathcal{A}}(id_V) &= M\left(\Phi_{\mathcal{B}}^{-1} \circ id_V \circ \Phi_{\mathcal{A}}\right) \\ &= M\left(\Phi_{\mathcal{B}}^{-1} \circ \Phi_{\mathcal{A}}\right) \\ &= \mathbf{B}^{-1}\mathbf{A}. \end{aligned}$$

Satz 4.2.20. *Sei* $(\mathbf{v}_1, ..., \mathbf{v}_n)$ *eine Basis* $\mathcal{A}$ *von* V *und* $\mathbf{A}$ *die Matrix, so dass* $\mathbf{a}_j = \mathbf{v}_j$.
Dann gilt:

$$M_{\mathcal{A}}(id_V) = M\left(\Phi_{\mathcal{A}}^{-1} \circ id_V \circ \Phi_{\mathcal{A}}\right) = M\left(\Phi_{\mathcal{A}}^{-1} \circ \Phi_{\mathcal{A}}\right) = \mathbf{A}^{-1}\mathbf{A} = \mathbf{E}$$

□

4.3 Basiswechsel

Satz 4.3.1. *Seien* $\mathcal{A}, \mathcal{C}$ *Basen von* V *und* $\mathcal{D}, \mathcal{B}$ *Basen von* W *und* $f : V \to W$ *ein Homomorphismus. Dann gilt:*

$$\begin{aligned} M_{\mathcal{B}}^{\mathcal{A}}(f) &= M_{\mathcal{B}}^{\mathcal{D}}(id_V) \cdot M_{\mathcal{D}}^{\mathcal{C}}(f) \cdot M_{\mathcal{C}}^{\mathcal{A}}(id_W) \\ &= M\left(T_{\mathcal{B}}^{\mathcal{D}}\right) \cdot M_{\mathcal{D}}^{\mathcal{C}}(f) \cdot M\left(T_{\mathcal{C}}^{\mathcal{A}}\right) \end{aligned}$$

Beweis.

$$\begin{aligned} M_{\mathcal{B}}^{\mathcal{A}}(f) &= M\left(\Phi_{\mathcal{B}}^{-1} \circ f \circ \Phi_{\mathcal{A}}\right) \\ &= M\left(\Phi_{\mathcal{B}}^{-1} \circ \left(\Phi_{\mathcal{D}} \circ \Phi_{\mathcal{D}}^{-1}\right) \circ f \circ \left(\Phi_{\mathcal{C}} \circ \Phi_{\mathcal{C}}^{-1}\right) \circ \Phi_{\mathcal{A}}\right) \\ &= M\left(\Phi_{\mathcal{B}}^{-1} \circ id_V \circ \Phi_{\mathcal{D}} \circ \Phi_{\mathcal{D}}^{-1} \circ f \circ \Phi_{\mathcal{C}} \circ \Phi_{\mathcal{C}}^{-1} \circ id_W \circ \Phi_{\mathcal{A}}\right) \\ &= M\left(\Phi_{\mathcal{B}}^{-1} \circ id_V \circ \Phi_{\mathcal{D}}\right) \cdot M\left(\Phi_{\mathcal{D}}^{-1} \circ f \circ \Phi_{\mathcal{C}}\right) \cdot M\left(\Phi_{\mathcal{C}}^{-1} \circ id_W \circ \Phi_{\mathcal{A}}\right) \\ &= M_{\mathcal{B}}^{\mathcal{D}}(id_V) \cdot M_{\mathcal{D}}^{\mathcal{C}}(f) \cdot M_{\mathcal{C}}^{\mathcal{A}}(id_W) \end{aligned}$$

□

Mit den bisherigen Ergebnissen gilt damit:

Satz 4.3.2. *Seien zwei Matrizen $M_{\mathcal{B}}^{\mathcal{A}}(f)$ und $M_{\mathcal{D}}^{\mathcal{C}}(f)$ gegeben, die den Homomorphismus f bestimmen. Dann gibt es vollrangige (= reguläre) Matrizen $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ und $\mathbf{D}$, deren Spalten aus den Basisvektoren von $\mathcal{A}, \mathcal{B}, \mathcal{C}$ und $\mathcal{D}$ bestehen, so dass*

$$M_{\mathcal{B}}^{\mathcal{A}}(f) = \mathbf{B}^{-1}\mathbf{D} \cdot M_{\mathcal{D}}^{\mathcal{C}}(f) \cdot \mathbf{C}^{-1}\mathbf{A}$$

Umgekehrt erhält man durch Multiplikation mit den entsprechenden Inversen:

$$\left(\mathbf{B}^{-1}\mathbf{D}\right)^{-1} \cdot M_{\mathcal{B}}^{\mathcal{A}}(f) \cdot \left(\mathbf{C}^{-1}\mathbf{A}\right)^{-1} = M_{\mathcal{D}}^{\mathcal{C}}(f)$$
$$M_{\mathcal{D}}^{\mathcal{C}}(f) = \mathbf{D}^{-1}\mathbf{B} \cdot M_{\mathcal{B}}^{\mathcal{A}}(f) \cdot \mathbf{A}^{-1}\mathbf{C}$$

Mit $\mathbf{S} := \mathbf{D}^{-1}\mathbf{B} = M\left(\Phi_{\mathcal{D}}^{-1} \circ \Phi_{\mathcal{B}}\right)$ und $\mathbf{T} := \mathbf{C}^{-1}\mathbf{A} = M\left(\Phi_{\mathcal{C}}^{-1} \circ \Phi_{\mathcal{A}}\right)$ gibt es also reguläre Matrizen $\mathbf{S}$ und $\mathbf{T}$, so dass:

$$M_{\mathcal{B}}^{\mathcal{A}}(f) = \mathbf{S}^{-1} \cdot M_{\mathcal{D}}^{\mathcal{C}}(f) \cdot \mathbf{T}$$

und

$$M_{\mathcal{D}}^{\mathcal{C}}(f) = \mathbf{S} \cdot M_{\mathcal{B}}^{\mathcal{A}}(f) \cdot \mathbf{T}^{-1}$$

Das folgende Diagramm illustriert die Zusammenhänge:

$$\begin{array}{ccccccc}
\mathbb{R}^n & \longrightarrow & \longrightarrow & M_{\mathcal{B}}^{\mathcal{A}}(f) & \longrightarrow & \longrightarrow & \mathbb{R}^m \\
\downarrow & \searrow & & & & \swarrow & \downarrow \\
\mathbf{T} = \mathbf{C}^{-1}\mathbf{A} & & V & \xrightarrow{f} & W & & \mathbf{S} = \mathbf{D}^{-1}\mathbf{B} \\
\downarrow & \nearrow & & & & \nwarrow & \downarrow \\
\mathbb{R}^n & \longrightarrow & \longrightarrow & M_{\mathcal{D}}^{\mathcal{C}}(f) & \longrightarrow & \longrightarrow & \mathbb{R}^m
\end{array}$$

Beispiel 4.3.3. *Sei* $\mathcal{A} = \left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 4 \end{pmatrix}, \begin{pmatrix} 5 \\ 6 \\ 1 \end{pmatrix} \right\}, \mathcal{B} = \left\{ \begin{pmatrix} 3 \\ 4 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\}$,
$\mathcal{C} = \left\{ \begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix}, \begin{pmatrix} 1 \\ 7 \\ 8 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} \right\}$ *und* $\mathcal{D} = \left\{ \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 5 \\ 6 \end{pmatrix} \right\}$. *Ferner sei der Homomorphismus f definiert durch*

$$f\left(\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}\right) = \begin{pmatrix} 4 \\ 8 \end{pmatrix}$$
$$f\left(\begin{pmatrix} 2 \\ 1 \\ 4 \end{pmatrix}\right) = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$
$$f\left(\begin{pmatrix} 5 \\ 6 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 14 \\ 81 \end{pmatrix}$$

Damit ist

$$M_{\mathcal{B}}^{\mathcal{A}}(f) = \begin{pmatrix} \frac{6}{5} & -\frac{1}{5} & \frac{148}{5} \\ \frac{1}{5} & \frac{9}{5} & -\frac{187}{5} \end{pmatrix}$$

denn

$$\begin{pmatrix} 4 \\ 8 \end{pmatrix} = a_{11} \begin{pmatrix} 3 \\ 4 \end{pmatrix} + a_{21} \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

und damit

$$4 = 3a_{11} + 2a_{21}$$
$$5 = 4a_{11} + ba_{21}$$

$a_{11} = \frac{6}{5}, a_{21} = \frac{1}{5}$

$$\begin{pmatrix} 3 \\ 1 \end{pmatrix} = a_{12} \begin{pmatrix} 3 \\ 4 \end{pmatrix} + a_{22} \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

und damit

$$\begin{aligned} 3 &= 3a_{12} + 2a_{22} \\ 1 &= 4a_{12} + a_{22} \end{aligned}$$

$a_{12} = -\frac{1}{5}, a_{22} = \frac{9}{5}$ *und schliesslich*

$$\begin{pmatrix} 14 \\ 81 \end{pmatrix} = a_{13} \begin{pmatrix} 3 \\ 4 \end{pmatrix} + a_{23} \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

und damit

$$\begin{aligned} 14 &= 3a_{13} + 2a_{23} \\ 81 &= 4a_{13} + a_{23} \end{aligned}$$

$a_{13} = \frac{148}{5}, a_{23} = -\frac{187}{5}$.
Damit erhält man die Matrix $M_{\mathcal{D}}^{\mathcal{C}}(f)$ *gemäss Satz durch*

$$\begin{aligned} M_{\mathcal{D}}^{\mathcal{C}}(f) &= \begin{pmatrix} 1 & 5 \\ 3 & 6 \end{pmatrix}^{-1} \begin{pmatrix} 3 & 2 \\ 4 & 1 \end{pmatrix} \cdot \begin{pmatrix} \frac{6}{5} & -\frac{1}{5} & \frac{148}{5} \\ \frac{1}{5} & \frac{9}{5} & -\frac{187}{5} \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & 5 \\ 2 & 1 & 6 \\ 3 & 4 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 3 & 1 & 3 \\ 1 & 7 & 2 \\ 5 & 8 & 1 \end{pmatrix} \\ &= \begin{pmatrix} -\frac{38}{153} & -\frac{398}{153} & \frac{2399}{153} \\ \frac{121}{153} & \frac{454}{153} & -\frac{307}{153} \end{pmatrix} \end{aligned}$$

Als Beispiel betrachten wir den Funktionswert $f \begin{pmatrix} 11 \\ 1 \\ 3 \end{pmatrix}$:

$$\begin{aligned} f \begin{pmatrix} 11 \\ 1 \\ 3 \end{pmatrix} &= \mathbf{B} M_{\mathcal{B}}^{\mathcal{A}}(f) \mathbf{A}^{-1} \begin{pmatrix} 11 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 3 & 2 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} \frac{6}{5} & -\frac{1}{5} & \frac{148}{5} \\ \frac{1}{5} & \frac{9}{5} & -\frac{187}{5} \end{pmatrix} \begin{pmatrix} 1 & 2 & 5 \\ 2 & 1 & 6 \\ 3 & 4 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 11 \\ 1 \\ 3 \end{pmatrix} \\ &= \begin{pmatrix} \frac{169}{17} \\ 88 \end{pmatrix} \end{aligned}$$

und mit den alternativen Basen:

$$\begin{aligned} f \begin{pmatrix} 11 \\ 1 \\ 3 \end{pmatrix} &= \mathbf{D} M_{\mathcal{D}}^{\mathcal{C}}(f) \mathbf{C}^{-1} \begin{pmatrix} 11 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 & 5 \\ 3 & 6 \end{pmatrix} \begin{pmatrix} -\frac{38}{153} & -\frac{398}{153} & \frac{2399}{153} \\ \frac{121}{153} & \frac{454}{153} & -\frac{307}{153} \end{pmatrix} \begin{pmatrix} 3 & 1 & 3 \\ 1 & 7 & 2 \\ 5 & 8 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 11 \\ 1 \\ 3 \end{pmatrix} \\ &= \begin{pmatrix} \frac{169}{17} \\ 88 \end{pmatrix} \end{aligned}$$

$\Diamond$

Entsprechend gilt für Endomorphismen:

Satz 4.3.4. *Seien zwei Matrizen* $M_{\mathcal{A}}(f)$ *und* $M_{\mathcal{B}}(f)$ *gegeben, die den Endomorphismus* f *bestimmen. Dann gibt es Matrizen* $\mathbf{A}$ *und* $\mathbf{B}$*, deren Spalten aus den Basisvektoren von* $\mathcal{A}$ *und* $\mathcal{B}$ *bestehen, so dass mit* $\mathbf{S} := \mathbf{B}^{-1}\mathbf{A} = M\left(\Phi_{\mathcal{B}}^{-1} \circ \Phi_{\mathcal{A}}\right)$

$$M_{\mathcal{A}}(f) = \mathbf{S}^{-1} \cdot M_{\mathcal{B}}(f) \cdot \mathbf{S}$$

und

$$M_{\mathcal{B}}(f) = \mathbf{S} \cdot M_{\mathcal{A}}(f) \cdot \mathbf{S}^{-1}$$

Beweis. Es gilt:

$$M_{\mathcal{A}}(f) = M_{\mathcal{A}}^{\mathcal{B}}(id_V) \cdot M_{\mathcal{B}}(f) \cdot M_{\mathcal{B}}^{\mathcal{A}}(id_V) \quad \left(= M\left(T_{\mathcal{A}}^{\mathcal{B}}\right) \cdot M_{\mathcal{B}}(f) \cdot M\left(T_{\mathcal{B}}^{\mathcal{A}}\right)\right)$$

und damit

$$M_{\mathcal{A}}(f) = \mathbf{A}^{-1}\mathbf{B} \cdot M_{\mathcal{B}}(f) \cdot \mathbf{B}^{-1}\mathbf{A},$$

und da $\left(\mathbf{A}^{-1}\mathbf{B}\right)^{-1} = \mathbf{B}^{-1}\mathbf{A}$, gilt mit $\mathbf{S} := \mathbf{B}^{-1}\mathbf{A}$ der Satz. □

Beispiel 4.3.5. *Sei* $\mathcal{A} = \left\{\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \end{pmatrix}\right\}, \mathcal{B} = \left\{\begin{pmatrix} 3 \\ 4 \end{pmatrix}, \begin{pmatrix} 6 \\ 5 \end{pmatrix}\right\}$. *Ferner sei der Endomorphismus* f *definiert durch*

$$f\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}\right) = \begin{pmatrix} 4 \\ 7 \end{pmatrix}$$
$$f\left(\begin{pmatrix} 3 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

Damit ist

$$M_{\mathcal{A}}(f) = \begin{pmatrix} 3 & -3 \\ \frac{1}{3} & \frac{7}{3} \end{pmatrix}$$

denn

$$\begin{pmatrix} 4 \\ 7 \end{pmatrix} = a_{11}\begin{pmatrix} 1 \\ 2 \end{pmatrix} + a_{21}\begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

und damit

$$4 = a_{11} + 3a_{21}$$
$$7 = 2a_{11} + 3a_{21}$$

$a_{11} = 3, a_{21} = \frac{1}{3}$ *und*

$$\begin{pmatrix} 4 \\ 1 \end{pmatrix} = a_{12}\begin{pmatrix} 1 \\ 2 \end{pmatrix} + a_{22}\begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

und

$$4 = a_{12} + 3a_{22}$$
$$1 = 2a_{12} + 3a_{22}$$

$a_{12} = -3, a_{22} = \frac{7}{3}$.
Gemäss Satz gilt

$$M_{\mathcal{B}}(f) = \mathbf{S} \cdot M_{\mathcal{A}}(f) \cdot \mathbf{S}^{-1}$$

wobei

$$\mathbf{S} = \mathbf{B}^{-1}\mathbf{A} = \begin{pmatrix} 3 & 6 \\ 4 & 5 \end{pmatrix}^{-1}\begin{pmatrix} 1 & 3 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} \frac{7}{9} & -1 \\ -\frac{2}{9} & 1 \end{pmatrix}$$

Entsprechend ist

$$M_{\mathcal{B}}(f) = \begin{pmatrix} \frac{7}{9} & -1 \\ -\frac{2}{9} & 1 \end{pmatrix}\begin{pmatrix} 3 & -3 \\ \frac{1}{3} & \frac{7}{3} \end{pmatrix}\begin{pmatrix} \frac{7}{9} & -1 \\ -\frac{2}{9} & 1 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{26}{15} & -\frac{44}{15} \\ \frac{3}{5} & \frac{18}{5} \end{pmatrix}$$

Am Beispiel eines konkreten Funktionswertes gilt:

$$f\begin{pmatrix} 11 \\ 8 \end{pmatrix} = \mathbf{A}M_{\mathcal{A}}(f)\,\mathbf{A}^{-1}\begin{pmatrix} 11 \\ 8 \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 3 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 3 & -3 \\ \frac{1}{3} & \frac{7}{3} \end{pmatrix}\begin{pmatrix} 1 & 3 \\ 2 & 1 \end{pmatrix}^{-1}\begin{pmatrix} 11 \\ 8 \end{pmatrix} = \begin{pmatrix} \frac{108}{5} \\ \frac{31}{5} \end{pmatrix}$$

und mit der alternativen Basis:

$$f\begin{pmatrix}11\\8\end{pmatrix} = \mathbf{B}M_{\mathcal{B}}(f)\,\mathbf{B}^{-1}\begin{pmatrix}11\\8\end{pmatrix}$$
$$= \begin{pmatrix}3&6\\4&5\end{pmatrix}\begin{pmatrix}\frac{26}{15}&-\frac{44}{15}\\\frac{3}{5}&\frac{18}{5}\end{pmatrix}\begin{pmatrix}3&6\\4&5\end{pmatrix}^{-1}\begin{pmatrix}11\\8\end{pmatrix} = \begin{pmatrix}\frac{108}{5}\\\frac{31}{5}\end{pmatrix}$$

◊

Definition 4.3.6. *Gibt es für die Matrizen* $\mathbf{A}$ *und* $\mathbf{B}$ *ein* $\mathbf{S}$*, so dass* $\mathbf{A} = \mathbf{S}^{-1}\cdot\mathbf{B}\cdot\mathbf{S}$*, so nennen wir* $\mathbf{A}$ *und* $\mathbf{B}$ *ähnlich.* ◊

Satz 4.3.7. $\mathbf{A}$ *und* $\mathbf{B}$ *sind ähnlich genau dann, wenn Sie bezüglich verschiedener Basen denselben Endomorphismus beschreiben.*

Beweis. (i) Wir nehmen an, $M_{\mathcal{A}}(f) = \mathbf{A}$ und $M_{\mathcal{B}}(f) = \mathbf{B}$ beschreiben bezüglich verschiedener Basen denselben Endomorphismus f. Es wurde oben bereits gezeigt, dass es dann $\mathbf{S}$ gibt mit $M_{\mathcal{A}}(f) = \mathbf{S}^{-1}\cdot M_{\mathcal{B}}(f)\cdot\mathbf{S}$. Damit gilt auch $\mathbf{A} = \mathbf{S}^{-1}\mathbf{B}\mathbf{S}$ sowie $\mathbf{B} = \mathbf{S}\mathbf{A}\mathbf{S}^{-1}$
(ii) Gelte umgekehrt $\mathbf{B} = \mathbf{S}^{-1}\cdot\mathbf{A}\cdot\mathbf{S}$. $\mathbf{S}$ ist damit invertierbar und die durch $\mathbf{S}$ definierbare Abbildung

$$f_{\mathbf{S}}(\mathbf{v}_j) = \sum_{i=1}^{n} s_{ij}\mathbf{v}_i = \sum_{i=1}^{n} s_{ij}\sum_{k=1}^{n} v_{ki}\mathbf{e}_k$$
$$= \sum_{k=1}^{n}\sum_{i=1}^{n}(v_{ki}s_{ij})\,\mathbf{e}_k =: \mathbf{c}_j \tag{4.3.1}$$

bildet die Basisvektoren $(\mathbf{v}_1,...,\mathbf{v}_n)$ in die Basisvektoren $\mathcal{C} = (\mathbf{c}_1,...,\mathbf{c}_n)$ ab. Es gilt gemäss (4.3.1) mit $\mathbf{A} := (\mathbf{v}_1,...,\mathbf{v}_n)$ und $\mathbf{C} = (\mathbf{c}_1,...,\mathbf{c}_n)$

$$\mathbf{C} = \mathbf{A}\mathbf{S}$$

und

$$\mathbf{S} = \mathbf{A}^{-1}\mathbf{C}$$

Daraus folgt mit

$$M_{\mathcal{A}}^{\mathcal{C}}(id_V) = M\left(\Phi_{\mathcal{A}}^{-1}\circ id_V\circ\Phi_{\mathcal{C}}^{-1}\right)$$
$$= \mathbf{A}^{-1}\mathbf{C} = \mathbf{S}$$

$$\begin{aligned}\mathbf{B} &= \mathbf{S}^{-1}\cdot\mathbf{A}\cdot\mathbf{S}\\ &= M_{\mathcal{C}}^{\mathcal{A}}(id_V)\cdot M_{\mathcal{A}}^{\mathcal{A}}(f)\cdot M_{\mathcal{A}}^{\mathcal{C}}(id_V)\\ &= M_{\mathcal{C}}^{\mathcal{C}}(id_V\circ f\circ id_V)\\ &= M_{\mathcal{C}}(f).\end{aligned}$$

$\mathbf{B}$ ist also eine Matrixdarstellung von f und es gilt $\mathbf{B} = \mathbf{C}$. □

Im Speziellen folgt

$$M_{\mathcal{B}}^{\mathcal{A}}(f) = M\left(\Phi_{\mathcal{B}}^{-1}\circ f\circ\Phi_{\mathcal{A}}\right)$$
$$= M\left(\Phi_{\mathcal{B}}^{-1}\right)\cdot M(f)\cdot M(\Phi_{\mathcal{A}})$$
$$= \mathbf{B}^{-1}M(f)\,\mathbf{A}\quad(=\mathbf{B}^{-1}\mathbf{E}M(f)\,\mathbf{E}^{-1}\mathbf{A})$$

sowie, wie bereits oben gezeigt,

$$M(f) = \mathbf{B} \cdot M_{\mathcal{B}}^{\mathcal{A}}(f) \cdot \mathbf{A}^{-1}.$$

Damit kann $M_{\mathcal{B}}^{\mathcal{A}}(f)$ immer in $M(f)$ umgerechnet werden. Bei einem Endomorphismus, der mit Hilfe der Basis $\mathcal{A}$ definiert wurde, gilt:

$$M_{\mathcal{A}}(f) = \mathbf{A}^{-1} \cdot M(f) \cdot \mathbf{A}$$

Damit ist $M_{\mathcal{A}}(f)$ und $M(f)$ ähnlich. Schliesslich gilt damit auch

$$M(f) = \mathbf{A} \cdot M_{\mathcal{A}}(f) \cdot \mathbf{A}^{-1}.$$

Satz 4.3.8. *Sind zwei Matrizen ähnlich, dann haben sie denselben Rang.*

Beweis. Multipliziert man eine Matrix $\mathbf{A}$ mit regulären Matrizen, so verändert sich der Rang von $\mathbf{A}$ nicht. □

Kapitel 5

Euklidische Vektorräume

I:\dateien\Mathematik\111Math\Lineare_Algebra\004linal_singulaer_werte.tex
Ziel: Definition Skalarprodukt, Längen von Vektoren (= Abstand eines Punktes vom Koordinatenursprung; Abstand zweier Punkte); Definition Orthogonalität; orthogonale Projektion; Drehung des Koordinatensystems
Vorausgesetzt ist für $(\mathbf{u}_1, ..., \mathbf{u}_n)$ und $\mathbf{v} = (v_1, ..., v_n) \in \mathbb{R}^n$ sowie $\alpha \in \mathbb{R}$: $\alpha\mathbf{u} := (\alpha\mathbf{u}_1, ..., \alpha\mathbf{u}_n)$ und $\mathbf{u} + \mathbf{v} := (u_1 + v_1, ..., u_n + v_n)$. $(v_1, ..., v_n)^T \in \mathbb{R}^{n\times 1} \Longleftrightarrow (v_1, .., v_n) \in \mathbb{R}^{1\times n}$. $\mathbf{v}^T \in \mathbb{R}^{1\times n} \Longleftrightarrow \mathbf{v} \in \mathbb{R}^{n\times 1}$. Die Matrix $\mathbf{X} \in \mathbb{R}^{n\times m}$ genau dann, wenn $\mathbf{X}$ Graph einer Abbildung $\mathbb{N}_n^* \times \mathbb{N}_m^* \to \mathbb{R}$ ist. $\mathbb{N}_n^* = \{k | k \in \mathbb{N} \text{ und } 1 \leq k \leq n\}$. $\mathbf{0} = \{0\}^n$ für $\mathbf{0} \in \mathbb{R}^n$

5.1 Das Skalarprodukt

Definition 5.1.1. *(euklidischer Raum und Skalarprodukt) Seien* $\mathbf{u} = (u_1, ..., u_n)$ *und* $\mathbf{v} = (v_1, ..., v_n) \in \mathbb{R}^n$. *Dann ist das euklidische Produkt von* $\mathbf{u}$ *und* $\mathbf{v}$ *(euklidisches Skalarprodukt) definiert durch*

$$\cdot : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$
$$\mathbf{u} \cdot \mathbf{v} := \sum_{i=1}^{n} u_i v_i$$

Der Vektorraum $(\mathbb{R}^n, \cdot)$ *wird n-dimensionaler euklidischer Raum genannt.* $\mathbf{u}^2 := \mathbf{u} \cdot \mathbf{u}$. ◊

Bemerkung 5.1.2. *Man kann das Skalarprodukt als Matrixmultiplikation auffassen:* $\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T\mathbf{v}$ *mit* $\mathbf{v}, \mathbf{u} \in \mathbb{R}^{n\times 1}$ *auf der rechten Seite der Gleichung. Üblich ist auch die Schreibweise* $\langle \mathbf{u}, \mathbf{v} \rangle$ ◊

Satz 5.1.3. *Seien* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$. $\alpha \in \mathbb{R}$. *Dann gilt*

1. $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$ *(Kommutativität, Symmetrie)*
2. $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$ *(Addititvität oder Distributivität)*
3. $(\mathbf{u} + \mathbf{v}) \cdot (\mathbf{w} + \mathbf{x}) = ((\mathbf{u} + \mathbf{v}) \cdot \mathbf{w}) + ((\mathbf{u} + \mathbf{v}) \cdot \mathbf{x})$
4. $(\alpha\mathbf{u}) \cdot \mathbf{v} = \alpha\,(\mathbf{u} \cdot \mathbf{v})$ *(Homogenität)*
5. $\mathbf{v} \cdot \mathbf{v} \geq 0$; $\mathbf{v} \cdot \mathbf{v} = 0$ *genau dann, wenn* $\mathbf{v} = \mathbf{0}$ *(positive Definitheit)*
6. $(\alpha\mathbf{u} + \beta\mathbf{v}) \cdot \mathbf{w} = \alpha\,(\mathbf{u} \cdot \mathbf{w}) + \beta\,(\mathbf{u} \cdot \mathbf{w})$
7. $\mathbf{u} \cdot (\alpha\mathbf{v} + \beta\mathbf{w}) = \alpha\,(\mathbf{u} \cdot \mathbf{v}) + \beta\,(\mathbf{v} \cdot \mathbf{w})$

Beweis. Es gilt

1. $\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{n} u_i v_i = \sum_{i=1}^{n} v_i u_i = \mathbf{v} \cdot \mathbf{u}$
2. $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \sum_{i=1}^{n} (u_i + v_i) w_i = \sum_{i=1}^{n} u_i w_i + \sum_{i=1}^{n} v_i w_i = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$
3. $(\mathbf{u} + \mathbf{v}) \cdot (\mathbf{w} + \mathbf{x}) = \sum_{i=1}^{n} (u_i + v_i)(w_i + x_i) = \sum_{i=1}^{n} ((u_i + v_i) w_i + ((u_i + v_i) x_i)) = \sum_{i=1}^{n} (u_i + v_i) w_i + \sum_{i=1}^{n} ((u_i + v_i) x_i)$
4. $(\alpha \mathbf{u}) \cdot \mathbf{v} = \sum_{i=1}^{n} \alpha u_i v_i = \alpha \sum_{i=1}^{n} u_i v_i = \alpha (\mathbf{u} \cdot \mathbf{v})$
5. $\mathbf{v} \cdot \mathbf{v} = \sum_{i=1}^{n} v_i v_i = \sum_{i=1}^{n} v_i^2 \geq 0$; $\mathbf{v} \cdot \mathbf{v} = 0$ genau dann, wenn $v_i^2 = 0$ für alle v_i, d.h. wenn $\mathbf{v} = 0$.
6. Durch Anwendung der bereits bewiesenen Regeln
7. Durch Anwendung der bereits bewiesenen Regeln

□

Definition 5.1.4. *(euklidische Norm $\|\|$ eines Vektors (Punktes) = Länge der Strecke vom Punkt zum Ursprung des Koordinatensystems; euklidische Distanz d zwischen zwei Punkten (Vektoren) = Länge der Strecke zwischen zwei Punkten) Sei $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. Dann ist*

$$\begin{aligned} \|\| &: \mathbb{R}^n \to \mathbb{R} \\ \|\mathbf{u}\| &:= \sqrt{\mathbf{u} \cdot \mathbf{u}} = \sqrt{u_1^2 + ... + u_n^2} \end{aligned}$$

$$\begin{aligned} d &: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \\ d(\mathbf{u}, \mathbf{v}) &:= \sqrt{(u_1 - v_1)^2 + ... + (u_n - v_n)^2} \end{aligned}$$

◊

Bemerkung 5.1.5. *Für*

$$\begin{aligned} \|\| &: \mathbb{R} \to \mathbb{R} \\ \|\mathbf{u}\| &:= \sqrt{\mathbf{u} \cdot \mathbf{u}} \end{aligned}$$

ist $\|\|$ mit der Betragsfunktion $|| : \mathbb{R} \to \mathbb{R}, |u| := \begin{cases} u & \text{für } u \geq 0 \\ -u & \text{für } u < 0 \end{cases}$ identisch.
Für $\mathbf{v} = \mathbf{0}$ gilt $d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u}\|$. ◊

Satz 5.1.6. *(Cauchy-Schwarzsche Ungleichung im $\mathbb{R}^n$. Für $\mathbf{u}, \mathbf{v}, \mathbf{0} \in \mathbb{R}^n$ gilt*

$$\|\mathbf{u} \cdot \mathbf{v}\| \leq \|\mathbf{u}\| \|\mathbf{v}\|$$

Beweis. Für $\mathbf{u} = 0$ oder $\mathbf{v} = 0$ gilt der Satz unmittelbar. Für $\mathbf{u} \neq 0$ und $\mathbf{v} \neq 0$ erhält man für $x \in \mathbb{R}$

$$\begin{aligned} 0 &\leq (\mathbf{u}x + \mathbf{v}) \cdot (\mathbf{u}x + \mathbf{v}) \\ &= (\mathbf{u} \cdot \mathbf{u}) x^2 + (\mathbf{u} \cdot \mathbf{v}) x + (\mathbf{v} \cdot \mathbf{u}) x + \mathbf{v}^2 \\ &= (\mathbf{u} \cdot \mathbf{u}) x^2 + 2 (\mathbf{u} \cdot \mathbf{v}) x + \mathbf{v} \cdot \mathbf{v} \end{aligned}$$

Dies ist ein Polynom 2. Grades ($\mathbf{u}\cdot\mathbf{u}, \mathbf{u}\cdot\mathbf{v}, \mathbf{v}\cdot\mathbf{v} \in \mathbb{R}$), das wegen $0 \leq (\mathbf{u}x+\mathbf{v})\cdot(\mathbf{u}x+\mathbf{v})$ höchstens eine reelle Nullstelle besitzt, d.h. für die Diskriminante gilt

$$4(\mathbf{u}\cdot\mathbf{v})^2 - 4(\mathbf{u}\cdot\mathbf{u})(\mathbf{v}\cdot\mathbf{v}) \leq 0$$
$$(\mathbf{u}\cdot\mathbf{v})^2 \leq (\mathbf{u}\cdot\mathbf{u})(\mathbf{v}\cdot\mathbf{v})$$

und durch Wurzelziehen auf beiden Seiten

$$\|\mathbf{u}\cdot\mathbf{v}\| \leq \|\mathbf{u}\|\,\|\mathbf{v}\|$$

($\mathbf{u}\cdot\mathbf{u}$, $\mathbf{u}\cdot\mathbf{v} \in \mathbb{R}$ und damit ist $|\mathbf{u}\cdot\mathbf{v}| = \|\mathbf{u}\cdot\mathbf{v}\|$). □

Satz 5.1.7. *Seien $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ und $\alpha \in \mathbb{R}$.*

1. $\|\mathbf{u}\| \geq 0$
2. $\|\mathbf{u}\| = 0$ *genau dann, wenn* $\mathbf{u} = 0$
3. $\|\alpha\mathbf{u}\| = |\alpha|\,\|\mathbf{u}\|$
4. $\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$ *(Dreiecksungleichung)*

Beweis. 1. $\|\mathbf{u}\| = \sqrt{\mathbf{u}^2} \geq 0$ genau dann, wenn $\mathbf{u}^2 \geq 0$ (was gemäss Satz 5.1.3 gilt)

2. $\|\mathbf{u}\| = \sqrt{\mathbf{u}^2} = 0$ genau dann, wenn $\mathbf{u}^2 = 0$ (was gemäss Satz 5.1.3 gilt)
3. $\|\alpha\mathbf{u}\| = \sqrt{\alpha\mathbf{u}\cdot\alpha\mathbf{u}} = \sqrt{\alpha^2(\mathbf{u}\cdot\mathbf{u})} = |\alpha|\sqrt{\mathbf{u}\cdot\mathbf{u}} = |\alpha|\,\|\mathbf{u}\|$.
4. $\|\mathbf{u}+\mathbf{v}\|^2 = (\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v}) = \mathbf{u}^2 + \mathbf{u}\cdot\mathbf{v} + \mathbf{v}\cdot\mathbf{u} + \mathbf{v}^2 = \|\mathbf{u}\|^2 + 2(\mathbf{u}\cdot\mathbf{v}) + \|\mathbf{v}\|^2 \leq \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\,\|\mathbf{v}\| + \|\mathbf{v}\|^2 = (\|\mathbf{u}\| + \|\mathbf{v}\|)^2$ (vorletzter Schritt: Cauchy-Schwarz. $\mathbf{u}\cdot\mathbf{v}$ ist eine reelle Zahl und es gilt $\mathbf{u}\cdot\mathbf{v} \leq |\mathbf{u}\cdot\mathbf{v}| = \|\mathbf{u}\cdot\mathbf{v}\| \leq \|\mathbf{u}\|\,\|\mathbf{v}\|$))

□

Satz 5.1.8. *(Eigenschaften euklidischer Abstand) Seien $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$.*

1. $d(\mathbf{u},\mathbf{v}) \geq 0$
2. $d(\mathbf{u},\mathbf{v}) = 0$ *genau dann, wenn* $\mathbf{u} = \mathbf{v}$
3. $d(\mathbf{u},\mathbf{v}) = d(\mathbf{u},\mathbf{v})$
4. $d(\mathbf{u},\mathbf{v}) \leq d(\mathbf{u},\mathbf{w}) + d(\mathbf{w},\mathbf{v})$

Beweis. Die Eigenschaften ergeben sich direkt aus Satz 5.1.7. □

Satz 5.1.9. $\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2(\mathbf{x}\cdot\mathbf{y}) + \|\mathbf{y}\|^2$

Beweis.

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

□

Definition 5.1.10. *(Orthogonale Vektoren oder Punkte) Zwei Vektoren $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ heissen orthogonal genau dann, wenn* $\mathbf{u}\cdot\mathbf{v} = 0$. ◊

Satz 5.1.11. *(des Pythagoras) Sind $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ orthogonal, dann gilt*

$$\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$$

Beweis. Dies folgt unmittelbar aus Satz 5.1.9. □

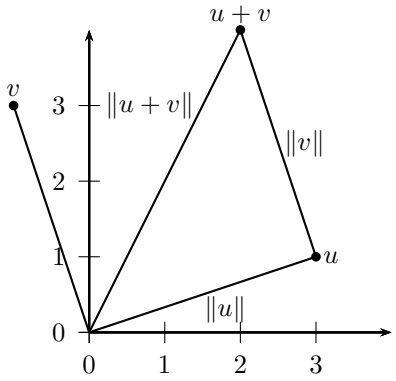


Abbildung 5.1.1: Graphik zum Satz von Pythogoras; $u \cdot v = 0$

Bemerkung 5.1.12. *Der Satz von Pythagoras besagt, dass in einem rechtwinkligen Dreieck das Quadrat der Hypotenuse gleich der Summe der Quadrate der Katheten ist - eine intuitive Herleitung des Satzes erfolgt mit Hilfe der Abbildung 5.1.2, wo* $c^2 = a^2 + b^2$.

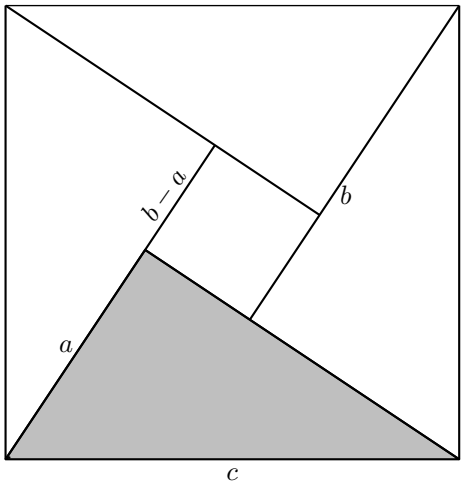


Abbildung 5.1.2: Graphik zum Satz von Pythagoras $c^2 = a^2 + b^2$

Die Fläche c^2 *ist zusammengesetzt aus vier gleichgrossen Dreiecken der Höhe* a *und der Basis* b *sowie aus dem Quadrat der Fläche* $(b-a)^2$. *Damit gilt* $c^2 = 4\frac{ab}{2} + (b-a)^2 = 2ab + b^2 - 2ab + a^2 = b^2 + a^2$. ◊

Satz 5.1.13. *Die Menge der zu einem Punkt* $\mathbf{x} \in \mathbb{R}^n$ *orthogonalen Punkte bilden einen Untervektorraum* $U^\perp$ *des* $\mathbb{R}^n$.

Beweis. (i) Sei $\mathbf{y} \in U^\perp$. Damit gilt $\mathbf{x} \cdot \mathbf{y} = 0$ und für $a \in \mathbb{R}$ ist $a\mathbf{y} \in U^\perp$, denn

$$\begin{aligned} \mathbf{x} \cdot a\mathbf{y} &= a\,(\mathbf{x} \cdot \mathbf{y}) \\ &= a \cdot 0 \\ &= 0 \end{aligned}$$

(ii) Seien $\mathbf{y}, \mathbf{z} \in U^\perp$. Damit ist auch $\mathbf{y} + \mathbf{z} \in U^\perp$, denn

$$\begin{aligned}\mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) &= \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z} \\ &= 0 + 0\end{aligned}$$

□

Entsprechend kann man sagen, dass ein Punkt $\mathbf{x}$ orthogonal zu einen Untervektorraum $U^\perp$ ist, nämlich auf die Menge der Punkte, zu denen er orthogonal ist. Zum Vektorraum $U^\perp$ steht aber nicht nur der Punkt $\mathbf{x}$ orthogonal, sondern auch alle Punkte des eindimensionalen Vektorraums $U = \{\mathbf{z} | \mathbf{z} = a\mathbf{x}, a \in \mathbb{R}\}$. Der Beweis verläuft analog zu dem von Satz 5.1.13. Wir sagen deshalb, das U senkrecht auf $U^\perp$ steht. Im Falle des $\mathbb{R}^2$ erhält man zwei aufeinander senkrechte Geraden, was die entsprechenden Definitionen verständlich macht. Zu jedem Punkt $\mathbf{x}$ in $\mathbb{R}^2$ gibt es entsprechend einen orthogonalen Unterraum von $\mathbb{R}^2$, zu dem der von $\mathbf{x}$ aufgespannte Vektorraum orthogonal ist.

Satz 5.1.14. *(lineare Unabhängigkeit orthogonaler Mengen) Eine orthogonale Menge $S = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ von $\mathbf{0}$ verschiedener $\mathbf{v}_i$ eines euklidischen Vektorraums ist linear unabhängig.*

Beweis. Man zeigt, dass $\sum_{i=1}^{n} \alpha_i \mathbf{v}_i = \mathbf{0}$ impliziert, dass $\alpha_i = 0$ für alle $i \in \mathbb{N}_n^*$. Sei also $\sum_{i=1}^{n} \alpha_i \mathbf{v}_i = \mathbf{0}$. Damit gilt für alle $\mathbf{v}_k$

$$\mathbf{0} = \mathbf{v}_k \cdot \sum_{i=1}^{n} \alpha_i \mathbf{v}_i = \sum_{i=1}^{n} \alpha_i \left(\mathbf{v}_k \cdot \mathbf{v}_i\right) = \alpha_k \left(\mathbf{v}_k \cdot \mathbf{v}_k\right)$$

Da $\mathbf{v}_k \cdot \mathbf{v}_k \neq \mathbf{0}$ (Vektoren sind von $\mathbf{0}$ verschieden), muss $\alpha_k = 0$ sein. □

Satz 5.1.15. *Sei U ein Untervektorraum von $\mathbb{R}^n = V$ und $U^\perp$ der zu U orthogonale Untervektoraum von V. Sei ferner $(\mathbf{v}_1, ..., \mathbf{v}_k)$ eine Basis von U und $(\mathbf{v}_{k+1}, ..., \mathbf{v}_n)$ eine Basis von $U^\perp$. Dann ist $(\mathbf{v}_1, ...\mathbf{v}_k, \mathbf{v}_{k+1}, ..., \mathbf{v}_n)$ eine Basis von V.*

Im $\mathbb{R}^n$ bezeichnet man Untervektorräume der Dimension $n-1$ als „Hyperebenen“. Durch die Berechnung der zu einem Punkt $\mathbf{x}$ oder dem von diesem aufgespannten eindimensionalen Vektorraum U orthogonalen Untervektorraums $U^\perp$ findet man entsprechend eine Hyperebene des $\mathbb{R}^n$. Die Hyperebene kann durch die Angabe einer passenden Basis bestehend aus $n-1$ unabhängigen, zu $\mathbf{x}$ orthogonalen Vektoren angegeben werden. Jeder Untervektorraum von $U^\perp$ ist orthogonal zu $\mathbf{x}$. Wir können einen Punkt in der Hyperebene $U^\perp$ wählen und den dazu orthogonalen Untervektorraum $U_2^\perp$ von $U^\perp$ berechnen, etc. bis wir n paarweise orthogonale Vektoren gefunden haben, die den Vektorraum $\mathbb{R}^n$ aufspannen. Jede Zerlegung der Menge dieser Basisvektoren produziert aufeinander paarweise orthogonale Untervektorräume von $\mathbb{R}^n$.

Definition 5.1.16. *(Winkel zwischen zwei Vektoren oder Punkten) Sei $(\mathbb{R}^n, \cdot)$ ein Euklidischer Vektorraum und $\mathbf{v}, \mathbf{u} \in \mathbb{R}^n, \mathbf{v}, \mathbf{u} \neq 0$. Dann definiert man*

$$\begin{aligned}\cos &: [0, \pi] \to [-1, 1] \\ \cos\gamma &:= \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \, \|\mathbf{v}\|}\end{aligned}$$

γ ist der Winkel im Bogenmass zwischen $\mathbf{v}$ und $\mathbf{u}$. ◊

Bemerkung 5.1.17. *Denkt man sich durch zwei Punkte $\mathbf{u}$ und $\mathbf{v} \in \mathbb{R}^2$ zwei Geraden g_1 und g_2, die durch den Ursprung des Koordinatensystems verlaufen, so ist γ in $\cos\gamma = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}$ der Winkel (im Bogenmass) zwischen diesen zwei Geraden. Alle Punkte $\mathbf{w}_1 \in g_1$ und $\mathbf{w}_2 \in g_2$, die in denselben Quadranten liegen wie $\mathbf{u}$ ($\mathbf{w}_1$ im selben wie $\mathbf{u}$) und $\mathbf{v}$ ($\mathbf{w}_2$ im selben wie $\mathbf{v}$), weisen*

denselben Winkel auf, da für alle solche Punkte gilt $\mathbf{w}_1 = \alpha\mathbf{u}$ *und* $\mathbf{w}_2 = \beta\mathbf{v}$ *für* $\alpha, \beta \in \mathbb{R}^+\backslash\{0\}$. *Damit gilt*

$$\begin{aligned}\frac{\mathbf{w}_1 \cdot \mathbf{w}_2}{\|\mathbf{w}_1\| \|\mathbf{w}_2\|} &= \frac{\alpha\mathbf{u} \cdot \beta\mathbf{v}}{\|\alpha\mathbf{u}\| \|\beta\mathbf{v}\|} \\ &= \frac{\alpha\mathbf{u} \cdot \beta\mathbf{v}}{|\alpha| \, |\beta| \, \|\mathbf{u}\| \|\mathbf{v}\|} \\ &= \frac{\alpha\beta \, (\mathbf{u} \cdot \mathbf{v})}{\alpha\beta \, \|\mathbf{u}\| \|\mathbf{v}\|} \\ &= \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|}\end{aligned}$$

◊

Bemerkung 5.1.18. *Der* $\cos\alpha$ *des Winkels* α *kann definiert werden durch* $\frac{a}{c}$ *im rechtwinkligen Dreieck mit den Seiten* (a, b, c) *so dass* $a \perp b$ *und der Winkel zwischen* a *und* c α *beträgt.* a *ist also die Ankathete des Winkels* α *und* c *ist die Hypothenuse - b ist die Gegenkathete. Für das Verhältnis* $\frac{a}{c}$ *gilt* $\frac{da}{dc} = \frac{a}{c}$. *Entsprechend könnte das Verhältnis am Einheitskreis (Kreis mit Radius 1 untersucht werden, s. Abbildung 5.1.3).Man kann zeigen, dass im* $\mathbb{R}^2$ *die Kosinusfunktion, die*

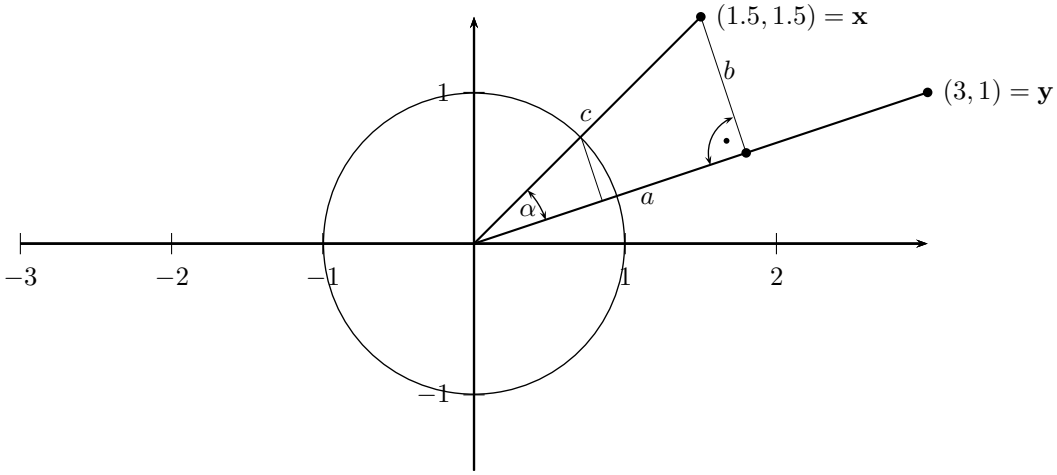


Abbildung 5.1.3: Winkel zwischen den Punkten (Vektoren) $(3, 1)$ und $(1.5, 1.5)$

mittels $\frac{Ankathete}{Hypothenuse}$ *definiert ist, dieselben Werte annimmt wie* $\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}$. *Zeichnet man eine auf* g_1 *Senkrechte durch* $\mathbf{w}_1 \in g_1$, *erhält man den Schnittpunkt* $\lambda\mathbf{w}_2 \in g_2$ *mit* $\lambda \neq 0$, *sofern* g_1 *nicht schon senkrecht auf* g_2 *steht. In letzterem Falle gilt* $\gamma = \frac{\pi}{2}$ *und* $\cos\frac{\pi}{2} = 0$. *Ebenso gilt* $\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} = 0$. *Die Werte stimmen also für diesen Fall überein. Im anderen Fall sei der Einfachheit halber* $\lambda > 0$. *Es gibt ein* $\mathbf{x}$, *so dass* $\mathbf{w}_1 = \lambda\mathbf{w}_2 + \mathbf{x}$ *und* $\mathbf{w}_1 \cdot \mathbf{x} = 0$. *Für dieses* $\mathbf{x}$ *gilt*

$$\mathbf{x} = \mathbf{w}_1 - \lambda\mathbf{w}_2$$

und man erhält

$$\mathbf{w}_1 \cdot (\mathbf{w}_1 - \lambda\mathbf{w}_2) = \mathbf{w}_1 \cdot \mathbf{w}_1 - \lambda \, (\mathbf{w}_1 \cdot \mathbf{w}_2) = 0,$$

d.h. $\lambda = \frac{\mathbf{w}_1 \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_2}$. *Damit gilt*

$$\|\lambda\mathbf{w}_2\| = \lambda \|\mathbf{w}_2\| = \frac{\mathbf{w}_1 \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_2} \|\mathbf{w}_2\| = \frac{\|\mathbf{w}_1\|^2 \|\mathbf{w}_2\|}{\mathbf{w}_1 \cdot \mathbf{w}_2}$$

Im rechtwinkligen Dreieck gilt damit (Ankathete über Hypothenuse)

$$\cos\gamma = \frac{\|\mathbf{w}_1\|}{\|\lambda\mathbf{w}_2\|} = \frac{\|\mathbf{w}_1\|}{\frac{\|\mathbf{w}_1\|^2 \|\mathbf{w}_2\|}{\mathbf{w}_1 \cdot \mathbf{w}_2}} = \|\mathbf{w}_1\| \frac{\mathbf{w}_1 \cdot \mathbf{w}_2}{\|\mathbf{w}_1\|^2 \|\mathbf{w}_2\|} = \frac{\mathbf{w}_1 \cdot \mathbf{w}_2}{\|\mathbf{w}_1\| \|\mathbf{w}_2\|}.$$

◊

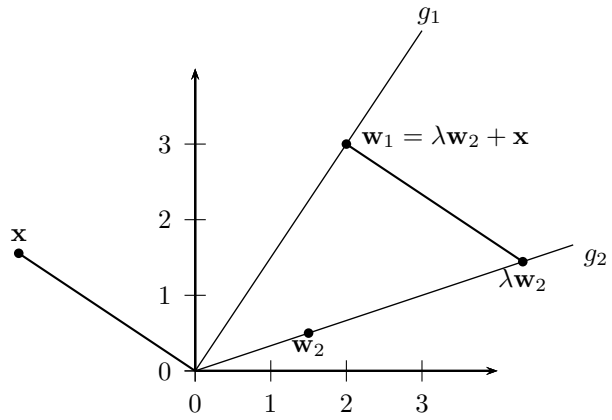


Abbildung 5.1.4: Graphik zum Nachweis der Verträglichkeit der Vektor-Cosinus-Definition mit der geometrischen Definition im $\mathbb{R}^2$

Bemerkung 5.1.19. *Gemäss Satz von Cauchy-Schwarz gilt*

$$\|\mathbf{u}\cdot\mathbf{v}\| \leq \|\mathbf{u}\|\,\|\mathbf{v}\|$$

und damit wegen $\|\mathbf{u}\|\,\|\mathbf{v}\| \geq 0$

$$0 \leq \frac{\|\mathbf{u}\cdot\mathbf{v}\|}{\|\mathbf{u}\|\,\|\mathbf{v}\|} = \left|\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\,\|\mathbf{v}\|}\right| \leq 1$$

Entsprechend liegt

$$-1 \leq \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\,\|\mathbf{v}\|} \leq 1$$

im Wertebereich der durch $\cos\gamma := \frac{Ankathete}{Hypothenuse}$ *auf* $[0,\pi]$ *definierten Funktion.* ◊

Bemerkung 5.1.20. *Im graphischen Beispiel werden die Punkte* $\mathbf{w}_1 = (2,3)$ *und* $\mathbf{w}_2 = (1.5, 0.5)$ *verwendet.* $\lambda = \frac{2^2+3^2}{2\cdot 1.5+3\cdot 0.5} = 2.8889$ *und damit der Punkt, durch den die Gerade, die auf* g_2 *in* $\mathbf{w}_1 = (2,3)$ *senkrecht steht, verläuft:* $2.8889\,(1.5, 0.5) = (4.33333, 1.444444)$ *(*$\mathbf{x} = \mathbf{w}_1 - \lambda\mathbf{w}_2 = (2,3) - 2.8889\,(1.5, 0.5) = \begin{pmatrix} -2.3334 & 1.5556 \end{pmatrix}$*).* ◊

Bemerkung 5.1.21. *Sind* $\mathbf{v}, \mathbf{u} \neq 0$ *orthogonal, so gilt* $\cos\gamma = 0$ *und damit* $\gamma = \frac{\pi}{2}$*. In Grad:* $\frac{x}{360} = \frac{\frac{\pi}{2}}{2\pi}$*, d.h.* $x = 90°$*.* ◊

Beispiel 5.1.22. $(5,3)$ *und* $(-6,10)$ *sind orthogonale Vektoren. Die Punkte* $(5,3)$ *und* $(-6,10)$ *liegen auf orthogonalen Geraden* g_1 *und* g_2*, die durch den Ursprung des Koordinatensystem verlaufen. Für alle* $(x,y) \in g_1$ *und* $(z,r) \in g_2$ *gilt*

$$(x,y)\cdot(z,r) = 0.$$

Für solche Punkte gilt nämlich $(x,y) = \alpha\,(5,3)$ *und* $\beta\,(-6,10)$ *und entsprechend*

$$\alpha\,(5,3)\cdot\beta\,(-6,10) = \alpha\beta((5,3)\cdot(-6,10)) = 0.$$

◊

5.2 Orthonormale Basen und Matrizen

Definition 5.2.1. *(Orthogonal- und Orthonormalbasis) Mengen von orthogonalen Vektoren eines euklidischen Vektorraums werden orthogonale Mengen genannt, wenn die Vektoren paarweise orthogonal sind. Weisen diese die Norm 1 auf, so spricht man von orthonormalen Mengen. Stellt eine orthogonale Menge eine Basis eines euklidischen Vektorraums dar, so wird sie Orthogonalbasis genannt, bei einer orthonormierten Basis spricht man von einer Orthonormalbasis.* ◊

Satz 5.2.2. *(Koordinatendarstellung eines Vektors in einer Orthonormalbasis). Sei* $S := \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ *eine Orthonormalbasis eines euklidischen Vektorraums* $(V, \cdot)$*. Dann gilt für* $\mathbf{u} \in V$

$$\mathbf{u} = \sum_{i=1}^{n} (\mathbf{u} \cdot \mathbf{v}_i)\, \mathbf{v}_i$$

In Matrixdarstellung:

$$\mathbf{u} = \mathbf{V}\mathbf{V}^T\mathbf{u}$$

Beweis. Da S eine Basis von V darstellt, gibt es die eindeutig bestimmte Darstellung ($\alpha_i \in \mathbb{R}$)

$$\mathbf{u} = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i$$

Damit gilt für jedes $k \in N_n^*$

$$\begin{aligned}
\mathbf{u} \cdot \mathbf{v}_k &= \left(\sum_{i=1}^{n} \alpha_i \mathbf{v}_i \right) \mathbf{v}_k \\
&= \mathbf{v}_k \sum_{i=1}^{n} \alpha_i \mathbf{v}_i \\
&= \sum_{i=1}^{n} \alpha_i \left(\mathbf{v}_k \cdot \mathbf{v}_i \right) \\
&= \alpha_k
\end{aligned}$$

(denn $\mathbf{v}_k \cdot \mathbf{v}_i = 0$ für $k \neq i$ und $\mathbf{v}_k \cdot \mathbf{v}_i = 1$ für $k = i$). In Matrixdarstellung ist der Satz offensichtlich: es gilt ja $\mathbf{V}\mathbf{V}^T = \mathbf{E}$. □

Beispiel 5.2.3. $\mathbf{v}_1 = (0, 1, 0)$, $\mathbf{v}_2 = \left(-\frac{4}{5}, 0, \frac{3}{5}\right)$ *und* $\mathbf{v}_3 = \left(\frac{3}{5}, 0, \frac{4}{5}\right)$ *bilden eine Orthonormalbasis des euklidischen Raums* $\mathbb{R}^3$*. Für* $\mathbf{u} = (1, 2, 7)$ *erhält man mit*

$$\begin{aligned}
\mathbf{u} \cdot \mathbf{v}_1 &= 2 \\
\mathbf{u} \cdot \mathbf{v}_2 &= \frac{17}{5} \\
\mathbf{u} \cdot \mathbf{v}_3 &= \frac{31}{5}
\end{aligned}$$

die Darstellung

$$(1, 2, 7) = 2\,(0, 1, 0) + \frac{17}{5}\left(-\frac{4}{5}, 0, \frac{3}{5}\right) + \frac{31}{5}\left(\frac{3}{5}, 0, \frac{4}{5}\right).$$

In Matrixschreibweise:

$$\begin{pmatrix} 0 & -\frac{4}{5} & \frac{3}{5} \\ 1 & 0 & 0 \\ 0 & \frac{3}{5} & \frac{4}{5} \end{pmatrix} \left(\begin{pmatrix} 0 & -\frac{4}{5} & \frac{3}{5} \\ 1 & 0 & 0 \\ 0 & \frac{3}{5} & \frac{4}{5} \end{pmatrix}^T \begin{pmatrix} 1 \\ 2 \\ 7 \end{pmatrix} \right) = \begin{pmatrix} 1 \\ 2 \\ 7 \end{pmatrix}$$

◊

Bemerkung 5.2.4. *Liegt eine orthogonale Basis vor, so kann diese problemlos in eine orthonormale Basis verwandelt werden. Man teilt jeden Basisvektor durch seine Norm. Entsprechend kann der obige Satz auch vorteilhaft auf orthogonale Basen angewendet werden.* ◊

Bemerkung 5.2.5. *Für orthonormale Basen lassen sich also für beliebige Vektoren besonders einfach die Koordinaten bezüglich der Basis angeben.* ◊

Satz 5.2.6. *(Gram und Schmidt) Jede Basis* $(\mathbf{u}_1, ..., \mathbf{u}_n)$ *eines euklidischen Raumes kann in eine Orthogonalbasis umgerechnet werden.*

Beweis. Orthogonalisierungsalgorithmus von Gram und Schmidt:

1. Schritt: $\mathbf{v}_1 := \mathbf{u}_1$
2. Schritt: Man setzt $\mathbf{v}_2 := \mathbf{u}_2 - \lambda_{21}\mathbf{v}_1$ und bestimmt λ_{21} so, dass $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$.

$$\begin{aligned} 0 &= \mathbf{v}_1 \cdot \mathbf{v}_2 \\ &= \mathbf{v}_1 \cdot (\mathbf{u}_2 - \lambda_{21}\mathbf{v}_1) \\ &= \mathbf{v}_1 \cdot \mathbf{u}_2 - \lambda_{21} (\mathbf{v}_1 \cdot \mathbf{v}_1) \end{aligned}$$

   also

$$\lambda_{21} = \frac{\mathbf{v}_1 \cdot \mathbf{u}_2}{\mathbf{v}_1 \cdot \mathbf{v}_1}.$$

   Man erhält für $\mathbf{v}_2$

$$\mathbf{v}_2 = \mathbf{u}_2 - \frac{\mathbf{v}_1 \cdot \mathbf{u}_2}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \mathbf{u}_2 - \frac{\mathbf{v}_1 \cdot \mathbf{u}_2}{\|\mathbf{v}_1\|^2}\mathbf{v}_1.$$

3. Schritt: Seien $(\mathbf{v}_1, ..., \mathbf{v}_{n-1})$ orthogonal. Man setzt $\mathbf{v}_n := \mathbf{u}_n - \sum_{i=1}^{n-1} \lambda_{ni}\mathbf{v}_i$, so dass $\mathbf{v}_n \cdot \mathbf{v}_j = 0$ für $j \in \mathbb{N}^*_{n-1}$, d.h.

$$\begin{aligned} 0 = \mathbf{v}_n \cdot \mathbf{v}_j &= \\ &= \left(\mathbf{u}_n - \sum_{i=1}^{n-1} \lambda_{ni}\mathbf{v}_i\right) \cdot \mathbf{v}_j \\ &= \mathbf{u}_n \cdot \mathbf{v}_j - \sum_{i=1}^{n-1} \lambda_{ni} (\mathbf{v}_i \cdot \mathbf{v}_j) \\ &= \mathbf{u}_n \cdot \mathbf{v}_j - \lambda_{nj} (\mathbf{v}_j \cdot \mathbf{v}_j) \Longrightarrow \\ \lambda_{nj} &= \frac{\mathbf{u}_n \cdot \mathbf{v}_j}{\mathbf{v}_j \cdot \mathbf{v}_j} \end{aligned}$$

   Also ist

$$\mathbf{v}_n = \mathbf{u}_n - \sum_{i=1}^{n-1} \frac{\mathbf{u}_n \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i}\mathbf{v}_i$$

Beispiel 5.2.7. *Zu orthogonalisieren sei*

$$\begin{pmatrix} 1 & 3 & 7 \\ 2 & 4 & 2 \\ 4 & 8 & 3 \end{pmatrix}$$

Man setzt:

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}$$

$$\mathbf{v}_2 = \begin{pmatrix} 3 \\ 4 \\ 8 \end{pmatrix} - \frac{\begin{pmatrix} 3 \\ 4 \\ 8 \end{pmatrix}^T \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}}{\begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}^T \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}} \cdot \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix} = \begin{pmatrix} \frac{20}{21} \\ -\frac{2}{21} \\ -\frac{4}{21} \end{pmatrix}$$

Es gilt

$$\begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}^T \begin{pmatrix} \frac{20}{21} \\ -\frac{2}{21} \\ -\frac{4}{21} \end{pmatrix} = 0$$

$$\mathbf{v}_3 = \begin{pmatrix} 7 \\ 2 \\ 3 \end{pmatrix} - \left(\frac{\begin{pmatrix} 7 \\ 2 \\ 3 \end{pmatrix}^T \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}}{\begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}^T \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}} \cdot \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix} + \frac{\begin{pmatrix} 7 \\ 2 \\ 3 \end{pmatrix}^T \begin{pmatrix} \frac{20}{21} \\ -\frac{2}{21} \\ -\frac{4}{21} \end{pmatrix}}{\begin{pmatrix} \frac{20}{21} \\ -\frac{2}{21} \\ -\frac{4}{21} \end{pmatrix}^T \begin{pmatrix} \frac{20}{21} \\ -\frac{2}{21} \\ -\frac{4}{21} \end{pmatrix}} \cdot \begin{pmatrix} \frac{20}{21} \\ -\frac{2}{21} \\ -\frac{4}{21} \end{pmatrix} \right)$$

$$= \begin{pmatrix} 0 \\ \frac{2}{5} \\ -\frac{1}{5} \end{pmatrix}$$

Es gilt

$$\begin{pmatrix} \frac{20}{21} \\ -\frac{2}{21} \\ -\frac{4}{21} \end{pmatrix}^T \begin{pmatrix} 0 \\ \frac{2}{5} \\ -\frac{1}{5} \end{pmatrix} = 0$$

$$\begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}^T \begin{pmatrix} 0 \\ \frac{2}{5} \\ -\frac{1}{5} \end{pmatrix} = 0$$

◊
□

Satz 5.2.8. *Jeder euklidische Vektorraum besitzt eine orthogonale Basis.*

Beweis. Folgt direkt aus dem vorausgehenden Satz □

Definition 5.2.9. *Sei $(V, \cdot)$ eine euklidischer Vektorraum und W ein Unterraum von V. Eine Abbildung $P : V \to W$ heisst Projektion genau dann, wenn $P(P(\mathbf{v})) = P(\mathbf{v})$.* ◊

Satz 5.2.10. *(orthogonale Projektion auf Unterräume) Sei $(V, \cdot)$ ein euklidischer Vektorraum. Sei $\mathbf{u} \in V$ und W ein Unterraum von V mit der Orthogonalbasis $(\mathbf{v}_1, ..., \mathbf{v}_n)$. Dann gilt für*

$$p_W(\mathbf{u}) := \sum_{i=1}^{n} \frac{\mathbf{u} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i} \mathbf{v}_i$$

(1) $p_W(\mathbf{u}) \in W$, d.h. p_W ist eine Abbildung von V nach W und
(2) p_W ist eine Projektion.
(3) für alle $\mathbf{w} \in W$ gilt: $(\mathbf{u} - p_W(\mathbf{u})) \cdot \mathbf{w} = 0$.

Beweis. (1) $\mathbf{w} := \sum_{i=1}^{n} \frac{\mathbf{u} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i} \mathbf{v}_i$ ist eine Linearkombination der Basisvektoren, die W aufspannen ($\frac{\mathbf{u} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i}$ sind reelle Zahlen!). Also ist $\mathbf{w} \in W$.

(2) Man zeigt zuerst, dass für $\mathbf{w} \in W : p_W(\mathbf{w}) = \mathbf{w}$. Da $\mathbf{w} \in W$, gibt es eine eindeutige Darstellung von $\mathbf{w}$ als $\mathbf{w} = \sum_{i=j}^{n} \alpha_j \mathbf{v}_j$ mit $\alpha_j \in \mathbb{R}$.

$$\begin{aligned} p_W(\mathbf{w}) &= p_W\left(\sum_{i=j}^{n} \alpha_j \mathbf{v}_j\right) \\ &= \sum_{i=1}^{n} \frac{\sum_{j=1}^{n} \alpha_j (\mathbf{v}_j \cdot \mathbf{v}_i)}{\mathbf{v}_i \cdot \mathbf{v}_i} \mathbf{v}_i \\ &= \sum_{i=1}^{n} \frac{\alpha_i (\mathbf{v}_i \cdot \mathbf{v}_i)}{\mathbf{v}_i \cdot \mathbf{v}_i} \mathbf{v}_i \\ &= \sum_{i=1}^{n} \alpha_i \mathbf{v}_i \\ &= \mathbf{w}. \end{aligned}$$

Damit gilt dann, da $p_W(\mathbf{u}) \in W$, $p_W(p_W(\mathbf{u})) = p_W(\mathbf{u}) \in W$ unmittelbar.
(3) Gemäss Verfahren von Gram und Schmidt gilt $\mathbf{u} - \sum_{i=1}^{n} \frac{\mathbf{u} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i} \mathbf{v}_i = \mathbf{u} - p_W(\mathbf{u})$ steht senkrecht auf $(\mathbf{v}_1, ..., \mathbf{v}_n)$. Entsprechend steht $\mathbf{u} - p_W(\mathbf{u})$ auch senkrecht auf $\mathbf{w} \in W$, denn dann ist $\mathbf{w} = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i$ (es gibt eine solche Darstellung mit $\alpha_j \in \mathbb{R}$ für $\mathbf{w} \in W$). Damit gilt

$$\begin{aligned} (\mathbf{u} - p_W(\mathbf{u})) \cdot \mathbf{w} &= (\mathbf{u} - p_W(\mathbf{u})) \cdot \sum_{i=1}^{n} \alpha_i \mathbf{v}_i \\ &= \sum_{i=1}^{n} \alpha_i ((\mathbf{u} - p_W(\mathbf{u})) \cdot \mathbf{v}_i) \\ &= 0. \end{aligned}$$

Dies gilt auch für $\mathbf{u} \in W$. Denn dann gilt $\mathbf{u} - p_W(\mathbf{u}) = \mathbf{u} - \mathbf{u} = 0$ und damit $(\mathbf{u} - p_W(\mathbf{u})) \cdot \mathbf{w} = 0$. □

Bemerkung 5.2.11. *In Matrixschreibweise:* $p_W(\mathbf{u}) = \mathbf{B}\left(\mathbf{A}\mathbf{A}^T\right)\mathbf{u}$ *wobei die Diagonalmatrix* $\mathbf{B}$ *die Kehrwerte der Diagonalkomponenten von* $\mathbf{A}\mathbf{A}^T$ *enthält und* $\mathbf{A}$ *spaltenweise aus den Orthogonal-Basisvektoren* $(\mathbf{v}_1, ..., \mathbf{v}_n)$ *von* W *besteht. Dabei gilt:* $\mathbf{B}\left(\mathbf{A}\mathbf{A}^T\right)$ *ist eine Diagonalmatrix mit Nullen und Einsen auf der Diagonale.* ◊

Beispiel 5.2.12. *Sei* $V = \mathbb{R}^3$ *und* W *der durch die Vektoren* $(3, 0, 1)$ *und* $(1, 0, -3)$ *gebildete Unter-Vektorraum. Die Vektoren stellen eine Orthogonalbasis von* W *dar. Der Punkt* $\mathbf{u} := (2, 4, 1)$ *liegt nicht in* W*. Man erhält mit obiger Formel:*

$$\begin{aligned} p_W(\mathbf{u}) &= \frac{(2,4,1) \cdot (3,0,1)}{(3,0,1) \cdot (3,0,1)} \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix} + \frac{(2,4,1) \cdot (1,0,-3)}{(1,0,-3) \cdot (1,0,-3)} \begin{pmatrix} 1 \\ 0 \\ -3 \end{pmatrix} \\ &= \frac{7}{10} \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix} + \frac{-1}{10} \begin{pmatrix} 1 \\ 0 \\ -3 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} \end{aligned}$$

Und mit Matrixoperationen:

$$\mathbf{A}^T\mathbf{A} = \begin{pmatrix} 3 & 1 \\ 0 & 0 \\ 1 & -3 \end{pmatrix} \begin{pmatrix} 3 & 0 & 1 \\ 1 & 0 & -3 \end{pmatrix} = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 10 \end{pmatrix}$$

$$\begin{pmatrix} 0.1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0.1 \end{pmatrix} \left(\begin{pmatrix} 3 & 1 \\ 0 & 0 \\ 1 & -3 \end{pmatrix} \begin{pmatrix} 3 & 0 & 1 \\ 1 & 0 & -3 \end{pmatrix} \right) \begin{pmatrix} 2 \\ 4 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}$$

Es gilt dabei

$$\begin{pmatrix} 0.1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0.1 \end{pmatrix} \left(\begin{pmatrix} 3 & 1 \\ 0 & 0 \\ 1 & -3 \end{pmatrix} \begin{pmatrix} 3 & 0 & 1 \\ 1 & 0 & -3 \end{pmatrix} \right) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

◊

Bei orthonormaler Basis erhält man $p_W(\mathbf{u}) := \sum_{i=1}^{n} (\mathbf{u} \cdot \mathbf{v}_i)\, \mathbf{v}_i$, in Matrixschreibweise also: $p_W(\mathbf{u}) = \mathbf{A}\left(\mathbf{u}^T\mathbf{A}\right)^T = \mathbf{A}\mathbf{A}^T\mathbf{u}$ wobei $(\mathbf{v}_1, ..., \mathbf{v}_n)$ eine Orthonormalbasis von W ist und die Matrix $\mathbf{A}$ spaltenweise aus den Basisvektoren besteht.

Beispiel 5.2.13. *Sei $V = \mathbb{R}^3$ und W der durch $\left(\frac{1}{10}\sqrt{10}, 0, \frac{3}{10}\sqrt{10}\right)$ und $\left(-\frac{3}{10}\sqrt{10}, 0, \frac{1}{10}\sqrt{10}\right)$ gebildete Unter-Vektorraum (die Vektoren stellen eine Orthonormalbasis von W dar - aus der Orthogonalbasis das vorangehenden Beipiels gewonnen mittels Normierung - Division der Komponenten der Basisvektoren des Beispiels durch $\sqrt{10}$). Der Punkt $\mathbf{u} := (2, 4, 1)$ liegt nicht in W.*

$$\begin{aligned} p_W(\mathbf{u}) &= \left((2,4,1) \begin{pmatrix} \frac{1}{\sqrt{10}} \\ 0 \\ \frac{3}{\sqrt{10}} \end{pmatrix} \right) \begin{pmatrix} \frac{1}{\sqrt{10}} \\ 0 \\ \frac{3}{\sqrt{10}} \end{pmatrix} + \left((2,4,1) \begin{pmatrix} -\frac{3}{\sqrt{10}} \\ 0 \\ \frac{1}{\sqrt{10}} \end{pmatrix} \right) \begin{pmatrix} -\frac{3}{\sqrt{10}} \\ 0 \\ \frac{1}{\sqrt{10}} \end{pmatrix} \\ &= \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} \end{aligned}$$

Mittels Matrixmultiplikation:

$$\begin{pmatrix} \frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{10}} \\ 0 & 0 \\ \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{pmatrix} \begin{pmatrix} \frac{1}{10}\sqrt{10} & 0 & \frac{3}{10}\sqrt{10} \\ -\frac{3}{10}\sqrt{10} & 0 & \frac{1}{10}\sqrt{10} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

und damit

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 4 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}$$

◊

Zu zeigen ist noch die Eindeutigkeit der Projektion.

Definition 5.2.14. *(orthogonales Komplement). Sei $(V, \cdot)$ ein euklidischer Vektorraum und W ein Unterraum von V, dann wird*

$$W^{\perp} := \{\mathbf{v} \in V | \mathbf{v} \cdot \mathbf{w} = 0 \text{ für alle } \mathbf{w} \in W\}$$

als orthogonales Komplement von W bezeichnet. ◊

Satz 5.2.15. *(Projektionssatz) Sei W ein Unterraum des euklidischen Vektorraums $(V,\cdot)$. Dann besitzt jedes $\mathbf{v}\in V$ eine eindeutige Darstellung*

$$\mathbf{v}=\mathbf{w}_1+\mathbf{w}_2$$

so dass $\mathbf{w}_1\in W$ und $\mathbf{w}_2\in W^\perp$.

Beweis. Die Existenz einer solchen Zerlegung ergibt sich aus dem Gram-Schmidt-Algorithmus, denn es gilt für $\mathbf{v}\in V$

$$\mathbf{w}_1:=p_W(\mathbf{v})\in W.$$

Zudem ist

$$\mathbf{w}_2:=\mathbf{v}-p_W(\mathbf{v})$$

zu $p_W(\mathbf{v})$ orthogonal, also ist $\mathbf{w}_2\in W^\perp$. Nun gilt

$$\mathbf{w}_1+\mathbf{w}_2=p_W(\mathbf{v})+\mathbf{v}-p_W(\mathbf{v})=\mathbf{v}.$$

Eindeutigkeit: seien $\mathbf{w}_1,\mathbf{w}_1'\in W$ und $\mathbf{w}_2,\mathbf{w}_2'\in W^\perp$ mit

$$\mathbf{w}_1+\mathbf{w}_2=\mathbf{w}_1'+\mathbf{w}_2'=\mathbf{v}$$

$$\mathbf{w}_1-\mathbf{w}_1'=\mathbf{w}_2-\mathbf{w}_2'$$

Damit gilt wegen der Abgeschlossenheit von Vektorräumen

$$\mathbf{w}_2-\mathbf{w}_2'\in W^T$$

und wegen $\mathbf{w}_1-\mathbf{w}_1'=\mathbf{w}_2-\mathbf{w}_2'$ auch

$$\mathbf{w}_2-\mathbf{w}_2'\in W.$$

Damit gilt

$$(\mathbf{w}_2'-\mathbf{w}_2)\cdot(\mathbf{w}_2'-\mathbf{w}_2)=0$$

womit

$$\mathbf{w}_2'-\mathbf{w}_2=\mathbf{0}=\mathbf{w}_1-\mathbf{w}_1'$$

Also ist $\mathbf{w}_2'=\mathbf{w}_2$ und $\mathbf{w}_1'=\mathbf{w}_1$. □

Satz 5.2.16. *(Approximationssatz) Sei $(V,\cdot)$ ein euklidischer Vektorraum und W ein Untervektorraum. Zu jedem $\mathbf{v}\in V$ ist $p_W(\mathbf{v})$ die beste Approximation von $\mathbf{v}$ in W, d.h.*

$$\|\mathbf{v}-p_W(\mathbf{v})\|<\|\mathbf{v}-\mathbf{w}\|$$

für alle $\mathbf{w}\in W$ mit $\mathbf{w}\neq p_W(\mathbf{v})$ (d.h. $p_W(\mathbf{v})$ liegt $\mathbf{v}$ näher als alle anderen, von $p_W(\mathbf{v})$ verschiedenen $\mathbf{w}\in W$).

Beweis.

$$\|\mathbf{v}-\mathbf{w}\|^2=\|\mathbf{v}-p_W(\mathbf{v})+p_W(\mathbf{v})-\mathbf{w}\|^2$$

Es gilt $\mathbf{v}-p_W(\mathbf{v})\in W^\perp$ und $p_W(\mathbf{v})-\mathbf{w}\in W$ (s. Satz 5.2.10), d.h. die beiden Vektoren sind orthogonal zueinander. Damit gilt mit dem Satz von Pythagoras

$$\begin{aligned}\|\mathbf{v}-\mathbf{w}\|^2&=\|\mathbf{v}-p_W(\mathbf{v})+p_W(\mathbf{v})-\mathbf{w}\|^2\\&=\|\mathbf{v}-p_W(\mathbf{v})\|^2+2(\mathbf{v}-p_W(\mathbf{v}))\cdot(p_W(\mathbf{v})-\mathbf{w})+\|p_W(\mathbf{v})-\mathbf{w}\|^2\\&=\|\mathbf{v}-p_W(\mathbf{v})\|^2+\|p_W(\mathbf{v})-\mathbf{w}\|^2\\&>\|\mathbf{v}-p_W(\mathbf{v})\|^2\end{aligned}$$

für $\mathbf{w}\neq p_W(\mathbf{v})$.
Damit ist $\|\mathbf{v}-p_W(\mathbf{v})\|<\|\mathbf{v}-\mathbf{w}\|$ für alle $\mathbf{w}\neq\mathbf{v}$. □

Definition 5.2.17. *Eine Matrix* $\mathbf{Q} \in \mathbb{R}^{n\times n}$ *aus orthonormalen Spaltenvektoren wird orthogonale Matrix (besser wäre orthonormale, habe sich aber nicht durchgesetzt) genannt. Man definiert ferner*

$$O(n) := \{\mathbf{Q} \in \mathbb{R}^{n\times n} | \mathbf{Q} \text{ ist orthogonal}\}$$

$\Diamond$

Die Spalten einer orthogonalen Matrix bilden eine Orthonormalbasis des $\mathbb{R}^n$.

Satz 5.2.18. *Folgende Aussagen sind äquivalent*

1. $\mathbf{Q} \in O(n)$
2. $\mathbf{Q}$ *ist invertierbar und* $\mathbf{Q}^{-1} = \mathbf{Q}^T$
3. $(\mathbf{Qu}) \cdot (\mathbf{Qv}) = \mathbf{u} \cdot \mathbf{v}$
4. $\|\mathbf{Qv}\| = \|\mathbf{v}\|$ *(die durch* $\mathbf{Q}$ *definierte Abbildung wird deshalb „Isometrie" genannt).*

Beweis.

1. (1) $\Longrightarrow$ (2) Es gilt $\mathbf{Q}^T\mathbf{Q} = \mathbf{E}$.
2. (2) $\Longrightarrow$ (1) Mit $\mathbf{Q}^{-1} = \mathbf{Q}^T$ folgt $\mathbf{Q}^T\mathbf{Q} = \mathbf{E}$ und damit sind die Spaltenvektoren von $\mathbf{Q}$ zueinander orthonormal.
3. (1) $\Longrightarrow$ (3) $(\mathbf{Qu}) \cdot (\mathbf{Qv}) = (\mathbf{Qu})^T (\mathbf{Qv}) = \mathbf{u}^T\mathbf{Q}^T\mathbf{Qv} = \mathbf{u}^T\mathbf{v} = \mathbf{u} \cdot \mathbf{v}$
4. (3) $\Longrightarrow$ (1) Gilt $(\mathbf{Qu}) \cdot (\mathbf{Qv}) = \mathbf{u} \cdot \mathbf{v}$ für beliebige $\mathbf{u}$ und $\mathbf{v}$, so gilt dies auch für die Standardbasisvektoren e_n^i. Damit gilt dann
$$\left(\mathbf{Q}e_n^i\right) \cdot \left(\mathbf{Q}e_n^j\right) = e_n^i \cdot e_n^j = \left\{ \begin{array}{cc} 1 & \text{für } i = j \\ 0 & \text{sonst.} \end{array} \right|$$
Da zudem für $\mathbf{Q} = (\mathbf{q}_1, ..., \mathbf{q}_n)$ gilt
$$\left(\mathbf{Q}e_n^i\right) = \mathbf{q}_i$$
gilt
$$\left(\mathbf{Q}e_n^i\right) \cdot \left(\mathbf{Q}e_n^j\right) = \mathbf{q}_i \cdot \mathbf{q}_j = \left\{ \begin{array}{cc} 1 & \text{für } i = j \\ 0 & \text{sonst.} \end{array} \right|$$
womit $\mathbf{Q} \in O(n)$.
5. (3) $\Longrightarrow$ (4) $\|\mathbf{Qv}\| = \sqrt{(\mathbf{Qu}) \cdot (\mathbf{Qu})} = \sqrt{\mathbf{u} \cdot \mathbf{u}} = \|\mathbf{u}\|$
6. (4) $\Longrightarrow$ (3) Wegen (4) gilt
$$(\mathbf{Q}(\mathbf{u} + \mathbf{v})) \cdot (\mathbf{Q}(\mathbf{u} + \mathbf{v})) = (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v})$$
$$(\mathbf{Q}(\mathbf{u} - \mathbf{v})) \cdot (\mathbf{Q}(\mathbf{u} - \mathbf{v})) = (\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v})$$
und damit
$$(\mathbf{Qu}) \cdot (\mathbf{Qu}) + 2(\mathbf{Qu}) \cdot (\mathbf{Qv}) + (\mathbf{Qv}) \cdot (\mathbf{Qv}) = \mathbf{u} \cdot \mathbf{u} + 2\mathbf{u} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{v}$$
$$(\mathbf{Qu}) \cdot (\mathbf{Qu}) - 2(\mathbf{Qu}) \cdot (\mathbf{Qv}) + (\mathbf{Qv}) \cdot (\mathbf{Qv}) = \mathbf{u} \cdot \mathbf{u} - 2\mathbf{u} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{v}$$
und mittels Subtraktion
$$4(\mathbf{Qu}) \cdot (\mathbf{Qv}) = 4\mathbf{u} \cdot \mathbf{v}$$

$\square$

Satz 5.2.19. *Für* $\mathbf{A}, \mathbf{B} \in O(n)$ *gilt:* $\mathbf{AB} \in O(n)$

Beweis. Auf Grund der Voraussetzung gilt: $\mathbf{A}^T = \mathbf{A}^{-1}$. Damit gilt für $\mathbf{A}, \mathbf{B} \in O(n)$

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} = \mathbf{B}^T\mathbf{A}^T = (\mathbf{AB})^T .$$

□

Satz 5.2.20. *Ist* $\mathbf{A} \in O(n)$*, ist* $\mathbf{A}^T \in O(n)$.

Beweis. Auf Grund der Voraussetzung gilt: $\mathbf{A} = \left(\mathbf{A}^T\right)^T$ und $\mathbf{A} = \left(\mathbf{A}^{-1}\right)^{-1}$ und damit wegen $\mathbf{A}^T = \mathbf{A}^{-1}$ die Gleichung $\left(\mathbf{A}^T\right)^T = \left(\mathbf{A}^T\right)^{-1}$, d.h. $\mathbf{A}^T \in O(n)$. □

Satz 5.2.21. *Für* $\mathbf{Q} \in O(n)$ *gilt* $|\det \mathbf{Q}| = |1|$

Beweis. Aus $\mathbf{Q}^T\mathbf{Q} = \mathbf{E}$ folgt

$$1 = \det \mathbf{E} = \det\left(\mathbf{Q}^T\mathbf{Q}\right) = \det\left(\mathbf{Q}^T\right) \det \mathbf{Q} = (\det \mathbf{Q})^2$$

(gemäss den Sätzen: Die Determinante eines Produktes von Matrizen ist mit dem Produkt der Determinanten der Matrizen identisch. Die Determinante der transponierten Matrix ist mit der Determinante der Matrix identisch). □

Definition 5.2.22. *Ist* $\det \mathbf{Q} = 1$*, so wird eine orthogonale Matrix „speziell orthogonal" genannt.* $\mathbf{Q} \in SO(n)$. ◊

Bemerkung 5.2.23. *Eine Matrix aus* $O(n)$ *beschreibt eine Drehung des euklidischen Raumes* $\mathbb{R}^n$ *(Ursprung 0 bleibt fest). Matrizen aus* $SO(n)$ *definieren orientierungserhaltende Transformationen (der Umlaufsinn im* $\mathbb{R}^2$ *oder die „Händigkeit" im* $\mathbb{R}^3$ *bleiben erhalten. Es handelt sich um Transformationen, die eine stetige Überführung des Anfangs in den Endzustand ermöglichen. Das sind beliebige Drehungen um* 0 *im* $\mathbb{R}^n$*. Matrizen aus* $O(n) \setminus SO(n)$ *definieren hingegen orientierungsumkehrende Transformationen (Drehungen verknüpft mit einer Spiegelung).* ◊

Satz 5.2.24. *Die Matrix, welche die Koordinatentransformation bezüglich orthonormaler Basen beschreibt, ist orthogonal.*

Beweis. Da mit den Orthonormalbasen $V_B := (\mathbf{v}_1, ..., \mathbf{v}_n)$ und $W_B := (\mathbf{w}_1, ..., \mathbf{w}_n)$ des $\mathbb{R}^n$ die Transformationsmatrix $\mathbf{Q} = \mathbf{W}^T\mathbf{V} = \mathbf{W}^{-1}\mathbf{V}$ (die Matrizen W und V aus den Basisvektoren gebildet) ist und $\mathbf{W}$ und $\mathbf{V}$ orthogonale Matrizen sind, ist $\mathbf{Q}$ als deren Produkt ebenfalls orthogonal. □

Beispiel 5.2.25. *Multipliziert man Punkte, die im Standardkoordinatensystem gegeben sind, mit einer orthogonalen Matrix* $\mathbf{A}$ *mit* $\det \mathbf{A} = 1$*, wird jeder Punkt auf einem Kreis mit dem Koordinatenursprung als Mittelpunkt verschoben. Im folgenden Beispiel wird der Punkte* $(5, 3)$ *um 10 Grad im Uhrzeigersinn verschoben. Die Abbildung liefert die Koordinaten des Punktes im Standardkoordinatensystem. Diese Koordinaten sind identisch mit den Koordinaten des alten Punktes (des Argumentes) bezüglich der Achsen, die durch die orthogonale Matrix gegeben sind (man kann sich die Sache also auch so vorstellen, dass man den alten Punkt im neuen Koordinatensystem samt diesem ins Standardkoordinatensystem dreht). Im Beispiel betrachte man die orthogonale Matrix*

$$\mathbf{T} := \begin{pmatrix} \cos\left(10^\circ\right) & -\sin\left(10^\circ\right) \\ \sin\left(10^\circ\right) & \cos\left(10^\circ\right) \end{pmatrix}$$

Diese ist für beliebige Winkeln φ *orthogonal. Es gilt nämlich spaltenweise*

$$(\cos\varphi)^2 + (\sin\varphi)^2 = 1$$
$$(\cos\varphi)^2 + (-\sin\varphi)^2 = 1.$$

Zudem gilt

$$\det\begin{pmatrix} \cos\left(10^{\circ}\right) & -\sin\left(10^{\circ}\right) \\ \sin\left(10^{\circ}\right) & \cos\left(10^{\circ}\right) \end{pmatrix} = \cos^2 10^{\circ} + \sin^2 10^{\circ} = 1.$$

Zuletzt ist

$$\begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix}^T \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Im Bogenmass erhält man für $\varphi = 10^{\circ}$*:*

$$\begin{pmatrix} \cos\left(\frac{10}{180}\pi\right) & -\sin\left(\frac{10}{180}\pi\right) \\ \sin\left(\frac{10}{180}\pi\right) & \cos\left(\frac{10}{180}\pi\right) \end{pmatrix} = \begin{pmatrix} 0.984\,81 & -0.173\,65 \\ 0.173\,65 & 0.984\,81 \end{pmatrix}$$

Wir multiplizieren $\mathbf{T}^T$ *mit* $(5,3)^T$ *und man erhält:*

$$\mathbf{T}^T\mathbf{x} = \begin{pmatrix} 0.984\,81 & -0.173\,65 \\ 0.173\,65 & 0.984\,81 \end{pmatrix}^T \begin{pmatrix} 5 \\ 3 \end{pmatrix} = \begin{pmatrix} 5.445 \\ 2.086\,2 \end{pmatrix}$$

Wir tragen den alten (nicht ausgefülltem Punkt) und den neuen Punkt (ausgefüllter Punkt) samt den neuen Achsen

$$a_1 = \left\{ a_1^* | a_1^* = b \begin{pmatrix} 0.984\,81 \\ 0.173\,65 \end{pmatrix} \text{ und } b \in \mathbb{R} \right\}$$
$$a_2 = \left\{ a_2^* | a_2^* = b \begin{pmatrix} -0.173\,65 \\ 0.984\,81 \end{pmatrix} \text{ und } b \in \mathbb{R} \right\}$$

in ein Koordinatensystem:

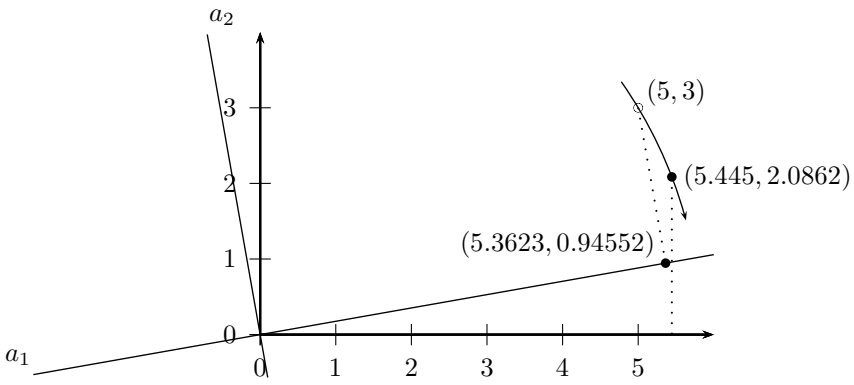


Abbildung 5.2.5: Drehung des Punktes $(5,3)$ um 10 Grad mit Hilfe einer orthogonalen Matrix

Man kann feststellen, dass der Punkt $(5,3)$ *durch die Multiplikation mit* $\mathbf{T}$ *in den Punkt* $(5.445, 2.086)$ *abgebildet wird. Die Projektion von* $(5,3)$ *auf die* a_1*-Achse ist (in Standardkoordinaten)*

$$P_{a_1}(\mathbf{x}_3) := \left(\begin{pmatrix} 5 & 3 \end{pmatrix} \begin{pmatrix} 0.984\,81 \\ 0.173\,65 \end{pmatrix} \right) \begin{pmatrix} 0.984\,81 \\ 0.173\,65 \end{pmatrix} = \begin{pmatrix} 5.362\,3 \\ 0.945\,52 \end{pmatrix}$$

Die orthogonale Distanz des Punktes $(5,3)$ *zur Achse* a_1 *(Länge der Strecke zwischen* $(5,3)$ *und* $(5.362, 0.946)$*) ist identisch mit der Länge der Strecke zwischen* $(5.445, 2.086)$ *und* $(5.445, 0)$*. Dies ist eine Folge der Isometrieeigenschaft von Funktionen, die durch orthogonale Matrizen definiert sind. Die Distanz zwischen dem Nullpunkt und* $(5.362, 0.946)$ *beträgt* $\sqrt{5.362^2 + 0.946^2} = 5.445$*, ist also mit der ersten Koordinate von* $(5.445, 2.086)$ *identisch.* ◊

Bemerkung 5.2.26. *Werden mehrere Punkte abgebildet, so bleibt die Summe der quadrierten Koordinaten der Daten pro Achse durch die Transformation erhalten. Die Streuung wird aber unterschiedlich auf die Achsen aufgeteilt. Im Beispiel gilt für den einzelnen Punkt*

$$5^2 + 3^2 = 34$$

Für das transformierte Datum:

$$5.445^2 + 2.086^2 = 34$$

Dies ist intuitiv klar, da die Punkte auf demselben Kreis verbleiben und damit x^2+y^2 konstant ist. ◊

5.3 Eigenvektoren und Eigenwerte

Definition 5.3.1. *Sei f ein Endomorphismus auf den $K-$Vektorraum V. Dann ist $\lambda \in K$ eine Eigenwert von f genau dann, wenn es ein $\mathbf{v} \neq 0$ gibt, so dass $f(\mathbf{v}) = \lambda\mathbf{v}$.*
$\mathbf{v} \neq \mathbf{0}$ wird Eigenvektor von f zu λ genannt. ◊

Definition 5.3.2. *$Eig(f, \lambda) = \{\mathbf{v} \mid f(\mathbf{v}) = \lambda\mathbf{v}\}$ (Eigenraum von f bezüglich λ, der Eigenraum enthält nicht nur Eigenvektoren, sondern zusätzlich den Vektor $\mathbf{0}$)* ◊

Satz 5.3.3. *a) $Eig(f, \lambda) \subset V$ ist ein Untervektorraum von V*
b) λ ist Eigenwert von f genau dann, wenn $Eig(f, \lambda) \neq \{\mathbf{0}\}$
c) $Eig(f, \lambda)\backslash\{\mathbf{0}\}$ ist die Menge der zu λ gehörigen Eigenvektoren von f.
d) $Kern(f - \lambda Id_V) = Eig(f, \lambda)$
e) Für $\lambda_i \neq \lambda_j$ gilt $Eig(f, \lambda_i) \cap Eig(f, \lambda_j) = \{\mathbf{0}\}$

Beweis. a) für $a\mathbf{v} + b\mathbf{w}$ mit $\mathbf{v}, \mathbf{w} \in Eig(f, \lambda)$ gilt:
$f(a\mathbf{v} + b\mathbf{w}) = af(\mathbf{v}) + bf(\mathbf{w}) = a\lambda\mathbf{v} + b\lambda\mathbf{w} = \lambda(a\mathbf{v} + b\mathbf{w})$. Somit ist $a\mathbf{v} + b\mathbf{w} \in Eig(f, \lambda)$
b) Sei λ Eigenwert von f. Damit gibt es gemäss Definition $\mathbf{v} \neq \mathbf{0}$, so dass $f(\mathbf{v}) = \lambda\mathbf{v}$. Somit ist $Eig(f, \lambda) \neq \{\mathbf{0}\}$. Sei umgekehrt $Eig(f, \lambda) \neq \{\mathbf{0}\}$. Dann gibt es gemäss Definition ein λ und ein $\mathbf{v}$ so dass $f(\mathbf{v}) = \lambda\mathbf{v}$. Also ist λ ein Eigenwert von f.
c) Gilt gemäss Definition.
d) Sei $\mathbf{v} \in Eig\,(f, \lambda)$. Damit gilt $(f - \lambda Id_V)\,(\mathbf{v}) = f(\mathbf{v}) - \lambda Id_V(\mathbf{v}) = f(\mathbf{v}) - \lambda\mathbf{v} = \lambda\mathbf{v} - \lambda\mathbf{v} = \mathbf{0}$
e) Sei $\mathbf{v} \in Eig(f, \lambda_i) \cap Eig(f, \lambda_j)$. Damit gilt $f(\mathbf{v}) = \lambda_i\mathbf{v}$ und $f(\mathbf{v}) = \lambda_j\mathbf{v}$ und damit $f(\mathbf{v}) - f(\mathbf{v}) = \mathbf{0} = \lambda_i\mathbf{v} - \lambda_j\mathbf{v} = (\lambda_i - \lambda_j)\mathbf{v}$. Dies führt auf einen Widerspruch, da $\mathbf{v} \neq \mathbf{0}$ und $\lambda_i - \lambda_j \neq 0$, denn $\mathbf{v}$ ist Eigenvektor und damit von $\mathbf{0}$ verschieden. Zudem ist λ_i von λ_j verschieden, gemäss Voraussetzung. □

Satz 5.3.4. *Sei $\mathcal{A} = (\mathbf{v}_1, ..., \mathbf{v}_n)$ eine Basis von V bestehend aus Eigenvektoren $\mathbf{v}_i$ zu den Eigenwerten $(\lambda_1, ..., \lambda_n)$ von f. Dann ist*

$$M_{\mathcal{A}}(f) = (\lambda_i \cdot \delta_{ij})_n$$

($M_{\mathcal{A}}(f)$ ist Diagonalmatrix mit den Eigenwerten auf der Diagonalen; δ_{ij} ist das Kroneckersymbol: $\delta_{ij} = 1$ wenn $i = j$, sonst $\delta_{ij} = 0$).

Beweis. Es gilt $f(\mathbf{v}_j) = \lambda_j\mathbf{v}_j = \sum_{i=1}^{n} (\lambda_i \cdot \delta_{ij})\,\mathbf{v}_i$, womit der Satz auf Grund der Definition der Matrix $M_{\mathcal{A}}(f)$ unmittelbar gilt. □

Satz 5.3.5. *Sei $\mathcal{A} = (\mathbf{v}_1, ..., \mathbf{v}_n)$ eine Basis von V bestehend aus Eigenvektoren $\mathbf{v}_i$ zu den Eigenwerten $\lambda := (\lambda_1, ..., \lambda_n)$ von f und sei $\mathbf{C} := M_{\mathcal{A}}(f)$, dann gilt für Eigenvektoren $\mathbf{v} \in Eig(f, \lambda)$: $f(\mathbf{v}) = \mathbf{ACA}^{-1}\mathbf{v} = \lambda\mathbf{v}$*

Beweis. Gilt gemäss Satz 4.3.4 von S. 91. □

Bemerkung 5.3.6. *Eigenwerte und Eigenvektoren erlauben es, die Auswirkungen einer Abbildung besonders einfach zu beschreiben. Multiplikation einer Matrix mit einem Eigenvektor ist mit dem Produkt eines Skalars mit dem Eigenvektor identisch. Im $\mathbb{R}^n$ liegt damit $f(\mathbf{v})$ für $\mathbf{v} \in Eig(f, \lambda)$ auf derselben Geraden wie $\mathbf{v}$.* ◊

Bemerkung 5.3.7. *Auf dem Hintergrund einer Basis aus Eigenvektoren $(\mathbf{v}_1, .., \mathbf{v}_n)$ mit $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$ gilt für $\mathbf{u} \in \mathbf{V}$*

$$\mathbf{u} = \sum_{i=1}^{n} a_i \mathbf{v}_i$$

und damit

$$\begin{aligned} \mathbf{A}\mathbf{u} &= \mathbf{A}\left(\sum_{i=1}^{n} a_i \mathbf{v}_i\right) \\ &= \sum_{i=1}^{n} a_i \mathbf{A}\mathbf{v}_i = \sum_{i=1}^{n} a_i \lambda_i \mathbf{v}_i. \end{aligned}$$

◊

Berechnung der Eigenwerte Auf Grund der Definition gilt

$$\mathbf{A}\mathbf{v} - \lambda\mathbf{v} = 0$$

und damit

$$(\mathbf{A} - \lambda\mathbf{E})\,\mathbf{v} = 0$$

Diese Gleichung besitzt eine nicht-triviale Lösung genau dann, wenn $\det(\mathbf{A} - \lambda\mathbf{E}) = 0$ (Wäre $\det(\mathbf{A} - \lambda\mathbf{E}) \neq 0$, wäre $\mathbf{A} - \lambda\mathbf{E}$ injektiv, wobei dann $(\mathbf{A} - \lambda\mathbf{E})\,\mathbf{x} = \mathbf{0}$ nur dann gilt, wenn $\mathbf{x} = \mathbf{0}$. Es gilt nämlich $(\mathbf{A} - \lambda\mathbf{E})\,\mathbf{0} = \mathbf{0}$. Ist nun $\mathbf{x} \neq \mathbf{0}$, ist bei einer injektiven Abbildung $(\mathbf{A} - \lambda\mathbf{E})\,\mathbf{x} \neq \mathbf{0}$. Da $\mathbf{x} = \mathbf{0}$ ausgeschlossen wird - nicht-triviale Lösung - so muss $\det(\mathbf{A} - \lambda\mathbf{E}) = 0$ sein). $\det(\mathbf{A} - \lambda\mathbf{E})$ ist ein Polynom n-ten Grades in λ (das charakteristische Polynom). Seine Nullstellen sind die gesuchten Eigenwerte. Kommen im faktorisierten charakteristischen Polynom

$$\prod_{i=1}^{n} (x - \lambda_i)$$

identische Ausdrücke $(x - \lambda_i)$ vor, so wird die Anzahl der identischen Ausdrücke „Vielfachheit von λ_i“ genannt.

Beispiel 5.3.8.

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 6 & 1 \end{pmatrix}$$

$$\begin{aligned} \mathbf{A} - \lambda\mathbf{E} &= \begin{pmatrix} 2 & 1 \\ 6 & 1 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \\ &= \begin{pmatrix} 2-\lambda & 1 \\ 6 & 1-\lambda \end{pmatrix} \end{aligned}$$

und

$$\begin{aligned} \det(\mathbf{A} - \lambda\mathbf{E}) &= (2-\lambda)(1-\lambda) - 6 \\ &= \lambda^2 - 3\lambda - 4 \end{aligned}$$

mit den Nullstellen $4, -1$, d.h. $\lambda_1 = 4$, $\lambda_2 = -1$ ◊

Beispiel 5.3.9.

$$\mathbf{A} = \begin{pmatrix} 3 & 4 \\ -4 & 3 \end{pmatrix}$$

$$\begin{aligned} \mathbf{A} - \lambda\mathbf{E} &= \begin{pmatrix} 3 & 4 \\ -4 & 3 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \\ &= \begin{pmatrix} 3-\lambda & 4 \\ -4 & 3-\lambda \end{pmatrix} \end{aligned}$$

und

$$\begin{aligned} \det(\mathbf{A} - \lambda\mathbf{E}) &= (3-\lambda)(3-\lambda) + 16 \\ &= \lambda^2 - 6\lambda + 25 \end{aligned}$$

ohne reelle Nullstellen. Es gibt also keine reellen Eigenwerte. Das ist plausibel: $\mathbf{A}$ *ist eine Rotationsmatrix. Alle Vektoren werden gedreht. Entsprechend wird keinem Vektor ein Vielfaches des Vektors zugeordnet. Es gibt allerdings zwei konjugiert komplexe Eigenwerte:* $3 + 4i, 3 - 4i$*. Eine Matrix mit nur reellen Einträgen kann also komplexe Eigenwerte aufweisen. Im Falle* $\mathbb{K} = \mathbb{C}$ *kann man zeigen, dass* $\det \mathbf{A}$ *das Produkt der Eigenwerte ist. Allgemein ist* $\mathbf{A}$ *genau dann invertierbar, wenn* 0 *kein Eigenwert von* $\mathbf{A}$ *ist.* ◊

Um den bestimmten Artikel in der Definition zu rechtfertigen, ist noch der folgende Satz nötig:

Satz 5.3.10. *Ähnliche Matrizen haben dasselbe charakteristische Polynom und damit dieselben Eigenwerte.*

Beweis. Sei $\mathbf{B} = \mathbf{SAS}^{-1}$. Damit ist $\mathbf{B}$ zu $\mathbf{A}$ ähnlich. Es gilt

$$\mathbf{SES}^{-1} = \mathbf{E}$$

$$\begin{aligned} \mathbf{B} - \mathbf{E} &= \mathbf{SAS}^{-1} - \mathbf{E} \\ &= \mathbf{SAS}^{-1} - \mathbf{SES}^{-1} \\ &= \mathbf{S}(\mathbf{A} - \mathbf{E})\mathbf{S}^{-1} \end{aligned}$$

Für Determinanten gilt nun:

$$\begin{aligned} \det(\mathbf{B} - \mathbf{E}) &= \det\left(\mathbf{S}(\mathbf{A} - \mathbf{E})\mathbf{S}^{-1}\right) \\ &= \det \mathbf{S} \cdot \det(\mathbf{A} - \mathbf{E}) \cdot \det \mathbf{S}^{-1} \\ &= \det(\mathbf{A} - \mathbf{E}) \end{aligned}$$

Damit haben $\mathbf{B}$ und $\mathbf{A}$ dasselbe charakteristische Polynom und damit dieselben Eigenwerte. □

Beispiel 5.3.11. *Sei die Basis*

$$\mathcal{B} = \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \end{pmatrix} \right)$$

gegeben und der Endomorphismus f *definiert durch*

$$\begin{aligned} f\begin{pmatrix} 1 \\ 2 \end{pmatrix} &= \begin{pmatrix} 5 \\ 8 \end{pmatrix} \\ f\begin{pmatrix} 3 \\ 4 \end{pmatrix} &= \begin{pmatrix} 7 \\ 6 \end{pmatrix} \end{aligned}$$

Damit erhält man mit

$$5 = a_{11} + 3a_{21}$$
$$8 = 2a_{11} + 4a_{21}$$

$a_{11} = 2, a_{21} = 1$ *und*

$$3 = a_{12} + 3a_{22}$$
$$4 = 2a_{12} + 4a_{22}$$

$a_{12} = 0, a_{22} = 1$

$$M_{\mathcal{B}} = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}$$

Man erhält also die ähnliche Matrix

$$\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}^{-1} = \begin{pmatrix} -7 & 6 \\ -12 & 10 \end{pmatrix}$$

Es gilt

$$\begin{pmatrix} -7 & 6 \\ -12 & 10 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -\lambda - 7 & 6 \\ -12 & 10 - \lambda \end{pmatrix}$$

und damit das charakteristische Polynome: $(-\lambda - 7)(10 - \lambda) - (6 \cdot -12) = \lambda^2 - 3\lambda + 2$. *Die Eigenwerte sind also die Lösungen von*

$$\lambda^2 - 3\lambda + 2 = 0,$$

nämlich $2, 1$. *Analog erhält man für*

$$\begin{pmatrix} 0 & 1 \\ -2 & 3 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -\lambda & 1 \\ -2 & 3 - \lambda \end{pmatrix}$$

das charakteristische Polynom $(-\lambda)(3 - \lambda) - (1 \cdot -2) = \lambda^2 - 3\lambda + 2$. *Eine weitere ähnliche Matrix wäre z.B. - mit der Basis*

$$\mathcal{B} = \begin{pmatrix} 3 & 4 \\ 1 & , & 5 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 4 \\ 1 & 5 \end{pmatrix}^{-1} \begin{pmatrix} 0 & 1 \\ -2 & 3 \end{pmatrix} \begin{pmatrix} 3 & 4 \\ 1 & 5 \end{pmatrix} = \begin{pmatrix} \frac{17}{11} & -\frac{3}{11} \\ -\frac{10}{11} & \frac{16}{11} \end{pmatrix}$$

Nachrechnen ergibt wiederum das charakteristische Polynom $\lambda^2 - 3\lambda + 2$. ◊

Bemerkung 5.3.12. *Der Satz bedeutet, dass jedem Endomorphismus eindeutig ein charakteristisches Polynom zugeordet wird.* ◊

Satz 5.3.13. *(Eigenwerte einer Dreiecksmatriz) Sei* $\mathbf{A} \in \mathbb{K}$ *eine obere oder untere Dreiecksmastrix. So sind die Eigenwerte durch die Diagonaleinträge gegeben.*

Beweis. Die Determinante einer Dreiecksmatrix ist das Produkte ihrer Diagonaleinträge. Ist $\mathbf{A}$ eine Dreiecksmatrix, so ist $\mathbf{A} - \lambda\mathbf{E}$ eine Dreiecksmatrix. Die Determinante ist damit $(a_{11} - \lambda) \cdot ... \cdot (a_{nn} - \lambda)$. Nullsetzen ergibt: $\lambda_1 = a_{11}$,..., $\lambda_n = a_{nn}$. □

Sind die Eigenwerte bekannt, so sind die dazugehörenden Eigenvektoren nichttriviale Lösungen von

$$(\mathbf{A} - \lambda\mathbf{E})\,\mathbf{v} = 0.$$

Eigenvektoren sind damit nicht eindeutig bestimmt. Ist $\mathbf{v}$ ein Eigenvektor, ist auch $\alpha\mathbf{v}$ ein Eigenvektor. Man sucht entsprechend für jeden Eigenwert Basisvektoren, die den Vektorraum der Eigenvektoren aufspannen (Eigenraum).

Beispiel 5.3.14.

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{pmatrix}$$

Eigenwerte:

$$\begin{aligned} \det \begin{pmatrix} -\lambda & 0 & -2 \\ 1 & 2-\lambda & 1 \\ 1 & 0 & 3-\lambda \end{pmatrix} &= -\lambda(2-\lambda)(3-\lambda) + 0 + 0 - (-2(2-\lambda) + 0 + 0) \\ &= -\lambda(2-\lambda)(3-\lambda) + 2(2-\lambda) \\ &= (2-\lambda)(-\lambda(3-\lambda) + 2)) \\ &= (2-\lambda)\left(\lambda^2 - 3\lambda + 2\right) \\ &= (2-\lambda)(2-\lambda)(\lambda - 1) \\ &= (2-\lambda)^2(\lambda - 1) \end{aligned}$$

Damit ist $\lambda_1 = 2$ und $\lambda_2 = 1$. λ_1 hat die Vielfachheit 2. Der Eigenraum von λ_1 ist der Lösungsraum von

$$\begin{pmatrix} -2 & 0 & -2 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Dies entspricht dem Gleichungssystem

$$\begin{aligned} -2x_1 - 2x_3 &= 0 \\ x_1 + x_3 &= 0 \\ x_1 + x_3 &= 0 \end{aligned}$$

d.h. $x_1 = -x_3$. x_2 kann frei gewählt werden. Damit ist die Lösungmenge

$$\left\{ \begin{pmatrix} x_1 \\ x_2 \\ -x_1 \end{pmatrix} \middle| x_1, x_2 \in \mathbb{R} \right\}$$

und eine Basis des Eigenraums von λ_1 ist

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$$

Der Eigenraum zu $\lambda_2 = 1$ ist demgegenüber bestimmt durch

$$\begin{pmatrix} -1 & 0 & -2 \\ 1 & 1 & 1 \\ 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

was dem Gleichungssystem

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

- mit den Lösungen $x_1 = -2x_3, x_2 = x_3$ - entspricht, d.h. der Eigenraum von λ_2 ist

$$\left\{ \begin{pmatrix} -2x_3 \\ x_3 \\ x_3 \end{pmatrix} \middle| x_3 \in \mathbb{R} \right\}.$$

Eine Basis des Eigenraums von λ_2 ist damit gegeben durch

$$\begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}$$

$\Diamond$

Oben wurde erwähnt, dass man mit Hilfe einer Basis von Eigenvektoren die Wirkung einer linearen Abbildung besonders einfach beschreiben kann. Es stellt sich die Frage, unter welchen Bedingungen man zu einer Matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ eine Basis des $\mathbb{R}^n$ angeben kann, die aus Eigenvektoren von $\mathbf{A}$ besteht.

Definition 5.3.15. *(Diagonalisierbarkeit) Eine Matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ heisst diagonalisierbar genau dann, wenn es eine invertierbare Matrix $\mathbf{P}$ gibt, so dass $\mathbf{D} := \mathbf{P}^{-1}\mathbf{AP}$ eine Diagonalmatrix ist.* $\Diamond$

Bemerkung 5.3.16. *(1) Ist $\mathbf{A}$ diagonalisierbar, gibt es entsprechend eine Darstellung von $\mathbf{A}$ mittels $\mathbf{A} = \mathbf{PDP}^{-1}$ mit der Diagonalmatrix $\mathbf{D}$ und der invertierbaren Matrix $\mathbf{P}$.*
(2) Die Matrix $\mathbf{A}$ und $\mathbf{D}$ sind gemäss Definition der Ähnlichkeit ähnlich und beschreiben denselben Endomorphismus auf dem Hintergrund verschiedener Basen, die Matrix $\mathbf{D}$ auf dem Hintergrund der Eigenvektorbasis. $\Diamond$

Satz 5.3.17. *Eine Matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ ist genau dann diagonalisierbar, wenn sie n linear unabhängige Eigenvektoren besitzt.*

Beweis. Ist $\mathbf{P}^{-1}\mathbf{AP} = \mathbf{D}$ eine Diagonalmatrix, so folgt für $\mathbf{AP} = \mathbf{PD}$, dass die i-te Spalte von $\mathbf{P}$ ein Eigenvektor von $\mathbf{A}$ ist mit $\lambda_i = d_{ii}$. Umgekehrt kann man aus n unabhängigen Eigenvektoren eine invertierbare Matrix $\mathbf{P}$ bauen und die entsprechenden Eigenwerte auf die Diagonale von $\mathbf{D}$ setzen. Damit ist dann $\mathbf{AP} = \mathbf{PD}$ und man kann beide Seite von links her mit $\mathbf{P}^{-1}$ multiplizieren, um $\mathbf{P}^{-1}\mathbf{AP} = \mathbf{D}$ zu erhalten. $\square$

Bemerkung 5.3.18. *Wenn eine Matrix $\mathbf{P} \in \mathbb{R}^{n\times n}$ n linear unabhängige Eigenvektoren besitzt, gibt es eine Basis des $\mathbb{R}^n$, die aus Eigenvektoren besteht.* $\Diamond$

Beispiel 5.3.19. *Mit der Matrix von Beispiel 5.3.14*

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{pmatrix}$$

erhält man die aus linear unabhängigen Eigenvektoren bestehende Matrix

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \\ -1 & 0 & 1 \end{pmatrix}$$

Damit ist

$$\mathbf{AP} = \begin{pmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \\ -1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 & -2 \\ 0 & 2 & 1 \\ -2 & 0 & 1 \end{pmatrix}$$

Andererseits ist mit (s. Eigenwerte, 2 ist zweifacher Eigenwert)

$$\mathbf{D} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

ebenfalls

$$\mathbf{PD} = \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 & -2 \\ 0 & 2 & 1 \\ -2 & 0 & 1 \end{pmatrix}$$

Damit stellt

$$\mathbf{P}^{-1}\mathbf{AP} = \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \\ -1 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \\ -1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

eine Diagonalisierung von $\mathbf{A}$ *dar und* $\mathbf{A}$ *lässt sich darstellen mit*

$$\begin{pmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \\ -1 & 0 & 1 \end{pmatrix}^{-1}$$

◊

Satz 5.3.20. *Sind die Eigenwerte* $\lambda_1, ..., \lambda_n$ *einer Matrix* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *paarweise verschieden, so sind die Eigenvektoren* $\mathbf{v}_1, ..., \mathbf{v}_n$ *linear unabhängig.*

Beweis. Der Satz gilt im Falle eines einzigen Eigenwertes (wegen $\mathbf{v}_1 \neq 0$). Der Satz gelte für $m-1$ Eigenwerte und Eigenvektoren ($m > 1$). Man betrachte die Bedingung

$$a_1\mathbf{v}_1 + ... + a_m\mathbf{v}_m = 0. \tag{5.3.1}$$

Man kann rechnen

$$\begin{aligned} \mathbf{A}(a_1\mathbf{v}_1 + ... + a_m\mathbf{v}_m) &= a_1\lambda_1\mathbf{v}_1 + ... + a_m\lambda_m\mathbf{v}_m = 0 \\ \lambda_m(a_1\mathbf{v}_1 + ... + a_m\mathbf{v}_m) &= a_1\lambda_m\mathbf{v}_1 + ... + a_m\lambda_m\mathbf{v}_m = 0 \end{aligned}$$

Durch Subtraktion erhält man

$$\begin{aligned} &a_1\lambda_1\mathbf{v}_1 + ... + a_m\lambda_m\mathbf{v}_m - (a_1\lambda_m\mathbf{v}_1 + ... + a_m\lambda_m\mathbf{v}_m) \\ &= a_1(\lambda_1 - \lambda_m)\mathbf{v}_1 + ... + a_{m-1}(\lambda_{m-1} - \lambda_m)\mathbf{v}_{m-1} = 0 \end{aligned}$$

Wegen der linearen Unabhängigkeit der $\mathbf{v}_1, ..., \mathbf{v}_{m-1}$ gilt, $a_1(\lambda_1 - \lambda_m)+...+a_{m-1}(\lambda_{m-1} - \lambda_m) = 0$ genau dann, wenn $a_1(\lambda_1 - \lambda_m) = ... = a_{m-1}(\lambda_{m-1} - \lambda_m) = 0$. Da die Eigenwerte paarweise verschieden sind, ist dies nur möglich, wenn $a_1 = ... = a_{m-1} = 0$. Eingesetzt in (5.3.1) ergibt sich: $a_m\mathbf{v}_m = 0$ und damit wegen $\mathbf{v}_m \neq 0$ $a_m = 0$, d.h. $\mathbf{v}_m$ ist von $(\mathbf{v}_1, ..., \mathbf{v}_{m-1})$ linear unabhängig. □

Bemerkung 5.3.21. *Es wird nicht ausgeschlossen, dass einer der Eigenwerte* $\lambda_i = 0$ *ist. In diesem Fall gilt für den von* $\mathbf{0}$ *verschiedenen Eigenvektor* $\mathbf{v}_i$*:* $\mathbf{A}\mathbf{v}_i = \lambda_i\mathbf{v}_i = \mathbf{0}$*. Entsprechend ist der Kern von* $\mathbf{A}$ *von* $\mathbf{0}$ *verschieden und die durch* $\mathbf{A}$ *gegebene Abbildung ist nicht injektiv:* $\det\mathbf{A} = 0$ *und* $rang(\mathbf{A}) < n$*. Der Eigenvektor* $\mathbf{v}_i$ *ist trotzdem von den übrigen linear unabhängig. Die Matrix*

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$$

hat die Eigenwerte 5 und 0, sowie die orthogonalen und damit unabhängigen Eigenvektoren

$$\begin{pmatrix} 0.4472136 \\ 0.8944272 \end{pmatrix}, \begin{pmatrix} -0.8944272 \\ 0.4472136 \end{pmatrix}.$$

$\det\mathbf{A} = 1\cdot 4 - 2\cdot 2 = 0$. *rang*$(\mathbf{A}) = 1$. *rang(A)=n, wenn alle paarweise verschiedenen Eigenwerte von 0 verschieden sind.* ◊

Satz 5.3.22. *Ist* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *und weist* $\mathbf{A}$ *n paarweise verschiedene Eigenwerte auf, so ist* $\mathbf{A}$ *diagonalisierbar.*

Beweis. Sind die Eigenwerte $\lambda_1, ..., \lambda_n$ paarweise verschieden, so sind die Eigenvektoren $\mathbf{v}_1, ..., \mathbf{v}_n$ linear unabhängig. Damit ist gemäss obigem Satz $\mathbf{A}$ diagonalisierbar. □

Beispiel 5.3.23. *Am obigen Beispiel sieht man, dass Diagonalierbarkeit auch bei einem* $\lambda_i = 0$ *gegeben ist:*

$$\begin{pmatrix} 0.4472136 & -0.8944272 \\ 0.8944272 & 0.4472136 \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0.4472136 & -0.8944272 \\ 0.8944272 & 0.4472136 \end{pmatrix}^T = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$$

◊

Bevor der nächste Satz bewiesen wird, ein paar Regeln für komplexe Zahlen und das euklidische Skalarprodukt $\mathbf{z} \cdot \mathbf{w}$,$\mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$: Für Vektoren und Matrizen ist $\overline{\mathbf{z}}$ beziehungsweise $\overline{\mathbf{A}}$ komponentenweise definiert:

$$\begin{aligned}
v_i := a + bi \Longleftrightarrow \overline{v}_i := a - bi \quad (a, b \in \mathbb{R}; v_i \in \mathbb{C}) \\
|z| := \sqrt{(a+bi)(a-bi)} = \sqrt{a^2+b^2} \in \mathbb{R} \\
\mathbf{z} \cdot \mathbf{w} := \sum_{i=1}^{n} z_i \overline{w}_i = \sum_{i=1}^{n} \overline{z}_i w_i \in \mathbb{C}, \text{ d.h. } \mathbf{z} \cdot \mathbf{w} = \overline{\mathbf{z}}^T \mathbf{w} = \mathbf{z}^T \overline{\mathbf{w}} \\
(\mathbf{z} + \mathbf{z}') \cdot \mathbf{w} = \mathbf{z} \cdot \mathbf{w} + \mathbf{z}' \cdot \mathbf{w} \\
\mathbf{z} \cdot (\mathbf{w} + \mathbf{w}') = \mathbf{z} \cdot \mathbf{w} + \mathbf{z} \cdot \mathbf{w}' \\
\mathbf{w} \cdot \mathbf{z} = \overline{\mathbf{z} \cdot \mathbf{w}} = \overline{\mathbf{z}} \cdot \overline{\mathbf{w}} \\
\mathbf{v} \cdot (\lambda \mathbf{v}) = \bar{\lambda} (\mathbf{v} \cdot \mathbf{w}) \text{ für } \lambda \in \mathbb{C} \\
(\lambda \mathbf{v}) \cdot \mathbf{v} = \lambda (\mathbf{v} \cdot \mathbf{w}) \text{ für } \lambda \in \mathbb{C} \\
\mathbf{z} \cdot \mathbf{z} \in \mathbb{R} \ ; \ \mathbf{z} \cdot \mathbf{z} = 0 \Longleftrightarrow \mathbf{z} = 0. \\
\|\mathbf{z}\| := \sqrt{\mathbf{z} \cdot \mathbf{z}}
\end{aligned}$$

Für Matrizen $\mathbf{A}$ mit komplexen Einträgen bedeutet Orthogonalität (Isometrie!!), dass $\mathbf{x} \cdot \mathbf{y} = \mathbf{A}\mathbf{x} \cdot \mathbf{A}\mathbf{y} = (\mathbf{A}\mathbf{x})^t \overline{\mathbf{A}\mathbf{y}} = \mathbf{x}^t \mathbf{A}^t \overline{\mathbf{A}} \overline{\mathbf{y}}$, also $\mathbf{A}^t \overline{\mathbf{A}} = \mathbf{E} = \overline{\mathbf{A}}^t \mathbf{A}$ und damit $\overline{\mathbf{A}}^{-1} = \mathbf{A}^t$ und $\mathbf{A}^{-1} = \overline{\mathbf{A}^t}$. Eine Matrix heisst hermitsch, wenn $\overline{\mathbf{A}} = \mathbf{A}^t$ und damit $\mathbf{A} = \overline{\mathbf{A}}^t$.

Satz 5.3.24. *(Eigenwerte symmetrischer Matrizen) Sei* $\mathbf{A} \in \mathbb{C}^{n\times n}$ *eine symmetrische oder hermitsche Matrix d.h.* $\mathbf{A} = \mathbf{A}^t$ *oder* $\overline{\mathbf{A}} = \mathbf{A}^t$ *und damit* $\mathbf{A} = \overline{\mathbf{A}}^t$ *. Dann hat (a) A nur reelle Eigenwerte und (b) Eigenvektoren zu verschiedenen Eigenwerten sind orthogonal.*

Beweis. (a) Mit $\mathbf{v} \in \mathbb{C}^n$ und $\mathbf{A}$ hermitsch sowie $\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$ gilt: $(\lambda \mathbf{v}) \cdot \mathbf{v} = \left(\overline{\lambda}\overline{\mathbf{v}}\right)^t \cdot \mathbf{v} = \left(\overline{\mathbf{A}\mathbf{v}}\right)^t \mathbf{v} = \left(\overline{\mathbf{v}}^t \overline{\mathbf{A}}^t\right) \mathbf{v} = \overline{\mathbf{v}}^t \mathbf{A}\mathbf{v} = \overline{\mathbf{v}}^t \lambda \mathbf{v} = \overline{\lambda} \overline{\mathbf{v}}^t \mathbf{v}$. Da $\overline{\mathbf{v}}^t \mathbf{v}$ reell und ungleich 0 ist, folgt $\lambda = \overline{\lambda}$. Also ist λ reell.
(b) Seien $\mathbf{v}_1, \mathbf{v}_2$ Eigenvektoren von $\mathbf{A}$ zu verschiedenen Eigenwerten λ_1, λ_2. Damit gilt

$$\begin{aligned}
\lambda_1 \mathbf{v}_1^T \mathbf{v}_2 = (\mathbf{A}\mathbf{v}_1)^T \mathbf{v}_2 = \mathbf{v}_1^T \mathbf{A}^T \mathbf{v}_2 \\
= \mathbf{v}_1^T \mathbf{A}\mathbf{v}_2 = \mathbf{v}_1^T \lambda_2 \mathbf{v}_2
\end{aligned}$$

und damit

$$\begin{aligned}
\lambda_1 \mathbf{v}_1^T \mathbf{v}_2 - \mathbf{v}_1^T \lambda_2 \mathbf{v}_2 = 0 \\
(\lambda_1 - \lambda_2) \mathbf{v}_1^T \mathbf{v}_2 = 0
\end{aligned}$$

Da $\lambda_1 \neq \lambda_2$, ist $\mathbf{v}_1^T \mathbf{v}_2 = 0$. □

Beispiel 5.3.25. *Für*

$$\mathbf{A} = \begin{pmatrix} 4 & 12 \\ 12 & 11 \end{pmatrix}$$

erhält man

$$\det(\mathbf{A} - \lambda\mathbf{E}) = \det\left(\begin{pmatrix} 4 & 12 \\ 12 & 11 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}\right) = \det\begin{pmatrix} 4-\lambda & 12 \\ 12 & 11-\lambda \end{pmatrix} = (4-\lambda)(11-\lambda) - 144$$

Die Nullstellen von $(4-\lambda)(11-\lambda) - 144$ *sind* $20, -5$. *Ein Eigenvektor zu 20 ist eine Lösung von*

$$\begin{pmatrix} 4-20 & 12 \\ 12 & 11-20 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

mit dem Gleichungssystem

$$\begin{aligned} -16x_1 + 12x_2 &= 0 \\ 12x_1 - 9x_2 &= 0 \end{aligned}$$

und den Lösungen. $x_1 = \frac{3}{4}x_2$. *Ein Eigenvektor zu 20 ist damit*

$$\begin{pmatrix} \frac{3}{4} \\ 1 \end{pmatrix}$$

Für -5 erhält man mit

$$\begin{pmatrix} 4+5 & 12 \\ 12 & 11+5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

die Lösung $x_1 = -\frac{4}{3}x_2$. *Ein Eigenvektor zu -5 ist damit*

$$\begin{pmatrix} -\frac{4}{3} \\ 1 \end{pmatrix}$$

Es gilt

$$\begin{pmatrix} -\frac{4}{3} \\ 1 \end{pmatrix}^T \begin{pmatrix} \frac{3}{4} \\ 1 \end{pmatrix} = 0$$

◊

Bemerkung 5.3.26. *Hat ein Eigenwert* λ *von* $\mathbf{A}$ *die Vielfachheit* k *so hat* λ *auch einen Eigenraum der Dimension* j *mit* $1 \leq j \leq k$. *Ist die Dimension aller Eigenräume von* $\mathbf{A}$ *mit der Vielfachheit der entsprechenden Eigenwerte identisch, besitzt* $\mathbb{R}^n$ *eine Basis aus Eigenvektoren von* $\mathbf{A}$. *Für symmetrische Matrizen vom Rang n stellen diese eine Orthogonalbasis dar, wenn man die Eigenvektoren von Eigenräumen höherer Dimension als 1 orthogonalisiert (Gram-Schmidt). Da sich aus einer Orthogonalbasis eine Orthonormabasis herstellen lässt, ist eine symmetrische Matrix nicht nur stets diagonalisierbar, sondern die Diagonalisierung ist immer durch eine orthogonale Matrix möglich.* ◊

Definition 5.3.27. $\mathbf{x}^\top\mathbf{A}\mathbf{x}$ *mit* $x \in \mathbb{R}^n$ *und* $A \in \mathbb{R}^{n\times n}$ *heisst „die zu* $\mathbf{A}$ *gehörende quadratische Form in* $\mathbf{x}$*“* ◊

Der nächste Satz erweist sich als nützlich bei der Bestimmung von Maxima und Minima von Abbildungen $\mathbb{R}^n \to \mathbb{R}$. Die Matrix der zweiten Ableitungen kann damit auf positive oder negative Definitheit überprüft werden.

Definition 5.3.28. $\mathbf{A}$ *ist*

- *positiv definit genau dann, wenn* $\mathbf{x}^\top\mathbf{A}\mathbf{x} > 0$ *für* $\mathbf{x} \in V, \mathbf{x} \neq \mathbf{0}$.

- *positiv semi-definit genau dann, wenn* $\mathbf{x}^\top \mathbf{A}\mathbf{x} \geq 0$ *für* $\mathbf{x} \in V$ *und* $\mathbf{x}^\top \mathbf{A}\mathbf{x} = \mathbf{0}$ *für ein* $\mathbf{x} \neq \mathbf{0}$
- *negativ definit genau dann, wenn* $\mathbf{x}^\top \mathbf{A}\mathbf{x} < 0$ *für* $\mathbf{x} \in V, \mathbf{x} \neq \mathbf{0}$.
- *negativ semi-definit genau dann, wenn* $\mathbf{x}^\top \mathbf{A}\mathbf{x} \geq 0$ *für* $\mathbf{x} \in V$ *und* $\mathbf{x}^\top \mathbf{A}\mathbf{x} = \mathbf{0}$ *für ein* $\mathbf{x} \neq \mathbf{0}$
- *indefinit, genau dann wenn es gibt* $\mathbf{x}$ *und* $\mathbf{y}$*, so dass* $\mathbf{x}^\top \mathbf{A}\mathbf{x} < \mathbf{0}$ *und* $\mathbf{y}^\top \mathbf{A}\mathbf{y} > \mathbf{0}$

Satz 5.3.29. *Sei* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *eine symmetrische Matrix mit den Eigenwerten* $(\lambda_1, ..., \lambda_n)$*. Dann ist* $\mathbf{A}$

1. *positiv definit genau dann, wenn* $\lambda_i > 0$ *für alle* $i \in \mathbb{N}_n^*$
2. *positiv semidefinit genau dann, wenn* $\lambda_i \geq 0$ *und* $\lambda_i = 0$ *für mindestens ein* $i \in \mathbb{N}_n^*$.
3. *negativ definit genau dann, wenn* $\lambda_i < 0$ *für alle* $i \in \mathbb{N}_n^*$
4. *negativ semidefinit genau dann, wenn* $\lambda_i \leq 0$ *und* $\lambda_i = 0$ *für mindestens ein* $i \in \mathbb{N}_n^*$
5. *indefinit genau dann, wenn es ein* $\lambda_i < 0$ *und* $\lambda_j > 0$ *gibt,* $i, j \in \mathbb{N}_n^*$.
6. *Eine positiv semidefinite Matrix* $\mathbf{A}$ *ist genau dann positiv definit, wenn* $\mathbf{A}$ *regulär ist (d.h.* $\mathtt{rg}(\mathbf{A}) = n$*).*
7. *Mit* $\mathbf{A}$ *ist auch* $\mathbf{A}^{-1}$ *positiv definit*

Beweis. 1. Stellt man Vektoren $\mathbf{x}$ von $\mathbb{R}^n$ mit Hilfe einer Orthonormalbasis $(\mathbf{w}_1, ..., \mathbf{w}_n)$ aus Eigenvektoren dar, so ergibt sich $\mathbf{x} = \sum_{i=1}^n \mu_i \mathbf{w}_i$ und es gilt für $\mathbf{x}^\top \mathbf{A}\mathbf{x} = \mathbf{x}^\top \sum_{i=1}^n a_{ji}\mu_i \mathbf{w}_i = \sum_{i=1}^n \lambda_i \mu_i^2 > 0$, für alle $\mathbf{x}$, nur wenn $\lambda_i > 0$ für alle Eigenwerte von $\mathbf{A}$. Die Beweise für die übrigen Aussagen bezüglich Definitheit verlaufen analog.
6. Eine positiv semidefinite Matrix hat nur Eigenwerte, die grösser gleich 0 sind. Diese sind aber von 0 verschieden, da $\mathtt{rg}(\mathbf{A}) = n$ (wie oben gezeigt gilt ja: $\mathbf{A}$ hat Rang n genau dann wenn $\lambda_i \neq 0$ für $i \in \mathbb{N}_n^*$). Also ist die Matrix positiv definit.
7. Die Eigenwerte der transponieren Matrix sind mit den Eigenwerten der ursprünglichen Matrix identisch. Damit ist die Transponierte einer positiv definiten Matrix auch positiv definit. □

◊

5.4 Hauptachsentransformation

Satz 5.4.1. *(Hauptachsentransformation) Sei* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *symmetrisch, dann ist* $\mathbf{A} = \mathbf{Q}\Lambda\mathbf{Q}^T$ *mit* $\Lambda = diag\,(\lambda_1, ...\lambda_n)$ *und der Matrix* $\mathbf{Q}$ *der normalisierten Eigenvektoren der* λ_i *als Spaltenvektoren eine Diagonalisierung von* $\mathbf{A}$ *- mit orthogonalen Spaltenvektoren von* $\mathbf{Q}$ *und reellen Einträgen in* Λ.

Beweis. Da $\mathbf{A}$ symmetrisch ist, gibt es eine Basis von $\mathbb{R}^n$ aus orthonormalen Eigenvektoren von $\mathbf{A}$. Daraus ist eine Diagonalisierung von $\mathbf{A}$ konstruierbar:

$$\mathbf{D} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$$

wobei die invertierbare Matrix $\mathbf{P}$ aus den Eigenvektoren von $\mathbf{A}$ gebildet wird und $\mathbf{D}$ die Diagonalmatrix der zugehörigen Eigenwerte ist, also $\mathbf{D} = \Lambda$. Wegen der Orthonormalität der Eigenvektoren ist $\mathbf{P} = \mathbf{Q}$ eine orthogonale Matrix und es folgt

$$\mathbf{A} = \mathbf{Q}\Lambda\mathbf{Q}^{-1}$$

und wegen $\mathbf{Q}^{-1} = \mathbf{Q}^T$

$$\mathbf{A} = \mathbf{Q}\Lambda\mathbf{Q}^T.$$

□

Bemerkung 5.4.2. *Man nennt die eindimensionalen Vektorräume der Eigenvektoren, die* $\mathbf{Q}$ *bilden, die Hauptachsen. Die Diagonalisierung durch eine orthogonale Matrix bedeutet, dass die durch die Matrix* $\mathbf{A}$ *beschriebene Abbildung von Vektoren in einem durch die Eigenvektoren gegebenen Koordinatensystem erfolgt. In*

$$\mathbf{A}\mathbf{u} = \mathbf{Q}\Lambda\mathbf{Q}^T\mathbf{u}$$

werden dem Vektor $\mathbf{u}$ *(in Standardbasis) durch* $\mathbf{Q}^{-1}\mathbf{u} = \mathbf{Q}^T\mathbf{u}$ *die Koordinaten von* $\mathbf{u}$ *bezüglich der Eigenvektorbasis zugeordnet.* $\Lambda\mathbf{Q}^T\mathbf{u}$ *liefert dann die Koordinaten von* $\mathbf{A}\mathbf{u}$ *bezüglich der Eigenvektorbasis.* $\mathbf{Q}\Lambda\mathbf{Q}^T\mathbf{u}$ *stellt die Rücktransformation in die Standardbasis dar. In der Eigenvektor-Basis ist die Abbildung, die* $\mathbf{A}$ *entspricht, besonders einfach, da sie durch die Diagonalmatrix* Λ *gegeben ist. Geometrisch bedeutet diese Abbildung, dass in jede Hauptachsenrichtung eine Streckung der Koordinaten von* $\mathbf{u}$ *im Eigenvektorkoordinatensystem um die Eigenwerte erfolgt.* $\Diamond$

Beispiel 5.4.3. *Sei* $\mathbf{A} = \begin{pmatrix} 4 & 4 \\ 4 & -2 \end{pmatrix}$. *Die Matrix weist die Eigenwerte* $6, -4$ *auf. Ein Eigenvektor von* 6 *ist* $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ *und von* -4 $\begin{pmatrix} -\frac{1}{2} \\ 1 \end{pmatrix}$. *Normierung ergibt*

$$\mathbf{q}_1 = \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix} = \begin{pmatrix} \frac{2}{5}\sqrt{5} \\ \frac{1}{5}\sqrt{5} \end{pmatrix} = \begin{pmatrix} 0.894\,43 \\ 0.447\,21 \end{pmatrix}$$

sowie

$$\mathbf{q}_2 = \begin{pmatrix} \frac{-\frac{1}{2}}{\frac{1}{2}\sqrt{5}} \\ \frac{1}{\frac{1}{2}\sqrt{5}} \end{pmatrix} = \begin{pmatrix} -\frac{1}{5}\sqrt{5} \\ \frac{2}{5}\sqrt{5} \end{pmatrix} = \begin{pmatrix} -0.447\,21 \\ 0.894\,43 \end{pmatrix}.$$

Man erhält die orthonormale Matrix

$$\mathbf{Q} = \begin{pmatrix} \frac{2}{5}\sqrt{5} & -\frac{1}{5}\sqrt{5} \\ \frac{1}{5}\sqrt{5} & \frac{2}{5}\sqrt{5} \end{pmatrix}$$

und es gilt

$$\mathbf{Q}^T\mathbf{A}\mathbf{Q} = \begin{pmatrix} \frac{2}{5}\sqrt{5} & -\frac{1}{5}\sqrt{5} \\ \frac{1}{5}\sqrt{5} & \frac{2}{5}\sqrt{5} \end{pmatrix}^T \begin{pmatrix} 4 & 4 \\ 4 & -2 \end{pmatrix} \begin{pmatrix} \frac{2}{5}\sqrt{5} & -\frac{1}{5}\sqrt{5} \\ \frac{1}{5}\sqrt{5} & \frac{2}{5}\sqrt{5} \end{pmatrix} = \begin{pmatrix} 6 & 0 \\ 0 & -4 \end{pmatrix} = diag\,(\lambda_1, \lambda_2)$$

Umgekehrt gilt

$$\mathbf{A} = \mathbf{Q}\Lambda\mathbf{Q}^T = \begin{pmatrix} \frac{2}{5}\sqrt{5} & -\frac{1}{5}\sqrt{5} \\ \frac{1}{5}\sqrt{5} & \frac{2}{5}\sqrt{5} \end{pmatrix} \begin{pmatrix} 6 & 0 \\ 0 & -4 \end{pmatrix} \begin{pmatrix} \frac{2}{5}\sqrt{5} & -\frac{1}{5}\sqrt{5} \\ \frac{1}{5}\sqrt{5} & \frac{2}{5}\sqrt{5} \end{pmatrix}^T = \begin{pmatrix} 4 & 4 \\ 4 & -2 \end{pmatrix}$$

Sei $\mathbf{u} = (0.5, 0.8)$ *und damit*

$$\mathbf{A}\mathbf{u} = \begin{pmatrix} 4 & 4 \\ 4 & -2 \end{pmatrix} \begin{pmatrix} 0.5 \\ 0.8 \end{pmatrix} = \begin{pmatrix} 5.2 \\ 0.4 \end{pmatrix}$$

Mit

$$\mathbf{Q}^T\mathbf{u} = \begin{pmatrix} \frac{2}{5}\sqrt{5} & -\frac{1}{5}\sqrt{5} \\ \frac{1}{5}\sqrt{5} & \frac{2}{5}\sqrt{5} \end{pmatrix}^T \begin{pmatrix} 0.5 \\ 0.8 \end{pmatrix} = \begin{pmatrix} 0.36\sqrt{5} \\ 0.22\sqrt{5} \end{pmatrix} = \begin{pmatrix} 0.804\,98 \\ 0.491\,93 \end{pmatrix}$$

erhält man die Koordinaten von $\mathbf{u}$ *bezüglich der durch die Eigenvektoren gegebenen Achsen, denn*

$$0.804\,98 \begin{pmatrix} \frac{2}{5}\sqrt{5} \\ \frac{1}{5}\sqrt{5} \end{pmatrix} + 0.491\,93 \begin{pmatrix} -\frac{1}{5}\sqrt{5} \\ \frac{2}{5}\sqrt{5} \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.8 \end{pmatrix}$$

(formal offensichtlich: $\mathbf{Q}\mathbf{Q}^T\mathbf{u} = \mathbf{u}$*). Mittels* $\Lambda\mathbf{Q}^T\mathbf{u}$ *erhält man die Koordinaten von* $\mathbf{Au}$ *bezüglich der Eigenvektorachsen:*

$$\begin{pmatrix} 6 & 0 \\ 0 & -4 \end{pmatrix} \begin{pmatrix} \frac{2}{5}\sqrt{5} & -\frac{1}{5}\sqrt{5} \\ \frac{1}{5}\sqrt{5} & \frac{2}{5}\sqrt{5} \end{pmatrix}^T \begin{pmatrix} 0.5 \\ 0.8 \end{pmatrix} = \begin{pmatrix} 2.16\sqrt{5} \\ -0.88\sqrt{5} \end{pmatrix} = \begin{pmatrix} 4.829\,9 \\ -1.967\,7 \end{pmatrix}$$

Durch

$$\mathbf{Q}\Lambda\mathbf{Q}^T\mathbf{u} = \begin{pmatrix} \frac{2}{5}\sqrt{5} & -\frac{1}{5}\sqrt{5} \\ \frac{1}{5}\sqrt{5} & \frac{2}{5}\sqrt{5} \end{pmatrix} \begin{pmatrix} 4.829\,9 \\ -1.967\,7 \end{pmatrix} = \begin{pmatrix} 5.2 \\ 0.4 \end{pmatrix}.$$

wechselt man die Basis und drückt $\mathbf{Au}$ *durch die Standardbasis aus.*

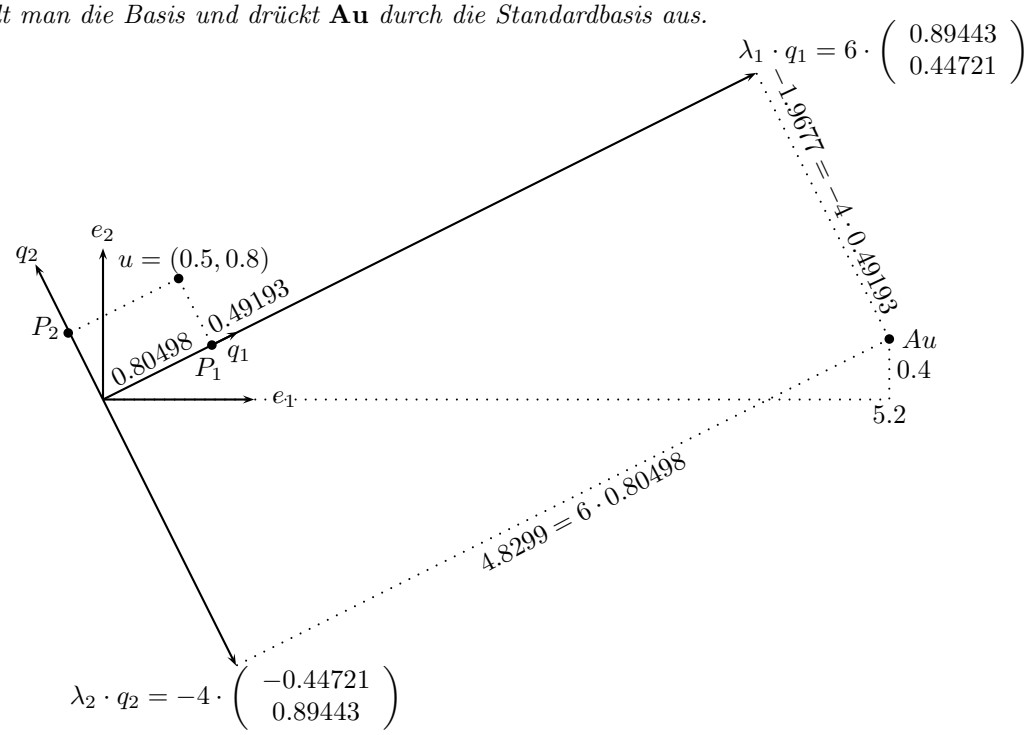


Abbildung 5.4.6: Abbildung des Vektors u mittels $A = Q\Lambda Q^T$

$\Diamond$

Bemerkung 5.4.4. *Man kann die Koordinaten von Punkten* P_1 *und* P_2 *bezüglich der Standardbasis berechnen, die den jeweiligen orthogonalen Projektionen von* $\mathbf{u}$ *auf die Geraden* $g_1 := \{\alpha\mathbf{q}_1 | \alpha \in \mathbb{R}\}$ *und* $g_2 := \{\beta\mathbf{q}_2 | \beta \in \mathbb{R}\}$ *sind. Die Komponenten von* $\mathbf{Q}^T\mathbf{u}$ *liefern die Streckungsfaktoren für die Eigenvektoren, um die Punkte* P_1 *und* P_2 *zu erhalten. Mit*

$$\mathbf{Q}^T\mathbf{u} = \begin{pmatrix} 0.804\,98 \\ 0.491\,93 \end{pmatrix}$$

ist

$$P_1 = 0.804\,98\mathbf{q}_1 = 0.804\,98 \begin{pmatrix} \frac{2}{5}\sqrt{5} \\ \frac{1}{5}\sqrt{5} \end{pmatrix} = \begin{pmatrix} 0.72 \\ 0.36 \end{pmatrix}$$

und

$$P_2 = 0.491\,93\mathbf{q}_2 = 0.491\,93 \begin{pmatrix} -\frac{1}{5}\sqrt{5} \\ \frac{2}{5}\sqrt{5} \end{pmatrix} = \begin{pmatrix} -0.22 \\ 0.44 \end{pmatrix}$$

Formal: $\mathbf{Q} \cdot diag\left(\mathbf{Q}^T\mathbf{u}\right)$ *liefert die Matrix mit den Spalten für* P_1 *und* P_2*.*

$$\mathbf{Q} \cdot diag\left(\mathbf{Q}^T\mathbf{u}\right) = \begin{pmatrix} \frac{2}{5}\sqrt{5} & -\frac{1}{5}\sqrt{5} \\ \frac{1}{5}\sqrt{5} & \frac{2}{5}\sqrt{5} \end{pmatrix} \begin{pmatrix} 0.36\sqrt{5} & 0 \\ 0 & 0.22\sqrt{5} \end{pmatrix} = \begin{pmatrix} 0.72 & -0.22 \\ 0.36 & 0.44 \end{pmatrix}$$

Die Punkte P_1 *und* P_2 *erhält man auch mit* $\frac{\mathbf{u}\cdot\mathbf{q}_i}{\mathbf{q}_i\cdot\mathbf{q}_i}\mathbf{q}_i$*. Multipliziert man die Punkte* P_1 *und* P_2 *mit den Eigenwerten, erhält man Punkte auf den Eigenvektorachsen, auf denen* $\mathbf{Au}$ *senkrecht steht. Die Punkte* $6P_1$ *und* $-4P_2$ *werden dabei in Koordinaten bezüglich der Standardbasis geliefert:*

$$6\begin{pmatrix} 0.72 \\ 0.36 \end{pmatrix} = \begin{pmatrix} 4.32 \\ 2.16 \end{pmatrix}$$

$$-4\begin{pmatrix} -0.22 \\ 0.44 \end{pmatrix} = \begin{pmatrix} 0.88 \\ -1.76 \end{pmatrix}.$$

Formal erhält man $6P_1$ *und* $-4P_2$ *durch die Spalten von :* $\mathbf{Q} \cdot diag\left(\mathbf{Q}^T\mathbf{u}\right) \cdot \Lambda$*.*

$$\mathbf{Q} \cdot diag\left(\mathbf{Q}^T\mathbf{u}\right) \cdot \Lambda = \begin{pmatrix} 0.72 & -0.22 \\ 0.36 & 0.44 \end{pmatrix} \begin{pmatrix} 6 & 0 \\ 0 & -4 \end{pmatrix} = \begin{pmatrix} 4.32 & 0.88 \\ 2.16 & -1.76 \end{pmatrix}$$

Die Koordinaten von $\mathbf{Au}$ *bezüglich der Eigenvektorachsen erhält man dann durch* $\left\|\begin{pmatrix} 4.32 \\ 2.16 \end{pmatrix}\right\| =$ $4.829\,9$ *und* $-\left\|\begin{pmatrix} 0.88 \\ -1.76 \end{pmatrix}\right\| = -1.967\,7.$ ◊

5.5 Singulärwertzerlegung

Symmetrische Matrizen lassen sich vollständig mit Hilfe ihrer Eigenwerte und Eigenvektoren beschreiben. Diese Darstellung kann zur geometrischen Beschreibung ihrer Wirkung auf Vektoren benutzt werden. Diese Beschreibung möchte man auf beliebige Matrizen erweitern. Dies leistet die Singulärwertzerlegung. Singulärwerte könne ähnlich gut interpretiert werden wie Eigenwerte symmetrischer Matrizen. Die Singulärwertzerlegung ist nicht auf quadratische Matrizen beschränkt. In der Singulärwertzerlegung einer reellen Matrix treten nur reelle Matrizen auf.

Satz 5.5.1. *Sei* $\mathbf{A} \in \mathbb{R}^{m\times n}$*. Dann gibt es orthogonale Matrizen* $\mathbf{U} \in O(m)$ *und* $\mathbf{V} \in O(n)$ *sowie eine Matrix* $\Sigma = (s_{ij}) \in \mathbb{R}^{m\times n}$*, mit* $s_{ij} = 0$ *für* $i \neq j$ *und nichtnegativen Diagonaleinträgen* $s_{11} \geq s_{22} \geq ...$*, so dass*

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$$

Diese Darstellung von $\mathbf{A}$ *heisst Singärwertzerlegung von* $\mathbf{A}$*. Die Werte* s_{ii} *heissen Singulärwerte von* $\mathbf{A}$*.*

Beweis. Der Beweis erfolgt mittels eine Konstruktionsanleitung mit darauffolgendem Nachweis der Eigenschaften der konstruierten Matrizen. In einem ersten Schritt setzte man

$$\mathbf{B} := \mathbf{A}^T\mathbf{A}$$

Man erhält eine symmetrische $n \times n-$Matrix $\mathbf{B}$ vom Rang r der Matrix $\mathbf{A}$. Man berechnet die Eigenwerte λ_i und enstprechende orthonormale Eigenvektoren $\mathbf{v}_i$ von $\mathbf{B}$. Die Eigenvektoren werden der Grösse der entsprechenden Eigenwerte nach geordnet. Da

$$\mathbf{v}_i \cdot \mathbf{B}\mathbf{v}_i = \lambda_i\mathbf{v}_i^T\mathbf{v}_i$$

und

$$\mathbf{v}_i \cdot \mathbf{B}\mathbf{v}_i = \mathbf{v}_i^T\mathbf{A}^T\mathbf{A}\mathbf{v}_i = (\mathbf{A}\mathbf{v}_i) \cdot (\mathbf{A}\mathbf{v}_i) = \|\mathbf{A}\mathbf{v}_i\|^2 \geq 0$$

sind die $\lambda_i \geq 0$. Da der Rang von $\mathbf{B}$ gleich dem Rang von $\mathbf{A}$ ist, sind genau die ersten r Eigenwerte $\lambda_1, ..., \lambda_r$ positiv.
In einem zweiten Schritt definiert man für $i \in \mathbb{N}_r^*$

$$\mathbf{u}_i := \frac{1}{\sqrt{\lambda_i}} \mathbf{A}\mathbf{v}_i$$

und man bestimmt $m - r$ orthonormale Vektoren $\mathbf{u}_{r+1}, ..., \mathbf{u}_m$, die zu $\mathbf{u}_1, ..., \mathbf{u}_r$ orthogonal sind. Dann bildet man die Matrizen

$$\mathbf{U} := (\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_m)$$
$$\mathbf{V} := (\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n)$$

aus den Vektoren $\mathbf{u}_i$ und $\mathbf{v}_i$ als Spaltenvektoren. Die Matrix $\Sigma = (s_{ij}) \in \mathbb{R}^{m \times n}$ wird gebildet nach der Vorschrift

$$s_{ij} := \begin{cases} \sqrt{\lambda_i} & \text{für } i = j \leq r \\ 0 & \text{sonst} \end{cases}$$

Nun zeigt man, dass $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$ eine Singulärwertzerlegung von $\mathbf{A}$ ist. $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ ist nach Konstruktion eine Orthonormalbasis von $\mathbb{R}^n$. $\{\mathbf{u}_1, ..., \mathbf{u}_m\}$ ist nach Konstruktion eine Orthonormalbasis von $\mathbb{R}^m$, denn für $i, j = 1, ..., r$ ist (da $\mathbf{A}^T\mathbf{A}\mathbf{v}_j = \mathbf{B}\mathbf{v}_j = \lambda_i\mathbf{v}_j$)

$$\mathbf{u}_i^T\mathbf{u}_j = \frac{1}{\sqrt{\lambda_i\lambda_j}}\mathbf{v}_i^T\mathbf{A}^T\mathbf{A}\mathbf{v}_j = \frac{1}{\sqrt{\lambda_i\lambda_j}}\mathbf{v}_i^T\lambda_i\mathbf{v}_j = \frac{\lambda_i}{\sqrt{\lambda_i\lambda_j}}\mathbf{v}_i^T\mathbf{v}_j = \begin{cases} \frac{\lambda_i}{\lambda_i} = 1 & \text{für } i = j \leq r \\ 0 & \text{sonst} \end{cases}$$

womit die $\mathbf{u}_1, ..., \mathbf{u}_r$ orthonormal sind. $\mathbf{u}_{r+1}, ..., \mathbf{u}_n$ sind orthnormal auf Grund der Konstruktion. Damit gilt gemäss Definition von $\mathbf{u}_i$ (die Singulärwerte für $i > r$ sind 0, d.h. $0\mathbf{u}_i\mathbf{v}_i^T = \mathbf{0}$ und zur Matrix $\sum_{i=1}^{r} \sqrt{\lambda_i}\mathbf{u}_i\mathbf{v}_i^T$ wird jeweils von $r+1$ bis n die Null-Matrix hinzugezählt).

$$\begin{aligned} \mathbf{U}\Sigma\mathbf{V}^T &= \sum_{i=1}^{r} \sqrt{\lambda_i}\mathbf{u}_i\mathbf{v}_i^T + \sum_{i=r+1}^{n} 0\mathbf{u}_i\mathbf{v}_i^T = \sum_{i=1}^{r} \mathbf{A}\mathbf{v}_i\mathbf{v}_i^T \\ &= \sum_{i=1}^{n} \mathbf{A}\mathbf{v}_i\mathbf{v}_i^T \\ &= \mathbf{A}\sum_{i=1}^{n} \mathbf{v}_i\mathbf{v}_i^T \\ &= \mathbf{A}\mathbf{V}\mathbf{V}^T \\ &= \mathbf{A}\mathbf{E} = \mathbf{A}. \end{aligned}$$

□

Beispiel 5.5.2.

$$\mathbf{A}_1 = \begin{pmatrix} 4 & 12 \\ 12 & 11 \end{pmatrix} = \begin{pmatrix} \frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & -\frac{3}{5} \end{pmatrix} \begin{pmatrix} 20 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} \frac{3}{5} & \frac{4}{5} \\ -\frac{4}{5} & \frac{3}{5} \end{pmatrix}$$

Mit R:

```
m=matrix(c(4,12,12,11),ncol=2,byrow=T)
sdv(m)
```

ergibt

$'d'	
[1]	*20 5*
$u	
	[,1] [,2]
[1,]	*-0.6 -0.8*
[2,]	*-0.8 0.6*
$v	
	[,1] [,2]
[1,]	*-0.6 0.8*
[2,]	*-0.8 -0.6*

Die gelieferte Zerlegung ist bis auf die Vorzeichen identisch. Es gilt

$$\begin{pmatrix} -0.6 & -0.8 \\ -0.8 & 0.6 \end{pmatrix} \begin{pmatrix} 20 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} -0.6 & 0.8 \\ -0.8 & -0.6 \end{pmatrix}^T = \begin{pmatrix} 4 & 12 \\ 12 & 11 \end{pmatrix}$$

◊

Beispiel 5.5.3.

$$\mathbf{A}_2 = \begin{pmatrix} 0.36 & 1.6 & 0.48 \\ 0.48 & -1.2 & 0.64 \end{pmatrix} = \begin{pmatrix} 0.8 & 0.6 \\ -0.6 & 0.8 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0.6 & 0 & 0.8 \\ -0.8 & 0 & 0.6 \end{pmatrix}$$

◊

Beispiel 5.5.4. *Man bestimme die Singulärwertzerlegung von*

$$\mathbf{A} = \frac{1}{9} \begin{pmatrix} -8 & 10 & 14 \\ 2 & 2 & 1 \\ -6 & -6 & -3 \\ -16 & 2 & 10 \end{pmatrix}$$

Es ist

$$\mathbf{B} = \frac{1}{9} \begin{pmatrix} -8 & 10 & 14 \\ 2 & 2 & 1 \\ -6 & -6 & -3 \\ -16 & 2 & 10 \end{pmatrix}^T \frac{1}{9} \begin{pmatrix} -8 & 10 & 14 \\ 2 & 2 & 1 \\ -6 & -6 & -3 \\ -16 & 2 & 10 \end{pmatrix} = \frac{1}{9} \begin{pmatrix} 40 & -8 & -28 \\ -8 & 16 & 20 \\ -28 & 20 & 34 \end{pmatrix}$$

Die charakterische Gleichung

$$\begin{aligned} 0 = -\det(\mathbf{B} - \lambda\mathbf{E}) &= \left(\frac{40}{9} - \lambda\right)\left(\frac{16}{9} - \lambda\right)\left(\frac{34}{9} - \lambda\right) - 2\frac{8}{9}\frac{20}{9}\frac{28}{9} + \\ &\frac{8}{9}\frac{8}{9}\left(\frac{34}{9} - \lambda\right) + \frac{28}{9}\frac{28}{9}\left(\frac{16}{9} - \lambda\right) + \frac{20}{9}\frac{20}{9}\left(\frac{40}{9} - \lambda\right) \\ &= \lambda^3 - 10\lambda^2 + 16\lambda \end{aligned}$$

weist die drei Lösungen

$$\begin{aligned} \lambda_1 &= 5 + \sqrt{25 - 16} = 8 \\ \lambda_2 &= 5 - \sqrt{28 - 16} = 2 \\ \lambda_3 &= 0 \end{aligned}$$

auf. Einen Eigenvektor zu λ_1 erhält man aus

$$(\mathbf{B} - 8\mathbf{E}) \begin{pmatrix} x \\ y \\ 7 \end{pmatrix} = 0$$

d.h.

$$\begin{aligned} -32x - 8y - 28z &= 0 \\ -8x - 56y + 20z &= 0 \\ -28x + 20y - 38z &= 0 \end{aligned}$$

Eine Lösung ist $x = -z, y = \frac{1}{2}z$ (oder äquivalent $x = -2s, y = s, z = 2s$). Ein normierte Eigenvektor ist daher

$$\mathbf{v}_1 = \begin{pmatrix} -\frac{2}{3} \\ \frac{1}{3} \\ \frac{2}{3} \end{pmatrix}$$

Zu $\lambda_2 = 2$ erhält man

$$\begin{aligned} 22x - 8y - 28z &= 0 \\ -8x - 2y + 20z &= 0 \\ -28x + 20y + 16z &= 0 \end{aligned}$$

mit einer Lösung $x = 2z, y = 2z$. Ein normierte Eigenvektor ist daher

$$\mathbf{v}_2 = \begin{pmatrix} \frac{2}{3} \\ \frac{2}{3} \\ \frac{1}{3} \end{pmatrix}.$$

Für $\lambda_3 = 0$ erhält man für

$$\mathbf{B} \begin{pmatrix} x \\ y \\ 7 \end{pmatrix} = 0$$

das Gleichungssystem

$$\begin{aligned} 40x - 8y - 28z &= 0 \\ -8x + 16y + 20z &= 0 \\ -28x + 20y + 34z &= 0 \end{aligned}$$

mit einer Lösung $x = \frac{1}{2}z, y = -z$ (oder äquivalent $x = 1, y = -2, z = 2$) Ein normierte Eigenvektor ist

$$\mathbf{v}_3 = \begin{pmatrix} \frac{1}{3} \\ -\frac{2}{3} \\ \frac{2}{3} \end{pmatrix}$$

Man berechnet

$$\mathbf{u}_1 = \frac{1}{\sqrt{\lambda_1}}\mathbf{A}\mathbf{v}_1 = \frac{1}{2\sqrt{2}}\frac{1}{9}\begin{pmatrix} -8 & 10 & 14 \\ 2 & 2 & 1 \\ -6 & -6 & -3 \\ -16 & 2 & 10 \end{pmatrix}\begin{pmatrix} -\frac{2}{3} \\ \frac{1}{3} \\ \frac{2}{3} \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\sqrt{2} \\ 0 \\ 0 \\ \frac{1}{2}\sqrt{2} \end{pmatrix} = \frac{\sqrt{2}}{2}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

$$\mathbf{u}_2 = \frac{1}{\sqrt{\lambda_2}}\mathbf{A}\mathbf{v}_2 = \frac{1}{\sqrt{2}}\frac{1}{9}\begin{pmatrix} -8 & 10 & 14 \\ 2 & 2 & 1 \\ -6 & -6 & -3 \\ -16 & 2 & 10 \end{pmatrix}\begin{pmatrix} \frac{2}{3} \\ \frac{2}{3} \\ \frac{1}{3} \end{pmatrix} = \begin{pmatrix} \frac{1}{3}\sqrt{2} \\ \frac{1}{6}\sqrt{2} \\ -\frac{1}{2}\sqrt{2} \\ -\frac{1}{3}\sqrt{2} \end{pmatrix} = \frac{1}{3\sqrt{2}}\begin{pmatrix} 2 \\ 1 \\ -3 \\ -2 \end{pmatrix}$$

Als nächstes bestimmt man zwei orthonormale Vektoren $\mathbf{u}_3$, $\mathbf{u}_4$, *die* $\mathbf{u}_1$ *und* $\mathbf{u}_2$ *zu einer orthonormalen Basis des* $\mathbb{R}^4$ *ergänzen, z.B.*

$$\mathbf{u}_3 = \frac{1}{3\sqrt{2}} \begin{pmatrix} -2 \\ -1 \\ -3 \\ 2 \end{pmatrix}$$

$$\mathbf{u}_4 = \frac{1}{3\sqrt{2}} \begin{pmatrix} 1 \\ -4 \\ 0 \\ -1 \end{pmatrix}$$

Mit diesen Ergebnissen erhält man

$$\mathbf{A} = \frac{1}{3\sqrt{2}} \begin{pmatrix} 3 & 2 & -2 & 1 \\ 0 & 1 & -1 & -4 \\ 0 & -3 & -3 & 0 \\ 3 & -2 & 2 & -1 \end{pmatrix} \begin{pmatrix} 2\sqrt{2} & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \frac{1}{3} \begin{pmatrix} -2 & 1 & 2 \\ 2 & 2 & 1 \\ 1 & -2 & 2 \end{pmatrix}.$$

$\Diamond$

Bemerkung 5.5.5. *1. Die Singulärwerte von* $\mathbf{A}$ *und damit die rechteckige Diagonalmatrix* Σ *sind eindeutig bestimmt, die orthogonalen Matrizen* $\mathbf{U}$ *und* $\mathbf{V}$ *hingegen nicht. R liefert etwa eine andere Zerlegung.*

2. Es gibt eine ökonomischere Darstellung, wobei dann $\mathbf{U}$ *oder* $\mathbf{V}$ *nicht mehr quadratisch sind, d.h. es gilt*

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T = \sum_{i=1}^{r} s_{ii}\mathbf{u}_i\mathbf{v}_i^T$$

mit dem i-ten Spalten Vektor $\mathbf{u}_i$ *von* $\mathbf{U}$ *und dem i-ten Spaltenvektore* $\mathbf{v}_i$ *von* $\mathbf{V}$. *Im Beispiel gilt z.B.*

$$\mathbf{A} = \frac{1}{3\sqrt{2}} \begin{pmatrix} 3 & 2 & -2 \\ 0 & 1 & -1 \\ 0 & -3 & -3 \\ 3 & -2 & 2 \end{pmatrix} \begin{pmatrix} 2\sqrt{2} & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & 0 \end{pmatrix} \frac{1}{3} \begin{pmatrix} -2 & 1 & 2 \\ 2 & 2 & 1 \\ 1 & -2 & 2 \end{pmatrix}$$

oder

$$\mathbf{A} = \frac{1}{3\sqrt{2}} \begin{pmatrix} 3 & 2 \\ 0 & 1 \\ 0 & -3 \\ 3 & -2 \end{pmatrix} \begin{pmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{pmatrix} \frac{1}{3} \begin{pmatrix} -2 & 1 & 2 \\ 2 & 2 & 1 \end{pmatrix}$$

R liefert eine Variante mit r = 3 (auf Grund von Rundungsfehlern ist der dritte Eigenwert nicht gleich 0).

3. Es gilt $Kern\,(\mathbf{A}) = span\{\mathbf{v}_{r+1}, ..., \mathbf{v}_n\}$ *und* $Bild(\mathbf{A}) = span\{\mathbf{u}_1, ..., \mathbf{u}_r\}$.

4. Die obige Bestimmung der Singulärwerte als Quadratwurzeln der Eigenwerte von $\mathbf{A}^T\mathbf{A}$ *ist numerisch nicht immer stabil. Entsprechend wurden numerisch stabilere Verfahren entwickelt (s. z.B. (Golub & Kahan, 1956)).*

5. Für symmetrische Matrizen $\mathbf{A}$ *sind die Singulärwerte die Beträge der Eigenwerte. Sind alle Eigenwerte nicht-negativ, so ist die Hauptachsentransformation*

$$\mathbf{A} = \mathbf{Q}\Lambda\mathbf{Q}^T$$

auch eine Singulärwertzerlegung.

6. Die Anzahl der von $\mathbf{0}$ *verschiedenen Singulärwerte ist gleich dem Rang* r *von* $\mathbf{A}$*.*

$\Diamond$

Kapitel 6

Hauptkomponentenanalyse

(prinicipal components analysis)

6.1 Grundlagen

Sei $\mathbf{Y} = (Y_1, ..., Y_m)$ ein m-dimensionaler Zufallsvektor mit Erwartungswertvektor $\boldsymbol{\mu}$ und Kovarianzmatrix Σ $(= Var(\mathbf{Y}) := Cov(\mathbf{Y}, \mathbf{Y}))$. Zu finden sind neue Variablen $X_1, ..., X_m$, die sogenannten Hauptkomponenten, die

- unkorreliert sind (damit sind sie unabhängig und bilden eine Basis des m-dimensionalen Vektorraums $\mathbb{R}^m$, denn Variablen sind orthogonal, wenn sie unkorreliert sind und damit sind sie unabhängig, s. Bemerkung 6.4.1 S. 145),
- deren Varianzen mit wachsendem Index $j = 1, ..., m$ kleiner werden.
- deren Varianzen bezüglich der verbleibenden Dimensionen jeweils maximal sind
- die Linearkombination der alten Variablen sind:

$$Y_j = a_{1j}X_1 + ... + a_{mj}X_m = \mathbf{a}_j^t \mathbf{X}$$

Gesucht ist also in einem ersten Schritt die Hauptkomponente Y_1 mit maximaler Varianz

$$Var(Y_1) = Var\left(\mathbf{a}_1^t \mathbf{X}\right) = \mathbf{a}_1^t Var(\mathbf{X})\, \mathbf{a}_1 = \mathbf{a}_1^t \Sigma \mathbf{a}_1$$

Setzt man die Nebenbedingung $\mathbf{a}_i^t \mathbf{a}_i = 1$, so ergeben sich orthogonale Y_i, wie sich in Kürze zeigen wird. Es ist also eine Maximierungsaufgabe unter Nebenbedingung zu lösen. Dazu verwendet man die Lagrange-Methode:

Bemerkung 6.1.1. *Lagrange-Methode: Man maximiert $f(\mathbf{y})$ unter der Nebenbedigung $g(\mathbf{y}) = c$, indem man $L(\mathbf{y}, \lambda) = f(\mathbf{y}) - \lambda[g(\mathbf{y}) - c]$ maximiert - mit dem Lagrangemultiplikatoren λ. Man leitet $L(\mathbf{y}, \lambda)$ ab und bestimmt die Nullstellen.* ♢

Im vorliegenden Fall erhält man:

$$L(\mathbf{a}_1) = \mathbf{a}_1^t \Sigma \mathbf{a}_1 - \lambda_1 \left(\mathbf{a}_1^t \mathbf{a}_1 - 1\right)$$

$$\frac{\partial L(\mathbf{a}_1)}{\partial \mathbf{a}_1} = 2\Sigma \mathbf{a}_1 - 2\lambda_1 \mathbf{a}_1$$

und mit

$$2\Sigma \mathbf{a}_1 - 2\lambda_1 \mathbf{a}_1 = 0$$

$$(\Sigma - \lambda_1 \mathbf{I})\,\mathbf{a}_1 = 0$$

das Eigenwertproblem, denn damit gilt $\Sigma\mathbf{a}_1 = \lambda_1\mathbf{a}_1$. Die Eigenvektoren $\mathbf{a}_i$ der λ_i werden normiert. Hat ein Eigenwertraum mehr als einen linear unabhängigen Eigenvektor, so sind diese zu orthonormalisieren.

Bemerkung 6.1.2. *Wenn wir orthogonalen - und damit linear unabhängigen - normierten Eigenvektoren eine $n \times n$-Matrix $\mathbf{A}$ bilden, so ist $\mathbf{A}^t\mathbf{A} = \mathbf{I}$ und $\Sigma\mathbf{A} = \mathbf{A}\mathbf{\Lambda}$. Entsprechend ist dann $\mathbf{A}^t\Sigma\mathbf{A} = \mathbf{\Lambda}$ respektive $\Sigma = \mathbf{A}\mathbf{\Lambda}\mathbf{A}^t$. Dies wird die kanonische Darstellung von Σ genannt.* $\Diamond$

Die Matrix Σ wird im allgemeinen m Eigenwerte haben (m = Rang der Matrix Σ). Man nehme an, dass alle verschieden sind und nach Grösse absteigend $\lambda_1 > \lambda_2 > ... > \lambda_m$ geordnet sind. Welcher Eigenwert muss nun zur Bestimmung der ersten Hauptkomponenten gewählt werden? Wegen $\mathbf{a}_1^t\mathbf{a}_1 = 1$, gilt

$$Var\,(Y_1) = Var\left(\mathbf{a}_1^t\mathbf{X}\right) = \mathbf{a}_1^t\Sigma\mathbf{a}_1 = \mathbf{a}_1^t\lambda_1\mathbf{a}_1 = \lambda_1\mathbf{a}_1^t\mathbf{a}_1 = \lambda_1$$

Die Varianz auf der ersten Hauptkomponenten wird also am grössten, wenn für λ_1 den grössten Eigenwert wählt. Diesem entspricht dann der Eigenvektor $\mathbf{a}_1$, der zur Berechnung der ersten Hauptkomponente $Z_1 = \mathbf{a}_1^t\mathbf{Y}$ verwendet wird.
Als nächstes wird die zweite Hautpkomponenten Z_2 mit der zweitgrössten Varianz berechnet. Es soll wieder gelten $\mathbf{a}_2^t\mathbf{a}_2 = 1$. Zudem sollen Z_1 und Z_2 unkorreliert sein, d.h. der folgende Ausdruck soll 0 ergeben:

$$Cov\,(Y_1, Y_2) = Cov\left(\mathbf{a}_2^t\mathbf{X}, \mathbf{a}_1^t\mathbf{X}\right) = \mathbf{a}_2^t Cov\,(\mathbf{X}, \mathbf{X})\,\mathbf{a}_1 = \mathbf{a}_2^t\Sigma\mathbf{a}_1 \tag{6.1.1}$$

Da $\Sigma\mathbf{a}_1 = \lambda_1\mathbf{a}_1$, erhält man

$$\mathbf{a}_2^t\Sigma\mathbf{a}_1 = \lambda_1\mathbf{a}_2^t\mathbf{a}_1$$

was 0 ergibt, wenn $\mathbf{a}_2^t\mathbf{a}_1 = 0$ (zweite Nebenbedigung). Man erhält die zu maximierende Funktion

$$L\,(\mathbf{a}_2) = \left(\mathbf{a}_2^t\Sigma\mathbf{a}_2\right) - \lambda_2\left(\mathbf{a}_2^t\mathbf{a}_2 - 1\right) - \delta\left(\mathbf{a}_2^t\mathbf{a}_1 - 0\right)$$

und deren Ableitung

$$\begin{aligned}\frac{\partial L}{\partial \mathbf{a}_2} &= 2\Sigma\mathbf{a}_2 - 2\lambda_2\mathbf{a}_2 - \delta\mathbf{a}_1 \\ &= 2\,(\Sigma - \lambda_2\mathbf{I})\,\mathbf{a}_2 - \delta\mathbf{a}_1\end{aligned}$$

Man setzt mit 0 gleich:

$$2\,(\Sigma - \lambda_2\mathbf{I})\,\mathbf{a}_2 - \delta\mathbf{a}_1 = 0$$

Man multipliziert von links her mit $\mathbf{a}_1^t$ und erhält:

$$\begin{aligned}\mathbf{a}_1^t\left(2\,(\Sigma - \lambda_2\mathbf{I})\,\mathbf{a}_2 - \delta\mathbf{a}_1\right) &= 2\mathbf{a}_1^t\Sigma\mathbf{a}_2 - 2\lambda_2\mathbf{a}_1^t\mathbf{I}\mathbf{a}_2 - \delta\mathbf{a}_1^t\mathbf{a}_1 \\ &= 2\mathbf{a}_1^t\Sigma\mathbf{a}_2 - \delta \\ &= 0\end{aligned}$$

Da (s. (6.1.1)) $\mathbf{a}_1^t\Sigma\mathbf{a}_2 = Cov\,(Y_2, Y_1) = Cov\,(Y_1, Y_2) = \mathbf{a}_2^t\Sigma\mathbf{a}_1 = 0$, gilt $\delta = 0$. Man erhält als zu lösende Gleichung:

$$(\Sigma - \lambda_2\mathbf{I})\,\mathbf{a}_2 = 0$$

Es ist der zweitgrösste Eigenwert mit seinem normierten Eigenvektor zu wählen, um $Y_2 = \mathbf{a}_2^t\mathbf{X}$ zu berechnen. Mit analogen Überlegungen erhält man die übrigen Hauptachsen.
Fasst man die Eigenvektoren als Spalten zu einer Matrix $\mathbf{A}$ und die Hauptkomponenten zu einer Matrix $\mathbf{Z}$ zusammen, gilt

$$\mathbf{Y} = \mathbf{A}^t\mathbf{X}$$

wobei

$$Var\left(\mathbf{Y}\right)=\mathbf{\Lambda}$$

mit der Diagonalmatrix $\mathbf{\Lambda}$ und den Eigenvektoren $\lambda_1,...,\lambda_m$ auf der Diagonalen, denn die Y_i sind unkorreliert ($Y_i^tY_j$ für $i\neq j$ ist also 0 und die Varianzen von Z_i sind die λ_i). Zudem gilt

$$Var\left(\mathbf{Y}\right)=Var\left(\mathbf{A}^t\mathbf{X}\right)=\mathbf{A}^tVar\left(\mathbf{X}\right)\mathbf{A}=\mathbf{A}^t\Sigma\mathbf{A}$$

womit

$$\mathbf{\Lambda}=\mathbf{A}^t\Sigma\mathbf{A}$$

und damit wiederum

$$\Sigma=\mathbf{A\Lambda A^t}$$

gilt. Die Matrix der Eigenvektoren stellt damit einen Zusammenhang zwischen den beiden Kovarianzmatrizen Σ und $\mathbf{\Lambda}$ her.
Da (bei quadratischen Matrizen gilt: Spur$(BC)=Spur\left(CB\right)$, die Spur ist die Summe der Diagonaleinträge einer quadratischen Matrix, künftig manchmal auch abgekürzt durch $\mathtt{tr}(\mathbf{A})$) :

$$Spur\left(\Lambda\right)=Spur\left(\mathbf{A}^t\Sigma\mathbf{A}\right)=Spur\left(\Sigma\mathbf{A}\mathbf{A}^t\right)=Spur\left(\Sigma\right)=\sum_{i=1}^{m}Var\left(Y_i\right)$$

ist die Summe der Varianzen der ursprünglichen Vairiablen mit der Summe der Varianzen der Hauptkomponenten identisch. Damit kann man sagen, dass die i-te Hauptkomponente den Anteil

$$\frac{\lambda_i}{\sum_{j=1}^{m}\lambda_j}$$

an der Gesamtvarianz der ursprünglichen Variablen auf sich trägt. Oder man sagt, dass die ersten p Hauptkomponenten

$$\frac{\sum_{j=1}^{p}\lambda_j}{\sum_{j=1}^{m}\lambda_j}100\%$$

der Gesamtvarianz erklären.
In der Praxis wird Σ durch die empirische Kovarianzmatrix der Datenmatrix geschätzt. Entsprechend sind alle abgeleiteten Werte Schätzer (Eigenwerte und Eigenvektoren).

Bemerkung 6.1.3. *Wird auf eine Datenmatrix die Hauptachsentransformation angewendet, so werden die Punkte im durch die Eigenvektoren aufgespannten Vektoraum beschrieben. Die Werte der Hauptkomponenten stellen die Koordinaten der Punkte bezüglich der neuen Achsen dar - diese Werte auf englisch auch „scores“ genannt. Die orthogonale Projektion der durch die Zeilen $\mathbf{y}_i$ der Matrix $\mathbf{Y}$ gegebenen Punkte auf die die neue Achse $\mathbf{t}_j$ ist nämlich geben durch*

$$\frac{\mathbf{y}_i^t\mathbf{t}_j}{\mathbf{t}_j^t\mathbf{t}_j}\mathbf{t}_j=\left(\mathbf{y}_i^t\mathbf{t}_j\right)\mathbf{t}_j$$

Der Abstand von $\left(\mathbf{y}_i^t\mathbf{t}_j\right)\mathbf{t}_j$ vom Nullpunkt des Koordinatensystems ist dann gegeben durch

$$\left\|\left(\mathbf{y}_i^t\mathbf{t}_j\right)\mathbf{t}_j\right\|=\left|\mathbf{y}_i^t\mathbf{t}_j\right|\left\|\mathbf{t}_j\right\|=\left|\mathbf{y}_i\mathbf{t}_j\right|=\left|z_i\right|$$

Die Varianz der Hauptkomponente Z_i wird entsprechend maximal, wenn die Summe der quadrierten, orthogonalen Distanzen $\left\|d\right\|^2$ auf die Hauptachsen minimal wird. Ist die Streuung maximal, sind diese Distanzen minimal. Für $\left\|d\right\|^2$ gilt nämlich wegen des Satze von Pythagoras $\left\|d\right\|^2=\left\|P\right\|^2-\left\|a\right\|^2$ mit der quadrierten Distanz $\left\|P\right\|^2$ des Punktes vom Ursprung des Koordinatensystems und der quadrierten Distanz $\left\|a\right\|^2$ des auf die Hauptachse projezierten Punktes vom Ursprung. Bei maximaler Streuung ist die Summe der $\left\|a\right\|^2$ maximal und damit die Summe der $\left\|d\right\|^2$ minimal.

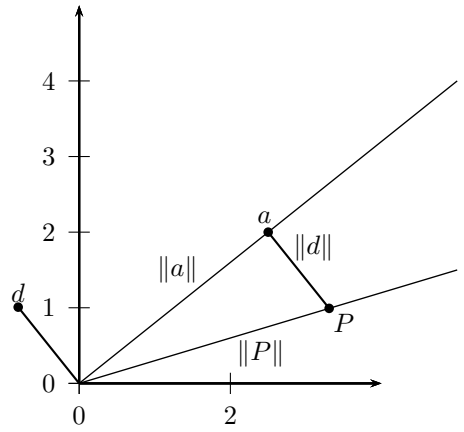


Abbildung 6.1.1: Graphik zur Bemerkung, dass Maximierung der Streuung zu Minimierung der orthogonalen Abständen führt. $\|a\|^2 + \|d\|^2 = \|P\|^2$

Im Gegensatz zur Regression werden bei der Hauptkomponentenmethode also nicht die vertikalen Summenquadrate von einer Gerade minimiert, sondern die kürzesten Abstände der Punkt zu den Hauptachsen gesucht. Dazu noch ein Beispiel. Die folgenden Daten werden orthogonal auf die erste Hauptachste projiziert (PC2 gibt dann die Koordinaten der Punkte bezüglich der zweiten Hauptachse an, deren Betrag die Distanzen der Punkte von der ersten Hauptachse sind):

id	*Y1*	*Y2*	*PC1=Z_1*	*PC2=Z_2*
1	*-0.14798*	*-0.01135*	*-0.12484617*	*-0.05011611*
2	*-1.19731*	*-0.23361*	*1.16296997*	*-0.04507080*
3	*0.99292*	*0.15288*	*-1.29985595*	*0.16958621*
4	*0.53165*	*0.10032*	*-0.86504457*	*0.04051044*
5	*1.81554*	*0.20739*	*-1.96120794*	*0.51385965*
6	*-0.37939*	*-0.25100*	*0.71360788*	*0.50544696*
7	*-1.30782*	*-0.26820*	*1.33112043*	*-0.01201512*
8	*-0.07482*	*0.02778*	*-0.28337141*	*-0.11920569*
9	*0.55637*	*0.13864*	*-0.99160582*	*-0.05583144*
10	*1.43925*	*0.37713*	*-2.22488623*	*-0.20982057*
11	*-0.16901*	*-0.02625*	*-0.06865612*	*-0.01963454*
12	*-1.41266*	*-0.42594*	*1.85398006*	*0.38268111*
13	*-1.95917*	*-0.41652*	*2.16062762*	*0.02123847*
14	*0.99973*	*0.26339*	*-1.62543058*	*-0.14766342*
15	*-0.96308*	*-0.08666*	*0.59240498*	*-0.32929745*
16	*-2.10678*	*-0.33517*	*2.01424986*	*-0.30558758*
17	*0.50802*	*0.19644*	*0.74610475*	*-0.08558747*
18	*-0.88945*	*-0.15498*	*-1.13016075*	*-0.25349263*

Man erhält die Graphik:

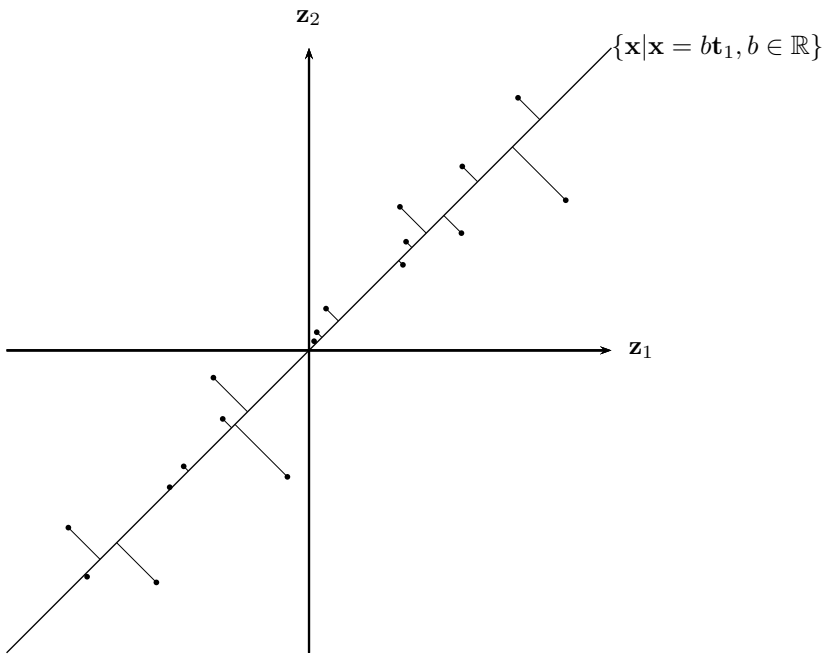


Abbildung 6.1.2: Orthogonale Projektion der Daten Z auf die Achse $\{x|x = bt_1, b \in \mathbb{R}\}$

Die Hauptkomponentenmethode besteht also darin, in einem ersten Schritt das orthogonale Koordinatensystem so zu drehen, das nachher auf der ersten Achse (durch 1. Eigenvektor aufgespannter Vektorraum) die Koordinaten der Punkte bezüglich dieser Achse maximale Varianz aufweisen. Im zweiten Schritt wird es um diese erste Achse so gedreht, dass die Koordinaten der Punkte bezüglich der zweiten Achse und bezüglich der restlichen Achsen maximale Varianz aufweist, etc. (s. folgende Graphik 6.1.3

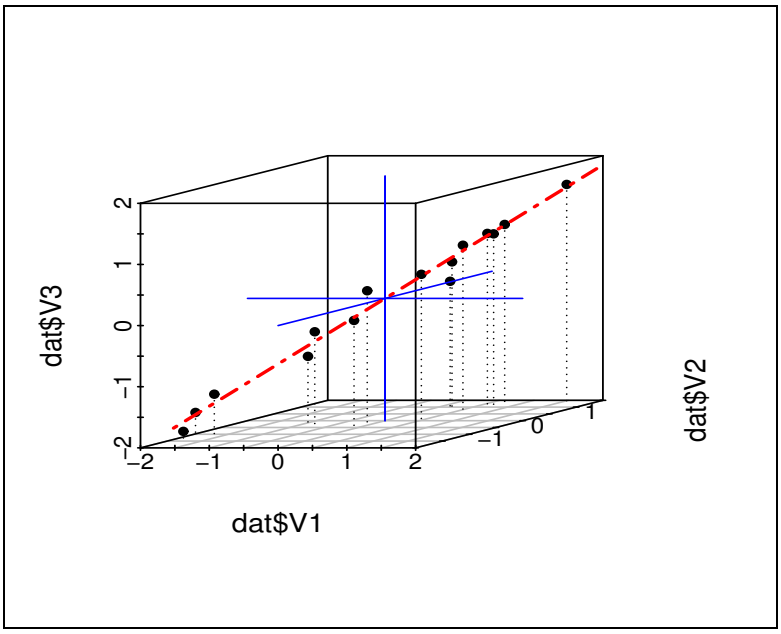


Abbildung 6.1.3: Die horizontale x-Achse wird im dreidimensionalen Raum in die gestrichelte Achse gedreht, die durch den ersten normierten Eigenvektor gegeben ist

Die Punktwolke ist gegeben durch (Name des Dataframes „dat“):

	V1	V2	V3
1	0.24	0.91	0.08
2	-0.55	-0.61	-0.41
3	0.88	0.78	0.89
4	-1.74	-1.30	-1.58
5	1.54	1.41	1.55
6	0.43	0.12	0.37
7	-0.06	-0.50	-0.25
8	0.13	-0.50	0.24
9	-0.68	-0.56	-0.82
10	0.47	0.65	0.45
11	-1.62	-1.69	-1.80
12	0.90	0.87	0.86
13	0.46	0.86	0.68
14	0.99	0.97	1.00
15	-1.39	-1.40	-1.26

Die Graphik erhält man durch

```
library(scatterplot3d)
s3d=scatterplot3d(x=dat$V1,y=dat$V2,z=dat$V3, type = "h",pch = 20,lty.hplot=3,angle=20,
 x.ticklabs=c(2",,1",,"0",,"1",,"2"), y.ticklabs=c(,1",,"0",,"1",),
 z.ticklabs=c(2",,1",,"0",,"1",,"2"))
s3d$points3d(c(-2,2), c(0,0), c(0,0), col="blue", pch=16, type="l")
s3d$points3d(c(0,0), c(-2,2), c(0,0), col="blue", pch=16, type="l")
s3d$points3d(c(0,0), c(0,0), c(-2,2), col="blue", pch=16, type="l")
s3d$points3d(c(0,0), c(0,0), c(-2,2), col="blue", pch=16, type="l")
a=prcomp(dat)$rotation[,1]
k=3
s3d$points3d(c(k*a[1],-k*a[1]), c(k*a[2],-k*a[2]), c(k*a[3],-k*a[3]), col="red", type="l",lty=30,lwd=2)
```

◊

Beispiel 6.1.4. *mtcars-R-Datensatz mit den folgenden Variablen:*

- **mpg: Fuel consumption (Miles per (US) gallon): more powerful and heavier cars tend to consume more fuel.*
- **cyl: Number of cylinders: more powerful cars often have more cylinders*
- **disp: Displacement (cu.in.): the combined volume of the engine's cylinders*
- **hp: Gross horsepower: this is a measure of the power generated by the car*
- **drat: Rear axle ratio (Hinterachsübersetzung): this describes how a turn of the drive shaft (Antriebsachse) corresponds to a turn of the wheels. Higher values will decrease fuel efficiency.*
- **wt: Weight (1000 lbs): pretty self-explanatory!*
- **qsec: 1/4 mile time: the cars speed and acceleration*
- **vs: Engine block: this denotes whether the vehicle's engine is shaped like a "V", or is a more common straight shape.*
- **am: Transmission: this denotes whether the car's transmission (Übertragung, vermutlich Gangschaltung) is automatic (0) or manual (1).*
- **gear: Number of forward gears (Getriebe): sports cars tend to have more gears.*
- **carb: Number of carburetors (Vergaser): associated with more powerful engines*

Note that the units used vary and occupy different scales.
Die Variablen 8 und 9 sind für die PCA ungeeignet, da nicht metrisch skaliert. Mit

```
matcars=mtcars[,-c(8:9)]
sigma=cov(matcars)
eigen(sigma)
```

erhält man die der Grösse nach absteigend geordneten Eigenwerte mit

```
eigen(sigma)$values
```

und die entsprechenden Eigenvektoren

```
(ev=eigen(sigma)$vectors)
```

Die Matrix der Hauptkomponenten erhält man dann mit (beim Rechnen mit Daten ist $\mathbf{Y} = (Y_1, ..., Y_m)$ *ja kein Zufallsvektor, sondern eine* $n \times m-$*Matrix mit* n *Werten der jeweiligen* n *Zufallsvariablen des Typs* Y_i*;* n *= Anzahl der Zeilen der Datenmatrix. Entsprechend ist nicht* $\mathbf{A}^T\mathbf{Y}$ *zu rechnen, sondern* $\mathbf{YA}$*):*

```
as.matrix(matcars)%*%ev
```

Mit

```
sum(eigen(sigma)$values[1:2])/sum(eigen(sigma)$values)*100
```

erhält man das Resultat, dass 99.9379% der Varianz durch die beiden ersten Hauptkomponenten erfasst wird. Effizienter geht es mit dem Befehl

```
(erg=prcomp(matcars,center=T,scale=F))
```

Zu beachten ist beim prcomp-Befehl, dass die Matrix der Daten, nicht die Kovarianzmatrix eingegeben werden muss. Geliefert werden die Wurzeln aus den Eigenwerten (erg$sdev) und die normierten Eigenvektoren (erg$rotation). Die Eigenwerte erhält man mit

```
(erg$sdev)^2
```

Mit

```
summary(prcomp())
```

erhält man die Standardabweichungen, die Anteile und die kumulative Anteile der Varianzen. Man kann auch mit princomp rechnen (Vorteil: man kann Teilmengen von Variablen analysieren). ◊

Bemerkung 6.1.5. *Gewöhnlich wird statt mit der Kovarianzmatrix mit der Korrelationsmatrix gearbeitet, wobei die Kovarianzmatrix der standardisierten Daten mit der Korrelationsmatrix identisch ist. Dadurch ergeben sich folgende Änderungen: die Varianz für alle Zufallsvariablen wird 1. Die Eigenwerte und Eigenvektoren verändern sich also: in der Korrelationsmatrix sind alle Diagonalelemente 1. Die Summe der Diagonalelemente oder die Summe der Varianzen der standardisierten Variablen ist damit mit m identisch. Die Summe der Eigenwerte ergibt ebenfalls m. Man würde berechnen:*

```
sigma1=cor(matcars)
eigen(sigma1)$values
eigen(sigma1)$vectors
```

Mit

```
sum(eigen(sigma1)$values[1:2])/sum(eigen(sigma1)$values)*100
```

erhält man das Resultat, dass 85.97822% (= 62.8437719+23.1344477) der Varianz durch die beiden ersten Hauptkomponenten erfasst wird. Es ergeben sich bezüglich Varianzerklärung also bedeutende Unterschiede. Dies liegt daran, das durch die Standardisierung die Varianzunterschiede, die verschiedenen Skalierungen der Variablen zuzuschreiben sind, ausgeglichen werden (misst man in cm statt in m, werden die Varianzen um 100^2 *grösser!)*
Für die Berechnung der Hauptkomponenten müssen die ursprünglichen Variablen standardisiert werden (mit scale(matcars)): .

```
hkst=as.matrix(scale(matcars)%*%eigen(cor(matcars))$vectors
```

Dieselben Hauptkomponenten erhält man auch mit

```
-prcomp(matcars,center=T,scale=T)$x
```

Es ist also ein Vorzeichenwechsel nötig, um dieselben Resultate zu erhalten. ◊

Bemerkung 6.1.6. *Üblicherweise werden die Daten mindestens zentriert (d.h. von jeder Variable wird das Mittel der Variable abgezählt). Dies wirkt sich im Gegensatz zur Standardisierung mittels der Standardabweichung nicht auf die Kovarianzmatrix aus. Es wird also gerechnet:*

$$\mathbf{Z} = \mathbf{A}^t (\mathbf{Y} - \boldsymbol{\mu})$$

Die Hauptkomponenten können in die ursprünglichen Variablen zurückgerechnet werden:

$$\begin{aligned} \mathbf{AZ} &= \mathbf{Y} - \boldsymbol{\mu} \\ \mathbf{AZ} + \boldsymbol{\mu} &= \mathbf{Y} \end{aligned}$$

◊

Bemerkung 6.1.7. *Wenn die Originalvariablen linear abhängig sind, werden einige Eigenwerte der Kovarianzmatrix Σ 0 sein. Die Dimension des Raumes, der die Beobachtungen enthält ist gleich dem Rang von Σ und dieser ist $m - k$, wenn k die Anzahl der Eigenwerte ist, die Null sind. Es gibt dann $m - k$ unabhängige lineare Beziehungen zwischen den Variablen. Die Existenz von exakt linearen Abhängigkeiten ist selten. Es geht vielmehr darum, beinahe lineare Beziehungen zwischen den Variablen aufzudecken. Wenn der kleinste Eigenwert λ_m beinahe Null ist, dann wird die m-te Hauptkomponente $\mathbf{h}_m$ beinahe konstant sein, denn ihre Varianz ist ja fast Null. Damit ist die Dimension des Datenraums beinahe kleiner als m. Wenn man die letzten Eigenwerte als klein betrachtet und damit die Dimension auf $p < m$ beschränkt, so kann der Anteil der esten p Hauptkomponenten an der Totalvariation als Gütekriterium dieses Vorgehens betrachtet werden.* ◊

Beispiel 6.1.8. *EWU-Daten (X1=Inflationsrate 1997 in %, X2=langfristiger Zinssatz 1997 in %, X3=öffentliche Neuverschulding 1997 in % des BIP 1997, X4=öffentlicher Schuldenstand 1997 in % des BIP (s. pca_mtcars_beispiel.r).*

```
EWU=data.frame(
Staat=c(" B"," DK"," D"," Fin"," F"," GR"," IRL"," I"," L"," NL"," A"," P"," S"," E"," GB"),
X1=c(1.4,1.9,1.4,1.3,1.2,5.2,1.2,1.8,1.4,1.8,1.1,1.8,1.9,1.8,1.8),
X2=c(5.7,6.2,5.6,5.9,5.5,9.8,6.2,6.7,5.6,5.5,5.6,6.2,6.5,6.3,7.0),
X3=c(2.1,-0.7,2.7,0.9,3.0,4.0,-0.9,2.7,-1.7,1.4,2.5,2.5,0.8,2.6,1.9),
X4=c(122.2,65.1,61.3,55.8,58.0,108.0,66.3,121.6,6.7,72.7,66.1,62.0,76.6,68.8,53.4),
Beitritt=c(1957,1973,1957,1995,1957,1981,1973,1957,1957,1957,1995,1986,1995,1986,1973))
EWUdat= EWU[,2:5]
prcomp(EWUdat)$sdevˆ2
summary(prcomp(EWUdat))
```

Die Eigenwerte ohne Standardisierung sind 836.17493418; 2.14621404; 1.37649151; 0.06121742. Die Anteile an der Varianz pro Hauptkomponente sind "Proportion of Variance 0.9957, 0.00256, 0.00164, 0.00007". Die Standardabweichungen sind" Standard deviation 28.9167, 1.46500, 1.17324, 0.24742". Im Beispiel würde man damit wohl nur die erste Hauptkomponente verwenden, das sie fast die gesamte Varianz trägt. Die Sache sieht aber anders aus, wenn man standardisiert:

```
prcomp(EWUdat,scale=T)$sdevˆ2
summary(prcomp(EWUdat,scale=T))
```

Die Eigenwerte sind [1] 2.56392387 0.93370674 0.44438531 0.05798409 und die
Proportion of Variance: 0.641 0.2334 0.1111 0.0145
Cumulative Proportion: 0.641 0.8744 0.9855 1.0000
Man würde nun vermutlich die ersten drei Hauptkomponenten weiterverwenden. ◊

6.2 Ladungen

Betrachtet man

$$Z_j = a_{1j}Y_1 + ... + a_{mj}Y_m = \mathbf{a}_j^t \mathbf{Y}$$

so sieht man, dass die Komponenten $(a_{1j}, ..., a_{mj})$ der Eigenvektoren $\mathbf{a}_j$ die ursprünglichen Variablen $Y_1, ..., Y_m$ bei der Berechnung der Hauptkomponente Z_j gewichten. Man spricht deshalb auch von *Ladungen (in der Beschreibung des R-Befehls prcomp „loadings")*. Gewöhnlich wird der Ladungsbegriff jedoch anders verwendet. Um zu einer Sicht auf die Ladungen zu kommen, welche den Zusammenhang der ursprünglichen Variablen mit den Hauptkomponenten leichter interpretierbar macht, kann man für die Berechnung der Eigenwerte und der Eigenvektoren von der Korrelationsmatrix der standardisierten Variablen ausgehen und betrachtet statt der Vektoren $\mathbf{a}_j$ die Vektoren

$$\mathbf{a}_j^* := \mathbf{a}_j \sqrt{\lambda_j}$$

Die obigen Ladungen werden also mit den Wurzeln aus den Eigenwerten gewichtet. Während

$$\mathbf{a}_j^t \mathbf{a}_j = 1$$

ist

$$\left(\mathbf{a}_j^*\right)^t \mathbf{a}_j^* = \left(\mathbf{a}_j \sqrt{\lambda_j}\right)^t \mathbf{a}_j \sqrt{\lambda_j} = \lambda_j \mathbf{a}_j^t \mathbf{a}_j = \lambda_j.$$

Setzt man

$$\mathbf{C} := (\mathbf{a}_1^*, ..., \mathbf{a}_m^*) \in \mathbb{R}^{m \times m}$$

so gilt

$$\mathbf{C} = \mathbf{A}\Lambda^{\frac{1}{2}}$$

Damit wird ersichtlich, wie $\mathbf{C}$ mit $\Sigma = \mathbf{A}\Lambda\mathbf{A}^t$ zusammenhängt:

$$\Sigma = \mathbf{A}\Lambda\mathbf{A}^t = \mathbf{A}\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}\mathbf{A}^t = \mathbf{A}\Lambda^{\frac{1}{2}}\left(\mathbf{A}\Lambda^{\frac{1}{2}}\right)^t = \mathbf{C}\mathbf{C}^t$$

wobei Σ die Korrelationsmatrix ist. Wegen

$$\begin{aligned} \mathbf{Z} &= \mathbf{A}^t\mathbf{Y} \\ \mathbf{Y} &= \mathbf{A}\mathbf{Z} \\ Y_i &= \sum_{k=1}^{m} a_{ik}Z_k \end{aligned}$$

gilt damit

$$cov\left(Z_j, Y_i\right) = cov\left(Z_j, \sum_{k=1}^{m} a_{ik}Z_k\right) = \sum_{k=1}^{m} a_{ik} cov\left(Z_j, Z_k\right) = a_{ij} Var\left(Z_j\right) = a_{ij}\lambda_j$$

Damit ist dann, da die Y_i standardisiert sind und entsprechend die Standardabweichung 1 aufweisen,

$$Kor\left(Z_j, Y_i\right) = \frac{cov\left(Z_j, Y_i\right)}{\sqrt{Var\left(Z_j\right)}\sqrt{Var\left(Y_i\right)}} = \frac{\lambda_j a_{ij}}{\sqrt{\lambda_j}} = a_{ij}\sqrt{\lambda_j} = a_{ij}^*$$

Die Matrix der Korrelationskoeffizienten zwischen $\mathbf{Z}$ und $\mathbf{Y}$ ist also gegeben durch

$$Kor\left(\mathbf{Z}, \mathbf{Y}\right) = \mathbf{A}\Lambda^{\frac{1}{2}} = \mathbf{C}$$

Damit messen die Komponenten von $\mathbf{C}$ die Korrelationen zwischen den Hauptkomponenten und den standardisierten Originalvariablen. Zudem gilt, dass die Varianz der j-ten Hauptkomponente

gleich der Summe der quadrierten Korrelationskoeffzienten zwischen dieser Hauptkomponente und allen m Originalvariablen ist (die Eigenvektoren sind normiert, d.h. $\sum_{i=1}^{m} a_{ij}^2 = 1$):

$$\sum_{i=1}^{m} Kor\left(Z_j, Y_i\right)^2 = \sum_{i=1}^{m} \left(a_{ij}^*\right)^2 = \lambda_j \sum_{i=1}^{m} a_{ij}^2 = \lambda_j$$

Damit kann man auf dem Hintergrund einer interpretierbaren Grösse (Pearson-Korrelationskoeffzienten) den Zusammenhang zwischen den Originalvariablen und den Hauptkomponenten beurteilen, was für die Interpretation nützlich sein kann. Gewöhnlich werden diese Korrelationen „Ladungen“ genannt.

Beispiel 6.2.1. *Man betrachte wieder die reduzierten R-mtcars-Daten.*

```
eigzer=prcomp(matcars,scale=T,center=T)
A=eigzer$rotation
L=eigzer$sdev^2
Cm=A%*%sqrt(L)
```

Damit ist

```
cov(scale(matcars))
```

mit

```
A%*%diag(L)%*%t(A)
```

identisch. Zudem gilt für die Hauptkomponenten HK

```
HK=as.matrix(scale(matcars))%*%A
```

dass

```
cor(HK,scale(matcars))
```

mit

```
Cm
```

identisch ist. Zuletzt gilt:

```
sum((Cm[,1])^2)
```

ist mit dem ersten Eigenwert identisch, d.h. die Summe der quadrierten Korrelationskoeffzienten der Variablen mit der ersten Hauptkomponente ist mit dem ersten Eigenwert identisch - analog für die anderen analogen Summen. Übrigens gilt sum((Cm[i,])^2)=1, für alle Zeilen i (s. (6.3.2), Seite 140). ◊

Mit dem folgenden R-Befehl kann man die Korrelationsmatrix zwischen Hauptkomponenten und Ursprungsvariablen berechnen (einzugeben ist das Objekt, das durch prcomp berechnet wird):

```
Korr_Y_PC=function(x){x$rotation%*%diag(x$sdev)}
```

Bemerkung 6.2.2. *Wenn eine ursprüngliche Variable Y_i mit den übrigen Variablen unkorreliert ist, dann stimmt ihre Varianz mit dem entsprechenden Eigenwert überein und der entsprechende Eigenvektor hat an der i-ten Stelle eine Eins und sonst überall eine Null. Damit wird Y_i selber zu einer Hauptkomponenten. Wenn alle Variablen paarweise unkorreliert sind, dann stimmen alle Hauptkomponenten mit den ursprünglichen Variablen überein. Eine Hauptkomponentenanalyse reproduziert dann also nur die ursprüngliche Datenmatrix - bis auf die Abweichungen, die den Zufälligkeiten der Zufallsvariablen zuzuschreiben sind.* ◊

6.3 Auswahl der Hauptkomponenten

Sei $\mathbf{P}$ die für die Hauptkomponentenanalyse verwendete Korrelationsmatrix. Dann gilt, wie gezeigt,

$$\Lambda = \mathbf{A}^t\mathbf{P}\mathbf{A}$$
$$\mathbf{P} = \mathbf{A}\Lambda\mathbf{A}^t$$

Reduziert man die Anzahl der Hauptkomponenten von m auf m_1, weil man die Hauptkomponenten mit zu kleinen Eigenwerten weglässt, so wird die Gesamtvarianz nicht mehr vollständig wiedergeben, da dann für $m_1 < m$

$$m = \sum_{j=1}^{m} Var\left(Y_j\right) > \sum_{j=1}^{m_1} Var\left(Y_j\right)$$

gilt. $\mathbf{P} = \mathbf{A}\Lambda\mathbf{A}^t$ lässt sich auch schreiben als

$$\mathbf{P} = \mathbf{A}\Lambda\mathbf{A}^t = \sum_{j=1}^{m} \lambda_j \mathbf{a}_j \mathbf{a}_j^t$$

Die rechte Summe wird als Spektralzerlegung der Korrelationsmatrix $\mathbf{P}$ bezeichnet. Jeder Summand

$$\mathbf{P}_j := \lambda_j \mathbf{a}_j \mathbf{a}_j^t$$

ist eine Matrix und es gilt entsprechend

$$\mathbf{P} = \sum_{j=1}^{m} \mathbf{P}_j$$

Dies wird Spektraldarstellung von $\mathbf{P}$ genannt. Die j-te Matrix $\mathbf{P}_j$ liefert den Beitrag der j-ten Hauptkomponente zur Korrelationsmatrix $\mathbf{P}$. Entsprechend kann sich die Korrelationsmatrix als Überlagerung von m Schichten vorstellen, wobei die j-te Schicht die Matrix $\mathbf{P}_j$ ist. Es gilt für jedes i

$$\sum_{j=1}^{m} Kor\left(Z_j, Y_i\right)^2 = \sum_{j=1}^{m} \left(a_{ij}^*\right)^2 = 1 \tag{6.3.2}$$

denn

$$\mathbf{P} = \begin{pmatrix} \left(a_{11}^*\right)^2 & a_{11}^* a_{21}^* & \cdots & a_{11}^* a_{m1}^* \\ a_{21}^* a_{11}^* & \left(a_{21}^*\right)^2 & \cdots & a_{21}^* a_{m1}^* \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}^* a_{11}^* & a_{m1}^* a_{21}^* & \cdots & \left(a_{m1}^*\right)^2 \end{pmatrix} + ... + \begin{pmatrix} \left(a_{1m}^*\right)^2 & a_{1m}^* a_{2m}^* & \cdots & a_{1m}^* a_{mm}^* \\ a_{2m}^* a_{1m}^* & \left(a_{2m}^*\right)^2 & \cdots & a_{2m}^* a_{mm}^* \\ \vdots & \vdots & \ddots & \vdots \\ a_{mm}^* a_{1m}^* & a_{mm}^* a_{2m}^* & \cdots & \left(a_{mm}^*\right)^2 \end{pmatrix}$$

Damit ist z.B. $p_{11} = \left(a_{11}^*\right)^2 + ... + \left(a_{1m}^*\right)^2$. Die Varianz der i-ten standardisierten Variable ist also die Summe der quadrierten Korrelationskoeffzienten dieser i-ten Variable mit allen Hauptkomponenten.
Wählt man nur die ersten m_1 Hauptkomponenten aus, so erhält man die Korrelationsmatrix des reduzierten Modells durch

$$\mathbf{P}_{m_1} := \sum_{j=1}^{m_1} \mathbf{P}_j$$

$\mathbf{P}_{m_1}$ gibt die Korrelationen zwischen den m standardisierten Originalvariablen an, wenn diese nur durch die m_1 grössten Hauptkomponenten erklärt würden. Man bezeichnet dann

$$\mathbf{P}^* := \mathbf{P} - \mathbf{P}_{m_1}$$

als die Rest- oder Fehlermatrix. Man zerlegt die Matrix $\mathbf{P}$ also in eine Summe von zwei Matrizen, d.h.

$$\mathbf{P} = \mathbf{P}_{m1} + \mathbf{P}^*$$

wobei $\mathbf{P}_{m1}$ die Korrelationen enthält, welche den wichtigen Hauptkomponenten zuzuschreiben sind, während $\mathbf{P}^*$ die restlichen, eher unwichtigen Korrelationen enthält.
Die Komponente ρ_{ii} der Diagonale von $\mathbf{P}_{m_1}$ wird „Kommunalität der i-ten standardisierten Originalvariable“ genannt. Sie drückt jenen Teil der Varianz der standardisierten Originalvariablen aus, der durch die m_1 wichtigsten Hauptkomponenten reproduziert wird. Sie können durch den folgenden Befehl berechnet werden: (x ist das durch prcomp produzierte Objekt, n ist die Anzahl der Hauptkomponenten, die zu berücksichtigen sind - also $n = m_1$)

```
Komm=function(x,n){m=length(x$rotation[,1])
if ( n > m) print("n ist groesser als die Dimensionen der Eigenvektormatrix") else
 {mat=matrix(rep(0,m^2),ncol=m)
 for (i in 1: n) mat=x$sdev[i]^2*x$rotation[,i]%*%t(x$rotation[,i])+mat
 commu=diag(mat)
 sdiag=sum(diag(mat))/m
 list('reduzierte Korrelationsmatrix'=mat,Kommunalitäten=commu,
 'Gesamtkommunalität (Anteil an Gesamtvarianz)'=sdiag,
Restmatrix=x$rotation%*%diag(x$sdev^2)%*%t(x$rotation)-mat)
 }
}
```

Für die WWU-Daten erhält man für die Berücksichtigung von 3 Hauptkomponenten mittels

```
EWUdat= EWU[,2:5]
mod=prcomp(EWUdat,scale=T)
Komm(mod,3)
```

$‘reduzierte Korrelationsmatrix‘	X1	X2	X3	X4
[1,]	0.9699710	0.9703619	0.3860371	0.4081558
[2,]	0.9703619	0.9707598	0.3840932	0.4062687
[3,]	0.3860371	0.3840932	0.7835760	0.7780344
[4,]	0.4081558	0.4062687	0.7780344	0.7733237
$Kommunalitäten				
[1]	0.9699710	0.9707598	0.7835760	0.7733237
$‘Gesamtkommunalität (Anteil an Gesamtvarianz)‘				
[1]	0.8744077			
$Restmatrix				
	X1	X2	X3	X4
X1	0.03002897	-0.029519037	0.013974052	-0.014400349
X2	-0.02951904	0.029240162	-0.006904028	0.007164628
X3	0.01397405	-0.006904028	0.216423954	-0.221490655
X4	-0.01440035	0.007164628	-0.221490655	0.226676311

Bei den ersten zwei Hauptkomponenten sind die Kommunalitäten gross. Bei den letzten zwei relativ klein (und im Beispiel fast identisch).
Es stellt sich die Frage, wie m_1 zu wählen ist. Dazu gibt es verschiedene Vorschläge, die selten zur selben Entscheidung führen:

- Man stellt die Eigenwerte λ_j auf der Ordinate und den Index j auf der Abszisse dar. Weist der Graph einen deutlichen Knick auf, so werden die Eigenwerte links vom Knick als wichtig

betrachtet. Dieses Verfahren heisst Scree-Test. Mit R für die EWU-Daten:

```
mod=prcomp(EWUdat,scale=T)
screeplot(mod,type="l")
```

Lässt man type="l"weg, so wird ein Säulendiagramm gezeichnet. Man sieht allfällige Knicks weniger gut. Es geht auch mit plot(mod,type="l")

- Man wählt die Hauptkomponenten, deren Eigenwerte grösser als 1 sind. Die Varianzen dieser Hauptkomponenten sind dabei grösser als die Varianzen der standardisierten Ursprungsvariablen, die ja mit 1 identisch sind.

- Man setzt eine prozentuale Grösse als Schwelle. Man verwendet die Hauptkomponenten, so dass der prozentuale Anteil der Summe derer Eigenwerte diese Schwelle knapp überschreitet. Man wählt also das $\min(m_1)$, so dass z.B.

$$\frac{1}{m}\sum_{i}^{m_1}\lambda_i > 0.9$$

- unter der Voraussetzung, dass die Daten eine multivariate Normalverteilung besitzen, kann ein Test von Barlett durchgeführt werden. Es handelt sich um einen Test auf Gleichheit von Varianzen. Es wird überprüft, ob sich die $m - m_1$ kleinsten Eigenwerte noch signifikant unterscheiden.

$$H_0 : \lambda_{m_1+1} = ... = \lambda_m$$

Die Teststatistik ist gegeben durch

$$B = (n-1)\left[-\ln\left(\prod_{j=1}^{m}\lambda_j\right) + \ln\left(\prod_{j=1}^{m_1}\lambda_j\right) + (m-m_1)\ln\left(\frac{m-\sum_{i}^{m_1}\lambda_i}{m-m_1}\right)\right]$$

Die Teststatistik ist chi-quadrat-verteilt mit $\frac{(m-m_1+2)(m-m_1-1)}{2}$ Freiheitsgraden. n = Anzahl Datensätze. Jede der Hauptkomponenten muss mit einem Q-Q-Plot auf Normalverteilung überprüft werden. Die folgende Funktion berechnet die Tests für aufsteigende m_1. Für x ist das durch prcomp produzierte Objekt einzugeben.

```
barlett.pca=function(x){n=length(x$rotation[,1])
 N=length(x$x[,1])
 ew=x$sdevˆ2
 a=c(Testwert=0,df=0,pwert=0)
 for (i in 1:(n-2)) {
    testw=(N-1)*(-log(prod(ew))+log(prod(ew[1:i]))+(n-i)*log((n-sum(ew[1:i]))/(n-i)))
    df=0.5*(n-i+2)*(n-i-1)
    pwert=1-pchisq(testw,df)
    a=rbind(a,c(testw,df,pwert))
 }
a[-1,]}
```

Am Beispiel der EWU-Daten erhält man mit

```
EWUdat= EWU[,2:5]
mod=prcomp(EWUdat,scale=T)
barlett.pca(mod)
```

das Ergebnis:

Anzahl_PC	Testwert	df	pwert
4	21.24009	5	0.0007296633
3	12.53723	2	0.0018948519

Man würde die ersten drei Hauptkomponenten verwenden.
Sind die Werte der Hauptkomponenten nicht normalverteilt, könnte man den Levene-Test verwenden. Man wendet den Kruskal-Wallis-Test auf den Betrag der Werte der Hauptkomponenten an (x ist das durch prcomp produzierte Objekt).

```
levene.pca=function(x){n = length(x$x[1,])
aus=c(AnzahlPC=0,chisquared=0,df=0,p_Wert=0)
 for (i in (1:(n-1))){
   y=abs(x$x[,i:n])
   dat=as.vector(y)
   m=length(x$x[,1])
   g=rep(seq(i:n),each=m)
   te=kruskal.test(dat,g)
   aus=rbind(aus,c(n-i+1,te$statistic,te$parameter,te$p.value))
 }
aus[-1,]
}
```

Auf die EWU-Daten angewendet ergibt sich mit levene.pca(mod)

AnzahlPC	chisquared	df	p_Wert
4	18.33202	3	0.0003756623
3	17.20348	2	0.0001837859
2	10.06839	1	0.0015083502

Die Testwerte sind im Beispiel durchgehend grösser und die p-Werte durchgehend kleiner als beim Barlett-Test. Man würde alle Hauptkomponenten verwenden. Das Problem mit diesen Tests ist, dass die die Unterschiede zwischen den Varianzen messen, unabhängig davon, ob diese Varianzen gross oder klein sind. Der Unterschied zwischen einer Varianz, die knapp kleiner als 1 ist, und einer Varianz, die nahe bei 0.1 liegt, kann statistisch signifikant sein. Trotzdem möchte man die Hauptkomponenten mit so kleinen Varianzen nicht berücksichtigen.

6.4 Graphische Darstellungen

Man kann die Korrelationen zwischen den ersten zwei Hauptkomponenten und den Ursprungsvariablen als Punkteplot darstellen. Im Beispiel der EWU-Daten erhält man mit

```
EWUdat= EWU[,2:5]
mod=prcomp(EWUdat,scale=T)
lad=Korr_Y_PC(mod)
ladeHK1=lad[,1]
ladeHK2=lad[,2]
plot(ladeHK1,ladeHK2,xlim=c(0.6,1),ylim=c(-1,1),lwd=0.5,
 xlab="Korrelationen erste Hauptkomponente",
 ylab="Korrelationen zweite Hauptkomponenten",
 main="Korrelationen Hauptkomponenten mit Variablen")
text(ladeHK1+0.04,ladeHK2+c(0.1,-0.1,0.1,-0.1),labels=c("Inflationsrate",
"Zinssatz","Neuverschuldung","Schuldenstand"), cex=0.5)
```

die Abbildung 6.4.4:

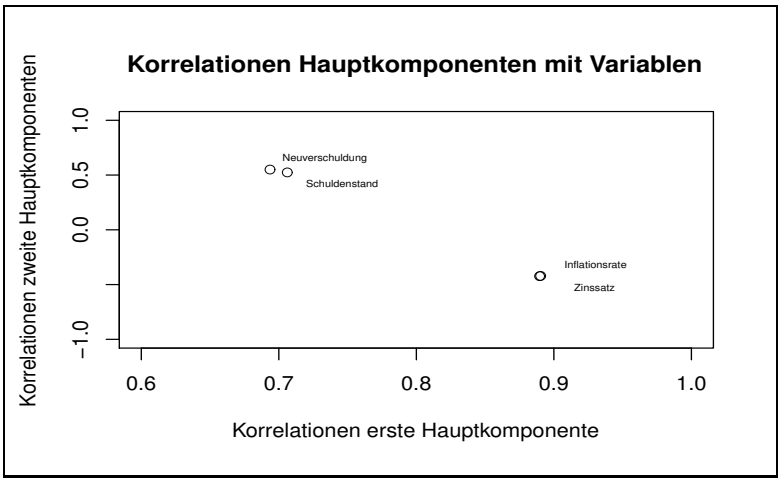


Abbildung 6.4.4: Korrelationen der standardisierten Ursprungsvariablen mit den ersten zwei Hauptkomponenten (Ladungen)

Man sieht deutlich zwei Gruppen von Variablen. Alle Variablen weisen eine positive Korrelation mit den ersten Hauptkomponenten auf. Zwei Variablen weisen eine positive Korrelation mit der zweiten Hauptkomponente (Neuverschuldung, Schuldenstand) und zwei Variablen eine negative Korrelation mit der zweiten Hauptkomponenten auf (Inflationsrate, Zinssatz). Die Frage ist, wie auf diesem Hintergrund die Hauptkomponenten inhaltlich zu interpretieren sind.
Eine Variante dieser Darstellung besteht darin, dass man statt der Punkte Pfeile vom Ursprung des Koordinatensystems in die Punkte darstellt. Mit

```
plot(ladeHK1,ladeHK2,type="n",xlim=c(0,1),ylim=c(-1,1),lwd=0.5,
 xlab="Korrelationen erste Hauptkomponente",
 ylab="Korrelationen zweite Hauptkomponenten",
 main="Korrelationen Hauptkomponenten mit Variablen")
text(ladeHK1+0.04,ladeHK2+c(0.1,-0.1,0.1,-0.1),labels=c("Inflationsrate",
"Zinssatz","Neuverschuldung","Schuldenstand"), cex=0.5)
null=rep(0,4)
abline(h=0)
arrows(null,null,ladeHK1,ladeHK2)
```

ergibt sich fürs Beispiel (s. Abbildung 6.4.5):

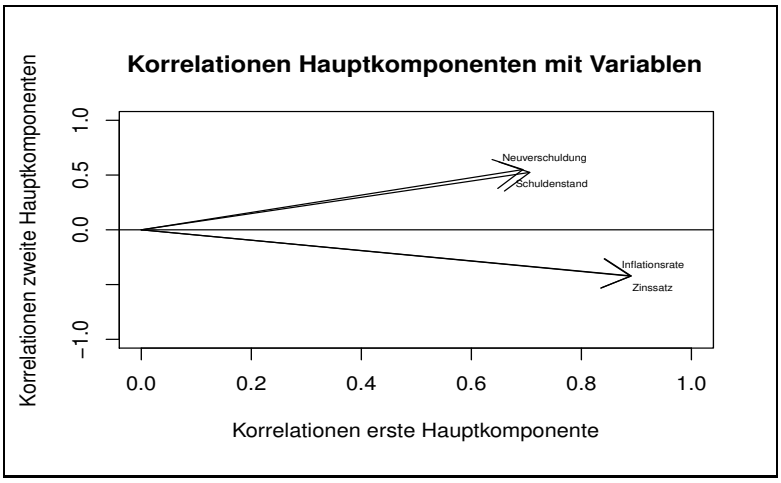


Abbildung 6.4.5: Darstellung der Korrelationen zwischen den Hauptkomponenten und den standardisierten Ursprungsvariablen (Ladungen) mit Pfeilen

Man sieht, welche Variablen nahe beieinanderliegen. Variablen, die näher bei der x-Achse liegen, korrelieren schwächer mit der 2. Hauptkomponente. Man sieht, ob die Korrelationen positiv oder negativ sind. Die zeilenweise Summe der Quadrate der Korrelationen ergibt 1. Die Varianz der standardisierten Originalvariablen ist damit gleich der Summe der Quadrate der Korrelationen der Hauptkomponenten mit dieser Variablen (siehe 6.3.2, S. 140). Die quadrierte Länge der Pfeile ist damit

$$\sum_{j=1}^{2} \left(a_{ij}^*\right)^2,$$

was den Kommunalitäten bei zwei Hauptkomponenten entspricht. Die Länge der Pfeile ist ≤ 1, da $\sum_{j=1}^{m} \left(a_{ij}^*\right)^2 = 1 = \sqrt{1}$. Zudem ist Im Beispiel erhält man:

```
lad=Korr_Y_PC(mod)
sum(lad[1,1:2]^2)
[1] 0.969971
sum(lad[2,1:2]^2)
[1] 0.9707598
sum(lad[3,1:2]^2)
[1] 0.783576
sum(lad[4,1:2]^2)
[1] 0.7733237
```

d.h. die ersten beiden Hauptkomponenten erklären z.B. einen Anteil von 0.969971 der Varianz der Variable Inflationsrate. etc.

Bemerkung 6.4.1. *Sei z_{ij} das i-te Datum der standardisierten Variable Z_j. Damit gilt einerseits, dass*

$$\sum_{i=1}^{n} z_{ij} z_{ik} = \sum_{i=1}^{n} \frac{x_{ij} - \bar{x}_j}{s_j} \frac{x_{jk} - \bar{x}_k}{s_k}$$

bis auf den Faktor $\frac{1}{n}$ die Korrelation der standardisierten Variablen Z_j und Z_k ist. Andererseits

ist dieser Ausdruck ein Skalarprodukt, womit wegen

$$\cos(\phi) = \frac{\mathbf{a}^t\mathbf{b}}{\|\mathbf{a}\|\,\|\mathbf{b}\|}$$
$$\mathbf{a}^t\mathbf{b} = \|\mathbf{a}\|\,\|\mathbf{b}\|\cos(\phi)$$

(ϕ ist der Winkel zwischen $\mathbf{a}$ *und* $\mathbf{b}$*) gilt*

$$r_{jk} := \frac{1}{n}\mathbf{z}_j^t\mathbf{z}_k = \frac{1}{n}\|\mathbf{z}_j\|\,\|\mathbf{z}_k\|\cos(\phi)$$

Zuletzt gilt

$$\|\mathbf{z}_j\|^2 = \sum_{i=1}^{n}\left(\frac{z_{ij}-\bar{z}_j}{s_j}\right)^2 = \frac{1}{s_j^2}\frac{n}{n}\sum_{i=1}^{n}(z_{ij}-\bar{z}_j)^2 = \frac{n}{s_j^2}s_j^2 = n$$

und damit

$$\begin{aligned} r_{jk} &= \frac{1}{n}\|\mathbf{z}_j\|\,\|\mathbf{z}_k\|\cos(\phi) \\ &= \frac{1}{n}\sqrt{n}\sqrt{n}\cos(\phi) \\ &= \cos(\phi) \end{aligned}$$

Die Korrelation zwischen den standardisierten Variablen Z_j und Z_k ist also mit dem Winkel zwischen diesen Variablen identisch. Stellt man die Korrelationen im Koordinatensystem dar, so liegen diese in der m-dimensionalen Einheitskugel (m = Anzahl Variablen) - je stärker die Korrelation, desto näher bei der Kugelhülle. Die maximale Korrelation entspricht damit $\cos(k\pi)$ *(gerade k; der Winkel ist* 0^o*) und die minimale* $\cos(k\pi)$ *(ungerade k; der Winkel ist* 180^o*). Die Korrelation ist null bei* $\cos\left(\frac{k}{2}\pi\right)$ *(Winkel von* 90^o *oder* 270^o*).* ◊

Zweitens kann man die Werte der ersten zwei Hauptkomponenten in einem Punktplot darstellen mittels

```
plot(mod$x[,1],mod$x[,2],xlab="erste Hauptkomponente",
 ylab="zweite Hauptkomponente",
 main="1. Hauptkomponente * 2. Hauptkomponente",
 ylim=c(-2.1,1.6))
text(mod$x[,1]+0.1,mod$x[,2]+0.1,EWU$Staat,cex=0.4)
```

(s. Abbildung 6.4.6).

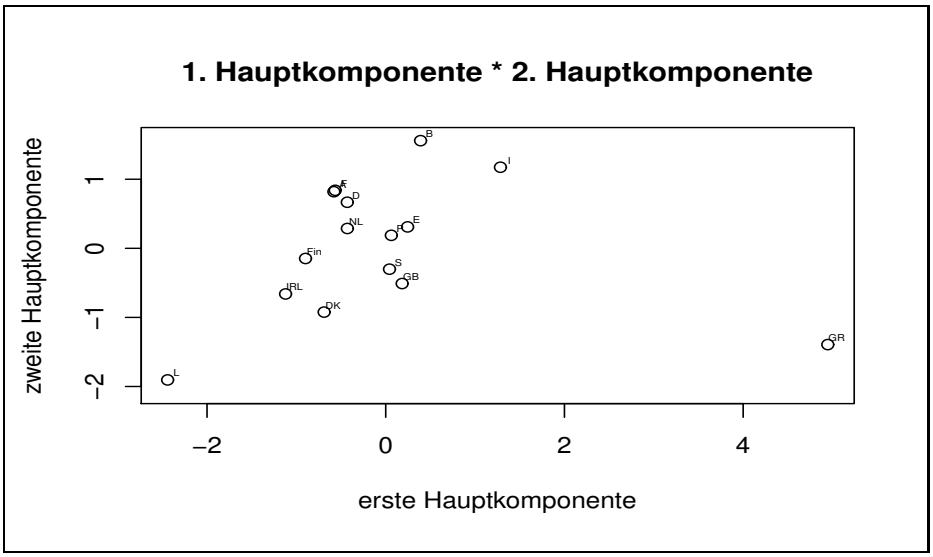


Abbildung 6.4.6: Punkte der ersten und zweiten Hauptkomponente

Man sieht, dass es zwei Ausreisser hat. Manchmal kann man in diesen Darstellungen auch Gruppen ausmachen, indem man nach einer nominalen Variable die Punkte verschiedene Farben oder Formen aufweisen lässt (z.B. nach Geschlecht). Man sieht dann, ob eine Hauptkomponente die Gruppen unterscheidet.
Schneller kann man die Werte der ersten zwei Hauptkomponenten in einem Punktplot darstellen mittels:

```
install.packages("ggplot2")
install.packages("ggfortify")
library(ggfortify)
autoplot(mod)
```

wobei die Daten transformiert werden und zusätzlich angegeben wird, wieviel der Varianz die beiden ersten Komponenten erklären (s. https://cran.r-project.org/web/packages/ggfortify/vignettes/plot_pca.html).
Drittens gibt es Darstellungen, welche die beiden bisherigen Darstellungsarten transformiert in einer Graphik vereinigen, sogenannte Biplots. In diesem Zusammenhang spielt die Singulärwertzerlegung eine wichtige Rolle. So erhält man z.B mit

```
library(ggfortify)
autoplot(mod,label = TRUE, label.size = 3,
loadings = TRUE, loadings.label = TRUE, loadings.label.size = 2)
```

die Graphik

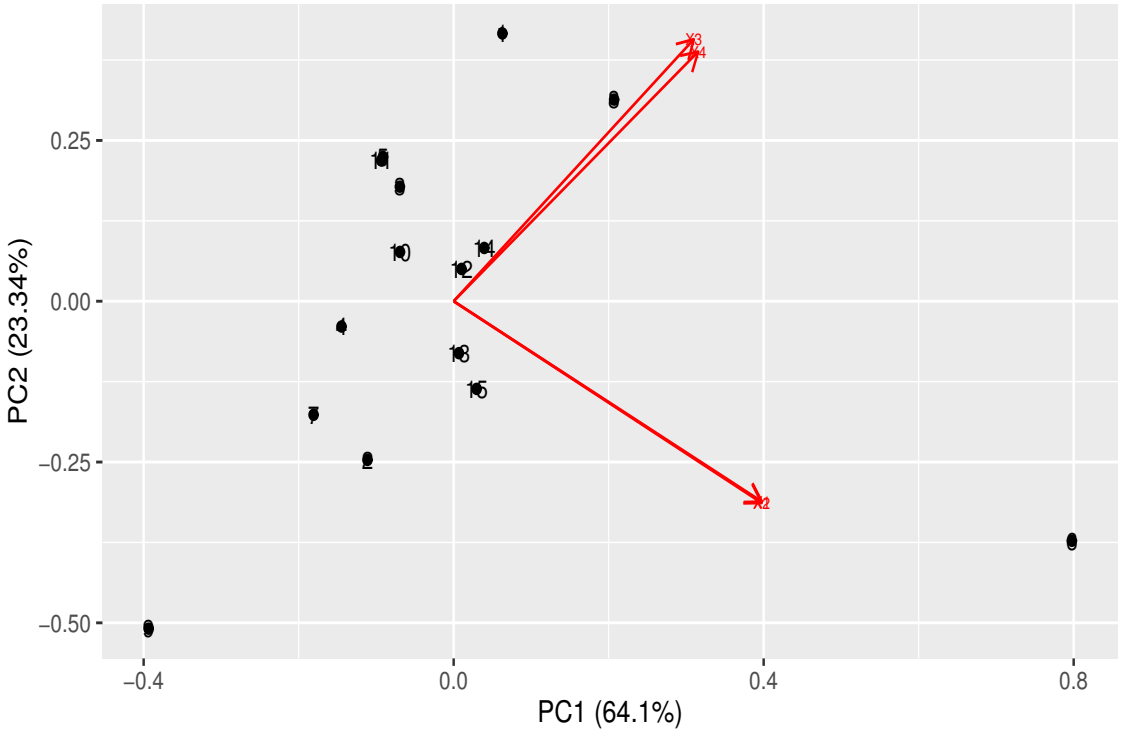


Abbildung 6.4.7:

Man beachte, dass die Skalierung der Punkte und Pfeile nicht den Skalierungen in den obigen Darstellungen entspricht (s. zu den Biplots Greenacre (2010)).

Bemerkung 6.4.2. *Wenn die Beobachtungen* $\mathbf{Y}$ *multivariat normalverteilt sind, d.h.*

$$\mathbf{Y} \sim N(\boldsymbol{\mu}; \Sigma),$$

so gilt für die Hauptkomponenten $\mathbf{Z} : \mathbf{Z} = \mathbf{A}^t(\mathbf{Y} - \boldsymbol{\mu})$ *ist multivariat normalverteilt ist und zwar gilt*

$$\mathbf{Z} \sim N(\mathbf{0}; \Lambda)$$

$\mathbf{Z}$ *ist ja eine Linearkombination der Spalten von* $\mathbf{Y}$. ◊

6.5 Rotation

Manchmal werden die Hauptkomponenten transformiert, um sie interpretierbarer zu machen (s. (Fahrmeir, Hamerle & Tutz, 1996, S. 677 ff.)). Es werden orthogonale und oblique Transformationen unterschieden. Bei der orthogonalen Rotation werden die Hauptkomponenten mittels einer orthogonalen Matrix

$$\begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix}$$

rotierte. Die erklärten Varianzanteile der Variablen (=Kommunalitäten) bleiben dadurch unverändert und die Hauptkomponenten bleiben unabhängig. Man kann sich die Frage stellen, wieso Rotationen von Vorteil sein können, wenn doch die Hauptachsentransformation die Achsen jeweils in eine Richtung dreht, so dass die Abstände der Punkte zu den Achsen minimal werden. Dies garantiert doch eine optimale Anpassung der Achsen an die Daten. Die Hauptachsentransformation

erfolgt allerdings im m-dimensionalen Raum. Lässt man unwichtige Achsen wegfallen, läuft dies auf eine Projektion der Punkt auf einen tieferdimensionalen Raum hinaus. Im tieferdimensionalen Raum müssen die mitprojizierten Achsen nicht mehr optimal durch die projizierten Punktwolken verlaufen (s. Abbildung 6.5.8). Die Graphik wurde produziert mit den dat-Daten Beispiel S. 134

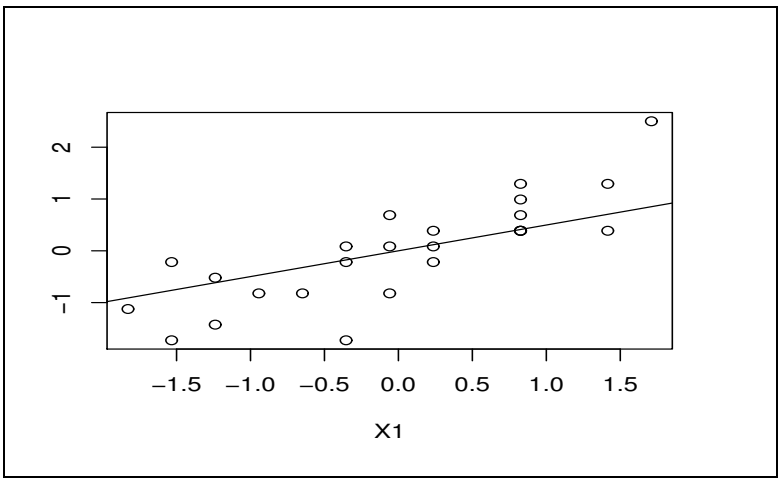


Abbildung 6.5.8: Projektion der standardisierten Punkte des Autobeispiels 7.2.1, Seite 159 in den zweidimensionalen Raum der ersten zwei Hauptkomponenten. Man sieht, dass die projizierte erste Hauptachse nicht optimal durch die Punkte verläuft.

```
plot(scale(dat[,1:2]))
lines(c(k*a[1],-k*a[1]), c(k*a[2],-k*a[2]))
```

Es gibt unterschiedliche Verfahren. Beim Varimaxverfahren (Kaiser, 1958) geht es darum, die Unterschiede bei den Korrelationen zwischen den Hauptkomponenten und den standardisierten Ausgangsvariablen möglichst gross werden zu lassen, d.h. die Varianz dieser Korrelationen soll möglichst gross werden. Es wird mit einem iterativen Verfahren gearbeitet. Mit R: varimax(). Einzugeben ist die Matrix der Korrelationen zwischen den Hauptkomponenten und den standardisierten Ausgangsvariablen, die in diesem Zusammenhang auch in R „matrix of loadings“ genannt wird. Im WWU-Beispiel ergeben sich in R nicht dieselben Werte wie mit SPSS, im Iris-Beispiel jedoch die gleichen (Iris-Daten von R):

```
irisX =iris[,1:4]
IPCA=prcomp(irisX,center=T,scale=T)
(vm=varimax(IPCA$rotation[,1:2]%*%diag(IPCA$sdev[1:2]),eps=10^(-20)))
scores= scale(IPCA$x[,1:2]) %*% vm$rotmat
```

In der Beschreibung des varimax-Befehls von R wird noch die Beschreibung des promax-Befehls geliefert. Diese liefert eine oblique Transformation (also nicht orthogonal, s. (Hendrickson & White, 1964)). Die Hauptkomponenten sind nachher nicht mehr orthogonal und damit nicht mehr unabhängig.

Bemerkung 6.5.1. *Mit SPSS: Analysieren, Dimensionsreduktion, Faktoranalyse, Extraktion: Hauptkomponenten oder mit Syntax-Befehl (für obige EWU-Beispieldaten)*
FACTOR
/VARIABLES X1 X2 X3 X4
/MISSING LISTWISE
/ANALYSIS X1 X2 X3 X4
/PRINT INITIAL CORRELATION EXTRACTION FSCORE

/CRITERIA MINEIGEN(0.5) ITERATE(25)
/EXTRACTION PC
/ROTATION VARIMAX
/SAVE AR(ALL)
/METHOD=CORRELATION.
Erhält man, wenn man entsprechendes anklickt, unter anderem:

- *Bei „Korrelationsmatrix" dieselbe Matrix wie mit R mittels cor(EWUdat).*
- *Bei „Extraktion" unter dem Titel „Kommunalitäten" die Kommunalitäten (Bei der Voreinstellung in SPSS werden nur Hauptkomponenten berücksichtigt, deren entsprechenden Eigenwerte grösser als 1 sind. Oben wurde /CRITERIA MINEIGEN(0.5) auf 0.5 gestellt, damit zwei Hauptkomponenten verwendet werden).*
- *Es wird die folgende Tabelle geliefert*

Erklärte Gesamtvarianz						
Anfängliche Eigenwerte				*Summen von quadrierten Faktorladungen für Extraktion*		
Komponente	*Gesamt*	*% der Varianz*	*Kumulierte %*	*Gesamt*	*% der Varianz*	*Kumuliert %*
1	*2.564*	*64.098*	*64.098*	*2.564*	*64.098*	*64.098*
2	*.062*	*23.343*	*87.441*	*0.934*	*23.343*	*87.441*
3	*.44*	*11.110*	*98.550*			
4	*.58*	*1.450*	*100.000*			
Extraktionsmethode: Hauptkomponentenanalyse						

beim Rotieren wird dieser Tabelle zusätzlich rechts angefügt:

Rotierte Summe der quadrierten Ladungen		
Gesamt	*% der Varianz*	*Kumulierte %*
2.700	*67.489*	*67.489*
1.133	*28.325*	*95.813*

- *Es wird eine sogenannte Komponentenmatrix geliefert. Diese besteht in den Korrelationen zwischen den Hauptkomponenten und den standardisierten Ausgangsvariablen (selbes Resultat wie Korr_Y_PC(mod), aber nur für die verwendeten Hauptkomponenten; wird also gewöhnlich „Ladungsmatrix" genannt).*
- *Zudem werden unter „Koeffizientenmatrix der Komponentenwerte" die standardisierten Werte der Eigenvektoren geliefert: ($\mathbf{T}\Lambda^{-\frac{1}{2}}$, mit R am Beispiel mod$rotation%*%solve(diag(mod$sdev)). $\mathbf{YT}\Lambda^{-\frac{1}{2}}$ liefert dann standardisierte Hauptkomponentenwerte (scores).*
- *Wird „Faktorscores" „Regression" angeklickt, werden in der Datentabelle unter FAC1_1 und FAC2_1 die Hauptkomponentenwerte (scores) standardisiert geliefert, sofern keine Rotation der Hauptkomponenten erfolgt (/ROTATION NOROTATE). In R kann man dies erreichen durch scale(mod$x) oder durch*

```
scale(EWUdat)%*%mod$rotation%*%solve(diag(mod$sdev))
```

- *bei /ROTATION VARIMAX und „Faktorscores" „Regression" werden die rotierten Hauptkomponentenwerte in der Datentabelle geliefert. In der „Ausgabe" werden die neuen „Ladungen" (Korrelationen) geliefert - unter dem Titel „Rotierte Komponentenmatrix". Zudem wird die Rotationsmatrix geliefert: die orthogonale Matrix, welche die Hauptkomponenten umrechnet (unter dem Titel „Komponententransformationsmatrix").*
- *https://www.methodenberatung.uzh.ch/de/datenanalyse_spss/interdependenz/reduktion/faktor.html*

$\Diamond$

6.6 Schlussbemerkungen

Die Hauptkomponentenanalyse hat einige Nachteile:

- Die Ergebnisse sind abhängig von der Skalierung und daher nicht eindeutig.
- Die Hauptkomponenten sind schwierig zu interpretieren.
- Die Interpretation ist subjektiv.
- Auch die Anzahl der auszuwählenden bedeutenden Hauptkomponenten ist nicht eindeutig.

Vorteile sind:

- Sie liefert Einsichten in die Korrelationsstruktur der Variablen
- Sie kann zu einer Reduktion der Anzahl Variablen führen.
- Die graphischen Darstellungen können zu Einsichten in die Daten führen (z.B. Cluster)
- eliminiert irrelevante Information
- eliminiert die Gefahr von Artefakten z.B. durch Ausreisser

Die wesentlichen Schritte einer Hauptkomponentenanalyse sind:

- Man entscheidet, ob manche oder alle Variablen zu transformieren sind (zentrieren,standardisieren)
- Man überprüft, ob die Beziehungen zwischen den Variablen affin sind (sonst kann man keine Pearson-Korrelationen zwischen den Variablen berechnen)
- Man berechnet die Korrelationsmatrix und überprüft, ob es offensichtliche Gruppen in den Variablen gibt. Wenn alle Korrelationen annähernd 0 sind, ist eine Hauptkomponentenanalyse nicht angebracht.
- Man entscheidet, welche Eigenwerte genügend gross sind. Die Zahl der zu berücksichtigenden Eigenwerte gibt die vermutliche Dimension der Daten an.
- Man überprüft, ob die Hauptkomponenten Hinweise auf Gruppierungen der Variablen geben und versucht die Hauptkomponenten zu interpretieren (Analyse der Korrelationen zwischen den standardisierten Variablen und den Hauptkomponenten).
- Benutzung der als wichtig erachteten Hauptkomponenten in anderweitigen Analysen (Dimensionsreduktion durch Variablenreduktion).

Home-pages mit Interpretationsbeispielen:

- http://stat.cmu.edu/~cshalizi/350/lectures/10/lecture-10.pdf
- https://onlinecourses.science.psu.edu/stat505/node/54
- https://mmi.psycho.unibas.ch/r-toolbox/faktorenanalyse.html

Bemerkung 6.6.1. *Die Hauptkomponentenanalyse wird in der Theorie gewöhnlich nicht der Faktoranalyse zugerechnet, in der Praxis wird unter dem Titel Faktoranalyse aber offenbar gewöhnlich die Hauptkomponentenanalyse angewendet. Sie hat den Vorteil, numerisch stabil zu sein. Die Hauptkomponentenanalyse verlangt metrisch skalierte Variablen. In der Praxis wird dies allerdings selten beachtet, was den Wert der entsprechnenden Ergebnisse fragwürdig macht.* ◊

Kapitel 7

Faktorenanalyse

7.1 Grundlagen

Man geht vom folgenden Modell aus

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{Lf} + \mathbf{e} \tag{7.1.1}$$

mit dem Zufallsvektor

$$\mathbf{y} = (Y_1, ..., Y_p)^t \in \mathbb{R}^p$$

(Vektor der Variablen Y_i, welche die manifesten Werte $y_i \in \mathbb{R}$ annehmen. Zudem verwendet man den Erwartungswertvektor

$$\boldsymbol{\mu} = (E(Y_1), ..., E(Y_p))^t \in \mathbb{R}^{p\times 1},$$

den Vektor der Faktoren

$$\mathbf{f} \in \mathbb{R}^{k\times 1}$$

(die latenten Zufallsvariablen) mit $k < p$, der Ladungsmatrix

$$\mathbf{L} \in \mathbb{R}^{p\times k}$$

und dem Fehlervektor

$$\mathbf{e} \in \mathbb{R}^{p\times 1}.$$

$\mathbf{Lf} \in \mathbb{R}^p$. Der Fehlervektor erfasst, was durch $\boldsymbol{\mu} + \mathbf{Lf}$ an $\mathbf{y}$ nicht erklärt wird und was entsprechend individuell den jeweiligen Zufallsvariablen Y_i zuzuschreiben ist (darum wird e_i auch „Einzelrestfaktor der Variable Y_i“ genannt). Die Ladungsmatrix kann man mit dem Vektor der Koeffizienten bei der Regression vergleichen. Sie gibt an, wieviel ein Faktor f_j zu einem Y_i beiträgt. Man spricht von der „Ladung l_{ij} der Variable Y_i auf den Faktor f_j“. Man erhält also

$$\begin{aligned}
Y_1 &= \mu_1 + l_{11}f_1 + l_{12}f_2 + ... + l_{1k}f_k + e_1\\
Y_2 &= \mu_2 + l_{21}f_1 + l_{22}f_2 + ... + l_{2k}f_k + e_2\\
&\vdots\\
Y_p &= \mu_p + l_{p1}f_1 + l_{p2}f_2 + ... + l_{pk}f_k + e_p
\end{aligned}$$

Das Modell kann umgewandelt werden in

$$\mathbf{y} - \boldsymbol{\mu} = \mathbf{Lf} + \mathbf{e}.$$

Damit ist $E(\mathbf{y} - \boldsymbol{\mu}) = E(\mathbf{y}) - E(\boldsymbol{\mu}) = \boldsymbol{\mu} - \boldsymbol{\mu} = \mathbf{0}$, womit auch $E(\mathbf{Lf} + \mathbf{e}) = \mathbf{0}$ sein muss.
Man geht von folgenden *Voraussetzungen* aus:

Condition 7.1.1. • $\mathbf{e}$ *und* $\mathbf{f}$ *sind intervallskaliert (metrisch skaliert, ohne absoluten Nullpunkt, nur Abstandsverhältnisse sind konstant). Der Nullpunkt und die Streuung sind damit durch Konvention festzulegen. Für* $\mathbf{f}$ *wird diesbezüglich folgendes festgelegt:*

- $\mathbf{f}$ *ist standardisiert, d.h.* $E(\mathbf{f}) = (E(f_1), ..., E(f_k))^t = \mathbf{0}$ *und* $E(f_i f_i) = 1$ *für* $i \in \mathbb{N}_k^*$.
- $\mathbf{f}$ *besteht aus unkorrelierten Faktoren, d.h.* $E(f_i f_j) = 0$ *für* $i \neq j$. *Damit gilt* $\mathbf{kov}(\mathbf{f}) := E\left(\mathbf{f}\mathbf{f}^t\right) = \mathbf{I}$ *(*$\mathbf{I}$ *= die Einheitsmatrix).*
- $E(\mathbf{e}) = \mathbf{0}$
- e_i *und* e_j *sind unkorreliert für* $i \neq j$. $\mathbf{V} := \mathbf{kov}(\mathbf{e}, \mathbf{e}) := E(\mathbf{e}\mathbf{e}^t) = \mathbf{diag}\left(v_1^2, ..., v_p^2\right)$ *ist damit eine Diagonalmatrix (überall 0, ausser auf der Diagonalen. Auf der Diagonalen stehen die Varianzen der* e_i*).* $\mathbf{V}$ *wird „Einzelrestvarianzmatrix" genannt.*
- $\mathbf{f}$ *und* $\mathbf{e}$ *sind unkorreliert, d.h.* $\mathbf{kov}(\mathbf{f}, \mathbf{e}) := E\left(\mathbf{f}\mathbf{e}^t\right) = \mathbf{0}$.

Für die Daten ergibt dies folgende Gleichungen. Für jede der Variablen Y_i werden N Werte (Daten) erhoben (p verschiedene Messdaten für N Untersuchungsobjekte). Man erhält die Datenmatrix für die Variablen Y_i:

$$\mathbf{Y} \in \mathbb{R}^{N \times p}.$$

Damit erhält man

$$\mathbf{Y} - \mathbf{M} = \mathbf{F}\mathbf{L}^t + \mathbf{E} \tag{7.1.2}$$

mit

- $\mathbf{M} \in \mathbb{R}^{N \times p}$ mit geschätzten i-ten Zeilen $(m_{i1}, ..., m_{ip}) = \left(\widehat{E(Y_1)}, ..., \widehat{E(Y_p)}\right)$ für $j \in \mathbb{N}_p^*$. In jeder Spalte steht damit durchgängig jeweils dieselbe geschätzte Zahl.
- $\mathbf{F} \in \mathbb{R}^{N \times k}$ (die Werte von $\mathbf{F}$ werden „Faktorwerte" genannt. Für jedes Untersuchungsobjekt gibt es pro Faktor einen Faktorwert. Diesen kann man als Messwert des Untersuchungsobjektes bezüglich der latenten Variable, dem Faktor, betrachten).
- $\mathbf{L} \in \mathbb{R}^{p \times k}$ ($\Longrightarrow \mathbf{L}^t \in \mathbb{R}^{k \times p}$ und $\mathbf{F}\mathbf{L}^t \in \mathbb{R}^{N \times p}$), für jede Variable Y_i k Faktorladungen - eine pro Faktor).
- $\mathbf{E} \in \mathbb{R}^{N \times p}$

Bemerkung 7.1.2. *Vergleich mit der multivariaten Regression (auf der Datenebene, s. (Fahrmeir et al., 1996, S. 132))*

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}$$

mit $\mathbf{Y} \in \mathbb{R}^{N \times p}$ *und* $\mathbf{X} \in \mathbb{R}^{N \times k}, \mathbf{B} \in \mathbb{R}^{k \times p}$, $\mathbf{E} \in \mathbb{R}^{N \times p}$. *Die Matrix der Faktorwerte* $\mathbf{F}$ *entspricht der Design-Matrix* $\mathbf{X}$ *in der multivariaten Regression. Während* $\mathbf{X}$ *bei der multivariaten Regression gegeben ist, ist* $\mathbf{F}$ *in der Faktorenanalyse nicht gegeben und muss geschätzt werden. In der Faktorenanalyse gibt es entsprechend bedeutend mehr zu schätzende Parameter und die Schätzprobleme sind erheblich grösser. Es müssen weitere Annahmen getroffen werden, um das Schätzproblem zu lösen.* ◊

Im Modell

$$\mathbf{y} - \boldsymbol{\mu} = \mathbf{L}\mathbf{f} + \mathbf{e}.$$

werden p Differenzen $Y_i - \mu_i$ durch $k + p$ latente Variablen erklärt (k Anzahl Faktoren, p Anzahl Einzelrestfaktoren e_i). Zur Verfügung stehen dabei nur die Rohdaten $\mathbf{Y}$. Das Modell ist durch die Daten entsprechend unterbestimmt. Die oben getroffenen Annahmen erlauben es jedoch, eine Hypothese über die Kovarianzmatrix von $\mathbf{y}$

$$\boldsymbol{\Sigma} := \mathbf{kov}(\mathbf{y}, \mathbf{y}) := E\left((\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})^t\right) \in \mathbb{R}^{p \times p}$$

zu testen. Es gilt nämlich gemäss Voraussetzungen (s. 7.1.1)

$$\begin{aligned}
\mathbf{\Sigma} &= E\left((\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})^t\right) \\
&= E\left((\mathbf{Lf}+\mathbf{e})(\mathbf{Lf}+\mathbf{e})^t\right) \\
&= E\left((\mathbf{Lf}+\mathbf{e})(\mathbf{f}^t\mathbf{L}^t+\mathbf{e}^t)\right) \\
&= E\left(\mathbf{Lff}^t\mathbf{L}^t+\mathbf{Lfe}^t+\mathbf{ef}^t\mathbf{L}^t+\mathbf{ee}^t\right) \\
&= E\left(\mathbf{Lff}^t\mathbf{L}^t\right)+E\left(\mathbf{Lfe}^t\right)+E\left(\mathbf{ef}^t\mathbf{L}^t\right)+E\left(\mathbf{ee}^t\right) \\
&= \mathbf{L}E\left(\mathbf{ff}^t\right)\mathbf{L}^t+\mathbf{L}E\left(\mathbf{fe}^t\right)+E\left(\mathbf{ef^t}\right)\mathbf{L}^t+\mathbf{V} \\
&= \mathbf{LIL}^t+\mathbf{0}+\mathbf{0}+\mathbf{V} \\
&= \mathbf{LL}^t+\mathbf{V}
\end{aligned}$$

Die Gleichung

$$\mathbf{\Sigma} = \mathbf{LL}^t + \mathbf{V} \tag{7.1.3}$$

wird in der Faktorenanalyse „Grundgleichung der Faktorenanalyse“ oder „Fundamentalsatz der Faktorenanalyse“ genannt. Für die Varianzen σ_{ii} der Variablen Y_i ($=(i,i)$-Komponenten der Diagonale von $\mathbf{\Sigma}$) gilt entsprechend

$$\sigma_{ii}^2 = \sum_{j=1}^{k} l_{ij}^2 + v_i^2$$

d. h. die Varianz σ_{ii}^2 setzt sich aus den Faktorladungen l_{ij}^2 der Variable Y_i bezüglich der verschiedenen Faktoren f_j und der Einzelrestvarianz v_i^2 additiv zusammen. So ist z.B. mit $p=3$ und $k=2$

$$\mathbf{LL}^t = \begin{pmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \\ l_{31} & l_{32} \end{pmatrix} \begin{pmatrix} l_{11} & l_{21} & l_{31} \\ l_{12} & l_{22} & l_{32} \end{pmatrix} = \begin{pmatrix} l_{11}^2+l_{12}^2 & l_{11}l_{21}+l_{12}l_{22} & l_{11}l_{31}+l_{12}l_{32} \\ l_{11}l_{21}+l_{12}l_{22} & l_{21}^2+l_{22}^2 & l_{21}l_{31}+l_{22}l_{32} \\ l_{11}l_{31}+l_{12}l_{32} & l_{21}l_{31}+l_{22}l_{32} & l_{31}^2+l_{32}^2 \end{pmatrix}$$

Man nennt die Summanden l_{ij}^2, $j \in N_k^*$, „Kommunalität der Variablen Y_i bezüglich des Faktors f_j“. Die Summe der Kommunalitäten ist der Teil der Varianz, der aus dem allen Variablen *gemeinsamen* Faktorenmodell resultiert.
Aus den faktoranalytischen Voraussetzungen 7.1.1 folgt zudem

$$\begin{aligned}
\mathbf{kov}(\mathbf{y},\mathbf{f}) &= E\left((\mathbf{y}-\boldsymbol{\mu})(\mathbf{f}-E(\mathbf{f}))^t\right) \\
&= E\left((\mathbf{Lf}+\mathbf{e})\mathbf{f}^t\right) \\
&= E\left(\mathbf{Lff}^t+\mathbf{ef}^t\right) \\
&= E\left(\mathbf{Lff}^t\right)+E\left(\mathbf{ef}^t\right) \\
&= \mathbf{L}E\left(\mathbf{ff}^t\right)+\mathbf{0} \\
&= \mathbf{LI} = \mathbf{L}
\end{aligned}$$

Die Ladungsmatrix $\mathbf{L}$ ist damit die Kovarianzmatrix zwischen den manifesten Variablen $\mathbf{y}$ und den latenten Faktorvariablen $\mathbf{f}$. Die Ladungsmatrix drückt damit (alle Grössen sind standardisiert) die Korrelationen zwischen den Faktoren und den gegebenen Datenvariablen aus, was für die Interpretation der Ergebnisse der Faktoranalyse von Nutzen ist.
$\boldsymbol{\mu}$ und $\mathbf{\Sigma}$ kann man erwartungstreu schätzen durch $\bar{\mathbf{y}} = (\bar{y}_1, ..., \bar{y}_p)^t \in \mathbb{R}^{p\times 1}$ mit

$$\bar{y}_i = \frac{1}{N}\sum_{i=1}^{N} y_i$$

(es handelt sich um die spaltenweisen Mittel der Matrix $\mathbf{Y}$) und die Kovarianzmatrix

$$\mathbf{S}^2 := \left(s_{jk}^2\right) := \frac{1}{N-1}\sum_{i=1}^{N}\left(y_{ij}-\bar{y}_j\right)\left(y_{ik}-\bar{y}_k\right) \in \mathbb{R}^{p\times p}.$$

Auf dieser Basis ist $\mathbf{L}$, $\mathbf{F}$ und k zu bestimmen. Man sagt auch, es sei eine Darstellung von $\mathbf{Y}-\mathbf{M}$ mittels $\mathbf{FL}^t+\mathbf{E}$ zu finden. Auf dieser Basis kann man dann die Einzelrestmatrix berechnen, womit alle Grössen der Gleichung

$$\mathbf{Y}-\mathbf{M} = \mathbf{FL}^t + \mathbf{E}$$

bestimmt sind.
$\mathbf{L}$ ist nur bis auf orthonormale lineare Transformationen bestimmt. Sei $\mathbf{O}$ eine orthogonale Matrix. Damit gilt $\mathbf{OO}^t = \mathbf{E}$. Entsprechend gilt bezüglich des faktorenanalytischen Modells.

$$\begin{aligned}\mathbf{y} &= \boldsymbol{\mu} + \mathbf{Lf} + \mathbf{e}\\ &\boldsymbol{\mu} + \mathbf{LEf} + \mathbf{e}\\ &= \boldsymbol{\mu} + \mathbf{LOO}^t\mathbf{f} + \mathbf{e}\\ &= \boldsymbol{\mu} + (\mathbf{LO})\left(\mathbf{O}^t\mathbf{f}\right) + \mathbf{e}\end{aligned}$$

Wenn $\mathbf{L}$ eine Lösung ist, ist also auch $\mathbf{LO}$ eine Lösung des faktgorenanalytischen Modells. In der Tat gilt:

$$\begin{aligned}E\left(\mathbf{O}^t\mathbf{f}\right) &= \mathbf{O}^t\mathbf{E}\left(\mathbf{f}\right) = \mathbf{O}^t\mathbf{0} = \mathbf{0}\\ cov\left(\mathbf{O}^t\mathbf{f}\right) &= \mathbf{O}^t cov\left(\mathbf{f}\right)\mathbf{O} = \mathbf{O}^t\mathbf{EO} = \mathbf{O}^t\mathbf{O} = \mathbf{E}\end{aligned}$$

Die orthonormal transformierten Faktoren haben also die gleichen statistischen Eigenschaften wie die nicht transformierten. Schliesslich ergibt sich auch der Fundamentalsatz für die transformierten Ladungen:

$$\Sigma = \mathbf{LL}^t + \mathbf{V} = \Sigma = \mathbf{LEL}^t + \mathbf{V} = \mathbf{LOO}^t\mathbf{L}^t + \mathbf{V} = (\mathbf{LO})\left(\mathbf{O}^t\mathbf{L}^t\right) + \mathbf{V}$$

Da orthogonale Matrizen geometrisch Rotationen des Koordinatensystems bedeuten, ist diese Tatsache die Grundlage der Faktorrotation. Gibt es also eine Lösung, so gibt es unendlich viele Lösungen des faktorenanalytischen Modells.
Aus der Diagonalisierungsmöglichkeit symmetrischer Matrizen folgt, dass für $k = p$ die symmetrisch Matrix $\boldsymbol{\Sigma}$ vom Rang p exakt in der Form $\mathbf{L}^t\mathbf{L} + \mathbf{V}$ darstellbar ist, nämlich mit $\mathbf{V} = \mathbf{0}$ (Hauptachsentransformation) - denn es gibt eine Darstellung $\boldsymbol{\Omega\Lambda\Omega}^t$ mit der Matrix $\boldsymbol{\Omega}$ der orthonormalen Eigenvektoren und der Diagonalmatrix $\boldsymbol{\Lambda}$ der Eigenwerte von $\boldsymbol{\Sigma}$. Damit erhält man

$$\boldsymbol{\Sigma} = \boldsymbol{\Omega\Lambda}^{\frac{1}{2}}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\Omega}^t = \left(\boldsymbol{\Omega\Lambda}^{\frac{1}{2}}\right)\left(\boldsymbol{\Omega\Lambda}^{\frac{1}{2}}\right)^t.$$

Aus demselben Grund ist bei gegebenem $\mathbf{V}$ und $k \leq p$ eine solche Darstellung dann gesichert, wenn $\boldsymbol{\Sigma} - \mathbf{V}$ positiv semidefinit ist mit $\mathtt{rg}\left(\boldsymbol{\Sigma} - \mathbf{V}\right) = k$. Sind diese Bedingungen nicht erfüllt, ist eine solche Darstellung nicht immer möglich und wenn mathematisch möglich, nicht immer statistisch sinnvoll. Dies ist etwa dann der Fall, wenn bei einem standardisierten $\mathbf{y}$-Variablenvektor einzelne Element von $\mathbf{L}$ betragsmässig grösser als Eins sind oder bei negativen Einzelrestvarianzen.
Es stellt sich die Frage, wie man auf eine der unendlich vielen Lösungen kommt. Um ein spezifisches $\mathbf{L}$ herauszupicken, sind Zusatzforderungen an $\mathbf{L}$ zu stellen. Man kann fordern, dass

$$\mathbf{L}^t\mathbf{V}^{-1}\mathbf{L} = \mathbf{diag}\{d_1, ..., d_k\} \text{ mit } d_1 > ... > d_k > 0.$$

Damit wird aus der Menge aller möglichen $\mathbf{L}$ genau eine Lösung eindeutig herausgegriffen, wie in der Folge zu zeigen ist:

Die Matrix

$$\mathbf{V}^{-\frac{1}{2}}\left(\mathbf{\Sigma}-\mathbf{V}\right)\mathbf{V}^{-\frac{1}{2}}$$

ist symmetrisch, da $\mathbf{\Sigma}-\mathbf{V}=\mathbf{L}\mathbf{L}^t$, sowie $\mathbf{V}$ symmetrisch sind. Entsprechend gibt es eine Darstellung von

$$\mathbf{V}^{-\frac{1}{2}}\left(\mathbf{\Sigma}-\mathbf{V}\right)\mathbf{V}^{-\frac{1}{2}}$$

mittels

$$\mathbf{V}^{-\frac{1}{2}}\left(\mathbf{\Sigma}-\mathbf{V}\right)\mathbf{V}^{-\frac{1}{2}}=\mathbf{\Omega}\mathbf{\Lambda}\mathbf{\Omega}^t$$

mit der Diagonalmatrix der Eigenwerte $\mathbf{\Lambda}$ (mit positiven Eigenwerten, die der Grösse nach auf der Diagonalen geordnet sind) und der entsprechenden orthogonalen Matrix der normalisierten Eigenvektoren $\mathbf{\Omega}$. Das bei gegebenen $\mathbf{V}$ und obigen $\mathbf{\Omega}$ und $\mathbf{\Lambda}$ eindeutig bestimmte

$$\mathbf{L}:=\mathbf{V}^{\frac{1}{2}}\mathbf{\Omega}\mathbf{\Lambda}^{\frac{1}{2}}$$

ist eine Lösung der Grundgleichung, denn

$$\begin{aligned}\mathbf{L}\mathbf{L}^t&=\mathbf{V}^{\frac{1}{2}}\mathbf{\Omega}\mathbf{\Lambda}^{\frac{1}{2}}\left(\mathbf{V}^{\frac{1}{2}}\mathbf{\Omega}\mathbf{\Lambda}^{\frac{1}{2}}\right)^t=\mathbf{V}^{\frac{1}{2}}\mathbf{\Omega}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{\Omega}^t\mathbf{V}^{\frac{1}{2}}\\&=\mathbf{V}^{\frac{1}{2}}\mathbf{\Omega}\mathbf{\Lambda}\mathbf{\Omega}^t\mathbf{V}^{\frac{1}{2}}\\&=\mathbf{V}^{\frac{1}{2}}\left(\mathbf{V}^{-\frac{1}{2}}\left(\mathbf{\Sigma}-\mathbf{V}\right)\mathbf{V}^{-\frac{1}{2}}\right)\mathbf{V}^{\frac{1}{2}}\\&=\mathbf{\Sigma}-\mathbf{V}\end{aligned}$$

Die Eigenwerte müssen paarweise verschieden und positiv sein, was mit Wahrscheinlichkeit 1 eintrifft, wenn $\mathbf{L}$ und $\mathbf{V}$ stetig verteilt sind. Mit dem oben festgelegten $\mathbf{L}$ ist nun

$$\begin{aligned}\mathbf{L}^t\mathbf{V}^{-1}\mathbf{L}&=\left(\mathbf{V}^{\frac{1}{2}}\mathbf{\Omega}\mathbf{\Lambda}^{\frac{1}{2}}\right)^t\mathbf{V}^{-1}\left(\mathbf{V}^{\frac{1}{2}}\mathbf{\Omega}\mathbf{\Lambda}^{\frac{1}{2}}\right)\\&=\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{\Omega}^{\mathbf{t}}\mathbf{V}^{\frac{1}{2}}\mathbf{V}^{-1}\mathbf{V}^{\frac{1}{2}}\mathbf{\Omega}\mathbf{\Lambda}^{\frac{1}{2}}\\&=\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{\Omega}^t\mathbf{\Omega}\mathbf{\Lambda}^{\frac{1}{2}}\\&=\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{\Lambda}^{\frac{1}{2}}\\&=\mathbf{\Lambda}.\end{aligned}$$

Damit ist die Bedingung $\mathbf{L}^t\mathbf{V}^{-1}\mathbf{L}=\mathbf{diag}\{d_1,...,d_k\}$ erläutert, wobei $\mathbf{diag}\{d_1,...,d_k\}=\mathbf{\Lambda}$ zu sein hat. Weitere Bedingungen für die Schätzbarkeit sind, dass die Zufallsvariablen stetig verteilt sind und dass $p\geq 2k+1$ (s. für eine ausführlichere Diskussion Fahrmeir et al. (1996, S. 646 ff.)).

Bemerkung 7.1.3. *Vergleich mit der Hauptkomponentenmethode: Man startet bei der Hauptkomponentenmethode mit* $\mathbf{Z}=\mathbf{A}^t\mathbf{Y}$*, wobei* $\mathbf{Y}$ *die Datenvariablen sind. Bezüglich der Korrelationsmatrix* $\mathbf{R}$ *ergibt sich* $\mathbf{R}=\mathbf{L}\mathbf{L}^t$ *mit den Ladungsmatrizen* $\mathbf{L}=\mathbf{A}\Lambda^{\frac{1}{2}}$*. Nach dem Weglassen der unwichtigen Hauptkomponenten entsteht die Zerlegung* $\mathbf{R}=\mathbf{L}_r\mathbf{L}_r^t+\mathbf{E}$ *mit der Restmatrix* $\mathbf{E}$*, wobei die Restmatrix keine Diagonalmatrix ist und mit der obigen Matrix* $\mathbf{V}$ *nichts zu tun hat. Die Problemstellung sieht also ziemlich anders aus, wobei man in der Gleichung der Hauptkompomentenmethode* $\mathbf{Y}$ *isolieren kann:* $\mathbf{Y}=\mathbf{A}\mathbf{Z}$*)* ◊

7.2 Schätzverfahren der Hauptfaktoranalyse

(Prinicipal Factor Analysis - Principal Factor Method)
Bei der Hauptfaktoranalyse handelt es sich um eine approximative Methode bezüglich des faktorenanalytischen Modells. Man schätzt zuerst die Korrelationsmatrix $\mathbf{R}$ durch die Datenmatrix $\mathbf{Y}$.

$$\mathbf{R}=Cor(\mathbf{Y})=Cov\left(\mathbf{Z}\right)$$

wobei $\mathbf{Z}$ die spaltenweise Standardisierung von $\mathbf{Y}$ ist. Man schätzt dann die Kommunalitäten $\mathbf{h}$ durch

$$\mathbf{h} = \mathbf{1} - diag(\mathbf{R}^{-1})$$

($\mathbf{1} \in \{1\}^p$, für eine Begründung s. Fahrmeir et al. (1996, S. 673 f.)) und man identifiert die Einzelrestvarianzmatrix $\mathbf{V}$ vorläufig mit

$$\mathbf{V} = \mathbf{E} - diag\,(\mathbf{h})$$

Dann zählt man $\mathbf{h}$ von der Diagonalen von $\mathbf{R}$ ab:

$$\mathbf{R}_h = \mathbf{R} - diag(\mathbf{h})$$

Auf $\mathbf{R}_h$ wird eine Hauptkomponentenanalyse angewendet. Dabei ist zu darauf zu achten, dass $\mathbf{R}_h$ nicht unbedingt positiv definit ist. Es können neben positiven auch negative Eigenwerte vorkommen. Eigenvektoren mit negativen oder mit 0 identischen Eigenwerten fallen weg. Zusätzlich lässt man Eigenvektoren mit zu kleinen Eigenwerten weg. Es werden dann die Ladungen $\mathbf{L}_k$ berechnet (k = Anzahl berücksichtigter Eigenwerte): Sei $\mathbf{T}_k$ die Matrix der normalisierten Eigenvektoren von $\mathbf{R}'$ mit k positiven, genügend grossen Eigenwerten und $\mathbf{\Lambda}_k$ die Diagonalmatrix dieser k ersten Eigenwerte von $\mathbf{R}_h$. Dann berechnet man

$$\mathbf{L}_k = \mathbf{T}_k \mathbf{\Lambda}_k^{-\frac{1}{2}}$$

und schätzt $\mathbf{V}$ durch

$$\mathbf{V} = diag(\mathbf{R} - \mathbf{L}_k \mathbf{L}_k^t)$$

Es gilt nun:

$$\mathbf{R} = \mathbf{L}_k \left(\mathbf{L}_k\right)^t + \mathbf{U}_k$$

$\mathbf{U}_k$ ist im Gegensatz zur den Voraussetzungen der Faktorenanalyse im Allgemeinen keine Diagonalmatrix. Es gilt

$$\mathbf{U}_k = \mathbf{l}_{k+1}\mathbf{l}_{k+1}^t + ... + \mathbf{l}_{k+1}\mathbf{l}_{k+1}^t + \mathbf{V}$$

Dies belegt, dass es sich bei der Hauptfaktormethode um eine approximative Vorgehensweise handelt. Im Allgemeinen wird diese Matrix aber besser einer Diagonalmatrix angepasst sein als die Restkorrelationsmatrix bei der Hauptkomponentenmethode.
Man kann nun die neuen Kommunaläten verwenden, um den Prozess zu wiederholen. Dabei ist bei jedem Durchgang darauf zu achten, dass nur positive Eigenvektoren mit positiven Eigenwerten verwendet werden. Man rechnet

$$\mathbf{R}_h^{neu} = \mathbf{R} - \mathbf{V}$$

und nimmt wiederum eine Hauptkomponentenanalyse vor. etc. Der Prozess konvergiert nicht immer. Entsprechend muss man neben einem Konvergenzkriterium (z.B. $\|\mathbf{V}_n - \mathbf{V}_{n-1}\| < \varepsilon$) noch eine maximale Zahl von Durchläufen festlegen. Es können insbesondere komplexe Zahlen ($h_i^2 < 0$) auftauchen oder Kommunalitäten $h_i^2 > 1$ (Heywood-cases). Dann muss die Iteration vorher abgebrochen werden. Mit der folgenden Routine kann man die entsprechende Analyse vornehmen (wobei in der ersten Schlaufe nur überprüft wird, ob die Eigenwerte grösser als 1 sind)
Eingabe:

- x = Daten;
- mF = Anzahl Faktoren - bei der Voreinstellung NULL wird die Anzahl Eigenwerte > 1 verwendet;
- n = Anzahl Runden - gerechnet wird mit der angebenenen Zahl (Voreinstellung: 20). Abgebrochen wird vorher, wenn die Norm der Differenz zweier aufeinanderfolgender Vs kleiner

ist als die Zahl hinter prec (Voreinstellung: 10^(-5)).

```
hauptfaktor=function(x,mF=NULL,n=20,prec=10^(-5)){ R=cor(x); m=length(R[,1])
V = 1 / diag(solve(R)); V1=1-V; l=0; Rh=R-V;
k=0; ev=eigen(Rh)$values; if (is.null(mF)) {for (i in 1:m) if (ev[i]>1) k=k+1 } else k = mF
repeat{Vvor=diag(V)
       L=eigen(Rh)$vectors[,1:k]%*%diag(eigen(Rh)$values[1:k]^(1/2))
       V=diag(R-L%*%t(L))
       Rh=R-diag(V)
       l=1+l
       if (l==n | norm(Vvor-diag(V))< prec) break}
ew=eigen(R)$values; ewe=eigen(Rh)$values
list(Kommunalitaeten=cbind(Start=V1,End=diag(Rh)), Anzahl_Runden=l,
    Anzahl_Faktoren = k, Ladungsmatrix=L[,1:k],
    EigenwerteStart=cbind(Werte=ew,Prozent=ew/sum(ew), kumProzent=cumsum(ew)/sum(ew)),
EigenwerteSchluss=cbind(Werte=ewe[1:2],ProzentanAnfangsEW=ewe[1:2]/sum(ew),
kumProzent=ewe[1:2]/sum(abs(ew))))}
```

Mit dem R-Paket psych (s. unten) werden drei Faktoren beibehalten. Gemäss Beschreibung lautet der Befehl "fa(auto, 3, n.obs=length(auto[,1]), rotate="none")".

Beispiel 7.2.1. *Es wurde die Wichtigkeit verschiedener Faktoren beim Autokauf erhoben X1: Anschaffungspreis; X2: Betriesbkosten; X3: Umfang der Serienausstattung; X4: Styling der Karosserie; X5: Prestige der Marke; X6: Fahrkomfort; X7: Raumangebot (Beispiel aus Rinne (2000)). Die Daten sind also kaum alle metrisch skaliert und eignen sich eigentlich nicht für eine Faktorenanalyse.*

id	X1	X2	X3	X4	X5	X6	X7
1	7	7	6	7	11	8	9
2	4	9	4	9	11	5	9
3	10	10	7	6	6	3	4
4	5	5	7	15	17	13	19
5	12	11	12	11	11	13	12
6	12	14	10	9	9	10	9
7	12	11	10	5	6	8	5
8	5	8	7	11	14	9	14
9	3	6	6	16	20	13	16
10	12	12	10	0	3	4	0
11	8	9	9	10	11	11	11
12	14	14	11	8	10	11	7
13	4	4	7	10	13	8	15
14	6	7	7	2	6	5	3
15	14	11	13	4	5	7	3
16	12	11	11	14	8	10	9
17	9	7	8	10	7	9	8
18	8	4	7	4	6	5	6
19	9	10	9	10	12	10	8
20	10	9	9	6	8	9	8
21	8	10	8	10	9	11	11
22	9	12	10	8	7	6	5
23	15	18	17	14	14	17	14
24	10	11	10	8	11	10	12
25	12	13	12	10	8	10	9

```
auto=auto[,-1]    #Weglassen der id-Spalte
```

Mit

```
hauptfaktor(auto)
```

erhält man:

$'Kommunalitaeten'	Start	End
X1	0.9151113	0.9739554
X2	0.7544061	0.7099052
X3	0.8717081	0.9018061
X4	0.8005869	0.8107675
X5	0.8946779	0.8933994
X6	0.8883035 0	0.911138
X7	0.9058643	0.9430744

(Die Startkommunalitäten sind mit den Ergebniss von SPSS identisch, die Endkommunalitäten unterscheiden sich ein wenig).

$Anzahl_Runden	
[1]	18

(Bei SPSS werden 7 Runden durchgeführt).

$Anzahl_Faktoren	
[1]	2

(Gleich wie SPSS)

$Ladungsmatrix	[,1]	[,2]
[1,]	0.5887184	-0.79206439
[2,]	0.3346763	-0.77323796
[3,]	0.2642402	-0.91213116
[4,]	-0.8349573	-0.33706638
[5,]	-0.9411418	-0.08747322
[6,]	-0.6553062	-0.69405455
[7,]	-0.9588541	-0.15386073

(Bis auf die Vorzeichen identisch mit den Ergebnissen von SPSS, die dort aber unter dem Titel Faktoren geliefert werden).

$EigenwerteStart	Werte	Prozent	kumProzent
[1,]	3.56680837	0.509544052	0.5095441
[2,]	2.81329785	0.401899693	0.9114437
[3,]	0.25857504	0.036939291	0.9483830
[4,]	0.16300397	0.023286282	0.9716693
[5,]	0.08178774	0.011683963	0.9833533
[6,]	0.06551600	0.009359429	0.9927127
[7,]	0.05101103	0.007287290	1.0000000

(Die selben Werte wie bei SPSS)

$EigenwerteSchluss	Werte	ProzentanAnfangsEW	kumProzent
[1,]	3.460149	0.4943070	0.4943070
[2,]	2.683896	0.3834137	0.3834137

(Dieselben Werte wie bei SPSS). Eine Rotation liefert schliesslich ein deutlicheres Bild:

```
ha=hauptfaktor(auto)
varimax(ha$Ladungsmatrix)
```

ergibt:

	[,1]	*[,2]*
[1,]	*0.254*	*-0.954*
[2,]		*-0.842*
[3,]		*-0.954*
[4,]	*-0.900*	
[5,]	*-0.907*	*0.267*
[6,]	*-0.866*	*-0.402*
[7,]	*-0.948*	*0.212*

(schwache Ladungen < 0.1 werden nicht angezeigt. Man sieht, dass die ersten drei Variablen auf den zweiten Faktor laden, die letzten vier auf den ersten (mögliche Intepretation: die ersten drei Variablen haben etwas mit der Wirtschaftlichkeit zu tun, die zweiten mit dem Auto als Wohlfühlgut). ◊

Bemerkung 7.2.2. *SPSS:*
GET DATA
/TYPE=XLSX
/FILE='I:\dateien\Statistik\Faktorenanalyse\auto.frame.xlsx'
SAVE OUTFILE='I:\dateien\Statistik\Faktorenanalyse\auto.frame.sav'
/COMPRESSED.
FACTOR
/VARIABLES X1 X2 X3 X4 X5 X6 X7
/MISSING LISTWISE
/ANALYSIS X1 X2 X3 X4 X5 X6 X7
/PRINT INITIAL EXTRACTION ROTATION
/CRITERIA MINEIGEN(1) ITERATE(25)
/EXTRACTION PAF
/CRITERIA ITERATE(25)
/ROTATION VARIMAX
/SAVE REG(ALL)
/METHOD=CORRELATION. ◊

7.3 Die ML-Faktorenanalyse

Die ML-Faktorenanalyse beruht auf Maximum-Likelihood-Schätzern der Grössen $\mathbf{L}$ und $\mathbf{V}$. Für die passende Anzahl k können auf diesem Hintergrund Likelihood-Quotienten-Tests verwendet werden. Die Faktorenwerte werden ebenfalls mit ML-Methoden geschätzt. Vorausgesetzt ist, dass die Daten normalverteilt sind, d.h.

$$\mathbf{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

mit $\operatorname{rg}(\boldsymbol{\Sigma}) = p$. Ohne die Normalitätsvoraussetzung erhält man Quasi-ML-Schätzer, die asymptotisch dieselben Eigenschaften eines ML-Schätzers aufweisen, nämlich Konsistenz und asymptotische Normalität. Dafür ist nur vorauszusetzen, dass $\mathbf{y}$ erste und zweite Momente $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ besitzt und $\boldsymbol{\Sigma}$ positiv definit ist (d.h. nur positive Eigenwerte hat). Für die geschätzten Werte können Vertrauensintervalle geliefert werden. Zudem können statistische Tests durchgeführt werden.

7.3.1 ML-Schätzung für L und V

Zu einer vorgegebenen Faktorenzahl $k < p$ sind die Parameter $\mathbf{L}$ und $\mathbf{V}$ eines k-Faktorenmodells für die Kovarianzmatrix $\boldsymbol{\Sigma} = \mathbf{kov}(\mathbf{Y})$, wobei

$$\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^t + \mathbf{V}$$

mit $\mathbf{L} \in \mathbb{R}^{p\times k}$, $\mathbf{V} = \mathbf{diag}\{v_1^2, ..., v_p^2\}$, $v_i^2 > 0$, unter der Zusatzforderung

$$\mathbf{L}^t\mathbf{V}^{-1}\mathbf{L} \text{ diagonal}$$

zu schätzen. Es liegen N unabhängige, normalverteilte Datensätze vor (Werte des Zufallsvektors $\mathbf{Y}$):

$$\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$\boldsymbol{\Sigma}$ wird geschätzt durch die empirische Kovarianzmatrix $\mathbf{S}$ (s. oben). $\mathbf{S}$ besitzt - da $\mathbf{Y}$ normalverteilt ist - eine Wishart-Verteilung:

$$(N-1)\,\mathbf{S} \sim W_p(\boldsymbol{\Sigma}, N-1)$$

mit der Dichte L

$$L(\boldsymbol{\Sigma}|\mathbf{S}) = \mathtt{const} \cdot |\mathbf{S}|^{\frac{N-p-2}{2}} \cdot |\boldsymbol{\Sigma}|^{-\frac{N-1}{2}} \cdot \exp\left(\left(-\frac{N-1}{2}\right) \cdot \mathtt{tr}\left(\mathbf{S}\boldsymbol{\Sigma}^{-1}\right)\right)$$

mit $|\mathbf{S}|$ der Determinante von $\mathbf{S}$ und $\mathtt{tr}\left(\mathbf{S}\boldsymbol{\Sigma}^{-1}\right)$ der Spur von $\mathbf{S}\boldsymbol{\Sigma}^{-1}$.
Die ML-Schätzer $\hat{\mathbf{L}}, \hat{\mathbf{V}}$ für $\mathbf{L}$ und $\mathbf{V}$ aufgrund der Statistik $\mathbf{S}$ erhält man durch Maximierung der Likelihoodfunktion $L(\boldsymbol{\Sigma}|\mathbf{S}) = L(\mathbf{L}, \mathbf{V}|\mathbf{S})$ (denn $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^t + \mathbf{V}$) unter der Nebenbedingung $\mathbf{L}^t\mathbf{V}^{-1}\mathbf{L}$ diagonal. Wie üblich wird L logarithmiert und man erhält

$$\ln L(\mathbf{L}, \mathbf{V}|\mathbf{S}) = \ln \mathtt{const} + \frac{N-p-2}{2}\ln|\mathbf{S}| - \frac{N-1}{2}\ln|\boldsymbol{\Sigma}| - \frac{N-1}{2} \cdot \mathtt{tr}\left(\mathbf{S}\boldsymbol{\Sigma}^{-1}\right)$$

Die Maximierung dieser Funktion unter der Nebenbedingung ist gleichwertig zur Minimierung von

$$f(\mathbf{L}, \mathbf{V}) = \ln|\boldsymbol{\Sigma}| + \mathtt{tr}\left(\mathbf{S}\boldsymbol{\Sigma}^{-1}\right) \tag{7.3.4}$$

$\mathbf{S}$ muss von vollem Rang sein, d.h. $\mathtt{rg}(\mathbf{S}) = p$. Dies ist vorgängig zu prüfen.

Bemerkung 7.3.1. *Gewöhnlich geht man bei ML-Schätzungen von den Daten aus und maximiert die Dichtefunktion unter Konstanthaltung der Daten (Dichtefunktion als Funktion der zu schätzenden Werte). In der Faktorenanalyse interessiert man sich jedoch in erster Linie für das Zustandekommen der Kovarianzen $\boldsymbol{\Sigma}$ (s. Grundgleichung) und nur in zweiter Linie für die Erklärung der Daten. Man betrachtet deshalb einen Schätzwert für $\boldsymbol{\Sigma}$ (nämlich $\mathbf{S}$) und maximiert die Likelihoodfunktion, so dass die Schätzer $\mathbf{L}$ und $\mathbf{V}$ für fixes $\mathbf{S}$ höchstwahrscheinlich werden. Trotzdem sind alle asymptotischen Resultate der gewöhnlichen ML-Theorie auf den vorliegenden Fall übertragbar.* ◊

7.3.2 Die impliziten Gleichungen für die ML-Schätzer für L und V

Durch Nullsetzen der partiellen Ableitungen von f nach $\mathbf{L}$ und $\mathbf{V}$ gelangt man zu impliziten Bestimmungsgleichungen für $\hat{\mathbf{L}}$ und $\hat{\mathbf{V}}$ (Herleitung s. Lawley/Maxwell 1971, S. 26)

1. $\hat{\mathbf{L}} = \mathbf{S}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\mathbf{L}}$ mit $\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{L}}\hat{\mathbf{L}}^t + \hat{\mathbf{V}}$

2. $\mathbf{diag}\left(\hat{\boldsymbol{\Sigma}}\right) = \mathbf{diag}(\mathbf{S})$ mit $\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{L}}\hat{\mathbf{L}}^t + \hat{\mathbf{V}}$

3. $\hat{\mathbf{L}}^t\hat{\mathbf{V}}\hat{\mathbf{L}}$ diagonal

7.3.3 Konsistenz und asymptotische Normalität

Die durch die eben angegebenen impliziten Gleichungen erklärten ML-Schätzer $\hat{\mathbf{L}}$ und $\hat{\mathbf{V}}$ für $\mathbf{L}$ und $\mathbf{V}$ sind bei eindeutiger Zerlegbarkeit von

$$\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^t + \mathbf{V}$$

mit

$$\mathbf{L} \in \mathbb{R}^{p\times k}, \mathbf{V} = \mathbf{diag}\{v_1^2, ..., v_p^2\} \in \mathbb{R}^{p\times p},$$
$$v_i^2 > 0,\ i = 1, ..., p, \mathtt{rg}\,(\mathbf{\Sigma}) = p$$

und einer stetigen Verteilung für $\mathbf{L}$ und $\mathbf{V}$ mit Wahrscheinlichkeit 1 konsistent und asymptotisch normalverteilt. Das heisst, es existiert mit Wahrscheinlichkeit 1 eine Umgebung U von $\mathbf{\Sigma}$ so, dass die Gleichungen für alle $\mathbf{S} \in U$ eindeutig nach $\hat{\mathbf{L}}$ und $\hat{\mathbf{V}}$ auflösbar sind und $\hat{\mathbf{L}}$ und $\hat{\mathbf{V}}$ konsistente und asymptotisch normalverteilte Schätzer für $\mathbf{L}$ und $\mathbf{V}$ bilden, sofern k die wirkliche Faktorenzahl der Population darstellt. Dieses Ergebnis gilt auch für nicht normalverteilte Daten, sofern die beiden ersten Momente existieren.

Praktische Ausführung der Minimierung

Die ML-Funktion $f(\mathbf{L}, \mathbf{V})$ (s. Gleichung 7.3.4) wird in zwei Stufen iterativ minimiert.

1. Bei festgehaltenem $\mathbf{V} > 0$ liefert die Matrix

$$\mathbf{L}_0(\mathbf{V}) = \mathbf{V}^{\frac{1}{2}}\mathbf{\Omega}\,(\mathbf{\Theta} - \mathbf{I})^{\frac{1}{2}} \tag{7.3.5}$$

mit $\mathbf{\Theta} = \mathbf{diag}\{\theta_1, ..., \theta_k\}$ (= Matrix der k grössten Eigenwerte von $\mathbf{V}^{-\frac{1}{2}}\mathbf{S}\mathbf{V}^{-\frac{1}{2}}$) und
$\mathbf{\Omega} = (\omega_1, ..., \omega_k)$ (= Matrix der zugehörigen orthonormierten Eigenvektoren)
das Minimum von $f(\mathbf{L}, \mathbf{V})$ in $\mathbf{L} \in \mathbf{R}^{p\times k}, \mathbf{L}^t\mathbf{V}^{-1}\mathbf{L}$ diagonal, nämlich :

$$\min f(\mathbf{L}, \mathbf{V}) = f_0(\mathbf{V}) = -\ln(\theta_{k+1} \cdot ... \cdot \theta_p) + (\theta_{k+1} + ... + \theta_p) + \ln|\mathbf{S}| + k,$$

ausgedrückt in den $p - k$ kleinsten Eigenwerten von $\mathbf{V}^{-\frac{1}{2}}\mathbf{S}\mathbf{V}^{-\frac{1}{2}}$. Die Darstellung (7.3.5) ist nur möglich, wenn keiner der k grössten Eigenwerte θ_i kleiner als 1 ist. Das ist bei grossem N selten der Fall. Für $N \to \infty$ strebt die entsprechende Wahrscheinlichkeit gegen 0 (Lawley, Maxwell, S. 28). Ebenso tritt der Fall mehrfacher empirischer Eigenwerte, so dass die Eigenvektoren nicht eindeutig sind, nur mit Wahrscheinlichkeit 0 auf ($\mathbf{S}$ besitzt eine stetige Verteilung).

2. $f_0(\mathbf{V})$ wird nach $\mathbf{V}$ minimiert. Dazu bildet man die 1. partiellen Ableitungen von f_0 nach v_i^2

$$\frac{\partial f_0}{\partial v_i^2} = \frac{1}{v_i^2}\left(\sum\nolimits_{r=1}^{k}(\theta_r - 1)\,\omega_{ir}^2 + 1 - \frac{s_{ii}}{v_i^2}\right)$$

und die zweite partiellen Ableitungen

$$\frac{\partial^2 f_0}{\partial v_i^2 \partial v_j^2} = \phi_{ij}^2 - 2\delta_{ij}v_i^{-2}\left(\frac{\partial f_0}{\partial v_i^2}\right) - \frac{2}{v_i^2 v_j^2}\sum_{r=1}^{k}\left((\theta_r\omega_{ir}\omega_{jr})\sum_{m=k+1}^{p}(\theta_m - 1)\,\frac{\omega_{im}\omega_{jm}}{\theta_r - \theta_m}\right)$$

mit $\mathbf{\Phi} = (\phi_{ij}) = \mathbf{V}^{-\frac{1}{2}}\left(\mathbf{I} - \mathbf{\Omega}\mathbf{\Omega}^t\right)\mathbf{V}^{-\frac{1}{2}}$. Lawley/Maxwell (1971, S. 38) verwenden für die 2. partiellen Ableitungen bei grossem N die Näherung

$$\frac{\partial^2 f_0}{\partial v_i^2 \partial v_j^2} \approx \phi_{ij}^2.$$

Dabei bedeutet ω_{ir} die r-te Komponenten von $\boldsymbol{\omega}_i$ mit $\mathbf{\Omega} = (\boldsymbol{\omega}_1, ..., \boldsymbol{\omega}_k)$, $\mathbf{\Theta} = \mathbf{diag}\{\theta_i, ..., \theta_k\}$, die Matrix der Eigenvektoren und der grössten Eigenwerte von $\mathbf{V}^{-\frac{1}{2}}\mathbf{S}\mathbf{V}^{-\frac{1}{2}}$.

7.3.4 Iterationen

Ausgehend von einem Startwert für $\mathbf{V}$, z.B. $v_i^2 = s_{ij}$ oder $v_i^2 = \left(1 - \frac{1}{2}\frac{k}{p}\right)\left(\frac{1}{s^{ii}}\right)$ mit dem i-ten Diagonalelement s^{ii} von $\mathbf{S}^{-1}$, wird $f_0(\mathbf{V})$ mit einem Quasi-Newton-Verfahren oder dem Newton-Verfahren nach $\mathbf{V}$ minimiert. Jedesmal werden $\mathbf{\Theta}$ und $\mathbf{\Omega}$ aus $\mathbf{V}^{-\frac{1}{2}}\mathbf{S}\mathbf{V}^{-\frac{1}{2}}$ berechnet sowie im Falle des Newton-Algorithmus die ersten und zweiten partiellen Ableitungen von $f_0(\mathbf{V})$. $\mathbf{L}_0(\mathbf{V})$ braucht nur am Schluss bestimmt zu werden, wenn die Iteration in $\hat{\mathbf{V}}$ ruht. $\hat{\mathbf{L}} = \mathbf{L}_0(\mathbf{V})$ und $\hat{\mathbf{V}}$ bilden die ML-Schätzer für $\mathbf{L}$ und $\mathbf{V}$. Wegen der zweiten impliziten Gleichung gilt

$$\mathbf{diag}\left(\hat{\mathbf{L}}\hat{\mathbf{L}}^t + \hat{\mathbf{V}}\right) = \mathbf{diag}(\mathbf{S})$$

so dass $\hat{\mathbf{V}}$ einfach aus $\mathbf{S}$ und $\hat{\mathbf{L}}$ gewonnen werden können:

$$\hat{\mathbf{V}} = \mathbf{diag}\left(\mathbf{S} - \hat{\mathbf{L}}\hat{\mathbf{L}}^t\right).$$

$f_o(\mathbf{V})$ kann im zulässigen Bereich $0 < v_i^2 < 1$ mehr als ein lokales Minimum besitzen: je nach Startpunkt des Näherungsverfahrens ergeben sich jeweils andere Minima. Man wählt das kleinste Minimum. Bei korrektem k, eindeutigem $\mathbf{V} > \mathbf{0}$ und bei $N \to \infty$ geht die Wahrscheinlichkeit für mehr als ein Minumum gegen 0.

7.3.5 Test des Modells, Bestimmung von k

Bei Erfüllung der Normalverteilungsannahme lässt sich testen, ob die Daten einem Modell mit k gemeinsamen Faktoren entsprechend. Nach einer ML-Schätzung der Modellparameter bietet sich der Likelihood-Quotienten-Test an. Wir testen, ob für die gewählte Faktorenzahl k Modell und Daten übereinstimmen.

$$H_0 : \mathbf{\Sigma} = \mathbf{L}\mathbf{L}^t + \mathbf{V} \text{ mit } \mathbf{L} \in \mathbb{R}^{p\times k}$$
$$H_1 : \mathbf{\Sigma} \in \mathbb{R}^{p\times p} \text{ beliebig positiv definit } (H_0 \subset H_1)$$

(ergibt sich ein Unterschied, ob $\mathbf{\Sigma}$ durch $\mathbf{L}\mathbf{L}^t + \mathbf{V}$ dargestellt wird oder nicht).
Unter H_0 wird die Likelihoodfunktion $L(\mathbf{\Sigma}|\mathbf{S})$ durch den ML-Schätzer $\hat{\mathbf{\Sigma}} = \hat{\mathbf{L}}\hat{\mathbf{L}}^t + \hat{\mathbf{V}}$ ($\hat{\mathbf{L}}, \hat{\mathbf{V}}$ ML-Schätzer für $\mathbf{L}$ und $\mathbf{V}$) maximiert mit

$$L_0 = \max_{\Sigma\in H_0} L(\mathbf{\Sigma}|\mathbf{S}) = \mathtt{const}|\mathbf{S}|^{\frac{N-p-2}{2}}|\hat{\mathbf{\Sigma}}|^{-\frac{N-1}{2}}\exp\left(-\frac{N-1}{2}\right)\mathtt{tr}\left(\mathbf{S}\hat{\mathbf{\Sigma}}^{-1}\right).$$

Daraus folgt für das Verhältnis $\frac{L_0}{L_1}$

$$\ln\left(\frac{L_0}{L_1}\right) = -\frac{N-1}{2}\left(\ln|\hat{\mathbf{\Sigma}}| - \ln|\mathbf{S}| + \mathtt{tr}\left(\mathbf{S}\hat{\mathbf{\Sigma}}^{-1}\right) - p\right),$$

und H_0 ist abzulehnen, wenn das Verhältnis zu klein ist. Als Prüfstatistik berechnet man das -2-fache des Loglikelihoodverhältnisses

$$\begin{aligned} U_k &= (N-1)\left(\ln|\hat{\mathbf{\Sigma}}| - \ln|\mathbf{S}| + \mathtt{tr}\left(\mathbf{S}\hat{\mathbf{\Sigma}}^{-1}\right) - p\right) \\ &= (N-1)\left(\ln|\hat{\mathbf{\Sigma}}| - \ln|\mathbf{S}|\right) \end{aligned}$$

(da $\mathtt{tr}\left(\mathbf{S}\hat{\mathbf{\Sigma}}^{-1}\right) = p$ für den ML-Schätzer $\hat{\mathbf{\Sigma}}$ des Faktorenmodells). Die exakte Verteilung von U_k ist unbekannt. Für grosse Stichproben gilt

$$U_k \approx \chi^2(d_k)$$

mit

$$d_k = \left(\frac{p\,(p+1)}{2} - \left(pk + p - \frac{k\,(k-1)}{2}\right)\right) = \frac{1}{2}\left((p-k)^2 - (p+k)\right)$$

Freiheitsgraden. d_k misst die Zahl der Freiheitsgrade in der Gleichung

$$\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^t + \mathbf{V}$$

(Anzahl $\frac{p(p+1)}{2}$ von Elementgleichungen minus Anzahl $pk + p$ der unbekannten Elemente von $\mathbf{L}$ und $\mathbf{V}$, vermindert um die Anzahl $\frac{k(k-1)}{2}$ der Restriktionen von $\mathbf{L}^t\mathbf{V}^{-1}\mathbf{L} = \mathbf{diag}$. Gemäss Bartlett (1954) kann die Approximation an die χ^2-Verteilung verbessert werden, wenn im Ausdruck für U_k N durch

$$N' = N - \frac{2p+5}{6} - \frac{2}{3}k$$

ersetzt wird. Zu einem gegebenen Signifikanzniveau α wird die Nullhypothese verworfen, wenn

$$U_k > \chi^2\,(d_k; 1-\alpha)\,.$$

Aus der Ablehnung der Nullhypothese kann man schliessen, dass wenigstens $k+1$ Faktoren erforderlich sind.
Zwei Interpretationen von U_k sind möglich (Lawley):

1. U_k misst den Abstand der $p-k$ kleinsten Eigenwerte der Matrix $\hat{\mathbf{V}}^{-\frac{1}{2}}\mathbf{S}\hat{\mathbf{V}}^{-\frac{1}{2}}$ von 1:

$$U_k \approx \frac{N-1}{2}\left(\left(\hat{\theta}_{k+1} - 1\right)^2 + ... + \left(\hat{\theta}_p - 1\right)^2\right)$$

2. U_k misst die Residuenkovarianz der Variablen $\mathbf{y}$ nach Extraktion von k gemeinsamen Faktoren,

$$U_k \approx (N-1)\sum\nolimits_{i<j}\left(\frac{(s_{ij} - \sigma_{ij})^2}{v_i^2 v_j^2}\right)$$

Für $k = 0$ erhält man den Sphärentest von Barlett, welcher die Hypothese

$$H_0 : \mathbf{\Sigma} = \mathbf{V} = \mathbf{diag}, \text{ d.h. } y_1, ..., y_p \text{ unabhängig},$$

prüft und dadurch festzustellen erlaubt, ob überhaupt signifikante Abhängigkeiten zwischen den Variablen bestehen, so dass eine Faktorenanalyse angebracht ist.

$$U_0 = N - 1 - \frac{2p+5}{6}\left(\ln|\hat{\mathbf{\Sigma}}| - \ln|\mathbf{S}|\right)$$

oder

$$U_0 = -\left(N - 1 - \frac{2p+5}{6}\right)\ln|\mathbf{R}| \tag{7.3.6}$$

ist asymptotisch χ^2-verteilt mit $d_0 = \frac{p(p-1)}{2}$ Freiheitsgraden ($\mathbf{R}$ ist die Korrelationsmatrix). H_0 wird zum Niveau α abgelehnt und eine Faktorenanalyse erscheint nur sinnvoll, wenn

$$U_0 > \chi^2\,(d_0, 1-\alpha)\,.$$

Die Gleichheit (7.3.6) folgt bei $k = 0$ aus

$$\mathbf{R} = \hat{\mathbf{V}}^{-\frac{1}{2}}\mathbf{S}\hat{\mathbf{V}}^{-\frac{1}{2}},$$

also

$$\ln|\mathbf{R}| = \frac{\ln|\mathbf{S}|}{\ln|\hat{\mathbf{V}}|} = -\left(\ln|\hat{\mathbf{V}}| - \ln|\mathbf{S}|\right) \text{ mit } \hat{\mathbf{\Sigma}} = \hat{\mathbf{V}}.$$

Bestimmung der Faktorenzahl: Ist k bekannt oder vorgegeben, so kann man mit der Statistik U_k die Hypothese testen: $H(k)$: „k Faktoren vermögen die Daten zu erklären“. Will man aber k erst bestimmen, dann ist die folgende Vorgehensweise möglich: Beginnend mit einem kleinen k_0 testet man, ob $k = k_0$ angenommen werden kann ($k_0 = 0$ bedeutet, dass die Hypothese „Variablen unabhängig, Faktorenanalyse nicht sinnvoll“ mitgetestet wird). Wird $H(k_0)$ abgelehnt, er erhöht man schrittweise k um 1, und testet $H(k_0 + 1j)$, $j \in \mathbb{N}$, bis $H(\hat{k})$als erste dieser Hypothesen nicht abgelehnt wird. $\hat{k}$ bildet einen Schätzwert für die wirkliche Faktorenzahl k. Jedesmal muss die gesamte ML-Schätzung der Parameter $\mathbf{L}$ und $\mathbf{V}$ ausgeführt werden.

Grosse Werte von k können dazu führen, dass die Freiheitsgrade negativ werden. Dann ist das Modell nicht mehr wohldefiniert. Ausserdem ist U_k stark vom Stichprobenumfang abhängig. Erhält man keinen akzeptablen Wert für U_k mit positivem Freiheitsgrad, ist das ein Zeichen dafür, dass die Daten nicht durch das Modell der Faktorenanalyse beschrieben werden können. Gründe dafür können starke Abweichung der Daten von der Normalverteilung, korrelierte Faktoren oder nichtlineare Beziehungen zwischen Faktoren und beobachteten Variablen sein.

7.3.6 Verteilung der ML-Schätzer: Vertrauensintervalle

Die ML-Verfahren bieten die Möglichkeit, Aussagen über Vertrauensintervalle der ML-Schätzer $\hat{\mathbf{V}}$ und $\hat{\mathbf{L}}$ zu machen. Asymptotisch gelten die asymptotischen Aussagen auch für nichtnormalverteilte Daten. Bei festen k und eindeutig definierten Parametern $\mathbf{L}$ und $\mathbf{V}$ gilt:

1. $\hat{v}_1^2, ..., \hat{v}_p^2$ besitzen asymptotisch eine multivariate Normalverteilung mit $E\left(\hat{v}_i^2\right) = v_i^2$ und der Kovarianzmatrix

$$kov\begin{pmatrix} \hat{v}_1^2 \\ \vdots \\ \hat{v}_p^2 \end{pmatrix} = \frac{1}{N-1}\mathbf{Q}$$

   mit

$$\begin{aligned} \mathbf{Q} &= \mathbf{G}^{-1} \\ \mathbf{G} &= \left(\phi_{ij}^2\right)_{ij}^p \\ \mathbf{\Phi} &= \left(\phi_{ij}\right)_{ij}^p = \mathbf{V}^{-\frac{1}{2}}\left(\mathbf{I} - \mathbf{\Omega}\mathbf{\Omega}^t\right)\mathbf{V}^{-\frac{1}{2}} \end{aligned}$$

   (Lawley, Maxwell). Damit lassen sich Vertrauensintervalle für v_i^2 berechnen, wenn $\mathbf{\Phi}$ mit Hilfe der ML-Schätzwerte für $\mathbf{V}$ und $\mathbf{\Omega}$ ausgedrückt wird:

$$\hat{\mathbf{\Phi}} = \left(\hat{\phi}_{ij}\right)_{ij}^p = \hat{\mathbf{V}}^{-\frac{1}{2}}\left(\mathbf{I} - \hat{\mathbf{\Omega}}\hat{\mathbf{\Omega}}^t\right)\hat{\mathbf{V}}^{-\frac{1}{2}}$$

   mit $\hat{\mathbf{\Omega}}$, der Matrix der Eigenvektoren von $\hat{\mathbf{V}}^{-\frac{1}{2}}\mathbf{S}\hat{\mathbf{V}}^{-\frac{1}{2}}$

2. Bei bekanntem $\mathbf{V}$ gilt für den Schätzer $\hat{\mathbf{L}}_0 = (l_{ij0}) = L_0(\mathbf{V})$, dass die Elemente der Differenz

$$\mathbf{Z} = \hat{\mathbf{L}}_0 - \mathbf{L}$$

   gemeinsam asymptotisch multivariat normalverteilt sind mit Erwartungswert 0 und Kovari-

anzen

$$
\begin{aligned}
E\left(z_{ir}z_{js}\right) &= \frac{1}{N-1}a_{ir,js} \\
a_{ir,js} &= -\frac{\theta_r\theta_s}{\left(\theta_r-\theta_s\right)^2}l_{is}l_{jr} \text{ für } r \neq s \text{ und} \\
a_{ir,jr} &= v_r\left(\sigma_{ij} - \frac{1}{2}v_r l_{ir}l_{jr} + \sum_{\substack{m=1\\m\neq r}}^{k}\left(v_m\gamma_{rm}l_{im}l_{jm}\right)\right) \\
v_r &= \frac{\theta_r}{\theta_r-1} \\
\gamma_{rm} &= \left(\frac{\theta_r-1}{\theta_r-\theta_m}\right)^2 - 1
\end{aligned}
$$

(für die Definitionen von θ_r und σ_{ij} s. o.)

3. Um die totale Kovarianz des ML-Schätzers $\hat{\mathbf{L}} = \mathbf{L}_0\left(\hat{\mathbf{V}}\right)$ zu erhalten, muss die Komponente addiert werden, da sie von der linearen Regression der Elemente von $\hat{\mathbf{L}}$ auf die des Schätzers $\hat{\mathbf{V}}$ herrührt (Lawley).

$$\hat{l}_{ir} - l_{ir} = \left(\hat{l}_{ir} - \hat{l}_{ir0}\right) + \left(\hat{l}_{ir0} - l_{ir}\right) = z_{ir} + \sum_j b_{j,ir}\left(\hat{v}_j^2 - v_j^2\right).$$

Für die Koeffizienten $b_{j,ir}$ rechnet man

$$b_{j,ir} = -\left(\frac{v_i^2}{\theta_r-1}\right)^{-\frac{1}{2}}\omega_{jr}v_j^{-2}\left(\delta_{ij} - \frac{1}{2}\omega_{ir}\omega_{jr} + \sum_{\substack{m=1\\m\neq r}}\omega_{im}\omega_{jm}\frac{(\theta_m-1)}{\theta_r-\theta_m}\right)$$

Gemeinsam resultiert (asymptotisch) für $\hat{\mathbf{L}}$ die pk-reihige Kovarianzmatrix

$$kov\begin{pmatrix}\hat{l}_{11}\\ \vdots \\ \hat{l}_{p1}\\ \vdots \\ \hat{l}_{1k}\\ \vdots \\ \hat{l}_{pk}\end{pmatrix} = \frac{1}{N-1}\left(\mathbf{A} + 2\mathbf{B}^t\mathbf{Q}\mathbf{B}\right),$$

wenn die Elemente $a_{ir,js}$ und $b_{j,ir}$ in folgender Weise in Matrizen $\mathbf{A}$ und $\mathbf{B}$ zusammengefasst werden:

$a_{ir,js}$ steht in der Zeile $p\,(r-1)+i$ und Spalte $p\,(s-1)+j$ von $\mathbf{A} \in \mathbb{R}^{pk\times pk}$

$b_{i,js}$ steht in der Ziele i und der Spalte $p(s-1)+j$ von $\mathbf{B} \in \mathbb{R}^{p\times pk}$

$\hat{\mathbf{L}}$ ist mit dieser Kovarianzmatrix und Erwartungswert $\mathbf{L}$ asymptotisch multivariat normalverteilt, so das Vertrauensintervall für die Ladungswerte bei hinreichend grossem Stichprobenumfang N angegeben werden können.

Bemerkung 7.3.2. *Gelangt die Iteration in* $\mathbf{V}$ *bei der Berechnung der ML-Schätzer* $\hat{\mathbf{V}}$ *und* $\hat{\mathbf{L}}$ *zu einem* $\hat{v}_i^2 < \varepsilon$ *nahe 0, dann deutet diese darauf hin, dass für die wirkliche Einzelrestvarianz*

der i-ten Variable y_i gilt: $v_i^2 = 0$. Um weiterrechen zu können, muss $\hat{v}_i^2 = \varepsilon$ festgehalten werden. Man spricht von einer „unsauberen Lösung“ oder „Heywood-case“. Ist vor einem Start des ML-Verfahren das Verschwinden von Einzelrestvarianzen zu vermuten, kann man die Variablen y_i mit $v_i^2 = 0$ eliminieren (für Verfahren s. Fahrmeir et al. (1996, S. 656)). ◊

7.3.7 Skaleninvarianz der ML-Faktorenanalyse

Das faktorenanalytische Modell verhält sich gegenüber Transformationen

$$\tilde{\mathbf{y}} = \mathbf{D}y$$

mit $\mathbf{D} = \mathbf{diag}\,(d_1, ..., d_p)$, $d_i > 0$, insofern unabhängig, als in gleichem Masse die Ladungsmatrix transformiert wird, ebenso die Einzelrestvarianzen. Die Faktoren bleiben unverändert:

$$\tilde{\mathbf{y}} - \tilde{\boldsymbol{\mu}} = \mathbf{D}\,(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{DLF} + \mathbf{De} = \tilde{\mathbf{L}}\mathbf{f} + \tilde{\mathbf{e}} \text{ mit } \tilde{\mathbf{L}} = \mathbf{DL}, \tilde{\mathbf{e}} = \mathbf{De}$$

$$\tilde{\boldsymbol{\Sigma}} = \mathbf{D\Sigma D} = \mathbf{DLL}^t\mathbf{D} + \mathbf{DVD} = \tilde{\mathbf{L}}\tilde{\mathbf{L}}^t + \tilde{\mathbf{V}} \text{ mit } \tilde{\mathbf{V}} = \mathbf{DVD}.$$

Zudem gilt

$$f\,(\mathbf{L}, \mathbf{V}) = \ln|\boldsymbol{\Sigma}| + \mathtt{tr}\left(\mathbf{S}\boldsymbol{\Sigma}^{-1}\right) = \min \text{ genau dann, wenn } f\left(\tilde{\mathbf{L}}, \tilde{\mathbf{V}}\right) = \ln|\tilde{\boldsymbol{\Sigma}}| + \mathtt{tr}\left(\tilde{\mathbf{S}}\tilde{\boldsymbol{\Sigma}}^{-1}\right) = \min.$$

Wegen dieser Massstabsunabhängigkeit des faktorenanalytischen Modells liefert also eine Analyse der Kovarianzmatrix $\tilde{\mathbf{S}} = \mathbf{DSD}$ der umskalierten empirischen Daten im wesentlichen die gleichen Schätzergebnisse wie eine Analyse von $\mathbf{S}$. Die Tests zur Wahl von k bleiben unverändert. Entsprechend werden die Daten üblicherweise standardisiert.

Bemerkung 7.3.3. *Die ML-Faktoranalyse wird in R unter factanal(). Angeboten. Für die obigen Auto-Daten erhält man mit:*

```
factanal(auto.frame,2)
```

(Die Zahl der zu extrahierenden Faktoren muss angegeben werden, im Beispiel 2). Unter Uniquenesses stehen die Einzelrestvarianzen. Hohe Einzelrestvarianzen geben an, dass die Varianz der entsprechenden Variable durch die Faktoren schlecht erklärt wird.

Uniquenesses:						
X1	*X2*	*X3*	*X4*	*X5*	*X6*	*X7*
0.026	*0.288*	*0.110*	*0.191*	*0.094*	*0.074*	*0.072*

Unter Loadings stehen die mit varimax rotierten Ladungen pro Faktor (die ML-Analyse führt also für die Daten zum selben Schluss wie die Haupt-Faktorenanalyse). Unrotierte Ladungen erhält man mit „rotation = "none"“.

	Loadings:	
	Factor1	*Factor2*
X1	*-0.252*	*0.954*
X2		*0.843*
X3		*0.938*
X4	*0.899*	
X5	*0.909*	*-0.281*
X6	*0.871*	*0.408*
X7	*0.942*	*-0.205*

Die nächste Tabelle gibt die Summe der quadrierten Ladungen pro Faktor (SS loadings; Kaiser's Regel: sollte grösser als 1 sein) - erhält man auch mit

```
sum(factanal(auto.frame,2)$loadings[,1]^2)
```

und den Anteil der erklärten Varianz pro Faktor an - Proportion Var, erhält man mit

```
sum(factanal(auto.frame,2)$loadings[,1]^2)/7
```

Die letzte Zeile gibt die Varianzanteile kumuliert an ($0.479 + 0.398 = 0.877$).

	Factor1	*Factor2*
SS loadings	*3.355*	*2.789*
Proportion Var	*0.479*	*0.398*
Cumulative Var	*0.479*	*0.878*

Der gelieferte Test geht von der Nullhypothese aus, dass die Anzahl der Faktoren die Dimensionen der Daten erfassen. Die Nullhypothse wird verworfen, wenn der p-Wert kleiner als 0.05 ist. Der hohe p-Wert im Beisiel zeigt also an, dass die Nullhypothese nicht verworfen werden muss. Das Modell ist also angemessen.

```
Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 4.96 on 8 degrees of freedom.
The p-value is 0.762
```

Die graphische Darstellung der rotierten und unrotierten Ladungen (s. Abbildungen 7.3.1 und 7.3.2): Befehle für die unrotierte Darstellung:

```
ladeHK1=factanal(auto,2,rotation="none")$loadings[,1]
ladeHK2=factanal(auto,2,rotation="none")$loadings[,2]
plot(ladeHK1,ladeHK2,type="n",xlim=c(-1,1),ylim=c(-1,1),lwd=0.5,
xlab="Korrelationen 1. Hauptkomponente",
ylab="Korrelationen 2. Hauptkomponente", main="Ladungen unrotiert")
text(ladeHK1+c(0.01,-0.01,0.04,0.05,0.1,0.03,-0.1),
ladeHK2+c(0.05,0.05,0.05,0.05,0.1,0.05,-0.1),labels=c("X1","X2","X3","X4","X5","X6","X7"),
cex=0.5)
null=rep(0,7);abline(h=0);abline(v=0)
arrows(null,null,ladeHK1,ladeHK2)
```

Befehle für die rotierte Darstellung:

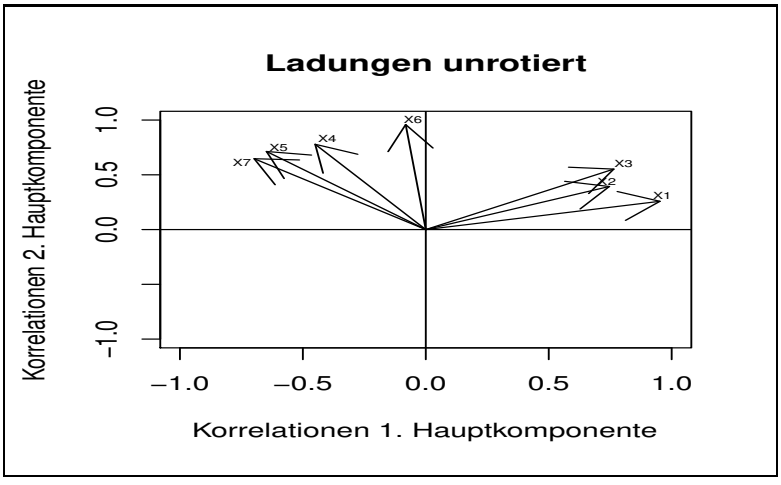


Abbildung 7.3.1: Ladungen unrotiert

```
ladeHK1=factanal(auto,2)$loadings[,1]
ladeHK2=factanal(auto,2)$loadings[,2]
```

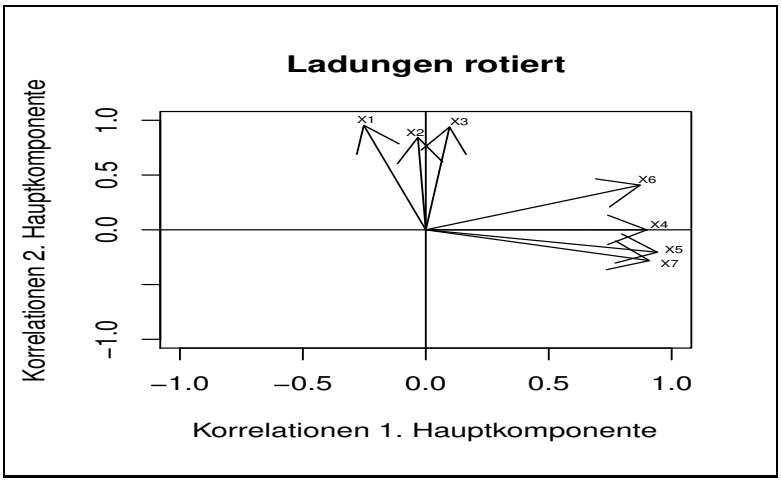


Abbildung 7.3.2: Ladungen unrotiert

Der Rest wie oben - bis auf den Titel der Graphik: ◊

Bemerkung 7.3.4. *Interessant ist die Bemerkung beim factanal-Befehle: Note: There are so many variations on factor analysis that it is hard to compare output from different programs. Further, the optimization in maximum likelihood factor analysis is hard, and many other examples we compared had less good fits than produced by this function. In particular, solutions which are 'Heywood cases' (with one or more uniquenesses essentially zero) are much more common than most texts and some other programs would lead one to believe.* ◊

Bemerkung 7.3.5. *https://www.geo.fu-berlin.de/en/v/soga/Geodata-analysis/factor-analysis/A-simple-example-of-FA/index.html* ◊

Bemerkung 7.3.6. *Die ML-Faktorenanalyse (und andere Methoden) werden in mehreren R-Paketen angeboten, z.B. lavaan (http://lavaan.ugent.be/tutorial/tutorial.pdf), psych (http://personality-project.org/r/psych/HowTo/factor.pdf)* ◊

Bemerkung 7.3.7. *Mit SPSS: Analysieren, Dimensionsreduktion, Faktorenanalyse, Extraktion, Maximum Likelihood.*
Die SS Loadings heissen hier „Summen von quadrierten Faktorladungen für Extraktion" und stimmen nicht genau mit den R-Ergebnissen überein (SPSS kleiner, Grössenordnungen ähnlich, für Auto.frame-Beispiel). Die Varianzanteile (unter „Rotierte Summe der quadrierten Ladungen") sind identisch. Der Test gibt dieselben Werte an. ◊

7.4 Faktorenanalyse mit kategorialen Variablen

Oft werden dichotome und ordinale Variablen in der Faktorenanalyse wie metrische Variablen behandelt. Nominale Variablen werden mit Hilfe von Dummy-Variablen dichotomisiert. Die Notwendigkeit metrischer Skalierung für die Faktorenanalyse wird unter Umständen erwähnt - in den Beispielen kümmert man sich dann aber nicht um die Anforderung. Oft wird davon ausgegangen, dass kategorialen Variablen eine darunterliegende metrische Skala zu Grunde liegt. So würde man z.B. postulieren, dass Zufriedenheit (mit etwas) kontinuierlich ist. Diese Idee rührt vermutlich daher, dass man eine metrische Skala wie die Längemessung dichotomisieren kann, indem man einen

Trennpunkt bestimmt und Werte darunter mit 0 kodiert und Werte darüber mit 1. Ebenso kann man aus einer metrischen Skala eine ordinale erzeugen, indem man mehrere Trennpunkt fixiert und damit Klassen definiert. Den Objekten der Klassen werden dann in aufsteigender Ordnung z.B. die Zahlen 1, 2, .., m zuordnet, wobei die Ordnung durch die Ordnung der Trennpunkte gegeben ist. Daraus zu schliessen, dass dichotomen oder metrischen Daten eine selber nicht messbare Eigenschaft zugrunde liegt, die eigentlich metrisch messbar ist, macht aber wenig Sinn: man betrachtet etwas als metrisch messbar, das man nicht metrisch messen kann. Es handelt sich also um ein nicht belegbares Postulat.

Will man nur die Korrelationsstruktur der Variablen analysieren (das Ziel der Variablenreduktion wird geopfert, da man keine Faktoren berechnen kann), so kann man eine Korrelationsmatrix erstellen, die je nach Daten unterschiedliche Korrelationen verwendet. So kann man z.B. bei dichotomen Variablen die übliche Pearson-Korrelation verwenden. Dabei ergibt sich das Problem, dass die Pearsonkorrelation bei ungleichen Randverteilungen nur mehr in einem echten Teilintervall von [-1,1] variiert (s. z.B. (Hartung, 1999, S. 446 ff.)). Dadurch werden Korrelationen eventuell unterschätzt und die Anzahl der relevanten latenten Variablen wird durch die Faktorenanalyse überschätzt. Ist die dichotomisierte Variable metrisch messbar und normalverteilt, kann man die tetrachorische Korrelation verwenden. Ist die ordinale Variable Resultat einer ordinalisierten metrischen Variable, kann man die polychorische Korrelaton verwenden. Es wird aber kaum Sinn machen, metrisch skalierte Variablen in informationsärmere Variablen zu verwandeln. Postulierte man einfach die metrische Skaliertheit einer nicht-metrisch- messbaren Variable und setzt zudem noch voraus, dass sie normalverteilt ist, wird man auf das bereits erwähnte Problem nicht weiter belegbarer Postulate geführt.

Bei ordinalen Daten besteht für die Korrelationsanalyse eine Möglichkeit darin, als Korrelationsmass etwa Goodmans- und Kruskals Gamma verwenden.

Schliesslich kann man bei dichotomen oder ordinalen Daten IRT-Modelle verwenden. Vorgängig werden die Daten mit Hilfe geeigneter Korrelationsmasse - ordinale Daten mit Gamma - gruppiert, um dann mit Hilfe von IRT-Methoden eine darunterliegende postulierte Variable berechnen. Die dabei berechneten, latenten Variablen werden allerdings nicht orthogonal sein.

Zuletzt kann man bezüglich Faktoranalyse mit ordinalen Daten die folgende Literatur konsultieren (Jöreskog & Moustaki, 2001) (Bartholomew, Steele, Galbraith & Moustaki, 2008)

Literatur

Bartholomew, D. J., Steele, F., Galbraith, J. & Moustaki, I. (2008). *Analysis of Multivariate Social Science Data* (2. Aufl.). London: Chapman and Hall/CRC.

Dümbgen, L. (2005). *Lineare Algebra 1 - 2*. Bern: Universität Bern.

Fahrmeir, L., Hamerle, A. & Tutz, G. (1996). *Multivariate statistische Verfahren* (2. Aufl.). Berlin: Walter de Gruyter.

Fischer, G. (2002). *Lineare Algebra* (13. Aufl.). Braunschweig/Wiesbaden: vieweg.

Golub, G. H. & Kahan, W. (1956). Calculating the Singular Values and Pseudo-Inverse of a Matrix. *SIAM J. Numer. Anal.*, *2* (2), 205-224.

Greenacre, M. (2010). *Biplots in Practice*. Bilbao: Fundación BBVA. Verfügbar unter `https://www.fbbva.es/wp-content/uploads/2017/05/dat/DE_2010_biplots_in_practice.pdf`

Hartung, J. (1999). *Statistik: Lehr- und Handbuch der angewandten Statistik*. München: Oldenbourg.

Hendrickson, A. E. & White, P. O. (1964). Promax: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology*, *17* (1), 65-70.

Jöreskog, K. G. & Moustaki, I. (2001). Factor Analysis of Ordinal Variables: A Comparison of Three Approaches. *Multivariate Behavioral Research*, *36* (3), 347-387.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*, 187–200.

Mortensen, U. (2013). *Einführung in die Faktorenanalyse mit einer Einführung in die Matrixrechnung.* Internet. Verfügbar unter `http://www.uwe-mortensen.de/fakanalysews0506b.pdf`
Pumplün, D. (1986). *Lineare Algebra I und II.* Hagen: FernUniversität Hagen.
Rinne, H. (2000). *Statistische Analyse multivariater Daten.* München: Oldenbourg Wissensch.Vlg.
Unger, L. (2009). *Lineare Algebra.* Hagen: FernUniversität Hagen.