Mathematik III - Lineare Algebra für Ökonomen

Prof. Dr. D. Klatte

Vorlesung im Frühjahrssemester 2008

Studierenden der Wirtschaftswissenschaftlichen Fakultät, die in ihrem Bachelor- und Masterstudium mit quantitativen Methoden ihres Studienfachs zu tun haben werden, sei dringend empfohlen, sich im Verlaufe ihres weiteren Studiums mathematische Kenntnisse und Fähigkeiten anzueignen, die deutlich über den obligatorischen Grundkurs *Mathematik I und II* hinausgehen. *Eine* Möglichkeit dazu bietet die Lehrveranstaltung *Mathematik III - Lineare Algebra für Ökonomen* (zukünftig kurz *Lineare Algebra für Ökonomen* genannt), die in engem Zusammenhang mit der Lehrveranstaltung *Mathematik III - Analysis für Ökonomen* im Sommersemester steht.

Natürlich werden gegenüber der Mathematik II, in der eine Einführung in die lineare Algebra gegeben wurde, auch zahlreiche neue Inhalte vermittelt, aber wir greifen die dort behandelten Themen und Methoden auf, vertiefen sie, vor allem *begründen* wir sie und bauen sie aus. Das erfordert einen höheren Grad an *formalem* Herangehen im Vergleich zum Grundkurs.

Wichtige Inhalte der Lehrveranstaltung *Lineare Algebra für Ökonomen* sind u.a. Vektorräume (ohne und mit Skalarprodukt), lineare Abbildungen, Matrizen, Eigenwerte und Anwendungen dieser Kapitel. Als grundlegendes Lehrbuch wird *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984* empfohlen, weitere Literaturhinweise folgen weiter unten.

Sehr wichtig ist die aktive Mitarbeit in den Übungen, vor allem auch das selbstständige Lösungen der wöchentlichen Übungsaufgaben für zu Hause.

Literaturhinweise

P. Kall, *Lineare Algebra für Ökonomen*, Teubner Studienbücher Mathematik, B.G. Teubner, Stuttgart, 1984.

G. Fischer, *Lineare Algebra* - eine Einführung für Studienanfänger, vieweg studium - Grundkurs Mathematik, Vieweg, Braunschweig-Wiesbaden, 2002.

G. Strang, *Lineare Algebra*, Springer-Lehrbuch, Springer, Berlin-Heidelberg-etc., 2003.

Sydsaeter, Hammond, *Mathematik für Wirtschaftswissenschaftler*, Pearson Studium, 2004.

K. Burg, H. Haf, F. Wille, *Höhere Mathematik für Ingenieure, Band II: Lineare Algebra*, Teubner, Stuttgart-Leipzig-Wiesbaden, 2002.

A. Fischer, W. Schirotzek, K. Vetters, *Lineare Algebra - Eine Einführung für Ingenieure und naturwissenschaftler*, Teubner, Stuttgart-Leipzig-Wiesbaden, 2003.

B. Luderer, U. Würker, *Einstieg in die Wirtschaftsmathematik*, Teubner, 2001.

F. Riedel, P. Wichardt, *Mathematik für Ökonomen*, Springer 2007.

H. Rommelfanger, *Mathematik für Wirtschaftswissenschaftler II*, Spektrum, Heidelberg-Berlin, 2002.

Inhaltsverzeichnis

1 Vektorräume **5**
1.1 Gruppen . . . 5
1.1.1 Motivation . . . 5
1.1.2 Definition einer Gruppe . . . 7
1.1.3 Eigenschaften von Gruppen . . . 9
1.2 Definition eines Vektorraums und Beispiele . . . 10
1.3 Erzeugendensystem, lineare Unabhängigkeit und Basis . . . 15
1.4 Endlichdimensionale Vektorräume . . . 20
1.5 Unterräume . . . 22

2 Lineare Abbildungen und Matrizen **27**
2.1 Matrizen . . . 27
2.2 Lineare Gleichungssysteme und Rang . . . 35
2.3 Lineare Abbildungen . . . 41
2.4 Koordinatentransformation . . . 53
2.5 Transformationsmatrizen der Gauss-Elimination . . . 57

3 Vektorräume mit Skalarprodukt **61**
3.1 Skalarprodukt und Norm . . . 61
3.2 Orthonormalsysteme . . . 67
3.3 Lineare Approximationsprobleme . . . 69

4 Eigenwerte **75**
4.1 Eigenwerte und Diagonalisierung von Matrizen . . . 75
4.2 Eigenwerte symmetrischer Matrizen . . . 83
4.3 Anhang: Repetition zu Determinanten . . . 87

5 Spezielle Themen und Anwendungen **95**
5.1 Definitheit von Matrizen . . . 95
5.2 Konvexe Funktionen . . . 101
5.3 Geometrische Reihe und Verbrauchsmatrizen . . . 103
5.4 Markov-Matrizen . . . 107
5.5 Matrizenrechnung und Regressionsanalyse . . . 111
5.5.1 Das einfache lineare Modell . . . 111
5.5.2 Multiple lineare Regressionsanalyse . . . 114
5.6 Hinreichende Optimalitätsbedingung 2. Ordnung bei einer Gleichungsrestriktion . . . 115

Kapitel 1

Vektorräume

1.1 Gruppen

1.1.1 Motivation

Was haben die folgenden Beispiele gemeinsam?

1.1.1 Beispiel. Zwei reellen Zahlen x und y sind

$$\begin{array}{ll} \text{ihre Summe } x+y & \text{(wieder eine reelle Zahl),} \\ \text{ihr Produkt } x \cdot y & \text{(wieder eine reelle Zahl),} \end{array}$$

zugeordnet, und es gelten für $x, y, z \in \mathbb{R}$ bekanntlich Gesetze wie

$$\begin{array}{l} (x+y)+z = x+(y+z),\ (x \cdot y) \cdot z = x \cdot (y \cdot z),\ x+y = y+x \\ x+0 = 0+x = x,\ x \cdot 1 = 1 \cdot x = x,\ x \neq 0 \Rightarrow x \cdot x^{-1} = 1 \end{array}$$

und andere Eigenschaften. ◇

1.1.2 Beispiel. Es seien $(i_1, \ldots, i_m)$ ein Vektor von (reellen) Produktionsinputs und $(a_1, \ldots, a_n)$ ein Vektor von (reellen) Produktionsoutputs - etwa in Periode 1 -, zusammengefasst zu einem Vektor von Produktionsaktivitäten

$$\underline{p} = (i_1, \ldots, i_m, a_1, \ldots, a_n).$$

Die Menge der Vektoren von Produktionsaktivitäten dieser Form sei mit P bezeichnet, d.h., es gilt $\underline{p} \in P$. In Periode 2 sei analog ein Vektor von Produktionsaktivitäten

$$\underline{p}' = (i_1', \ldots, i_m', a_1', \ldots, a_n')$$

gegeben, d.h., es ist wieder $\underline{p}' \in P$. Offenbar liegt nun die

$$\text{Summe } \underline{p} + \underline{p}' = (i_1 + i_1', \ldots, i_m + i_m', a_1 + a_1', \ldots, a_m + a_m')$$

wieder in der Menge P der Produktionsaktivitäten, und es gelten für $\underline{p}, \underline{p}', \underline{p}'' \in P$

$$(\underline{p} + \underline{p}') + \underline{p}'' = \underline{p} + (\underline{p}' + \underline{p}''), \quad \underline{p} + \underline{p}' = \underline{p}' + \underline{p}, \quad \underline{p} + \underline{o} = \underline{p}$$

und andere Gesetze, wobei $\underline{o} = (0, \ldots, 0, 0, \ldots, 0)$ ist. $\diamond$

1.1.3 Beispiel. Sei $\mathcal{M}$ die Menge der regulären quadratischen Matrizen der Ordnung n. Zwei Matrizen $A \in \mathcal{M}$ und $B \in \mathcal{M}$ ordne man

$$\text{das Matrizenprodukt} \quad AB$$

in der üblichen Weise zu. Offenbar haben wir wieder

$$AB \in \mathcal{M}.$$

Eine besondere Rolle spielt bekanntlich die Einheitmatrix I, d.h., die Diagonalmatrix der Ordnung n, deren Hauptdiagonalelemente stets gleich der Zahl 1 sind. Welche der in den vorigen Beispielen genannten Gesetze gelten hier ebenfalls?

Zusatzfrage. In *Mathematik II* hatten wir eine inverse Matrix zu $A \in \mathcal{M}$ so definiert: Eine Matrix X heisst die zu $A \in \mathcal{M}$ inverse Matrix, falls

$$XA = I.$$

Wir hatten in der *Mathematik II* ohne Beweis behauptet, dass die Inverse eindeutig bestimmt ist und dass gilt

$$AX = I. \tag{1.1}$$

Warum gilt das? $\diamond$

Allen drei Beispielen ist folgendes Prinzip gemeinsam:

> zwei Elemente a und b irgendeiner gegebenen Menge G
> *verknüpft* man zu einem neuen Element $a * b$ in G,
> und man fragt nach "Rechengesetzen" für diese Verknüpfung.

Mathematisch ist also eine Verknüpfung eine Abbildung von der Produktmenge $G \times G$ in G, d.h.,

$$(a, b) \in G \times G \ \mapsto \ a * b \ \in G,$$

und es entsteht die Frage, welche Rechengesetze man bereits aus gewissen Grundannahmen, sogenannten *Axiomen* (Duden: Axiom = "keines Beweises bedürfender Grundsatz"), über die Verknüpfung erhält - unabhängig von der konkreten Interpretation. Man nennt das das Studium *algebraischer Strukturen.*

1.1.2 Definition einer Gruppe

1.1.4 Definition. Gegeben sei eine Menge G. Gegeben sei ferner eine *Verknüpfung* $*$ *auf* G, d.h., eine Abbildung von $G \times G$ in G.

Die Menge G zusammen mit der Verknüpfung $*$ (wir schreiben das als ein Paar $(G, *)$) heisst ***Gruppe***, wenn folgende Axiome erfüllt sind:

A1. (Assoziativgesetz) $\quad (a * b) * c = a * (b * c)$ für alle $a, b, c \in G$.

A2. Es gibt ein *neutrales Element* $e \in G$ bezüglich der Verknüpfung $*$ mit den folgenden Eigenschaften:
  (i) $e * a = a$ für alle $a \in G$,
  (ii) zu jedem $a \in G$ gibt es ein $a' \in G$, *inverses Element von* a bezüglich der Verknüpfung $*$ genannt, so dass $a' * a = e$.

Man spricht von einer ***kommutativen Gruppe*** oder ***abelschen Gruppe***, wenn $(G, *)$ ausserdem dem

A3. (Kommutativgesetz) $\quad a * b = b * a$ für alle $a, b \in G$

genügt. ◇

In einem konkreten Kontext heisst das neutrale Element meist *Nullelement* (falls $*$ als Addition interpretiert wird) oder *Einselement* (falls $*$ als Multiplikation interpretiert wird).

Offenbar gilt in den obigen Beispielen:

$(\mathbb{R}, +)$ ist eine kommutative Gruppe mit dem neutralen Element $0 \in \mathbb{R}$,

$(\mathbb{R} \setminus \{0\}, \cdot)$ ist eine kommutative Gruppe mit dem neutralen Element $1 \in \mathbb{R}$,

$(P, +)$ ist eine kommutative Gruppe mit dem neutralen Element $\underline{o} \in \mathbb{R}^{m+n}$.

1.1.5 Übung. Eine algebraische Struktur $(G, *)$ sei wie folgt gegeben:

$$G = \{0, 1\} \;\text{ mit }\; 0 * 0 = 0,\; 0 * 1 = 1 * 0 = 1,\; 1 * 1 = 0.$$

Ist $(G, *)$ eine Gruppe? Woran erinnert Sie das Beispiel? ◇

1.1.6 Übung. Man untersuche für die algebraische Struktur $(\mathbb{R}, *)$ mit

$$x * y = \frac{1}{2}(x + y) \;\text{ (arithmetisches Mittel)}$$

die Axiome (A1), (A2) und (A3). ◇

1.1.7 Übung. Man zeige (vgl. Beispiel 1.1.3), dass für die Menge $\mathcal{M}$ der regulären quadratischen Matrizen der Ordnung n und die Verknüpfung

$$A \odot B := AB \quad \text{(gewöhnliches Matrizenprodukt)}$$

durch $(\mathcal{M}, \odot)$ eine Gruppe definiert ist, die für $n \geq 2$ nicht-kommutativ ist. Man verwende als Definition der Regularität einer Matrix $A \in \mathcal{M}$ die Eigenschaft, dass es eine quadratische Matrix X der Ordnung n gibt, so dass $XA = I$ ist.

Lösung: Sei $M(n,n)$ die Menge der quadratischen Matrizen der Ordnung n. Offenbar gilt nach den Gesetzen der Matrizenmultiplikation für alle $A, B \in M(n,n)$, dass

$$AB \in M(n,n) \quad \text{und} \quad IA = A.$$

Die zweite Eigenschaft folgt, weil für $A = (a_{ij})_{i,j=1,\ldots,n}$ das Skalarprodukt der i-ten Zeile von I mit der j-ten Spalte von A das Element a_{ij} ergibt.

Die Studenten mögen selbstständig das Assoziativgesetz in $M(n,n)$ bezüglich der gewöhnlichen Matrizenmultiplikation nachprüfen. Falls das geschehen ist, sind die Ausführbarkeit der Verknüpfung $\odot$ und die Axiome (A1), (A2 i) sogar für $M(n,n)$ bewiesen.

Bleibt also zu zeigen, dass die Verknüpfung $\odot$ nicht aus $\mathcal{M}$ herausführt und auch Axiom (A2 ii), aber nicht (A3) erfüllt ist.

Wenn $A \in \mathcal{M}$ und $B \in \mathcal{M}$, so existieren nach Definition der Regularität quadratische Matrizen X_1 und X_2 der Ordnung n, so dass

$$X_1 A = I \quad \text{und} \quad X_2 B = I. \tag{1.2}$$

Folglich gilt nach dem Assoziativgesetz in $M(n,n)$ sowie wegen $IB = B$ und (1.2)

$$(X_2 X_1)(AB) = X_2(X_1(AB)) = X_2((X_1 A)B) = X_2(IB) = X_2 B = I,$$

d.h., auch AB ist regulär. Die Existenz eines inversen Elements ist per definitionem gesichert (vgl. (1.2)), somit gilt auch (A2 ii).

Ein Gegenbeispiel zur Kommutativität (A3) für $n = 2$ ist z.B.:

$$\text{Mit } A := \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \text{ gilt } \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = AA^\mathsf{T} \neq A^\mathsf{T} A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix},$$

wobei das Symbol $^\mathsf{T}$ das Transponieren einer Matrix bedeutet. Im Falle $n \geq 3$ bilde man mit der Einheitmatrix $I \in M(n-2, n-2)$ und der Nullmatrix O passender Dimension die Blockmatrix

$$B := \begin{pmatrix} A & O \\ O^\mathsf{T} & I \end{pmatrix},$$

und es gilt $BB^\mathsf{T} \neq B^\mathsf{T} B$. $\diamond$

1.1.8 Hinweis: Beim Nachweis, ob eine algebraische Struktur $(G, *)$ eine Gruppe ist, sind zwei Dinge nachzuprüfen, nämlich

(1) ob mit $a, b \in G$ auch $a * b$ in der Menge G liegt,

(2) ob die Gruppenaxiome (A1) und (A2) erfüllt sind.

1.1.3 Eigenschaften von Gruppen

Unabhängig von der konkreten Interpretation, lassen sich für alle Gruppen gewisse Eigenschaften der Verknüpfung herleiten.

1.1.9 Satz. *Ist $(G, *)$ eine Gruppe, so gelten folgende Eigenschaften:*

1. *Das neutrale Element $e \in G$ ist eindeutig bestimmt und hat auch die Eigenschaft $a * e = a$ für alle $a \in G$.*

2. *Zu jedem $a \in G$ ist das inverse Element a' eindeutig bestimmt und hat auch die Eigenschaft $a * a' = e$.*

Das zu $a \in G$ eindeutig bestimmte inverse Element wird von nun ab mit a^{-1} bezeichnet. Dann gilt auch

3. $(a^{-1})^{-1} = a$ *und* $(a * b)^{-1} = b^{-1} * a^{-1}$ *für alle* $a, b \in G$.

4. $a * x = a * y \Rightarrow x = y$ *und* $u * a = v * a \Rightarrow u = v$ *für alle* $a, x, y, u, v \in G$.

Eigenschaft 4. umfasst sogenannte Kürzungsregeln. ◇

Beweis: Um Schreibarbeit zu sparen, benutzen wir

$$ab \text{ an Stelle von } a * b.$$

Zu 1. und 2.:

Sei e *ein* neutrales Element und $a \in G$ beliebig. Zu a gibt es ein inverses Element a', zu diesem wiederum ein inverses Element, das wir a'' nennen. Dann folgt aus den Axiomen

$$\begin{aligned} aa' = e(aa') = (a''a')(aa') &= a''(a'(aa')) \\ &= a''((a'a)a') \\ &= a''(ea') \\ &= a''a' \\ &= e\,, \end{aligned}$$

und es folgt

$$ae = a(a'a) = (aa')a = ea.$$

Damit sind bereits die Vertauschbarkeitsaussagen in 1. und 2. bewiesen.

Die jeweilige Eindeutigkeit folgt nun so. Ist n ein anderes neutrales Element in G, dann haben wir

$$ne = e \text{ (da } n \text{ neutrales Element) und } en = n \text{ (da } e \text{ neutrales Element),}$$

wegen $ne = en$ folgt dann $e = n$, also ist e eindeutig bestimmt.

Ist a° ein anderes inverses Element zu $a \in G$, dann folgt

$$a^\circ = a^\circ e = a^\circ(aa') = (a^\circ a)a' = ea' = a',$$

also ist a' eindeutig zu a bestimmt. 1. und 2. sind bewiesen.

Zu 3. und 4.:

Aus $aa^{-1} = e$ folgt, dass a invers zu a^{-1} ist, also $(a^{-1})^{-1} = a$. Weiter gilt

$$(b^{-1}a^{-1})(ab) = b^{-1}(a^{-1}(ab)) = b^{-1}((a^{-1}a)b) = b^{-1}(eb) = b^{-1}b = e$$

und damit 3., während die Kürzungsregeln 4. sich nach Verknüpfung ('Multiplikation') von links bzw. rechts mit a^{-1} ergeben. □

1.1.10 Übung. Formulieren Sie die in Satz 1.1.9 hergeleiteten Gesetze speziell für die Gruppe $(\mathcal{M}, \odot)$ der regulären quadratischen Matrizen der Ordnung n (mit der gewöhnlichen Matrizenmultiplikation $A \odot B := AB$).

Erkennen Sie einige der im Fach *Mathematik II* gelernten Gesetze wieder? Haben wir nun insbesondere Eigenschaft (1.1) bewiesen? ◇

1.2 Definition eines Vektorraums und Beispiele

Analog zum Begriff der Gruppe wollen wir nun axiomatisch den Begriff des Vektorraums einführen. Im Unterschied zur Gruppe wird eine Menge V nicht nur mit einer Verknpufung ("Addition") *in* V versehen, sondern auch mit einer weiteren Verknüpfung zwischen reellen Zahlen ("Skalaren") und Elementen aus V.

1.2.1 Definition. Eine Menge V, die mit einer *Addition* $\dot{+} : V \times V \to V$ und einer *Multiplikation mit Skalaren* $\otimes : \mathbb{R} \times V \to V$ versehen ist, heisst ***(reeller) Vektorraum*** bzw. ***(reeller) linearer Raum***, falls folgende Bedingungen erfüllt sind:

(B1) $(V, \dot{+})$ ist eine kommutative Gruppe,

(B2) und es gelten als weitere Rechenregeln

$$\begin{aligned}
(\lambda + \mu) \otimes \underline{v} &= \lambda \otimes \underline{v} \;\dot{+}\; \mu \otimes \underline{v} \\
\lambda \otimes (\underline{v} \dot{+} \underline{w}) &= \lambda \otimes \underline{v} \;\dot{+}\; \lambda \otimes \underline{w} \\
\lambda \otimes (\mu \otimes \underline{v}) &= (\lambda\mu) \otimes \underline{v} \\
1 \otimes \underline{v} &= \underline{v}.
\end{aligned}$$

für alle $\underline{v}, \underline{w} \in V$ und alle $\lambda, \mu \in \mathbb{R}$.

Das neutrale Element bezüglich $\dot{+}$ heisst *Nullelement* des Vektorraums und wird in der Regel mit $\underline{o}$ bezeichnet. Das Inverse zu $\underline{v} \in V$ bezüglich der Addition wird mit $-\underline{v}$ bezeichnet. Ferner schreiben wir $\underline{v} - \underline{w}$ statt $\underline{v} \dot{+} (-\underline{w})$. ◇

1.2.2 Definition. Elemente eines Vektorraums V heissen ***Vektoren***. Wir heben sie in der Regel durch Unterstreichen hervor, also $\underline{v}$, $\underline{w}$, $\underline{o}$ etc. ◇

1.2.3 Erneuter Hinweis: Wie beim Thema *Gruppen* darf beim Nachprüfen der Vektorraumeigenschaften nie vergessen werden, auch nach $\underline{v} \dot{+} \underline{w} \in V$ (falls $\underline{v}, \underline{w} \in V$) und $\lambda \otimes \underline{v} \in V$ (falls $\lambda \in \mathbb{R}$, $\underline{v} \in V$) zu schauen!

1.2.4 Wir beschränken hier auf die Betrachtung *reeller Vektorräume* - und lassen daher auch den Zusatz *reell* bei der Benennung eines Vektorraums dann weg -, obwohl in zahlreichen Anwendungen auch die Multiplikation $\otimes$ von Vektoren mit komplexen Zahlen von Interesse ist. Allgemein lässt sich der Begriff des Vektorraums sogar auf die Multiplikation $\otimes$ mit Elementen aus einem beliebigen sogenannten *Körper* ausdehnen. Für Interessenten sei z.B. auf *G.Fischer, Lineare Algebra, Vieweg, 2002* verwiesen.

1.2.5 Die Symbole $\dot{+}$ und $\otimes$ in der Definition eines Vektorraums haben wir nur zur deutlichen Unterscheidung von den Operationen in der Menge der reellen Zahlen $\mathbb{R}$ verwendet. Wir schreiben ab sofort (in der Regel) einfacher

$$\lambda\underline{v} \quad \text{und} \quad \underline{v} + \underline{w} \quad \text{statt } \lambda \otimes \underline{v} \ \text{ bzw. } \underline{v} \dot{+} \underline{w}.$$

1.2.6 Beispiele für Vektorräume sind, wie der Leser leicht selbstständig nachprüft,

(i) der Vektorraum $\mathbb{R}^n$, d.h., die Menge der n-Tupel reeller Zahlen, versehen mit der Addition bzw. Multiplikation (mit $\lambda \in \mathbb{R}$)

$$\begin{aligned} (a_1, \ldots, a_n)^\mathsf{T} + (b_1, \ldots, b_n)^\mathsf{T} &:= (a_1 + b_1, \ldots, b_n + a_n)^\mathsf{T}, \\ \lambda\,(a_1, \ldots, a_n)^\mathsf{T} &:= (\lambda a_1, \ldots, \lambda a_n)^\mathsf{T}. \end{aligned}$$

Bemerkung: Wie in der *Mathematik II* wollen wir Elemente des Vektorraums $\mathbb{R}^n$ in der Regel als Spaltenvektoren schreiben, deshalb steht hier das Transponiertheitssymbol $^\mathsf{T}$.

(ii) der Vektorraum $C[a, b]$ der auf einem Intervall $[a, b]$ stetigen reellwertigen Funktionen, versehen mit der Addition bzw. Multiplikation (mit $\lambda \in \mathbb{R}$)

$$\begin{aligned} (f + g)(x) &:= f(x) + g(x) \ \text{ für alle } x \in [a, b], \\ (\lambda f)(x) &:= \lambda f(x) \ \text{ für alle } x \in [a, b], \end{aligned}$$

für gegebene stetige Funktionen f und g.

(iii) der Vektorraum $C^1[a,b]$ der auf einem Intervall $[a,b]$ stetig differenzierbaren reellwertigen Funktionen, versehen mit der Addition bzw. Multiplikation (mit $\lambda \in \mathbb{R}$)

$$\begin{aligned}(f+g)(x) &:= f(x)+g(x) \text{ für alle } x \in [a,b],\\ (\lambda f)(x) &:= \lambda f(x) \text{ für alle } x \in [a,b],\end{aligned}$$

für gegebene stetig differenzierbare Funktionen f und g.

(iv) der Vektorraum $\mathcal{P}^n$ der Polynome f höchstens n-ten Grades, d.h., von Funktionen der Form

$$x \in \mathbb{R} \mapsto f(x) := \alpha_0 + \sum_{j=1}^{n} \alpha_j x^j, \quad \alpha_j \in \mathbb{R}\,(j=0,1,\ldots,n) \text{ gegeben,}$$

versehen mit der Addition bzw. Multiplikation (mit $\lambda \in \mathbb{R}$)

$$\begin{aligned}(f+g)(x) &:= f(x)+g(x) \text{ für alle } x \in \mathbb{R},\\ (\lambda f)(x) &:= \lambda f(x) \text{ für alle } x \in \mathbb{R},\end{aligned}$$

für Polynome f und g in $\mathcal{P}^n$.

(v) der Vektorraum $M(m,n)$ der reellen Matrizen mit m Zeilen und n Spalten und komponentenweiser Addition bzw. Multiplikation (mit $\lambda \in \mathbb{R}$) analog zu (i).

1.2.7 Aus den Axiomen abgeleitete Rechengesetze: Aus den Axiomen für das Rechnen in einem Vektoraum V folgt für beliebige $\underline{v} \in V$ und $\lambda \in \mathbb{R}$:

1. $0\underline{v} = \underline{o}$
2. $\lambda\underline{o} = \underline{o}$
3. $\lambda\underline{v} = \underline{o} \quad \Rightarrow \quad \lambda = 0 \text{ oder } \underline{v} = \underline{o}$
4. $(-1)\underline{v} = -\underline{v}$
5. $(-\lambda)\underline{v} = -(\lambda\underline{v}) = \lambda(-\underline{v})$.

Beweis:

1. Mit den Axiomen gilt

$$\underline{o} + 0\underline{v} = 0\underline{v} = (0+0)\underline{v} = 0\underline{v} + 0\underline{v}$$

und folglich $\underline{o} = 0\underline{v}$ wegen der Kürzungsregel für die Gruppe $(V,+)$, vgl. Satz 1.1.9.

2. Mit den Axiomen gilt

$$\underline{o} + \lambda\underline{o} = \lambda\underline{o} = \lambda(\underline{o}+\underline{o}) = \lambda\underline{o} + \lambda\underline{o}$$

und somit 2. wiederum wegen der Kürzungsregel für die Gruppe $(V, +)$.

3. Ist $\lambda\underline{v} = \underline{o}$, aber $\lambda \neq 0$, dann haben wir

$$\underline{v} = 1\underline{v} = (\lambda^{-1}\lambda)\underline{v} = \lambda^{-1}(\lambda\underline{v}) = \lambda^{-1}\underline{o} = \underline{o},$$

letzteres wegen 2.

4. Es gilt nach den Axiomen und Eigenschaft 1.

$$\underline{v} + (-1)\underline{v} = 1\underline{v} + (-1)\underline{v} = (1-1)\underline{v} = 0\underline{v} = \underline{o},$$

und damit ist $(-1)\underline{v}$ das inverse Element zu $\underline{v}$ bezüglich der Addition.

5. Es gilt nach 4. und den Axiomen

$$-(\lambda\underline{v}) = (-1)(\lambda\underline{v}) = ((-1)\lambda)\underline{v} = (-\lambda)\underline{v} = (\lambda(-1))\underline{v} = \lambda((-1)\underline{v}) = \lambda(-\underline{v}),$$

was zu zeigen war. □

1.2.8 Definition. Sind V ein Vektorraum, $\underline{v}^i \in V$ und $\lambda_i \in \mathbb{R}$ für $i = 1, \ldots, m$, dann heisst

$$\sum_{i=1}^{m} \lambda_i \underline{v}^i = \lambda_1 \underline{v}^1 + \lambda_2 \underline{v}^2 + \cdots + \lambda_m \underline{v}^m$$

eine (endliche) ***Linearkombination*** der Vektoren $\underline{v}^1$, $\underline{v}^2$, ..., $\underline{v}^m$.

Ist X eine nichtleere Teilmenge von V, so heisst die Menge aller endlichen Linearkombinationen von Vektoren aus X die ***lineare Hülle*** von X und wird mit $\operatorname{lin} X$ (in vielen Büchern auch mit $\operatorname{span} X$) bezeichnet. Man setzt $\operatorname{lin} \emptyset = \{\underline{o}\}$.

Ist J eine beliebige Indexmenge, die möglicherweise aus unendlich vielen Elementen besteht, so bedeutet *per definitionem* die Schreibweise

$$\sigma = \sum_{j \in J} \lambda_j \underline{v}^j,$$

dass es eine endliche Teilmenge $I = \{j_1, \ldots, j_m\}$ von J gibt, so dass $\sigma = \sum_{i=1}^{m} \lambda_{j_i} \underline{v}^{j_i}$ gilt, also die restlichen Koeffizienten gleich 0 sind. ◇

Durch vollständige Induktion (vgl. *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984*, Lemma 1.3) zeigt man

1.2.9 Lemma. *Sind $\underline{v}^0$, $\underline{v}^1$, ..., $\underline{v}^m$ Elemente eines Vektorraums V und $\lambda_i \in \mathbb{R}$ für $i = 0, 1, \ldots, m$, dann gelten die Klammerregeln*

$$\begin{aligned} \Big(\sum_{i=1}^{m} \lambda_i\Big)\, \underline{v}^0 &= \sum_{i=1}^{m} (\lambda_i \underline{v}^0) \\ \lambda_0\, \Big(\sum_{i=1}^{m} \underline{v}^i\Big) &= \sum_{i=1}^{m} (\lambda_0 \underline{v}^i). \end{aligned}$$

◇

Wir vereinbaren, dass die Teilmengenbeziehungen im folgenden so geschrieben werden:

$$A \subset B \text{ bedeutet: } x \in A \quad \Rightarrow \quad x \in B,$$

$$A \subsetneqq B \text{ bedeutet: } A \subset B \text{ und } A \neq B,$$

im Unterschied zur Vorlesung *Mathematik II.*

Man kann ebenfalls durch vollständige Induktion unmittelbar zeigen, dass jede endliche Linearkombination von Elementen eines Vektorraums V wieder in V liegt, d.h., es gilt

$$X \subset V \quad \Rightarrow \quad \operatorname{lin} X \subset V. \tag{1.3}$$

Überdies gilt

1.2.10 Satz. *Die Abbildung, die jeder Teilmenge X eines Vektorraums V ihre lineare Hülle zuordnet, ist ein sogenannter Hüllenoperator, d.h., es gelten die folgenden Eigenschaften:*

(i) $X \subset \operatorname{lin} X$ *für alle* $X \subset V$,

(ii) $X \subset Y \Rightarrow \operatorname{lin} X \subset \operatorname{lin} Y$ *für alle* $X, Y \subset V$,

(iii) $\operatorname{lin}(\operatorname{lin} X) = \operatorname{lin} X$ *für alle* $X \subset V$. ◇

Beweis: Der Fall $X = \emptyset$ ist offensichtlich. Sei also $X \neq \emptyset$. (i) Jedes $\underline{v} \in X$ ist auch gleich $1\underline{v}$, also in $\operatorname{lin} X$. (ii) Da jede endliche Menge von Vektoren aus X auch Teilmenge von Y ist, liegen alle aus ihr gebildeten Linearkombinationen nicht nur in $\operatorname{lin} X$, sondern auch in $\operatorname{lin} Y$. (iii) Wegen (i) und (ii) sowie (1.3) ist nur noch $\operatorname{lin}(\operatorname{lin} X) \subset \operatorname{lin} X$ zu zeigen. Sei $\underline{v}$ eine endliche Linearkombination von Elementen aus $\operatorname{lin} X$, d.h., es habe die Darstellung

$$\underline{v} = \sum_{i=1}^{m} \lambda_i \underline{v}^i \quad \text{mit} \quad \lambda_i \in \mathbb{R},\ \underline{v}^i \in \operatorname{lin} X, \quad (i = 1, \dots, m).$$

Jedes $\underline{v}^i$ ist aber endliche Linearkombination aus Elementen aus X. Also sind an der Erzeugung aller dieser Vektoren $\underline{v}^i$ insgesamt endlich viele Elemente $\underline{x}^1, \dots, \underline{x}^N \in X$ beteiligt. Ohne Beschränkung der Allgemeinheit gilt also (man setze ggf. einige Koeffizienten gleich 0) für jedes $\underline{v}^i$ eine Darstellung

$$\underline{v}^i = \sum_{j=1}^{N} \mu_{ij} \underline{x}^j,\ \ \mu_{ij} \in \mathbb{R},\ (i = 1, \dots, m; j = 1, \dots, N).$$

Das liefert

$$\underline{v} = \sum_{i=1}^{m} \lambda_i (\sum_{j=1}^{N} \mu_{ij} \underline{x}^j)$$

und besagt mit den Klammerregeln von Lemma 1.2.9, dass

$$\underline{v} \;=\; \sum_{i=1}^{m}\sum_{j=1}^{N}(\lambda_i\mu_{ij})\underline{x}^j \;=\; \sum_{j=1}^{N}\sum_{i=1}^{m}(\lambda_i\mu_{ij})\underline{x}^j \;=\; \sum_{j=1}^{N}(\sum_{i=1}^{m}\lambda_i\mu_{ij})\underline{x}^j, \tag{1.4}$$

was zu zeigen war. □

1.3 Erzeugendensystem, lineare Unabhängigkeit und Basis

Wir wissen bereits aus der *Mathematik II*, dass es von grossem Interesse ist, ob ein Vektorraum V bereits durch Kenntnis endlich vieler seiner Vektoren vollständig beschrieben werden kann, genauer: ob jeder Vektor in V sich als Linearkombination endlich vieler ausgewählter Elemente aus V darstellen (*"erzeugen"*) lässt. Das steckte z.B. hinter der Bestimmung einer allgemeinen Lösung mit dem Algorithmus von Gauss.

Eine bekannte Tatsache ist z.B.

$$\mathbb{R}^n = \operatorname{lin}\{\underline{e}^1, \dots, \underline{e}^n\},$$

d.h., jedes Element des Vektorraums der n–Tupel reeller Zahlen lässt sich durch eine Linearkombination aus den Einheitsvektoren

$$\underline{e}^i = (0, \dots, 1, \dots, 0)^\mathsf{T}, \quad 1 \text{ an der } i\text{-ten Stelle},$$

erzeugen. Noch einfacher ist das Beispiel $V = \mathbb{R}$: Jeder "Vektor" aus V lässt sich durch irgend*ein* beliebiges Element $a \neq 0$ *erzeugen.*

Im allgemeinen ist es nicht möglich, einen beliebigen Vektorraum durch endlich viele Elemente zu erzeugen. Deshalb fassen wir den Begriff des Erzeugendensystems allgemeiner als in der *Mathematik II.*

1.3.1 Definition. $\mathcal{A} \subset V$ heisst ***Erzeugendensystem*** (oder ***Menge von Erzeugenden***) von V, wenn jedes Element von V als endliche Linearkombination von Elementen in $\mathcal{A}$ darstellen lässt, in anderen Worten: wenn $V = \operatorname{lin}\mathcal{A}$. ◇

Wir haben bei der Umformulierung 'in anderen Worten' benutzt, dass wir schon wissen, dass $\mathcal{A} \subset V \Rightarrow \operatorname{lin}\mathcal{A} \subset V$ gilt. Speziell folgt aus der Definition, dass jede Teilmenge $\mathcal{B}$ von V, die ein Erzeugendensystem $\mathcal{A}$ von V umfasst, selbst wieder ein Erzeugendensystem von V ist. Es gilt etwas allgemeiner

1.3.2 Satz. *Seien V ein Vektorraum und $\mathcal{A}, \mathcal{B} \subset V$ wie folgt:*

$$\mathcal{A} \textit{ ist Erzeugendensystem von } V,$$

$$\textit{jedes Element aus } \mathcal{A} \textit{ gehört zu } \operatorname{lin}\mathcal{B},$$

dann ist auch $\mathcal{B}$ Erzeugendensystem von V. ◇

Zum Beweis ist nach Satz 1.2.10 und der Definition eines Erzeugendensystems nur zu bemerken, dass offenbar gilt (beachte wieder $\mathcal{B} \subset V \Rightarrow \operatorname{lin} \mathcal{B} \subset V$):

$$\mathcal{A} \subset \operatorname{lin} \mathcal{B} \;\Rightarrow\; \operatorname{lin} \mathcal{A} \subset \operatorname{lin} \mathcal{B} \;\Rightarrow\; V = \operatorname{lin} \mathcal{A} \subset \operatorname{lin} \mathcal{B} \subset V \Rightarrow\; \operatorname{lin} \mathcal{B} = V.$$

Interessant ist nun natürlich die Frage, ob man einzelne Elemente aus einem Erzeugendensystem eliminieren kann, ohne dass die Eigenschaft, Erzeugendensystem zu sein, verloren geht oder ob es sogar ein "kleinstes" Erzeugendensystem gibt. Bei der Beantwortung hilft ein zentraler Begriff der linearen Algebra, der der *linearen Unabängigkeit.*

1.3.3 Definition. Eine endliche Teilmenge $\{\underline{v}^1, \ldots, \underline{v}^m\}$ von Vektoren aus einem Vektorraum V heisst ***linear unabhängig***, falls die Implikation

$$\sum_{i=1}^{m} \lambda_i \underline{v}^i = \underline{o} \qquad \Longrightarrow \qquad \lambda_1 = \ldots = \lambda_m = 0$$

gilt, d.h., falls sich der Nullvektor $\underline{o}$ **nur** als *triviale* Linearkombination der Vektoren $\underline{v}^1, \ldots, \underline{v}^m$ darstellen lässt. ◇

Natürlich ist der Nullvektor stets Linearkombination aus endlich vielen Vektoren $\underline{v}^1, \ldots, \underline{v}^m$, nämlich $\underline{o} = 0\underline{v}^1 + \ldots + 0\underline{v}^m$ (die *triviale* Linearkombination). Wenn das aber *die einzige Möglichkeit* ist, den Nullvektor aus $\underline{v}^1, \ldots, \underline{v}^m$ zu erzeugen, dann spricht man von linearer Unabhängigkeit.

Sofort aus der Definition folgt, dass jede Teilmenge einer linear unabhängigen Menge $\{\underline{v}^1, \ldots, \underline{v}^m\} \subset V$ selbst wieder linear unabhängig ist.

1.3.4 Beispiel. Seien $V = \mathbb{R}^4$ und

$$\underline{v}^1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \underline{v}^2 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \quad \underline{v}^3 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix},$$

dann ist die Menge $\{\underline{v}^1, \underline{v}^2, \underline{v}^3\}$ linear unabhängig, denn das homogene lineare Gleichungssystem

$$\begin{array}{ccccccc} \lambda_1 & + & \lambda_2 & + & \lambda_3 & = & 0 \\ \lambda_1 & + & \lambda_2 & + & \lambda_3 & = & 0 \\ \lambda_1 & + & \lambda_2 & & & = & 0 \\ \lambda_1 & & & & & = & 0 \end{array}$$

hat nur die triviale Lösung $\lambda_i = 0$ $(i = 1, \ldots, 3)$. ◇

1.3.5 Übung. Seien $V = \mathcal{P}^2$, d.h., der Vektorraum der reellen Polynome höchstens zweiten Grades, und $\underline{v}^i = v_i(\cdot)$ $(i = 1, 2, 3)$ mit

$$v_1(x) = x^2, \; v_2(x) = x, \; v_3(x) = 7x - 8x^2 \qquad (\forall x \in \mathbb{R}).$$

Zeigen Sie: Die Mengen $\{\underline{v}^1, \underline{v}^2\}$, $\{\underline{v}^1, \underline{v}^3\}$ und $\{\underline{v}^2, \underline{v}^3\}$ sind linear unabhängig, aber die Menge $\{\underline{v}^1, \underline{v}^2, \underline{v}^3\}$ ist nicht linear unabhängig. ◇

1.3.6 Definition. Eine beliebige Teilmenge $\mathcal{A}$ von Vektoren aus einem Vektorraum V heisst ***linear unabhängig***, falls jede endliche Teilmenge von $\mathcal{A}$ linear unabhäng ist.

Eine beliebige Teilmenge $\mathcal{A}$ von Vektoren aus einem Vektorraum V heisst ***linear abhängig***, falls sie nicht linear unabhäng ist. ◇

Eine Menge $\mathcal{A} \subset V$ ist also linear abhängig, falls es eine endliche Teilmenge $\{\underline{v}^1, \ldots, \underline{v}^m\}$ von $\mathcal{A}$ gibt, so dass das System

$$\sum_{i=1}^{m} \lambda_i \underline{v}^i \;=\; \underline{o} \text{ eine Lösung } \underline{\lambda} = (\lambda_1, \ldots, \lambda_m) \neq \underline{o} \text{ hat.}$$

1.3.7 Es folgt sofort aus der Definition, dass jede Teilmenge einer *beliebigen* linear unabhängigen Menge in V selbst wieder linear unabhängig ist. Damit gilt auch, dass $\mathcal{B} \subset V$ linear abhängig ist, wenn $\mathcal{A} \subset \mathcal{B}$ linear abhängig ist (Kontraposition).

1.3.8 Enthält $\mathcal{A} \subset V$ den Nullvektor $\underline{o}$ von V, so ist $\mathcal{A}$ linear abhängig. Begründung: Wegen $\underline{o} = 1\underline{o}$ ist die einelementige Menge $\{\underline{o}\}$ stets linear abhängig. Damit folgt die Aussage aus 1.3.7.

1.3.9 Beispiel. Betrachten wir im Vektorraum $V = C[a, b]$ der auf einem Intervall $[a, b]$ stetigen reellwertigen Funktionen die Teilmenge (von sogenannten *Monomen*)

$$\mathcal{A} = \{\underline{v}^i = v_i(\cdot) | i \in \mathbb{N} \cup \{0\}\} \quad \text{mit } v_0(\cdot) \equiv 1 \text{ und } v_i(x) = x^i \text{ für } i = 1, 2, \ldots,$$

alle $v_i(\cdot)$ (eingeschränkt) definiert auf $x \in [a, b]$. Wir zeigen: Die Menge $\mathcal{A}$ ist eine linear unabhängige Teilmenge des Vektorraums V. Man nehme irgendeine endliche Indexmenge $I \subset \mathbb{N} \cup \{0\}$ und betrachte die dazugehörige Menge von Monomen $\{v_i(\cdot) | i \in I\}$. Da I endlich ist, gibt es einen grössten Grad n für alle diese Monome. Um zu zeigen, dass $\{v_i(\cdot) | i \in I\}$ linear unabhängig ist, reicht es hin zu zeigen, dass die (im allgemeinen grössere!) endliche Teilmenge von $\mathcal{A}$,

$$\mathcal{A}_{(n)} = \{v_i(\cdot) | 0 \leq i \leq n\},$$

linear unabhängig ist. Bezeichnet $0(\cdot)$ das Nullelement von $C[a, b]$, d.h., die Funktion $x \mapsto 0$ für $x \in [a, b]$, müssen wir also das System

$$\sum_{i=0}^{n} \lambda_i \, v_i(\cdot) = 0(\cdot) \text{ mit den Unbekannten } \lambda_0, \lambda_1, \ldots, \lambda_n$$

anschauen, d.h., man suche einen Vektor $\underline{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_n)$, so dass

$$\varphi_{\underline{\lambda}}(x) := \lambda_0 + \sum_{i=1}^{n} \lambda_i x^i = 0 \text{ für alle } x \in [a, b] \text{ gilt.} \tag{1.5}$$

Gäbe es ein $\underline{\lambda} \neq \underline{o}$, dann definierte $\varphi_{\underline{\lambda}}(\cdot)$ ein reelles Polynom in x höchstens n-ten Grades. Das hat aber nach dem Fundamentalsatz der Algebra höchstens n reelle Nullstellen im Widerspruch zu (1.5). Damit ist $\mathcal{A}$ linear unabhängig. $\diamond$

1.3.10 Satz. *Seien V ein Vektorraum und $\mathcal{A}$ eine Teilmenge von V mit mindestens 2 Elementen. Dann ist $\mathcal{A}$ genau dann linear abhängig, wenn für mindestens ein Element $\underline{v}$ von $\mathcal{A}$ die Inklusion $\underline{v} \in \operatorname{lin}(\mathcal{A} \setminus \{\underline{v}\})$ gilt.* $\diamond$

Beweis in der Vorlesung, vgl. auch *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984*, Satz 1.8. $\square$

Mit anderen Worten: $\mathcal{A}$ *ist linear abhängig* bedeutet äquivalent, dass mindestens ein Element $\underline{v} \in \mathcal{A}$ eine Linearkombination von *anderen* Elementen aus $\mathcal{A}$ ist.

In einer linear abhängigen Menge kann man im allgemeinen nicht jedes Element aus den verbleibenden erzeugen. Man betrachte in $V = \mathbb{R}^2$ (Machen Sie eine Zeichnung!) das Beispiel

$$\underline{c} \notin \operatorname{lin}\{\underline{a}, \underline{b}\} \quad \text{für} \quad \underline{a} = (1,1),\ \underline{b} = (-1,-1),\ \underline{c} = (0,1).$$

1.3.11 Satz. *Seien V ein Vektorraum und $\mathcal{A}$ eine Teilmenge von V. Dann ist $\mathcal{A}$ genau dann linear unabhängig, wenn jedes Element des Vektorraums V auf höchstens eine Weise als endliche Linearkombination von Vektoren aus $\mathcal{A}$ dargestellt werden kann.* $\diamond$

Beweis in der Vorlesung, vgl. auch *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984*, Satz 1.9. $\square$

1.3.12 Beispiel. Für $V = \mathbb{R}^3$ und $\mathcal{A} = \{(1,1,1)^\mathsf{T}, (-1,-1,0)^\mathsf{T}\}$ gilt die Darstellung

$$\underline{v} = (v_1, v_2, v_3)^\mathsf{T} = \lambda_1 (1,1,1)^\mathsf{T} + \lambda_2 (-1,-1,0)^\mathsf{T} \quad \text{mit gewissen } \lambda_1, \lambda_2 \in \mathbb{R}$$

entweder gar nicht, z.B. für $\underline{v} = (1,0,0)^\mathsf{T}$, oder eindeutig - man löse das inhomogene lineare Gleichungssystem in den Variablen λ_1, λ_2,

$$\begin{array}{ccccc} \lambda_1 & - & \lambda_2 & = & v_1 \\ \lambda_1 & - & \lambda_2 & = & v_2 \\ \lambda_1 & & & = & v_3. \end{array}$$

Das illustriert den vorhergehenden Satz. Natürlich ist $\mathcal{A}$ nach Definition linear unabhängig. (Warum?) $\diamond$

Nun wenden wir uns der Frage nach einem minimalen Erzeugendensystem in einem Vektorraum V zu. Es handelt sich dabei gerade um ein *linear unabhängiges Erzeugendensystem.*

Ist ein Erzeugendensystem $\mathcal{A}$ nämlich linear abhängig, so gibt es nach Satz 1.3.10 ein Element $\underline{v}$ von $\mathcal{A}$ mit $\underline{v} \in \operatorname{lin}(\mathcal{A} \setminus \{\underline{v}\})$, also $\mathcal{A} \subset \operatorname{lin}(\mathcal{A} \setminus \{\underline{v}\})$ und somit nach den Hülleneigenschaften von "lin" auch $V = \operatorname{lin} \mathcal{A} \subset \operatorname{lin}(\mathcal{A} \setminus \{\underline{v}\}) \subset V$, d.h., $\mathcal{A}$ ist kein minimales Erzeugendensystem.

Hat man andererseits eine linear unabhängige Menge $\mathcal{A}$ von Erzeugenden von V, so führt das Hinzufügen eines Elements $\underline{v} \in V \setminus \mathcal{A}$ zu einer linear abhängigen Menge $\mathcal{A} \cup \{\underline{v}\}$ (denn es gilt $\underline{v} \in V = \operatorname{lin} \mathcal{A} = \operatorname{lin}[(\mathcal{A} \cup \{\underline{v}\}) \setminus \{\underline{v}\}]$ und man kann Satz 1.3.10 anwenden).

1.3.13 Definition. Ist V ein Vektorraum und $\mathcal{B}$ ein linear unabhängiges Erzeugendensystem von V, dann heisst $\mathcal{B}$ *eine* ***Basis*** von V. $\diamond$

Es gibt im Falle eines Vektorraums V mit mindestens 2 Elementen zu einer gegebenen Basis $\mathcal{B}$ stets unendlich viele Basen von V, da auch $\mathcal{B}(\alpha) = \{\alpha\underline{v} \mid \underline{v} \in \mathcal{B}\}$ für beliebiges $\alpha \neq 0$ wieder eine Basis von V ist. Wir werden aber zeigen, dass im Falle von Vektorräumen mit endlichen Erzeugendensystemen dennoch ein übereinstimmendes Merkmal existiert: die Anzahl der Elemente zweier verschiedener Basen ist stets gleich.

Zitat *G.Fischer, Lineare Algebra, Vieweg, 2002*, Abschnitt 1.5, trifft auch auf uns zu: "Die Begriffe sind so erklärt, dass die leere Menge eine Basis des Nullvektorraums ist. Diese kleine Freude kann man ihr gönnen." (*Nullvektorraum* = Vektorraum mit nur einem Element, und das muss nach den Axiomen das Nullelement sein.)

1.3.14 Satz. *Sei V ein Vektorraum. $\mathcal{B} \subset V$ ist eine Basis von V genau dann, wenn jedes Element $\underline{v}$ von V auf genau eine Weise als endliche Linearkombination von Vektoren aus $\mathcal{B}$ dargestellt werden kann.* $\diamond$

Beweis: Sei $\mathcal{B}$ Basis von V. Wegen $V = \operatorname{lin} \mathcal{B}$ ist jedes $\underline{v} \in V$ auf mindestens eine Weise aus $\mathcal{B}$ erzeugbar, nach Satz 1.3.11 auf höchstens eine Weise. Ist umgekehrt Element $\underline{v}$ von V auf genau eine Weise als endliche Linearkombination von Vektoren aus $\mathcal{B}$ darstellbar, so ist $V = \operatorname{lin} \mathcal{B}$ (also $\mathcal{B}$ Erzeugendensystem von V) und $\mathcal{B}$ nach Satz 1.3.11 linear unabhängig. $\square$

1.3.15 Definition. Sind I eine beliebige Indexmenge, die Familie $\mathcal{B} = \{\underline{v}^i \mid i \in I\}$ eine Basis eines Vektorraums V und hat $\underline{v} \in V$ die Darstellung (zur Definition vgl. 1.2.8)

$$\underline{v} = \sum_{i \in I} \lambda_i \underline{v}^i,$$

dann heissen die Koeffizienten λ_i, $i \in I$, ***Komponenten*** von $\underline{v}$ ***bezüglich der Basis*** $\mathcal{B}$. $\diamond$

Wir haben den Zusatz "bezüglich der Basis $\mathcal{B}$" fett unterstrichen, da die Komponenten natürlich von der gewählten Basis abhängen. Selbst im trivialen Fall $V = \mathbb{R}$ hat z.B. der "Vektor" (= die Zahl) 10 bezüglich der Basis $\mathcal{B} = \{1\}$, die (einzige) Komponente 10, bezüglich der Basis $\mathcal{B} = \{-5\}$ die (einzige) Komponente -2.

1.4 Endlichdimensionale Vektorräume

In diesem Abschnitt betrachten wir die Situation, dass der betrachtete Vektorraum V ein endliches Erzeugendensystem hat. Wir beschreiben zunächst eine Konstruktion zur Bestimmung einer endlichen Basis von V, die sogar die Eigenschaft hat, eine vorgegebene linear unabhängige Teilmenge von V zu umfassen.

1.4.1 Lemma. *Sei V ein Vektorraum mit endlichem Erzeugendensystem $\mathcal{A} = \{\underline{v}^1, \ldots, \underline{v}^n\}$. Ist $\mathcal{C}$ eine linear unabhängige Teilmenge von $\mathcal{A}$, so existiert stets eine Menge $\mathcal{B}$ mit*

$$\mathcal{C} \subset \mathcal{B} \subset \mathcal{A},$$

so dass $\mathcal{B}$ Basis von V ist. ◇

Beweis: Wir definieren induktiv eine (endliche) Folge von Mengen, indem wir zu $\mathcal{C}$ sukzessive ein Element von $\mathcal{A}$ entweder hinzunehmen oder nicht, und zwar gemäss

$$\begin{aligned} \mathcal{B}_0 &:= \mathcal{C} \\ \mathcal{B}_{i+1} &:= \begin{cases} \mathcal{B}_i, & \text{falls } \underline{v}^{i+1} \in \operatorname{lin} \mathcal{B}_i \ \ \text{(d.h., } \mathcal{B}_i \cup \{\underline{v}^{i+1}\} \text{ linear abhängig)}, \\ \mathcal{B}_i \cup \{\underline{v}^{i+1}\}, & \text{falls } \mathcal{B}_i \cup \{\underline{v}^{i+1}\} \text{ linear unabhängig}, \end{cases} \end{aligned}$$

für $i = 0, 1, \ldots, n-1$. Die Menge $\mathcal{B}_n$ ist nach dieser Konstruktion linear unabhängig und umfasst $\mathcal{C}$. Ferner gilt nach Voraussetzung und nach Konstruktion

$$\mathcal{A} \text{ ist Erzeugendensystem von } V,$$

$$\text{jedes Element aus } \mathcal{A} \text{ gehört zu } \operatorname{lin} \mathcal{B}_n,$$

Das war aber gerade die Voraussetzung von Satz 1.3.2, also ist nach diesem Satz die Menge $\mathcal{B} := \mathcal{B}_n$ ein Erzeugendensystem von V. Da diese Menge $\mathcal{B}$ auch linear unabhängig ist, ist $\mathcal{B}$ eine Basis von V. □

1.4.2 Satz. (Basisergänzungssatz) *Besitzt V ein endliches Erzeugendensystem und ist $\mathcal{C} = \{\underline{v}^1, \ldots, \underline{v}^m\} \subset V$ linear unabhängig, dann existiert eine endliche Basis $\mathcal{B}$ von V, die $\mathcal{C}$ umfasst (d.h., man kann $\mathcal{C}$ zu einer endlichen Basis $\mathcal{B}$ von V ergänzen).* ◇

Beweis: Der Beweis ist eine einfache Anwendung von Lemma 1.4.1 und kann dem Leser getrost als Übungsaufgabe überlassen werden. Hinweis: Man schaue sich die Menge $\mathcal{A}$ an, die aus der Vereinigung des gegebenen endlichen Erzeugendensystems von V mit der Menge $\mathcal{C}$ entsteht. □

1.4.3 Da die Konstruktion von Lemma 1.4.1 speziell auch für den Fall $\mathcal{C} = \emptyset$ richtig ist, haben wir insbesondere erhalten, dass jeder Vektorraum mit endlichem Erzeugendensystem auch eine endliche Basis besitzt. Können unterschiedliche (endliche) Basen eventuell unterschiedlich viele Elemente enthalten? Damit befassen wir uns nun.

1.4.4 Satz. (Austauschsatz von Steinitz) *Sei $\mathcal{B} = \{\underline{v}^1, \underline{v}^2, \ldots, \underline{v}^n\}$ eine Basis von V und habe $\underline{w}$ die Darstellung*

$$\underline{w} = \sum_{i=1}^{n} \lambda_i \underline{v}^i \quad \textit{mit } \lambda_1 \neq 0, \tag{1.6}$$

dann ist $\widetilde{\mathcal{B}} = \{\underline{w}, \underline{v}^2, \ldots, \underline{v}^n\}$ auch eine Basis von V. ◇

Beweis wird in der Vorlesung gegeben, vgl. auch *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984* Satz 1.13 oder *G.Fischer, Lineare Algebra, Vieweg, 2002* Abschnitt 1.5.4. □

Natürlich ist der Austausch von $\underline{w}$ auch gegen ein Element $\underline{v}^j$ von $\mathcal{B}$ mit $j \neq 1$ möglich, man muss dann in (1.6) $\lambda_j \neq 0$ fordern.

1.4.5 Übung. Machen Sie sich das Prinzip des Beweises des Steinitzschen Austauschsatzes klar, indem Sie ihn auf die kanonische Basis $\mathcal{B} = \{\underline{e}^1, \underline{e}^2, \underline{e}^3\}$ des $\mathbb{R}^3$ (mit den Einheitsvektoren $\underline{e}^i$) und den Vektor $\underline{w} = (1, -1, 1)^\mathsf{T}$ (oder alternativ auf $\underline{w} = (0, -1, 1)^\mathsf{T}$ oder wiederum anders mit beliebigem $\underline{w} \in \mathbb{R}^3$ - Vorsicht! -) anwenden. ◇

1.4.6 Satz. *Sind $\mathcal{A}$ und $\mathcal{B}$ zwei endliche Basen eines Vektorraums V, so enthalten $\mathcal{A}$ und $\mathcal{B}$ gleich viele Elemente.* ◇

Beweis wird in der Vorlesung gegeben, vgl. auch *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984* Satz 1.14 oder *G.Fischer, Lineare Algebra, Vieweg, 2002* Abschnitt 1.5.5. □

1.4.7 Korollar. *Hat ein Vektorraum V eine Basis mit n Elementen und ist $\mathcal{C}$ eine linear unabhängige Teilmenge von V mit n Elementen, so ist $\mathcal{C}$ eine Basis von V.* ◇

Beweis: Sei $\mathcal{B}_0$ Basis von V. Es gilt offenbar $\mathcal{C} \subset \mathcal{B}_0 \cup \mathcal{C}$ und $\mathcal{A} := \mathcal{B}_0 \cup \mathcal{C}$ ist endliches Erzeugendensystem von V. Damit erfüllen $\mathcal{A}$ und $\mathcal{C}$ die Voraussetzungen von Lemma 1.4.1, und es existiert also eine Basis $\mathcal{B}$ von V mit

$$\mathcal{C} \subset \mathcal{B} \subset \mathcal{A}.$$

Nach Satz 1.4.6 hat die Basis $\mathcal{B}$ n Elemente, also folgt - da auch $\mathcal{C}$ n Elemente enthält - sogar $\mathcal{B} = \mathcal{C}$. □

Folgende Definition ist nun offenbar gerechtfertigt. Wir werden deshalb nun nicht mehr kompliziert von einem *Vektorraum mit endlichem Erzeugendensystem*, sondern von einem *endlich-dimensionalen Vektorraum* sprechen.

1.4.8 Definition. Hat ein Vektorraum V eine Basis mit n Elementen, so heisst V ***n-dimensional*** (oder V habe ***die Dimension n***). **Symbol:** $\dim V = n$. $V = \{\underline{o}\}$ heisst ***nulldimensional***. Ein Vektorraum $V \neq \{\underline{o}\}$, der keine endliche Basis hat, heisst ***unendlichdimensional***. $\diamond$

1.4.9 Wenn der Vektorraum V eine linear unabhängige Teilmenge $\mathcal{A}$ mit unendlich vielen Elementen enthält, ist V offenbar unendlichdimensional. Es gilt auch die Umkehrung: Sei V unendlichdimensional und $\mathcal{A} = \{\underline{v}^1, \ldots, \underline{v}^n\} \subset V$ linear unabhängig. Dann gibt es stets ein $\underline{w} \in V$, so dass auch $\mathcal{A} \cup \{\underline{w}\}$ linear unabhängig ist. In der Tat: Wäre $\mathcal{A} \cup \{\underline{w}\}$ *für jedes* $\underline{w} \in V$ linear abhängig, dann wäre $\mathcal{A}$ ein Erzeugendensystem und damit (wegen seiner linearen Unabhängigkeit) auch eine Basis von V - im Widerspruch zur Voraussetzung der Unendlichdimensionalität.

1.4.10 Übung. Zeigen Sie mit Hilfe von Beispiel 1.3.9, dass der Vektorraum $V = C[a, b]$ der auf einem Intervall $[a, b]$ stetigen reellwertigen Funktionen unendlichdimensional ist. $\diamond$

1.4.11 Übung. Die Aussagen von Satz 1.4.6 und Korollar 1.4.7 hatten Sie bereits in der *Mathematik II* ohne Beweis kennengelernt und vielfach in Übungsaufgaben angewendet. Erinnern Sie sich? Zum Beispiel: Man betrachte z.B. die Lösungsmenge L eines homogenen linearen Gleichungssystems (das ist ein Vektorraum, wie wir aus der *Mathematik II* wissen, repetieren Sie das bitte in Übung 1.5.3 unten), die man aus dem Gauss-Algorithmus ermittelt hat. Man kennt also eine Basis $\mathcal{B}$ von L. Man kann nun für eine endliche Menge von Vektoren X aus dem Vektorraum L leicht nachprüfen, ob X eine Basis oder ein Erzeugendensystem von L ist. $\diamond$

1.5 Unterräume

1.5.1 Definition. Es seien V ein Vektorraum und W eine Teilmenge von V. Ist W mit der Addition und der Multiplikation (mit Skalaren) aus V selbst ein Vektorraum, dann heisst W ***Unterraum von V*** (in der Literatur auch als *Untervektorraum* oder *linearer Teilraum* von V bezeichnet). $\diamond$

1.5.2 Satz. (Unterraumkriterium) *Seien V ein Vektorraum und $W \subset V$. W ist ein Unterraum genau dann, wenn*

(i) $\underline{o} \in W$ *($\underline{o}$ Nullelement von V),*

(ii) $\underline{v}, \underline{w} \in W \;\Rightarrow\; \underline{v} + \underline{w} \in W$ *(d.h., W ist abgeschlossen bezüglich der Addition),*

(iii) $\underline{v} \in W, \lambda \in \mathbb{R} \;\Rightarrow\; \lambda \underline{v} \in W$ *(d.h., W ist abgeschlossen bezüglich der Multiplikation mit Skalaren).*

$\diamond$

Beweis: Wenn $W \subset V$ selbst ein Vektorraum ist, gelten natürlich (i)–(iii). Wenn (i)–(iii) erfüllt sind, bleiben Addition und Multiplikation mit Skalaren für Elemente aus W wieder in W. Ferner ist $\underline{o} \in W$ nach (i), und es gelten Assoziativ- und Kommutativgesetz der Addition sowie die Rechenregeln (B2) aus Definition 1.2.1 wegen der entsprechenden Gesetze in V. Schliesslich folgt mit $\underline{w} \in W$ auch für sein negatives Element $-\underline{w}$ in V

$$-\underline{w} = (-1)\underline{w} \quad \text{(Regel 4 in 1.2.7)} \quad \text{mit } (-1)\underline{w} \in W \quad \text{wegen (iii)},$$

also $-\underline{w} \in W$. Damit sind alle Eigenschaften eines Vektorraums für W gültig unter (i)–(iii).

$\square$

Das Beispiel in der folgenden Übung ist Ihnen wohlbekannt und erklärt einleuchtend, warum man sich für Unterräume interessiert: Man kann beliebige Linearkombinationen von Lösungen eines homogenen linearen Gleichungssystems bilden, und man bleibt in der Menge der Lösungen.

1.5.3 Übung. Wir betrachten ein homogenes lineares Gleichungssystem mit m Gleichungen in n Unbekannten,

$$\sum_{j=1}^{n} a_{ij} x_j = 0, \; i = 1, \ldots, m, \tag{1.7}$$

(a_{ij} $\forall i \forall j$ gegeben). Das System (1.7) ist stets lösbar (trivialerweise ist der Nullvektor des $\mathbb{R}^n$ eine Lösung). Zeigen Sie, dass die Lösungsmenge L von (1.7), d.h.,

$$L = \{\underline{x} = (x_1, \ldots, x_n)^{\mathsf{T}} \mid \underline{x} \text{ löst das System (1.7)}\}$$

ein Unterraum des $\mathbb{R}^n$ ist. $\diamond$

Allgemein gilt die folgende Aussage, die sofort aus der Definition der linearen Hülle und dem Unterraumkriterium folgt.

1.5.4 Lemma. *Sind V ein Vektorraum und $W \subset V$, dann ist die lineare Hülle $\operatorname{lin} W$ ein Unterraum von V.* $\diamond$

Ebenfalls sofort aus dem Unterraumkriterium folgt

1.5.5 Lemma. *Sind V ein Vektorraum und $\{W_i, i \in I\}$ eine beliebige Familie von Unterräumen W_i von V, dann ist auch der Durchschnitt*

$$W = \bigcap_{i \in I} W_i$$

ein Unterraum von V. $\diamond$

1.5.6 Satz. *Sind V ein Vektorraum und W ein Unterraum von V, dann gilt* $\dim W \leq \dim V$. ◇

Beweis: Ist W unendlichdimensional, so hat W nach 1.4.9 eine unendliche linear unabhängige Teilmenge $\mathcal{A}$, die wiederum in V liegt, also ist auch V unendlichdimensional. Sind $\dim W$ endlich und $\dim V$ unendlich, so ist nichts zu zeigen. Sind $\dim W$ und $\dim V$ endlich, so hat W eine endliche Basis $\{\underline{v}^1, \ldots, \underline{v}^n\}$ (d.i. speziell eine linear unabhängige Teilmenge von V), die nach dem Basisergänzungssatz zu einer Basis von V ergänzt werden kann. Also ist auch in diesem Fall $\dim W \leq \dim V$. □

In der Modellierung komplizierter Zusammenhänge versucht man sehr oft, schwierig zu handhabende Funktionen durch einfachere anzunähern (z.B. differenzierbare Funktionen durch lineare Funktionen). Insbesondere versucht man in zahlreichen Anwendungen, stetige Funktionen durch Polynome zu approximieren oder periodische Funktionen durch Linearkombinationen spezieller trigonometrischer Funktionen (Elementarschwingungen) anzunähern. Dazu sind die folgenden beiden Beispiele nützlich.

1.5.7 Übung. Zeigen Sie, dass der Vektorraum $\mathcal{P}^n$ der Polynome höchstens n-ten Grades (eingeschränkt auf das Intervall $[a, b]$) ein Unterraum des Vektorraums $C[a, b]$ ist. Zeigen Sie weiterhin, dass $\mathcal{P}^n$ die Dimension $n + 1$ hat und geben Sie eine Basis von $\mathcal{P}^n$ an. Zur Erinnerung: $\mathcal{P}^n$ und $C[a, b]$ wurden in 1.2.6 eigeführt. ◇

1.5.8 Übung. Zeigen Sie, dass die Menge der Funktionen auf $[-\pi, \pi]$ der Form

$$f(x) = \sum_{\nu=0}^{n} \alpha_\nu \cos \nu x \; + \; \sum_{\mu=1}^{n} \beta_\mu \sin \mu x,$$

$\alpha_0, \alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n$ gegeben, einen Unterraum von $C[-\pi, \pi]$ bildet. ◇

Aus dem Physikunterricht der Schule wissen Sie vielleicht noch, dass Kräfte in Komponenten zerlegt werden, die ihre eigene physikalische Bedeutung haben. Oder: Wenn ein dreidimensionaler Datenvektor gegeben ist, interessiert man sich häufig dafür, welchen "Abstand" er von einer Referenzebene hat. Oder: Ist (wie in Übung 1.5.7) g eine stetige Funktion über dem Intervall $[a, b]$, so interessiert - im Sinne einer "guten Näherung" - , ihr "Abstand" zum Unterraum $\mathcal{P}^n$.

Bei diesen Problemen ist die Summe von Unterräumen und die (möglichst eindeutige) Zerlegung eines Vektorraums in eine Summe von Unterräumen von Nutzen. Wir definieren für Teilmengen W_1 und W_2 eines Vektorraums V die *algebraische Summe* (auch: *Minkowski-Summe*)

$$W_1 + W_2 := \{\underline{w} \,|\, \underline{w} = \underline{w}^1 + \underline{w}^2 \;\text{ mit } \underline{w}^1 \in W_1,\, \underline{w}^2 \in W_2\} \; .$$

Dann gilt (vgl. *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984*, Lemma 1.19)

1.5.9 Lemma. *Sind V ein Vektorraum und W_1, W_2 Unterräume von V, dann ist auch die algebraische Summe $W_1 + W_2$ ein Unterraum von V.* $\diamond$

Im allgemeinen kann natürlich ein Element aus $W_1 + W_2$ mehrere Darstellungen haben. Man betrachte z.B. $V = \mathbb{R}^3$ und $W_1 = \operatorname{lin}\{\underline{e}^1, \underline{e}^2\}$, $W_2 = \operatorname{lin}\{\underline{e}^1, \underline{e}^3\}$. Dann ist z.B. $(1,1,1)^\mathsf{T} = (1,1,0)^\mathsf{T} + (0,0,1)^\mathsf{T} = (0,1,0)^\mathsf{T} + (1,0,1)^\mathsf{T}$. Einen Ausweg liefert

1.5.10 Lemma. *Seien V ein Vektorraum und W_1, W_2 Unterräume von V und sei $\underline{v} \in W_1 + W_2$. Die Darstellung $\underline{v} = \underline{w}^1 + \underline{w}^2$ mit $\underline{w}^1 \in W_1$ und $\underline{w}^2 \in W_2$ ist eindeutig genau dann, wenn $W_1 \cap W_2 = \{\underline{o}\}$.* $\diamond$

Beweis wird in der Vorlesung geführt, vgl. auch *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984*, Lemma 1.20. $\square$

1.5.11 Definition. Sind W_1 und W_2 Unterräume eines Vektorraums V, dann heisst die Menge $W_1 + W_2$ ***direkte Summe*** von W_1 und W_2, wenn $W_1 \cap W_2 = \{\underline{o}\}$. Die direkte Summe schreiben wir als $W_1 \oplus W_2$. $\diamond$

1.5.12 Satz. *Sind V ein endlichdimensionaler Vektorraum und W_1 ein Unterraum von V, dann existiert ein Unterraum W_2 von V, so dass $V = W_1 \oplus W_2$.* $\diamond$

Beweis wird in der Vorlesung geführt, vgl. auch *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984*, Satz 1.22. $\square$

1.5.13 Ohne Beweis sei an dieser Stelle bemerkt, dass gegebene Basen von Unterräumen W_1 und W_2 eines endlichdimensionalen Vektorraums V zu einer Basis von $W_1 \oplus W_2$ vereinigt werden können und dass dann $\dim(W_1 \oplus W_2) = \dim W_1 + \dim W_2$ gilt.

Kapitel 2

Lineare Abbildungen und Matrizen

2.1 Matrizen

Im Folgenden wiederholen wir aus der Lehrveranstaltung *Mathematik I* die Definitionen einer Matrix und der Operationen zwischen Matrizen sowie weitere Begriffe und elementare Eigenschaften in diesem Kontext.

2.1.1 Definition. Gegeben sind $m \cdot n$ reelle Zahlen a_{ij}, $i = 1, \ldots, m$, $j = 1, \ldots, n$. Das daraus gebildete Zahlenschema

$$A \;=\; \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}$$

heisst ***(reelle) Matrix*** mit m *Zeilen* und n *Spalten* (oder kurz ***(m×n)–Matrix*** bzw. ***Matrix vom Typ m×n***). Die Menge aller (reellen) $(m \times n)$–Matrizen wird mit dem Symbol $M(m, n)$ bezeichnet.

Eine $(n \times n)$–Matrix wird auch ***quadratische Matrix der Ordnung n*** oder kurz ***n-reihige Matrix*** genannt. ◇

2.1.2 Bezeichnungen. Die Grössen a_{ij} heissen *Elemente von A*, wir schreiben eine Matrix kürzer auch

$$A = (a_{ij}; 1 \leq i \leq m,\ 1 \leq j \leq n) \quad \text{bzw.} \quad A = (a_{ij}),$$

letztere Schreibweise verwenden wir, wenn der Typ aus dem Zusammenhang klar ist.

Die i-te Zeile einer Matrix A mit m Zeilen schreiben wir als $A_{i\bullet}$, d.h.,

$$A = \begin{pmatrix} A_{1\bullet} \\ A_{2\bullet} \\ \vdots \\ A_{m\bullet} \end{pmatrix}.$$

Die j-te Spalte einer Matrix A mit n Spalten schreiben wir als $A_{\bullet j}$, d.h.,

$$A = \begin{pmatrix} A_{\bullet 1} & A_{\bullet 2} & \ldots & A_{\bullet n} \end{pmatrix}.$$

2.1.3 Satz. *Die Menge $M(m,n)$ der $(m \times n)$–Matrizen bildet mit den Verknüpfungen*

$$\begin{array}{lcl} A+B & := & (a_{ij}+b_{ij};\ 1 \le i \le m,\ 1 \le j \le n) \\ \lambda A & := & (\lambda a_{ij};\ 1 \le i \le m,\ 1 \le j \le n) \end{array}$$

(für beliebige $A = (a_{ij};\ 1 \le i \le m,\ 1 \le j \le n)$, $B = (b_{ij};\ 1 \le i \le m,\ 1 \le j \le n)$ und $\lambda \in \mathbb{R}$) einen Vektorraum der Dimension $m \cdot n$ mit der Nullmatrix als Nullelement. $\diamond$

Der Beweis ist offensichtlich und sei dem Leser als Übung überlassen.

2.1.4 Wir erinnern ferner an die folgenden Bezeichnungen und Begriffe.

Sei $A \in M(m,n)$ gegeben. Die mit $A^\top$ bezeichnete $(n \times m)$–Matrix entsteht aus der Matrix A dadurch, dass für alle i und j die i–te Zeile von A zur i–ten Spalte von $A^\top$ und die j–te Spalte von A zur j–ten Zeile von $A^\top$ wird. $A^\top$ heisst *Transponierte von A*.

Eine $(1 \times n)$–Matrix heisst *Zeilenmatrix* (oder *Zeilenvektor*), geschrieben $(a_1, \ldots, a_n)$, eine $(m \times 1)$–Matrix heisst *Spaltenmatrix* (oder *Spaltenvektor*), geschrieben

$$\begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

Die Elemente a_{ii} $(i = 1, \ldots, n)$ einer quadratischen Matrix $A = (a_{ij})$ der Ordnung n bilden die sogenannte *Hauptdiagonale von A*. Hat eine Matrix $A \in M(n,n)$ ausserhalb der Hauptdiagonalen nur Elemente gleich 0, so heisst A *Diagonalmatrix*; hat sie unterhalb der Hauptdiagonalen nur Elemente gleich 0, heisst sie *obere Dreicksmatrix*; hat sie oberhalb der Hauptdiagonalen nur Elemente gleich 0, heisst sie *untere Dreicksmatrix*.

Eine quadratische Matrix A mit der Eigenschaft $A = A^\top$ heisst *symmetrisch*; mit der Eigenschaft $A = -A^\top$ heisst sie *schiefsymmetrisch* (oder *alternierend*).

Eine Diagonalmatrix $A = (a_{ij})$ mit $a_{ii} = 1\ \forall i$ heisst *Einheitsmatrix*, Bezeichnung: I.

2.1.5 Beispiel. Beispiele für die Begriffe aus 2.1.4:

- Transponierte von A

$$A = \begin{pmatrix} 4 & -2 & 1 & 0 \\ 3 & 1 & -3 & 2 \\ 0 & -1 & 0 & -2 \end{pmatrix} \implies A^\top = \begin{pmatrix} 4 & 3 & 0 \\ -2 & 1 & -1 \\ 1 & -3 & 0 \\ 0 & 2 & -2 \end{pmatrix}$$

- Diagonalmatrix D bzw. Einheitsmatrix I der Ordnung 3

$$D = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- Obere Dreiecksmatrix B bzw. untere Dreiecksmatrix C der Ordnung 3

$$B = \begin{pmatrix} 4 & 12 & 5 \\ 0 & 0 & -3 \\ 0 & 0 & -7 \end{pmatrix}, \qquad C = \begin{pmatrix} 9 & 0 & 0 \\ 2 & 3 & 0 \\ -3 & 5 & 6 \end{pmatrix}$$

- Symmetrische Matrix A bzw. schiefsymmetrische Matrix Q der Ordnung 3

$$A = \begin{pmatrix} -1 & 0 & -3 \\ 0 & 7 & 4 \\ -3 & 4 & 8 \end{pmatrix}, \qquad Q = \begin{pmatrix} 0 & 1 & -3 \\ -1 & 0 & 4 \\ 3 & -4 & 0 \end{pmatrix}$$

◇

2.1.6 Definition. (Matrizenmultiplikation)

1. Eine reelle Zeilenmatrix $A_{1\bullet} = (a_1, \ldots, a_n)$ darf mit einer reellen Spaltenmatrix gleicher Länge

$$B_{\bullet 1} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

zu einer reellen Zahl c verknüpft werden, und zwar nach der Vorschrift

$$c = A_{1\bullet} . B_{\bullet 1} := \sum_{j=1}^{n} a_j b_j .$$

2. Eine reelle Matrix A vom Typ $m \times n$ und der Darstellung

$$A = \begin{pmatrix} A_{1\bullet} \\ A_{2\bullet} \\ \vdots \\ A_{m\bullet} \end{pmatrix}.$$

darf mit einer reellen Matrix B vom Typ $n \times r$ und der Darstellung

$$B = \begin{pmatrix} B_{\bullet 1} & B_{\bullet 2} & \ldots & B_{\bullet r} \end{pmatrix}.$$

zu einer $(m \times r)$-Matrix C verknüpft werden, und zwar nach der Vorschrift

$$C = (c_{ik}; 1 \leq i \leq m,\ 1 \leq k \leq r) \quad \text{mit} \quad c_{ik} = A_{i\bullet} . B_{\bullet k}.$$

Diese Verknüpfung zwischen zwei Matrizen heisst ***Multiplikation*** und ist nur erklärt, wenn im oben vorgeschriebenen Sinne ihre ***Typen verkettbar*** sind. ◇

2.1.7 Beispiel. Ein Landwirt setzt für vier Haustiergruppen H1, ..., H4 pro Woche drei Futtermischungen F1 , ..., F3 ein, und zwar folgenden Mengen (in Mengeneinheiten ME):

	$F1$	$F2$	$F3$
$H1$	40	10	5
$H2$	20	30	5
$H3$	20	30	10
$H4$	10	30	20

Pro ME kosten die Futtermischungen F1 1.50 CHF, F2 1 CHF und F3 2 CHF. Ermitteln Sie die wöchentlichen Futterkosten c_i für jede Haustiergruppe Hi. Offenbar berechnen sich die Kosten für H1

$$c_1 = (40, 10, 5)\begin{pmatrix} 1.5 \\ 1 \\ 2 \end{pmatrix} = 40 \cdot 1.5 + 10 \cdot 1 + 5 \cdot 2 = 80 \qquad \text{(in CHF).}$$

Die Gesamtübersicht liefert die Rechnung

$$\begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} = \begin{pmatrix} 40 & 10 & 5 \\ 20 & 30 & 5 \\ 20 & 30 & 10 \\ 10 & 30 & 20 \end{pmatrix} \begin{pmatrix} 1.5 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 80 \\ 70 \\ 80 \\ 85 \end{pmatrix}.$$

Die Verkettung der Matrizentypen sieht hier wie folgt aus:

$$(4 \times \underbrace{3) \quad (3}_{(4 \times 1)} \times 1)$$

Der Landwirt prüft die Offerte eines anderen Futtermittelhändlers, der die gleichen Futtermischungen zu anderen Preisen anbietet. Bei diesem Händler kosten F1 1 CHF, F2 1 CHF und F3 3 CHF (jeweils pro ME). Einen übersichtlichen Vergleich der alten Kosten c_i mit den neuen Kosten d_i für jede Haustiergruppe Hi liefert

$$\begin{pmatrix} c_1 & d_1 \\ c_2 & d_2 \\ c_3 & d_3 \\ c_4 & d_4 \end{pmatrix} = \begin{pmatrix} 40 & 10 & 5 \\ 20 & 30 & 5 \\ 20 & 30 & 10 \\ 10 & 30 & 20 \end{pmatrix} \begin{pmatrix} 1.5 & 1 \\ 1 & 1 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} 80 & 65 \\ 70 & 65 \\ 80 & 80 \\ 85 & 100 \end{pmatrix}.$$

Die Verkettung der Matrizentypen sieht hier wie folgt aus:

$$(4 \times \underbrace{3) \quad (3}_{(4 \times 2)} \times 2)$$

$\diamond$

2.1.8 Beispiel. Die Matrizenmultiplikation im oben definierten Sinne ist nicht erlaubt, wenn die Typen nicht verkettbar sind, z.B. für die Zeilenmatrix $A = (0, 1, 5)$ existiert AA nicht. $\diamond$

Wir erinnern daran, dass nach Satz 2.1.3 für Matrizen A, B, C gleichen Typs bezüglich der Addition die Gesetze

$$\begin{aligned}(A+B)+C &= A+(B+C)\\ A+B &= B+A\end{aligned}$$

gelten. Wir werden jetzt weitere Gesetze für das Rechnen mit "rechteckigen" Matrizen (also Zeilenanzahl kann ungleich der Spaltenanzahl sein) zusammenstellen und beweisen.

2.1.9 Satz. (Weitere Rechengesetze für Matrizen)

1. *Es gilt* $(A^\mathsf{T})^\mathsf{T} = A$ *für alle* $A \in M(m, n)$.
2. *Es gilt* $(A + B)^\mathsf{T} = A^\mathsf{T} + B^\mathsf{T}$ *für alle* $A, B \in M(m, n)$.
3. *Es gilt* $(AB)^\mathsf{T} = B^\mathsf{T} A^\mathsf{T}$, *falls* AB *definiert ist.*
4. *Seien die Matrizen* A, B, C *und die Einheitsmatrix* I *jeweils so gewählt, dass die jeweilige Summe bzw. das jeweilige Produkt definiert sind. Dann gilt:*

   4.1. *(AB)C = A(BC)*
   4.2. *A(B+C) = AB + AC*
   4.3. *(A+B)C = AC + BC*
   4.4. *IA = A*
   4.5. *AI = A*

$\diamond$

Beweis: Die Gesetze 1. und 2. folgen sofort aus den Definitionen der Transponierten bzw. der Summe.

Zu 3.: Damit AB definiert ist, muss mit gewissen m, n und r gelten: $A \in M(m, n)$ und $B \in M(n, r)$, also $AB \in M(m, r)$. Dann sind aber $B^\mathsf{T} \in M(r, n)$ und $A^\mathsf{T} \in M(n, m)$, folglich existiert $B^\mathsf{T} A^\mathsf{T} = (d_{ki})$ und liegt in $M(r, m)$. Mit den Zeilen $A_{i\bullet}$ von A und den Spalten $B_{\bullet k}$ von B gilt dann aber für $AB = (c_{ik})$ nach Definition der Multiplikation von Zeilen- und Spaltenmatrizen gleicher Länge, dass

$$c_{ik} = A_{i\bullet} B_{\bullet k} = (B^\mathsf{T})_{k\bullet} (A^\mathsf{T})_{\bullet i} = d_{ki},$$

d.h., $(AB)^\mathsf{T} = B^\mathsf{T} A^\mathsf{T}$, womit 3. bewiesen ist.

Zu 4.: 4.4. und 4.5 sind sofort klar aus der Definition. Ferner: Mit den Zeilen $A_{i\bullet}$ von A sowie den Spalten $B_{\bullet k}$ von B und $C_{\bullet k}$ von C gilt nach Definition der Multiplikation bzw. Addition von Zeilen- und Spaltenmatrizen gleicher Länge

$$A_{i\bullet}(B_{\bullet k}+C_{\bullet k}) = A_{i\bullet}B_{\bullet k} + A_{i\bullet}C_{\bullet k},$$

also gilt 4.2. Gesetz 4.3. beweist man analog. Es bleibt, das Assoziativgesetz der Multiplikation, also 4.1., zu beweisen. Seien $A = (a_{ij}) \in M(m,n)$, $B = (b_{jk}) \in M(n,r)$ und $C = (c_{k\nu}) \in M(r,p)$. Dann gilt

$$AB = (\delta_{ik}) \quad \text{mit} \quad \delta_{ik} = \sum_{j=1}^{n} a_{ij}b_{jk},$$

also für

$$(AB)C = (\beta_{i\nu}) \quad \text{mit} \quad \beta_{i\nu} = \sum_{k=1}^{r} \delta_{ik}c_{k\nu} = \sum_{k=1}^{r}(\sum_{j=1}^{n} a_{ij}b_{jk})c_{k\nu}.$$

Andererseits gilt

$$BC = (\gamma_{j\nu}) \quad \text{mit} \quad \gamma_{j\nu} = \sum_{k=1}^{r} b_{jk}c_{k\nu},$$

folglich

$$A(BC) = (\widetilde{\beta}_{i\nu}) \quad \text{mit} \quad \widetilde{\beta}_{i\nu} = \sum_{j=1}^{n} a_{ij}(\sum_{k=1}^{r} b_{jk}c_{k\nu}).$$

Nach Ausklammern und Vertauschen folgt sofort $\beta_{i\nu} = \widetilde{\beta}_{i\nu}$, was zu zeigen war. □

2.1.10 Bemerkung. Wir erinnern nochmals daran, dass im allgemeinen die Produkte AB und BA nicht gleich sind. Das kann aus drei Gründen der Fall sein:

a) weil AB oder BA gar nicht existieren oder

b) weil zwar AB und BA existieren, aber von unterschiedlichem Typ sind (z.B. für eine $(1 \times n)$-Matrix A und eine $(n \times 1)$-Matrix B mit $n \neq 1$) oder

c) wenn zwar AB und BA existieren und sogar von gleichem Typ sind (d.h., A und B sind dann notwendigerweise quadratische Matrizen gleicher Ordnung), aber dennoch unterschiedliche Matrizen ergeben, vgl. Übung 1.1.7.

◇

2.1.11 Ist A eine quadratische Matrix, so definiert man

$$A^0 = I, \quad A^n = A^{n-1}A \ (n = 1, 2, \ldots),$$

und es gilt $A^{m+n} = A^m A^n$ für $m, n \in \mathbb{N} \cup \{0\}$.

2.1.12 Beispiel. In einer Volkswirtschaft mit n Produktionszweigen kann der Output jedes Zweiges als Produktionsfaktor (Input) im eigenen wie in den anderen Zweigen eingesetzt und *verbraucht* werden. Nehmen wir das Beispiel $n = 3$ und nennen wir die Zweige En (Energieproduktion), Ch (Chemieproduktion) und Ba (Bauwesen). Es sei folgende *Verbrauchsmatrix* gegeben:

$$\begin{pmatrix} \text{En-Verbrauch} \\ \text{Ch-Verbrauch} \\ \text{Ba-Verbrauch} \end{pmatrix} = \begin{pmatrix} 0.2 & 0.4 & 0.3 \\ 0.3 & 0.4 & 0.2 \\ 0.4 & 0.3 & 0.1 \end{pmatrix} \begin{pmatrix} \text{Input für En} \\ \text{Input für Ch} \\ \text{Input für Ba} \end{pmatrix}.$$

So sagt z.B. die erste Zeile der Verbrauchsmatrix, dass der Verbrauch an Energie zur Produktion von 1 Einheit in En 0.2 Einheiten, zur Produktion von 1 Einheit in Ch 0.4 Einheiten sowie zur Produktion von 1 Einheit in Ba 0.3 Einheiten beträgt.

Nennen wir die Verbrauchsmatrix A, den Produktionsvektor (Input) $\underline{p} = (p_1, p_2, p_3)^{\mathsf{T}}$ und den Nachfragevektor $\underline{y} = (y_1, y_2, y_3)^{\mathsf{T}}$, so sucht man bei gegebenem $\underline{y}$ in dem Modell

$$\underline{p} - A\underline{p} = \underline{y}$$

einen geeigneten Input $\underline{p}$, so dass die Nachfrage $\underline{y}$ befriedigt werden kann, wobei berücksichtigt wird, dass ein Teil des Inputs $\underline{p}$, nämlich $A\underline{p}$, während der Produktion verbraucht wird.

Es ist also ein Gleichungssystem in den Variablen p_1, p_2, p_3 zu lösen. Interessant für

$$\underline{p} - A\underline{p} = \underline{y}, \quad \text{d.h.,} \ (I - A)\underline{p} = \underline{y},$$

sind die Fragen, ob dieses System

- überhaupt lösbar ist (Kriterien kennen Sie aus der *Mathematik I*, siehe auch den nächsten Abschnitt) oder

- sogar *eindeutig* lösbar ist (Kriterien kennen Sie ebenfalls aus der *Mathematik I*, siehe auch die folgende Definition) – was auf die Frage nach der Existenz der Inversen $(I - A)^{-1}$ führt – und

- ob (genau) eine *nichtnegative* Lösung p existiert (Bedingungen dafür liefern wir in einem späteren Kapitel).

$\diamond$

2.1.13 Definition. Seien A eine n-reihige (quadratische) Matrix und I die $(n \times n)$-Einheitsmatrix. Falls eine $(n \times n)$-Matrix X existiert, so dass $XA = I$ gilt, dann heisst X ***Inverse*** von A und wird mit A^{-1} symbolisiert. Die Matrix A heisst in diesem Fall *invertierbar*. $\diamond$

2.1.14 Weitere Gesetze für invertierbare Matrizen. Aus der Übung 1.1.7 sind zusätzlich zu den oben für allgemeinere Fälle bewiesenen Gesetzen folgende Eigenschaften für invertierbare n-reihige Matrizen A, B bekannt: A^{-1} und AB sind wieder invertierbar, und es gilt

$$\begin{aligned}
(i)\quad & AA^{-1} = I \quad (= A^{-1}A \;\text{ nach Definition 2.1.13)} \\
(ii)\quad & (A^{-1})^{-1} = A \\
(iii)\quad & (AB)^{-1} = B^{-1}A^{-1}.
\end{aligned}$$

Offenbar ist für jede invertierbare Matrix A und $\lambda \in \mathbb{R} \setminus \{0\}$ auch λA invertierbar mit $(\lambda A)^{-1} = \lambda^{-1}A^{-1}$. Aus (iii) folgt durch vollständige Induktion leicht, dass für invertierbare Matrizen $A_1, A_2, \ldots, A_m \in M(n, n)$ auch das Produkt $A_1A_2 \ldots A_m$ invertierbar ist mit

$$(A_1A_2 \ldots A_m)^{-1} = A_m^{-1} \ldots A_2^{-1}A_1^{-1}.$$

Ferner ist zu jeder invertierbaren Matrix A auch ihre Transponierte $A^\top$ invertierbar, und es gilt

$$(A^\top)^{-1} = (A^{-1})^\top.$$

Zur Begründung sei nur angemerkt, dass $(A^{-1})^\top A^\top = (AA^{-1})^\top = I^\top = I$ nach den schon bekannten Regeln über das Transponieren und Invertieren von Matrizen gilt. $\diamond$

2.1.15 Sind A eine invertierbare n-reihige Matrix und $\underline{b} \in \mathbb{R}^n$, so hat das lineare Gleichungssystem (LGS) $A\underline{x} = \underline{b}$ eine eindeutige Lösung $\widehat{\underline{x}}$ mit

$$\widehat{\underline{x}} = A^{-1}\underline{b}.$$

Der Beweis ist einfach: $\widehat{\underline{x}} = A^{-1}\underline{b}$ eingesetzt in $A\underline{x} = \underline{b}$ ergibt sofort $A\widehat{\underline{x}} = AA^{-1}\underline{b} = \underline{b}$, also ist $\widehat{\underline{x}}$ Lösung. Andererseits: *Jede* Lösung $\underline{x}$ von $A\underline{x} = b$ erfüllt

$$\underline{x} = I\underline{x} = A^{-1}A\underline{x} = A^{-1}\underline{b}.$$

Die rechte Seite ist eindeutig, deshalb ist auch die Lösung des LGS eindeutig.

Auch die Umkehrung gilt in gewissem Sinne für $A \in M(n, n)$: Ist das LGS $A\underline{x} = \underline{b}$ eindeutig lösbar für jede rechte Seite $\underline{b} \in \mathbb{R}^n$, so ist A invertierbar. Man betrachte zur Begründung das Matrizengleichungssystem $AX = I$, das nach Voraussetzung lösbar sein muss. $\diamond$

2.2 Lineare Gleichungssysteme und Rang

2.2.1 Definition. Gegeben sind $m \cdot n + m$ reelle Zahlen a_{ij} und b_i für $i = 1, \ldots, m$, $j = 1, \ldots, n$. Ein System der Form

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \ldots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \ldots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \ldots & + & a_{mn}x_n & = & b_m \end{array} \tag{2.1}$$

heisst ***lineares Gleichungssystem*** (kurz: LGS) in den Variablen $x_1, x_2, \ldots, x_n$.

Die Matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}$$

heisst ***Koeffizientenmatrix*** von (2.1), die Matrix

$$[A, \underline{b}\,] = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{pmatrix}$$

heisst ***erweiterte Koeffizientenmatrix*** von (2.1).

Wenn alle Komponenten des Vektors $\underline{b} = (b_1, b_2, \ldots, b_m)^\mathsf{T}$ gleich 0 sind, heisst das LGS (2.1) ***homogen***, andernfalls ***inhomogen***.

Ein n-Tupel $\underline{x} = (x_1, x_2, \ldots, x_m)^\mathsf{T}$, das dem LGS (2.1) genügt, heisst (spezielle) ***Lösung*** dieses LGS. Falls das LGS mindestens eine Lösung hat, heisst es ***lösbar***. ◇

2.2.2 Mit den im Abschnitt 2.1 und in Definition 2.2.1 eingeführten Symbolen, kann man nun das LGS (2.1) äquivalent in folgenden Kurzversionen schreiben:

$$A\underline{x} = \underline{b}$$

bzw.

$$A_{i\bullet}\,\underline{x} = b_i, \quad i = 1, \ldots, m$$

bzw.

$$\sum_{j=1}^{n} x_j A_{\bullet j} = \underline{b}\ .$$

◇

2.2.3 Definition. Sei $A \in M(m,n)$ eine gegebene Matrix.

Wir sprechen von einer ***elementaren Zeilenoperation in A*** bei

- Vertauschen der i-ten mit der k-ten Zeile von A unter Beibehaltung der restlichen Zeilen von A,
- Multiplikation der i-ten Zeile von A mit einer reellen Zahl $\lambda \neq 0$ unter Beibehaltung der restlichen Zeilen von A,
- Addition eines reellen Vielfachen der i-ten Zeile zur k-ten Zeile unter Beibehaltung der i-ten wie der restlichen Zeilen von A. ◇

2.2.4 Gauss-Elimination. Wir erinnern an den in der Vorlesung *Mathematik II* hergeleiteten Fakt, dass man durch sukzessive Anwendung elementarer Zeilenoperationen nach *endlich vielen Schritten* die erweiterte Koeffizientenmatrix $[A, \underline{b}]$ eines LGS $A\underline{x} = \underline{b}$ mit m Gleichungen in n Variablen auf die folgende - bis auf Spaltenvertauschungen gültige - Form bringen kann:

$$[\widetilde{A}, \widetilde{\underline{b}}] = \begin{pmatrix} 1 & \tilde{a}_{12} & \dots & \tilde{a}_{1r} & \tilde{a}_{1\,r+1} & \dots & \tilde{a}_{1n} & \tilde{b}_1 \\ 0 & 1 & \dots & \tilde{a}_{2r} & \tilde{a}_{2\,r+1} & \dots & \tilde{a}_{2n} & \tilde{b}_2 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & \tilde{a}_{r\,r+1} & \dots & \tilde{a}_{rn} & \tilde{b}_r \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & \tilde{b}_{r+1} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & \tilde{b}_m \end{pmatrix}, \tag{2.2}$$

vorausgesetzt, dass A nicht die Nullmatrix ist. Das LGS zur neuen Koeffizientenmatrix $\widetilde{A}\underline{x} = \widetilde{\underline{b}}$ ist offenbar genau dann lösbar, wenn $\tilde{b}_i = 0$ für alle $i \in \{r+1, \dots, m\}$ gilt. Wir zeigen weiter unten (nochmals), dass elementare Zeilenoperationen die Lösungsmenge eines LGS nicht ändern, woraus dann folgt, dass die LGS $A\underline{x} = \underline{b}$ und $\widetilde{A}\underline{x} = \widetilde{\underline{b}}$ die gleiche Lösungsmenge (eventuell $= \emptyset$) haben. Vgl. §2.5. ◇

Wir erinnern daran, dass die lineare Hülle einer Teilmenge eines gegebenen Vektorraums V ein *Unterraum* von V ist. Die Zeilen einer gegebenen $(m \times n)$-Matrix A können wir als Elemente des Vektorraums $\mathbb{R}^n$, ihre Spalten als Elemente des $\mathbb{R}^m$ auffassen. Das führt zu folgender Definition.

2.2.5 Definition. Bezeichnen $A_{i\bullet}$ die i-te Zeile einer Matrix $A \in M(m,n)$ und $A_{\bullet j}$ die j-te Spalte von A, so heissen

$$\operatorname{lin}\{A_{1\bullet}, A_{2\bullet}, \dots, A_{m\bullet}\} \text{ der } \textbf{\textit{Zeilenraum von A}}$$

und

$$\operatorname{lin}\{A_{\bullet 1}, A_{\bullet 2}, \dots, A_{\bullet n}\} \text{ der } \textbf{\textit{Spaltenraum von A}}.$$

◇

2.2.6 Definition. Sei $A \in M(m,n)$ eine gegebene Matrix. Dann heissen die Dimension des Zeilenraums von A ***Zeilenrang von A*** (Bezeichnung: $z(A)$) und die Dimension des Spaltenraums von A ***Spaltenrang von A*** (Bezeichnung: $s(A)$).

Der Zeilenrang der Matrix A ist also die maximale Anzahl linear unabhängiger Zeilen von A, ihr Spaltenrang ist die maximale Anzahl linear unabhängiger Zeilen von A. $\diamond$

2.2.7 Lemma. *Ist $[A, \underline{b}]$ die erweiterte Koeffizientenmatrix eines LGS $A\underline{x} = \underline{b}$ und ist $[\widehat{A}, \widehat{\underline{b}}]$ aus $[A, \underline{b}]$ durch eine elementare Zeilenoperation hervorgegegangen, so haben die linearen Gleichungssysteme $A\underline{x} = \underline{b}$ und $\widehat{A}\underline{x} = \widehat{\underline{b}}$ die gleiche Lösungsmenge (wobei es sich gegebenenfalls um die leere Menge handeln kann).* $\diamond$

2.2.8 Lemma. *Ist $\widehat{A}$ die aus $A \in M(m,n)$ durch eine elementare Zeilenoperation hervorgegangene Matrix, so sind der Zeilenraum von A und der Zeilenraum von $\widehat{A}$ gleich.* $\diamond$

2.2.9 Satz. *Der Zeilenrang $z(A)$ und der Spaltenrang $s(A)$ einer $(m \times n)-Matrix$ A sind gleich.* $\diamond$

Die Beweise dieses Satzes und der vorangehenden Lemmata werden in der Vorlesung gegeben. Alternative Beweise des wichtigen Satzes 2.2.9 finden sich z.B. in *G.Fischer, Lineare Algebra, Vieweg, 2002*, §1.5, und *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984*, §2.2.

2.2.10 Definition. Sei $A \in M(m,n)$. Der ***Rang von A*** ist $\operatorname{rg} A := s(A) = z(A)$. $\diamond$

2.2.11 Elementare Folgerungen: Aus der Definition des Rangs folgt für $A \in M(m,n)$ sofort

$$\operatorname{rg} A \leq m, \ \operatorname{rg} A \leq n \quad \text{und} \quad \operatorname{rg} A = \operatorname{rg} A^{\mathsf{T}}.$$

$\diamond$

2.2.12 Satz. (Lösbarkeit und Lösungsmenge eines LGS) *Wir betrachten das LGS (2.1) $A\underline{x} = \underline{b}$ mit m Gleichungen in n Unbekannten.*

1. *Das LGS (2.1) $A\underline{x} = \underline{b}$ ist lösbar genau dann, wenn* $\operatorname{rg}[A, \underline{b}] = \operatorname{rg} A$.

2. *Ist das LGS (2.1) $A\underline{x} = \underline{b}$ lösbar und $\underline{x}^0$ eine spezielle Lösung von (2.1), so ist die Lösungsmenge M von (2.1) gegeben durch die algebraische Summe*

$$M = \{\underline{x}^0\} + L \quad \textit{mit } L = \{\underline{u} \in \mathbb{R}^n \mid A\underline{u} = \underline{o}\}.$$

L ist ein Unterraum, und es gilt $\dim L = n - \operatorname{rg} A$. $\diamond$

2.2.13 Beispiel. Machen Sie sich die Definitionen, Aussagen und Beweise (!) dieses Abschnitts am Beispiel des folgenden LGS klar:

$$\begin{array}{rcrcrcr} & & x_2 & - & x_3 & = & 1 \\ x_1 & - & 2x_2 & + & x_3 & = & -1 \\ x_1 & + & x_2 & - & 2x_3 & = & 2\ . \end{array}$$

2.2.14 Korollar. *Wir betrachten wieder das LGS (2.1) $A\underline{x} = \underline{b}$ mit m Gleichungen in n Unbekannten.*

(i) Das LGS (2.1) $A\underline{x} = \underline{b}$ hat genau eine Lösung dann und nur dann, wenn $\operatorname{rg}[A, \underline{b}] = \operatorname{rg} A$ und $\operatorname{rg} A = n$.

(ii) Es gilt $\operatorname{rg} A = m$ dann und nur dann, wenn das LGS (2.1) $A\underline{x} = \underline{b}$ für jede rechte Seite $\underline{b} \in \mathbb{R}^m$ lösbar ist. ◇

Beweis: Der Beweis von (i) ergibt sich sofort aus Aussage 1 von Satz 2.2.12 und der Dimensionsbeziehung in Aussage 2 des gleichen Satzes.

Zum Beweis von Aussage (ii): Falls $\operatorname{rg} A = m$ und $\underline{b} \in \mathbb{R}^m$ beliebig, so

$$m = \operatorname{rg} A \leq \operatorname{rg}[A, \underline{b}] \leq \text{Zeilenanzahl von } \operatorname{rg}[A, \underline{b}] = m,$$

also $\operatorname{rg} A = \operatorname{rg}[A, \underline{b}]$, d.h., nach Satz 2.2.12 ist $A\underline{x} = \underline{b}$ lösbar.

Umgekehrt, wenn $A\underline{x} = \underline{b}$ für alle $\underline{b} \in \mathbb{R}^m$ lösbar ist, so bedeutet das, dass $\mathbb{R}^m$ eine Teilmenge des Spaltenraums $L = \operatorname{lin}\{A_{\bullet 1}, \ldots, A_{\bullet n}\}$ von A ist und folglich mit L zusammenfällt. Da nach Definition $\operatorname{rg} A = \dim L$ ist, folgt sofort $\operatorname{rg} A = m$. □

Im Ergebnis des bisher Gezeigten folgt

2.2.15 Satz. *Sei A eine quadratische Matrix A der Ordnung n. A ist genau dann invertierbar, wenn $\operatorname{rg} A = n$. Folglich gilt $\operatorname{rg} A = n$ genau dann, wenn das entsprechende inhomogene LGS $A\underline{x} = \underline{b}$ mit n Gleichungen in n Unbekannten für jedes $\underline{b} \in \mathbb{R}^n$ eindeutig lösbar ist.* ◇

2.2.16 Definition. Eine quadratische Matrix A der Ordnung n heisst ***regulär*** oder ***nichtsingulär***, falls $\operatorname{rg} A = n$. ◇

2.2.17 Bemerkung: Es folgt sofort, dass invertierbare n-reihige Matrizen regulär und umgekehrt reguläre n-reihige Matrizen invertierbar sind.

Für reguläre n-reihige Matrizen gelten also insbesondere – neben allen anderen Rechenregeln für die Addition und Multiplikation von Matrizen – die unter Punkt 2.1.14 zusammengefassten Gesetze. ◇

Anhang: Beweis von Satz 2.2.9

Zunächst geben wir die Beweise zu den Lemmata 2.2.7 und 2.2.8.

Für die elementaren Zeilenoperationen *Zeilenvertauschung* und *Multiplikation einer Zeile mit* $\lambda \neq 0$ ist es offensichtlich, dass sich weder die Lösungsmenge des LGS noch der Zeilenraum ändern.

Nehmen wir nun die elementare Zeilenoperation *Addition des* λ*-fachen der* i*-ten Zeile zur* k*-ten Zeile unter Beibehaltung der Zeilen mit der Nummer* $l \neq k$. Offenbar löst der Spaltenvektor $\underline{x}$ das System

$$\begin{array}{rcll} A_{l\bullet}\underline{x} &=& b_l, & \quad l \in \{1,\ldots,m\} \setminus \{k\}, \\ (A_{k\bullet} + \lambda A_{i\bullet})\underline{x} &=& b_k + \lambda b_i & \end{array}$$

genau dann, wenn $\underline{x}$ (man nutze $A_{i\bullet}\underline{x} = b_i$ aus) das System

$$\begin{array}{rcll} A_{l\bullet}\underline{x} &=& b_l, & \quad l \in \{1,\ldots,m\} \setminus \{k\}, \\ A_{k\bullet}\underline{x} &=& b_k & \end{array}$$

löst. Damit ist Lemma 2.2.7 gezeigt. Ferner gilt für den Zeilenraum

$$Z = \mathrm{lin}\,\{A_{1\bullet},\ldots,A_{m\bullet}\},$$

dass $\underline{x}$ dann und nur dann zu Z gehört, falls Zahlen $\lambda_1,\ldots,\lambda_m \in \mathbb{R}$ existieren, so dass

$$\begin{aligned} \underline{x} &= \lambda_k A_{k\bullet} + \sum_{l=1,\, l\neq k}^{m} \lambda_l A_{l\bullet} \\ &= \lambda_k (A_{k\bullet} + \lambda A_{i\bullet}) + (\lambda_i - \lambda_k \lambda) A_{i\bullet} + \sum_{l=1,\, l\neq k,\, l\neq i}^{m} \lambda_l A_{l\bullet}, \end{aligned}$$

also $\underline{x} \in \mathrm{lin}\,\{A_{1\bullet},\ldots,A_{i\bullet},\ldots,A_{k\bullet} + \lambda A_{i\bullet},\ldots,A_{m\bullet}\}$, gilt. Damit ist

$$Z = \mathrm{lin}\,\{A_{1\bullet},\ldots,A_{i\bullet},\ldots,A_{k\bullet} + \lambda A_{i\bullet},\ldots,A_{m\bullet}\},$$

erhalten worden, was für Lemma 2.2.8 zu zeigen war.

Kommen wir nun zum eigentlichen Beweis von Satz 2.2.9. Die Matrix

$$\widetilde{A} \text{ aus } (2.2)$$

kann nach dem Gauss-Algorithmus aus A nach endlich vielen elementaren Zeilenoperationen und gegebenenfalls endlich vielen Spaltenvertauschungen erzeugt werden.

Wir nehmen o.B.d.A. an, dass keine Spaltenvertauschungen notwendig waren. Andernfalls wird der Gauss-Algorithmus in der bekannten Weise modifiziert, dass statt der Form (2.2) die sogenannte Zeilenstufenform entsteht, für die die folgenden Überlegungen mit einer anderen Variablennummerierung analog gelten.

Ferner sei o.B.d.A. die Matrix $\widetilde{A}$ bereits in der folgenden Form gegeben, die man aus (2.2) nach endlich vielen weiteren elementaren Zeilenoperationen und dem Streichen der letzten $m-r$ Zeilen (Nullzeilen!) erhält:

$$\widetilde{A} = \begin{pmatrix} 1 & 0 & \dots & 0 & \tilde{a}_{1\,r+1} & \dots & \tilde{a}_{1n} \\ 0 & 1 & \dots & 0 & \tilde{a}_{2\,r+1} & \dots & \tilde{a}_{2n} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 & \tilde{a}_{r\,r+1} & \dots & \tilde{a}_{rn} \end{pmatrix},$$

Nach einer endlichen Anzahl von Anwendungen der beiden soeben bewiesenen Lemmata folgt sofort:

$$\underline{x} \text{ löst } A\underline{x} = \underline{o} \;\Leftrightarrow\; \underline{x} \text{ löst } \widetilde{A}\underline{x} = \underline{o}$$

sowie

$$\operatorname{lin}\{A_{1\bullet}, \dots, A_{m\bullet}\} = \operatorname{lin}\{\widetilde{A}_{1\bullet}, \dots, \widetilde{A}_{r\bullet}\} \subset \mathbb{R}^n,$$

wobei $\widetilde{A}_{i\bullet}$ die i-te Zeile von $\widetilde{A}$ bezeichnet. Also folgt

$$z(A) = z(\widetilde{A}) = r. \tag{2.3}$$

Seien jetzt $\underline{e}^i$, $i = 1, \dots, n-r$, die Einheitsvektoren in $\mathbb{R}^{n-r}$. Die Lösungsmenge L von $A\underline{x} = \underline{o}$ bzw. (äquivalent) $\widetilde{A}\underline{x} = \underline{o}$ ergibt sich mit der Definition der Spaltenvektoren (das sind die bekannten *Fundamentallösungen* aus der allgemeinen Lösung des LGS)

$$\widehat{A}_{\bullet k} = \begin{pmatrix} -\widetilde{A}_{\bullet k} \\ \underline{e}^{k-r} \end{pmatrix} \in \mathbb{R}^n, \quad k = r+1, \dots, n,$$

(d.h., an den Spaltenvektor $-\widetilde{A}_{\bullet k}$ ist unten der Spaltenvektor $\underline{e}^{k-r}$ angehängt worden) als

$$L = \{\underline{x} \,|\, A\underline{x} = \underline{o}\} = \operatorname{lin}\{\widehat{A}_{\bullet r+1}, \dots, \widehat{A}_{\bullet n}\}.$$

Die Menge $\mathcal{B} = \{\widehat{A}_{\bullet r+1}, \dots, \widehat{A}_{\bullet n}\} \subset \mathbb{R}^n$ ist also ein Erzeugendensystem von L und nach Definition der Vektoren $\widehat{A}_{\bullet k}$ linear unabhängig. Also ist $\mathcal{B}$ eine Basis von L. Nach dem Basisergänzungssatz lässt sich diese zu einer Basis $\mathcal{A}$ des $\mathbb{R}^n$ ergänzen:

$$\mathcal{A} = \{\underline{a}^1, \dots, \underline{a}^n\} \text{ mit } \underline{a}^k = \widehat{A}_{\bullet k} \text{ und } A\underline{a}^k = \underline{o} \text{ für } k = r+1, \dots, n. \tag{2.4}$$

Ist $S = \operatorname{lin}\{A_{\bullet 1}, \dots, A_{\bullet n}\} \subset \mathbb{R}^m$ der Spaltenraum von A, so gilt

$$\begin{aligned} \underline{b} \in S \;&\Leftrightarrow\; \underline{b} = A\underline{x} \text{ für ein } \underline{x} \in \mathbb{R}^n \\ &\Leftrightarrow\; \underline{b} = A(\sum_{j=1}^{n} \xi_j \underline{a}^j) \text{ für gewisse } \xi_j, \text{ weil } \mathcal{A} \text{ Basis des } \mathbb{R}^n \\ &\Leftrightarrow\; \underline{b} \in \operatorname{lin}\{A\underline{a}^1, \dots, A\underline{a}^r\} \text{ wegen (2.4).} \end{aligned}$$

Also folgt $s(A) = \dim S \le r$, d.h., $s(A) \le z(A)$ gemäss (2.3). Wenden wir nun die gleichen Argumente (statt auf A) auf die Matrix A^T und das LGS $A^\mathsf{T}\underline{y} = \underline{o}$ an, so folgt $s(A^\mathsf{T}) \le z(A^\mathsf{T})$ und somit auch $s(A) \le z(A) = s(A^\mathsf{T}) \le z(A^\mathsf{T}) = s(A)$. Daraus folgt $s(A) = z(A)$, was zu zeigen war.

2.3 Lineare Abbildungen

2.3.1 Einführung. Betrachten wir das Beispiel 2.1.7. Dort haben wir zwei Preisvektoren $\underline{\pi} = (\pi_1, \pi_2, \pi_3)^\mathsf{T}$ zu Futtermischungen F1, F2 und F3 mittels einer Futterplan-Matrix - nennen wir sie P - , die die an 4 Haustiergruppen H1, ..., H4 pro Woche verfütterten Mengen enthält, zwei Vektoren $\underline{c} = (c_1, c_2, c_3, c_4)^\mathsf{T}$ der wöchentlichen Futterkosten zugeordnet.

Für den ersten Preisvektor sah das so aus

$$\begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{pmatrix} = \begin{pmatrix} 1.5 \\ 1 \\ 2 \end{pmatrix} \longmapsto \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} = \begin{pmatrix} 40 & 10 & 5 \\ 20 & 30 & 5 \\ 20 & 30 & 10 \\ 10 & 30 & 20 \end{pmatrix} \begin{pmatrix} 1.5 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 80 \\ 70 \\ 80 \\ 85 \end{pmatrix},$$

für den zweiten Preisvektor ergab sich

$$\begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} \longmapsto \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} = \begin{pmatrix} 40 & 10 & 5 \\ 20 & 30 & 5 \\ 20 & 30 & 10 \\ 10 & 30 & 20 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 65 \\ 65 \\ 80 \\ 100 \end{pmatrix}.$$

Verallgemeinern wir diese Zuordnung und abstrahieren wir davon, dass Preise und Kosten in diesem praktischen Beispiel sicher nichtnegativ sein sollten, ergibt sich folgende *Abbildung*

$$\underline{\pi} \in \mathbb{R}^3 \longmapsto \underline{c} = P\underline{\pi} \in \mathbb{R}^4.$$

Offenbar folgt im Beispiel für einen "Mischpreis" die Zuordnung proportional gemischter Kosten, und zwar,

$$\begin{aligned} &\tfrac{1}{2}(1.5, 1, 2) + \tfrac{1}{2}(1, 1, 3) = (1.25, 1, 2.5) \\ \mapsto\ &\tfrac{1}{2}(80, 70, 80, 85) + \tfrac{1}{2}(65, 65, 80, 100) = (72.5, 67.5, 80, 92.5). \end{aligned}$$

Das könnten Sie nach Multiplikation von P mit dem "Mischpreisvektor" leicht ausrechnen. Aber das ist gar nicht nötig. Allgemein ergibt sich nämlich für beliebige reelle Zahlen λ, μ nach den Rechengesetzen für Matrizen und Vektoren, dass

$$\lambda\underline{\pi}^1 + \mu\underline{\pi}^2 \longmapsto P(\lambda\underline{\pi}^1 + \mu\underline{\pi}^2) = \lambda P\underline{\pi}^1 + \mu P\underline{\pi}^2.$$

Eine Abbildung mit dieser Eigenschaft nennt man linear. $\diamond$

2.3.2 Definition. Gegeben seien zwei Vektorräume V und W sowie eine Abbildung φ von V in W. Die Abbildung φ heisst ***linear***, wenn für beliebige $\underline{v}^1, \underline{v}^2 \in V$ und $\lambda_1, \lambda_2 \in \mathbb{R}$ stets

$$\varphi(\lambda_1\underline{v}^1 + \lambda_2\underline{v}^2) = \lambda_1\varphi(\underline{v}^1) + \lambda_2\varphi(\underline{v}^2)$$

gilt. Eine lineare Abbildung wird in der Literatur auch als *lineare Transformation* oder *Homomorphismus* bezeichnet. Ist $W = V$, so spricht man von einem *Endomorphismus*. $\diamond$

2.3.3 Vereinbarung. Wenn wir zukünftig kurz von der *linearen Abbildung* $\varphi : V \to W$ sprechen, ist naturgemäss immer eingeschlossen, dass V und W Vektorräume sind. ◇

2.3.4 Beispiel. Ist A eine reelle $(m \times n)$-Matrix, dann definiert

$$\underline{x} \in \mathbb{R}^n \mapsto \varphi(\underline{x}) := A\underline{x} \in \mathbb{R}^m$$

eine lineare Abbildung von $\mathbb{R}^n$ in $\mathbb{R}^m$. ◇

2.3.5 Beispiel. (Differenzieren als lineare Abbildung) Sei $x \mapsto f(x)$ irgendeine stetig differenzierbare Funktion auf dem Intervall $[a, b]$. Durch Differenzieren entsteht die auf $[a, b]$ stetige Funktion $x \mapsto f'(x)$ (das ist die 1. Ableitung von f). Dabei gilt die übliche Verabredung, dass auf den Randpunkten von (a, b) jeweils die einseitigen Ableitungen bzw. Stetigkeiten zu betrachten sind.

Also können wir f und f' auffassen als

$$f \in C^1[a, b] \qquad \text{und} \qquad f' \in C[a, b].$$

Die Abbildung $\varphi : C^1[a, b] \to C[a, b]$, die durch

$$f \in C^1[a, b] \mapsto \varphi(f) = f' \in C[a, b]$$

definiert wird, ist offenbar eine *lineare Abbildung* vom Vektorraum $C^1[a, b]$ in den Vektorraum $C[a, b]$.

Begründung: Für $f_1 \in C^1[a, b]$ und $f_2 \in C^1[a, b]$ sowie $\lambda_1, \lambda_2 \in \mathbb{R}$ gilt bekanntlich

$$(\lambda_1 f_1 + \lambda_2 f_2)'(x) = \lambda_1 f_1'(x) + \lambda_2 f_2'(x) \quad \forall x \in [a, b],$$

also $\varphi(\lambda_1 f_1 + \lambda_2 f_2) = \lambda_1 \varphi(f_1) + \lambda_2 \varphi(f_2)$. ◇

2.3.6 Beispiel. (Integrieren als lineare Abbildung) Sei nun $x \mapsto f(x)$ irgendeine stetige Funktion auf dem Intervall $[a, b]$. Nach dem Hauptsatz der Differential- und Integralrechnung ist dann die Funktion

$$F(x) := \int_a^x f(t)dt\,, \quad x \in [a, b],$$

auf (a, b) differenzierbar (mit $F'(x) = f(x)$ für $x \in (a, b)$), und in den Randpunkten gilt $f(a) = F'_+(a)$ und $f(b) = F'_-(b)$. Also ist F auf jeden Fall wieder stetig auf $[a, b]$. F ist eine sogenannte Stammfunktion.

Ähnlich wie im vorangehenden Beispiel haben wir also eine Abbildung

$$f \in C[a, b] \mapsto \psi(f) = F \in C[a, b].$$

Nach den Gesetzen der Integration gilt

$$(\lambda F + \mu G)(x) = \lambda F(x) + \mu G(x), \quad x \in [a,b],$$

wenn F wie oben aus $f \in C[a,b]$ und G analog aus $g \in C[a,b]$ gebildet werden. Damit ist ψ eine lineare Abbildung vom Vektorraum $C[a,b]$ in sich selbst. $\diamond$

2.3.7 Elementare Eigenschaften linearer Abbildungen. Seien V und W Vektorräume mit den Nullvektoren $\underline{o}_V$ bzw. $\underline{o}_W$ und $\varphi : V \to W$ linear. Dann gilt:

(i) $\varphi(\underline{o}_V) = \underline{o}_W$.

(ii) $\varphi(\underline{v}^1 - \underline{v}^2) = \varphi(\underline{v}^1) - \varphi(\underline{v}^2)$ für $\underline{v}^1, \underline{v}^2 \in V$.

(iii) $\varphi(\sum_{i=1}^m \lambda_i \underline{v}^i) = \sum_{i=1}^m \lambda_i \varphi(\underline{v}^i)$ für $\underline{v}^i \in V$, $\lambda_i \in \mathbb{R}$ $(i = 1, ..., m)$.

(iv) Wenn $\{\underline{v}^1, \ldots, \underline{v}^m\} \subset V$ linear abhängig ist, so ist auch $\{\varphi(\underline{v}^1), \ldots, \varphi(\underline{v}^m)\} \subset W$ linear abhängig.

(v) Wenn $\{\varphi(\underline{v}^1), \ldots, \varphi(\underline{v}^m)\} \subset W$ linear unabhängig ist, so ist auch $\{\underline{v}^1, \ldots, \underline{v}^m\} \subset V$ linear unabhängig.

Beweis:

(i) folgt wegen $\varphi(\underline{o}_V) = \varphi(0\, \underline{o}_V) = 0\, \varphi(\underline{o}_V) = \underline{o}_W$ nach den Gesetzen für Vektorräume und lineare Abbildungen.

(ii) folgt wegen $\varphi(\underline{v}^1 - \underline{v}^2) = \varphi(\underline{v}^1 + (-1)\underline{v}^2) = \varphi(\underline{v}^1) + (-1)\varphi(\underline{v}^2) = \varphi(\underline{v}^1) - \varphi(\underline{v}^2)$ nach den Gesetzen für Vektorräume und lineare Abbildungen.

(iii) folgt nach wiederholtem Anwenden der Definition einer linearen Abbildung.

(iv) Die Beziehung

$$(*) \qquad \sum_{i=1}^m \lambda_i \underline{v}^i = \underline{o}_V \ \text{ für } \ \{\underline{v}^1, \ldots, \underline{v}^m\} \in V$$

impliziert – nach (iii) und (i) –

$$(**) \qquad \sum_{i=1}^m \lambda_i \varphi(\underline{v}^i) = \varphi(\sum_{i=1}^m \lambda_i \underline{v}^i) = \underline{o}_W.$$

Ist $(*)$ eine nichttriviale Linearkombination, so auch $(**)$, woraus (iv) folgt.

(v) ist die Kontraposition von (iv). $\square$

Bemerkung. Wir haben hier der Klarheit wegen die Nullvektoren $\underline{o}_V$ von V und $\underline{o}_W$ von W unterschiedlich bezeichnet. Wir wollen zukünftig in der Regel wieder einheitlich das Symbol $\underline{o}$ verwenden. Beachten Sie aber immer sorgfältig, aus welchem Vektorraum $\underline{o}$ genommen ist. Bei der linearen Abbildung $\varphi : \mathbb{R}^2 \to \mathbb{R}^1$ mittels $\varphi(x_1, x_2) = x_1 + x_2$ zum Beispiel wird der Nullvektor $(0,0)^\mathsf{T}$ des Vektorraums $\mathbb{R}^2$ auf den Nullvektor des Vektorraums $\mathbb{R}$, das ist die reelle Zahl 0, abgebildet. $\diamond$

2.3.8 Definition. Sind φ und ψ lineare Abbildungen von V in W und $\lambda \in \mathbb{R}$, so definieren wir die Abbildungen $(\lambda\varphi)$ und $(\varphi + \psi)$ mittels

$$(\lambda\varphi)(\underline{v}) := \lambda\varphi(\underline{v}) \quad \text{und} \quad (\varphi+\psi)(\underline{v}) := \varphi(\underline{v}) + \psi(\underline{v}) \quad \forall \underline{v} \in V.$$

Sind $\varphi : V \to W$ und $\varrho : W \to X$ lineare Abbildungen, dann ist die Hintereinanderausführung $(\varrho \circ \varphi) : V \to X$ definiert durch $(\varrho \circ \varphi)(\underline{v}) := \varrho(\varphi(\underline{v})) \quad \forall \underline{v} \in V$. $\diamond$

2.3.9 Übung. Die eben definierten Abbildungen sind wieder linear. $\diamond$

2.3.10 Definition. Sei $\varphi : V \to W$ eine lineare Abbildung. Dann heissen

$$\varphi(V) := \{\underline{w} \in W \mid \exists \underline{v} \in V : \ \varphi(\underline{v}) = \underline{w}\}$$

das ***Bild von V unter der Abbildung*** φ (oder auch kurz ***Bild der Abbildung*** φ),

$$\operatorname{kern}\varphi := \{\underline{v} \in V \mid \varphi(\underline{v}) = \underline{o}\}$$

der ***Kern der Abbildung*** φ sowie für gegebenes $\underline{w} \in W$ die Gleichung

$$\varphi(\underline{v}) = \underline{w}$$

lineare Gleichung (in der Variablen $\underline{v} \in V$). $\diamond$

Aus dem englischsprachigen Raum sind für das Bild der Abbildung φ auch die Bezeichnung $\operatorname{Im}\varphi$ (*image*) und für den Kern von φ auch die Bezeichnung $\operatorname{Ker}\varphi$ (*kernel*) geläufig. Die Lösungsmenge der linearen Gleichung $\varphi(\underline{v}) = \underline{w}$ heisst auch *Faser über* $\underline{w} \in W$.

2.3.11 Satz. *Seien $\varphi : V \to W$ eine lineare Abbildung und $\underline{w} \in W$ gegeben. Für die Menge*

$$M = \{\underline{v} \in V \mid \varphi(\underline{v}) = \underline{w}\},$$

d.h., die Lösungsmenge der linearen Gleichung $\varphi(\underline{v}) = \underline{w}$, gilt

(i) M ist nichtleer genau dann, wenn $\underline{w} \in \varphi(V)$.

(ii) Ist $\underline{v}^0$ irgendeine Lösung der linearen Gleichung $\varphi(\underline{v}) = \underline{w}$, so ist M als algebraische Summe $M = \{\underline{v}^0\} + \operatorname{kern}\varphi$ darstellbar. $\diamond$

Bemerkung zu Satz 2.3.11. Für eine gegebene $(m \times n)$-Matrix A und die lineare Abbildung $\underline{x} \in \mathbb{R}^n \mapsto A\underline{x} \in \mathbb{R}^m$ fällt die Aussage von (ii) offenbar mit Aussage 2 in Satz 2.2.12 zusammen.

Beweis von Satz 2.3.11. Aussage (i) ergibt sich sofort aus der Definition von $\varphi(V)$. Zur Aussage (ii): Sei $\underline{v}^0 \in M$ gegeben. Für jedes $\underline{u} \in \operatorname{kern}\varphi$ gilt $\varphi(\underline{v}^0 + \underline{u}) = \varphi(\underline{v}^0) + \varphi(\underline{u}) = \underline{w} + \underline{o} = \underline{w}$, d.h., $\underline{v}^0 + \underline{u} \in M$. Ist umgekehrt $\underline{v} \in M$ beliebig, so gilt mit der Definition $\underline{u} := \underline{v} - \underline{v}^0$ die Beziehung $\varphi(\underline{u}) = \varphi(\underline{v}) - \varphi(\underline{v}^0) = \underline{w} - \underline{w} = \underline{o}$, also $\underline{u} \in \operatorname{kern}\varphi$ und somit $\underline{v} \in \{\underline{v}^0\} + \operatorname{kern}\varphi$. □

2.3.12 Satz. *Sei $\varphi : V \to W$ eine lineare Abbildung. Dann ist $\varphi(V)$ ein Unterraum von W, und es ist* $\operatorname{kern}\varphi$ *ein Unterraum von V.* ◇

Beweis: Offenbar ist der Nullvektor von V in $\operatorname{kern}\varphi$ enthalten, und es ist der Nullvektor von W als Bild des Nullvektors von V unter der linearen Abbildung φ in $\varphi(V)$ enthalten. Es muss nur gezeigt werden, dass die Linearkombinationen von je zwei Elementen in $\varphi(V)$ bzw. $\operatorname{kern}\varphi$ wieder in $\varphi(V)$ bzw. $\operatorname{kern}\varphi$ liegen. Dieser Beweis sei dem Leser als Übungsaufgabe überlassen. □

2.3.13 Definition. Seien V und W Vektorräume und $\varphi : V \to W$.

- Gilt für $\underline{v}^1, \underline{v}^2 \in V$ mit $\underline{v}^1 \neq \underline{v}^2$ stets $\varphi(\underline{v}^1) \neq \varphi(\underline{v}^2)$, so heisst φ ***injektiv***. Ist φ linear und injektiv, so wird φ auch ***regulär*** bzw. ***nichtsingulär*** genannt.
- Gilt $W = \varphi(V)$ so heisst φ ***surjektiv*** oder ***Abbildung auf W*** .
- Ist φ injektiv und surjektiv, so heisst φ ***bijektiv***. Ist φ linear und bijektiv, so heisst φ auch ***Isomorphismus*** zwischen V und W; man sagt dann auch, die Vektorräume V und W seien zueinander ***isomorph***.
- Ist φ eine injektive Abbildung von V in W, so ist φ eine bijektive Abbildung zwischen den Vektorräumen V und $\varphi(V)$ (vgl. Satz 2.3.12), zu der eine inverse Abbildung existiert; man bezeichnet diese wie üblich mit φ^{-1} und nennt sie ***Inverse*** von φ. ◇

2.3.14 Satz. *Eine lineare Abbildung $\varphi : V \to W$ ist injektiv dann und nur dann, wenn* $\operatorname{kern}\varphi = \{\underline{o}\}$. ◇

Beweis: Ist φ injektiv, so gilt nach Definition der Injektivität, dass für jedes $\underline{v} \in V \setminus \{\underline{o}\}$

$$\varphi(\underline{v}) \neq \varphi(\underline{o}) = \underline{o}$$

gilt, also $\operatorname{kern}\varphi = \{\underline{o}\}$ erfüllt ist. Umgekehrt gilt

$$\operatorname{kern}\varphi = \{\underline{o}\} \quad \Rightarrow \quad \varphi(\underline{v}) \neq \underline{o} \ \forall \underline{v} \in V \setminus \{\underline{o}\}.$$

Folglich gilt für alle $\underline{v}^1, \underline{v}^2 \in V$ mit $\underline{v}^1 \neq \underline{v}^2$ auch $\varphi(\underline{v}^2) - \varphi(\underline{v}^1) = \varphi(\underline{v}^2 - \underline{v}^1) \neq \underline{o}$ wegen $\underline{v}^2 - \underline{v}^1 \neq \underline{o}$, was die Rückrichtung beweist. □

2.3.15 Satz. *Sei $\varphi : V \to W$ eine injektive lineare Abbildung mit der Inversen φ^{-1}. Dann ist $\varphi^{-1} : \varphi(V) \subset W \to V$ wiederum eine injektive lineare Abbildung.* ◇

Der Beweis wird in der Vorlesung gegeben, vgl. auch *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984*, Satz 2.3.

2.3.16 Bemerkung. Sind $\{\underline{v}^1, \ldots, \underline{v}^m\} \subset V$ linear unabhängig und ist $\varphi : V \to W$ eine injektive lineare Abbildung, so ist auch die Menge $\{\varphi(\underline{v}^1), \ldots, \varphi(\underline{v}^m)\} \subset \varphi(V)$ linear unabhängig. Zur Begründung wende man nur die Aussage (v) in 2.3.7 auf die lineare Abbildung φ^{-1} an.

2.3.17 Satz. *Sei $\varphi : V \to W$ eine lineare Abbildung, und setzen wir voraus, dass V die (endliche) Dimension n hat. Dann ist $\varphi(V)$ endlichdimensional, und es gilt:*

(i) Sind

$$\text{Basen } \{\underline{v}^1, \ldots, \underline{v}^k\} \text{ von } \operatorname{kern}\varphi \text{ sowie } \{\underline{w}^1, \ldots, \underline{w}^r\} \text{ von } \varphi(V)$$

gegeben und sind $\underline{u}^1, \ldots, \underline{u}^r$ Elemente von V mit

$$\varphi(\underline{u}^i) = \underline{w}^i, \quad i = 1, \ldots, r,$$

dann ist $k = n - r$ und

$$\mathcal{B} = \{\underline{u}^1, \ldots, \underline{u}^r, \underline{v}^1, \ldots, \underline{v}^k\} \text{ eine Basis von } V.$$

(ii) Es gilt die Dimensionsformel

$$\dim V \;=\; \dim \operatorname{kern}\varphi \;+\; \dim \varphi(V),$$

insbesondere also $\dim \varphi(V) \le \dim V$.

(iii) φ ist injektiv genau dann, wenn $\dim \varphi(V) = n$.

(iv) Ist $\varphi : V \to W$ bijektiv, so ist W auch endlichdimensional, und es gilt

$$\dim V = \dim W.$$

◇

Der Beweis wird in der Vorlesung gegeben, vgl. auch *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984*, §2.1 oder *G.Fischer, Lineare Algebra, Vieweg, 2002*, §2.2. Offenbar sind dabei (iii) und (iv) einfache Folgerungen aus der Dimensionsformel (ii) und Satz 2.3.14.

2.3.18 Beispiel. Wir betrachten die lineare Abbildung $\varphi : \mathbb{R}^n \to \mathbb{R}^m$, die mit Hilfe einer $(m \times n)$-Matrix A durch

$$\varphi(\underline{x}) = A\underline{x}, \quad \underline{x} \in \mathbb{R}^n,$$

definiert ist. Die folgenden Aussagen ergeben sich aus den vorhergehenden Definitionen und Sätzen.

Dann ist der Kern von φ (d.h., $\ker \varphi \subset \mathbb{R}^n$) die Lösungsmenge L des LGS $A\underline{x} = \underline{o}$. Die Abbildung φ ist dann und nur dann injektiv, wenn das LGS nur die triviale Lösung hat.

Das Bild von φ (d.h., $\varphi(\mathbb{R}^n) \subset \mathbb{R}^m$) ist der Spaltenraum von A. Seine Dimension ist immer kleiner oder gleich m (wegen $\varphi(\mathbb{R}^n) \subset \mathbb{R}^m$), aber auch kleiner gleich n (nach (ii) im vorhergehenden Satz). Die Abbildung φ ist genau dann injektiv, wenn $\dim \varphi(\mathbb{R}^n) = n$ ist. Also kann φ nur injektiv sein, wenn $m \geq n$ ist.

Vereinigt man eine Basis $\{\underline{v}^1, \ldots, \underline{v}^k\} \subset \mathbb{R}^n$ der Lösungsmenge L des LGS $A\underline{x} = \underline{o}$ (d.i. der Kern von φ) mit je einer speziellen Lösung $\underline{u}^i \in \mathbb{R}^n$ der inhomogenen linearen Gleichungssysteme $A\underline{u} = \underline{w}^i$ $(i = 1, ..., r)$ bezüglich einer Basis $\{\underline{w}^1, \ldots, \underline{w}^r\}$ des Spaltenraums von A (= Bild von φ), dann erhält man eine Basis $\{\underline{v}^1, \ldots, \underline{v}^k, \underline{u}^1, \ldots, \underline{u}^r\}$ des $\mathbb{R}^n$. Also erhalten wir die aus Satz 2.2.12 bekannte Formel

$$\dim L = n - \operatorname{rg} A,$$

denn $k = \dim L$ (siehe oben) und $r = \operatorname{rg} A$ ist die Dimension des Spaltenraums (wie des Zeilenraums) von A. $\diamond$

2.3.19 Übung. (vgl. *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984*, Beispiel 2.2 b)

Seien $C[a,b]$ und $C^1[a,b]$ - wie oben definiert - die Vektorräume der auf dem Intervall $[a,b]$ stetigen bzw. stetig differenzierbaren reellwertigen Funktionen mit der Nullfunktion o. (Da es sich um reellwertige Funktionen handelt, benutzen wir nicht unterstrichene Symbole für die Vektoren aus diesen Räumen.)

Mit $y \in C^1[a,b]$ gilt $y' \in C[a,b]$, also ist zu einer festen Zahl $\alpha \in \mathbb{R}$ mittels

$$\varrho(y) = y' + \alpha y$$

eine lineare Abbildung von $C^1[a,b]$ in $C[a,b]$ definiert.

Wir betrachten nun folgende Aufgabe: Gesucht sind alle $y \in C^1[a,b]$, so dass $\varrho(y) = o$, d.h., so dass

$$y'(x) + \alpha y(x) \equiv 0 \qquad \forall x \in [a,b]. \tag{2.5}$$

Diese Gleichung heisst *homogene lineare Differentialgleichung mit konstanten Koeffizienten* und hat z.B. auch Anwendungen bei Wachstumsprozessen (Kapital, Bevölkerung etc.).

Durch Einsetzen stellt man fest, dass die Funktionen

$$y_\gamma(x) = \gamma e^{-\alpha x}, \quad x \in [a,b], \tag{2.6}$$

(mit beliebigem $\gamma \in \mathbb{R}$) Lösungen von (2.5) sind. Offenbar gilt also $y_\gamma \in \operatorname{kern} \varrho$ für alle $\gamma \in \mathbb{R}$, also $\dim \operatorname{kern} \varrho \geq 1$.

Folgen Sie *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984*, um zu zeigen dass $\dim \operatorname{kern} \varrho = 1$.

Durch Vorgabe eines *Anfangswerts* $y(a) = \mu$ für die Lösungen der Differentialgleichung (2.5) hat man eine sogenannte *Anfangswertaufgabe*, und in der Familie (2.6) von Lösungen ist genau eine Lösung y_{γ_0} ausgezeichnet, die diese Anfangswertaufgabe löst: Es muss

$$\gamma_0 e^{-\alpha a} = \mu, \quad \text{d.h., } \gamma_0 = \mu e^{\alpha a}$$

gelten. ◇

Die folgende Aussage wird in der Vorlesung bewiesen, vgl. auch *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984*, Satz 2.8 und Korollar 2.9.

2.3.20 Satz. (Prinzip der linearen Fortsetzung)

Seien V und W Vektorräume und zumindest V endlichdimensional. Ist

$$\{\underline{v}^1, \ldots, \underline{v}^n\} \ \text{ eine Basis von } V \text{ und } \ \{\underline{w}^1, \ldots, \underline{w}^n\} \subset W,$$

dann gibt es genau eine lineare Abbildung $\varphi : V \to W$ mit

$$\varphi(\underline{v}^i) = \underline{w}^i \ \text{ für alle } i \in \{1, \ldots, n\}.$$

Speziell gilt also:

Hat V eine endliche Basis $\{\underline{v}^1, \ldots, \underline{v}^n\}$ und ist $\varphi : V \to W$ linear, so ist φ durch die Bilder der Basiselemente $\varphi(\underline{v}^i)$, $i = 1, \ldots, n$, bereits vollständig bestimmt. ◇

2.3.21 Übung. Sei $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ die lineare Abbildung, die durch

$$\varphi(\underline{e}^1) = \underline{e}^1 \quad \text{und} \quad \varphi(\underline{e}^2) = -\underline{e}^2$$

definiert ist, wobei $\underline{e}^1$ und $\underline{e}^2$ die Einheitsvektoren in $\mathbb{R}^2$ sind. Bestimmen Sie

$$\varphi(\underline{x}) \ \text{ für } \underline{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \ \underline{x} = \begin{pmatrix} -5 \\ 5 \end{pmatrix},$$

sowie für beliebiges $\underline{x} = (x_1, x_2)^\mathsf{T} \in \mathbb{R}^2$. Was bedeutet diese Abbildung geometrisch? ◇

2.3.22 Übung. Für die lineare Abbildung $\varphi : \mathbb{R}^2 \to \mathbb{R}^3$ sei nur bekannt, dass

$$\varphi(\mathbb{R}^2) = \operatorname{lin}\{(1,1,1)^\mathsf{T}, (1,1,0)^\mathsf{T}\}.$$

Zeigen Sie, dass φ injektiv ist. Finden Sie eine derartige Abbildung. Ist es die einzige? Wie verträgt sich diese Erkenntnis mit dem Prinzip der linearen Fortsetzung? $\diamond$

2.3.23 Übung. (vgl. Kall, §2.1) Sei $\mathcal{P}^n$ der Vektorraum der Polynome $p(t)$ höchstens n-ten Grades und $p_0 \in \mathcal{P}^n$ das lineare Polynom $p_0(t) = t$. Wir definieren eine Abbildung φ von $\mathcal{P}^n$ in $\mathcal{P}^{n+1}$ durch

$$\varphi(p) = p_0 \cdot p \quad \forall p \in \mathcal{P}^n,$$

d.h., $p_0(t)$ wird mit $p(t)$ multipliziert. Zeigen Sie, dass φ linear und injektiv, aber nicht bijektiv ist. $\diamond$

2.3.24 Übung. Wir erinnern an das Beispiel 2.1.12 zur Verbrauchsmatrix:

$$\begin{pmatrix} \text{En-Verbrauch} \\ \text{Ch-Verbrauch} \\ \text{Ba-Verbrauch} \end{pmatrix} = \begin{pmatrix} 0.2 & 0.4 & 0.3 \\ 0.3 & 0.4 & 0.2 \\ 0.4 & 0.3 & 0.1 \end{pmatrix} \begin{pmatrix} \text{Input für En} \\ \text{Input für Ch} \\ \text{Input für Ba} \end{pmatrix}.$$

Wie in Beispiel 2.1.12 nennen wir die Verbrauchsmatrix A, den Produktionsvektor (Input) $\underline{p} = (p_1, p_2, p_3)^\mathsf{T}$ und den Nachfragevektor $\underline{y} = (y_1, y_2, y_3)^\mathsf{T}$. Es ist ein Gleichungssystem in den Variablen p_1, p_2, p_3 zu lösen, und zwar

$$\underline{p} - A\underline{p} = \underline{y}, \quad \text{d.h., } (I - A)\underline{p} = \underline{y}.$$

Lösen Sie das Zahlenbeispiel und diskutieren Sie die Eigenschaften der Matrix $I - A$ und der linearen Abbildung

$$\varphi(\underline{p}) = (I - A)\underline{p}$$

und gegebenenfalls ihrer Umkehrabbildung. $\diamond$

2.3.25 Lineare Abbildungen zwischen endlichdimensionalen Vektorräumen.

Seien V und W endlichdimensionale Vektorräume und

$$\{\underline{v}^1, \ldots, \underline{v}^n\} \text{ eine Basis von } V \text{ und } \{\underline{w}^1, \ldots, \underline{w}^m\} \text{ eine Basis von } W.$$

Sei $\varphi : V \to W$ linear. Die Bilder der Basisvektoren in V sind durch

$$\varphi(\underline{v}^j) = \sum_{i=1}^{m} a_{ij}\underline{w}^i, \; j = 1, \ldots, n, \tag{2.7}$$

mit reellen a_{ij} darstellbar. Nach dem Prinzip der linearen Fortsetzung ist durch die Bilder der Basisvektoren in V die lineare Abbildung φ schon vollständig bestimmt. Mit anderen Worten: φ ist durch die Matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

bereits vollständig bestimmt. Die j-te Spalte von A ist also der Komponentenvektor von $\varphi(\underline{v}^j)$ bezüglich der gegebenen Basis in W.

Sei nun $\underline{v} \in V$ beliebig und habe die Komponenten x_j bezüglich der Basisvektoren $\underline{v}^j$ $\forall j$, also

$$\underline{v} = \sum_{j=1}^{n} x_j \, \underline{v}^j .$$

Dann folgt, da φ linear,

$$\varphi(\underline{v}) = \sum_{j=1}^{n} x_j \, (\sum_{i=1}^{m} a_{ij} \underline{w}^i) = \sum_{i=1}^{m} y_i \underline{w}^i \quad \text{mit} \quad y_i = \sum_{j=1}^{n} a_{ij} x_j, \; i = 1, \dots, m,$$

d.h., y_i definieren die Komponenten des Vektors $\varphi(\underline{v})$ bezüglich der gegebenen Basis in W, und wir haben für die Spaltenvektoren (Spaltenmatrizen) $\underline{x} = (x_1, \dots, x_n)^\mathsf{T}$ und $\underline{y} = (y_1, \dots, y_m)^\mathsf{T}$ die Beziehung

$$\underline{y} = A\underline{x} \tag{2.8}$$

nach den uns bekannten Gesetzen der Matrizenmultiplikation.

Fazit: Bei gegebenen Basen in V und W ist jeder linearen Abbildung $\varphi : V \longrightarrow W$ eine Matrix A zugeordnet, deren j-te Spalte der Komponentenvektor von $\varphi(\underline{v}^j)$ ist und die eine lineare Abbildung vom Raum $\mathbb{R}^n$ (der Komponentenvektoren zu $\underline{v} \in V$) in den Raum $\mathbb{R}^m$ (der Komponentenvektoren zu $\varphi(\underline{v}) \in W$) vermittelt.

Umgekehrt definiert (bei gegebenen Basen in V und W) jede Matrix $A \in M(m, n)$ eine lineare Abbildung von V in W, wenn man diese gemäss (2.7) erklärt.

Wichtig: Die Matrix A hängt von den vorgegebenen Basen in V und W ab. $\diamond$

2.3.26 Übung. (vgl. Kall, §2.3) Sei $\varphi : \mathbb{R}^3 \to \mathbb{R}^3$ mit einer festen Zahl $\theta \in \mathbb{R}$ definiert durch

$$\varphi(x, y, z) = (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta, z).$$

Zeigen Sie, dass φ linear und bijektiv (also ein Isomorphismus) ist und geben Sie die zugeordnete Matrix A bezüglich der kanonischen Basis $\{\underline{e}^1, \underline{e}^2, \underline{e}^3\}$ des $\mathbb{R}^3$ an. Ist A regulär? Welche Matrix ergibt sich, wenn im Urbildraum $\mathbb{R}^3$ die Basis $\{\underline{e}^1, \underline{e}^1 + \underline{e}^2, \underline{e}^1 + \underline{e}^2 + \underline{e}^3\}$, aber im Bildraum $\mathbb{R}^3$ die kanonische Basis gewählt wird? $\diamond$

2.3.27 Isomorphismen und reguläre Matrizen. Seien V und W n-dimensionale Vektorräume und $\varphi : V \to W$ ein Isomorphismus. Dann ist jedem Paar von Basen in V und W eine reguläre n-reihige Matrix zugeordnet. Beweis: siehe Vorlesung. $\diamond$

2.3.28 Verknüpfungen von linearen Abbildungen bzw. Matrizen.

Wegen der eineindeutigen Zuordnung von linearen Abbildungen zwischen endlichdimensionalen Vektorräumen und Matrizen passenden Typs sowie den gleichberechtigten Charakterisierungen (2.7) und (2.8) haben wir für die bekannten Verknüpfungen folgende Zusammenhänge.

Seien $\dim V = n$, $\dim W = m$, $\varphi : V \to W$ und $\psi : V \to W$ mit zugeordneten $(m \times n)$-Matrizen A_φ und A_ψ, dann gehört zu $\lambda\varphi$ (mit $\lambda \in \mathbb{R}$ ist die Abbildung bekanntlich wieder linear) die Matrix λA_φ und zu $\varphi + \psi$ (bekanntlich ebenfalls wieder linear) die Matrix $A_\varphi + A_\psi$.

Seien $\dim V = n$, $\dim W = m$, $\dim X = r$, $\varphi : V \to W$ und $\psi : W \to X$ mit zugeordneten $(m \times n)$- bzw. $(r \times m)$-Matrizen A_φ und A_ψ, dann gehört zur Hintereinanderausführung $\psi \circ \varphi$ (bekanntlich ebenfalls wieder linear) das Matrixprodukt $A_\psi A_\varphi$. Beachten Sie die Dimensionszusammenhänge

$$V \to W \to X \quad \text{mit} \quad \dim V = n,\ \dim W = m,\ \dim X = r$$

für die linearen Abbildungen und

$$A_\varphi \in M(m,n),\ A_\psi \in M(r,m), \quad \text{verknüpft zu} \quad A_\psi A_\varphi \in M(r,n)$$

für die zugeordneten Matrizen.

Die Definition des Matrizenprodukts, wie wir sie vorgenommen haben, ist also gerechtfertigt durch die Hintereinanderausführung von linearen Abbildungen. $\diamond$

2.3.29 Definition. Sind V und W endlichdimensionale Vektorräume und $\varphi : V \to W$ linear, dann heisst die Dimension des Bildes $\varphi(V)$ auch ***Rang*** der Abbildung φ, d.h.,

$$\operatorname{rg}(\varphi) = \dim \varphi(V).$$

$\diamond$

2.3.30 Bemerkung. Diese Definition ist gerechtfertigt im Vergleich zur schon bekannten Definition des Rangs einer $(m \times n)$-Matrix: Seien nämlich Basen in V und W gegeben und A die dazugehörige Matrix bezüglich der Abbildung φ. Dann ist $\dim \varphi(V)$ gleich der Dimension des Spaltenraums von A, also

$$\operatorname{rg}(\varphi) = \operatorname{rg} A.$$

$\diamond$

Es folgt nun unmittelbar aus den vorangegangenen Aussagen und Überlegungen:

2.3.31 Satz. *Seien V und W Vektorräume mit* $\dim V = n$ *und* $\dim W = m$, $\varphi : V \to W$ *eine lineare Abbildung und A diejenige Matrix, die der Abbildung φ bezüglich eines festen Paars von Basen in V und W eindeutig zugeordnet ist.*

1. *Dann sind die folgenden Bedingungen gleichwertig:*

   (i) φ *ist injektiv.*

   (ii) $\operatorname{kern}\varphi = \{\underline{o}\}$ *(d.h.,* $\dim \operatorname{kern}\varphi = 0$*).*

   (iii) $\operatorname{rg}(\varphi) = n$ *(wobei* $\operatorname{rg}(\varphi) = \dim \varphi(V)$*).*

   (iv) $\operatorname{rg} A = n$ *(= Anzahl der Spalten von A).*

2. *Ferner sind die folgenden Bedingungen gleichwertig:*

   (i) φ *ist surjektiv (d.h.,* $W = \varphi(V)$*).*

   (ii) $\operatorname{rg}(\varphi) = m$ *(wobei* $\operatorname{rg}(\varphi) = \dim \varphi(V)$*).*

   (iii) $\operatorname{rg} A = m$ *(= Anzahl der Zeilen von A).*

3. *Schliesslich sind die folgenden Bedingungen gleichwertig:*

   (i) φ *ist bijektiv.*

   (ii) $\dim V = \dim W$ *und* φ *ist injektiv.*

   (iii) $\dim V = \dim W$ *und* φ *ist surjektiv.* ◇

2.3.32 Folgerung. Sei A eine gegebene $(m \times n)$-Matrix. Im Spezialfall der durch $\varphi(\underline{x}) = A\underline{x}$ definierten linearen Abbildung $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ gilt also

$$\begin{aligned} \varphi \text{ injektiv} &\Leftrightarrow \operatorname{rg} A = n. \\ \varphi \text{ surjektiv} &\Leftrightarrow \operatorname{rg} A = m. \\ \varphi \text{ bijektiv} &\Leftrightarrow m = n = \operatorname{rg} A. \end{aligned}$$

◇

2.3.33 Folgerung. Seien A eine $(m \times n)$-Matrix sowie S eine reguläre n-reihige Matrix und T eine eine reguläre m-reihige Matrix. Dann gilt

$$\operatorname{rg} TA = \operatorname{rg} A \quad \text{und} \quad \operatorname{rg} AS = \operatorname{rg} A.$$

In der Sprache der linearen Abbildungen gelten also für eine beliebige lineare Abbildung $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ sowie Isomorphismen $\varrho : \mathbb{R}^m \to \mathbb{R}^m$ und $\psi : \mathbb{R}^n \to \mathbb{R}^n$ die Beziehunegn

$$\operatorname{rg}(\varrho \circ \varphi) = \operatorname{rg}(\varphi) \quad \text{und} \quad \operatorname{rg}(\varphi \circ \psi) = \operatorname{rg}(\varphi).$$ ◇

2.4 Koordinatentransformation

2.4.1 Beispiel. (Ellipse in verschiedenen Koordinatensystemen) In jedem Tafelwerk findet sich für eine Ellipse $\mathcal{E}$ in Standarddarstellung die Formel

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

wobei $a > 0$ und $b > 0$ die Längen der beiden Halbachsen von $\mathcal{E}$ sowie der Koordinatenursprung $(0,0)$ der Mittelpunkt von $\mathcal{E}$ sind. Im Spezialfall $a = b = r > 0$ erhalten wir die Gleichung eines Kreises mit Mittelpunkt in $(0,0)$ und Radius r.

Was können wir aber tun, wenn z.B. eine Menge $\mathcal{X}$ durch

$$\mathcal{X} = \{(x,y) \in \mathbb{R}^2 \,|\, 5x^2 + 8xy + 5y^2 = 1\}$$

gegeben ist, um festzustellen, ob es sich um eine Ellipse handelt, und wenn ja, wo ihr Mittelpunkt liegt und wie lang ihre Halbachsen sind?

Wenn man bei *G. Strang, Lineare Algebra, Springer, 2003* nachschaut, von wo das Beispiel stammt (siehe dort Beispiel 6.5.6), sieht man, dass eine "geschickte Umformung"

$$5x^2 + 8xy + 5y^2 = 9\left(\frac{x+y}{\sqrt{2}}\right)^2 + \left(\frac{x-y}{\sqrt{2}}\right)^2$$

(nachrechnen!) auf die neuen Koordinaten

$$\xi := \frac{x+y}{\sqrt{2}}, \quad \eta := \frac{x-y}{\sqrt{2}}$$

führt, und man erhält die Formel $9\xi^2 + \eta^2 = 1$ für eine Ellipse in Standarddarstellung. Folglich ist $(0,0)$ der Mittelpunkt von $\mathcal{E}$, die Halbachsen haben die Längen $1/3$ und 1.

Diese *Koordinatentransformation* kann man als eine lineare Abbildung von $\mathbb{R}^2$ in $\mathbb{R}^2$ auffassen, sie ist durch

$$\begin{pmatrix} \xi \\ \eta \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

definiert und repräsentiert eine Drehung plus Spiegelung.

Die systematische Behandlung quadratischer Gleichungen in n Variablen mit Hilfe der sogenannten Hauptachsentransformation, einer speziellen Koordinatentransformation, behandeln wir im Kapitel "Eigenwerte". ◇

2.4.2 Definition. Es sei V ein Vektorraum mit $\dim V = n$, und es seien $\mathcal{B} = \{\underline{p}^1, \ldots, \underline{p}^n\}$ und $\mathcal{C} = \{\underline{q}^1, \ldots, \underline{q}^n\}$ zwei Basen von V. Die lineare Abbildung $\varphi : V \to V$, die durch $\varphi(\underline{p}^i) = \underline{q}^i \;\forall i$ gegeben ist, heisst ***Koordinatentransformation***. Schreibt man die Bilder von V wieder als Linearkombination der Basis $\mathcal{B}$, so nennt man die Matrix $T = (t_{ij})$, die durch

$$\underline{q}^j = \sum_{i=1}^{n} t_{ij}\underline{p}^i, \quad j = 1, \ldots, n, \tag{2.9}$$

gegeben ist, die ***zugehörige Transformationsmatrix bezüglich der Basis*** $\mathcal{B}$. ◇

2.4.3 Wir werden uns im folgenden auf den Fall $V = \mathbb{R}^n$ beschränken. Seien wie oben $\mathcal{B} = \{\underline{p}^1, \ldots, \underline{p}^n\}$ und $\mathcal{C} = \{\underline{q}^1, \ldots, \underline{q}^n\}$ zwei Basen dieses Vektorraums.

Die Formel (2.9) kann man ausführlicher so schreiben:

$$\underline{q}^j = t_{1j}\underline{p}^1 + t_{2j}\underline{p}^2 + \ldots + t_{nj}\underline{p}^n, \quad j = 1, \ldots, n. \tag{2.10}$$

Also entspricht für die Matrix

$$T := \begin{pmatrix} t_{11} & t_{12} & \ldots & t_{1n} \\ t_{21} & t_{22} & \ldots & t_{2n} \\ \vdots & \vdots & & \vdots \\ t_{n1} & t_{n2} & \ldots & t_{nn} \end{pmatrix}$$

die j-te Spalte $T_{\bullet j}$ von T den Koeffizienten zur Darstellung von $\underline{q}^j$ in (2.10).

Schreibt man die Elemente $\underline{p}^i$ bzw. $\underline{q}^j$ der Basen $\mathcal{B}$ bzw. $\mathcal{C}$ des $\mathbb{R}^n$ als Spaltentupel der Länge n, so kann man sie zu n-reihigen Matrizen

$$P = [\underline{p}^1 \; \underline{p}^2 \; \ldots \; \underline{p}^n] \quad \text{und} \quad Q = [\underline{q}^1 \; \underline{q}^2 \; \ldots \; \underline{q}^n]$$

zusammenfassen. Da $\mathcal{B}$ und $\mathcal{C}$ Basen des $\mathbb{R}^n$ sind, sind die Matrizen P und Q regulär, und man kann (2.10) auch schreiben als

$$\underline{q}^j = PT_{\bullet j}, \quad j = 1, \ldots, n,$$

also in Matrixschreibweise

$$Q = PT. \tag{2.11}$$

Da P und Q reguläre n-reihige Matrizen sind, gilt:

$$T \text{ ist eine reguläre } n\text{-reihige Matrix.}$$

Betrachten wir nun die *Rücktransformation*, also die Koordinatentransformation

$$\psi : \mathbb{R}^n \to \mathbb{R}^n \quad \text{mit} \quad \psi(\underline{q}^i) = \underline{p}^i, \quad i = 1, \ldots, n,$$

so definiert die j-te Spalte $S_{\bullet j}$ der Matrix

$$S := \begin{pmatrix} s_{11} & s_{12} & \ldots & s_{1n} \\ s_{21} & s_{22} & \ldots & s_{2n} \\ \vdots & \vdots & & \vdots \\ s_{n1} & s_{n2} & \ldots & s_{nn} \end{pmatrix}$$

die Koeffizienten zur Darstellung von $\underline{p}^j$ als Linearkombination aus den Vektoren $\underline{q}^j$:

$$\underline{p}^j = s_{1j}\underline{q}^1 + s_{2j}\underline{q}^2 + \ldots + s_{nj}\underline{q}^n, \quad j = 1, \ldots, n.$$

In Matrixschreibweise ergibt das
$$P = Q\,S.$$
Zusammen mit (2.11) folgt daraus
$$S = Q^{-1}P = (T^{-1}P^{-1})P = T^{-1}, \tag{2.12}$$
also ist die zur Rücktransformation ψ gehörende Matrix (bezüglich $\mathcal{C}$) die Inverse der Matrix zur Koordinatentransformation φ (bezüglich $\mathcal{B}$).

Sei nun $\underline{v} \in \mathbb{R}^n$ beliebig und seien $\underline{x}$ und $\underline{y}$ die Komponentenvektoren von $\underline{v}$ bezüglich $\mathcal{B}$ bzw. $\mathcal{C}$, d.h.,
$$\underline{v} \;=\; \sum_{i=1}^{n} x_i \underline{p}^i \;=\; \sum_{i=1}^{n} y_i \underline{q}^i. \tag{2.13}$$
Dann ergibt sich - wenn $\underline{v}$, $\underline{x}$ und $\underline{y}$ als Spaltenvektoren geschrieben werden - die Matrixdarstellung
$$\underline{v} \;=\; P\underline{x} \;=\; Q\underline{y}.$$
Somit haben wir wegen (2.12) die Beziehung
$$\underline{y} = Q^{-1}P\underline{x} = T^{-1}\underline{x}. \tag{2.14}$$

Zusammengefasst ergibt sich also

2.4.4 Satz. *Seien $\mathcal{B} = \{\underline{p}^1, \dots, \underline{p}^n\}$ und $\mathcal{C} = \{\underline{q}^1, \dots, \underline{q}^n\}$ zwei Basen des $\mathbb{R}^n$ und φ die Koordinatentransformation $\varphi(\underline{p}^i) = \underline{q}^i\ \forall i$ mit der zugehörigen Transformationsmatrix T bezüglich $\mathcal{B}$. Dann gilt:*

1. *Ist ψ die Koordinatentransformation $\psi(\underline{q}^i) = \underline{p}^i\ \forall i$, so ist $S = T^{-1}$ für die zugehörige Transformationsmatrix S bezüglich $\mathcal{C}$ erfüllt.*

2. *Ist $\underline{v} \in \mathbb{R}^n$ beliebig und sind $\underline{x}$ und $\underline{y}$ die Komponentenvektoren von $\underline{v}$ bezüglich $\mathcal{B}$ bzw. $\mathcal{C}$ gemäss (2.13), dann gilt $\underline{y} = T^{-1}\underline{x}$.* ◇

2.4.5 Satz. *Es seien $\mathcal{B} = \{\underline{p}^1, \dots, \underline{p}^n\}$ und $\mathcal{C} = \{\underline{q}^1, \dots, \underline{q}^n\}$ zwei Basen des $\mathbb{R}^n$ sowie $\mathcal{G} = \{\underline{g}^1, \dots, \underline{g}^m\}$ und $\mathcal{H} = \{\underline{h}^1, \dots, \underline{h}^m\}$ zwei Basen des $\mathbb{R}^m$. Ferner seien in $\mathbb{R}^n$ und $\mathbb{R}^m$ Koordinatentransformationen mit den Transformationsmatrizen $T = (t_{ij})$ und $S = (s_{ij})$*
$$\text{gemäss}\quad \underline{q}^j = \sum_{i=1}^{n} t_{ij}\underline{p}^i, \quad j = 1, \dots, n, \quad \text{bzw.}\ \ \underline{h}^j = \sum_{i=1}^{m} s_{ij}\underline{g}^i, \quad j = 1, \dots, m,$$
gegeben. Der linearen Abbildung
$$\varphi : \mathbb{R}^n \to \mathbb{R}^m$$
sei bezüglich der Basen $\mathcal{B}$ und $\mathcal{G}$ die Matrix
$$A = (a_{ij};\ 1 \le i \le m,\ 1 \le j \le n\}$$
zugeordnet. Dann wird φ bezüglich der Basen $\mathcal{C}$ und $\mathcal{H}$ beschrieben durch die Matrix
$$B = S^{-1}AT.$$
◇

Beweis: Alle hier auftretenden Vektoren werden als Spaltentupel aufgefasst. Sei $\underline{v} \in \mathbb{R}^n$ beliebig. Gemäss (2.8) gilt für die Komponentenvektoren $\underline{x}$ von $\underline{v}$ in der Basis $\mathcal{B}$ bzw. $\underline{z}$ von $\varphi(\underline{v})$ in der Basis $\mathcal{G}$ die Beziehung

$$\underline{z} = A\underline{x}.$$

Andererseits gilt nach Aussage 2 in Satz 2.4.4 für den Komponentenvektor $\underline{y}$ von $\underline{v}$ gemäss der Basis $\mathcal{C}$ und den Vektor $\underline{x}$ die Beziehung

$$\underline{y} = T^{-1}\underline{x}$$

Bezeichnen wir den Komponentenvektor von $\varphi(\underline{v})$ in der Basis $\mathcal{H}$ mit $\underline{\pi}$, so gilt wieder nach Aussage 2 in Satz 2.4.4, dass

$$\underline{\pi} = S^{-1}\underline{z}.$$

Um die gewünscht Aussage zu beweisen, ist eine Matrix B gesucht, die φ bezüglich der Basen $\mathcal{C}$ und $\mathcal{H}$ beschreibt. Nach dem Abschnitt 2.3.25 muss B also

$$\underline{\pi} = B\underline{y}$$

erfüllen. Mit den obigen Gleichungen ermitteln wir

$$\underline{\pi} = S^{-1}\underline{z} = S^{-1}A\underline{x} = S^{-1}AT\underline{y},$$

also $B = S^{-1}AT$, was zu zeigen war. □

2.4.6 Korollar. *Sei die lineare Abbildung φ in Satz 2.4.5 eine Abbildung von $\mathbb{R}^n$ in $\mathbb{R}^n$. Ferner sei*

$$\mathcal{B} = \mathcal{G} \quad \text{und} \quad \mathcal{C} = \mathcal{H},$$

d.h., für den Urbild- und Bildraum $\mathbb{R}^n$ von φ wird zum einen jeweils die Basis $\mathcal{B}$ verwendet, zum anderen jeweils die Basis $\mathcal{C}$. Die Koordinatentransformation von $\mathcal{B}$ in $\mathcal{C}$ sei wie in Satz 2.4.5 durch die Transformationsmatrix T beschrieben. Dann gilt: Ist die Abbildung φ bezüglich der Basis $\mathcal{B}$ durch die Matrix A beschrieben, so wird φ bezüglich der Basis $\mathcal{C}$ durch die Matrix

$$B = T^{-1}AT$$

beschrieben. ◇

2.4.7 Definition. Die mit der Koordinatentransformation T gemäss Korollar 2.4.6 verbundene Transformation von A in $B = T^{-1}AT$ heisst ***Ähnlichkeitstransformation*** der Matrix A. Zwei n-reihige Matrizen A und B heissen ***ähnlich***, falls es eine reguläre n-reihige Matrix T gibt, so dass $B = T^{-1}AT$ gilt. ◇

2.4.8 Übung. Man veranschauliche sich die Aussage von Korollar 2.4.6 an dem Beispiel in Aufgabe 2.3.26. ◇

2.5 Transformationsmatrizen der Gauss-Elimination

2.5.1 Beispiel. Betrachten wir die Matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{pmatrix} := \begin{pmatrix} 2 & 2 & -1 & -1 \\ 2 & 0 & 2 & 4 \\ 4 & 2 & 1 & 3 \end{pmatrix}$$

Durch welche Transformationsmatrizen werden Gauss-Schritte repräsentiert?

1. Vertauschung der 1. und 2. Zeile von A:

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

2. Erzeugung von Nullen unter dem *Pivotelement* $a_{11} = 2$ (Pivotelement bleibt unverändert):

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

3. Vertauschung der 2. und 4. Spalte von A

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

Die Multiplikation einer Zeile von A mit einem Faktor $\lambda \neq 0$ lässt sich ähnlich darstellen, wir benötigen diese Transformation im folgenden aber nicht. $\diamond$

2.5.2 Für eine beliebige $(m \times n)$-Matrix $A = (a_{ij})$ sind die Transformationsmatrizen entsprechend:

1. Die Vertauschung der i-ten und k-ten Zeile von A erfolgt durch die Multiplikation von links

$$PA,$$

und zwar mit der $(m\times m)$-Permutationsmatrix P, in der im Vergleich mit der $(m\times m)$-Einheitsmatrix die i-te und k-te Einheitsspalte vertauscht sind.

2. Ausführlich aufgeschrieben, hat A die Form

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} & a_{1k+1} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2k} & a_{2k+1} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & \dots & a_{kk} & a_{kk+1} & \dots & a_{kn} \\ a_{k+11} & a_{k+12} & \dots & a_{k+1k} & a_{k+1k+1} & \dots & a_{k+1n} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mk} & a_{mk+1} & \dots & a_{mn} \end{pmatrix}.$$

Zur Erzeugung von Nullen unter dem *Pivotelement* a_{kk} mit $a_{kk} \neq 0$ (!), wobei die ersten k Zeilen (insbesondere das Pivotelement) unverändert bleiben, multipliziere man von links

$$T^{(k)}A,$$

und zwar mit der Matrix

$$T^{(k)} = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & -t^k_{k+1k} & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & -t^k_{mk} & 0 & \dots & 1 \end{pmatrix}, \tag{2.15}$$

wobei $t^k_{ik} = a_{ik}/a_{kk}$ für $i > k$. Stehen unter a_{11}, a_{22}, ... , a_{k-1k-1} bereits Nullen, bleibt diese Eigenschaft in $T^{(k)}A$ offenbar erhalten.

3. Die Vertauschung der j-ten und l-ten Spalte von A erfolgt durch die Multiplikation von rechts

$$AQ,$$

und zwar mit der $(n \times n)$-Permutationsmatrix Q, in der im Vergleich mit der $(m \times m)$-Einheitsmatrix die j-te und l-te Einheitsspalte vertauscht sind.

2.5.3 Inverse einer Permutationsmatrix. Sei P eine Permutationsmatrix. Dann gilt, wie man leicht sieht, $P^\mathsf{T} P = PP^\mathsf{T} = I$, also $P^{-1} = P^\mathsf{T}$.

Offenbar gilt

2.5.4 Satz. *Die in Abschnitt 2.5.2 definierten quadratischen Matrizen P, $T^{(k)}$ und Q sind regulär.* ◇

2.5.5 Das Lösen eines inhomogenen LGS $A\underline{x} = \underline{b}$ mit $A = (a_{ij};\ i = 1, \ldots, m;\ j = 1, \ldots, n)$ und $\underline{b} = (b_1, \ldots, b_m)^\mathsf{T}$ führt bekanntlich nach endlich vielen Schritten vom Typ 1 und 2, die auf die sukzessive transformierte erweiterte Koeffizientenmatrix $[A, \underline{b}]$ angewendet werden, bzw. Spaltenvertauschungen in der sukzessive transformierten Matrix A (Typ 3) auf die *Trapezform*

$$[\widetilde{A}, \widetilde{\underline{b}}] \;=\; \begin{pmatrix} \tilde{a}_{11} & \tilde{a}_{12} & \ldots & \tilde{a}_{1r} & \tilde{a}_{1\,r+1} & \ldots & \tilde{a}_{1n} & \tilde{b}_1 \\ 0 & \tilde{a}_{22} & \ldots & \tilde{a}_{2r} & \tilde{a}_{2\,r+1} & \ldots & \tilde{a}_{2n} & \tilde{b}_2 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \ldots & \tilde{a}_{rr} & \tilde{a}_{r\,r+1} & \ldots & \tilde{a}_{rn} & \tilde{b}_r \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 & \tilde{b}_{r+1} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 & \tilde{b}_m \end{pmatrix} \tag{2.16}$$

mit $\tilde{a}_{ii} \neq 0$ für $i = 1, \ldots, r$.

Transformationen vom Typ 1 bzw. 2 entsprechen nach Satz 2.5.4 jeweils der Umformung

$$PA\underline{x} = P\underline{b} \qquad \text{bzw.} \qquad T^{(k)}A\underline{x} = T^{(k)}\underline{b}$$

auf ein äquivalentes LGS (d.h., die Lösungsmenge bleibt gleich).

Transformationen vom Typ 3 entsprechen der Umformung von $A\underline{x} = \underline{b}$ in

$$AQ\underline{y} = \underline{b} \qquad \text{mit} \qquad \underline{x} = Q\underline{y},$$

was also einer Vertauschung von Variablenindizes entspricht und als spezielle Koordinatentransformation im $\mathbb{R}^n$ interpretiert werden kann.

Kapitel 3

Vektorräume mit Skalarprodukt

Inhaltsverzeichnis

3.1 Skalarprodukt und Norm 61
3.2 Orthonormalsysteme 66
3.3 Lineare Approximationsprobleme 69

3.1 Skalarprodukt und Norm

Bisher haben wir noch nicht die Möglichkeit genutzt, in Vektorräumen Entfernungen, Winkel oder Flächen- und Rauminhalte zu messen.

Dazu führen wir zunächst eine Verknüpfung ein, die zwei Vektoren eines Vektorraums V auf solche Weise eine reelle Zahl zuordnet, dass man Winkel und Entfernungen zwischen Vektoren in bekannter Weise betrachten kann.

3.1.1 Definition. Ist in einem Vektorraum V zu je zwei Vektoren $\underline{v}, \underline{w} \in V$ eindeutig eine reelle Zahl $\langle \underline{v}, \underline{w} \rangle$ mit den folgenden Eigenschaften E1 – E4 zugeordnet, so heisst $\langle \underline{v}, \underline{w} \rangle$ ***Skalarprodukt*** oder ***inneres Produkt*** der Vektoren $\underline{v}$ und $\underline{w}$:

E1. $\langle \underline{v}, \underline{w} \rangle = \langle \underline{w}, \underline{v} \rangle$,

E2. $\langle \lambda \underline{v}, \underline{w} \rangle = \lambda \langle \underline{v}, \underline{w} \rangle$ für beliebige $\lambda \in \mathbb{R}$,

E3. $\langle \underline{u} + \underline{v}, \underline{w} \rangle = \langle \underline{u}, \underline{w} \rangle + \langle \underline{v}, \underline{w} \rangle$ für beliebige $\underline{u} \in V$,

E4. $\langle \underline{v}, \underline{v} \rangle > 0$, falls $\underline{v} \neq \underline{o}$. $\diamond$

3.1.2 Beispiele.

1. Im Vektorraum $V = \mathbb{R}^n$ ist zu Elementen $\underline{v}, \underline{w} \in \mathbb{R}^n$ mit den Komponenten v_j und w_j, wobei $\underline{v}$ und $\underline{w}$ als Spaltenvektoren aufgefasst werden, durch

$$\langle \underline{v}, \underline{w} \rangle := \underline{v}^{\mathsf{T}} \underline{w} = \sum_{j=1}^{n} v_j w_j \tag{3.1}$$

ein Skalarprodukt zugeordnet, das ***euklidische Skalarprodukt*** von $\underline{v}$ und $\underline{w}$. Ökonomisch interpretierbar ist das z.B. durch die Zuordnung des Erlöses (reelle Zahl) zu einem vorgegebenen (Verkaufs-)Preisvektor $\underline{v}$ und einem (verkauften) Güterbündel $\underline{w}$.

2. Betrachten wir wieder den Vektorraum $V = \mathbb{R}^n$ und schreiben seine Elemente als Spaltenvektoren. Sei Q eine symmetrische n-reihige Matrix mit der Eigenschaft

$$\underline{v}^{\mathsf{T}} Q \underline{v} > 0 \quad \text{für alle } \underline{v} \in \mathbb{R}^n \setminus \{\underline{o}\}. \tag{3.2}$$

Derartige Matrizen heissen *positiv definit*. Dann ist durch

$$\langle \underline{v}, \underline{w} \rangle := \underline{v}^{\mathsf{T}} Q \underline{w} \qquad \forall \underline{v}, \underline{w} \in \mathbb{R}^n, \tag{3.3}$$

ein Skalarprodukt zugeordnet: Es gilt E1 wegen

$$\underline{v}^{\mathsf{T}} Q \underline{w} = (\underline{v}^{\mathsf{T}} Q \underline{w})^{\mathsf{T}} \text{ (eine Zahl!) } = \underline{w}^{\mathsf{T}} Q^{\mathsf{T}} \underline{v} = \underline{w}^{\mathsf{T}} Q \underline{v},$$

da Q symmetrische Matrix ist (d.h., $Q^{\mathsf{T}} = Q$). Ferner gelten E2 wegen

$$(\lambda \underline{v})^{\mathsf{T}} Q \underline{w} = \lambda\, \underline{v}^{\mathsf{T}} Q \underline{w}$$

und E3 wegen

$$(\underline{u} + \underline{v})^{\mathsf{T}} Q \underline{w} = \underline{u}^{\mathsf{T}} Q \underline{w} + \underline{v}^{\mathsf{T}} Q \underline{w},$$

jeweils nach den Rechengesetzen für Matrizen. Letztlich folgt E4 aus den Definitionen (3.3) und (3.2). Insbesondere sind damit auch die Skalarproduktaxiome für das Beispiel 1 nachgewiesen, denn mit $Q = I$ ist Beispiel 1 nur ein Spezialfall von Beispiel 2.

3. Sei nun $V = C[a,b]$, der Vektorraum der auf $[a,b]$ (mit $a < b$) stetigen Funktionen. Dann ist zu $f, g \in C[a,b]$ mittels

$$\langle f, g \rangle := \int_a^b f(x) g(x) dx \tag{3.4}$$

das Skalarprodukt von f und g zugeordnet. E1 gilt, da man unter dem Integral $f(x)$ und $g(x)$ vertauschen kann, E2 und E3 gelten nach den Rechengesetzen für Integrale. Zu E4: Wenn $f(x_0) \neq 0$ für mindestens ein $x_0 \in [a,b]$ gilt, so gibt es ein $\varepsilon > 0$, so dass

$$f^2(x) > 0 \quad \forall x \in [a,b] \cap B(x_0, \varepsilon)$$

gilt, wobei $B(x_0, \varepsilon) = \{x \mid \|x - x_0\| \le \varepsilon\}$. Wegen $f^2(x) \ge 0 \; \forall x \in [a, b]$ folgt dann

$$\int_a^b f^2(x)dx > 0,$$

was für E4 zu zeigen war. $\diamond$

3.1.3 Definition. Ist V ein Vektorraum mit Skalarprodukt $\langle\cdot,\cdot\rangle$, dann heissen zwei Vektoren $\underline{v} \ne \underline{o}$ und $\underline{w} \ne \underline{o}$ aus V ***orthogonal*** zueinander, falls $\langle\underline{v}, \underline{w}\rangle = 0$ gilt. Wir schreiben dann $\underline{v} \perp \underline{w}$ $\diamond$

Man sagt auch, dass zwei zueinander orthogonale Vektoren einen *rechten Winkel* bezüglich des Skalarprodukts $\langle\cdot,\cdot\rangle$ bilden. Statt "orthogonal" ist natürlich auch die deutsche Bezeichnung "senkrecht" üblich. Somit ist in V ein Einheitswinkel definiert, und man kann damit eine Winkelmessung in $(V, \langle\cdot,\cdot\rangle)$ durch Unterteilung des rechten Winkels durchführen.

3.1.4 Beispiele.

1. In $V = \mathbb{R}^2$ mit dem euklidischen Skalarprodukt

$$\underline{x}^\mathsf{T}\underline{y} = x_1y_1 + x_2y_2$$

für $\underline{x} = (x_1, x_2)^\mathsf{T}$, $\underline{y} = (y_1, y_2)^\mathsf{T}$ ist Ihnen aus der Schule und aus der *Mathematik II* bekannt, dass

$$\underline{x}^\mathsf{T}\underline{y} = x_1y_1 + x_2y_2 = \|\underline{x}\| \, \|\underline{y}\| \cos\alpha$$

gilt, wobei α der Winkel zwischen $\underline{v}$ und $\underline{w}$ ist und $\|\underline{x}\| = \sqrt{x_1^2 + x_2^2}$ (analog $\|\underline{y}\|$) die (euklidische) Länge von $\underline{x}$ bezeichnet.

Damit erhalten wir die Standarddefinition der Orthogonalität, da $\underline{x}^\mathsf{T}\underline{y} = 0$ genau dann gilt, wenn $\cos\alpha = 0$ ist.

Ebenso erhalten wir die Standarddefinition des rechten Winkles, da bei Beschränkung auf $\alpha \in [0, \pi]$ gerade $\cos\alpha = 0$ für $\alpha = \pi/2$ gilt.

2. Sei nun $V = C[-\pi, \pi]$. Dann sind (für gegebenes $n \in \mathbb{N}$) in der Menge

$$\mathcal{W}(n) := \{1, \cos x, \cos 2x, \ldots, \cos nx, \sin x, \sin 2x, \ldots, \sin nx\} \subset C[-\pi, \pi]$$

je zwei Vektoren zueinander orthogonal bezüglich des Skalarprodukts (3.4), $a = -\pi$, $b = \pi$.

Allgemein gilt für ganze Zahlen $\mu, \nu \geq 1$ nämlich, dass

$$\left.\begin{array}{l}
\int_{-\pi}^{\pi} \sin \mu x \ \sin \nu x dx = \begin{cases} 0, & \text{falls } \mu \neq \nu \\ \pi & \text{falls } \mu = \nu \end{cases} \\
\int_{-\pi}^{\pi} \sin \mu x \ \cos \nu x dx = 0 \\
\int_{-\pi}^{\pi} \cos \mu x \ \cos \nu x dx = \begin{cases} 0, & \text{falls } \mu \neq \nu \\ \pi & \text{falls } \mu = \nu \end{cases} \\
\int_{-\pi}^{\pi} 1 \cdot \sin \mu x dx \quad = \quad 0 \\
\int_{-\pi}^{\pi} 1 \cdot \cos \mu x dx \quad = \quad 0 \\
\int_{-\pi}^{\pi} 1 dx \quad = \quad 2\pi \,,
\end{array}\right\} \tag{3.5}$$

vgl. *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984*, Beispiel 1.11 b. Zur Übung sei Ihnen empfohlen, diese Integrale (zumindest für die Menge $\mathcal{W}(2)$) nachzurechnen. ◇

3.1.5 Übung. Sei C die folgende Diagonalmatrix

$$C = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 4 \end{pmatrix}.$$

(i) Zeigen Sie, ohne auf Beispiel 2. in Punkt 3.1.2 zurückzugreifen, dass durch

$$\langle \underline{x}, \underline{y} \rangle := \underline{x}^\mathsf{T} C \underline{y} \tag{3.6}$$

für $\underline{x} = (x_1, x_2)^\mathsf{T}$, $\underline{y} = (y_1, y_2)^\mathsf{T}$ ein Skalarprodukt im $\mathbb{R}^2$ definiert ist.

(ii) Machen Sie sich klar, dass die Vektoren $(1,1)^\mathsf{T}$ und $(1,-1)^\mathsf{T}$ bezüglich des Skalarprodukts (3.6) nicht aufeinander senkrecht stehen und bestimmen Sie alle Vektoren, die senkrecht zu $(1,1)^\mathsf{T}$ bezüglich des Skalarprodukts (3.6) sind.

◇

3.1.6 Definition. Ist V ein beliebiger Vektorraum, dann heisst eine auf V definierte reellwertige Funktion $\underline{v} \in V \mapsto \|\underline{v}\| \in \mathbb{R}$ eine ***Norm auf V*** und die Zahl $\|\underline{v}\|$ ***Norm von*** $\underline{\boldsymbol{v}}$, falls die folgenden Eigenschaften erfüllt sind:

P1. $\|\underline{v}\| > 0 \quad \forall \underline{v} \in V \setminus \{\underline{o}\}$ (*Definitheit*)

P2. $\|\lambda \underline{v}\| \ = \ |\lambda| \, \|\underline{v}\|$ für alle $\lambda \in \mathbb{R}$ und $\underline{v} \in V$, (*Homogenität*)

P3. $\|\underline{v} + \underline{w}\| \leq \|\underline{v}\| + \|\underline{w}\|$ für alle $\underline{v}, \underline{w} \in V$, (*Dreiecksungleichung*).

Aus P1 und P2 folgt, dass $\|\underline{v}\| = 0$ genau dann, wenn $\underline{v} = \underline{o}$. ◇

Eine wichtige Klasse von Normen sind die durch ein Skalarprodukt erzeugten. Dass für sie die Axiome P1-P3 erfüllt sind, werden wir anschliessend gleich begründen.

3.1.7 Definition. Sei V ein beliebiger Vektorraum mit einem Skalarprodukt $\langle \cdot, \cdot \rangle$. Dann heisst für $\underline{v} \in V$

$$\|\underline{v}\| = \sqrt{\langle \underline{v}, \underline{v} \rangle}$$

die ***durch das Skalarprodukt*** $\langle \cdot, \cdot \rangle$ ***induzierte Norm*** von $\underline{v}$. $\diamond$

3.1.8 Satz. *Sei V ein beliebiger Vektorraum mit einem Skalarprodukt $\langle \cdot, \cdot \rangle$ und dadurch induzierter Norm $\|\cdot\|$. Dann gilt für beliebige $\underline{v}, \underline{w} \in V$ die Cauchy-Schwarz'sche Ungleichung*

$$|\langle \underline{v}, \underline{w} \rangle| \leq \|\underline{v}\| \|\underline{w}\|.$$

$\diamond$

Zum Beweis vgl. Vorlesung und *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984*, Satz 1.25.

3.1.9 Lemma. (Rechtfertigungslemma) *Jede durch ein Skalarprodukt in einem Vektorraum V induzierte Norm ist eine Norm im Sinne von Definition 3.1.6.* $\diamond$

Beweis: P1 folgt sofort aus E4 in der Definition des Skalarprodukts. Aus E2 und E1 in der Definition des Skalarprodukts folgt sofort

$$\langle \lambda\underline{v}, \lambda\underline{v} \rangle = \lambda \langle \underline{v}, \lambda\underline{v} \rangle = \lambda \langle \lambda\underline{v}, \underline{v} \rangle = \lambda^2 \langle \underline{v}, \underline{v} \rangle$$

und somit P2. Zum Beweis der Dreiecksungleichung P3 nutzen wir die Cauchy-Schwarz'sche Ungleichung und die Definitionen eines Skalarprodukts und seiner induzierten Norm wie folgt aus:

$$\begin{aligned}
\|\underline{v} + \underline{w}\|^2 &= \langle \underline{v} + \underline{w}, \underline{v} + \underline{w} \rangle \\
&= \|\underline{v}\|^2 + 2\langle \underline{v}, \underline{w} \rangle + \|\underline{w}\|^2 \\
&\leq \|\underline{v}\|^2 + 2|\langle \underline{v}, \underline{w} \rangle| + \|\underline{w}\|^2 \\
&\leq \|\underline{v}\|^2 + 2\|\underline{v}\|\|\underline{w}\| + \|\underline{w}\|^2,
\end{aligned}$$

was $\|\underline{v} + \underline{w}\|^2 \leq (\|\underline{v}\| + \|\underline{w}\|)^2$ und damit P3 ergibt. $\square$

3.1.10 Beispiele.

1. Sei $V = \mathbb{R}^n$ mit Elementen $\underline{x} = (x_1, \ldots, x_n)^\mathsf{T}$. Normen auf $\mathbb{R}^n$ sind zum Beispiel die ***euklidische Norm***

$$\|\underline{x}\|_2 := \sqrt{\underline{x}^\mathsf{T}\underline{x}} = \sqrt{\sum_{j=1}^{n} {x_j}^2},$$

aber auch die ***Maximumnorm***

$$\|\underline{x}\|_\infty := \max_{1\le j\le n} |x_j|$$

sowie die ***Summennorm***

$$\|\underline{x}\|_1 := \sum_{j=1}^{n} |x_j|.$$

Die Funktion $\|\cdot\|_2$ ist in der Tat eine Norm, nämlich nach dem vorangegangenen Rechtfertigungslemma, denn sie basiert auf dem euklidischen Skalarprodukt. Zur Übung sei Ihnen empfohlen, auch für die Funktionen $\|\cdot\|_\infty$ und $\|\cdot\|_1$ die Normeigenschaften nachzuprüfen. Das ist mit elementaren Mittel möglich. Durch ein Skalarprodukt sind sie aber nicht induziert, für die Maximumnorm vergleiche man dazu *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984*, Beispiel 1.10a.

2. Sei nun $V = C[a,b]$. Zu $f,g \in C[a,b]$ ist durch das Skalarprodukt (vgl. 3.4)

$$\langle f,g\rangle := \int_a^b f(x)g(x)dx$$

die sogenannte ***L_2-Norm***

$$\|f\|_2 := \left(\int_a^b f^2(x)\,dx\right)^{\frac{1}{2}}$$

induziert. Weitere Normen in $C[a,b]$ (die sich allerdings nicht durch ein Skalarprodukt erzeugen lassen, vgl. *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984*, Beispiel 1.10b) sind die sogenannte ***L_1-Norm***

$$\|f\|_1 := \int_a^b |f(x)|\,dx$$

und die sogenannte ***L_∞-Norm***

$$\|f\|_\infty := \sup_{a\le x\le b} |f(x)|.$$

Der Nachweis der Normeigenschaften sei wieder den Lesern überlassen. ◇

3.1.11 Übung. Definieren Sie die durch das Skalarprodukt in Übung 3.1.5 induzierte Norm im $\mathbb{R}^2$. ◇

Interessant ist, wann die Cauchy-Schwarz'sche Ungleichung und die Dreicksungleichung als Gleichung erfüllt sind, vgl. *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984*, §1.5 und die Vorlesung.

3.1.12 Satz. *Sei V ein beliebiger Vektorraum mit einem Skalarprodukt $\langle\cdot,\cdot\rangle$ und dadurch induzierter Norm $\|\cdot\|$. Dann gilt für $\underline{v},\underline{w}\in V$ in der Cauchy-Schwarz'schen Ungleichung die Gleichheit*

$$|\langle\underline{v},\underline{w}\rangle| = \|\underline{v}\|\|\underline{w}\|$$

genau dann, wenn $\{\underline{v},\underline{w}\}$ linear abhängig ist. ◇

3.1.13 Satz. *Sei V ein beliebiger Vektorraum mit einem Skalarprodukt $\langle\cdot,\cdot\rangle$ und dadurch induzierter Norm $\|\cdot\|$. Dann gilt für $\underline{v},\underline{w}\in V$ in der Dreiecksungleichung die Gleichheit*

$$\|\underline{v}+\underline{w}\| = \|\underline{v}\| + \|\underline{w}\|$$

genau dann, wenn es eine Zahl $\lambda \geq 0$ gibt, so dass $\underline{v}=\lambda\underline{w}$ oder $\underline{w}=\lambda\underline{v}$ gilt (d.h., wenn $\{\underline{v},\underline{w}\}$ positiv linear abhängig sind). ◇

3.2 Orthonormalsysteme

3.2.1 Satz. *Sei V ein Vektorraum mit Skalarprodukt $\langle\cdot,\cdot\rangle$. Ist I eine beliebige Indexmenge und erfüllt*

$$\mathcal{B} = \{\underline{v}^i \,|\, i\in I\}\subset V \quad \text{mit} \quad \underline{v}^i \neq \underline{o}\ \forall i\in I \quad \text{und} \quad \langle\underline{v}^i,\underline{v}^j\rangle = 0\ \forall i,j\in I : i\neq j,$$

dann ist $\mathcal{B}$ linear unabhängig. ◇

Beweis: Sei $\mathcal{C}$ eine endliche Teilmenge von $\mathcal{B}$, o.B.d.A. $\mathcal{C}=\{\underline{v}^1,\ldots,\underline{v}^n\}$. Zu zeigen ist:

$$\sum_{j=1}^n \lambda_j\underline{v}^j = \underline{o} \;\Rightarrow\; \lambda_1 = \ldots = \lambda_n = 0.$$

Aus $\sum_{j=1}^n \lambda_j\underline{v}^j = \underline{o}$ folgt für ein beliebiges festgehaltenes $i\in\{1,\ldots,n\}$ nach Voraussetzung und den Gesetzen für Skalarprodukte

$$0 = \langle\underline{o},\underline{v}^i\rangle = \langle\sum_{j=1}^n \lambda_j\underline{v}^j,\underline{v}^i\rangle = \sum_{j=1}^n \lambda_j\langle\underline{v}^j,\underline{v}^i\rangle = \lambda_i\langle\underline{v}^i,\underline{v}^i\rangle,$$

was wegen $\langle\underline{v}^i,\underline{v}^i\rangle > 0$ (da $\underline{v}^i\neq\underline{o}$ vorausgesetzt ist) sofort $\lambda_i = 0$ ergibt. Da $i\in\{1,\ldots,n\}$ beliebig war, folgt $\lambda_1=\ldots=\lambda_n=0$, was zu zeigen war. □

3.2.2 Definition. Sei V ein Vektorraum mit Skalarprodukt $\langle\cdot,\cdot\rangle$ und dadurch induzierter Norm $\|\cdot\|$. Eine Menge

$$\mathcal{B} = \{\underline{v}^i \,|\, i\in I\}\subset V \quad \text{mit} \quad \|\underline{v}^i\| = 1\ \forall i\in I \quad \text{und} \quad \langle\underline{v}^i,\underline{v}^j\rangle = 0\ \forall i,j\in I : i\neq j,$$

heisst ein ***Orthonormalsystem*** (kurz ***ONS***) in V. Ist $\mathcal{B}$ überdies eine Basis von V so heisst $\mathcal{B}$ ***Orthonormalbasis*** von V. ◇

3.2.3 Komponentenvektor in einer Orthonormalbasis. Sei $\mathcal{B} = \{\underline{v}^1, \ldots, \underline{v}^n\}$ eine Basis eines n-dimensionalen Vektorraums V, die zugleich ONS ist. Dann gilt für die bezüglich $\mathcal{B}$ eindeutig bestimmten Komponenten $x_1, \ldots, x_n$ eines beliebigen Vektors $\underline{w}$ aus V

$$\underline{w} = \sum_{j=1}^{n} x_j \underline{v}^j$$

und somit für ein beliebiges, aber festes $i \in \{1, \ldots, n\}$

$$\langle \underline{w}, \underline{v}^i \rangle = \langle \sum_{j=1}^{n} x_j \underline{v}^j, \underline{v}^i \rangle = \sum_{j=1}^{n} x_j \langle \underline{v}^j, \underline{v}^i \rangle = x_i \|\underline{v}^i\|^2,$$

d.h., da i beliebig war

$$x_i = \langle \underline{w}, \underline{v}^i \rangle, \ i = 1, \ldots, n.$$

Damit haben die Komponenten eines Vektors in einer Orthonormalbasis eine besonders einfache Darstellung. Das benutzt man bei der orthogonalen Projektion eines Punktes auf einen Unterraum sowie bei der Konstruktion eines ONS aus einem beliebigen linear unabhängigen System.

3.2.4 Übung. Durch $\underline{x} \mapsto F(\underline{x}) := A\underline{x}$ mit

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}, \quad \text{also} \quad A_{\bullet 1} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \ A_{\bullet 2} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \ A_{\bullet 3} = \begin{pmatrix} 2 \\ 1 \\ 1 \\ 1 \end{pmatrix},$$

ist eine lineare Abbildung von $\mathbb{R}^3$ in $\mathbb{R}^4$ definiert. Bestimmen Sie eine Orthonormalbasis von $F(\mathbb{R}^3)$.

Wir sind hier in der Situation, dass der fragliche Vektorraum $V = F(\mathbb{R}^3)$ ein Unterraum des $\mathbb{R}^4$ ist, wir arbeiten hier also mit dem euklidischen Skalarprodukt und der euklidischen Norm $\|\cdot\| = \|\cdot\|_2$ im $\mathbb{R}^4$. Offenbar ist $\operatorname{rg} A = 2$, da $A_{\bullet 3} = A_{\bullet 1} + A_{\bullet 2}$ und $\{A_{\bullet 1}, A_{\bullet 2}\}$ linear unabhängig. Somit ist z.B. $\mathcal{C} = \{A_{\bullet 1}, A_{\bullet 2}\}$ eine Basis des Spaltenraums von A, also von $F(\mathbb{R}^3)$. Offenbar ist $\mathcal{C}$ kein ONS.

Beim Lösen dieser Aufgabe kann man so vorgehen. Wir betrachten $\underline{x} = A_{\bullet 1}/\|A_{\bullet 1}\|$ als ersten Vektor der gesuchten ONS. Da die Dimension von $F(\mathbb{R}^3)$ gleich 2 ist, suchen wir also einen Vektor

$$\underline{\widetilde{y}} = (y_1, y_2, y_3, y_4)^\mathsf{T} \text{ mit } \underline{\widetilde{y}} \perp \underline{x}, \ \|\underline{\widetilde{y}}\| = 1 \ \text{ und } \ \underline{\widetilde{y}} \in \operatorname{lin}\{A_{\bullet 1}, A_{\bullet 2}\},$$

also lösen wir das LGS in den Variablen $y_1, \ldots, y_4, \lambda_1, \lambda_2$:

$$y_1 + y_3 + y_4 = 0 \ \text{ und } \ y_1 = \lambda_1 + \lambda_2, \ y_2 = \lambda_2, \ y_3 = \lambda_1, \ y_4 = \lambda_1.$$

Man erhält als allgemeine Lösung mit $t = \lambda_1$:

$$\lambda_2 = -3t \quad \text{und} \quad \underline{y} = (y_1, y_2, y_3, y_4)^\mathsf{T} = t(-2,-3,1,1)^\mathsf{T},\ t \in \mathbb{R},$$

so dass der gesuchte normierte Vektor $\widetilde{\underline{y}}$ sich mit

$$\widetilde{\underline{y}} = (-2,-3,1,1)^\mathsf{T}/\sqrt{15}$$

ergibt (da $\sqrt{15} = \|(-2,-3,1,1)^\mathsf{T}\|$ in der eukidischen Norm).

$\diamond$

3.2.5 Übung. Wir betrachten wie in Teil 2 von Beispiel 3.1.4 die Menge $\mathcal{W}(n) \subset C[-\pi,\pi]$ und den Unterraum

$$U = \operatorname{lin} \mathcal{W}(n)$$

von $C[-\pi,\pi]$. Zeigen Sie, dass die folgende Menge $\widetilde{\mathcal{W}}(n)$ eine Orthonormalbasis von U ist:

$$\widetilde{\mathcal{W}}(n) := \left\{ \frac{1}{\sqrt{2\pi}} \right\} \cup \left\{ \frac{1}{\sqrt{\pi}} \cos kx, \frac{1}{\sqrt{\pi}} \sin kx \right\}_{k=1}^{n} \subset C[-\pi,\pi]$$

Sie dürfen dabei die Formeln (3.5) benutzen. $\diamond$

3.2.6 Übung. Sei A eine $(n \times n)$-Matrix, deren Zeilen ein ONS bilden. Berechnen Sie

$$AA^\mathsf{T}\ , A^{-1}\ , \text{die Determinante von } A.$$

$\diamond$

3.3 Lineare Approximationsprobleme

3.3.1 Sei V ein Vektorraum mit Norm $\|\cdot\|$ und W ein endlichdimensionaler Umterraum von V. Dann lautet das sogenannte ***lineare Approximationsproblem*** wie folgt:

Gegeben ist ein Vektor $\underline{v} \in V$.

Bestimme einen Vektor $\underline{w}^* \in W$, so dass

$$\|\underline{v} - \underline{w}^*\| \le \|\underline{v} - \underline{w}\| \quad \text{für alle } \underline{w} \in W.$$

Formal schreiben wir das als Optimierungsaufgabe

$$\min_{\underline{w} \in W} \|\underline{v} - \underline{w}\|. \tag{3.7}$$

3.3.2 Beispiel. Lösen Sie die Aufgabe (3.7) für

$$V = \mathbb{R}^2,\ W = \{(x,y) \in \mathbb{R}^2 \,|\, x + 2y = 0\} \text{ und } \underline{v} = (4,1)^\mathsf{T}$$

bei Annahme verschiedener Normen, und zwar $\|\cdot\|_1$ (Summennorm), $\|\cdot\|_2$ (euklidische Norm) sowie $\|\cdot\|_\infty$ (Maximumnorm). ◇

3.3.3 Satz. *Sei V ein Vektorraum mit Skalarprodukt $\langle\cdot,\cdot\rangle$ und dadurch induzierter Norm $\|\cdot\|$. Seien $\underline{v} \in V$ sowie $\mathcal{B} = \{\underline{w}^1, \dots \underline{w}^n\}$ ein ONS in V und $W = \operatorname{lin}\mathcal{B}$. Dann hat das lineare Approximationsproblem (3.7) die eindeutige Lösung $\underline{w}^* = \underline{w}(\underline{v})$ mit*

$$\underline{w}(\underline{v}) = \textstyle\sum_{j=1}^n \langle \underline{v}, \underline{w}^j\rangle\, \underline{w}^j. \qquad \diamond$$

3.3.4 Definition. Unter den Voraussetzungen von Satz 3.3.3 heissen

- die Zahlen $\alpha_j = \langle \underline{v}, \underline{w}^j\rangle\ \forall j$ ***Fourierkoeffizienten***,
- der Vektor $\underline{w}(\underline{v})$ (orthogonale) ***Projektion*** von $\underline{v}$ auf W,
- der Vektor $\underline{v} - \underline{w}(\underline{v})$ das ***Lot*** von $\underline{v}$ auf W. ◇

Beweis von Satz 3.3.3: Zu lösen ist das Optimierungsproblem

$$\min_{(\alpha_1,\dots,\alpha_n)} \langle \underline{v} - \sum_{j=1}^n \alpha_j \underline{w}^j\ ,\ \underline{v} - \sum_{j=1}^n \alpha_j \underline{w}^j\rangle,$$

wobei $\mathcal{B} = \{\underline{w}^1, \dots \underline{w}^n\}$ das gegebene ONS in V ist. Man berechnet - unter Ausnutzung von $\langle \underline{w}^i, \underline{w}^j\rangle = 0$ $(i \neq j)$ und $\langle \underline{w}^j, \underline{w}^j\rangle = 1$, dass

$$\begin{aligned} f(\alpha_1,\dots,\alpha_n) &:= \langle \underline{v} - \sum_{j=1}^n \alpha_j \underline{w}^j\ ,\ \underline{v} - \sum_{j=1}^n \alpha_j \underline{w}^j\rangle. \\ &= \langle \underline{v}, \underline{v}\rangle + \sum_{i=1}^n\sum_{j=1}^n \alpha_i\alpha_j \langle \underline{w}^i, \underline{w}^j\rangle - 2\sum_{j=1}^n \alpha_j \langle \underline{v}, \underline{w}^j\rangle \\ &= \langle \underline{v}, \underline{v}\rangle + \sum_{j=1}^n \alpha_j^2 - 2\sum_{j=1}^n \alpha_j \langle \underline{v}, \underline{w}^j\rangle. \end{aligned}$$

Da die Zielfunktion f eine in $(\alpha_1,\dots,\alpha_n)$ konvexe Funktion ist, ist notwendig und hinreichend für die Optimalität, dass $\frac{\partial f}{\partial \alpha_j} = 0$ für alle j gilt, also

$$\frac{\partial f}{\partial \alpha_j} = 2\alpha_j - 2\langle \underline{v}, \underline{w}^j\rangle = 0 \quad\Leftrightarrow\quad \alpha_j = \langle \underline{v}, \underline{w}^j\rangle \qquad (j = 1,\dots,n),$$

was zu zeigen war. □

3.3.5 Gram-Schmidt'sche Orthonormierung - geometrische Idee.

Seien V ein Vektorraum mit Skalarprodukt $\langle\cdot,\cdot\rangle$ und dadurch induzierter Norm $\|\cdot\|$ sowie

$$\mathcal{A} = \{\underline{w}^1, \ldots, \underline{w}^n\} \subset V \text{ linear unabhängig.}$$

Konstruktion eines ONS $\mathcal{B} = \{\underline{u}^1, \ldots, \underline{u}^n\}$ nach folgender Idee:

- Man normiere zunächst den Vektor $\underline{w}^1$, d.h., man bilde den Vektor

$$\underline{u}^1 := \frac{1}{\|\underline{w}^1\|}\underline{w}^1.$$

- Nun betrachte man den Unterraum (von V)

$$W^1 = \operatorname{lin}\{\underline{u}^1\}$$

und projiziere den Vektor $\underline{w}^2$ orthogonal auf W^1, man nenne die Projektion $\underline{w}(\underline{w}^2)$, d.h.,

$$\underline{w}(\underline{w}^2) = \langle\underline{w}^2, \underline{u}^1\rangle\underline{u}^1. \qquad \text{Setze } \underline{u}^2 = \frac{1}{\|\underline{w}^2 - \underline{w}(\underline{w}^2)\|}\left(\underline{w}^2 - \underline{w}(\underline{w}^2)\right).$$

Offenbar ist $\{\underline{u}^1, \underline{u}^2\}$ nach Konstruktion eine Basis des Unterraums (von V)

$$W^2 = \operatorname{lin}\{\underline{w}^1, \underline{w}^2\},$$

denn $\underline{u}^1$, $\underline{u}^2$ sind orthogonal und somit linear unabhängig, ausserdem Linearkombinationen von $\underline{w}^1, \underline{w}^2$. Es ist offenbar eine Orthonormalbasis.

- Nun betrachte man den Unterraum W^2 und projiziere den Vektor $\underline{w}^3$ orthogonal auf W^2, man nenne die Projektion $\underline{w}(\underline{w}^3)$, d.h.,

$$\underline{w}(\underline{w}^3) = \sum_{j=1}^{2}\langle\underline{w}^3, \underline{u}^j\rangle\underline{u}^j. \qquad \text{Setze } \underline{u}^3 = \frac{1}{\|\underline{w}^3 - \underline{w}(\underline{w}^3)\|}\left(\underline{w}^3 - \underline{w}(\underline{w}^3)\right).$$

Wieder ist offenbar $\widehat{\mathcal{B}} = \{\underline{u}^1, \underline{u}^2, \underline{u}^3\}$ eine Basis des Unterraums (von V)

$$W^3 = \operatorname{lin}\{\underline{w}^1, \underline{w}^2, \underline{w}^3\},$$

denn die Vektoren von $\widehat{\mathcal{B}}$ stehen paarweise aufeinander senkrecht, sind also linear unabhängig, und sie sind nach Konstruktion Linearkombinationen von $\underline{w}^1, \underline{w}^2, \underline{w}^3$.

- Diese Konstruktion setzen wir fort, indem wir jeweils einen weiteren Vektor aus $\mathcal{A}$ hinzunehmen, am Ende haben wir eine Orthonormalbasis von $\operatorname{lin}\mathcal{A}$ erzeugt.

3.3.6 Existenz von Orthonormalbasen. Wir haben soeben bewiesen:

> In jedem endlichdimensionalen Vektorraum mit Skalarprodukt existiert eine Orthonormalbasis, man kann sie z.B. mit Hilfe des *Orthonormierungsverfahrens von Gram und Schmidt* aus einer gegebenen Basis konstruieren.

Die Idee liesse sich auch auf eine abzählbar unendliche Menge $\mathcal{A}$ übertragen. Dazu und zu einer algorithmischen Darstellung der Gram-Schmidt'schen Orthonormierung vergleiche man *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984*, Satz 1.30. $\diamond$

3.3.7 Beispiel. Die lineare Approximation einer stetigen Funktion $f \in C[-\pi, \pi]$ bezüglich des Unterraums $W = \operatorname{lin} \mathcal{W}(n)$ mit

$$\mathcal{W}(n) := \{1, \cos x, \cos 2x, \ldots, \cos nx, \sin x, \sin 2x, \ldots, \sin nx\}$$

wird häufig verwendet im Zusammenhang mit Schwingungen, saisonalen Schwankungen, Konjunkturzyklen und anderen periodischen Vorgängen.

Fourierreihen und lineare Approximation: Für gegebenes $f \in C[-\pi, \pi]$ und bei Beachtung von $W = \operatorname{lin} \widetilde{\mathcal{W}}(n)$, wobei $\widetilde{\mathcal{W}}(n)$ das Orthonormalsystem aus Übung 3.2.5 ist, hat das lineare Approximationsproblem (in der L_2-Norm $\|\cdot\|_2$)

$$\min_{g \in W} \|f - g\|_2$$

die eindeutige Lösung $g = f_n$ mit

$$f_n(x) = \frac{a_0}{2} + \sum_{k=1}^{n} a_k \cos kx + \sum_{k=1}^{n} b_k \sin kx$$

(die n-te Partialsumme einer sogenannten *Fourierreihe*) mit

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos kx \, dx \quad (k = 0, 1, \ldots, n), \tag{3.8}$$

$$b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin kx \, dx \quad (k = 1, \ldots, n). \tag{3.9}$$

Das kann aus Satz 3.3.3 mit Hilfe der Formeln (3.5) und Übung 3.2.5 abgeleitet werden, wir verzichten hier auf die Details. Interessenten verweisen wir auf Beispiel 1.13 in *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984*.

Man könnte allgemeiner auch den Vektorraum $C[a, b]$ statt $C[-\pi, \pi]$ betrachten. Für $t \in [a, b]$ führt die Variablensubstitution

$$x = \frac{2\pi}{b - a}(t - a) - \pi$$

auf das Intervall $[-\pi, \pi]$, womit sich die Ergebnisse von $C[-\pi, \pi]$ auf $C[a, b]$ direkt übertragen lassen.

3.3.8 Orthogonales Komplement.

Sei V ein Vektorraum mit Skalarprodukt $\langle\cdot,\cdot\rangle$ und W ein Unterraum von V. Dann heisst

$$W^\perp = \{\underline{v} \in V \mid \langle \underline{v}, \underline{w}\rangle = 0 \text{ für alle } \underline{w} \in W\}$$

das ***orthogonale Komplement*** von W.

Die Menge $W^\perp$ ist wieder ein Unterraum von V. Jeder Vektor $\underline{v} \in V$ lässt sich mittels orthogonaler Projektion zerlegen in

$$\underline{v} = \underline{w}(\underline{v}) + (\underline{v} - \underline{w}(\underline{v})) \quad \text{mit} \quad \underline{w}(\underline{v}) \in W,\ \underline{v} - \underline{w}(\underline{v}) \in W^\perp.$$

Umgekehrt löst jede Zerlegung

$$\underline{v} = \underline{w} + \underline{u} \quad \text{mit} \quad \underline{w} \in W,\ \underline{u} \in W^\perp$$

das lineare Approximationsproblem (3.7), vgl. *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984*, Seite 60. ◇

3.3.9 Orthogonale Projektion als lineare Abbildung.

Durch

$$\varphi(\underline{v}) := \underline{v} - \underline{w}(\underline{v}),\ \underline{v} \in V,$$

wobei $\underline{w}(\underline{v})$ wieder die orthogonale Projektion von $\underline{v} \in V$ auf den Unterraum W eines Vektorraums V mit Skalarprodukt $\langle\cdot,\cdot\rangle$ bezeichnet, ist eine lineare Abbildung von V in V definiert. Begründung: φ ist die Differenz der linearen Abbildung $\underline{v} \mapsto \underline{v}$ und der linearen Abbildung $\underline{w}(\cdot)$: Es gilt ja

$$\begin{aligned}
\underline{w}(\lambda_1\underline{v}^1 + \lambda_2\underline{v}^2) &= \sum_{j=1}^{n} \langle \lambda_1\underline{v}^1 + \lambda_2\underline{v}^2, \underline{w}^j\rangle\, \underline{w}^j \\
&= \lambda_1 \sum_{j=1}^{n} \langle \underline{v}^1, \underline{w}^j\rangle\, \underline{w}^j + \lambda_2 \sum_{j=1}^{n} \langle \underline{v}^2, \underline{w}^j\rangle\, \underline{w}^j \\
&= \lambda_1\underline{w}(\underline{v}^1) + \lambda_2\underline{w}(\underline{v}^2)
\end{aligned}$$

nach den Rechengesetzen für das Skalarprodukt. Offenbar gilt

$$\varphi(V) = W^\perp \qquad \text{und} \qquad \operatorname{kern}\varphi = W,$$

wobei die erste Beziehung aus der letzten Bemerkung unter 3.3.8 folgt und sich die zweite direkt aus der Definition von φ ergibt. ◇

3.3.10 Dimensionsformel. Ist V endlichdimensional, so gilt nach Punkt 3.3.9 und dem Dimensionssatz für lineare Abbildungen, vgl. Satz 2.3.17,

$$\dim W^{\perp} = \dim V - \dim W.$$

Ein bekanntes Beispiel für orthogonale Komplemente im $\mathbb{R}^n$ sind der Lösungsraum

$$L = \{\underline{x} \in \mathbb{R}^n \mid A\underline{x} = \underline{o}\}, \;\; A \in M(m,n) \text{ gegeben,}$$

eines homogenen LGS und der Zeilenraum der Koeffizientenmatrix A dieses LGS. $\diamond$

Kapitel 4

Eigenwerte

4.1 Eigenwerte und Diagonalisierung von Matrizen

Bevor wir den Begriff des Eigenwerts einer Matrix definieren, betrachten wir zunächst ein populäres ökonomisches Beispiel. Mathematisch handelt es sich um ein Beispiel einer sogenannten Markov-Kette.

4.1.1 Beispiel (Kundenwanderung). Auf einem Markt mit $n = 3$ Herstellern H_i $(i = 1, 2, 3)$ und z Kunden wird ein gewisses Produkt angeboten, z.B. Bier. Zu Beginn von Periode 1 seien die Kunden gemäss (z_1^0, z_2^0, z_3^0) mit $z_1^0 + z_2^0 + z_3^0 = z$ aufgeteilt, d.h., die Marktanteile x_i^0 der Hersteller berechnen sich zu

$$x_i^0 = z_i^0/z, \quad i = 1, 2, 3.$$

Der Vektor $\underline{x} = (x_1, x_2, x_3) \geq \underline{o}$ (mit $x_1 + x_2 + x_3 = 1$) heisst *Marktaufteilung*, $\underline{x}^0$ ist also die Marktaufteilung zu Beginn von Periode 1. Innerhalb der betrachteten Periode findet eine Kundenwanderung statt, und zwar

$$z_{ij} \text{ Kunden wandern von Hersteller } H_i \text{ zu Hersteller } H_j$$

(bzw. bleiben ihm treu, falls $i = j$), $i, j = 1, 2, 3$. Die *Wanderungszahl*

$$a_{ij} = z_{ij}/z_i^0, \; i, j = 1, 2, 3,$$

gibt also den Anteil der Kunden des Herstellers H_i an, die in der betrachteten Periode zum Herstellers H_j wandern (bzw. H_i treu bleiben, falls $i = j$), es sei z.B.

$$A = (a_{ij}) = \begin{pmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.4 & 0.3 \\ 0.1 & 0.4 & 0.5 \end{pmatrix}.$$

Da es sich um Anteile handelt, ist die Zeilensumme notwendig gleich 1. Die Anfangsaufteilung sei

$$\underline{x}^0 = \begin{pmatrix} 0.25 \\ 0.40 \\ 0.35 \end{pmatrix}.$$

Probleme:

1. Berechnen Sie sie Marktaufteilung $\underline{x}^1$ am Ende der Periode 1.

2. Auch in den folgenden $k \in \mathbb{N}$ Perioden sei die Kundenwanderung konstant. Berechnen Sie die Marktaufteilung $\underline{x}^3$ am Ende der Periode 3. Wie berechnet sich die Marktaufteilung $\underline{x}^k$ am Ende der Periode k?

3. Gibt es eine *stationäre* Marktaufteilung zu der gegebenen Wanderungsmatrix A, d.h. ein $\underline{x}^* \geq \underline{o}$ mit $x_1^*+x_2^*+x_3^* = 1$, so dass die Kundenwanderung innerhalb einer Periode die Marktaufteilung $\underline{x}^*$ nicht ändert?

Die Lösung der Probleme 1. und 2. ergibt sich, wie man sich leicht überlegt, mittels der Formel

$$\underline{x}^k = (A^T)^k \underline{x}^0, \; k = 1, 2, 3, \ldots.$$

Problem 3. führt auf die Frage der Existenz eines *Eigenvektors* $\underline{x}^*$ zum *Eigenwert* $\lambda = 1$ der Matrix A^T (d.h., die Frage ob es ein $\underline{x}^*$ mit

$$A^\mathsf{T} \underline{x}^* = 1 \cdot \underline{x}^*$$

gibt), wobei die Zusatzbedingungen $\underline{x}^* \geq \underline{o}$ und $x_1^*+x_2^*+x_3^* = 1$ ebenfalls erfüllt sind. Unser Beispiel erfüllt die Voraussetzungen eines Modells, in dem die gewünschten Eigenschaften gelten. Wir kommen weiter unten darauf zurück. $\diamond$

4.1.2 Definition. Sei A eine quadratische Matrix der Ordnung n. Gilt für eine reelle Zahl λ und einen Vektor $\underline{x}^* \in \mathbb{R}^n$ mit $\underline{x}^* \neq \underline{o}$ die Gleichung

$$A\underline{x}^* = \lambda \underline{x}^*,$$

so heissen λ ***(reeller) Eigenwert*** und $\underline{x}^*$ ein ***Eigenvektor*** zum Eigenwert λ. $\diamond$

Offenbar ist mit jedem Eigenvektor $\underline{x}^*$ zu einem Eigenwert λ auch $\alpha\underline{x}^*$ mit $\alpha \neq 0$ ein Eigenvektor zu λ.

Achtung! Gemäss Definition ist ein Eigenvektor stets verschieden vom Nullvektor. Der Eigenwert Null ist dagegen nicht verboten. Allerdings kann eine reguläre Matrix A nicht den Eigenwert Null haben, da das LGS $A\underline{x} = \underline{o}$ dann nur die triviale Lösung $\underline{x} = \underline{o}$ hat.

4.1.3 Beispiel. (Geometrische Interpretation) Sei

$$A = \begin{pmatrix} \cos\frac{\pi}{2} & \sin\frac{\pi}{2} \\ \sin\frac{\pi}{2} & -\cos\frac{\pi}{2} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Offenbar definiert A die Spiegelung des $\mathbb{R}^2$ an der Geraden $\{(x_1, x_2) | x_2 = x_1\}$ (Zeichnung!), und es gibt nur zwei Möglichkeiten, dass das Bild eines Vektors $\underline{x} \neq \underline{o}$ auf der von $\underline{x}$ erzeugten Gerade liegt (das ist dann ein Eigenvektor): $\underline{x}^1 = (1, 1)$ ist ein Eigenvektor zum Eigenwert $\lambda_1 = 1$, und $\underline{x}^2 = (1, -1)$ ist ein Eigenvektor zum Eigenwert $\lambda_1 = -1$ (Übungsaufgabe!). $\diamond$

4.1.4 Bemerkung. Nicht jede n-reihige Matrix besitzt einen reellen Eigenwert. Man betrachte die Matrix
$$A = \begin{pmatrix} 1 & -2 \\ 1 & -1 \end{pmatrix}.$$
Für jedes $\lambda \in \mathbb{R}$ und jeden Vektor $\underline{x} = (x_1, x_2)^\mathsf{T} \in \mathbb{R}^2$ gilt dann
$$A\underline{x} = \begin{pmatrix} x_1 - 2x_2 \\ x_1 - x_2 \end{pmatrix} \quad \text{und} \quad \lambda\underline{x} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \end{pmatrix}.$$
Die Beziehung $A\underline{x} = \lambda\underline{x}$ mit $\underline{x} \neq \underline{o}$ impliziert
$$(1-\lambda)x_1 = 2x_2, \quad (1+\lambda)x_2 = x_1$$
und damit $x_1 \neq 0$ und $x_2 \neq 0$. Die Multiplikation beider Gleichungen gibt
$$(1-\lambda^2)x_1x_2 = 2x_1x_2 \quad \text{mit} \quad x_1x_2 \neq 0,$$
also $\lambda^2 + 1 = 0$, d.h., λ kann nicht reell sein. $\diamond$

4.1.5 Bemerkung. Die Schwierigkeiten des Beispiels in 4.1.4 kann man überwinden, wenn die Vektoren $\underline{x}$ als Elemente des Vektorraums $\mathbb{C}^n$ der n-Tupel komplexer Zahlen (wobei die Multiplikation mit Skalaren die Multiplikation mit komplexen Zahlen ist) aufgefasst wird und man komplexe Eigenwerte betrachtet. Da in dieser Vorlesung die komplexe Analysis nicht vorausgesetzt werden kann, beschränken wir uns auf die Theorie reeller Eigenwerte.

Eine andere Verallgemeinerung ergibt sich, indem man die Begriffe *Eigenwert* und *Eigenvektor* auf *Endomorphismen*, also lineare Abbildungen eines Vektorraums V in sich selbst, bezieht: Ist $\varphi : V \to V$ linear, so heissen λ *Eigenwert von* φ und $\underline{v}^* \in V \setminus \{\underline{o}\}$ ein *Eigenvektor* zum Eigenwert λ, falls
$$\varphi(\underline{v}^*) = \lambda\underline{v}^*.$$
Im Spezialfall $V = \mathbb{R}^n$ mit $\varphi(\underline{x}) = A\underline{x}$ erhält man gerade die Definition 4.1.2. $\diamond$

4.1.6 Seien $\lambda \in \mathbb{R}$ und $A \in M(n,n)$ gegeben und
$$\operatorname{Eig}(A;\lambda) := \{\underline{x} \mid A\underline{x} = \lambda\underline{x}\}.$$
Offenbar ist stets $\underline{o} \in \operatorname{Eig}(A;\lambda)$ und
$$\operatorname{Eig}(A;\lambda) \text{ ist ein Unterraum von } \mathbb{R}^n. \text{ (Übungsaufgabe!)}$$
Nach Definition ist λ Eigenwert von A genau dann, wenn $\dim \operatorname{Eig}(A;\lambda) \geq 1$. Die Menge $\operatorname{Eig}(A;\lambda)$ heisst dann ***Eigenraum*** von A bezüglich des Eigenwerts λ. Da - mit der Einheitmatrix $I \in M(n,n)$ -
$$\operatorname{Eig}(A;\lambda) = \{\underline{x} \mid (A - \lambda I)\underline{x} = \underline{o}\}$$
gilt, ist $\dim \operatorname{Eig}(A;\lambda) = n - \operatorname{rg}(A - \lambda I)$. $\diamond$

4.1.7 Lemma. *Seien* $\|\cdot\|$ *irgendeine Norm im* $\mathbb{R}^n$, $A = (a_{ij}; i = 1, \dots, n,\ j = 1, \dots, n)$ *eine gegebene* n*-reihige Matrix und* λ *ein Eigenwert von* A. *Dann gilt*

$$|\lambda| \leq \max\{\|A\underline{x}\| \mid \|\underline{x}\| = 1\}. \tag{4.1}$$

(Die Zahl $\max\{\|A\underline{x}\| \mid \|\underline{x}\| = 1\}$ heisst *die der Vektornorm* $\|\cdot\|$ *zugeordnete Matrixnorm.*) *Speziell gilt*

$$|\lambda| \leq \max_{1\leq j\leq n} \sum_{i=1}^{n} |a_{ij}| \quad \textit{und} \quad |\lambda| \leq \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|, \tag{4.2}$$

d.h., $|\lambda|$ *ist kleiner als die grössten absoluten Spalten- und Zeilensummen von* A. ◇

Beweis: Ist λ ein Eigenwert und $\underline{x}^*$ ein dazugehöriger Eigenvektor, so gilt

$$\|A\underline{x}^*\| = \|\lambda\underline{x}^*\| = |\lambda|\|\underline{x}^*\|.$$

Setzen wir $\underline{z}^* = \|\underline{x}^*\|^{-1}\underline{x}^*$, so folgt

$$|\lambda| = \|A\underline{z}^*\| \quad \text{mit} \quad \|\underline{z}^*\| = 1,$$

was (4.1) impliziert. Dabei existiert $\max\{\|A\underline{x}\| \mid \|\underline{x}\| = 1\}$ nach dem Satz von Weierstrass (vgl. 4.2.2). Die Aussagen in (4.2) folgen, wenn man speziell $\|\cdot\| = \|\cdot\|_1$ (Summennorm) und $\|\cdot\| = \|\cdot\|_\infty$ (Maximumnorm) wählt, siehe die folgende Übung. □

4.1.8 Übung. Beweisen Sie die Aussagen in (4.2). ◇

4.1.9 Sei $A \in M(n, n)$. Die Zahl

$$P(t) = \det(A - tI)$$

ist ein Polynom n-ten Grades und heisst ***charakteristisches Polynom***.

In der Tat kommt in der Leibniz-Formel für $\det(A - tI)$ der Summand

$$(a_{11} - t) \cdot (a_{22} - t) \cdot \ldots \cdot (a_{nn} - t) \tag{4.3}$$

vor, also lässt sich dieser Summand als Summe des Terms $(-1)^n t^n$ und von Termen mit niedrigeren Potenzen von t schreiben. Alle anderen Summanden in der Leibniz-Formel sind ebenfalls Polynome (und zwar vom Grad $p \leq n - 2$). Damit ist $P(t)$ ein Polynom n-ten Grades.

Das charakteristische Polynom einer Diagonalmatrix Λ mit Einträgen λ_j auf der Hauptdiagonalen hat nach den Determinantengesetzen die Form

$$\det(\Lambda - tI) = \Pi_{j=1}^{n}(\lambda_j - t).$$

Als elementare Schlussfolgerung erhalten wir wegen $A^\mathsf{T} - tI = (A - tI)^\mathsf{T}$ und $\det(A - tI)^\mathsf{T} = \det(A - tI)$ sofort, dass A und A^T das gleiche charakteristische Polynom haben. ◇

4.1.10 Satz. *Die Zahl $\lambda \in \mathbb{R}$ ist genau dann Eigenwert der Matrix $A \in M(n,n)$, wenn $\det(A - \lambda I) = 0$ gilt, d.h., wenn λ eine Nullstelle des charakteristischen Polynoms ist.* $\diamond$

Beweis: Ist $\underline{x}^*$ ein Eigenvektor zu λ, so ist $\underline{x}^*$ nichttriviale Lösung des LGS $(A-\lambda I)\underline{x} = \underline{o}$. Also gilt $\operatorname{rg}(A-\lambda I) < n$ und somit $\det(A-\lambda I) = 0$. Umgekehrt impliziert $\det(A-\lambda I) = 0$, dass $\dim \operatorname{Eig}(A;\lambda) \geq 1$, d.h., λ ist - wie oben bemerkt - Eigenwert von A. $\square$

Für Beispiel 4.1.3 und das Beispiel in Bemerkung 4.1.4 lassen sich nach dem vorhergehenden Satz jetzt die Aussagen anhand der Nullstellen des charakteristischen Polnoms überprüfen. Aus 4.1.9 erhalten wir, dass A und A^T die gleichen Eigenwerte haben.

4.1.11 Satz. *Das charakteristische Polynom und damit auch die Eigenwerte einer Matrix sind gegenüber Ähnlichkeitstransformationen invariant.* $\diamond$

Beweis: Seien $A, B, T \in M(n,n)$ derart, dass T regulär ist und $B = T^{-1}AT$ gilt. Wegen

$$B - tI \;=\; T^{-1}AT \;-\; tT^{-1}T \;=\; T^{-1}(AT - tT) \;=\; T^{-1}(A - tI)T \qquad (t \in \mathbb{R})$$

folgt aus Satz 4.3.8 sofort $\det(B - tI) = \det(A - tI)$. $\square$

4.1.12 Definition. Eine Matrix $A \in M(n,n)$ heisst ***diagonalisierbar***, wenn eine zu A ähnliche Diagonalmatrix $\Lambda \in M(n,n)$ existiert. $\diamond$

Denken wir an die Repräsentation einer linearer Abbildung φ eines n-dimensionalen Vektorraums V in sich (bezüglich einer festen Basis in V) durch eine n-reihige Matrix, dann ist die Bedeutung der Diagonalisierbarkeit klar: Es gibt eine Koordinatentransformation (d.h., einen Basiswechsel), die es erlaubt, φ in den neuen Koordinaten durch eine Diagonalmatrix zu repräsentieren. Die Eigenwerttheorie gibt einen konstruktiven Weg zur Bestimmung dieser Koordinatentransformation (soweit sie existiert).

4.1.13 Lemma. *A ist genau dann diagonalisierbar, wenn es (nicht notwendig verschiedene) Eigenwerte $\lambda_1, \ldots, \lambda_n$ und eine Basis des $\mathbb{R}^n$ aus zugehörigen Eigenvektoren $\underline{x}^1, \ldots, \underline{x}^n$ gibt, so dass*

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix} = T^{-1}AT \quad \textit{mit} \quad T = [\underline{x}^1 \ldots \underline{x}^n] \ \textit{(spaltenweise)}$$

gilt. $\diamond$

Beweis. "wenn"-Richtung: Für die Eigenwerte λ_j von A und die zugehörigen Eigenvektoren $\underline{x}^j$, $j = 1, \ldots, n$ gilt

$$A\underline{x}^j = \lambda_j \underline{x}^j, \tag{4.4}$$

also mit T und Λ wie oben

$$AT = A[\underline{x}^1 \dots \underline{x}^n] = [\lambda_1 \underline{x}^1 \dots \lambda_n \underline{x}^n] = T\Lambda. \tag{4.5}$$

Wegen der linearen Unabhängigkeit von $\{\underline{x}^1, \dots, \underline{x}^n\}$ ist T regulär, und es folgt $T^{-1}AT = \Lambda$, also ist A diagonalisierbar.

"dann, wenn"-Richtung: Ist A diagonalisierbar, so gibt es eine reguläre $(n \times n)$-Matrix T und eine Diagonalmatrix Λ mit den Einträgen λ_j auf der Hauptdiagonalen, so dass (*) $T^{-1}AT = \Lambda$. Bezeichnen wir die n Spaltenvektoren von T mit $\underline{x}^1$, ..., $\underline{x}^n$, so folgt nach Multiplikation mit T auf beiden Seiten der Gleichung (*) die Formel (4.5) und damit auch (4.4) für alle j, was zu zeigen war. □

4.1.14 Lemma. *Gegeben seien $A \in M(n,n)$ und Eigenvektoren $\underline{x}^1, \dots, \underline{x}^k$ zu paarweise verschiedenen Eigenwerten $\lambda_1, \dots, \lambda_k$ von A. Dann ist $\{\underline{x}^1, \dots, \underline{x}^k\}$ linear unabhängig. Insbesondere ist notwendig $k \leq n$.* ◇

Beweis: durch vollständige Induktion. Für $k = 1$ folgt die Aussage nach Definition eines Eigenvektors (als Nicht-Nullvektor). Sei die Aussage für $k = m-1$ $(m \geq 2)$ bewiesen. Wir betrachten nun das LGS in $\underline{\alpha}$

$$\alpha_1 \underline{x}^1 + \alpha_2 \underline{x}^2 + \dots \alpha_m \underline{x}^m = \underline{o}. \tag{4.6}$$

Zu zeigen ist $\underline{\alpha} = \underline{o}$. Multiplikationen des LGS liefern einerseits

$$A(\alpha_1 \underline{x}^1 + \alpha_2 \underline{x}^2 + \dots \alpha_m \underline{x}^m) = \underline{o}$$

und

$$\lambda_1(\alpha_1 \underline{x}^1 + \alpha_2 \underline{x}^2 + \dots \alpha_m \underline{x}^m) = \underline{o}$$

andererseits. Durch Subtraktion folgt wegen $A\underline{x}^j = \lambda_j \underline{x}^j$ für alle j sofort

$$(\lambda_2 - \lambda_1)\alpha_2 \underline{x}^2 + \dots + (\lambda_m - \lambda_1)\alpha_m \underline{x}^m = \underline{o}.$$

Nach Induktionsvoraussetzung sind die $m-1$ Vektoren $\{\underline{x}^2, \dots, \underline{x}^m\}$ linear unabhängig, also folgt

$$(\lambda_2 - \lambda_1)\alpha_2 = 0, \ \dots \ , (\lambda_m - \lambda_1)\alpha_m = 0$$

und somit wegen $\lambda_j - \lambda_1 \neq 0$ für $j \neq 1$ auch

$$\alpha_2 = 0, \ \dots \ , \alpha_m = 0. \tag{4.7}$$

Da $\underline{x}^1$ Eigenvektor ist, gilt $\underline{x}^1 \neq \underline{o}$, d.h., (4.6) und (4.7) liefern auch $\alpha_1 = 0$, was zu zeigen war. Wegen $\dim \mathbb{R}^n = n$ folgt $k \leq n$. □

Die letzte Aussage im vorigen Satz (d.h., $k \leq n$) ist auch nach dem Fundamentalsatz der Algebra klar, da das charakteristische Polynom von Grad n ist.

4.1.15 Bemerkung. Hat eine quadratische Matrix A nicht-reelle Eigenwerte, so ist A nicht ähnlich zu einer reellen Diagonalmatrix, also im Sinne unserer Definition nicht diagonalisierbar, vgl. Lemma 4.1.13.

Nicht diagonalisierbar kann eine n-reihige Matrix A auch sein, wenn es mehrfache (reelle) Nullstellen des charakteristischen Polynoms gibt und es nicht n linear unabhängige Eigenvektoren von A gibt. Beispiel:

$$A = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}, \text{ d.h. } \det(A - \lambda I) = \begin{vmatrix} 1-\lambda & -1 \\ 1 & -1-\lambda \end{vmatrix} = -1 + \lambda^2 + 1,$$

also ist $\lambda = 0$ doppelter Eigenwert und jeder Eigenvektor (x_1, x_2) erfüllt $x_1 - x_2 = 0$, also hat der Eigenraum zum einzigen Eigenwert nur die Dimension 1.

Es ist aber nicht zwingend, dass mehrfache Eigenwerte die Diagonalisierbarkeit verhindern. **Zum einen werden wir im nächsten Abschnitt sehen, dass symmetrische Matrizen stets diagonalisierbar sind.** Zum anderen gibt es auch nichtsymmetrische Matrizen mit mehrfachen Eigenwerten und ausreichender Dimension der Eigenräume, man löse die folgende Übung. ◇

4.1.16 Übung. Man berechne für die Matrix

$$A = \begin{pmatrix} 0 & -1 & 1 \\ -3 & -2 & 3 \\ -2 & -2 & 3 \end{pmatrix}$$

die Eigenwerte und bestimme eine Basis des $\mathbb{R}^3$, bestehend aus Eigenvektoren des $\mathbb{R}^3$. ◇

Als Zusammenfassung bisheriger Erkenntnisse erhalten wir das folgende notwendige bzw. hinreichende Kriterium für die Diagonalisierbarkeit.

Die erste Aussage (notwendige Bedingung) folgt aus Satz 4.1.11 sowie Lemma 4.1.13 und der Formel für charakteristische Polynome von Diagonalmatrizen.

Die zweite Aussage (hinreichende Bedingung) folgt aus Lemma 4.1.14 und Satz 4.1.11, wenn wir beachten, dass n linear unabhängige Vektoren im $\mathbb{R}^n$ eine Basis des $\mathbb{R}^n$ bilden.

4.1.17 Satz. *Sei $A \in M(n,n)$. Dann gilt:*

(i) Ist A diagonalisierbar, so ist $\det(A - tI) = \Pi_{j=1}^n(\lambda_j - t)$, d.h., das charakteristische Polynom zerfällt in (reelle) Linearfaktoren.

(ii) Ist $\det(A - tI) = \Pi_{j=1}^n(\lambda_j - t)$ mit paarweise verschiedenen $\lambda_1, \ldots, \lambda_n$, so ist A diagonalisierbar.

Dabei sind jeweils $\lambda_1, \ldots, \lambda_n$ die Eigenwerte von A (in (i) nicht notwendig paarweise verschieden). ◇

Der vorhergehende Satz kann noch allgemeiner gefasst werden: A ist diagonalisierbar genau dann, wenn das charakteristische Polynom $P(t)$ in (reelle) Linearfaktoren zerfällt und zu jedem Eigenwert λ die Dimension des Eigenraums $\mathrm{Eig}\,(\lambda; A)$ gleich der algebraischen Vielfachheit der Nullstelle λ von $P(t)$ ist. Dann fällt der $\mathbb{R}^n$ mit der direkten Summe der Eigenräume $\mathrm{Eig}\,(\lambda_i; A)$ zu den $(k \leq n)$ paarweise verschiedenen Eigenwerten von A zusammen. Für Details vgl. *G.Fischer, Lineare Algebra, Vieweg, 2002*, Theorem 4.3.3.

4.1.18 Bemerkung. Für eine diagonalisierbare Matrix $A \in M(n, n)$ folgt aus Satz 4.1.17 sofort

$$\det A = \lambda_1 \cdot \ldots \cdot \lambda_n, \tag{4.8}$$

setze $t = 0$ in (i) von Satz 4.1.17.

Vergleicht man den Koeffizienten zu t^{n-1} in dieser Formel (i) mit den Koeffizienten zu t^{n-1} im ersten Summanden der Leibniz-Formel von $\det(A - tI)$ (nur dort kommt t^{n-1} vor), vgl. Formel (4.3), so folgt

$$\lambda_1 + \lambda_2 + \ldots + \lambda_n = \ \mathrm{spur}\, A := a_{11} + a_{22} + \ldots + a_{nn}. \tag{4.9}$$

Die Zahl $\mathrm{spur}\, A$ heisst ***Spur*** von A.

Die Formeln (4.8) und (4.9) können als Probe für die richtige Berechnung der Eigenwerte dienen. $\diamond$

4.1.19 Anwendung der Diagonalisierbarkeit: Potenzen von A. In unserem Beispiel 4.1.1 zur Kundenwanderung waren Potenzen einer Matrix zu berechnen. Besonders einfach lässt sich A^k berechnen, wenn A diagonalisierbar ist (das ist in Beispiel 4.1.1 der Fall), d.h., wenn

$$\Lambda = T^{-1}AT \ \text{ mit der Diagonalmatrix } \Lambda \text{ der Eigenwerte } \lambda_1, \ldots, \lambda_n.$$

Dann gilt

$$A^2 = AA = (T\Lambda T^{-1})(T\Lambda T^{-1}) = T\Lambda\Lambda T^{-1} = T\Lambda^2 T^{-1},$$

und so weiter bis

$$A^k = T\Lambda^k T^{-1}.$$

Die Eigenwerte von A^k sind $(\lambda_j)^k \ \forall j$, die Eigenräume sind die gleichen wie die von A. $\diamond$

4.2 Eigenwerte symmetrischer Matrizen

In diesem Abschnitt setzen wir durchweg voraus, dass A eine n-reihige symmetrische Matrix ist, also $A^{\mathsf{T}} = A$ ist.

4.2.1 Satz. *Sei A eine symmetrische n-reihige Matrix. Sind λ und μ zwei voneinander verschiedene Eigenwerte von A mit den zugehörigen Eigenvektoren $\underline{x}$ und $\underline{y}$, so sind $\underline{x}$ und $\underline{y}$ orthogonal.* $\diamond$

Zum Beweis sei nur angemerkt, dass $\underline{y}^{\mathsf{T}} A\underline{x} = \underline{x}^{\mathsf{T}} A\underline{y}$ wegen der Symmetrie von A gilt und deshalb mit $\underline{y}^{\mathsf{T}}\underline{x} = \underline{x}^{\mathsf{T}}\underline{y}$ die Beziehungen

$$0 = \underline{y}^{\mathsf{T}} A\underline{x} - \underline{x}^{\mathsf{T}} A\underline{y} = \underline{y}^{\mathsf{T}}(\lambda\underline{x}) - \underline{x}^{\mathsf{T}}(\mu\underline{y}) = (\lambda - \mu)\underline{x}^{\mathsf{T}}\underline{y}$$

folgen, was wegen $\lambda \neq \mu$ auf $\underline{x} \perp \underline{y}$ führt.

4.2.2 Satz. *Ist $A = (a_{ij})$ eine symmetrische n-reihige Matrix, so existiert mindestens ein reeller Eigenwert.* $\diamond$

Beweis: Wir verwenden zwei zentrale Sätze aus der Analysis: den Satz von Weierstrass über die Existenz von Extremalstellen für stetige Funktionen über beschränkten, abgeschlossenen Teilmengen des $\mathbb{R}^n$ (wird in der Vorlesung *Analysis für Ökonomen* im Sommersemester bewiesen) und den Satz von Lagrange (Lagrangemethode - ist aus der Vorlesung *Mathematik I* bekannt).

Die Funktion

$$f(x_1, \ldots, x_n) := \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j$$

ist offenbar stetig. Die Menge

$$K := \{\underline{x} \in \mathbb{R}^n | \underline{x}^{\mathsf{T}}\underline{x} = 1\}$$

ist beschränkt (denn in der euklidischen Norm gilt $\|\underline{x}\|_2 = 1$ für alle $\underline{x} \in K$) und abgeschlossen (das heisst nach Definition, dass für jede Folge $\{\underline{x}^k\} \subset K$ mit $\underline{x}^k \to \underline{x}^0$ auch $\underline{x}^0$ in K liegt). Nach dem Satz von Weierstrass existiert dann ein $\underline{x}^* \in K$ mit Komponenten x_j^*, so dass

$$f(x_1^*, \ldots, x_n^*) \geq f(x_1, \ldots, x_n) \quad \forall \underline{x} = (x_1, \ldots, x_n)^{\mathsf{T}} \in K,$$

d.h., $\underline{x}^*$ ist globaler Maximalpunkt von f bezüglich K.

Da für

$$g(x_1, \ldots, x_n) = \sum_{i=1}^{n}(x_i)^2$$

die partiellen Ableitungen von g in $\underline{x}$ sich als

$$\frac{\partial g}{\partial x_i}(x_1, \ldots, x_n) = 2x_i \qquad \forall i$$

berechnen, gilt wegen $\underline{x}^* \in K$

$$\frac{\partial g}{\partial x_\nu}(x_1^*, \ldots, x_n^*) \neq 0 \qquad \text{für mindestens ein } \nu \in \{1, \ldots, n\}.$$

Somit kann der Satz von Lagrange angewendet werden, und es existiert ein $\lambda \in \mathbb{R}$, so dass für alle i unter Ausnutzung der Symmetrie von A (d.h., $a_{ij} = a_{ji}$)

$$0 = \frac{\partial f}{\partial x_i}(x_1^*, \ldots, x_n^*) - \lambda \frac{\partial g}{\partial x_i}(x_1^*, \ldots, x_n^*) = 2 \sum_{j=1}^{n} a_{ij} x_j^* - 2\lambda x_i^*$$

gilt. In Vektorschreibweise mit $\underline{x}^* = (x_1^*, \ldots, x_n^*)^\mathsf{T}$ bedeutet das

$$A\underline{x}^* = \lambda \underline{x}^*,$$

was zu zeigen war. □

Wir zeigen nun zunächst für $n = 2$ den Satz über die Diagonalisierbarkeit symmetrischer Matrizen: Das charakteristische Polynom einer n-reihigen symmetrischen Matrix A hat, mit ihrer Vielfachheit gezählt, genau n reelle Nullstellen (das sind die Nullstellen von A), und es gibt zu diesen Eigenwerten ein Orthonormalsystem von n zugehörigen Eigenvektoren (die dann also eine Orthonormalbasis des $\mathbb{R}^n$ bilden). Gemäss Lemma 4.1.13 ist A folglich diagonalisierbar, d.h., ähnlich zur Diagonalmatrix aus den Eigenwerten. Die Ähnlichkeitstransformation erfolgt dann mit einer Orthogonalmatrix T (d.h., $T^\mathsf{T} T = I$).

4.2.3 Satz. *Jede 2-reihige symmetrische Matrix hat, mit ihrer Vielfachheit gezählt, zwei (reelle) Eigenwerte, und es gibt ein Orthonormalsystem von zwei zugehörigen Eigenvektoren.* ◇

Der Beweis wird in der Vorlesung gegeben, vgl. auch *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984*, Satz 3.23.

4.2.4 Übung. Man illustriere Satz 4.2.3 durch Beispiel 4.1.3 (Spiegelung an der Geraden $\{\underline{x} | x_1 = x_2\}$). ◇

4.2.5 Übung. Man illustriere Satz 4.2.3 für die durch

$$f(x_1, x_2) = 5x_1^2 + 8x_1 x_2 + 5x_2^2$$

gegebene quadratische Form in Beispiel 2.4.1. Offenbar gilt mit $\underline{x}^\mathsf{T} = (x_1, x_2)$ die Beziehung

$$f(x_1, x_2) = \underline{x}^\mathsf{T} A \underline{x} \qquad \text{mit } A = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}.$$

◇

4.2.6 Übung. Berechnen Sie ein ONS von Eigenvektoren der Matrix $A = \begin{pmatrix} 3 & \sqrt{2} \\ \sqrt{2} & 1 \end{pmatrix}$. ◇

Die Beweise der folgenden beiden Sätze werden in der Vorlesung gegeben, vgl. auch *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984*, Satz 3.24 und Korollar 3.25.

4.2.7 Satz. *Jede n-reihige symmetrische Matrix hat, mit ihrer Vielfachheit gezählt, genau n (reelle) Eigenwerte, und es gibt ein Orthonormalsystem von n zugehörigen Eigenvektoren.* ◇

4.2.8 Satz. (Hauptachsentransformation) *Ist A eine n-reihige symmetrische Matrix mit den Eigenwerten λ_1, ..., λ_n (nicht notwendig verschieden), dann gibt es eine orthogonale Matrix $T \in M(n,n)$ (d.h., $T^\mathsf{T}T = I$) derart, dass*

$$\Lambda = T^\mathsf{T} A T = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}.$$

Dabei bilden die Spalten von T ein ONS von Eigenvektoren zu λ_1, ..., λ_n. ◇

4.2.9 Anwendung: Klassifizierung quadratischer Formen. Die Hauptachsentransformation ist ein wichtiges Hilfsmittel zur geometrischen Interpretation quadratischer Gleichungen

$$\underline{x}^\mathsf{T} A \underline{x} + \underline{c}^\mathsf{T} \underline{x} + \gamma = 0, \tag{4.10}$$

wobei die n-reihige Matrix A sowie $\underline{c} \in \mathbb{R}^n$ und $\gamma \in \mathbb{R}$ gegeben seien. Ist A die Nullmatrix, reduziert sich (4.10) auf eine lineare Gleichung. Ist A nicht die Nullmatrix, kann man ohne Beschränkung der Allgemeinheit annehmen, dass A symmetrisch ist, denn man kann die quadratische Form in (4.10) offenbar umschreiben in

$$\underline{x}^\mathsf{T} A \underline{x} = \underline{x}^\mathsf{T} \left(\tfrac{1}{2}A + \tfrac{1}{2}A^\mathsf{T}\right) \underline{x}.$$

Dann ist aber die neue Matrix $B = \frac{1}{2}A + \frac{1}{2}A^\mathsf{T}$ symmetrisch.

Sei also nun

$$A \text{ symmetrisch.}$$

Verwenden wir die Matrix T der Hauptachsentransformation (d.h. $\Lambda = T^\mathsf{T} A T$) als Matrix einer Koordinatentransformation

$$\underline{y} = T^{-1}\underline{x} = T^\mathsf{T}\underline{x}, \quad \text{bzw. } \underline{x} = T\underline{y}$$

und setzen das in (4.10) ein, so folgt

$$\underline{y}^\mathsf{T} \Lambda \underline{y} + \underline{c}^\mathsf{T} T \underline{y} + \gamma = 0.$$

Ausgeschrieben ergibt das mit $(\alpha_1, \dots, \alpha_m) = \underline{c}^\mathsf{T} T$ die Gleichung

$$\sum_{j=1}^{n} \lambda_j y_j^2 + \sum_{j=1}^{n} \alpha_j y_j + \gamma = 0. \tag{4.11}$$

Man nennt das auch die *Normalform* der quadratischen Gleichung (4.10).

Für $n = 2$ und $\lambda_1 > 0$ und $\lambda_2 > 0$ ergibt sich als Lösungsmenge (soweit sie nichtleer ist) eine Ellipse, für $\lambda_1 \lambda_2 < 0$ ein Paar von Geraden oder ein Paar von Hyperbeln. Die vollständige Charakterisierung im Falle $n = 2$ finden Sie in *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984*, S. 147-150. ◇

4.2.10 Übung. Man berechne die Normalform der quadratischen Gleichung

$$5x_1^2 + 8x_1x_2 + 5x_2^2 = 0$$

aus Beispiel 4.2.5.

4.3 Anhang: Repetition zu Determinanten

4.3.1 Aus der *Mathematik II* wissen wir, dass die Determinante einer (2×2)-Matrix der vorzeichenbehaftete Flächeninhalt desjenigen Parallelogramms ist, das durch die Zeilenvektoren der Matrix aufgespannt wird.

Wir rechnen das zur Erinnerung noch einmal vor. Sei

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad \text{mit den Zeilenvektoren} \quad A_{1\bullet} = (a_{11}, a_{12}),\ A_{2\bullet} = (a_{21}, a_{22}).$$

Ferner sei α der Winkel zwischen $A_{1\bullet}$ und $A_{2\bullet}$. Dann gilt nach dem Sinussatz für den Flächeninhalt $F(A)$ des durch $A_{1\bullet}$ und $A_{2\bullet}$ aufgespannten Parallelogramms

$$[F(A)]^2 = \|A_{1\bullet}\|^2 \|A_{2\bullet}\|^2 \sin^2 \alpha,$$

wobei

$$\|\cdot\| \ \text{ die euklidische Norm in } \mathbb{R}^2 \text{ bezeichnet.}$$

Daraus folgt – mit dem euklidischen Skalarprodukt (für Zeilenvektoren) $A_{1\bullet} A_{2\bullet}{}^{\mathsf{T}}$ –, dass

$$\begin{aligned} [F(A)]^2 = \|A_{1\bullet}\|^2 \|A_{2\bullet}\|^2 (1 - \cos^2 \alpha) &= \|A_{1\bullet}\|^2 \|A_{2\bullet}\|^2 - (\, A_{1\bullet} A_{2\bullet}{}^{\mathsf{T}}\,)^2 \\ &= (a_{11}^2 + a_{12}^2)(a_{21}^2 + a_{22}^2) - (a_{11}a_{21} + a_{12}a_{22})^2 \\ &= (a_{11}a_{22} - a_{12}a_{21})^2, \end{aligned}$$

die letzte Gleichung folgt dabei leicht durch Ausmultiplizieren. Die Zahl

$$\det A = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} := a_{11}a_{22} - a_{12}a_{21} \tag{4.12}$$

heisst dann *Determinante von A*. Offenbar gelten die folgenden Eigenschaften:

(i)
$$\begin{vmatrix} a_{11} + a'_{11} & a_{12} + a'_{12} \\ a_{21} & a_{22} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} a'_{11} & a'_{12} \\ a_{21} & a_{22} \end{vmatrix}, \quad \begin{vmatrix} \lambda a_{11} & \lambda a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \lambda \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix},$$

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} + a'_{21} & a_{22} + a'_{22} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} \\ a'_{21} & a'_{22} \end{vmatrix}, \quad \begin{vmatrix} a_{11} & a_{12} \\ \lambda a_{21} & \lambda a_{22} \end{vmatrix} = \lambda \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}.$$

(ii) $\begin{vmatrix} a_{11} & a_{12} \\ a_{11} & a_{12} \end{vmatrix} = 0$ (2 gleiche Zeilen).

(iii) $\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1$ (Determinante der Einheitsmatrix).

◇

Die Eigenschaften (i)-(iii) werden - anders als in der Mathematik II - direkt zur Definition der Determinante einer $(n \times n)$-Matrix benutzt.

Wie bisher bezeichnen wir mit $A_{i\bullet}$ die i-te Zeile einer Matrix $A \in M(n,n)$ und mit I die Einheitsmatrix in $M(n,n)$.

4.3.2 Definition. Sei $M(n,n)$ die Menge der n-reihigen quadratischen Matrizen. Eine Abbildung $\det : M(n,n) \to \mathbb{R}$, d.h.,

$$A \in M(n,n) \mapsto \det A \in \mathbb{R},$$

heisst ***Determinante***, wenn die folgenden Eigenschaften D1, D2 und D3 erfüllt sind.

D1. Die Abbildung $A \mapsto \det A$ ist ***linear in jeder Zeile***, d.h., für jeden Zeilenindex $i \in \{1, \ldots, n\}$ gilt: Ist $A_{i\bullet} = A'_{i\bullet} + A''_{i\bullet}$ bzw. $A_{i\bullet} = \lambda A'_{i\bullet}$, $\lambda \in \mathbb{R}$, unter Beibehaltung der Zeilen $A_{k\bullet}$ für $k \neq i$, so ist

$$\text{a)} \qquad \det \begin{pmatrix} A_{1\bullet} \\ \vdots \\ A'_{i\bullet} + A''_{i\bullet} \\ \vdots \\ A_{n\bullet} \end{pmatrix} = \det \begin{pmatrix} A_{1\bullet} \\ \vdots \\ A'_{i\bullet} \\ \vdots \\ A_{n\bullet} \end{pmatrix} + \det \begin{pmatrix} A_{1\bullet} \\ \vdots \\ A''_{i\bullet} \\ \vdots \\ A_{n\bullet} \end{pmatrix},$$

$$\text{b)} \qquad \det \begin{pmatrix} A_{1\bullet} \\ \vdots \\ \lambda A'_{i\bullet} \\ \vdots \\ A_{n\bullet} \end{pmatrix} = \lambda \det \begin{pmatrix} A_{1\bullet} \\ \vdots \\ A'_{i\bullet} \\ \vdots \\ A_{n\bullet} \end{pmatrix}.$$

D2. Die Abbildung $A \mapsto \det A$ ist ***alternierend***, d.h., hat A zwei gleiche Zeilen, so ist $\det A = 0$.

D3. Die Abbildung $A \mapsto \det A$ ist ***normiert***, d.h., $\det I = 1$.

Die Abbildung $A \mapsto \det A$ ist damit eine sogenannte *normierte alternierende Multilinearform.* Zu gegebenem $A \in M(n,n)$ heisst die Zahl $\det A$ ***Determinante von A***. $\diamond$

Eine andere Schreibweise für die Determinante einer Matrix $A = (a_{ij};\, i,j = 1, \ldots, n)$ ist

$$\begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} = \det \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}.$$

Die senkrechten Striche deuten nicht auf einen Absolutbetrag hin, $\det A$ kann negativ sein, z.B. $\det A = -1$ für A mit $a_{11} = -1$, $a_{22} = \ldots = a_{nn} = 1$ und $a_{ij} = 0$ für alle i, j mit $i \neq j$ gemäss D3 und D1.

4.3.3 Für (2×2)-Matrizen haben wir oben die Existenz einer solchen Abbildung bewiesen, für $(n \times n)$-Matrizen verzichten wir auf den Beweis.

Wir leiten nun interessante weitere Eigenschaften aus den Axiomen D1-D3 ab, wobei der Rückgriff auf bekannte elementare Zeilenoperationen von Nutzen sein wird.

4.3.4 Zur Erinnerung: Bei einer quadratischen Matrix A der Ordnung n kann man durch endlich viele *elementare Zeilenoperationen* der Form

Typ I. Vertauschen zweier Zeilen (unter Beibehaltung der übrigen Zeilen),

Typ II. Addition des λ-fachen der k-ten Zeile zur i-ten Zeile
(unter Beibehaltung der k-ten wie der übrigen Zeilen),

entweder feststellen, dass $\operatorname{rg} A < n$ ist oder eine regluläre obere Dreiecksmatrix

$$\widetilde{A} \;=\; \begin{pmatrix} \tilde{a}_{11} & \cdots & \tilde{a}_{1n} \\ & \ddots & \vdots \\ & & \tilde{a}_{nn} \end{pmatrix} \tag{4.13}$$

erzeugen (d.h., $\tilde{a}_{kk} \neq 0$ für $1 \leq k \leq n$ und somit $\operatorname{rg} A = \operatorname{rg} \widetilde{A} = n$), der Leser möge sich die Begründung dafür noch einmal genau überlegen. Falls $\widetilde{A}$ regulär ist, kann man dann offenbar in einem zweiten Iterationszyklus nur mit Iterationen vom Typ II die Diagonalmatrix

$$\widetilde{D} \;=\; \begin{pmatrix} \tilde{a}_{11} & & \\ & \ddots & \\ & & \tilde{a}_{nn} \end{pmatrix} \tag{4.14}$$

erzeugen, wobei sich $\tilde{a}_{kk} \neq 0$ für $1 \leq k \leq n$ nicht mehr ändern.

Man benötigt also nicht die elementare Zeilenoperation vom Typ "Multipliziere die i-te Zeile mit $\lambda \neq 0$" und auch keine Spaltenvertauschungen! $\diamond$

4.3.5 Satz (Weitere Eigenschaften von Determinanten).

Eine Determinante $\det : M(n,n) \to \mathbb{R}$ hat die folgenden weiteren Eigenschaften:

D4. Für jedes $\lambda \in \mathbb{R}$ ist $\det(\lambda A) = \lambda^n \det A$.

D5. Enthält A eine Nullzeile, so gilt $\det A = 0$.

D6. Entsteht A' aus A durch genau eine Zeilenvertauschung (d.h., eine elementare Zeilenoperation vom Typ I), dann ist

$$\det A' \;=\; -\det A.$$

D7. Entsteht A' aus A durch die Addition eines Vielfachen der k-ten Zeile zur i-ten Zeile (d.h., eine elementare Zeilenoperation vom Typ II), dann ist

$$\det A' \ = \ \det A.$$

D8. Es gilt $\det A = 0 \;\Leftrightarrow\; \operatorname{rg} A < n$.

D9. Ist A eine obere bzw. untere Dreiecksmatrix

$$A \ = \ \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ & \ddots & \vdots \\ & & a_{nn} \end{pmatrix} \quad \text{bzw.} \quad A \ = \ \begin{pmatrix} a_{11} & & \\ \vdots & \ddots & \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

so ist $\det A = a_{11} \cdot a_{22} \cdot \ldots \cdot a_{nn}$.

D10. Für alle $A, B \in M(n,n)$ gilt

$$\det(AB) \ = \ \det A \cdot \det B.$$

D11. Ist $A \in M(n,n)$ regulär, so gilt

$$\det A^{-1} \ = \ \frac{1}{\det A}.$$

D12. Für alle $A \in M(n,n)$ gilt

$$\det A^{\mathsf{T}} \ = \ \det A.$$

D13. Ist $A \in M(n,n)$ gegeben durch

$$A = \begin{pmatrix} B & R \\ O & C \end{pmatrix},$$

wobei B und C quadratische Matrizen der Ordnungen $m < n$ bzw. $n - m$ sowie O die Nullmatrix in $M(n-m, m)$ und R irgendeine Matrix in $M(m, n-m)$ ist, so gilt

$$\det A = \det B \cdot \det C.$$

◇

4.3.6 Übung. D14. Man zeige, dass für $n \geq 2$ und $A, B \in M(n,n)$ die Beziehung $\det(A + B) = \det A + \det B$ im allgemeinen nicht gilt. ◇

4.3.7 Spaltenoperationen bei Determinanten. Aussage D12 des vorigen Satzes hat eine bedeutsame Konsequenz:

Die Eigenschaften D1-D14 sind auch gültig, wenn die Determinante als normierte alternierende Multilinearform bezüglich der Spalten von n-reihigen Matrizen definiert ist. Mit anderen Worten: Man kann statt der Zeilenoperationen auch die entsprechenden Spaltenoperationen betrachten, und es bleiben die bewiesenen Eigenschaften erhalten.

Wir sprechen bei den Elementarumformungen in diesem Falle sinngemäss von *elementaren Spaltenoperationen vom Typ I bzw. Typ II.* ◇

4.3.8 Satz. *Die Determinante ist gegenüber Ähnlichkeitstransformationen invariant, d.h., sind $A, B \in M(n,n)$ ähnlich, so gilt* $\det A = \det B$. ◇

Zum Beweis sei nur bemerkt, dass nach Definition A und B ähnlich sind, falls es eine reguläre n-reihige Matrix T gibt, so dass $B = T^{-1}AT$ gilt. Aus D10 und D11 folgt aber

$$\begin{aligned} \det T^{-1}AT &= \det T^{-1} \det A \det T \\ &= \det T^{-1} \det T \det A \\ &= \det A, \end{aligned}$$

was die Aussage beweist.

4.3.9 Laplace-Entwicklung mittels Kofaktoren. Der einfacheren Darstellung wegen entwickeln wir die Determinante nach der j-ten Spalte. Wegen Eigenschaft D12 für Determinanten ist das gleichberechtigt zur Entwicklung nach der i-ten Zeile. Dabei ist j bzw. i beliebig.

Sei mit $A = (a_{ij})$ wie stets in diesem Kapitel eine n-reihige Matrix bezeichnet. Dann gilt für die j-te Spalte von A die folgende Darstellung mit Hilfe von Einheitsvektoren $\underline{e}^i$ (als Spalte aufgefasst):

$$A_{\bullet j} = \sum_{i=1}^{n} a_{ij}\underline{e}^i.$$

Eigenschaft D1 liefert dann

$$\begin{aligned} \det A &= a_{1j} \det(A_{\bullet 1}, \ldots, \underline{e}^1, \ldots, A_{\bullet n}) \\ &\quad + a_{2j} \det(A_{\bullet 1}, \ldots, \underline{e}^2, \ldots, A_{\bullet n}) \\ &\quad + \ldots + a_{nj} \det(A_{\bullet 1}, \ldots, \underline{e}^n, \ldots, A_{\bullet n}) \\ &= \sum_{i=1}^{n} a_{ij} \det \widetilde{A}_{ij}, \end{aligned}$$

d.h., im i-ten Summanden schreiben wir

$$\widetilde{A}_{ij} = (A_{\bullet 1}, \ldots, \underline{e}^i, \ldots, A_{\bullet n}).$$

Wir betrachten nun $\det \widetilde{A}_{ij}$ für ein festes Paar (i,j). Die erste der folgenden Umformungen ergibt sich nach $n-1$ elementaren Spaltenoperationen vom Typ II, die zweite nach $j-1$ Spaltenvertauschungen und $i-1$ Zeilenvertauschungen, bei der dritten Umformung wird

D13 angewendet:

$$
\begin{aligned}
\det \widetilde{A}_{ij} \quad &:= \begin{vmatrix} a_{11} & \dots & a_{1j-1} & 0 & a_{1j+1} & \dots & a_{1n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{i-11} & \dots & a_{i-1j-1} & 0 & a_{i-1j+1} & \dots & a_{i-1n} \\ a_{i1} & \dots & a_{i1j-1} & 1 & a_{ij+1} & \dots & a_{in} \\ a_{i+11} & \dots & a_{i+1j-1} & 0 & a_{i+1j+1} & \dots & a_{i+1n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & \dots & a_{nj-1} & 0 & a_{nj+1} & \dots & a_{nn} \end{vmatrix} \\
\text{(1. Umformung)} \quad &= \begin{vmatrix} a_{11} & \dots & a_{1j-1} & 0 & a_{1j+1} & \dots & a_{1n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{i-11} & \dots & a_{i-1j-1} & 0 & a_{i-1j+1} & \dots & a_{i-1n} \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ a_{i+11} & \dots & a_{i+1j-1} & 0 & a_{i+1j+1} & \dots & a_{i+1n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & \dots & a_{nj-1} & 0 & a_{nj+1} & \dots & a_{nn} \end{vmatrix} \\
\text{(2. Umformung)} \quad &= (-1)^{i+j} \cdot \begin{vmatrix} 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & a_{11} & \dots & a_{1j-1} & a_{1j+1} & \dots & a_{1n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & a_{i-11} & \dots & a_{i-1j-1} & a_{i-1j+1} & \dots & a_{i-1n} \\ 0 & a_{i+11} & \dots & a_{i+1j-1} & a_{i+1j+1} & \dots & a_{i+1n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & a_{n1} & \dots & a_{nj-1} & a_{nj+1} & \dots & a_{nn} \end{vmatrix} \\
\text{(3. Umformung)} \quad &= (-1)^{i+j} \cdot 1 \cdot \det A_{ij},
\end{aligned}
$$

wobei A_{ij} die Matrix ist, die aus A nach Streichen der i-ten Zeile und j-ten Spalte entsteht. Wie oben erwähnt, wurde im letzten Schritt Eigenschaft D13 angewendet, und zwar auf die Blockmatrizen $B = (1)$, $C = A_{ij}$, O Nullspalte und R Nullzeile jeweils in $\mathbb{R}^{n-1}$. $\diamond$

4.3.10 Definition. Sei $A = (a_{ij}) \in M(n, n)$ gegeben und sei A_{ij} die Matrix die aus A durch Streichen der i-ten Zeile und j-ten Spalte von A entsteht, dann heisst die Zahl

$$(-1)^{i+j} \cdot \det A_{ij}$$

Kofaktor des Elements a_{ij}. $\diamond$

In 4.3.9 haben wir den folgenden Satz bewiesen, den wir schon aus der *Mathematik II* (ohne Beweis) kannten, dort als Definition einer Determinante und praktische Rechenvorschrift verwendet haben:

4.3.11 Satz. ***(Laplace-Entwicklung)*** *Für $A = (a_{ij}) \in M(n,n)$ gilt für beliebiges, aber festes $j \in \{1,\ldots,n\}$*

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} \cdot a_{ij} \cdot \det A_{ij} \qquad \textit{(Entwicklung nach Spalte } j\textit{)}$$

und für beliebiges, aber festes $i \in \{1,\ldots,n\}$

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} \cdot a_{ij} \cdot \det A_{ij} \qquad \textit{(Entwicklung nach Zeile } i\textit{).}$$

$\diamond$

Aus dem Satz über die Laplace-Entwicklung folgen nun zwei Ihnen bereits wohlbekannte Sätze, nämlich die *Cramersche Regel* zum Lösen eines inhomogenen LGS mit regulärer Koeffizientenmatrix A und die daraus folgende Berechnungsvorschrift für A^{-1}.

4.3.12 Satz. ***(Cramersche Regel)***
Ist $A = (a_{ij}) \in M(n,n)$ regulär und ein Spaltenvektor $\underline{b} \in \mathbb{R}^n$ gegeben, so ist $\underline{x}^0 \in \mathbb{R}^n$ mit

$$x_j^0 = \frac{1}{\det A} \cdot \det(A_{\bullet 1}, \ldots, A_{\bullet j-1}, \underline{b}, A_{\bullet j+1}, \ldots, A_{\bullet n}), \quad j = 1, \ldots, n,$$

die eindeutige Lösung des inhomogenen LGS $A\underline{x} = \underline{b}$. $\diamond$

4.3.13 Satz. ***(Berechnung der Inversen mit Hilfe von Kofaktoren)***
Ist $A = (a_{ij}) \in M(n,n)$ regulär, so ist $A^{-1} = (c_{ij})$ mit Hilfe der Kofaktoren von A gemäss

$$c_{ij} = \frac{1}{\det A} \cdot (-1)^{i+j} \cdot \det A_{ji}$$

darstellbar. Man beachte, dass c_{ij} durch den Kofaktor zum Element a_{ji} definiert ist. $\diamond$

Beweis der Sätze 4.3.12 und 4.3.13. Einsetzen von $\underline{b} = A\underline{x}^0$ in die j-te Spalte ergibt

$$\begin{aligned} \det(A_{\bullet 1}, \ldots, \underline{b}, \ldots, A_{\bullet n}) &= \det(A_{\bullet 1}, \ldots, \sum_{k=1}^{n} x_k^0 A_{\bullet k}, \ldots, A_{\bullet n}) \\ &= \sum_{k=1}^{n} x_k^0 \det(A_{\bullet 1}, \ldots, A_{\bullet k}, \ldots, A_{\bullet n}) \\ &= x_j^0 \det A, \end{aligned}$$

da

$$\det(A_{\bullet 1}, \ldots, A_{\bullet k}, \ldots, A_{\bullet n}) = \det A, \text{ falls } k = j \text{ und } = 0 \text{ sonst.}$$

Um die j-te Spalte $\underline{x}^0 := C_{\bullet j} = (c_{1j}, \ldots, c_{nj})^\mathsf{T}$ von $C = A^{-1}$ zu berechnen, ist die eindeutige Lösung $\underline{x}^0$ des inhomogenen LGS

$$A\underline{x} = \underline{e}^j, \quad \underline{e}^j \text{ } j\text{-te Einheitsspalte,}$$

zu bestimmen, also folgt nach Cramerscher Regel für die i-te Komponente $x_i^0 = c_{ij}$

$$\det A \cdot c_{ij} = \det(A_{\bullet 1}, \ldots, A_{\bullet i-1}, \underline{e}^j, A_{\bullet i+1}, \ldots, A_{\bullet n}) = \det \widetilde{A_{ji}} = (-1)^{i+j} \det A_{ji}$$

(d.h., $\underline{e}^j$ ersetzt i-te (!!) Spalte von A) - mit $\widetilde{A}_{ji}$ gemäss Herleitung in 4.3.9. $\square$

Kapitel 5

Spezielle Themen und Anwendungen

5.1 Definitheit von Matrizen

5.1.1 Repetition 1. Hat $f : \mathbb{R}^2 \to \mathbb{R}$ stetige zweite partielle Ableitungen nach x_1 und x_2 und ist $\underline{x}^0$ ein stationärer Punkt von f, d.h., der Gradient $\nabla f(\underline{x}^0) := (f_{x_1}(\underline{x}^0), f_{x_2}(\underline{x}^0))^\mathsf{T}$ erfüllt

$$\nabla f(\underline{x}^0) = \underline{o},$$

dann gelten mit der Hesse-Matrix $H(\underline{x}^0) = (f_{x_i x_j}(\underline{x}^0)\,;\; 1 \le i, j \le 2)$ (sie ist symmetrisch!) die folgenden Aussagen:

1. Falls $\det H(\underline{x}^0) > 0$ und $f_{x_1 x_1}(\underline{x}^0) > 0$, so hat f in $\underline{x}^0$ ein relatives Minimum.

2. Falls $\det H(\underline{x}^0) > 0$ und $f_{x_1 x_1}(\underline{x}^0) < 0$, so hat f in $\underline{x}^0$ ein relatives Maximum.

3. Falls $\det H(\underline{x}^0) < 0$, so besitzt f in $\underline{x}^0$ kein relatives Extremum (und wir sprechen bei $\underline{x}^0$ von einem Sattelpunkt).

Beim Beweis der Aussagen 1 und 2 in der Vorlesung *Mathematik I* hatten wir davon Gebrauch gemacht, dass folgende Äquivalenzen gelten

$$\begin{array}{lcl} (\;\det H(\underline{x}^0) > 0 \text{ und } f_{x_1 x_1}(\underline{x}^0) > 0\;) & \Leftrightarrow & \underline{v}^\mathsf{T} H(\underline{x}^0)\underline{v} > 0 \;\forall \underline{v} \neq \underline{o}, \\ (\;\det H(\underline{x}^0) > 0 \text{ und } f_{x_1 x_1}(\underline{x}^0) < 0\;) & \Leftrightarrow & \underline{v}^\mathsf{T} H(\underline{x}^0)\underline{v} < 0 \;\forall \underline{v} \neq \underline{o}. \end{array} \tag{5.1}$$

◇

5.1.2 Verallgemeinerung. Hat $f : \mathbb{R}^n \to \mathbb{R}$ stetige zweite partielle Ableitungen nach allen n Variablen, so bezeichnen wir für $\underline{x}^0 = (x_1^0, \ldots, x_n^0)^\mathsf{T}$ den Gradienten von f in $\underline{x}^0$ mit

$$\nabla f(\underline{x}^0) := (f_{x_1}(\underline{x}^0), \ldots, f_{x_n}(\underline{x}^0))^\mathsf{T}$$

und die Hesse-Matrix (sie ist wieder symmetrisch!) von f in $\underline{x}^0$ mit

$$H(\underline{x}^0) := (f_{x_i x_j}(\underline{x}^0)\,;\; 1 \le i, j \le n).$$

Die Aussagen 1, 2 aus Abschnitt 5.1.1 gelten dann in der äquivalenten Form (5.1), Aussage 3 aus Abschnitt 5.1.1 gilt in einer modifizierten Form.

Satz. Sei wieder $\underline{x}^0$ stationärer Punkt von f, d.h., $\nabla f(\underline{x}^0) = \underline{o}$.

1' Falls $\underline{v}^\mathsf{T} H(\underline{x}^0)\underline{v} > 0 \ \forall \underline{v} \neq \underline{o}$, so hat f in $\underline{x}^0$ ein relatives Minimum.

2' Falls $\underline{v}^\mathsf{T} H(\underline{x}^0)\underline{v} < 0 \ \forall \underline{v} \neq \underline{o}$, so hat f in $\underline{x}^0$ ein relatives Maximum.

3' Falls sowohl $\underline{u}^\mathsf{T} H(\underline{x}^0)\underline{u} > 0$ für mindestens ein $\underline{u} \neq \underline{o}$ als auch $\underline{v}^\mathsf{T} H(\underline{x}^0)\underline{v} < 0$ für mindestens ein $\underline{v} \neq \underline{o}$ gilt, so besitzt f in $\underline{x}^0$ kein relatives Extremum (d.h., $\underline{x}^0$ ist ein Sattelpunkt).

Dieser Satz wird in der Vorlesung *Mathematik III - Analysis für Ökonomen* bewiesen. ◇

5.1.3 Repetition 2. Aus der *Mathematik I* ist auch ein Kriterium für die Überprüfung der Konvexität/Konkavität einer zweimal stetig differenzierbaren Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mittels der Hesse-Matrizen $H(\underline{x}) = (f_{x_i x_j}(\underline{x})\,;\ 1 \leq i, j \leq 2)$ bekannt:

4. f ist konvex genau dann, wenn <u>für alle x</u> gilt
$$\underline{v}^\mathsf{T} H(\underline{x})\underline{v} \geq 0 \ \forall \underline{v}.$$

5. f ist konkav genau dann, wenn <u>für alle x</u> gilt
$$\underline{v}^\mathsf{T} H(\underline{x})\underline{v} \leq 0 \ \forall \underline{v}.$$

Die Verallgemeinerung auf Funktionen in n Variablen wird in Kapitel 5.2 behandelt. ◇

Man spricht bei den Kriterien in den Aussagen 1 und 2 (bzw. 1' und 2') von positiver bzw. negativer Definitheit, bei Aussage 3 (bzw. 3') von Indefinitheit der symmetrischen Matrix $H(\underline{x}^0)$ bzw. der betreffenden quadratischen Form $\underline{v} \mapsto \underline{v}^\mathsf{T} H(\underline{x}^0)\underline{v}$. In den Kriterien 4. und 5. heissen die Matrizen $H(\underline{x})$ positiv bzw. negativ *semi*-definit.

Diese Begriffe werden wir nun in der Sprache der linearen Algebra für beliebige symmetrische n-reihige Matrizen formulieren.

5.1.4 Definition. Eine n-reihige symmetrische Matrix A heisst

positiv definit, wenn $\underline{v}^\mathsf{T} A\underline{v} > 0$ für alle $\underline{v} \in \mathbb{R}^n \setminus \{\underline{o}\}$ gilt;
positiv semidefinit, wenn $\underline{v}^\mathsf{T} A\underline{v} \geq 0$ für alle $\underline{v} \in \mathbb{R}^n$ gilt;
negativ definit, wenn $\underline{v}^\mathsf{T} A\underline{v} < 0$ für alle $\underline{v} \in \mathbb{R}^n \setminus \{\underline{o}\}$ gilt;
negativ semidefinit, wenn $\underline{v}^\mathsf{T} A\underline{v} \leq 0$ für alle $\underline{v} \in \mathbb{R}^n$ gilt;
indefinit, wenn keine dieser vier Eigenschaften gilt. ◇

Offenbar ist A negativ definit genau dann, wenn $-A$ positiv definit ist. Analog ist A negativ semidefinit genau dann, wenn $-A$ positiv semidefinit ist. Der folgende Satz wird in der Vorlesung bewiesen, vgl. auch *P. Kall, Lineare Algebra für Ökonomen, Teubner, 1984*, Satz 3.26.

5.1.5 Satz. *Seien A eine n-reihige symmetrische Matrix und B eine n-reihige reguläre Matrix. Dann gilt:*
(i) A ist positiv definit genau dann, wenn $B^{\mathsf{T}}AB$ positiv definit ist.
(ii) A ist positiv semidefinit genau dann, wenn $B^{\mathsf{T}}AB$ positiv semidefinit ist.

◇

Wenn man als reguläre Matrix B speziell die Transformationsmatrix T der Hauptachsentransformation von A wählt, so übertragen sich die Definitheitseigenschaften eineindeutig auf die Diagonalmatrix Λ aus den n Eigenwerten λ_j der symmetrischen Matrix A: Mit

$$T\Lambda T^{\mathsf{T}} = A \quad \text{und} \quad \underline{y} = (y_1, \dots, y_n)^{\mathsf{T}} = T^{\mathsf{T}}\underline{x}$$

gilt

$$0 < \underline{x}^{\mathsf{T}} A \underline{x} = (\underline{x}^{\mathsf{T}} T)\Lambda(T^{\mathsf{T}}\underline{x}) = \underline{y}^{\mathsf{T}}\Lambda\underline{y} = \sum_{j=1}^{n} \lambda_j y_j^2,$$

analog mit $0 \leq$ ... oder $0 >$ ... oder $0 \geq$ .... Also gilt wegen der Regularität der Matrix T

$$\underline{x}^{\mathsf{T}} A \underline{x} > 0 \ \forall \underline{x} \neq \underline{o} \quad \Leftrightarrow \quad \sum_{j=1}^{n} \lambda_j y_j^2 > 0 \ \ \forall \underline{y} = (y_1, \dots, y_n) \neq \underline{o},$$

analog mit $0 \leq$ ... oder $0 >$ ... oder $0 \geq$ .... Setzt man in der rechten Seite der letzten Äquivalenz $\underline{y}$ sukzessive gleich den Einheitsvektoren im $\mathbb{R}^n$, folgt sofort

$$\underline{x}^{\mathsf{T}} A \underline{x} > 0 \ \forall \underline{x} \neq \underline{o} \quad \Leftrightarrow \quad \lambda_j > 0 \ \ \forall j \in \{1, \dots, n\},$$

analog mit $0 \leq$ ... oder $0 >$ ... oder $0 \geq$ .... Daraus folgen Charakterisierungen der verschiedenen Definitheitstypen symmetrischer Matrizen mit Hilfe der Eigenwerte:

5.1.6 Satz. *Eine n-reihige symmetrische Matrix mit den Eigenwerten $\lambda_1, \dots, \lambda_n$ ist*

positiv definit	*genau dann, wenn*	$\lambda_j > 0 \ \ \forall j \in \{1, \dots, n\}$;
positiv semidefinit	*genau dann, wenn*	$\lambda_j \geq 0 \ \ \forall j \in \{1, \dots, n\}$;
negativ definit	*genau dann, wenn*	$\lambda_j < 0 \ \ \forall j \in \{1, \dots, n\}$;
negativ semidefinit	*genau dann, wenn*	$\lambda_j \leq 0 \ \ \forall j \in \{1, \dots, n\}$.

◇

5.1.7 Bemerkungen.

1. Da bei der Hauptachsentransformation die Determinante einer symmetrischen Matrix A invariant ist, folgt nach Satz 5.1.6 speziell, dass positiv definite und negativ definite Matrizen regulär sind.

2. Ferner folgt, dass eine positiv semidefinite (bzw. negativ semidefinite) symmetrische Matrix genau dann positiv definit (negativ definit) ist, wenn sie regulär ist.

3. Eine symmetrische Matrix A ist genau dann indefinit, wenn sie zwei von Null verschiedene Eigenwerte mit unterschiedlichen Vorzeichen besitzt. ◇

5.1.8 Für symmetrische Matrizen kleinerer Dimension ist manchmal ein Kriterium für die positive bzw. negative Definitheit mit Hilfe der Vorzeichen sogenannter *Hauptabschnittsdeterminanten* nützlich. Entsteht $A^{[k]}$ aus einer n-reihigen symmetrischen Matrix $A = (a_{ij})$ durch Streichen sowohl der letzten k Zeilen als auch der letzten k Spalten von A, dann heisst

$$\det A^{[k]} \quad \text{eine Hauptabschnittsdeterminante von } A.$$

Alle Teilmatrizen $A^{[k]}$ sind wieder symmetrisch. Insbesondere gilt also $\det A = \det A^{[0]}$ und $a_{11} = \det A^{[n-1]}$. Zum Beispiel sind für eine (3×3)-Matrix $A = (a_{ij})$ die Zahlen

$$a_{11}, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \quad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

die Hauptabschnittsdeterminanten von A. So sind auch in der Repetition 5.1.1 die hinreichenden Optimalitätskriterien in der Sprache von Hauptabschnittsdeterminanten formuliert. ◇

5.1.9 Satz. *Sei A eine n-reihige symmetrische Matrix. Dann ist*

A positiv definit genau dann, wenn
alle Hauptabschnittsdeterminanten von A positiv sind;

A negativ definit genau dann, wenn
die Hauptabschnittsdeterminanten $\det A^{[k]}$ von A positiv im Falle gerader Ordnung von $A^{[k]}$, aber negativ im Falle ungerader Ordnung von $A^{[k]}$ sind. ◇

Zum Beispiel ist eine symmetrische (3×3)-Matrix $A = (a_{ij})$ negativ definit, falls

$$a_{11} < 0, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \quad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} < 0$$

gilt. Eine symmetrische (2×2)-Matrix $A = (a_{ij})$ ist negativ definit, falls

$$a_{11} < 0, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0$$

gilt, vgl. Bedingung 2 in Repetition 5.1.1.

Beweis von Satz 5.1.9 (optional). Wenn A positiv definit ist, gilt

$$\underline{x}^\mathsf{T} A \underline{x} > 0 \quad \text{für alle } \underline{x} = (x_1, \ldots, x_n)^\mathsf{T} \neq \underline{o}. \tag{5.2}$$

Wählen wir zur Matrix $A^{[k]}$ solche Vektoren $\underline{x} = (x_1, \ldots, x_n)^\mathsf{T}$ aus, dass

$$\underline{z} = (x_1, \ldots, x_{n-k})^\mathsf{T} \neq \underline{o} \text{ und } x_i = 0 \text{ für } i = n-k+1, \ldots, n$$

erfüllt ist, dann muss wegen (5.2)

$$\underline{z}^\mathsf{T} A^{[k]} \underline{z} = \underline{x}^\mathsf{T} A \underline{x} > 0$$

für diese Vektoren gelten, also ist auch $A^{[k]}$ positiv definit. Damit gilt für jedes $k \in \{0, 1, \ldots, n-1\}$: Die Eigenwerte von $A^{[k]}$ (und der jeweils mittels Hauptachsentransformation zugeordneten ähnlichen (!) Diagonalmatrix) sind positiv, also $\det A^{[k]} > 0$, was zu zeigen war. Analog schliesst man für eine negativ definite Matrix, dass alle Matrizen $A^{[k]}$ negativ definit sind. Folglich sind die Eigenwerte von $A^{[k]}$ (und der jeweils mittels Hauptachsentransformation zugeordneten Diagonalmatrix) negativ, was auf die alternierenden Vorzeichen für $\det A^{[k]}$ – je nach Ordnung von $A^{[k]}$ – in der Behauptung führt.

Die Rückrichtung beweisen wir mit Hilfe vollständiger Induktion. Für eine (1×1)-Matrix sind die beiden Aussagen trivial. Als Induktionsvoraussetzung nehmen wir nun an, dass unter den Vorzeichenbedingungen für die Hauptabschnittsdeterminanten jeder $((n-1) \times (n-1))$-Matrix ihre positive (bzw. negative) Definitheit gilt.

Sei nun A eine $(n \times n)$-Matrix mit den vorausgesetzten Vorzeichenbedingungen für alle Hauptabschnittsdeterminanten von A. Dann ist nach Induktionsvoraussetzung die Matrix $A^{[1]}$ (letzte Zeile und Spalte von A gestrichen) positiv bzw. negativ definit, und es existiert eine orthogonale Matrix T der Ordnung $n-1$ und $n-1$ (sämtlich positive oder sämtlich negative) reelle Eigenwerte $\lambda_1, \ldots, \lambda_{n-1}$ – zusammengefasst in einer Diagonalmatrix Λ –, so dass

$$T^\mathsf{T} A^{[1]} T = \Lambda.$$

Wir definieren mit dem $(n-1)$-dimensionalen Nullvektor $\underline{o}$ die reguläre (sogar orthogonale) Matrix

$$S := \begin{pmatrix} T & \underline{o} \\ \underline{o}^\mathsf{T} & 1 \end{pmatrix}.$$

Bilden wir $S^\mathsf{T} A S$ so entstehen ein Vektor $\underline{b} \in \mathbb{R}^{n-1}$ und eine Zahl β, so dass

$$S^\mathsf{T} A S = \begin{pmatrix} \Lambda & \underline{b} \\ \underline{b}^\mathsf{T} & \beta \end{pmatrix}.$$

Nach Satz 5.1.5 ist A positiv (negativ) definit genau dann, wenn $S^\mathsf{T} A S$ positiv (negativ) definit ist. Da alle Hauptdiagonalelemente $\lambda_1, \ldots, \lambda_{n-1}$ von Λ verschieden von Null sind, existieren endlich viele elementare Zeilenoperationen vom Typ "Addition eines Vielfachen der i-ten Zeile zur letzten Zeile" (bekanntlich mit zugeordneten regulären Transformationsmatrizen, deren Produktmatrix wir mit $M^\top$ bezeichnen wollen), so dass

$$M^\top S^\top A S = \begin{pmatrix} \Lambda & \underline{b} \\ \underline{o}^\top & \delta \end{pmatrix}$$

mit einem gewissen δ. Wenden wir nun die analogen elementaren Spaltenoperationen auf diese Matrix an, so entsteht

$$M^\top S^\top A S M = \begin{pmatrix} \Lambda & \underline{o} \\ \underline{o}^\top & \delta \end{pmatrix}.$$

Die Matrix M ist nach Definition regulär. Wieder aus Satz 5.1.5 folgt dann, dass A positiv (negativ) definit genau dann ist, wenn

$$B := (M^\top S^\top) A (SM)$$

positiv (negativ) definit ist. Die genannten elementaren Zeilen- bzw. Spaltenoperationen lassen die Determinante von $S^\top A S$ unverändert. Da S und $S^\top$ orthogonale Matrizen sind, folgt nach den Determinantengesetzen, dass

$$\det A = \det B = \lambda_1 \cdot \ldots \cdot \lambda_{n-1} \cdot \delta\,.$$

Sind alle Hauptabschnittsdeterminanten von A als positiv vorausgesetzt (Fall I), so folgt also (nach Induktionsvoraussetzung ist dann $A^{[1]}$ positiv definit)

$$\lambda_1 > 0, \ldots, \lambda_{n-1} > 0, \quad \text{somit} \quad \delta > 0.$$

Sind die Vorzeichen der Hauptabschnittsdeterminanten von A als alternierend vorausgesetzt (Fall II), so gilt (nach Induktionsvoraussetzung ist dann $A^{[1]}$ negativ definit)

$$\lambda_1 < 0, \ldots, \lambda_{n-1} < 0, \quad \text{somit} \quad \delta < 0.$$

B ist als Diagonalmatrix im Falle I positiv definit, damit auch A, und es ist B im Falle II negativ definit, damit auch A, was zu zeigen war. □

5.1.10 Übung. Untersuchen Sie die folgenden symmetrischen Matrizen auf positive bzw. negative Definitheit/Semidefinitheit oder Indefinitheit:

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} -2 & 1 \\ 1 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix},$$

Welchen Typ von Lösungsmenge erhalten Sie für die quadratische Gleichung $\underline{x}^\top A \underline{x} = 1$, wenn Sie für A jede der 5 Matrizen einsetzen? ◇

5.1.11 Übung. Untersuchen Sie die folgenden Matrizen auf positive bzw. negative Definitheit/Semidefinitheit oder Indefinitheit:

$$\begin{pmatrix} 4 & -2 & 1 \\ -2 & 2 & -1 \\ 1 & -1 & 2 \end{pmatrix}, \begin{pmatrix} 2 & -2 & 1 \\ -2 & 2 & -1 \\ 1 & -1 & 2 \end{pmatrix}.$$

◇

5.2 Konvexe Funktionen

5.2.1 Definition. Eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$ heisst ***konvex***, falls

$$\underline{x}, \underline{y} \in \mathbb{R}^n \text{ und } \lambda \in [0,1] \qquad \Longrightarrow \qquad f(\lambda \underline{x} + (1-\lambda)\underline{y}) \le \lambda f(\underline{x}) + (1-\lambda) f(\underline{y})$$

gilt. f heisst ***strikt konvex***, wenn das strenge Ungleichheitszeichen gilt. $g : \mathbb{R}^n \to \mathbb{R}$ heisst (strikt) ***konkav***, falls $f := -g$ (strikt) konvex ist. $\diamond$

Für Funktionen in einer Variablen wurden die Begriffe *konvex* und *konkav* in der *Mathematik I* eingeführt, und wir gaben Kriterien für die Überprüfung einer differenzierbaren bzw. zweimal differenzierbaren Funktion auf Konvexität/Konkavität bezüglich eines Intervalls. Zur Erinnerung: Sind $I \subset \mathbb{R}$ ein Intervall und $\varphi : I \to \mathbb{R}$ eine gegebene Funktion, so ist φ konvex auf I genau dann, wenn

$$x, y \in I \text{ und } \lambda \in [0,1] \qquad \Longrightarrow \qquad \varphi(\lambda x + (1-\lambda)y) \le \lambda \varphi(x) + (1-\lambda)\varphi(y)$$

erfüllt ist – analog zur obigen Definition.

5.2.2 Lemma. *Die Funktion $f : \mathbb{R}^n \to \mathbb{R}$ ist genau dann konvex, wenn zu je zwei Punkten $\underline{x}, \underline{y} \in \mathbb{R}^n$ die jeweilige Funktion in einer Variablen*

$$t \to F^{x,y}(t) := f(t\underline{x} + (1-t)\underline{y}),$$

auf dem Intervall $I = [0,1]$ konvex ist. $\diamond$

Beweis. Seien $\underline{x}, \underline{y} \in \mathbb{R}^n$ beliebig. Ist $F := F^{x,y}$ konvex auf $[0,1]$, so folgt für beliebige $\lambda \in [0,1]$

$$\begin{aligned} f(\lambda \underline{x} + (1-\lambda)\underline{y}) = F(\lambda) &= F(\lambda \cdot 1 + (1-\lambda) \cdot 0) \\ &\le \lambda F(1) + (1-\lambda) F(0) \\ &= \lambda f(\underline{x}) + (1-\lambda) f(\underline{y}), \end{aligned}$$

d.h., f ist konvex, da $\underline{x}, \underline{y} \in \mathbb{R}^n$ und $\lambda \in [0,1]$ beliebig waren.

Ist umgekehrt f konvex, so gilt für $F := F^{x,y}$ sowie $\lambda, t, s \in [0,1]$, dass

$$\begin{aligned} F(\lambda t + (1-\lambda)s) &= f(\ (\lambda t + (1-\lambda)s)\,\underline{x} + \underbrace{(\ 1 \ - \lambda t - (1-\lambda)s\ )\ \underline{y}}_{(\lambda + 1 - \lambda - \lambda t - (1-\lambda)s)\ \underline{y}}\ ) \\ &= f(\ (\lambda t + (1-\lambda)s)\underline{x} + \lambda(1-t)\underline{y} + (1-\lambda)(1-s)\underline{y}\ ) \\ &= f(\ \lambda[t\underline{x} + (1-t)\underline{y}] + (1-\lambda)[s\underline{x} + (1-s)\underline{y}]\ ) \\ &\le \lambda f(\ t\underline{x} + (1-t)\underline{y}) + (1-\lambda) f(s\underline{x} + (1-s)\underline{y}) \\ &= \lambda F(t) + (1-\lambda) F(s), \end{aligned}$$

was zu beweisen war. $\square$

5.2.3 Satz. (Repetition Mathematik I) *Sei $I \subset \mathbb{R}$ ein Intervall und $F : I \to \mathbb{R}$ zweimal differenzierbar. Dann gilt:*

F ist genau dann konvex, wenn $F''(x) \geq 0$ für alle $x \in I$ gilt. ◇

5.2.4 Satz. (Konvexitäts-/Konkavitätskriterium) *Sei f eine auf dem Raum $\mathbb{R}^n$ definierte reellwertige Funktion, die stetige zweite partielle Ableitungen $f_{x_i x_j}$ nach allen n Variablen x_1, ..., x_n besitzt. Dann gilt:*

f ist konvex (auf $\mathbb{R}^n$) genau dann, wenn für jeden Punkt $\underline{x} \in \mathbb{R}^n$ die Hesse-Matrix $H(\underline{x}) := (f_{x_i x_j}(\underline{x})\,;\ 1 \leq i, j \leq n)$ von f in $\underline{x}$ positiv semidefinit ist. ◇

Folglich ist f genau dann konkav, wenn für jeden Punkt $\underline{x} \in \mathbb{R}^n$ die Hesse-Matrix $H(\underline{x})$ negativ semidefinit ist.

Beweis. Es genügt, das Konvexitätskriterium zu zeigen. Nach Lemma 5.2.2 ist f genau dann konvex (auf $\mathbb{R}^n$), wenn f auf jeder Verbindungsstrecke zwischen zwei Punkten des $\mathbb{R}^n$ konvex ist. Dazu betrachten wir beliebige $\underline{x}, \underline{h} \in \mathbb{R}^n$ sowie die (von $\underline{x}$ und $\underline{h}$ abhängige) Funktion

$$F(t) := f(\underline{x} + t\underline{h}),\ \ t \in [0, 1].$$

Nach der Kettenregel gilt dann

$$F'(t) \ = \ \sum_{i=1}^{n} f_{x_i}(\underline{x} + t\underline{h})\, h_i \ = \ \underline{h}^{\mathsf{T}} \nabla f(\underline{x} + t\underline{h}).$$

Nochmalige Differentiation und Anwendung der Kettenregel ergibt

$$F''(t) \ = \ \sum_{i=1}^{n} \sum_{j=1}^{n} f_{x_i x_j}(\underline{x} + t\underline{h})\, h_i h_j \ = \ \underline{h}^{\mathsf{T}} H(\underline{x} + t\underline{h})\underline{h}.$$

Wenn $H(\underline{z})^{\mathsf{T}}$ für jedes $\underline{z}$ positiv semidefinit ist, gilt speziell $F''(t) \geq 0$ für alle $t \in [0, 1]$ und somit nach Satz 5.2.3 die Konvexität von F auf $[0, 1]$. Da $\underline{x}$ und $\underline{h}$ beliebige Vektoren des $\mathbb{R}^n$ waren, folgt dann nach Lemma 5.2.2 die Konvexität von f.

Ist umgekehrt f konvex, so ist für beliebige Vektoren $\underline{x}$ und $\underline{h}$ des $\mathbb{R}^n$ die zugeordnete Funktion $t \mapsto F(t) = f(\underline{x} + t\underline{h})$ konvex, also gilt speziell $F''(0) = \underline{h}^{\mathsf{T}} H(\underline{x})\underline{h} \geq 0$. Das bedeutet gerade, dass $H(\underline{x})$ für jedes $\underline{x}$ positiv semidefinit ist. □

5.2.5 Wenn $H(\underline{x})$ für jedes $\underline{x}$ sogar positiv definit ist, kann man zeigen, dass f strikt konvex ist. Es ist aus der *Mathematik I* bekannt, dass in diesem Falle jede stationäre Lösung der Aufgabe $f(\underline{x}) \to \min$ ein eindeutiger *globaler Minimalpunkt* von f ist. ◇

Für 2× differenzierbare Funktionen kann also die Überprüfung der Konvexität/Konkavität auf ein Problem der linearen Algebra zurückgeführt werden: den Nachweis von Definitheitseigenschaften.

5.3 Geometrische Reihe und Verbrauchsmatrizen

5.3.1 Bereits im Kapitel "Eigenwerte" hatten wir den Begriff einer Matrixnorm eingeführt. Sei zunächst $\|\cdot\|$ irgendeine Norm im $\mathbb{R}^n$. Gegeben sei eine quadratische n-reihige Matrix A. Die Zahl

$$\|A\| = \max_{\underline{x} \neq \underline{o}} \frac{\|A\underline{x}\|}{\|\underline{x}\|}$$

heisst die *(der Norm $\|\cdot\|$ zugeordnete)* ***Matrixnorm von A***. Dass diese Zahl wohldefiniert ist, hatten wir bereits im Punkt 4.1.7 begründet.

Aus der Definition folgt sofort

$$\|A\underline{x}\| \le \|A\| \, \|\underline{x}\| \quad (\forall \underline{x} \in \mathbb{R}^n). \tag{5.3}$$

Daraus ergibt sich dann für $A, B \in M(n,n)$ auch

$$\|AB\| \le \|A\| \, \|B\|,$$

denn es gilt $\|AB\underline{x}\| \le \|A\| \, \|B\underline{x}\| \le \|A\| \, \|B\| \, \|\underline{x}\|$ für beliebige $\underline{x} \in \mathbb{R}^n$.

$\diamond$

5.3.2 Lemma. *Sei $\|\cdot\|$ irgendeine Norm im $\mathbb{R}^n$. Die zugeordnete Matrixnorm ist eine Norm (im Sinne der üblichen Normdefinition) im Vektorraum $M(n,n)$ der n-reihigen Matrizen.* $\diamond$

Der Beweis bleibt dem Leser überlassen.

5.3.3 Bemerkung. Die der euklidischen Norm zugeordnete Norm einer gegebenen Matrix $A \in M(n,n)$ ist die Wurzel aus dem maximalen Eigenwert $\lambda_{\max}(A^\mathsf{T} A)$ von $A^\mathsf{T} A$, d.h.,

$$(\,\|A\|_2\,)^2 = \max_{\underline{x} \neq \underline{o}} \frac{\underline{x}^\mathsf{T} A^\mathsf{T} A \underline{x}}{\underline{x}^\mathsf{T} \underline{x}} = \lambda_{\max}(A^\mathsf{T} A). \tag{5.4}$$

Zum Beweis vgl. z.B. *Strang, Lineare Algebra, Springer 2003, §9.2, Aussage 9A.*

Ist $A \in M(n,n)$ eine symmetrische Matrix mit den Eigenwerten λ_1, ..., λ_n, so gilt

$$\|A\|_2 \; = \max_{1 \le i \le n} |\lambda_i|.$$

Diese Aussage folgt sofort aus (5.4), wenn man bedenkt, dass $A^\mathsf{T} A = AA = A^2$ gerade die Eigenwerte λ_i^2 $(i = 1, \ldots, n)$ hat. $\diamond$

5.3.4 Lemma. *Für die Matrix $Q \in M(n,n)$ gelte in irgendeiner Matrixnorm $\|\cdot\|$*

$$\|Q\| < 1.$$

Dann ist $I - Q$ invertierbar, und die Inverse genügt der Abschätzung

$$\|(I - Q)^{-1}\| \le (1 - \|Q\|)^{-1}.$$

$\diamond$

Die Voraussetzung $\|Q\| < 1$ ist hinreichend für die Existenz der Inversen von $I - Q$, nicht notwendig, vgl. Beispiel 5.3.7.

Beweis von Lemma 5.3.4: Für jedes $\underline{x} \in \mathbb{R}^n$ gilt nach Dreiecksungleichung und der Abschätzung (5.3) – angewendet auf $A = Q$ –

$$\begin{aligned} \|(I-Q)\underline{x}\| &\geq \|\underline{x}\| - \|Q\underline{x}\| \\ &\geq \|\underline{x}\| - \|Q\| \, \|\underline{x}\| \\ &= (1 - \|Q\|) \, \|\underline{x}\|. \end{aligned} \tag{5.5}$$

Nach Voraussetzung ist $1 - \|Q\| > 0$, so dass aus (5.5)

$$\text{für } \underline{x} \neq \underline{o} \text{ sofort } (I-Q)\underline{x} \neq \underline{o} \text{ folgt.}$$

Also ist $I - Q$ regulär. Wenn $\underline{x} = (I-Q)^{-1}\underline{y}$ gesetzt wird, geht (5.5) über in

$$\|\underline{y}\| \quad \geq \quad (1 - \|Q\|) \, \|(I-Q)^{-1}\underline{y}\| \quad \forall \underline{y} \in \mathbb{R}^n.$$

Nach Definition der Matrixnorm folgt dann die Behauptung. □

5.3.5 Ist A eine n-reihige Matrix mit Norm kleiner als 1, so ist die Existenz der Inversen der Matrix $I - A$ nach Lemma 5.3.4 gesichert. Unten beweisen wir die folgende Formel

$$(I-A)^{-1} = I + A + A^2 + \ldots + A^k + \ldots \tag{5.6}$$

in Analogie zur Summenformel der geometrischen Reihe

$$1/(1-q) = 1 + q + q^2 + \ldots + q^k + \ldots \quad \text{für } -1 < q < 1.$$

Man spricht deshalb bei der Summe auf der rechten Seite von (5.6) von der ***geometrischen Reihe für Matrizen***. Diese Reihe heisst auch *Neumannsche Reihe* nach dem Mathematiker Carl Neumann (1832-1925). Sie war uns bereits im Beispiel 2.1.12 zur Verbrauchmatrix in Input-Output-Modellen begegnet, wir kommen auf diese Anwendung unten zurück.

5.3.6 Satz. *Sei $\| \cdot \|$ eine Matrixnorm in $M(n,n)$ und $A \in M(n,n)$ mit $\|A\| < 1$. Dann existiert $(I-A)^{-1}$, und es gilt $(I-A)^{-1} = I + A + A^2 + \ldots + A^k + \ldots$.* ◇

Beweis: Die Existenz von $(I-A)^{-1}$ folgt aus Lemma 5.3.4. Wir betrachten nun die Partialsumme S_m und das Produkt AS_m

$$\begin{aligned} S_m &= I + A + A^2 + \ldots + A^m \\ AS_m &= A + A^2 + \ldots + A^m + A^{m+1} \end{aligned}$$

und bilden die Differenz $S_m - AS_m$. Als Ergebnis erhalten wir

$$(I-A)S_m = I - A^{m+1},$$

also
$$S_m = (I-A)^{-1} - (I-A)^{-1}A^{m+1}.$$
Wegen $\|A\| < 1$ haben wir für $m \to \infty$ die Konvergenz $\|A\|^{m+1} \to 0$ und somit
$$\|(I-A)^{-1}A^{m+1}\| \leq \|(I-A)^{-1}\|\|A\|^{m+1} \to 0,$$
also $S_m \to (I-A)^{-1}$, was zu zeigen war. □

5.3.7 Beispiel. Die Voraussetzung $\|A\| < 1$ im vorigen Satz ist wichtig dafür, dass die Partialsummen S_m konvergieren. Betrachten Sie z.B.
$$A = \begin{pmatrix} 0 & 4 \\ 1 & 0 \end{pmatrix}, \quad \text{d.h.} \quad I - A = \begin{pmatrix} 1 & -4 \\ -1 & 1 \end{pmatrix}.$$
Dann existiert $(I-A)^{-1}$, denn es gilt
$$-\frac{1}{3}\begin{pmatrix} 1 & 4 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & -4 \\ -1 & 1 \end{pmatrix} = I, \text{ d.h. } \ (I-A)^{-1} = -\frac{1}{3}\begin{pmatrix} 1 & 4 \\ 1 & 1 \end{pmatrix}.$$
Offenbar hat A die Eigenwerte -2 und +2, lässt sich daher diagonalisieren mit einer regulären Transformationsmatrix T zu
$$\Lambda := \begin{pmatrix} -2 & 0 \\ 0 & 2 \end{pmatrix} = T^{-1}AT,$$
somit $A^m = T\Lambda^m T^{-1}$ und folglich $\|A^m\| \to +\infty$, also divergieren auch die Partialsummen $S_m = I + A + A^2 + \ldots + A^m$. ◇

5.3.8 Definition. Eine Matrix $A = (a_{ij}) \in M(n,n)$ heisst *nichtnegativ* (bzw. *positiv*), falls $a_{ij} \geq 0$ (bzw. $a_{ij} > 0$) für alle i, j gilt.

5.3.9 Lemma. *Sei $\|\cdot\|_\infty$ die der Maximumnorm im $\mathbb{R}^n$ zugeordnete Matrixnorm. Dann gilt für jede nichtnegative Matrix $A = (a_{ij}) \in M(n,n)$, dass*
$$\|A\|_\infty = \max_{1\leq i\leq n} \sum_{j=1}^{n} a_{ij},$$
d.h., $\|A\|_\infty$ ist die maximale Zeilensumme. (Bemerkung: Enthält A negative Elemente, muss man zu den absoluten Zeilensummen übergehen, vgl. Lemma 4.1.7.) ◇

Beweis: Sei $\|\underline{x}\|_\infty = 1$, also speziell $x_j \leq 1$ für jede Komponente. Dann gilt für jede Zeile $A_{i\bullet}$ von A:
$$A_{i\bullet}\underline{x} = \sum_{j=1}^{n} a_{ij}x_j \leq \sum_{j=1}^{n} a_{ij}.$$
Für $\underline{x} = (1,\ldots,1)^\mathsf{T}$ wird Gleichheit realisiert. Also gilt wegen $\|A\|_\infty = \max_{\|\underline{x}\|_\infty=1}\{\|A\underline{x}\|_\infty\}$ die Behauptung. □

Sofort aus Satz 5.3.6 und Lemma 5.3.9 erhalten wir

5.3.10 Folgerung. Ist $A = (a_{ij}) \in M(n,n)$ eine nichtnegative Matrix mit $\sum_{j=1}^{n} a_{ij} < 1$ für alle Zeilenindizes i, dann existiert $(I - A)^{-1}$ und ist eine nichtnegative Matrix. $\diamond$

5.3.11 Anwendung auf Verbrauchsmatrizen. In Beispiel 2.1.12 haben wir eine Volkswirtschaft mit n Produktionszweigen betrachtet, in der der Output jedes Zweiges als Produktionsfaktor (Input) im eigenen wie in den anderen Zweigen eingesetzt und verbraucht werden kann. Dort spielte folgende Verbrauchmatrix eine Rolle (mit En = Energieproduktion, Ch = Chemieproduktion und Ba = Bauwesen):

$$\begin{pmatrix} \text{En-Verbrauch} \\ \text{Ch-Verbrauch} \\ \text{Ba-Verbrauch} \end{pmatrix} = \underbrace{\begin{pmatrix} 0.2 & 0.4 & 0.3 \\ 0.3 & 0.4 & 0.2 \\ 0.4 & 0.3 & 0.1 \end{pmatrix}}_{=:\,A} \begin{pmatrix} \text{Input für En} \\ \text{Input für Ch} \\ \text{Input für Ba} \end{pmatrix}.$$

So sagt z.B. die erste Zeile der Verbrauchsmatrix, dass der Verbrauch an Energie zur Produktion von 1 Einheit in En 0.2 Einheiten, zur Produktion von 1 Einheit in Ch 0.4 Einheiten sowie zur Produktion von 1 Einheit in Ba 0.3 Einheiten beträgt.

Nennen wir den Produktionsvektor (Input) $\underline{p} = (p_1, p_2, p_3)^{\mathsf{T}}$ und den Nachfragevektor $\underline{y} = (y_1, y_2, y_3)^{\mathsf{T}}$, so sucht man bei gegebenem $\underline{y} \geq \underline{o}$ in dem Modell

$$\underline{p} - A\underline{p} = \underline{y}, \quad \text{d.h., } (I - A)\underline{p} = \underline{y}, \tag{5.7}$$

einen geeigneten Input $\underline{p} \geq \underline{o}$, so dass die Nachfrage $\underline{y}$ befriedigt werden kann, wobei berücksichtigt wird, dass ein Teil des Inputs $\underline{p}$, nämlich $A\underline{p}$, während der Produktion verbraucht wird.

Da die gegebene Matrix A die Voraussetzungen von Folgerung 5.3.10 erfüllt, hat das LGS (5.7) die eindeutige Lösung $\underline{p}^0 = (I - A)^{-1}\underline{y}$, und die Matrix

$$(I - A)^{-1} = I + A + A^2 + \ldots + A^k + \ldots$$

ist nichtnegativ. Also gilt bei $\underline{y} \geq \underline{o}$ auch $\underline{p}^0 \geq \underline{o}$. Liegt eine solche Lösung vor, nennt man die Volkswirtschaft ***produktiv***.

Mathematica liefert als Inverse von $I - A$ in unserem Beispiel die Matrix

$$(I - A)^{-1} = \frac{1}{29} \begin{pmatrix} 96 & 90 & 52 \\ 70 & 120 & 50 \\ 66 & 80 & 72 \end{pmatrix}.$$

Interpretiert man die Matrix in Beispiel 5.3.7 als Verbrauchsmatrix einer Volkswirtschaft mit zwei Zweigen, dann sieht man: Der Verbrauch ist zu gross. Eine Nachfrage kann nicht bedient werden, weil die Produktion mehr verbraucht als sie liefert. $\diamond$

5.4 Markov-Matrizen

5.4.1 Einführung. Im Beispiel 4.1.1 haben wir die Kundenwanderung auf einem Markt mit 3 Herstellern betrachtet, wobei die Matrix der Kundenwanderung durch

$$A = (a_{ij}) = \begin{pmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.4 & 0.3 \\ 0.1 & 0.4 & 0.5 \end{pmatrix}$$

gegeben war. Die Marktaufteilung zu Beginn der ersten Zeitperiode war

$$\underline{x}^0 = \begin{pmatrix} 0.25 \\ 0.40 \\ 0.35 \end{pmatrix}.$$

Wir hatten dann folgende Probleme gestellt:

1. Berechnen Sie sie Marktaufteilung $\underline{x}^1$ am Ende der Periode 1.

2. Auch in den folgenden $k \in \mathbb{N}$ Perioden sei die Kundenwanderung konstant. Berechnen Sie die Marktaufteilung $\underline{x}^3$ am Ende der Periode 3. Wie berechnet sich die Marktaufteilung $\underline{x}^k$ am Ende der Periode k?

3. Gibt es eine *stationäre* Marktaufteilung zu der gegebenen Wanderungsmatrix A, d.h. ein $\underline{x}^* \geq \underline{o}$ mit $x_1^*+x_2^*+x_3^* = 1$, so dass die Kundenwanderung innerhalb einer Periode die Marktaufteilung $\underline{x}^*$ nicht ändert?

Die Lösung der Probleme 1. und 2. ergibt sich, wie wir uns überlegt hatten, mittels der Formel

$$\underline{x}^k = A^{\mathsf{T}}\underline{x}^{k-1} = (A^T)^k\underline{x}^0, \; k = 1, 2, 3, \ldots. \tag{5.8}$$

Problem 3. führt auf die Frage der Existenz eines *Eigenvektors* $\underline{x}^*$ zum *Eigenwert* $\lambda = 1$ der Matrix A^{T}, wobei die Zusatzbedingungen $\underline{x}^* \geq \underline{o}$ und $x_1^* + x_2^* + x_3^* = 1$ ebenfalls erfüllt sind.

Die Matrix

$$B = A^{\mathsf{T}}$$

ist positiv und hat in allen Spalten die Spaltensumme 1. Solche Matrizen spielen eine besondere Rolle und werden (positive) *Markov-Matrizen* genannt, der Prozess (5.8) liefert eine sogenannte Markov-Kette $\{\underline{x}^k\}$. Wir werden für Matrizen dieses Typs zeigen, dass sie stets einen Eigenwert 1 haben und dass bei beliebigem Anfangsvektor $\underline{x}^0 \geq \underline{o}$ mit $\sum x_j^0 = 1$ die Folge $\underline{x}^k$ gegen einen eindeutig bestimmten Vektor $\underline{y}$ konvergiert, wobei $\underline{y}$ Eigenvektor zum Eigenwert 1 der Matrix ist. Angewandt auf das obige Beispiel liefert dann $\underline{y}$ die gesuchte stationäre Marktaufteilung. ◇

5.4.2 Definition. Eine quadratische n-reihige Matrix $B = (b_{ij})$ heisst ***Markov-Matrix***, falls B nichtnegativ ist und $\sum_i^n b_{ij} = 1$ für alle Spaltenindizes $j \in \{1 \ldots, n\}$ gilt. ◇

5.4.3 Satz. *Sei $B = (b_{ij}) \in M(n,n)$ eine positive Markov-Matrix. Dann gelten folgende Aussagen:*

(i) *B hat den Eigenwert 1, der dazugehörige Eigenraum* $\operatorname{Eig}(B;1)$ *hat die Dimension 1.*

(ii) *Ist $\underline{x}^0 \in \mathbb{R}^n$ irgendein Vektor mit $\underline{x}^0 \geq \underline{o}$ und $\sum_{j=1}^n x_j^0 = 1$, dann konvergiert die durch*

$$\underline{x}^k = B\underline{x}^{k-1}, \quad k = 1, 2, 3, \ldots$$

erzeugte Folge $\{\underline{x}^k\}$ gegen einen Punkt $\underline{y} \in$ $\operatorname{Eig}(B;1)$ *mit $\underline{y} \geq \underline{o}$ und $\sum_{j=1}^n y_j = 1$, wobei $\underline{y}$ von der Wahl des Anfangspunktes $\underline{x}^0$ unabhängig ist.*

(iii) $\lim_{k\to\infty} B^k = [\underline{y} \ldots \underline{y}]$. $\diamond$

Beweis: (optional) Der Beweis folgt dem Buch *R. Bellman, Introduction to Matrix Analysis, 2nd edition, SIAM, 1997.* Wir betrachten zunächst einen beliebigen Vektor

$$\underline{b} \in \mathbb{R}^n \ \text{ mit } 0 \leq b_i \leq 1 \ \forall i$$

sowie die Folge

$$\underline{z}^0 = \underline{b}, \quad \underline{z}^{k+1} = B^\top \underline{z}^k = (B^\top)^k \underline{b}, \ k = 0, 1, 2, \ldots. \tag{5.9}$$

Nehmen wir an, es ist schon gezeigt, dass ein $\mu \in \mathbb{R}$ existiert, so dass

$$\lim_{k\to\infty} \underline{z}^k = \underline{z} := (\mu, \ldots, \mu)^\top. \tag{5.10}$$

gilt (das werden wir am Ende des Beweises tun).

Beweis von (i) und (ii): Seien nun gemäss Voraussetzung des Satzes

$$\underline{x}^0 \in \mathbb{R}^n \text{ irgendein Vektor mit } \underline{x}^0 \geq \underline{o} \text{ und } \textstyle\sum_{j=1}^n x_j^0 = 1$$

sowie

$$\underline{x}^k = B\underline{x}^{k-1}, \quad k = 1, 2, 3, \ldots. \tag{5.11}$$

Betrachten wir jetzt für $j = 1, \ldots, n$ sukzessive die Folgen $\underline{z}^k = \underline{z(j)}^k$ gemäss (5.9) mit $\underline{b} = \underline{e}^j$ (Einheitsvektoren) und Grenzwerten $\underline{z(j)} = \lim_{k\to\infty} \underline{z(j)}^k$ gemäss (5.10), dann erhalten wir

$$x_j^k = (\underline{e}^j)^\top \underline{x}^k = (\underline{e}^j)^\top B^k \underline{x}^0 = (\underline{x}^0)^\top (B^\top)^k \underline{e}^j \to (\underline{x}^0)^\top \underline{z(j)} \quad \text{falls } k \to \infty. \tag{5.12}$$

Nach (5.10) haben die Vektoren $\underline{z(j)}$ stets gleiche Komponenten, sagen wir

$$\underline{z(j)} = (\mu(j), \ldots, \mu(j))^\top.$$

Dann folgt im letzten Term von (5.12)

$$(\underline{x}^0)^\top \underline{z(j)} = \sum_{j=1}^n x_j^0 \mu(j) = \mu(j)(\sum_{j=1}^n x_j^0) = \mu(j) \quad (\text{wegen } \textstyle\sum_{j=1}^n x_j^0 = 1 \text{ nach Voraussetzung}).$$

Damit liefert (5.12)

$$\lim_{k\to\infty} x_j^k = \mu(j) \;\; \forall j,$$

d.h., $\underline{y} = \lim_{k\to\infty} x^k$ ist unabhängig vom Anfangsvektor $\underline{x}^0$. Wir bemerken noch, dass die Spaltensummen von B gleich 1 sind, also folgt mit $\underline{e} = (1,\ldots,1)^\mathsf{T}$ auch $\underline{e}^\mathsf{T} B = \underline{e}^\mathsf{T}$ und somit

$$\underline{e}^\mathsf{T}\underline{x}^k = \underline{e}^\mathsf{T} B \underline{x}^{k-1} = \underline{e}^\mathsf{T}\underline{x}^{k-1} = 1,\; k = 1,2,\ldots \quad \text{wegen } \underline{e}^\mathsf{T}\underline{x}^0 = 1.$$

Damit gilt

$$\underline{y} = \lim_{k\to\infty} x^k \geq \underline{o} \;\; \text{mit} \;\; \sum_{j=1}^n y_j = 1.$$

Zum Nachweis von (i) und (ii) fehlt noch zu zeigen, dass $\underline{y}$ Eigenvektor zum Eigenwert 1 von B ist und den Eigenraum $\operatorname{Eig}(B;1)$ aufspannt. Zunächst bemerken wir dazu, dass

$$\underline{y} = \lim_{k\to\infty} B^k \underline{y} = B \lim_{k\to\infty} B^{k-1}\underline{y} = B\underline{y}$$

gilt, und zwar für $\underline{x}^k = B^k \underline{y}$, wenn wir $\underline{x}^0$ speziell gleich $\underline{y}$ wählen (was möglich ist, da der Grenzwert $\underline{y}$ unabhängig vom Ausgangsvektor $\underline{x}^0$ war, für den nur $\underline{x}^0 \geq \underline{o}$ und $\underline{e}^\mathsf{T}\underline{x}^0 = 1$ erfüllt sein musste). Also ist $\underline{y}$ Eigenvektor zum Eigenwert 1 von B. Ist $\underline{c}$ ein anderer Eigenvektor zum Eigenwert 1, dann ist

$$\underline{c} + t\underline{y} \quad \text{ein positiver Vektor für ein hinreichend grosses } t > 0.$$

Dann erfüllt bezüglich der Summennorm $\|\cdot\|_1$

$$\underline{c}' = (\|\underline{c} + t\underline{y}\|_1)^{-1}(\underline{c} + t\underline{y}) \quad \text{die Bedingung } \; \underline{c}' \geq \underline{o},\; \underline{e}^\mathsf{T}\underline{c}' = 1$$

und ist ebenfalls Eigenvektor zu 1 (Zur Erinnerung: $\operatorname{Eig}(B;1)$ ist ein Unterraum). Also hätten wir - wie oben gezeigt, ist $\underline{y}$ eindeutiger Grenzwert von Folgen vom Typ (5.11) -

$$\underline{c}' \equiv B^k \underline{c}' = \lim_{k\to\infty} B^k \underline{c}' = \underline{y},$$

also ist $\underline{c}$ ein Vielfaches von $\underline{y}$. Mit anderen Worten: $\underline{y}$ spannt den Eigenraum zum Eigenwert 1 auf.

Als letzter Beweisschritt erfolgt nun der Nachweis von Aussage (5.10). Seien dazu

$$u(k) = \max_{1\leq i\leq n} z_i^k \;\; \text{und} \;\; v(k) = \min_{1\leq i\leq n} z_i^k$$

(das sind also eine maximale und minimale Komponente des Vektors $\underline{z}^k$ aus (5.9)), und sei $\underline{b} \geq \underline{o}$ beliebig mit Komponenten kleiner oder gleich 1. Es genügt dann zu zeigen, dass

$$u(k) - v(k) \to 0 \;\; \text{für } k \to \infty.$$

Schreibt man die Definitionsgleichung (5.9) komponentenweise auf, ergibt sich

$$z_i^{k+1} = \sum_{j=1}^{n} b_{ji} z_j^k, \; i = 1, \ldots, n.$$

Folglich gilt

$$z_i^{k+1} \leq \sum_{j=1}^{n} b_{ji} u(k) \; \forall i, \text{ also speziell } u(k+1) \leq u(k)$$

wegen $\sum_{j=1}^{n} b_{ji} = 1 \; \forall i$. Analog folgt

$$v(k+1) \geq v(k).$$

Beachte $0 \leq u(k) \leq u(0)$ und $0 \leq v(k) \leq 1$ wegen der Voraussetzungen an B und $\underline{b}$. Damit sind $u(k)$ und $v(k)$ monotone beschränkte Folgen, haben also Grenzwerte u bzw. v. Zu zeigen ist $u = v$.

Da B eine positive Markov-Matrix ist, existiert für die Dimension $n \geq 2$ eine Zahl d mit

$$b_{ij} \geq d > 0 \; \forall i \, \forall j \quad \text{und} \quad d \leq \tfrac{1}{2}.$$

Ohne Beschränkung der Allgemeinheit sei für den Folgenindex k

$$u(k+1) = z_1^{k+1}, \; v(k) = z_n^k.$$

Dann gilt wegen $\sum_{j=1}^{n} b_{j1} = 1$

$$\begin{aligned} u(k+1) = z_1^{k+1} &= \sum_{j=1}^{n} b_{j1} z_j^k \\ &= dv(k) + (b_{n1} - d) z_n^k + \sum_{j=1}^{n-1} b_{j1} z_j^k \\ &\leq dv(k) + (b_{n1} - d) u(k) + \sum_{j=1}^{n-1} b_{j1} u(k) \\ &= dv(k) + (1 - d) u(k). \end{aligned}$$

Analog zeigt man

$$v(k+1) \geq du(k) + (1-d) v(k).$$

Das ergibt für $k \to \infty$ wegen $d \leq \frac{1}{2}$

$$u(k+1) - v(k+1) \leq (1-2d)(u(k) - v(k)) \leq \ldots \leq (1-2d)^{k+1}(u(0) - v(0)) \to 0,$$

was zu zeigen war.

Zum Beweis von (iii) hat man in (ii) nur sukzessive als Anfangsvektor $\underline{x}^0$ die Spaltenvektoren von B zu nehmen, was auf $B^{k+1} = B^k B \to [\underline{y} \ldots \underline{y}]$ führt. □

5.5 Matrizenrechnung und Regressionsanalyse

Literatur: Schira *Statistische Methoden der VWL und BWL*, Pearson 2003.

5.5.1 Das einfache lineare Modell

Das folgende *lineare* Modell widerspiegele eine ökonomische Hypothese für die Beziehung zwischen zwei Variablen X und Y:

$$Y = \beta_0 + \beta_1 X.$$

Die *Parameter* β_0 und β_1 sind auf der Grundlage von Beobachtungen zu "bestimmen".

5.5.1 Beispiel (Konsumhypothese von Keynes).
Der gesamtwirtschaftliche Konsum C sei eine Funktion des verfügbaren Einkommens Y^{verf}:

$$C = \beta_0 + \beta_1 Y^{verf}.$$

Dabei wird $\beta_1 \in (0,1)$ als *marginale Konsumneigung*, $\beta_0 > 0$ als *autonomer Konsum* (einkommensunabhängiger Konsum) interpretiert.
Im allgemeinen Modell oben sind dann $Y := C$ und $X := Y^{verf}$. $\diamond$

5.5.2 Der ökonometrische Modellansatz berücksichtigt weitere Einflüsse auf Y im ökonomischen Modell $Y = \beta_0 + \beta_1 X$ und fügt eine *Störvariable* U (eine *Zufallsvariable*) hinzu,

$$Y = \beta_0 + \beta_1 X + U.$$

Geht man zu einer Beobachtung der Werte von X und Y über, so gilt für eine *Stichprobe* aus n Werten analog

$$y_i = \beta_0 + \beta_1 x_i + u_i \qquad (i = 1, \ldots, n), \tag{5.13}$$

dabei heissen

y_i **endogene** Variablen (beobachtbar),

x_i **exogene** Variablen (beobachtbar),

u_i **latente** Variablen (nicht beobachtbar),

β_0, β_1 **Modellparameter** oder **Koeffizienten**, ("wahre" Werte)

$\hat{\beta}_0$, $\hat{\beta}_1$ **Schätzparameter** oder **Schätzer** (geschätzte Werte).

Die wahren Werte der Parameter β_0 und β_1 sind unbekannt, sie können nur *geschätzt* werden - das ist die Aufgabe der ***Regressionsanalyse***.

Die Wahrscheinlichkeitsverteilungen der latenten Variablen sind unbekannt (oft wird aber Normalverteilung vorausgesetzt), aber man unterstellt für Erwartungswert E, Varianz V und Kovarianz Cov, dass

$$E(u_i) = 0 \text{ und } V(u_i) = \sigma^2\text{=const. für alle } i$$

sowie

$$Cov(u_i, u_k) = 0 \text{ für alle } i, k = 1, \ldots, n,\ i \neq k.$$

Bei der Schätzung geht es darum, eine Schätzgerade $y = \hat{\beta}_0 + \hat{\beta}_1 x$ zu finden, die der unbekannten Modellgeraden $y = \beta_0 + \beta_1 x$ nahe kommt. Man vergleicht

$$\begin{array}{ll} y_i = \beta_0 + \beta_1 x_i + u_i & \text{(Beobachtungswerte)}, \\ \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i & \text{(geschätzte Werte)}. \end{array}$$

◇

5.5.3 Methode der kleinsten Quadrate.

Gegeben beobachtete x_i, y_i, $i = 1, \ldots, n$, finde $(\hat{\beta}_0, \hat{\beta}_1)$, so dass

$$\varphi(\hat{\beta}_0, \hat{\beta}_1) := \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \tag{5.14}$$

minimal wird.

Schreiben wir nun das System (5.13)

$$y_i = \beta_0 + \beta_1 x_i + u_i \qquad (i = 1, \ldots, n)$$

in Matrixschreibweise. Man setze - **wir schreiben hier Vektoren fett** -

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}.$$

Dann hat (5.13) die völlig gleichbedeutende Darstellung

$$\mathbf{y} = X\boldsymbol{\beta} + \mathbf{u}. \tag{5.15}$$

Die Zielfunktion (5.14) in der Methode der kleinsten Quadrate lautet dann, äquivalent umgeschrieben mit Matrizen und Vektoren,

$$\varphi(\widehat{\boldsymbol{\beta}}) = (\mathbf{y} - X\widehat{\boldsymbol{\beta}})^\mathsf{T}(\mathbf{y} - X\widehat{\boldsymbol{\beta}}) = \mathbf{y}^\mathsf{T}\mathbf{y} - 2\widehat{\boldsymbol{\beta}}^\mathsf{T} X^\mathsf{T}\mathbf{y} + \widehat{\boldsymbol{\beta}}^\mathsf{T} X^\mathsf{T} X\widehat{\boldsymbol{\beta}},$$

wobei $\widehat{\boldsymbol{\beta}}$ der (gesuchte!) geschätzte $\boldsymbol{\beta}$ -Vektor ist.

Der Gradient einer gegebenen quadratischen Funktion

$$\begin{aligned} g(p,q) &= a_1 p + a_2 q \quad + \quad b_{11}p^2 + 2b_{12}pq + b_{22}q^2 \\ &= (p,q)\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + (p,q)\begin{pmatrix} b_{11} & b_{12} \\ b_{12} & b_{22} \end{pmatrix}\begin{pmatrix} p \\ q \end{pmatrix} \end{aligned}$$

lautet offenbar (ausrechnen!)

$$\nabla g(p,q) = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + 2\begin{pmatrix} b_{11} & b_{12} \\ b_{12} & b_{22} \end{pmatrix}\begin{pmatrix} p \\ q \end{pmatrix}.$$

Damit gilt für den Gradienten der Kleinste-Quadrate-Funktion φ im Minimum $\widehat{\boldsymbol{\beta}}^{\,*}$ der Aufgabe (5.14)

$$\nabla\varphi(\widehat{\boldsymbol{\beta}}^{\,*}) = -2X^{\mathsf{T}}\mathbf{y} + 2X^{\mathsf{T}}X\widehat{\boldsymbol{\beta}}^{\,*} = \mathbf{o}.$$

Die Matrix

$$X^{\mathsf{T}}X = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix}\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}$$

(mit $\sum = \sum_{i=1}^{n}$) ist genau dann invertierbar, wenn

$$\det X^{\mathsf{T}}X = n\sum x_i^2 - (\sum x_i)^2 \neq 0\,.$$

Es gilt in diesem Falle sogar $\det X^{\mathsf{T}}X > 0$, wie man leicht ausrechnet, also ist dann wegen $n > 0$ die Matrix $X^{\mathsf{T}}X$ positiv definit und somit φ (strikt) konvex, d.h., der Punkt $\widehat{\boldsymbol{\beta}}^{\,*}$ aus

$$\mathbf{o} = -2X^{\mathsf{T}}\mathbf{y} + 2X^{\mathsf{T}}X\widehat{\boldsymbol{\beta}}^{\,*} \;\Leftrightarrow\; \widehat{\boldsymbol{\beta}}^{\,*} = (X^{\mathsf{T}}X)^{-1}X^{\mathsf{T}}\mathbf{y} \tag{5.16}$$

definiert den eindeutig bestimmten Minimalpunkt der Aufgabe (5.14).

Die Komponenten $\widehat{\beta_0}^{*}$ und $\widehat{\beta_1}^{*}$ von $\widehat{\boldsymbol{\beta}}^{\,*}$ kann man als Lösung des LGS $X^{\mathsf{T}}X\boldsymbol{\beta} = X^{\mathsf{T}}\mathbf{y}$ mit der Cramerschen Regel leicht ausrechnen:

$$X^{\mathsf{T}}\mathbf{y} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}$$

ergibt

$$\widehat{\beta_0}^* = \frac{\begin{vmatrix} \sum y_i & \sum x_i \\ \sum x_i y_i & \sum x_i^2 \end{vmatrix}}{\det X^\mathsf{T} X} = \frac{(\sum y_i)(\sum x_i^2) - (\sum x_i)(\sum x_i y_i)}{n \sum x_i^2 - (\sum x_i)^2}$$

$$\widehat{\beta_1}^* = \frac{\begin{vmatrix} n & \sum y_i \\ \sum x_i & \sum x_i y_i \end{vmatrix}}{\det X^\mathsf{T} X} = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2}$$

Mit den in der Statistik üblichen Bezeichnungen für Mittelwerte

$$\overline{x} = \frac{1}{n}\sum x_i, \quad \overline{y} = \frac{1}{n}\sum y_i, \quad \overline{x^2} = \frac{1}{n}\sum x_i^2, \quad \overline{xy} = \frac{1}{n}\sum x_i y_i$$

folgen sofort die in Lehrbüchern oft angegebenen Formeln

$$\widehat{\beta_0}^* = \frac{\overline{y}\,\overline{x^2} - \overline{x}\,\overline{xy}}{\overline{x^2}\,\overline{x}^2}, \quad \widehat{\beta_1}^* = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2}\,\overline{x}^2}.$$

◇

5.5.2 Multiple lineare Regressionsanalyse

Lineare ökonomische Modelle mit zwei Variablen sind häufig zu einschränkend, deshalb spezifiziert man oft Modelle mit zwei oder mehr exogenen Variablen, d.h.,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + U.$$

wobei U wieder als Störvariable (Zufallsvariable) angesehen wird.

5.5.4 Beispiel (Erweiterte Konsumhypothese). Der gesamtwirtschaftliche Konsum C wird als Funktion des verfügbaren Einkommens Y^{verf} und des Vermögens W angesehen und wir spezifizieren deshalb

$$C = \beta_0 + \beta_1 Y^{verf} + \beta_2 W + U, \quad 0 < \beta_1 < 1, \ \beta_0, \beta_2 > 0.$$

Mit einem solchen Modell kann man argumentieren, dass die Aktienkurse einen Einfluss auf die private Konsumnachfrage ausüben. ◇

5.5.5 Beobachtet man nun wieder Stichprobenvektoren $\mathbf{x}^j$, $j = 1, \ldots, k$, und $\mathbf{y}$ aus je n Elementen, so definiert man via

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \ X = \begin{pmatrix} 1 & x^1{}_1 & \ldots & x^k{}_1 \\ 1 & x^1{}_2 & \ldots & x^k{}_2 \\ \vdots & \vdots & & \vdots \\ 1 & x^1{}_n & \ldots & x^k{}_n \end{pmatrix}, \ \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \ \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

das multiple lineare Regressionsmodell in Matrixschreibweise

$$\mathbf{y} = X\boldsymbol{\beta} + \mathbf{u}. \tag{5.17}$$

Zu den oben getroffenen Annahmen über die latenten Variablen u_i wird zusätzlich gefordert, dass sie unabhängig und normalverteilt seien. Ferner setzen wir voraus:

$$X^\top X \text{ ist regulär,}$$

offenbar ist dabei $X^\top X$ eine $(k+1, k+1)$-Matrix. ◇

5.5.6 Methode der kleinsten Quadrate (MKQ)

Gegeben $\mathbf{y}$ und X, finde $\widehat{\boldsymbol{\beta}} \in \mathbb{R}^{k+1}$, so dass

$$\Phi(\widehat{\boldsymbol{\beta}}) = (\mathbf{y} - X\widehat{\boldsymbol{\beta}})^\top(\mathbf{y} - X\widehat{\boldsymbol{\beta}})$$

minimal wird. In den Klammern stehen Vektoren der Länge n, $\Phi(\widehat{\boldsymbol{\beta}})$ ist also für jedes $\widehat{\boldsymbol{\beta}}$ eine reelle Zahl !

Ausgerechnet, erhalten wir

$$\Phi(\widehat{\boldsymbol{\beta}}) = \mathbf{y}^\top\mathbf{y} - 2\widehat{\boldsymbol{\beta}}^\top X^\top\mathbf{y} + \widehat{\boldsymbol{\beta}}^\top X^\top X\widehat{\boldsymbol{\beta}},$$

somit hat die notwendige Optimalitätsbedingung die Form

$$\nabla\Phi(\widehat{\boldsymbol{\beta}}) = -2X^\top\mathbf{y} + 2X^\top X\widehat{\boldsymbol{\beta}} = \mathbf{o}.$$

Die Voraussetzungen sichern, dass $X^\top X$ positiv definit ist: $\beta^\top X^\top X\beta = \|X\beta\|^2 \geq 0$ $(\forall\beta)$ in der euklidischen Norm, also ist $X^\top X$ positiv semidefinit, daher - weil regulär nach Voraussetzung - auch positiv definit. Damit hat der eindeutige Optimalpunkt in der MKQ die gleiche Form wie bei (5.14), (5.16) - nur der Typ von X bzw. die Länge von $\boldsymbol{\beta}$ unterscheiden sich -, nämlich

$$\widehat{\boldsymbol{\beta}}^{*} = (X^\top X)^{-1}X^\top\mathbf{y}.$$ ◇

5.6 Hinreichende Optimalitätsbedingung 2. Ordnung bei einer Gleichungsrestriktion

Literatur: Rommelfanger Bd. 2 §7.4, Sydsaeter-Hammond §14.4

Wir betrachten die Optimierungsaufgabe

$$\text{(P)} \qquad \text{Minimiere } f(x_1, x_2) \quad \text{bezüglich} \quad g(x_1, x_2) = 0,$$

wobei f und g zweimal stetig differenzierbar auf $\mathbb{R}^2$ sind.

Lagrangefunktion der Aufgabe (P):

$$L(x_1, x_2, \lambda) := f(x_1, x_2) - \lambda g(x_1, x_2)$$

Hessematrix von L bezüglich (x_1, x_2) im Punkt (x_1^0, x_2^0, λ), λ fest:

$$H := \begin{pmatrix} f_{x_1x_1}(x_1^0, x_2^0) - \lambda g_{x_1x_1}(x_1^0, x_2^0) & f_{x_1x_2}(x_1^0, x_2^0) - \lambda g_{x_1x_2}(x_1^0, x_2^0) \\ f_{x_1x_2}(x_1^0, x_2^0) - \lambda g_{x_1x_2}(x_1^0, x_2^0) & f_{x_2x_2}(x_1^0, x_2^0) - \lambda g_{x_2x_2}(x_1^0, x_2^0) \end{pmatrix}$$

Der Nullpunkt im $\mathbb{R}^2$ sei mit $\underline{o}$ bezeichnet.

5.6.1 Satz (hinreichende Bedingung 2. Ordnung für lokales Minimum).

Seien in (x_1^0, x_2^0, λ) die Lagrange-Bedingungen der Aufgabe (P) erfüllt und bezeichne $\nabla g(x_1^0, x_2^0)$ wieder den Gradienten von g in (x_1^0, x_2^0).

Falls $\underline{y}^\mathsf{T} H \underline{y} \;> 0$ für alle $\underline{y} = (y_1, y_2)^\mathsf{T} \neq \underline{o}$ mit

$$\nabla g(x_1^0, x_2^0)^\mathsf{T} \underline{y} \;= 0 \tag{5.18}$$

gilt, so ist (x_1^0, x_2^0) strikter lokaler Minimalpunkt von (P), d.h., es gilt

$$f(x_1, x_2) \;> f(x_1^0, x_2^0)$$

für alle Punkte (x_1, x_2) "nahe" (x_1^0, x_2^0), die $g(x_1, x_2) = 0$ erfüllen.

Im Falle strikter lokaler Maxima dreht sich das Zeichen $>$ jeweils um. $\diamond$

Der Beweis wird in der Vorlesung gegeben.

5.6.2 Definitheit unter linearen Nebenbedingungen:

Die hinreichende Bedingung für lokale Minima ist von der Form

$$\underline{y}^\mathsf{T} H \underline{y} > 0 \quad \forall \underline{y} \neq \underline{o} : \; \underline{c}^\mathsf{T} \underline{y} = 0$$

mit einem Vektor $\underline{c}^\mathsf{T} = (c_1, c_2)$ und symmetrischer $(2,2)$-Matrix H, denn H ist Hesse-Matrix mit stetigen zweiten partiellen Ableitungen.

Die hinreichende Bedingung für lokale Maxima ist von der Form

$$\underline{y}^\mathsf{T} H \underline{y} < 0 \quad \forall \underline{y} \neq \underline{o} : \; \underline{c}^\mathsf{T} \underline{y} = 0$$

Man sagt im ersten Falle, dass die Matrix H *positiv definit* - im zweiten Falle *negativ definit* - *auf der Geraden* $\{\underline{y} \in \mathbb{R}^2 \,|\, \underline{c}^\mathsf{T} \underline{y} = 0\}$ ist. $\diamond$

5.6.3 Beispiel: Man betrachte die Aufgabe

$$\begin{array}{ll} \min & f(x_1,x_2) := \tfrac{1}{2}(x_1^2 - x_2^2) - x_1 \\ \text{bezüglich} & g(x_1,x_2) := -x_1 + 2x_2 = 0 \end{array} \tag{5.19}$$

Lösung durch Substitution: Es gilt $x_1 = 2x_2$, eingesetzt in die Zielfunktion ergibt sich

$$\varphi(x_2) := f(2x_2, x_2) = \frac{1}{2}(4x_2^2 - x_2^2) - 2x_2 = \frac{3}{2}x_2^2 - 2x_2.$$

Folglich

$$\varphi'(x_2) = 3x_2 - 2 = 0 \;\Leftrightarrow\; x_2 = \frac{2}{3}, \quad \text{somit } x_1 = \frac{4}{3}.$$

Die hinreichende Bedingung 2. Ordnung für freie Extrema lautet

$$\varphi''(\tfrac{2}{3}) = 3 > 0,$$

d.h., $\underline{x}^0 = (\frac{4}{3}, \frac{2}{3})$ ist strikter lokaler Minimalpunkt.

Die Lagrange-Funktion ist im Beispiel gleich

$$L(x_1,x_2,\lambda) = \tfrac{1}{2}(x_1^2 - x_2^2) - x_1 - \lambda(-x_1 + 2x_2)$$

Die Lagrange-Bedingungen lauten

$$\begin{array}{rrrrcl} (i) & x_1 & & +\lambda & = & 1 \\ (ii) & & -x_2 & -2\lambda & = & 0 \\ (iii) & -x_1 & +2x_2 & & = & 0 \end{array}$$

Die Lösung des Systems der Lagrange-Bedingungen ist

$$x_1^0 = \frac{4}{3}, \quad x_2^0 = \frac{2}{3}, \quad \lambda = -\frac{1}{3}.$$

Die Bedingung (5.18) ist hier sehr einfach:

$$-y_1 + 2y_2 = 0, \quad \text{d.h. } y_2 = \tfrac{1}{2}y_1, \tag{5.20}$$

da $\nabla g(x_1,x_2) = (-1,2)^{\mathsf{T}}$ für alle (x_1,x_2).

Die Hesse-Matrix von L bezüglich (x_1,x_2) in (x_1^0, x_2^0, λ) lautet hier

$$H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad \text{(Hesse-Matrix).}$$

Sei nun $\underline{y} = (y_1,y_2)^{\mathsf{T}} \neq \underline{o}$ ein beliebiger Vektor mit (5.20). Dann gilt

$$\underline{y}^{\mathsf{T}} H \underline{y} = (y_1,y_2)\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = y_1^2 - y_2^2$$

und nach Einsetzen von (5.20)

$$\underline{y}^{\mathsf{T}} H \underline{y} = y_1^2 - y_2^2 = y_1^2 - \frac{1}{4} y_1^2 = \frac{3}{4} y_2^2 > 0 \quad \forall \underline{y} \neq \underline{o} \text{ mit } y_2 = \tfrac{1}{2} y_1.$$

Folglich ist $\underline{x}^0 = (\frac{4}{3}, \frac{2}{3})$ strikter lokaler Minimalpunkt. ◇

5.6.4 Spezialfall eines Kriteriums von Debreu

In der ökon. Literatur sind von Gerard Debreu (1983 Nobelpreis für Ökonomie) bewiesene Kriterien zum Checken der positiven/negativen Definitheit einer symmetrischen Matrix H unter linearen Nebenbedingungen

$$B\underline{y} = \underline{o}$$

(B muss dabei eine Matrix passender Ordnung sein) sehr beliebt, die mit Determinanten/Unterdeterminanten einer "geänderten" Matrix

$$\begin{pmatrix} O & B \\ B^{\mathsf{T}} & H \end{pmatrix}$$

arbeiten, vgl. am besten die Originalarbeit *G. Debreu, Definite and semidefinite quadratic forms, Econometrica 20 (1952) 295-300.*

Wir betrachten hier nur den Spezialfall für die Aufgabe (P).

Sei (x_1^0, x_2^0, λ) ein Punkt, der den Lagrange-Bedingungen von (P) genügt, und sei $\nabla g(x_1^0, x_2^0) \neq \underline{o}$.

Sei wieder H die Hesse-Matrix der Lagrange-Funktion L bezüglich (x_1, x_2) in (x_1^0, x_2^0, λ):

$$H = \begin{pmatrix} h_{11} & h_{12} \\ h_{12} & h_{22} \end{pmatrix}$$

(H symmetrisch!) und

$$\underline{c} = \nabla g(x_1^0, x_2^0).$$

Satz.
Unter den gestellten Voraussetzungen gelten folgende Kriterien:

1. Die Bedingung

$$\det \begin{pmatrix} 0 & \underline{c}^{\mathsf{T}} \\ \underline{c} & H \end{pmatrix} < 0$$

   ist notwendig und hinreichend dafür, dass $\underline{y}^{\mathsf{T}} H \underline{y} > 0$ für alle $\underline{y} \neq \underline{o}$ mit $\underline{c}^{\mathsf{T}} \underline{y} = 0$ gilt. Also ist unter dieser Bedingung (x_1^0, x_2^0) strikter lokaler Minimalpunkt von (P).

2. Die Bedingung
$$\det \begin{pmatrix} 0 & \underline{c}^\mathsf{T} \\ \underline{c} & H \end{pmatrix} > 0,$$
ist notwendig und hinreichend dafür, dass $\underline{y}^\mathsf{T} H \underline{y} < 0$ für alle $\underline{y} \neq \underline{o}$ mit $\underline{c}^\mathsf{T}\underline{y} = 0$ gilt. Also ist unter dieser Bedingung (x_1^0, x_2^0) strikter lokaler Maximalpunkt von (P).

Beweis des Satzes. Wir beweisen nur Aussage 1., das heisst,
$$\det \begin{pmatrix} 0 & c^\mathsf{T} \\ c & H \end{pmatrix} < 0 \quad \Leftrightarrow \quad y^\mathsf{T} H y > 0 \;\forall y \neq 0 : \; c^\mathsf{T} y = 0.$$
Der Beweis von Aussage 2. ist analog. Wir führen den Beweis für eine beliebige symmetrische Matrix $H = (h_{ij})$ der Ordnung 2 und einen beliebigen Vektor $c \in \mathbb{R}^2 \setminus \{0\}$.

Wir definieren eine $(3,3)$-Matrix $R(\mu)$ mit nützlichen Eigenschaften:
$$R(\mu) := \begin{pmatrix} -1 & o^\mathsf{T} \\ \mu c & H + \mu c c^\mathsf{T} \end{pmatrix} = \begin{pmatrix} -1 & c^\mathsf{T} \\ \mu c & H \end{pmatrix} \begin{pmatrix} 1 & c^\mathsf{T} \\ o & I \end{pmatrix}, \tag{5.21}$$
wobei $\mu > 0$, $o = (0,0)^\mathsf{T}$ und $I =$ Einheitsmatrix. Dann gilt offenbar
$$\det R(\mu) = \det \begin{pmatrix} -1 & c^\mathsf{T} \\ \mu c & H \end{pmatrix} \quad \text{und} \quad \det R(\mu) = -\det(H + \mu c c^\mathsf{T}). \tag{5.22}$$
Sei zunächst $\det \left(\begin{smallmatrix} 0 & c^\mathsf{T} \\ c & H \end{smallmatrix}\right) < 0$. Dann folgt

$\exists \mu_0 > 0 : \; \det \left(\begin{smallmatrix} -1/\mu & c^\mathsf{T} \\ c & H \end{smallmatrix}\right) < 0 \;\; \forall \mu \geq \mu_0$ wegen der Stetigkeit der Determinante,

$\Rightarrow$ $\det R(\mu) < 0 \;\; \forall \mu \geq \mu_0$ nach (5.22) und wegen $\det R(\mu) = \mu \det \left(\begin{smallmatrix} -1/\mu & c^\mathsf{T} \\ c & H \end{smallmatrix}\right)$, $\mu > 0$,

$\Rightarrow$ $\det(H + \mu c c^\mathsf{T}) = \det(h_{ij} + \mu c_i c_j) > 0 \;\; \forall \mu \geq \mu_0$ nach (5.22),

$\Rightarrow$ $\exists \widetilde{\mu} \geq \mu_0 : \det(H + \widetilde{\mu} c c^\mathsf{T}) > 0$ und $h_{ii} + \widetilde{\mu} c_i^2 > 0$ $(i = 1, 2)$, denn wegen $c \neq 0$ ist z.B. $c_1 \neq 0$, also wird für grosses $\mu > 0$ auch $\mu c_1^2 > 0$ gross – und für ein gewisses $\widetilde{\mu} \geq \mu_0$ gilt somit $h_{11} + \widetilde{\mu} c_1^2 > 0$, schliesslich muss dann auch $h_{22} + \widetilde{\mu} c_2^2$ grösser Null sein wegen der Rechenregeln für $(2,2)$-Determinanten,

$\Rightarrow$ die Matrix $(H + \widetilde{\mu} c c^\mathsf{T})$ ist positiv definit (Argument der linearen Algebra),

$\Rightarrow$ $y^T H y = y^T (H + \widetilde{\mu} c c^\mathsf{T}) y > 0 \;\; \forall y \neq 0 : \; c^\mathsf{T} y = 0,$

was für diese Richtung zu zeigen war.

Sei nun $y^T H y > 0 \;\; \forall y \neq 0 : \; c^\mathsf{T} y = 0$. Wir wählen ein festes $\bar{y} \neq 0$ mit $c^\mathsf{T} \bar{y} = 0$. Dann bildet $\{\bar{y}, c\}$ eine orthogonale Basis des $\mathbb{R}^2$, die wir zu einer $(2,2)$-Matrix
$$B = [\bar{y} \; c] \quad \text{(spaltenweise geschrieben)}$$

zusammenfassen. Dann folgt wegen $\bar{y}^\mathsf{T} H \bar{y} > 0$ sowie wegen $c \neq 0$ und somit $c^\mathsf{T} c > 0$: Es existiert ein $\mu_0 > 0$, so dass für alle $\mu \geq \mu_0$ gilt

$$\varrho(\mu) := \det \begin{pmatrix} \bar{y}^\mathsf{T} H \bar{y} & \bar{y}^\mathsf{T} H c \\ \bar{y}^\mathsf{T} H c & c^\mathsf{T} H c + \mu (c^\mathsf{T} c)^2 \end{pmatrix} = \mu (\bar{y}^\mathsf{T} H \bar{y})(c^\mathsf{T} c)^2 + (\bar{y}^\mathsf{T} H \bar{y})(c^\mathsf{T} H c) - (\bar{y}^\mathsf{T} H c)^2 > 0.$$

Folglich gilt unter Beachtung von $c^\mathsf{T} \bar{y} = 0$

$$\det\left(B^\mathsf{T} (H + \mu c c^\mathsf{T}) B \right) = \det \left(\begin{bmatrix} \bar{y}^\mathsf{T} \\ c^\mathsf{T} \end{bmatrix} (H + \mu c c^\mathsf{T}) [\bar{y}\ c] \right) = \varrho(\mu) > 0 \quad \forall \mu \geq \mu_0. \tag{5.23}$$

Nach den Determinantenregeln gilt $\det\left(B^\mathsf{T} (H + \mu c c^\mathsf{T}) B \right) = (\det B)^2 (\det(H + \mu c c^\mathsf{T}))$ und somit folgt aus (5.23) wegen der Regularität von B

$$\det(H + \mu c c^\mathsf{T}) > 0 \quad \forall \mu \geq \mu_0.$$

Die Beziehungen in (5.22) liefern dann

$$\det \begin{pmatrix} -1 & c^\mathsf{T} \\ \mu c & H \end{pmatrix} = \det R(\mu) = -\det(H + \mu c c^\mathsf{T}) < 0 \quad \forall \mu \geq \mu_0. \tag{5.24}$$

Die Matrix $\left(\begin{smallmatrix} 0 & c^\mathsf{T} \\ c & H \end{smallmatrix}\right)$ ist regulär - wie man es sich leicht überlegt -, also ist $\det \left(\begin{smallmatrix} 0 & c^\mathsf{T} \\ c & H \end{smallmatrix}\right) \neq 0$. Aus (5.24) folgt dann nach der Laplace-Entwickung für Determinanten

$$0 > \det \begin{pmatrix} -1 & c^\mathsf{T} \\ \mu c & H \end{pmatrix} = -\det H + \mu \det \begin{pmatrix} 0 & c^\mathsf{T} \\ c & H \end{pmatrix} \quad \forall \mu \geq \mu_0 \tag{5.25}$$

und folglich

$$\det \begin{pmatrix} 0 & c^\mathsf{T} \\ c & H \end{pmatrix} < 0,$$

da im Falle $\det \left(\begin{smallmatrix} 0 & c^\mathsf{T} \\ c & H \end{smallmatrix}\right) > 0$ die rechte Seite in (5.25) mit $\mu \to \infty$ positive Werte annähme. Damit ist auch die zweite Richtung in Aussage 1. bewiesen. $\square$