

Mathematik für Wirtschaftswissenschaftler

(K. Steffen, Heinrich-Heine-Universität Düsseldorf, WS 2006/07)

Kapitel 1: Zahlen und Rechnen

1.1 Die Zahlbereiche und die Grundrechenarten

Die für die Wirtschaftswissenschaft relevanten **Zahlbereiche** sind:

$\mathbb{N} = \{1, 2, 3, \dots\}$ die Menge der **natürliche Zahlen** (ohne Null),

$\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ die Menge der **ganze Zahlen**,

\mathbb{Q} die Menge der **rationalen Zahlen** $\frac{m}{n}$ mit $m \in \mathbb{Z}$ und $n \in \mathbb{N}$,

\mathbb{R} die Menge der **reellen Zahlen** (Punkte auf der *Zahlengeraden*).

Jeder dieser Zahlbereiche erweitert den vorangehenden,

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}.$$

Die Erweiterungen sind jeweils nötig, weil gewisse Operationen in den kleineren Bereichen nicht ausführbar sind: Subtraktion nicht in \mathbb{N} , Division nicht in \mathbb{Z} , Wurzelziehen und Grenzwertbildung nicht in \mathbb{Q} . (Auch in \mathbb{R} kann man nicht alle Rechenoperationen ausführen, z.B. nicht das Quadratwurzelziehen aus negativen Zahlen. Daher wird in der Mathematik noch ein größerer Zahlbereich eingeführt, der Bereich der sog. *komplexen Zahlen* $\mathbb{C} \supset \mathbb{R}$, in dem auch die Zahl $i = \sqrt{-1}$ existiert; aber dieser Zahlbereich ist für die Wirtschaftswissenschaft von geringer Bedeutung.)

BEISPIELE: 1) Zahlen aus \mathbb{Q} :

$\frac{1}{2}, \frac{7}{8}, \frac{-3}{15}$, alle abbrechenden Dezimalbrüche wie $12.394 = 12 + \frac{394}{1000} = \frac{12394}{1000}$, alle periodischen Dezimalbrüche wie $0.2\bar{3} = 0.2333\dots = \frac{7}{30}$ und $-9.\bar{09} = -9.090909\dots = -\frac{100}{11}$.

2) Zahlen aus $\mathbb{R} \setminus \mathbb{Q}$ (sog. *irrationale Zahlen*):

$\sqrt{2} = 1.4142\dots$, die *natürliche Basis* $e = 2.7182\dots$, die *Kreiszahl* $\pi = 3.1415\dots$ und überhaupt alle nichtabbrechenden und nichtperiodischen Dezimalbrüche. ■

Konkrete Zahlen in der Wirtschaftswissenschaft sind immer abbrechende Dezimalbrüche, also rational. Für die mathematische Theorie aber sind irrationale Zahlen unverzichtbar, — sonst könnte man z.B. nicht einmal quadratische Gleichungen lösen, und sehr einfache Funktionen hätten keine Extremstellen mehr. Mit einer solch unbefriedigenden Theorie könnte man auch in der Wirtschaftswissenschaft nichts anfangen.

Für Teilbereiche der Zahlbereiche sind Notationen wie die folgenden praktisch:

$\mathbb{N}_0 = \mathbb{Z}_{\geq 0} = \{0, 1, 2, 3, \dots\}$ (natürliche Zahlen mit Null),

$\mathbb{Q}_{\neq 0} = \{r \in \mathbb{Q} : r \neq 0\}$ (rationale Zahlen ungleich Null),

$\mathbb{R}_{> 0} = \{x \in \mathbb{R} : x > 0\}$ (positive reelle Zahlen).

Die Grundrechenarten **Addition** und **Multiplikation** kann man mit je zwei Zahlen a, b innerhalb jedes Zahlbereichs ausführen. Das Ergebnis heißt die **Summe** $a + b$ der *Summanden* a, b bzw. das **Produkt** $a \cdot b$ (oft einfach ab notiert) der *Faktoren* a, b . Dabei gilt $0 + a = a$ und $1 \cdot a = a$ für alle Zahlen a , d.h. *in einer Summe darf man Summanden mit Wert Null weglassen und in einem Produkt Faktoren mit Wert Eins*, ohne das Ergebnis zu ändern. Man kann auch mehr als zwei Zahlen addieren und multiplizieren. Die wichtigste Rechenregel dafür ist:

- Die Summe ist unabhängig von der Reihenfolge und von der Art der Zusammenfassung (Klammerung) der Summanden;
- das Produkt ist unabhängig von der Reihenfolge und von der Art der Zusammenfassung (Klammerung) der Faktoren.

BEISPIEL: Was ist günstiger: Erst ein Rabatt auf den Preis und danach die MwSt aufschlagen, oder nach Aufschlag der MwSt den prozentual gleichen Rabatt abziehen?

Was ist günstiger: Sparen mit wachsendem Zins 3%, 4%, 5% oder Sparen mit fallendem Zins 5%, 4%, 3%?

Die Antwort ist sofort klar, wenn man das tut, was bei vielen ähnlichen Fragestellungen angeraten ist, nämlich

- *multiplikativ denken!*

Den Rabattabzug von $r\%$ berechnet man nämlich durch Multiplikation des Preises mit dem "Verbilligungsfaktor" $1 - \frac{r}{100}$, den Mehrwertsteuerzuschlag durch Multiplikation mit dem "Verteuerungsfaktor" 1.16 (zur Zeit noch), den jährlichen Zinszuschlag bei Verzinsung mit $p\%$ durch Multiplikation des Kapitals mit dem sog. *Aufzinsungsfaktor* $q = 1 + \frac{p}{100}$. Die Alternativen sind daher jeweils gleichwertig, keine ist günstiger; denn die Reihenfolge bei der Multiplikation des Preises bzw. Anfangskapitals mit den diversen Faktoren ist ja egal! ■

Für Summen und Produkte von n Zahlen a_1, a_2, \dots, a_n sind folgende Notationen mit dem **Summenzeichen** bzw. **Produktzeichen** üblich und praktisch:

$$\sum_{i=1}^n a_i = a_1 + a_2 + \dots + a_n, \quad \prod_{i=1}^n a_i = a_1 a_2 \cdot \dots \cdot a_n.$$

Dabei ist der sog. *Laufindex* i ein Buchstabensymbol für die Nummern, mit denen die Summanden oder Faktoren numeriert werden; a_i ist der Summand / Faktor mit Nummer i . Der angegebene *Indexlaufbereich* $i = 1$ bis n zeigt an, dass hier die Zahlen $1, 2, \dots, n$ als Nummern verwendet werden. Die Wahl des Buchstaben für den Laufindex ist dabei irrelevant, und als Laufbereich kann auch eine andere Nummernmenge wie etwa $0, 1, \dots, n-1$ verwendet werden. Also ist die obige Summe bzw. das obige Produkt gleich

$$\sum_{j=1}^{n-1} b_j = b_0 + b_1 + \dots + b_{n-1} \quad \prod_{k=0}^{n-1} b_k = b_0 b_1 \cdot \dots \cdot b_{n-1},$$

wenn immer die Zahlen b_0, b_1, \dots, b_{n-1} dieselben sind wie a_1, a_2, \dots, a_n in irgendeiner Reihenfolge (wobei mehrfach auftretende Zahlen unter den a_i und b_j gleich oft vorkommen). In welcher Reihenfolge man die einzelnen Additionen / Multiplikationen beim Ausrechnen durchführt, ist egal; man darf (und sollte) also möglichst geschickt zusammenfassen.

BEISPIELE: (Berechnung endlicher Summen)

$$1) \sum_{i=1}^5 i = 1 + 2 + 3 + 4 + 5 = 15;$$

$$2) \sum_{j=1}^4 j^2 = 1^2 + 2^2 + 3^2 + 4^2 = 1 + 4 + 9 + 16 = 30;$$

$$3) \sum_{\substack{k=-100 \\ k \neq 0}}^{100} \frac{1}{k} = \frac{1}{-100} + \frac{1}{-99} + \dots + \frac{1}{-2} + \frac{1}{-1} + \frac{1}{1} + \frac{1}{2} + \dots + \frac{1}{99} + \frac{1}{100} = 0;$$

4) bei lauter gleichen Summanden $a_i = a$ gilt $\sum_{i=1}^n a = \underbrace{a + a + \dots + a}_n = n \cdot a$, d.h. man kann n gleiche Summanden mit Wert a zu einem Produkt $n \cdot a$ zusammenfassen. ■

Summen, deren Summanden nach einem einfachen Bildungsgesetz gebildet sind, lassen sich unter Umständen durch eine "geschlossene Formel" berechnen, in der kein Summenzeichen mehr vorkommt. Ein einfaches Beispiel für diese Situation, das auch für die Wirtschaftswissenschaft wichtig ist (vgl. Kap. 2) ist folgendes:

BEISPIEL: Die Zahlen a_i bilden eine **arithmetische Progression** (oder: *arithmetische Folge*), wenn die Differenz aufeinander folgender Glieder stets dieselbe ist, $a_{i+1} - a_i = d$. Dann gilt die

• **arithmetische Summenformel:**

$$\sum_{i=1}^n a_i = n \cdot \frac{a_1 + a_n}{2}$$

in Worten: *Die Summe einer arithmetischen Progression ist der Mittelwert von erstem und letztem Glied multipliziert mit der Anzahl der Summanden.*

(Das gilt auch bei anderer Nummerierung etwa beginnend mit $i = 0$. Aber Achtung, bei Nummerierung mit $i = 0, 1, \dots, n$ hat man $n + 1$ Summanden, also ist der Wert der Summe dann $\frac{1}{2}(n+1)(a_0 + a_n)$.) Man erkennt die Gültigkeit dieser Summenformel sofort, wenn man dieselbe Summe mit umgekehrter Reihenfolge der Summanden addiert:

$$\begin{array}{cccccccc} a_1 & + & a_2 & + & \dots & + & a_{n-1} & + & a_n \\ + & a_n & + & a_{n-1} & + & \dots & + & a_2 & + & a_1 \\ \hline = & (a_1 + a_n) & + & (a_2 + a_{n-1}) & + & \dots & + & (a_{n-1} + a_2) & + & (a_n + a_1) & = & n \cdot (a_1 + a_n) \end{array}$$

Die letzte Gleichung gilt, weil in der letzten Zeile lauter gleiche Summanden $(a_i + a_{n+1-i}) = (a_1 + a_n)$ stehen; denn wenn man in einer arithmetischen Progression von i zu $i+1$ übergeht, so verändert sich a_i um d , aber a_{n+1-i} um $-d$, die Summe $(a_i + a_{n+1-i})$ also überhaupt nicht. Das Doppelte der zu berechnenden arithmetischen Summe ist also $n(a_1 + a_n)$, und die arithmetische Summenformel folgt mit Division durch 2. Hier noch einige konkrete arithmetische Summen:

$$\sum_{i=1}^{100} i = 100 \cdot \frac{1 + 100}{2} = 100 \cdot 50.5 = 5050 \quad (\text{Gauß}),$$

$$\text{allgemeiner } 1 + 2 + \dots + n = \sum_{i=1}^n i = n \cdot \frac{1 + n}{2} = \frac{1}{2}n(n + 1),$$

$$\sum_{j=0}^{100} \left(\frac{1}{2}j - 2\right) = 101 \cdot \frac{-2 + 48}{2} = 101 \cdot 23 = 2323. \quad \blacksquare$$

BEISPIELE: (endliche Produkte)

$$1) \prod_{i=1}^5 i = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120;$$

allgemein heißt $\prod_{i=1}^n i = 1 \cdot 2 \cdot \dots \cdot n$ die **Fakultät** von $n \in \mathbb{N}$, notiert $n!$;

$$2) \prod_{j=1}^4 j^2 = 1^2 \cdot 2^2 \cdot 3^2 \cdot 4^2 = (1 \cdot 2 \cdot 3 \cdot 4)^2 = 24^2 = 576;$$

$$3) \prod_{k=0}^{100} (k-50)(k+50) = 0, \quad \text{weil der Faktor zu } k=50 \text{ null ist und es gilt:}$$

- Ein Produkt hat den Wert Null, genau wenn einer seiner Faktoren Null ist.

4) Bei lauter gleichen Faktoren $a_i = a$ heißt $\prod_{i=1}^n a = \underbrace{a \cdot a \cdot \dots \cdot a}_n$ die **Potenz** der Basis a mit dem **Exponenten** $n \in \mathbb{N}$. Dafür gelten dann die **Potenzrechenregeln**:

$$\begin{aligned} a^{m+n} &= \underbrace{a \cdot a \cdot \dots \cdot a}_{m+n} = \underbrace{(a \cdot \dots \cdot a)}_m \cdot \underbrace{(a \cdot \dots \cdot a)}_n = a^m \cdot a^n, \\ a^{m \cdot n} &= \underbrace{a \cdot a \cdot \dots \cdot a}_{m \cdot n} = \underbrace{(a \cdot \dots \cdot a)}_m \cdot \dots \cdot \underbrace{(a \cdot \dots \cdot a)}_m = (a^m)^n, \\ (ab)^n &= \underbrace{(ab) \cdot (ab) \cdot \dots \cdot (ab)}_n = \underbrace{(a \cdot \dots \cdot a)}_n \cdot \underbrace{(b \cdot \dots \cdot b)}_n = a^n \cdot b^n. \end{aligned}$$

Produkte a^k , $a^k b^l$, $a^k b^l c^m$, ... von Potenzen von Zahlen oder Variablen a, b, c, \dots nennt man **Potenzprodukte**. Addiert man Potenzen oder Potenzprodukte, die evtl. zuvor noch mit gegebenen Zahlen multipliziert wurden, so spricht man von einer **Linearkombination** von Potenzen bzw. Potenzprodukten mit den gegebenen Zahlen als **Koeffizienten**. Solche Terme kommen oft vor, z.B. sog. *Polynomausdrücke* wie $3 - 2x + x^3 - 5x^4$. (Dies steht für $3 + (-2)x + 0x^2 + 1x^3 + (-5)x^4$, die Koeffizienten bei $x^0 = 1, x, x^2, x^3, x^4$ sind also der Reihe nach 3, -2, 0, 1, -5. Beachte, dass der nicht aufgeschriebene Summand mit x^2 als Null interpretiert werden muss. Dagegen ist der nicht aufgeschriebene Koeffizient von x^3 als 1 zu interpretieren und nicht etwa als 0. Wäre er 0, so würde man den x^3 Term gar nicht aufschreiben!) Ein Polynomausdruck in drei Variablen ist z.B. $2x^3 y z^2 - 7xy^4 + \frac{1}{2}y^2 z^3 - x^4$ (mit den Koeffizienten 2, -7, $\frac{1}{2}$, -1 vor den aufgeführten Potenzprodukten).

Die Grundrechenart **Subtraktion** kann man im Bereich \mathbb{N} nicht uneingeschränkt ausführen, wohl aber innerhalb von \mathbb{Z} , \mathbb{Q} und \mathbb{R} . In diesen Bereichen gibt es zu jeder Zahl b genau eine Zahl c mit $b + c = 0$. Man nennt c das *additive Inverse*, das *Negative* oder die *Gegenzahl* zu b und bezeichnet diese Zahl mit $-b$. (Auf der Zahlengeraden wird $-b$ gegenüber von b in gleicher Entfernung vom Nullpunkt gezeichnet. Achtung: $-b$ muss nicht unbedingt eine positive Zahl sein, auch wenn das vorangestellte Minuszeichen dies suggeriert; wenn b negativ ist, also links vom Nullpunkt liegt, so liegt $-b$ rechts von Null, ist also positiv! Zur deutlichen Unterscheidung von b und seiner Gegenzahl $-b$ schreibt man manchmal auch $+b$ für b ; dann muss natürlich $+b$ auch nicht unbedingt eine positive Zahl sein.) Allgemein ist $-b$ das Produkt von b mit der Gegenzahl zu 1, also $-b = (-1) \cdot b$. Daraus folgen mit $1 \cdot (-1) = -1$ und $(-1) \cdot (-1) = +1$ die bekannten **Vorzeichenregeln** $(-a) \cdot b = -(ab) = a \cdot (-b)$ ("Minus mal Plus gleich Minus") und $(-a) \cdot (-b) = +(ab)$ ("Minus mal Minus gleich Plus"). Die **Differenz** $a - b$ von zwei Zahlen a, b wird nun erklärt als die Summe von a und der Gegenzahl zu b , also $a - b := a + (-b) = a + (-1) \cdot b$; dann ist $x = a - b$ die eindeutige Lösung der Gleichung $x + b = a$.

Ganz analog erklärt man die Grundrechenart **Division**, die man in \mathbb{N} und \mathbb{Z} nur eingeschränkt ausführen kann, in \mathbb{Q} und \mathbb{R} aber ohne Einschränkungen (außer der einen, dass nicht durch Null dividiert werden darf). In \mathbb{Q} und \mathbb{R} gibt es zu jeder Zahl $b \neq 0$ genau eine Zahl c mit $bc = 1$. Man nennt diese Zahl c das *multiplikative Inverse*, das *Reziproke* oder den *Kehrwert* von b und schreibt dafür $\frac{1}{b}$, $1/b$, $1:b$ oder b^{-1} . Der **Quotient** von zwei Zahlen a und $b \neq 0$ wird nun definiert als das Produkt $a \cdot \frac{1}{b}$ von a mit dem Kehrwert von b und er wird $\frac{a}{b}$, a/b , $a:b$ oder ab^{-1} notiert. Der Quotient $x = \frac{a}{b}$ ist dann die eindeutige Lösung der Gleichung $bx = a$. Beachte, dass dies nur für $b \neq 0$ gilt; denn *Division durch Null ist nicht erklärt und daher nicht erlaubt*. (Für $b = 0$, $a \neq 0$ hat die Gleichung $bx = a$ keine Lösung, für $b = 0$, $a = 0$ ist jede Zahl x eine Lösung.)

Im Unterschied zu Addition und Multiplikation gilt:

- Bei Subtraktion und Division kommt es auf die Reihenfolge der Zahlen an, bei mehr als zwei Zahlen auch auf die Art der Zusammenfassung (Klammerung). Wenn das Rechenergebnis von der Klammerung abhängt, so muss man auch Klammern setzen!

Also Vorsicht! Es ist $a - b \neq b - a$ und $a/b \neq b/a$ (außer wenn $a = b$); und $(a - b) + c \neq a - (b + c)$ (außer wenn $c = 0$) sowie $(a/b) \cdot c \neq a/(b \cdot c)$ und $(a/b)/c \neq a/(b/c)$ (außer wenn $c = 1$). Man darf also z.B. nicht $a/b/c$ ohne Klammern schreiben, weil dieser Term je nach Klammerung verschiedene Werte haben kann. Für die Übersichtlichkeit ist es gut, Vereinbarungen zum Einsparen von Klammern zu treffen. Z.B. vereinbart man über *Summen mit wechselnden Vorzeichen*

$$a - b + c - d - e + f + \dots - z := a + (-b) + c + (-d) + (-e) + f + \dots + (-z),$$

d.h. man "arbeitet die Summe von links nach rechts ab". Weitere Vereinbarungen zum Klammernsparen sind: *Punktrechnung* $(\cdot, :)$ geht vor *Strichrechnung* $(+, -)$, also z.B. $a + bc = a + (b \cdot c)$ (nicht etwa $= (a + b) \cdot c$) und $a/b - c = \frac{a}{b} - c$ (nicht etwa $= \frac{a}{b-c}$), und *Multiplikation* geht vor *Division*, also z.B. $a/bc = \frac{a}{bc}$ (nicht etwa $= \frac{a}{b} \cdot c$). Aber " $a/b/c$ " darf man nicht schreiben, weil es als $(a/b)/c = \frac{a}{bc}$ oder als $a/(b/c) = \frac{ac}{b}$ interpretiert werden könnte, was verschieden ist (außer für $c = 1$).

Oft wird ein Quotient $\frac{a}{b}$ als "Bruch" bezeichnet. Genau genommen ist ein *Bruch* aber ein Paar von Zahlen a und b , genannt *Zähler* und *Nenner* des Bruchs, und wenn $b \neq 0$ ist, so ordnet man einem solchen Bruch eine Zahl als *Wert des Bruches* zu, nämlich das Ergebnis $\frac{a}{b}$ der Division des Zählers durch den Nenner. Brüche mit verschiedenen Zählern und Nennern können durchaus denselben Wert haben, z.B. ist $\frac{1}{2} = \frac{2}{4} = \frac{500}{1000} = \frac{0.25}{0.5} = \frac{\sqrt{2}}{\sqrt{8}}$. (Wir beschränken uns keineswegs darauf, dass Zähler und Nenner immer ganze Zahlen sein sollen, wie man hier sieht.) Für das Umgehen mit Brüchen gelten dann die bekannten **Bruchrechenregeln**, die das Ergebnis von Rechenoperationen mit Bruchwerten wieder als Wert eines Bruchs darstellen (wobei selbstverständlich alle Nenner $\neq 0$ sein müssen):

$$\begin{aligned} \frac{a}{b} &= \frac{a \cdot c}{b \cdot c}, & \frac{a}{b} &= \frac{a/d}{c/d} && (\text{Erweitern und Kürzen}), \\ \frac{a}{b} \cdot c &= \frac{a \cdot c}{b}, & \frac{a}{b} \cdot \frac{c}{d} &= \frac{a \cdot c}{b \cdot d} && (\text{Multiplikation von Brüchen}), \\ \frac{1}{\frac{c}{d}} &= \frac{d}{c}, & \frac{\frac{a}{b}}{\frac{c}{d}} &= \frac{a \cdot d}{b \cdot c} && (\text{Kehrwert und Quotient von Brüchen; "Doppelbruchregeln"}), \\ \frac{a}{d} \pm \frac{c}{d} &= \frac{a \pm c}{d} &&&& (\text{Addition/Subtraktion "gleichnamiger" Brüche}), \\ \frac{a}{b} \pm \frac{c}{d} &= \frac{a \cdot d}{b \cdot d} \pm \frac{b \cdot c}{b \cdot d} = \frac{ad \pm bc}{b \cdot d} &&&& (\text{Addition/Subtraktion beliebiger Brüche durch "gleichnamig machen"}) \end{aligned}$$

Noch eine Warnung vor der (in der Schule üblichen) Schreibweise $l\frac{m}{n}$ für sog. *gemischte Brüche*. Damit ist nämlich nicht das Produkt $l \cdot \frac{m}{n}$ gemeint, sondern die Summe $l + \frac{m}{n}$, wobei l, m, n natürliche Zahlen mit $m < n$ sind (“ l Ganze und m n -tel”). Also bedeutet $2\frac{1}{3}$ in dieser Schreibweise $\frac{7}{3}$ und nicht $\frac{2}{3}$. Da das zu Missverständnissen führen kann, sollte man die Schreibweise der gemischten Brüche vermeiden.

Brüche mit Nenner 100 kommen in der Wirtschaftswissenschaft besonders oft vor, weil sie Grundlage der **Prozentrechnung** sind. $p\%$ (p Prozent) ist nämlich eine Abkürzung für $\frac{p}{100}$. Diese Zahl kann man dezimal darstellen, indem man in der (von links mit Nullen aufgefüllten) Dezimalbruchdarstellung von p den Dezimalpunkt um zwei Stellen nach links verschiebt. Also ist $12\% = 0.12$, $5\% = 0.05$, $2.5\% = 0.025$ etc. (Mehr zur Prozentrechnung sagen wir in Kap. 2.

Die Dezimalbruchschreibweise reeller Zahlen ist eine Darstellung als Linearkombination von Potenzen der “Basis” 10 mit “Ziffern” $0, 1, \dots, 8$ oder 9 als Koeffizienten. Bei nicht-ganzen Zahlen treten dabei auch Potenzen von 10 mit negativem Exponenten auf, d.h. Potenzen von $\frac{1}{10}$. Allgemein erklärt man die **Potenzen mit ganzen Exponenten** zu einer Basis $b \neq 0$ durch $b^n := \underbrace{b \cdot b \cdot \dots \cdot b}_n$ für natürliche Zahlen $n \in \mathbb{N}$ wie früher, durch $b^0 := 1$, wenn der Exponent Null ist, und durch $b^{-n} := \underbrace{\frac{1}{b} \cdot \frac{1}{b} \cdot \dots \cdot \frac{1}{b}}_n = \left(\frac{1}{b}\right)^n = 1/b^n$, wenn

der Exponent eine negative ganze Zahl $-n \in \mathbb{Z}_{<0}$ ist (also $n \in \mathbb{N}$). (“ 0^0 ” bleibt undefiniert.) Man kann sich dann leicht davon überzeugen, dass die drei **Potenzgesetze für beliebige ganze Exponenten** genau so gelten, wie wir sie früher aufgeführt haben, und dass außerdem auch noch das weitere Potenzgesetz $(a/b)^n = a^n/b^n = a^n b^{-n}$ für beliebige $n \in \mathbb{Z}$ gültig ist (vorausgesetzt natürlich die Basen a, b sind nicht Null).

BEISPIEL: Die Potenzen der Zahl 10 mit positiven und negativen Exponenten spielen eine besondere Rolle bei der Darstellung von positiven Zahlen als **Dezimalbruch**.

$$a = d_m \dots d_2 d_1 d_0 \quad \text{mit Ziffern } d_i \in \{0, 1, \dots, 9\}$$

ist nämlich die Dezimalschreibweise für die natürliche Zahl

$$a = d_m 10^m + \dots + d_2 10^2 + d_1 10 + d_0 = \sum_{i=0}^m d_i 10^i,$$

$$b = 0.z_1 z_2 \dots z_n \quad \text{mit Ziffern } z_j \in \{0, 1, \dots, 9\}$$

ist die dezimale Darstellung der rationalen Zahl

$$b = z_1 10^{-1} + z_2 10^{-2} + \dots + z_n 10^{-n} = \sum_{j=1}^n z_j 10^{-j},$$

und ein allgemeiner **abbrechender Dezimalbruch** (*endlicher Dezimalbruch*)

$$d_m \dots d_1 d_0 . z_1 z_2 \dots z_n \quad \text{mit Ziffern } d_i, z_j \in \{0, 1, \dots, 9\}$$

ist nichts anderes als der dezimale Code für die Summe von a und b . In dieser Form kann man genau die rationalen Zahlen schreiben, die als Bruch darstellbar sind mit Zähler aus \mathbb{N}_0 und mit einer natürlichen Zahl als Nenner, die nur von den Primzahlen 2 und 5 geteilt wird; denn durch Erweitern kann man dann auch eine Zehnerpotenz als Nenner erreichen.

Alle anderen positiven reellen Zahlen, auch so einfache rationale Zahlen wie $\frac{1}{3}$, $\frac{1}{7}$, $\frac{1358}{1100}$, ... sind nur als **unendlicher Dezimalbruch**

$$r = d_m \dots d_1 d_0 . z_1 z_2 \dots z_n z_{n+1} \dots$$

darstellbar, mit endlich vielen Ziffern d_i vor und unendlich vielen Ziffern z_j hinter dem Dezimalpunkt. Diese Darstellung ist eine Abkürzung ist für

$$r = \sum_{i=0}^m d_i 10^i + \lim_{n \rightarrow \infty} \sum_{j=1}^n z_j 10^{-j},$$

d.h. für den Grenzwert bei $n \rightarrow \infty$ des Dezimalbruchs, der durch Abbrechen mit der n -ten Stelle nach dem Dezimalpunkt entsteht. (Dieser unterscheidet sich von r höchstens noch um 10^{-n} . Da 10^{-n} mit wachsendem n beliebig klein wird, approximiert der abgebrochene Dezimalbruch die Zahl r beliebig genau, wenn n groß genug gewählt wird. Mehr über Grenzwerte muss man hier nicht wissen.)

Rationale Zahlen kann man übrigens daran erkennen, dass sie eine *periodische Dezimalbruchdarstellung* haben, d.h. ab einer gewissen Stelle nach dem Dezimalpunkt wiederholt sich eine endliche Ziffernfolge (die oft durch Überstreichung gekennzeichnet wird) immer wieder. Zum Beispiel gilt $\frac{1}{3} = 0.\overline{3} = 0.333\dots$, $\frac{1}{7} = 0.\overline{142857} = 0.142857142857142857\dots$ und $\frac{1358}{1100} = 1.23\overline{45} = 1.23454545\dots$. Irrationale Zahlen wie $\sqrt{2}$, e , π haben dagegen eine unendliche Dezimalbruchdarstellung ohne Periodizität (auch wenn der Taschenrechner bzw. Computer nur endlich viele Stellen zeigt).

Alles, was hier gesagt wurde, gilt übrigens ganz analog, wenn man statt 10 eine andere Zahl aus $\mathbb{N}_{\geq 2}$ als "*Basis des Zahlensystems*" nimmt, wobei als "Ziffern" dann die Zahlen aus \mathbb{N}_0 verwendet werden, die kleiner als die gewählte Basis sind. Eine Rolle spielen hier, besonders in der Informatik, vor allem die Basen 2 (*dyadische Zahldarstellung*; Ziffern 0,1) und 16 (*hexadezimale Zahldarstellung*; Ziffern 0,1,...,15; statt der letzten sechs "Ziffern" 10,11,12,13,14,15 verwendet man dabei die Großbuchstaben A,B,C,D,E,F). ■

Ein wichtiges Rechengesetz, das Addition und Multiplikation verbindet, müssen wir noch ansprechen. Das **Distributivgesetz** besagt in seiner einfachsten Form:

- *Man multipliziert eine Summe mit einem Faktor, indem man jeden Summanden damit multipliziert und die entstehenden Produkte aufsummiert.*

Also: $a(b + c) = ab + ac$ und allgemeiner

$$a \cdot \sum_{i=1}^n b_i = \sum_{i=1}^n ab_i, \quad a \cdot (b_1 + \dots + b_n) = ab_1 + \dots + ab_n.$$

Dies ist nichts anderes als die Regel über das **Ausklammern**: Faktoren, die allen Summanden gemeinsam sind, kann man vor die Klammer (vor die Summe) ziehen. Ein Spezialfall ist die Regel über die **Vorzeichenänderung bei Summen**: Das Negative einer Summe bildet man, indem man jeden einzelnen Summanden durch sein Negatives ersetzt, also bei allen Summanden das Vorzeichen ändert:

$$-(a - b + c - e - f + \dots - z) = -a + b - c + e + f - \dots + z.$$

(Beachte, dass das links nicht aufgeschriebene Vorzeichen zu a ein "+" ist und auch geändert werden muss. Es wäre aber ein krasser Fehler, allein bei diesem ersten Summanden a das Vorzeichen umzukehren! Das Minuszeichen vor der Klammer bedeutet ja die Multiplikation mit -1 , und nach dem Distributivgesetz muss dann zur Auflösung der Klammer *jeder* Summand mit -1 multipliziert werden.)

Eine allgemeinere Version des Distributivgesetzes ist folgende:

- Man multipliziert Summen miteinander, indem man die Produkte aufaddiert, die entstehen, wenn man auf alle möglichen Arten aus jeder Summe jeweils einen Summanden herausgreift und die herausgegriffenen Zahlen miteinander multipliziert.

Also:

$$\begin{aligned}(a+b)(c+d) &= ac + ad + bc + bd, \\ (a+b)(c+d+e) &= ac + ad + ae + bc + bd + be, \\ (a+b)(c+d)(e+f) &= ace + acf + ade + adf + bce + bcf + bde + bdf.\end{aligned}$$

BEISPIEL: Eine Folge von Zahlen $a_i \neq 0$ bildet eine **geometrische Progression** (oder: *geometrische Folge*), wenn der Quotient aufeinander folgender Zahlen stets derselbe ist, $a_{i+1}/a_i = q$, also $a_2 = a_1q$, $a_3 = a_1q^2$, $a_4 = a_1q^3, \dots$. Die Summe aus den Gliedern einer solchen geometrischen Folge kann man berechnen, wenn $q \neq 1$ ist (im Fall $q = 1$ hat man lauter gleiche Summanden, also kein Problem), und zwar mit folgendem Trick, der das Distributivgesetz und die Potenzgesetze verwendet:

$$\begin{aligned}& (1-q) \cdot (a_1 + a_2 + a_3 + \dots + a_n) \\ &= (1-q) \cdot (a_1 + a_1q + a_1q^2 + \dots + a_1q^{n-1}) \\ &= a_1 \cdot (1-q) \cdot (1 + q + q^2 + \dots + q^{n-1}) \\ &= a_1 \cdot (1 + q + q^2 + \dots + q^{n-1} - q - q^2 - q^3 - \dots - q^n) \\ &= a_1 \cdot (1 - q^n) = a_1 - a_1q^n = a_1 - a_{n+1}.\end{aligned}$$

Also gilt die (auch in ökonomischen Anwendungen wichtige; siehe Kap. 2)

- **geometrische Summenformel:**

$$\sum_{i=1}^n a_i = \frac{a_1 - a_{n+1}}{1 - q}$$

in Worten: Die Summe einer geometrischen Progression zum Quotienten $q \neq 1$ ist die mit $1 - q$ dividierte Differenz des ersten Summanden und des ersten nicht mehr berücksichtigten Gliedes.

Konkret ist z.B.

$$1 + 2 + 4 + 8 + 16 + 32 = \frac{1 - 64}{1 - 2} = 63 \quad (q = 2),$$

$$\sum_{k=2}^n 3^k = \frac{3^2 - 3^{n+1}}{1 - 3} = \frac{3^{n+1} - 3^2}{2} = \frac{3}{2}(3^n - 3) \quad (q = 3),$$

$$1 - \frac{1}{4} + \frac{1}{16} - \frac{1}{64} = \frac{1 - \frac{1}{256}}{1 + \frac{1}{4}} = \frac{255 \cdot 4}{256 \cdot 5} = \frac{51}{64} \quad (q = -\frac{1}{4}).$$

Wenn q echt zwischen -1 und 1 liegt, so wird $a_{n+1} = a_1q^n$ mit wachsendem n dem Betrag nach beliebig klein, d.h. die geometrische Summe strebt dann mit $n \rightarrow \infty$ gegen den Grenzwert $\frac{a_1}{1-q}$. In der Mathematik schreibt man dafür symbolisch

$$\sum_{i=1}^{\infty} a_i = \sum_{i=1}^{\infty} a_1q^{i-1} = \frac{a_1}{1-q}$$

und nennt $\frac{a_1}{1-q}$ den Wert der *unendlichen Reihe* $a_1 + a_2 + a_3 + \dots$. Konkret ist z.B.

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \dots = \frac{1/2}{1 - 1/2} = 1 \quad (q = \frac{1}{2}).$$

Das bedeutet natürlich nicht, dass man unendlich viele Summanden addieren kann, sondern nur, dass die Summen, die man durch Addition der ersten n Glieder erhält, den angegebenen Werten beliebig nahe kommen, wenn n groß genug ist. ■

Ein anderes Beispiel für die Anwendung des allgemeinen Distributivgesetzes liefern die binomischen Formeln:

BEISPIEL: Die bekannten **binomischen Formeln**

$$(a + b)^2 = (a + b) \cdot (a + b) = a \cdot a + a \cdot b + b \cdot a + b \cdot b = a^2 + 2ab + b^2$$

und entsprechend (mit $-b$ statt b)

$$(a - b)^2 = a^2 - 2ab + b^2$$

sowie

$$(a + b) \cdot (a - b) = a \cdot a - a \cdot b + b \cdot a - b \cdot b = a^2 - b^2$$

sind einfache Anwendungen des Distributivgesetzes. Auch

$$\begin{aligned} (a + b)^3 &= (a + b) \cdot (a + b) \cdot (a + b) \\ &= a \cdot a \cdot a + a \cdot a \cdot b + a \cdot b \cdot a + a \cdot b \cdot b \\ &\quad + b \cdot a \cdot a + b \cdot a \cdot b + b \cdot b \cdot a + b \cdot b \cdot b \\ &= a^3 + 3a^2b + 3ab^2 + b^3 \end{aligned}$$

kann man (mit etwas Geduld) noch durch direktes Ausmultiplizieren berechnen. Bei Potenzen eines "Binoms" (damit ist eine Summe von zwei Summanden gemeint) mit größeren Exponenten $n \in \mathbb{N}$,

$$(a + b)^n = \underbrace{(a + b) \cdot (a + b) \cdot \dots \cdot (a + b)}_n,$$

ist das Ausmultiplizieren problematisch. Klar ist aber nach dem Distributivgesetz, dass das Ergebnis eine Summe von Potenzprodukten der Form $a^{n-i}b^i$ sein wird. Dabei erhält man jedes Potenzprodukt $a^{n-i}b^i$ gerade so oft, wie man aus den n Klammern i -mal den zweiten Summanden b auswählen kann (und aus den restlichen $n - i$ Klammern dann den ersten Summanden a). Eine Zählung ergibt den sog. **Binomialkoeffizienten**

$$\binom{n}{i} := \frac{n \cdot (n-1) \cdot \dots \cdot (n-i+1)}{i \cdot (i-1) \cdot \dots \cdot 1} = \frac{n!}{i!(n-i)!}$$

als Anzahl der Möglichkeiten, aus n Objekten (Klammern) i Stück ohne Berücksichtigung der Reihenfolge auszuwählen (mit $\binom{n}{i} := 1$ für $i = 0$ oder $i = n$). Daher gilt der

• **Binomialsatz:**
$$(a + b)^n = \sum_{i=0}^n \binom{n}{i} a^{n-i} b^i$$

$$= a^n + \binom{n}{1} a^{n-1} b + \binom{n}{2} a^{n-2} b^2 + \dots + \binom{n}{n-1} a b^{n-1} + b^n.$$

(In dieser Formel sind evtl. auftauchende Terme der Form 0^0 als 1 zu interpretieren.)

Die Binomialkoeffizienten sind natürliche Zahlen und können angeordnet werden im sog. *Pascalschen Dreieck*, in dem jede Zahl die Summe der zwei darüber stehenden ist.

In der binomischen Formel zum Exponenten n treten gerade die Zahlen als Binomialkoeffizienten auf, die in der n -ten Zeile des Dreiecks stehen. ■

$$\begin{array}{ccccccc} & & & & & & 1 \\ & & & & & & 1 & 1 \\ & & & & & 1 & 2 & 1 \\ & & & & 1 & 3 & 3 & 1 \\ & & & 1 & 4 & 6 & 4 & 1 \\ & 1 & 5 & 10 & 10 & 5 & 1 \\ & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{array}$$

Es gibt (seltener gebrauchte) Verallgemeinerungen für Potenzen von "Multinomen" (Summen von mehr als zwei Summanden) $a + b + c$, $a + b + c + d$, $a_1 + a_2 + \dots + a_m$. Zum Beispiel ist

$$(a + b + c)^n = \sum \frac{n!}{i!j!k!} a^i b^j c^k,$$

wobei über alle Wahlen von i, j, k aus $\{0, 1, \dots, n\}$ mit $i + j + k = n$ zu summieren ist, und der allgemeine *Multinomial*satz besagt für n -te Potenzen von Summen mit m Summanden:

$$(a_1 + a_2 + \dots + a_m)^n = \sum \frac{n!}{i_1! \cdot i_2! \cdot \dots \cdot i_m!} a_1^{i_1} a_2^{i_2} \cdot \dots \cdot a_m^{i_m},$$

wo über alle Wahlen der Exponenten i_1, i_2, \dots, i_m aus $\{0, 1, \dots, n\}$ mit $i_1 + i_2 + \dots + i_m = n$ zu summiert wird. All dies ist nichts weiter als das allgemeine Distributivgesetz zusammen mit einer Abzählung, auf wieviele verschiedene Arten dasselbe Potenzprodukt von a_1, a_2, \dots, a_m entstehen kann, wenn man sich aus den n Klammern $(a_1 + a_2 + \dots + a_m)$ jeweils einen Summanden herausgreift und die herausgegriffenen Zahlen miteinander multipliziert. Die so bestimmten Anzahlen

$$\binom{n}{i_1 i_2 \dots i_m} = \frac{n!}{i_1! \cdot i_2! \cdot \dots \cdot i_m!} \quad (i_1 + i_2 + \dots + i_m = n)$$

sind die Koeffizienten, die im Multinomialssatz auftreten, und heißen dementsprechend *Multinomialkoeffizienten*. (Das sind natürliche Zahlen; die Faktoren der Fakultäten im Nenner lassen sich also allesamt gegen Teiler der Fakultät im Zähler kürzen.)

1.2 Auflösung von Gleichungen

Mathematische Modelle der Finanz- und Wirtschaftswissenschaft stellen das Ergebnis eines ökonomischen Vorgangs oft durch einen "Rechenterm" dar, in den viele diesen Vorgang charakterisierenden Parameter / Variablen eingehen. In der Finanzmathematik (s. Kap. 2) wird z.B. der Endkapitalstand K_t nach Ablauf der Zeitspanne t (Jahre) bei einem Ratenparvorgang mit Zinseszins im einfachsten Fall (Zinsperioden = Ratenperioden, z.B. Monate, Quartale oder Jahre) dargestellt durch $K_t = K_0 \cdot q^n + R \cdot \frac{q^n - 1}{q - 1}$, und die rechte Seite hängt von den Parametern K_0 (Anfangskapital), R (Ratenhöhe), n (Anzahl der Zins- und Ratenperioden im Zeitraum t) und $q = 1 + \frac{p \cdot t}{100 \cdot n}$ (Aufzinsungsfaktor für eine Ratenperiode zum Zinsfuß $p\%$) ab. Auf das Problem der Auflösung einer solchen **Gleichung** nach einer der darin vorkommenden Variablen stößt man immer dann, wenn ein bestimmtes Ergebnis gewünscht ist und gefragt wird, welchen Wert man für eine bestimmte Variable zu wählen hat, damit dieses Ergebnis auch genau herauskommt. Diese Variable wird dann die **Unbekannte** der Gleichung genannt, und jeden Wert der Unbekannten, für den die Gleichung richtig ist, nennt man eine **Lösung** der Gleichung. Von den anderen in der Gleichung auftauchenden Variablen hängt die Gleichung und ihre Lösung(en) natürlich auch ab, aber für diese Variablen kennt man entweder aus dem ökonomischen Kontext die Zahlenwerte oder man stellt sich bei der Behandlung der Gleichung auf den Standpunkt, dass diese Werte als bekannt anzusehen sind; diese anderen Variablen werden **Parameter** in der Gleichung genannt.

Fragt man z.B. danach, wie obiger Ratenparvorgang verzinst sein muss, um ein bestimmtes Endkapital zu produzieren, so ist q die Unbekannte in der Rentengleichung und K_t, K_0, R, n sind Parameter. In einer parameterabhängigen Situation hat man gewissermaßen nicht nur eine Gleichung, sondern eine ganze Schar von Gleichungen – für jede mögliche Wahl der Parameterwerte eine. Dann hängt auch die Lösung von den jeweiligen Parameterwerten ab, kann also nicht als Zahlenwert angegeben werden, sondern allenfalls mit einer Lösungsformel, einem "Rechenterm", der die Werte der Lösung(en) in Abhängigkeit von den jeweiligen Parameterwerten angibt. Unter der expliziten **Auflösung der Gleichung** nach der Unbekannten versteht man die Angabe eines solchen Rechenterms, der die Lösungen der Gleichung parametrisiert. Allerdings ist das schon bei relativ einfachen Gleichungen, wie z.B. bei der Rentengleichung oben mit der Unbekannten q , nicht möglich. Man kann dann allenfalls für konkrete Parameterwerte durch geeignete mathematische Verfahren die Lösung(en) näherungsweise berechnen.

Es gibt auch ökonomische Situationen, in denen man mehrere Variablen als gleichberechtigte Unbekannte betrachten möchte, ohne eine von ihnen als (beliebiger Werte fähige) Unbekannte auszuzeichnen und die andern als (auf feste Werte fixierte) Parameter anzusehen. Wenn man dann nur eine Gleichung hat, welche die Unbekannten verbindet, so wird diese im Allgemeinen eine ganze Schar von Lösungen haben: Man kann z.B. alle Unbekannten außer einer auf beliebige Werte festsetzen und dann mit der Gleichung dazu einen Wert der letzten Unbekannten so bestimmen, dass die Gleichung erfüllt ist. Eine Situation, bei der man aufgrund ökonomischer Überlegungen genau eine Lösung erwartet (etwa genau eine optimale Wahl der Variablen), kann von einer solchen Gleichung für mehrere Unbekannte also nicht mathematisch modelliert werden; vielmehr braucht man weitere Gleichungen (ökonomische Gesetze), welche die Unbekannten in Beziehung setzen und so die Vielfalt der Lösungen reduzieren. Bei einer sinnvollen Problemstellung ist dabei (von Ausnahmen abgesehen) *die Anzahl der Gleichungen gleich der Anzahl der Unbekannten*. Hat man weniger Gleichungen als Unbekannte, so gibt es im allgemeinen "zu viele" Lösungen, nämlich ganze Lösungsscharen, hat man aber mehr Gleichungen als Unbekannte, so gibt es im Allgemeinen "zu wenige" Lösungen, nämlich gar keine.

In diesem Abschnitt betrachten wir aber nur *eine* Gleichung für *eine* Unbekannte, die wir, wie in der Mathematik allgemein üblich, oft mit “ x ” bezeichnen und die Werte in Bereich der reellen Zahlen \mathbb{R} annehmen kann oder auch in einer vorgegebenen Teilmenge $G \subset \mathbb{R}$, der *Grundmenge* für die Gleichung. Für ökonomische Variablen sind z.B. oft nur positive Werte sinnvoll, und wenn x eine solche Variable ist, so wird man die Menge $\mathbb{R}_{>0}$ der positiven reellen Zahlen als Grundmenge wählen. Eine Gleichung für die Unbekannte x hat nun die allgemeine Form

$$T(x) = 0 \quad \text{oder} \quad T_{\text{links}}(x) = T_{\text{rechts}}(x) \quad (x \in G),$$

wobei $T_{\text{links}}(x)$ und $T_{\text{rechts}}(x)$ reellwertige “Rechterme” sind, genannt die **linke Seite** bzw. **rechte Seite** der Gleichung. Die kürzere erste Form $T(x) = 0$ lässt sich immer herstellen, indem man den Term $T(x)$ als Differenz $T_{\text{links}}(x) - T_{\text{rechts}}(x)$ definiert; doch taucht bei vielen Gleichungen die Unbekannte zunächst auf beiden Seiten auf, daher ist die zweite Form vielleicht vorzuziehen. In den Termen können dabei neben der Unbekannten x auch noch die Parameter a, b, c, \dots der Gleichung vorkommen; das kann man durch die genauere Schreibweise $T(x; a, b, c, \dots)$ ausdrücken, wenn es erforderlich ist. Andererseits braucht die Variable x in den Termen überhaupt nicht aufzutreten, auch wenn die Notation Abhängigkeit von x suggeriert, sondern es kann sich auch um sog. *Konstanten* handeln, d.h. reelle Zahlen oder Ausdrücke, die nur von den Parameterwerten abhängen, nicht vom Wert der Unbekannten x . (Das ist sinnvoll; denn man kann ja eine Konstante c immer auch als Term $c - x + x$ schreiben, in dem x formal vorkommt.)

Lösungen der Gleichung sind diejenigen reellen Zahlen (in der Grundmenge), die bei Einsetzen für die Unbekannte x auf der linken und auf der rechten Seite der Gleichung dieselbe reelle Zahl ergeben, für die also die Gleichung tatsächlich “gilt”. Dazu gehört auch die Bedingung, dass sich diese Werte in die Terme überhaupt einsetzen lassen, also zum Definitionsbereich der Terme auf beiden Seiten gehören. Nicht einsetzen kann man z.B. Werte von x , für die ein in den Termen vorkommender Nenner Null wird; denn durch Null darf man ja nicht dividieren. Nicht einsetzen kann man weiter Werte von x , die einen vorkommenden Radikanden, also einen Ausdruck, aus dem die Quadratwurzel zu ziehen ist, negativ machen; denn Quadratwurzeln haben nur nichtnegative Zahlen. Nicht einsetzen kann man schließlich Werte von x , die einen Ausdruck nichtpositiv machen, der zu logarithmieren ist; denn nur positive Zahlen haben Logarithmen. Das sind die wichtigsten Restriktionen für das Einsetzen von Werten der Variablen in Rechenterme.

Die Gesamtheit aller Lösungen heißt die **Lösungsmenge der Gleichung**. Wenn eine Grundmenge spezifiziert ist, so gehören zur Lösungsmenge nur die Lösungen, die auch in der Grundmenge liegen; es gibt dann vielleicht noch weitere Lösungen außerhalb der Grundmenge, aber die sind für die ins Auge gefasste Problemstellung eben nicht interessant oder unsinnig (etwa eine negative Zahl als Lösung für ein Problem, in dem eine unbekannte Laufzeit zu bestimmen ist). Im Prinzip kann als Lösungsmenge jede Teilmenge der Grundmenge auftreten: Die Lösungsmenge kann leer sein, d.h. die Gleichung hat überhaupt keine Lösung (wie z.B. die nicht erfüllbare Gleichung $x = x + 1$), sie kann genau eine Zahl enthalten, d.h. die Gleichung hat genau eine Lösung (das ist eine günstige Situation und gerade bei sinnvollen ökonomischen Fragestellungen erwartet, die eine eindeutige Antwort haben sollten), sie kann aus zwei, drei oder einer beliebigen endlichen Anzahl von Lösungen bestehen, sie kann auch unendlich viele Lösungen enthalten, und überhaupt kann jede Teilmenge der Grundmenge im Prinzip als Lösungsmenge auftreten, auch die ganze Grundmenge selbst (wie z.B. bei der allgemeingültigen Gleichung $x = x$).

Wie findet man nun die Lösung — oder die Lösungen — einer Gleichung? Ziel ist die **Auflösung nach der Unbekannten**, d.h. die Transformation der gegebenen Gleichung durch eine Reihe von Umformungen, welche die Lösungsmenge nicht verändern oder jedenfalls nicht verkleinern, in eine Gleichung der “nach x aufgelösten Form” $x = K$ wobei auf der rechten Seite eine Konstante K steht, also einer Zahl oder einem Term, der von den Parametern der Gleichung abhängt, nicht aber von der Unbekannten. Die eindeutige Lösung der Gleichung $x = K$ ist dann natürlich K , und einzig diese Zahl kommt als Lösung der ursprünglichen Gleichung in Frage; durch Einsetzen des Wertes K für x in der ursprünglichen Gleichung prüft man dann leicht nach, ob K tatsächlich Lösung ist. Leider ist die Auflösung der Gleichung nach der Unbekannten nur in sehr einfachen — aber gerade in der Wirtschaftsmathematik häufig auftretenden — Fällen möglich; in den anderen Fällen ist man auf mathematische Verfahren zur näherungsweise Berechnung der Lösungen angewiesen. Auch enden die Umformungen der ursprünglichen Gleichung nicht immer mit einer Gleichung der aufgelösten Form, sondern man kann auf eine nicht erfüllbare Gleichung stoßen (wie $x + 1 = x$), in welchem Fall die ursprüngliche Gleichung gar keine Lösung hat, oder auf eine allgemeingültige Gleichung (wie $x = x$), in welchem Fall evtl. auch die ursprüngliche Gleichung von jeder Zahl gelöst wird. Manchmal sind auch Fallunterscheidungen notwendig, die zu $N \geq 2$ verschiedenen nach x aufgelösten Gleichungen führen; dann hat auch die ursprüngliche Gleichung unter Umständen N verschiedene Lösungen.

Eine **zulässige Umformung** einer Gleichung besteht in der Anwendung derselben Rechenoperation auf beide Seiten der Gleichung. Dabei gehen offenbar keine Lösungen verloren, d.h. die umgeformte Gleichung hat mindestens dieselben Lösungen wie die ursprüngliche — unter Umständen aber mehr (z.B. wenn man beide Seiten der ursprünglichen Gleichung mit Null multipliziert). Kommt man nach einer Serie von Umformungen auf eine Gleichung, deren Lösungen man alle bestimmen kann, so muss man durch sog. “Rückwärtseinsetzen” in die ursprüngliche Gleichung also noch überprüfen, welche davon auch die ursprüngliche Gleichung lösen. Damit hat man dann alle Lösungen der ursprünglichen Gleichung gefunden. Das Rückwärtseinsetzen kann man sich sparen (obwohl es als Probe immer sinnvoll ist), wenn man nur **Äquivalenzumformungen** vorgenommen hat, d.h. zulässige Umformungen, die durch eine weitere zulässige Umformung wieder rückgängig gemacht werden können. Die durch Äquivalenzumformungen entstandene Gleichung hat nämlich dieselbe Lösungsmenge wie die ursprüngliche. Hier eine (unvollständige) Liste der *Rechenoperationen, die beim Umformen von Gleichungen vorgenommen werden können*:

- Addition oder Subtraktion derselben Zahl oder desselben Terms auf beiden Seiten (Äquivalenzumformung);
- Multiplikation mit derselben Zahl oder mit demselben Term (Äquivalenzumformung, wenn die Zahl bzw. der Term $\neq 0$ ist; aber Nullstellen des Terms sind Lösungen der neuen Gleichung, nicht unbedingt der alten);
- Division mit derselben Zahl oder mit demselben Term $\neq 0$ (Äquivalenzumformung, wo der Term nicht Null ist; Nullstellen des Terms sind nicht im Definitionsbereich der neuen Gleichung, können aber Lösungen der ursprünglichen sein);
- Bildung des Kehrwertes von beiden Seiten (Äquivalenzumformung, wo beide Seiten $\neq 0$ sind; Lösungen der ursprünglichen Gleichung sind aber auch die Werte von x , für die beide Seiten $= 0$ sind);
- Quadrieren beider Seiten oder Erhebung zur Potenz mit demselben Exponenten $n \in \mathbb{N}$ (für gerade n keine Äquivalenzumformung, z.B. hat $x = 1$ nur eine, $x^2 = 1^2$ aber zwei Lösungen $+1, -1$).

Im Vorgriff auf Abschnitt 1.3 (Wurzeln und Logarithmen) erwähnen wir noch folgende Operationen zur zulässigen Umformung von Gleichungen:

- Ziehen der nichtnegativen Wurzel auf beiden Seiten (wenn die Seiten nichtnegative Werte haben, sonst erst beide Seiten mit -1 multiplizieren, das erfordert also eine Fallunterscheidung; außerdem darf man natürlich nicht etwa auf einer Seite die positive, auf der anderen die negative Quadratwurzel ziehen — es muss ja dieselbe Rechenoperation ausgeführt werden!);
- Logarithmieren beider Seiten (wenn die Seiten positiv sind, sonst erst mit -1 multiplizieren und Fallunterscheidung; die ursprüngliche Gleichung kann auch noch Lösungen haben, für die beide Seiten $= 0$ sind).

Zu der Frage, welche Umformungen bei einer gegebenen Gleichung man denn vornehmen müsse, um sie aufzulösen, gibt es kein allgemeines Rezept als Antwort. Ziel der Umformungen ist in jedem Fall, dass die Unbekannte rechts nicht mehr vorkommt und links nur in einem möglichst einfachen Term, wie eben “ x ” in der aufgelösten Form. Es ist eine Sache der Erfahrung und des Geschicks bei Termumformungen, einen Weg zu finden, der zum Ziel führt. Und einen solchen Weg gibt es ja auch nicht immer; denn schon recht einfach aussehende Gleichungen sind, wie gesagt, durch Umformungen nicht auflösbar.

BEISPIELE (zum Auflösen von Gleichungen):

1) Ein sehr einfacher Gleichungstyp für eine Unbekannte x ist die

lineare Gleichung: $ax = b$,

wobei die sog. *Koeffizienten* a, b konkrete Zahlen sind (etwa $2x = 3$ oder $\frac{1}{3}x = \sqrt{2}$) oder reelle Parameter. Der *Normalfall* ist $a \neq 0$; dann dividiert man beide Seiten der Gleichung durch a und erhält die äquivalente aufgelöste Form $x = \frac{b}{a}$. Also ist

$$x = \frac{b}{a} \quad \text{die eindeutige Lösung zu } ax = b, \quad \text{wenn } a \neq 0.$$

Im *Ausnahmefall* $a = 0$ sind dagegen alle reellen Zahlen Lösungen der linearen Gleichung, nämlich wenn auch $b = 0$ ist, oder es gibt überhaupt keine Lösung, nämlich wenn $b \neq 0$.

Lineare Gleichungen resultieren oft aus **Proportionalitäten** (Dreisatz), d.h. aus einem (z.B. ökonomischen) Gesetz, demzufolge eine Größe b zu einer anderen Größe a in demselben Verhältnis steht wie die Unbekannte x zu einer dritten Größe c , also $b : a = x : c$ oder (nach Multiplikation mit $a \neq 0$ und $c \neq 0$) $ax = bc$, $x = bc/a$.

Eine ganze Reihe ökonomischer Fragestellungen führt auf eine lineare Gleichung: Die Aufzinsungsformel und die Rentenformel bei einfacher Verzinsung (siehe Kap. 2)

$$K_t = \left(1 + \frac{p \cdot t}{100}\right) \cdot K_0, \quad K_t = K_0 - mR + \frac{p \cdot t}{100} \left(K_0 - \frac{m+1}{2} R\right)$$

sind z.B. lineare Gleichungen für jede der darin vorkommenden Variablen K_0 , K_t , p , t , R , m , wenn man die anderen Variablen jeweils als Parameter auffasst. (Bei der zweiten Gleichung muss man rechts erst die Terme mit dem Faktor K_0 und die Terme mit dem Faktor R jeweils zusammenfassen, um das zu sehen.) Die Gleichungen sind also nach jeder der vorkommenden Variablen auflösbar. Der Ausnahmefall, dass der Koeffizient vor der Unbekannten Null ist, kommt zwar formal auch vor, hat aber keine ökonomische Bedeutung. Z.B. hat die erste Gleichung, wenn $K_0 = 0$ ist und p oder t die Unbekannte, entweder keine Lösung (bei $K_t \neq 0$) oder alle Zahlen als Lösungen ($K_t = 0$) — aus einem Anfangskapital Null kann man eben außer dem Endkapital Null nichts machen, und dabei sind Zinsfuß und Laufzeit ganz egal!

2) Zu den Gleichungstypen, die sich auf lineare zurückführen lassen, gehört z.B. die

$$\text{gebrochen-lineare Gleichung: } \frac{ax + b}{cx + d} = r.$$

Dabei nehmen wir $c \neq 0$ an, sonst hat man ja einfach eine lineare Gleichung $\frac{a}{d}x + \frac{b}{d} = r$, also $\frac{a}{d}x = r - \frac{b}{d}$ (wenn auch $d = 0$ ist, so ist die Gleichung sinnlos, da die linke Seite für kein x definiert ist). Multiplikation beider Seiten mit dem Term $cx + d$ (der für Lösungen $\neq 0$ ist) gibt $ax + b = rcx + rd$, oder $(a - rc)x = rd - b$, und dies ist eine äquivalente lineare Gleichung, die im Normalfall $(a - rc) \neq 0$ genau eine Lösung $x = \frac{rd - b}{a - rc}$ hat, im Ausnahmefall $a - rc = 0$ aber alle Zahlen als Lösung oder keine. Was die ursprüngliche gebrochen-lineare Gleichung angeht, so muss man noch überprüfen, ob für die gefundenen Lösungen x der Nenner $cx + d \neq 0$ ist; dann und nur dann ist x auch Lösung der gebrochen-linearen Gleichung. Diese hat also genau eine Lösung, wenn $a - rc \neq 0$ und $ad - bc \neq 0$ ist, sie hat alle Zahlen $x \neq -\frac{d}{c}$ als Lösung, wenn $a - rc = 0 = rd - b$ ist, und in allen anderen Fällen hat sie überhaupt keine Lösung. Hier einige konkrete Beispiele:

$$\frac{x+1}{x-1} = 2 \iff x+1 = 2(x-1) \text{ und } x \neq 1 \iff x = 3,$$

$$\frac{x+1}{x-1} = 1 \iff x+1 = x-1 \text{ und } x \neq 1, \text{ hat keine Lösung,}$$

$$\frac{2x-2}{x-1} = 2 \iff 2x-2 = 2(x-1) \text{ und } x \neq 1, \text{ hat alle } x \neq 1 \text{ als Lösungen.}$$

Bei der letzten Gleichung hätte man das Ergebnis natürlich sofort sehen können: Wenn der Zähler das Doppelte des Nenners ist, so hat der Bruch immer den Wert 2 (außer der Nenner ist Null und der Bruchwert nicht definiert). Ebenso bei der mittleren Gleichung: Wenn der Zähler um 2 größer ist als der Nenner, so kann der Bruchwert nie 1 sein.

3) Kompliziertere Gleichungen, in denen gebrochen-lineare Terme auftreten, wie etwa

$$\frac{ax + b}{cx + d} = \frac{kx + l}{mx + n} + r,$$

können unter Umständen in lineare Gleichungen umgeformt werden. Lösungen können natürlich höchstens solche Zahlen x sein, für die beide auftretenden Nenner $\neq 0$ sind. Multipliziert man dann beide Seiten der Gleichung mit dem Produkt der Nenner, so entsteht $(ax+b)(mx+n) = (kx+l)(cx+d) + r(cx+d)$. Hier treten beim Ausmultiplizieren zwar Quadrate von x auf, aber wenn man Glück hat, so ist der Koeffizient von x^2 auf beiden Seiten derselbe, $am = kc$, und dann kann man die Quadrate durch Subtrahieren von $amx^2 = kcx^2$ auf beiden Seiten entfernen und hat eine lineare Gleichung für x übrig, deren Lösung(en) auch die Lösung(en) der ursprünglichen Gleichung ist (sind), soweit dafür nicht ein Nenner Null wird. Hier ein konkretes Beispiel:

$$\frac{x+1}{x-1} = \frac{2x+1}{2x-1} + r \iff (2x-1)(x+1) = (x-1)(2x+1) + (x-1)r \text{ und } 1 \neq x \neq \frac{1}{2}$$

$$\iff (r-2)x = r \text{ und } 1 \neq x \neq \frac{1}{2} \iff x = \frac{r}{r-2}, \text{ wenn } r \neq \pm 2, \text{ sonst keine Lösung.}$$

Wenn sich allerdings die quadratischen Terme nicht herausheben, so bleibt nach Multiplikation mit den Nennern eine quadratische Gleichung über, keine lineare. Diese kann mit Quadratwurzeln gelöst werden (s. 1.3).

4) Bei einer Doppelbruch-Gleichung wie

$$\frac{1}{1 + \frac{1}{1+x}} = r$$

liegt es nahe, auf beiden Seiten den Kehrwert zu bilden, um die Gleichung zu vereinfachen. Wenn $r \neq 0$ ist (sonst hat die Gleichung keine Lösung, weil ja die linke Seite nie Null werden kann), so ergibt sich die äquivalente Gleichung $1 + \frac{1}{1+x} = \frac{1}{r}$. Nach Subtraktion von 1 auf beiden Seiten ist dies eine gebrochen-lineare Gleichung wie in 2). Man kann aber, wenn $r \neq 1$ ist (sonst hat auch diese Gleichung keine Lösung), durch nochmaligen Übergang zu Kehrwerten gleich die äquivalente Form $1+x = (\frac{1}{r}-1)^{-1}$ herstellen. Ergebnis: Für $r = 0$ oder $r = 1$ hat die Gleichung keine Lösung (das hätte man von vorneherein gleich sehen können), ansonsten ist $x = (\frac{1}{r} - 1)^{-1} - 1 = \frac{2r-1}{1-r}$ die eindeutige Lösung. ■

Unter einer **algebraischen Gleichung** (oder *Polynomgleichung*) vom Grad $n \in \mathbb{N}$ (man sagt auch *Ordnung* statt Grad) versteht man eine Gleichung der Form

$$a_n x^n + a_{n-1} x^{n-1} + \dots + a_2 x^2 + a_1 x + a_0 = 0$$

mit gegebenen Zahlen $a_n \neq 0$ und $a_0, a_1, a_2, \dots, a_{n-1}$ als sog. *Koeffizienten* der Gleichung. Nach Division mit dem führenden Koeffizienten a_n kann man ohne Einschränkung annehmen, dass dieser gleich 1 ist; die Gleichung beginnt dann $x^n + \dots = 0$. Obwohl es sich hier noch um einen relativ einfachen Gleichungstyp handelt, kann man eine solche Gleichung im nichtlinearen Fall $n \geq 2$ nicht mehr mit den Grundrechenarten auflösen. Für *quadratische Gleichungen* ($n = 2$) gibt es noch Lösungsformeln, welche die höhere Rechenoperation des Wurzelziehens verwenden (s. 1.3), auch bei Grad $n = 3$ und $n = 4$ gibt es noch Lösungsformeln mit "verschachtelten Wurzelausdrücken", die aber weniger nützlich sind als die bekannte Lösungsformel für quadratische Gleichungen. Für Ordnungen $n \geq 5$ gibt es aber — das ist mathematisch bewiesen — keine Lösungsformeln mehr, die nur Grundrechenarten und Wurzelziehen beinhalten. Man muss sich also damit abfinden, dass algebraische Gleichungen im Allgemeinen nicht auflösbar sind, auch solche nicht, die in ökonomischem Zusammenhang auftauchen.

Z.B. ist die Effektivzinsgleichung für Annuitätsdarlehen (siehe Kap. 2; man sucht hier den unbekanntem effektiven Aufzinsungsfaktor $q_* > 1$ und α, β sind Parameter)

$$q_*^n - \alpha \frac{q_*^n - 1}{q_* - 1} = \beta$$

durch Multiplikation mit $q_* - 1 > 0$ äquivalent zu der algebraischen Gleichung $(n+1)$ -ter Ordnung

$$q_*^{n+1} - (\alpha+1)q_*^n - \beta q_* + \alpha + \beta = 0$$

für die Unbekannte q_* . Obwohl hier die Potenzen q_*^2, \dots, q_*^{n-1} der Unbekannten nicht auftreten, ist diese Gleichung für $n \geq 4$ nicht mehr durch eine Lösungsformel nach q_* auflösbar.

Es gibt eine mathematische Theorie der Polynomgleichungen, aus der wir zur Information die folgenden Fakten mitteilen:

- Eine algebraische Gleichung hat höchstens so viele reellen Lösungen wie ihr Grad n ;
- bei ungerader Ordnung n gibt es stets mindestens eine reelle Lösung;
- bei gerader Ordnung n aber existiert unter Umständen keine reelle Lösung.

So hat zum Beispiel $x^n + 1 = 0$ für gerade $n = 2m \in \mathbb{N}$ keine reelle Lösung, weil $x^{2m} = (x^m)^2 \geq 0$ ist, die linke Seite der Gleichung also immer ≥ 1 . Dass bei ungerader Ordnung $n = 2m+1$ immer mindestens eine reelle Lösung existiert, liegt daran, dass die linke Seite $x^{2m+1} + a_{2m}x^{2m} + \dots$ dann für große Werte des Betrags $|x|$ dasselbe Vorzeichen wie x hat, also positiv ist für große positive Werte von x und negativ für absolut große negative Werte (die Potenzen niedriger Ordnung von x fallen im Vergleich zur führenden Potenz x^{2m+1} nicht ins Gewicht). Nach dem "Zwischenwertsatz" gibt es dann auch einen Wert von x (oder mehrere solche Werte), für den die linke Seite der Gleichung Null wird. Eine Gleichung n -ter Ordnung, die n gegebene Zahlen x_1, x_2, \dots, x_n als Lösungen hat, erhält man durch Ausmultiplizieren des Produkts $(x - x_1) \cdot (x - x_2) \cdot \dots \cdot (x - x_n)$ und Zusammenfassen der Potenzen von x mit gleichen Exponenten. Es entsteht dann ein polynomialer Term $x^n + a_{n-1}x^{n-1} + \dots + a_1x + a_0$, der genau dann $= 0$ wird, wenn x eine der Zahlen x_1, x_2, \dots, x_n ist; denn ein Produkt hat genau dann den Wert Null, wenn mindestens einer seiner Faktoren, hier also ein Faktor $x - x_i$, den Wert Null hat.

Die letzte Bemerkung ist auch die Grundlage für ein *Reduktionsverfahren*, das man anwenden kann, wenn man eine Lösung ξ der algebraischen Gleichung $a_nx^n + \dots + a_1x + a_0 = 0$ geraten hat (indem man einige Zahlen ausprobiert hat oder aus dem Kontext der Aufgabenstellung schon eine Lösung kennt). Man kann dann die linke Seite der Gleichung in der Form $(x - \xi) \cdot (b_{n-1}x^{n-1} + \dots + b_1x + b_0)$ faktorisieren. Diese Prozedur nennt man **Abspalten des Linearfaktors** $x - \xi$ oder **Polynomdivision durch den Linearfaktor**. Die Koeffizienten b_{n-1}, b_{n-2}, \dots , bestimmt man dabei in dieser Reihenfolge so, dass nach Ausmultiplizieren des Produkts $(x - \xi) \cdot (b_{n-1}x^{n-1} + \dots + b_1x + b_0)$ und Sammeln der Terme mit gleichen Potenzen von x gerade die Koeffizienten a_n, a_{n-1}, \dots vor x^n, x^{n-1}, \dots entstehen. Da die linke Seite der Gleichung nun in der Produktform $(x - \xi) \cdot (b_{n-1}x^{n-1} + \dots + b_1x + b_0)$ geschrieben ist, sieht man, dass sie den Wert Null genau dann annimmt, wenn $x = \xi$ ist oder wenn x eine Lösung der Polynomgleichung $b_{n-1}x^{n-1} + \dots + b_1x + b_0 = 0$ ist. Die Reduktion besteht darin, dass der Grad der letzten Gleichung um 1 niedriger ist als bei der ursprünglichen Gleichung. Wenn die neue Gleichung keine reelle Lösung besitzt, so war $x = \xi$ die einzige Lösung der Ausgangsgleichung. Wenn man andererseits zu dieser neuen Gleichung wieder eine Lösung raten kann, so lässt sich das Verfahren erneut anwenden, und mit Glück kommt man so fortfahrend vielleicht bis zu einer Gleichung vom Grad 1, also zu einer linearen Gleichung, und hat dann alle Lösungen der ursprünglichen algebraischen Gleichung gefunden. Hier zwei konkrete

BEISPIELE (zur Lösung algebraischer Gleichungen mit dem Reduktionsverfahren):

1)
$$x^2 - x - 2 = 0.$$

Bei dieser quadratischen Gleichung erkennt man die Lösung $x = 2$ und findet die Faktorisierung $x^2 - x - 2 = (x - 2) \cdot (x + 1)$. (Der erste Koeffizient im letzten Linearfaktor muss 1 sein, damit beim Ausmultiplizieren $1 \cdot x^2$ entsteht, der zweite muss $+1$ sein, damit beim Ausmultiplizieren der konstante Term -2 entsteht). Also sind $x = 2$ und $x = -1$ die Lösungen der quadratischen Gleichung.

2)
$$x^3 - x^2 + x - 1 = 0.$$

Bei dieser kubischen (d.h. Grad 3) Gleichung erkennt man sofort die Lösung $x = 1$ und findet die Faktorisierung der linken Seite $(x - 1) \cdot (x^2 + 1)$ (die man hier z.B. sieht, indem man $x^3 - x^2 = (x - 1) \cdot x^2$ schreibt; man kann für den zweiten Faktor der Zerlegung auch $ax^2 + bx + c$ ansetzen und dann a, b, c der Reihe nach so bestimmen, dass beim Ausmultiplizieren die linke Seite der Gleichung entsteht). Da $x^2 + 1$ für reelle x nie Null wird, ist folglich $x = 1$ auch die einzige Lösung der kubischen Gleichung. ■

Zum Schluss dieses Abschnitts geben wir zur Warnung noch drei Beispiele von Fehlern, die beim Lösen von Gleichungen nicht selten gemacht werden:

$$\frac{2x+6}{x+3} = 1.$$

Bei dieser Gleichung multipliziert man mit $x+3$ und erhält $2x+6 = x+3$, also $x = -3$ als eindeutige Lösung??? — Aber für $x = -3$ ist der Nenner in der ursprünglichen Gleichung Null, also handelt es sich hier nicht um eine Lösung, und die Gleichung ist überhaupt nicht lösbar! (Der Fehler war, das Rückwärtseinsetzen zu unterlassen; Multiplikation mit einem Term, der Null werden kann, ist keine Äquivalenzumformung, sondern kann zu einer Gleichung mit *mehr* Lösungen führen als die Ausgangsgleichung.)

Bei der Gleichung

$$x^3 + 1 = x^2 + 2x + 1,$$

wird man als Kenner der binomischen und trinomischen Formeln sofort sehen, dass die rechte Seite $(x+1)^2$ ist und der Faktor $x+1$ auch auf der linken Seite $(x+1)(x^2-x+1)$ ausgeklammert werden kann. Nach Kürzen dieses Faktors erhält man dann die Gleichung $x^2 - x + 1 = x + 1$, also $x^2 - 2x = 0$ bzw. $x \cdot (x-2) = 0$, d.h. $x = 0$ und $x = 2$ sind die Lösungen??? — Aber $x = -1$ ist auch eine Lösung der ursprünglichen Gleichung, die hier übersehen wurde! (Der Fehler war hier, zu vergessen, dass Division mit einem Term, der Null werden kann, keine Äquivalenzumformung ist, sondern zu einer neuen Gleichung führen kann, die *weniger* Lösungen hat als die ursprüngliche.)

Bei der dritten Gleichung

$$(x+1)^2 = (x-1)^2$$

liegt es nahe, aus beiden Seiten die Quadratwurzel zu ziehen. Man erhält dann die lineare Gleichung $x+1 = x-1$, die keine Lösung hat, also besitzt auch die ursprüngliche Gleichung keine Lösung??? — Aber $x = 0$ ist offenbar doch eine Lösung! (Der Fehler war hier, dass man auf beiden Seiten nicht *dieselbe* Rechenoperation angewendet hat; denn für $-1 < x < 1$ ist $x+1$ die positive Quadratwurzel aus $(x+1)^2$, aber $x-1$ die negative Quadratwurzel aus $(x-1)^2$. Dieser Fehler bei der Rechnung mit Quadratwurzeln wird drastisch durch die unkorrekte Gleichungskette $-1 = \sqrt{(-1)^2} = \sqrt{1^2} = 1$ demonstriert.) Wurzelziehen auf beiden Seiten einer Gleichung ist nur dann zulässig, wenn man auf beiden Seiten die positive Wurzel zieht (oder auf beiden Seiten die negative Wurzel, sofern beide Seiten nicht Null sind). Für den Bereich $-1 < x < 1$ der Unbekannten x hätte man die radizierte Gleichung in der äquivalenten Form $x+1 = -(x-1)$ schreiben müssen und dann die Lösung $x = 0$ auch gefunden. (Für die Bereiche $x \geq 1$ und $x \leq -1$ ist dagegen $x+1 = x-1$ äquivalent, daher gibt es in diesen Bereichen keine weitere Lösung der ursprünglichen Gleichung und $x = 0$ ist die einzige.) Korrekt wäre es auch gewesen, für alle Werte von x die nichtnegativen Wurzeln in Form von Beträgen zu schreiben und dann die Gleichung $|x+1| = |x-1|$ durch Fallunterscheidungen zu lösen (vgl. 1.4).

Mathematik für Wirtschaftswissenschaftler

(K. Steffen, Heinrich-Heine-Universität Düsseldorf, WS 2006/07)

1.3 Wurzeln, Potenzen mit beliebigen Exponenten, Logarithmen

DISKUSSION (Wurzelrechnung):

1) Es ist eine fundamentale Tatsache, dass jede *Potenzgleichung* $x^n = a$ mit natürlichem Exponenten $n \geq 2$ und mit positiver rechter Seite $a \in \mathbb{R}_{>0}$ genau eine positive reelle Lösung $x > 0$ hat. Man kann dies so verstehen: Für einen kleinen positiven Wert von x ist x^n so klein wie man will, also auch kleiner als a ; für einen großen Wert von x ist aber x^n so groß wie man will, also größer als a . Daher muss es irgendwo dazwischen einen Wert von x geben, für den x^n genau so groß ist wie a , eben eine Lösung der Potenzgleichung. (Dass dieses "Zwischenwertargument" gültig ist, liegt an einer fundamentalen Vollständigkeitseigenschaft des Bereichs der reellen Zahlen \mathbb{R} ; es gibt darin eben "alle Zahlen, die man braucht". Im Bereich der rationalen Zahlen hat dagegen schon eine so einfache Gleichung wie $x^2 = 2$ keine Lösung.) Weil nun x^n bei Vergrößerung von x größer wird und bei Verkleinerung von x kleiner (siehe 1.4), kann es auch keine weiteren positiven Lösungen geben. Für $x^n = 0$ ist natürlich $x = 0$ die eindeutige Lösung.

- Die eindeutige positive reelle Lösung der Potenzgleichung $x^n = a$ mit Exponent $2 \leq n \in \mathbb{N}$ und rechter Seite $a > 0$ heißt die **n -te Wurzel** aus a und wird $\sqrt[n]{a}$ oder $a^{1/n}$ notiert; außerdem setzt man noch $\sqrt[n]{0} := 0$. Im Fall $n = 2$ spricht man auch von der **Quadratwurzel** aus a und schreibt dafür einfach $\sqrt{a} = a^{1/2}$.

Demnach gilt $(\sqrt[n]{a})^n = a$ für $a \geq 0$ und $\sqrt[n]{a}$ ist die einzige nichtnegative Zahl mit n -ter Potenz a . Für $a = b^n$ mit $b \geq 0$ hat auch b diese Eigenschaften (nichtnegativ mit n -ter Potenz a), daher gilt auch $\sqrt[n]{b^n} = b$, wenn $b \geq 0$ ist. Der Name "Wurzeln" wird manchmal auch für Lösungen von allgemeineren Gleichungen als der obigen Potenzgleichung gebraucht, z.B. für Lösungen von algebraischen Gleichungen.

2) Um mit Wurzeln richtig umzugehen, ist es ganz wichtig, folgendes zu beachten:

- Wurzeln $\sqrt[n]{a}$ werden niemals aus negativen Zahlen a gezogen und haben niemals negative Werte.

Die Potenzgleichung $x^n = a$ kann zwar auch negative Lösungen haben und auch für negative rechte Seiten lösbar sein, aber in diesem Fall bezeichnen wir die Lösungen nicht mit dem Symbol $\sqrt[n]{a}$. Das ist sinnvoll, weil sonst die Wurzelrechengesetze wie $\sqrt[n]{ab} = \sqrt[n]{a}\sqrt[n]{b}$ falsch werden. (Das Beispiel $-1 = \sqrt{(-1)^2} = \sqrt{1^2} = 1$ haben wir ja schon gesehen; die Rechnung ist ungültig, weil im ersten Schritt eine *negative* Quadratwurzel gezogen wird!) Außerdem machen negative Wurzeln gerade in der Ökonomie, wo die relevanten Größen oft positiv sind, meist gar keinen Sinn.

Im Übrigen kann man auch die negativen Lösungen der Potenzgleichung durch positive Wurzeln aus positiven Zahlen ausdrücken. Für gerade $n = 2m$ z.B. hat $x^n = a$ keine Lösung, wenn $a < 0$, genau die Lösung $x = 0$, wenn $a = 0$, und genau die zwei Lösungen

$\sqrt[n]{a} > 0$ und $-\sqrt[n]{a} < 0$, wenn $a > 0$ ist. (Das erkennt man sofort aus $x^n - a = (x^m)^2 - (\sqrt{a})^2 = (x^m - \sqrt{a})(x^m + \sqrt{a})$.) Für ungerade $n = 2m + 1$ hat zwar $x^n = a$ genau eine Lösung für jede rechte Seite $a \in \mathbb{R}$, aber diese Lösung ist negativ für $a < 0$ (weil Potenzen nichtnegativer Zahlen wieder nichtnegativ sind) und sie kann mit der positiven n -ten Wurzel aus der (positiven) Gegenzahl $|a| = -a$ in der Form $-\sqrt[n]{|a|} < 0$ ausgedrückt werden (wie man aus $(-x)^n = (-1)^n x^n = -a$ für $x^n = a$ und ungerade n abliest.) Obwohl beispielsweise $x^3 = -8$ genau eine reelle Lösung hat, nämlich $x = -2$, bezeichnen wir diese Lösung also *nicht* mit $\sqrt[3]{-8}$, sondern mit $-\sqrt[3]{8}$. (*Warnung:* Das handhaben nicht alle Autoren so und schreiben durchaus $\sqrt[3]{-8}$. Das Wurzelrechengesetz $\sqrt[2]{\sqrt[3]{a}} = \sqrt[6]{a}$ gilt dann aber für $a = -8$ nicht mehr, weil -8 überhaupt keine 6-te Wurzel hat. Daher ist es besser, unserer Konvention zu folgen und $\sqrt[n]{a}$ immer nur für die *nichtnegativen* Wurzeln aus *nichtnegativen* Zahlen a zu schreiben.)

3) Rechenregeln für Wurzeln lassen sich aus ihrer Definition in 1) und aus den Potenzgesetzen herleiten für nichtnegative(!) Zahlen a und b :

$$\begin{aligned}\sqrt[n]{0} &= 0, & \sqrt[n]{1} &= 1, \\ (\sqrt[n]{a})^n &= a = \sqrt[n]{a^n}, \\ (\sqrt[n]{a})^m &= \sqrt[n]{a^m}, \\ \sqrt[m]{\sqrt[n]{a}} &= \sqrt[m \cdot n]{a} = \sqrt[n]{\sqrt[m]{a}}, \\ \sqrt[n]{a \cdot b} &= \sqrt[n]{a} \cdot \sqrt[n]{b}, \\ \sqrt[n]{a/b} &= \sqrt[n]{a} / \sqrt[n]{b},\end{aligned}$$

wobei zuletzt natürlich $b > 0$ sein muss. Insbesondere ist die n -te Wurzel des Kehrwertes gleich dem Kehrwert aus der n -ten Wurzel, $\sqrt[n]{1/b} = 1/\sqrt[n]{b}$, und dafür wird auch $b^{-1/n}$ geschrieben. Die ersten beiden Regeln haben wir schon gesehen, die dritte gilt, weil beide Seiten die n -te Potenz a^m haben, die vierte, weil alle drei Terme die $(m \cdot n)$ -te Potenz a besitzen, die fünfte und sechste schließlich, weil beide Seiten n -te Potenz ab bzw. a/b haben. Die Wurzelrechengesetze sind Spezialfälle der Rechengesetze für Potenzen mit gebrochenen Exponenten, man braucht sie sich daher nicht zu merken, wenn man alle Wurzelausdrücke systematisch als Potenzen umschreibt. Die ganze Wurzelrechnung ist im Grunde Teil der allgemeineren Potenzrechnung, die wir anschließend behandeln. ■

BEISPIELE: (zur Wurzelrechnung:)

1) Die allgemeine **quadratische Gleichung**

$$ax^2 + bx + c = 0 \quad (\text{mit } a, b, c \in \mathbb{R} \text{ und } a \neq 0)$$

löst man durch sog. *quadratische Ergänzung* und Wurzelziehen so:

$$0 = ax^2 + bx + c = a\left(x + \frac{b}{2a}\right)^2 + c - \frac{b^2}{4a} \iff \left(x + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a^2}.$$

Fall 1 $b^2 - 4ac < 0$: Dann gibt es keine reelle Lösung (weil die linke Seite der letzten Gleichung immer nichtnegativ ist);

Fall 2 $b^2 - 4ac = 0$: Dann gibt es genau eine Lösung $x = -\frac{b}{2a}$;

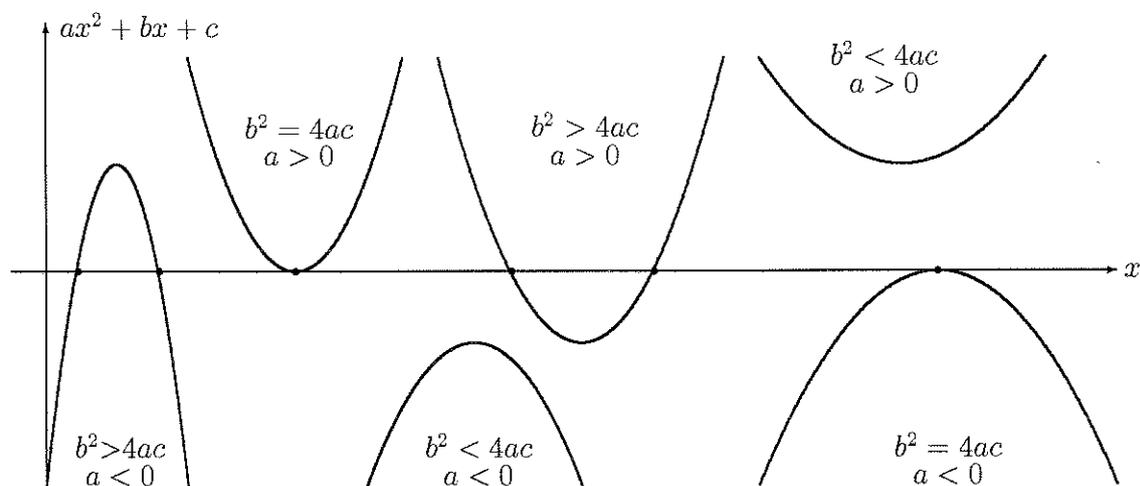
Fall 3 $b^2 - 4ac > 0$: Dann gibt es genau zwei Lösungen, nämlich die beiden Zahlen x , für die $x + \frac{b}{2a}$ die positive bzw. die negative Quadratwurzel aus $\frac{b^2 - 4ac}{4a^2}$ ist, also

$$x_1, x_2 = -\frac{b}{2a} \pm \frac{1}{2a} \sqrt{b^2 - 4ac}.$$

Das ist die bekannte Lösungsformel für quadratische Gleichungen; sie gilt auch im zweiten Fall, liefert dann jedoch nur eine Lösung $x_1 = x_2 = -\frac{b}{2a}$. Schreibt man die quadratische Gleichung mit Division durch den führenden Koeffizienten a um, so erhält man die sog. "p-q-Formel"

$$x^2 + px + q = 0 \iff x_1, x_2 = -\frac{1}{2}p \pm \sqrt{\frac{1}{4}p^2 - q}, \text{ wenn } p^2 \geq 4q.$$

Das Erste, was man bei einer vorgelegten quadratischen Gleichung zu tun hat, ist, das Vorzeichen der Diskriminante $b^2 - 4ac$ bzw. $p^2 - 4q$ festzustellen; denn dieses Vorzeichen entscheidet darüber, ob die Gleichung zwei, eine oder keine Lösung besitzt. Geometrisch kann man die verschiedenen Fälle darstellen, indem man den Graphen der Funktion von x zeichnet, die durch die linke Seite der Gleichung gegeben ist. Man erhält dann nach oben ($a > 0$) bzw. unten ($a < 0$) geöffnete Parabeln, welche die horizontale x -Achse nicht treffen (Fall 1), in einem Punkt berühren (Fall 2), oder in zwei Punkten schneiden, die den Nullstellen entsprechen (Fall 3).



2) Für algebraische Gleichungen vom Grad 3 oder 4 gibt es, wie schon gesagt, auch noch Lösungsformeln, in denen ineinander verschachtelte Wurzel­ausdrücke vorkommen; diese Formeln sind aber zu kompliziert, um von praktischem Nutzen zu sein. Für Gleichungen von noch größerem Grad $n \geq 5$ gibt es aber (beweisbar) auch keine solchen Lösungsformeln mehr. Man ist dann auf Verfahren der numerischen Mathematik zur näherungsweisen Berechnung der Lösungen (wenn es welche gibt!) angewiesen. Das schließt natürlich nicht aus, dass man in speziellen Fällen doch alle Lösungen einer algebraischen Gleichung von höherem Grad bestimmen kann. Bei einer kubischen Gleichung (Grad 3) z.B., bei der man eine Lösung schon kennt, führt das in 1.2 beschriebene Reduktionsverfahren auf eine quadratische Gleichung, aus der man dann die weiteren Lösungen berechnen kann (bzw. erkennt, dass es keine weiteren gibt). Ein explizit lösbarer Typ von Gleichungen vierten Grades, der gelegentlich vorkommt, ist die sog. **biquadratische Gleichung**

$$ax^4 + bx^2 + c = 0.$$

Dies kann man als quadratische Gleichung $ay^2 + by + c = 0$ für die Variable $y := x^2$ auffassen. Von deren nichtnegativen Lösungen y (Quadrate $y = x^2$ von reellen Zahlen sind ja nie negativ) nimmt man dann die Wurzeln \sqrt{y} und $-\sqrt{y}$ und erhält damit alle der maximal 4 Lösungen der biquadratischen Gleichung. (Wenn die quadratische Gleichung für y keine, oder nur negative Lösungen hat, so besitzt die biquadratische Gleichung keine Lösungen.)

3) Die Auflösung von **Gleichungen, die Wurzelterme enthalten** ist unter Umständen möglich, indem man die Wurzeln durch Quadrieren bzw. Potenzieren beseitigt. Beispiel:

$$\begin{aligned} x + \sqrt{2x+3} &= 0 \\ \Leftrightarrow \sqrt{2x+3} &= -x \\ \Rightarrow 2x+3 &= x^2, \quad \text{also } x^2 - 2x - 3 = 0 \\ \Leftrightarrow x &= 1 \pm \sqrt{1+3} = 1 \pm 2. \end{aligned}$$

Im zweiten Schritt haben wir quadriert; das ist aber keine Äquivalenzumformung, also müssen wir durch Einsetzen in die ursprünglich Gleichung noch nachprüfen, ob die gefundenen Werte Lösungen sind: Für $x = -1$ ist die Ausgangsgleichung erfüllt, für $x = 3$ aber nicht, also ist $x = -1$ ihre einzige Lösung. ($x = 3$ löst die Gleichung $x - \sqrt{2x+3}$, also die ursprüngliche Gleichung mit der negativen Wurzel aus $2x+3$; wir bezeichnen mit $\sqrt{\dots}$ aber immer nur nichtnegative Zahlen! Man hätte in der obigen Rechnung daher gleich mit notieren können, dass $-x \geq 0$ sein muss, und damit $x = 3$ ausgeschlossen.) In komplizierteren Fällen hilft manchmal mehrfaches Quadrieren:

$$\begin{aligned} \sqrt{x} + \sqrt{x+3} - \sqrt{x+8} &= 0 & \Leftrightarrow & \sqrt{x} + \sqrt{x+3} = \sqrt{x+8} \\ \Rightarrow (\sqrt{x} + \sqrt{x+3})^2 &= x+8 & \Leftrightarrow & x + 2\sqrt{x}\sqrt{x+3} + x+3 = x+8 \\ \Leftrightarrow 2\sqrt{x}\sqrt{x+3} &= 5-x & \Rightarrow & 4x(x+3) = (5-x)^2 \\ \Leftrightarrow 3x^2 + 22x - 25 &= 0 & \Leftrightarrow & x = -\frac{22}{6} \pm \frac{1}{6}\sqrt{22^2 + 300} = -\frac{22}{6} \pm \frac{28}{6}. \end{aligned}$$

Von den beiden gefundenen Werten für x ist nur $x = 1$ Lösung der ursprünglichen Gleichung, der andere Wert $x = -\frac{50}{6}$ nicht. (Man hätte von vorneherein bemerken können, dass $x \geq 0$ sein muss, da sonst \sqrt{x} nicht definiert ist.) Wenn vier Wurzelausdrücke auftreten, so kann man sie im Allgemeinen nicht mehr durch mehrfaches Quadrieren wegschaffen. In der Algebra aber wird bewiesen, dass sich eine "Wurzelgleichung" immer (auf kompliziertere Weise) in eine algebraische Gleichung umformen lässt.

4) Anwendungen der Lösungsformeln für quadratische Gleichungen und der Wurzelrechnung in der Wirtschaftswissenschaft behandeln wir in Kap. 2. Allgemeine n -te Wurzeln benötigt man z.B. für die Berechnung des *Effektivzinssatzes* vor für einen Vorgang, bei dem ein Anfangskapital K_0 durch Verzinsung zu evtl. wechselnden Zinsfüßen, wie z.B. beim Sparen mit wachsendem Zins, in einer gegebenen Anzahl n von Jahren auf einen Kapitalendwert K_n anwächst. Der Effektivzinssatz für diesen Vorgang ist derjenige Zinsfuß p_* , bei dem n -fache jährliche Aufzinsung dasselbe Ergebnis liefert. Für den zu p_* gehörenden Aufzinsungsfaktor $q_* = 1 + \frac{1}{100}p_*$ bedeutet das

$$K_n = q_*^n K_0 \quad \text{bzw.} \quad q_*^n = \frac{K_n}{K_0}.$$

Das ist eine Potenzgleichung n -ter Ordnung für die Unbekannte q_* , und die einzige positive Lösung (Aufzinsungsfaktoren sind ja stets positiv) ist

$$q_{\text{eff}} = \sqrt[n]{\frac{K_n}{K_0}} \quad \text{also} \quad p_{\text{eff}} = 100 \cdot \left(\sqrt[n]{\frac{K_n}{K_0}} - 1 \right) \quad (\text{Prozent}).$$

In komplizierteren Fällen führt die Effektivzinsberechnung aber auf eine algebraische Gleichung, die nicht mehr durch Wurzelziehen auflösbar ist. Das haben wir ja schon zur Effektivzinsgleichung für Annuitätendarlehen $q_*^n - \alpha \frac{q_*^n - 1}{q_* - 1} = \beta$ bemerkt. ■

Potenzen a^s mit beliebigen reellen Exponenten s treten in mathematischen Modellen für ökonomische Vorgänge häufig auf, insbesondere deswegen, weil Vorgänge uneingeschränkten Wachstums (oder Verfalls) durch *Exponentialfunktionen* $f(x) = a^x$ von einer reellen Variablen x beschrieben werden. Dabei will man die Funktionswerte selbstverständlich nicht nur für ganze Zahlen x berechnen können, sondern eben für alle reellen Zahlen. Wenn man eine sinnvolle und nützliche Potenzrechnung für Potenzen mit beliebigen reellen Exponenten erklären will, so müssen natürlich die üblichen Potenzrechengesetze gelten, die wir für ganze Exponenten in Abschnitt 1.1 wiederholt haben. Aus dieser Forderung ergibt sich aber schon zwangsläufig, wie man die allgemeinen Potenzen zu definieren hat. Betrachten wir zum Beispiel einen Stammbruch $s = 1/n$ als Exponent (mit $n \in \mathbb{N}$), so sollte für die zu erklärende Zahl $a^{1/n}$ gelten $(a^{1/n})^n = a^{(1/n) \cdot n} = a^1 = a$, also müssen wir festsetzen $a^{1/n} := \sqrt[n]{a}$ (wobei natürlich $\sqrt[n]{a} := a$ zu verstehen ist; die Notation $a^{1/n}$ für $\sqrt[n]{a}$ haben wir ja auch schon bei der Wurzelrechnung angegeben). Wegen der Schwierigkeiten beim Wurzelziehen aus negativen Zahlen sind wir an dieser Stelle übrigens gezwungen, uns auf das Potenzieren von *positiven Basen* $a > 0$ zu beschränken.

Im nächsten Schritt wollen wir nun $a^{m/n}$ mit einer rationalen Zahl m/n als Exponenten erklären (mit $m \in \mathbb{Z}$ und $n \in \mathbb{N}$). Dafür sollte dann $(a^{m/n})^n = a^{(m/n) \cdot n} = a^m$ gelten, also haben wir $a^{m/n} := \sqrt[n]{a^m} = (\sqrt[n]{a})^m$ festzusetzen. Hier gibt es ein kleines Problem, weil man dieselbe rationale Zahl r ja (durch Erweitern bzw. Kürzen) auf verschiedene Weise als Bruch $\frac{m}{n} = \frac{k}{l}$ schreiben kann (mit ganzen Zählern k, m und natürlichen Nennern l, n), und der Wert der Potenz mit diesem Exponenten darf natürlich nicht davon abhängen, wie man den Exponenten als Bruch geschrieben hat. Mit $m \cdot l = k \cdot n$ und den Rechenregeln für das Potenzieren mit ganzen Exponenten findet man aber $(\sqrt[n]{a^m})^{l \cdot n} = ((\sqrt[n]{a^m})^n)^l = (a^m)^l = a^{m \cdot l} = a^{k \cdot n} = (a^k)^n = ((\sqrt[l]{a^k})^l)^n = (\sqrt[l]{a^k})^{l \cdot n}$, d.h. sowohl $\sqrt[n]{a^m}$ als auch $\sqrt[l]{a^k}$ sind die $(l \cdot n)$ -te Wurzel aus $a^{k \cdot n}$ und damit gleich. Das Problem ist also zum Glück nicht vorhanden und unsere Definition von a^r für $r \in \mathbb{Q}$ unabhängig von der Darstellung von r als Bruch.

Im letzten Schritt müssen wir jetzt noch a^s für irrationale Zahlen s erklären. Da man jede reelle Zahl s beliebig gut zwischen rationale Zahlen r, t einschachteln kann, z.B. indem man die Dezimalbruchentwicklung von s erst sehr weit nach dem Dezimalpunkt abbricht (um r zu erhalten) und dann die letzte Dezimale um 1 erhöht (um t zu erhalten), werden wir das natürlich so machen wollen, dass die zu definierende Zahl a^s zwischen den schon definierten Potenzen a^r und a^t liegt, wenn $r < s < t$ ist. Es lässt sich zeigen — allerdings mit einigem mathematischem Aufwand —, dass durch diese Bedingung die Zahl a^s schon eindeutig festgelegt ist. Wir brauchen das nicht näher zu erklären, weil man für praktische Zwecke doch immer nur mit rationalen Exponenten rechnet (wie auch die allgemeine Potenzfunktion in einem Taschenrechner). Statt a^s genau zu kennen, genügt es eben einen guten Näherungswert a^r mit rationalem Exponenten r zu wissen, und der Unterschied zwischen a^s und a^r ist so klein, wie man will, wenn man nur die rationale Zahl r nahe genug bei der reellen Zahl s gewählt hat. Wir fassen die Definition der Potenzen mit allgemeinen Exponenten zusammen und besprechen die Rechenregeln dafür in folgender

DISKUSSION (Potenzrechnung mit beliebigen reellen Exponenten):

- 1) • Für positive Zahlen $a > 0$ und rationale Zahlen $r = m/n$ mit $m \in \mathbb{Z}$ und $n \in \mathbb{N}$ wird die Potenz a^r der Basis a mit dem gebrochenen Exponenten r erklärt durch

$$a^r = a^{\frac{m}{n}} := \sqrt[n]{a^m} = (\sqrt[n]{a})^m.$$

Für beliebige reelle Zahlen $s \in \mathbb{R}$ ist dann die Potenz a^s der Basis a mit dem reellen Exponenten s erklärt als die eindeutige reelle Zahl, die zwischen a^r und a^t liegt, wenn immer der Exponent s zwischen den rationalen Zahlen r und t ist.

Für einen irrationalen Exponenten wie etwa $\sqrt{2} = 1.4142\dots$ bedeutet diese Definition also, dass $a^{\sqrt{2}}$ zwischen $a^{1.4} = a^{7/5} = (\sqrt[5]{a})^7$ und $a^{1.5} = a^{3/2} = (\sqrt{a})^3$ liegt, genauer zwischen $a^{1.41} = a^{141/100} = (\sqrt[100]{a})^{141}$ und $a^{1.42} = a^{71/50} = (\sqrt[50]{a})^{71}$, noch genauer zwischen $a^{1.414} = a^{707/500} = (\sqrt[500]{a})^{707}$ und $a^{1.415} = a^{283/200} = (\sqrt[200]{a})^{283}$, usw.; je besser man $\sqrt{2}$ durch rationale Zahlen r wie 1.4, 1.41, 1.414, 1.4142, ... approximiert hat, desto besser approximieren die Potenzen der Basis a mit diesen rationalen Exponenten (also Potenzen, die man mit Wurzelziehen berechnen kann) auch die Potenz $a^{\sqrt{2}}$.

- 2) • *Potenzen mit allgemeinen Exponenten s sind nur für positiven Basen $a > 0$ erklärt und haben stets positive Werte a^s .*

Das muss betont werden; denn sonst können die Potenzrechengesetze nicht gelten. (Z.B. kann für $a > 0$ dann $-a = (-a)^{1/2+1/2} = ((-a)^{1/2})^2$ nicht richtig sein, weil Quadrate nie negativ sind, und aus demselben Grund könnte $b^s = b^{s/2} \cdot b^{s/2}$ nicht gelten, wenn $b^s < 0$ wäre.) Eine Ausnahme hiervon machen wir nur bei ganzzahligen Exponenten, wo wir $a^n = a \cdot a \cdot \dots \cdot a$ (n Faktoren) für beliebige reelle Basen definiert haben, wenn $n \in \mathbb{N}$ ist, sowie $a^{-n} := 1/a^n$ für reelle Basen $a \neq 0$. Es ist auch sinnvoll, noch $0^s := 0$ zu setzen, allerdings nur für positive Exponenten $s > 0$. Die Potenz " 0^0 " bleibt undefiniert, weil es genau so gute Gründe gibt, diesem Ausdruck den Wert 0 zuzuweisen (es ist nämlich $0^s = 0$ für $s > 0$), wie es Argumente für den Wert 1 gibt (nämlich $a^0 = 1$ für Basen $a \neq 0$). Man sagt: " 0^0 " ist ein "unbestimmter Ausdruck". Für negative Exponenten $-s < 0$ kann man 0^{-s} auch nicht sinnvoll erklären, weil es das Reziproke zu 0^s sein müsste, wenn die Potenzrechenregeln gelten sollen, aber der Kehrwert von $0^s = 0$ kann nicht gebildet werden. Die Ausnahmen ändern aber nichts am Grundsatz: *Sobald in einer Rechnung nichtganze Exponenten auftauchen, müssen die Basen positiv sein*, sonst kann Unfug entstehen wie eben $-1 = (-1)^1 = (-1)^{1/2+1/2} = (-1)^{1/2} \cdot (-1)^{1/2} = ((-1)^{1/2})^2 \geq 0$.

- 3) Die Rechenregeln für den Umgang mit Potenzen von positiven Basen $a > 0$ und $b > 0$ zu beliebigen reellen Exponenten s, t sind folgende **Potenzgesetze**:

- *Potenzen mit Exponent 0 und Potenzen von 1 mit beliebigem Exponenten haben den Wert 1:*

$$a^0 = 1, \quad 1^s = 1;$$

- *man multipliziert/dividiert Potenzen derselben Basis, indem man die Exponenten addiert/subtrahiert:*

$$a^s \cdot a^t = a^{s+t}, \quad a^s / a^t = a^{s-t};$$

- *man bildet den Kehrwert einer Potenz, indem man das Vorzeichen des Exponenten umkehrt:*

$$1/a^s = a^{-s};$$

- *man potenziert eine Potenz, indem man die Exponenten multipliziert:*

$$(a^s)^t = a^{s \cdot t};$$

- *man multipliziert/dividiert zwei Potenzen mit demselben Exponenten, indem man das Produkt/den Quotienten der Basen mit diesem Exponenten potenziert:*

$$a^s \cdot b^s = (a \cdot b)^s, \quad a^s / b^s = (a/b)^s.$$

All dies muss man nur für rationale Exponenten zeigen, für reelle Exponenten folgt es dann aufgrund der Definition der Potenz durch Approximation mit Potenzen zu rationalen Exponenten automatisch. Für rationale Exponenten $s = \frac{k}{l}$ und $t = \frac{m}{n}$ (mit $k, m \in \mathbb{Z}$ und $l, n \in \mathbb{N}$) aber sind die angegebenen Potenzgesetze unmittelbare Folgerungen aus den Gesetzen der Wurzelrechnung und den bereits bekannten Potenzrechenregeln bei ganzen Exponenten. Zum ersten Gesetz $a^0 = 1$ und $1^s = 1$ brauchen wir nichts mehr zu sagen,

und die andern Potenzgesetze ergeben sich dann so:

$$\begin{aligned} a^{\frac{m}{n}} \cdot b^{\frac{m}{n}} &= (\sqrt[n]{a})^m \cdot (\sqrt[n]{b})^m = (\sqrt[n]{a} \cdot \sqrt[n]{b})^m = (\sqrt[n]{a \cdot b})^m = (a \cdot b)^{\frac{m}{n}}, \\ (a^{\frac{k}{l}})^{\frac{m}{n}} &= \left(\sqrt[l]{(\sqrt[l]{a})^k} \right)^m = \left(\left(\sqrt[l]{\sqrt[l]{a}} \right)^k \right)^m = \left(\sqrt[l \cdot n]{a} \right)^{k \cdot m} = a^{\frac{k \cdot m}{l \cdot n}} = a^{\frac{k}{l} \cdot \frac{m}{n}}, \\ a^{\frac{k}{l}} \cdot a^{\frac{m}{n}} &= a^{\frac{kn}{ln}} \cdot a^{\frac{lm}{ln}} = (\sqrt[l \cdot n]{a})^{kn} \cdot (\sqrt[l \cdot n]{a})^{lm} = (\sqrt[l \cdot n]{a})^{kn+lm} = a^{\frac{kn+lm}{ln}} = a^{\frac{k}{l} + \frac{m}{n}}, \end{aligned}$$

und dieselben Rechnungen sind auch gültig für Division statt Multiplikation. Das Gesetz über den Kehrwert einer Potenz ist ein Spezialfall der Division, nämlich $1/a^s = 1^s/a^s = (1/a)^s = (a^{-1})^s = a^{-s}$.

4) Wurzeln $\sqrt[n]{a}$ hatten wir definiert als eindeutige positive Lösung der Potenzgleichung $x^n = a$. Nun können wir solche Gleichungen auch für beliebige reelle Exponenten $s \neq 0$ betrachten und lösen:

- Die eindeutige positive Lösung der Potenzgleichung $x^s = a$ zu gegebenem Exponenten $s \in \mathbb{R}_{\neq 0}$ und rechter Seite $a \in \mathbb{R}_{>0}$ ist $x = a^{1/s}$; hierfür schreibt man auch $\sqrt[s]{a}$.

Auf Taschenrechnern sind diese Wurzeln zu nichtganzen Exponenten s (mit der Funktionsbezeichnung " $\sqrt[s]{y}$ ") meistens vorhanden und alle früher angegebenen Wurzelrechengesetze gelten für solche allgemeineren Wurzeln $\sqrt[s]{a}$, bei denen s nicht mehr eine natürliche Zahl sein muss. Diese Gesetze für das Wurzelrechnen braucht man sich aber nicht zu merken; denn es handelt sich einfach um Spezialfälle der Potenzgesetze, bei denen die Exponenten in der Form $\frac{1}{s}$ geschrieben sind:

$$\begin{aligned} (\sqrt[s]{a})^s &= a = \sqrt[s]{a^s} &\iff & (a^{1/s})^s = a^1 = (a^s)^{1/s}, \\ (\sqrt[s]{a})^t &= \sqrt[s]{a^t} &\iff & (a^{1/s})^t = a^{t/s} = (a^t)^{1/s}, \\ \sqrt[s]{\sqrt[t]{a}} &= \sqrt[st]{a} = \sqrt[t]{\sqrt[s]{a}} &\iff & (a^{1/t})^{1/s} = a^{1/st} = (a^{1/s})^{1/t}, \\ \sqrt[s]{a} \cdot \sqrt[s]{b} &= \sqrt[s]{a \cdot b} &\iff & a^{1/s} \cdot b^{1/s} = (a \cdot b)^{1/s}, \end{aligned}$$

und entsprechend für Division statt Multiplikation. Sogar "neue" Wurzelrechengesetze, auf die man nicht ohne Weiteres kommt, lassen sich aus den Potenzrechengesetzen ableiten:

$$\sqrt[s]{a} \cdot \sqrt[t]{a} = \sqrt[st]{a^{s+t}} = (\sqrt[st]{a})^{s+t} \iff a^{1/s} \cdot a^{1/t} = a^{\frac{1}{s} + \frac{1}{t}} = a^{\frac{s+t}{st}}.$$

Statt mit solchen komplizierten Wurzelrechengesetzen zu operieren, sollte man aber besser

- immer alle Wurzeln $\sqrt[s]{a}$ als Potenzen $a^{1/s}$ umschreiben und dann die einfacheren und leicht zu merkenden Rechenregeln für Potenzen verwenden. ■

Es liegt auf der Hand, dass die Basen 10 (Dezimalsystem) und 2 (dyadisches System, wird für die Zahlendarstellung in Computern benutzt ebenso wie das Hexadezimalsystem, also die Basis 16) von besonderer Bedeutung sind. Keineswegs offensichtlich und eine bemerkenswerte mathematische Entdeckung ist, dass eine andere Basis für die Potenzrechnung besonders günstig ist, nämlich die Eulersche Zahl $e = 2.71828182\dots$. Diese Zahl ist nicht rational und auch nicht mit wiederholtem Wurzelziehen aus einer rationalen Zahl zu erhalten; es ist vielmehr bekannt, dass sie *transzendent* ist, d.h. sie löst keine algebraische Gleichung $a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 = 0$ mit rationalen Koeffizienten a_k . Wie kann es sein, dass eine so "abgelegene" reelle Zahl in gewisser Weise die beste Basis für die Potenzrechnung ist, so dass sie als "natürliche Basis" bezeichnet wird? Bemerkenswerterweise hängt die Antwort mit Finanzmathematik zusammen, nämlich mit dem Problem der kontinuierlichen Verzinsung. Dazu greifen wir in der folgenden Diskussion auf Kap. 2 vor.

DISKUSSION (kontinuierliche Verzinsung und die natürliche Basis):

1) Wir betrachten die Verzinsung eines Kapitals in einem Zeitraum t Jahre zu einem gegebenen Zinsfuß $p\%$. Bei einfacher Verzinsung ist der Aufzinsungsfaktor dann $1 + \frac{pt}{100}$. Teilen wir den Zeitraum in n gleich lange Zinsperioden ein und schlagen die Zinsen am Ende jeder Periode dem Kapital zu, so ist der Aufzinsungsfaktor $(1 + \frac{pt}{100n})^n$ natürlich größer. Das ist ökonomisch klar wegen des Zinseszinsseffekts und mathematisch auch leicht beweisbar; z.B. ergibt sich aus der binomischen Formel, wenn man nur die ersten beiden Summanden berücksichtigt, sofort $(1 + \frac{pt}{100n})^n > 1 + \binom{n}{1} \frac{pt}{100n} = 1 + \frac{pt}{100}$.

Nun kann man sich auf den Standpunkt stellen, dass die Einteilung in endlich viele Zinsperioden, z.B. in Quartale oder Monate, "ungerecht" ist — vielmehr müsste man, um die anfallenden Zinsen "gerecht" mitzuverzinsen, die Laufzeit in Tage, Stunden, Sekunden, ja eigentlich in beliebig kurze Zinsperioden einteilen. Das bedeutet, dass die Zahl n der Zinsperioden immer größer gemacht werden muss und schließlich über alle Grenzen wächst. Es ist ökonomisch klar und mathematisch auch beweisbar, dass die Aufzinsungsfaktoren $q_n := (1 + \frac{pt}{100n})^n$ größer werden, wenn man die Zahl n der Zinsperioden (innerhalb derselben Laufzeit t) vergrößert; denn dabei wirkt sich der Zinseszinsseffekt ja stärker aus. Die Frage ist nun: Wie groß werden diese Aufzinsungsfaktoren? Etwa beliebig groß, so dass der Zinsertrag bei "gerechter" Verzinsung ebenfalls beliebig groß wird? Ökonomisch betrachtet erscheint das zumindest unwahrscheinlich, und ein Test, etwa mit $p\% = 5\%$ und $t = 1$ (Jahr), gibt für $n = 1, 2, 10, 100, 1000, \dots$ die Aufzinsungsfaktoren $1.05, 1.025^2 = 1.050625, 1.005^{10} \approx 1.051135, 1.0005^{100} \approx 1.051258, 1.00005^{1000} \approx 1.0512695, \dots$, was auch nicht gerade auf ein Anwachsen über alle Schranken hindeutet.

Tatsächlich bleiben die q_n beschränkt, und auch dies kann man mit einer ökonomischen Überlegung einsehen. Dazu betrachten wir für $n > \frac{pt}{100}$ die Abzinsungsfaktoren $\tilde{q}_n := (1 - \frac{pt}{100n})^n$ für die vorschüssige Abzinsung eines Kapitals in n gleich langen Zinsperioden der Gesamtdauer t . Es ist ökonomisch klar und wiederum mathematisch beweisbar, dass auch diese Faktoren mit wachsendem n größer werden, weil die abzuziehenden Zinsen nach jeder Zinsperiode von einem verkleinerten Kapital berechnet werden, ein Effekt, der sich bei mehr Zinsperioden früher und stärker auswirkt. Nun gilt hier aber $q_n \cdot \tilde{q}_n = [(1 + \frac{pt}{100n}) \cdot (1 - \frac{pt}{100n})]^n = [1 - (\frac{pt}{100n})^2]^n < 1$, d.h. $q_n < 1/\tilde{q}_n$. Weil hier die Zahlen q_n und \tilde{q}_n mit wachsendem n anwachsen, die Reziproken $1/\tilde{q}_n$ also abnehmen, können die q_n mit wachsendem n nicht beliebig groß werden. Nun ist es eine fundamentale Eigenschaft des Bereichs der reellen Zahlen, dass jede wachsende Folge $q_1 < q_2 < q_3 < \dots$ von reellen Zahlen, die von oben beschränkt ist (d.h. alle q_n liegen unterhalb einer festen Zahl) einen Grenzwert q besitzt, was hier bedeutet, dass die q_n dieser Zahl q so nahe kommen wie man will (aber stets kleiner bleiben), wenn man nur n genügend groß wählt. Diese Zahl wird $q := \lim_{n \rightarrow \infty} q_n$ bezeichnet ("Limes von q_n bei n gegen Unendlich") und hat die ökonomische Bedeutung des Aufzinsungsfaktors zum Zinsfuß $p\%$ in der Laufzeit t Jahre bei Zinszuschlag nach beliebig kurzen Zinsperioden, also gewissermaßen bei ständigem Zinszuschlag. Sie heißt daher

- **Aufzinsungsfaktor bei kontinuierlicher Verzinsung** zu $p\%$ in der Laufzeit t :

$$q = \lim_{n \rightarrow \infty} \left(1 + \frac{p \cdot t}{100 \cdot n}\right)^n.$$

2) Die Eulersche Zahl e erhalten wir nun, wenn wir oben p und t so wählen, dass $\frac{pt}{100} = 1$ ist, also z.B. die kontinuierliche Verzinsung für 100 Jahre zu 1% betrachten oder für 20 Jahre zu 5%. Es ergibt sich dann die sogenannte

- natürliche Basis (Eulersche Zahl):

$$e := \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = 2.718281828459\dots$$

An dieser Stelle ist noch nicht einzusehen, warum diese Zahl eine "gute Basis" für die Potenzbildung sein soll. Aber wenn wir annehmen, dass $\frac{pt}{100} = \frac{k}{l}$ rational ist mit $k, l \in \mathbb{N}$ und für n Vielfache $n = m \cdot k$ des Zählers k einsetzen, so ergibt sich $q_n^l = \left[\left(1 + \frac{1}{l \cdot m}\right)^{l \cdot m}\right]^k$. Hier strebt mit wachsendem m nun der Term [...] gegen die Eulersche Zahl e , weil $l \cdot m$ natürliche Zahlen durchläuft und beliebig groß wird, und man überlegt sich leicht, dass die Potenzen [...] e^k beliebig nahe kommen für hinreichend große m . Auf der anderen Seite strebt q_n mit wachsendem $n = m \cdot k$ gegen den kontinuierlichen Aufzinsungsfaktor q aus 1) und die Potenz q_n^l damit gegen q^l . Das Resultat ist, dass $q^l = e^k$ sein muss, also $q = e^{k/l} = e^{pt/100}$. Mit Worten:

- Der Aufzinsungsfaktor bei kontinuierlicher Verzinsung ist die Potenz der Eulerschen Zahl mit dem Laufzeit-anteiligen Zinsfuß als Exponenten,

$$q = \lim_{n \rightarrow \infty} \left(1 + \frac{p \cdot t}{100 \cdot n}\right)^n = e^{\frac{p \cdot t}{100}}$$

Das haben wir zwar nur für rationale Werte von $\frac{pt}{100}$ vorgerechnet, es gilt aber für beliebige positive reelle Werte x des Laufzeit-anteiligen Zinsfußes, weil der dazu gehörende Aufzinsungsfaktor offenbar zwischen den Aufzinsungsfaktoren zu rationalen Werten von $\frac{pt}{100}$ liegt, die kleiner bzw. größer als x sind.

Außerdem gilt eine entsprechende Aussage für das vorschüssige kontinuierliche Abzinsen. Wir haben in 1) schon gesehen, dass die zugehörigen Abzinsungsfaktoren $\tilde{q}_n < 1$ mit n anwachsen, wobei $q_n \cdot \tilde{q}_n < 1$ gilt. Daraus folgt, dass auch die Folge $\tilde{q}_1, \tilde{q}_2, \tilde{q}_3, \dots$ einen Grenzwert \tilde{q} hat, wobei $q \cdot \tilde{q} \leq 1$ ist. Für große n haben wir andererseits auch $q_n \cdot \tilde{q}_n = \left[1 - \left(\frac{pt}{100n}\right)^2\right]^n > 1 - \frac{1}{n} \left(\frac{pt}{100}\right)^2$, was man erkennt, wenn man den vorletzten Term als vorschüssigen kontinuierlichen Abzinsungsfaktor zu $p\%$ für n Zinsperioden in der Laufzeit $pt^2/100n$ auffasst und mit dem aus ökonomischen Gründen offensichtlich kleineren vorschüssigen Diskontierungsfaktor bei einfacher Verzinsung vergleicht. Da $\frac{1}{n}$ beliebig klein wird mit wachsendem n , muss daher auch $q \cdot \tilde{q} \geq 1$ sein, d.h. es gilt sogar $\tilde{q} = q^{-1} = e^{-pt/100}$. Da dies das Reziproke des Aufzinsungsfaktors $e^{pt/100}$ ist, sind also bei kontinuierlicher Verzinsung die vor- und die nachschüssigen Aufzinsungsfaktoren gleich, und wir können die erhaltenen Ergebnisse wie folgt zusammenfassen:

- Aufzinsungsfaktor bzw. Abzinsungsfaktor zum Laufzeit-anteiligen Zinsfuß $x = \frac{p \cdot t}{100} > 0$ sind bei kontinuierlicher (nach- oder vorschüssiger) Verzinsung gegeben durch die Potenzen der natürlichen Basis mit Exponent x bzw. $-x$,

$$e^x = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n \quad \text{bzw.} \quad e^{-x} = \lim_{n \rightarrow \infty} \left(1 - \frac{x}{n}\right)^n$$

Es versteht sich, dass zu einem strengen Beweis der obigen Grenzwertformeln für die Potenzen von e mathematische Grenzwerttheorie erforderlich ist, um die Rechnungen mit Grenzwerten, die wir angedeutet haben, zu rechtfertigen. Diese Grenzwerttheorie stellt die Mathematik auch zur Verfügung, d.h. man kann alle obigen Argumente zu exakten mathematischen Beweisen ausarbeiten. Das brauchen wir aber hier nicht zu tun; denn für die Mathematik in der Wirtschaftswissenschaft genügt eine intuitive Vorstellung von Grenzwerten und Grenzübergängen, und damit kann man auch den Zusammenhang von kontinuierlicher Verzinsung mit den Potenzen der Eulerschen Zahl schon gut verstehen.

3) Mit der Interpretation der Potenzen von e als Aufzinsungsfaktoren für das Anwachsen eines Kapitals bei kontinuierlicher Verzinsung ist nur eine von vielen Eigenschaften beschrieben, welche die natürliche Basis gegenüber anderen Basen auszeichnet. Dieselbe Eigenschaft ist übrigens auch die Ursache für das Auftauchen der Potenzen dieser Zahl in vielen Gesetzen der Naturwissenschaften; denn wenn man einen Prozess unbeschränkten Wachstums oder Zerfalls hat (wie z.B. radioaktiver Zerfall), so lässt sich genau dieselbe Argumentation anwenden wie oben beim Anwachsen eines Kapitals durch kontinuierliche Aufzinsung bzw. beim Abnehmen des Kapitalstands durch kontinuierliche Abzinsung, d.h. man erhält als Wachstums- bzw. Zerfallsfaktoren für die Laufzeit t die Potenzen $e^{\pm kt} = a^t$ mit einer gewissen für den Prozess charakteristischen Wachstumsrate $k > 0$ und $a = e^{\pm k}$. Ein solches "Exponentialgesetz" ist auch deswegen zu erwarten, weil bei demselben kontinuierlichem Wachstum in n aufeinander folgenden Zeitspannen t_1, \dots, t_n der Wachstumsfaktor für die Gesamtlaufzeit $t = t_1 + \dots + t_n$ gleich dem Produkt der Wachstumsfaktoren für die einzelnen Zeitabschnitte t_k sein sollte, und in der Tat ist ja auch gemäß Potenzgesetzen

$$a^{t_1} \cdot a^{t_2} \cdot \dots \cdot a^{t_n} = a^{t_1+t_2+\dots+t_n}.$$

4) Ein besonderer Vorteil der natürlichen Basis e ist, dass man die Potenzen e^x mit beliebigen reellen Exponenten x sehr gut und schnell näherungsweise berechnen kann, viel besser als die Potenzen anderer Basen wie 2 oder 10. Die Darstellung von e^x als Grenzwert der Ausdrücke $(1 + \frac{x}{n})^n$ bei $n \rightarrow \infty$ ist dazu allerdings wenig geeignet; für $x = 1$ etwa ist der Fehler $e^x - (1 + \frac{x}{n})^n$ etwa so groß wie $\frac{1}{n}$, was bedeutet, dass man n sehr groß wählen muss (z.B. 10^6), um durch Berechnung von $(1 + \frac{x}{n})^n$ den Wert von e^x auf einige Dezimalen (z.B. 6 Dezimalen) genau zu erhalten. Mit der binomischen Entwicklung

$$\left(1 + \frac{x}{n}\right)^n = \sum_{k=0}^n \binom{n}{k} \left(\frac{x}{n}\right)^k$$

kann man eine für die Berechnung von e^x besser geeignete Darstellung erhalten. Dazu überlegt man, dass die Faktoren $\binom{n}{k} \left(\frac{1}{n}\right)^k = \frac{1}{k!} \cdot \frac{n}{n} \cdot \frac{n-1}{n} \cdot \dots \cdot \frac{n-k+1}{n}$ für große n ungefähr gleich $\frac{1}{k!}$ sind, wobei die Übereinstimmung um so besser ist, je größer n gewählt wird. Mit wachsendem n strebt daher die linke Seite der binomischen Entwicklung gegen e^x und die rechte gegen die Summe der Summanden $\frac{1}{k!} x^k$, d.h. man erhält die für alle $x \in \mathbb{R}$ gültige Darstellung der Potenz e^x durch die sog. **Exponentialreihe**

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + \frac{1}{1!}x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \dots$$

Dies ist wie bei der in Abschnitt 1.1 erwähnten unendlichen geometrischen Reihe so zu verstehen, dass die Summe der ersten n Summanden rechts ein Näherungswert für e^x ist, wobei der Fehler kleiner ist als jede vorgegebene Fehlerschranke (z.B. 10^{-6}), wenn man nur n gross genug wählt. Weil die Fakultäten $k!$ der natürlichen Zahlen k sehr schnell anwachsen mit wachsendem k — viel schneller als die Potenzen x^k jeder festen Zahl x , selbst wenn x groß ist —, werden die Glieder der Exponentialreihe sehr schnell klein. Man braucht daher nur wenige Summanden der Reihe zu berücksichtigen, um e^x schon mit sehr großer Genauigkeit durch die Summe dieser Glieder annähern zu können. Eine genauere Analyse zeigt, dass der Fehler, den man bei Abbruch der Exponentialreihe mit dem n -ten Summanden macht, nicht größer ist als $\frac{1}{(n+1)!} |x|^{n+1} e^{|x|}$; für $|x| \leq 1$ genügen daher die ersten 10 Summanden der Reihe, um e^x bis auf 6 Dezimalstellen genau zu erhalten. *Die natürliche Basis e ist dadurch ausgezeichnet, dass man mit der Exponentialreihe ein exzellentes Verfahren zur Berechnung ihrer Potenzen e^x zur Verfügung hat* — ein derart einfaches und genaues Verfahren zur Berechnung von Potenzen a^x gibt es für andere Basen a nicht!

5) Die sog. **natürliche Exponentialfunktion** $\exp(x) := e^x$ zur Basis e hat weitere besondere Eigenschaften, die sie vor den Exponentialfunktionen $\exp_a(x) := a^x$ zu anderen Basen a auszeichnet. Zum Beispiel ist sie die einzige Exponentialfunktion, die Steigung 1 an der Stelle $x = 0$ hat, d.h. es gilt $e^x \approx 1 + x$ für kleine Werte des Betrags $|x|$, wobei der Fehler im Verhältnis zur Größe von $|x|$ beliebig klein wird, wenn $|x|$ klein genug ist. Daraus folgt, dass die natürliche Exponentialfunktion die einzige Exponentialfunktion — und bis auf einen Faktor sogar die einzige Funktion überhaupt — ist, deren Ableitung die Funktion selbst ist (wir kommen darauf in der Differentialrechnung zurück). Die Eulersche Zahl e tritt ferner in vielen sehr verschiedenen Zusammenhängen in der Mathematik auf, es handelt sich in diesem Sinne wirklich um ein fundamentale “Naturkonstante”, wie z.B. auch die Kreiszahl π , und zwischen beiden Zahlen gibt es tiefe Zusammenhänge. Wir erwähnen die *Stirlingsche Formel*

$$n! \sim \sqrt{2\pi n} \frac{n^n}{e^n},$$

die für große natürliche Zahlen n die Größe der Fakultäten $n!$ sehr genau beschreibt. Für große Zahlen n sind die Fakultäten, eben weil sie riesig groß sind, schwer in den Griff zu bekommen; ein gewöhnlicher Taschenrechner kommt z.B. nur bis $(69)!$, weil $(70)!$ schon größer ist als die auf dem Rechner nicht mehr darstellbare Zahl 10^{100} . Die Stirlingsche Formel gibt Näherungswerte für die Fakultäten in dem Sinne, dass der relative Fehler, das ist der Quotient aus linker und rechter Seite minus 1, sehr klein ist für große n , und zwar sogar kleiner als eine beliebig klein gegebene Fehlerschranke, wenn nur n groß genug ist. (Genauer ist der relative Fehler größer als Null und kleiner als $e^{1/12n} - 1$.) Die Näherungsformel für die Werte der Fakultäten $n!$ wird also gewissermaßen um so besser, je größer n ist! Eine gute näherungsweise Kenntnis der Fakultäten großer Zahlen n und der hierdurch ausdrückbaren Binomialkoeffizienten $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ ist wichtig in der Statistik bei großen Fallzahlen. Dies ist ein durchaus auch für die Wirtschaftswissenschaft relevantes mathematisches Fachgebiet, und insofern ist auch die Stirlingsche Formel für die Ökonomie von Bedeutung. Statistik gehört aber nicht mehr zum Lehrstoff dieses Kurses, sondern ist ein eigener spezieller Kurs im Studium der Wirtschaftswissenschaft.

Ein anderer tiefer Zusammenhang zwischen e und π zeigt sich, wenn man den Bereich der reellen Zahlen \mathbb{R} zum Bereich der komplexen Zahlen \mathbb{C} erweitert, in dem auch eine Quadratwurzel aus -1 vorhanden ist, die sog. “imaginäre Einheit” \mathring{i} . Man kann dann auch die Potenzen a^z von positiven Basen $a \in \mathbb{R}_{>0}$ mit komplexen Zahlen $z \in \mathbb{C}$ als Exponenten sinnvoll erklären, und Euler hat eine tiefe Beziehung der Potenzen von e mit Vielfachen $\mathring{i}t$ der imaginären Einheit als Exponenten ($t \in \mathbb{R}$) zu den Kreisfunktionen Sinus und Kosinus und damit zur Kreiszahl π gefunden, nämlich die *Eulersche Formel*

$$e^{\mathring{i}t} = \cos t + \mathring{i} \sin t,$$

speziell

$$e^{\mathring{i}\pi/2} = \mathring{i}, \quad e^{\mathring{i}\pi} = -1, \quad e^{\mathring{i}3\pi/2} = -\mathring{i}, \quad e^{\mathring{i}2\pi} = 1.$$

In der Wirtschaftsmathematik kommen freilich komplexe Zahlen nicht (oder selten) vor, daher ist auch die bemerkenswerte Eulersche Formel dort nicht von Bedeutung. Wir haben sie nur erwähnt, um die besondere Rolle der natürlichen Basis e zu unterstreichen, die sich — wie oben gesehen — schon bei der mathematischen Modellierung bestimmter “idealer” ökonomischer Vorgänge wie der kontinuierlichen Verzinsung zeigt, mehr aber noch, wenn man in andere Bereiche der Mathematik hineinschaut. ■

Wir kommen nun zum letzten Thema dieses Abschnittes, den Logarithmen. Auf diesen Begriff stößt man, wenn man die Lösung von *Exponentialgleichungen* $a^x = y$ untersucht, d.h. zu gegebener Basis a und rechter Seite y ist der Exponent x gesucht, für den die Potenz a^x den Wert y hat. (Natürlich muss, wie immer, $a > 0$ und $y > 0$ sein, damit die Fragestellung sinnvoll ist; die Basis $a = 1$ ist ebenfalls ein auszuschließender Sonderfall, weil ja $1^x = 1$ ist für alle x , so dass die Exponentialgleichung dann für $y \neq 1$ keine und für $y = 1$ alle Zahlen x als Lösung hat.) Betrachten wir eine Basis $a > 1$, so wird a^x beliebig groß für große Werte von x , also größer als y . Das erkennt man daran, dass für die nächstkleinere ganze Zahl $n := \lfloor x \rfloor$ die Ungleichung $a^x = a^{x-n} \cdot a^n \geq a^n = (1 + (a-1))^n > 1 + n(a-1)$ gilt (wobei wir im letzten Schritt nur die beiden ersten Summanden der binomischen Entwicklung berücksichtigt haben; außerdem ist $a^{x-n} \geq 1$, weil die Basis > 1 ist und der Exponent nichtnegativ, siehe 1.4). Für negative Werte von x mit sehr großem Betrag $|x| = -x$ ist andererseits a^x das Reziproke der sehr großen Zahl $a^{|x|}$ und damit sehr klein, also auch kleiner als $y > 0$. Mit der Zwischenwerteigenschaft der reellen Zahlen folgt daher, dass es einen Wert x des Exponenten geben muss, für den a^x genau gleich y ist. (Der exakte mathematische Beweis benutzt die Stetigkeit der Exponentialfunktion, d.h. a^s und a^t unterscheiden sich beliebig wenig, wenn s und t in \mathbb{R} nahe genug beieinander liegen. Das ergibt sich z.B. aus der für $n \in \mathbb{N}_{\geq 2}$ und $0 < t-s < \frac{1}{n}$ gültigen Abschätzung $a^t = a^{t-s} \cdot a^s < a^{1/n} \cdot a^s = \sqrt[n]{1+(a-1)} \cdot a^s < [1 + \frac{1}{n}(a-1)] \cdot a^s$; siehe 1.4.) Es kann auch nur einen einzigen solchen Wert geben; denn ist $a^x = y = a^{\tilde{x}}$, so folgt $a^{x-\tilde{x}} = 1$ und damit $x = \tilde{x}$, weil Potenzen von Basen $\neq 0$ mit Exponenten $\neq 0$ nicht den Wert 1 haben können (für $a > 1$ ist der Wert > 1 , wenn der Exponent positiv ist, und < 1 , wenn der Exponent negativ ist; siehe 1.4). Damit haben wir also für Basen $a > 1$ gefunden, dass die Exponentialgleichung $a^x = y$ für alle $y > 0$ genau eine Lösung hat. Diese Feststellung ist die Grundlage der folgenden Definition der Logarithmen. Für Basen $0 < a < 1$ gilt übrigens wegen $a^x = (1/a)^{-x}$ mit $1/a > 1$ dieselbe Aussage.

DISKUSSION (Logarithmenrechnung):

1) Zunächst ermöglichen uns die obigen Vorüberlegungen folgende Definition:

- Für eine gegebene Basis $1 \neq a \in \mathbb{R}_{>0}$ und eine gegebene Zahl $y \in \mathbb{R}_{>0}$ heißt die eindeutige Lösung $x \in \mathbb{R}$ der Exponentialgleichung $a^x = y$ der **Logarithmus des Numerus y zur Basis a** und wird $\log_a y$ notiert. Bei Wahl der Basis $a = 10$ schreibt man einfach $\log y$ statt $\log_{10} y$ und nennt diese Zahl den **dekadischen Logarithmus von y** (auch: Briggscher Logarithmus und gelegentlich $\lg y$ notiert). Bei Wahl der natürlichen Basis $a = e$ schreibt man $\ln y$ für $\log_e y$ und nennt dies den **natürlichen Logarithmus des Numerus y** ("ln" für "logarithmus naturalis").

Als Logarithmus wird also immer ein gesuchter *Exponent* bezeichnet, und der Numerus ist der Wert der Potenz der Basis zu diesem Exponenten. Dabei *muss der Numerus stets positiv sein*; für $y \leq 0$ ist $\log_a y$ nicht definiert, weil a^x nur positive Werte annehmen kann. Außerdem *muss die Basis positiv und verschieden von 1 sein*, weil sonst die Potenzen a^x überhaupt nicht definiert sind bzw. nur den Wert 1 annehmen können. Üblicherweise werden nur Zahlen $a > 1$ als Basis gewählt, mit Basen $0 < a < 1$ arbeitet man besser nicht.

2) Aus der Definition der Logarithmen und den Potenzgesetzen ergeben sich sofort Rechenregeln für Logarithmen, die nur eine Übersetzung der Potenzgesetze in die Logarithmensprache" sind. Dabei nehmen wir natürlich an, dass alle Numeri und die Basis positiv sind und dass die Basis außerdem von 1 verschieden ist. Dann haben wir die folgenden

Logarithmusgesetze:

- die Potenz eines Logarithmus zu derselben Basis ist der Numerus, der Logarithmus einer Potenz zu derselben Basis ist der Exponent:

$$a^{\log_a y} = y, \quad \log_a(a^x) = x;$$

- Der Logarithmus von 1 ist stets 0, der Logarithmus der Basis zu sich selbst ist 1:

$$\log_a 1 = 0, \quad \log_a a = 1;$$

- der Logarithmus eines Produkts/Quotienten ist die Summe/Differenz der Logarithmen der Faktoren zur gleichen Basis:

$$\log_a(y \cdot z) = \log_a y + \log_a z, \quad \log_a \frac{y}{z} = \log_a y - \log_a z;$$

- der Logarithmus einer Potenz des Numerus ist das Produkt des Exponenten mit dem Logarithmus des Numerus zur gleichen Basis:

$$\log_a(y^s) = s \cdot \log_a y;$$

- der Logarithmus des Kehrwertes eines Numerus ist das Negative, der Logarithmus der n -ten Wurzel des Numerus ist das $\frac{1}{n}$ -Fache des Logarithmus vom Numerus:

$$\log_a \frac{1}{y} = -\log_a y, \quad \log_a \sqrt[n]{y} = \frac{1}{n} \log_a y.$$

Die erste Gleichung ist nichts anderes als die Definition des Logarithmus, $\log_a y$ ist schließlich gerade der Exponent, mit dem man a potenzieren muss, um das Ergebnis y zu erhalten. Und die zweite Gleichung besagt nichts anderes, als dass man eben a mit dem Exponenten x potenzieren muss, wenn man das Ergebnis a^x erhalten will. Diese beiden Gesetze kann man auch so zusammenfassen:

- Das Logarithmieren zur Basis a ist die Umkehroperation des Exponentierens zur Basis a ,

d.h. wenn man erst den Logarithmus zu einer positiven Zahl bildet und anschließend die Potenz der Basis mit diesem Exponenten, so erhält man die Ausgangszahl zurück, und wenn man umgekehrt erst zu einem reellen Exponenten die Potenz der Basis bildet und anschließend davon den Logarithmus, so erhält man den Ausgangsexponenten zurück. Für die speziellen Basen 10 und e gilt also

$$10^{\log y} = y \quad \log(10^x) = x, \quad e^{\ln y} = y, \quad \ln(e^x) = x.$$

Das zweite Logarithmusgesetz ist ein Spezialfall des ersten ($x = 0$ bzw. $x = 1$ wählen). Das dritte und vierte angegebene Gesetz ist die Entsprechung der Potenzgesetze $a^{x+\tilde{x}} = a^x \cdot a^{\tilde{x}}$ (entsprechend bei Subtraktion der Exponenten links und Division rechts) und $(a^x)^s = a^{x \cdot s}$; setzt man hier $x = \log_a y$ und $\tilde{x} = \log_a z$ ein, so erhält man sofort die behaupteten Gleichungen. Das letzte Gesetz schließlich ist ein Spezialfall des vorletzten ($x = -1$ bzw. $x = \frac{1}{n}$ einsetzen).

3) Die Logarithmusgesetze bezogen sich auf das Logarithmieren zu einer festen Basis a . Hier diskutieren wir Rechenregeln für den Wechsel der Basis. Wir betrachten dazu zwei von 1 verschiedene Basen $a, b \in \mathbb{R}_{>0}$. Aus

$$y = b^x = \left(a^{\log_a b}\right)^x = a^{x \cdot \log_a b}$$

lesen wir dann unmittelbar die folgenden Regeln für den Basiswechsel ab.

Umrechnen von Logarithmen und Potenzen bei Wechsel der Basis:

- Die Potenz einer neuen Basis mit einem Exponenten erhält man, indem man in der Potenz der alten Basis diesen Exponenten mit dem Logarithmus der neuen Basis multipliziert:

$$b^x = a^{x \cdot \log_a b};$$

- der Logarithmus eines Numerus bzgl. einer neuen Basis ist der Quotient des alten Logarithmus dieses Numerus und des alten Logarithmus der neuen Basis:

$$\log_b y = \frac{\log_a y}{\log_a b}.$$

Es versteht sich, dass hier der Logarithmus der neuen Basis b bzgl. der alten Basis a gemeint ist; denn der Logarithmus von b bzgl. b ist ja gleich 1. Statt mit dem Logarithmus der neuen Basis bzgl. der alten zu dividieren, kann man auch mit dem Logarithmus der alten Basis bzgl. der neuen multiplizieren; denn mit der speziellen Wahl $y = a$ erhält man aus der letzten Gleichung

$$\log_b a = \frac{1}{\log_a b}.$$

Für das häufig benötigte **Umrechnen auf die natürliche Basis** gilt insbesondere:

$$a^x = e^{x \cdot \ln a}, \quad \ln y = \frac{\log_a y}{\log_a e} = (\log_a y) \cdot (\ln a).$$

Beim Wechsel der Basis verändern sich die Logarithmen aller (positiven) Zahlen stets um denselben, nur von den beteiligten Basen abhängigen Faktor, daher gilt:

- Der Quotient zweier Logarithmen zu derselben Basis, ist von der Wahl der Basis unabhängig,

$$\frac{\log_a y}{\log_a z} = \frac{\log_b y}{\log_b z} = \frac{\log y}{\log z} = \frac{\ln y}{\ln z}.$$

Aus dieser leicht zu merkenden Aussage kann man sich die Umrechnungsformeln für Logarithmen bei Basiswechsel immer schnell herleiten; man braucht dazu für z nur eine der Basen zu nehmen und $\log_z z = 1$ zu beachten. Da $\log_a a^{-1} = -\log_a a = -1$ ist, zeigt die Umrechnungsformel für den Übergang zur reziproken Basis (was auch aufgrund der Logarithmusdefinition klar ist: Man muss $1/a$ mit $-x$ potenzieren, um dasselbe Ergebnis zu erhalten wie bei der Potenzierung von a mit x):

$$\log_{1/a} y = -\log_a y.$$

Deshalb ist es keine Einschränkung sich nur mit Basen $a > 1$ zu befassen: Die Logarithmen bzgl. einer Basis a mit $0 < a < 1$ erhält man einfach durch Vorzeichenumkehr bei den Logarithmen zur reziproken Basis $1/a > 1$.

4) Auf jedem wissenschaftlichen Taschenrechner sind die dekadische und die natürliche Logarithmusfunktion verfügbar sowie die natürliche Exponentialfunktion und die Exponentialfunktionen zu allgemeinen Basen. Wir wollen abschließend noch einige Worte dazu sagen, wie man Logarithmen und Potenzen mit allgemeinen Exponenten gut (näherungsweise) berechnen kann. Wie die natürliche Exponentialfunktion so hat auch die Logarithmusfunktion zur natürlichen Basis besondere Eigenschaften, die sie vor allen andern Logarithmusfunktionen auszeichnet und sie leichter berechenbar macht. Zu erwähnen ist

hier vor allem die für $0 < y \leq 2$ gültige **Logarithmusreihe**

$$\begin{aligned} \ln y &= \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} (y-1)^k && (0 < y \leq 2) \\ &= (y-1) - \frac{1}{2}(y-1)^2 + \frac{1}{3}(y-1)^3 - \frac{1}{4}(y-1)^4 + \dots \end{aligned}$$

Wie generell bei unendlichen Reihen ist dies so zu verstehen, dass die Teilsumme der ersten n Summanden rechts ein Näherungswert für die Zahl $\ln y$ links ist, wobei die Abweichung kleiner ist als eine beliebig klein gegebene Fehlerschranke, wenn man nur die Anzahl der erfassten Summanden groß genug macht. Für y nahe bei 0 oder bei 2 ist die Reihe zur Berechnung von $\ln y$ nicht gut geeignet, man braucht sehr viele Summanden (etwa 10^6), um $\ln y$ mit akzeptabler Genauigkeit (etwa 10^{-6}) zu erhalten. Für $y > 2$ ist die Reihe überhaupt nicht mehr brauchbar, die Summanden werden mit wachsendem k immer größer und die Teilsummen nähern sich mit wachsender Summandenzahl n keiner reellen Zahl mehr an; man sagt, dass die Reihe für $y > 2$ "divergiert". Für y zwischen $\frac{1}{2}$ und $\frac{3}{2}$ aber ist die Logarithmusreihe ganz brauchbar; mit n Summanden erhält man dann den Wert $\ln y$ schon mit einem Fehler kleiner als $\frac{1}{n}2^{-n}$. Und die Berechnung der Logarithmen anderer Zahlen z kann man mit den Logarithmusgesetzen auf die Berechnung der Logarithmen zu Zahlen zwischen $\frac{1}{2}$ und $\frac{3}{2}$ zurückführen, z.B. mit $\ln 2 = \ln(\frac{3}{2} \cdot \frac{4}{3}) = \ln \frac{3}{2} + \ln \frac{4}{3}$ und $\ln z = m \cdot \ln 2 + \ln(z \cdot 2^{-m})$, wobei $m \in \mathbb{Z}$ so gewählt wird, dass $\frac{1}{2} \leq z \cdot 2^{-m} \leq \frac{3}{2}$ ist.

Eine andere Formel, mit der man den natürlichen Logarithmus von $y > 0$ durch Ziehen von Wurzeln aus y approximieren kann, ist

$$\ln y = \lim_{n \rightarrow \infty} n \cdot (\sqrt[n]{y} - 1),$$

was bedeutet, dass der Ausdruck rechts ein Näherungswert für $\ln y$ ist, wobei der Fehler kleiner ist als jede gegebene Fehlerschranke, wenn man n groß genug wählt. Die Formel kann man verstehen, wenn man $\sqrt[n]{y} = y^{1/n} = e^{(1/n)\ln y}$ schreibt und mit der Exponentialreihe $1 + \frac{1}{1!}\frac{1}{n}\ln y + \frac{1}{2!}(\frac{1}{n}\ln y)^2 + \dots$ berechnet. Für große Werte von n machen wir hier einen Fehler von der Größenordnung $(\frac{1}{n})^2$, wenn wir die Summanden nach den ersten beiden weglassen, daher ist $n \cdot (\sqrt[n]{y} - 1) \approx \ln y$ bis auf einen Fehler der Größenordnung $\frac{1}{n}$. Lässt man n nur Zweierpotenzen $n = 2^m$ durchlaufen, so kann man \sqrt{y} , $\sqrt[4]{y} = \sqrt{\sqrt{y}}$, $\sqrt[8]{y} = \sqrt{\sqrt[4]{y}}$, ... durch fortgesetztes Quadratwurzelziehen berechnen und erhält so ein einfaches Verfahren für die näherungsweise Berechnung von $\ln y$ durch $2^m(\sqrt[2^m]{y} - 1)$, das nur das Ziehen von Quadratwurzeln und die Potenzen von 2 mit natürlichen Exponenten benötigt.

Nachdem man Verfahren zur Berechnung natürlicher Logarithmen $\ln y$ hat, kann man mit 2) natürlich auch Logarithmen $\log_a y = \ln y / \ln a$ zu beliebigen Basen $1 \neq a > 0$ berechnen und mit der Exponentialreihe auch allgemeine Potenzen $a^x = e^{x \ln a}$ solcher Basen. In der Mathematik sind Verfahren entwickelt (und in Taschenrechnern und Rechenprogrammen implementiert) worden, die viel besser und genauer sind als das, was wir hier angedeutet haben. Es ging uns lediglich darum, deutlich zu machen, dass man allgemeine Potenzen a^x und Logarithmen $\log_a y$ nicht nur wie oben sinnvoll definieren kann, sondern dass sich diese Potenzen und Logarithmen bei konkreten Vorgaben der Parameter a, x, y auch praktisch und effektiv berechnen lassen (selbst in Bereichen der Variablen, die von einem normalen Taschenrechner nicht mehr erfasst werden). ■

Wir schließen diesen Abschnitt mit zwei Beispielen zur Anwendung der Logarithmenrechnung, einer mathematischen und einer ökonomischen, die zeigt, dass einfache ökonomische Fragestellungen durchaus auf Exponentialgleichungen führen können, deren Lösung das Logarithmieren erfordert.

BEISPIELE (*Anwendungen der Logarithmenrechnung*):

1) Der dekadische Logarithmus $\log y$ bestimmt die Anzahl der Stellen vor bzw. der Nullen nach dem Dezimalpunkt in der Dezimaldarstellung der Zahl $y \in \mathbb{R}_{>0}$. Ist nämlich $m = \lfloor \log y \rfloor$ die größte ganze Zahl kleiner oder gleich $\log y$, so gilt $m \leq \log y < m+1$ und daher $10^m \leq y < 10^{m+1}$. (Wir benutzen hier, dass die Potenzen von 10 bei Vergrößerung des Exponenten größer werden; siehe 1.4.) Ist m eine natürliche Zahl oder Null, so bedeutet dies, dass die Dezimaldarstellung der Zahl y vor dem Dezimalpunkt $m+1$ Stellen hat mit führender Ziffer $\neq 0$. Ist aber $m = -n$ eine negative ganze Zahl, so bedeutet es, dass die Dezimaldarstellung von y mit einer Null vor dem Dezimalpunkt beginnt und nach dem Dezimalpunkt $n-1$ Nullen hat gefolgt von einer Ziffer $\neq 0$. (Dabei sollen abbrechende Dezimalbrüche stets so dargestellt sein, dass sie mit $000\dots$ enden, nicht mit $999\dots$.) Dieselben Aussagen gelten für andere Basen, z.B. gibt $\lfloor \log_2 y \rfloor$ die Anzahl der Stellen vor bzw. der Nullen nach dem Punkt in der dyadischen Zahldarstellung von $y \in \mathbb{R}_{>0}$ an.

2) Laufzeitberechnungen bei Zinseszinsverzinsung erfordern die Bestimmung unbekannter Exponenten, also das Logarithmieren. Sind etwa ein Anfangskapital $K_0 > 0$ und ein Aufzinsungsfaktor $q > 1$ für eine Zinsperiode gegeben und ist gefragt, nach welcher Anzahl von Zinsperioden ein vorgegebener Zielwert $K_{\text{Ziel}} > K_0$ erreicht oder erstmals überschritten ist, so hat man

$$q^n K_0 = K_{\text{Ziel}} \quad \text{bzw.} \quad q^n = \frac{K_{\text{Ziel}}}{K_0}$$

nach n aufzulösen. Logarithmieren gibt $n = \log_q(K_{\text{Ziel}}/K_0) = \frac{1}{\log q} \log(K_{\text{Ziel}}/K_0)$. (Zuletzt haben wir den dekadischen Logarithmus verwendet; die Wahl der Basis spielt aber bei Quotienten von Logarithmen zu derselben Basis keine Rolle, also hätten wir genau so gut die natürlichen Logarithmen nehmen können.) Nun wird der Exponent n , zu dem die Potenz von q genau den Wert K_{Ziel}/K_0 erreicht, im Allgemeinen keine ganze Zahl sein. Gehen wir dann zu der nächstgrößeren ganzen Zahl über, die mit $\lceil n \rceil$ bezeichnet wird, so überschreitet die Potenz von q mit diesem ganzen Exponenten erstmals den Wert K_{Ziel}/K_0 (weil die Potenzen von $q > 1$ mit wachsendem Exponenten größer werden, siehe 1.4). Ergebnis ist (wenn wir wieder n für $\lceil n \rceil$ schreiben) die

• **Laufzeitformel bei Zinseszinsverzinsung:**

$$n = \left\lceil \frac{\log K_{\text{Ziel}} - \log K_0}{\log q} \right\rceil.$$

Etwas komplizierter ist die Sache, wenn auch noch in jeder Zinsperiode eine positive Rate (Rente) R vom Kapital ausgezahlt werden soll. Der Kapitalstand nach n Zins- und Ratenperioden ist dann durch eine "Rentenformel" gegeben, die wir in Kap. 2 herleiten. Eine typische Fragestellung ist hier, wie lange die Rente gezahlt werden kann, bis das Kapital aufgebraucht ist. Man hat hier also $K_{\text{Ziel}} = 0$ zu wählen. Die Formel kann dann wieder nach q^n aufgelöst werden, und durch Logarithmieren erhält man n . Da dies im Allgemeinen keine ganze Zahl ist, muss man hier die nächstkleinere ganze Zahl bilden, um die Anzahl der Rentenperioden zu erhalten, nach denen das Kapital gerade noch positiv ist, so dass es mit Zahlung einer kleineren "Restrente" auf Null gesetzt wird. (Genauerer sagen wir in Kap. 2.) ■

Mathematik für Wirtschaftswissenschaftler

(K. Steffen, Heinrich-Heine-Universität Düsseldorf, WS 2006/07)

1.4 Die Ordnung der Zahlen, Rechnen mit Ungleichungen

Der Bereich \mathbb{R} der reellen Zahlen hat neben seiner durch die Grundrechenarten gegebenen algebraischen Grundstruktur noch eine weitere fundamentale Struktur, die den *Größenvergleich von Zahlen* ermöglicht. Das ist naturgemäß für die Ökonomie von größter Bedeutung: Alle ökonomischen Beurteilungen und Entscheidungen sind letztlich auf den Größenvergleich zwischen Zahlen (Kosten, Umsätze, Gewinne, ...) gegründet!

DISKUSSION (die Ordnung auf dem Zahlbereich \mathbb{R}):

1) Der Bereich der reellen Zahlen ist mit einer Ordnungsstruktur versehen, die in der Mathematik **totale Ordnung** genannt wird. Das bedeutet folgendes:

- Für je zwei reelle Zahlen x, y ist erklärt, ob **x kleiner als y** ist, notiert $x < y$, oder nicht, und es gelten dabei für alle $x, y, z \in \mathbb{R}$ die folgenden Ordnungsgesetze:

entweder $x < y$ oder $y < x$ oder $x = y$;

und:

wenn $x < y$ und $y < z$ ist, so gilt auch $x < z$.

Für $x < y$ sagt man auch, dass **y größer als x** ist, und schreibt dafür $y > x$. Gelegentlich wird hierfür auch formuliert, dass x (im Sinne der Ordnung) *vor* y kommt und y *nach* x . Außerdem führt man die Notation $x \leq y$ ein, gelesen **x kleiner oder gleich y** , sowie $y \geq x$, gelesen **y größer oder gleich x** , um auszudrücken, dass $x < y$ oder $x = y$ gilt. Eine Aussage der Form $x \leq y$ oder $y \geq x$ wird auch **schwache Ungleichung** genannt, im Unterschied zu einer **starken Ungleichung** $x < y$ oder $y > x$ oder $x \neq y$, bei der Gleichheit ausgeschlossen ist (auch *strenge* oder *strikte* Ungleichung genannt).

Die entscheidende Eigenschaft einer totalen Ordnung auf einer Menge ist die *Vergleichbarkeit aller Elemente*; das ist das erste oben angegebene Ordnungsgesetz: Von zwei verschiedenen Elementen ist stets eines im Sinne der Ordnung kleiner als das andere. (Es gibt sog. partielle Ordnungen, bei denen das anders ist. Man kann z.B. an eine Einteilung der Menge in verschiedene "Hierarchieklassen" denken, wobei Elemente in niedrigeren Hierarchieklassen kleiner sind als solche in höheren Klassen, aber Elemente innerhalb derselben Klasse nicht vergleichbar sind, d.h. keines ist kleiner als das andere.) Das zweite oben angegebene Ordnungsgesetz, die sog. *Transitivität*, ist für jeden Ordnungsbegriff unverzichtbar: Wenn x vor y kommt und y vor z , so kommt eben x vor z . (Wenn das nicht gilt, wie etwa bei der Anordnung der Elemente auf einer Kreislinie, so liegt kein sinnvoller Ordnungsbegriff vor.) Bei einer totalen Ordnung kann man gewissermaßen alle Elemente in einer Linie aufreihen, so dass kleinere immer weiter links stehen als größere; deshalb spricht man hier auch von einer "*linearen Ordnung*".

Bei der Veranschaulichung der reellen Zahlen als Punkte auf dem Zahlenstrahl entsprechen die kleineren Zahlen x Punkten, die weiter links liegen als die Punkte, die zu größeren Zahlen y gehören. Für ganze Zahlen bedeutet $m < n$, dass man n aus m durch Addition einer natürlichen Zahl (nämlich $n - m$) erhält; auf dem Zahlenstrahl gelangt man dann von m aus durch ein- oder mehrfaches Abtragen einer Einheitslänge nach rechts zu n . Für

rationale Zahlen $r = \frac{k}{l}$ und $t = \frac{m}{n}$ mit natürlichen Nennern erfolgt der Größenvergleich dadurch, dass man die Brüche gleichnamig macht und die Zähler vergleicht; also ist von den beiden Zahlen $r = \frac{kn}{ln}$ und $t = \frac{lm}{ln}$ diejenige die kleinere, die den kleineren Zähler hat (sofern $r \neq t$). Von zwei verschiedenen reellen Zahlen $x > 0$ und $y > 0$ ist diejenige die größere, die eine *lexikographisch größere Ziffernfolge* hat, d.h. an der ersten Stelle von links, bei der ein Unterschied auftritt, hat sie die größere Ziffer. (Dabei sind die Dezimaldarstellungen nach links durch unendlich viele vorangestellte Nullen aufzufüllen und bei abbrechenden Dezimalbrüchen auch nach rechts; ein Dezimalbruch, der mit lauter Ziffern 9 endet, muss also so umgeschrieben werden, dass er am Ende lauter Nullen hat.) *Es ist gerade der Vorteil der Dezimaldarstellung, dass man den Größenvergleich für Zahlen direkt an den Ziffernfolgen ablesen kann!*

2) Reelle Zahlen, die größer als Null sind, nennt man **positiv**, solche die kleiner als Null sind, **negativ**. Im Fall $x \geq 0$, wenn also x positiv oder Null sein kann, aber nicht negativ, nennen wir die Zahl x **nichtnegativ**, im Fall $x \leq 0$ entsprechend **nichtpositiv**. Die Menge der positiven, nichtnegativen, negativen bzw. nichtpositiven reellen Zahlen notieren wir mit $\mathbb{R}_{>0}$, $\mathbb{R}_{\geq 0} = \mathbb{R}_{>0} \cup \{0\}$, $\mathbb{R}_{<0}$, bzw. $\mathbb{R}_{\leq 0} = \mathbb{R}_{<0} \cup \{0\}$. Auf dem Zahlenstrahl liegen die positiven Zahlen rechts von Null, die negativen links davon. Die Gegenzahlen zu positiven Zahlen sind negativ und umgekehrt. Alle natürlichen Zahlen sind positiv.

3) In einer total geordneten Menge wie \mathbb{R} existiert nicht nur zu je zwei Elementen ein größtes, sondern auch zu endlich vielen gegebenen Elementen x_1, \dots, x_n . Dazu nimmt man die größere Zahl y_2 von x_1 und x_2 (wenn $x_1 = x_2$ ist, so setze $y_2 := x_2$), dann die größere Zahl y_3 von y_2 und x_3 und so fort, bis man nach endlich vielen Schritten mit der größten Zahl y_n von x_1, \dots, x_n endet. Diese Zahl heißt das **Maximum** von x_1, \dots, x_n . Entsprechend erklärt man das **Minimum** von endlich vielen Zahlen. Als Notationen dafür verwendet man

$$\max(x_1, \dots, x_n) \quad := \quad \text{die größte Zahl von } x_1, \dots, x_n.$$

$$\min(x_1, \dots, x_n) \quad := \quad \text{die kleinste Zahl von } x_1, \dots, x_n.$$

4) Für unendliche Mengen von Zahlen sieht die Sache im Allgemeinen anders aus: Zwar hat jede nichtleere Menge von natürlichen Zahlen ein kleinstes Element (das ist eine fundamentale Eigenschaft der Ordnung auf \mathbb{N}), aber nicht unbedingt ein größtes, z.B. offenbar $\mathbb{N} = \{1, 2, 3, \dots\}$ selbst nicht. Und die Menge der Stammbrüche $\{1, \frac{1}{2}, \frac{1}{3}, \dots\}$ hat ein größtes Element, aber kein kleinstes, und es gibt natürlich auch Teilmengen von \mathbb{R} wie etwa \mathbb{Z} oder \mathbb{Q} oder \mathbb{R} selbst, die weder ein größtes, noch ein kleinstes Element besitzen. Wir sagen, dass eine Menge M von reellen Zahlen **von oben beschränkt** ist, wenn es darin nicht beliebig große Zahlen gibt, d.h. wenn ein Zahl $s \in \mathbb{R}$ existiert, eine sog. **obere Schranke** für M , mit $x \leq s$ für alle $x \in M$. Entsprechend erklärt man eine **untere Schranke** für M und nennt M **von unten beschränkt**, wenn es eine untere Schranke gibt. Die Menge M heißt **beschränkt**, wenn sie von oben und unten beschränkt ist. Es ist nun eine fundamentale sog. *Vollständigkeitseigenschaft* der reellen Zahlen, dass es zu jeder (nichtleeren) von oben beschränkten Menge $M \subset \mathbb{R}$ eine *kleinste obere Schranke* s^* gibt, die **obere Grenze** oder **das Supremum** der Menge M genannt wird. Dies bedeutet, dass alle Zahlen $x \in M$ kleiner oder gleich s^* sind, dass aber jeder kleinere Wert $s < s^*$ durch Zahlen x aus M übertroffen wird. Entsprechend hat eine von unten beschränkte nichtleere Teilmenge M von \mathbb{R} eine *größte untere Schranke* s_* , die **untere Grenze** oder **Infimum** von M heißt. Gebräuchliche Notationen für Supremum bzw. Infimum von M sind

$$\sup M \quad \text{oder} \quad \sup_{x \in M} x \quad \text{bzw.} \quad \inf M \quad \text{oder} \quad \inf_{x \in M} x.$$

Wenn eine Menge in \mathbb{R} ein Maximum hat, ein größtes Element, so ist dieses auch das Supremum, und wenn sie ein Minimum besitzt, ein kleinstes Element, so ist dies auch das Infimum. Das Umgekehrte gilt aber nicht, weil Supremum bzw. Infimum nicht zur fraglichen Menge gehören müssen. Zum Beispiel ist 0 das Infimum der Menge der Stammbrüche, die kein Minimum hat; denn alle Stammbrüche sind positiv, und zu jeder positiven Zahl s gibt es Stammbrüche $\frac{1}{n}$, die kleiner als s sind. Man hat dazu die natürliche Zahl n größer als $\frac{1}{s}$ zu wählen; das geht, weil es **beliebig große natürliche Zahlen** in \mathbb{N} gibt. Dies ist eine weitere fundamentale Eigenschaft der Ordnung der reellen Zahlen, genannt die *Archimedische Eigenschaft*. (Man kann sie auch aus der Vollständigkeitseigenschaft folgern, indem man bemerkt, dass \mathbb{N} keine obere Grenze in \mathbb{R} haben kann.)

5) Für die Messung der "Größe" von negativen reellen Zahlen wird oft deren positive Gegenzahl verwendet. Z.B. stellt man eine Schuld auf einem überzogenen Konto oft als negativen Kontostand $-K$ dar, spricht aber von einer Schuld der (positiven) Größe K . Diese Sichtweise wird formalisiert durch die Einführung der Größe $|x|$, genannt **Absolutbetrag** oder **Betrag** der reellen Zahl x und definiert durch

$$|x| = |-x| := \max(x, -x) = \begin{cases} x, & \text{wenn } x \geq 0, \\ -x, & \text{wenn } x \leq 0. \end{cases}$$

Der Betrag $|x|$ von x ist positiv, außer $|x| = 0$ im Falle $x = 0$, und er misst gewissermaßen die *absolute Größe* der Zahl x . Wir sagen daher, dass eine Zahl x *absolut größer* ist als eine andere Zahl y wenn $|x| > |y|$ gilt. Wenn man von "großen" oder "kleinen" reellen Zahlen spricht, so meint man dies meist im absoluten Sinn. Eine Zahl wie $-10^6 = -1\,000\,000$, die sehr viel kleiner als Null ist, wird man in diesem Sinne nicht "klein" nennen, sondern besser "negativ und absolut groß", weil ihr Betrag $10^6 = 1\,000\,000$ eben eine große positive Zahl ist. Der Betrag der Differenz $|x-y|$ heißt der **Abstand von x zu y** auf der Zahlengeraden, und gibt die Länge der Strecke von x nach y an (in der Längeneinheit, die dem Abstand von 0 zu 1 entspricht).

Als Rechenregel für den Betrag vermerken wir neben $|x| = |-x| \geq 0$ die

$$\text{Multiplikativität: } |x \cdot y| = |x| \cdot |y| \quad \text{und} \quad \left| \frac{x}{y} \right| = \frac{|x|}{|y|} \quad \text{wenn } y \neq 0.$$

Diese Regel ergibt sich sofort, indem man links $|x|$ oder $-|x|$ für x einsetzt und $|y|$ oder $-|y|$ für y , je nachdem ob die Zahlen positiv sind oder nicht. Ungleichungen für den Betrag und den Abstand diskutieren wir weiter unten. Weil schon die Definition des Betrags $|x|$ auf einer Fallunterscheidung beruht ($x \geq 0$ oder $x \leq 0$), *muss man beim Rechnen mit Absolutbeträgen sehr oft Fallunterscheidungen vornehmen*.

6) Wir sagen von drei reellen Zahlen x, y, z , dass $x \in \mathbb{R}$ **zwischen y und z liegt**, wenn $y \leq x \leq z$ oder $z \leq x \leq y$ gilt. Soll dabei noch Gleichheit $x = y$ oder $y = z$ ausgeschlossen werden, so nennen wir x *echt* (oder *strikt*) zwischen y und z liegend. Eine Teilmenge I von \mathbb{R} mit der Eigenschaft, dass I mit je zwei Zahlen y und z auch alle dazwischen liegenden Zahlen x enthält, heißt ein **Intervall**. Die Intervalle in \mathbb{R} sind wichtig, weil sie häufig als sinnvoller Definitionsbereich für (ökonomische) Funktionen verwendet werden. Wir beschreiben im Folgenden alle Intervalle, die es gibt:

Man unterscheidet beschränkte und unbeschränkte Intervalle. Jedes nichtleere beschränkte Intervall hat zwei (evtl. zusammenfallende) **Randpunkte** $a \leq b$ (auch *Endpunkte* des Intervalls genannt; a heißt auch sein *Anfangspunkt*) und enthält alle Zahlen x die echt zwischen a und b liegen; es ist bestimmt durch die Festlegung, ob die Randpunkte zu dem Intervall gehören sollen oder nicht. Es gibt also folgende Typen:

beschränkte Intervalle:	<i>offen</i>	$]a, b[= \{x \in \mathbb{R} : a < x < b\},$
	<i>halboffen</i>	$[a, b[= \{x \in \mathbb{R} : a \leq x < b\},$
	<i>halboffen</i>	$]a, b] = \{x \in \mathbb{R} : a < x \leq b\},$
	<i>abgeschlossen</i>	$[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}.$

Die abgeschlossenen beschränkten Intervalle $[a, b]$ nennt man auch *kompakt*. Der Fall $a = b$ ist dabei nicht ausgeschlossen; dann besteht $[a, b] = \{a\}$ aus einem einzigen Punkt a und die anderen Intervalle $]a, b[$, $]a, b]$, $[a, b[$ enthalten gar keine Punkte, sie beschreiben dann also das *leere Intervall* \emptyset . Der Abstand $|a - b| = b - a$ der Randpunkte wird die **Länge des Intervalls** genannt. (Statt der "gespiegelten" eckigen Klammern in obiger suggestiver Notation, die anzeigen, dass die betreffenden Endpunkte aus dem Intervall ausgeschlossen sind, werden oft auch runde Klammern geschrieben, also z.B. (a, b) statt $]a, b[$. Bei den offenen Intervallen (a, b) kann das aber zur Verwechslung mit dem Paar der Zahlen a und b führen, das üblicherweise auch mit (a, b) bezeichnet wird. Deshalb ziehen wir die Schreibweise mit den eckigen Klammern vor.)

Die unbeschränkten Intervalle haben keine endliche Länge und nur einen Endpunkt a oder b in \mathbb{R} oder gar keinen Endpunkt in \mathbb{R} ; die in \mathbb{R} nicht vorhandenen Endpunkte werden dabei mit dem Symbol $\infty = +\infty$ oder $-\infty$ bezeichnet. Es gibt folgende

halbbeschränkte Intervalle:	<i>offen</i>	$]a, \infty[= \{x \in \mathbb{R} : a < x\} = \mathbb{R}_{>a},$
	<i>offen</i>	$]-\infty, b[= \{x \in \mathbb{R} : x < b\} = \mathbb{R}_{<b},$
	<i>abgeschlossen</i>	$[a, \infty[= \{x \in \mathbb{R} : a \leq x\} = \mathbb{R}_{\geq a},$
	<i>abgeschlossen</i>	$]-\infty, b] = \{x \in \mathbb{R} : x \leq b\} = \mathbb{R}_{\leq b},$

und außerdem noch das beidseitig unbeschränkte Intervall

$$]-\infty, \infty[= \mathbb{R}. \quad \blacksquare$$

Bisher haben wir die Ordnung der reellen Zahlen diskutiert, ohne einen Bezug zu den Rechenoperationen in \mathbb{R} herzustellen. Aber natürlich muss man wissen, wie es sich mit dem Größenvergleich von Zahlen verhält, wenn man mit ihnen Rechenoperationen ausführt. Bei welchen Rechenoperationen, die man auf beide Seiten einer Ungleichung $x < y$ anwendet, bleibt die Ungleichung erhalten? (Das heißt natürlich, dass für die aus x, y durch die Operation hervorgegangenen neuen Zahlen \tilde{x}, \tilde{y} dieselbe Ungleichung $\tilde{x} < \tilde{y}$ gilt.) Bei welchen Rechenoperationen kehrt sich die Ungleichung um? (D.h. $\tilde{y} < \tilde{x}$.) Die Antwort auf solche Fragen geben die *Rechenregeln für Ungleichungen*, welche die Ordnung mit den Grundrechenarten verbinden. Eine verwandte Frage ist die *Monotoniediskussion für einen gegebenen Rechenterm* $T(x)$; wird der Term bei Vergrößerung von x größer oder kleiner? Es ist klar, dass derartige Fragen von fundamentaler Bedeutung für die mathematische Modellierung von Wirtschaftsvorgängen sind. Man stelle sich etwa vor, dass $T(x)$ den Gewinn aus einem Produktions- und Vermarktungsvorgang in Abhängigkeit von einem dafür relevanten Parameter x angibt (etwa die Menge der produzierten Einheiten oder der Marktpreis); dann will man natürlich wissen, ob der Gewinn größer oder kleiner wird, wenn man diesen Parameter in eine bestimmte Richtung verändert. Bei der Beantwortung derartiger Fragen helfen ebenfalls die Regeln für das Rechnen mit Ungleichungen, die wir nun besprechen.

DISKUSSION (Rechnen mit Ungleichungen):

1) Es gibt zwei einfache Grundregeln, welche die totale Ordnung auf \mathbb{R} mit Addition und Multiplikation verbinden und aus denen alle weiteren Rechenregeln für Ungleichungen folgen. Diese **Grundgesetze** sind:

$$\begin{aligned} a < b &\implies a + c < b + c, \\ a < b, \quad c > 0 &\implies a \cdot c < b \cdot c. \end{aligned}$$

In Worten: Eine strenge Ungleichung bleibt erhalten, wenn man zu beiden Seiten dieselbe Zahl addiert oder beide Seiten mit derselben positiven (!) Zahl multipliziert. Diese Regeln gelten für alle reellen Zahlen a, b, c (mit $c > 0$ in der zweiten Regel) und sind als Axiome, also nicht zu beweisende Grundannahmen anzusehen (die man in der Mathematik rechtfertigen kann). Sie gelten offenbar auch für schwache Ungleichungen $a \leq b$ links und entsprechend rechts und ebenso für Ungleichungen " $>$ " oder " \geq " im umgekehrten Sinne. Da Subtraktion einer Zahl die Addition der Gegenzahl ist, gilt die erste Regel analog für die Subtraktion. Außerdem ist mit c auch der Kehrwert $1/c$ positiv (s.u.), so dass die zweite Regel auch für Division mit einer positiven Zahl c gilt; denn das läuft ja auf Multiplikation mit $1/c$ hinaus. Wir können also folgende etwas allgemeinere **Grundgesetze** formulieren:

- Eine Ungleichung bleibt erhalten, wenn man auf beiden Seiten dieselbe Zahl addiert oder subtrahiert;
- Eine Ungleichung bleibt erhalten, wenn man beide Seiten mit derselben positiven (!) Zahl multipliziert oder dividiert.

Was ist bei Multiplikation mit einer negativen Zahl c ? Ist $a < b$, also $b - a > a - a = 0$, so folgt durch Multiplikation der Ungleichung $c < 0$ mit der positiven Zahl $b - a$ nun $cb - ca = c \cdot (b - a) < 0 \cdot (b - a) = 0$, also nach beidseitiger Addition von ca die *umgekehrte* Ungleichung $cb < ca$ bzw. $ac > bc$.

- Bei Multiplikation oder Division beider Seiten mit einer negativen Zahl kehrt sich dagegen die Ungleichung um.

Das heißt also, dass für $a < b$ (oder $a \leq b$) und $c < 0$ nun $ac > bc$ (bzw. $ac \geq bc$) gilt; die Art der Ungleichung (stark oder schwach) wird nicht verändert. **Warnung:** der häufigste Fehler beim Rechnen mit Ungleichungen besteht darin, dass beide Seiten mit einer negativen Zahl multipliziert werden, ohne die Ungleichung umzukehren. Das passiert insbesondere dann, wenn die negative Zahl ein "positiv aussehender" Term ist wie etwa " c " oder " $x + 1$ ". Ein *Beispiel*: Es ist falsch zu argumentieren, dass die Ungleichung $\frac{x}{x+1} \geq 1$ keine Lösung habe, weil nach Multiplikation mit $x+1$ die nicht erfüllbare Ungleichung $x \geq x + 1$ entstehe. So sieht die entstehende Ungleichung nur aus, wenn $x + 1 > 0$ ist; im Fall $x + 1 < 0$ entsteht aber die umgekehrte Ungleichung $x \leq x + 1$, die immer richtig ist! Also sind die Lösungen der Ungleichung genau die Zahlen $x < -1$. ($x = -1$ kommt nicht in Frage, weil dafür der Term auf der linken Seite der Ungleichung nicht definiert ist.) Fehler dieser Art beim Umformen einer Ungleichung durch beidseitige Multiplikation mit demselben Faktor vermeidet man durch sorgfältige **Fallunterscheidung**: Ist der Faktor positiv, oder ist er negativ, oder ist er Null? (Multiplikation mit Null zerstört strikte Ungleichungen und erhält schwache, weil dadurch ja die Gleichung " $0 = 0$ " entsteht.)

2) Wir ziehen einige einfache Folgerungen: Ist $a > 0$, so auch $a^2 = a \cdot a > a \cdot 0 = 0$; ist aber $a < 0$, so folgt auch $a^2 = a \cdot a > a \cdot 0 = 0$, weil sich die Ungleichung ja umkehrt.

- Quadrate von reellen Zahlen ungleich Null sind stets positiv.

(Das Umgekehrte gilt wegen der Existenz von Quadratwurzeln aus positiven Zahlen auch: Jede positive Zahl ist ein Quadrat.) Eine banale Folgerung, die wir der Vollständigkeit halber festhalten, ist

$$1 > 0 \quad \text{und} \quad -1 < 0;$$

letzteres, weil aus $-1 > 0$ der Widerspruch $0 = 1 + (-1) > 1 + 0 = 1$ folgte. Für $a > 0$ ist dann $-a = (-1) \cdot a < (-1) \cdot 0 = 0$ und $1/a > 0$, weil andernfalls der Widerspruch $1 = a \cdot \frac{1}{a} < a \cdot 0 = 0$ folgte. Für $a < 0$ ist entsprechend $-a > 0$ und $1/a < 0$.

- Die Gegenzahl $-a$ zu einer positiven Zahl a ist negativ, die Gegenzahl zu einer negativen Zahl ist positiv;
- das Reziproke einer positiven Zahl ist wieder positiv, das Reziproke einer negativen Zahl ist wieder negativ.

Eine weitere Folgerung aus 1) sind folgende **Vorzeichenregeln**:

- Der Bereich $\mathbb{R}_{>0}$ der positiven reellen Zahlen ist additiv und multiplikativ abgeschlossen; d.h. Summen und Produkte von positiven Zahlen sind wieder positiv.
- Summen von negativen Zahlen sind negativ; Produkte von negativen Zahlen sind positiv und Produkte von einer positiven mit einer negativen Zahl sind negativ.

Insbesondere sind alle Zahlen positiv, die man durch fortgesetzte Addition von 1 zu 1 erhält, also *sind alle natürlichen Zahlen positiv*. Dieser wohlbekannt Sachverhalt bedarf eigentlich keiner Begründung; die Herleitung sollte eher als Bestätigung dafür gesehen werden, dass die angegebenen Grundgesetze der Ordnung wirklich sinnvoll sind. Über die Summe einer positiven und einer negativen Zahl kann man natürlich keine generelle Aussage machen: Die Summe kann negativ, Null oder positiv sein; es kommt auf die absolute Größe der Summanden an.

Eine Konsequenz der obigen Überlegungen ist übrigens, dass es im Bereich \mathbb{C} der komplexen Zahlen keine totale Ordnung geben kann, für welche dieselben Grundgesetze wie in 1) gelten; denn dann müssten Quadrate komplexer Zahlen positiv sein. Aber man hat die komplexen Zahlen ja gerade eingeführt, um eine Quadratwurzel \mathfrak{i} aus -1 zur Verfügung zu haben, also ist $-1 = \mathfrak{i}^2$ ein Quadrat einer Zahl aus $\mathbb{C}_{\neq 0}$, kann aber nicht positiv sein. Das ist der Preis, den man für die Konstruktion eines Zahlbereichs zahlen muss, in dem -1 eine Quadratwurzel hat: Es kann in einem solchen Bereich keinen sinnvollen Größenvergleich für Zahlen geben. (Und deshalb ist \mathbb{C} auch für die Ökonomie uninteressant.)

3) Für den **Größenvergleich bei Summen** beliebig vieler Zahlen gilt folgendes:

- Eine Summe wird größer, wenn man mindestens einen Summanden vergrößert und die anderen nicht verkleinert.

Das folgt aus 1) sofort, indem man sukzessive jeweils genau einen Summanden vergrößert. Man kann das so formalisieren:

$$a_k \leq b_k \quad \text{für } k = 1 \dots n \quad \implies \quad a_1 + a_2 + \dots + a_n \leq b_1 + b_2 + \dots + b_n,$$

wobei rechts strenge Ungleichung " $<$ " gilt, wenn auch links in mindestens einem Fall $a_k < b_k$ war. Man sagt dazu: *Gleichsinnige Ungleichungen darf man addieren*. Insbesondere ist eine Summe von nichtnegativen Zahlen positiv, wenn mindestens ein Summand positiv ist, und gleich Null nur, wenn schon alle Summanden Null sind. Natürlich kann man aus gegensinnigen Ungleichungen $a_1 \leq b_1$ und $a_2 \geq b_2$ nichts über den Größenvergleich zwischen $a_1 + b_1$ und $a_2 + b_2$ folgern.

4) Für den **Größenvergleich bei Produkten** beliebig vieler positiver Zahlen gilt:

- *Ein Produkt positiver Zahlen wird größer, wenn man mindestens einen Faktor vergrößert und die anderen nicht verkleinert.*

Das folgt aus 1) und 3) unmittelbar, indem man jeweils genau einen Faktor vergrößert. Man sagt auch: *Gleichsinnige Ungleichungen zwischen positiven Zahlen darf man multiplizieren*, d.h.

$$0 < a_k \leq b_k \quad \text{für } k = 1 \dots n \quad \implies \quad a_1 \cdot a_2 \cdot \dots \cdot a_n \leq b_1 \cdot b_2 \cdot \dots \cdot b_n,$$

wobei rechts strenge Ungleichung “<” gilt, wenn $a_k < b_k$ ist für mindestens ein k . Bei gegensinnigen Ungleichungen $0 < a_1 \leq b_1$ und $a_2 \geq b_2 > 0$ kann man für den Größenvergleich der Produkt $a_1 b_1$ und $a_2 b_2$ natürlich nichts folgern. Auf die Positivitätsbedingung für die Faktoren kann man übrigens nicht verzichten: Z.B. ist $-2 \leq 1$ und $-1 \leq 1$, aber $(-2) \cdot (-1) = 2 > 1 = 1 \cdot 1$. Auch bleibt ein Produkt von nichtnegativen Zahlen Null, wenn man einen Faktor vergrößert und ein anderer Faktor noch Null ist; das Produkt wird also in dieser Situation nicht größer. Sind Produkte ohne Nullfaktoren zu vergleichen, die negative Faktoren enthalten, so muss man die negativen Faktoren zählen. Ist eine Anzahl gerade, die andere ungerade, so liegt ein Produkt mit negativem und eines mit positivem Wert vor, und der Vergleich ist klar. Andernfalls ersetzt man die negativen Faktoren durch ihre positiven Gegenzahlen (evtl. unter Umkehrung der Ungleichung). Maßgebend ist die aus 2) und 3) abgeleitete **Vorzeichenregel**:

- *Ein Produkt ohne Nullfaktoren ist positiv, wenn die Anzahl seiner negativen Faktoren gerade ist, und negativ, wenn die Anzahl seiner negativen Faktoren ungerade ist.*

5) Für den **Größenvergleich von Kehrwerten** gilt:

- *Zwischen den Kehrwerten von zwei positiven oder zwei negativen Zahlen besteht die entgegengesetzte Ungleichung wie zwischen diesen Zahlen.*

Das heißt also $\frac{1}{a} > \frac{1}{b}$ für $0 < a < b$. Zum Beweis multipliziert man die Ungleichung $a < b$ mit der gemäß 1) und 2) positiven Zahl $\frac{1}{ab}$ und erhält $\frac{1}{b} = \frac{a}{ab} < \frac{b}{ab} = \frac{1}{a}$. Das Argument gilt auch, wenn a, b beide negativ sind, weil auch dann $\frac{1}{ab} > 0$ ist. Aber wenn eine Zahl negativ, die andere positiv ist, also $a < 0 < b$, so besteht zwischen den Kehrwerten die Ungleichung $\frac{1}{a} < 0 < \frac{1}{b}$ in derselben Richtung, weil ja $\frac{1}{a}$ wie a negativ ist und $\frac{1}{b}$ wie b positiv. Bei Übergang zu Reziproken in einer Ungleichung ist also eine *Fallunterscheidung* erforderlich. Beispielsweise ist die Ungleichung $\frac{1}{x+1} > \frac{1}{x}$ äquivalent mit der umgekehrten Ungleichung für die Kehrwerte $x+1 < x$, wenn $x > 0$ oder $x < -1$ ist (weil dann x und $x+1$ beide positiv oder beide negativ sind), aber äquivalent mit der gleichsinnigen Ungleichung $x+1 > x$ für $-1 < x < 0$, und genau diese Zahlen x sind auch die Lösungen der Ungleichung.

6) Für den **Größenvergleich bei Brüchen** folgt aus 5), 1) und 2) nun sofort:

- *der Wert eines Bruchs ist positiv, genau wenn Zähler und Nenner beide positiv oder beide negativ sind;*
- *der Wert eines Bruchs mit positivem Nenner wird vergrößert, wenn der Zähler vergrößert wird;*
- *der Wert eines Bruchs mit positivem Zähler und Nenner wird vergrößert, wenn der Nenner verkleinert wird, aber noch positiv bleibt.*

Das alles gilt nicht nur, wenn Zähler und Nenner ganze Zahlen sind, sondern auch für beliebige reelle Zähler und Nenner. Was beim Vergrößern bzw. Verkleinern von Zähler und Nenner passiert, wenn diese nicht beide positiv sind, kann man sich dann sofort überlegen, indem man mit -1 multipliziert. Wichtig ist, dass der Nenner beim Verkleinern die Null nicht "überschreitet". Z.B. ist -1 kleiner als 2 , aber $\frac{-1}{-1} = -1$ nicht größer als $\frac{1}{2}$. Zum Größenvergleich von zwei Brüchen $\frac{a}{b}$, $\frac{c}{d}$ muss man diese *gleichnamig machen*, also in der Form $\frac{ad}{bd}$ und $\frac{bc}{bd}$ schreiben; danach braucht man nur noch die Zähler zu vergleichen und das Vorzeichen des Nenners zu beachten. Es gilt also:

$$\frac{a}{b} < \frac{c}{d} \iff \begin{cases} ad < bc & \text{im Fall } bd > 0, \\ ad > bc & \text{im Fall } bd < 0. \end{cases}$$

7) Beim **Größenvergleich von Potenzen** ist stets zu unterscheiden, ob die Basis größer oder kleiner als 1 ist (aber positiv) und ob der Exponent positiv oder negativ ist. Zunächst bemerken wir:

- Der Wert einer Potenz mit Basis > 1 ist größer als 1, wenn der Exponent positiv ist und kleiner als 1, wenn er negativ ist; bei Basen echt zwischen 0 und 1 verhält es sich umgekehrt;

$$a^s > 1 \quad \text{für} \quad \begin{cases} a > 1 \text{ und } s > 0 & \text{oder} \\ 0 < a < 1 \text{ und } s < 0; \end{cases}$$

$$a^s < 1 \quad \text{für} \quad \begin{cases} a > 1 \text{ und } s < 0 & \text{oder} \\ 0 < a < 1 \text{ und } s > 0; \end{cases}$$

Für $a > 1$, und $m \in \mathbb{N}$ folgt nämlich aus 4) $a^m > 1$ und $\sqrt[m]{a} > 1$ (sonst wäre ja $a = (\sqrt[m]{a})^m \leq 1$), also auch $a^{m/n} = \sqrt[n]{a^m} > 1$, das ist die erste Behauptung für rationale Exponenten > 0 , und für reelle Exponenten s folgt sie dann auch, da a^s zwischen a^r und a^t liegt für rationale r, t mit $0 < r < s < t$. Die Fälle mit $0 < a < 1$ oder $s < 0$ ergeben sich dann mit Übergang zu Reziproken gemäß 5), da $a^{-s} = (1/a)^s$ ist. Für Basen $a = 1$ oder Exponenten $s = 0$ (und $a > 0$) ist natürlich $a^s = 1$. Mit den Potenzgesetzen folgt jetzt sofort, indem wir $b^s = (\frac{b}{a})^s a^s$ und $a^t = a^{t-s} a^s$ schreiben:

- Der Wert einer Potenz mit positivem Exponenten vergrößert sich, wenn man die (positive) Basis vergrößert; bei negativem Exponenten verkleinert er sich dagegen;
- Der Wert einer Potenz mit Basis > 1 vergrößert sich, wenn man den Exponenten vergrößert; bei Basen echt zwischen 0 und 1 verkleinert er sich dagegen;

$$\begin{aligned} a^s < b^s & \quad \text{für } 0 < a < b \text{ und } s > 0, \\ a^s > b^s & \quad \text{für } 0 < a < b \text{ und } s < 0; \end{aligned}$$

$$\begin{aligned} a^s < a^t & \quad \text{für } s < t \text{ und } a > 1, \\ a^s > a^t & \quad \text{für } s < t \text{ und } 0 < a < 1. \end{aligned}$$

Wenn der Exponent Null ist, so passiert bei Basisvergrößerung natürlich gar nichts, ebensowenig bei Exponentenvergrößerung im Fall der Basis 1; die Potenz ist und bleibt 1.

Ein Spezialfall ist der **Größenvergleich bei Wurzeln** $\sqrt[n]{a} = a^{1/n}$ mit $n \in \mathbb{N}_{\geq 2}$. Die Wurzel wird größer, wenn man den Radikanden a in $\mathbb{R}_{\geq 0}$ vergrößert. Hält man andererseits $a > 1$ fest und vergrößert die Ordnung n der Wurzel, so wird ihr Wert kleiner, weil ja der Exponent $\frac{1}{n}$ kleiner wird; bei $0 < a < 1$ wird dagegen $\sqrt[n]{a}$ größer mit wachsendem n . (Bei $n \rightarrow \infty$ strebt die Folge $\sqrt[n]{a}$ übrigens in beiden Fällen gegen den Grenzwert 1.)

Wenn wir nur ganze Exponenten n betrachten, so sind Potenzen auch für negative Basen $-a < 0$ definiert. Am besten formt man diese mittels $(-a)^n = (-1)^n a^n$ in Potenzen von positiven Basen um, bevor man einen Größenvergleich vornimmt. Dabei ist zu unterscheiden, ob n gerade oder ungerade ist; denn für gerade n ist $a^n = (a^{n/2})^2$ stets positiv (wenn $a \neq 0$), für ungerade n aber hat a^n dasselbe Vorzeichen wie a (d.h. a und a^n sind beide positiv oder beide negativ, wenn $a \neq 0$). Eine Folgerung ist

$$\left. \begin{array}{l} a^n < b^n \iff a < b, \text{ wenn } a, b \in \mathbb{R} \text{ und } n \in \mathbb{N} \text{ ungerade;} \\ a^n < b^n \iff a < b, \text{ wenn } a, b \in \mathbb{R}_{\geq 0} \\ a^n < b^n \iff a > b, \text{ wenn } a, b \in \mathbb{R}_{\leq 0} \end{array} \right\} \text{ und } n \in \mathbb{N} \text{ gerade.}$$

Also *Vorsicht beim Größenvergleich von Potenzen mit negativen Basen!*

8) Für den **Größenvergleich von Logarithmen** zur gleichen Basis $a > 1$ folgt sofort:

- Der Wert eines Logarithmus zu einer Basis > 1 wird größer, wenn man den Numerus in $\mathbb{R}_{>0}$ vergrößert; der Wert ist positiv, genau wenn der Numerus größer als 1 ist und negativ, wenn der Numerus zwischen 0 und 1 liegt.

$$\log_a y < \log_a z \iff 0 < y < z, \text{ wenn } a > 1.$$

Das ist klar, weil aus $\log_a y \geq \log_a z$ ja $y = a^{\log_a y} \geq a^{\log_a z} = z$ folgt gemäß 7). Bei Basen zwischen 0 und 1 verhält sich alles umgekehrt wegen $\log_{1/a} y = -\log_a y$; aber mit solchen Basen arbeitet man besser nicht in der Logarithmenrechnung, wie gesagt. Der Vergleich von Logarithmen zu verschiedenen Basen a, b erfolgt mit der Umrechnungsformel $\log_b y \cdot \log_a b = \log_a y$ (wenn a, b positiv und $\neq 1$ sowie $y > 0$). ■

Für den Größenvergleich von Termen ist es wichtig, einige fundamentale Ungleichungen zu kennen, mit denen man kompliziertere Terme, deren Größe schwer zu beurteilen ist, durch einfachere Terme abschätzen kann, die für den betrachteten Bereich der Variablen stets einen kleineren Wert haben (Abschätzung nach unten) oder stets einen größeren (Abschätzung nach oben).

DISKUSSION (fundamentale Ungleichungen);

1) Die sog. **Bernoulli-Ungleichung** zum Abschätzen von Potenzen lautet:

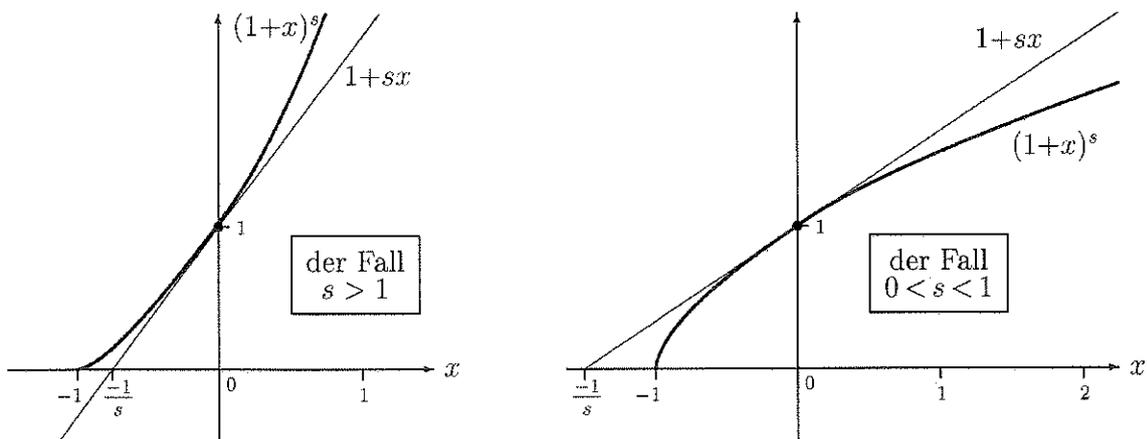
$$\left. \begin{array}{l} (1+x)^s > 1+sx, \text{ wenn } s > 1 \\ (1+x)^s < 1+sx, \text{ wenn } 0 < s < 1 \end{array} \right\} \text{ und } x \geq -1 \text{ mit } x \neq 0.$$

Die zweite Ungleichung mit Exponenten $0 < s < 1$ heißt auch *umgekehrte Bernoulli-Ungleichung*, weil sie im umgekehrter Richtung gilt wie die Bernoulli-Ungleichung mit Exponent $s > 1$.

Die Bernoulli-Ungleichung hat übrigens einen *ökonomischen Hintergrund*. Betrachten wir zunächst Exponenten $n \in \mathbb{N}_{\geq 2}$ und $x > 0$. Dann kann $1+x$ als Aufzinsungsfaktor für eine Zinsperiode aufgefasst werden, und $(1+x)^n$ ist der Aufzinsungsfaktor für n gleiche Perioden bei Zinsverzinsung, $1+nx$ der entsprechende Aufzinsungsfaktor bei einfacher Verzinsung. Also ist $(1+x)^n > 1+nx$ ökonomisch klar (und mathematisch auch: Multiplikation mit $1+x$ vergrößert $(1+x)^n$ um $x(1+x)^n > x$). Für $-1 \leq -x < 0$ kann man $1-x$ als vorschüssigen Abzinsungsfaktor auffassen, und dann ist die Ungleichung ökonomisch (und mathematisch) auch klar, weil Abzinsen für eine weitere Periode, also Multiplikation mit $1-x$ den Abzinsungsfaktor um $x(1-x)^n < x$ verkleinert. Damit ist die Bernoulli-Ungleichung für ganze Exponenten $n \geq 2$ schon bewiesen.

Für $n < m \in \mathbb{N}$ ist weiter ökonomisch klar, dass $(1 \pm \frac{1}{n}x)^n < (1 \pm \frac{1}{m}x)^m$ ist; denn der Auf- oder Abzinsungsfaktor wird größer, wenn man den Verzinsungszeitraum in mehr Zinsperioden einteilt. (Mathematisch sieht man das wie folgt: $(1 + \frac{x}{n+1})^{n+1}(1 + \frac{x}{n})^{-n} = (1 - \frac{1}{n+1} \frac{x}{n+x})^{n+1}(1 + \frac{x}{n}) > (1 - \frac{x}{n+x})(1 + \frac{x}{n}) = 1$ für $0 \neq x > -n$.) Wenn wir hier nun x durch mx ersetzen mit $0 \neq x > -1$ und die n -te Wurzel ziehen, so folgt $1 + \frac{m}{n}x < (1+x)^{m/n}$, und das ist die (offenbar auch noch für $x = -1$ gültige) Bernoulli-Ungleichung für rationale Exponenten $s = \frac{m}{n} > 1$. Für reelle Exponenten $s > 1$ folgt sie dann durch Approximation mit rationalen Exponenten (wobei man zunächst eine schwache Ungleichung erhält, mit einer kleinen Zusatzüberlegung aber dann auch wieder die strenge Ungleichung beweist), für Exponenten $0 < 1/s < 1$ ergibt sich schließlich die Ungleichung in umgekehrter Richtung, indem man in der Bernoulli-Ungleichung mit $s > 1$ zuerst x durch $\frac{1}{s}x$ ersetzt und dann beide Seiten mit Exponent $\frac{1}{s}$ potenziert.

Die Bernoulli-Ungleichung hat eine einfache geometrische Interpretation, die in der Differentialrechnung klar wird: Die Graphen von Potenzfunktionen mit Exponent $s > 1$ verlaufen oberhalb ihrer Tangenten, die von "Wurzelfunktionen", d.h. Potenzfunktionen mit Exponent $0 < s < 1$ verlaufen unter ihren Tangenten. Die Bernoulli-Ungleichung besagt das für die (um 1 nach links verschobenen) Potenzfunktion $(1+x)^s$ und die Graphentangente über dem Punkt $x = 0$. Mit $(a+x)^s = a^s(1 + \frac{x}{a})^s > a^s(1 + s\frac{x}{a})$ für $s > 1$ bzw. $< a^s(1 + s\frac{x}{a})$ für $0 < s < 1$ (sowie jeweils $a > 0$ und $0 \neq x \geq -a$) findet man aber, dass die Aussage auch für die Graphentangente über jedem anderen Punkt a gilt.



geometrische Interpretation der Bernoulli-Ungleichungen

2) Wenn wir in 1) den Grenzübergang zu kontinuierlicher Verzinsung zum Laufzeitanteiligen Zinsfuß x vornehmen, also $(1 + \frac{1}{n}x)^n$ bei $n \rightarrow \infty$ betrachten, so erhalten wir die **fundamentale Ungleichung für die natürliche Exponentialfunktion**:

$$e^x > 1 + x \quad \text{für alle } x \in \mathbb{R}_{\neq 0};$$

denn $(1 + \frac{1}{n}x)^n$ wächst mit wachsendem n , wie wir in 1) sahen (wenigstens für $n > -x$), hat als Grenzwert den Aufzinsungsfaktor bei kontinuierlicher Verzinsung e^x und ist gemäß Bernoulli-Ungleichung größer als $1 + n \cdot (\frac{1}{n}x) = 1 + x$. (Aus der Exponentialreihe $e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 \dots$ kann man auch $e^x > 1 + x$ ablesen, allerdings direkt nur für $x > 0$; für $-1 < x < 0$ muss man etwas genauer hinschauen, um die Ungleichung an der Reihendarstellung zu erkennen; für $x \leq -1$ ist die Ungleichung klar, weil dann $1+x < 0 < e^x$ gilt.) Ökonomisch interpretiert besagt die Ungleichung im Fall $x > 0$ einfach, dass der kontinuierliche Aufzinsungsfaktor zu gegebener Laufzeit und gegebenem Zinsfuß größer ist als der Aufzinsungsfaktor bei einfacher Verzinsung, und im Fall $-1 < x < 0$ entspricht sie einer entsprechenden Aussage für Abzinsungsfaktoren.

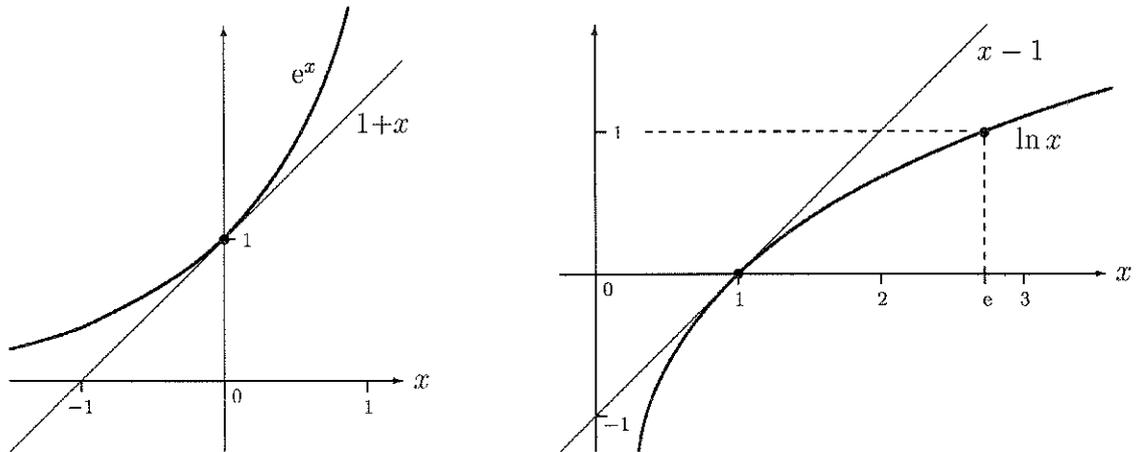
Setzen wir in der fundamentalen Ungleichung für die natürliche Exponentialfunktion $\ln x$ statt x ein, so erhalten wir die äquivalente **fundamentale Ungleichung für den natürlichen Logarithmus**:

$$\ln x < x - 1 \quad \text{für alle } x \in \mathbb{R}_{>0} \text{ mit } x \neq 1.$$

Eine Abschätzung für $\ln x$ in der anderen Richtung erhält man dann mit $\ln x = -\ln \frac{1}{x} > -\frac{1}{x} + 1$ und die Potenzen von e kann man mit $e^x = 1/e^{-x} < 1/(1-x)$, solange $x > -1$ ist (!), von oben abschätzen:

$$\ln x > 1 - \frac{1}{x} \quad \text{für } 1 \neq x \in \mathbb{R}_{>0}, \quad e^x < \frac{1}{1-x} \quad \text{für } 1 \neq x \in \mathbb{R}_{>-1}.$$

Die geometrische Interpretation der fundamentalen Ungleichungen für die Exponentialfunktion und die Logarithmusfunktion ist, dass der Graph der ersteren über und der Graph der letzteren unter seinen Tangenten verläuft. Dabei sind die Funktionen zur natürlichen Basis e dadurch ausgezeichnet, dass die Tangente im Punkt $(0, 1)$ bzw. im Punkt $(1, 0)$ gerade die Steigung 1 hat. Da \ln die Umkehrfunktion zu \exp ist, erhält man das Bild für \ln durch Spiegelung des Bildes für \exp an der Diagonalen der Koordinatenachsen.



geometrische Interpretation der Ungleichungen für Exponentialfunktion und Logarithmus

3) Eine unmittelbare Folgerung aus den Ungleichungen in 1) und 2) ist, dass Potenzen x^s mit positivem Exponenten s über alle Schranken wachsen, wenn x gegen Unendlich geht, und ebenso auch die Potenzen e^x (und allgemeiner a^x für $a > 1$). Man schreibt dafür $\lim_{x \rightarrow \infty} x^s = \infty$ bzw. $\lim_{x \rightarrow \infty} e^x = \infty$, was genau bedeutet, dass x^s und e^x jede beliebig groß vorgegebene Zahl S übertreffen, wenn nur x groß genug ist. Ebenso gilt auch $\lim_{x \rightarrow \infty} \ln x = \infty$; denn wenn x größer als e^S ist, so gilt ja $\ln x > S$. Es gibt aber große Unterschiede in der Geschwindigkeit, mit der die Potenzen von x , e^x , $\ln x$ groß werden:

- Die Potenz x^t strebt schneller gegen Unendlich als x^s bei $x \rightarrow \infty$, wenn der Exponent t größer ist als $s > 0$; und die Werte einer Exponentialfunktion a^x mit Basis $a > 1$ gehen schneller gegen Unendlich als die Potenz x^t (wie groß der Exponent t auch sei), die Werte der Logarithmusfunktion $\log_a x$ aber langsamer als jede Potenz x^s (mit noch so kleinem positiven Exponenten s).

In diesem Sinne wächst also die Folge n^2 schneller als die Folge der natürlichen Zahlen n , die Kubikzahlen wachsen noch schneller u.s.w., aber die geometrische Folge 2^n wächst noch sehr viel schneller an als alle diese Potenzfolgen. Auf der anderen Seite wächst die Folge \sqrt{n} langsamer als die Folge der natürlichen Zahlen, die Folge $\sqrt[3]{n}$ noch langsamer, und die Folge der Logarithmen $\log n$ wiederum noch sehr viel langsamer als alle Wurzelfolgen. (Das tausendste Glied der Logarithmenfolge ist 3, das millionste erst 6; die Folgenglieder werden aber mit wachsendem n doch beliebig groß!)

Um den Beweis der obigen Aussage über das Wachstum von Potenz- und Exponentialfolgen anzudeuten, die ja nicht evident ist, bemerken wir zunächst, dass der Quotient $x^t/x^s = x^{t-s}$ immer noch mit x gegen Unendlich geht, wenn $t > s$ ist, und dies ist gerade die Bedeutung der Aussage, dass x^t schneller als x^s gegen Unendlich geht. Für $a > 1$ und $m > t$ haben wir aus der fundamentalen Ungleichung für die Exponentialfunktion $a^x = e^{x \ln a} = (e^{(x/m) \ln a})^m > (1 + \frac{x}{m} \ln a)^m > (\frac{1}{m} \ln a)^m x^m$. Daher ist $a^x/x^t > (\frac{1}{m} \ln a)^m x^{m-t}$, und wegen $m > t$ und $\ln a > 0$ wird der letzte Ausdruck immer noch beliebig groß für hinreichend große x ; also geht a^x schneller gegen Unendlich als x^t . Auf der anderen Seite ist für $x > 1$ und $0 < \frac{1}{m} < s$ gemäß der fundamentalen Ungleichung für den Logarithmus $x^s/\ln x = \frac{1}{m} x^s / \ln x^{1/m} > \frac{1}{m} x^s / (x^{1/m} - 1) > \frac{1}{m} x^s / x^{1/m} = \frac{1}{m} x^{s-1/m}$, was wegen $s > \frac{1}{m}$ mit $x \rightarrow \infty$ beliebig groß wird; also geht x^s schneller gegen Unendlich als $\log_a x = \ln x / \ln a$.

4) Zwei oft verwendete Ungleichungen für Terme, in denen Absolutbeträge vorkommen, sind die

• **Dreiecksungleichung:** $|x \pm y| \leq |x| + |y|$

und die

• **umgekehrte Dreiecksungleichung:** $|x \pm y| \geq \left| |x| - |y| \right|$

für reelle Zahlen x, y . Man beweist sie — wie meistens beim Rechnen mit Absolutbeträgen — durch Fallunterscheidungen: Wenn x, y beide nichtnegativ oder beide nichtpositiv sind, so gilt $|x + y| = |x| + |y|$; wenn aber x positiv und y negativ ist, so gilt $|x| + |y| > x + y = |x + y|$, falls $x \geq -y$ ist, und $|x| + |y| > -(x + y) = |x + y|$, falls $x \leq -y$ ist. Die umgekehrte Dreiecksungleichung folgt dann mit $|x| = |(x - y) + y| \leq |x - y| + |y|$ und $|y| = |(y - x) + x| \leq |y - x| + |x|$, also $|x - y| \geq \max\{|x| - |y|, |y| - |x|\} = \left| |x| - |y| \right|$. Die Argumentation liefert auch Information über das *Eintreten der Gleichheit*: In der Ungleichung $|x + y| \leq |x| + |y|$ tritt Gleichheit ein, genau wenn x und y nicht entgegengesetztes Vorzeichen haben, d.h. wenn beide Zahlen nichtnegativ oder beide nichtpositiv sind (mit anderen Worten: 0 liegt nicht echt zwischen x und y). Genau dann gilt auch Gleichheit in der umgekehrten Ungleichung $|x - y| \geq \left| |x| - |y| \right|$.

Der Name “Dreiecksungleichung” wird verständlich, wenn man $|a - b|$ als Abstand der reellen Zahlen a und b auffasst und sie in der Form

$$|a - c| \leq |a - b| + |b - c|$$

schreibt (also oben $x := a - b$ und $y := b - c$ einsetzt). Die Ungleichung besagt dann, dass die Länge der Strecke von a nach c in dem Dreieck mit Ecken a, b, c höchstens so groß ist wie die Summe der Längen der beiden anderen Dreiecksseiten von a nach b und von b nach c . Und Gleichheit tritt genau dann ein, wenn b schon zwischen a und c liegt, so dass man auf dem Weg von a nach c gewissermaßen ohne Umweg über b kommt. Auf der Zahlengeraden \mathbb{R} gibt es freilich nur entartete Dreiecke, für die zwei Innenwinkel Null sind. Ganz genau dieselben Ungleichungen gelten aber auch für drei Punkte a, b, c in der Ebene \mathbb{R}^2 und den Euklidischen Abstand $|a - b| = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$ von Punkten $a = (a_1, a_2)$ und $b = (b_1, b_2)$, und von dieser geometrischen Situation wurde der Name “Dreiecksungleichung” auch auf die Zahlengerade übertragen. ■

Die Rechenregeln für Ungleichungen werden benutzt beim **Auflösen von Ungleichungen**. Dabei ist ähnlich wie bei Gleichungen eine zu erfüllende Ungleichung

$$T_{\text{links}}(x) < T_{\text{rechts}}(x)$$

vorgelegt mit Termen, die ein reelle Unbekannte x enthalten (und daneben noch Konstanten und evtl. Parameter; die Variable x muss auch nicht unbedingt in beiden Termen vorkommen), welche in einer gewissen Grundmenge in \mathbb{R} läuft (meist ist $\mathbb{R}_{>0}$ oder \mathbb{R} selbst diese Grundmenge). Statt des “Kleiner-Zeichens“ kann natürlich auch “>“ stehen (dann vertauschen nur die beiden Seiten der Ungleichung ihre Rollen) oder ein schwaches Ungleichungssymbol “ \leq “, “ \geq “ oder auch “ \neq “ (dann hat man zwei Ungleichungen zu betrachten, eine mit “<“ und eine mit “>“). Eine **Lösung der Ungleichung** ist ein Wert von x in der Grundmenge, für den die Ungleichung nach Einsetzen in die Terme gültig ist (insbesondere müssen die Werte der Terme für dieses x definiert sein), und die **Lösungsmenge der Ungleichung** ist die Menge aller Lösungen. Anders als bei Gleichungen, wo man genau eine oder endlich viele Lösungen erwartet, treten bei Ungleichungen typischerweise Intervalle oder Vereinigungen von mehreren Intervallen als Lösungsmenge auf.

Unter dem **Auflösen einer Ungleichung** versteht man ihre Umformung in eine Gestalt, bei der die Variable x auf einer Seite steht und auf der anderen nicht mehr vorkommt, also in die Gestalt $x < c$, $x \leq c$, $x > c$ oder $x \geq c$, aus der man die Lösungen direkt ablesen kann (eben alle x in der Grundmenge, welche kleiner, kleiner oder gleich, größer bzw. größer oder gleich der Konstanten c sind). Dabei benutzt man meist **Äquivalenzumformungen**, welche die Lösungsmenge der Ungleichung nicht verändern. (Zulässige Umformungen, welche die Lösungsmenge nicht verkleinern, aber evtl. vergrößern, werden bei Ungleichungen selten eingesetzt, weil sie die Lösungsmenge unter Umständen um ganze Intervalle vergrößern; man müsste dann beim “Rückwärtseinsetzen“ am Ende auch für alle Zahlen aus diesen Intervallen überprüfen, ob sie die ursprüngliche Ungleichung erfüllen.) Die Äquivalenzumformungen bei Ungleichungen beruhen auf den Rechenregeln für Ungleichungen, die wir oben besprochen haben, und sie erhalten im Allgemeinen die Art der Ungleichung (streng oder schwach). Äquivalenzumformungen sind zum Beispiel:

- Addition oder Subtraktion derselben Zahl oder desselben Terms auf beiden Seiten;
- Multiplikation beider Seiten mit derselben positiven Zahl oder mit demselben positiven Term $T(x)$; (*Achtung*: Die Multiplikation mit negativen Zahlen kehrt die Ungleichung um; daher ist eine *Fallunterscheidung* $T(x) > 0$ bzw. $T(x) < 0$ bzw. $T(x) = 0$ nötig!)
- Bildung des Kehrwerts beider Seiten der Ungleichung, wenn beide Seiten positiv oder beide negativ sind; (Wenn eine Seite positiv, die andere negativ ist, so ist die Gültigkeit oder Ungültigkeit der Ungleichung klar.)
- Potenzieren beider Seiten mit demselben positiven Exponenten, wenn beide Seiten der Ungleichung nichtnegativ sind;
- Exponentieren beider Seiten der Ungleichung zu derselben Basis > 1 ;
- Logarithmieren beider Seiten zu einer Basis > 1 , wenn beide Seiten der Ungleichung positiv sind.

Man erkennt die Notwendigkeit von **Fallunterscheidungen**. Das gilt insbesondere, wenn Absolutbeträge $|t(x)|$ von Termen in der Ungleichung (oder auch in einer Gleichung) vorkommen; denn durch die Fallunterscheidung $t(x) \geq 0$ bzw. $t(x) < 0$ kann man *die Absolutbetragssymbole beseitigen*; im ersten Fall ersetzt man einfach $|t(x)|$ durch $t(x)$, im zweiten durch $-t(x)$.

Eine **graphische Methode** zum Auflösen von Ungleichungen der Form $f(x) > c$ verläuft so: Man zeichnet den Graphen der Funktion $y = f(x)$ und bringt ihn mit der horizontalen Geraden mit Hochwert c zum Schnitt; die Lösungsmenge wird dann von den Intervallen auf der x -Achse gebildet, über denen der Graph echt oberhalb dieser Geraden verläuft. (Entsprechend bei Ungleichungen $f(x) \geq c$, $f(x) < c$, $f(x) \leq c$.)

BEISPIELE (zum Auflösen von Ungleichungen):

1) Die lineare Ungleichung

$$ax > b \quad (\text{oder } \geq, <, \leq)$$

mit $a \neq 0$ löst man so: $ax > b \iff x > \frac{b}{a}$ im Fall $a > 0$, aber $\iff x < \frac{b}{a}$ im Fall $a < 0$ (weil Multiplikation mit $\frac{1}{a} < 0$ die Ungleichung umkehrt!). Die Lösungsmenge ist also das Intervall $]\frac{b}{a}, \infty[$, wenn $a > 0$, und $]-\infty, \frac{b}{a}[$, wenn $a < 0$. Bei der schwachen Ungleichung $ax + b \geq 0$ erhält man zusätzlich noch $\frac{b}{a}$ als Lösung, also ist die Lösungsmenge $[\frac{b}{a}, \infty[$, wenn $a > 0$, und $]-\infty, \frac{b}{a}]$, wenn $a < 0$. Die Ungleichungen $ax < b$ hat als Lösungsmenge natürlich die komplementären Intervalle $]-\infty, \frac{b}{a}[$ im Fall $a > 0$ bzw. $[\frac{b}{a}, \infty[$ im Fall $a < 0$; bei $ax \leq b$ kommt der Intervallendpunkt $\frac{b}{a}$ zur Lösungsmenge hinzu.

Man kann eine lineare Ungleichung auch einfach so behandeln, dass man zunächst die Lösung x_0 der entsprechenden linearen Gleichung $ax = b$ bestimmt. Die Lösungsmenge ist dann ein halbbeschränktes Intervall mit Randpunkt x_0 , und x_0 gehört zur Lösungsmenge, wenn die Ungleichung schwach ist, aber x_0 gehört nicht dazu, wenn sie streng ist. Man braucht also nur noch zu festzustellen, ob sich das Lösungsintervall von x_0 aus nach $+\infty$ oder nach $-\infty$ erstreckt, und das kann man entscheiden, indem man prüft, ob die Ungleichung für große Werte von x erfüllt ist (bei $ax > b$ z.B., wenn $a > 0$, nicht aber, wenn $a < 0$). Die beschriebene graphische Methode führt auch sofort zum Ziel.

Auf eine lineare Ungleichung in der Ökonomie stößt man z.B. bei Rentenzahlung mit einfacher Verzinsung, wenn gefragt wird, welches Anfangskapital K_0 oder welcher Zinsfuß $p\%$ oder welche Ratenhöhe R erforderlich ist, damit der Kapitalendstand K_{Ziel} nach einer gegebenen Zahl m von Rentenperioden im gegebenen Zeitraum t (Jahre) eine bestimmte Zielgröße nicht unterschreitet (z.B. die Größe 0 bei Auszahlung von Renten aus einem Guthaben). Die bei Ratenzahlung am Periodenbeginn zuständige Rentenformel (siehe Kap. 2) führt bei dieser Fragestellung auf eine lineare Ungleichung für die Unbekannte K_0 oder p oder R :

$$K_0 - mR + \frac{p \cdot t}{100} \left(K_0 - \frac{m+1}{2} R \right) \geq K_{\text{Ziel}}.$$

Allerdings ist hier zu beachten, dass nur Lösungen in einer ökonomisch sinnvollen Grundmenge sinnvoll sind; z.B. sind alle Variablen in der Rentenformel nichtnegativ. Und obwohl die lineare Ungleichung mathematische Lösungen haben kann, so hat sie bei unsinnigen Zielvorgaben doch keine ökonomisch sinnvolle Lösung. Wenn wir z.B. im Falle der Auszahlung von Renten aus einem Guthaben K_0 eine Zielvorgabe $K_{\text{Ziel}} > (1 + \frac{1}{100}pt)K_0$ machen, so hat die Ungleichung nur negative Lösungen R und die Menge der ökonomisch sinnvollen Lösungen ist leer; denn mit Auszahlungen $R \geq 0$ ist die Vorgabe nicht zu erfüllen.

2) Gebrochen-lineare Ungleichungen der Form $\frac{ax+b}{cx+d} > z$ (oder $\geq, <, \leq$) kann man durch Multiplikation mit dem Nenner auf lineare Ungleichungen zurückführen. Doch **Achtung!** Man muss hier die *Fallunterscheidung* $cx + d > 0$ bzw. $cx + d < 0$ machen, weil Ungleichungen in verschiedener Richtung entstehen. Zum Beispiel geht

$$\frac{x}{2x-1} < \frac{1}{3}$$

durch Multiplikation mit $2x-1$ über in $x < \frac{1}{3}(2x-1)$ oder äquivalent $\frac{1}{3}x < -\frac{1}{3}$ mit den Lösungen $x < -1$; aber das gilt nur, wenn $2x-1 > 0$ ist, also $x > \frac{1}{2}$ und daher gibt es überhaupt keine Lösung mit $x > \frac{1}{2}$. Im Fall $2x-1 < 0$, also $x < \frac{1}{2}$, dagegen geht die umgeformte Ungleichung in die umgekehrte Richtung und lautet $x > \frac{1}{3}(2x-1)$ oder äquivalent $\frac{1}{3}x > -\frac{1}{3}$ mit den Lösungen $x > -1$; also sind in diesem Fall alle x mit $-1 < x < \frac{1}{2}$ Lösungen. Die Lösungsmenge der Ungleichung in \mathbb{R} ist also das Intervall $]-1, \frac{1}{2}[$. ($x = \frac{1}{2}$ ist natürlich keine Lösung, weil hierfür der Nenner Null ist.)

3) Auch in anderen Fällen, in denen eine Gleichung auf eine lineare Gleichung reduziert werden kann, läßt sich eine entsprechende Ungleichung auf eine oder mehrere lineare Ungleichungen zurückführen. Betrachten wir z.B.

$$\frac{x-2}{x-1} \leq \frac{x+2}{x+1}$$

Hier multipliziert man mit dem Faktor $(x-1)(x+1) = x^2-1$ und hat zwei Fälle zu unterscheiden. Im Fall 1 ist $x > 1$ oder $x < -1$, der Faktor also positiv; die Ungleichung ist dann äquivalent mit $(x-2)(x+1) \leq (x+2)(x-1)$ oder $x^2-x-2 \leq x^2+x-2$ oder $-x \leq x$; letzteres hat alle $x \geq 0$ als Lösungen, also erhalten wir im Fall 1 genau die Lösungen $x > 1$. Im Fall 2 ist $-1 < x < 1$ und der Faktor negativ; nach Multiplikation endet man also mit der umgekehrten Ungleichung $x^2-x-2 \geq x^2+x-2$ mit den Lösungen $x \leq 0$; im Fall 2 erhalten wir also genau die x mit $-1 < x \leq 0$ als Lösungen. Da $x = 1$ und $x = -1$ keine Lösungen sein können (weil dann Nenner in der ursprünglichen Ungleichung Null sind), ist also die Lösungsmenge eine Vereinigung von zwei Intervallen $]-1, 0] \cup]1, \infty[$.

4) Wenn Beträge in den Ungleichungen vorkommen, so kann man ganz genau wie oben vorgehen, nachdem man noch zusätzlich unterschieden hat, ob der Betrag von einer nicht-negativen oder von einer nichtpositiven Zahl gebildet wird, um so "die Betragsstriche zu eliminieren". Bei

$$|x-2| > 2 - |3-2x|$$

zum Beispiel haben wir vier Fälle $x \geq 2$, $x < 2$, $x \geq \frac{3}{2}$, und $x < \frac{3}{2}$ zu betrachten, was drei Fallunterscheidungen ergibt:

$$|x-2| > 2 - |3-2x| \iff \begin{cases} x-2 > 2 + (3-2x), & \text{wenn } x \geq 2, \\ -(x-2) > 2 + (3-2x), & \text{wenn } \frac{3}{2} \leq x < 2, \\ -(x-2) > 2 - (3-2x), & \text{wenn } x < \frac{3}{2}. \end{cases}$$

Im Fall $x \geq 2$ finden wir die Lösungen $x > \frac{7}{3}$, im Fall $\frac{3}{2} \leq x < 2$ keine Lösungen (weil nur $x > 3$ die lineare Ungleichung löst), im Fall $x < \frac{3}{2}$ die Lösungen $x < 1$. Die Lösungsmenge ist also $]-\infty, 1[\cup]\frac{7}{3}, \infty[$. Genau auf dieselbe Weise hätten wir übrigens vorgehen müssen, um die "Betragsgleichung" $|x-2| = 2 - |3-2x|$ zu behandeln, und hätten die (einzigen) Lösungen $x = 1$ und $x = \frac{7}{3}$ gefunden.

Analog wird folgende Ungleichung mit drei Fallunterscheidungen gelöst:

$$\left| \frac{x-2}{x+3} \right| \leq 2 \iff \begin{cases} x-2 \leq 2(x+3), & \text{wenn } x \geq 2, \\ -(x-2) \leq 2(x+3), & \text{wenn } -3 < x < 2, \\ -(x-2) \leq -2(x+3), & \text{wenn } x < -3. \end{cases}$$

Die oberste Ungleichung rechts wird erfüllt von $x \geq -8$, dieser Fall liefert also die Lösungen $x \geq 2$; die mittlere Ungleichung wird erfüllt von $x \geq -\frac{4}{3}$ und liefert alle x mit $-\frac{4}{3} \leq x < 2$ als Lösungen; die untere Ungleichung wird erfüllt von $x < -8$ und diese x sind auch Lösungen in diesem Fall. $x = -3$ ist keine Lösung, da hierfür der Nenner Null wird. Die Lösungsmenge ist damit $]-\infty, -8] \cup [-\frac{4}{3}, 2[\cup [2, \infty[=]-\infty, -8] \cup [-\frac{4}{3}, \infty[$.

4) Wie man quadratische Gleichungen mit quadratischer Ergänzung behandelt, so lassen sich auch quadratische Ungleichungen diskutieren. Zunächst hat die *spezielle quadratische Ungleichung* $x^2 < d^2$ die Lösungsmenge $] -d, d[$, wenn $d \geq 0$ ist; denn diese Ungleichung ist ja äquivalent mit $|x| = \sqrt{x^2} < d$. Entsprechend hat die Ungleichung $x^2 > d^2$ die Lösungsmenge $] -\infty, -d[\cup] d, \infty[$. Bei schwacher Ungleichung sind $x = d$ und $x = -d$ auch noch Lösungen. Die allgemeine **quadratische Ungleichung** lässt sich mit quadratischer Ergänzung und Division durch den führenden Koeffizienten $a \neq 0$ (Vorzeichen von a beachten!) auf den speziellen Fall zurückführen:

$$ax^2 + bx + c > 0 \iff \left(x + \frac{b}{2a}\right)^2 \begin{cases} > \frac{1}{4a^2}(b^2 - 4ac), & \text{wenn } a > 0, \\ < \frac{1}{4a^2}(b^2 - 4ac), & \text{wenn } a < 0. \end{cases}$$

Man liest ab: Im Fall $a > 0$ und $b^2 - 4ac < 0$ sind alle $x \in \mathbb{R}$ Lösungen, im Fall $a < 0$ und $b^2 - 4ac \leq 0$ gibt es keine Lösung, in den anderen Fällen ist die Lösungsmenge entweder das Intervall $]x_1, x_2[$ zwischen den beiden Nullstellen $x_1 = -\frac{b}{2a} - \frac{1}{2a}\sqrt{b^2 - 4ac}$ und $x_2 = -\frac{b}{2a} + \frac{1}{2a}\sqrt{b^2 - 4ac}$ der quadratischen Funktion $ax^2 + bx + c$ (nämlich wenn $a < 0$ ist), oder das Äußere $] -\infty, x_1[\cup] x_2, \infty[$ dieses Intervalls (nämlich wenn $a > 0$ ist; dabei ist auch $x_1 = x_2$ möglich, also Lösungsmenge $\mathbb{R}_{\neq x_1}$). Bei schwacher Ungleichung $ax^2 + bx + c \geq 0$ kommen die beiden Nullstellen x_1, x_2 noch zur Lösungsmenge hinzu.

Wenn die quadratische Funktion $ax^2 + bx + c$ reelle Nullstellen $x_1 \leq x_2$ besitzt, so kann man die Lösungen der quadratischen Ungleichung $ax^2 + bx + c > 0$ (oder ≥ 0) auch einfach durch **Zerlegung in Linearfaktoren**

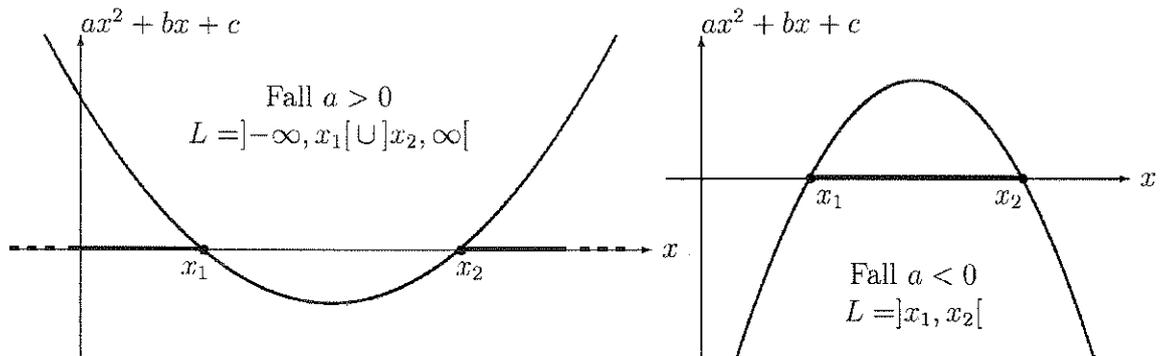
$$ax^2 + bx + c = a \cdot (x - x_1) \cdot (x - x_2)$$

ablesen. Das Produkt ist nämlich positiv, genau wenn alle drei Faktoren positiv sind oder wenn zwei der Faktoren $a, x - x_1, x - x_2$ negativ sind und der dritte positiv. Da die Faktoren $x - x_1$ bzw. $x - x_2$ ihr Vorzeichen nicht ändern, wenn die Variable x im Intervall zwischen den Nullstellen x_1, x_2 oder links bzw. rechts von x_1, x_2 läuft, ist dann sofort klar, dass die Lösungsmenge sich aus beschränkten oder unbeschränkten Intervallen zusammensetzt, deren Endpunkte Nullstellen sind. Bei der strikten Ungleichung $ax^2 + bx + c > 0$ gehören die Nullstellen selbst nicht zur Lösungsmenge, also handelt es sich um die Intervalle $] -\infty, x_1[$, $] x_1, x_2[$, $] x_2, \infty[$, und da $ax^2 + bx + c$ für große Werte von $|x|$ dasselbe Vorzeichen wie a hat, ist die Lösungsmenge die Vereinigung der beiden unbeschränkten Intervalle, wenn $a > 0$, bzw. das beschränkte Intervall $] x_1, x_2[$, wenn $a < 0$. (Dabei kann auch $x_1 = x_2$ sein; dann ist die Lösungsmenge $\mathbb{R}_{\neq x_1}$ bzw. leer.) Die Lösungsmenge der schwachen Ungleichung $ax^2 + bx + c \geq 0$ entsteht natürlich aus der Lösungsmenge der strengen Ungleichung durch Hinzunahme der Nullstellen x_1, x_2 .

Wir halten das Ergebnis fest, wobei wir uns auf Ungleichungen der Form $\dots > 0$ bzw. $\dots \geq 0$ beschränken, die man ja immer herstellen kann:

- *Hat die quadratische Funktion $ax^2 + bx + c$ die Nullstellen $x_1 \leq x_2$ in \mathbb{R} , so ist die Lösungsmenge der strengen quadratischen Ungleichung $ax^2 + bx + c > 0$ das offene Intervall $] x_1, x_2[$ zwischen den Nullstellen, wenn $a < 0$, bzw. das Äußere $] -\infty, x_1[\cup] x_2, \infty[$ dieses Intervalls, wenn $a > 0$. Hat die quadratische Funktion keine reelle Nullstelle, so ist die Lösungsmenge leer, wenn $a < 0$, und sie ist ganz \mathbb{R} , wenn $a > 0$. Die Lösungsmenge der schwachen Ungleichung $ax^2 + bx + c \geq 0$ ergibt sich jeweils durch Hinzunahme der Nullstellen.*

Mit der beschriebenen graphischen Methode sieht man das auch leicht ein. Der Graph der quadratischen Funktion $ax^2 + bx + c$ ist ja eine Parabel, die nach oben geöffnet ist im Fall $a > 0$ bzw. nach unten geöffnet im Fall $a < 0$. Wenn es Nullstellen $x_1 \leq x_2$ gibt, so sind diese Punkte auf der x -Achse die Schnittpunkte der Achse mit der Parabel, und die Lösungsmenge der strengen Ungleichung $ax^2 + bx + c > 0$ besteht aus den Intervallen auf der x -Achse, über denen die Parabel oberhalb der Achse verläuft.



die Lösungsmenge L der quadratischen Ungleichung $ax^2 + bx + c > 0$

Wenn es keine reellen Nullstellen gibt, d.h. wenn die Parabel die x -Achse nicht schneidet, so sind alle $x \in \mathbb{R}$ Lösungen ($a > 0$, die Parabel verläuft ganz oberhalb der x -Achse), oder die Ungleichung hat gar keine Lösung ($a < 0$, die Parabel verläuft ganz unterhalb der Achse).

5) Auf eine quadratische Ungleichung stößt man z.B. bei der *Laufzeitberechnung für Renten bei einfacher Verzinsung* bis zum Erreichen eines vorgegebenen Zielwerts K_{Ziel} (z.B. $K_{\text{Ziel}} = 0$). Ist die Laufzeit $t = m\tau$ (Jahre) in m Ratenperioden der Dauer τ eingeteilt, so lautet die oben schon angegebene Ungleichung

$$K_0 - mR + \frac{p \cdot m \cdot \tau}{100} \left(K_0 - \frac{m+1}{2} R \right) \geq K_{\text{Ziel}}.$$

Dies ist eine quadratische Ungleichung für m , und gesucht ist hier die kleinste nichtnegative ganze Lösung. Ist $K_{\text{Ziel}} < K_0$, so ist die Ungleichung für $m = 0$ gültig; für große m gilt sie aber offenbar nicht. Daher gibt es zwei Lösungen $m_1 < 0$ und $m_2 > 0$ der entsprechenden quadratischen Gleichung für m , und die Lösungsmenge der Ungleichung ist $]m_1, m_2[$. Die Lösung unseres Problems ist daher $m = \lfloor m_2 \rfloor$, also die größte ganze Zahl, die kleiner oder gleich m_2 ist.

6) Auf eine quadratische Ungleichung führt z.B. auch

$$x + 1 \geq \frac{2}{x}.$$

Hier wird man natürlich mit x multiplizieren und erhält die quadratische Ungleichung $x^2 + x \geq 2$ — aber nur im Fall $x > 0$! Im anderen Fall $x < 0$ ergibt sich gerade die umgekehrte quadratische Ungleichung $x^2 + x \leq 2$! Aus $x^2 + x - 2 = (x-1)(x+2)$ entnimmt man, dass $x^2 + x - 2 \geq 0$ die Zahlen x mit $x \leq -2$ oder $x \geq 1$ als Lösungen hat, das ergibt im Fall $x > 0$ also die Lösungen $x \geq 1$. Die quadratische Ungleichung $-x^2 - x + 2 \geq 0$ hat dagegen die Lösungsmenge $[-2, 1]$, und das gibt die Lösungen $-2 \leq x < 0$ im Fall $x < 0$. Die Lösungsmenge der obigen Ungleichung ist damit $[-2, 0[\cup [1, \infty[$.

Ohne die sorgfältige *Fallunterscheidung* hätte man die Lösungen im Intervall $[-2, 0[$ leicht übersehen!

Ähnlich führt eine Ungleichung wie

$$\frac{x-2}{x+2} < \frac{2x-1}{x+1}$$

auf verschiedene quadratische Ungleichungen, wenn man mit den Nennern multipliziert und dabei Fälle entsprechend den Vorzeichen der Nenner unterscheidet. Für $-2 < x < -1$ gilt $(x+2)(x+1) < 0$, daher ist dann die Ungleichung äquivalent zu $(x-2)(x+1) > (2x-1)(x+2)$ bzw. $x^2 - x - 2 > 2x^2 + 3x - 2$, d.h. $x(x+4) < 0$, was für alle x mit $-2 < x < -1$ erfüllt ist. Für $x > -1$ oder $x < -2$ ist die Ungleichung dagegen äquivalent zu $x(x+4) > 0$, was genau für $x > 0$ oder für $x < -4$ erfüllt ist. Die Lösungsmenge ist daher die Vereinigung von drei Intervallen $]-\infty, -4[\cup]-2, -1[\cup]0, \infty[$.

Auch hier werden bei fehlender Sorgfalt bei der Fallunterscheidung die Lösungen zwischen -2 und -1 werden leicht übersehen!

7) Schließlich betrachten wir noch eine "Betragsungleichung"

$$|x-1| \cdot |x+1| > \frac{24}{25}$$

Die linke Seite können wir $|(x-1) \cdot (x+1)| = |x^2-1|$ umschreiben, daher ist die Ungleichung im Falle $x^2 \geq 1$ äquivalent zu $x^2 - 1 > \frac{24}{25}$ bzw. $x^2 > (\frac{7}{5})^2$ mit den Lösungen $x > \frac{7}{5}$ und $x < -\frac{7}{5}$; im Fall $x^2 < 1$ ist sie dagegen äquivalent zu $1 - x^2 > \frac{24}{25}$ bzw. $x^2 < (\frac{1}{5})^2$ mit den Lösungen $-\frac{1}{5} < x < \frac{1}{5}$. Die Lösungsmenge ist also $]-\infty, -\frac{7}{5}[\cup]-\frac{1}{5}, \frac{1}{5}[\cup]\frac{7}{5}, \infty[$. Es ist klar, dass mit x auch $-x$ eine Lösung der ursprünglichen Ungleichung ist; daher muss die Lösungsmenge in \mathbb{R} symmetrisch bzgl. des Nullpunkts sein.

8) Die Methode der **Zerlegung in Faktoren** funktioniert auch bei algebraischen Ungleichungen $a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 > 0$ vom Grad $n > 2$, wenn man alle reellen Nullstellen $x_1 \leq x_2 \leq \dots \leq x_m$ kennt. Man kann dann nämlich zerlegen

$$a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 = a_n \cdot (x - x_1) \cdot (x - x_2) \cdot \dots \cdot (x - x_m) \cdot q(x)$$

mit $q(x) = x^{n-m} + \dots + b_1 x + b_0 > 0$ für alle $x \in \mathbb{R}$. Die Lösungsmenge der Ungleichung besteht somit aus den $x \in \mathbb{R}$, für die eine gerade Anzahl der Linearfaktoren $x - x_j$ negativ ist, wenn $a_n > 0$, bzw. eine ungerade Anzahl, wenn $a_n < 0$. So hat z.B.

$$x^3 - x^2 + x - 1 < 0 \quad \text{bzw.} \quad x^3 - x^2 - x + 1 \leq 0$$

die Faktorzerlegung $(x-1)(x^2+1) < 0$ bzw. $(x+1)(x-1)^2 \leq 0$ und damit die Lösungsmenge $]-\infty, 1[$ bzw. $]-\infty, -1[\cup \{1\}$. ■

Mit den Rechenregeln für Ungleichungen, die wir angegeben haben, kann man auch noch andere Aufgaben erledigen als die Bestimmung der Lösungsmenge von Ungleichungen. Eine solche Aufgabe ist z.B. die **Anordnung verschiedener Terme der Größe nach**. Hier ist gemeint, dass verschiedene reelle "Rechenterte" $T_1(a, b, \dots)$, $T_2(a, b, \dots)$, ... gegeben sind, die von Parametern a, b, \dots mit bekanntem Größenvergleich abhängen (z.B. $a < b < \dots$), und die Aufgabe ist, diese Terme der Größe nach zu vergleichen, also z.B. in eine aufsteigende Reihenfolge zu bringen. Beispiele dazu findet man in 1) der nächsten Beispielserie.

Eine verwandte Aufgabenstellung liegt vor bei der **Monotonieuntersuchung eines Terms** $T(x)$. Hier handelt es sich darum, zu entscheiden, ob der Term bei Vergrößerung von x (in einer gewissen Grundmenge in \mathbb{R}) größer wird oder kleiner (oder gleich bleibt). Beispiele hierzu bringen wir ebenfalls in der nächsten Serie. In einfachen Fällen kann man die Monotonieuntersuchung eines Terms direkt durch Inspektion und Anwendung der behandelten Rechenregeln für Ungleichungen erledigen (z.B. wird eine Potenz mit Basis > 1 größer, wenn der Exponent vergrößert wird, und ein Bruch mit positivem Zähler und Nenner wird kleiner, wenn man den Zähler verkleinert oder den Nenner vergrößert). In komplizierteren Fällen muss man untersuchen, ob aus $x < y$ die Ungleichung $T(x) < T(y)$ folgt (oder umgekehrt), was evtl. durch Auflösen dieser Ungleichung nach x entschieden werden kann. Man kann auch versuchen, aus $T(x) \geq T(y)$ durch Umformung die Ungleichung $x \geq y$ herzuleiten; für $x < y$ folgt dann $T(x) < T(y)$ (weil die Alternative $T(x) \geq T(y)$ ja ausscheidet).

BEISPIELE (*Größenvergleich von Termen, Monotonieuntersuchung bei Termen*):

1) Für gegebene Zahlen $0 < a < 1 < b < 1/a$ sollen die Ausdrücke

$$1, a, b, a^a, a^b, b^a, b^b, b^{-a}, \sqrt[a]{b}, \sqrt[b]{a}, \sqrt[a]{a}, \sqrt[b]{b}, \frac{1}{\sqrt[a]{a}}$$

der Größe nach angeordnet werden. Hier schreibt man am besten alle Terme als Potenzen der Basen a und b um. Weil mit $1 < b < 1/a$ auch $a < 1/b < 1$ ist, weiß man über die dabei auftretenden Exponenten $0 < a < 1/b < 1 < b < 1/a$. Da Potenzen mit der Basis $b > 1$ bei Vergrößerung des Exponenten größer werden, während Potenzen der Basis $0 < a < 1$ dabei kleiner werden, haben wir sofort folgende Anordnung der Größe nach:

$$a^{1/a} < a^b < a = a^1 < a^{1/b} < a^a < a^0 = 1 = b^0 < b^a < b^{1/b} < b = b^1 < b^b < b^{1/a}.$$

Jetzt sind nur noch die zwei Terme b^{-a} und $1/\sqrt[a]{a}$ einzuordnen. Weil $a > 0$ und $a < 1/b < 1$ ist, liegt $b^{-a} = (1/b)^a$ zwischen a^a und $1^a = 1$ und wegen $1/a > 0$ und $0 < b < 1/a$ ist $1/\sqrt[a]{a} = (1/a)^{1/a} > b^{1/a}$, also lautet die Anordnung aller Terme der Größe nach

$$a^{1/a} < a^b < a < a^{1/b} < a^a < b^{-a} < 1 < b^a < b^{1/b} < b < b^b < b^{1/a} < \frac{1}{\sqrt[a]{a}}.$$

2) Der Term

$$\frac{1}{1 + \frac{1}{1+x}}$$

soll im Bereich $\mathbb{R}_{>0}$ auf sein Monotonieverhalten untersucht werden. Die Antwort kann man direkt ablesen: Bei Vergrößerung von $x > 0$ wird $1+x$ größer, also der Kehrwert $\frac{1}{1+x}$ und damit $1 + \frac{1}{1+x}$ kleiner, also der Kehrwert hierzu wieder größer — und das ist gerade der zu untersuchende Term.

Diese Argumentation war gültig, weil für $x > 0$ alle auftretenden Nenner positiv sind, so dass es keine Probleme gibt. Die Argumentation gilt sogar für alle $x > -1$. Für $x = -1$ und für $x = -2$ ist der Term nicht definiert, da ein Nenner Null wird. Mit einer ähnlichen Diskussion unter Berücksichtigung des Vorzeichens von $\frac{1}{1+x}$ und von $1 + \frac{1}{1+x}$ kann man sehen, dass der Term auf dem Intervall $]-\infty, -2[$ ebenfalls größer wird bei Vergrößerung von x , auf dem Intervall $]-2, -1[$ aber ist das Monotonieverhalten anders: Der Term wird kleiner, wenn man x in diesem Intervall vergrößert.

3) So einfach wie in 2) ist die Monotoniediskussion meistens nicht, weil man *entgegengesetzte Monotonietendenzen* in dem zu beurteilenden Term hat. Betrachten wir z.B.

$$\frac{1}{x} - \frac{1}{x+1}, \quad \frac{x-1}{x+1}, \quad \frac{e^x}{1+e^x}, \quad \frac{\sqrt{1+x}}{x}.$$

In der ersten Differenz nehmen *beide* Terme ab, wenn man $x > 0$ vergrößert, in den drei Brüchen nehmen Zähler *und* Nenner zu, wenn man $x > 0$ vergrößert, also kann man das Monotonieverhalten des gesamten Terms nicht direkt erkennen — es liegen eben entgegengesetzte Monotonietendenzen vor, und dann kommt es darauf an, welche Tendenz überwiegt. Der Ausweg ist hier *Termumformung* in eine Gestalt, aus der man das Monotonieverhalten ablesen kann. Der erste Term geht z.B. durch Gleichnamig-Machen über in $\frac{x+1-x}{x(x+1)} = \frac{1}{x(x+1)}$, und an dieser Form des Terms ist alles klar: Der Nenner wird größer, wenn $x > 0$ vergrößert wird, also wird der Term insgesamt dabei kleiner. Beim zweiten Term liegt eine Umformung nahe, mit der "x" aus dem Zähler verschwindet. Das erreicht man durch

$$\frac{x-1}{x+1} = \frac{x+1-2}{x+1} = \frac{x+1}{x+1} - \frac{2}{x+1} = 1 - \frac{2}{x+1}.$$

Hier nimmt $\frac{2}{x+1}$ ab bei Vergrößerung von $x > 0$, also wird $1 - \frac{2}{x+1}$ größer bei Vergrößerung von x , und das gilt sogar für x aus $\mathbb{R}_{>-1}$. Beim dritten Term ist es vielleicht am einfachsten, e^x zu kürzen, dann entsteht $1/(e^{-x} + 1)$. Hier nimmt der Nenner ab bei Vergrößerung von $x \in \mathbb{R}$, also wird der Term größer. (Da e^{-x} mit wachsendem x sehr schnell klein wird, strebt der Term sogar sehr schnell von unten dem Grenzwert 1 zu. Mit Funktionen dieser Bauart modelliert man in der Ökonomie wachsende Größen, die einen endlichen "Sättigungswert erreichen", wenn die unabhängige Variable gegen Unendlich geht; sie heißen *logistische Funktionen*.) Der letzte Term geht mit Erweitern durch den Faktor $\sqrt{1+x}$ über in $\frac{1+x}{x\sqrt{1+x}} = \frac{1}{x\sqrt{1+x}} + \frac{1}{\sqrt{1+x}}$. Hier haben wir nun eine Summe von zwei Termen, die beide kleiner werden bei Vergrößerung von $x > 0$, also wird auch die Summe kleiner. (Andere Möglichkeit: $\frac{\sqrt{1+x}}{x} = \sqrt{\frac{1+x}{x^2}} = \sqrt{\frac{1}{x^2} + \frac{1}{x}}$.) Übrigens modelliert $\frac{\sqrt{1+x}}{x}$ eine typische *Stückkostenfunktion* in Abhängigkeit von der produzierten Menge x eines Gutes.

4) Wer solche trickreichen Termumformungen wie in 3) nicht findet (es gibt auch nicht immer welche, die zum Ziel führen, also das Monotonieverhalten des Terms direkt zeigen), der ist auf den Vergleich der Terme $T(x)$, $T(y)$ für Werte $x < y$ angewiesen. Für den zweiten Term in 3) und $x, y \in \mathbb{R}_{>-1}$ z.B. geht das so:

$$\begin{aligned} \frac{x-1}{x+1} < \frac{y-1}{y+1} &\iff (y+1)(x-1) < (x+1)(y-1) \\ \iff xy + x - y - 1 < xy + y - x - 1 &\iff 2x < 2y \iff x < y, \end{aligned}$$

also wächst der Term mit wachsendem $x > -1$. (Dasselbe Ergebnis hätten wir erhalten, wenn wir überall mit ">" statt "<" argumentiert hätten.) Für den vierten Term in 3) verläuft die Rechnung wie folgt, wenn $x, y > 0$ sind:

$$\begin{aligned} \frac{\sqrt{1+x}}{x} < \frac{\sqrt{1+y}}{y} &\iff y\sqrt{1+x} < x\sqrt{1+y} \\ \iff y^2(1+x) < x^2(1+y) &\iff x^2 - y^2 + yx^2 - xy^2 > 0 \\ \iff (x-y)(x+y+xy) > 0 &\iff x-y > 0, \end{aligned}$$

also wird dieser Term kleiner, wenn man $x > 0$ vergrößert (weil die letzte Ungleichung ja $y < x$ ist; in der Rechnung haben wir benutzt, dass Quadrieren beider Seiten einer Ungleichung eine Äquivalenzumformung ist, wenn beide Seiten positiv sind, und dass $x+y+xy$ positiv ist für positive x, y).

5) Für **quadratische Terme** ist die Monotoniediskussion mit quadratischer Ergänzung einfach:

$$ax^2 + bx + c = a\left(x + \frac{b}{2a}\right)^2 + c - \frac{b^2}{4a}$$

wächst mit zunehmendem $x \geq -\frac{b}{2a}$ und mit abnehmendem $x \leq -\frac{b}{2a}$, wenn $a > 0$ ist, und fällt bei solchen Veränderungen von x , wenn $a < 0$ ist. Das ergibt sich einfach daraus, dass y^2 mit wachsendem $y \geq 0$ zunimmt, bei Vergrößerung von y im Bereich $\mathbb{R}_{\leq 0}$ aber abnimmt. Wie immer bei quadratischen Funktionen gibt also die Methode der quadratischen Ergänzung auch hier sofort die Antwort auf die gestellte Frage. Geometrisch interpretiert besagt das Ergebnis, dass für eine nach oben geöffnete Parabel ($a > 0$) der rechte Ast ab dem Scheitelpunkt aufsteigend und der linke Ast bis zum Scheitelpunkt absteigend ist, während es sich bei einer nach unten geöffneten Parabel ($a < 0$) gerade umgekehrt verhält.

6) Zum Schluss noch eine schwer zu knackende "Nuss":

$$\sqrt{x+h} - \sqrt{x}$$

soll für festes $h > 0$ auf sein Monotonieverhalten im Bereich $x > 0$ untersucht werden. Es handelt sich um eine Differenz von zwei wachsenden Termen, also ist das Monotonieverhalten nicht unmittelbar klar. Der Vergleich von $\sqrt{x+h} - \sqrt{x}$ mit $\sqrt{y+h} - \sqrt{y}$ wie in 4) führt hier auch zu nichts. (Man kann die Ungleichung $\sqrt{x+h} - \sqrt{x} < \sqrt{y+h} - \sqrt{y}$ zwar durch zweifaches Quadrieren in eine quadratische Ungleichung für x umrechnen, doch wird das sehr unübersichtlich.) Aber eine geschickte Termumformung hilft auch hier: Man fasst den Term als Bruch mit Nenner 1 auf und erweitert dann mit $\sqrt{x+h} + \sqrt{x}$. Dann entsteht im Zähler $(\sqrt{x+h})^2 - (\sqrt{x})^2 = x+h-x = h$ gemäß einer binomischen Formel, und der Term ist daher

$$= \frac{h}{\sqrt{x+h} + \sqrt{x}}.$$

Hier sind nun im Nenner beide Summanden wachsend mit $x > 0$, also wird der Term kleiner bei Vergrößerung von $x > 0$. ■

In der *Differentialrechnung* werden wir später eine sehr effektive Methode zur Feststellung des Monotonieverhaltens einer gegebenen reellen Funktion $f(x)$ einer reellen Variablen x behandeln: Man braucht nur die Ableitung $f'(x)$ zu berechnen und das Vorzeichen der Ableitungswerte $f'(x)$ zu untersuchen (was oft eine einfache Sache ist, aber auch nicht immer). Auf Intervallen, auf denen die Ableitung nur positive Werte hat, ist die Funktion $f(x)$ wachsend, auf Intervallen, auf denen die Ableitung nur negative Werte hat, ist die Funktion $f(x)$ fallend. Damit kann man dann auch Monotonieuntersuchungen wie bei dem letzten Term oben einfach und systematisch erledigen (denn die Ableitung dieser Funktion ist negativ auf $\mathbb{R}_{>0}$).

Mathematik für Wirtschaftswissenschaftler

(K. Steffen, Heinrich-Heine-Universität Düsseldorf, WS 2006/07)

1.5 Mittelwerte

Eine erste Information über einen Satz erhobener Daten — und oft auch die einzige darin enthaltene Information, die ausgenutzt wird — liefert die Bildung eines Mittelwerts zu diesen Daten. Auch und gerade bei der Beurteilung ökonomischer Daten spielen Mittelwerte eine wichtige Rolle. Nun gibt es verschiedene Mittelwertbildungen, die auch verschiedene Werte liefern. Daher muss man, um die Aussagekraft von ausgewiesenen Mittelwerten beurteilen zu können, etwas über verschiedene Mittelungsverfahren und über den Größenvergleich der Mittelwerte wissen, welche diese Verfahren auf derselben Datengrundlage berechnen. Darum geht es in diesem Abschnitt. Zunächst geben wir eine sehr allgemeine Definition von Mittelungsverfahren für einen endlichen Datensatz, d.h. für endlich viele gegebene reelle Zahlen.

DEFINITION: Unter einem **Mittelungsverfahren** (einer **Mittelwertbildung**) für n reelle Zahlen aus einem Intervall I in \mathbb{R} verstehen wir eine Vorschrift, die je n Zahlen x_1, x_2, \dots, x_n aus I eine reelle Zahl $M(x_1, x_2, \dots, x_n)$ zuordnet und die folgenden Bedingungen genügt:

- (i) $\min(x_1, x_2, \dots, x_n) \leq M(x_1, x_2, \dots, x_n) \leq \max(x_1, x_2, \dots, x_n)$;
- (ii) (*Monotonie*) Vergrößerung einer Zahl x_i in I bei Festhalten der Werte aller anderen Zahlen $x_j, j \neq i$, verkleinert den Wert von $M(x_1, x_2, \dots, x_n)$ nicht;
- (iii) (*Symmetrie*) Vertauschung der Zahlen x_j untereinander ändert den Wert von $M(x_1, x_2, \dots, x_n)$ nicht.

Die Zahl $M(x_1, x_2, \dots, x_n)$ heißt dann der **Mittelwert** von x_1, x_2, \dots, x_n bezüglich des Mittelungsverfahrens M . ■

Das ist leicht zu verstehen: Die erste Bedingung besagt einfach, dass $M(x_1, x_2, \dots, x_n)$ immer zwischen der größten und der kleinsten der zu mittelnden Zahlen liegen muss. Insbesondere liegt $M(x_1, x_2, \dots, x_n)$ dann auch in dem Intervall I , dem die zu mittelnden Zahlen entnommen wurden. Als Intervalle in den Anwendungen kommen hier übrigens hauptsächlich $I = \mathbb{R}$ oder $I = \mathbb{R}_{>0}$ in Frage. Die zweite Bedingung bedeutet, dass der Wert $M(x_1, x_2, \dots, x_n)$ größer wird, oder jedenfalls nicht kleiner, wenn man eine der zu mittelnden Zahlen vergrößert. (Bei gewissen in der Praxis durchaus verwendeten Mittelungsverfahren braucht der Wert nicht echt größer zu werden.) Die dritte Bedingung schließlich besagt, dass der Wert $M(x_1, x_2, \dots, x_n)$ von der Reihenfolge der Zahlen x_1, x_2, \dots, x_n nicht abhängen soll. (Sind einige der Zahlen untereinander gleich, so kommt es aber sehr wohl darauf an, wie oft jede vorkommt; wenn sich diese "Vielfachheiten" ändern, so wird sich im Allgemeinen auch der Wert $M(x_1, x_2, \dots, x_n)$ ändern.) Diese drei Bedingungen sind unmittelbar einleuchtende Forderungen, die man an jedes vernünftige Mittelungsverfahren stellen wird.

Der Punkt ist, dass wir, um keine praktisch relevanten Mittelungsverfahren mit der Definition auszuschließen, nicht noch *mehr* Bedingungen gefordert haben. Das hat als Konsequenz, dass es im Sinne der Definition sehr viele und sehr unterschiedliche Mittelwertbildungen gibt, auch solche, die einem eher abwegig erscheinen. Z.B. ist die Bildung des Maximums $\max(x_1, x_2, \dots, x_n)$ zu n gegebenen Zahlen $x_i \in \mathbb{R}$ ein zulässiges Mittelungsverfahren im Sinne der Definition, ebenso auch die Bildung des Minimums $\min(x_1, x_2, \dots, x_n)$. Die Maximumbildung oder Minimumbildung wird man in der Praxis wohl kaum als Mittelungsverfahren benutzen, es gibt aber, wie wir sehen werden, auch durchaus eindrucksvolle "wissenschaftliche" Mittelwertformeln, welche Mittelwerte liefern, die immer ziemlich nahe beim Maximum der zu mittelnden Zahlen liegen (oder immer nahe beim Minimum).

Außerdem kann man das Ergebnis einer Mittelbildung erheblich dadurch verändern, dass man (mit mehr oder weniger guten Begründungen) Gewichte einführt oder die Daten manipuliert. Gewichte bewirken, dass die einzelnen zu mittelnden Zahlen das Ergebnis unterschiedlich stark beeinflussen. Die Manipulation von Daten erfolgt oft so, dass einzelne aus dem Rahmen fallende Werte zu "Ausreißern" deklariert und dann bei der Mittelbildung schlicht weggelassen werden. Aber auch ohne solche Maßnahmen im Auge zu haben, kann man sagen (wir werden das noch präzisieren): Für jedes gewünschte Ergebnis zwischen Minimum und Maximum der zu mittelnden Zahlen gibt es auch eine "Mittelwertformel", welche dieses Ergebnis liefert.

- *Die Angabe eines Mittelwertes ist daher ohne Informationswert, wenn das dabei verwendete Mittelungsverfahren nicht beschrieben wird (einschließlich vorgenommener Gewichtungen und Gewinnung der Daten). Die Wahl eines bestimmten Mittelungsverfahrens muss durch anwendungsbezogene Überlegungen gerechtfertigt werden.*

Bevor wir zu Beispielen von Mittelungsverfahren kommen, nennen wir noch eine Eigenschaft, die viele wichtige Mittelungsverfahren auf dem Grundintervall \mathbb{R} oder $\mathbb{R}_{>0}$ besitzen. Diese Eigenschaft besagt, dass sich der Mittelwert $M(x_1, x_2, \dots, x_n)$ zum Beispiel verdoppelt, wenn wir jeden der zu mittelnden Zahlenwerte x_j durch den doppelten Wert $2x_j$ ersetzen; entsprechend auch bei Multiplikation mit anderen Faktoren. Man spricht, wenn eine Mittelwertbildung diese Eigenschaft hat, von der **Homogenität** der Mittelwertbildung, d.h. für alle $s > 0$ und alle x_1, x_2, \dots, x_n gilt

$$M(sx_1, sx_2, \dots, sx_n) = s \cdot M(x_1, x_2, \dots, x_n).$$

Da aber nicht alle in der Praxis relevanten Mittelungsverfahren homogen sind, haben wir diese Eigenschaft nicht zum Bestandteil der allgemeinen Definition von Mittelwertbildungen gemacht.

BEISPIELE und DISKUSSION von Mittelungsverfahren:

Für alle nachfolgend angegebenen Mittelungsverfahren sind die Bedingungen (i), (ii) und (iii) aus der obigen Definition mit Hilfe einer Monotoniediskussion für die entsprechenden Terme leicht nachzuprüfen; darauf gehen wir nicht mehr ein. Die Mittelwertverfahren sind auch alle homogen.

1) Am häufigsten verwendet — aber nicht immer zu Recht — wird das sogenannte **arithmetische Mittel**:

$$AM(x_1, x_2, \dots, x_n) := \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}.$$

Hier addiert man einfach die zu mittelnden Zahlen x_j und dividiert durch die Anzahl der Summanden. Der arithmetische Mittelwert wird relativ stark durch sog. *Ausreißer* beeinflusst, d.h. durch Werte x_i , die erheblich größer oder kleiner sind als die meisten anderen x_j (und die vielleicht auf fehlerhafter Datenerhebung beruhen). Daher ist arithmetische Mittelbildung oft nicht sachgerecht und sollte nicht blindlings angewendet werden, nur weil dieses Mittel leicht zu berechnen ist. Zum Beispiel ist es nicht sinnvoll, durchschnittliche Studienzeiten durch arithmetische Mittelbildung zu berechnen; denn wenige Fälle von extrem langer Studiendauer beeinflussen den Wert des arithmetischen Mittels erheblich, die Ursachen liegen aber meist im außeruniversitären Bereich, während man mit der mittleren Studiendauer eher einen Parameter im Auge hat, der eine Information über Studienbedingungen und Studienverhalten geben soll. Beispiele für arithmetische Mittelbildung in der Ökonomie sind zahllos: Durchschnittliche Umsätze, Kosten, Gewinne, Preise, ... in mehreren Perioden oder für mehrere Produkte bzw. für mehrere Produktionsstätten eines Unternehmens usw.

2) Häufig sind die zu mittelnden Zahlen nicht "gleichberechtigt", sondern mit unterschiedlichem Gewicht zu berücksichtigen. Unter einem Satz von n **Gewichten** verstehen wir immer die Vorgabe von n positiven Zahlen s_1, s_2, \dots, s_n mit Summe $s_1 + s_2 + \dots + s_n = 1$. (Insbesondere sind dann alle Gewichte s_j Zahlen zwischen 0 und 1, wenn $n \geq 2$ ist. Man könnte auch den Wert Null für Gewichte zulassen; das läuft darauf hinaus, dass die Zahlen mit Gewicht Null bei der Mittelbildung einfach weggelassen werden.) Gewichte werden oft als Prozentsätze angegeben, $s_j = p_j\% = \frac{1}{100}p_j$; die Zahlen p_1, \dots, p_n müssen sich dann natürlich zu 100 addieren. Das mit den Gewichten s_1, \dots, s_n aus den reellen Zahlen x_1, x_2, \dots, x_n gebildete **gewichtete arithmetische Mittel** ist

$$AM(x_1, x_2, \dots, x_n; s_1, s_2, \dots, s_n) := \sum_{j=1}^n s_j x_j = s_1 x_1 + s_2 x_2 + \dots + s_n x_n.$$

Bei dieser Mittelbildung multipliziert man also jede der zu mittelnden Zahlen mit ihrem zugehörigen Gewicht und addiert dann die entstandenen Produkte auf. Wählt man alle Gewichte gleich, also $s_1 = s_2 = \dots = s_n = \frac{1}{n}$, so ergibt sich das gewöhnliche ("ungewichtete") arithmetische Mittel der x_j .

Mitunter sind die Gewichte s_j nicht direkt so vorgegeben, dass sie sich zu 1 summieren, sondern man kennt nur gewisse positive "Referenzgrößen" r_j und möchte die gegebenen Zahlen mit Gewichten s_j mitteln, die in denselben Verhältnissen $s_1 : s_2 : \dots : s_n$ stehen wie die Referenzgrößen $r_1 : r_2 : \dots : r_n$ (d.h. die Quotienten $s_i : s_j$ und $r_i : r_j$ sind für alle $i \neq j$ gleich). Dies bedeutet, dass man die Gewichte $s_j := r_j/r$ zu wählen hat, wobei $r := r_1 + r_2 + \dots + r_n$ die Summe der Referenzgrößen ist (denn genau bei dieser Wahl der s_j stehen die s_j in denselben Verhältnissen zueinander wie die r_j und haben die Summe 1). Für je n gegebene reelle Zahlen x_1, x_2, \dots, x_n und positive Zahlen r_1, r_2, \dots, r_n ist also das **arithmetische Mittel im Verhältnis der Referenzgrößen** $r_j > 0$ gleich

$$\frac{\sum_{j=1}^n r_j x_j}{\sum_{j=1}^n r_j} = \frac{r_1 x_1 + r_2 x_2 + \dots + r_n x_n}{r_1 + r_2 + \dots + r_n}.$$

Das arithmetische Mittel der Zahlen x_j im Verhältnis $1 : 1 : \dots : 1$ ist also ihr gewöhnliches ("ungewichtetes") arithmetisches Mittel.

Soll das arithmetische Mittel zu einer endlichen Zahlenfolge gebildet werden, in denen Zahlen mehrfach auftreten können (Z.B. Klausurergebnisse), so kann man stattdessen auch die verschiedenen auftretenden Zahlen arithmetisch im Verhältnis ihrer Fallzahlen mitteln, d.h. im Verhältnis der Häufigkeiten ihres Auftretens in der Folge. Das Ergebnis ist dasselbe, weil die Summe der mit ihren Fallzahlen multiplizierten verschiedenen Zahlen gleich der Summe aller Zahlen der ursprünglichen Zahlenfolge mit "Wiederholungen" ist und die Summe aller Fallzahlen gleich der Anzahl der Glieder der ursprünglichen Folge.

In der Wahrscheinlichkeitstheorie treten gewichtete arithmetische Mittel $s_1x_1 + \dots + s_nx_n$ als *Erwartungswert einer Zufallsgröße* auf, welche die n verschiedenen Werte x_1, \dots, x_n annehmen kann und keine anderen. Das Gewicht $s_j \geq 0$ ist dabei die Wahrscheinlichkeit dafür, dass der Wert x_j auftritt, und die Summe dieser Wahrscheinlichkeiten ist 1, weil die Zufallsgröße mit Sicherheit einen der Werte x_j annimmt. Der Erwartungswert ist im Allgemeinen kein Wert, den die Zufallsgröße annehmen kann; insofern ist die Benennung etwas irreführend. (Beim Würfeln werden z.B. die Werte $1, \dots, 6$ der Augenzahl jeweils mit Wahrscheinlichkeit $\frac{1}{6}$ angenommen, der Erwartungswert, also hier das arithmetische Mittel der möglichen Augenzahlen, ist aber $\frac{7}{2}$ und somit keine mögliche Augenzahl.) Vielmehr gibt der Erwartungswert einen in diesem Kontext sinnvollen Mittelwert der möglichen Werte an, welche die Zufallsgröße annehmen kann, nämlich das gewichtete arithmetische Mittel mit den entsprechenden Wahrscheinlichkeiten als Gewichten.

3) Ein Beispiel aus der Ökonomie für die Bildung eines gewichteten arithmetischen Mittels ist der **Durchschnittspreis** für eine auf n Märkten abgesetzte Ware: Sind p_1, \dots, p_n die auf diesen Märkten jeweils erzielten Preise (pro abgesetzter Einheit), so darf man natürlich nicht die Zahlen p_j direkt arithmetisch mitteln, sondern man muss sie im Verhältnis der jeweils abgesetzten Mengen q_1, \dots, q_n (Einheiten) mitteln, um einen ökonomisch sinnvollen Mittelwert zu erhalten. (Ein hoher erzielter Preis auf einem Markt, auf dem fast nichts abgesetzt wurde, darf den Mittelwert nicht so stark beeinflussen, wie ein niedriger Preis auf einem andern Markt mit großem Umsatz.) Analog werden die **Durchschnittskosten** eines Produktionsfaktors (z.B. eines Rohstoffs), dessen Preis von Monat zu Monat schwankt, als arithmetisches Mittel der in n Monaten zu zahlenden Preise p_1, \dots, p_n im Verhältnis der jeweils eingekauften Mengen q_1, \dots, q_n definiert; etwaige Restbestände des eingekauften Produktionsfaktors werden am Jahresende mit diesen Durchschnittskosten bewertet und bilanziert. In beiden Situationen wird hier also das gewichtete arithmetische Mittel $\sum_{j=1}^n q_j p_j / \sum_{j=1}^n q_j$ gebildet.

Indexzahlen sind Verhältnisse von zwei (gewichteten oder ungewichteten) arithmetischen Mitteln ökonomischer Variablen und werden benutzt, um die zeitliche Entwicklung wichtiger ökonomischer Parameter zu beschreiben (Preisindex, Umsatzindex, Aktienindex, ...). Dabei erhebt man die Werte x_1, \dots, x_n von gewissen positiven ökonomischen Variablen in einer aktuellen "Berichtsperiode" und stellt diese den "Basiswerten" $\hat{x}_1, \dots, \hat{x}_n$ dieser Variablen aus einer früheren "Basisperiode" gegenüber, indem man den Quotienten der gewichteten arithmetischen Mittel der beiden Sätze von Variablenwerten bildet. Dabei muss man natürlich für beide Mittelungen dieselben Gewichte nehmen, sonst ist der Vergleich der Mittelwerte sinnlos. Eine Indexzahl sieht also, wenn die Gewichte im gleichen Verhältnis wie bei den positiven Referenzgrößen r_1, \dots, r_n gewählt sind, folgendermaßen aus

$$I = \frac{\sum_{j=1}^n r_j x_j / \sum_{j=1}^n r_j}{\sum_{j=1}^n r_j \hat{x}_j / \sum_{j=1}^n r_j} = \frac{\sum_{j=1}^n r_j x_j}{\sum_{j=1}^n r_j \hat{x}_j}$$

Statt der Zahl I wird oft der Prozentsatz $100 \cdot I\%$ angegeben.

Die Indexzahl I fällt größer als 1 aus, wenn der Mittelwert vom Bezugszeitpunkt bis zum aktuellen Zeitpunkt zugenommen hat, bzw. kleiner als 1, wenn er abgenommen hat. Genauer ist der aktuelle Mittelwert gleich $100 \cdot I\%$ vom Mittelwert aus der Basisperiode, und $I - 1$ gibt die relative Zunahme bzw. Abnahme des Mittelwertes von der Basisperiode bis zur Berichtsperiode an (d.h. die Zunahme/Abnahme beträgt $100 \cdot (I - 1)\%$ des Mittelwertes aus der Basisperiode).

Betrachten wir z.B. einen "Warenkorb" von n Gütern, mit Marktpreisen p_j (Geldeinheiten pro Mengeneinheit), umgesetzten Mengen q_j (Mengeneinheiten) und Umsätzen $u_j = p_j \cdot q_j$ (Geldeinheiten) im aktuellen Jahr. Für die Bildung eines mittleren Umsatzes ist ungewichtete arithmetische Mittelbildung sinnvoll, d.h. man mittelt die Umsätze der einzelnen Güter im Verhältnis $1 : 1 : \dots : 1$. Dementsprechend ist der

$$\text{Umsatzindex: } \frac{\sum_{j=1}^n u_j}{\sum_{j=1}^n \hat{u}_j} = \frac{\sum_{j=1}^n p_j q_j}{\sum_{j=1}^n \hat{p}_j \hat{q}_j}.$$

Bei der Bildung eines mittleren Preises ist es aber sicher nicht sinnvoll, einfach das arithmetische Mittel der Preise der einzelnen Güter zu nehmen; denn eine Verteuerung bei einem wenig verkauften Produkt sollte nicht so ins Gewicht fallen wie eine Verteuerung bei einem viel umgesetzten Produkt. Also wird man die Preise der n Güter im Verhältnis der umgesetzten Mengen mitteln. Damit entfällt auch das Problem bei der Definition der Mengenangaben q_j , wenn die umgesetzte Mengen der Güter des Warenkorbs in verschiedenen Einheiten gemessen werden (Stückzahlen, Gewichtseinheiten, ...); denn die Produkte $q_j p_j = u_j$ sind unabhängig von der Wahl der Mengeneinheiten. Man hat nun zwei verschiedene Möglichkeiten der Gewichtung, die sich anbieten: Man kann sich auf die umgesetzten Mengen in der aktuellen Berichtsperiode beziehen (*Index nach Paasche*) oder auf die umgesetzten Mengen in der früheren Basisperiode (*Index nach Laspeyres*). Demnach ist der

$$\text{Preisindex: } \frac{\sum_{j=1}^n q_j p_j}{\sum_{j=1}^n q_j \hat{p}_j} \quad (\text{Paasche}), \quad \text{bzw.} \quad \frac{\sum_{j=1}^n \hat{q}_j p_j}{\sum_{j=1}^n \hat{q}_j \hat{p}_j} \quad (\text{Laspeyres}).$$

Diese Preisindizes haben wir in Abschnitt 1.1 schon im Zusammenhang mit den Rechenregeln für Summen diskutiert. Beide bestimmen die mittleren Preise so, als hätten sich die umgesetzten Mengen seit der Basisperiode nicht geändert, wobei im ersten Index die aktuellen umgesetzten Mengen, im zweiten die umgesetzten Mengen in der Basisperiode angesetzt werden. Analog hat man auch zwei Möglichkeiten, einen Mengenindex zu definieren, der die Änderung der mittleren umgesetzten Mengen von der Basisperiode bis zur Berichtsperiode beschreibt. Hier ist die einfache arithmetische Mittelung der umgesetzten Mengen ebenfalls nicht sinnvoll, schon wegen der unterschiedlichen Einheiten, in denen die umgesetzten Mengen gemessen werden. (Es macht schließlich einen beträchtlichen zahlenmäßigen Unterschied, ob man z.B. verkaufte Streichhölzer in Stückzahlen oder in Kilogramm angibt.) Das Problem entfällt, wenn man die umgesetzten Mengen im Verhältnis der Preise mittelt, und so erhält man den

$$\text{Mengenindex: } \frac{\sum_{j=1}^n p_j q_j}{\sum_{j=1}^n p_j^{\circ} q_j} \quad (\text{Paasche}), \quad \text{bzw.} \quad \frac{\sum_{j=1}^n \overset{\circ}{p}_j q_j}{\sum_{j=1}^n \overset{\circ}{p}_j \overset{\circ}{q}_j} \quad (\text{Laspeyres}).$$

Bei beiden Mengenindizes werden die mittleren Umsätze so ermittelt, als hätten sich die Preise seit der Basisperiode nicht verändert, wobei im ersten Fall die aktuellen, im zweiten Fall die Basispreise angenommen werden. Insofern sind beide Indizes ein Maß für die Entwicklung der durchschnittlich abgesetzten Mengen von der Basis- bis zur Berichtsperiode.

Die Vor- und Nachteile der Index-Definitionen nach Paasche und Laspeyres liegen auf der Hand: Bei der Definition nach Laspeyres kann man die Gewichte aus den Daten der Basisperiode entnehmen, das spart Erhebungskosten und ermöglicht, die Gewichte für mehrere aufeinander folgende Berichtsperioden beizubehalten. So verfährt z.B. das statistische Bundesamt bei der Ermittlung eines Lebenshaltungskostenindex. Allerdings entspricht die Gewichtung nach einigen Perioden dann möglicherweise kaum noch der aktuellen Situation, während die Indizes nach Paasche sich eben immer an der aktuellen Situation orientieren. Deshalb müssen die angenommenen Umsätze für die Güter des "Lebensmittelkorbs" von Zeit zu Zeit aktualisiert werden (wie auch die Zusammensetzung des Korbs überhaupt den eingetretenen Entwicklungen angepasst werden muss).

4) Das geometrische Mittel von n positiven (!) Zahlen x_1, \dots, x_n ist das multiplikative Analogon des arithmetischen Mittels. Statt die Zahlen x_j zu addieren und durch die Anzahl n zu dividieren, multipliziert man also die x_j und zieht aus dem Produkt die n -te Wurzel (denn weil die Erhebung in die n -te Potenz das multiplikative Analogon der Vervielfachung mit dem Faktor n ist, hat man das Ziehen der n -ten Wurzel als multiplikatives Analogon der Division durch n anzusehen). Somit ist das **geometrische Mittel** definiert durch

$$GM(x_1, x_2, \dots, x_n) := \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} = \prod_{j=1}^n x_j^{1/n} \quad (x_j > 0).$$

Der Name "geometrisches Mittel" kommt von folgender geometrischen Interpretation des geometrischen Mittels \sqrt{ab} von zwei positiven Zahlen: Dieser Mittelwert \sqrt{ab} ist gerade die Seitenlänge eines Quadrats mit demselben Flächeninhalt ab wie ein Rechteck mit Seitenlängen a und b . Die Positivität der Zahlen x_j , die man hier mittelt, ist wichtig. Bei negativen Faktoren kann das Produkt negativ werden und das Wurzelziehen ist dann nicht möglich. Bei nichtnegativen Faktoren ist das Produkt Null und damit das geometrische Mittel gleich dem Minimum der zu mittelnden Zahlen, wenn eine davon den Wert Null hat; deshalb ist es nicht besonders sinnvoll, für die x_j hier auch den Wert Null zuzulassen, und wir tun das auch nicht. Ein wesentlicher Unterschied zwischen dem arithmetischen und dem geometrischen Mittel ist, dass letzteres von "Ausreißern nach oben" viel weniger stark beeinflusst wird als die arithmetische Mittelbildung. Ein extremes Beispiel mag das verdeutlichen: Mittelt man 10 Zahlen, von denen 9 gleich 1 sind und eine gleich 1000, so ist das arithmetische Mittel größer als 100, das geometrische Mittel aber kleiner als 2!

Da Logarithmieren die Multiplikation in Addition verwandelt, kann man die Bildung des geometrischen Mittels der x_j auch so beschreiben: Man schreibt die Zahlen als Potenzen einer Basis $a > 1$, also $x_j = a^{\xi_j}$ mit $\xi_j = \log_a x_j$, und mittelt dann die Exponenten ξ_j arithmetisch:

$$GM(x_1, x_2, \dots, x_n) = a^{AM(\xi_1, \xi_2, \dots, \xi_n)} \quad (x_j = a^{\xi_j} > 0).$$

Hier sieht man nun, wie man zu vorgegebenen positiven Gewichten $s_j > 0$ mit $s_1 + s_2 + \dots + s_n = 1$ das **gewichtete geometrische Mittel** der positiven Zahlen x_j zu definieren hat: Man bildet einfach das entsprechend gewichtete arithmetische Mittel der Exponenten, was man wegen $a^{s_j \xi_j} = x_j^{s_j}$ auch so schreiben kann:

$$GM(x_1, x_2, \dots, x_n; s_1, s_2, \dots, s_n) := \prod_{j=1}^n x_j^{s_j} \quad (x_j > 0).$$

Das geometrische Mittel im Verhältnis der Referenzgrößen $r_j > 0$ mit Summe r ist dann

$$\left(\prod_{j=1}^n x_j^{r_j} \right)^{\frac{1}{r}} \quad (x_j > 0, \quad r = r_1 + r_2 + \dots + r_n).$$

Wählt man alle r_j gleich, also $s_j = \frac{1}{n}$ für $j = 1 \dots n$, so ergibt sich das gewöhnliche ("ungewichtete") geometrische Mittel.

5) Eine Anwendung in der Ökonomie findet das geometrische Mittel bei der Berechnung des effektiven Zinssatzes für einen Vorgang, bei dem ein Kapital in n Zinsperioden gleicher Dauer $\frac{1}{m}$ (Jahre, $m \in \mathbb{N}$) mit evtl. unterschiedlichen Aufzinsungsfaktoren q_1, \dots, q_n verzinst wird. Der **mittlere Aufzinsungsfaktor** $q_* > 1$ für eine Periode ist dann derjenige, der bei Anwendung in jeder Zinsperiode nach n Perioden dasselbe Ergebnis liefert. Die Gleichung für q_* ist also $q_*^n \cdot K_0 = q_1 \cdot q_2 \cdot \dots \cdot q_n \cdot K_0$, und die Lösung ist $q_* = \sqrt[n]{q_1 \cdot q_2 \cdot \dots \cdot q_n} = GM(q_1, q_2, \dots, q_n)$. Der äquivalente Jahresbezogene Aufzinsungsfaktor ist dann q_*^m und der dadurch bestimmte Jahreszinsfuß, der sogenannte **effektive Zinssatz** $p_{\text{eff}}\%$ für den Verzinsungsvorgang, ist dann gegeben durch $p_{\text{eff}} = 100 \cdot (q_*^m - 1)\%$. Bei einfacher Verzinsung wäre dagegen das arithmetische Mittel der q_j der mittlere Aufzinsungsfaktor für eine Periode gewesen, und als Effektivzinssatz hätte man das arithmetische Mittel der in den einzelnen Perioden verwendeten Jahreszinsfüße erhalten.

Hat man unterschiedliche Aufzinsungsfaktoren q_j , die evtl. in mehreren Zinsperioden Anwendung finden, so ist der mittlere Aufzinsungsfaktor q_* entsprechend das geometrische Mittel der q_j im Verhältnis der Anzahlen m_j der Perioden, in denen mit Aufzinsungsfaktor q_j verzinst wird. Bei kontinuierlicher Verzinsung zu Zinsfüßen p_j in n Perioden beliebiger Dauer t_j ist der Jahresbezogene mittlere Aufzinsungsfaktor das geometrische Mittel der Jahresbezogenen Aufzinsungsfaktoren $e^{p_j/100}$ im Verhältnis der Dauern t_j , also

$$\left[(e^{p_1/100})^{t_1} \cdot \dots \cdot (e^{p_n/100})^{t_n} \right]^{\frac{1}{t}} = e^{(t_1 p_1 + \dots + t_n p_n)/t \cdot 100} =: e^{\tilde{p}/100} \quad (t = t_1 + \dots + t_n),$$

und der entsprechende mittlere Jahreszinsfuß $\tilde{p}\%$ ist, wie man sieht, das arithmetische Mittel der in den einzelnen Perioden verwendeten Zinsfüße im Verhältnis der Länge dieser Perioden. (Der effektive Zinssatz, d.h. der konforme Zinsfuß in einem Vergleichsverfahren mit jährlichem Zinszuschlag, ist hier $p_{\text{eff}} = 100 \cdot (e^{\tilde{p}/100} - 1)$ Prozent, und das ist natürlich größer als \tilde{p} Prozent, weil bei kontinuierlicher Verzinsung ja ein kleinerer Zinsfuß schon dasselbe Jahresergebnis liefert.)

6) Für n gegebene nichtnegative Zahlen x_1, x_2, \dots, x_n heißt die Quadratwurzel aus dem arithmetischen Mittel ihrer Quadrate das **quadratische Mittel**,

$$QM(x_1, x_2, \dots, x_n) := \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}} = \left(\frac{1}{n} \sum_{j=1}^n x_j^2\right)^{\frac{1}{2}}.$$

Bildet man das arithmetische Mittel der Quadrate zu Gewichten $s_1, s_2, \dots, s_n > 0$ mit $s_1 + s_2 + \dots + s_n = 1$, so erhält man das entsprechende **gewichtete quadratische Mittel**:

$$QM(x_1, x_2, \dots, x_n; s_1, s_2, \dots, s_n) := \sqrt{s_1 x_1^2 + s_2 x_2^2 + \dots + s_n x_n^2} = \left(\sum_{j=1}^n s_j x_j^2\right)^{\frac{1}{2}}.$$

Zwar sind diese Ausdrücke auch definiert, wenn einige der Zahlen x_j negativ sind, aber dann liegen sie nicht mehr unbedingt zwischen dem Minimum und dem Maximum der x_j , können also nicht mehr als Mittelwert aufgefasst werden. (Wenn z.B. alle x_j negativ sind, so ist $QM(x_1, x_2, \dots, x_n) > 0$ sicher größer als das Maximum.) Daher beschränken wir uns beim quadratischen Mitteln auf nichtnegative Zahlen. Das quadratische Mittel wird von "Ausreißern nach oben" noch mehr beeinflusst als das arithmetische. Bei dem schon betrachteten extremen Beispiel, in dem 10 Zahlen gemittelt werden, von denen eine gleich 1000 ist und 9 gleich 1 sind, ist z.B. das arithmetische Mittel 100.9, das quadratische aber ≈ 316.23 .

7) Quadratische Mittel sind von großer Bedeutung für die Wahrscheinlichkeitsrechnung und Statistik sowie für die sog. *Fehlerausgleichsrechnung*. Dabei sind z.B. Werte x_j einer mit zufälligen Schwankungen behafteten Variablen erhoben (etwa Meßgrößen in der Qualitätskontrolle), und man sucht einen vernünftig erscheinenden Wert \bar{x} , den man als "wahren Wert" der Größe interpretieren kann. Nach einem Vorschlag von Gauß kann man z.B. \bar{x} durch die Bedingung festlegen, dass das quadratische Mittel aller "Fehler" $|x_j - \bar{x}|$ möglichst klein wird; dies ist die sog. **Methode der kleinsten Quadrate**. Mit quadratischer Ergänzung können wir \bar{x} dann leicht bestimmen: Aus

$$\frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2 = \frac{1}{n} \sum_{j=1}^n x_j^2 - \frac{2\bar{x}}{n} \sum_{j=1}^n x_j + \bar{x}^2 = \frac{1}{n} \sum_{j=1}^n x_j^2 + \left(\bar{x} - \frac{1}{n} \sum_{j=1}^n x_j\right)^2 - \left(\frac{1}{n} \sum_{j=1}^n x_j\right)^2$$

lesen wir nämlich ab, dass $\frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2$ genau dann am kleinsten wird, wenn wir für \bar{x} das arithmetische Mittel der x_j einsetzen. Bei dieser Festsetzung des "wahren Wertes" \bar{x} ist dann das quadratische Mittel der Fehler $|x_j - \bar{x}|$ gleich

$$\sqrt{\frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2} = \sqrt{\frac{1}{n} \sum_{j=1}^n x_j^2 - \bar{x}^2} \quad (\bar{x} := \frac{1}{n} \sum_{j=1}^n x_j).$$

Diese Größe heißt **mittlerer Fehler** oder **Standardabweichung** der x_j von ihrem arithmetischen Mittel. Das einfacher zu handhabende Quadrat des Ausdrucks nennt man die **Varianz** der Werte x_1, \dots, x_n ,

$$\frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2 = \frac{1}{n} \sum_{j=1}^n x_j^2 - \bar{x}^2 \quad (\bar{x} := \frac{1}{n} \sum_{j=1}^n x_j).$$

Mittlerer Fehler und Varianz sind ein Maß für die **Streuung** der Werte x_1, \dots, x_n der Zufallsgröße. Eine große Varianz deutet z.B. auf große Qualitätsunterschiede hin, wenn es um Qualitätskontrolle geht, oder auf große Meßfehler. Meist ist weniger die Standardabweichung von Interesse als ihr Verhältnis zur Größe des "wahren Wertes" \bar{x} , jedenfalls in Situationen, in denen $\bar{x} > 0$ groß ist gegenüber den Fehlern $|x_j - \bar{x}|$. Dieses Verhältnis heißt die **relative Standardabweichung** oder der **Variationskoeffizient**

$$\frac{1}{\bar{x}} \sqrt{\frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2} = \sqrt{\frac{1}{n} \sum_{j=1}^n \frac{x_j^2}{\bar{x}^2} - 1} \quad (\bar{x} := \frac{1}{n} \sum_{j=1}^n x_j).$$

Eine etwas allgemeinere Aufgabe besteht darin, aus Erhebungen der Werte x_1, \dots, x_n und y_1, \dots, y_n zu zwei Variablen, für die ein linearer Zusammenhang $y = ax + b$ angenommen wird, die unbekanntenen Koeffizienten a, b möglichst gut zu bestimmen. Man kann dies so sehen, dass in der (x, y) -Ebene eine Gerade gesucht ist, welche möglichst nahe bei den Punkten (x_j, y_j) , $j = 1 \dots n$, verläuft. Nimmt man als Maß für die Abweichung der Geraden von den Punkten wieder das quadratische Mittel der "Fehler" $|y_j - ax_j - b|$, d.h. der Entfernung der Punkte (x_j, y_j) von den vertikal darüber oder darunter liegenden Punkten auf der Geraden, so kann man die Aufgabe ganz ähnlich wie oben mit quadratischer Ergänzung lösen. Sind \bar{x} und \bar{y} wie oben die arithmetischen Mittel der x_j bzw. der y_j , so lesen wir aus

$$\frac{1}{n} \sum_{j=1}^n (y_j - ax_j - b)^2 = \sum_{j=1}^n \frac{(y_j - ax_j)^2}{n} - 2b(\bar{y} - a\bar{x}) + b^2 = \sum_{j=1}^n \frac{(y_j - ax_j)^2}{n} + (b - \bar{y} + a\bar{x})^2 - (\bar{y} - a\bar{x})^2$$

direkt ab, dass $b = \bar{y} - a\bar{x}$ sein muss, wenn der quadratische Fehler minimal sein soll. Und wenn wir diesen Wert für b einsetzen, so zeigt

$$\begin{aligned} \sum_{j=1}^n [(y_j - \bar{y}) - a(x_j - \bar{x})]^2 &= \sum_{j=1}^n (y_j - \bar{y})^2 - 2a \sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y}) + a^2 \sum_{j=1}^n (x_j - \bar{x})^2 \\ &= \sum_{j=1}^n (y_j - \bar{y})^2 + \left(\frac{\sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_{j=1}^n (x_j - \bar{x})^2}} - a \sqrt{\sum_{j=1}^n (x_j - \bar{x})^2} \right)^2 - \frac{\left(\sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y}) \right)^2}{\sum_{j=1}^n (x_j - \bar{x})^2}, \end{aligned}$$

dass der quadratische Fehler minimal wird, genau wenn wir die gesuchten Zahlen a und b wie folgt wählen:

$$a = \frac{\sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y})}{\sum_{j=1}^n (x_j - \bar{x})^2}, \quad b = \bar{y} - a\bar{x} = \frac{1}{n} \sum_{j=1}^n (y_j - ax_j)$$

Die hierdurch festgelegte Gerade, also der Graph der linearen Funktion $y = ax + b$, heißt die **Regressionsgerade** zu den Werten x_1, \dots, x_n und y_1, \dots, y_n . Aufgabenstellungen wie die hier behandelte, und allgemeinere ähnlicher Art, zur Auswertung erhobener Daten (Meßwerte) fasst man zusammen unter dem Begriff **Regressionsanalyse**, was ein wichtiges Teilgebiet der Statistik und auch der Ökonometrie ist.

8) Der **Potenzmittelwert zum Exponenten t** von $x_1, x_2, \dots, x_n \in \mathbb{R}_{>0}$ ist definiert durch

$$PM^t(x_1, x_2, \dots, x_n) := \left(\frac{1}{n} \sum_{j=1}^n x_j^t \right)^{1/t} = \left(\frac{x_1^t + x_2^t + \dots + x_n^t}{n} \right)^{1/t},$$

also als die t -te Wurzel aus dem arithmetischen Mittel der t -ten Potenzen der x_j ; dabei darf t ein beliebiger reeller Exponent $\neq 0$ sein, und die zu mittelnden Zahlen müssen positiv sein (für $t > 0$ kann man auch den Wert Null noch zulassen). Für $t = 1$ ergibt sich z.B. das arithmetische Mittel, für $t = 2$ das quadratische Mittel, und für $t = -1$ erhält man das Reziproke des arithmetischen Mittels der Kehrwerte der x_j , das sogenannte **harmonische Mittel**

$$HM(x_1, x_2, \dots, x_n) := \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}.$$

Für $t = 0$ sind die Potenzmittelwerte nicht definiert, wir werden aber sehen, dass es sinnvoll ist, das geometrische Mittel als Potenzmittelwert zum Exponenten Null festzulegen, also PM^0 als GM zu definieren. Die Potenzmittelwerte berücksichtigen "Ausreißer nach oben" um so stärker/weniger, je größer/kleiner der Exponent t ist. In unserem Beispiel der Mittelung von 10 Zahlen, von denen 9 gleich 1 sind und eine gleich 1000 ist, ergeben sich für die Exponenten $t = -2, -1, 0, 1, 2, 3$ die Potenzmittelwerte $1.054092\dots$, $1.110987\dots$, $1.995263\dots$, 100.9 , $316.2291\dots$, $464.1588\dots$.

Bildet man statt des gewöhnlichen arithmetischen Mittels der Potenzen x_j^t oben das gewichtete arithmetische Mittel zu Gewichten $s_1, s_2, \dots, s_n > 0$ mit $s_1 + s_2 + \dots + s_n = 1$, so erhält man die **gewichteten Potenzmittelwerte zum Exponenten t** ,

$$PM^t(x_1, x_2, \dots, x_n; s_1, s_2, \dots, s_n) := \left(s_1 x_1^t + s_2 x_2^t + \dots + s_n x_n^t \right)^{1/t}.$$

9) Eine Mittelbildung ganz anderer Art ist folgende: Man numeriert die zu mittelnden Zahlen $x_1 \leq x_2 \leq \dots \leq x_n$ der Größe nach und wählt im Fall einer ungeraden Anzahl n die Zahl $x_{(n+1)/2}$ in der mittleren Position, im Fall gerader Anzahl n einen Wert zwischen den beiden Zahlen $x_{n/2}$ und $x_{1+n/2}$ in der mittleren Position. Den so erhaltenen Mittelwert nennt man einen **Median** der Zahlen x_1, x_2, \dots, x_n . Mediane sind dadurch gekennzeichnet, dass ebenso viele der zu mittelnden Zahlen über dem Median liegen wie unter ihm. Die genaue Festlegung bei einer geraden Anzahl n von Zahlen erfolgt meist als **mittlerer Median**, d.h. als arithmetischer Mittelwert $\frac{1}{2}(x_{n/2} + x_{1+n/2})$ der beiden Zahlen in mittlerer Position bei Anordnung der Größe nach. Wir schreiben

$$Med(x_1, x_2, \dots, x_n)$$

für den Median der Zahlen x_j bei ungerader Anzahl n bzw. für ihren mittleren Median bei gerader Anzahl n . Andere Möglichkeiten der Festlegung bei geradem n sind der kleinstmögliche Median $x_{n/2}$ oder der größtmögliche Median $x_{1+n/2}$. Da es bei der Bildung eines Medians nur darauf ankommt zu zählen, wieviele der Zahlen x_j unter einem gewissen Wert liegen und wieviele darüber, nicht aber auf die Größe der Zahlen, sind **Mediane völlig unsensibel gegen Ausreißer** — darin liegt ihre praktische Bedeutung. In unserem Beispiel der Mittelung von 9 Zahlen mit Wert 1 und einer mit Wert 1000 ist der Median z.B. gleich 1, und das würde auch so bleiben, wenn wir vier der Zahlen beliebig vergrößern würden.

Wegen ihrer Unempfindlichkeit gegen Ausreißer sind Mediane in vielen Situationen die sachlich angemessenen Mittelwerte, z.B. auch bei der Angabe durchschnittlicher Studierendauern. Wie andererseits Unsinn auch mit Medianen getrieben werden kann, zeigt eine

Statistik des Wissenschaftsrates über Studiendauern (1985). Dort wurde als Mittelwert der um ein halbes Semester erhöhte kleinstmögliche Median verwendet. In einem Diplomstudiengang einer Hochschule ergab sich so eine durchschnittliche Studiendauer von 11.5 Semestern, womit die Hochschule zu den drei Universitäten mit den kürzesten Studiendauern im betreffenden Fach zählte. Tatsächlich hatten im fraglichen Jahr dort aber nur 4 Studierende den Studiengang abgeschlossen, und zwar zwei nach 11 Semestern und zwei nach 17 Semestern. Schon beim mittleren Median wäre die Hochschule im letzten Drittel, beim genau so gerechtfertigten Median 16.5 gar das Schlusslicht in bezug auf die Studiendauer in diesem Fach gewesen. Soviel zum "Ranking von Hochschulen" auf der Basis unsinniger Mittelwertbildungen. (Zur Ehrenrettung des Wissenschaftsrates sei gesagt, dass er sein Mittelungsverfahren und die zugrunde liegenden Daten genau angegeben hat, so dass es möglich war, die unsinnige Mittelwertangabe überhaupt zu erkennen.)

Die Medianbildung zu Gewichten kommt in der Praxis kaum vor. Wenn man an verschiedene der Größe nach numerierte Zahlen x_j denkt, die mit Fallzahlen $m_j \geq 1$ auftreten, welche sich zur Gesamtzahl m addieren, so zählt man einfach für jedes j die Zahl x_j mit der Vielfachheit m_j . Der Median ist dann die Zahl x_{k+1} , wenn $\sum_{j=1}^k m_j < \frac{1}{2}m < \sum_{j=1}^{k+1} m_j$ ist, und eine Zahl zwischen x_k und x_{k+1} (z.B. das arithmetische Mittel dieser beiden Werte), wenn $\sum_{j=1}^k m_j = \frac{1}{2}m$ ist. Da die Gewichte in dieser Situation $s_j = m_j/m$ sind, wird man allgemein für Zahlen $x_1 \leq x_2 \leq \dots \leq x_n$ und gegebene Gewichte $s_j > 0$ mit $s_1 + s_2 + \dots + s_n = 1$ den **gewichteten Median** als x_{k+1} definieren, wenn $\sum_{j=1}^k s_j < \frac{1}{2} < \sum_{j=1}^{k+1} s_j$ ist und als eine Zahl zwischen x_k und x_{k+1} (z.B. als $\frac{1}{2}x_k + \frac{1}{2}x_{k+1}$ beim mittleren gewichteten Median), wenn $\sum_{j=1}^k s_j = \frac{1}{2}$ ist. ■

Nachdem wir nun viele verschiedene (aber in der Praxis gebräuchliche) Mittelwertbildungen eingeführt haben, ist es angebracht, etwas zum Größenvergleich für die verschiedenen Mittelwerte zu sagen. Der folgende in der Mathematik bewiesene Lehrsatz enthält alle Vergleichsaussagen, die man in dieser Hinsicht generell machen kann.

SATZ (über den Vergleich verschiedener Mittelwerte):

Für positive reelle Zahlen $x_1, \dots, x_n > 0$ gilt:

(i) die Ungleichung zwischen dem harmonischen, geometrischen und arithmetischen Mittel:

$$HM(x_1, \dots, x_n) \leq GM(x_1, \dots, x_n) \leq AM(x_1, \dots, x_n)$$

mit Gleichheit in einer der beiden Ungleichungen nur wenn $x_1 = \dots = x_n$;

(ii) die Ungleichung zwischen Potenzmittelwerten:

$$PM^r(x_1, \dots, x_n) \leq PM^t(x_1, \dots, x_n) \quad \text{für } r < t \text{ in } \mathbb{R}$$

mit Gleichheit nur wenn $x_1 = \dots = x_n$;

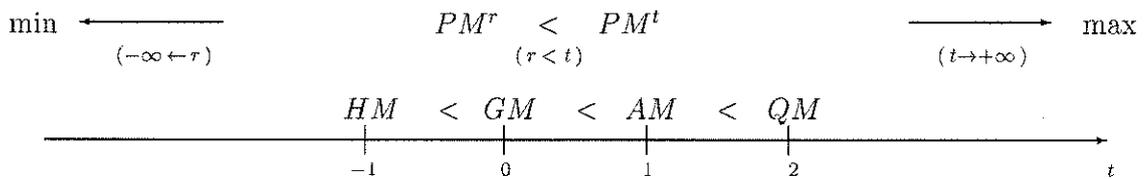
(iii) der Wertebereich der Potenzmittelwerte $PM^t(x_1, \dots, x_n)$ überdeckt das ganze Intervall zwischen dem Minimum der x_j und dem Maximum der x_j wenn man den Exponenten t von $-\infty$ bis $+\infty$ laufen läßt.

Dabei ist $PM^0(x_1, \dots, x_n)$ als geometrisches Mittel $GM(x_1, \dots, x_n)$ definiert. Die Aussagen (i) – (iii) gelten auch, wenn man alle Mittelwerte zu denselben vorgegebenen Gewichten $s_1, \dots, s_n > 0$ mit $s_1 + \dots + s_n = 1$ bildet. ■

Die Aussage (iii) erklärt, was wir zu Beginn dieses Abschnitts gesagt haben: Zu jeder gewünschten Zahl zwischen dem Minimum und dem Maximum der zu mittelnden Zahlen gibt es eine Mittelwertformel, welche genau die gewünschte Zahl als Ergebnis liefert, nämlich ein Potenzmittelwert $(x_1^t + \dots + x_n^t)^{1/t}$ mit passendem reellen Exponenten t . Natürlich würde die Wahl eines extremen Exponenten wie $t = 100$ auffallen, wenn man manipulieren möchte. Aber schon die Auswahl aus den Exponenten $-1, 0, 1$ und 2 , also aus harmonischem, geometrischen, arithmetischem und quadratischem Mittel, die alle gebräuchlich sind, kann das Ergebnis der Mittelung erheblich in die gewünschte Richtung beeinflussen. In Abwandlung eines Ausspruchs, der von Churchill kolportiert wird ("Traue keiner Statistik, die du nicht selbst gefälscht hast"), könnte man daher sagen:

"Traue keiner Mittelwertangabe, die du nicht selbst (durch Auswahl einer geeigneten Mittelwertformel) manipuliert hast."

Das folgende Diagramm stellt die Aussagen des Satzes graphisch dar für den Fall, dass nicht alle der zu mittelnden positiven Zahlen gleich sind (andernfalls sind auch alle Mittel gleich):



Für die Ungleichung " $GM \leq AM$ " gibt es eine einleuchtende ökonomische Begründung: Für Aufzinsungsfaktoren q_1, \dots, q_n in n aufeinander folgenden Zinsjahren ist nämlich $AM(q_1, \dots, q_n)$ der mittlere jährliche Aufzinsungsfaktor bei einfacher Verzinsung (Zinsen werden entnommen) und $GM(q_1, \dots, q_n)$ bei Zinseszinsverzinsung (vgl. Bsp. 5) oben). Die Ungleichung " $GM \leq AM$ " besagt hier also einfach, dass bei Zinseszinsverzinsung in n Jahren derselbe Zinsgewinn mit einem niedrigeren konstanten Zinsfuß erreicht wird als bei einfacher Verzinsung.

Die Ungleichung klärt aber auch weniger evidente Sachverhalte. Wird z.B. in 12 Monaten mit unterschiedlichen Zinsfüßen $p_1\%, \dots, p_{12}\%$ (p.a.) einfach verzinst, so ist $AM(p_1, \dots, p_{12})$ der entsprechende Jahres-Zinsfuß. Betrachtet man aber Zinseszinsverzinsung und setzt für den j -ten Monat den Monats-bezogenen Aufzinsungsfaktor $\sqrt[12]{q_j}$ an, der bei Anwendung in allen 12 Monaten den Aufzinsungsfaktor $q_j = 1 + \frac{p_j}{100}$ für das Jahr bringt, so ist $\sqrt[12]{q_j}$ natürlich kleiner als der Monats-bezogene Aufzinsungsfaktor bei einfacher Verzinsung $1 + \frac{1}{12} \frac{p_j}{100}$. (Das zeigt auch die Bernoulli-Ungleichung.) Daher ist nun für den Gesamtzeiteaum von einem Jahr nicht klar, ob die Zinseszinsverzinsung mit den Monats-bezogenen Aufzinsungsfaktoren $\sqrt[12]{q_j}$ günstiger ist oder die einfache Verzinsung mit den Monats-bezogenen Aufzinsungsfaktoren $1 + \frac{1}{12} \frac{p_j}{100}$. Die " $GM \leq AM$ " Ungleichung beantwortet das: $\sqrt[12]{q_1} \cdot \dots \cdot \sqrt[12]{q_{12}} \leq \frac{1}{12}(q_1 + \dots + q_{12}) = 1 + \frac{1}{100} \frac{p_1 + \dots + p_{12}}{12}$, also ist Letzteres günstiger (und zwar echt günstiger, wenn die Zinsfüße p_1, \dots, p_{12} nicht alle gleich sind). Das ist ökonomisch nicht ohne Weiteres klar, weil wir dabei zwar größere monatliche Aufzinsungsfaktoren haben, aber keinen Zinsverzinsungseffekt.

Auch wenn es für das Verständnis der Mittelwertungleichungen nicht nötig ist, wollen wir hier zum Schluss für Interessierte den Beweis noch skizzieren; denn die Ungleichungen sind ja keineswegs evident.

Für den *Beweis von (i)* schreiben wir $x_j = e^{y_j}$ und $\bar{x} := \sum_{j=1}^n s_j x_j$, $\bar{y} := \sum_{j=1}^n s_j y_j$. Mit der fundamentalen Ungleichung für die Exponentialfunktion aus 1.4 erhalten wir dann $e^{y_j - \bar{y}} \geq 1 + y_j - \bar{y}$, und Multiplikation der Ungleichungen mit s_j sowie Addition gibt $\bar{x}/e^{\bar{y}} \geq 1$ mit Gleichheit nur, wenn $y_j = \bar{y}$ ist für alle j , also auch $x_j = \bar{x}$. Das ist die Ungleichung " $GM \leq AM$ ", und " $HM \leq GM$ " folgt durch Übergang zu Kehrwerten.

Für den *Beweis von (ii)* im Fall $r = 1 < t$ kann man $\sum_{j=1}^n s_j x_j^t = 1$ annehmen (sonst jedes x_j durch die t -te Wurzel aus dieser Summe dividieren). Aus der " $GM \leq AM$ " Ungleichung für zwei Zahlen erhalten wir $s_j x_j = s_j^{1-1/t} (s_j x_j^t)^{1/t} \leq (1 - \frac{1}{t}) s_j + \frac{1}{t} s_j x_j^t$ und damit $\sum_{j=1}^n s_j x_j \leq (1 - \frac{1}{t}) + \frac{1}{t} \cdot 1 = 1$. Das ist die Ungleichung $PM^1 \leq PM^t$ in diesem Fall. Indem man darin x_j durch x_j^r ersetzt und t durch $\frac{t}{r}$, folgt $PM^r \leq PM^t$ für $0 < r < t$. Für $r < t < 0$ erhält man dieselbe Ungleichung durch Übergang zu Kehrwerten. Für $r < 0 < t$ ist schließlich noch $PM^t = (\sum_{j=1}^n s_j x_j^t)^{1/t} \geq (\prod_{j=1}^n x_j^{s_j t})^{1/t} = GM = PM^0 = (\prod_{j=1}^n x_j^{s_j r})^{1/r} \geq (\sum_{j=1}^n s_j x_j^r)^{1/r} = PM^r$ gemäß (i). Inspektion der Argumente zeigt auch, dass Gleichheit $PM^r = PM^t$ für $r < t$ nur eintritt, wenn alle x_j gleich sind.

Zum *Beweis von (iii)* nehmen wir an, dass etwa x_n die größte der Zahlen x_j ist und bemerken $x_n \geq PM^t \geq s_n^{1/t} x_n$. Für hinreichend große t liegt $s_n^{1/t} = e^{(1/t) \ln s_n}$ beliebig nahe bei 1 und daher PM^t beliebig nahe beim Maximum x_n . Auf der anderen Seite liegen für hinreichend kleine $t > 0$ die Zahlen $x_j^t = e^{t \ln x_j}$ beliebig nahe bei 1, und mit der fundamentalen Ungleichung aus 1.4 für die Logarithmusfunktion folgt daher $\ln PM^t = \frac{1}{t} \ln(\sum_{j=1}^n s_j x_j^t) \leq \frac{1}{t} (\sum_{j=1}^n s_j x_j^t - 1) = \frac{1}{t} \sum_{j=1}^n s_j (x_j^t - 1) = \frac{1}{t} \sum_{j=1}^n x_j^t s_j (1 - x_j^{-t}) \leq \frac{1}{t} \sum_{j=1}^n s_j x_j^t (-\ln x_j^{-t}) = \sum_{j=1}^n s_j x_j^t \ln x_j \approx \sum_{j=1}^n s_j \ln x_j = \ln GM$, wobei der im vorletzten Schritt gemachte Fehler beliebig klein ist, wenn $t > 0$ klein genug. Mit Exponentieren ergibt sich somit, dass PM^t für hinreichend kleine $t > 0$ beliebig wenig größer als GM ist. Damit nimmt $PM^t = (\sum_{j=1}^n s_j x_j^t)^{1/t}$ Werte beliebig nahe beim Maximum der x_j und beliebig nahe beim geometrischen Mittel $GM = PM^0$ an, wenn t im Intervall $]0, \infty[$ variiert. Weil PM^t auch stetig von t abhängt (d.h. die Änderung von PM^t ist kleiner als eine beliebig vorgegebene positive Größe, wenn man $t > 0$ hinreichend wenig verändert — das kann man ähnlich, aber einfacher überlegen), folgt aus dem Zwischenwertsatz der Analysis (siehe 2.1), dass PM^t alle Werte zwischen $GM = PM^0$ und dem Maximum der x_j annimmt, wenn t von 0 bis ∞ läuft. Durch Übergang zu den Reziproken der Mittel für die reziproken Zahlen $1/x_j$ folgt dann, dass PM^t auch alle Werte zwischen dem Minimum und $GM = PM^0$ annimmt, wenn t von $-\infty$ nach 0 läuft. Damit sind alle Aussagen des Satzes bewiesen.

Wie gesagt, es ist für das Verständnis der Mittelwertungleichungen nicht nötig, den gerade skizzierten Beweis nachzuvollziehen — das kann man den Mathematikern überlassen. Aber man sollte sich mit dem Größenvergleich zwischen den verschiedenen in der Ökonomie gebräuchlichen Arten von Mittelwerten auskennen und, wenn ein Argument auf Mittelwerte von Daten gestützt wird, immer fragen (ggf. auch sich selbst), ob die gewählte Art der Mittelbildung sachlich angemessen ist, oder ob möglicherweise eine weniger angemessene Mittelwertbildung verwendet wurde mit der Absicht, den berechneten Mittelwert in eine genehme Richtung größer oder kleiner ausfallen zu lassen.

Kapitel 3: Lineare Algebra

Funktionen, mit denen Wirtschaftsvorgänge modelliert werden, hängen im Allgemeinen von von vielen Variablen oder Parametern ab. Ökonomische Zusammenhänge bzw. Restriktionen bzw. Zielvorstellungen werden daher meistens beschrieben durch Gleichungen bzw. Ungleichungen bzw. Optimierungsaufgaben für Funktionen $f(x_1, \dots, x_n)$ von mehreren reellen Veränderlichen x_1, \dots, x_n .

In der Linearen Algebra befassen wir uns nur mit den (abgesehen von Konstanten) einfachsten Funktionen von mehreren Veränderlichen, den sog. **linearen Funktionen**. Diese haben die Form

$$\ell(x_1, x_2, \dots, x_n) = a_1x_1 + a_2x_2 + \dots + a_nx_n$$

mit gegebenen reellen Zahlen a_1, a_2, \dots, a_n , den *Koeffizienten* der linearen Funktion. Genauer nennt man Funktionen dieser Form **homogen lineare Funktion**, wobei die Homogenität die Eigenschaft bezeichnet, dass man einen gemeinsamen Faktor $r \in \mathbb{R}$ aller Argumente x_j "herausziehen" kann, $\ell(rx_1, rx_2, \dots, rx_n) = r \cdot \ell(x_1, x_2, \dots, x_n)$. Auch Funktionen der allgemeineren Form $f(x_1, \dots, x_n) = a_1x_1 + \dots + a_nx_n + b$, die sich von einer homogen linearen Funktion nur durch eine Konstante $b \in \mathbb{R}$ unterscheiden, werden als *lineare Funktionen* bezeichnet oder als **affin lineare Funktion**, wenn man hervorheben will, dass die allgemeinere Form mit einer nicht unbedingt verschwindenden Konstanten $b \in \mathbb{R}$ zugelassen ist.

Es wird oft gesagt, lineare Funktionen seien solche, in denen "alle Variablen nur in der ersten Potenz auftreten", aber das ist nicht korrekt: Auch bei der Funktion $f(x_1, x_2) = x_1 \cdot x_2$ treten die Variablen x_1 und x_2 nur in der ersten Potenz auf (keine Quadrate der x_j etc.), aber sie ist nicht in der Form $a_1x_1 + a_2x_2 + b$ darstellbar mit reellen Konstanten a_1, a_2, b , also eine nichtlineare Funktion! Produkte der Variablen dürfen eben auch nicht auftreten und ebensowenig nichtlineare Funktionen wie e^{x_j} , $\ln x_j$, $\sin x_j$, $x_j^{x_k}$,

Lineare Funktionen sind also sehr speziell — zu speziell für die mathematische Beschreibung vieler ökonomischer Zusammenhänge, die eben häufig nichtlinear sind, d.h. nicht durch lineare Funktionen mathematisch zu beschreiben. Dennoch ist Lineare Algebra wichtig: Erstens gibt es eben doch viele konkrete ökonomische Problemstellungen, die mit linearen Funktionen mathematisch modelliert und mit Linearer Algebra gelöst werden können. Zweitens ist die Lineare Algebra eine unverzichtbare Vorstufe für die Analysis von differenzierbaren Funktionen von mehreren Veränderlichen. Dort approximiert man nämlich allgemeine nichtlineare Funktionen durch lineare, um mit Hilfe von Kenntnissen aus der linearen Situation Aussagen über die nichtlinearen Probleme zu erhalten. Das geht natürlich nur, wenn man schon über Kenntnisse aus der Linearen Algebra verfügt.

3.1 Lineare Gleichungssysteme und Matrizen

DEFINITIONEN und TERMINOLOGIE (Lineare Gleichungssysteme):

1) Eine **lineare Gleichung** für n reelle **Unbekannte** x_1, x_2, \dots, x_n ist eine Gleichung der Form $\ell(x_1, x_2, \dots, x_n) = b$ mit einer homogen linearen Funktion ℓ , lautet also ausgeschrieben

$$a_1x_1 + a_2x_2 + \dots + a_nx_n = b \quad \text{oder} \quad \sum_{j=1}^n a_jx_j = b,$$

wobei die sog. **Koeffizienten** a_1, a_2, \dots, a_n der linearen Gleichung gegebene reelle Zahlen sind und ebenso die **rechte Seite** b der Gleichung eine gegebene reelle Zahl ist. Im Fall $b = 0$ spricht man von einer **homogenen linearen Gleichung**, im Fall $b \neq 0$ nennt man die Gleichung **inhomogen**; die rechte Seite b wird daher auch die **Inhomogenität** der linearen Gleichung genannt. Gesucht sind hier die Werte der Unbekannten $x_1, x_2, \dots, x_n \in \mathbb{R}$, für welche die Gleichung erfüllt ist. Eine **Lösung** der Gleichung ist also eine n -gliedrige Folge von reellen Zahlen x_1, x_2, \dots, x_n , wofür man (x_1, x_2, \dots, x_n) schreibt und **Lösungs- n -tupel** sagt, und die **Lösungsmenge** der linearen Gleichung ist die Menge all ihrer Lösungs- n -tupel, wozu man auch **allgemeine Lösung** sagt.

2) Allgemein nennt man eine n -gliedrige Folge (x_1, \dots, x_n) von Elementen x_1, \dots, x_n aus einer Grundmenge ein **n -tupel** (bzw. ein **Paar**, **Tripel**, **Quadrupel**, \dots , wenn $n = 2, 3, 4, \dots$) und die x_j seine **Komponenten**, **Glieder** oder **Einträge**. Die Komponenten des n -tupels müssen dabei nicht unbedingt paarweise verschieden sein, sondern Übereinstimmungen zwischen Gliedern sind erlaubt. Ein n -tupel ist dadurch bestimmt, dass man für jede seiner n Positionen (man stellt sich Plätze vor, die von 1 bis n nummeriert sind) festlegt, welcher Eintrag auf dieser Position vorgenommen wird. Dabei kommt es nicht nur auf die eingetragenen Elemente an, sondern auch auf die Positionen, an denen sie stehen. Zwei Tupel sind also dann und nur dann gleich, wenn sie erstens dieselbe Anzahl n von Gliedern haben und zweitens dieselben Einträge an denselben Positionen, $(x_1, \dots, x_n) = (y_1, \dots, y_n) \iff x_j = y_j$ für alle $j \in \{1, \dots, n\}$, und sie sind verschieden, wenn sie an mindestens einer Position verschiedene Glieder haben, $(x_1, \dots, x_n) \neq (y_1, \dots, y_n) \iff x_j \neq y_j$ für mindestens ein $j \in \{1, \dots, n\}$. Also ist das Paar $(1, 2)$ verschieden vom Paar $(2, 1)$, obwohl beide 2-tupel die Einträge 1 und 2 haben (aber an verschiedenen Positionen!), und $(1, 2) \neq (1, 2, 2)$ gilt schon deswegen, weil diese beiden Tupel nicht dieselbe Zahl von Gliedern haben.

Die Menge aller n -tupel (x_1, \dots, x_n) von reellen Zahlen heißt der **n -dimensionale reelle Zahlenraum** und wird notiert

$$\mathbb{R}^n := \{(x_1, \dots, x_n) : x_j \in \mathbb{R} \text{ für } j = 1 \dots n\}.$$

Dieser Raum \mathbb{R}^n ist die Grundmenge, in der wir die Lösungen einer linearen Gleichung für n reelle Unbekannte (oder eines linearen Gleichungssystems für n Unbekannte, s.u.) suchen; die Lösungsmenge ist eine Teilmenge von \mathbb{R}^n .

3) Wenn man es mit mehreren Unbekannten zu tun hat, so meist auch mit mehreren Gleichungen. Man muss dann nicht nur die Unbekannten, sondern auch die Gleichungen nummerieren. Dies bedeutet bei mehreren linearen Gleichungen — unvermeidlicherweise —, dass wir es mit doppelt nummerierten Koeffizienten $a_{ij} \in \mathbb{R}$ zu tun bekommen: Das erste Subskript " i " gibt die Nummer der Gleichung an, das zweite Subskript " j " die

Ein Subtraktionsterm $\dots -ax_j \dots$ in einer linearen Gleichung ist natürlich zu lesen als $\dots + (-a)x_j \dots$; der Koeffizient vor x_j in dieser Gleichung ist also $-a$. Die drei Koeffizienten in der oberen der beiden linearen Gleichungen des oben schon aufgeschriebenen Systems

$$\begin{array}{rcl} 2x & - & y & & = & 0 \\ & & y & + & 3z & = & 1 \end{array}$$

sind somit (in der Reihenfolge von links nach rechts) 2, -1 , 0 und die Koeffizienten der zweiten linearen Gleichung sind 0, 1, 3. (Es wäre ein *krasser Fehler*, als Koeffizienten von y in der ersten oder zweiten Gleichung die Zahl 0 anzugeben, weil ja vor y "kein Koeffizient vorhanden ist"!)

Spitzfindige können nun einwenden, dass dieses Gleichungssystem auch eines für vier Unbekannte w, x, y, z sein könnte, wobei der Koeffizient vor der Unbekannten w in jeder Gleichung Null ist. Das ist formal richtig; jedoch treten bei sinnvollen konkreten Problemstellungen Gleichungssysteme nicht auf, in denen alle Koeffizienten zu einer Unbekannten Null sind. Im Übrigen ergibt sich aus dem Kontext, was die Unbekannten sind. ■

FRAGEN: Die grundlegenden Fragen, deren Beantwortung man von einer guten Theorie über Gleichungssysteme (einer bestimmten Form, z.B. linear) verlangen wird, sind:

- **Existenz:** *Gibt es (mindestens) eine Lösung?*
- **Eindeutigkeit:** *Gibt es (wenn überhaupt) genau eine Lösung, oder mehrere?*
- **Beschreibung der allgemeinen Lösung:** *Läßt sich die Lösungsgesamtheit, wenn es mehrere Lösungen gibt, einfach beschreiben?*
- **Berechnung:** *Wie kann man die Lösungen konkret berechnen bzw. die Nichtexistenz von Lösungen rechnerisch feststellen?*

Mit "Lösung" ist hier natürlich, wenn es sich um ein Gleichungssystem für n Unbekannte handelt, stets ein Lösungs- n -tupel (x_1, \dots, x_n) gemeint. Wir werden sehen, dass für lineare Gleichungssysteme alle vier Fragen in sehr befriedigender Weise beantwortet werden können. Für homogene lineare Gleichungssysteme bemerken wir vorab, dass die Existenzfrage immer positiv zu beantworten ist; denn wenn alle rechten Seiten Null sind, so kann man einfach alle Unbekannten Null setzen und hat dann offenbar eine Lösung. Diese heißt die **Null-Lösung** oder, weil sie uninteressant ist, die **triviale Lösung** des homogenen linearen Gleichungssystems. ■

Für allgemeine (nichtlineare) Gleichungssysteme ist die Lage dagegen sehr viel schlechter: Zu den ersten drei Fragen gibt es, selbst wenn man die Struktur der zugelassenen Gleichungen stark einschränkt, kaum befriedigende Antworten, und diese erfordern unter Umständen extrem hohen mathematischen Aufwand. (Wenn man z.B. Systeme quadratischer Gleichungen oder algebraischer Gleichungen betrachtet, d.h. die linke Seite jeder Gleichung ist eine Linearkombination von Potenzprodukten der Unbekannten mit nicht-negativen ganzen Exponenten, so ist die sog. Algebraische Geometrie zuständig; das ist eine sehr anspruchsvolle und umfangreiche mathematische Theorie, die weit über das hinausgeht, was man in einem Mathematikstudium mit anschließendem Promotionsstudium lernen kann.) Und die Berechnung der Lösungen von allgemeinen (nichtlinearen) Gleichungssystemen ist meist nur näherungsweise möglich mit Verfahren der numerischen Mathematik, die oft auf einer Approximation mit einem linearen Gleichungssystem und auf den effektiven Methoden zur Lösung linearer Gleichungssysteme beruhen.

Bevor wir uns der allgemeinen Theorie linearer Gleichungssysteme zuwenden, betrachten wir, um einen ersten Eindruck davon zu bekommen, was man erwarten kann, einige

BEISPIELE (*lineare Gleichungssysteme mit kleinem m oder n*):

1) $m = 1, n = 1$: Hier haben wir es mit *einer* linearen Gleichung für *eine* reelle Unbekannte x zu tun,

$$ax = b.$$

Die einzige Lösung ist $x = \frac{b}{a}$ wenn der Koeffizient $a \neq 0$ ist. Das ist der "Normalfall". Wenn wir aber ganz genau sind, so müssen wir auch noch den (zugegebenermaßen nicht sehr sinnvollen) "Ausnahmefall" $a = 0$ betrachten: Dann sind entweder alle $x \in \mathbb{R}$ Lösungen, nämlich wenn auch die rechte Seite $b = 0$ ist, oder es gibt gar keine Lösung, nämlich wenn $b \neq 0$. Mehr gibt es hier nicht zu sagen.

2) $m > 1, n = 1$: Hier handelt es sich um *mehrere* lineare Gleichungen für *eine* reelle Unbekannte x ,

$$\begin{aligned} a_1x &= b_1 \\ &\vdots \\ a_mx &= b_m. \end{aligned}$$

Das ist natürlich keine sehr sinnvolle Aufgabenstellung: Die erste Gleichung legt ja die Unbekannte x schon fest (wenn $a_1 \neq 0$ ist), und wir können nicht erwarten, dass die Lösung $x = b_1/a_1$ der ersten Gleichung auch noch eine der anderen Gleichungen $a_ix = b_i$ löst. Genauer gesagt ist letzteres genau dann der Fall, wenn $b_i = (a_i/a_1)b_1$ ist, d.h. wenn man die i -te Gleichung aus der ersten erhält, indem man jede Seite mit dem Faktor a_i/a_1 multipliziert. "Normalerweise" wird das nicht der Fall sein; wenn man z.B. die Koeffizienten a_i und/oder die rechten Seiten b_i zufällig wählt, so ist die Wahrscheinlichkeit dafür gleich Null. Im "Normalfall", d.h. wenn nicht alle Gleichungen Vielfache einundderselben Gleichung sind, gibt es also für dieses Gleichungssystem keine Lösung.

Der "Ausnahmefall" liegt vor, wenn alle m Gleichungen Vielfache derselben Gleichung sind. Ist mindestens ein Koeffizient a_i nicht Null, so kann man, nötigenfalls nach Vertauschung der ersten mit der i -ten Gleichung, $a_1 \neq 0$ annehmen. Dann ist $x = b_1/a_1$ die eindeutige Lösung des Gleichungssystems. Sind aber alle a_i gleich Null, so gibt es entweder keine Lösung, nämlich wenn nicht alle b_i Null sind, oder alle $x \in \mathbb{R}$ sind Lösungen, nämlich wenn auch alle b_i Null sind.

3) $m = 1, n > 1$: Dies ist der Fall *einer* linearen Gleichung für *mehrere* Unbekannte,

$$a_1x_1 + a_2x_2 + \dots + a_nx_n = b.$$

Die Lösung ist einfach: Ist ein Koeffizient $a_j \neq 0$ — das ist hier der "Normalfall" —, so kann man die Unbekannten x_k mit Nummern $k \neq j$ beliebig als reelle Zahlen r_k (sog. *Parameter*) vorgeben und dann x_j eindeutig aus der Gleichung bestimmen, so dass

$$\begin{aligned} x_1 &= r_1, \quad \dots, \quad x_{j-1} = r_{j-1}, \\ x_j &= \frac{1}{a_j}(b - a_1r_1 - \dots - a_{j-1}r_{j-1} - a_{j+1}r_{j+1} - \dots - a_nr_{n-1}), \\ x_{j+1} &= r_{j+1}, \quad \dots, \quad x_n = r_{n-1} \end{aligned}$$

ein Lösungs- n -tupel ist. Im "Normalfall" gibt es also eine unendliche Schar von Lösungen, nämlich für jede Wahl der Parameter r_1, \dots, r_{n-1} genau eine.

Da man $n-1$ Parameter in der Lösung frei wählen kann, ist die Lösungsmenge in einem gewissen Sinne $(n-1)$ -dimensional (ein $(n-1)$ -dimensionaler affiner Unterraum von \mathbb{R}^n in der Fachsprache, d.h. eine Gerade für $n-1 = 1$, eine Ebene für $n-1 = 2$ usw.).

Im "Ausnahmefall", in dem alle Koeffizienten a_j Null sind, gibt es entweder gar keine Lösung, nämlich wenn die rechte Seite $b \neq 0$ ist, oder alle n -tupel $(x_1, \dots, x_n) \in \mathbb{R}^n$ sind Lösungen, nämlich wenn $b = 0$ ist.

4) $m = 2, n = 2$: Dies ist der einfachste Fall eines "echten" linearen Gleichungssystems, nämlich eines Systems von *zwei* Gleichungen für *zwei* Unbekannte. Nennen wir die Unbekannten x, y , die Koeffizienten a, b, c, d und die rechten Seiten r, s , so lautet es

$$(1) \quad \begin{cases} ax + by = r \\ cx + dy = s \end{cases}$$

Um eine Unbekannte zu eliminieren, multiplizieren wir die obere Gleichung von (1) mit d (d.h. beide Seiten werden mit d multipliziert) und die untere Gleichung mit b :

$$\Rightarrow (2) \quad \begin{cases} dax + dby = dr \\ bcx + bdy = bs \end{cases}$$

und wenn wir hier die untere von der oberen Gleichung subtrahieren (d.h. die linke untere von der linken oberen Seite und die rechte untere von der rechten oberen Seite), so ergibt sich folgende lineare Gleichung, in der nur noch die Unbekannte x vorkommt:

$$\Rightarrow (3) \quad (ad - bc)x = dr - bs.$$

Wenn nun $ad - bc \neq 0$ ist, so können wir aus der linearen Gleichung (3) die erste Unbekannte eindeutig berechnen zu $x = \frac{dr - bs}{ad - bc}$. Setzen wir diesen Wert für x in die obere oder untere Gleichung des ursprünglichen Systems (1) ein, so ergibt sich (weil $b \neq 0$ oder $d \neq 0$ ist, wenn $ad - bc \neq 0$) eine eindeutig lösbare lineare Gleichung für die zweite Unbekannte mit dem Ergebnis $y = \frac{as - cr}{ad - bc}$.

Die Logik bei dieser Herleitung war folgende: Wenn das System (1) erfüllt ist, so auch das System (2) und die Gleichung (3), also müssen dann x und y die angegebenen Werte haben. Damit ist zwar noch nicht gesagt, dass diese Werte von x und y wirkliche eine Lösung des ursprünglichen Systems (1) liefern, aber durch Einsetzen in die Gleichungen von (1) überzeugt man sich davon, dass dies tatsächlich der Fall ist. Wir haben also folgendes Ergebnis:

Ist die sog. Determinante $ad - bc$ des linearen Gleichungssystems (1) $\begin{cases} ax + by = r \\ cx + dy = s \end{cases}$ von Null verschieden (das ist der "Normalfall"), so besitzt das Gleichungssystem die

eindeutige Lösung $\boxed{x = \frac{dr - bs}{ad - bc}}$ und $\boxed{y = \frac{as - cr}{ad - bc}}$.

Die Zahl $ad - bc$ heißt die **Determinante** des Gleichungssystems mit den Koeffizienten a, b, c, d , weil sie bestimmt (determiniert), ob der Fall einer eindeutigen Lösung vorliegt oder nicht. Ist nämlich $ad - bc = 0$, so kann man überlegen, dass das Gleichungssystem entweder gar keine oder unendlich viele Lösungen hat.

Das geht wie folgt: Sind alle vier Koeffizienten a, b, c, d gleich Null, so hat (1) offenbar gar keine Lösung, wenn $r \neq 0$ oder $s \neq 0$ ist, und (1) hat alle Paare $(x, y) \in \mathbb{R}^2$ als Lösung, wenn $r = s = 0$. Ist aber $ad - bc = 0$ und einer der Koeffizienten $\neq 0$, etwa $a \neq 0$, so entsteht die linke Seite der unteren Gleichung aus der oberen durch Multiplikation mit $\frac{c}{a}$, weil ja $d = \frac{c}{a}b$ ist. Daher gelten beide Gleichungen simultan, genau wenn die erste gilt und wenn auch $s = \frac{c}{a}r$ ist. Das System hat also in diesem Fall entweder gar keine Lösung, nämlich wenn $s \neq \frac{c}{a}r$, oder es ist äquivalent zu einer einzigen linearen Gleichung $ax + by = r$, welche unendlich viele Lösungen $y \in \mathbb{R}$ beliebig, $x = \frac{1}{a}(r - by)$, besitzt. Völlig analog schließt man, wenn $b \neq 0$ oder $c \neq 0$ oder $d \neq 0$ ist, dass dann entweder keine Lösung von (1) existiert, oder dass es unendlich viele Lösungen (x, y) gibt, bei denen man x oder y beliebig als reellen Parameter wählen kann. Wir fassen diese Diskussion zusammen:

Ist aber die Determinante $ad - bc$ des linearen Gleichungssystems (1) $\begin{cases} ax + by = r \\ cx + dy = s \end{cases}$ gleich Null (das ist der "Ausnahmefall"), so besitzt das Gleichungssystem entweder:

- **keine Lösung**, wenn nämlich mindestens ein Koeffizient $\neq 0$ ist, aber keine der beiden Gleichungen Vielfaches der anderen ist, oder wenn alle Koeffizienten $= 0$ sind, aber nicht beide rechten Seiten r, s ; oder:
- **eine einparametrische Lösungsmenge**, wenn nämlich mindestens ein Koeffizient $\neq 0$ ist und eine der beiden Gleichungen Vielfaches der anderen (man kann dann in der Lösung (x, y) den Wert von x oder von y beliebig vorgeben und die jeweils andere Unbekannte eindeutig berechnen); oder:
- **alle $(x, y) \in \mathbb{R}^2$ als Lösungen**, wenn nämlich alle Koeffizienten und auch beide rechte Seiten $= 0$ sind. ■

Die Systeme von zwei linearen Gleichungen für zwei Unbekannte haben wir deswegen so ausführlich unter Beachtung aller möglichen auftretenden Fälle besprochen, weil das Ergebnis schon in gewisser Weise typisch für die allgemeine Situation von m linearen Gleichungen für n Unbekannte ist. Ganz allgemein bestehen nämlich, wie wir sehen werden, die *Alternativen*: Es gibt entweder genau eine Lösung, oder gar keine Lösung, oder eine Schar von unendlich vielen Lösungen, wobei einige der Unbekannten als beliebige Parameter wählbar sind und die anderen dann durch die Gleichungen des Systems eindeutig festgelegt sind. Und im Fall $m = n$, also bei ebenso vielen Gleichungen wie Unbekannten, ist die Existenz genau einer Lösung der "Normalfall", der sich, wenn er nicht von vorneherein vorliegt, nach beliebig kleinen zufälligen Änderungen der Koeffizienten einstellt.

Um beliebig große lineare Gleichungssysteme übersichtlich behandeln zu können, ist eine abkürzende Schreibweise nötig, in der nur noch die relevanten Parameter des Gleichungssystems, also die Koeffizienten und die rechten Seiten, notiert werden. Da die Koeffizienten als Zahlenschema in rechteckiger Anordnung erscheinen, wie es auch sonst in der Mathematik und in der Ökonomie in vielfältigen Zusammenhängen auftritt, hat man für solche rechteckig angeordneten Zahlensysteme eine eigene Begriffsbildung eingeführt; man nennt sie "Matrizen". Die einschlägige Terminologie hierzu beschreiben wir als Nächstes:

Mathematik für Wirtschaftswissenschaftler

(K. Steffen, Heinrich-Heine-Universität Düsseldorf, WS 2006/07)

DEFINITIONEN und TERMINOLOGIE (*Matrizen*):

1) Eine **Matrix** A vom Format (m, n) , kurz auch $m \times n$ -**Matrix** genannt (lies: "m-Kreuz-n-Matrix"), ist ein System von Zahlen (oder Elementen einer Grundmenge) a_{ij} , genannt die **Einträge** der Matrix, die nummeriert sind durch Paare von Zahlen $i \in \{1, \dots, m\}$ und $j \in \{1, \dots, n\}$. Es wird als Zahlenschema in rechteckiger Anordnung wie folgt notiert:

$$A = (a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & & \dots & a_{3n} \\ \vdots & \vdots & & a_{ij} & \vdots \\ a_{m1} & a_{m2} & & \dots & a_{mn} \end{pmatrix} \begin{array}{l} \leftarrow i\text{-te Zeile} \\ \\ \\ \uparrow j\text{-te Spalte} \end{array}$$

Dabei heißt m die **Zeilenzahl** und n die **Spaltenzahl** der Matrix. Sind beide Zahlen gleich, $m = n$, so spricht man von einer **quadratischen Matrix**. Den ersten zur Nummerierung der Einträge verwendeten Index (hier "i") nennt man den **Zeilenindex** und $\{1, \dots, m\}$ seinen **Laufbereich**, den zweiten (hier "j") nennt man den **Spaltenindex** und $\{1, \dots, n\}$ seinen Laufbereich. Die Wahl der Buchstabensymbole zur Bezeichnung der Matrixindizes ist dabei irrelevant, und die Laufbereiche werden oft nicht angegeben, wenn sie aus dem Kontext hervorgehen; man schreibt dann also einfach (a_{ij}) für die Matrix (mit Klammern, um sie von einem einzelnen Eintrag a_{ij} zu unterscheiden). Eine Matrix ist bestimmt durch ihr Format (m, n) und durch die Festlegung ihrer Einträge a_{ij} für alle "Positionen" (i, j) zu diesem Format, d.h. für alle Paare von Indizes i und j mit $1 \leq i \leq m$ und $1 \leq j \leq n$.

- *Zwei Matrizen sind gleich, genau wenn sie dasselbe Format haben und wenn ihre Einträge für alle Positionen zu diesem Format übereinstimmen.*

Das heißt also $(a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n} = (b_{kl})_{1 \leq k \leq p, 1 \leq l \leq q}$, genau wenn $m = p$, $n = q$ und $a_{ij} = b_{ij}$ für alle $i \in \{1, \dots, m\}$ und $j \in \{1, \dots, n\}$. Die **Menge aller $m \times n$ -Matrizen** mit reellen Einträgen bezeichnen wir mit $\mathbb{R}^{m \times n}$. Die $m \times n$ -Matrix mit lauter Nulleinträgen heißt **$m \times n$ -Nullmatrix** 0 .

2) Eine einzeilige Matrix ($m = 1$) bzw. eine einspaltige Matrix ($n = 1$) bezeichnet man auch als

$$\text{Zeilenvektor } (a_1 \ a_2 \ \dots \ a_n) \quad \text{bzw.} \quad \text{Spaltenvektor } \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}$$

und, um das Format mit anzugeben, genauer als n -gliedrigen Zeilenvektor bzw. Spaltenvektor mit m Einträgen / Komponenten oder ähnlich. Dabei wird in der Notation jeweils ein Matrixindex unterdrückt (der sowieso nur den Wert 1 annehmen kann). Sind alle Einträge Null, so sprechen wir vom **Nullvektor** 0 .

und den aus den rechten Seiten gebildeten Spaltenvektor die

$$\text{Inhomogenität oder rechte Seite} \quad b := \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

des Gleichungssystems. Für das gesamte lineare Gleichungssystem (LGS) schreiben wir dann einfach symbolisch

$$Ax = b,$$

was später (nach Einführung der Multiplikation von Matrizen) noch seine Rechtfertigung erhält, hier aber einfach nur als Abkürzung für (LGS) zu verstehen ist. Fügen wir der Koeffizientenmatrix A als $(n+1)$ -te Spalte die Inhomogenität b an, so erhalten wir eine Matrix $(A|b)$ vom Format $(m, n+1)$, die alle Parameter enthält, welche das lineare Gleichungssystem bestimmen, die sogenannte

$$\text{erweiterte Koeffizientenmatrix} \quad (A|b) := \left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array} \right)$$

des linearen Gleichungssystems (LGS). Der vertikale Strich vor der letzten Spalte hebt hervor, dass diese Spalte (die rechte Seite) eine andere Funktion hat als die anderen Spalten (Koeffizientenspalten), kann aber auch weggelassen werden. Eine andere Form, in der die erweiterte Koeffizientenmatrix oft notiert wird, ist das sog.

Tableau	x_1	x_2	\dots	x_n	b_1
	a_{11}	a_{12}	\dots	a_{1n}	b_1
	a_{21}	a_{22}	\dots	a_{2n}	b_2
	\vdots	\vdots	\dots	\vdots	\vdots
	a_{m1}	a_{m2}	\dots	a_{mn}	b_m

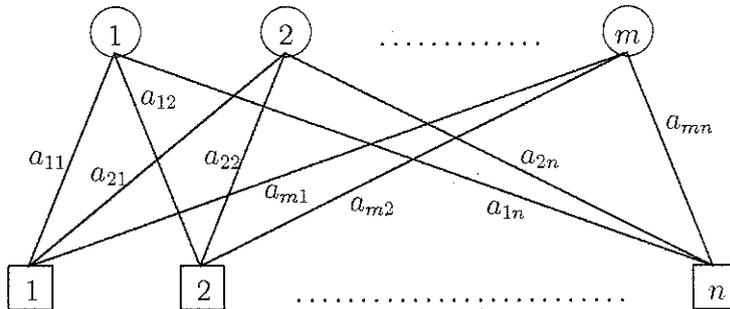
des Gleichungssystems. Dabei sind die Buchstabensymbole für die Unbekannten über die entsprechenden Spalten der Koeffizientenmatrix geschrieben. Das ist nicht nötig, wenn die Unbekannten ebenso nummeriert sind wie die Spalten, aber es ist zweckmäßig, wenn die Unbekannten x, y, \dots, z nicht nummeriert sind, und unbedingt notwendig, wenn die Nummerierung oder Reihenfolge der Unbekannten geändert wird. Die über den Koeffizientenspalten stehende Symbole der Unbekannten zeigen dann an, mit welchen Koeffizienten die jeweiligen Unbekannten in den linearen Gleichungen zu multiplizieren sind. Mathematisch gesehen ist das Tableau eines linearen Gleichungssystems nichts anderes als seine erweiterte Koeffizientenmatrix. Nur die Notation ist verändert, nicht die darin enthaltene Information. Insbesondere sind beim Tableau die manchmal unpraktischen großen "Matrixklammern" weggelassen. ■

BEISPIELE (*Matrizen in der Ökonomie*):

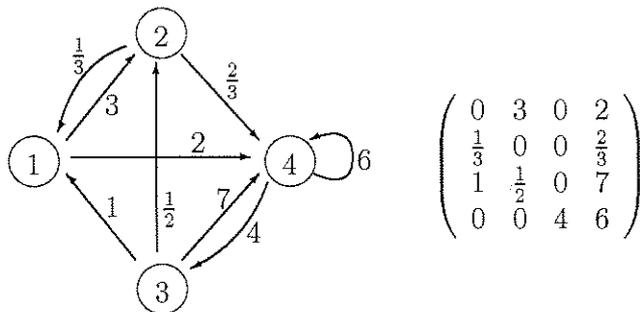
In ökonomischem Kontext treten Matrizen meistens als sog. **Verflechtungsmatrizen** auf, die einen durch Zahlenangaben quantifizierten Zusammenhang zwischen zwei Arten (oder auch derselben Art) von ökonomischen Größen darstellen. Die Größen der ersten Art entsprechen dabei z.B. den Zeilen, die der zweiten Art den Spalten der Matrix, und der Eintrag a_{ij} in Position (i, j) (i -te Zeile und j -te Spalte) gibt die quantitative Beziehung zwischen der i -ten Größe erster Art und der j -ten Größe zweiter Art an. Wenn wir die m Größen erster Art durch $\textcircled{1}, \textcircled{2}, \dots, \textcircled{m}$ symbolisieren und die n Größen zweiter Art durch $\boxed{1}, \boxed{2}, \dots, \boxed{n}$, so können wir die Zusammenhänge tabellarisch so darstellen:

	$\boxed{1}$	$\boxed{2}$...	\boxed{n}
$\textcircled{1}$	a_{11}	a_{12}	...	a_{1n}
$\textcircled{2}$	a_{21}	a_{22}	...	a_{2n}
\vdots	\vdots	\vdots		\vdots
\textcircled{m}	a_{m1}	a_{m2}	...	a_{mn}

Das ist nichts anderes als die $m \times n$ -Matrix mit den Einträgen a_{ij} , wobei lediglich an die Zeilen und Spalten noch Symbole geschrieben sind, die an die ökonomische Bedeutung der Einträge erinnern. Alternativ kann man die Beziehungen durch einen *bewerteten Graphen* darstellen, in dem die Größen als "Knoten" dargestellt und durch "Kanten" von \textcircled{i} nach \boxed{j} verbunden sind, welche durch die Anschrift der Zahlen a_{ij} "bewertet" werden.



Diese Darstellung wird allerdings recht unübersichtlich, wenn man viele Knoten hat. Daher lässt man Verbindungskanten mit der Zahl 0 als Bewertung weg. Handelt es sich um Beziehungen zwischen Größen derselben Art $\textcircled{1}, \textcircled{2}, \dots, \textcircled{m}$, so kann man diese auch durch einen **bewerteten und gerichteten Graphen** (in der Ökonomie auch "*Gozintograph*" genannt) darstellen, in dem die Knoten die Größen \textcircled{i} sind und ein Pfeil von \textcircled{i} nach \textcircled{j} mit einer angeschriebenen Zahl (das ist die "Bewertung") dem Eintrag dieser Zahl in die Zeile i und Spalte j der Verflechtungsmatrix entspricht. Pfeile mit der Zahl 0 als Bewertung werden dabei wieder weggelassen; Pfeile, die von einem Knoten zu demselben Knoten zurückführen entsprechen den Einträgen auf der Diagonalen der Matrix (gleiche Zeilen- und Spaltennummer). Die Abbildung rechts, die einen bewerteten und gerichteten Graphen mit 4 Knoten und die zugehörige Verflechtungsmatrix zeigt, sollte klar machen, wie die Information in dem Graphen kodiert ist.



Einige konkrete ökonomische Situationen, in denen die interessierenden Zusammenhänge durch Verflechtungsmatrizen beschrieben werden können, sind folgende:

1) **Transportmatrix:**

	1	2	...	n	← Ankunftsorte
1	a_{11}	a_{12}	...	a_{1n}	
2	a_{21}	a_{22}	...	a_{2n}	
⋮	⋮	⋮		⋮	
m	a_{m1}	a_{m2}	...	a_{mn}	

Absendeorte ↗

Der Eintrag a_{ij} gibt hier die Kosten für den Versand einer Mengeneinheit vom Absendeort i zum Ankunftsort j an.

2) **Produktionsmatrix:**

	1	2	...	n	← Endprodukte bzw. Zwischenprodukte
1	r_{11}	r_{12}	...	r_{1n}	
2	r_{21}	r_{22}	...	r_{2n}	
⋮	⋮	⋮		⋮	
m	r_{m1}	r_{m2}	...	r_{mn}	

Rohstoffe bzw. Vorprodukte ↗

Hier gibt der Eintrag r_{ij} den Verbrauch von Mengeneinheiten des Rohstoffs i für die Herstellung einer Einheit des Produkts j an (*Rohstoffverbrauchscoeffizient*) bzw. den Verbrauch von Mengeneinheiten des Vorprodukts i für die Herstellung einer Einheit des Endprodukts j (*Produktionscoeffizient*).

3) **Produktionskostenmatrix:**

	1	2	...	n	← Produktionsanlagen
1	k_{11}	k_{12}	...	k_{1n}	
2	k_{21}	k_{22}	...	k_{2n}	
⋮	⋮	⋮		⋮	
m	k_{m1}	k_{m2}	...	k_{mn}	

Produkte ↗

Diese Matrix ist von Interesse, wenn man Produkte auf verschiedenen Produktionsanlagen mit unterschiedlichen Kosten fertigen kann. Der Eintrag k_{ij} gibt dann die Kosten (in Geldeinheiten) für die Produktion einer Mengeneinheit des Produkts i auf der Produktionsanlage j an.

4) **Innerbetriebliche Leistungsverrechnung:** Hier stehen die einzelnen Kostenstellen des Betriebs an den Zeilen und an den Spalten der Matrix, die Matrix ist also quadratisch.

	①	②	...	②	②	②	②	← abgebende Kostenstelle
①	a_{11}	a_{12}	...	a_{1n}				
②	a_{21}	a_{22}	...	a_{2n}				
⋮	⋮	⋮		⋮				
②	a_{n1}	a_{n2}	...	a_{nn}				

empfangende
Kostenstelle ↗

Der Eintrag a_{ij} gibt hier die bei der Kostenstelle ② empfangene Leistung der Kostenstelle ② an (in Leistungseinheiten). Die Diagonaleinträge a_{ii} beschreiben also die von der Kostenstelle ② erbrachte und dort selbst verbrauchte Eigenleistung. Zum innerbetrieblichen Kostenausgleich muss man die erbrachten Leistungen mit Verrechnungspreisen bewerten — dazu später mehr.

5) **Volkswirtschaftliche Verflechtungsmatrix:**

	①	②	...	②	②	②	← Industriezweige (eines Wirtschaftssektors)
①	a_{11}	a_{12}	...	a_{1n}			
②	a_{21}	a_{22}	...	a_{2n}			
⋮	⋮	⋮		⋮			
②	a_{n1}	a_{n2}	...	a_{nn}			

Industriezweige (eines
Wirtschaftssektors) ↗

Der Eintrag a_{ij} in dieser quadratischen Matrix beschreibt den Verbrauch von Output-Einheiten des i -ten Industriezweigs durch den j -ten Industriezweig.

6) **Marktforschung, Übergangsmatrix:**

	①	②	...	②	②	②	← konkurrierende Produkte (z.B. Waschmittel)
①	p_{11}	p_{12}	...	p_{1n}			
②	p_{21}	p_{22}	...	p_{2n}			
⋮	⋮	⋮		⋮			
②	p_{n1}	p_{n2}	...	p_{nn}			

konkurrierende Produkte ↗

Hier ist p_{ij} der Anteil der Kunden des i -ten Produkts, die im Beobachtungszeitraum zum Produkt ② gewechselt sind ($i \neq j$) bzw. die dem Produkt ② treu geblieben sind ($i = j$).

Die Einträge in der i -ten Zeile geben also die vollständige Aufteilung des Kundenkreises von Produkt \textcircled{i} zu Beginn des Beobachtungszeitraums auf die konkurrierenden Produkte am Ende dieses Zeitraums an. Dabei werden die Anteile in 100 % der jeweiligen Kundenzahl angegeben. Insbesondere sind also alle Einträge Zahlen im Intervall $[0, 1]$, und die Einträge einer jeden Zeile addieren sich zu 1. Man sagt, dass die **Zeilensummen** $\sum_{j=1}^n p_{ij}$ alle den Wert 1 haben (für $i = 1 \dots n$), und nennt Matrizen mit dieser Eigenschaft **stochastische Matrizen**, weil die Einträge p_{ij} einer jeden Zeile als Wahrscheinlichkeiten interpretiert werden können (hier die Wahrscheinlichkeiten für einen Wechsel vom i -ten zum j -ten Produkt bzw. im Fall $i = j$ für das Verbleiben beim i -ten Produkt), die sich zur Gesamtwahrscheinlichkeit 1 addieren. Haben auch alle **Spaltensummen** $\sum_{i=1}^n p_{ij}$ den Wert 1 (für $j = 1 \dots n$), so liegt eine sog. **doppelt-stochastische Matrix** vor; das ist aber in der hier beschriebenen Situation im Allgemeinen nicht der Fall.

7) Häufigkeitsmatrix, Wahrscheinlichkeitstheorie: Beobachtet werden bei einem Produkttyp zwei Merkmale \mathcal{A} (z.B. Preis) und \mathcal{B} (z.B. Qualität), die in unterschiedlichen Ausprägungen / Abstufungen $\textcircled{1}, \dots, \textcircled{m}$ bei \mathcal{A} und $\boxed{1}, \dots, \boxed{n}$ bei \mathcal{B} auftreten. Dann interessiert die Häufigkeit p_{ij} (in 100 %), mit der das Merkmal \mathcal{A} in der Ausprägung \textcircled{i} simultan mit dem Merkmal \mathcal{B} in der Ausprägung \boxed{j} auftritt. Man kann auch sagen, dass $p_{ij} \in [0, 1]$ die Wahrscheinlichkeit für das Auftreten dieser Konstellation angibt. Die p_{ij} sind die Einträge der $m \times n$ -Matrix

	1	2	...	n	← Ausprägungen von Merkmal \mathcal{B}
$\textcircled{1}$	p_{11}	p_{12}	\dots	p_{1n}	
$\textcircled{2}$	p_{21}	p_{22}	\dots	p_{2n}	
\vdots	\vdots	\vdots	\dots	\vdots	
\textcircled{m}	p_{m1}	p_{m2}	\dots	p_{mn}	

Ausprägungen
↑
Ausprägungen von Merkmal \mathcal{A}

Hier ist offenbar die Summe aller Matrixeinträge

$$\sum_{i=1}^m \sum_{j=1}^n p_{ij} = 1;$$

denn diese Summe ist die Wahrscheinlichkeit dafür, dass ein Produkt mit irgendeiner Ausprägung des Merkmals \mathcal{A} und mit irgendeiner Ausprägung des Merkmals \mathcal{B} auftritt. (Natürlich ist dabei vorausgesetzt, dass jedes betrachtete Produkt die Merkmale \mathcal{A} und \mathcal{B} in einer der erfassten Ausprägungen aufweist.) ■

Diese Beispiele für das Auftreten von Matrizen bei der Beschreibung von Abhängigkeiten zwischen ökonomischen Größen ließen sich noch weiter fortsetzen. Es handelt sich dabei im Grunde immer nur um die Darstellung eines komplexen ökonomischen Beziehungsgeflechts durch eine übersichtliche Tabelle, also ein rechteckiges Zahlenschema. Natürlich hilft das noch nicht viel bei der Lösung von Problemen, die sich im ökonomischen Kontext stellen. Es kommt vielmehr darauf an, einen Kalkül, also ein Rechenverfahren, für Matrizen zu entwickeln, mit dem man dann die ökonomisch relevanten Probleme auch lösen kann. In vielen Fällen lassen sich diese Problemstellungen als lineare Gleichungssysteme formulieren, und solche Gleichungssysteme lassen sich mit einfachen Matrixoperationen tatsächlich lösen, wie wir im nächsten Abschnitt sehen werden.

3.2 Zeilenoperationen und Zeilen–Stufen–Form

Wir beschreiben nun, wie man mit einfachen Operationen mit Matrizen, den sog. Zeilenoperationen, lineare Gleichungssysteme in eine besonders einfache Gestalt äquivalent umformen und daran die Lösung(en) unmittelbar ablesen kann.

DEFINITION: (Zulässige) Zeilenoperationen mit einer gegebenen Matrix (oder einem Tableau) sind folgende Manipulationen, die eine andere (evtl. auch dieselbe) Matrix vom gleichen Format erzeugen:

- (i) **Vertauschung zweier Zeilen**, was wir symbolisch mit $\textcircled{i} \leftrightarrow \textcircled{h}$ notieren, wenn die i -te mit der h -ten Zeile vertauscht wird;
- (ii) **Multiplikation einer Zeile mit einem Faktor** $r \in \mathbb{R}_{\neq 0}$, d.h. jeder Eintrag der Zeile wird mit dem Faktor $r \neq 0$ multipliziert; dafür schreiben wir symbolisch $\textcircled{i} \rightarrow r \cdot \textcircled{i}$, wenn es sich um die i -te Zeile handelt;
- (iii) **Addition einer Zeile zu einer anderen**, d.h. jeder Eintrag einer Zeile mit Nummer h wird ersetzt durch die Summe dieses Eintrags und des Eintrags an derselben Position in einer Zeile mit anderer gegebener Nummer i ; für diese Operation schreiben wir symbolisch $\textcircled{h} \rightarrow \textcircled{h} + \textcircled{i}$.

Eine Operation vom Typ (iii) wird (von geübten Rechnern) oft zusammengefasst mit einer vorangegangenen Operation vom Typ (ii) zu einer

- (iv) **kombinierten Zeilenoperation**, d.h. man addiert das r -fache einer Zeile zu einer andern, was wir symbolisch $\textcircled{h} \rightarrow \textcircled{h} + r \cdot \textcircled{i}$ notieren, wenn das r -fache der i -ten zur h -ten Zeile addiert wird.

Ist $(a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$ die betrachtete Matrix, so wird bei der kombinierten Operation also jeder Eintrag a_{hj} der h -ten Zeile ersetzt durch $a_{hj} + ra_{ij}$ für $j = 1 \dots n$. Hier darf der Faktor r auch Null sein; denn das ändert an der Matrix ja gar nichts. ■

Ist nun $(A|b)$ die erweiterte Koeffizientenmatrix eines linearen Gleichungssystems und führt man damit eine zulässige Zeilenoperation aus, so ist jede Lösung des ursprünglichen Gleichungssystems auch eine Lösung des umgeformten Gleichungssystems, das als erweiterte Koeffizientenmatrix die neue Matrix hat, die sich aus $(A|b)$ durch die Zeilenoperation ergab. So sind eben die Zeilenoperationen gerade eingerichtet. Nun kann man aber offenbar jede Zeilenoperation durch eine zulässige Zeilenoperation wieder rückgängig machen. Deshalb ist auch jede Lösung des neuen linearen Gleichungssystems eine Lösung des alten; mit anderen Worten: Beide Systeme haben dieselbe Lösungsmenge. Es gilt also der folgende Satz, der der Schlüssel zur Lösung linearer Gleichungssysteme ist:

SATZ: *Zulässige Zeilenoperationen bei der erweiterten Koeffizientenmatrix eines linearen Gleichungssystems führen stets auf die erweiterte Koeffizientenmatrix eines äquivalenten linearen Gleichungssystems, d.h. beide Systeme haben dieselbe Lösungsmenge.* ■

DISKUSSION: 1) Man kann die Zeilenoperationen direkt am Gleichungssystem, an der Matrix oder am Tableau vornehmen, das läuft alles auf dasselbe hinaus. Wichtig ist, dass man in jedem Fall die *erweiterte* Koeffizientenmatrix nimmt, d.h. die *Zeilenoperationen müssen auch mit den rechten Seiten des Gleichungssystems vorgenommen werden!*

2) **Spaltenoperationen sind keine Äquivalenzumformungen**, d.h. wenn man Spaltenoperationen an der Koeffizientenmatrix ausführt, die natürlich ganz analog erklärt

werden, so erhält man im Allgemeinen ein neues Gleichungssystem, das nicht dieselbe Lösungsmenge hat. Zwar kann man aus den Lösungen des ursprünglich Systems die des neuen leicht berechnen und auch umgekehrt, aber das neue System ist eben im Allgemeinen nicht äquivalent in dem strikten Sinne, dass es dieselbe Lösungsmenge besitzt wie das alte. Also Hände weg von Spaltenoperationen! Ganz abwegig wäre es, etwa eine Spalte der Koeffizientenmatrix mit der Spalte der rechten Seiten zu vertauschen; denn die Koeffizienten haben eine andere Funktion in dem Gleichungssystem als die rechten Seiten.

Eine Ausnahme von der Regel sind Spaltenvertauschungen bei der Koeffizientenmatrix:

- *Spaltenvertauschungen bei der Koeffizientenmatrix sind zulässig, wenn man die Zuordnung der Unbekannten zu den Spalten beibehält, die Unbekannten also entsprechend mitvertauscht.*

Das läuft nämlich nur auf eine andere Reihenfolge der Summanden in jeder einzelnen linearen Gleichung des Systems hinaus, bzw. auf eine Ummummerierung der Unbekannten; es ändert also überhaupt nichts an den Gleichungen. Statt

x_1	x_2	x_3	\dots	x_n	b_1	kann man also genau so gut schreiben	x_2	x_1	x_3	\dots	x_n	b_1
a_{11}	a_{12}	a_{13}	\dots	a_{1n}	b_1		a_{12}	a_{11}	a_{13}	\dots	a_{1n}	b_1
a_{21}	a_{22}	a_{23}	\dots	a_{2n}	b_2		a_{22}	a_{21}	a_{23}	\dots	a_{2n}	b_2
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots		\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
a_{m1}	a_{m2}	a_{m3}	\dots	a_{mn}	b_m		a_{m2}	a_{m1}	a_{m3}	\dots	a_{mn}	b_m

denn beide Tableaus stellen dasselbe lineare Gleichungssystem dar. Ist dann zum Beispiel $(r_1, r_2, r_3, \dots, r_n)$ Lösungs- n -tupel zum zweiten Tableau, so zeigt die Notation der Unbekannten über den Spalten, dass $x_1 = r_2, x_2 = r_1, x_3 = r_3, \dots, x_n = r_n$ das Gleichungssystem zum ersten Tableau löst. (Beide Systeme haben also nicht dieselben Lösungs- n -tupel, sondern die des zweiten gehen aus denen des ersten durch die Komponentenvertauschung hervor, die der vorgenommenen Spaltenvertauschung entspricht.)

3) Das Ziel der Ausführung von Zeilenoperationen an einem gegebenen linearen Gleichungssystem ist, dieses in ein äquivalentes von so einfacher Form umzuformen, dass man dessen Lösungsmenge direkt ablesen kann. Weil Zeilenoperationen Äquivalenzumformungen sind, ist diese dann auch die Lösungsmenge des ursprünglichen Gleichungssystems. Das Ziel lässt sich, wie wir sehen werden, immer erreichen. ■

Bevor wir das im allgemeinen Fall diskutieren, betrachten wir erst ein

BEISPIEL:

$$\begin{array}{rcl} x + 2y & = & 0 \\ -x & + & z = 2 \\ -x + 4y & = & 6 \end{array}$$

Wir schreiben die erweiterte Koeffizientenmatrix dieses Systems auf und formen sie durch Zeilenoperationen um mit dem Ziel, Nulleinträge unter der Diagonalen zu erreichen:

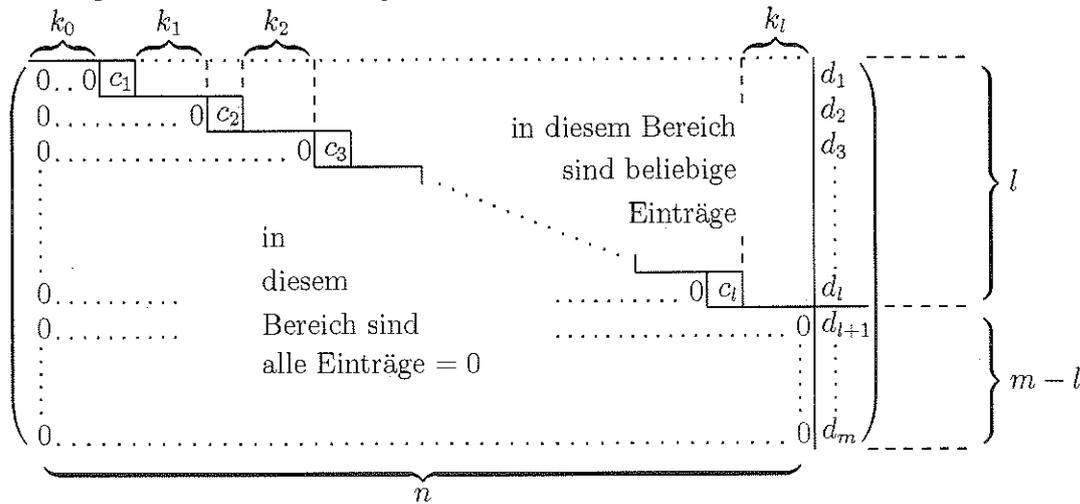
$$\begin{array}{ccc} \begin{array}{ccc|c} x & y & z & \\ \hline 1 & 2 & 0 & 0 \\ -1 & 0 & 1 & 2 \\ -1 & 4 & 0 & 6 \end{array} & \textcircled{2} \rightarrow \textcircled{2} + \textcircled{1} & \begin{array}{ccc|c} x & y & z & \\ \hline 1 & 2 & 0 & 0 \\ 0 & 2 & 1 & 2 \\ -1 & 4 & 0 & 6 \end{array} \\ \\ \textcircled{3} \rightarrow \textcircled{3} + \textcircled{1} & \begin{array}{ccc|c} x & y & z & \\ \hline 1 & 2 & 0 & 0 \\ 0 & 2 & 1 & 2 \\ 0 & 6 & 0 & 6 \end{array} & \textcircled{3} \rightarrow \textcircled{3} - 3 \cdot \textcircled{2} & \begin{array}{ccc|c} x & y & z & \\ \hline 1 & 2 & 0 & 0 \\ 0 & 2 & 1 & 2 \\ 0 & 0 & -3 & 0 \end{array} \end{array}$$

Im letzten Tableau hat die Koeffizientenmatrix sog. obere **Dreiecksform**, und das bedeutet, dass man die Lösungen des zugehörigen linearen Gleichungssystems von unten nach oben fortschreitend ablesen kann: Die unterste Gleichung lautet nun $-3z = 0$, also ist $z = 0$, die mittlere Gleichung ist $2y + z = 2$, also $y = \frac{1}{2}(2 - z) = 1$, und die erste Gleichung $x + 2y = 0$ gibt schließlich noch $x = -2y = -2$. Somit besitzt das letzte — und das ursprüngliche — Gleichungssystem die eindeutige Lösung $x = -2, y = 1, z = 0$. Statt der letzten Zeilenoperation hätten wir auch die zweite und dritte Spalte im Tableau vertauschen und direkt ablesen können $y = 1, z = 2 - 2y = 0, x = -2y = -2$. ■

Es zeigt sich, dass die Methode aus dem letzten Beispiel *immer* zum Ziel führt: Mit Zeilenoperationen lässt sich eine Art Dreiecksgestalt der Koeffizientenmatrix erreichen.

SATZ (über die Zeilen-Stufen-Form):

(i) Durch zulässige Zeilenoperationen kann die erweiterte Koeffizientenmatrix $(A|b)$ eines linearen Gleichungssystems stets auf eine Form gebracht werden, bei der (von links) der erste von Null verschiedene Eintrag in jeder Koeffizientenzeile später kommt als in der darüber liegenden Zeile. Diese sogenannte **Zeilen-Stufen-Form** hat also die Gestalt



mit Zahlen $0 \leq l \leq \min(m, n)$ und $k_0, k_1, \dots, k_l \in \mathbb{N}_0$, so dass $l + k_0 + \dots + k_l = n$ ist, und mit von Null verschiedenen Einträgen c_1, c_2, \dots, c_l in der vordersten Position jeder Stufe. (Rechts von den c_i können in jeder Zeile beliebige Einträge stehen; auch die Einträge d_i in der Spalte der rechten Seite sind beliebig.) Dabei hängen die Zahlen l (die Anzahl der Stufen) und k_0, k_1, \dots, k_l (die Breite der einzelnen Stufen) nur von der ursprünglichen Koeffizienten- $m \times n$ -Matrix A ab und nicht von der Art und Reihenfolge der Zeilenoperationen, mit denen die Zeilen-Stufen-Form hergestellt wurde. (Die Einträge in der Zeilen-Stufen-Form hängen dagegen von den gewählten Zeilenoperationen ab.)

(ii) Aus der Zeilen-Stufen-Form des linearen Gleichungssystems erhält man die Lösungen folgendermaßen: Für die sog. **Nicht-Basisvariablen**, das sind die Unbekannten, die zu einer Spalte gehören, welche nicht durch einen Stufenbeginn geht (also nicht durch eine der Positionen, auf denen die c_i eingetragen sind), kann man beliebige reelle Zahlen als Parameter vorgeben. Dazu lassen sich dann durch Auflösen der Gleichungen Nr. l bis Nr. 1 (von unten nach oben) eindeutige Werte der **Basisvariablen** berechnen, also der Unbekannten, die zu einer der Positionen der Einträge $c_i \neq 0$ gehören, derart dass sich insgesamt ein Lösungs- n -tupel ergibt. Im **Konsistenzfall** $d_{l+1} = 0, \dots, d_m = 0$ oder $l = m$ erhält man in dieser Weise die allgemeine, von $n-l$ Parametern abhängende bzw. im Fall $l = n$ eindeutige Lösung des Gleichungssystems. Im **Inkonsistenzfall** $l < m$ und $d_i \neq 0$ für ein $i > l$ hat das Gleichungssystem keine Lösung. ■

Dieses Ergebnis ist vielleicht nicht so einfach, wie man es sich wünschen würde; aber es gilt eben für *beliebige* lineare Gleichungssysteme und erfasst *alle* Fälle, die beim Lösen linearer Gleichungssysteme überhaupt eintreten können. Da man die Zeilen-Stufen-Form auch rechnerisch leicht herstellen kann (von Hand bei kleinen linearen Gleichungssystemen, mit einem Computer-Programm bei größeren), liefert der Satz zugleich eine effektive *Rechenmethode*, mit der man für beliebige lineare Gleichungssysteme die allgemeine Lösung berechnen bzw. die Nichtexistenz einer Lösung feststellen kann. In diesem Sinne beantwortet der Satz alle Fragen, die wir in 3.1 gestellt haben!

Der *Beweis* des Satzes besteht in folgendem auch praktisch anwendbaren Verfahren, das **Gaußsches Eliminationsverfahren** heißt, weil es zu Gleichungen führt, in denen eine Unbekannte "eliminiert" ist, d.h. nicht mehr auftritt (bzw. den Koeffizienten Null hat). Man wählt in der erweiterten Koeffizientenmatrix

$$(A|b) = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array} \right)$$

in Schritt 1 eine Zeile mit Eintrag $\neq 0$ an erster Position aus. (Die ausgewählte Zeile wird auch *Pivot-Zeile* genannt.) Durch Zeilenvertauschung bringt man nun diese Zeile nach oben, so dass danach der (neue) Eintrag $a_{11} \neq 0$ ist. Dann subtrahiert man für $i = 2 \dots n$ das a_{i1}/a_{11} -fache der ersten Zeile von der i -ten Zeile und erhält eine neue erweiterte Koeffizientenmatrix, in der die erste Spalte unter a_{11} nur noch Nulleinträge hat. Hat die erste Spalte von vorneherein nur Nulleinträge, so bestimmt man die Zahl k_0 der Nullspalten, die vor dem erstem Matrixkoeffizienten $\neq 0$ kommen, und führt die Prozedur mit der (k_0+1) -ten Spalte statt mit der 1-ten durch. (Sind alle Einträge von A Null, so ist man schon fertig.) Das Ergebnis dieses ersten Schrittes des Eliminationsverfahrens ist eine erweiterte Koeffizientenmatrix der Gestalt

$$\left(\begin{array}{cccc|ccc} 0 & \dots & 0 & c_1 & * & \dots & * & * \\ 0 & \dots & 0 & 0 & \boxed{\begin{array}{c} \tilde{A} \\ \tilde{b} \end{array}} & & & \\ \vdots & & \vdots & \vdots & & & & \\ 0 & \dots & 0 & 0 & & & & \end{array} \right) \quad \begin{array}{l} \text{mit } k_0 \in \mathbb{N}_0, c_1 \in \mathbb{R}_{\neq 0}, \text{ einer} \\ (m-1) \times (n-k_0-1)\text{-Matrix } \tilde{A}, \\ \text{einem Spaltenvektor } \tilde{b} \in \mathbb{R}^{m-1} \\ \text{und irgendwelchen Einträgen } * . \end{array}$$

Damit sind die ersten $k_0 + 1$ Unbekannten aus den Gleichungen Nr. 2 bis Nr. m eliminiert. An der ersten Gleichung, also an der obersten Zeile dieser Matrix, wird in den folgenden Schritten nichts mehr geändert.

In Schritt 2 wiederholt man nun die Prozedur mit der um eine Zeile und k_0+1 Spalten kleineren Matrix $(\tilde{A}|\tilde{b})$. Es ist klar, dass man so in weiteren Schritten fortfahrend nach endlich vielen Schritten die Zeilen-Stufen-Form erreicht hat und damit die Lösung(en) des Gleichungssystems wie im Satz angegeben berechnen kann. Die Anzahl l der erforderlichen Schritte ist offenbar nicht größer als die Zahl m der Gleichungen und auch nicht größer als die Zahl n der Unbekannten.

Bleibt noch zu überlegen, dass l und k_0, k_1, \dots, k_l nur von der ursprünglichen Koeffizientenmatrix A abhängen. Wäre das nicht so, so könnte man durch Zeilenoperationen aus dem homogenen Gleichungssystem $Ax = 0$ zwei verschiedene Zeilen-Stufen-Formen dieses Systems herstellen (die dann natürlich auch homogen sind), so dass eine Unbekannte x_j Nicht-Basisvariable bezüglich der ersten Zeilenstufenform ist, aber Basisvariable bezüglich der zweiten. Letzteres bedeutet, dass in jedem Lösungs- n -tupel x_1, \dots, x_n

mit $x_{j+1} = 0, \dots, x_n = 0$ auch $x_j = 0$ sein muss, ersteres aber, dass man Lösungs- n -tupel mit $x_{j+1} = 0, \dots, x_n = 0$ finden kann, bei denen x_j einen beliebig gegebenen reellen Wert annimmt. Da es es sich jeweils um dieselbe Lösungsmenge handelt, eben die zu $Ax = 0$, kann nicht beides zugleich richtig sein, d.h. alle Zeilen-Stufen-Formen von A haben dieselben Basis- und Nicht-Basisvariablen und damit auch dieselben Zahlen l und k_0, k_1, \dots, k_n . Zwei verschiedene Rechner können durchaus Matrizen mit unterschiedlichen Einträgen als Zeilen-Stufen-Form zu derselben erweiterten Koeffizientenmatrix $(A|b)$ erhalten; aber die Anzahl der Stufen und die Breite der einzelnen Stufen müssen — wenn korrekt gerechnet wurde — bei beiden dieselben sein!

Die Zahl l der Stufen (also der Koeffizientenzeilen mit mindestens einer von Null verschiedenen Eintragung) in der Zeilen-Stufen-Form beschreibt in gewisser Weise, wieviele der Gleichungen des ursprünglichen linearen Gleichungssystems $Ax = b$ unabhängig sind. Besteht das System z.B. aus einer einzigen linearen Gleichung, die man m -fach aufgeschrieben hat, so ergibt sich $l = 1$ (sofern nicht alle Koeffizienten Null sind und damit $l = 0$). Und besteht das System aus drei Gleichungen, von denen die dritte die Summe der beiden ersten Gleichungen ist oder daraus durch eine Zeilenoperation hervorgeht, so ist $l = 2$ (sofern nicht schon eine der ersten beiden Gleichungen Vielfaches der anderen ist, in welchem Falle man $l = 1$ erhält oder, wenn alle Koeffizienten Null sind, sogar $l = 0$). Allgemein sind alle linearen Gleichungssysteme, die man mit der Koeffizientenmatrix A aufstellen kann, schon äquivalent zu einem linearen Gleichungssystem, in dem nur Koeffizientenzeilen aus l geeignet ausgesuchten Zeilen von A auftreten, aber im Allgemeinen nicht zu einem mit weniger als l Zeilen von A . Weil die Zahl l von fundamentaler Bedeutung für die Diskussion des linearen Gleichungssystems ist, führt man einen Namen dafür ein:

DEFINITION: Die Zahl l der Stufen, also der Zeilen mit mindestens einer von Null verschiedenen Eintragung, die sich bei Herstellung der Zeilen-Stufen-Form der Koeffizientenmatrix A eines linearen Gleichungssystems ergibt, heißt die **(maximale) Anzahl der linear unabhängigen Gleichungen** des Gleichungssystems oder der **Zeilenrang** der Koeffizientenmatrix A . ■

Bevor wir einige konkrete Beispiele zur Herstellung der Zeilen-Stufen-Form und Lösung von linearen Gleichungssystemen durchrechnen, diskutieren wir noch Konsequenzen für die allgemeine Theorie linearer Gleichungssysteme.

DISKUSSION (*Konsequenzen der Zeilen-Stufen-Form für lineare Gleichungssysteme*):

1) Gegeben sei ein System $Ax = b$ von m linearen Gleichungen für n Unbekannte, und die Zahl der linear unabhängigen Gleichungen dabei sei $l \leq \min(m, n)$. Dann besteht folgende dreiteilige Alternative:

- Entweder liegt ein **wohlgestelltes Problem** vor, d.h. $Ax = b$ hat genau eine Lösung;
- oder ein **inkonsistentes Problem**, d.h. $Ax = b$ hat keine Lösung;
- oder ein **konsistentes, aber nicht wohlgestelltes Problem**, und in diesem Fall hat $Ax = b$ eine unendliche Schar von Lösungen.

Das liest man aus dem Satz über die Zeilen-Stufen-Form unmittelbar ab. Die Information, die diese Alternative zunächst einmal beinhaltet, ist, dass der Fall von mehr als einer,

aber insgesamt nur endlich vielen Lösungen bei einem linearen Gleichungssystem nicht auftreten kann. Anders als z.B. bei Systemen von quadratischen Gleichungen gibt es also bei linearen Gleichungssystemen nicht den Fall, dass man genau zwei, oder genau drei, oder genau vier ... Lösungen hat! An der Zeilen-Stufen-Form kann man außerdem unmittelbar ablesen, welcher der drei Fälle eintritt.

Der Fall, der für Anwendungen am wichtigsten ist, ist natürlich der erste — wenn es keine Lösung gibt, so kann man mit dem Gleichungssystem in einer Anwendung nichts anfangen (außer der Erkenntnis, dass man eine unmögliche Situation beschrieben hat); wenn es unendlich viele Lösungen gibt, so hat man das Problem, welche von diesen die für die ins Auge gefasste Anwendung die "richtige" oder "beste" ist. Zur Modellierung eines Problems, das aus einem Anwendungsgebiet kommt, taugt also ein lineares Gleichungssystem nur, wenn es wohlgestellt ist, d.h. genau eine Lösung hat. Der Satz über die Zeilen-Stufen-Form sagt uns, dass dies genau dann der Fall ist, wenn erstens $n = l$ ist, wenn man also ebenso viele unabhängige Gleichungen wie Unbekannte hat, und wenn zweitens auch $m = l$ ist oder alle rechten Seiten d_j mit $j > m$ verschwinden. Insbesondere kann ein wohlgestelltes Problem nur vorliegen, wenn man mindestens so viele lineare Gleichungen wie Unbekannte hat und, genauer gesagt, ebenso viele unabhängige Gleichungen wie Unbekannte. Die Zeilen-Stufen-Form hat bei einem wohlgestellten Problem die Gestalt

$$\left(\begin{array}{ccc|c} c_1 & & & d_1 \\ & c_2 & * & d_2 \\ & & \ddots & \vdots \\ 0 & & & c_n \\ & & & d_n \end{array} \right) \quad \begin{array}{l} \text{mit } c_i \in \mathbb{R}_{\neq 0}, d_i \in \mathbb{R}, \\ \text{beliebigen Einträgen } * \\ \text{und Nulleinträgen } 0 \end{array}$$

eventuell noch mit $(n+1)$ -gliedrigen Nullzeilen darunter. Man sagt, dass die Koeffizientenmatrix eine **obere Dreiecksmatrix** mit Diagonaleinträgen $\neq 0$ ist. Die Auflösung eines Gleichungssystems dieser Form von unten nach oben ist offenbar eindeutig möglich.

Der Inkonsistenzfall kann im Prinzip immer eintreten, wenn das lineare Gleichungssystem inhomogen ist, mit Ausnahme des Falls $l = m = n$, bei dem man ebenso viele Gleichungen wie Unbekannte hat und die Gleichungen alle linear unabhängig sind. In diesem letzteren Fall hat man ja für jede Wahl der rechten Seiten b ein wohlgestelltes Problem. Bei einem homogenen linearen Gleichungssystem kommt der Inkonsistenzfall dagegen nie vor; denn es gibt ja immer die triviale Lösung.

Im dritten Fall gibt uns der Satz über die Zeilen-Stufen-Form eine sehr viel präzisere Information als nur Unendlichkeit der Lösungsmenge. Dieser Fall kann nur eintreten, wenn $l < n$ ist, wenn man also *weniger unabhängige Gleichungen als Unbekannte hat*, und dann sind $n-l$ der Unbekannten (die Nicht-Basisvariablen) beliebig als reelle Parameter wählbar und dazu die restlichen Unbekannten (die Basisvariablen) eindeutig berechenbar, so dass sich jeweils ein Lösungs- n -tupel ergibt. Man sagt, dass die Lösungen eine **spezielle $(n-l)$ -parametrische lineare Schar L** bilden; für $n-l = 1, 2, 3, \dots$ ist die Lösungsmenge L geometrisch gesprochen eine Gerade, eine Ebene, ein 3-dimensionaler Raum ... in \mathbb{R}^n .

2) Wenn man fragt, wann ein lineares Gleichungssystem $Ax = b$ mit Koeffizientenmatrix A vom Format (m, n) für *alle* rechten Seiten b lösbar ist, so lautet die Antwort:

- *Das System $Ax = b$ ist genau dann für jede Vorgabe der rechten Seite $b \in \mathbb{R}^m$ lösbar, wenn $l = m$ ist, wenn also alle Gleichungen des Systems unabhängig sind; das ist nur möglich, wenn $m \leq n$ ist, wenn man also höchstens so viele Gleichungen hat wie Unbekannte.*

Das ist klar, weil man genau im Fall $l < m$ ja rechte Seiten d_i in der Zeilen-Stufen-Form wählen kann, die zu einem inkonsistenten Gleichungssystem führen (nämlich $d_i \neq 0$ für ein $i > l$) und weil man daraus durch Rückrechnen der Zeilentransformationen ein inkonsistentes System $Ax = b$ erhält. Wenn man andererseits fragt, unter welchen Bedingungen an A die Lösung immer eindeutig ist, wenn sie existiert, so lautet die Antwort:

- *Das System $Ax = b$ hat genau dann für jede Vorgabe der rechten Seite $b \in \mathbb{R}^m$ höchstens eine Lösung, wenn $l = n$ ist, wenn es also genau so viele unabhängige Gleichungen wie Unbekannte gibt; das ist nur möglich, wenn $m \geq n$ ist, wenn man also mindestens so viele Gleichungen hat wie Unbekannte.*

Das ist klar, weil es genau im Fall $l < n$ Nicht-Basisvariablen gibt, deren Werte im Lösungs- n -tupel frei wählbar sind. Die Lösungsmenge zu $Ax = b$ ist für jede rechte Seite b dann entweder leer oder eine unendliche $(n-l)$ -parametrische Schar.

3) Besonders wichtig ist der **Spezialfall $m = n$** , d.h. *genau so viele Gleichungen wie Unbekannte*. Dann lautet die Alternative so:

- *Sind die Gleichungen linear unabhängig ($l = n = m$), so hat das System $Ax = b$ für jede rechte Seite $b \in \mathbb{R}^n$ eine eindeutige Lösung.*
- *Sind die Gleichungen aber linear abhängig ($l < n = m$), so ist das System nicht für jede rechte Seite b lösbar und die Lösungen bilden, wenn welche existieren, eine spezielle $(n-l)$ -parametrische lineare Schar in \mathbb{R}^n .*

4) Ein anderer wichtiger Spezialfall ist der **homogene Fall**, d.h. es liegt ein System $Ax = 0$ von m linearen Gleichungen für n Unbekannte vor, bei dem alle rechten Seiten Null sind. Dann hat man stets mindestens die triviale Lösung $x = (0, \dots, 0)$, also lautet die Alternative aus 1) nun, wenn wieder l die Zahl der unabhängigen Gleichungen bezeichnet:

- *Entweder: Es ist $l = n$, d.h. es gibt ebenso viele unabhängige Gleichungen wie Unbekannte, und dann ist die triviale Lösung die einzige Lösung. Dieser Fall kann nur eintreten, wenn $m \geq n$ ist, wenn man also nicht weniger homogene lineare Gleichungen hat als Unbekannte.*
- *Oder: Es ist $l < n$, d.h. es gibt weniger unabhängige Gleichungen als Unbekannte, und dann bilden die Lösungen eine unendliche $(n-l)$ -parametrische lineare Schar. Dieser Fall liegt immer vor, wenn $m < n$ ist, wenn man also insgesamt weniger homogene lineare Gleichungen hat als Unbekannte.*

5) Zwischen homogenen und inhomogenen linearen Gleichungssystemen mit derselben Koeffizientenmatrix A besteht folgender Zusammenhang:

- *Die allgemeine Lösung des inhomogenen Systems erhält man (wenn es überhaupt eine Lösung hat), indem man zu einer speziellen Lösung die allgemeine Lösung des homogenen Systems addiert.*

Das bedeutet Folgendes: Ist $x^* = (x_1^*, \dots, x_n^*)$ irgendein Lösungs- n -tupel zu $Ax = b$ (das ist mit einer "speziellen Lösung" gemeint), so erhält man alle Lösungen dieses Gleichungssystems in der Form $x = x^* + y := (x_1^* + y_1, \dots, x_n^* + y_n)$ (man nennt das die *komponentenweise gebildete Summe* der Spaltenvektoren x^* und y in \mathbb{R}^n), wobei y ein Lösungs- n -tupel des homogenen Gleichungssystems $Ay = 0$ ist. Der Beweis ist einfach: $Ax = b \iff Ax = Ax^* \iff a_{i1}x_1 + \dots + a_{in}x_n = a_{i1}x_1^* + \dots + a_{in}x_n^*$ für $i = 1 \dots m \iff a_{i1}(x_1 - x_1^*) + \dots + a_{in}(x_n - x_n^*) = 0$ für $i = 1 \dots m \iff y := (x_1 - x_1^*, \dots, x_n - x_n^*)$ löst $Ay = 0 \iff x_j = x_j^* + y_j$ für $j = 1 \dots n$ und ein $y \in \mathbb{R}^n$ mit $Ay = 0$.

6) Allgemein nennen wir eine nichtleere Menge $L \subset \mathbb{R}^n$ von n -tupeln eine **k -parametrische lineare Schar oder Menge in \mathbb{R}^n** , wenn es n affin lineare Funktionen $\ell_j(r_1, \dots, r_k)$ gibt, so dass die n -tupel $(\ell_1(r_1, \dots, r_k), \dots, \ell_n(r_1, \dots, r_k))$ alle Elemente der Menge L durchlaufen (und keine anderen), wenn die Parameter (r_1, \dots, r_k) alle reellen k -tupel durchlaufen. Dabei ist die Anzahl k der Parameter durch die Menge L nicht eindeutig bestimmt, weil man sich immer zusätzliche Parameter r_{k+1}, r_{k+2}, \dots denken kann, die in den linearen Funktionen ℓ_j nur mit Nullkoeffizienten auftreten. Ist die Zahl k der Parameter minimal, also dieselbe Menge L nicht auch als $(k-1)$ -parametrische lineare Schar darstellbar, so sprechen wir von einer **k -dimensionalen linearen Schar**. Die **Dimension** einer linearen Menge ist also die minimale Zahl k der Parameter, die erforderlich ist, um sie als lineare Schar zu parametrisieren. Die Dimension k liegt z.B. vor, wenn wir eine **spezielle k -parametrische lineare Schar** haben, bei der die Parameter Komponenten der n -tupel in L sind, d.h. es ist $\ell_{j_h}(r_1, \dots, r_k) = r_h$ für gewisse Zahlen $j_1 < j_2 < \dots < j_k$ in $\{1, \dots, n\}$. Diese Situation haben wir gerade mit $k := n - l$ bei der Lösungsmenge eines konsistenten linearen Gleichungssystems für n Unbekannte, wenn die Anzahl l der unabhängigen Gleichungen kleiner als die der Unbekannten ist und man somit $k = n - l$ Nicht-Basisvariable als freie Parameter im Lösungs- n -tupel wählen kann. Mit affin linearen Funktionen $\tilde{\ell}_j(s_1, \dots)$ von weniger als k Parametern s_i kann man nämlich L nicht beschreiben, weil sonst $\tilde{\ell}_{j_h}(s_1, \dots) = r_h$, $h = 1 \dots k$, ein für alle Wahlen der rechten Seiten r_h lösbares System von k linearen Gleichungen für weniger als k Unbekannte wäre, was nach 2) unmöglich ist. Wenn wir noch für $k = 0$ vereinbaren, unter einer 0-dimensionalen linearen Schar in \mathbb{R}^n ein Teilmenge zu verstehen, die aus genau einem n -tupel besteht, so können wir also sagen:

- Die Lösungsmenge L eines konsistenten linearen Gleichungssystems mit l unabhängigen Gleichungen für n Unbekannte ist eine $(n-l)$ -dimensionale lineare Menge,

d.h. L lässt sich als $(n-l)$ -parametrische lineare Menge in \mathbb{R}^n beschreiben, aber nicht als lineare Schar mit weniger als $n-l$ Parametern. In der mathematischen Fachterminologie nennt man eine k -dimensionale lineare Teilmenge L von \mathbb{R}^n einen **k -dimensionalen affinen Unterraum** von \mathbb{R}^n und, wenn der Nullvektor $0 = (0, \dots, 0)$ in L liegt, einfach einen **k -dimensionalen (Vektor-)Unterraum** von \mathbb{R}^n . Geometrisch gesprochen handelt es sich dabei für $k = 0, 1, 2, 3, \dots$ um einen Punkt, eine Gerade, eine Ebene, einen 3-dimensionalen Raum, ... in \mathbb{R}^n , worin im Fall eines Vektorunterraums der Ursprung $0 = (0, \dots, 0)$ von \mathbb{R}^n enthalten ist.

7) Der Satz über die Zeilen-Stufen-Form ermöglicht auch eine genaue Beschreibung der Menge Z der **konsistenten rechten Seiten** zu einer gegebenen Koeffizienten- $m \times n$ -Matrix A , d.h. der Menge der $b \in \mathbb{R}^m$, für die eine Lösung x zu $Ax = b$ existiert. Wenn man die Zeilenoperationen, die zu einer Zeilen-Stufen-Form führen, mit einem allgemeinen Spaltenvektor (b_1, \dots, b_m) von rechten Seiten durchführt, so ergeben sich die Einträge d_{l+1}, \dots, d_m der letzten Spalte der Zeilen-Stufen-Form als lineare Funktionen von b_1, \dots, b_m in der Form $d_{l+h} = \sum_{i=1}^m c_{hi} b_i$, wobei die Koeffizienten c_{hi} durch die ausgeführten Zeilen-Operationen bestimmt sind. Die $b \in \mathbb{R}^m$, für die $Ax = b$ lösbar ist, sind dann genau die Lösungen des linearen Gleichungssystems $Cb = 0$ mit der $(m-l) \times m$ -Koeffizientenmatrix $C = (c_{hi})$. Da man die Zeilenoperationen wieder rückgängig machen kann, lässt sich das System $Cb = d$ für jede rechte Seite $d \in \mathbb{R}^{m-l}$ lösen, also sind nach 2) die Gleichungen dieses Systems unabhängig. Die Lösungsmenge ist daher eine unendliche l -dimensionale lineare Schar von Spaltenvektoren $b \in \mathbb{R}^m$, bzw. besteht im Fall $l = 0$ (d.h. A ist die Nullmatrix) nur aus der Nullspalte $0 \in \mathbb{R}^m$. Eine konkrete Parametrisierung der

konsistenten rechten Seiten b in der Gestalt $b_i = \sum_{k=1}^l b_{ik}d_k$ für $i = 1 \dots m$ mit Parametern $d_1, \dots, d_l \in \mathbb{R}$ erhält man, wenn man mit rechten Spalten der Form $(d_1, \dots, d_l, 0, \dots, 0)$ in der Zeilen-Stufen-Form die ausgeführten Zeilenoperationen zurückrechnet.

- Die konsistenten rechten Seiten für ein System $Ax = b$ von m linearen Gleichungen bilden eine l -dimensionale lineare Schar Z in \mathbb{R}^m , wobei l der Zeilenrang von A ist, also die maximale Zahl der unabhängigen Gleichungen. Eine lineare Parametrisierung von Z mit l Parametern und ein System von $m-l$ unabhängigen homogenen linearen Gleichungen für m Unbekannte, das Z als Lösungsmenge hat, kann mit den Zeilenoperationen, die zur Zeilen-Stufen-Form führen, explizit berechnet werden.

8) Zusammen mit der Aussage aus 6), dass die Lösungsmenge L zu $Ax = 0$ eine lineare Menge der Dimension $n-l$ ist, ergibt sich ein fundamentaler Satz der Linearen Algebra: Die Dimensionen von L und Z addieren sich stets zu $(n-l)+l = n$. Dies ist der sogenannte

- **Rangatz:** Für jede $m \times n$ -Matrix A addieren sich die Dimensionen des Lösungsraumes L des homogenen Gleichungssystems $Ax = 0$ und des Raumes Z der konsistenten rechten Seiten für $Ax = b$ zur Spaltenzahl n von A .

Grob gesprochen kann man sagen, dass $Ax = 0$ um so weniger / mehr Lösungen hat, je mehr / weniger zulässige rechte Seiten b existieren. ■

BEISPIELE (zur Zeilen-Stufen-Form und Lösung linearer Gleichungssysteme):
Zunächst eine Vorbemerkung darüber,

- was man zuerst tut,

wenn ein lineares Gleichungssystem vorgelegt ist: Man zählt zunächst, *wieviele Gleichungen man hat und wieviele Unbekannte* und dann schaut man auf die rechte Seite, um festzustellen, ob das Gleichungssystem *homogen ist oder nicht*. Als Zweites sollte man überlegen, welche Informationen man aus der allgemeinen Theorie linearer Gleichungssysteme nun über das vorgelegte Gleichungssystem schon hat. Nur wenn man mindestens so viele Unbekannte wie Gleichungen hat, ist Lösbarkeit für jede rechte Seite zu erhoffen. Nur wenn man höchstens so viele Unbekannte wie Gleichungen hat, kann man die Eindeutigkeit der Lösung erhoffen. "Erhoffen" muss hier gesagt werden, weil es nicht sicher ist. Das Problem ist, dass die Gleichungen linear abhängig sein könnten. Das wird bei einem Problem aus den Anwendungen zwar kaum vorkommen, weil es im Konsistenzfall bedeutet, dass man eigentlich überflüssige Gleichungen aufgeschrieben hat, die aus den anderen folgen, und im Inkonsistenzfall, dass man Gleichungen aufgeschrieben hat, die anderen widersprechen, es ist aber (zumindest bei einem System mit mehr als zwei Gleichungen) nicht unmittelbar durch Inspektion der Koeffizientenmatrix zu erkennen. Daher muss man als Drittes die Zeilen-Stufen-Form herstellen, mit der sich die Frage der Abhängigkeit und der Konsistenz des Gleichungssystems entscheiden und die Lösung (bzw. die lineare Schar der Lösungen) auch berechnen lässt.

$$\begin{array}{rcl}
 1) & 2x & - z = 0 \\
 & -x + 2y & = -2 \\
 & -3x - y + 2z & = 2
 \end{array}$$

ist ein System von 3 linearen Gleichungen für 3 Unbekannte x, y, z (also $m = n = 3$). Wir erwarten also, dass es genau eine Lösung gibt — es sei denn, die drei Gleichungen sind abhängig. (Das sieht zwar nicht so aus, muss aber noch nachgeprüft werden.) Wir stellen das Tableau der erweiterten Koeffizientenmatrix auf, wobei wir gleich die erste

mit der zweiten Zeile vertauschen, und machen naheliegende Zeilentransformationen zur Herstellung der Zeilen-Stufen-Form:

$$\begin{array}{ccc} \begin{array}{ccc|c} -1 & 2 & 0 & -2 \\ 2 & 0 & -1 & 0 \\ -3 & -1 & 2 & 2 \end{array} & \begin{array}{c} \iff \\ \textcircled{2} \rightarrow \textcircled{2} + 2 \cdot \textcircled{1} \\ \textcircled{3} \rightarrow \textcircled{3} - 3 \cdot \textcircled{1} \end{array} & \begin{array}{ccc|c} -1 & 2 & 0 & -2 \\ 0 & 4 & -1 & -4 \\ 0 & -7 & 2 & 8 \end{array} & \begin{array}{c} \iff \\ \textcircled{3} \rightarrow \textcircled{3} + \frac{7}{4} \cdot \textcircled{2} \end{array} & \begin{array}{ccc|c} -1 & 2 & 0 & -2 \\ 0 & 4 & -1 & -4 \\ 0 & 0 & \frac{1}{4} & 1 \end{array} \end{array}$$

Die anfängliche Vertauschung der beiden ersten Zeilen, also die Wahl der zweiten als Pivot-Zeile, haben wir vorgenommen, weil so das Auftreten von Brüchen bei den ersten beiden Zeilenoperationen vermieden wird. (Es ist immer günstig, als Pivot-Element 1 oder -1 zu haben.) Das letzte System ist in Zeilen-Stufen-Form, wobei die Koeffizientenmatrix obere Dreiecksgestalt mit Diagonaleinträgen $\neq 0$ hat. Also sind die drei gegebenen linearen Gleichungen tatsächlich unabhängig, und wir haben eine eindeutige Lösung, die wir aus dem letzten Tableau von unten nach oben ablesen können:

$$z = 4, \quad y = \frac{1}{4}(-4 + z) = 0, \quad x = 2 + 2y = 2.$$

Um die Nullen in der ursprünglichen Koeffizientenmatrix besser auszunutzen, hätten wir auch zu Beginn die erste mit der dritten Spalte vertauschen und wie folgt rechnen können, wobei wir die Vertauschung der Variablen in der Kopfzeile festhalten:

$$\begin{array}{ccc} \begin{array}{ccc|c} z & y & x & \\ -1 & 0 & 2 & 0 \\ 0 & 2 & -1 & -2 \\ 2 & -1 & -3 & 2 \end{array} & \begin{array}{c} \iff \\ \textcircled{3} \rightarrow \textcircled{3} + 2 \cdot \textcircled{1} \end{array} & \begin{array}{ccc|c} z & y & x & \\ -1 & 0 & 2 & 0 \\ 0 & 2 & -1 & -2 \\ 0 & -1 & 1 & 2 \end{array} & \begin{array}{c} \iff \\ \textcircled{3} \rightarrow 2 \cdot \textcircled{3} + \textcircled{2} \end{array} & \begin{array}{ccc|c} z & y & x & \\ -1 & 0 & 2 & 0 \\ 0 & 2 & -1 & -2 \\ 0 & 0 & 1 & 2 \end{array} \end{array}$$

Hieraus lesen wir erneut von unten nach oben die eindeutige Lösung $x = 2, y = 0, z = 4$ ab, weil uns die Kopfzeile daran erinnert, dass die letzte Spalte nun zur Unbekannten x gehört und die erste zu z . (Ohne die Kopfzeile festzuhalten, hätte es hier leicht passieren können, dass man die falsche Lösung $z = 2, y = 0, x = 4$ angibt!)

2)

$$\begin{array}{rcl} 2x & - & z = 0 \\ -x + y & & = -2 \\ -3x - y + 2z & & = 2 \end{array}$$

Gegenüber 1) ist hier nur der Koeffizient a_{22} , also der Koeffizient von y in der zweiten Gleichung, von 2 auf 1 abgeändert. Wenn wir genau dieselben Zeilenoperationen wie in 1) durchführen, wobei wir nur zu verfolgen brauchen, wie diese Änderung sich auswirkt, so verläuft die Rechnung nun so:

$$\begin{array}{ccc} \begin{array}{ccc|c} -1 & 1 & 0 & -2 \\ 2 & 0 & -1 & 0 \\ -3 & -1 & 2 & 2 \end{array} & \begin{array}{c} \iff \\ \textcircled{2} \rightarrow \textcircled{2} + 2 \cdot \textcircled{1} \\ \textcircled{3} \rightarrow \textcircled{3} - 3 \cdot \textcircled{1} \end{array} & \begin{array}{ccc|c} -1 & 1 & 0 & -2 \\ 0 & 2 & -1 & -4 \\ 0 & -4 & 2 & 8 \end{array} & \begin{array}{c} \iff \\ \textcircled{3} \rightarrow \textcircled{3} + 2 \cdot \textcircled{2} \end{array} & \begin{array}{ccc|c} -1 & 1 & 0 & -2 \\ 0 & 2 & -1 & -4 \\ 0 & 0 & 0 & 0 \end{array} \end{array}$$

Die Zeilen-Stufen-Form enthält nun eine Nullzeile, d.h. von den ursprünglichen drei linearen Gleichungen sind nur zwei unabhängig. (Tatsächlich sieht man jetzt — nachdem die Abhängigkeit festgestellt ist —, dass z.B. die erste Gleichung das $(-\frac{1}{2})$ -fache der Summe von zweiter und dritter Gleichung ist.) Da auch die rechte Seite der letzten Zeile Null ist, liegt der Konsistenzfall vor, wir können die Nicht-Basisvariable z als beliebigen Parameter $t \in \mathbb{R}$ wählen und erhalten dann $y = \frac{1}{2}(t - 4) = \frac{1}{2}t - 2$ und $x = y + 2 = \frac{1}{2}t$ aus der zweiten und ersten Gleichung des letzten Tableaus. Die allgemeine Lösung ist also die 1-dimensionale lineare Schar $(x, y, z) = (\frac{1}{2}t, \frac{1}{2}t - 2, t)$ ($t \in \mathbb{R}$), die geometrisch eine Gerade im 3-dimensionalen Zahlenraum \mathbb{R}^3 beschreibt.

Wenn wir wie in 1) zuerst die Variablen x und z vertauschen, so liefert die Rechnung

$$\begin{array}{c|ccc|c} z & y & x & \\ \hline -1 & 0 & 2 & 0 \\ 0 & 1 & -1 & -2 \\ 2 & -1 & -3 & 2 \end{array} \quad \begin{array}{l} \iff \\ \textcircled{3} \rightarrow \textcircled{3} + 2 \cdot \textcircled{1} \end{array} \quad \begin{array}{c|ccc|c} z & y & x & \\ \hline -1 & 0 & 2 & 0 \\ 0 & 1 & -1 & -2 \\ 0 & -1 & 1 & 2 \end{array} \quad \begin{array}{l} \iff \\ \textcircled{3} \rightarrow \textcircled{3} + \textcircled{2} \end{array} \quad \begin{array}{c|ccc|c} z & y & x & \\ \hline -1 & 0 & 2 & 0 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 \end{array}$$

nun die allgemeine Lösung $x = r$, $y = r - 2$, $z = 2x = 2r$ und damit eine andere 1-Parameter-Darstellung derselben Lösungsmenge $(x, y, z) = (r, r - 2, 2r)$ ($r \in \mathbb{R}$).

Es gibt immer unendlich viele verschiedene solche linearen Parameterdarstellungen derselben unendlichen Lösungsmenge; die Anzahl der auftretenden Parameter ist aber dieselbe — kleinstmögliche —, wenn man die Parameterdarstellungen wie hier aus einer Zeilen-Stufen-Form mit Wahl der Nicht-Basisvariablen als Parameter ausrechnet. Man kann als Parameter auch die Variable $y = s$ wählen und die allgemeine Lösung damit in der Form $(x, y, z) = (s + 2, s, 2s + 4)$ ($s \in \mathbb{R}$) darstellen. Nach Spaltenvertauschungen ergibt sich nicht immer, wie hier, die Zeilen-Stufen-Form mit genau denselben Stufenlängen. Die Anzahl l der Stufen ist aber stets dieselbe, weil diese ja nur von der Lösungsmenge abhängt. $(n-l)$ ist ja die Dimension der Lösungsmenge im konsistenten Fall, also die erforderliche Mindestzahl von Parametern in einer linearen Parameterdarstellung der Lösungsmenge.

3) Wir bleiben bei dem Gleichungssystem 2), ändern aber die rechte Seite ab:

$$\begin{array}{rcl} 2x & -z & = 0 \\ -x + y & & = -1 \\ -3x - y + 2z & & = 2 \end{array}$$

Wenn wir genau dieselben Zeilenoperationen wie in 1) durchführen, wobei wir wieder nur zu verfolgen brauchen, wie diese Änderung sich auswirkt, so ergibt sich nun:

$$\begin{array}{c|ccc|c} -1 & 1 & 0 & -1 \\ 2 & 0 & -1 & 0 \\ -3 & -1 & 2 & 2 \end{array} \quad \begin{array}{l} \iff \\ \textcircled{2} \rightarrow \textcircled{2} + 2 \cdot \textcircled{1} \\ \textcircled{3} \rightarrow \textcircled{3} - 3 \cdot \textcircled{1} \end{array} \quad \begin{array}{c|ccc|c} -1 & 1 & 0 & -1 \\ 0 & 2 & -1 & -2 \\ 0 & -4 & 2 & 8 \end{array} \quad \begin{array}{l} \iff \\ \textcircled{3} \rightarrow \textcircled{3} + 2 \cdot \textcircled{2} \end{array} \quad \begin{array}{c|ccc|c} -1 & 1 & 0 & -1 \\ 0 & 2 & -1 & -2 \\ 0 & 0 & 0 & 4 \end{array}$$

Die letzte Gleichung hat lauter Nullkoeffizienten, aber rechte Seite $\neq 0$, also liegt der Inkonsistenzfall vor, das Gleichungssystem hat keine Lösung. Dasselbe Ergebnis hätten wir auch für jede andere rechte Seite als -2 in der zweiten Gleichung (entspricht der ersten Zeile des Tableaus) erhalten, wenn wir die rechten Seiten in den anderen Gleichungen beibehalten. Führen wir die Rechnung mit allgemeinen rechten Seiten b_1, b_2, b_3 in den ursprünglichen Gleichungen durch, so verläuft sie so:

$$\begin{array}{c|ccc|c} -1 & 1 & 0 & b_2 \\ 2 & 0 & -1 & b_1 \\ -3 & -1 & 2 & b_3 \end{array} \quad \iff \quad \begin{array}{c|ccc|c} -1 & 1 & 0 & b_2 \\ 0 & 2 & -1 & b_1 + 2b_2 \\ 0 & -4 & 2 & b_3 - 3b_2 \end{array} \quad \iff \quad \begin{array}{c|ccc|c} -1 & 1 & 0 & b_2 \\ 0 & 2 & -1 & b_1 + 2b_2 \\ 0 & 0 & 0 & 2b_1 + b_2 + b_3 \end{array}$$

Der Konsistenzfall liegt also genau dann vor, wenn $2b_1 + b_2 + b_3 = 0$ ist, und diese Gleichung beschreibt die konsistenten rechten Seiten für das betrachtete Gleichungssystem. Wenn wir andererseits im letzten Tableau als rechte Seite $(r, s, 0)$ einsetzen und die Zeilentransformationen zurückrechnen, so bekommen wir:

$$\begin{array}{c|ccc|c} -1 & 1 & 0 & r \\ 0 & 2 & -1 & s \\ 0 & 0 & 0 & 0 \end{array} \quad \iff \quad \begin{array}{c|ccc|c} -1 & 1 & 0 & r \\ 0 & 2 & -1 & s \\ 0 & -4 & 2 & -2s \end{array} \quad \iff \quad \begin{array}{c|ccc|c} -1 & 1 & 0 & r \\ 2 & 0 & -1 & -2r + s \\ -3 & -1 & 2 & 3r - 2s \end{array}$$

In dieser Weise haben wir aus der Zeilen-Stufen-Form sowohl eine Beschreibung durch eine lineare Gleichung

$$Z = \{(b_1, b_2, b_3) \in \mathbb{R}^3 : 2b_1 + b_2 + b_3 = 0\}$$

als auch eine lineare 2-parametrische Darstellung

$$Z = \{(s-2r, r, 3r-2s) : r, s \in \mathbb{R}\}$$

für die Menge Z der konsistenten rechten Seiten des Gleichungssystems hergeleitet, also für

$$Z := \left\{ \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} : \begin{array}{l} 2x - z = b_1 \\ -x + y = b_2 \\ -3x - y + 2z = b_3 \end{array} \text{ hat eine Lösung} \right\}.$$

Die Menge Z ist hier eine Ebene im Raum \mathbb{R}^3 aller möglichen rechten Seiten (b_1, b_2, b_3) . Wenn wir also eine rechte Seite zufällig wählen oder eine konsistente rechte Seite wie $(0, -2, 2)$ in zufälliger Weise stören, so ist die Wahrscheinlichkeit dafür, dass wir ein lösbares Gleichungssystem erhalten, gleich Null. Das ist immer so, wenn ein lineares Gleichungssystem nicht für jede rechte Seite gelöst werden kann.

$$4) \quad \begin{array}{r} 3x - 2y + z = 2 \\ x - 2z = 3 \end{array}$$

Hier haben wir ein System von 2 linearen Gleichungen für 3 Unbekannte x, y, z (also $m = 2, n = 3$). Da es weniger Gleichungen als Unbekannte sind, wissen wir von vornherein, dass es keine eindeutige Lösung haben kann, sondern entweder gar keine oder eine unendliche Schar von Lösungen. Der Inkonsistenzfall kann nur vorliegen, wenn die Gleichungen abhängig sind. Bei nur zwei Gleichungen wie hier kann man das "mit bloßem Auge" erkennen: Abhängigkeit liegt genau dann vor, wenn man die Koeffizienten der einen Gleichung durch Multiplikation der Koeffizienten der anderen Gleichung mit demselben Faktor erhält —, was hier offenbar nicht geht. Also gibt es für jede rechte Seite eine 1-dimensionale lineare Schar von Lösungen. Die Zeilen-Stufen-Form bestätigt das:

$$\begin{array}{ccc} \begin{array}{c|c} 3 & -2 & 1 & | & 2 \\ 1 & 0 & -2 & | & 3 \end{array} & \begin{array}{c} \iff \\ \textcircled{1} \leftrightarrow \textcircled{2} \end{array} & \begin{array}{c|c} 1 & 0 & -2 & | & 3 \\ 3 & -2 & 1 & | & 2 \end{array} & \begin{array}{c} \iff \\ \textcircled{2} \rightarrow \textcircled{2} - 3 \cdot \textcircled{1} \end{array} & \begin{array}{c|c} 1 & 0 & -2 & | & 3 \\ 0 & -2 & 7 & | & -7 \end{array} \end{array}$$

(Der erste Schritt vermeidet Brüche.) Die Nicht-Basisvariable ist hier z , und wenn wir diese als beliebigen Parameter $z = t$ wählen, so ergibt sich $y = \frac{1}{2}(7z + 7) = \frac{7}{2}(t + 1)$ aus der unteren Gleichung und $x = 2z + 3 = 2t + 3$ aus der oberen. Die Lösungsmenge ist also

$$L = \{(x, y, z) \in \mathbb{R}^3 : x = 2t + 3, y = \frac{7}{2}(z+1), z = t \text{ für ein } t \in \mathbb{R}\}.$$

Dies ist eine Gerade im 3-dimensionalen Raum \mathbb{R}^3 . Geometrisch kann man das so verstehen: Jede der beiden linearen Gleichungen beschreibt eine Ebene in \mathbb{R}^3 , und die Lösungsmenge des Gleichungssystems ist der Durchschnitt der beiden Ebenen. Die Unabhängigkeit der Gleichungen bedeutet, dass die beiden Ebenen nicht zusammenfallen oder parallel sind; daher ist der Durchschnitt eine Gerade. Wären die beiden Gleichungen aber linear abhängig, so hätten wir (wenn die Koeffizientenmatrix keine Nullzeile enthält) entweder zwei verschiedene parallele Ebenen und damit einen leeren Durchschnitt (Inkonsistenzfall, keine Lösung), oder zwei gleiche Ebenen und damit auch eine Ebene als Durchschnitt (Konsistenzfall, 2-dimensionale lineare Lösungsmenge).

Die obige Rechnung war nicht die einfachste, die zur Lösung des gegebenen Gleichungssystems führt. Da die zweite Gleichung einen Nullkoeffizienten hat, führt hier nämlich eine Spaltenvertauschung direkt zu einer Zeilen-Stufen-Form:

$$\begin{array}{ccc|c} x & y & z & \\ \hline 3 & -2 & 1 & 2 \\ 1 & 0 & -2 & 3 \end{array} \iff \begin{array}{ccc|c} y & x & z & \\ \hline -2 & 3 & 1 & 2 \\ 0 & 1 & -2 & 3 \end{array}$$

Auch hier ist z die Nicht-Basisvariable, und wir erhalten dieselbe Parameterdarstellung der Lösung wie zuvor. Aber wir können auch eine andere Variable zur Nicht-Basisvariablen und damit zum Parameter einer Parameterdarstellung machen:

$$\begin{array}{ccc|c} x & y & z & \\ \hline 3 & -2 & 1 & 2 \\ 1 & 0 & -2 & 3 \end{array} \iff \begin{array}{ccc|c} z & y & x & \\ \hline 1 & -2 & 3 & 2 \\ -2 & 0 & 1 & 3 \end{array} \iff \begin{array}{ccc|c} z & y & x & \\ \hline 1 & -2 & 3 & 2 \\ 0 & -4 & 7 & 7 \end{array} \quad \textcircled{2} \rightarrow \textcircled{2} + 2 \cdot \textcircled{1}$$

Hier ist jetzt x Nicht-Basisvariable, und wir erhalten für L die Parameterdarstellung $x = r$, $y = \frac{7}{4}(r-1)$, $z = \frac{1}{2}r - \frac{3}{2}$. Auch y können wir zur Nicht-Basisvariablen machen, was die Parameterdarstellung $x = \frac{4}{7}s + 1$, $y = s$, $z = \frac{2}{7}s - 1$ für L ergibt. Natürlich kann man die Parameterdarstellungen auch direkt ineinander umrechnen, indem man $r := 2t + 3$ oder $s := \frac{7}{2}(t-1)$ setzt. Grundsätzlich lassen sich bei zwei unabhängigen Gleichungen alle Unbekannten zur Nicht-Basisvariablen machen, für die nach Streichung der zugehörigen Koeffizientenspalte immer noch ein System von zwei unabhängigen linearen Gleichungen für eine Unbekannte weniger verbleibt.

5)
$$a_1x_1 + \dots + a_nx_n = b$$

ist eine lineare Gleichung ($m = 1$) für n Unbekannte. Dieses System hat schon Zeilen-Stufen-Form, und zwar mit einer Stufe, wenn ein Koeffizient $a_j \neq 0$ ist (und mit Null Stufen sonst). Man kann mit Variablenvertauschung dann diesen Koeffizienten an die vorderste Position bringen, was die anderen Unbekannten zu Nicht-Basisvariablen macht. Wählt man diese als beliebige reelle Parameter, etwa $x_1 = t_1, \dots, x_{j-1} = t_{j-1}, x_{j+1} = t_j, \dots, x_n = t_{n-1}$, so ergibt sich $x_j = \frac{1}{a_j}(b - a_1t_1 - \dots - a_{j-1}t_{j-1} - a_{j+1}t_j - \dots - a_nt_{n-1})$ und damit eine $(n-1)$ -parametrische Darstellung der Lösungsmenge. Je nach Wahl des Koeffizienten $a_j \neq 0$ bekommt man so unterschiedliche Parameterdarstellungen der $(n-1)$ -dimensionalen Lösungsmenge (und auf den ersten Blick ist nicht zu erkennen, dass solche verschiedenen Parameterdarstellungen tatsächlich dieselbe Menge beschreiben). Sind alle Koeffizienten Null und auch die rechte Seite, so ist die Lösungsmenge natürlich der ganze Raum \mathbb{R}^n , d.h. man kann alle n Unbekannten als beliebige reelle Parameter wählen und erhält stets eine Lösung. Sind alle Koeffizienten Null, aber $b \neq 0$, so liegt Inkonsistenz vor.

6) Wir betrachten ein allgemeines System von 2 linearen Gleichungen für 2 Unbekannte:

$$\begin{array}{l} ax + by = b_1 \\ cx + dy = b_2 \end{array}$$

Die möglichen Zeilen-Stufen-Formen sind

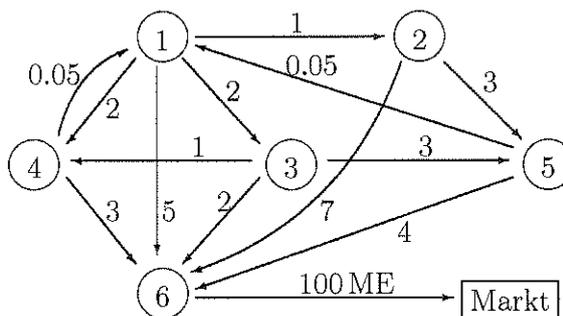
$$\left(\begin{array}{cc|c} c_1 & * & * \\ 0 & c_2 & * \end{array} \right) \quad \left(\begin{array}{cc|c} c_1 & * & * \\ 0 & 0 & * \end{array} \right) \quad \left(\begin{array}{cc|c} 0 & c_1 & * \\ 0 & 0 & * \end{array} \right) \quad \left(\begin{array}{cc|c} 0 & 0 & * \\ 0 & 0 & * \end{array} \right)$$

mit $c_1 \neq 0$, $c_2 \neq 0$ und irgendwelchen Einträgen " $*$ ". Die erste oder zweite Form ergibt sich, wenn $a \neq 0$ ist, dann $c_1 = a$ und $c_2 = \frac{1}{a}(ad - bc)$, oder wenn $c \neq 0$ ist, dann $c_1 = c$ und $c_2 = -\frac{1}{c}(ad - bc)$; je nachdem ob $ad - bc \neq 0$ ist oder nicht, hat man die erste oder zweite Form. Im Fall $a = 0 = c$ und $b \neq 0$ oder $d \neq 0$ ergibt sich die dritte Form, bei $a = b = c = d = 0$ die vierte.

Hier wird erneut die Bedeutung der **Determinante** $ad - bc$ der Koeffizientenmatrix deutlich, die wir schon früher diskutiert haben: Genau wenn die Determinante von Null verschieden ist, ist das Problem für jede rechte Seite wohlgestellt, also eindeutig lösbar.

Andernfalls existiert entweder keine Lösung, nämlich wenn die Koeffizienten in einer Zeile der Zeilen-Stufen-Form Null sind, der zugehörige letzte Eintrag * aber nicht, oder es existiert eine unendliche Lösungsmenge, die sich als 1-dimensionale lineare Menge parametrisieren lässt im Fall der Zeilen-Stufen-Formen zwei oder drei, oder die alle $(x, y) \in \mathbb{R}^2$ enthält im Fall der letzten Form mit Nulleinträgen auch in der letzten Spalte.

7) Als Beispiel für das Auftreten linearer Gleichungssysteme in ökonomischem Problemstellungen betrachten wir folgenden gerichteten und bewerteten Graphen, der eine (z.B. chemische) **Produktion mit Verflechtung** beschreibt:



Wir haben schon erklärt, welche Information in einem solchen "Gozintographen" codiert ist: Ein Pfeil $(i) \xrightarrow{r} (j)$ bedeutet, dass für die Produktion einer Mengeneinheit (ME) des Produkts (j) genau r Mengeneinheiten des Vorprodukts (i) benötigt werden. Vom Endprodukt (6) sind 100 ME an den Markt zu liefern.

Gesucht sind hier natürlich die Outputs x_1, \dots, x_6 (in ME) der Produkte $(1), \dots, (6)$, die hergestellt werden müssen, damit die Lieferung erfolgen kann. Wenn wir annehmen, dass ohne Überschuss produziert, von jedem Produkt also nur die benötigte Menge hergestellt werden soll, so können wir ein **lineares Gleichungssystem für die Produktion** wie folgt aufstellen: Für jeden Knoten (i) bestimmt man alle von dort ausgehenden Pfeile $(i) \xrightarrow{r_{ij}} (j)$ mit ihren Bewertungen r_{ij} . Für die Produktion von x_j Einheiten des Produkts Nummer (j) werden $r_{ij}x_j$ Einheiten des Vorprodukts (i) benötigt. Die Summe dieser Werte über alle von (i) ausgehenden Pfeile ist also die von (i) insgesamt benötigte Menge, und diese Summe setzt man $= x_i$, weil ja von dem Produkt (i) nicht mehr als nötig hergestellt werden soll. Für den obigen Graphen gibt das folgendes Gleichungssystem:

$$\begin{aligned} 2x_4 + 5x_6 + 2x_3 + x_2 &= x_1 \\ 7x_6 + 3x_5 &= x_2 \\ x_4 + 2x_6 + 3x_5 &= x_3 \\ 0.05x_1 + 3x_6 &= x_4 \\ 0.05x_1 + 4x_6 &= x_5 \\ 100 &= x_6 \end{aligned}$$

Die letzte Gleichung beschreibt, dass 100 ME des Endprodukts an den Markt zu liefern sind. Andere Unbekannte als x_6 treten in dieser letzten Gleichung nicht auf, weil hier im Produktionsablauf kein Anteil des Endprodukts für die Produktion eines Vorprodukts benötigt wird (was im Prinzip, bei biotechnologischen oder chemischen Produktionsabläufen, durchaus auch vorkommen kann). Wenn wir das System nach Unbekannten sortieren, so ergibt sich folgendes Tableau:

$$\begin{array}{cccccc|c}
 1 & -1 & -2 & -2 & 0 & -5 & 0 \\
 0 & 1 & 0 & 0 & -3 & -7 & 0 \\
 0 & 0 & 1 & -1 & -3 & -2 & 0 \\
 -0.05 & 0 & 0 & 1 & 0 & -3 & 0 \\
 -0.05 & 0 & 0 & 0 & 1 & -4 & 0 \\
 0 & 0 & 0 & 0 & 0 & 1 & 100
 \end{array}$$

Die Einträge "1" auf der Diagonalen hat man offenbar immer, wenn der Verflechtungsgraph keine Schleifen hat (keine Pfeile von \textcircled{i} nach \textcircled{i}). Außerdem hat das entstehende Gleichungssystem immer gleich viele Gleichungen wie Unbekannte, eben für jedes beteiligte Vor-, Zwischen- oder Endprodukt eine Gleichung. Man kann also hoffen, dass es eine eindeutige Lösung gibt. Im vorliegenden Fall hat die Koeffizientenmatrix schon nahezu Dreiecksgestalt. Die Herstellung der Zeilen-Stufen-Form ist daher mit wenigen Zeilenoperationen möglich: Man addiert die zweite und das Doppelte der dritten Zeile zur ersten, subtrahiert die fünfte von der vierten, addiert dann das Vierfache der vierten zur ersten Zeile und schließlich das 0.05-fache der ersten zur fünften. Resultat ist das Dreiecks-Tableau

$$\begin{array}{cccccc|c}
 1 & 0 & 0 & 0 & -13 & -12 & 0 \\
 0 & 1 & 0 & 0 & -3 & -7 & 0 \\
 0 & 0 & 1 & -1 & -3 & -2 & 0 \\
 0 & 0 & 0 & 1 & -1 & 1 & 0 \\
 0 & 0 & 0 & 0 & 0.35 & -4.6 & 0 \\
 0 & 0 & 0 & 0 & 0 & 1 & 100
 \end{array}$$

mit der eindeutigen Lösung $x_6 = 100$, $x_5 = (4.6/.35)x_6 \approx 1314.29$, $x_4 = x_5 - x_6 \approx 1214.29$, $x_3 = x_4 + 3x_5 + 2x_6 \approx 5357.14$, $x_2 = 3x_5 + 7x_6 \approx 4642.87$ und $x_1 = 13x_5 + 12x_6 \approx 18285.71$ (jeweils Mengeneinheiten).

Wir schließen noch einige Beobachtungen an: Dass die Koeffizientenmatrix in diesem Beispiel schon nahezu Dreiecksgestalt hat, liegt natürlich daran, dass hier nur wenige Zwischenprodukte in die Herstellung der eigenen Vorprodukte eingehen. Bei einfachen Produktionsabläufen kommt so etwas überhaupt nicht vor, und dann hat man schon direkt Dreiecksgestalt der Koeffizientenmatrix. Es gibt aber in der Realität noch sehr viel kompliziertere Produktionsverflechtungen als die hier behandelte. Auch können durchaus Schleifen vorkommen, d.h. ein Teil eines End- oder Zwischenprodukts wird zur Herstellung des Produkts selbst benötigt. (Für den Produktionsbeginn muss man sich diesen Teil dann natürlich zunächst "leihen" und kann ihn am Ende wieder zurückgeben.)

Obwohl sich stets ein lineares Gleichungssystem mit ebenso vielen Unbekannten wie Gleichungen ergibt, ist dessen eindeutige Lösbarkeit für beliebig ausgedachte Produktionspläne keineswegs garantiert. Aus der allgemeinen Theorie wissen wir, dass eindeutige Lösbarkeit genau dann vorliegt, wenn das homogene System nur die triviale Lösung hat. Andernfalls gäbe es einen Produktionsablauf, bei dem alle produzierten Mengen für die Produktion selbst wieder verbraucht werden. Eine solche Situation liegt z.B. vor, wenn man zwei Produkte hat und den Produktionsplan $\textcircled{1} \xrightarrow{1} \textcircled{2}$ sowie $\textcircled{2} \xrightarrow{1} \textcircled{1}$, d.h. für die Produktion einer ME des einen wird jeweils eine ME des anderen Produkts benötigt. Derartige ist natürlich ökonomisch unsinnig, und das entsprechende inhomogene Gleichungssystem ist niemals lösbar, d.h. Produktion unter Abgabe einer nichtleeren Produktmenge an den Markt ist nicht möglich.

Bedenken muss man auch, dass *nur Lösungen mit positiven Werten der Outputs x_i ökonomisch sinnvoll sind*. Selbst wenn eine eindeutige Lösung existiert, so braucht diese Positivitätsbedingung nicht erfüllt zu sein. Eine solche Situation liegt z.B. vor, wenn für die Produktion einer ME des Vorprodukts (i) von einem späteren Produkt (j) mehr Einheiten gebraucht werden, als mit einer ME von (i) überhaupt produziert werden kann, also z.B. bei einer Schleife mit Bewertung > 1 oder bei einem Produktionsplan $(1) \xrightarrow{1} (2)$ sowie $(2) \xrightarrow{2} (1)$. Das entsprechende Gleichungssystem $x_1 - x_2 = b_1$, $x_2 - 2x_1 = b_2$ hat dann zwar für alle rechten Seiten b_1, b_2 eine eindeutige Lösung, aber für $b_1 \geq 0$ und $b_2 \geq 0$ niemals eine Lösung mit positiver Komponente x_1 oder x_2 (denn es folgt ja $x_1 = x_2 + b_1 \geq x_2 = 2x_1 + b_2 \geq 2x_1$). Ökonomisch ist ein derartiger Produktionsplan natürlich von vorneherein unsinnig, aber bei einem komplizierten Produktionsverflechtungsgraphen ist nicht ohne weiteres zu erkennen, ob die beschriebene Produktionssituation sinnvoll ist und tatsächlich zur Produktion von Mengen führt, die an den Markt abgegeben werden können und nicht intern im Produktionsablauf verbraucht werden.

Es ist nicht nur für die hier betrachtete, sondern auch für diverse andere ökonomische Problemstellungen eine wichtige und mathematisch gesehen auch interessante Frage, wie eine quadratische Koeffizientenmatrix beschaffen sein muss, damit das zugehörige System von n linearen Gleichungen für n Unbekannte bei jeder Wahl von nichtnegativen rechten Seiten (Mengen, die an den Markt gehen sollen), die nicht alle Null sind, eine eindeutige Lösung mit lauter positiven Komponenten hat. Man nennt derartige quadratische Koeffizientenmatrizen *invers positive Matrizen*. Matrizen mit positive Einträgen auf der Diagonalen und nichtpositiven Einträgen von genügend kleinem Betrag sonst haben z.B. diese Eigenschaft. Die weitere Diskussion würde hier aber zu weit führen (s. 3.4). Für ökonomische Anwendungen wie oben kann man einfach das Gleichungssystem aufstellen, durch Herstellung der Zeilenstufen-Form feststellen, ob es eine eindeutige Lösung gibt, und durch Berechnung der Lösung dann überprüfen, ob diese Lösung auch positive Komponenten hat. Wenn ja, so hat man eine eindeutige ökonomisch sinnvolle Lösung; wenn nicht, so weiß man, dass die gestellte Aufgabe keine ökonomisch sinnvolle Lösung hat.

8) Noch ein ökonomisches Beispiel, die **Innerbetriebliche Leistungsverrechnung**: Hier geht es darum, innerbetriebliche Vorleistungen an Teilbetriebe angemessen mit fiktiven Preisen zu bewerten, damit die Selbstkosten und die Rentabilität der Teilbetriebe ermittelt werden können. Man unterscheidet *primäre Kosten*, die den Betriebsteilen / Kostenstellen direkt entstehen (Löhne, Materialkosten, Abschreibungen, ...), und *sekundäre Kosten*, das sind die fiktiven Kosten von unentgeltlich gewährten innerbetrieblichen Leistungen (Vorprodukte, Energie, Wartungs- und Reparaturarbeiten, ...). Zur Erfassung der tatsächlichen Kosten eines Teilbetriebs sind dessen primären Kosten die mit angemessenen *Verrechnungspreisen* bewerteten Leistungen hinzuzurechnen, die ihm von anderen Betriebsteilen gewährt werden; die so ermittelten Kosten sind als *Verrechnungskosten* zu bilanzieren. Ebenso müssen natürlich die von dem Teilbetrieb an andere Betriebsteile bzw. an den Gesamtbetrieb abgegebenen Leistungen mit angemessenen Verrechnungspreisen bewertet werden, wenn die Rentabilität des Teilbetriebs beurteilt werden soll.

Das *Problem* ist, dass die angemessenen Verrechnungspreise für die Leistungen eines Teilbetriebs nicht nur von dessen primären Kosten abhängen, sondern auch von dessen sekundären Kosten, also von den Verrechnungspreisen für die von anderen Betriebsteilen empfangenen Leistungen. Man kann daher die "richtigen" Verrechnungspreise nicht separat für jede einzelne Kostenstelle im Betrieb bestimmen, sondern nur simultan für alle Betriebsteile gemeinsam. Dies wird auf ein lineares Gleichungssystem führen.

Der Einfachheit halber nehmen wir an, dass es n Teilbetriebe $\textcircled{1}, \dots, \textcircled{n}$ gibt, die an den Betrieb jeweils Leistungen einer einzigen Art abgeben und mit Verrechnungspreisen p_1, \dots, p_n (Geldeinheiten pro Leistungseinheit) bewertet werden sollen. Die *interne Leistungsbilanz* kann dann z.B. durch einen bewerteten gerichteten Graphen wie in 7) dargestellt werden, wobei ein Pfeil $\textcircled{j} \xrightarrow{a_{ij}} \textcircled{i}$ bedeutet, dass a_{ij} Leistungseinheiten von \textcircled{j} an \textcircled{i} abgegeben werden. Eine andere Möglichkeit ist die Darstellung durch eine *Verflechtungsmatrix* (a_{ij}) , wobei ein Nulleintrag a_{ij} bedeutet, dass von \textcircled{j} keine Leistung an \textcircled{i} abgegeben wird, und ein Diagonaleintrag $a_{ii} \neq 0$ die von \textcircled{i} selbst verbrauchte Eigenleistung angibt (das entspricht einer Schleife im Verflechtungsgraphen). Wir haben das schon früher diskutiert. Weitere relevante Daten sind hier noch die bei der Stelle \textcircled{i} anfallenden primären Kosten k_i (Geldeinheiten) und die von \textcircled{i} an den Betrieb insgesamt gelieferte Leistung l_i (Leistungseinheiten). Der Geldwert dieser erbrachten Leistung ist dann mit $l_i p_i$ anzusetzen, der Wert der von \textcircled{j} an \textcircled{i} abgegebenen Leistung mit $a_{ij} p_j$, die gesamten Kosten des Teilbetriebs \textcircled{i} also mit $k_i + a_{i1} p_1 + a_{i2} p_2 + \dots + a_{in} p_n$. Wenn man davon ausgeht, dass Teilbetriebe nicht Gewinne auf Kosten anderer Teilbetriebe machen sollen, so wird man postulieren, dass für jeden Teilbetrieb \textcircled{i} gilt:

$$\left. \begin{array}{l} \text{Wert der von } \textcircled{i} \text{ an den} \\ \text{Betrieb erbrachten Leistung} \end{array} \right\} = \left\{ \begin{array}{l} \text{Summe der bei } \textcircled{i} \text{ anfallenden} \\ \text{primären und sekundären Kosten,} \end{array} \right.$$

d.h. in Formeln

$$l_i p_i = k_i + a_{i1} p_1 + a_{i2} p_2 + \dots + a_{in} p_n \quad \text{für } i = 1 \dots n.$$

Dies ist das angekündigte Gleichungssystem. Nach Sortieren der Variablen erhalten wir das folgende **lineare Gleichungssystem für die Verrechnungspreise** p_1, \dots, p_n :

$$\begin{array}{rcccc} (l_1 - a_{11})p_1 & - a_{12} p_2 - \dots & - a_{1n} p_n & = k_1 \\ -a_{21} p_1 + (l_2 - a_{22})p_2 & - \dots & - a_{2n} p_n & = k_2 \\ \vdots & & \vdots & \\ -a_{n1} p_1 & - a_{n2} p_2 - \dots + (l_n - a_{nn})p_n & & = k_n \end{array}$$

(Zur Erinnerung: a_{ij} ist die von Betriebsteil \textcircled{j} an \textcircled{i} abgegebene Leistung, k_i sind die primären Kosten bei \textcircled{i} , l_i ist die von \textcircled{i} an den Gesamtbetrieb abgegebene Leistung.)

Die Koeffizienten- $n \times n$ -Matrix dieses Gleichungssystems hat positive Diagonaleinträge ($l_i > a_{ii}$, sonst würde der Teilbetrieb \textcircled{i} ja mehr Leistung verbrauchen als er erbringt) und hat ansonsten nichtpositive Einträge. Diese Struktur ist typisch für *invers positive Matrizen*, wie wir in 7) schon erwähnt haben, d.h. man kann erwarten, dass bei positiven rechten Seiten $k_i > 0$ wie hier genau eine Lösung (p_1, \dots, p_n) existiert, die zudem lauter positive Komponenten $p_j > 0$ hat, also ökonomisch sinnvoll ist. (Es gibt aber auch interne Leistungsbilanzen, die auf ein Gleichungssystem für die Verrechnungspreise führen, das keine eindeutige Lösung mit positiven Komponenten besitzt. Das ist dann ein Hinweis darauf, dass bei einem solchen Ablauf etwas ökonomisch Unsinniges passiert, z.B. dass ein Teilbetrieb bei jeder Wahl der Verrechnungspreise mehr Kosten verursacht, als seine abgegebene Leistung wert ist.) Hat man die Verrechnungspreise durch Lösung des Gleichungssystems bestimmt und sind sie ökonomisch sinnvoll (positiv, eindeutig bestimmt), so kann man anschließend auch noch die *Kosten der Betriebsteile, die keine Leistungen an den Betrieb abgeben*, sondern nur Leistungen von Betriebsteilen empfangen, als deren primäre plus sekundäre Kosten berechnen.

Ein konkretes Beispiel (nach Tietze, §9.2): In einem Unternehmen gibt es vier Hilfskostenstellen ①, ..., ④, die an sich gegenseitig und an drei Hauptkostenstellen ①, ②, ③ Leistungen erbringen. Die Hauptkostenstellen liefern ihre Leistungen an den Markt und nicht an irgendwelche Betriebsteile. Die Leistungsbilanz sieht wie folgt aus:

	①	②	③	④	← Lieferant
①	10	40	100	80	
②	30	10	80	20	
③	40	50	0	20	
④	50	100	40	30	
①	80	100	180	250	
②	90	150	150	200	
③	100	150	30	200	

↑ Empfänger

	Primärkosten	Gesamtleistung
①	2020	400
②	3700	600
③	1960	580
④	7700	800
①	15200	—
②	21000	—
③	45000	—

Diese Tabellen geben die relevanten Daten zur Leistungsbilanz an; die Einträge in der letzten Spalte der zweiten Tabelle sind die Spaltensummen der Einträge in der ersten Tabelle. Das lineare Gleichungssystem für die unbekanntenen Verrechnungspreise, welche die von ①, ..., ④ abgegebenen Leistungen bewerten sollen, hat hier das folgende Tableau:

390	-40	-100	-80	2020
-30	590	-80	-20	3700
-40	-50	580	-20	1960
-50	-100	-40	770	7700

Zur Herstellung der Zeilen-Stufen-Form dividiere alle Zeilen durch 10, vertausche die erste mit der vierten und die zweite mit der dritten Zeile, addiere dann geeignete Vielfache der (neuen) ersten Zeile zur zweiten, dritten und vierten Zeile, um Nullen in den Positionen 2, 3 und 4 der ersten Spalte zu erzeugen, sodann addiere Vielfache der zweiten zur dritten und vierten Zeile, um auch in der zweiten Spalte Nullen in den Positionen 3 und 4 zu erzeugen, schließlich erzeuge noch eine Null in Position 4 der dritten Spalte und addiere die vorherige vierte Zeile zur dritten, um diese noch zu vereinfachen. Das Resultat ist — wenn wir korrekt gerechnet haben — folgendes Tableau in Dreiecksgestalt:

-5	-10	-4	77	770
0	3	61.2	-63.6	-420
0	0	300	184	3736
0	0	0	483.59...	5765.43...

mit der von unten nach oben eindeutig zu berechnenden Lösung $p_4 \approx 11.92$, $p_3 \approx 5.14$, $p_2 \approx 7.87$, $p_1 \approx 9.75$ (jeweils Geldeinheiten pro Leistungseinheit). Damit sind die Verrechnungspreise bestimmt, und das Ergebnis ist auch ökonomisch sinnvoll, da alle p_j positiv sind. Mit diesen Verrechnungspreisen sind nun die Kosten aller Betriebsteile berechenbar. Für die Hilfskostenstelle ① zum Beispiel sind die Kosten 2020 (Primärkosten) $+10p_1+40p_2+100p_3+80p_4$ (sekundäre Kosten) ≈ 3899.90 Geldeinheiten; für die Hauptkostenstelle ③ sind es 45 000 (Primärkosten) $+100p_1+150p_2+30p_3+200p_4$ (Sekundärkosten) $\approx 49 693.70$ Geldeinheiten insgesamt. ■

Mathematik für Wirtschaftswissenschaftler

(K. Steffen, Heinrich-Heine-Universität Düsseldorf, WS 2006/07)

Die obigen Beispiele für das Auftreten linearer Gleichungssysteme bei ökonomischen Fragestellungen waren noch sehr einfach gehalten. In der Praxis kann man es mit sehr viel mehr Unbekannten und sehr viel mehr linearen Gleichungen zu tun haben. Was die Theorie angeht, ist das kein prinzipieller Unterschied: Die Zeilen–Stufen–Matrix ist dann eben auch sehr groß. Für die konkrete Berechnung der Lösungen ist aber die Herstellung der Zeilen–Stufen–Form, also die Ausführung des Eliminationsverfahrens, ab einer gewissen Größe von Hand kaum noch durchzuführen. Natürlich gibt es für solche Fälle Rechenprogramme zum Lösen linearer Gleichungssysteme. Bei sehr großen linearen Gleichungssystemen (mit mehr als 100 Unbekannten und Gleichungen etwa) ist aber auch für Computer die exakte Ausführung des Eliminationsverfahrens zu aufwendig. Man benutzt dann numerische Verfahren, welche die Lösung näherungsweise mit akzeptabler Genauigkeit berechnen. Solche Verfahren können allerdings nur befriedigend funktionieren, wenn ein wohlgestelltes Problem vorliegt, also eine eindeutige Lösung existiert. (Wenn es keine Lösung gibt, kann auch ein numerisches Verfahren keine berechnen; gibt es aber viele Lösungen, so wird das Verfahren das Problem haben, welche davon es berechnen soll.)

Wir wollen nun noch eine “Verbesserung” der Zeilen–Stufen–Form besprechen, die nützlich ist wenn man *Gleichungssysteme mit derselben Koeffizientenmatrix für mehrere rechte Spaltenvektoren* zu lösen hat. Man denke z.B. an einen Produktionsplan wie im vorangegangenen Beispiel 7), der für unterschiedliche Produktlieferungen an den Markt durchgerechnet werden soll. In solchen Fällen ist es zweckmäßig, die Koeffizientenmatrix gleich um alle Spalten von interessierenden rechten Seiten zu erweitern und diese Spalten bei der Herstellung der Zeilen–Stufen–Form mit den Zeilentransformationen mit umzuformen. Man betrachtet also dann erweiterte Koeffizientenmatrizen der Form

$$\left(\begin{array}{cccc|cccc} a_{11} & a_{12} & \dots & a_{1n} & b_1 & c_1 & \dots & z_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 & c_2 & \dots & z_2 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m & c_m & \dots & z_m \end{array} \right)$$

und kann nach Erreichen der Zeilen–Stufen–Form für die Koeffizientenmatrix (a_{ij}) dann die Lösungen für jede einbezogene Spalte von rechten Seiten von unten nach oben berechnen, ohne für jede einzelne Spalte das ganze Eliminationsverfahren von vorne beginnen zu müssen.

Hat man allerdings für sehr viele rechte Spalten zu lösen (für mehr als $\min(m, n)$ Stück), so gibt es eine bessere Methode. Mit etwas mehr Rechenaufwand kann man nämlich die Zeilen–Stufen–Form noch so vereinfachen, dass sich die Auflösung des Gleichungssystems von unten nach oben ohne jede Rechnung ergibt. Man muss dazu nur bemerken, dass man durch weitere Zeilenoperationen auch über den “Stufenanfängen” Nulleinträge erzeugen kann. Danach steht in den Spalten, die den Basisvariablen entsprechen, jeweils nur noch ein Eintrag $\neq 0$, und indem man die entsprechende Zeile durch diesen Eintrag dividiert, erreicht man, dass sein Wert 1 ist. Vertauscht man schließlich noch die Spalten so, dass die Spalten der Basisvariablen zuerst kommen, so hat man folgendes Ergebnis:

SATZ (kanonische Zeilen-Stufen-Form):

(i) Durch zulässige Zeilentransformationen kann die erweiterte Koeffizientenmatrix $(A|b)$ eines Systems von m linearen Gleichungen für n Unbekannte stets auf eine Zeilen-Stufen-Form gebracht werden, in der die zu den Basisvariablen gehörenden Spalten kanonische Einheitsvektoren sind. Durch zusätzliche Spaltenvertauschungen in der Koeffizientenmatrix, welche die Basisvariablen-Spalten nach vorne bringen, kann diese Matrix also in folgende sog. **kanonische Zeilen-Stufen-Form** gebracht werden:

$$\begin{array}{l}
 l \\
 \left\{ \begin{array}{c} \left(\begin{array}{ccc|ccc} 1 & & & & & d_1 \\ & 1 & 0 & & & d_2 \\ & & \ddots & & & \vdots \\ 0 & & & & * & d_l \\ \hline & & & 0 & & d_{l+1} \\ & 0 & & & & \vdots \\ & & & & & d_m \end{array} \right) \end{array} \right. \\
 m-l \\
 \underbrace{\hspace{10em}}_l \quad \underbrace{\hspace{10em}}_{n-l}
 \end{array}$$

($0 \leq l \leq \min(m, n)$ die maximale Zahl der unabhängigen Gleichungen, $d_1, \dots, d_m \in \mathbb{R}$, beliebige reelle Einträge im Rechteck “*” und Nulleinträge in den mit “0” gekennzeichneten Dreiecken und Rechtecken).

(ii) Wenn eine der rechten Seiten d_{l+1}, \dots, d_m von Null verschieden ist, so ist das Gleichungssystem inkonsistent. Andernfalls erhält man die Lösungen, indem man die Nicht-basisvariablen, die nach den in (i) vorgenommenen Spaltenvertauschungen den Spalten mit Nummern $l+1, \dots, n$ entsprechen, als beliebige reelle Parameter wählt, und die Werte der Basisvariablen, die dann den Spalten mit Nummern $1, \dots, l$ zugeordnet sind, aus den ersten l Gleichungen direkt entnimmt (ohne das System in Zeilen-Stufen-Form von unten nach oben auflösen zu müssen).

(iii) Bei vorgegebener Zuordnung der Unbekannten zu den Koeffizientenspalten sind Form und Einträge der Koeffizientenmatrix in der kanonischen Zeilen-Stufen-Form eindeutig bestimmt durch die Lösungsmenge des homogenen Gleichungssystems $Ax = 0$. Im Konsistenzfall sind auch die Einträge d_i der rechten Seite eindeutig bestimmt durch die Lösungsmenge des inhomogenen Systems $Ax = b$. ■

Dabei heißt ein Spaltenvektor (oder ein m -tupel) in \mathbb{R}^m ein **kanonischer Einheitsvektor** oder ein **kanonischer Basisvektor**, wenn er genau einen Eintrag 1 hat und sonst lauter Einträge 0. Befindet sich die eingetragene 1 an der i -ten Position, so schreiben wir dafür (oder für den entsprechenden m -gliedrigen Spaltenvektor)

$$e_i := \underbrace{(0, \dots, 0)}_{i-1}, 1, \underbrace{(0, \dots, 0)}_{m-i-1} \in \mathbb{R}^m,$$

und man nennt e_1, e_2, \dots, e_m die **kanonische Basis** des m -dimensionalen Zahlenraums \mathbb{R}^m . (Wenn nötig, so muss man die Gliederzahl der kanonischen Basisvektoren in der Notation kenntlich machen, indem man z.B. genauer $e_i^{(m)}$ für die kanonischen Basisvektoren in \mathbb{R}^m schreibt.) In der Literatur zur Mathematik für die Wirtschaftswissenschaften findet sich auch die Bezeichnung “Einheitsvektor” für die kanonischen Einheitsvektoren. Das ist aber nicht korrekt, weil man unter einem Einheitsvektor allgemein einen Vektor der Länge 1 versteht, und davon gibt es viel mehr als nur die ganz speziellen kanonischen Basisvektoren. Zum Beispiel sind auch $(-1, 0, \dots, 0)$, $(\pm \frac{1}{\sqrt{2}}, \pm \frac{1}{\sqrt{2}}, 0, \dots, 0)$, $(\pm \frac{1}{2}, \pm \frac{1}{2}, \pm \frac{1}{2}, \pm \frac{1}{2}, 0, \dots, 0), \dots$ Einheitsvektoren (Länge 1), aber keine kanonischen Einheitsvektoren.

Der *Beweis* der beiden ersten Aussagen ist aufgrund der Vorbemerkungen eigentlich schon klar. Wegen seiner praktischen Bedeutung beschreiben wir aber nochmals konkret das Verfahren, das zur kanonischen Zeilen-Stufen-Form führt:

Man sucht sich zunächst irgendeinen Eintrag a_{ij} der Koeffizienten-Matrix A aus, der von Null verschieden ist. (Wenn alle Koeffizienten Null sind, so ist nichts mehr zu beweisen; links steht schon die kanonische Zeilen-Stufen-Form mit $l = 0$, und je nachdem, ob $b \in \mathbb{R}^m$ eine Nullspalte ist oder nicht, sind alle $x \in \mathbb{R}^n$ Lösungen oder es gibt keine.) Dies ist das sog. erste *Pivot-Element*. In der Praxis wählt man es natürlich so, dass die Rechnungen möglichst einfach werden; ein Pivot-Element mit Wert ± 1 ist z.B. für Rechnungen von Hand oft günstig. Bei numerischen Verfahren zur Lösung großer linearer Gleichungssysteme wählt man oft ein Pivot-Element von möglichst großem Betrag, weil durch das Pivot-Element im weiteren Verlauf des Verfahrens zu dividieren ist und weil die Division mit Zahlen von kleinem Betrag große Rundungsfehler mit sich bringt. Nun dividiert man die i -te Zeile der erweiterten Koeffizientenmatrix durch das Pivot-Element a_{ij} und subtrahiert danach für $h = 1 \dots m$ mit $h \neq i$ das a_{hj} -fache der (dividierten) i -ten Zeile von der h -ten Zeile. Resultat ist eine Matrix, die als j -te Spalte den kanonischen Einheitsspaltenvektor e_i hat. Durch eine Zeilenvertauschung bringt man die i -te Zeile nach oben, durch eine Spaltenvertauschung die j -te Spalte nach vorne. Danach hat man als erste Spalte den kanonischen Basisvektor e_1 , also eine erweiterte Koeffizientenmatrix der Form

$$\left(\begin{array}{c|c|c} 1 & & * \\ 0 & & * \\ \vdots & B & \vdots \\ 0 & & * \end{array} \right)_{n-1}$$

mit einer $m \times (n-1)$ -Matrix B (und beliebigen reellen Einträgen "*"). Mit dieser durch die letzte Spalte erweiterten Matrix ($B|*$) führt man nun dasselbe Verfahren durch, wobei man aber das Pivot-Element nicht in der ersten Zeile wählen darf, weil sonst die Nullen, die man in der ersten Spalte schon erzeugt hat, wieder zerstört werden würden. (Wenn B von Null verschiedene Einträge nur in der ersten Zeile hat, so ist man schon fertig.) Resultat ist dann eine erweiterte Koeffizientenmatrix, deren beide erste Spalten kanonische Einheitsvektoren sind:

$$\left(\begin{array}{cc|c|c} 1 & 0 & & * \\ 0 & 1 & & * \\ 0 & 0 & C & * \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & * \end{array} \right)_{n-2}$$

Nun arbeitet man mit ($C|*$) weiter, wobei das Pivot-Element nicht in den beiden oberen Zeilen gewählt werden darf. Nach endlich vielen Schritten gelangt man so offenbar zu der in Teil (i) des Satzes angegebenen kanonischen Zeilen-Stufen-Form. Wichtig ist bei dem Verfahren: *Pivot-Zeilen, die zur Herstellung von Nullen in einer Spalte verwendet wurden, sind danach tabu!* Sonst würde man ja bereits erzeugte Nulleinträge wieder zerstören.

Um die Eindeutigkeitsaussage (iii) des Satzes einzusehen, muss man nur bemerken, dass zwei Gleichungen $x_h + c_{hl+1}x_{l+1} + \dots + c_{hn}x_n = d_h$ und $x_h + \tilde{c}_{hl+1}x_{l+1} + \dots + \tilde{c}_{hn}x_n = \tilde{d}_h$ genau dann für jede Wahl von x_{l+1}, \dots, x_n dieselbe Lösung x_h haben, wenn erstens $d_h = \tilde{d}_h$ ist (wähle $x_{l+1} = \dots = x_n := 0$) und zweitens $c_{hj} = \tilde{c}_{hj}$ für $j = l+1 \dots n$ (wähle $x_j := 1$ und $x_k := 0$ für $l+1 \leq k \leq n$ mit $k \neq j$).

DISKUSSION: 1) Das beschriebene Verfahren heißt auch **vollständige Gauß-Elimination**, weil am Ende jede Basisvariable aus allen Gleichungen außer einer vollständig eliminiert ist. Der Vorteil ist, dass man dann, nach Festsetzung der beliebig wählbaren Werte der Nicht-Basisvariablen, die Werte der Basisvariablen unmittelbar erhält, ohne wie bei der gewöhnlichen (nichtkanonischen) Zeilen-Stufen-Form das vereinfachte Gleichungssystem noch von unten nach oben lösen zu müssen, weil die Werte jeder Basisvariablen von denen der Basisvariablen mit größerer Nummer abhängen können. Die Berechnung der Lösungen ist also einfacher, wenn man die kanonische Zeilen-Stufen-Form des Gleichungssystems hergestellt hat. Allerdings erfordert die Erzeugung der kanonischen Zeilen-Stufen-Form mehr Zeilenoperationen, also einen größeren Aufwand, als für die gewöhnliche Zeilen-Stufen-Form nötig ist. Und dieser Mehraufwand überwiegt meistens nicht die dadurch erzielte Arbeitersparnis bei der Lösungsberechnung. Wenn es also nur darum geht, die Lösungen eines konkreten linearen Gleichungssystems zu bestimmen, so ist die "unvollständige" Gauß-Elimination, die zu einer gewöhnlichen Zeilen-Stufen-Form führt, der einfachere Weg mit dem kleineren Rechenaufwand.

2) Die kanonische Zeilen-Stufen-Form hat also weniger praktische, sondern vor allem theoretische Bedeutung. Ein wichtiger Gesichtspunkt, den wir als Teil (iii) des Satzes schon hervorgehoben haben, ist ihre *Eindeutigkeit*. Im Unterschied zur gewöhnlichen Zeilen-Stufen-Form, ergeben zwei Rechnungen, die mit verschiedenen Wahlen von Pivot-Elementen zur kanonischen Zeilen-Stufen-Form führen, nicht nur gleiche Stufenzahl l (die Anzahl der unabhängigen Gleichungen), sondern auch gleiche Matrixeinträge in gleichen Positionen der kanonischen Zeilen-Stufen-Form — vorausgesetzt man hat in beiden Rechnungen am Ende dieselbe Zuordnung der Unbekannten zu den Koeffizientenspalten. (Wenn man die Spalten in der Reihenfolge der Nummerierung der Unbekannten anordnet, also überhaupt keine Spaltenvertauschungen vornimmt, so erhält man also unabhängig von den ausgeführten Zeilenoperationen dieselbe Zeilen-Stufen-Matrix mit kanonischen Basisvektoren in den Spalten der Basisvariablen, vorausgesetzt man hat dieselben Unbekannten als Basisvariablen gewählt.)

Ein anderer Vorteil der kanonischen Zeilen-Stufen-Form ist, dass sie im Fall von unendlich vielen Lösungen eine spezielle Parametrisierung der Lösungsmenge liefert, in der die Nicht-Basisvariablen die Parameter sind. Hat man x_1, \dots, x_l als Basisvariable, so lautet für $h = 1 \dots l$ die h -te Gleichung in kanonischer Form $x_h + c_{hl+1}x_{l+1} + \dots + c_{hn}x_n = d_h$. Wenn man also die Nicht-Basisvariablen als Parameter r_k wählt, so erhält man die *spezielle lineare Parametrisierung der Lösungsmenge* mit der kleinstmöglichen Zahl von $n-l$ Parametern:

$$\begin{aligned} x_h &= d_h - c_{hl+1}r_1 - \dots - c_{hn}r_{n-l} \quad \text{für } h = 1 \dots l, \\ x_k &= r_{k-l} \quad \text{für } k = l+1 \dots n. \end{aligned}$$

Hier sind die d_h und die Koeffizienten c_{hj} aus der kanonischen Zeilen-Stufen-Form der erweiterten Koeffizientenmatrix des gegebenen linearen Gleichungssystems $Ax = b$ zu entnehmen. Wenn nicht die ersten l Unbekannten die Basisvariablen sind, sondern die Unbekannten x_{j_1}, \dots, x_{j_l} mit $1 \leq j_1 < \dots < j_l \leq n$, so ändert sich nur an der Nummerierung etwas: Statt x_h hat man oben x_{j_h} zu schreiben und statt x_k dann x_{j_k} , wobei $1 \leq j_{l+1} < \dots < j_n \leq n$ die Nummern der Nicht-Basisvariablen sind.

3) Der **Austausch einer Basisvariablen gegen eine Nicht-Basisvariable** ist möglich, wenn die Koeffizienten für beide Unbekannten in einer Zeile des Gleichungssystems in kanonischer Zeilen-Stufen-Form von Null verschieden sind (d.h. in der Zeile des Systems, die den Koeffizienten 1 für die Basisvariable hat, ist auch der Koeffizient vor der Nicht-Basisvariablen $\neq 0$). Ist das in der h -ten Zeile der Fall und etwa i die Spaltennummer der Basisvariablen, also $c_{hi} = 1$, und j die Spaltennummer einer Nicht-Basisvariablen mit Koeffizient $c_{hj} \neq 0$, so kann man offenbar durch Zeilenoperationen die j -te Spalte in den kanonischen Basisvektor e_h verwandeln, wobei die anderen kanonischen Basisspalten, außer der mit Nummer i , nicht verändert werden. Vertauscht man dann noch die i -te mit der j -ten Spalte, so hat man wieder die kanonische Zeilen-Stufen-Form, wobei allerdings jetzt die vorher der Spaltennummer j zugeordnete Nicht-Basisvariable zur Basisvariablen-Spalte Nummer i gehört.

4) Im Fall eines Systems von n linear unabhängigen Gleichungen für n Unbekannte hat die kanonische Zeilen-Stufen-Form die Gestalt

$$\left(\begin{array}{ccc|c} 1 & & & d_1 \\ & 1 & 0 & d_2 \\ & & \ddots & \vdots \\ & 0 & & 1 \\ & & & d_n \end{array} \right)$$

mit der sog. **$n \times n$ -Einheitsmatrix** I_n als Koeffizientenmatrix, das ist die quadratische $n \times n$ -Matrix mit Einträgen 1 auf der Diagonalen und Nulleinträgen überall sonst. Diese Form des Gleichungssystems ist dann ohne Spaltenvertauschungen alleine durch Zeilenoperationen herzustellen, d.h. die Unbekannten x_j sind den Spalten der Einheitsmatrix in der Reihenfolge ihrer ursprünglichen Nummerierung zugeordnet. Die eindeutig bestimmte Lösung des Systems ist dann natürlich einfach die Spalte der rechten Seiten, die sich ergeben hat, also $x_1 = d_1, x_2 = d_2, \dots, x_n = d_n$. Ähnlich ist es bei $m > n$ Gleichungen für n Unbekannte, wenn davon n linear unabhängig sind und der Konsistenzfall vorliegt. Die erweiterte Zeilen-Stufen-Matrix hat dann die obige Form mit $m-n$ zusätzlichen Nullzeilen darunter. ■

BEISPIELE (zur kanonischen Zeilen-Stufen-Form):

1)
$$\begin{array}{r} 3x - 2y + z = 2 \\ x - 2z = 3 \end{array}$$

Dieses Beispiel haben wir schon früher bei den Beispielen zur gewöhnlichen Zeilen-Stufen-Form behandelt.

$$\begin{array}{c|c} \begin{array}{ccc|c} 3 & -2 & 1 & 2 \\ 1 & 0 & -2 & 3 \end{array} & \begin{array}{c} \iff \\ \textcircled{1} \rightarrow \textcircled{1} - 3 \cdot \textcircled{2} \end{array} & \begin{array}{ccc|c} 0 & -2 & 7 & -7 \\ 1 & 0 & -2 & 3 \end{array} & \begin{array}{c} \textcircled{1} \xrightarrow{-\frac{1}{2}} \textcircled{1} \\ \textcircled{1} \leftrightarrow \textcircled{2} \end{array} & \begin{array}{ccc|c} 1 & 0 & -2 & 3 \\ 0 & 1 & -\frac{7}{2} & \frac{7}{2} \end{array} \end{array}$$

Damit ist die kanonische Zeilen-Stufen-Form ohne Spaltenvertauschungen erreicht, und die 1-parametrische Lösungsdarstellung $x = 3 + 2t, y = \frac{7}{2}(t+1), z = t$ ablesbar. Wir können aber auch anders vorgehen:

$$\begin{array}{c|c} \begin{array}{ccc|c} 3 & -2 & 1 & 2 \\ 1 & 0 & -2 & 3 \end{array} & \begin{array}{c} \iff \\ \textcircled{2} \rightarrow -\frac{1}{2}\textcircled{2} \\ \textcircled{1} \rightarrow -\frac{1}{2}\textcircled{1} + \frac{1}{2}\textcircled{2} \end{array} & \begin{array}{ccc|c} x & y & z & \\ -\frac{7}{4} & 1 & 0 & -\frac{7}{4} \\ -\frac{1}{2} & 0 & 1 & -\frac{3}{2} \end{array} & \iff & \begin{array}{ccc|c} y & z & x & \\ 1 & 0 & -\frac{7}{4} & -\frac{7}{4} \\ 0 & 1 & -\frac{1}{2} & \frac{3}{2} \end{array} \end{array}$$

Dies ist eine andere kanonische Zeilen-Stufen-Form desselben Gleichungssystems — widerspricht das nicht der behaupteten Eindeutigkeit der kanonischen Zeilen-Stufen-Form?!

Nein, weil die Zuordnung der Unbekannten zu den Spalten hier eine andere ist. Jetzt ist x die Nicht-Basisvariable, und wir erhalten die Parameterdarstellung $x = r$, $y = \frac{7}{4}(r-1)$, $z = \frac{1}{2}r - \frac{3}{2}$. Da der Koeffizient von x in der obersten Zeile nicht verschwindet, kann man auch y zur Nicht-Basisvariablen machen und erhält die Parameterdarstellung der Lösung $x = \frac{4}{7}s + 1$, $y = s$, $z = \frac{2}{7}s - 1$. Jede andere Wahl von Zeilenoperationen und Spaltenvertauschungen, um eine kanonische Zeilen-Stufen-Form zu erzielen, führt auf eine dieser drei Parameterdarstellungen der Lösungsmenge und bei gegebener Zuordnung der Unbekannten zu den Spalten zu derselben kanonischen Zeilen-Stufen-Matrix!

2) Bei
$$\begin{array}{r} 3x - 2y + z = 2 \\ -1.5 + y - .5z = b \end{array}$$

soll in Abhängigkeit vom Parameter $b \in \mathbb{R}$ die Lösbarkeit diskutiert und die Lösungsmenge bestimmt werden. Hier rechnen wir

$$\begin{array}{c} \begin{array}{ccc|c} 3 & -2 & 1 & 2 \\ -1.5 & 1 & -0.5 & b \end{array} \\ \textcircled{2} \rightarrow \textcircled{2} + \frac{1}{2}\textcircled{1} \end{array} \iff \begin{array}{ccc|c} 3 & -2 & 1 & 2 \\ 0 & 0 & 0 & b+1 \end{array} ,$$

die beiden Gleichungen sind also linear abhängig (im Nachhinein sieht man jetzt natürlich, dass die untere Koeffizientenzeile aus der oberen durch Multiplikation mit $-\frac{1}{2}$ hervorgeht). Durch Multiplikation der ersten Zeile mit $\frac{1}{3}$ oder $-\frac{1}{2}$ nebst Spaltenvertauschungen erreichen wir die kanonischen Zeilen-Stufen-Formen

$$\begin{array}{ccc|c} x & y & z & \\ \hline 1 & -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ 0 & 0 & 0 & b+1 \end{array} , \quad \begin{array}{ccc|c} y & x & z & \\ \hline 1 & -\frac{3}{2} & -\frac{1}{2} & -1 \\ 0 & 0 & 0 & b+1 \end{array} , \quad \begin{array}{ccc|c} z & y & x & \\ \hline 1 & -2 & 3 & 2 \\ 0 & 0 & 0 & b+1 \end{array} .$$

Nur für $b = -1$ existiert eine Lösung, und dann ist die Lösungsmenge eine 2-dimensionale lineare Schar, wobei wir hier zwei beliebige der drei Unbekannten als Parameter wählen können, z.B. $x = r$, $y = s$, $z = -3r + 2s + 2$.

3)
$$\begin{array}{ccc|c} x & y & z & \\ \hline 2 & 0 & -1 & 0 \\ -1 & 2 & 0 & -2 \\ -3 & -1 & 2 & 2 \end{array} \quad \text{hatten wir früher schon} \quad \begin{array}{ccc|c} z & y & x & \\ \hline -1 & 0 & 2 & 0 \\ 0 & 2 & -1 & -2 \\ 0 & 0 & 1 & 2 \end{array} \quad \text{als eine}$$

Zeilen-Stufen-Form berechnet. Hier subtrahieren wir nun das 2-fache der dritten von der ersten Zeile, addieren die dritte zur zweiten Zeile und multiplizieren die erste Zeile noch mit -1 , die zweite mit $\frac{1}{2}$, um die kanonische Zeilen-Stufen-Form zu erhalten (die zweite ergibt sich aus der ersten durch Spalten und Zeilenvertauschung):

$$\begin{array}{ccc|c} z & y & x & \\ \hline 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \end{array} \iff \begin{array}{ccc|c} x & y & z & \\ \hline 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 4 \end{array} ,$$

woraus man die eindeutige Lösung $x = 2$, $y = 0$, $z = 4$ unmittelbar abliest. Die Koeffizientenmatrix in der kanonischen Zeilen-Stufen-Form ist hier die Einheitsmatrix, und das ist immer so, wenn man n unabhängige lineare Gleichungen für n Unbekannte hat.

4)
$$\begin{array}{ccc|c} x & y & z & \\ \hline 2 & 0 & -1 & 0 \\ -1 & 1 & 0 & -2 \\ -3 & -1 & 2 & 2 \end{array}$$
 hatten wir früher schon
$$\begin{array}{ccc|c} z & y & x & \\ \hline -1 & 0 & 2 & 0 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 \end{array}$$
 als eine Zeilen-Stufen-Form erhalten. Nach Multiplikation der ersten Zeile mit -1 hat man schon eine kanonische Zeilen-Stufen-Form:

$$\begin{array}{ccc|c} z & y & x & \\ \hline 1 & 0 & -2 & 0 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 \end{array}$$

Hier ist die Zahl der unabhängigen Gleichungen $l = 2$, daher hat die Koeffizientenmatrix eine Nullzeile. Weil aber auch der Eintrag auf der rechten Seite dieser Zeile Null ist, gibt es Lösungen. Man erhält alle, indem man x als Parameter wählt, in der Form $x = r$, $y = r - 2$, $z = 2r$ aus der mittleren und oberen Gleichung. Da die Koeffizienten von x in der ersten und zweiten Zeile $\neq 0$ sind, kann man auch y oder z als Parameter wählen: $x = s + 2$, $y = s$, $z = 2s + 4$ beziehungsweise $x = \frac{1}{2}t$, $y = \frac{1}{2}t - 2$, $z = t$. Wäre die kanonische Zeilen-Stufen-Form z.B.

$$\begin{array}{ccc|c} z & y & x & \\ \hline 1 & 0 & -2 & 0 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 0 \end{array},$$

so könnte man nur x oder z als Nicht-Basisvariable (Parameter) wählen, da die Koeffizienten von x und z in der zweiten Zeile Null sind. Die zweite Gleichung $y = -2$ bestimmt ja auch y eindeutig, so dass y nicht beliebige Werte annehmen kann.

5)
$$\begin{array}{l} x_1 + 2x_2 - x_3 + x_4 = 0 \\ 2x_1 + 4x_2 + x_3 - x_4 = 0 \\ -x_1 - 2x_2 + 2x_3 = 0 \\ x_3 + x_4 = 0 \end{array}$$
 mit Koeffizientenmatrix
$$\begin{pmatrix} 1 & 2 & -1 & 1 \\ 2 & 4 & 1 & -1 \\ -1 & -2 & 2 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

erreichen wir durch Addition der ersten zur dritten Zeile und Subtraktion des Doppelten der ersten von der zweiten Zeile zunächst einen kanonischen Basisvektor in der ersten Spalte (die rechten Seiten bei diesem homogenen Gleichungssystem sind alle Null und werden nicht notiert):

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

Ein weiterer, davon verschiedener, kanonischer Basisvektor kann jetzt nur in der dritten oder vierten Spalte erzeugt werden. Mit Subtraktion der dritten von der vierten und des Dreifachen der dritten von der zweiten Zeile und mit anschließender Division der zweiten Zeile durch -6 und Subtraktion der entstehenden Zeile von der dritten erreichen wir leicht die folgende kanonische Zeilen-Stufen-Form der Koeffizientenmatrix:

$$\begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ \hline 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \iff \begin{pmatrix} x_1 & x_4 & x_3 & x_2 \\ \hline 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Dies ist eine kanonische Zeilen-Stufen-Form mit $l = 3$ unabhängigen Gleichungen. Die zugehörige Nullspalte der rechten Seiten haben wir nicht notiert. Um die 1-dimensionale Lösungsmenge zu parametrisieren, kann man x_2 oder, weil der Koeffizient von x_2 in der ersten Zeile nicht Null ist, x_1 als Parameter wählen, nicht aber x_3 oder x_4 . Die eine Parametrisierung der Lösungen ist dann $x_1 = -2s$, $x_2 = s$, $x_3 = 0$, $x_4 = 0$ und die andere $x_1 = r$, $x_2 = -\frac{1}{2}r$, $x_3 = 0$, $x_4 = 0$. Da x_3 und x_4 hier nur Basisvariable sein können, ist die Herstellung der kanonischen Zeilen-Stufen-Form mit den drei ersten kanonischen Einheitsvektoren als vorderen Spalten nicht ohne Spaltenvertauschung möglich. (Das wollten wir mit diesem Beispiel zeigen. Die Lösungen kann man ohne Rechnung schon aus der Koeffizientenmatrix oben mit kanonischer erster Spalte direkt erkennen, weil die zweite und dritte Gleichung offenbar nur von $x_3 = 0$ und $x_4 = 0$ gelöst werden.) ■

BEMERKUNGEN: 1) Es versteht sich, dass man bei der Herstellung der kanonischen Zeilen-Stufen-Form *bereits in der Koeffizientenmatrix vorhandene Nulleinträge zur Vereinfachung der Rechnungen ausnutzt*, indem man die Spalten, in denen kanonische Basisvektoren erzeugt werden sollen, so aussucht, dass sie schon viele Nulleinträge haben.

2) Das letzte Beispiel zeigt, dass man aus den n Unbekannten nicht einfach $n-l$ beliebige als Nicht-Basisvariable herausgreifen kann, wenn l die Zahl der unabhängigen Gleichungen des linearen Gleichungssystems ist. An der kanonischen Zeilen-Stufen-Form sieht man:

- *Genau dann sind $n-l$ herausgegriffene Unbekannte als Nicht-Basisvariable wählbar, also als Parameter in einer linearen Parametrisierung der Lösungsmenge im Konsistenzfall, wenn durch Zeilenoperationen l verschiedene kanonische Einheitsvektoren in den Spalten zu den anderen Unbekannten herstellbar sind.*

Leider kann man der ursprünglichen Koeffizientenmatrix meist nicht direkt ansehen, wieviele Nicht-Basisvariablen es gibt und welche der Unbekannten man dafür wählen kann, sondern das ergibt sich erst im Verlauf der Berechnung einer kanonischen Zeilen-Stufen-Form, wenn nämlich erkennbar wird, dass gewisse Spalten nur in den bereits verwendeten Pivot-Zeilen Einträge $\neq 0$ haben und somit für die Erzeugung neuer kanonischer Basisvektoren nicht mehr in Frage kommen. ■

DISKUSSION (*lineare Gleichungssysteme und Spaltenoperationen*):

1) (**Zulässige**) **Spaltenoperationen** mit einer Matrix werden natürlich ganz analog zu den Zeilenoperationen erklärt: Vertauschung von zwei Spalten, Multiplikation einer Spalte mit einer reellen Zahl $\neq 0$ (d.h. Multiplikation aller Einträge mit dieser Zahl), Addition einer Spalte zu einer anderen (d.h. Addition der Einträge einer gegebenen Spalte zu denen mit gleicher Position in einer anderen gegebenen Spalte) und Kombinationen davon. Bei Zeilenoperationen mit der erweiterten Koeffizientenmatrix $(A|b)$ eines Gleichungssystems $Ax = b$ bleibt, wie wir gesehen haben, die Lösungsmenge L erhalten. Hier gilt nun:

- *Bei zulässigen Spaltenoperationen mit der Koeffizientenmatrix bleibt die Menge Z der konsistenten rechten Seiten im zugehörigen Gleichungssystem erhalten.*

Wenn also $y \in \mathbb{R}^m$ zulässig ist, d.h. wenn $Ax = y$ eine Lösung hat, und wenn \tilde{A} aus A durch Spaltenoperationen hervorgeht, so hat auch $\tilde{A}\tilde{x} = y$ eine Lösung \tilde{x} und umgekehrt. Das braucht man nur für die elementaren Spaltenoperationen zu überprüfen: Bei Vertauschung zweier Spalten erhält man aus x die Lösung \tilde{x} einfach durch entsprechende Vertauschung der Komponenten von $x \in \mathbb{R}^n$, bei Multiplikation der j -ten Spalte mit $r \neq 0$ erhält man \tilde{x} durch Multiplikation der j -ten Komponente x_j mit $\frac{1}{r}$ und bei Addition der j -ten zur k -ten Spalte erhält man \tilde{x} , indem man bei x die k -te von der j -ten Komponente subtrahiert.

2) Um Informationen über die Menge der konsistenten rechten Seiten $y \in \mathbb{R}^m$ zu einer $m \times n$ -Matrix A zu gewinnen, können wir also A mit Spaltenoperationen beliebig umformen und auf eine einfache Gestalt bringen. Erreichbar ist z.B. wie bei Zeilentransformationen eine **Spalten-Stufen-Form** und, wenn man noch Zeilenvertauschungen zulässt (das verändert Z nur so, dass bei allen Vektoren in Z die Komponenten entsprechend vertauscht werden), sogar die **kanonische Spalten-Stufen-Form** der $m \times n$ -Matrix A :

$$\begin{array}{l} k \\ m-k \end{array} \left\{ \left(\begin{array}{ccc|c} 1 & & & \\ & 1 & 0 & \\ & & \ddots & \\ & 0 & & 1 \\ \hline & * & & 0 \end{array} \right) \right.$$

$\underbrace{\hspace{10em}}_k \quad \underbrace{\hspace{10em}}_{n-k}$

Hier ist die ganze Zahl $0 \leq k \leq \min(m, n)$ eindeutig durch A bestimmt und heißt der **Spaltenrang** der Matrix; seine Bedeutung ist die (**maximale**) **Anzahl der linear unabhängigen Spalten**, ganz analog zum früher eingeführten Zeilenrang l von A .

3) Aus der kanonischen Spalten-Stufen-Form C erhalten wir nun sofort eine Parametrisierung der Menge Z der konsistenten rechten Seiten $y \in \mathbb{R}^m$, indem wir die ersten k Komponenten als Parameter $y_1 = r_1, \dots, y_k = r_k$ wählen, das legt dann auch die ersten k Komponenten $x_1 = r_1, \dots, x_k = r_k$ der Lösungen $x \in \mathbb{R}^n$ zu $Cx = y$ fest, und y ist zulässig, genau wenn die restlichen Komponenten von y durch $y_i = c_{i1}r_1 + \dots + c_{ik}r_k$ gegeben sind. Das ist eine *spezielle k -parametrische lineare Parametrisierung von Z* und k ist hier die kleinstmögliche Parameterzahl. Gleichzeitig erhält man auch ein System von $m-k$ unabhängigen linearen Gleichungen $y_i = c_{i1}y_1 + \dots + c_{ik}y_k$, $k < i \leq m$, dessen Lösungsmenge Z ist. (Wenn zur Herstellung der kanonischen Zeilen-Stufen-Form Zeilenvertauschungen durchgeführt wurden, so ändert das nur die Nummerierung der Komponenten.)

4) Nun haben wir aber schon früher gesehen, dass die Menge Z der konsistenten rechten Seiten zu einer $m \times n$ -Matrix A eine l -dimensionale lineare Menge ist, wobei l der Zeilenrang von A ist. Lineare Parametrisierungen von Z haben also die minimale Parameterzahl l , und durch Vergleich mit 3) folgt $k = l$, d.h. wir haben folgendes fundamentale Ergebnis der Linearen Algebra:

- Für jede Matrix A gilt: **Zeilenrang(A) = Spaltenrang(A)**.

Diese ganze Zahl nennt man dann einfach den **Rang der Matrix**.

5) Eine beliebige homogene lineare Parametrisierung einer Teilmenge Z von \mathbb{R}^m mit n Parametern kann man immer als Beschreibung von Z als Menge der konsistenten rechten Seiten eines Gleichungssystems $Ax = y$ von m Gleichungen für n Unbekannte auffassen (x_1, \dots, x_n sind hier die Parameter). Aus 3) ergibt sich dann, dass dieselbe Menge Z auch mit der minimalen Zahl k von Parametern beschrieben werden kann, wenn k der (Spalten-)Rang von A ist, und dass man ein System von $m-k$ unabhängigen homogenen linearen Gleichungen angeben kann, dessen Lösungsmenge Z ist. Es folgt (mit n statt m) die **äquivalente Beschreibung k -dimensionaler affiner Unterräume E von \mathbb{R}^n** :

- E lässt sich *affin-linear* mit k Parametern parametrisieren, aber nicht mit weniger;
- E ist Lösungsmenge eines konsistenten Systems von $n-k$ unabhängigen, aber nicht eines Systems von mehr oder weniger unabhängigen, linearen Gleichungen. ■

3.3 Das Rechnen mit Vektoren und Matrizen

Matrizen haben wir bisher nur benutzt, um die Zeilenoperationen beim Lösen linearer Gleichungssysteme übersichtlich darstellen zu können. Man kann aber auch sinnvolle Rechenoperationen für Matrizen und Vektoren definieren und dafür einen leistungsfähigen Kalkül entwickeln, der sich zur Lösung linearer Gleichungssysteme einsetzen lässt. Darum geht es in diesem Abschnitt. Dabei werden alle Rechenoperationen und Rechenregeln letztlich auf das Rechnen mit reellen Zahlen, mit den Einträgen der Matrizen, zurückgeführt. Der hauptsächliche Vorteil des Kalküls ist, dass man damit viele gleichartige Rechnungen / Gleichungen mit reellen Zahlen übersichtlich zu einer einzigen Rechnung / Gleichung für Matrizen oder Vektoren zusammenfassen kann. Daher wird die Matrix- und Vektorrechnung auch für die Modellierung komplexer (aber immer noch "linearer") ökonomischer Abläufe eingesetzt.

DEFINITION: (i) Die **Summe zweier Matrizen** ist positionsweise definiert, wenn beide dasselbe Format haben, d.h. man addiert die Einträge beider Matrizen in gleicher Position und trägt die Summen in die entsprechenden Positionen der Ergebnismatrix ein. Sind also $A = (a_{ij})$ und $B = (b_{ij})$ Matrizen desselben Formats (m, n) , so ist ihre Summe die Matrix $C = A + B$ desselben Formats (m, n) , die Einträge $c_{ij} = a_{ij} + b_{ij}$ hat. Analog ist die **Differenz zweier Matrizen** $A = (a_{ij})$, $B = (b_{ij})$ von gleichem Format (m, n) die Matrix $D = A - B$ desselben Formats (m, n) , die Einträge $d_{ij} = a_{ij} - b_{ij}$ hat.

(ii) Die **Vervielfachung einer Matrix mit einem Faktor** $r \in \mathbb{R}$, auch **Multiplikation der Matrix mit einem Skalar** $r \in \mathbb{R}$ genannt (weil die Zahlen aus dem zu Grunde liegenden Zahlbereich in der Vektor- und Matrixrechnung auch als *Skalare* bezeichnet werden), ist definiert durch Multiplikation eines jeden Eintrags der Matrix A mit der Zahl r . Das Produkt rA ist also die Matrix desselben Formats wie $A = (a_{ij})$, welche die Einträge ra_{ij} hat. Die Matrix $(-1)A = (-a_{ij})$ heißt **das Negative der Matrix A** und wird $-A$ notiert. ■

DISKUSSION: 1) Man kann also nicht beliebige Matrizen addieren und subtrahieren:

- *Summe und Differenz sind nur für Matrizen gleichen Formats erklärt, und das Ergebnis ist wieder eine Matrix von diesem Format.*

Ebenso haben Vielfache rA einer Matrix A mit Faktoren $r \in \mathbb{R}$ dasselbe Format wie A . Wir schreiben skalare Faktoren meistens links vor die Matrizen bzw. die Zeilen- oder Spaltenvektoren. Das muss man nicht so halten; Ar ist dann dasselbe wie rA .

2) Ausführlicher geschrieben gehen die Rechenoperationen so:

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & b_{ij} & \vdots \\ b_{m1} & \cdots & b_{mn} \end{pmatrix} = \begin{pmatrix} a_{11}+b_{11} & \cdots & a_{1n}+b_{1n} \\ \vdots & a_{ij}+b_{ij} & \vdots \\ a_{m1}+b_{m1} & \cdots & a_{mn}+b_{mn} \end{pmatrix}$$

$$r \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} ra_{11} & \cdots & ra_{1n} \\ \vdots & ra_{ij} & \vdots \\ ra_{m1} & \cdots & ra_{mn} \end{pmatrix}$$

und entsprechend natürlich für die Differenzbildung (überall "−" statt "+") und für die Bildung des Negativen (überall "−" statt "r").

3) Für **Zeilen-** bzw. **Spaltenvektoren** sind diese Rechenoperationen natürlich auch erklärt, weil das ja spezielle Matrizen sind (mit nur einer Zeile bzw. nur einer Spalte). Dafür gilt also:

$$(a_1 \ a_2 \ \dots \ a_n) \pm (b_1 \ b_2 \ \dots \ b_n) = (a_1 \pm b_1 \ a_2 \pm b_2 \ \dots \ a_n \pm b_n),$$

$$r(a_1 \ a_2 \ \dots \ a_n) = (ra_1 \ ra_2 \ \dots \ ra_n),$$

bzw.

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} \pm \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} = \begin{pmatrix} a_1 \pm b_1 \\ a_2 \pm b_2 \\ \vdots \\ a_m \pm b_m \end{pmatrix}, \quad r \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} ra_1 \\ ra_2 \\ \vdots \\ ra_m \end{pmatrix}.$$

Den Unterschied zwischen Zeilen- und Spaltenvektoren muss man nicht zu genau nehmen; schließlich enthält ein Zeilenvektor genau dieselbe Information wie der Spaltenvektor, der dieselben Einträge (in gleicher Reihenfolge) hat. Die Summe und Differenz ist aber nur für Vektoren gleichen Formats erklärt (also gleiche Komponentenzahl und beides Zeilenvektoren oder beides Spaltenvektoren), und das Ergebnis der Rechenoperationen ist wieder ein Vektor dieses Formats. Wir erinnern daran, dass wir statt Spaltenvektoren auch m -tupel schreiben, deren Einträge durch Kommas abgetrennt sind. Dafür sehen die Rechenoperationen dann aus wie bei Zeilenvektoren: $(a_1, a_2, \dots, a_m) \pm (b_1, b_2, \dots, b_m) = (a_1 \pm b_1, a_2 \pm b_2, \dots, a_m \pm b_m)$ und $r(a_1, a_2, \dots, a_m) = (ra_1, ra_2, \dots, ra_m)$.

4) Natürlich kann man analog auch mehr als zwei Matrizen A_1, A_2, \dots, A_l addieren, wenn alle Summanden dasselbe Format haben. Und vor der Addition kann man sie noch mit reellen Faktoren r_1, r_2, \dots, r_l multiplizieren. So erhält man die sog. **Linearkombination der Matrizen A_k mit Koeffizienten r_k** , notiert

$$r_1 A_1 + r_2 A_2 + \dots + r_l A_l = \sum_{k=1}^l r_k A_k.$$

Dies ist eine Matrix desselben Formats wie alle A_k , und sie hat in Zeile i und Spalte j den Eintrag $r_1 a_{ij}^{(1)} + r_2 a_{ij}^{(2)} + \dots + r_l a_{ij}^{(l)}$, wenn A_k die Einträge $a_{ij}^{(k)}$ besitzt für $k = 1 \dots l$. Man bildet also auch Linearkombinationen positionsweise. Besonders oft kommen **Linearkombinationen von Vektoren** gleichen Formats vor, z.B. ist

$$ra + sb + tc = r \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} + s \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} + t \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix} = \begin{pmatrix} ra_1 + sb_1 + tc_1 \\ ra_2 + sb_2 + tc_2 \\ \vdots \\ ra_m + sb_m + tc_m \end{pmatrix}$$

eine Linearkombination von drei Vektoren $\mathbf{a}, \mathbf{b}, \mathbf{c}$ aus \mathbb{R}^m . (Das Ergebnis ist wieder ein Spaltenvektor mit m Komponenten, auch wenn er recht "breit" geraten ist.) Lässt man die Parameter r, s, t hier beliebige reelle Zahlen durchlaufen, so bildet die entsprechende Menge aller Linearkombinationen der drei Vektoren das, was wir in 3.2 eine 3-parametrische lineare Schar in \mathbb{R}^m genannt hatten. Entsprechend bildet die Gesamtheit der Linearkombinationen $r_1 \mathbf{a}_1 + \dots + r_l \mathbf{a}_l$ von l Vektoren $\mathbf{a}_1, \dots, \mathbf{a}_l \in \mathbb{R}^m$ eine l -parametrische lineare Schar in \mathbb{R}^m (mit der speziellen Eigenschaft, dass sie den m -gliedrigen Nullvektor enthält; denn der entsteht, wenn man alle Koeffizienten $r_k = 0$ setzt). ■

BEISPIELE (zur Matrix-/Vektoraddition und Multiplikation mit Skalaren):

$$1) \quad \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \begin{pmatrix} 6 & 5 & 4 \\ 3 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 7 & 7 & 7 \\ 7 & 7 & 7 \end{pmatrix} = 7 \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

$$2) \text{ Es ist} \quad (1 \ 2) + (-1 \ 1) + (0 \ -3) = 0$$

$$\text{und} \quad 0 \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 0,$$

aber man darf *nicht* schreiben

$$(1 \ 2) + (-1 \ 1) + (0 \ -3) = 0 = 0 \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (???) ;$$

denn die erste "0" steht für die 1×2 -Nullmatrix (Zeilenvektor), die zweite aber für die 2×1 -Nullmatrix (Spaltenvektor). *Nullmatrizen von verschiedenem Format sind verschieden* — wie generell Matrizen verschiedenen Formats! Wenn nötig, so kann man das Format einer Nullmatrix als Subskript angeben, etwa hier $0_{1 \times 2}$, $0_{2 \times 1}$ und allgemein $0_{m \times n}$ bei der $m \times n$ -Nullmatrix.

3) Einige Linearkombinationen:

$$2(1 \ -1) - \frac{1}{2}(0 \ 4) + 3(2 \ 0) = (8 \ -4),$$

$$\begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix} - \frac{1}{3} \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix} - \frac{1}{6} \begin{pmatrix} 0 & -1 \\ 1 & 6 \end{pmatrix} = 0 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

$$x_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \dots + x_m \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_m \end{pmatrix} \in \mathbb{R}^m.$$

Die letzte Rechnung zeigt:

- Jeder Vektor in $x \in \mathbb{R}^m$ ist eine Linearkombination der kanonischen Basisvektoren e_1, \dots, e_m von \mathbb{R}^m ; die Koeffizienten sind dabei die Komponenten x_i von x und eindeutig bestimmt,

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} = \sum_{i=1}^m x_i e_i.$$

4) Die Linearkombination von l Vektoren / Matrizen mit gleichen Koeffizienten $r_k = \frac{1}{l}$ für $k = 1 \dots l$ nennt man ihr **arithmetisches Mittel**. Zum Beispiel sind

$$\frac{1}{2} \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 1/2 \\ 1/2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

und

$$\frac{1}{3} \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} + \frac{1}{3} \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} + \frac{1}{3} \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix} = \begin{pmatrix} 2/3 \\ 1/3 \\ -1/3 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}$$

arithmetische Mittel von zwei bzw. drei Vektoren aus \mathbb{R}^3 .

Die geometrische Bedeutung des arithmetischen Mittels von zwei Vektoren $a, b \in \mathbb{R}^m$ ist der *Mittelpunkt* der Verbindungsstrecke von a nach b in \mathbb{R}^m . Das arithmetische Mittel von drei Vektoren in \mathbb{R}^m ist der *Schwerpunkt* des Dreiecks, das diese Vektoren als Eckpunkte hat. Das arithmetische Mittel der kanonischen Basisvektoren in \mathbb{R}^m ist

$$\frac{1}{m} \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \frac{1}{m} \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \dots + \frac{1}{m} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1/m \\ 1/m \\ 1/m \\ \vdots \\ 1/m \end{pmatrix} = \frac{1}{m} \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^m.$$

5) Wir betrachten eine *ökonomische Anwendung* (formaler Art, d.h. ökonomische Probleme werden nicht gelöst): In einem Unternehmen mit drei Teilbetrieben wird für jedes Quartal eines Jahres die **Verflechtungsmatrix** aufgestellt, welche die Leistungen der liefernden Stellen (Spalten) an die empfangenden Stellen (Zeilen) in Leistungseinheiten angibt. Die vier Verflechtungsmatrizen für die vier Quartale seien z.B.

$$A_1 = \begin{pmatrix} 0 & 20 & 0 \\ 30 & 0 & 10 \\ 10 & 10 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 15 & 0 \\ 25 & 0 & 5 \\ 10 & 5 & 0 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0 & 20 & 0 \\ 25 & 0 & 15 \\ 15 & 10 & 0 \end{pmatrix}, \quad A_4 = \begin{pmatrix} 0 & 25 & 0 \\ 40 & 0 & 10 \\ 5 & 15 & 0 \end{pmatrix}.$$

Dann gibt die Summe der vier Matrizen

$$A_1 + A_2 + A_3 + A_4 = \begin{pmatrix} 0 & 80 & 0 \\ 120 & 0 & 40 \\ 40 & 40 & 0 \end{pmatrix}$$

die Verflechtungsmatrix für das ganze Jahr an und z.B. die Differenz

$$A_4 - A_3 = \begin{pmatrix} 0 & 5 & 0 \\ 15 & 0 & -5 \\ -10 & 5 & 0 \end{pmatrix}$$

die Mehr- bzw. Minderleistungen im vierten Quartal gegenüber dem dritten Quartal. (Beachte, dass hier kein Teilbetrieb Eigenleistungen verbraucht, also die A_k Nulleinträge auf der Diagonalen haben; das gilt dann natürlich auch für die Summe und für Differenzen.) Hätte man in jedem Quartal dieselben Leistungen verzeichnet wie im ersten, so hätte sich als Verflechtungsmatrix für das gesamte Jahr $4A_1$ ergeben. Hier ist "zufällig" $4A_1$ gleich der Summe $A_1 + A_2 + A_3 + A_4$, d.h. die im Vergleich mit dem ersten Quartal festzustellenden Mehr- und Minderleistungen im zweiten, dritten und vierten Quartal heben sich gerade auf. A_1 ist daher auch das arithmetische Mittel der vier Verflechtungsmatrizen.

6) Noch eine (formale) *ökonomische Anwendung*: Ein Betrieb stellt an Betriebsstätten $\textcircled{1} \dots \textcircled{m}$ dasselbe Produkt her. Die monatlich produzierten Mengen erfasst ein Spaltenvektor (x_1, \dots, x_m) , der sog. **Produktionsvektor**. Der Eintrag in Position \textcircled{i} gibt dabei die an der Betriebsstätte \textcircled{i} produzierte Produktmenge (in Mengeneinheiten) an. Für die 12 Monate eines Jahres hat man dann 12 solche Produktionsvektoren $\mathbf{x}_1, \dots, \mathbf{x}_{12}$ (jeder mit m Einträgen). Werden die produzierten Mengen im k -ten Monat am Markt zum Preis p_k abgesetzt, so ist die Linearkombination

$$p_1 \mathbf{x}_1 + p_2 \mathbf{x}_2 + \dots + p_{12} \mathbf{x}_{12}$$

ein Spaltenvektor mit m Komponenten, dessen i -ter Eintrag den von der i -ten Produktionsstätte erzielten Jahreserlös angibt; man könnte ihn also den *Jahreserlösvektor* des in die Stätten $\textcircled{1} \dots \textcircled{m}$ aufgegliederten Betriebs nennen. Seine Komponentensumme (die Summe all seiner Einträge) ist dann der Jahreserlös des gesamten Betriebs. ■

Natürlich braucht man nicht wirklich Vektor- und Matrizenrechnung, um die ökonomischen Sachverhalte in den beiden letzten Beispielen zu erfassen. Man kann ja den Jahresaldo auch separat für jeden Teilbetrieb bzw. für jede Produktionsstätte ermitteln, und das läuft auf Dasselbe hinaus wie die in den Beispielen ausgeführte Addition der Verflechtungsmatrizen bzw. Linearkombination der Produktionsvektoren. Der Vorteil des Gebrauchs von Matrizen und Vektoren liegt hier nur in der konzentrierten Schreibweise und der Zusammenfassung vieler gleichartiger Vorgänge oder Berechnungen:

- *Mehrere gleichartige Rechnungen mit reellen Zahlen kann man oft zu einer einzigen Vektor- oder Matrixrechnung zusammenfassen.*

Aus den Rechengesetzen im Zahlbereich \mathbb{R} und aus der Definition der Addition und Vervielfachung von Matrizen / Vektoren ergeben sich unmittelbar entsprechende Rechengesetze, die man zur Grundlage der Definition eines allgemeinen Vektorraums gemacht hat. Diese abstrakte Begriffsbildung ist in der Mathematik fundamental für die Lineare Algebra. Für Anwendungen der Linearen Algebra in der Ökonomie ist sie aber nicht so wichtig, weil man es dort immer mit konkreten Vektorräumen zu tun hat, z.B. mit dem Raum \mathbb{R}^m der m -gliedrigen Spaltenvektoren (bzw. m -tupel) oder dem Raum $\mathbb{R}^{m \times n}$ der $m \times n$ -Matrizen mit reellen Einträgen. Wir stellen deshalb hier auch nicht den Begriff des Vektorraums in den Vordergrund, sondern erklären ihn nur kurz (unter anderem deswegen, weil sich daraus die Antwort auf die oft gestellte Frage ergibt: "Was ist eigentlich ein Vektor?").

DISKUSSION (*Rechenregeln für die Addition und Vervielfachung; Vektorräume*)

1) Die **Rechenregeln für die Addition** von Matrizen sind:

$$\begin{aligned} A + B &= B + A && \text{(Kommutativgesetz)} \\ (A + B) + C &= A + (B + C) && \text{(Assoziativgesetz)} \\ A + 0 &= A = 0 + A && \text{(neutrales Element)} \\ A + B = 0 &\iff B = -A && \text{(additives Inverses)} \end{aligned}$$

Hier sind — selbstverständlich — A, B, C und die Nullmatrix 0 Matrizen desselben Formats. Die beiden ersten Gesetze haben zur Folge, dass die Summe von beliebig vielen gleichformatigen Matrizen definiert ist und von der Klammerung und der Reihenfolge der Summanden nicht abhängt. Die Summe hat natürlich dasselbe Format wie alle Summanden. Die Rechengesetze hier gelten auch für Vektoren, also für einzeilige oder einspaltige Matrizen. Nur schreibt man dann meistens kleine Buchstaben a, b, \dots, x, y, \dots (manchmal auch fett $\mathbf{a}, \mathbf{b}, \dots, \mathbf{x}, \mathbf{y}, \dots$) statt der für Matrizen üblichen großen Buchstaben.

2) Die **Rechenregeln für die Multiplikation mit einem Skalar** sind:

$$\begin{aligned} 1A &= A, \quad 0A = 0 \\ r(sA) &= (rs)A && \text{(Assoziativgesetz)} \\ (r + s)A &= rA + sA && \text{(linkes Distributivgesetz)} \\ r(A + B) &= rA + rB && \text{(rechtes Distributivgesetz)} \end{aligned}$$

Hierbei sind $0, 1, r, s \in \mathbb{R}$ und A, B Matrizen desselben Formats wie die Nullmatrix 0 . Für das Rechnen mit Zeilen- oder Spaltenvektoren gelten diese Gesetze natürlich auch. Da sie ganz analog zu Rechengesetzen für reelle Zahlen sind, ist es nicht schwer, damit richtig umzugehen.

3) Hat man eine Menge V mit einem ausgezeichneten Element 0 und ist für je zwei Elemente $v, w \in V$ sowie für Zahlen $r \in \mathbb{R}$ eine Summe $v + w \in V$ und das Vielfache $rv \in V$ definiert, derart dass die in 1) und 2) angegebenen Rechenregeln analog gelten, so nennt man V mit dieser Struktur einen **reellen Vektorraum**, die Elemente v, w, \dots von V **Vektoren** und 0 den **Nullvektor**. Der m -dimensionale Zahlenraum \mathbb{R}^m der Spaltenvektoren mit m reellen Komponenten und der Raum $\mathbb{R}^{m \times n}$ der $m \times n$ -Matrizen mit reellen Einträgen sind die hier wichtigsten Beispiele eines Vektorraums, aber es gibt viele weitere. Zum Beispiel bilden die Polynomfunktionen auf \mathbb{R} oder die differenzierbaren Funktionen aus einem offenen Intervall I in \mathbb{R} einen Vektorraum, wenn dafür die Addition und Vervielfachung wie üblich (wertweise) definiert und die Nullfunktion als Nullvektor genommen wird. Auch der Bereich der reellen Zahlen \mathbb{R} selbst bildet einen reellen Vektorraum (weil die geforderten Rechenregeln in \mathbb{R} alle gelten). Daher ist es nicht ganz korrekt zu sagen, ein Vektor sei ein "Objekt mit mehreren Komponenten" oder Einträgen wie eben ein m -tupel (mit $m > 1$) oder eine Matrix (mit mehr als einem Eintrag). Man kann eigentlich nicht definieren, was ein einzelner "Vektor" ist, sondern der richtige Begriff ist der des *Vektorraums* mit seinen Rechenoperationen und Rechengesetzen. Wenn man einen Vektorraum hat, dann nennt man seine Elemente eben auch "Vektoren", worum immer es sich dabei auch handeln mag (m -tupel, Matrizen, Polynome, Funktionen, ...).

4) Der Vorteil der Begriffsbildung des abstrakten Vektorraums ist, dass man damit gleichartige Konstruktionen und Rechnungen in vielen ganz verschiedenen Situationen einheitlich erfassen und ausführen kann. So lassen sich z.B. Linearkombinationen nicht nur von Spaltenvektoren und Matrizen, sondern auch von Polynomen, Funktionen, ..., eben von Vektoren in einem beliebigen Vektorraum V erklären, nämlich

$$r_1 v_1 + r_2 v_2 + \dots + r_l v_l \in V$$

als **Linearkombination der Vektoren** $v_1, v_2, \dots, v_l \in V$ mit **Koeffizienten** $r_1, r_2, \dots, r_l \in \mathbb{R}$, und man kann damit genau so rechnen wie mit Linearkombinationen von Spaltenvektoren in \mathbb{R}^m oder von $m \times n$ -Matrizen.

Auch viele geometrische Vorstellungen, die man mit Vektoroperationen in der durch \mathbb{R}^2 modellierten "Zeichenebene" oder in dem durch \mathbb{R}^3 modellierten "Anschauungsraum" verbindet, lassen sich auf beliebige Vektorräume übertragen und erleichtern dann den Umgang mit und das Verständnis von Vektorrechnungen. Zum Beispiel bildet die Menge aller Linearkombinationen der Form $(1-t)v + tw$ zu zwei verschiedenen Vektoren $v, w \in V$ mit $t \in \mathbb{R}$, also die Menge aller Linearkombinationen dieser beiden Vektoren mit Koeffizientensumme 1, die sog. **Verbindungsgerade** der Punkte v, w in V . Und wenn man noch verlangt, dass die Koeffizienten nichtnegativ sind, so erhält man die **Verbindungsstrecke** $\{(1-t)v + tw : 0 \leq t \leq 1\}$ der beiden Punkte mit dem **Mittelpunkt** $\frac{1}{2}v + \frac{1}{2}w$. Allgemeiner nennt man eine Linearkombination von $v_1, v_2, \dots, v_l \in V$ eine **affine Kombination** dieser Vektoren, wenn die Koeffizientensumme 1 ist, und eine **konvexe Kombination**, wenn außerdem noch alle Koeffizienten nichtnegativ sind. Eine konvexe Kombination kann man auch als **gewichtetes arithmetisches Mittel der Vektoren** ansehen,

$$s_1 v_1 + s_2 v_2 + \dots + s_l v_l \in V \quad (s_k \geq 0 \text{ für } k = 1 \dots l, s_1 + s_2 + \dots + s_l = 1).$$

Für drei verschiedene Punkte $u, v, w \in V$, von denen keiner auf der Verbindungsgeraden der beiden anderen liegt, bilden die affinen Kombinationen anschaulich gesprochen die Ebene in V , die durch diese drei Punkte bestimmt ist, und die konvexen Kombinationen füllen das Dreieck darin mit den Ecken u, v, w aus; das arithmetische Mittel $\frac{1}{3}u + \frac{1}{3}v + \frac{1}{3}w$ ist z.B. der Schwerpunkt dieses Dreiecks.

All diese geometrischen Konstruktionen und Sprechweisen machen nicht nur in der Zeichenebene \mathbb{R}^2 und im Anschauungsraum \mathbb{R}^3 Sinn, sondern — weil sie nur die Rechenoperationen und Rechengesetze eines Vektorraums erfordern — in *jedem* Vektorraum, also auch im m -dimensionalen Zahlenraum \mathbb{R}^m mit $m > 3$, im Raum der $m \times n$ -Matrizen $\mathbb{R}^{m \times n}$, in Vektorräumen von Funktionen usw. Das ist eine der nützlichen Erkenntnisse, welche das Konzept des allgemeinen Vektorraums mit sich bringt. ■

Wichtiger als die Addition, aber nicht so einfach zu definieren, ist die Multiplikation von Matrizen (deren Formate in gewisser Weise zusammen “passen”). Zur Vorbereitung definieren wir das für die Anwendungen ebenfalls sehr wichtige Skalarprodukt zweier Zeilen- oder Spaltenvektoren mit gleicher Komponentenzahl. Der Name dieses Produkts kommt daher, dass das Ergebnis ein Skalar ist, also eine reelle Zahl. Das Skalarprodukt ist auch wichtig für die Geometrie, weil mit seiner Hilfe Grundgrößen wie Abstand und Winkel erklärt werden. Hier aber interessieren uns nur die formalen Eigenschaften dieses Produkts.

DEFINITION: Sind a und b zwei Zeilen- oder Spaltenvektoren mit gleich vielen reellen Einträgen a_1, \dots, a_n und b_1, \dots, b_n so ist das **Skalarprodukt der beiden Vektoren** (auch **inneres Produkt** genannt) definiert als die reelle Zahl

$$a \cdot b := a_1 b_1 + a_2 b_2 + \dots + a_n b_n \in \mathbb{R}.$$

Um das Skalarprodukt $a \cdot b$ zu erhalten, multipliziert man also die Einträge von a und b mit gleicher Positionsnummer und addiert alle so entstehenden Produkte. ■

DISKUSSION: 1) Bei der Bildung des Skalarprodukts ist es gleichgültig, ob die Faktoren Zeilen- oder Spaltenvektoren bzw. n -tupel sind. Die Faktoren müssen auch nicht von gleichem Format sein, d.h. es ist durchaus erlaubt, dass ein Faktor ein Zeilenvektor, der andere Faktor ein Spaltenvektor ist. Einzige Bedingung ist, dass die Faktoren Vektoren mit gleich vielen Komponenten sind.

- *Das Skalarprodukt ist nur für zwei (Zeilen- oder Spalten-) Vektoren mit gleicher Anzahl von Einträgen definiert; das Ergebnis ist dann ein Skalar, also eine reelle Zahl.*

Dementsprechend kann man auch nicht das Skalarprodukt $a \cdot b \cdot c$ von drei Vektoren mit gleicher Komponentenzahl bilden, weil $a \cdot b$ eine reelle Zahl und das Skalarprodukt von $a \cdot b$ mit c nicht erklärt ist. (Ausnahme: a, b, c haben jeweils nur eine Komponente, sind also reelle Zahlen; dann ist ihr Skalarprodukt das gewöhnliche Produkt reeller Zahlen.) Allenfalls kann man $(a \cdot b)c$ als Produkt sc des Skalars $s := a \cdot b$ mit dem Vektor c erklären.

2) Analog zur Addition könnte man ein Produkt “ $*$ “ für Spaltenvektoren mit gleicher Komponentenzahl definieren durch positionsweise Multiplikation der Einträge, also durch $(a_1, a_2, \dots, a_n) * (b_1, b_2, \dots, b_n) := (a_1 b_1, a_2 b_2, \dots, a_n b_n)$. Dieses Produkt würde auch den üblichen Rechenregeln für Produkte genügen. Nur ist es für die Anwendungen völlig bedeutungslos im Gegensatz zum Skalarprodukt von Vektoren und der damit später definierten Matrixmultiplikation. Deshalb wird dieses Produkt auch nicht eingeführt.

3) Wir haben die *Notation* $a \cdot b$ mit einem fetten Multiplikationspunkt für das Skalarprodukt gewählt, damit man es vom Produkt $a \cdot b = ab$ reeller Zahlen a, b unterscheiden und an der Notation sofort erkennen kann, dass es sich um das Skalarprodukt von Vektoren mit gleicher Komponentenzahl handelt. Andere gebräuchliche Bezeichnungen für das Skalarprodukt sind $\langle a, b \rangle$ oder (a, b) . (Letzteres ist aber ungünstig, weil die Gefahr der Verwechslung mit dem Paar der Vektoren a und b besteht, das ebenfalls (a, b) bezeichnet wird.) Bei zwei Spaltenvektoren mit gleicher Komponentenzahl findet man auch die

Notation $a^T b$ für ihr Skalarprodukt, wobei a^T der transponierte Vektor zu a ist, d.h. der Zeilenvektor mit denselben Einträgen (in gleicher Reihenfolge) wie der Spaltenvektor a . (Eigentlich ist $a^T b$ dann das Produkt einer einzeiligen Matrix a^T mit einer einspaltigen Matrix b , und dieses Produkt hat als Ergebnis eine 1×1 -Matrix mit dem Skalarprodukt $a \cdot b$ als einzigem Eintrag. Aber zwischen reellen Zahlen s und 1×1 -Matrizen (s) macht man gewöhnlich keinen Unterschied.)

4) Aus der Definition des Skalarprodukts und den Rechenregeln im Zahlbereich \mathbb{R} ergeben sich unmittelbar folgende **Rechengesetze für das Skalarprodukt**:

$$\begin{aligned} a \cdot b &= b \cdot a && (\text{Symmetrie}) \\ (ra) \cdot b &= r(a \cdot b) = a \cdot (rb) && (\text{Homogenität}) \\ (a \pm \tilde{a}) \cdot b &= a \cdot b \pm \tilde{a} \cdot b && (\text{linkes Distributivgesetz}) \\ a \cdot (b \pm \tilde{b}) &= a \cdot b \pm a \cdot \tilde{b} && (\text{rechtes Distributivgesetz}) \end{aligned}$$

Hier ist r eine reelle Zahl und $a, \tilde{a}, b, \tilde{b}$ sind Zeilen- oder Spaltenvektoren mit gleicher Zahl von Einträgen, wobei a und \tilde{a} auch gleiches Format haben (d.h. beide sind Zeilenvektoren oder beide Spaltenvektoren; sonst wäre ja $a \pm \tilde{a}$ nicht definiert) und ebenso auch b und \tilde{b} . Die Tatsache, dass man Summen und reelle Zahlen aus jedem Faktor des Skalarprodukts "herausziehen" kann (die Homogenität und die Distributivität), wird kurz als *Bilinearität* des Skalarprodukts bezeichnet, und man sagt, das Skalarprodukt sei eine *symmetrische Bilinearform*, um diese Bilinearität und das Symmetriengesetz hervorzuheben. Wie beim Rechnen mit Zahlen folgen weitere Rechenregeln, wie z.B. die *binomischen Formeln für Skalarprodukte* von Zeilen- oder Spaltenvektoren a, b desselben Formats

$$(a \pm b) \cdot (a \pm b) = a \cdot a \pm 2a \cdot b + b \cdot b, \quad (a + b) \cdot (a - b) = a \cdot a - b \cdot b,$$

wobei es hier nicht üblich ist, etwa a^2 für $a \cdot a$ zu schreiben, und wie das *Herausziehen beliebiger Linearkombinationen*

$$\begin{aligned} (r_1 \mathbf{a}_1 + \dots + r_k \mathbf{a}_k) \cdot \mathbf{b} &= r_1 (\mathbf{a}_1 \cdot \mathbf{b}) + \dots + r_k (\mathbf{a}_k \cdot \mathbf{b}), \\ (r_1 \mathbf{a}_1 + \dots + r_k \mathbf{a}_k) \cdot (s_1 \mathbf{b}_1 + \dots + s_l \mathbf{b}_l) &= \sum_{i=1}^k \sum_{j=1}^l r_i s_j (\mathbf{a}_i \cdot \mathbf{b}_j), \end{aligned}$$

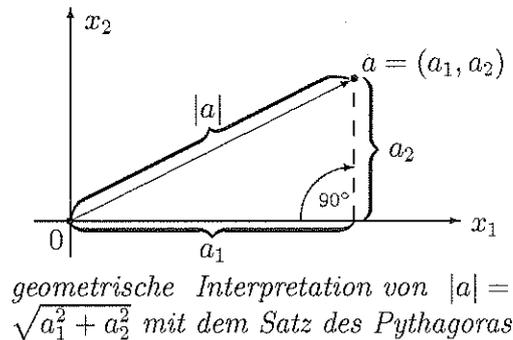
wobei die r_i und s_j reelle Zahlen sind, alle Vektoren \mathbf{a}_i dasselbe Format haben, alle \mathbf{b}_j ebenfalls dasselbe Format besitzen und alle \mathbf{a}_i und \mathbf{b}_j dieselbe Anzahl von Komponenten haben. (Die Klammern rechts hätte man auch weglassen können, weil die Ausdrücke wegen der Homogenität des Skalarproduktes nicht von der Art der Klammerung abhängen.)

4) Das Skalarprodukt $a \cdot a = a_1^2 + a_2^2 + \dots + a_n^2$ eines Vektors a mit Komponenten a_1, a_2, \dots, a_n ist nichtnegativ und nur dann gleich Null, wenn alle Komponenten Null sind. Die Bildung der (nichtnegativen) Wurzel aus $a \cdot a$ hat viele Eigenschaften mit dem Absolutbetrag reeller Zahlen gemeinsam und wird deshalb auch genau so notiert:

$$|a| := \sqrt{a \cdot a} = \sqrt{a_1^2 + a_2^2 + \dots + a_n^2},$$

genannt die (**Euklidische**) **Norm** oder die **Länge** des Vektors a . (Falls $n = 1$ ist, also a eine reelle Zahl, so ist das der gewöhnliche Absolutbetrag von a . Wenn man die Norm in der Notation vom Absolutbetrag unterscheiden möchte, so kann man $\|a\|$ dafür schreiben. Nötig ist das aber nicht, weil sich aus dem Kontext ergibt, ob a eine reelle Zahl ist oder

ein Vektor mit $n \geq 2$ Komponenten, und im letzteren Fall kann $|a|$ ja nur die Euklidische Norm bedeuten.) Mit dem elementargeometrischen Satz des Pythagoras kann man sehen, dass $|a|$ für $a = (a_1, a_2) \in \mathbb{R}^2$ tatsächlich die Länge der Strecke vom Ursprung $0 = (0, 0)$ zum Punkt a ist. Oft stellt man Vektoren als Pfeile dar, hier z.B. den Vektor a als Pfeil mit Ende in 0 und Spitze in a ; dann ist $|a|$ die elementargeometrische Länge dieses Pfeils. (Von dieser Darstellungsweise kommt auch der Name "Vektor" her, der eigentlich "gerichtete Größe" heißt.)



Es gibt andere Möglichkeiten, die "Größe" von Vektoren a mit Komponenten a_1, a_2, \dots, a_n zu messen, z.B. sind für diesen Zweck auch die Summe der Komponentenbeträge $|a|_1 := |a_1| + |a_2| + \dots + |a_n|$ oder ihr Maximum $|a|_\infty := \max(|a_1|, |a_2|, \dots, |a_n|)$ gebräuchlich, oder allgemeiner die sog. p -Norm $|a|_p := (|a_1|^p + |a_2|^p + \dots + |a_n|^p)^{1/p}$ zu einem Exponenten $p \in [1, \infty[$, die für $p = 2$ gerade die Euklidische Norm ist. Die Euklidische Norm $|a|$ ist ausgezeichnet durch ihre geometrische Bedeutung und ihren Zusammenhang mit dem Skalarprodukt. So kann man nicht nur $|a| = \sqrt{a \cdot a}$ durch das Skalarprodukt von a mit sich selbst ausdrücken, sondern auch umgekehrt Skalarprodukte von Vektoren $a, b \in \mathbb{R}^n$ durch Euklidische Normen

$$a \cdot a = |a|^2, \quad a \cdot b = \frac{1}{2}|a + b|^2 - \frac{1}{2}|a|^2 - \frac{1}{2}|b|^2.$$

6) Sind a, b zwei vom Nullvektor verschiedene Vektoren in \mathbb{R}^n , so gibt die binomische Formel für Skalarprodukte folgenden Ausdruck für das Quadrat des Abstands von a zu einem Punkt tb auf der Geraden durch den Ursprung 0 und durch b :

$$|a - tb|^2 = |a|^2 - 2ta \cdot b + t^2|b|^2.$$

Diese quadratische Funktion von $t \in \mathbb{R}$ hat ihr Minimum bei $t = 0$, genau wenn $a \cdot b = 0$ ist. Genau dann ist also der Nullpunkt der zu a nächste Punkt auf der Geraden durch 0 und b . Aus der elementaren Geometrie in der Zeichenebene \mathbb{R}^2 bzw. im Anschauungsraum \mathbb{R}^3 weiß man aber, dass die Verbindungsstrecke von einem Punkt a zum nächsten Punkt auf einer Geraden das Lot von a auf diese Gerade ist, d.h. auf der Geraden senkrecht steht. Also gilt dort $a \cdot b = 0$, genau wenn die Pfeile, welche die Vektoren a und b darstellen, senkrecht zueinander sind. Daher ist es sinnvoll, allgemein zwei **Vektoren orthogonal zueinander** oder **senkrecht aufeinander** zu nennen, wenn ihr Skalarprodukt definiert ist und den Wert Null hat. Geometrisch stellt man sich einen rechten Winkel zwischen den die Vektoren darstellenden Pfeilen vor.

7) Eine zum Skalarprodukt gehörende *Division von Vektoren ist nicht definiert*. Der Quotient von b und a müsste ja, wenn er sinnvoll zu definieren wäre, die eindeutige Lösung x der Gleichung $a \cdot x = b$ sein. Dann müsste also x ein Vektor mit derselben Zahl n von Einträgen wie a sein und b ein Skalar. Aber es gibt (außer in dem hier nicht interessierenden Fall $n = 1$, wo a, b, x reelle Zahlen sind und $a \neq 0$) immer unendlich viele Lösungen x dieser Gleichung, wenn a mindestens eine von Null verschiedene Komponente a_j hat; denn es handelt sich ja um eine lineare Gleichung $a_1x_1 + a_2x_2 + \dots + a_nx_n = b$ für die n Komponenten von x . Und deshalb ist keine sinnvolle Definition der Division des Skalars b durch den Vektor a möglich, auch nicht wenn a kein Nullvektor ist. ■

BEISPIELE (zum Skalarprodukt von Vektoren):

1)

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ -2 \\ 1 \end{pmatrix} = 1 \cdot 3 + 2 \cdot (-2) + 3 \cdot 1 = 2,$$

$$(2 \ 0 \ -1 \ 1) \cdot \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \end{pmatrix} = 2 \cdot 0 + 0 \cdot 1 + (-1) \cdot 2 + 1 \cdot 3 = 1,$$

$$(1, 1, 1) \cdot (1, -2, 1) = 1 \cdot 1 + 1 \cdot (-2) + 1 \cdot 1 = 0.$$

Natürlich ist ein Skalarprodukt Null, wenn einer der Faktoren ein Nullvektor ist, also nur Einträge Null hat. Das letzte Beispiel zeigt, dass die Umkehrung nicht gilt. Geometrisch bedeutet $a \cdot b = 0$ für $a, b \in \mathbb{R}^n$ ja nur, dass die Vektoren a und b , als Pfeile vom Ursprung aus gedacht, orthogonal sind, also senkrecht zueinander; und das bedeutet (in \mathbb{R}^n mit $n \geq 2$) natürlich nicht, dass a oder b der Nullvektor sein muss.

2) Das Skalarprodukt von einem Vektor a mit m Komponenten und einem Vektor, der m gleiche Einträge 1 hat, ist gerade die Summe der Einträge von a :

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = a_1 + a_2 + \dots + a_m.$$

Daher heißen Vektoren wie $(1, 1, \dots, 1)$ auch *summierende Vektoren*. Entsprechend ist das Skalarprodukt von a mit $\frac{1}{m}(1, 1, \dots, 1) \in \mathbb{R}^m$ das arithmetische Mittel der Einträge von a . Sind allgemeiner Gewichte $s_1, s_2, \dots, s_m \geq 0$ mit $s_1 + s_2 + \dots + s_m = 1$ gegeben, so ist das Skalarprodukt von a mit dem *Gewichtsvektor* (s_1, s_2, \dots, s_m) das gewichtete arithmetische Mittel der m Einträge von a mit den gegebenen Gewichten:

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} \cdot \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_m \end{pmatrix} = s_1 a_1 + s_2 a_2 + \dots + s_m a_m.$$

3) Das *Skalarprodukt eines Vektors x mit einem kanonischen Basisvektor* ist eine Komponente von x . Genauer ist $x \cdot e_i = x_i$, wenn x die Komponenten x_1, \dots, x_m hat und e_i wie üblich den kanonischen Basisvektor mit Eintrag 1 in i -ter Position und mit weiteren $m-1$ Einträgen 0 bezeichnet, also

$$x \cdot e_i = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \cdot \overbrace{(0, \dots, 0, 1, 0, \dots, 0)}^{m} = x_i.$$

i -te Position

Insbesondere haben wir für *das Skalarprodukt von zwei kanonischen Basisvektoren e_i und e_j in \mathbb{R}^m* :

$$e_i \cdot e_j = \delta_{ij} := \begin{cases} 1, & \text{wenn } i = j, \\ 0, & \text{wenn } i \neq j. \end{cases}$$

Dabei ist das sog. *Kronecker-Delta-Symbol* δ_{ij} eine in der Mathematik übliche Abkürzung für eine Größe, die den Wert 1 hat, wenn beide Indizes gleich sind, und den Wert 0 sonst.

4) Eine *ökonomische Anwendung* (formaler Art): Für m gefertigte Produkte können wir die in einem gewissen Zeitraum abgesetzten Mengen x_1, x_2, \dots, x_m (in Mengeneinheiten) zu einem **Absatzvektor** \mathbf{x} zusammenfassen und die beim Verkauf dieser Produkte erzielten Preise p_1, p_2, \dots, p_m zu einem **Preisvektor** \mathbf{p} . Dann gibt das Skalarprodukt den

$$\text{Umsatz / Erlös} \quad \mathbf{x} \cdot \mathbf{p} = x_1 p_1 + x_2 p_2 + \dots + x_m p_m$$

an, der mit den m Produkten im betrachteten Zeitraum insgesamt erzielt wurde. Entsprechend erhält man aus einem **Produktionsvektor** \mathbf{x} , dessen Einträge x_i die Outputs (in Mengeneinheiten) von m Produkten angeben, und dem zugehörigen **Stückkostenvektor** \mathbf{k} , dessen Einträge k_i die bei der Produktion pro Einheit des i ten Produkts entstandenen Kosten sind, durch Skalarproduktbildung die

$$\text{Gesamtkosten} \quad \mathbf{x} \cdot \mathbf{k} = x_1 k_1 + x_2 k_2 + \dots + x_m k_m.$$

Immer wenn, wie hier, die gleich positionierten Einträge zweier gleich langen Tabellenspalten miteinander multipliziert und die entstehenden Produkte aufsummiert werden, hat man in der Ökonomie eine Skalarproduktbildung.

5) *Anwendung auf eine lineare Gleichung*: Eine lineare Gleichung

$$a_1 x_1 + a_2 x_2 + \dots + a_n x_n = b$$

kann unter Verwendung des Skalarprodukts kurz

$$\mathbf{a} \cdot \mathbf{x} = b$$

geschrieben werden, wo $\mathbf{a} = (a_1, a_2, \dots, a_n)$ der *Koeffizientenvektor* der Gleichung ist und $\mathbf{x} = (x_1, x_2, \dots, x_n)$ der *Vektor der Unbekannten*. Die rechte Seite ist hier natürlich ein Skalar $b \in \mathbb{R}$. Die Lösungsmenge besteht also aus allen Vektoren $\mathbf{x} \in \mathbb{R}^n$, die mit dem Koeffizientenvektor \mathbf{a} ein Skalarprodukt mit dem gegebenen Wert b bilden. Im homogenen Fall $b = 0$ läßt sich die Lösungsmenge geometrisch als die Menge aller zu \mathbf{a} senkrechten Vektoren in \mathbb{R}^n beschreiben, genannt der zu \mathbf{a} *orthogonale Unterraum von \mathbb{R}^n* und notiert $\{\mathbf{a}\}^\perp$. Aus den Rechenregeln für das Skalarprodukt folgt für jede "spezielle" Lösung \mathbf{x}^* :

$$\mathbf{a} \cdot \mathbf{x} = b \iff \mathbf{a} \cdot \mathbf{x} = \mathbf{a} \cdot \mathbf{x}^* \iff \mathbf{a} \cdot (\mathbf{x} - \mathbf{x}^*) = 0,$$

d.h. die Lösungsmenge L zu $\mathbf{a} \cdot \mathbf{x} = b$ erhält man durch die Parallelverschiebung von $\{\mathbf{a}\}^\perp$, die den Nullvektor in \mathbf{x}^* überführt, $L = \mathbf{x}^* + \{\mathbf{a}\}^\perp$. In \mathbb{R}^2 ist das die zur Richtung von $\mathbf{a} \neq 0$ senkrechte Gerade durch \mathbf{x}^* , in \mathbb{R}^3 die zur Richtung von $\mathbf{a} \neq 0$ senkrechte Ebene durch \mathbf{x}^* .

Aus den Rechenregeln für das Skalarprodukt sieht man auch unmittelbar:

- *Linearkombinationen von Lösungsvektoren $\mathbf{x}_k \in \mathbb{R}^n$ der homogenen Gleichung sind wieder Lösungen der homogenen Gleichung;*
- *affine Kombinationen von Lösungsvektoren $\mathbf{x}_k \in \mathbb{R}^n$ der inhomogenen Gleichung sind wieder Lösungen der inhomogenen Gleichung.*

Beides folgt aus

$$\mathbf{a} \cdot \sum_{k=1}^l t_k \mathbf{x}_k = \sum_{k=1}^l t_k (\mathbf{a} \cdot \mathbf{x}_k),$$

weil im ersten Fall $\mathbf{a} \cdot \mathbf{x}_k = 0$ ist für alle k , also auch die rechte Seite der letzten Gleichung $= 0$, und im zweiten Fall $\mathbf{a} \cdot \mathbf{x}_k = b$ für alle k sowie $\sum_{k=1}^l t_k = 1$, also die rechte Seite der letzten Gleichung wieder $= b$. Die Aussagen hier gelten übrigens nicht nur für lineare Gleichungen, sondern auch analog für lineare Gleichungssysteme und können (mühsam) durch direkte Rechnungen mit den Gleichungssystemen oder (einfacher) mit den folgenden Rechenregeln für die Multiplikation von Matrizen an Spaltenvektoren gezeigt werden. ■

Mathematik für Wirtschaftswissenschaftler

(K. Steffen, Heinrich-Heine-Universität Düsseldorf, WS 2006/07)

Um zu motivieren, wie man das *Produkt einer Matrix mit einem Spaltenvektor* sinnvoll definiert, betrachten wir folgende ökonomische Situation: In einem Betrieb werden die Produkte $\boxed{1}, \dots, \boxed{n}$ gefertigt, wobei die Rohstoffe $\textcircled{1}, \dots, \textcircled{m}$ Verwendung finden. Die **Materialverflechtungsmatrix** $R = (r_{ij})$

	$\boxed{1}$	$\boxed{2}$...	\boxed{n}	← Produkte
$\textcircled{1}$	r_{11}	r_{12}	...	r_{1n}	
$\textcircled{2}$	r_{21}	r_{22}	...	r_{2n}	
⋮	⋮	⋮		⋮	
\textcircled{m}	r_{m1}	r_{m2}	...	r_{mn}	
Rohstoffe					↑

hat in der i -ten Zeile und j -ten Spalte den **Rohstoffverbrauchscoeffizienten** r_{ij} als Eintrag, der angibt, wieviele Mengeneinheiten des Rohstoffs \textcircled{i} für die Herstellung einer Einheit des Produkts \boxed{j} benötigt werden. Die Frage ist nun, wie man zu einem gegebenen **Produktionsoutputvektor** $\mathbf{x} = (x_1, x_2, \dots, x_n)$ den zugehörigen **Rohstoffbedarfsvektor** $\mathbf{z} = (z_1, z_2, \dots, z_m)$ berechnet, dessen Einträge z_i angeben, wieviele Einheiten vom Rohstoff \textcircled{i} gebraucht werden, wenn der gegebene Produktionsplan von x_1 Outputeinheiten des Produkts $\boxed{1}$, x_2 Outputeinheiten des Produkts $\boxed{2}$, ..., x_n Outputeinheiten des Produkts \boxed{n} realisiert werden soll. Da man für die Herstellung von x_j Outputeinheiten des Produkts \boxed{j} gerade r_{ij} Einheiten des Rohstoffs \textcircled{i} braucht, ist die Antwort

$$z_i = r_{i1}x_1 + r_{i2}x_2 + \dots + r_{in}x_n \quad \text{für } i = 1 \dots m,$$

die i -te Komponente von \mathbf{z} ist also das Skalarprodukt der i -ten Zeile der Materialverflechtungsmatrix R mit dem Produktionsvektor \mathbf{x} . Dies kann man durch eine einzige Vektorgleichung

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

ausdrücken, wenn man das Produkt von der Matrix R (mit n Spalten) mit dem Spaltenvektor \mathbf{x} (mit n Einträgen) gerade so definiert, wie wir es jetzt tun:

DEFINITION: Das **Produkt einer Matrix mit einem Spaltenvektor** ist definiert, wenn die Spaltenzahl n der Matrix A gleich der Anzahl der Komponenten des Spaltenvektors x ist; ist m die Zeilenzahl von A , so ist das Ergebnis Ax der Multiplikation der m -gliedrige Spaltenvektor, der als i -ten Eintrag das Skalarprodukt der i -ten Zeile von A mit der Spalte x hat für $1 \leq i \leq m$. ■

DISKUSSION: 1) Ausgeschrieben lautet die Definition:

$$Ax = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix} \in \mathbb{R}^m.$$

Dabei steht rechts ein Spaltenvektor mit ebenso vielen Einträgen, wie die Matrix links Zeilen besitzt, und die Spaltenzahl der Matrix ist gleich der Komponentenzahl des Spaltenvektors, an den sie multipliziert wird. Führt man die Abkürzung $\underline{a}_i := (a_{i1} \ a_{i2} \ \dots \ a_{in})$ für die i -te Zeile der Matrix A ein, so kann man entsprechend der Definition schreiben

$$Ax = \begin{pmatrix} \underline{a}_1 \\ \underline{a}_2 \\ \vdots \\ \underline{a}_m \end{pmatrix} x = \begin{pmatrix} \underline{a}_1 \cdot x \\ \underline{a}_2 \cdot x \\ \vdots \\ \underline{a}_m \cdot x \end{pmatrix}.$$

2) Eine andere Möglichkeit der Beschreibung des Produkts Ax ist folgende:

$$Ax = x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \dots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} = x_1 \bar{a}_1 + x_2 \bar{a}_2 + \dots + x_n \bar{a}_n,$$

wo \bar{a}_j die j -te Spalte der Matrix A bezeichnet. Es gilt also:

- Das Produkt Ax ist die Linearkombination der Spaltenvektoren $\bar{a}_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}$ von A mit den Komponenten x_j der Spalte x als Koeffizienten.

Insbesondere hat der Ergebnisvektor Ax dasselbe Format wie die Spalten von A , und x muss ebenso viele Komponenten haben wie die Matrix A Spalten, damit das Produkt Ax definiert ist.

3) Wir betonen nochmals ausdrücklich, dass man nicht beliebige Matrizen an beliebige Vektoren multiplizieren kann, sondern dass die Formate zusammen "passen" müssen:

- Das Produkt Ax ist nur definiert, wenn die Spaltenzahl der Matrix A gleich der Komponentenzahl des Spaltenvektors x ist;
- das Ergebnis der Multiplikation Ax ist dann ein Spaltenvektor mit so vielen Einträgen, wie A Zeilen hat;
- deshalb sind x und Ax im Allgemeinen Spalten mit unterschiedlich vielen Einträgen; nur wenn A eine quadratische Matrix ist, haben x und Ax dasselbe Format.

4) Die grundlegenden **Rechenregeln** für die Multiplikation von Matrizen an Spaltenvektoren sind folgende:

$$\begin{aligned} 0x &= 0, & A0 &= 0 \\ A(rx) &= r(Ax) = (rA)x & & \text{(Homogenität)} \\ (A+B)x &= Ax + Bx & & \text{(linkes Distributivgesetz)} \\ A(x + \tilde{x}) &= Ax + A\tilde{x} & & \text{(rechtes Distributivgesetz)} \end{aligned}$$

Dabei sind A, B und die Nullmatrix 0 Matrizen vom gleichen Format (m, n) , die Vektoren x, \tilde{x} und der Nullvektor im Produkt $A0$ sind n -gliedrige Spaltenvektoren, $r \in \mathbb{R}$ ist ein Skalar und die Gleichungen sind als Gleichheit von Spaltenvektoren mit m Komponenten zu verstehen. Die beiden Distributivgesetze und die Homogenität fasst man zusammen, indem man sagt, dass das Produkt von $m \times n$ -Matrizen mit n -gliedrigen Spaltenvektoren eine *bilineare Rechenoperation* ist. Die Bilinearität, also das erlaubte "Herausziehen" von Summenbildung und Multiplikation mit Skalaren aus den Faktoren, ist überhaupt die grundlegende Eigenschaft jeder Rechenoperation, die als "Produktbildung" bezeichnet wird. Natürlich lassen sich dann auch Summen mit mehr als zwei Summanden aus den Faktoren des Produkts herausziehen, und allgemeiner gilt:

- *Man kann bei der Produktbildung aus jedem Faktor beliebige Linearkombinationen "herausziehen",*

was hier für Matrizen $A, A_1, \dots, A_k \in \mathbb{R}^{m \times n}$, Spaltenvektoren $x, x_1, x_2, \dots, x_l \in \mathbb{R}^n$ und Skalare $r_1, r_2, \dots, r_k, s_1, s_2, \dots, s_l \in \mathbb{R}$ genau heißt:

$$\begin{aligned}(r_1 A_1 + r_2 A_2 + \dots + r_k A_k)x &= r_1(A_1 x) + r_2(A_2 x) + \dots + r_k(A_k x), \\ A(s_1 x_1 + s_2 x_2 + \dots + s_l x_l) &= s_1(Ax_1) + s_2(Ax_2) + \dots + s_l(Ax_l).\end{aligned}$$

Wenn beide Faktoren Linearkombinationen sind, so muss man diese nacheinander erst aus dem einen dann aus dem anderen Faktor "herausziehen", also

$$\begin{aligned}(A + B)(x + y) &= A(x + y) + B(x + y) = Ax + Ay + Bx + By, \\ (r_1 A_1 + r_2 A_2 + \dots + r_k A_k)(s_1 x_1 + s_2 x_2 + \dots + s_l x_l) &= \sum_{i=1}^k \sum_{j=1}^l r_i s_j (A_i x_j),\end{aligned}$$

wenn auch noch $B \in \mathbb{R}^{m \times n}$ und $y \in \mathbb{R}^n$ sind. ■

BEISPIELE (zum Produkt von Matrizen mit Spaltenvektoren):

$$1) \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_2 \\ x_1 \end{pmatrix},$$

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

$$\begin{pmatrix} 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Man beachte die Formate der Matrizen und Vektoren. Es darf einfach nicht passieren, dass man Matrizen an nicht passende Spaltenvektoren multipliziert oder Ergebnisvektoren mit falscher Komponentenzahl erhält! (Es kann auch nicht passieren, wenn man sich an die angegebene und einzig sinnvolle Definition des Produkts von Matrizen mit Spaltenvektoren hält und sich nicht seine eigene private Definition selbst "strickt".)

2) Im Fall einer einzeiligen Matrix $A = a = (a_1 \ a_2 \ \dots \ a_n)$ reduziert sich die Definition des Produkts Ax mit $x \in \mathbb{R}^n$ auf die Bildung des Skalarprodukts $a \cdot x$ von a mit x :

$$ax = (a_1 \ a_2 \ \dots \ a_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = (a_1x_1 + a_2x_2 + \dots + a_nx_n) = (a \cdot x).$$

Genau genommen steht rechts nicht das Skalarprodukt $a \cdot x$, sondern der 1-gliedrige Spaltenvektor bzw. die 1×1 -Matrix mit diesem Skalarprodukt als einzigem Eintrag. Diesen Unterschied, der rein formal ist, kann man ignorieren und daher die Matrixklammern rechts weglassen. (Beachte aber, dass für Skalarprodukte $a \cdot x = x \cdot a$ gilt, während das Matrixprodukt xa der Spalte x mit der Zeile a , das später eingeführt wird, nicht gleich der 1×1 -Matrix ax ist, sondern eine $n \times n$ -Matrix.)

3) Für das Produkt der $n \times n$ -**Einheitsmatrix**

$$\mathbb{I} = \mathbb{I}_n := (\delta_{ij})_{1 \leq i \leq n, 1 \leq j \leq n} = \left(\underbrace{\begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}}_n \right) \Bigg\}^n$$

(wobei das schon erwähnte Symbol δ_{ij} den Wert 1 hat, wenn $i = j$ ist und den Wert 0 sonst, und das Subskript "n" bei \mathbb{I} das Format der Einheitsmatrix anzeigt, wenn nötig) mit einem n -gliedrigen Spaltenvektor $x \in \mathbb{R}^n$ ist

$$\mathbb{I}x = x;$$

denn die i -te Zeile von \mathbb{I} ist der i -te kanonische n -gliedrige Basisvektor und dessen Skalarprodukt mit $x \in \mathbb{R}^n$ ist gerade die i -te Komponente von x .

- *Multiplikation der Einheitsmatrix an einen (passenden!) Spaltenvektor reproduziert diesen Vektor.*

4) Für eine **Diagonalmatrix**, d.h. eine quadratische Matrix der Form

$$D = (d_i \delta_{ij})_{1 \leq i \leq n, 1 \leq j \leq n} = \left(\underbrace{\begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}}_n \right) \Bigg\}^n$$

mit Einträgen d_i auf der Diagonalen und Nulleinträgen sonst, gilt

$$Dx = \left(\underbrace{\begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}}_n \right) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} d_1x_1 \\ d_2x_2 \\ \vdots \\ d_nx_n \end{pmatrix},$$

d.h. die Multiplikation von D an $x \in \mathbb{R}^n$ läuft auf Dasselbe hinaus, wie jede Komponente von x mit dem entsprechenden Diagonaleintrag zu multiplizieren.

Sind alle Diagonaleinträge $d_i = d$ gleich, so ist D das d -fache der Einheitsmatrix, und $Dx = (d\mathbb{I})x = d(\mathbb{I}x) = dx$ das d -fache des Spaltenvektors x . Die Multiplikation der Matrix $d\mathbb{I}$ an einen (passenden) Spaltenvektor x läuft also auf Dasselbe hinaus wie die Multiplikation von x mit dem Skalar d ; deshalb nennt man eine Matrix der Form $d\mathbb{I}$, also eine Diagonalmatrix mit lauter gleichen Diagonaleinträgen d , auch eine **skalare Matrix**.

5) Setzen wir für x den j -ten kanonischen Basisvektor $e_j \in \mathbb{R}^n$ ein, so ergibt sich

$$Ae_j = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}.$$

\uparrow j -te Position (von n)

- Das Produkt einer $m \times n$ -Matrix A mit dem j -ten kanonischen Basisvektor von \mathbb{R}^n ist gerade die j -te Spalte von A .

6) Eine ökonomische Anwendung findet die Multiplikation von Matrizen an Spaltenvektoren in der **Input-Output-Analyse**. Wir betrachten dazu wie vor der letzten Definition die Herstellung von Produkten $\boxed{1}, \dots, \boxed{n}$ aus Rohstoffen / Vorprodukten $\textcircled{1}, \dots, \textcircled{m}$, wobei die in der $m \times n$ -**Materialverflechtungsmatrix** $R = (r_{ij})$ zusammengefassten sog. **Rohstoffverbrauchskoeffizienten** r_{ij} die Anzahl der Einheiten von \textcircled{i} angeben, die für die Herstellung einer Einheit von \boxed{j} benötigt werden. Ist $x = (x_1, x_2, \dots, x_n)$ der **Produktionsvektor**, dessen Einträge die zu produzierenden Mengen der n Produkte angeben (in Mengeneinheiten), so ist der zugehörige **Rohstoffbedarfsvektor** $z = (z_1, z_2, \dots, z_m)$ gegeben durch das Produkt $z = Rx$, weil die Anzahl der für den Produktionsplan benötigten Einheiten des Rohstoffs / Vorprodukts \textcircled{i} ja gleich $z_i = r_{i1}x_1 + r_{i2}x_2 + \dots + r_{in}x_n$ ist.

Wir nehmen nun an, dass die hergestellten Produkte zum Teil am Markt verkauft werden ("Output"), was durch einen **Nachfragevektor** oder **Absatzvektor** $y = (y_1, y_2, \dots, y_n)$ beschrieben wird, und zum anderen Teil in die Produktion der anderen Produkte oder auch desselben Produkts als Vorprodukte eingehen ("Input"). Dieser Sachverhalt wird durch eine **Produktionsverflechtungsmatrix** $A = (a_{jk})$ vom Format (n, n) beschrieben, wobei der Eintrag a_{jk} angibt, wieviele Einheiten des Produkts \boxed{j} für die Herstellung einer Einheit des Produkts \boxed{k} gebraucht werden. Die a_{jk} heißen **Produktionskoeffizienten**, a_{jj} sind die **Eigenverbrauchskoeffizienten**. Die Gesamtproduktion des Produkts \boxed{j} setzt sich dann zusammen aus der am Markt abgesetzten Menge y_j (**Output**) und der für die Herstellung der anderen Produkte bzw. desselben Produkts verbrauchten Menge $a_{j1}x_1 + a_{j2}x_2 + \dots + a_{jn}x_n$ (**endogener Input**), also hat man die Bilanz

$$x_j = y_j + a_{j1}x_1 + a_{j2}x_2 + \dots + a_{jn}x_n \quad \text{für } j = 1 \dots n,$$

die man in Form einer Vektorgleichung $x = y + Ax$ schreiben kann. Mit der $n \times n$ -Einheitsmatrix \mathbb{I} und den Rechenregeln kann man die Vektorgleichung $y = x - Ax = \mathbb{I}x - Ax = (\mathbb{I} - A)x$ umformen, also gilt für Produktions- und Absatzvektor die Gleichung:

$$(\mathbb{I} - A)x = y.$$

Die hier auftretende Matrix $\mathbb{I} - A$ heißt auch **Technologie-Matrix**. Die Vektorgleichung $(\mathbb{I} - A)x = y$ selbst wird als **statisches Input-Output-Modell** oder **Leontief-Modell** bezeichnet ("statisch", weil hier die Koeffizienten r_{ij} und a_{jk} als zeitlich konstant angenommen sind). Das Modell lässt sich auch auf ein in Teilbetriebe / Abteilungen gegliedertes Unternehmen anwenden und auf eine in Sektoren gegliederte Volkswirtschaft.

Für einen gegebenen Produktionsvektor x kann man nun den Rohstoffverbrauch und die für den Markt verfügbaren Mengen durch einfache Matrixmultiplikationen an x berechnen gemäß

$$z = Rx \in \mathbb{R}^m, \quad y = (I - A)x \in \mathbb{R}^n.$$

Dazu wäre natürlich Matrixrechnung nicht unbedingt erforderlich gewesen. Ihr Vorteil ist die übersichtliche "indexfreie" Schreibweise, in der auch die lineare Natur der Abhängigkeit von z und y vom Produktionsvektor x zum Ausdruck kommt (Multiplikation von x mit skalaren Faktoren und Ersetzen von x durch eine Summe von Produktionsvektoren wirken sich entsprechend auf y und z aus).

Interessanter ist die Frage, ob man einen vorgegebenen Nachfragevektor y realisieren kann. Dies erfordert die Auflösung der Gleichung $(I - A)x = y$ nach $x \in \mathbb{R}^n$, was auf die Lösung eines Systems von n linearen Gleichungen für n Unbekannte mit Koeffizientenmatrix $I - A$ und Inhomogenität y hinausläuft. Die *Struktur der Technologie-Matrix* $I - A$ ist oft folgende: Die Produktionskoeffizienten a_{jk} sind nichtnegativ und klein, so dass in den Diagonalpositionen von $I - A$ Einträge mit einem Wert nahe bei 1 und in den anderen Positionen nichtpositive Einträge von kleinem Betrag stehen. In dieser Situation kann man mathematisch beweisen (wozu aber mehr Theorie gehört; s. 3.4), dass die Auflösung von $(I - A)x = y$ für jedes $y \in \mathbb{R}^n$ eindeutig möglich ist und dass zudem der Lösungsvektor $x \in \mathbb{R}^n$ nichtnegative Komponenten $x_k \geq 0$ hat, wenn alle Komponenten $y_j \geq 0$ sind, was ja allein ökonomisch sinnvoll ist. (Man kann schließlich eine Produktion nicht ohne Weiteres umkehren und durch Produktverbrauch wieder Rohstoffe zurückgewinnen.) Sind allerdings Produktionskoeffizienten zu groß, so lässt sich unter Umständen nicht mehr jeder Nachfragevektor y (mit $y_j \geq 0$ für alle j) durch einen Produktionsvektor x mit $x_k \geq 0$ für alle k realisieren, sondern die Lösung x hat — wenn überhaupt eine existiert — einen oder mehrere negative Einträge. Dies bedeutet, dass von dem entsprechenden Produkt für den endogenen Input mehr verbraucht wird, als überhaupt produziert werden kann. Das ist natürlich ökonomisch unsinnig, und die Konsequenz ist, dass in solchen Fällen eben nicht jede beliebige Nachfrage nach den Produkten realisiert werden kann.

In der Praxis kommen meist noch *Restriktionen für die Verfügbarkeit von Rohstoffen* hinzu, was durch *lineare Ungleichungen*

$$z_i \leq c_i \quad \text{für gewisse } i \in \{1, 2, \dots, m\}$$

mit gegebenen Schranken $c_i \in \mathbb{R}_{>0}$ auszudrücken ist. Im Sinne der ökonomischen Fragestellung **zulässige Produktionsvektoren** sind dann nur solche $x \in \mathbb{R}^n$, für die erstens alle Komponenten nichtnegativ sind und zweitens die Komponenten z_i von $z = Rx$ den Ungleichungen $z_i \leq c_i$ genügen, also

$$x_k \geq 0 \quad \text{für } k = 1 \dots n, \quad r_{i1}x_1 + r_{i2}x_2 + \dots + r_{in}x_n \leq c_i \quad \text{für gewisse } i \in \{1, 2, \dots, m\}.$$

Klar ist dann, dass Produktionsvektoren x mit hinreichend kleinen positiven Komponenten zulässig sind. Ob es aber zu einem gegebenen Nachfragevektor y (mit nichtnegativen Komponenten) einen *zulässigen* Produktionsvektor x gibt, der $(I - A)x = y$ löst, und welche Nachfragevektoren in diesem Sinne realisierbar sind bzw. welche \tilde{y} möglichst nahe bei y realisiert werden können, wenn das für y selbst nicht möglich ist, — das sind durchaus schwierige und komplexe Fragen, auch in der mathematischen Theorie, mit denen sich die sog. **Input-Output-Analyse** befasst.

7) Noch eine ökonomische Anwendung der Multiplikation von Matrizen an Spaltenvektoren: Ein Unternehmen stelle die Produkte $\boxed{1}, \dots, \boxed{n}$ her an verschiedenen Produktionsstätten $\textcircled{1}, \dots, \textcircled{m}$. Die Verteilung der Produktionsoutputs wird beschrieben durch die $m \times n$ -**Produktionsmatrix** $X = (x_{ij})$, deren Einträge x_{ij} angeben, wieviele Einheiten des Produkts \boxed{j} an der Stätte \textcircled{i} erzeugt werden (in einem betrachteten Zeitintervall, z.B. in einem Geschäftsjahr). Ist p_j der für den Verkauf einer Einheit von \boxed{j} am Markt erzielte Preis, so bilden wir den **Preisvektor** $p = (p_1, p_2, \dots, p_n)$ und können damit den **Umsatzvektor** (auch **Erlösvektor** genannt) $u = (u_1, u_2, \dots, u_m)$ als Produkt der Produktionsmatrix mit dem Preisvektor darstellen gemäß

$$u = Xp, \quad \text{also} \quad \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{pmatrix}.$$

Der i -te Eintrag von u ist der bei der Produktionsstätte \textcircled{i} erzielte Umsatz $u_i = \sum_{j=1}^n x_{ij}p_j = x_{i1}p_1 + x_{i2}p_2 + \dots + x_{in}p_n$.

Um die Kostenstruktur zu beschreiben, führt man die $m \times n$ -**Kostenproduktionsmatrix** $K = (k_{ij})$ ein, deren Einträge k_{ij} die bei Herstellung des Produkts \boxed{j} an der Produktionsstätte \textcircled{i} entstehenden Stückkosten (Kosten pro hergestellte Mengeneinheit) angeben. Das Produkt Kx der Produktionskostenmatrix K mit einem **Produktionsoutputvektor** $x = (x_1, x_2, \dots, x_n)$ ist dann der Spaltenvektor, dessen i -ter Eintrag die Kosten $k_{i1}x_1 + k_{i2}x_2 + \dots + k_{in}x_n$ für die Realisierung des durch x gegebenen Produktionsplans bei Produktion allein an der Stätte \textcircled{i} angibt. Die eigentliche ökonomische Aufgabe ist aber, die Produktion mit gewünschten Outputs x_1, x_2, \dots, x_n so auf die m Produktionsstätten zu verteilen, dass die Gesamtkosten minimiert werden. Gesucht ist also eine Produktionsmatrix $X = (x_{ij})$, deren Spaltensummen $x_{1j} + x_{2j} + \dots + x_{mj} = x_j$ vorgegeben sind, mit der Eigenschaft, dass die Summe $\sum_{i=1}^m \sum_{j=1}^n k_{ij}x_{ij}$, also die Gesamtkosten für die Produktion an allen Produktionsstätten, minimal wird. Dabei unterliegen die Einträge der zulässigen Produktionsmatrizen noch ökonomischen Restriktionen wie $x_{ij} \geq 0$ und $x_{ij} \leq c_{ij}$ mit gegebenen Kapazitätsgrenzen $c_{ij} \geq 0$ (maximaler möglicher Produktionsoutput für das Produkt \boxed{j} an der Produktionsstätte \textcircled{i} ; kann auch $= 0$ sein, nämlich wenn \boxed{j} an der Stätte \textcircled{i} gar nicht gefertigt wird). Die Bestimmung einer Produktionsmatrix X zu gegebenem Produktionsoutputvektor x , welche diese Restriktionen erfüllt und dabei zu kleinsten Gesamtkosten führt, ist dann eine anspruchsvolle mathematische Aufgabe, die in die sog. **Lineare Optimierung** gehört. Mehr dazu würde zu weit führen; das Rechnen mit Matrizen und Vektoren ist jedenfalls Voraussetzung für diese Theorie.

9) Ein letztes ökonomisches Beispiel für das Auftreten des Produkts einer Matrix mit einem Spaltenvektor: Wir betrachten eine $n \times n$ -**Übergangsmatrix** $P = (p_{kj})$, die den Käuferwechsel bei n konkurrierende Produkten in einem Beobachtungszeitraum beschreibt, d.h. $p_{kj} \cdot 100\%$ gibt den prozentualen Anteil der Kunden des Produkts \boxed{j} an, die zum Produkt \boxed{k} gewechselt (bzw. im Fall $k = j$ beim Produkt \boxed{j} geblieben) sind. Der **Marktverteilungsvektor** $x = (x_1, x_2, \dots, x_n)$ zu Anfang des Beobachtungszeitraums gibt die Anzahl x_j der Käufer des Produkts \boxed{j} zu diesem Zeitpunkt an für $j = 1 \dots n$. Die Käuferzahl des Produkts \boxed{k} am Ende des Beobachtungszeitraums ist dann $y_k := p_{k1}x_1 + p_{k2}x_2 + \dots + p_{kn}x_n$, d.h. der Marktverteilungsvektor $y = (y_1, y_2, \dots, y_n)$ am Ende ist gleich dem Matrixprodukt $y = Px$. (Vorausgesetzt ist dabei natürlich, dass sich die Gesamtzahl der Käufer im Beobachtungszeitraum nicht geändert hat.) ■

Die Liste der Beispiele für eine prägnante Formulierung ökonomischer Sachverhalte mittels des Produkts von Matrizen und Spaltenvektoren ließe sich noch fortsetzen. Wir verzichten darauf, weil es sich dabei nur um die *Formulierung*, nicht um die *Lösung* ökonomischer Probleme handelt. Für die Lösung muss man letztlich doch meistens lineare Gleichungssysteme auflösen, was wir in 3.2 besprochen haben. Es gibt aber — neben seiner Nützlichkeit für übersichtliche Formulierungen komplexer Zusammenhänge — noch einen tieferen Grund dafür, dass das Produkt von Matrizen und Vektoren in der mathematischen Beschreibung ökonomischer Sachverhalte zwangsläufig auftaucht: Immer wenn eine vektorielle ökonomische Größe $\mathbf{y} = (y_1, \dots, y_m)$ *lineare Abhängigkeit* von einer anderen vektoriellen Größe $\mathbf{x} = (x_1, \dots, x_n)$ zeigt, d.h. bei Vervielfachung von \mathbf{x} mit einem Faktor $r \in \mathbb{R}$ vervielfacht sich die abhängige Größe \mathbf{y} mit demselben Faktor und bei Ersetzung von \mathbf{x} durch eine Summe $\mathbf{x} + \tilde{\mathbf{x}}$ wird \mathbf{y} ersetzt durch die entsprechende Summe $\mathbf{y} + \tilde{\mathbf{y}}$ der \mathbf{x} und $\tilde{\mathbf{x}}$ zugeordneten abhängigen Größen, immer wenn eine derartige einfache Gesetzmäßigkeit besteht, wird der Zusammenhang zwischen der abhängigen und der unabhängigen vektoriellen Größe durch eine Matrix-an-Vektor-Multiplikation $\mathbf{y} = A\mathbf{x}$ beschrieben! Wir erklären das in folgender

DEFINITION und DISKUSSION (*Lineare Operatoren und Matrizen*):

1) Eine Zuordnung $T: V \rightarrow W$, $V \ni x \mapsto T(x) \in W$, zwischen zwei reellen Vektorräumen V und W heißt **linearer Operator** oder **lineare Abbildung** von V nach W , wenn sie mit den Rechenoperationen in folgender Weise verträglich ist:

$$T(rx) = rT(x) \quad \text{und} \quad T(x + \tilde{x}) = T(x) + T(\tilde{x}) \quad \text{in } W$$

für alle $x, \tilde{x} \in V$ und alle $r \in \mathbb{R}$. Man sagt, dass man die Multiplikation mit einem Skalar bzw. die Summenbildung aus der linearen Abbildung "herausziehen" bzw. mit dem linearen Operator vertauschen kann. (Es ist gleich, ob man erst diese Rechenoperationen ausführt und dann die Abbildung T anwendet, oder ob man erst T ausführt und danach die Rechenoperationen — das Ergebnis ist dasselbe.) Da $0x = 0$ der Nullvektor in V ist, gilt dann insbesondere $T(0) = T(0x) = 0T(x) = 0$, also

$$T(0) = 0,$$

wobei natürlich links der Nullvektor von V und rechts der von W gemeint ist. Die linearen Operatoren sind die strukturerhaltenden Abbildungen zwischen Vektorräumen und für die mathematische Lineare Algebra ein genau so fundamentales Konzept wie der Begriff des Vektorraums. In der mathematischen Modellierung ökonomischer Sachverhalte tauchen lineare Operatoren auf, weil sie in gewisser Weise die einfachste Form der Abhängigkeit einer (vektoriellen) Größe $\mathbf{y} = T(\mathbf{x})$ von einer unabhängigen (vektoriellen) Variablen \mathbf{x} darstellen.

2) Ist $T: V \rightarrow W$ linear, so gilt allgemeiner

$$T(rx + \tilde{r}\tilde{x}) = T(rx) + T(\tilde{r}\tilde{x}) = rT(x) + \tilde{r}T(\tilde{x})$$

für $x, \tilde{x} \in V$ und $r, \tilde{r} \in \mathbb{R}$ und entsprechend für Linearkombinationen von mehr als zwei Vektoren $\mathbf{x}_1, \dots, \mathbf{x}_k \in V$ mit Koeffizienten $r_1, \dots, r_k \in \mathbb{R}$:

$$T(r_1\mathbf{x}_1 + r_2\mathbf{x}_2 + \dots + r_k\mathbf{x}_k) = r_1T(\mathbf{x}_1) + r_2T(\mathbf{x}_2) + \dots + r_kT(\mathbf{x}_k).$$

- *Lineare Operatoren sind solche Abbildungen, aus denen man beliebige Linearkombinationsbildungen "herausziehen" kann.*

3) Aus den Rechengesetzen für die Multiplikation von $m \times n$ -Matrizen A an n -gliedrige Spaltenvektoren folgt, dass der Multiplikationsoperator $\mathbb{R}^n \ni x \mapsto Ax \in \mathbb{R}^m$ ein linearer Operator ist; denn es gilt ja $A(rx) = r(Ax)$ und $A(x+\tilde{x}) = Ax + A\tilde{x}$ für $x, \tilde{x} \in \mathbb{R}^n$ und $r \in \mathbb{R}$. Das Bemerkenswerte ist, dass wir auf diese Weise *jeden* linearen Operator T von \mathbb{R}^n nach \mathbb{R}^m erhalten. Dazu erinnern wir uns, dass sich jeder Vektor $x = (x_1, \dots, x_n)$ als Linearkombination $x = x_1e_1 + \dots + x_n e_n$ der kanonischen Basisvektoren e_j von \mathbb{R}^n schreiben lässt mit den Komponenten x_j von x als Koeffizienten. Daher gilt gemäß 2), wenn T linearer Operator von \mathbb{R}^n nach \mathbb{R}^m ist,

$$T(x) = T(x_1e_1 + x_2e_2 + \dots + x_n e_n) = x_1T(e_1) + x_2T(e_2) + \dots + x_nT(e_n).$$

Andererseits haben wir für $m \times n$ -Matrizen A gesehen, dass Ax gerade die Linearkombination der Spaltenvektoren von A ist mit den Komponenten von x als Koeffizienten. Wenn wir also die Bilder der kanonischen Basisvektoren $T(e_j)$ als Spalten der Matrix A nehmen, so haben wir $T(x) = Ax$ für alle $x \in \mathbb{R}^n$. Damit ist ein fundamentales Ergebnis bewiesen:

- Die linearen Operatoren $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ sind genau die Abbildungen von der Form $T(x) = Ax$ mit einer $m \times n$ -Matrix $A = (a_{ij})$.
- Dabei ist die j -te Spalte von A das Bild $T(e_j)$ des j -ten kanonischen Basisvektors in \mathbb{R}^n ; der Matrixeintrag a_{ij} ist also die i -te Komponente von $T(e_j) \in \mathbb{R}^m$

(für $i = 1 \dots m$ und $j = 1 \dots n$). Wegen dieses engen Zusammenhangs zwischen linearen Operatoren $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ und Matrizen $A \in \mathbb{R}^{m \times n}$ wird oft überhaupt kein Unterschied zwischen beiden Objekten gemacht, d.h. man bezeichnet eine Matrix auch als linearen Operator etc. Das ist natürlich nicht ganz korrekt — schließlich ist eine lineare Abbildung T von \mathbb{R}^n nach \mathbb{R}^m etwas anderes als eine $m \times n$ -Matrix, ein rechteckiges Zahlenschema also —, aber es ist gerechtfertigt, weil sich T und A in so einfacher Weise gegenseitig bestimmen: T ist der Multiplikationsoperator $x \mapsto Ax$, der A an n -gliedrige Spaltenvektoren multipliziert, und A ist die Matrix mit den Bildern $T(e_j) \in \mathbb{R}^m$ der kanonischen Basisvektoren von \mathbb{R}^n als Spalten. Man sagt, dass *der lineare Operator T durch die Matrix A dargestellt wird*, wenn dieser Zusammenhang zwischen T und A besteht.

Die eigentliche Bedeutung der Matrix-an-Vektor-Multiplikation und die mathematische Begründung der dazu getroffenen Definition liegt in der damit erreichten Darstellung linearer Operatoren. Die Multiplikation ist eben gerade so definiert, dass $x \mapsto Ax$ die linearen Operatoren von \mathbb{R}^n nach \mathbb{R}^m beschreibt. Gleichzeitig ist mit dem obigen Ergebnis auch die vor der Definition gemachte Aussage gerechtfertigt, dass ein linearer Zusammenhang $y = T(x)$ zwischen einer abhängigen vektoriellen Größe y (in \mathbb{R}^m) und einer unabhängigen vektoriellen Variablen x (in \mathbb{R}^n) *immer* durch eine Matrix-an-Vektor-Multiplikation $y = Ax$ beschrieben werden kann.

4) Eine Folge des Ergebnisses aus 3), die keineswegs von vorneherein klar war, ist:

- Jede lineare Abbildung $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ ist durch die Angabe von $m \cdot n$ reellen Zahlen eindeutig bestimmt, nämlich durch die Einträge der $m \times n$ -Matrix A , die T darstellt.

Daher lassen sich lineare Operatoren $T(x) = Ax$ auch einfach programmieren: Man braucht nur die Matrixeinträge von A einzugeben und die Multiplikation von Matrizen an Spaltenvektoren zu programmieren, also die Rechenvorschrift $y_i := a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n$ für $i = 1 \dots m$, dann liefert der Computer für jede Eingabe $x \in \mathbb{R}^n$ den zugehörigen Wert $y = T(x) = Ax$ der linearen Abbildung als Ausgabe. Beliebige Abbildungen von \mathbb{R}^n

schon die abkürzende Schreibweise $Ax = b$ gebraucht, wobei $A = (a_{ij})$ die $m \times n$ -Koeffizientenmatrix des Systems ist, $b = (b_1, b_2, \dots, b_m) \in \mathbb{R}^m$ die Spalte der rechten Seiten und $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ der Spaltenvektor der Unbekannten. Jetzt können wir aber $Ax \in \mathbb{R}^m$ als Produkt der Matrix A mit x auffassen, dann ist die i -te Komponente von Ax gerade die linke Seite der i -ten linearen Gleichung in (LGS). Somit ist

$$Ax = b$$

nicht nur eine symbolische Abkürzung für das Gleichungssystem, sondern, wenn wir Ax als Produkt auffassen und die Gleichung als Vektorgleichung in \mathbb{R}^m , eine sinnvolle äquivalente Formulierung von (LGS).

2) Aus der Linearität des Operators $\mathbb{R}^n \ni x \mapsto Ax \in \mathbb{R}^m$ ergeben sich nun unmittelbar folgende Aussagen über das lineare Gleichungssystem (LGS):

- (i) Gilt $Ax = b$ und $A\tilde{x} = \tilde{b}$, so löst die Linearkombination $rx + \tilde{r}\tilde{x}$ mit Koeffizienten $r, \tilde{r} \in \mathbb{R}$ das lineare Gleichungssystem mit derselben Koeffizientenmatrix A und mit rechter Seite $rb + \tilde{r}\tilde{b}$. (Entsprechend für Linearkombinationen $r_1x_1 + \dots + r_lx_l$ von Lösungen x_1, \dots, x_l zu $Ax_k = b_k$ für $k = 1 \dots l$.)
- (ii) Jede Linearkombination von Lösungen des homogenen Gleichungssystems $Ax = 0$ ist wieder eine Lösung des homogenen Systems. Jede affine Kombination von Lösungen des inhomogenen Gleichungssystems $Ax = b$ löst ebenfalls dieses System.
- (iii) Jede Linearkombination von zulässigen rechten Spalten zur Koeffizientenmatrix A ist wieder eine zulässige rechte Spalte.

Die erste Aussage folgt sofort aus $A(rx + \tilde{r}\tilde{x}) = r(Ax) + \tilde{r}(A\tilde{x}) = rb + \tilde{r}\tilde{b}$ und entsprechend für allgemeine Linearkombinationen. Die zweite erhält man daraus, indem man alle rechten Spalten Null setzt oder alle gleich b und dabei $r_1 + \dots + r_l = 1$ für affine Kombinationen beachtet. Die dritte Aussage schließlich folgt sofort aus der ersten, wenn man sich daran erinnert, dass eine zulässige rechte Spalte ein Vektor $b \in \mathbb{R}^m$ ist, für den $Ax = b$ (mindestens) eine Lösung $x \in \mathbb{R}^n$ besitzt. Alle drei Aussagen sind also unmittelbare Konsequenzen der Linearität des Operators $x \mapsto Ax$. Man kann sie natürlich auch ohne Matrix- und Vektorrechnung — aber viel unübersichtlicher und weniger elegant — durch direkte Rechnungen mit dem ursprünglichen Gleichungssystem (LGS) herleiten. Als weitere Folgerung aus (i) notieren wir noch, was schon mehrfach bemerkt wurde:

- (iv) Die allgemeine Lösung des inhomogenen Systems $Ax = b$ ist, wenn es überhaupt eine Lösung gibt, die Summe irgendeiner ("speziellen") Lösung $x^* \in \mathbb{R}^n$ und der allgemeinen Lösung des homogenen Systems $Ax = 0$;

denn wegen $Ax^* = b$ gilt ja $A(x+x^*) = b \iff Ax + Ax^* = b \iff Ax = 0$.

3) Man kann die Erkenntnisse aus 2) bei der Lösung linearer Gleichungssysteme praktisch nutzen. Hat man z.B. schon Lösungen für zwei rechte Spalten b und c und soll für eine weitere rechte Spalte d lösen, die man als Linearkombination $d = rb + sc$ von b und c erkennt, so braucht man nicht mehr zu rechnen, sondern kann gemäß (i) einfach die entsprechende Linearkombination der bereits bekannten Lösungen nehmen. Und wenn man zu $Ax = b$ schon die allgemeine Lösung bestimmt hat, so kennt man gemäß (iv) auch die allgemeine Lösung des homogenen Systems und damit wieder nach (iv) auch die allgemeine Lösung zu der neuen rechten Spalte d . ■

Die Eigenschaft der Lösungsmenge des homogenen linearen Gleichungssystems $Ax = 0$ und der Menge der zulässigen rechten Seiten b für $Ax = b$, dass beliebige Linearkombinationsbildung mit Elementen der Menge nicht aus der Menge herausführt, erweist sich als grundlegend für die Definition von "linearen Teilmengen" allgemeiner Vektorräume, der sogenannte *Unterräume*. Mit dieser Terminologie kann man dann (ii) dadurch ausdrücken, dass man sagt, die Lösungsmenge eines homogenen linearen Gleichungssystems $Ax = 0$ von m Gleichungen für n Unbekannte sei ein "Unterraum" von \mathbb{R}^n und die Lösungsmenge des inhomogenen Systems $Ax = b$ ein "affiner Unterraum" (oder leer). Und (iii) wird ausgedrückt, indem man sagt, dass die Menge der zulässigen rechten Seiten ein Unterraum von \mathbb{R}^m sei. Wir verschieben diese Diskussion aber auf einen späteren Abschnitt, weil wir nun zum *Produkt von zwei beliebigen, aber zusammen "passenden" Matrizen* kommen wollen.

Um diese Produktbildung zu motivieren, die alle bisher eingeführten Produkte im Matrixkalkül als Spezialfälle enthält, betrachten wir das Produkt $y = Ax$ einer $m \times n$ -Matrix A mit einem n -gliedrigen Spaltenvektor x und das Produkt $z = By$ einer weiteren $l \times m$ -Matrix B mit dem m -gliedrigen Spaltenvektor y (damit dieses Produkt By erklärt ist, muss B natürlich m Spalten haben!). Dann ist $z = B(Ax) \in \mathbb{R}^l$, und es ist klar, dass z linear von x abhängt. Also gibt es eine $l \times n$ -Matrix C , so dass wir direkt $z = Cx$ als Produkt von C mit x ausgedrückt haben. Mit anderen Worten: Wenn die linearen Operatoren $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ und $U: \mathbb{R}^m \rightarrow \mathbb{R}^l$ durch $T(x) = Ax$ und $U(y) = By$ gegeben sind, so suchen wir die Matrix C , welche die *Hintereinanderausführung* (*Verkettung, Komposition*) $U \circ T(x) := U(T(x))$ darstellt. Diese Verkettung ist offenbar wieder ein linearer Operator von \mathbb{R}^n nach \mathbb{R}^l . (Man kann skalare Faktoren und Summenbildung erst aus T und dann aus U "herausziehen" und damit aus $U \circ T$.) Also gibt es gemäß dem Zusammenhang zwischen linearen Operatoren und Matrizen auch genau eine Matrix C mit $U \circ T(x) = Cx$ für alle $x \in \mathbb{R}^n$, und diese Matrix C muss natürlich nun das Format (l, n) haben.

Um C zu bestimmen, genügt es, für x kanonische Basisvektoren e_j aus \mathbb{R}^n einzusetzen; denn $U \circ T(e_j) = B(Ae_j)$ ist ja die j -te Spalte von C . Dann ist Ae_j die j -te Spalte von A , und $B(Ae_j)$ hat gemäß der Definition des Produkts von B mit dieser Spalte als h -ten Eintrag das Skalarprodukt der h -ten Zeile von B mit der j -ten Spalte von A , also

$$b_{h1}a_{1j} + b_{h2}a_{2j} + \dots + b_{hm}a_{mj} = \sum_{i=1}^m b_{hi}a_{ij},$$

wenn $A = (a_{ij})$ und $B = (b_{hi})$ ist. Dies muss nun der Eintrag c_{hj} in der h -ten Zeile und j -ten Spalte der Matrix C sein, und damit ist sie bestimmt: c_{hj} ist das Skalarprodukt der h -ten Zeile von B mit der j -ten Spalte von A . Da die so definierte Matrix C bilinear von B und A abhängt, ist es sinnvoll, sie als Produkt $C = BA$ zu schreiben. Mit dieser Definition des Produkts BA (und nur mit dieser) gilt also $(BA)x = B(Ax)$, d.h. das Matrixprodukt BA stellt den linearen Operator dar, den man durch Hintereinanderausführung der linearen Operatoren zu A und B erhält.

Diese Überlegungen motivieren die folgende mathematische Definition des Matrixprodukts und haben durchaus auch ökonomische Bedeutung. Man stelle sich vor, dass gewisse Endprodukte Nr. $1 \dots n$ aus Zwischenprodukten Nr. $1 \dots m$ gefertigt werden gemäß einer Produktionsmatrix $S = (s_{ij})$ und dass die Zwischenprodukte ihrerseits aus $1 \dots l$ Rohstoffen hergestellt werden gemäß einer Produktionsmatrix $R = (r_{hi})$. Für einen gegebenen Outputvektor $\mathbf{x} = (x_1, \dots, x_n)$ zu den Endprodukten beschreibt dann $\mathbf{y} = S\mathbf{x}$ die dafür erforderliche Produktion der Zwischenprodukte und $\mathbf{z} = R\mathbf{y} = R(S\mathbf{x})$ wiederum

die für die Realisierung dieser Zwischenproduktion erforderlichen Mengen an Rohstoffen. Das Matrixprodukt RS gibt also zu jedem Produktionsplan \mathbf{x} direkt den zugehörigen Rohstoffbedarfsvektor $(RS)\mathbf{x}$ an und beschreibt damit auch in dieser realen ökonomischen Situation die Komposition von zwei linearen Abhängigkeiten $\mathbf{y} = S\mathbf{x}$ und $\mathbf{z} = R\mathbf{y}$. In analoger Weise kommen Matrixprodukte überall ins Spiel, wo (vektorielle) ökonomische Größen linear von anderen Größen zweiter Art abhängen, die ihrerseits wiederum von Größen einer dritten Art linear abhängig sind.

DEFINITION: Das **Produkt zweier Matrizen** ist definiert, wenn die Spaltenzahl des ersten Faktors gleich der Zeilenzahl des zweiten Faktors ist; das Ergebnis der Multiplikation ist dann die Matrix mit der Zeilenzahl des ersten und der Spaltenzahl des zweiten Faktors, deren Einträge die Skalarprodukte der Zeilen des ersten mit den Spalten des zweiten Faktors sind. Ist also $B = (b_{hi})$ eine $l \times m$ -Matrix und $A = (a_{ij})$ eine $m \times n$ -Matrix, so ist das Produkt BA definiert und gleich der $l \times n$ -Matrix $C = (c_{hj})$ mit den Einträgen

$$\begin{aligned} c_{hj} &= \text{Skalarprodukt der } h\text{-ten Zeile von } B \text{ mit der } j\text{-ten Spalte von } A \\ &= b_{h1}a_{1j} + b_{h2}a_{2j} + \dots + b_{hm}a_{mj} = \sum_{i=1}^m b_{hi}a_{ij}. \quad \blacksquare \end{aligned}$$

DISKUSSION: 1) Bevor man eine Matrixmultiplikation ausführt, sollte man sich zunächst

- Klarheit über die Formate der Matrixfaktoren und der Produktmatrix beschaffen!
- Das Produkt ist nur definiert, wenn die Matrixfaktoren "passen", d.h. wenn die Spaltenzahl des ersten Faktors gleich der Zeilenzahl des zweiten ist;
- das Matrixprodukt hat dann so viele Zeilen wie der erste Matrixfaktor und so viele Spalten wie der zweite.

Die *schlimmsten Fehler*, die man bei der Matrixmultiplikation machen kann, sind das Multiplizieren von Matrizen, deren Formate nicht zusammen passen, oder die Angabe einer Produktmatrix, die nicht das richtige Format hat. (Das kann natürlich nur passieren, wenn man sich nicht an die Definition hält, sondern irgendwelche Vorschriften zur Matrixmultiplikation frei erfindet.) Hat A das Format (m, n) und B das Format (l, \tilde{m}) , so muss also zunächst $\tilde{m} = m$ überprüft werden, damit das Produkt BA überhaupt gebildet werden kann, und dann ist das Ergebnis BA eine Matrix des Formats (l, n) . Für die Bedingung $\tilde{m} = m$ sagt man auch, dass die Matrizen B und A *komponierbar* sind oder *verkettet werden können*, weil diese Bedingung ja gerade bedeutet, dass die Hintereinanderausführung der linearen Operatoren $x \mapsto Ax$ und $y \mapsto By$ möglich ist. Dabei kommt es auf die Reihenfolge der Matrixfaktoren an; es ist durchaus möglich, dass B mit A verkettet werden kann ($\tilde{m} = m$), aber nicht A mit B ($n \neq l$)!

2) Ist einer der Matrixfaktoren quadratisch, so hat das Matrixprodukt — wenn es definiert ist — dasselbe Format wie der andere Faktor. Sind *beide Faktoren quadratisch*, so ist das Produkt genau dann definiert, wenn sie dasselbe Format haben, und das Ergebnis der Multiplikation ist dann wieder eine Matrix von diesem quadratischen Format.

3) Bei etwas Übung multipliziert man Matrizen (mit ganzen Zahlen oder einfachen Brüchen als Einträgen) im Kopf, indem man "schielend" mit dem linken Auge eine Zeile der ersten Matrix und mit dem rechten Auge eine Spalte der zweiten Matrix abtastet, dabei das

Skalarprodukt der Zeile mit der Spalte ermittelt und diesen Wert in die richtige Position der Ergebnismatrix einträgt (dieselbe Zeilennummer wie die abgetastete Zeile und dieselbe Spaltennummer wie die abgetastete Spalte). Für weniger Geübte mag folgendes **Rechenschema für die Matrixmultiplikation** hilfreich sein (in der Mathematik für Ökonomen wird das *Falksches Schema* genannt — es ist aber keine so großartige Erfindung, dass man den Namen eines Erfinders hierzu nennen müsste):

$$\begin{array}{c}
 \left. \begin{array}{c} \underbrace{\hspace{10em}} \\ m \end{array} \right\} \left(\begin{array}{c} \overbrace{\hspace{10em}}^n \\ \left(\begin{array}{cccc} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ \vdots & & a_{ij} & & \vdots \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mn} \end{array} \right) \end{array} \right) = A \\
 \\
 \left. \begin{array}{c} l \\ B = \left(\begin{array}{cccc} b_{11} & \cdots & \cdots & \cdots & b_{1m} \\ \vdots & & \vdots & & \vdots \\ b_{h1} & \cdots & b_{hi} & \cdots & b_{hm} \\ \vdots & & \vdots & & \vdots \\ b_{l1} & \cdots & \cdots & \cdots & b_{lm} \end{array} \right) \end{array} \right\} \left(\begin{array}{c} \left(\begin{array}{cccc} c_{11} & \cdots & \cdots & c_{1n} \\ \vdots & & \downarrow & \vdots \\ \rightarrow & c_{hj} & \cdots & \vdots \\ \vdots & \vdots & & \vdots \\ c_{l1} & \cdots & \cdots & c_{ln} \end{array} \right) \end{array} \right) = BA
 \end{array}$$

Das Schema zeigt an, dass die Eintragung c_{hj} in Zeile h und Spalte j der Produktmatrix BA alleine aus der h -ten Zeile des ersten Faktors B und der j -ten Spalte des zweiten Faktors A berechnet wird; allerdings muss man noch wissen *wie*, nämlich als Skalarprodukt von dieser Zeile und dieser Spalte.

4) Man kann die Matrixmultiplikation auch so beschreiben: *Der j -te Spaltenvektor des Produkts $C = BA$ ist das Produkt von B mit dem j -ten Spaltenvektor von A .*

$$\begin{pmatrix} c_{1j} \\ \vdots \\ c_{lj} \end{pmatrix} = \begin{pmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & & \vdots \\ b_{h1} & \cdots & b_{hm} \end{pmatrix} \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}$$

Für den Fall, dass der zweite Faktor eine $m \times 1$ -Matrix ist, also ein Spaltenvektor, stimmt das nun definierte Matrixprodukt mit dem zuvor definierten Produkt einer Matrix mit einem (passenden) Spaltenvektor natürlich überein. Für die Zeilen des Produkts gilt:

Der h -te Zeilenvektor des Produkts $C = BA$ ist das Produkt der h -ten Zeile von B mit der Matrix A .

$$(c_{h1} \ \cdots \ c_{hn}) = (b_{h1} \ \cdots \ b_{hm}) \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

5) **Rechenregeln für die Matrixmultiplikation:**

$$\begin{array}{ll}
 B(rA) = r(BA) = (rB)A & \text{(Homogenität)} \\
 (B + \tilde{B})A = BA + \tilde{B}A & \text{(linkes Distributivgesetz)} \\
 B(A + \tilde{A}) = BA + B\tilde{A} & \text{(rechtes Distributivgesetz)}
 \end{array}$$

Diese Regeln zusammenfassend sagt man, dass *die Matrixmultiplikation eine bilineare Rechenoperation ist*. Man kann dann auch Linearkombinationen aus jedem Faktor herausziehen, also z.B. $B(rA + \tilde{r}\tilde{A}) = rBA + \tilde{r}(B\tilde{A})$. Natürlich müssen die Matrizen dabei zusammen passen, also A und \tilde{A} von einem Format (m, n) und B, \tilde{B} von einem Format

(l, m) sein. Der Beweis der Bilinearität ergibt sich unmittelbar aus der schon festgestellten Bilinearität der Skalarproduktbildung oder (weniger elegant und mühsamer) durch direktes Nachrechnen. Für die Multiplikation mit Nullmatrizen und Einheitsmatrizen gilt:

$$0A = 0, \quad B0 = 0; \quad IA = A, \quad BI = B,$$

vorausgesetzt die Matrixprodukte sind definiert. (Dabei steht "0" für Nullmatrizen evtl. verschiedenen Formats und "I" für Einheitsmatrizen evtl. verschiedenen Formats.) Die *Multiplikation mit einer passenden Einheitsmatrix von links oder rechts an eine Matrix reproduziert diese*. Das folgt daraus, dass die Multiplikation einer Einheitsmatrix an einen passenden Spaltenvektor diesen Vektor reproduziert. Und dass das Produkt mit einer passenden Nullmatrix wieder eine Nullmatrix als Ergebnis hat, ist klar.

6) Wenn sie zusammen passen, so kann man auch mehr als zwei Matrizen miteinander multiplizieren. Dabei hängt das Ergebnis nicht von der Klammerung der einzelnen Faktoren ab, d.h. es gilt

$$C(BA) = (CB)A \quad (\text{Assoziativgesetz der Matrixmultiplikation}),$$

wenn die Matrixprodukte auf einer Seite der Gleichung definiert sind (dann sind die Produkte auf der anderen Seite auch definiert; wenn man also BA und $C(BA)$ bilden kann, so auch CB und $(CB)A$). Das kann direkt mit Matrixeinträgen nachgerechnet werden und sieht dann so aus:

$$\sum_{h=1}^l c_{gh} \left(\sum_{i=1}^m b_{hi} a_{ij} \right) = \sum_{h=1}^l \sum_{i=1}^m c_{gh} b_{hi} a_{ij} = \sum_{i=1}^m \left(\sum_{h=1}^l c_{gh} b_{hi} \right) a_{ij}.$$

Aber eigentlich ist diese Rechnung überflüssig; denn wir brauchen uns nur daran zu erinnern, dass die Matrixmultiplikation der Hintereinanderausführung der zugeordneten linearen Operatoren entspricht, und die Hintereinanderausführung von Operatoren/Abbildungen ist assoziativ. (Es ist gleich, ob man sich von drei nacheinander auszuführenden Operationen die beiden ersten oder die beiden letzten zu einer einzigen Operation zusammengefasst denkt.) Daher gilt für alle passenden Spaltenvektoren x auch $(C(BA))x = C((BA)x) = C(B(Ax)) = (CB)(Ax) = ((CB)A)x$, und das bedeutet $C(BA) = (CB)A$, weil die linearen Operatoren ja die zugehörigen Matrizen eindeutig bestimmen. (Setze kanonische Einheitsvektoren für x ein.)

7) Wir betonen nochmals, was wir gerade benutzt haben und was die Grundlage des tieferen Verständnisses (und nicht nur der rechnerischen Beherrschung) der Matrixmultiplikation ist:

- Das Produkt BA zweier Matrizen ist, wenn definiert, die Matrixdarstellung desjenigen linearen Operators, der durch Hintereinanderausführung des zu A gehörenden (ersten) und des zu B gehörenden (nachfolgenden) linearen Operators entsteht.

Da $x \mapsto Ax$ die zu A gehörende lineare Abbildung ist und $y \mapsto By$ die zu B gehörende, heißt das gerade $(BA)x = B(Ax)$ für alle zu A passenden Spaltenvektoren x , und genau durch diese Bedingung hatten wir ja die Matrixmultiplikation definiert. Man beachte, dass im Produkt BA der rechts stehende Faktor der zuerst auszuführenden linearen Abbildung entspricht. Das liegt daran, dass wir Abbildungen — wie allgemein üblich — links vor die Argumente schreiben, also wird bei einer Verkettung $U \circ T(x) = U(T(x))$ auch die rechts stehende Abbildung T zuerst ausgeführt.

(Manche Autoren schreiben deshalb $(x)T$ oder x^T statt $T(x)$; dann erscheinen bei Hintereinanderausführung $((x)T)U = (x^T)^U$ die Abbildungen T und U von links nach rechts gelesen in der Reihenfolge, in der sie ausgeführt werden. Im Matrixkalkül werden nun die Vektoren als Zeilenvektoren geschrieben und die Matrizen von rechts daran multipliziert, um lineare Operatoren zu beschreiben. Dann entspricht im Matrixprodukt AB der erste Faktor auch der zuerst ausgeführten linearen Abbildung. Unter dem Gesichtspunkt, dass wir von links nach rechts lesen, mag diese Vorgehensweise gerechtfertigt sein. Sie ist aber unüblich, und wir folgen der Konvention, zur Beschreibung linearer Operatoren die Vektoren $x \in \mathbb{R}^n$ immer als Spalten zu schreiben und die Matrizen links davor.)

8) Ganz dringende **Warnung**:

- *Für die Matrixmultiplikation gilt kein Kommutativgesetz; die Existenz und der Wert eines Matrixprodukts hängen im Allgemeinen von der Reihenfolge der Faktoren ab!*

Für Matrizen A und B mit beliebigen Formaten (m, n) und (k, l) kann $AB = BA$ im Allgemeinen schon deswegen nicht richtig sein, weil mit der Existenz des einen Produkts AB (d.h. $n = k$) nicht auch das andere Produkt BA definiert (d.h. $l = m$) sein muss. Und auch wenn beide Produkte AB und BA definiert sind (also $n = k$ und $l = m$), so haben sie im Allgemeinen verschiedenes Format (nämlich (n, n) bei BA , aber (m, m) bei AB) und können schon deswegen nicht gleich sein. Schließlich zeigen einfachste Beispiele mit 2×2 -Matrizen, dass selbst in dem Fall, wo A und B gleichformatige quadratische Matrizen sind und somit auch AB und BA dasselbe quadratische Format besitzen, trotzdem AB und BA nicht dieselben Einträge in denselben Positionen haben müssen, so dass $AB \neq BA$ ist, obwohl beide Produkte definiert und Matrizen des gleichen Formats sind.

Die Ungültigkeit des Kommutativgesetzes für die Matrizenmultiplikation ist eine häufige Fehlerquelle, weil geläufige Rechenregeln, welche auf der Vertauschbarkeit der Faktoren in einem Produkt beruhen, auf die Matrixmultiplikation angewendet werden, wo sie nicht gelten. Zum Beispiel ist für gleichformatige quadratische Matrizen $(A+B)^2 = (A+B)(A+B) = A^2 + BA + AB + B^2$, und das ist $\neq A^2 + 2AB + B^2$, wenn $AB \neq BA$ ist. (Bei quadratischen Matrizen A ist dabei A^2 natürlich eine Abkürzung für das Produkt AA von A mit sich selbst.) Aus analogem Grund ist für gleichformatige quadratische Matrizen im Allgemeinen $(A+B)(A-B) \neq A^2 - B^2$. Die geläufigen binomischen Formeln gelten also für die Matrixmultiplikation nicht! Und in einer Summe $AB + BC$ von Matrixprodukten kann man den Faktor B nicht einfach ausklammern (nur $AB + CB = (A+C)B$ und $BA + BC = B(A+C)$ sind korrekt). Eine *Ausnahme* ist die Multiplikation mit einer Einheitsmatrix \mathbb{I} oder einer skalaren Matrix $r\mathbb{I}$; dafür gilt immer $(r\mathbb{I})A = rA = A(r\mathbb{I})$, wenn die Matrixprodukte erklärt sind, wenn also A und \mathbb{I} dasselbe Format haben.

9) Noch eine Warnung: *Eine Division von Matrizen ist nicht definiert.* Der Quotient " B/A " zweier Matrizen A und B müsste nämlich eine Matrix-Lösung X der Gleichung $AX = B$ sein, aber diese Gleichung ist im Allgemeinen unlösbar, selbst wenn A keine Nullmatrix ist. Außerdem kann es unendlich viele Lösungsmatrizen X geben, und die gleichberechtigte Gleichung $XA = B$ kann ganz andere Lösungen haben. Welche dieser Lösungen X soll man dann zur Definition des Quotienten von B und A nehmen? Darauf gibt es keine sinnvolle Antwort, und deshalb bleibt die Division von Matrizen eben undefiniert.

In diesem Zusammenhang bemerken wir noch, dass *aus dem Verschwinden eines Matrixprodukts $AB = 0$ keineswegs das Verschwinden eines Faktors folgt.* Ein Beispiel erhalten

wir, wenn wir $A \neq 0$ als Koeffizientenmatrix eines homogenen linearen Gleichungssystems mit einer nichttrivialen Lösung $x \neq 0$ nehmen; ist B die aus der einzigen Spalte x bestehende Matrix, so gilt auch $B \neq 0$, aber $AB = Ax = 0$.

10) Die Menge der quadratischen Matrizen eines gegebenen Formats (n, n) ist versehen mit einer Addition und einer Multiplikation mit Skalaren, aber auch noch mit einer zusätzlichen Multiplikation *in sich*, weil das Produkt von zwei $n \times n$ -Matrizen wieder das Format (n, n) hat. Eine mit solchen Rechenoperationen versehene Menge, für die alle bei der Matrixrechnung oben angegebenen Rechengesetze gelten, heißt in der Mathematik eine *Algebra* (über dem Zahlbereich, dem die Skalare entnommen sind, hier also über \mathbb{R}). Der Raum $\mathbb{R}^{n \times n}$ der $n \times n$ -Matrizen bildet somit eine Algebra, und zwar für $n \geq 2$ eine sog. *nichtkommutative Algebra*, weil eben für die Multiplikation das Kommutativgesetz $AB = BA$ dann nicht allgemein gilt. Beispiele für kommutative Algebren, in denen man dieselben Rechenoperationen und -gesetze und zusätzlich noch das Kommutativgesetz hat, sind der Raum der Polynomfunktionen auf \mathbb{R} , und der Raum der differenzierbaren reellwertigen Funktionen auf einem offenen Intervall in \mathbb{R} ; hier gilt $fg = gf$ für das Produkt von Funktionen, also Kommutativität. ■

BEISPIELE (zur *Matrixmultiplikation*):

$$1) \quad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 1 & 1 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 8 \\ 8 & 4 \end{pmatrix}, \quad \begin{pmatrix} 2 & 0 \\ 1 & 1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 4 & 6 \\ 4 & 4 & 4 \\ 6 & 4 & 2 \end{pmatrix};$$

hier sind die Produkte AB und BA definiert, aber von verschiedenem Format.

$$\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 2 & -1 & 0 \\ 0 & 1 & -2 \end{pmatrix} = \begin{pmatrix} 2 & -2 & 2 \\ -2 & 2 & -2 \end{pmatrix}, \quad \text{aber} \quad \begin{pmatrix} 2 & -1 & 0 \\ 0 & 1 & -2 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \text{ ist nicht definiert.}$$

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix};$$

hier ist $A \neq 0$, $B \neq 0$, $AB = 0$, $BA = 0$, aber $AB \neq BA$ wegen unterschiedlichen Formats.

2) *Produkte mit kanonischen Einheitsvektoren:*

$$\overbrace{\begin{pmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{pmatrix}}^m \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{i1} & \dots & a_{in} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} = (a_{i1} \dots a_{in}) = i\text{-te Zeile von } A;$$

$$\begin{pmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \dots & a_{mj} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix} = j\text{-te Spalte von } A.$$

↑
j-te Position (von n)

3) Produkt eines Spalten- mit einem Zeilenvektor (sog. dyadisches Produkt):

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} (b_1 \ b_2 \ \dots \ b_n) = \begin{pmatrix} a_1 b_1 & a_1 b_2 & \dots & a_1 b_n \\ a_2 b_1 & a_2 b_2 & \dots & a_2 b_n \\ \vdots & \vdots & & \vdots \\ a_m b_1 & a_m b_2 & \dots & a_m b_n \end{pmatrix} \in \mathbb{R}^{m \times n}.$$

Dies ist für beliebige m und n definiert (passt immer, da der erste Matrixfaktor eine Spalte und der zweite eine Zeile hat). In der anderen Reihenfolge ist das Matrixprodukt dagegen nur definiert, wenn beide Vektoren gleich viele Komponenten haben, und das Ergebnis ist dann das Skalarprodukt der beiden Vektoren a und b (bzw. genauer: die 1×1 -Matrix mit dem Skalarprodukt als einzigem Eintrag):

$$(b_1 \ b_2 \ \dots \ b_n) \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = (a_1 b_1 + a_2 b_2 + \dots + a_n b_n) = (a \cdot b).$$

4) Die Multiplikation mit einer Diagonalmatrix an eine beliebige (passende!) Matrix A hat folgenden Effekt: Bei Multiplikation von D rechts an A

$$\begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}$$

werden die Spalten von A der Reihe nach mit den Diagonaleinträgen von D vervielfacht, bei Multiplikation von D links an A dagegen die Zeilen :

$$\begin{pmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \dots & a_{mj} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} d_1 & & & & 0 \\ & \ddots & & & \\ & & d_j & & \\ & & & \ddots & \\ 0 & & & & d_n \end{pmatrix} = \begin{pmatrix} d_1 a_{11} & \dots & d_j a_{1j} & \dots & d_n a_{1n} \\ \vdots & & \vdots & & \vdots \\ d_1 a_{m1} & \dots & d_j a_{mj} & \dots & d_n a_{mn} \end{pmatrix},$$

$$\begin{pmatrix} d_1 & & & & 0 \\ & \ddots & & & \\ & & d_i & & \\ & & & \ddots & \\ 0 & & & & d_m \end{pmatrix} \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{i1} & \dots & a_{in} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} = \begin{pmatrix} d_1 a_{11} & \dots & d_1 a_{1n} \\ \vdots & & \vdots \\ d_i a_{i1} & \dots & d_i a_{in} \\ \vdots & & \vdots \\ d_m a_{m1} & \dots & d_m a_{mn} \end{pmatrix}.$$

Es ist also bei einer $n \times n$ -Matrix A nicht gleichgültig, ob man A mit der Diagonalmatrix von rechts oder von links multipliziert. Ausnahme: D ist eine skalare Matrix, also eine Diagonalmatrix mit lauter gleichen Diagonaleinträgen $d_1 = \dots = d_n = r$; dann ist $AD = DA$ für alle $A \in \mathbb{R}^{n \times n}$ (und nur dann, wie man beweisen kann). Weitere Ausnahme: A ist selbst eine Diagonalmatrix vom gleichen Format wie D ; dann gilt $AD = DA$; denn:

- das Produkt von zwei $n \times n$ -Diagonalmatrizen ist wieder eine Diagonalmatrix, und ihre Einträge sind die Produkte der gleich positionierten Einträge beider Faktoren.

5) Zeilen- bzw. Spaltenoperationen mit einer Matrix A kann man durch Multiplikation passender quadratischer Matrizen von links bzw. von rechts an A beschreiben. Diese Matrizen sehen folgendermaßen aus (alle quadratisch vom gleichen Format (n, n) , die nicht angegebenen Einträge sind 0 außerhalb der Diagonalen und 1 auf der Diagonalen) und heißen

Elementarmatrizen:

$$M_i(r) := \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & r & & \dots \\ & & & & 1 & \\ & & & & & \ddots \\ & & & & & & 1 \end{pmatrix} \leftarrow i\text{-te Zeile, } r \in \mathbb{R}_{\neq 0}$$

$$A_{ij}(s) := \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & \ddots & & & \\ & & & s & & \\ & & & & \ddots & \\ & & & & & 1 \\ & & & & & & 1 \end{pmatrix} \leftarrow i\text{-te Zeile, } s \in \mathbb{R}$$

↑ j -te Spalte, $j \neq i$

$$V_{ij} := \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 0 & \dots & 1 & \dots \\ & & \vdots & \ddots & \vdots & \\ & & 1 & \dots & 0 & \dots \\ & & \vdots & & \vdots & \ddots \\ & & & & & & 1 \end{pmatrix} \leftarrow i\text{-te Zeile}$$

i -te Spalte ↑ ↑ j -te Spalte

← j -te Zeile ($j \neq i$)

$M_i(r)$ ist die Diagonalmatrix mit Eintrag $r \neq 0$ in i -ter Diagonalposition und anderen Diagonaleinträgen 1, $A_{ij}(s)$ entsteht aus der Einheitsmatrix, indem man den Nulleintrag in Position (i, j) durch den Skalar s ersetzt, und V_{ij} erhält man aus der Einheitsmatrix durch Vertauschung der i -ten mit der j -ten Zeile (oder der i -ten mit der j -ten Spalte, das läuft hier auf Dasselbe hinaus). Allgemein bezeichnet man eine Matrix, die aus der Einheitsmatrix durch Vertauschungen von Zeilen und / oder Spalten hervorgeht, als *Permutationsmatrix*, weil ihre Multiplikation an einen passenden Spaltenvektor bewirkt, dass dessen Komponenten entsprechend permutiert (vertauscht) werden. Die Multiplikation einer Elementarmatrix von links an eine passende Matrix A bewirkt nun eine Zeilenoperation bei A wie folgt:

$M_i(r)A$ entsteht aus A durch Multiplikation der i -ten Zeile mit r ,

$A_{ij}(s)A$ entsteht aus A durch Addition des s -fachen der j -ten Zeile zur i -ten;

$V_{ij}A$ entsteht aus A durch Vertauschung der i -ten mit der j -ten Zeile.

Entsprechend bewirkt die Multiplikation einer Elementarmatrix von rechts an eine passende Matrix B eine Spaltenoperation bei B :

$BM_j(r)$ entsteht aus B durch Multiplikation der j -ten Spalte mit r ,

$BA_{ij}(s)$ entsteht aus B durch Addition des s -fachen der i -ten Spalte zur j -ten;

BV_{ij} entsteht aus B durch Vertauschung der i -ten mit der j -ten Spalte.

Nun haben wir in 3.1 gesehen, dass jede $m \times n$ -Matrix A durch Zeilenoperationen auf Zeilen-Stufen-Form gebracht werden kann. Folglich gibt es $m \times m$ -Elementarmatrizen E_1, \dots, E_p , derart dass das Matrixprodukt $Z = E_p \dots E_2 E_1 A$ eine $m \times n$ -Matrix von Zeilen-Stufen-Form ist. Durch Spaltenoperationen, also durch Multiplikationen mit wei-

teren $n \times n$ -Elementarmatrizen F_1, \dots, F_q von rechts, kann man die Zeilen-Stufen-Matrix Z in die Form einer $m \times n$ -Blockmatrix bringen mit einer $l \times l$ -Einheitsmatrix \mathbb{I}_l als linker oberer Block und Nullmatrizen als weiteren Blöcken. Dabei ist l die Anzahl der Stufen (also der Zeilen mit gewissen Einträgen $\neq 0$) in der Zeilen-Stufen-Matrix Z , also der Rang der Matrix A . Resultat ist eine Darstellung

$$E_p \cdots E_2 E_1 A F_1 F_2 \cdots F_q = \left(\begin{array}{c|c} \mathbb{I}_l & 0_{l,n-l} \\ \hline 0_{m-l,l} & 0_{m-l,n-l} \end{array} \right),$$

wo die Subskripte die Formate der Nullmatrizen anzeigen. Nun kann man die Zeilen- und Spaltenoperationen durch Multiplikation mit Elementarmatrizen von links und rechts wieder rückgängig machen und erhält so Gleichheit von A mit der entsprechend umgeformten Matrix der rechten Seite. Auf diese Weise bekommt man die Darstellung

$$A = E_1 E_2 \cdots E_p \left(\begin{array}{c|c} \mathbb{I}_l & 0_{l,n-l} \\ \hline 0_{m-l,l} & 0_{m-l,n-l} \end{array} \right) F_q \cdots F_2 F_1$$

einer beliebigen $m \times n$ -Matrix A mit Rang l als Produkt von Elementarmatrizen E_1, \dots, E_p , F_1, \dots, F_q (andere als oben) und der Blockmatrix aus einem Block \mathbb{I}_l und Nullblöcken sonst. Im Fall $m = n = l$ ist diese Blockmatrix die Einheitsmatrix \mathbb{I}_n , und daher gilt:

- Jede $n \times n$ -Matrix von maximalem Rang $l = n$ lässt sich als Produkt von $n \times n$ -Elementarmatrizen schreiben.

6) Potenzen einer Matrix A kann man bilden, wenn sie quadratisch ist, also $A^1 := A$, $A^2 := AA$, $A^3 := AAA$ usw. Es gelten dann die Potenzgesetze $A^k A^l = A^{k+l}$ und $(A^k)^l = A^{kl}$ für natürliche Exponenten (aber nicht $(BA)^k = B^k A^k$, wenn $BA \neq AB$ ist, weil man dann z.B. in $(BA)^2 = BABA$ die beiden mittleren Faktoren nicht vertauschen darf; man definiert auch noch $A^0 := \mathbb{I}$ als die Einheitsmatrix desselben Formats wie A). Zum Beispiel gilt:

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \implies A^2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, A^3 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \dots, A^k = A \text{ für alle } k \in \mathbb{N};$$

Matrizen wie diese, deren Potenzen alle gleich der Matrix selbst sind, heißen *idempotente Matrizen*. Ein anderes Beispiel, in dem bei der Potenzbildung auch etwas Unerwartetes passiert, ist:

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \implies A^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbb{I}_2, A^3 = A, A^4 = \mathbb{I}_2, A^5 = A, A^6 = \mathbb{I}_2, \dots$$

Matrizen wie diese, deren Quadrat die Einheitsmatrix ist, heißen *involutorische Matrizen*. Ein drittes Beispiel:

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \implies A^2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, A^3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \implies A^k = 0 \text{ für } k \geq 3.$$

Hier haben wir eine *nilpotente Matrix*, d.h. eine, die nicht selbst die Nullmatrix ist, aber eine Nullmatrix als Potenz hat. Die $n \times n$ -Matrix A mit Einträgen $a_{i,i+1} = 1$ in der ersten Schrägzeile oberhalb der Diagonalen und Nulleinträgen sonst, hat allgemein den *Nilpotenzgrad* $n-1$, d.h. es gilt $A^k = 0$ für $k \geq n$, aber $A^k \neq 0$ für $1 \leq k < n$.

7) Bei sog. **gestaffelten Produktionsabläufen** werden gewisse Endprodukte Nr. $1 \dots n$ aus Zwischenprodukten Nr. $1 \dots m$ gefertigt gemäß einer Endproduktionsverflechtungsmatrix $S = (s_{ij})$, und die Zwischenprodukte werden ihrerseits aus $1 \dots l$ Rohstoffen hergestellt gemäß einer Zwischenproduktionsverflechtungsmatrix $R = (r_{hi})$. Die Einträge r_{hi} von R heißen *Rohstoffverbrauchskoeffizienten* und geben an, wieviele Mengeneinheiten des h -ten Rohstoffs für die Herstellung einer Einheit des Zwischenprodukts Nr. i benötigt werden. Entsprechend beschreiben die *Zwischenproduktverbrauchskoeffizienten* s_{ij} , wieviele Einheiten des i -ten Zwischenprodukts man zur Produktion einer Einheit des Endprodukts Nr. j braucht. Man kann dabei auch zulassen, dass ein Rohstoff direkt in die Endproduktion eingeht, indem man ihn einfach als Zwischenprodukt mit Verbrauchskoeffizient 1 für den betreffenden Rohstoff und mit Verbrauchskoeffizient 0 für die anderen Rohstoffe führt. Jedoch wollen wir hier den Fall ausschließen, dass ein Endprodukt in der Produktion von Zwischenprodukten oder von Endprodukten benötigt wird. (Eine derartige Situation haben wir in 3.2 für eine Produktion ohne Staffelung behandelt und ein lineares Gleichungssystem für den Produktionsvektor aufgestellt; ähnlich könnte man auch hier bei der gestaffelten Produktion verfahren, doch wird dann alles noch komplizierter.) Ist $\mathbf{x} \in \mathbb{R}^n$ ein gegebener *Outputvektor der Endproduktion (Produktionsplan)*, dessen Komponenten x_j die zu produzierenden Mengeneinheiten des j -ten Endprodukts angeben, so ist $y_i = s_{i1}x_1 + s_{i2}x_2 + \dots + s_{in}x_n$ die für die Endproduktion bereitzustellende Menge des i -ten Zwischenprodukts. Die y_i sind die Komponenten des *Inputvektors für die Endproduktion* $\mathbf{y} \in \mathbb{R}^m$, der gleichzeitig der *Outputvektor der Zwischenproduktion* ist (wenn ohne Überschuss zwischenproduziert wird). In Matrixschreibweise kann der lineare Zusammenhang zwischen \mathbf{x} und \mathbf{y} durch $\mathbf{y} = S\mathbf{x}$ ausgedrückt werden. Entsprechend ist der Bedarf z_h an Einheiten des h -ten Rohstoffs für die Zwischenproduktion gegeben durch $z_h = r_{h1}y_1 + r_{h2}y_2 + \dots + r_{hm}y_m$, und der *Rohstoffverbrauchsvektor (Rohstoffbedarfsvektor)* $\mathbf{z} \in \mathbb{R}^l$, also der Inputvektor für die Zwischenproduktion, berechnet sich gemäß $\mathbf{z} = R\mathbf{y}$ in linearer Weise aus \mathbf{y} . Da \mathbf{y} seinerseits linear von \mathbf{x} abhängt, ist \mathbf{z} eine mittelbar von \mathbf{x} linear abhängige Variable, d.h. man erhält \mathbf{z} aus \mathbf{x} durch Hintereinanderschaltung von zwei linearen Operatoren: $\mathbf{z} = R\mathbf{y} = R(S\mathbf{x})$. Nun ist das Matrixprodukt RS gerade so motiviert und definiert worden, dass $R(S\mathbf{x}) = (RS)\mathbf{x}$ gilt, d.h. die Hintereinanderausführung der beiden Multiplikationen mit S und R kann ersetzt werden durch eine einzige lineare Operation, nämlich die Multiplikation mit RS .

- Der unmittelbare lineare Zusammenhang zwischen dem Endproduktoutputvektor \mathbf{x} und dem Rohstoffverbrauchsvektor \mathbf{z} ist gegeben durch das Produkt der Zwischenproduktionsmatrix R mit der Endproduktionsmatrix S , d.h. $\mathbf{z} = (RS)\mathbf{x}$.

Die Überlegungen lassen sich ausdehnen auf einen *mehrstufigen Produktionsprozess*, bei dem man $k \geq 2$ Produktionsstufen hat, derart dass der Outputvektor \mathbf{x}_{h-1} der $(h-1)$ -ten Stufe jeweils der Inputvektor (Bedarfsvektor) für die Produktion der nächsten Stufe h ist. Der Inputvektor für die erste Produktionsstufe ist dann der Rohstoffverbrauchsvektor \mathbf{x}_0 , und der Outputvektor der letzten Stufe ist der Endproduktoutputvektor \mathbf{x}_k . Man hat dann k Produktionsverflechtungsmatrizen A_1, A_2, \dots, A_k , wobei die Einträge von $A_h = (a_{ij}^{(h)})$ angeben, wieviele Einheiten des i -ten Produkts der Stufe $h-1$ für die Herstellung einer Einheit des j -ten Produkts der Stufe h erforderlich sind. Der Zusammenhang zwischen \mathbf{x}_h und \mathbf{x}_{h-1} ist also wie oben $\mathbf{x}_{h-1} = A_h\mathbf{x}_h$ für $h = 1 \dots k$. Somit gilt

$$\mathbf{x}_0 = A_1\mathbf{x}_1 = A_1A_2\mathbf{x}_2 = \dots = A_1A_2 \cdots A_k\mathbf{x}_k,$$

d.h. der unmittelbare lineare Zusammenhang zwischen dem Endproduktoutputvektor \mathbf{x}_k und dem Rohstoffbedarfsvektor \mathbf{x}_0 ist durch das Produkt aller k Produktionsmatrizen (in der Reihenfolge von links nach rechts aufsteigender Produktionsstufen) gegeben.

8) In analoger Weise kommen, wie schon gesagt, Matrixprodukte überall ins Spiel, wo (vektorielle) ökonomische Größen mittelbar über eine Kette von linearen Abhängigkeitsgesetzen durch andere ökonomische (vektorielle) Variablen bestimmt sind. Noch ein Beispiel dieser Art: Wird die *Kundenwanderung zwischen n konkurrierenden Produkten* in einer ersten Beobachtungsperiode durch die Übergangsmatrix $P_1 \in \mathbb{R}^{n \times n}$ beschrieben und in den folgenden Perioden durch $P_2, \dots, P_k \in \mathbb{R}^{n \times n}$, so ist das Matrixprodukt $P = P_k \cdots P_2 P_1$ die Übergangsmatrix, welche den Kundenwechsel im gesamten Zeitraum vom Beginn der ersten bis zum Ende der k -ten Beobachtungsperiode darstellt. (Daraus folgt übrigens, dass ein Produkt von stochastischen Matrizen wie die P_h hier, also Matrizen mit der Zeilensumme 1, wieder eine stochastische Matrix ist. Das sieht man auch durch Multiplikation der Matrizen von links an einen summierenden Spaltenvektor, also einen mit lauter Einträgen 1; denn Stochastizität der Matrix bedeutet gerade, dass ein solches Matrix-an-Vektor-Produkt als Ergebnis wieder einen summierenden Spaltenvektor hat.)

9) Matrixprodukte treten in ökonomischem Zusammenhang auch dann auf, wenn man mehrere Vorgänge zusammenfasst, die durch Multiplikation einer festen Matrix an verschiedene Spaltenvektoren beschrieben werden. Sind z.B. in einer Produktion für k verschiedene Perioden Produktionsoutputvektoren $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \in \mathbb{R}^n$ vorgesehen und die zugehörigen Rohstoffverbrauchsvektoren $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_k \in \mathbb{R}^m$ berechnet, so erhebt sich die Frage nach dem *günstigsten Rohstoffeinkauf*. Kommen dafür Lieferanten L_1, \dots, L_l in Frage, wobei der h -te Lieferant für den i -ten Rohstoff den Preis p_{hi} verlangt, so gibt das Produkt der *Preismatrix* $P = (p_{hi})$ mit einem Rohstoffverbrauchsvektor \mathbf{z} den zugehörigen *Rohstoffkostenvektor* $P\mathbf{z}$, dessen h -te Komponente $p_{h1}z_1 + \dots + p_{hm}z_m$ die Kosten beim Bezug aller in \mathbf{z} angegebenen Rohstoffmengen vom Lieferanten L_h angibt für $h = 1 \dots l$.

Hier kann man nun die aus dem Produktionsplan resultierenden Rohstoffverbrauchsvektoren $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_k$ als Spalten zu einer $m \times k$ -*Verbrauchsmatrix* Z zusammenstellen. Dann gibt die j -te Eintragung in der h -ten Zeile des Matrixprodukts PZ die Kosten der für die Realisierung des j -ten Produktionsoutputvektors benötigten Rohstoffe an, wenn diese alle beim Lieferanten L_h bezogen werden. Unterstellt, der Bezug aller Rohstoffe bei einem einzigen Lieferanten ist (wegen Rabatten bei größeren Bestellungen) in jeder Periode günstiger als die Beschaffung bei verschiedenen Lieferanten, so wird man also in jeder Spalte von PZ die kleinste Zahl aufsuchen und dann, wenn diese bei der j -ten Spalte etwa in h -ter Position steht, die Rohstoffe für die j -te Produktionsperiode beim Lieferanten L_h ordern.

Bei Beispielen dieser Art ist die Verwendung eines Matrixprodukts allerdings weder notwendig, noch besonders hilfreich — man kann genau so gut die einzelnen Vorgänge für jede Periode separat durchrechnen. Die mathematische Grundlage der *Zusammenfassung von mehreren Matrix-an-Vektor-Multiplikationen zu einer Matrix-Matrix-Multiplikation* ist folgende schon früher gemachte Beobachtung:

- Die j -te Spalte des Matrixprodukts BA ist das Produkt der Matrix B mit der j -ten Spalte von A ;
- die i -te Zeile des Matrixprodukts BA ist das Produkt der i -ten Zeile von B mit der Matrix A .

10) Ein letztes Beispiel zur Matrixmultiplikation aus der Ökonomie: In einem Unternehmen werden an Produktionsstätten $\textcircled{1}, \dots, \textcircled{m}$ Produkte $\boxed{1}, \dots, \boxed{n}$ gefertigt gemäß einer $m \times n$ -*Mengenproduktionsmatrix* $X = (x_{ij})$, wobei x_{ij} die Zahl der von Produkt Nr. j an der Produktionsstätte Nr. i hergestellten Einheiten ist. Dazu gehört die gleichfor-

matige *Kostenproduktionsmatrix* $K = (k_{ij})$, wobei der *Kostenfaktor* k_{ij} die Kosten der Produktion einer Einheit des Produkts Nr. j an der Produktionsstätte (i) angibt, also die Stückkosten für das Produkt an dieser Produktionsstätte. Die Summe

$$k_{i1}x_{i1} + k_{i2}x_{i2} + \dots + k_{in}x_{in} = \sum_{j=1}^n k_{ij}x_{ij}$$

gibt dann die gesamten Kosten der Produktionsstätte (i) an, wenn gemäß dem durch X gegebenen Produktionsplan produziert wird.

Um diese Summe als Eintrag in einem Matrixprodukt zu interpretieren bilden wir die **transponierte Matrix** ("gekippte Matrix")

$$K^T = \begin{pmatrix} k_{11} & k_{21} & \dots & k_{m1} \\ k_{12} & k_{22} & \dots & k_{m2} \\ \vdots & \vdots & & \vdots \\ k_{1n} & k_{2n} & \dots & k_{mn} \end{pmatrix} \quad \text{zu} \quad K = \begin{pmatrix} k_{11} & k_{12} & \dots & k_{1n} \\ k_{21} & k_{22} & \dots & k_{2n} \\ \vdots & \vdots & & \vdots \\ k_{m1} & k_{m2} & \dots & k_{mn} \end{pmatrix}.$$

Die Zeilen von K^T sind also die Spalten von K , und die Spalten von K^T sind die Zeilen von K . (Hat K das Format (m, n) so ist K^T eine Matrix vom Format (n, m) . Nur bei quadratischen Matrizen K hat K^T dasselbe Format. Ist sogar $K = K^T$, also $k_{ij} = k_{ji}$ für alle i, j , so heißt K eine – zur Diagonalen – **symmetrische Matrix**. Andere gebräuchliche Bezeichnungen für die transponierte Matrix zu K sind K^t oder K' ; manchmal wird das Superskript auch vorangestellt, also ${}^T K$ bzw. ${}^t K$ geschrieben.)

Nun ist das Matrixprodukt

$$XK^T \in \mathbb{R}^{m \times m}$$

definiert, und sein i -ter Eintrag auf der Diagonalen ist gerade die obige Summe $\sum_{j=1}^n x_{ij}k_{ij}$, welche die Kosten der Produktionsstätte (i) angibt. In diesem Sinne können wir sagen;

- Die *Diagonaleinträge* der $m \times m$ -Matrix XK^T *schlüsseln die Kosten auf die m Produktionsstätten auf.*

(Die in nichtdiagonalen Positionen (h, i) befindlichen Einträge der Matrix sind weniger relevant; sie geben die Kosten an der Produktionsstätte (i) an, wenn dort nach dem eigentlich für (h) vorgesehenen Produktionsplan produziert würde.) Auch in der anderen Reihenfolge ist das Produkt von X und K^T definiert, nun allerdings eine $n \times n$ -Matrix:

$$K^T X \in \mathbb{R}^{n \times n}.$$

Der j -te Diagonaleintrag dieser Matrix ist

$$k_{1j}x_{1j} + k_{2j}x_{2j} + \dots + k_{mj}x_{mj} = \sum_{i=1}^m k_{ij}x_{ij}$$

und gibt die Kosten für die Erzeugung des Produkts Nr. j saldiert über alle m Produktionsstätten an. In diesem Sinne gilt also:

- Die *Diagonaleinträge* der $n \times n$ -Matrix $K^T X$ *schlüsseln die Kosten auf die n Produkte auf.*

(Die Nichtdiagonaleinträge beschreiben hier, was die Produktion des Produkts kosten würde, wenn man dafür die Stückkosten eines anderen Produkts hätte; diese Einträge haben also keine ökonomische Bedeutung.)

Die Summe aller Diagonaleinträge ist bei XK^T und bei $K^T X$ dieselbe,

$$\sum_{i=1}^m \left(\sum_{j=1}^n x_{ij} k_{ij} \right) = \sum_{j=1}^n \left(\sum_{i=1}^m k_{ij} x_{ij} \right),$$

und das ist auch ökonomisch klar; denn beide Summen geben ja die *Gesamtkosten* der Produktion an, nur einmal aufgliedert in die Kosten der einzelnen Produktionsstätten und einmal in die Gesamtkosten für die einzelnen Produkte. Man nennt die Summe der Diagonaleinträge einer quadratischen Matrix ihre *Spur* und kann daher formulieren

- Die Spur von XK^T und die von $K^T X$ geben die Gesamtkosten $\sum_{i=1}^m \sum_{j=1}^n k_{ij} x_{ij}$ der Produktion an. ■

DEFINITION: Die **Spur einer quadratischen Matrix** $A = (a_{ij})_{1 \leq i, j \leq n}$ ist die Summe ihrer Diagonaleinträge, $\text{Spur}(A) := a_{11} + a_{22} + \dots + a_{nn} = \sum_{i=1}^n a_{ii}$. ■

DISKUSSION: 1) In den Überlegungen im letzten Beispiel war X eine eigentlich ganz beliebige $m \times n$ -Matrix und K^T eine beliebige $n \times m$ -Matrix. Daher zeigen diese Überlegungen ganz allgemein den **Spursatz**:

$$\text{Spur}(AB) = \text{Spur}(BA) \quad \text{für alle } A \in \mathbb{R}^{m \times n} \text{ und } B \in \mathbb{R}^{n \times m}.$$

Obwohl also im Allgemeinen $AB \neq BA$ ist, im Fall $m \neq n$ haben diese beiden quadratischen Matrizen ja sogar verschiedenes Format, besitzen AB und BA doch stets dieselbe Summe von Diagonaleinträgen, wenn beide Matrixprodukte definiert sind. In der Tat, ist $A = (a_{ij})$ und $B = (b_{jk})$, so hat AB die Diagonaleinträge $\sum_{j=1}^n a_{ij} b_{ji}$ für $i = 1 \dots m$ und BA die Diagonaleinträge $\sum_{i=1}^m b_{ji} a_{ij}$ für $j = 1 \dots n$, und die Summe aller Diagonaleinträge ist in beiden Fällen

$$\text{Spur}(AB) = \sum_{i=1}^m \sum_{j=1}^n a_{ij} b_{ji}.$$

2) Bildet man zu einer Matrix $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ die Spur des Produkts AA^T , so erhält man die Quadratsumme aller Einträge der Matrix:

$$\text{Spur}(AA^T) = \sum_{i=1}^m \sum_{j=1}^n (a_{ij})^2.$$

Daher ist die Wurzel aus der Spur ein Analogon der Euklidischen Norm von Vektoren aus \mathbb{R}^n , die ja als Quadratwurzel aus der Quadratsumme der Komponenten definiert ist. Als ein sinnvolles Maß für die "absolute Gesamtgröße" einer Matrix (eines von vielen verschiedenen in der Mathematik gebräuchlichen) definieren wir daher die **Euklidische Norm einer Matrix** als Quadratwurzel aus der Summe der Quadrate aller Einträge:

$$|A| := \sqrt{\text{Spur}(AA^T)} = \sqrt{\sum_{i=1}^m \sum_{j=1}^n (a_{ij})^2}.$$

Man kann durch $A \bullet B := \text{Spur}(AB^T) = \sum_{i=1}^m \sum_{j=1}^n a_{ij} b_{ji}$ auch ein Skalarprodukt auf dem Raum $\mathbb{R}^{m \times n}$ der $m \times n$ -Matrizen einführen, das zum üblichen Skalarprodukt auf \mathbb{R}^n analog ist. Dann gilt auch hier, dass die Norm die Wurzel aus dem Skalarprodukt einer Matrix mit sich selbst ist, $|A| = \sqrt{A \bullet A}$. ■

3.4 Inverse Matrix und Determinante

In diesem Abschnitt behandeln wir nur *quadratische* Matrizen (von einigen Nebenbemerkungen abgesehen). Zur Motivation der Inversenbildung bei Matrizen betrachten wir ein lineares Gleichungssystem

$$Ax = y$$

mit Koeffizientenmatrix $A \in \mathbb{R}^{n \times n}$ und Inhomogenität $y \in \mathbb{R}^n$, also ein System mit ebenso vielen Gleichungen wie Unbekannten $x = (x_1, \dots, x_n)$. Angenommen wir kennen eine Matrix $B \in \mathbb{R}^{n \times n}$ mit

$$BA = \mathbb{I} \quad \text{und} \quad AB = \mathbb{I},$$

wobei \mathbb{I} die $n \times n$ -Einheitsmatrix ist. Dann folgt aus $BA = \mathbb{I}$ durch Multiplikation der Gleichung $y = Ax$ von links mit B die Gleichung $By = B(Ax) = (BA)x = \mathbb{I}x = x$, also kann nur $x = By$ die Lösung unseres Gleichungssystems sein. Dies heißt noch nicht, dass $x = By$ wirklich Lösung ist, aber das folgt aus $AB = \mathbb{I}$ mit $Ax = A(By) = (AB)y = \mathbb{I}y = y$. Also ist tatsächlich für jede rechte Seite $y \in \mathbb{R}^n$

$$x = By \quad \text{die eindeutige Lösung des Gleichungssystems} \quad Ax = y.$$

Das ist sehr praktisch; denn man kann dann die Lösung für jede rechte Seite y durch eine einfache Matrix-an-Vektor-Multiplikation ausrechnen, nämlich durch Bildung des Produkts By , und das ist im Allgemeinen viel einfacher als die Lösung des Gleichungssystems mit dem Eliminationsverfahren (Herstellung der Zeilen-Stufen-Form). Es klappt freilich nur, wenn wir auch eine Matrix B mit $BA = \mathbb{I} = AB$ kennen, und daher machen wir erst einmal eine

DEFINITION: Eine quadratische Matrix A heißt **invertierbare Matrix** oder **reguläre Matrix**, wenn es eine gleichformatige Matrix B gibt mit $BA = \mathbb{I} = AB$. Dann heißt B die zu A **inverse Matrix** und wird A^{-1} notiert. Ist A nicht invertierbar, so heißt A auch **singuläre Matrix**. ■

Man darf von *der* inversen Matrix sprechen, weil es nur eine einzige geben kann; denn aus $BA = \mathbb{I}$ und $A\tilde{B} = \mathbb{I}$ folgt die Gleichheit $B = B\mathbb{I} = B(A\tilde{B}) = (BA)\tilde{B} = \mathbb{I}\tilde{B} = \tilde{B}$. Die Notation A^{-1} für die Inverse ist sinnvoll, weil dann gilt $A^{-1}A^1 = A^0 = A^1A^{-1}$ wie bei einem Potenzgesetz (wobei $A^1 := A$ und $A^0 := \mathbb{I}$ gesetzt ist).

DISKUSSION: 1) Die Bedeutung der inversen Matrix für die Lösung linearer Gleichungssysteme haben wir vor der Definition schon gesehen:

- *Ist die Koeffizientenmatrix des linearen Gleichungssystems $Ax = y$ von n Gleichungen für n Unbekannte unvertierbar, so ist $x = A^{-1}y$ die eindeutige Lösung.*

Insbesondere liegt der "gute" Fall der eindeutigen Lösbarkeit für jede rechte Seite vor, wenn eine Inverse zur Koeffizientenmatrix existiert. Und man erhält die Lösung durch eine einfache Matrix-an-Vektor-Multiplikation, wenn man die inverse Matrix A^{-1} kennt.

2) *Nur quadratische Matrizen können eine Inverse haben;* denn wenn $BA = \mathbb{I} = AB$ ist, so sind die Produkte AB und BA beide definiert und von gleichem Format, und das geht nur, wenn A, B gleichformatige quadratische Matrizen sind. Aber natürlich ist nicht *jede* quadratische Matrix A invertierbar!

3) Bei quadratischen Matrizen $A, B \in \mathbb{R}^{n \times n}$ genügt es, *eine* der beiden Gleichungen $AB = \mathbb{I}$ bzw. $BA = \mathbb{I}$ nachzuweisen; dann folgt schon die jeweils andere Gleichung, und B ist die Inverse zu A . Ist z.B. $BA = \mathbb{I}$, so hat das homogene Gleichungssystem $Ax = 0$ nur die triviale Lösung $x = \mathbb{I}x = (BA)x = B(Ax) = B0 = 0$. Dann aber wissen wir aus 3.2, weil wir gleich viele Gleichungen und Unbekannte haben, dass das Gleichungssystem $Ax = y$ für jede rechte Seite $y \in \mathbb{R}^n$ eindeutig lösbar ist, und wegen $x = \mathbb{I}x = (BA)x = B(Ax) = By$ muss $x = By$ diese eindeutige Lösung sein, d.h. es gilt auch $(AB)y = A(By) = y$ für alle $y \in \mathbb{R}^n$ und damit $AB = \mathbb{I}$. Ist andererseits $AB = \mathbb{I}$ so wenden wir das Argument auf B statt A an und erhalten auch $BA = \mathbb{I}$.

4) Die inverse Matrix entspricht dem **linearen Umkehroperator** zu einem linearen Operator $T: \mathbb{R}^n \ni x \mapsto Ax \in \mathbb{R}^n$. Dass T eine Umkehrabbildung T^{-1} besitzt, heißt ja, dass es zu jedem $y \in \mathbb{R}^n$ genau ein $x \in \mathbb{R}^n$ gibt mit $T(x) = y$, und dann ist $T^{-1}(y) = x$. Besitzt A eine inverse Matrix A^{-1} , so ist $x = A^{-1}y$ die eindeutige Lösung zu $T(x) = Ax = y$, also existiert die Umkehrabbildung und ist durch $T^{-1}(y) = A^{-1}y$ gegeben. Ist andererseits T^{-1} Umkehrabbildung zu T , so ist T^{-1} auch wieder linear (denn $T^{-1}(sy) = sT^{-1}(y)$ gilt, weil $T(sx) = sy$ ist, und $T^{-1}(y+\tilde{y}) = T^{-1}(y) + T^{-1}(\tilde{y})$ gilt, weil $T(x+\tilde{x}) = y+\tilde{y}$ ist, wenn $T(x) = y$ und $T(\tilde{x}) = \tilde{y}$). Also wird T^{-1} durch eine Matrix $B \in \mathbb{R}^{n \times n}$ dargestellt, $T^{-1}(y) = By$ für alle y . Das aber heißt $BAx = T^{-1}(Ax) = T^{-1}(T(x)) = x$ für alle $x \in \mathbb{R}^n$, also ist $BA = \mathbb{I}_n$ und B inverse Matrix zu A .

- Eine inverse Matrix zu $A \in \mathbb{R}^{n \times n}$ existiert genau dann, wenn der durch A dargestellte lineare Operator $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$ umkehrbar ist; der inverse lineare Operator T^{-1} wird dann durch die inverse Matrix A^{-1} dargestellt.

5) **Rechenregeln für die Inversenbildung:** Die invertierbaren $n \times n$ -Matrizen bilden eine Gruppe, d.h. mit $A, B \in \mathbb{R}^{n \times n}$ sind auch A^{-1} , AB invertierbar. Außerdem sind mit A auch die transponierte Matrix A^T und Vielfache rA mit $r \in \mathbb{R}_{\neq 0}$ invertierbar. Es gilt:

$$(A^{-1})^{-1} = A, \quad (AB)^{-1} = B^{-1}A^{-1}, \quad (A^T)^{-1} = (A^{-1})^T, \quad (rA)^{-1} = \frac{1}{r}A^{-1} \quad (r \neq 0);$$

die Inverse der inversen Matrix ist also die Ausgangsmatrix, die Inverse eines Matrixprodukts ist das Produkt der inversen Faktoren, aber *in umgekehrter Reihenfolge* (!), das Inverse der transponierten Matrix ist die Transponierte der Inversen. Die erste Regel besagt $AA^{-1} = \mathbb{I}$ und $A^{-1}A = \mathbb{I}$ und ist klar ebenso wie die letzte. Die zweite folgt aus $(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}\mathbb{I}B = B^{-1}B = \mathbb{I}$. Die dritte beruht auf

$$(BA)^T = A^T B^T,$$

was man sich leicht überlegt. (Der Eintrag des letzten Produkts in Position (i, j) ist das Skalarprodukt der i -ten Spalte von A mit der j -ten Zeile von B , weil ja die Spalten der transponierten Matrix die Zeilen der ursprünglichen Matrix sind und die Zeilen der transponierten Matrix die Spalten der ursprünglichen. Dieses Skalarprodukt ist aber gerade der Eintrag von BA in Position (j, i) , also der von $(BA)^T$ in Position (i, j) .) Da offenbar $\mathbb{I}^T = \mathbb{I}$ ist, folgt daraus $(A^{-1})^T A^T = (AA^{-1})^T = \mathbb{I}^T = \mathbb{I}$.

6) *Matrixpotenzen mit ganzen Exponenten* A^k erklärt man für $k \in \mathbb{N}$ durch Multiplikation von k Matrixfaktoren A und, wenn A invertierbare Matrix ist, für $k = -l \in \mathbb{Z}_{<0}$ durch Multiplikation von l Matrixfaktoren A^{-1} ; für $k = 0$ wird noch $A^0 := \mathbb{I}$ gesetzt. Es gelten dann die üblichen Potenzgesetze

$$A^k A^l = A^{k+l}, \quad (A^k)^l = A^{kl} \quad \text{für alle } k, l \in \mathbb{Z},$$

aber $(AB)^k = A^k B^k$ ist im Allgemeinen nicht richtig, wenn $AB \neq BA$ ist; für $k = -1$ gilt sogar $(AB)^{-1} = B^{-1}A^{-1}$ gerade mit der umgekehrten Reihenfolge der Faktoren rechts. Eine *Division von Matrizen ist nicht erklärt*, selbst wenn beide quadratisch sind und der Nenner A invertierbar ist. Das Problem ist, dass man hier zwei mögliche Definitionen $A^{-1}B$ und BA^{-1} des Quotienten " $\frac{B}{A}$ " hat, die im Allgemeinen verschiedene Ergebnisse liefern, und dass für keine der beiden Definitionen vernünftige Rechenregeln gelten, die man von einer Quotientenbildung fordern würde. Zum Beispiel sollte $A \frac{B}{A} = B = \frac{B}{A} A$ gelten, aber $ABA^{-1} = B$ bzw. $A^{-1}BA = B$ stimmt nur, wenn $AB = BA$ ist, und die Matrixmultiplikation bei $n \times n$ -Matrizen ist ja (für $n \geq 2$) nicht kommutativ.

7) Für allgemeine $m \times n$ -Matrizen A heißt $B \in \mathbb{R}^{n \times m}$ eine **Linksinverse**, wenn $BA = \mathbb{I}_n$ ist, und $C \in \mathbb{R}^{n \times m}$ eine **Rechtsinverse** wenn $AC = \mathbb{I}_m$ ist. Existiert eine Linksinverse B zu A , so ist das Gleichungssystem $Ax = 0$ nur trivial lösbar, weil $x = \mathbb{I}_n x = (BA)x = B(Ax) = B0 = 0$ folgt; also muss gemäß 3.2 dann $m \geq n$ sein (mindestens so viele homogene Gleichungen wie Unbekannte). Existiert aber eine Rechtsinverse C , so gilt $A(Cy) = (AC)y = \mathbb{I}_m y = y$, d.h. das Gleichungssystem $Ax = y$ hat für jede rechte Seite $y \in \mathbb{R}^m$ eine Lösung $x = Cy$; gemäß 3.2 muss dann $m \leq n$ sein (höchstens so viele Gleichungen wie Unbekannte). Im Fall $m < n$ ist außerdem eine Rechtsinverse C — wenn eine existiert — nicht eindeutig bestimmt, weil man beliebige Lösungen $x \neq 0$ von $Ax = 0$ zu den Spalten von C addieren kann, ohne an der Gleichung $AC = \mathbb{I}_m$ etwas zu ändern. Entsprechend ist im Fall $m > n$ eine Linksinverse B nie eindeutig — wenn eine existiert —, weil man zu den Zeilen von B Lösungen $x \neq 0$ des homogenen Gleichungssystems $zA = 0$ von n linearen Gleichungen für m Unbekannte $z = (z_1 \dots z_m)$ addieren kann, ohne an der Gleichung $BA = \mathbb{I}_n$ etwas zu ändern. Nur im quadratischen Fall $m = n$ kann es also Rechtsinverse und Linksinverse zugleich geben, und dann zeigt 3), dass jede Rechtsinverse und jede Linksinverse automatisch schon die eindeutige beidseitige Inverse ist.

8) Eine linksinverse Matrix zu $A \in \mathbb{R}^{m \times n}$ kann man als Lösung $X = B$ der Matrixgleichung $XA = \mathbb{I}_n$ auffassen, eine rechtsinverse Matrix entsprechend als Lösung $X = C$ der Matrixgleichung $AX = \mathbb{I}_m$. Für allgemeine *lineare Matrixgleichungen* $XA = R$ bzw. $AX = S$ folgt, dass $X = RC$ bzw. $X = BS$ die Lösung sein muss, wenn es eine gibt. Ist A invertierbar und sind R, S quadratische Matrizen vom gleichen Format wie A , so hat $XA = R$ die eindeutige Lösung $X = RA^{-1}$ und $AX = S$ die eindeutige Lösung $X = A^{-1}S$. ■

FRAGEN, die sich nun aufdrängen, sind:

- 1) Wie sieht man, ob eine gegebene (quadratische!) Matrix A invertierbar ist?
- 2) Wie berechnet man dann die Inverse A^{-1} ?

Die Antwort auf beide Fragen liegt, wie wir sehen werden, in der *Herstellung der Zeilen-Stufen-Form* der Matrix mit Zeilentransformationen, und zwar genügt für die Beantwortung der ersten Frage eine gewöhnliche Zeilen-Stufen-Form, während man zur Berechnung der inversen Matrix noch weitere Zeilentransformationen durchführen muss bis zur kanonischen Zeilen-Stufen-Form, welche bei invertierbaren Matrizen nichts anderes als die Einheitsmatrix ist. ■

Im folgenden Satz geben wir verschiedene Kriterien für die Invertierbarkeit einer quadratischen Matrix an und beschreiben die Berechnung der Inversen genauer:

SATZ: 1) (Kriterien für Invertierbarkeit)

Für eine quadratische $n \times n$ -Matrix A sind folgende Aussagen äquivalent:

- (i) A ist regulär, d.h. die Inverse A^{-1} existiert;
- (ii) das homogene lineare Gleichungssystem $Ax = 0$ hat nur die triviale Lösung;
- (iii) das Gleichungssystem $Ax = y$ hat für jede rechte Seite $y \in \mathbb{R}^n$ eine Lösung;
- (iv) der Rang von A ist maximal, also gleich n ;
- (v) A hat als Zeilen-Stufen-Form eine $n \times n$ -Dreiecksmatrix mit Diagonaleinträgen $\neq 0$;
- (vi) A hat als kanonische Zeilen-Stufen-Form die Einheitsmatrix \mathbb{I}_n .

2) (Berechnung der Inversen) Ist A invertierbar, so erhält man die j -te Spalte von A^{-1} als Lösung des linearen Gleichungssystems $Ax = e_j$ mit dem j -ten kanonischen Basisvektor e_j von \mathbb{R}^n als rechte Seite ($j = 1 \dots n$). Man erhält die Inverse A^{-1} auch, indem man die Zeilenoperationen, die von A zur kanonischen Zeilen-Stufen-Form (vi) führen, in gleicher Reihenfolge auf die Einheitsmatrix \mathbb{I}_n anwendet.

Die Äquivalenz von (ii), (iii), (iv) und (v) für Systeme von n linearen Gleichungen mit n Unbekannten wissen wir schon aus 3.2. Aus einer Zeilen-Stufen-Form wie in (v) kann man offenbar durch weitere Zeilentransformationen die Einheitsmatrix \mathbb{I}_n erzeugen, daher ist (v) auch äquivalent zu (vi). Wir haben auch schon gesehen, dass die Existenz der inversen Matrix (i) gleichbedeutend ist mit der Umkehrbarkeit des linearen Operators $x \mapsto Ax$ auf \mathbb{R}^n , also mit der eindeutigen Lösbarkeit (ii),(iii) von $Ax = y$ für alle $y \in \mathbb{R}^n$. Damit ist Teil 1) des Satzes schon klar. Für Teil 2) erinnern wir uns, dass $A^{-1}e_j$ die j -te Spalte der Matrix A^{-1} ist. Andererseits gilt $A(A^{-1}e_j) = e_j$, also ist $A^{-1}e_j$ auch die Lösung zu $Ax = e_j$. Wenn man die erweiterte Koeffizientenmatrix $(A|y)$ durch Zeilenoperationen in die Form $(\mathbb{I}_n|z)$ umgeformt hat, so ist z die Lösung zu $Ax = y \iff \mathbb{I}_n x = z$. Daher ergibt die Anwendung dieser Zeilenoperationen auf e_j die j -te Spalte von A^{-1} , und die Anwendung auf \mathbb{I}_n , also auf die Matrix mit den Spalten e_1, \dots, e_n , gibt die Inverse A^{-1} .

DISKUSSION: 1) Aus dem Satz ergibt sich ein **Verfahren zur Feststellung der Invertierbarkeit und zur Berechnung der Inversen**: Man formt die aus A als linker Block und der Einheitsmatrix \mathbb{I}_n als rechter Block zusammengesetzte $n \times 2n$ -Matrix $(A|\mathbb{I}_n) \in \mathbb{R}^{n \times 2n}$ durch Zeilenoperationen so um, dass A auf Zeilen-Stufen-Form gebracht wird. Falls sich dabei eine Nullzeile in der linken $n \times n$ -Matrix ergibt, so hat A kleineren Zeilenrang als n , also existiert die Inverse nicht. Ergibt sich aber links eine Dreiecksmatrix mit Diagonaleinträgen $\neq 0$, so erzeugt man mit weiteren Zeilenoperationen in der $n \times 2n$ -Matrix links die Einheitsmatrix \mathbb{I}_n und erhält als rechten $n \times n$ -Block die Inverse $B = A^{-1}$.

$$\begin{array}{ccc}
 A = (a_{ij}) & \left(\begin{array}{cccc|cccc}
 a_{11} & a_{12} & \dots & a_{1n} & 1 & 0 & \dots & 0 \\
 a_{21} & a_{22} & \dots & a_{2n} & 0 & 1 & \dots & 0 \\
 \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
 a_{n1} & a_{n2} & \dots & a_{nn} & 0 & 0 & \dots & 1
 \end{array} \right) \\
 & \begin{array}{c} \downarrow \text{Zeilenoperationen} \downarrow \\ \end{array} \\
 & \left(\begin{array}{cccc|cccc}
 1 & 0 & \dots & 0 & b_{11} & b_{12} & \dots & b_{1n} \\
 0 & 1 & \dots & 0 & b_{21} & b_{22} & \dots & b_{2n} \\
 \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
 0 & 0 & \dots & 1 & b_{n1} & b_{n2} & \dots & b_{nn}
 \end{array} \right) & (b_{ij}) = A^{-1}
 \end{array}$$

Spaltenoperationen sind verboten bei diesem Berechnungsverfahren für die Inverse Matrix, auch Spaltenvertauschungen! Solche Operationen sind auch nicht erforderlich; denn die Zeilen-Stufen-Form von A kann man alleine mit Zeilenoperationen herstellen, und wenn sie keine Nullzeile hat, also hier eine Dreiecksmatrix ohne Nulleinträge auf der Diagonalen ist, so kann man mit weiteren Zeilenoperationen auch die Einheitsmatrix erreichen. Man kann das Verfahren auch verstehen, indem man die Zeilenoperationen durch Multiplikation von links mit entsprechenden Elementarmatrizen darstellt (siehe 3.3). Die Herstellung der Einheitsmatrix \mathbb{I} aus A durch Zeilenoperationen bedeutet demnach, dass es ein Produkt E von Elementarmatrizen gibt mit $EA = \mathbb{I}$. Dann ist $A^{-1} = E = E\mathbb{I}$, d.h. man erhält A^{-1} , indem man dieselben Zeilenoperationen in gleicher Reihenfolge auf die Einheitsmatrix anwendet. Hat man aber auch Spaltenoperationen bei A durchgeführt, um die Einheitsmatrix zu erhalten, so heißt das $EAF = \mathbb{I}$ mit einem weiteren Produkt F von Elementarmatrizen. Daraus folgt dann $E = F^{-1}A^{-1}$ und $A^{-1} = FE = FE\mathbb{I}$, d.h. man muss nach den durch E bestimmten Zeilenoperationen an der Einheitsmatrix noch weitere, durch Multiplikation mit F von links gegebene, Zeilenoperationen ausführen, um A^{-1} zu erhalten.

2) Kennt man die Inverse zu $A \in \mathbb{R}^{n \times n}$, so kann man die eindeutige Lösung eines linearen Gleichungssystems $Ax = b$ einfach berechnen, indem man A^{-1} an den Spaltenvektor b multipliziert. Das ist sicher weniger aufwendig, als das Gaußsche Eliminationsverfahren. Andererseits ist die Berechnung von A^{-1} aufwendig: Man hat dazu n lineare Gleichungssysteme $Ax = e_j$, $j = 1 \dots n$, zu lösen. Die Frage ist also: *Wann lohnt sich der Aufwand für die Berechnung von A^{-1} , wenn es um die Lösung vieler linearer Gleichungssysteme $Ax = b$ mit derselben Koeffizientenmatrix A geht?*

Die simultane Lösung von $Ax = \mathbf{b}_j$ für k verschiedene rechte Spalten $\mathbf{b}_1, \dots, \mathbf{b}_k$ mit dem Rechenschema aus 1) erfordert $\frac{1}{2}n(n+2k-1)$ Divisionen und $\frac{1}{2}n(n-1)(n+2k-1)$ Multiplikationen sowie $\frac{1}{2}n(n-1)(n+2k-1)$ Additionen im ungünstigsten Fall. (Man kann die erforderlichen Operationen zur Herstellung eines kanonischen Einheitsvektors in Spalte j zählen und mit der arithmetischen Summenformel für $j = 1 \dots n$ aufaddieren. Gibt man sich mit einer Dreiecksform zufrieden und löst die Gleichungssysteme dann von unten nach oben auf, so genügen $\frac{1}{6}n(n-1)(2n+6k-1)$ Multiplikationen und Additionen / Subtraktionen, das ist etwas weniger.) Kennt man die Inverse, so hat man für die Berechnung von k Produkten $A^{-1}\mathbf{b}_j$ offenbar kn^2 Multiplikationen und $kn(n-1)$ Additionen auszuführen. Der Aufwand für die Berechnung von A^{-1} ist nicht größer als der bei n Gleichungssystemen, erfordert also nicht mehr als $\frac{1}{2}n(3n-1)$ Divisionen und $\frac{1}{2}n(n-1)(3n-1)$ Additionen sowie $\frac{1}{2}n(n-1)(3n-1)$ Multiplikationen. Da die Division eine viel aufwendigere Rechenoperation ist als Multiplikation und Addition, erkennt man:

- *Sind mehrere lineare Gleichungssysteme $Ax = \mathbf{b}_j$ mit derselben invertierbaren Koeffizientenmatrix $A \in \mathbb{R}^{n \times n}$ zu lösen, so lohnt sich der Aufwand für die Berechnung der Inversen A^{-1} und der Lösungen als Matrixprodukt $A^{-1}\mathbf{b}_j$, wenn die Anzahl der zu lösenden Gleichungssysteme deutlich größer als n ist.*

Eine genauere Aussage darüber, ab welcher Zahl k von zu lösenden Gleichungssystemen sich die Inversenberechnung lohnt, ist nur möglich, wenn man quantitativ festlegt, wieviele Multiplikationen dem Rechenaufwand einer Division entsprechen. Werden (nicht realistisch) Divisionen und Multiplikationen als gleich aufwendig angesetzt, so kommt man auf eine Zahl $k \approx n^2$ erheblich größer als n . Je aufwendiger Divisionen im Verhältnis zu Multiplikationen tatsächlich sind, desto kleiner wird die Zahl $k \geq n$ von Gleichungssystemen, ab der sich die Lösungsberechnung mit Hilfe der inversen Matrix lohnt.

3) Kriterien für die Existenz von einseitigen Inversen zu einer gegebenen Matrix $A \in \mathbb{R}^{m \times n}$ kann man analog zum letzten Satz herleiten. Hat A eine Rechtsinverse C , also $AC = \mathbb{I}_m$, so ist $Ax = y$ für jedes $y \in \mathbb{R}^m$ lösbar, da man eine Lösung sofort in der Form $x = Cy$ angeben kann. Ist umgekehrt $Ax = y$ für alle y lösbar, so kann man aus Lösungen $x = c_i$ von $Ax = e_i$ zu den kanonischen Basisvektoren e_i von \mathbb{R}^m eine Matrix $C = (c_1 \dots c_m) \in \mathbb{R}^{n \times m}$ spaltenweise zusammenstellen und erhält $AC = \mathbb{I}_m$. Aus 3.2 wissen wir, dass Lösbarkeit von $Ax = y$ für alle $y \in \mathbb{R}^m$ genau dann gegeben ist, wenn A den Rang m hat (Zeilenrang oder Spaltenrang — am Ende von 3.2 haben wir ja gesehen, dass diese Ränge gleich sind). Für den linearen Operator $T(x) = Ax$ von \mathbb{R}^n nach \mathbb{R}^m ist gleichbedeutend, dass T *surjektiv* ist, d.h. zu jedem $y \in \mathbb{R}^m$ gibt es (mindestens) ein $x \in \mathbb{R}^n$ mit $T(x) = y$. Wir fassen zusammen:

- $A \in \mathbb{R}^{m \times n}$ besitzt genau dann eine Rechtsinverse C , d.h. $AC = \mathbb{I}_m$, wenn das lineare Gleichungssystem $Ax = y$ für jede rechte Seite $y \in \mathbb{R}^m$ lösbar ist, bzw. äquivalent, wenn $\text{Rang}(A) = m$ ist. (Dies ist nur im Fall $m \leq n$ möglich.)

Analog verläuft die Diskussion für Linksinverse $B \in \mathbb{R}^{n \times m}$ zu $A \in \mathbb{R}^{m \times n}$. Gilt $BA = \mathbb{I}_n$, so hat $Ax = 0$ nur die triviale Lösung, weil ja $x = BAx = B0 = 0$ folgt. Aus 3.2 wissen wir, dass dies genau im Fall $\text{Rang}(A) = n$ so ist. Für den linearen Operator $T(x) = Ax$ ist gleichbedeutend, dass er *injektiv* ist, d.h. $T(x) = T(\tilde{x})$ gilt nur im Fall $x = \tilde{x}$ (bzw. $T(x) = 0$ nur für $x = 0$; das ist äquivalent wegen $T(x) - T(\tilde{x}) = T(x - \tilde{x})$). Hat umgekehrt A den Rang n , so auch die transponierte Matrix A^T (weil Zeilenrang gleich Spaltenrang ist), also besitzt $A^T x = e_j$ Lösungen $x = c_j \in \mathbb{R}^m$ für jeden kanonischen Basisvektor e_j von \mathbb{R}^n . Stellen wir die c_j wieder zu einer $m \times n$ -Matrix C zusammen, so folgt $A^T C = \mathbb{I}_n$ und für $B := C^T$ dann $BA = C^T A = (A^T C)^T = \mathbb{I}_n^T = \mathbb{I}_n$, also ist B Linksinverse zu A . Wir fassen zusammen:

- $A \in \mathbb{R}^{m \times n}$ besitzt genau dann eine Linksinverse B , d.h. $BA = \mathbb{I}_n$, wenn das homogene lineare Gleichungssystem $Ax = 0$ nur die triviale Lösung hat, bzw. äquivalent, wenn $\text{Rang}(A) = n$ ist. (Dies ist nur im Fall $m \geq n$ möglich.) ■

BEISPIELE (zur Berechnung von inversen Matrizen):

1) Die Inverse zu einer 1×1 -Matrix (a) existiert genau dann, wenn der (einzige) Eintrag a nicht Null ist, und dann ist $A^{-1} = \left(\frac{1}{a}\right)$.

2) Die Inverse zu einer 2×2 -Matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ existiert genau dann, wenn ihre Determinante $ad - bc$ von Null verschieden ist. In 3.1 haben wir nämlich schon gesehen, dass genau dann das Gleichungssystem $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r \\ s \end{pmatrix}$ für alle rechten Seiten eindeutig lösbar ist. Die Lösung haben wir auch berechnet zu $\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{ad-bc} \begin{pmatrix} dr-bs \\ as-cr \end{pmatrix}$. Die Spalten der inversen Matrix erhalten wir, wenn wir für $\begin{pmatrix} r \\ s \end{pmatrix}$ die kanonischen Basisvektoren $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ und $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ einsetzen. Das Ergebnis ist:

- Die Inverse A^{-1} zu einer 2×2 -Matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ existiert genau dann, wenn ihre Determinante $ad - bc$ von Null verschieden ist, und dann gilt

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

Man erhält also die inverse Matrix, indem man in der Ausgangsmatrix die Diagonaleinträge vertauscht, die anderen Einträge mit -1 multipliziert und schließlich noch alle Einträge durch die Determinante der Ausgangsmatrix dividiert. Der letzte Schritt ist nur ausführbar, wenn die Determinante $\neq 0$ ist; sonst existiert die Inverse ja auch gar nicht.

Konkret ist z.B. $\begin{pmatrix} 1 & 2 \\ -2 & 3 \end{pmatrix}^{-1} = \frac{1}{1 \cdot 3 - 2 \cdot (-2)} \begin{pmatrix} 3 & -2 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 3/7 & -2/7 \\ 2/7 & 1/7 \end{pmatrix}$,

aber $\begin{pmatrix} 1 & 2 \\ -2 & -4 \end{pmatrix}^{-1}$ existiert nicht, weil die Determinante $1 \cdot (-4) - 2 \cdot (-2) = 0$ ist.

3) Wir wenden das Schema zur Berechnung von A^{-1} an auf $A := \begin{pmatrix} 1 & 4 & 3 \\ 2 & 5 & 4 \\ 1 & -3 & -2 \end{pmatrix}$:

$$\begin{array}{c} \left(\begin{array}{ccc|ccc} 1 & 4 & 3 & 1 & 0 & 0 \\ 2 & 5 & 4 & 0 & 1 & 0 \\ 1 & -3 & -2 & 0 & 0 & 1 \end{array} \right) \\ \downarrow \quad \textcircled{2} \rightarrow \textcircled{2} - 2 \cdot \textcircled{1}, \quad \textcircled{3} \rightarrow \textcircled{3} - \textcircled{1} \\ \left(\begin{array}{ccc|ccc} 1 & 4 & 3 & 1 & 0 & 0 \\ 0 & -3 & -2 & -2 & 1 & 0 \\ 0 & -7 & -5 & -1 & 0 & 1 \end{array} \right) \\ \downarrow \quad \textcircled{3} \rightarrow \textcircled{3} - 2 \cdot \textcircled{2}, \quad \textcircled{3} \rightarrow (-1)\textcircled{3}, \quad \textcircled{2} \leftrightarrow \textcircled{3} \\ \left(\begin{array}{ccc|ccc} 1 & 4 & 3 & 1 & 0 & 0 \\ 0 & 1 & 1 & -3 & 2 & -1 \\ 0 & -3 & -2 & -2 & 1 & 0 \end{array} \right) \\ \downarrow \quad \textcircled{1} \rightarrow \textcircled{1} - 4 \cdot \textcircled{2}, \quad \textcircled{3} \rightarrow \textcircled{3} + 3 \cdot \textcircled{2} \\ \left(\begin{array}{ccc|ccc} 1 & 0 & -1 & 13 & -8 & 4 \\ 0 & 1 & 1 & -3 & 2 & -1 \\ 0 & 0 & 1 & -11 & 7 & -3 \end{array} \right) \\ \downarrow \quad \textcircled{1} \rightarrow \textcircled{1} + \textcircled{3}, \quad \textcircled{2} \rightarrow \textcircled{2} - \textcircled{3} \\ \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 2 & -1 & 1 \\ 0 & 1 & 0 & 8 & -5 & 2 \\ 0 & 0 & 1 & -11 & 7 & -3 \end{array} \right) \quad \Rightarrow \quad A^{-1} = \begin{pmatrix} 2 & -1 & 1 \\ 8 & -5 & 2 \\ -11 & 7 & -3 \end{pmatrix}. \end{array}$$

Um jetzt z.B. das Gleichungssystem $Ax = b$ für die rechte Seite $b = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ zu lösen, rechnen wir einfach

$$Ax = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \Leftrightarrow \quad x = A^{-1} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 & -1 & 1 \\ 8 & -5 & 2 \\ -11 & 7 & -3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \\ -7 \end{pmatrix}.$$

Das hätten wir mit dem Eliminationsverfahren (Herstellung der Zeilen-Stufen-Form) natürlich schneller haben können. Der Vorteil ist jetzt aber, dass wir für jede andere rechte Seite $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$, $\begin{pmatrix} -1 \\ 2 \\ 5 \end{pmatrix}$, $\begin{pmatrix} \pi \\ \sqrt{2} \\ e \end{pmatrix}$, ... die Lösung nun genau so einfach durch Multiplikation mit der nun explizit bekannten Matrix A^{-1} von links erhalten.

4) Das Verfahren zur Inversenberechnung führt natürlich nicht immer zum Ziel; denn wenn keine Inverse existiert, so kann man auch keine berechnen — mit welchem Verfahren auch immer. Man erkennt diese Situation aber “auf dem Wege”, ohne dass vorher die Existenz der Inversen überprüft werden muss. Ein konkretes Beispiel dazu:

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 \\ -2 & 1 & 0 & 0 & 0 & 1 \end{array} \right) \longrightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 \\ 0 & 1 & 2 & 2 & 0 & 1 \end{array} \right) \longrightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 & -1 & 1 \end{array} \right).$$

Die ausgeführten Zeilenoperationen sind hier leicht zu erkennen. Da in der Diagonalen des linken 3×3 -Dreieck-Blocks der letzten Matrix ein Nulleintrag auftritt, existiert die Inverse zu der Matrix A (der linke 3×3 -Block der ersten Matrix) nicht. Das hätte man nach dem ersten Schritt schon sehen können: Da zwei gleiche Zeilen in der Koeffizientenmatrix entstanden sind, hat A höchstens zwei unabhängige Zeilen und somit kann der Rang nicht 3 sein, wie es bei invertierbaren 3×3 -Matrizen der Fall ist.

5) Die **Inverse einer Diagonalmatrix** existiert genau dann, wenn alle Diagonaleinträge $\neq 0$ sind, und dann ergibt sich die inverse Matrix einfach, indem man alle Diagonaleinträge durch ihr Reziprokes ersetzt. (*Warnung*: Das geht *nur* bei Diagonalmatrizen so!)

$$D = \begin{pmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ 0 & & & d_n \end{pmatrix} \implies D^{-1} = \begin{pmatrix} 1/d_1 & & & \\ & 1/d_2 & & \\ & & \ddots & \\ 0 & & & 1/d_n \end{pmatrix} \quad \text{wenn alle } d_i \neq 0.$$

Da bei der Multiplikation zweier Diagonalmatrizen die Diagonaleinträge in gleichen Positionen miteinander multipliziert werden, ist klar, dass das Produkt der rechten mit der linken Diagonalmatrix die Einheitsmatrix ist. Ein Spezialfall ist das *Inverse einer skalaren Matrix* $D = d\mathbb{I}$, also einer Diagonalmatrix mit lauter gleichen Diagonaleinträgen d . Ist $d \neq 0$, so ist die skalare Matrix $\frac{1}{d}\mathbb{I}$ die Inverse. Aus den Rechenregeln für die Matrixmultiplikation ist ja auch klar, dass $(\frac{1}{d}\mathbb{I})(d\mathbb{I}) = \frac{1}{d}d\mathbb{I}\mathbb{I} = \mathbb{I}$ ist. Die Einheitsmatrix \mathbb{I} ist ihre eigene Inverse, $\mathbb{I}^{-1} = \mathbb{I}$, und das gilt auch für ihr Negatives, $(-\mathbb{I})^{-1} = -\mathbb{I}$. (Es gibt noch andere Matrizen, die ihre eigene Inverse sind; siehe 8) unten.)

6) Die **Inverse einer Dreiecksmatrix** A existiert genau dann, wenn alle ihre Diagonaleinträge $\neq 0$ sind. Eine obere $n \times n$ -Dreiecksmatrix $A = (a_{ij})$ (also $a_{ij} = 0$ für $i > j$) hat nämlich schon Zeilen-Stufen-Form, und genau wenn $a_{ii} \neq 0$ gilt für alle i , so ist ihr Rang gleich n , also A^{-1} existent. Man kann dann die kanonische Zeilen-Stufen-Form leicht herstellen, und das Verfahren zur Berechnung der Inversen mit Zeilenoperationen zeigt, dass A^{-1} ebenfalls obere Dreiecksmatrix ist, wobei auf der Diagonalen von A^{-1} die Reziproken der Diagonaleinträge von A stehen. Für obere Dreiecksmatrizen gilt also schematisch:

$$A = \begin{pmatrix} a_{11} & & & * \\ & a_{22} & & \\ 0 & & \ddots & \\ & & & a_{nn} \end{pmatrix} \implies A^{-1} = \begin{pmatrix} 1/a_{11} & & & \tilde{*} \\ & 1/a_{22} & & \tilde{*} \\ 0 & & \ddots & \\ & & & 1/a_{nn} \end{pmatrix} \quad \text{wenn alle } a_{ii} \neq 0,$$

wobei “*” für irgendwelche Einträge oberhalb der Diagonalen in A steht und “ $\tilde{*}$ ” für irgendwelche (im Allgemeinen anderen) Einträge oberhalb der Diagonalen von A^{-1} . Es gibt auch eine allgemeine Formel dafür, wie man die Einträge in “ $\tilde{*}$ ” aus den Einträgen von A berechnet (dazu siehe auch 10) unten). Wir diskutieren diese Formel hier aber nicht, weil sie komplizierter ist als das Verfahren zur Inversenberechnung mit Zeilenoperationen.

Für untere Dreiecksmatrizen B ist der Sachverhalt übrigens völlig analog; die Inverse existiert, genau wenn keine Diagonaleinträge b_{ii} von B Null sind, und dann ist B^{-1} ebenfalls eine untere Dreiecksmatrix und hat Diagonaleinträge $1/b_{ii}$. Das kann man analog überlegen oder durch Übergang zu der transponierten (oberen) Dreiecksmatrix $A := B^T$ zeigen.

7) Die **Inversen von Elementarmatrizen sind wieder Elementarmatrizen**, und genauer gilt (mit den Bezeichnungen aus 3.3):

$$M_i(r)^{-1} = M_i(1/r), \quad A_{ij}(s)^{-1} = A_{ij}(-s), \quad V_{ij}^{-1} = V_{ij}.$$

Das ist sofort klar, wenn man die Elementarmatrizen als lineare Operatoren auffasst, die durch Multiplikation von links auf passende Spaltenvektoren x wirken. $M_i(r)$ ist die Diagonalmatrix mit Eintrag $r \neq 0$ in i -ter Diagonalposition und Einträgen 1 auf der Diagonalen sonst, und der entsprechende Operator wirkt auf x , indem die i -te Komponente von x mit dem Faktor r multipliziert wird. Dann ist natürlich $M_i(1/r)$ die Matrix des Umkehroperators, der die i -te Komponente mit $1/r$ multipliziert und so den ursprünglichen Spaltenvektor wieder herstellt. $A_{ij}(s)$ entsteht aus der Einheitsmatrix durch Ersetzung des Nulleintrags in Position (i, j) mit $i \neq j$ durch den Eintrag $s \in \mathbb{R}$ und wirkt auf Spaltenvektoren x , indem das s -fache der j -ten Komponente zur i -ten Komponente von x addiert wird. Um diese Operation rückgängig zu machen, muss man offenbar das s -fache der j -ten von der i -ten Komponente subtrahieren, also ist $A_{ij}(-s)$ die Matrix des Umkehroperators. V_{ij} entsteht aus der Einheitsmatrix durch Vertauschen der i -ten mit der j -ten Zeile für $i \neq j$, und Multiplikation von V_{ij} an x bewirkt eine Vertauschung der i -ten mit der j -ten Komponente von x . Wenn man diese Operation noch einmal anwendet, so erhält man also wieder x , und daher ist V_{ij} seine eigene Inverse.

Das Schema zur Inversenberechnung mit Zeilenoperationen ist übrigens nichts anderes als folgende Beobachtung: Wenn E_1, \dots, E_k Elementarmatrizen sind mit $E_k \cdots E_2 E_1 A = \mathbb{I}$ (d.h. man hat A durch Zeilenoperationen in die Einheitsmatrix umgeformt), so erhält man durch Multiplikation mit A^{-1} von rechts $E_k \cdots E_2 E_1 \mathbb{I} = A^{-1}$ (d.h. die Anwendung derselben Zeilenoperationen auf die Einheitsmatrix gibt die Inverse).

8) Eine **involutorische Matrix** A ist eine mit $A^2 = \mathbb{I}$, d.h. die Matrix *ist ihre eigene Inverse*, $A = A^{-1}$. Da bei Zahlen nur 1 und -1 ihre eigenen multiplikativen Inversen sind, könnte man zunächst denken, dass nur \mathbb{I}_n , und $-\mathbb{I}_n$ involutorische $n \times n$ -Matrizen sind. Es gibt aber sehr viel mehr. Involutorische 2×2 -Matrizen sind zum Beispiel

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix},$$

und allgemeiner

$$\begin{pmatrix} 0 & r \\ 1/r & 0 \end{pmatrix}, \begin{pmatrix} 1 & s \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ t & -1 \end{pmatrix}, \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ mit } ad - bc = -1 \text{ und } a + d = 0.$$

(Diese Beispiele erfassen tatsächlich *alle* involutorischen 2×2 -Matrizen.) Eine $n \times n$ -Diagonalmatrix D ist involutorisch, genau wenn alle Diagonaleinträge ± 1 sind. Jede zu D *ähnliche Matrix*, d.h. jede Matrix der Form $C^{-1}DC$ mit invertierbarer $n \times n$ -Matrix C , ist dann auch involutorisch; denn es gilt ja

$$(C^{-1}DC)^2 = C^{-1}DCC^{-1}DC = C^{-1}D\mathbb{I}_nDC = C^{-1}D^2C = C^{-1}DC.$$

Man kann beweisen, dass *jede* involutorische Matrix von dieser Form $C^{-1}DC$ ist mit Diagonaleinträgen ± 1 in der Diagonalmatrix D . Geometrisch betrachtet ist der zu D gehörige lineare Operator auf \mathbb{R}^n eine *Achsenpiegelung*. (Die den Diagonaleinträgen $+1$ entsprechenden Achsen bleiben fest, die anderen werden am Ursprung gespiegelt.) $C^{-1}DC$ kann dann als Achsenpiegelung bzgl. eines allgemeinen (nicht unbedingt orthogonalen) Achsensystems in \mathbb{R}^n interpretiert werden. In diesem Sinne entsprechen die involutorischen Matrizen A den Spiegelungen, und $A^2 = \mathbb{I}$ beschreibt die charakteristische Eigenschaft von Spiegelungen, dass man bei zweifacher Anwendung wieder das Original erhält.

9) Involutorische Matrizen A erfüllen $A^2 = \mathbb{I}$; sie sind also gewissermaßen Quadratwurzeln aus der Einheitsmatrix. Man kann fragen, ob es auch k -te Wurzeln A aus der Einheitsmatrix gibt für beliebige $k \in \mathbb{N}_{\geq 2}$, d.h. $A^k = \mathbb{I}$. Wegen $A^{k-1}A = A^k$ ist äquivalent, dass die Inverse von A gleich der $(k-1)$ -ten Potenz von A ist, $A^{-1} = A^{k-1}$. Noch allgemeiner kann man nach der Existenz von **k -ten Wurzeln A aus einer gegebenen quadratischen Matrix B** fragen, d.h. nach Matrizen A mit $A^k = B$.

Bei beliebigen quadratischen Matrizen B ist die Antwort im Allgemeinen nein. Z.B. hat $B = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ keine Quadratwurzel A . Wäre nämlich $A \in \mathbb{R}^{2 \times 2}$ eine, so gälte $AB = AA^2 = A^3 = A^2A = BA$ und A müsste Diagonalmatrix sein; denn Multiplikation mit B von rechts ersetzt die zweite Spalte in A durch ihr Negatives, aber Multiplikation von B mit links ersetzt die zweite Zeile in A durch ihr Negatives. Eine Diagonalmatrix kann aber nicht B als Quadrat haben, weil -1 keine Quadratwurzel in \mathbb{R} hat.

Die Einheitsmatrix \mathbb{I}_n hat dagegen für jedes $k \geq 2$ viele k -te Wurzeln, wenn $n > 1$ ist. Für $n = 2$ kann man z.B. die Matrix $D_{2\pi/k}$ der Drehung um den Ursprung mit Drehwinkel $\frac{2}{k}\pi = \frac{1}{k} \cdot 360^\circ$ nehmen; deren k -te Potenz entspricht dann der Drehung um einen Vollwinkel, die jeden Punkt festlässt, d.h. $(D_{2\pi/k})^k$ ist die 2×2 -Einheitsmatrix. Diese *Drehmatrix* ist

$$D_{2\pi/k} = \begin{pmatrix} \cos \frac{2\pi}{k} & -\sin \frac{2\pi}{k} \\ \sin \frac{2\pi}{k} & \cos \frac{2\pi}{k} \end{pmatrix} \quad \text{mit} \quad (D_{2\pi/k})^k = \mathbb{I}_2, \quad (D_{2\pi/k})^{-1} = (D_{2\pi/k})^{k-1} = D_{-2\pi/k}.$$

Das kann man mit Hilfe von Additionstheoremen für die trigonometrischen Funktionen nachrechnen, es ist aber aufgrund der geometrischen Beschreibung unmittelbar klar. Für $n > 2$ erhält man k -te Wurzeln aus \mathbb{I}_n , indem man eine "Block-Diagonalmatrix" aus m 2×2 -Diagonalblöcken $D_{2\pi/k}$ und aus $n-2m$ Blöcken (1) des Formats 1×1 zusammenstellt. Weitere k -te Wurzeln findet man mit folgender Überlegung: Ist $A^k = B$ und C irgendeine invertierbare Matrix desselben Formats wie A , so ist $C^{-1}AC$ eine k -te Wurzel aus $C^{-1}BC$; denn

$$\begin{aligned} (C^{-1}AC)^k &= \underbrace{(C^{-1}AC)(C^{-1}AC) \cdots (C^{-1}AC)}_k \\ &= C^{-1} \underbrace{A(CC^{-1})A(CC^{-1}) \cdots A(CC^{-1})}_k C \\ &= C^{-1} \underbrace{A|A| \cdots |A|}_k C = C^{-1}A^kC = C^{-1}BC. \end{aligned}$$

Speziell im Fall $B = \mathbb{I}_n$ gilt $C^{-1}\mathbb{I}_n C = C^{-1}C = \mathbb{I}_n$ also ist mit A auch jede zu A ähnliche Matrix $C^{-1}AC$ eine k -te Wurzel aus der Einheitsmatrix \mathbb{I}_n . Wenn $A \neq \pm \mathbb{I}_n$ ist, so bekommt man durch Variation von C auf diese Weise eine unendliche Schar von k -ten Wurzeln $C^{-1}AC$ aus \mathbb{I}_n .

10) Eine **nilpotente Matrix** ist eine quadratische Matrix N ungleich der Nullmatrix 0 mit $N^k = 0$ für ein $k \in \mathbb{N}_{\geq 2}$. Für den kleinsten derartigen Exponenten k heißt $k-1$ der *Nilpotenzgrad* von N . Eine nilpotente Matrix ist selbst nie invertierbar, aber es gilt die Formel für die **Inverse einer nilpotenten Störung der Einheitsmatrix $\mathbb{I} - N$** :

- $\mathbb{I} - N$ ist stets invertierbar, wenn N nilpotent ist, und zwar ist die Inverse gleich $(\mathbb{I} - N)^{-1} = \mathbb{I} + N + N^2 + \dots + N^{k-1}$, wenn $N^k = 0$ ist.

Das bestätigt man einfach durch Multiplikation mit $\mathbb{I} - N$:

$$\begin{aligned} & (\mathbb{I} - N)(\mathbb{I} + N + N^2 + \dots + N^{k-1}) \\ &= (\mathbb{I} + N + N^2 + \dots + N^{k-1}) - (N + N^2 + N^3 + \dots + N^k) = \mathbb{I} - N^k = \mathbb{I}. \end{aligned}$$

Typische nilpotente Matrizen sind *Dreiecksmatrizen A mit lauter Nulleinträgen auf der Diagonalen*. Im Fall einer oberen $n \times n$ -Dreiecksmatrix z.B. kann man dann nachrechnen, dass A^2 auch Nulleinträge in der ersten Oberdiagonalen hat, d.h. in den Positionen $(i, i+1)$, und A^3 auch Nulleinträge in den beiden ersten Oberdiagonalen, also in den Positionen $(i, i+1)$ und $(i, i+2)$ usw. Die Potenz A^n hat dann Nulleinträge in allen Positionen, also ist A nilpotent mit Nilpotenzgrad $\leq n-1$. Konkretes Beispiel:

$$N = \underbrace{\begin{pmatrix} 0 & 1 & & & 0 \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ 0 & & & & 0 \end{pmatrix}}_n \quad N^2 = \underbrace{\begin{pmatrix} 0 & 0 & 1 & & 0 \\ & 0 & 0 & 1 & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ 0 & & & & 0 \end{pmatrix}}_n \quad N^{n-1} = \underbrace{\begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ & 0 & 0 & & 0 \\ & & \ddots & \ddots & \\ & & & \ddots & 0 \\ 0 & & & & 0 \end{pmatrix}}_n$$

Bei Erhöhung des Exponenten um 1 rückt die Schrägzeile mit Einträgen 1 jeweils um eine Stufe nach oben, und N^n ist dann die $n \times n$ -Nullmatrix, also ist N nilpotent mit Nilpotenzgrad $n-1$. Für die Inverse von $\mathbb{I}_n - N$ gilt dann gemäß der obigen Inversionsformel für nilpotente Störungen der Einheitsmatrix:

$$\begin{pmatrix} 1 & -1 & & & 0 \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & -1 \\ 0 & & & & 1 \end{pmatrix}^{-1} = (\mathbb{I} - N)^{-1} = \mathbb{I} + N + N^2 + \dots + N^{n-1} = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ & 1 & 1 & 1 & \vdots \\ & & \ddots & \ddots & 1 \\ & & & \ddots & 1 \\ 0 & & & & 1 \end{pmatrix},$$

wobei die Matrix rechts auf und über der Diagonalen Einträge 1 hat und darunter 0.

Eine allgemeine invertierbare Dreiecksmatrix A kann man immer in der Form $A = D(\mathbb{I} - N)$ schreiben, wobei D die Diagonalmatrix mit denselben Diagonaleinträgen wie A ist und N die Einträge $-a_{ij}/a_{ii}$ in Positionen (i, j) mit $i < j$ hat sowie Nulleinträge in allen anderen Positionen. Dann ist N nilpotent mit Nilpotenzgrad höchstens $n-1$, und wir erhalten folgende **Formel für die Invertierung von Dreiecksmatrizen mit Diagonaleinträgen $\neq 0$** :

$$A^{-1} = [D(\mathbb{I} - N)]^{-1} = (\mathbb{I} - N)^{-1}D^{-1} = (\mathbb{I} + N + N^2 + \dots + N^{n-1})D^{-1},$$

Diese Formel für die Inversen von Dreiecksmatrizen ist, weil man man die Inverse der Diagonalmatrix D nach 5) einfach berechnen kann und sonst nur Matrixadditionen und Multiplikationen auszuführen hat, auch zur praktischen Berechnung von A^{-1} geeignet. ■

Das letzte Beispiel mit der Formel $(\mathbb{I}-A)^{-1} = \mathbb{I} + A + A^2 + \dots + A^{k-1}$, wenn $A^k = 0$, erinnert an die geometrische Reihe $(1-a)^{-1} = \frac{1}{1-a} = a + a + a^2 + a^3 + \dots$ für Zahlen a mit $|a| < 1$. Wir fragen uns daher, ob man nicht auch mit quadratischen $n \times n$ -Matrizen A die Matrix-Reihe

$$\mathbb{I} + A + A^2 + A^3 + \dots = \sum_{k=0}^{\infty} A^k \in \mathbb{R}^{n \times n}$$

erklären kann, wenigstens wenn A in einem passenden Sinne "Betrag" < 1 hat, und ob man dann, wenn dies gelungen ist, vielleicht nachweisen kann, dass die Reihe die inverse Matrix zu $\mathbb{I}-A$ ist. Immerhin stimmt das ja im Fall von nilpotenten Matrizen A , wo die Reihe ab einem gewissen Index mit lauter Nullmatrizen als Gliedern abbricht, so dass man nur endlich viele Summanden zu addieren hat. Tatsächlich funktioniert diese zunächst spekulativ erscheinende Idee, und sie hat auch ökonomische Anwendungen, weil z.B. Matrizen der Form $\mathbb{I}-A$ (mit "kleiner" Matrix A) bei Produktion mit Verflechtung und bei Input-Output-Modellen auftreten und solche der Form $D-A = (\mathbb{I}-AD^{-1})D$ (mit invertierbarer Diagonalmatrix D und "kleiner" Matrix AD^{-1}) bei der innerbetrieblichen Leistungsverrechnung (siehe 3.2 und 3.3). Deshalb geben wir hier für Interessierte einen Einblick in diese mathematische Theorie, der über den Rahmen der Vorlesung hinausführt.

DISKUSSION (*Neumannsche Reihe und invers positive Matrizen*):

1) Eine *formale Potenzreihe quadratischer Matrizen* ist eine Reihe der Form

$$\sum_{k=0}^{\infty} c_k A^k = c_0 \mathbb{I} + c_1 A + c_2 A^2 + c_3 A^3 + \dots$$

mit Koeffizienten $c_k \in \mathbb{R}$ und $A \in \mathbb{R}^{n \times n}$. (In der Mathematik werden auch komplexe Zahlen als Koeffizienten und quadratische Matrizen mit komplexen Einträgen zugelassen.) "Formal" heißt die Reihe, weil sie eigentlich nur eine Absichtserklärung ist; denn man kann natürlich nicht unendlich viele quadratische Matrizen addieren. Was man aber bilden kann, ist für jedes $l \in \mathbb{N}$ die *l-te Partialsumme* $S_l := \sum_{k=0}^l c_k A^k \in \mathbb{R}^{n \times n}$. Wir sagen, dass die *Potenzreihe gegen die $n \times n$ -Matrix B konvergiert* und schreiben

$$\sum_{k=0}^{\infty} c_k A^k = B \quad \text{oder} \quad c_0 \mathbb{I} + c_1 A + c_2 A^2 + c_3 A^3 + \dots = B,$$

wenn für jede gegebene Fehlerschranke $\varepsilon > 0$ ein $l_0 \in \mathbb{N}$ existiert, so dass die Einträge aller Partialsummenmatrizen S_l mit $l \geq l_0$ um weniger als ε von den Einträgen von B in gleicher Position abweichen. Man kann nachweisen, dass Konvergenz der Matrixreihe gegen eine eindeutig bestimmte Matrix B vorliegt, wenn z.B. die Einträge der Matrizen $c_1 A, c_2 A^2, c_3 A^3, \dots$ dem Betrage nach mindestens so schnell klein werden wie eine geometrische Folge q, q^2, q^3, \dots mit $0 \leq q < 1$, d.h. wenn es eine Konstante K gibt, so dass alle Einträge von $c_k A^k$ Betrag $\leq K q^k$ haben für $k = 1, 2, 3, \dots$. Mit konvergenten Matrixreihen darf man wie mit Reihen von reellen Zahlen rechnen — bei Rechnungen, die Matrixmultiplikationen beinhalten, allerdings nur, wenn die Matrixfaktoren A, C miteinander vertauschbar sind, d.h. $AC = CA$ (was ja im Allgemeinen für $A, C \in \mathbb{R}^{n \times n}$ nicht der Fall ist).

2) Das Matrix-Analogon zur geometrischen Reihe $\sum_{k=0}^{\infty} a^k$ ist die **Neumannsche Reihe**

$$\mathbb{I} + A + A^2 + A^3 + \dots = \sum_{k=0}^{\infty} A^k \quad (A \in \mathbb{R}^{n \times n}),$$

also die Matrix-Potenzreihe, bei der alle Koeffizienten 1 sind. Und analog zu der Formel $\frac{1}{1-a} = \sum_{k=0}^{\infty} a^k$ für den Wert der geometrischen Reihe, wenn sie konvergiert (was genau für $|a| < 1$ der Fall ist), gilt hier tatsächlich:

- Wenn die Neumannsche Reihe $\mathbb{I} + A + A^2 + A^3 + \dots$ zu $A \in \mathbb{R}^{n \times n}$ konvergent ist, so ist $\mathbb{I} - A$ invertierbar und die Reihe konvergiert gegen die Inverse

$$(\mathbb{I} - A)^{-1} = \mathbb{I} + A + A^2 + A^3 + \dots = \sum_{k=0}^{\infty} A^k.$$

Der Beweis hierfür verläuft wie bei der geometrischen Reihe: $(\mathbb{I} - A)(\mathbb{I} + A + A^2 + \dots) = \mathbb{I} + A + A^2 + \dots - A - A^2 - A^3 - \dots = \mathbb{I}$, und wenn die Matrix-Reihe konvergiert, so darf man so rechnen.

3) Die Frage ist nun, für welche quadratischen Matrizen A die Neumannsche Reihe konvergiert. Dazu müssen wir irgendwie die Größe der Einträge von A beschränken. Das kann man am besten tun durch Einführung einer Norm $\|A\|$ auf dem Matrizen-Raum $\mathbb{R}^{n \times n}$, d.h. eine Art Betragsfunktion mit $\|A\| \geq 0$ und $\|A\| = 0$ nur für die Nullmatrix A , mit $\|rA\| = |r|\|A\|$ für Skalare $r \in \mathbb{R}$, und mit der sog. *Dreiecksungleichung* $\|A+B\| \leq \|A\| + \|B\|$, was auch als *Subadditivität* der Norm bezeichnet wird. Wenn die Norm zusätzlich noch *submultiplikativ* ist, d.h. wenn $\|AB\| \leq \|A\|\|B\|$ erfüllt ist für das Matrix-Produkt von $A, B \in \mathbb{R}^{n \times n}$, so gilt:

- Die Neumannsche Reihe $\mathbb{I} + A + A^2 + A^3 + \dots$ konvergiert, wenn $\|A\| < 1$ ist;

denn für $q := \|A\| < 1$ hat man $\|A^k\| \leq q^k$ wegen der Submultiplikativität, und man kann zeigen, dass es dann eine Konstante K gibt, so dass alle Einträge von A^k Betrag $\leq Kq^k$ haben, woraus die Konvergenz (wie in 1) bemerkt) folgt.

4) Brauchbare Normen auf dem Matrizen-Raum $\mathbb{R}^{n \times n}$ sind zum Beispiel:

$$\|A\|_{\infty, \infty} := \max_{i=1}^n \sum_{j=1}^n |a_{ij}| \quad \text{Maximum der Zeilenbetragssummen,}$$

$$\|A\|_{1,1} := \sum_{i=1}^n \max_{j=1}^n |a_{ij}| \quad \text{Summe der Zeilenbetragsmaxima,}$$

$$\|A\|_2 := \sqrt{\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2} \quad \text{Euklidische Norm,}$$

$$\|A\| := \sup\{|Ax| : x \in \mathbb{R}^n, |x| = 1\} \quad \text{Operator-Norm,}$$

Die Euklidische Norm einer Matrix ist dabei ganz analog zur Euklidischen Norm $|x|$ von Vektoren in $x \in \mathbb{R}^n$ als die Quadratwurzel aus der Summe der Quadrate aller Einträge definiert. Die Operator-Norm $\|A\|$ ist die maximale Länge, die das Bild eines Einheitsvektors $x \in \mathbb{R}^n$ unter dem durch A beschriebenen linearen Operator auf \mathbb{R}^n haben kann, und gleichzeitig die kleinstmögliche Dehnungsschranke für diesen Operator, d.h. die kleinste Zahl L mit $|A\tilde{x} - Ax| \leq L|\tilde{x} - x|$ für alle $\tilde{x}, x \in \mathbb{R}^n$. Die Operator-Norm beschreibt also die maximale Längendehnung (wenn $\|A\| \geq 1$) bzw. die minimale Längenschauchung (wenn $\|A\| \leq 1$) von Vektoren in \mathbb{R}^n unter diesem linearen Operator. Man kann für die vier hier definierten Größemessungen von Matrizen (mit einigem Aufwand) die Normeigenschaften nachrechnen und auch die Submultiplikativität dieser Normen beweisen. Es gilt daher:

- Die Neumannsche Reihe $\mathbb{I} + A + A^2 + \dots$ von $A \in \mathbb{R}^{n \times n}$ konvergiert gegen $(\mathbb{I} - A)^{-1}$, wenn das Maximum der Zeilenbetragssummen oder die Summe der Zeilenbetragsmaxima oder die Euklidische Norm oder die Operator-Norm von A kleiner als 1 ist.

In dieser Feststellung kann man übrigens "Zeilen" ohne weiteres durch "Spalten" ersetzen, wenn man die Rechenregel $(\mathbb{I} - A^T)^{-1} = ((\mathbb{I} - A)^T)^{-1} = ((\mathbb{I} - A)^{-1})^T$ für die Inversion transponierter Matrizen berücksichtigt und überlegt, dass $\mathbb{I} + A^T + (A^T)^2 + (A^T)^3 + \dots$ gegen B^T konvergiert, wenn $\mathbb{I} + A + A^2 + A^3 + \dots = B$ ist, und umgekehrt.

5) Eine $m \times n$ -Matrix B heißt **positive Matrix** (bzw. **nichtnegative Matrix**), wenn für alle $x \in \mathbb{R}^n$ mit Komponenten $x_j > 0$ auch $y = Bx$ lauter positive (bzw. nichtnegative) Komponenten hat. Und B heißt **invers positive Matrix** (bzw. **invers nichtnegative Matrix**), wenn für alle $y \in \mathbb{R}^m$ mit Komponenten $y_i > 0$ das Gleichungssystem $Bx = y$ nur Lösungsvektoren $x \in \mathbb{R}^n$ besitzt, die lauter positive (bzw. nichtnegative) Komponenten haben. Man überlegt dann leicht:

- B nichtnegative Matrix \iff alle Einträge von B sind ≥ 0 ;
 B positive Matrix \iff alle Einträge von B sind ≥ 0 und B hat keine Nullzeile;
- B invers positiv (bzw. invers nichtnegativ) $\iff B^{-1}$ positiv (bzw. nichtnegativ), vorausgesetzt die Inverse B^{-1} existiert;
- B kann nur dann invers nichtnegativ sein, wenn $Bx = 0$ nur die triviale Lösung hat, also $\text{Rang}(B) = n \geq m$ ist

Letzteres ist klar, weil anderenfalls gewisse Komponenten der Lösung frei als Parameter, also auch negativ, wählbar sind.

Von ökonomischer Bedeutung sind diese Begriffsbildungen, weil man bei ökonomischen Vektorvariablen $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ sehr oft die Nebenbedingung hat, dass alle Komponenten positiv oder wenigstens nichtnegativ sind, da negative Werte der Komponenten sinnlos sind. Ist x z.B. ein Produktionsvektor, dessen Komponenten produzierte Mengeneinheiten mehrerer Produkte oder eines Produkts an mehreren Produktionsstätten angeben, so muss natürlich $x_j \geq 0$ sein für alle j . Ist y ein zugehöriger Rohstoffverbrauchsvektor und wird ein linearer Zusammenhang $y = Rx$ modelliert, so sind nur nichtnegative Verflechtungsmatrizen R sinnvoll, weil es ja keinen negativen Rohstoffverbrauch gibt. Ist andererseits y ein Produktions-Outputvektor für den Verkauf und x der dazu erforderliche Produktionsvektor, der auch einen internen Verbrauch der erzeugten Produkte im Produktionsprozess berücksichtigt (endogener Input), so hat man zu gegebenem y den Produktionsvektor x aus einem linearen Gleichungssystem $Bx = y$ zu bestimmen, für das nur invers nichtnegative Matrizen B sinnvoll sind; denn negative Produktionsmengen (Rückgewinnung der Rohstoffe aus den Endprodukten) werden ja nicht hergestellt.

6) Die Positivität einer Matrix ist nach 5) mit einem Blick auf ihre Einträge zu erkennen. Schwieriger, für ökonomische Anwendungen aber auch wichtiger, ist die Feststellung der inversen Positivität. Hier hilft folgende Beobachtung:

- Ist $A \in \mathbb{R}^{n \times n}$ nichtnegativ und die Neumannsche Reihe $\mathbb{I} + A + A^2 + A^3 + \dots = (\mathbb{I} - A)^{-1}$ konvergent, so ist $\mathbb{I} - A$ invers positiv, also $(\mathbb{I} - A)^{-1}$ positiv.

Das liegt daran, dass mit A offenbar auch A^2 , A^3 , ... lauter nichtnegative Einträge hat; daher haben die Partialsummen-Matrizen $S_l = \mathbb{I} + A + A^2 + \dots + A^l$ ebenfalls nichtnegative Einträge und auf der Diagonalen sogar Einträge ≥ 1 . Dasselbe gilt dann auch für den Grenzwert $(\mathbb{I} - A)^{-1}$ und hat zur Folge, dass $(\mathbb{I} - A)^{-1}y$ nur positive Komponenten hat, wenn dies für y der Fall ist.

7) Wir geben noch ein direktes Argument dafür an, dass $\mathbb{I}-A$ invers positiv ist, wenn eine der in 4) betrachteten Normen von $A \in \mathbb{R}^{n \times n}$ kleiner als 1 ist. Das Argument ist zwar weniger durchsichtig als 6), aber es benutzt keine unendliche Matrix-Reihe. Wir nehmen dazu an, dass für alle $x \in \mathbb{R}^n$ für die Euklidischen Normen von x und Ax gilt:

$$|Ax| \leq \kappa|x| \quad \text{mit einer Konstanten } \kappa < 1.$$

Wenn die Operator-Norm $\|A\| < 1$ ist, so gilt das mit $\kappa := \|A\|$, und wenn die Euklidische Norm $\|A\|_2 < 1$ ist, so gilt es erst recht, weil man $\|A\| \leq \|A\|_2$ zeigen kann.

Wir zerlegen nun jeden Vektor $x \in \mathbb{R}^n$ in der Form $x = x_+ + x_-$, wobei x_+ dieselben Einträge wie x hat in Positionen, in denen diese nichtnegativ sind, und Nulleinträge sonst, während x_- dieselben negativen Komponenten wie x hat und Nulleinträge sonst. Ist A nichtnegativ, so gilt $Ax = Ax_+ + Ax_-$, wobei Ax_+ nur nichtnegative Komponenten besitzt und Ax_- nur nichtpositive. Die negativen Komponenten von Ax haben daher keinen größeren Betrag als die entsprechenden Komponenten von Ax_- und es folgt $|(Ax)_-| \leq |Ax_-|$. Ist nun $(\mathbb{I}-A)x = y$, also $x = Ax + y$, wobei $y \in \mathbb{R}^n$ lauter positive Komponenten hat, so sind auch die negativen Komponenten von $Ax + y$ im Betrag kleiner als die entsprechenden negativen Komponenten von Ax , d.h. es gilt auch $|(Ax + y)_-| \leq |(Ax)_-|$. Setzen wir alles zusammen, so folgt $|x_-| = |(Ax + y)_-| \leq |(Ax)_-| \leq |Ax_-| \leq \kappa|x_-|$, und wegen $\kappa < 1$ ist das nur möglich, wenn x_- der Nullvektor ist. Das heißt zunächst nur, dass x selbst nichtnegative Komponenten hat, aber aus $x = Ax + y$ folgt dann sogar, dass alle Komponenten von x positiv sind, weil das für y gilt und weil A nichtnegativ ist. Damit ist gezeigt, dass $\mathbb{I}-A$ invers positiv ist, und mit 5) folgt dann auch die Existenz und Positivität von $(\mathbb{I}-A)^{-1}$. (Aus $|x| = |Ax| \leq \kappa|x|$ für $x = Ax$ sieht man auch direkt, dass $(\mathbb{I}-A)x = 0$ nur die triviale Lösung hat, wenn $\kappa < 1$ ist.)

Ist $\|A\|_{\infty, \infty} < 1$ bzw. $\|A\|_{1,1} < 1$, so kann man ganz analog argumentieren, wenn man die Euklidische Norm $|x|$ von Vektoren $x \in \mathbb{R}^n$ ersetzt durch das Komponentenbetragsmaximum $|x|_{\infty} := \max_{j=1}^n |x_j|$ bzw. durch die Komponentenbetragssumme $|x|_1 := \sum_{j=1}^n |x_j|$. Es gilt dann $|Ax|_{\infty} \leq \kappa|x|_{\infty}$ für $\kappa := \|A\|_{\infty, \infty} < 1$ bzw. $|Ax|_1 \leq \kappa|x|_1$ für $\kappa := \|A\|_{1,1} < 1$, und dieselben Überlegungen wie eben zeigen inverse Positivität von $\mathbb{I}-A$. *Ökonomische Anwendungen* der so gewonnenen Positivitätsaussagen über die Inverse zu $\mathbb{I}-A$ mit "kleiner" nichtnegativer Matrix Abesprechen wir in den nächsten Beispielen.

8) Eine andere Matrix-Potenzreihe wollen wir noch erwähnen, weil sie im Zusammenhang mit der Lösung von linearen Systemen gewöhnlicher Differentialgleichungen steht, die auch zur Modellierung ökonomischer Sachverhalte verwendet werden. Die **Exponentialreihe einer quadratischen Matrix** $A \in \mathbb{R}^{n \times n}$ ist definiert als

$$e^A := \sum_{k=0}^{\infty} \frac{1}{k!} A^k = \mathbb{I} + \frac{1}{1!} A + \frac{1}{2!} A^2 + \frac{1}{3!} A^3 + \dots \in \mathbb{R}^{n \times n}$$

und stellt das Matrix-Analogon der Exponentialreihe $e^a = \sum_{k=0}^{\infty} \frac{1}{k!} a^k = 1 + \frac{1}{1!} a + \frac{1}{2!} a^2 + \frac{1}{3!} a^3 + \dots \in \mathbb{R}$ von reellen Zahlen a dar. Mit einer submultiplikativen Matrix-Norm $\|\cdot\|$ wie in 4) kann man die Norm der Glieder durch $\|\frac{1}{k!} A^k\| \leq \frac{1}{k!} \|A\|^k$ abschätzen, und weil die Fakultäten $k!$ mit wachsendem k sehr viel schneller anwachsen als die Potenzen $\|A\|^k$, streben die Einträge der Glieder der Matrix-Exponentialreihe schneller gegen Null als geometrische Nullfolgen, so dass die Reihe tatsächlich für jede quadratische Matrix konvergiert. Es ist dann sinnvoll, den Matrix-Grenzwert der Reihe mit e^A zu bezeichnen in Analogie zum Wert e^a der Exponentialreihe der reellen Zahl a . Für eine Diagonalmatrix D mit Diagonaleinträgen d_1, \dots, d_n ist z.B. e^D die Diagonalmatrix mit den Einträgen e^{d_1}, \dots, e^{d_n} ; aber für allgemeine quadratische Matrizen A hängt jeder Eintrag von e^A im Allgemeinen in komplizierter Weise von *allen* Einträgen von A ab. ■

BEISPIELE (*inverse Positivität, ökonomische Anwendungen*):

1) Die 2×2 -Matrix $A = \begin{pmatrix} 0 & s \\ s & 0 \end{pmatrix}$ ist positiv für $s > 0$ (alle Einträge sind ≥ 0 , und es gibt keine Nullzeile), und $\mathbb{I} - A = \begin{pmatrix} 1 & -s \\ -s & 1 \end{pmatrix}$ ist invertierbar für $s \neq \pm 1$ mit

$$(\mathbb{I} - A)^{-1} = \begin{pmatrix} 1 & -s \\ -s & 1 \end{pmatrix}^{-1} = \frac{1}{1 - s^2} \begin{pmatrix} 1 & s \\ s & 1 \end{pmatrix}.$$

Diese Matrix ist nur für $0 \leq s < 1$ positiv, für $s > 1$ sind dagegen alle Einträge von $(\mathbb{I} - A)^{-1}$ negativ (und für $-1 \neq s < 0$ hat sie zwei negative Einträge). Man sieht hier, dass eine Kleinheitsbedingung wie $\|A\| < 1$ an die nichtnegative Matrix A nicht entbehrlich ist, um inverse Positivität von $\mathbb{I} - A$ schließen zu können. Für $s > 1$ konvergiert natürlich auch die Neumannsche Reihe nicht, obwohl $(\mathbb{I} - A)^{-1}$ existiert. (Sonst wäre ja $(\mathbb{I} - A)^{-1}$ positiv; hier sieht man Divergenz aber auch direkt an $A^{2k} = s^{2k}\mathbb{I}$.)

2) Bei den statischen **Input-Output-Modellen** (nach *Leontief*) hatten wir in 3.3 gesehen, dass zwischen dem *Produktionsvektor* $x \in \mathbb{R}^n$ und dem *Nachfragevektor* $y \in \mathbb{R}^n$ der lineare Zusammenhang

$$(\mathbb{I} - A)x = y$$

besteht, wobei $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ die *Matrix der Produktionskoeffizienten* $a_{ij} \geq 0$ ist, die angeben, wieviele Einheiten des Produkts (i) intern ("endogen") für die Herstellung einer Einheit des Produkts (j) gebraucht werden. In einem "normalen" Unternehmen bzw. in einer "gesunden" Volkswirtschaft wird es sich so verhalten, dass die a_{ij} relativ kleine Zahlen oder Null sind. Lassen wir Produkte außer Betracht, die im Produktionsprozess intern völlig verbraucht werden, und nehmen also an, dass bei laufender Produktion von jedem Produkt $(1), \dots, (n)$ eine positive Menge auf den Markt gebracht werden kann, so ist für jedes i die Summe der Produktionskoeffizienten $\sum_{j=1}^n a_{ij}$ kleiner als 1, d.h. die Zeilen(-Betrags-)Summen von A sind alle kleiner als 1. Aus 4) und 6) der vorangegangenen Diskussion folgt dann:

- Sind die Zeilensummen der Produktionskoeffizientenmatrix A alle kleiner als 1, so hat die Gleichung $(\mathbb{I} - A)x = y$ für alle ökonomisch sinnvollen Nachfragevektoren $y \in \mathbb{R}^n$ (d.h. alle Komponenten y_i sind positiv) genau eine Lösung $x \in \mathbb{R}^n$, und diese ist auch ökonomisch sinnvoll (d.h. auch alle Komponenten x_j sind positiv).

In der Praxis bedeutet "ökonomisch sinnvoll" für einen Produktionsvektor x allerdings meist nicht nur, dass seine Komponenten positiv sind, sondern man hat auch *Kapazitätsgrenzen* $x_j \leq c_j$ zu berücksichtigen. Das Problem, zu welchen Nachfragevektoren $y \in \mathbb{R}^n$ der Produktionsvektor x auch die Kapazitätsbeschränkungen erfüllt, kann man ebenfalls mit der inversen Positivität von $\mathbb{I} - A$ behandeln. Dazu sei $c := (c_1, \dots, c_n) \in \mathbb{R}^n$ der *Vektor der Kapazitätsgrenzen*. Dann gilt $(\mathbb{I} - A)(c - x) = (\mathbb{I} - A)c - y$, und aus der inversen Positivität von $\mathbb{I} - A$ folgt, dass $c - x$ lauter nichtnegative Komponenten hat, dass also die Ungleichungen $x_j \leq c_j$ für $j = 1 \dots n$ gelten, wenn $(\mathbb{I} - A)c - y$ keine negativen Komponenten besitzt.

- Hat dann das Produkt $(\mathbb{I} - A)c$ von $\mathbb{I} - A$ mit dem Vektor der Kapazitätsgrenzen c keine kleinere Komponente als der Nachfragevektor y , so erfüllt der zugehörige Produktionsvektor x auch die Kapazitätsrestriktionen $x_j \leq c_j$ für $j = 1 \dots n$.

Das ist ökonomisch plausibel; denn es besagt, dass man den Nachfragevektor y realisieren kann, wenn bei maximal möglicher Produktion aller Produkte der nicht intern verbrauchte Output eines jeden Produkts die entsprechende Nachfrage übertrifft oder erreicht.

Allerdings ist es nur ein hinreichendes Kriterium für die Realisierbarkeit eines gegebenen Nachfragevektors y . Es ist durchaus möglich, dass der Produktionsvektor zu y die Kapazitätsgrenzen einhält, auch wenn $(\mathbb{I}-A)c$ eine kleinere Komponente als y hat. Es kann sogar sein, dass $(\mathbb{I}-A)c$ eine nichtpositive Komponente hat, so dass kein ökonomisch sinnvoller Nachfragevektor eine kleinere Komponente besitzt, aber dennoch sind natürlich alle Nachfragevektoren mit hinreichend kleinen positiven Komponenten innerhalb der Kapazitätsgrenzen realisierbar. Genauer gilt mit den Bezeichnungen $|x|_\infty$ für das Maximum der Komponentenbeträge von x und $\|A\|_{\infty, \infty} < 1$ für das Maximum der Zeilenbetragssummen von A (vgl. 7) der vorangegangenen Diskussion):

$$\begin{aligned} (\mathbb{I}-A)x = y &\iff x = Ax + y \implies |x|_\infty \leq |Ax|_\infty + |y|_\infty \leq \|A\|_{\infty, \infty}|x|_\infty + |y|_\infty \\ &\implies |x|_\infty \leq \frac{|y|_\infty}{1 - \|A\|_{\infty, \infty}}, \end{aligned}$$

woraus man abliest, dass der Produktionsvektor x zu jedem Nachfragevektor y mit Komponenten $0 < y_i \leq (1 - \|A\|_{\infty, \infty}) \min_{j=1}^n c_j$ jedenfalls die Kapazitätsgrenzen einhält.

3) Bei der **innerbetrieblichen Leistungsverrechnung** hatten wir in 3.2 das lineare Gleichungssystem

$$(D - A)p = k$$

für den *Verrechnungspreisvektor* $p = (p_1, \dots, p_n)$ und den *Primärkostenvektor* $k = (k_1, \dots, k_n)$ aufgestellt, wobei D die *Diagonalmatrix der Gesamtleistungen* $l_i > 0$ der einzelnen Kostenstellen \textcircled{i} ist, welche die Diagonaleinträge von D bilden, und A die *Verflechtungsmatrix*, deren Einträge $a_{ij} \geq 0$ die von der Kostenstelle \textcircled{j} an den Betriebsteil \textcircled{i} abgegebene Zahl von Leistungseinheiten angeben. Ökonomisch sinnvoll sind hier nur positive Verrechnungspreise $p_j > 0$ und nichtnegative Primärkosten $k_i \geq 0$. Wir nehmen dabei an, dass jeder Teilbetrieb \textcircled{j} nur Leistungen einer einzigen Art erbringt, für die der Verrechnungspreis p_j zu bestimmen ist; das ist keine Einschränkung, weil man sich Betriebsteile, die Leistungen verschiedener Art liefern, in einzelne Stellen aufgespalten denken kann, die jeweils nur Leistungen von einer einzigen Sorte abgeben. Da die Gesamtleistung l_j eines jeden Betriebsteiles \textcircled{j} nicht kleiner ist als dessen an andere Betriebsteile abgegebene und selbst verbrauchte Leistung, gilt

$$a_{1j} + \dots + a_{nj} \leq l_j \quad \text{für } j = 1 \dots n,$$

d.h. keine Spaltensumme von A ist größer als der entsprechende Eintrag von D .

Wir betrachten nun zunächst den Fall, dass jeder Teilbetrieb seine Leistung nicht ausschließlich an andere Betriebsteile abgibt, sondern auch eine direkte Leistung den Gesamtbetrieb erbringt, d.h. $a_{1j} + \dots + a_{nj} < l_j$ gilt für $j = 1 \dots n$. Dann hat die Matrix AD^{-1} Spaltensummen $(a_{1j} + \dots + a_{nj})/l_j < 1$ und nach 6), 7) der vorangegangenen Diskussion existiert daher $(\mathbb{I} - AD^{-1})^{-1}$ und ist eine positive Matrix. Dasselbe gilt natürlich auch für

$$(D - A)^{-1} = [(\mathbb{I} - AD^{-1})D]^{-1} = D^{-1}(\mathbb{I} - AD^{-1})^{-1},$$

d.h. unsere Gleichung $(D - A)p = k$ hat eine eindeutige Lösung, und wenn $k_i \geq 0$ ist für alle i , so folgt auch $p_j \geq 0$ für alle j . Wegen $Dp = k + Ap$, also $l_i p_i = k_i + \sum_{j=1}^n a_{ij} p_j$, ist nur dann $p_i = 0$, wenn die Primärkosten k_i und die Sekundärkosten $\sum_{j=1}^n a_{ij} p_j$ von \textcircled{i} Null sind. (Das entspricht dem Ansatz, aus dem man das Gleichungssystem für die Verrechnungspreise erhalten hat: Die von \textcircled{i} erbrachte Leistung wird ja so bewertet, dass ihr Geldwert gleich den Gesamtkosten der Stelle \textcircled{i} ist.) Ist Leistungserbringung ohne Primärkosten nicht möglich, also $k_i > 0$ für alle i , so sind folglich auch alle p_j positiv.

4) Bei vielen Betriebsabläufen gibt es nun aber Betriebsteile \textcircled{j} , die ihre Leistungen ausschließlich an andere Teilbetriebe abgeben und keinen Leistungsüberschuss an den Gesamtbetrieb zur Vermarktung liefern (z.B. Reparatur- und Wartungsteilbetrieb). Dafür ist dann $a_{1j} + \dots + a_{nj} = l_j$, und somit hat in der Matrix AD^{-1} die Spaltensumme $(a_{1j} + \dots + a_{nj})/l_j$ den Wert 1. In dieser Situation ist die obige Argumentation nicht mehr gültig und inverse Positivität von $\mathbb{I} - AD^{-1}$ ohne weitere ökonomisch sinnvolle Annahmen über die Verflechtungsmatrix nicht mehr gesichert. Eine solche Annahme ist, dass ein System von Teilbetrieben, das keine Leistung nach außen erbringt, ausgeschlossen ist. Die Teile eines solchen Systems würden sich nämlich nur gegenseitig Leistungen erbringen, ohne unmittelbar oder mittelbar etwas zum Betriebsergebnis positiv beizutragen, und das ist eine in der Realität zwar manchmal nicht ausgeschlossene, aber ökonomisch sicher nicht sinnvolle Situation. Unsere Annahme verbietet z.B., dass ein Betriebsteil \textcircled{j} seine gesamte erbrachte Leistung selbst wieder verbraucht ($l_j = a_{jj}$) oder dass zwei Betriebsteile \textcircled{i} und \textcircled{j} ihre gesamte Leistung ausschließlich füreinander erbringen ($l_j = a_{ij} + a_{jj}$ und $l_i = a_{ji} + a_{ii}$). Allgemeiner ist mit unserer Annahme verboten, dass die Spalten mit gewissen Nummern j_k ($j_1 < \dots < j_m$, $1 \leq m \leq n$) von Null verschiedene Einträge höchstens in diesen Positionen j_1, \dots, j_m haben und die Spaltensummen gleich l_{j_k} sind.

Unter der Annahme, dass Systeme von Teilbetrieben, die keine Leistung nach außen erbringen, nicht vorhanden sind, können wir nun tatsächlich beweisen, dass $D - A$ invertierbar und invers positiv ist, so dass das Gleichungssystem $(D - A)\mathbf{p} = \mathbf{k}$ für jeden Primärkostenvektor $\mathbf{k} \in \mathbb{R}^n$ genau einen Verrechnungspreisvektor $\mathbf{p} \in \mathbb{R}^n$ als Lösung hat, wobei alle Komponenten p_j von \mathbf{p} nichtnegativ sind, wenn dies für die Komponenten k_i von \mathbf{k} gilt. Dazu betrachten wir eine Lösung $x \in \mathbb{R}^n$ zu $(D - A)x = \mathbf{k}$ bzw. $Dx - \mathbf{k} = Ax$, mit m negativen Komponenten, $1 \leq m \leq n$. Nach geeigneter Nummerierung der Betriebsteile können wir $x_1 < 0, \dots, x_m < 0$ und $x_{m+1} \geq 0, \dots, x_n \geq 0$ annehmen. Summation der Gleichungen $l_i x_i - k_i = \sum_{j=1}^n a_{ij} x_j$ über $i = 1 \dots m$ liefert:

$$\begin{aligned} \sum_{i=1}^m l_i x_i - \sum_{i=1}^m k_i &= \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij} x_j \right) \\ &\geq \sum_{i=1}^m \left(\sum_{j=1}^m a_{ij} x_j \right) \quad (\text{weil } a_{ij} x_j \geq 0 \text{ für } j > m) \\ &= \sum_{j=1}^m \left(\sum_{i=1}^m a_{ij} \right) x_j \geq \sum_{j=1}^m l_j x_j \quad (\text{weil } \sum_{i=1}^m a_{ij} \leq l_j \text{ und } x_j < 0 \text{ für } j = 1 \dots m). \end{aligned}$$

Sind alle $k_i \geq 0$, so ist das nur möglich, wenn $\sum_{i=1}^m k_i = 0$ ist und $\sum_{i=1}^m a_{ij} = l_j$ für $j = 1 \dots m$, also insbesondere $a_{ij} = 0$ für $j = 1 \dots m$ und $i > m$. Dies bedeutet aber, dass die Betriebsteile $\textcircled{1} \dots \textcircled{m}$ ein System bilden, das keine Leistung nach außen abgibt, und das ist durch unsere Annahme ausgeschlossen. Folglich hat keine Lösung x zu $(D - A)x = \mathbf{k}$ eine negative Komponente x_j , wenn $k_1, \dots, k_n \in \mathbb{R}_{\geq 0}$ sind. Daraus folgt erstens, dass $(D - A)x = 0$ nur die triviale Lösung hat (sonst gäbe es auch eine Lösung x mit negativer Komponente, weil ja mit x auch $-x$ Lösung ist), dass also die quadratische Matrix $D - A$ invertierbar ist, und zweitens, dass $(D - A)^{-1}$ eine positive Matrix ist, also nur nichtnegative Einträge und keine Nullzeile hat. Insbesondere sind die durch $(D - A)\mathbf{p} = \mathbf{k}$ gegebenen Verrechnungspreise eindeutig bestimmt und nichtnegativ, weil der Primärkostenvektor \mathbf{k} lauter nichtnegative Komponenten hat.

Sind die Primärkosten k_i für alle Betriebsteile \textcircled{i} positiv, so sind dann natürlich auch alle Verrechnungspreise p_j positiv und damit ökonomisch sinnvoll. Im Allgemeinen aber wird es Betriebsteile geben, bei denen die Primärkosten $k_i = 0$ sind und nur Sekundärkosten entstehen. In dieser Situation ist nicht ausgeschlossen, dass auch einige Verrechnungspreise $p_i = 0$ sind, was jedoch ökonomisch nicht sinnvoll wäre, weil es ja bedeutet, dass ein Betriebsteil Leistungen erbringt, ohne dass primäre oder sekundäre Kosten entstehen. Mathematisch ist das unter den bisherigen Voraussetzungen aber durchaus möglich, wie das Beispiel der Nullmatrix $A = 0$ zeigt, in dem alle Betriebsteile \textcircled{i} ohne jede Verflechtung nur Leistungen l_i an den Gesamtbetrieb abgeben. Hier ist dann $p_i = k_i/l_i$ und somit $p_i = 0$, wenn $k_i = 0$ ist. Um Derartiges auszuschließen, also positive Verrechnungspreise zu erhalten, müssen wir eine weitere ökonomisch sinnvolle Annahme an die Verflechtungsmatrix A und den Primärkostenvektor \mathbf{k} machen, nämlich dass ein *System von Teilbetrieben, das keine Leistung von außen bezieht*, ausgeschlossen ist. Dies bedeutet insbesondere, dass es keinen Teilbetrieb gibt, der von keinem anderen Teilbetrieb Leistungen erhält und dem auch keine Primärkosten entstehen. Und es heißt natürlich auch, dass der Primärkostenvektor nicht der Nullvektor sein kann. Allgemein verbietet diese Annahme, dass A Zeilen mit Nummern i_1, \dots, i_m ($i_1 < \dots < i_m, 1 \leq m \leq n$) besitzt, derart dass von Null verschiedene Einträge in diesen Zeilen höchstens in den Positionen i_1, \dots, i_m vorkommen und dass die Primärkosten k_{i_i} alle gleich Null sind (d.h. diese Teilbetriebe beziehen Leistungen ausschließlich voneinander und arbeiten als System ohne Kosten).

Damit können wir dann auch zeigen, dass alle $p_j > 0$ sind. Wäre das nämlich nicht der Fall, so hätten wir nach geeigneter Nummerierung der Teilbetriebe $p_1 = \dots = p_m = 0$ und $p_{m+1} > 0, \dots, p_n > 0$ für ein $m \in \{1, \dots, n\}$. Aus $(D-A)\mathbf{p} = \mathbf{k}$ folgte dann

$$0 = l_i p_i = k_i + \sum_{j=1}^n a_{ij} p_j = k_i + \sum_{j=m+1}^n a_{ij} p_j \quad \text{für } i = 1 \dots m$$

und damit $k_i = 0$ für $i = 1 \dots m$ sowie $a_{ij} = 0$ für $i = 1 \dots m$ und $j > m$. Das bedeutet aber, dass die Teilbetriebe $\textcircled{1} \dots \textcircled{m}$ ein System bilden, das keine Leistung von außen bezieht, und das hatten wir mit unserer Annahme ja ausgeschlossen. Zusammenfassung:

- *Gibt es bei der innerbetrieblichen Leistungsverrechnung mit Verflechtungsmatrix A und Gesamtleistungsmatrix D kein System von Teilbetrieben, das keine Leistungen nach außen erbringt, so ist das lineare Gleichungssystem $(A-D)\mathbf{p} = \mathbf{k}$ für jeden Primärkostenvektor \mathbf{k} eindeutig lösbar, und alle Verrechnungspreise p_j sind nichtnegativ (wenn alle Primärkosten k_i nichtnegativ sind).*
- *Gibt es zu A und \mathbf{k} auch kein System von Teilbetrieben, das keine Leistungen von außen bezieht, so sind sogar alle Verrechnungspreise p_j positiv.*

Damit ist das Problem der *Existenz, Eindeutigkeit und Positivität der Verrechnungspreise* in sehr befriedigender Weise, d.h. unter ökonomisch sinnvollen Voraussetzungen, gelöst. Wenn bei gegebenen A , D und \mathbf{k} die Auflösung des linearen Gleichungssystems $(D-A)\mathbf{p} = \mathbf{k}$ für die Verrechnungspreise ergibt, dass die Lösung nicht eindeutig ist, so hat der Betrieb ein System von Teilbetrieben, die nur füreinander arbeiten und keine Leistung nach außen abgeben — das muss dann natürlich durch Änderung der Betriebsabläufe abgestellt werden. Wenn sich zwar ein eindeutiger Verrechnungspreisvektor ergibt, der aber Nullkomponenten enthält, so muss ein Fehler bei der Aufstellung der Daten A , D und \mathbf{k} vorliegen, der z.B. auf der unvollständigen Erfassung aller entstehenden Kosten beruht; denn Systeme, die Leistungen ohne Kosten erbringen, existieren in der Realität nicht. Ist aber die Struktur von A , D und \mathbf{k} so, dass ökonomisch unsinnige Systeme von Teilbetrieben nicht existieren, so gibt es auch eindeutige positive Verrechnungspreise! ■

Nun zum zweiten Thema dieses Abschnitts, den Determinanten. Bei 2×2 -Matrizen $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ haben wir schon gesehen, dass man an der Zahl $ad - bc$ direkt ablesen kann, ob die Matrix invertierbar ist oder nicht: Wenn $ad - bc \neq 0$ ist, so existiert die Inverse, andernfalls nicht. Da diese aus den Matrixeinträgen leicht zu berechnende Zahl also bestimmt ("determiniert"), ob die Matrix invertierbar ist oder nicht, heißt sie die *Determinante* der 2×2 -Matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Auch bei der Berechnung der Inversen $\frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ taucht die Determinante auf, nämlich als Nenner der Einträge der inversen Matrix. Daran kann man zum Beispiel ablesen, dass die Inverse ganzzahlige Einträge hat, wenn das für die Matrix selbst gilt und wenn die Determinante gleich 1 oder gleich -1 ist; und wenn die Determinante den Wert $\pm k$ hat mit $k \in \mathbb{N}$, so lassen sich alle Einträge der Inversen einer ganzzahligen 2×2 -Matrix als rationale Zahlen mit ganzem Zähler und Nenner k schreiben.

Die Frage ist nun, ob wir auch für quadratische Matrizen A mit mehr als 2 Zeilen und Spalten eine Zahl $\det(A) \in \mathbb{R}$ aus den Matrixeinträgen berechnen können, die ähnliche Bedeutung hat, also z.B. genau dann ungleich Null ist, wenn die Matrix invertierbar ist. Um einen Anhaltspunkt zu haben, wie man dabei vorgehen könnte, schaut man sich die Eigenschaften der Determinantenfunktion $ad - bc$ von 2×2 -Matrizen $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ an. Eigenschaften, die ins Auge fallen, sind z.B. folgende: Die Determinante hängt linear ab von jeder Spalte $\begin{pmatrix} a \\ c \end{pmatrix}$, $\begin{pmatrix} b \\ d \end{pmatrix}$ der Matrix und sie ändert sich um den Faktor -1 , wenn man die zwei Spalten vertauscht. Es stellt sich heraus, dass durch die analogen Eigenschaften und die zusätzliche Bedingung $\det(\mathbb{I}_n) = 1$ eine Funktion auf dem Raum $\mathbb{R}^{n \times n}$ der $n \times n$ -Matrizen schon eindeutig bestimmt ist, und dass diese Funktion tatsächlich genau auf den nicht invertierbaren Matrizen den Wert Null annimmt. Dies ist dann die Determinantenfunktion von $n \times n$ -Matrizen.

SATZ und DEFINITION (der Determinante): *Es gibt für jede natürliche Zahl n genau eine Funktion $\det: \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$, die jeder $n \times n$ -Matrix A eine reelle Zahl $\det(A)$ zuordnet, genannt die **Determinante von A** , und folgende Eigenschaften hat:*

- (i) **Normierung:** $\det(\mathbb{I}_n) = 1$, die Determinante der $n \times n$ -Einheitsmatrix ist 1;
- (ii) **Antisymmetrie:** bei Vertauschung von zwei Spalten in A ändert sich die Determinante um den Faktor -1 ;
- (iii) **lineare Abhängigkeit von jeder Spalte, das heißt**
Homogenität: *multipliziert man eine Spalte von A mit einem Faktor $r \in \mathbb{R}$, so ändert sich auch die Determinante um den Faktor r ;*
Additivität: *ist eine Spalte von A die Summe von zwei Spaltenvektoren, so ist $\det(A)$ die Summe der Determinanten der beiden Matrizen, die man erhält, wenn man die betreffende Spalte durch jeweils einen der beiden Spaltenvektoren ersetzt.*

DISKUSSION: 1) Als Erstes betonen wir:

- Die Determinante ist nur für quadratische Matrizen definiert.

Was immer jemand tut, der mit irgendwelchen selbst erfundenen Rechenvorschriften die "Determinante" einer nichtquadratischen Matrix ausrechnet — es ist Unsinn!

2) Ohne die Normierungsbedingung (i) wäre die Determinantenfunktion nur bis auf einen Faktor festgelegt, da auch die Funktion $\mathbb{R}^{n \times n} \ni A \mapsto s \cdot \det(A)$ die Eigenschaften (ii) und (iii) hat für jedes $s \in \mathbb{R}$. Die Bedeutung von (i) ist also lediglich, diesen Faktor festzulegen, und die Normierungsbedingung (i) ist die sinnvollste Art und Weise, dies zu tun.

3) Eine Funktion von k (Spalten-)Vektoren aus \mathbb{R}^n nennt man *symmetrisch*, wenn sie ihren Wert bei Vertauschung von zwei Vektoren nicht ändert, und *antisymmetrisch*, wenn sie ihren Wert bei dieser Prozedur um den Faktor -1 ändert. In diesem Sinne ist also die Determinante $\det(A)$ eine antisymmetrische Funktion von n Vektoren, nämlich von den n Spalten der $n \times n$ -Matrix A . Die Regel gilt nur für die Vertauschung von *zwei* Spalten in der Matrix. Allgemein folgt:

- Wenn man mehrere Spalten der Matrix untereinander permutiert (d.h. in beliebige andere Reihenfolge bringt), so multipliziert sich die Determinante mit dem Faktor $(-1)^k$, wo k die notwendige Anzahl der Vertauschungen von je zwei Spalten ist, mit denen dieselbe Permutation erreicht wird.

Bringt man also z.B. die Spalten mit den Nummern 1, 2, 3 in die Reihenfolge 2, 3, 1, so ändert das die Determinante überhaupt nicht, weil man mit zwei Spaltenvertauschungen denselben Effekt erreicht (nämlich mit der Vertauschung von erster und zweiter Spalte und anschließender Vertauschung von zweiter und dritter Spalte in der nach der ersten Vertauschung entstandenen Matrix). Eine andere Konsequenz der Antisymmetrie ist, dass die Determinante verschwindet, wenn zwei Spalten der Matrix gleich sind; denn bei Vertauschung dieser Spalten multipliziert sie sich einerseits mit dem Faktor -1 , andererseits bleibt sie unverändert, weil die Matrix dieselbe bleibt.

- Die Determinante ist Null, wenn die Matrix zwei gleiche Spalten hat.

4) Als *Notation* für die Determinante einer Matrix $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ ist auch gebräuchlich

$$|A| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} := \det(A) = \det \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

Diese Bezeichnungsweise birgt die Gefahr in sich, dass man die Determinante $|A|$ wie den Betrag einer Zahl für eine stets nichtnegative Größe halten könnte, was sie natürlich nicht ist; denn sie ändert ja z.B. bei Spaltenvertauschung das Vorzeichen (wenn sie $\neq 0$ ist). Außerdem wird $|A|$ auch zur Bezeichnung der Euklidischen Norm der Matrix A verwendet, also der Quadratwurzel aus der Quadratsumme aller Matrixeinträge, und das ist eine ganz andere Größe als $\det(A)$. Von der Notation $|A|$ für die Determinante von A sollte man daher absehen. Bei Anordnung der Matrixeinträge in einem rechteckigen Schema ist es aber praktisch, die Bildung der Determinante einfach dadurch anzuzeigen, dass man die runden Matrix-Klammern durch gerade vertikale Striche ersetzt.

Eine andere Bezeichnungsweise, in der besser zum Ausdruck kommt, dass man die Determinante als eine Funktion von n (Spalten-)Vektoren ansehen kann, ist

$$\det(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) := \det(A)$$

wenn $A = (a_{ij})$ die $n \times n$ -Matrix mit den Spalten $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n \in \mathbb{R}^n$ ist, also $\mathbf{a}_j \in \mathbb{R}^n$ die Einträge $a_{1j}, a_{2j}, \dots, a_{nj}$ hat für $j = 1 \dots n$. Dann hängt $\det(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$ linear ab von jedem der n eingesetzten Vektoren \mathbf{a}_j , und bei Vertauschung von zwei der eingesetzten Vektoren ändert sich $\det(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$ um einen Faktor -1 .

5) Die Homogenität der Determinantenfunktion in jeder Matrixspalte bedeutet, dass das *Herausziehen eines skalaren Faktors aus einer Spalte* erlaubt ist. In der Notation von 4) geschrieben lautet diese Regel für jedes $r \in \mathbb{R}$ und alle $j \in \{1, \dots, n\}$:

$$\begin{vmatrix} a_{11} & \cdots & ra_{1j} & \cdots & a_{1n} \\ a_{21} & \cdots & ra_{2j} & \cdots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \cdots & ra_{nj} & \cdots & a_{nn} \end{vmatrix} = r \begin{vmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \cdots & a_{nj} & \cdots & a_{nn} \end{vmatrix}$$

oder

$$\det(\mathbf{a}_1, \dots, r\mathbf{a}_j, \dots, \mathbf{a}_n) = r \cdot \det(\mathbf{a}_1, \dots, \mathbf{a}_j, \dots, \mathbf{a}_n).$$

Eine *Warnung* vor einem nicht seltenen Fehler ist hier angebracht: Wenn mehrere Spalten der Matrix mit einem Faktor multipliziert werden, so muss natürlich aus *jeder* Spalte der entsprechende Faktor herausgezogen werden,

$$\det(r_1\mathbf{a}_1, r_2\mathbf{a}_2, \dots, r_n\mathbf{a}_n) = r_1 \cdot r_2 \cdot \dots \cdot r_n \cdot \det(A).$$

Wird insbesondere die ganze Matrix A , also jede ihrer n Spalten, mit dem Faktor $r \in \mathbb{R}$ multipliziert, so haben wir den Faktor r^n herauszuziehen:

$$\det(rA) = r^n \cdot \det(A) \quad \text{für } A \in \mathbb{R}^{n \times n}.$$

Das wird als *Homogenität vom Grad n der Determinantenfunktion auf $\mathbb{R}^{n \times n}$* bezeichnet. Der Spezialfall $r = -1$ lautet:

$$\det(-A) = (-1)^n \det(A) = \begin{cases} \det(A), & \text{wenn } n \text{ gerade ist;} \\ -\det(A), & \text{wenn } n \text{ ungerade ist.} \end{cases}$$

6) Die Additivität der Determinantenfunktion in jeder Spalte bedeutet analog, dass das *Herausziehen einer Summe aus einer Spalte* erlaubt ist. In der Notation von 4) geschrieben lautet diese Regel für alle $j \in \{1, \dots, n\}$ sowie alle Spaltenvektoren $\mathbf{a}'_j = (a'_{1j}, \dots, a'_{nj})$ und $\mathbf{a}''_j = (a''_{1j}, \dots, a''_{nj})$ aus \mathbb{R}^n :

$$\begin{vmatrix} a_{11} & \cdots & (a'_{1j} + a''_{1j}) & \cdots & a_{1n} \\ a_{21} & \cdots & (a'_{2j} + a''_{2j}) & \cdots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \cdots & (a'_{nj} + a''_{nj}) & \cdots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & \cdots & a'_{1j} & \cdots & a_{1n} \\ a_{21} & \cdots & a'_{2j} & \cdots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \cdots & a'_{nj} & \cdots & a_{nn} \end{vmatrix} + \begin{vmatrix} a_{11} & \cdots & a''_{1j} & \cdots & a_{1n} \\ a_{21} & \cdots & a''_{2j} & \cdots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \cdots & a''_{nj} & \cdots & a_{nn} \end{vmatrix},$$

oder

$$\det(\mathbf{a}_1, \dots, \mathbf{a}'_j + \mathbf{a}''_j, \dots, \mathbf{a}_n) = \det(\mathbf{a}_1, \dots, \mathbf{a}'_j, \dots, \mathbf{a}_n) + \det(\mathbf{a}_1, \dots, \mathbf{a}''_j, \dots, \mathbf{a}_n).$$

(Dabei stehen die Einträge mit Subskript "j" natürlich in der j-ten Spalte der Matrix; die Einträge in den anderen Spalten sind links und bei den Matrizen rechts jeweils dieselben.)

7) Wenn man die Homogenität und die Additivität der Determinante in jeder Spalte der eingesetzten Matrix kombiniert, so erkennt man, dass auch das *Herausziehen beliebiger Linearkombinationen aus jeder Spalte* erlaubt ist, also

$$\det(\mathbf{a}_1, \dots, \mathbf{a}_{j-1}, \sum_{k=1}^l s_k \mathbf{b}_k, \mathbf{a}_{j+1}, \dots, \mathbf{a}_n) = \sum_{k=1}^l s_k \cdot \det(\mathbf{a}_1, \dots, \mathbf{a}_{j-1}, \mathbf{b}_k, \mathbf{a}_{j+1}, \dots, \mathbf{a}_n)$$

für alle $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^n$, $j \in \{1, \dots, n\}$, $l \in \mathbb{N}$, $\mathbf{b}_1, \dots, \mathbf{b}_l \in \mathbb{R}^n$ und $s_1, \dots, s_l \in \mathbb{R}$. Dies ist die allgemeine Form der *Multilinearität der Determinantenfunktion* von n (Spalten-) Vektoren in \mathbb{R}^n , also der linearen Abhängigkeit von jedem der n Vektoren.

8) Es ist keineswegs selbstverständlich, dass es eine, und nur eine, Funktion \det mit den im vorigen Satz angegebenen Eigenschaften gibt. Das zu beweisen, ist Sache der Mathematik. Weil sich im Beweis aber eine explizite Formel ergibt, welche die Determinante einer quadratischen Matrix durch ihre Einträge ausdrückt und viele Eigenschaften der Determinantenfunktion verständlich macht, skizzieren wir den Beweis hier:

Wir zeigen zuerst die *Eindeutigkeit*, dass also die Determinantenfunktion die einzige mit den Eigenschaften (i), (ii) und (iii) ist. Ausgangspunkt ist die Darstellung der Spaltenvektoren \mathbf{a}_j von $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ als Linearkombination der kanonischen Basisvektoren e_i von \mathbb{R}^n ,

$$\mathbf{a}_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{pmatrix} = \sum_{i=1}^n a_{ij} e_i.$$

Mit der Linearität der Determinante in der ersten Spalte 7) folgt dann

$$\begin{aligned} \det(A) &= \det(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) \\ &= \det\left(\sum_{i=1}^n a_{i1} e_i, \mathbf{a}_2, \dots, \mathbf{a}_n\right) = \sum_{i=1}^n a_{i1} \cdot \det(e_i, \mathbf{a}_2, \dots, \mathbf{a}_n). \end{aligned}$$

Nutzen wir in den zuletzt auftretenden Determinanten die Linearität in der zweiten Spalte aus, so folgt weiter:

$$\det(A) = \sum_{i=1}^n \sum_{i'=1}^n a_{i1} a_{i'2} \cdot \det(e_i, e_{i'}, \mathbf{a}_3, \dots, \mathbf{a}_n).$$

So fahren wir fort, bis alle n Linearkombinationen der Basisvektoren herausgezogen sind. Wir erhalten dann eine n -fach iterierte Summe und brauchen dafür n verschiedene Summationsindizes i, i', i'', \dots , die wir mit i_1, i_2, \dots, i_n numerieren. Das Ergebnis ist

$$\det(A) = \sum_{i_1=1}^n \sum_{i_2=1}^n \dots \sum_{i_n=1}^n a_{i_1 1} a_{i_2 2} \dots a_{i_n n} \cdot \det(e_{i_1}, e_{i_2}, \dots, e_{i_n}).$$

In der letzten Determinante stehen nun lauter kanonische Basisvektoren. Die Determinante ist Null, wenn zwei von ihnen gleich sind, also muss nur über solche Indexkombinationen (i_1, i_2, \dots, i_n) summiert werden, die eine *Permutation* der Zahlen $1, 2, \dots, n$ darstellen, d.h. diese Zahlen in irgendeiner Reihenfolge aufführen. Dann kann die Matrix mit den kanonischen Spalten $e_{i_1}, e_{i_2}, \dots, e_{i_n}$ durch Spaltenvertauschungen in die Einheitsmatrix \mathbb{I}_n verwandelt werden, also ist ihre Determinante gleich $(-1)^m$, wobei m die Anzahl der Vertauschungen ist, mit denen (i_1, i_2, \dots, i_n) in $(1, 2, \dots, n)$ zu überführen ist. Bezeichnen wir dieses von i_1, i_2, \dots, i_n abhängige Vorzeichen ("Signum") mit $\text{sign}(i_1, i_2, \dots, i_n) = \pm 1$, so haben wir die Formel

$$\det(A) = \sum_{i_1, i_2, \dots, i_n} \text{sign}(i_1, i_2, \dots, i_n) a_{i_1 1} a_{i_2 2} \dots a_{i_n n},$$

wobei über alle Permutationen der Zahlen $1, 2, \dots, n$ zu summieren ist. Damit ist die Eindeutigkeit gezeigt: Wenn es eine Funktion $\det: \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ mit den Eigenschaften (i), (ii) und (iii) gibt, so muss $\det(A)$ aus den Einträgen der Matrix $A = (a_{ij})$ so berechnet werden können wie in dieser Formel.

Gleichzeitig können wir nun die *Existenz* der gesuchten Funktion zeigen, indem wir sie durch eben diese Formel definieren und dann die Eigenschaften (i),(ii),(iii) nachweisen. (i) gilt, weil im Fall der Einheitsmatrix nur der zu $i_1 = 1, i_2 = 2, \dots, i_n = n$ gehörende Summand von Null verschieden ist und den Wert 1 hat. (iii) gilt, weil jedes in der Summe auftretende Produkt von Matrixeinträgen aus jeder Spalte der Matrix genau einen Eintrag als Faktor enthält. Für (ii) muss man nur bemerken, dass die Vertauschung von Spalte Nr. j mit Spalte Nr. k (mit $j < k$), auf dasselbe hinausläuft wie die Ersetzung der Vorzeichen $\text{sign}(i_1, \dots, i_j, \dots, i_k, \dots, i_n)$ durch $\text{sign}(i_1, \dots, i_k, \dots, i_j, \dots, i_n)$; die beiden Vorzeichen unterscheiden sich aber gerade durch einen Faktor -1 , weil sich die beiden Permutationen eben um eine Vertauschung unterscheiden. Damit alles korrekt ist, muss man noch überlegen, dass das Vorzeichen $\text{sign}(i_1, i_2, \dots, i_n)$ einer Permutation eindeutig definiert ist, d.h. dass die Anzahl der Vertauschungen von je zwei Gliedern, mit denen man (i_1, i_2, \dots, i_n) in $(1, 2, \dots, n)$ bringen kann, unabhängig von der konkreten Wahl dieser Vertauschungen immer gerade oder immer ungerade ist. Das kann man z.B. einsehen, indem man überlegt, dass im Produkt $\prod_{1 < h < l \leq n} (i_l - i_h)$ gerade $2(k-j) + 1$ Faktoren ihr Vorzeichen ändern, wenn man i_j mit i_k vertauscht für ein Paar von Zahlen $j < k$. Man kann also $\text{sign}(i_1, i_2, \dots, i_n)$ einfach als das Vorzeichen dieses Produkts definieren (das offenbar positiv ist, wenn $i_1 = 1, i_2 = 2, \dots, i_n = n$). ■

Die explizite Formel, die sich für die Berechnung der Determinante aus den Matrixeinträgen ergeben hat, ist zwar für die Berechnung der Determinante von großen quadratischen Matrizen ungeeignet (weil die auftretende Summe zu viele Summanden hat, nämlich $n!$ (Fakultät von n) Summanden im Falle von $n \times n$ -Matrizen), aber bei 3×3 -Matrizen ist sie gut brauchbar, und sie liefert weitere interessante Eigenschaften der Determinantenfunktion. Deshalb besprechen wir die Formel noch etwas genauer in folgender

DISKUSSION mit BEISPIELEN: 1) Für $n \times n$ -Matrizen $A = (a_{ij})$ gilt die sog. **Leibnizsche Formel**:

$$\det(A) = \sum_{i_1, i_2, \dots, i_n} \text{sign}(i_1, i_2, \dots, i_n) a_{i_1 1} a_{i_2 2} \cdots a_{i_n n},$$

wobei über alle Permutationen (i_1, i_2, \dots, i_n) der Zahlen $1, 2, \dots, n$ zu summieren ist und das Vorzeichen $\text{sign}(i_1, i_2, \dots, i_n)$ gleich $(-1)^m$ ist, wenn (i_1, i_2, \dots, i_n) mit m Vertauschungen von jeweils zwei Gliedern in $(1, 2, \dots, n)$ überführt werden kann. Da diese Formel, die wir oben hergeleitet haben, für Nicht-Mathematiker schwer zu entschlüsseln ist, geben wir ihren Inhalt in Worten an:

- Die Determinante einer $n \times n$ -Matrix erhält man, indem man auf alle möglichen Arten n Matrixpositionen herausgreift, derart dass keine zwei aus derselben Spalte oder derselben Zeile genommen werden, die Produkte der in den herausgegriffenen Positionen befindlichen n Matrixeinträge bildet, sie mit dem Vorzeichenfaktor $+1$ bzw. -1 multipliziert, wenn die herausgegriffenen Positionen durch eine gerade bzw. ungerade Anzahl von Spaltenvertauschungen auf die Matrixdiagonale gebracht werden können, und schließlich alle so gebildeten und mit Vorzeichenfaktoren versehenen Produkte aufsummiert.

2) Eine unmittelbare Folgerung aus der Leibnizschen Formel ist:

- Die Determinantenfunktion $\det: \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ ist eine homogene Polynomfunktion der Matrixeinträge vom Grad n .

Dies bedeutet, dass $\det(A)$ für $A \in \mathbb{R}^{n \times n}$ durch eine Formel zu berechnen ist, bei der Produkte von jeweils n Matrixeinträgen gebildet werden, die nach einer festen Regel aus der Matrix herausgegriffen werden (hier sogar so, dass keine Faktoren eines solchen Produkts derselben Zeile oder derselben Spalte entnommen sind), und dann mit festgelegten Koeffizienten (hier nur $+1$ und -1) linear kombiniert werden. Die schon besprochene Rechenregel $\det(rA) = r^n \det(A)$ für $A \in \mathbb{R}^{n \times n}$ ist eine unmittelbare Folgerung.

3) Für 1×1 -Matrizen $A = (a)$ ist natürlich

$$\det(a) = a$$

der einzige Eintrag der Matrix. Mehr gibt es hierzu nicht zu sagen.

4) Für $n = 2$ gibt es genau zwei Permutationen von $(1, 2)$, nämlich $(1, 2)$ (die "identische Permutation", bei der gar nichts vertauscht wird) und die Vertauschung $(2, 1)$. Erstere hat das Vorzeichen $+1$, letztere das Vorzeichen -1 , also lautet die Leibnizsche Formel für die *Determinante von 2×2 -Matrizen*:

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11}a_{22} - a_{21}a_{12}.$$

Das ist genau dieselbe Formel $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$, mit der wir die Determinante von 2×2 -Matrizen schon früher definiert hatten.

5) Für $n = 3$ gibt es bereits 6 Permutationen $(1, 2, 3)$, $(3, 1, 2)$, $(2, 3, 1)$ mit dem Vorzeichen $+1$ und $(3, 2, 1)$, $(1, 3, 2)$, $(2, 1, 3)$ mit dem Vorzeichen -1 . Die Leibnizsche Formel für die *Determinante einer 3×3 -Matrix $A = (a_{ij})$* lautet also:

$$\det(A) = a_{11}a_{22}a_{33} + a_{31}a_{12}a_{23} + a_{21}a_{32}a_{13} - a_{31}a_{22}a_{13} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33}.$$

Mit folgendem Schema kann man sich diese Formel einprägen: Man schreibt die beiden ersten Spalten der Matrix nochmals rechts neben die Matrix, addiert die Produkte der jeweils drei Einträge auf jeder der drei Diagonalen von links oben nach rechts unten und subtrahiert die Produkte der drei Einträge auf jeder der drei Diagonalen von rechts oben nach links unten. Dies ist die sog. **Sarrussche Regel**:

$$\begin{array}{c}
 \begin{array}{ccc|cc}
 a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\
 a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\
 a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \\
 \hline
 \ominus & \ominus & \ominus & \oplus & \oplus & \oplus
 \end{array}
 &
 \begin{array}{c}
 \left| \begin{array}{ccc}
 a_{11} & a_{12} & a_{13} \\
 a_{21} & a_{22} & a_{23} \\
 a_{31} & a_{32} & a_{33}
 \end{array} \right| \\
 = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\
 - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33}
 \end{array}
 \end{array}$$

Wir betrachten als konkretes Beispiel die Determinante der "Telefon-Matrix":

$$\left| \begin{array}{ccc}
 1 & 2 & 3 \\
 4 & 5 & 6 \\
 7 & 8 & 9
 \end{array} \right| = \left\{ \begin{array}{l} 1 \cdot 5 \cdot 9 + 2 \cdot 6 \cdot 7 + 3 \cdot 4 \cdot 8 \\ -3 \cdot 5 \cdot 7 - 1 \cdot 6 \cdot 8 - 2 \cdot 4 \cdot 9 \end{array} \right\} = 0.$$

Die Determinante der Telefon-Matrix hat also den Wert Null.

6) Für $n = 4$ hat die Leibnizsche Formel für die Determinante einer $n \times n$ -Matrix schon 24 Summanden, für $n = 5$ gar 120; denn die Anzahl der Permutationen von $1, 2, \dots, n$ ist ja $n!$ (Fakultät von n). Das zur Sarrusschen Regel analoge Schema bei 4×4 -Matrizen gäbe aber nur 8 Summanden und kann schon deswegen keine korrekte Berechnung der Determinante sein. Das wird aber nicht selten falsch gemacht, daher die *Warnung*:

- *Bei quadratischen Matrizen vom Format 4×4 oder größer gilt keine "Sarrussche Regel" für die Berechnung der Determinante!*

7) Die Leibnizsche Formel für die Determinante hat eine Symmetrie, die zeigt, dass alle Eigenschaften der Determinante als Funktion der Spalten einer quadratischen Matrix analog für Zeilen gelten. Um das zu sehen, ordnen wir die Faktoren in einem nach Spaltennummern sortierten Produkt von Matrixeinträgen so um, dass die Faktoren mit aufsteigenden Zeilennummern aufeinander folgen:

$$a_{i_1 1} a_{i_2 2} \cdots a_{i_n n} = a_{1 j_1} a_{2 j_2} \cdots a_{n j_n} .$$

Dabei ist (j_1, j_2, \dots, j_n) die *Umkehrpermutation* der Permutation (i_1, i_2, \dots, i_n) von $1, 2, \dots, n$, d.h. die Permutation, die man erhält, wenn man dieselben Vertauschungen von Gliedern, die (i_1, i_2, \dots, i_n) in $(1, 2, \dots, n)$ überführen, auf $(1, 2, \dots, n)$ anwendet. Insbesondere hat (j_1, j_2, \dots, j_n) dasselbe Vorzeichen wie (i_1, i_2, \dots, i_n) , und deshalb kann man die Leibnizsche Formel auch schreiben:

$$\det(A) = \sum_{j_1, j_2, \dots, j_n} \text{sign}(j_1, j_2, \dots, j_n) a_{1 j_1} a_{2 j_2} \cdots a_{n j_n} ,$$

weil mit (i_1, i_2, \dots, i_n) auch die Umkehrpermutationen (j_1, j_2, \dots, j_n) alle Permutationen von $1, 2, \dots, n$ durchlaufen. Nun ist aber a_{ij} der Eintrag in Position (j, i) der transponierten Matrix A^T zur Matrix $A = (a_{ij})$ und man erkennt, dass die Summe in der letzten Formel nichts anderes ist als die Determinante von A^T .

- *Die $n \times n$ -Matrix A und ihre transponierte Matrix A^T (deren Spalten die Zeilen von A sind) haben dieselbe Determinante:*

$$\det(A) = \det(A^T) .$$

- *Folglich gelten alle bisher und künftig in Bezug auf die Matrixspalten formulierten Eigenschaften der Determinantenfunktion völlig analog auch in Bezug auf die Zeilen.*

Also ist z.B. $\det(A)$ eine lineare Funktion in jeder Zeile der Matrix A , und bei Vertauschung von zwei Zeilen in der Matrix ändert sich die Determinante auch um den Faktor -1 , genau wie bei Spaltenvertauschungen. ■

Die Leibnizsche Formel ist, wie gesagt, nicht gut zur Berechnung der Determinante von größeren quadratischen Matrizen geeignet. Aus den grundlegenden Eigenschaften der Determinantenfunktion folgen aber Rechenregeln für Determinanten, mit denen man auch die Determinante größerer quadratische Matrizen effektiv berechnen kann. Diese stellen wir zusammen in folgenden

RECHENREGELN für Determinanten: Hier sei A stets eine $n \times n$ -Matrix.

1) Wir haben als Folge der Antisymmetrie der Determinantenfunktion in den Spalten der Matrix schon vermerkt, dass die Determinante einer Matrix verschwindet, wenn sie zwei gleiche Spalten hat. Das gilt wegen der Homogenität dann natürlich auch, wenn eine Spalte Vielfaches einer anderen ist; denn nach Herausziehen eines Faktors aus einer Spalte verbleibt dann ja die Determinante einer Matrix mit zwei gleichen Spalten. Und wenn A eine Nullspalte hat, so kann man den Faktor 0 herausziehen, also ist auch dann die Determinante gleich Null. Alles gilt, wie gesagt, analog auch für Zeilen statt Spalten.

- $\det(A) = 0$, wenn A eine Nullspalte (bzw. eine Nullzeile) hat, oder zwei gleiche Spalten (bzw. Zeilen), oder zwei Spalten (bzw. Zeilen), die sich nur um einen skalaren Faktor unterscheiden.

2) Folgende Zeilen- und Spaltenoperationen mit Determinantentermen $\det(A)$ sind zulässig, d.h. sie verändern den Wert des Terms nicht:

- (i) Man addiert ein Vielfaches einer Zeile / Spalte von A zu einer anderen;
- (ii) man vertauscht zwei Zeilen / Spalten in A und multipliziert die Determinante mit -1 ;
- (iii) man zieht aus einer Zeile / Spalte von A einen Faktor $r \in \mathbb{R}$ vor die Determinante.

Damit ist natürlich gemeint $\det(A) = \det(A') = -\det(A'') = r \det(A''')$, wenn A' aus A durch Addition einer Zeile / Spalte zu einer anderen entsteht (natürlich darf man nie Zeilen zu Spalten addieren oder umgekehrt — das wäre völliger Unfug!), wenn A'' aus A durch eine Vertauschung von zwei Zeilen / Spalten hervorgeht, und wenn A aus A''' gebildet wird durch Multiplikation einer Zeile / Spalte mit dem Faktor r . Die Regel (iii) folgt unmittelbar aus der Homogenität der Determinantenfunktion in den Spalten (und den Zeilen) der Matrix, und die Regel (ii) ist nichts anderes als die Antisymmetrie der Determinante als Funktion der Spalten (bzw. Zeilen). Die Regel (i) gilt für die Spalten $\mathbf{a}_1, \dots, \mathbf{a}_n$ von A , weil z.B. $\det(A') = \det(\dots, \mathbf{a}_j + s\mathbf{a}_k, \dots, \mathbf{a}_k, \dots)$ ist für gewisse $j \neq k$ aus $\{1, \dots, n\}$ und $s \in \mathbb{R}$. (Die angegebenen Spalten befinden sich in den Positionen j und k ; die nicht angegebenen Spalten sind dieselben wie bei A ; der Fall $j > k$ ist analog.) Wegen der Linearität der Determinante in der j -ten Spalte ist folglich $\det(A')$ gleich der Summe von $\det(\dots, \mathbf{a}_j, \dots, \mathbf{a}_k, \dots) = \det(A)$ und von $s \cdot \det(\dots, \mathbf{a}_k, \dots, \mathbf{a}_k, \dots) = 0$ (weil es sich ja um die Determinante einer Matrix mit zwei gleichen Spalten handelt), d.h. es gilt $\det(A') = \det(A)$ wie behauptet.

3) Der Nutzen der Zeilen- und Spaltenoperationen bei der Determinantenberechnung besteht darin, dass man mit solchen Operationen die Matrix vereinfachen kann, ohne ihre Determinante zu ändern bzw. mit Änderung nur um genau bekannte Vorfaktoren. Ziel der Vereinfachungen ist natürlich, eine Form der Matrix zu erreichen, für die man die Determinante unmittelbar ablesen kann. Das ist zum Beispiel bei Dreiecksgestalt der Fall:

- Die Determinante einer Dreiecksmatrix ist das Produkt der Diagonaleinträge.

Ist nämlich ein Diagonaleintrag $= 0$, so kann man die Dreiecksmatrix A durch Zeilenoperationen wie in 2) auf eine Zeilen-Stufen-Form mit mindestens einer Nullzeile bringen, also ist dann die Determinante nach 1) gleich Null und ebenso auch das Produkt der Diagonaleinträge der Matrix. Sind aber alle Diagonaleinträge $a_{ii} \neq 0$, so kann man durch

Es gilt dann der sog. **Determinanten-Entwicklungssatz** von Laplace:

- Die Determinante von A ist die Linearkombination der Kofaktoren von A zu allen Positionen in einer beliebig gewählten Zeile oder Spalte mit den an diesen Positionen befindlichen Einträgen von A als Koeffizienten; in Formeln:

$$\det(A) = \sum_{j=1}^n a_{ij}(-1)^{i+j} \det(A_{ij}) \quad (\text{Entwicklung nach der } i\text{-ten Zeile}),$$

$$\det(A) = \sum_{i=1}^n a_{ij}(-1)^{i+j} \det(A_{ij}) \quad (\text{Entwicklung nach der } j\text{-ten Spalte}),$$

wobei in der ersten Formel i die Nummer der ausgewählten Zeile ist und in der zweiten Formel j die Nummer der ausgewählten Spalte. Mit dem Entwicklungssatz wird die Berechnung der Determinante einer $n \times n$ -Matrix zurückgeführt auf die Berechnung der Determinanten von maximal n Matrizen des Formats $(n-1) \times (n-1)$. In der Praxis wählt man die Zeile oder Spalte, nach der die Determinante "entwickelt" wird, natürlich so, dass die Rechnungen möglichst einfach werden, dass also z.B. viele Nulleinträge in der Zeile / Spalte stehen.

Um den Entwicklungssatz zu beweisen, schreiben wir z.B. die j -te Spalte von A als Linearkombination $\sum_{i=1}^n a_{ij}e_i$ der kanonischen Basisvektoren und ziehen die Linearkombination aus der Determinante heraus: $\det(A) = \sum_{i=1}^n a_{ij} \det(\tilde{A}_{ij})$, wobei \tilde{A}_{ij} aus A entsteht, indem man die j -te Spalte durch e_i ersetzt. Durch $j-1$ Vertauschungen unmittelbar nebeneinander liegender Spalten und $i-1$ Vertauschungen unmittelbar übereinander liegender Zeilen in \tilde{A}_{ij} bringt man die Position (i, j) , auf der nun 1 eingetragen ist, in die Position $(1, 1)$ und erhält so die Matrix \hat{A}_{ij} mit erster Spalte e_1 und mit der Restmatrix A_{ij} als diagonalen Block in Zeilen und Spalten Nr. 2... n . Nach 2) und 5) gilt dann $\det(\tilde{A}_{ij}) = (-1)^{i-1+j-1} \det(\hat{A}_{ij}) = (-1)^{i+j} \cdot 1 \cdot \det(A_{ij})$.

Es gibt noch eine allgemeinere Version des Entwicklungssatzes, wobei man nach k fest gewählten Spalten (bzw. nach k fest gewählten Zeilen) entwickelt ($1 \leq k < n$). Man greift dann auf alle möglichen Arten k Zeilen heraus (bzw. k Spalten) und bildet das Produkt der Determinante der durch die jeweils betrachteten k Spalten und k Zeilen gegebenen $k \times k$ -Untermatrix (gebildet aus den Positionen, in denen sich diese Spalten und Zeilen kreuzen) mit der Determinante der nach Streichen dieser Spalten und Zeilen verbleibenden $(n-k) \times (n-k)$ -Restmatrix. Das Produkt wird noch multipliziert mit dem Vorzeichen $(-1)^{\text{hoch Zahl der Vertauschungen benachbarter Spalten oder Zeilen, mit denen man die } k \times k\text{-Untermatrix in eine zur Diagonalen symmetrische Lage bringen kann, und dann ist die Summe der mit diesen Vorzeichen versehenen Determinantenprodukte über alle } \binom{n}{k} \text{ Möglichkeiten aus } n \text{ Spalten (bzw. Zeilen) } k \text{ herauszugreifen, gleich der Determinante der gegebenen } n \times n\text{-Matrix. Dies ist der } \textit{allgemeine Laplacesche Entwicklungssatz}$, der aber mehr theoretische als praktische Bedeutung hat; deshalb unterlassen wir es, dafür eine Formel aufzuschreiben (die schwerer zu dekodieren wäre als die gegebene Beschreibung).

7) Es gibt eine — auf den ersten Blick überraschend einfache — Regel für die Berechnung der Determinante eines Produkts von zwei gleichformatigen quadratischen Matrizen:

- **Determinanten-Produktsatz:** Die Determinante des Produkts zweier (gleichformatiger, quadratischer) Matrizen A, B ist das Produkt ihrer Determinanten,

$$\det(AB) = \det(A) \cdot \det(B).$$

Man sagt dazu auch: Die Determinantenfunktion ist multiplikativ auf $\mathbb{R}^{n \times n}$.

Unmittelbare Folgerungen sind (mit $k \in \mathbb{N}$):

$$\begin{aligned} \det(A^k) &= (\det(A))^k, & \det(AB) &= \det(BA), \\ \det(A^{-1}) &= \frac{1}{\det(A)}, & \det(A^{-k}) &= \frac{1}{(\det(A))^k}, & \text{wenn } A \text{ invertierbar ist,} \\ \det(C^{-1}AC) &= \det(A), & & & \text{wenn } C \text{ invertierbar ist.} \end{aligned}$$

Die letzte Gleichung besagt, dass *ähnliche Matrizen dieselbe Determinante haben*. Man kann den Produktsatz auf mehrere Arten beweisen, z.B. mit den in 3.3 eingeführten Elementarmatrizen $M_i(r)$, $A_{ij}(s)$ und V_{ij} , deren Multiplikation von links an A entsprechende Zeilenoperationen bei A bewirken. Aus 2) und 3) sieht man $\det(M_i(r)) = r$, $\det(A_{ij}(s)) = 1$, $\det(V_{ij}) = -1$ und daher $\det(EA) = \det(E)\det(A)$, wenn E irgendeine Elementarmatrix ist. Ist A invertierbar, so kann man A durch Zeilenoperationen in die Einheitsmatrix transformieren, d.h. es gibt Elementarmatrizen E_1, \dots, E_k mit $E_k \cdots E_1 A = \mathbb{I}$. Dann folgt

$$1 = \det(\mathbb{I}) = \det(E_k \cdots E_1 A) = \det(E_k) \cdots \det(E_1) \cdot \det(A)$$

und

$$\det(B) = \det(\mathbb{I}B) = \det(E_k \cdots E_1 AB) = \det(E_k) \cdots \det(E_1) \cdot \det(AB),$$

also $\det(B) = \frac{1}{\det(A)} \det(AB)$. Ist aber A nicht invertierbar, so hat A eine Zeilen-Stufen-Form mit einer Nullzeile, also ist $\det(A) = 0$, und dieselben Zeilentransformationen, welche diese Zeilen-Stufen-Matrix herstellen, ergeben bei Anwendung auf AB ebenfalls eine Matrix mit Nullzeile, so dass auch $\det(AB) = 0$ ist.

Eleganter ist der Beweis des Produktsatzes, wenn man die charakterisierenden Eigenschaften der Determinantenfunktion aus ihrer Definition benutzt. Man sieht sofort, dass $\widetilde{\det}(B) := \det(AB)$ bei fester Matrix A eine antisymmetrische und multilineare Funktion der Spalten von B ist. Aus dem früheren Beweis der Eindeutigkeit der Determinantenfunktion folgt dann schon $\widetilde{\det}(B) = s \cdot \det(B)$ mit dem Faktor $s = \widetilde{\det}(\mathbb{I}) = \det(A)$, d.h. $\det(AB) = \det(A) \cdot \det(B)$.

Bei Verwendung der Notation $|A|$ für die Determinante von A erhält der Produktsatz die Form $|AB| = |A||B|$ eines Rechengesetzes für Beträge. (Aber die Determinante $|A|$ ist kein Betrag! Sie kann ja auch negative Werte annehmen.) ■

BEISPIELE (zur Determinantenberechnung);

1) Wir berechnen nochmals die Determinante der Telefon-Matrix, einmal mit Transformation auf Dreiecksform durch Zeilenoperationen,

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -12 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 0 & 0 \end{vmatrix} = 1 \cdot (-3) \cdot 0 = 0,$$

einmal mit Entwicklung nach der ersten Zeile,

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} = 1 \cdot \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} - 2 \cdot \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} + 3 \cdot \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} = 1 \cdot (5 \cdot 9 - 6 \cdot 8) - 2 \cdot (4 \cdot 9 - 6 \cdot 7) + 3 \cdot (4 \cdot 8 - 5 \cdot 7) = 0.$$

Die zweite Berechnungsart ist hier aufwendiger, weil man keine Nulleinträge in der Matrix nutzen kann. Bei der ersten Berechnung hätte man auf die zweite Matrix auch den Kästchen-Satz anwenden können und $1 \cdot \begin{vmatrix} -3 & -6 \\ -6 & -12 \end{vmatrix} = 0$ als Ergebnis erhalten. Noch besser wäre es gewesen, gleich zu erkennen, dass die letzte Zeile dieser Matrix das 2-fache der mittleren Zeile ist und daher die Determinante verschwindet.

2) Wir berechnen die Determinante von

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

mit dem Entwicklungssatz. Es bietet sich die Entwicklung nach der zweiten Spalte ein, weil diese zwei Nulleinträge hat:

$$\begin{vmatrix} 1 & 0 & 2 & -1 \\ 3 & -2 & 0 & 2 \\ 1 & 2 & -1 & 0 \\ 3 & 0 & 2 & -2 \end{vmatrix} = (-2)(-1)^{2+2} \begin{vmatrix} 1 & 2 & -1 \\ 3 & 2 & -2 \end{vmatrix} + 2(-1)^{3+2} \begin{vmatrix} 1 & 2 & -1 \\ 3 & 0 & 2 \\ 3 & 2 & -2 \end{vmatrix}.$$

Die beiden noch zu berechnenden Determinanten von 3×3 -Matrizen kann man mit der Sarrusschen Regel ausrechnen oder z.B. durch Entwickeln nach der zweiten Zeile:

$$\begin{aligned} &= -2 \cdot \left[1(-1)^{2+1} \begin{vmatrix} 2 & -1 \\ 2 & -2 \end{vmatrix} + (-1)(-1)^{2+2} \begin{vmatrix} 1 & -1 \\ 3 & -2 \end{vmatrix} \right] \\ &\quad - 2 \cdot \left[3(-1)^{2+1} \begin{vmatrix} 2 & -1 \\ 2 & -2 \end{vmatrix} + 2(-1)^{2+3} \begin{vmatrix} 1 & 2 \\ 3 & 2 \end{vmatrix} \right] \\ &= -2 \cdot [(-1)(-4 + 2) + (-1)(-2 + 3)] - 2 \cdot [(-3)(-4 + 2) + (-2)(2 - 6)] \\ &= -2 \cdot [2 - 1] - 2 \cdot [6 + 8] = -30. \end{aligned}$$

3) Wir berechnen die Determinante der Matrix aus 2) nochmals, diesmal mit Transformation auf Dreiecksform durch Zeilen- und Spaltenoperationen. Zunächst addieren wir die zweite zur dritten Zeile, subtrahieren die erste von der vierten Zeile und klammern den Faktor -2 aus der zweiten Spalte aus (Letzteres muss man nicht tun; wir wollen aber auch diese Spaltenoperation vorführen),

$$\begin{vmatrix} 1 & 0 & 2 & -1 \\ 3 & -2 & 0 & 2 \\ 1 & 2 & -1 & 0 \\ 3 & 0 & 2 & -2 \end{vmatrix} = (-2) \cdot \begin{vmatrix} 1 & 0 & 2 & -1 \\ 3 & 1 & 0 & 2 \\ 4 & 0 & -1 & 2 \\ 2 & 0 & 0 & -1 \end{vmatrix},$$

dann addieren wir das Doppelte der letzten Spalte und das 8-fache der vorletzten zur ersten Spalte,

$$= (-2) \cdot \begin{vmatrix} 15 & 0 & 2 & -1 \\ 7 & 1 & 0 & 2 \\ 0 & 0 & -1 & 2 \\ 0 & 0 & 0 & -1 \end{vmatrix} = (-2)(-1)^2 \begin{vmatrix} 1 & 7 & 0 & 2 \\ 0 & 15 & 2 & -1 \\ 0 & 0 & -1 & 2 \\ 0 & 0 & 0 & -1 \end{vmatrix} = (-2)[1 \cdot 15 \cdot (-1) \cdot (-1)] = -30,$$

wobei wir im vorletzten Schritt die ersten beiden Spalten und die ersten beiden Zeilen vertauscht und im letzten Schritt die Determinante als Produkt der Diagonaleinträge der Dreiecksmatrix berechnet haben. Statt der beiden letzten Schritte hätten wir auch den Kästchen-Satz anwenden können und direkt $(-2)[(15 \cdot 1) \cdot (-1)^2] = -30$ erhalten. ■

Man sieht, dass man bei Determinantenberechnungen auf vielen verschiedenen Wegen zum Ziel kommen kann. Das Bestreben ist stets, die *Zeilen- und Spaltenoperationen so zu wählen, dass schon vorhandene Nulleinträge ausgenutzt werden*, und außerdem so, dass Brüche nach Möglichkeit gar nicht, oder erst möglichst spät in der Rechnung auftreten. Das letzte Beispiel demonstriert, dass die Transformation der Matrix auf Dreiecksgestalt mittels Zeilen- und Spaltenoperationen meist die einfachste Methode ist.

Mathematik für Wirtschaftswissenschaftler

(K. Steffen, Heinrich-Heine-Universität Düsseldorf, WS 2006/07)

Wir kommen nun zum Ausgangspunkt unserer Diskussion der Determinanten zurück: Wir hatten ja eine jeder quadratischen Matrix A zugeordnete und (verhältnismäßig) einfach zu berechnende Zahl $\det(A)$ gesucht, mit der man anhand des Kriteriums $\det(A) \neq 0$ die Invertierbarkeit der Matrix leicht feststellen kann. Dass dieses Invertierbarkeitskriterium tatsächlich gilt, hatten wir im Prinzip schon im Beweis des Produktsatzes gezeigt: Wenn A^{-1} existiert, so gilt $1 = \det(\mathbb{I}) = \det(AA^{-1}) = \det(A) \cdot \det(A^{-1})$, also muss dann $\det(A) \neq 0$ sein. Wenn andererseits A^{-1} nicht existiert, so lässt sich mit Zeilenoperationen eine Zeilen-Stufen-Form aus A herstellen, die eine Nullzeile hat, also ist dann $\det(A) = 0$. Man kann mit Determinanten aber nicht nur die Invertierbarkeit von A prüfen, sondern auch explizite Formeln für die Einträge von A^{-1} und damit auch für die Lösungen $x = A^{-1}b$ eines linearen Gleichungssystems $Ax = b$ angeben. Wir tun das in folgendem

SATZ (Determinante und inverse Matrix): Sei A eine $n \times n$ -Matrix. Dann gilt:

(i) **Determinantenkriterium für Invertierbarkeit:**

$$A \text{ invertierbar} \iff \det(A) \neq 0.$$

(ii) **Cramersche Regel für die Inverse:** Wenn sie existiert, so ist die Inverse A^{-1} die transponierte Kofaktor-Matrix zu A dividiert durch die Determinante von A ,

$$A^{-1} = (b_{ji}) \quad \text{mit} \quad b_{ji} = \frac{1}{\det(A)} (-1)^{i+j} \det(A_{ij}) \quad \text{für } i, j = 1 \dots n.$$

(iii) **Cramersche Regel für die Lösung von linearen Gleichungssystemen:** Wenn $\det(A) \neq 0$ ist, so ist die eindeutige Lösung $x \in \mathbb{R}^n$ des linearen Gleichungssystems $Ax = b$ zu $b \in \mathbb{R}^n$ gegeben durch

$$x_j = \frac{\det(A \text{ mit } j\text{-ter Spalte ersetzt durch } b)}{\det(A)} \quad \text{für } j = 1 \dots n.$$

Wir erinnern daran, dass A_{ij} die $(n-1) \times (n-1)$ -Restmatrix bezeichnet, die aus A durch Streichen der i -ten Zeile und der j -ten Spalte entsteht; $(-1)^{i+j} \det(A_{ij})$ heißt dann der **Kofaktor** von A zur Position (i, j) , und dies ist laut (ii) gerade der Eintrag in Position (j, i) bei der inversen Matrix, wenn sie existiert. Die Regel (iii) sagt, dass die j -te Komponente des Lösungsvektors der Quotient von zwei Determinanten ist, nämlich $\det(A)$ als Nenner (wenn $\neq 0$) und als Zähler die Determinante der Matrix, die aus A entsteht, indem man die j -te Spalte von A ersetzt durch den Spaltenvektor der rechten Seiten des Gleichungssystems.

Den Beweis von (i) haben wir oben schon gegeben. (ii) folgt aus dem Entwicklungssatz $\det(A) = \sum_{j=1}^n a_{ij}(-1)^{i+j} \det(A_{ij})$. Ersetzt man in der Summe a_{ij} durch a_{hj} mit einer Zeilennummer $h \neq i$, so läuft das auf die Bildung der Determinante einer Matrix mit zwei gleichen Zeilen hinaus, also hat die Summe dann den Wert Null. Somit gilt:

$$\sum_{j=1}^n a_{hj} \frac{(-1)^{i+j} \det(A_{ij})}{\det(A)} = \delta_{hi} = \begin{cases} 1 & \text{für } h = i, \\ 0 & \text{für } h \neq i, \end{cases}$$

d.h. für die Matrix $B = (b_{ji})$ mit den in (ii) angegebenen Einträgen gilt $AB = \mathbb{I}$ und damit (weil es sich um quadratische Matrizen handelt) $B = A^{-1}$. Die eindeutige Lösung $x = A^{-1}b$ zu $Ax = b$ hat dann die Komponenten

$$x_j = \sum_{i=1}^n b_{ji} b_i = \frac{1}{\det(A)} \sum_{i=1}^n b_i (-1)^{i+j} \det(A_{ij}).$$

Die letzte Summe kann aber nach dem Entwicklungssatz aufgefasst werden als die Entwicklung nach der j -ten Spalte der Matrix, die aus A mit Ersetzen der j -ten Spalte durch b entsteht, und das beweist (iii).

DISKUSSION: Die Cramersche Regel ist nicht geeignet zur Berechnung der inversen Matrix oder der Lösung eines quadratischen linearen Gleichungssystems — der Rechenaufwand ist viel zu groß! (Außer der Determinante von A braucht man noch n weitere Determinanten von $n \times n$ -Matrizen.) Die Bedeutung der Regel liegt vielmehr darin, dass sie Strukturaussagen über die Abhängigkeit der Inversen A^{-1} bzw. der Lösung zu $Ax = b$ von den Einträgen der quadratischen Matrix $A = (a_{ij})$ und der Spalte b der rechten Seiten erlaubt.

Da beim Gaußschen Eliminationsverfahren nur die Grundrechenarten angewendet werden, ist von vornherein klar, dass die Einträge von A^{-1} rationale Funktionen der Einträge von A sind. Die Cramersche Regel zeigt nun, dass man bei diesen rationalen Funktionen die Determinantenfunktion als Hauptnenner verwenden kann. Genauer sind die Einträge von A^{-1} homogene Polynomfunktionen des Grades $n-1$ in den Einträgen a_{ij} von A mit ganzen Koeffizienten dividiert durch die homogene Polynomfunktion $\det(A)$ des Grades n . Und die Komponenten des Lösungsvektors $x = A^{-1}b$ sind homogene Polynomfunktionen des Grades n in den Einträgen a_{ij} von A und b_i von b dividiert durch $\det(A)$. (Eine Funktion von mehreren Variablen heißt *Polynomfunktion von mehreren Veränderlichen*, wenn sie sich als Linearkombination von Produkten dieser Variablen schreiben läßt, wobei die Variablen in den Produkten auch mehrfach, also in Potenzen, auftreten dürfen. Kommen dabei immer nur Produkte mit derselben Zahl k von Faktoren vor, so heißt die Polynomfunktion *homogen vom Grad k* .)

Für Matrizen mit ganzen Zahlen als Einträgen folgt aus der Cramerschen Regel insbesondere:

- Hat $A \in \mathbb{R}^{n \times n}$ ganzzahlige Einträge, so ist auch die Determinante $q := \det(A)$ eine ganze Zahl, und wenn $q \neq 0$ ist, so sind alle Einträge von A^{-1} als Brüche mit ganzem Zähler und mit Nenner q darstellbar; dasselbe gilt dann auch für die Komponenten der Lösung x zu $Ax = b$, wenn auch b nur ganzzahlige Einträge hat.
- Hat A ganzzahlige Einträge und Determinante $\det(A) = \pm 1$, so hat auch A^{-1} ganzzahlige Einträge, und wenn alle Komponenten von b ganzzahlig sind, so hat auch die Lösung x zu $Ax = b$ lauter ganzzahlige Komponenten. ■

Eine (auch für die Ökonomie) wichtige Anwendung des Determinantenkriteriums für Invertierbarkeit von quadratischen Matrizen besteht in der Bestimmung der Eigenwerte einer quadratischen Matrix. Hierunter versteht man Folgendes:

DEFINITION. Eine Zahl λ heißt **Eigenwert** einer $n \times n$ -Matrix A (bzw. des zugehörigen linearen Operators $L: \mathbb{R}^n \rightarrow \mathbb{R}^n$, $L(x) = Ax$), wenn es einen Vektor $0 \neq x \in \mathbb{R}^n$ gibt mit $Ax = \lambda x$, d.h. x wird bei Multiplikation mit der Matrix A (bzw. bei Anwendung des linearen Operators L) bis auf den Faktor λ reproduziert. Jeder solche Vektor $x \neq 0$ heißt dann ein **Eigenvektor** von A (bzw. von L) zum Eigenwert λ . ■

DISKUSSION: 1) Die Begriffe "Eigenvektor" und "Eigenwert" sind *nur bei quadratischen Matrizen erklärt*. Ist A nicht quadratisch, so hat der Vektor Ax ein anderes Format als x , also ist eine Gleichung der Form $Ax = \lambda x$ unsinnig! (λ ist der kleine griechische Buchstabe "Lambda" und für Eigenwerte eine übliche Bezeichnung wie auch der kleine griechische Buchstabe "Mü" μ .)

2) Die Bedingung $x \neq 0$ für Eigenvektoren ist nötig, weil $Ax = \lambda x$ ja für jede Zahl λ stets die triviale Lösung $x = 0 \in \mathbb{R}^n$ hat (wenn $A \in \mathbb{R}^{n \times n}$). Wir sind hier aber an *nichttrivialen Lösungen* der Gleichung $Ax = \lambda x$ interessiert, und nur diese heißen "Eigenvektoren" und nur die Zahlen λ , zu denen Eigenvektoren existieren, "Eigenwerte". Es stellt sich heraus, dass $Ax = \lambda x$ für die meisten Zahlen λ nur die triviale Lösung hat. Nur für endlich viele "Ausnahmehzahlen" λ gibt es eine nichttriviale Lösung dieser Gleichung, und diese Ausnahmen sind gerade die Eigenwerte von A . ■

BEISPIELE (von Eigenwerten und Eigenvektoren):

1) Für die *Einheitsmatrix* \mathbb{I}_n sind alle $x \neq 0$ in \mathbb{R}^n Eigenvektoren zum Eigenwert 1; denn es gilt ja $\mathbb{I}_n x = x$. Von 1 verschiedene Eigenwerte gibt es nicht (da aus $\mathbb{I}_n x = \lambda x$ folgt $x = \lambda x$ und daher $\lambda = 1$, wenn x nicht lauter Nullkomponenten hat).

2) *Skalare Matrizen* $A = s\mathbb{I}_n$ erfüllen $Ax = sx$ für alle x , also ist $\lambda = s$ ihr einziger Eigenwert, und alle $x \neq 0$ sind Eigenvektoren.

3) *Diagonalmatrizen* D haben genau ihre Diagonaleinträge d_1, \dots, d_n als Eigenwerte; denn für den Spaltenvektor $x = (x_1, \dots, x_n)$ ist Dx der Spaltenvektor $(d_1 x_1, \dots, d_n x_n)$, und dies ist gleich $\lambda x = (\lambda x_1, \dots, \lambda x_n)$, nur wenn $\lambda = d_j$ ist für mindestens ein j oder wenn alle $x_j = 0$ sind. Für den i -ten kanonischen Basisvektor gilt $De_i = d_i e_i$, also ist er ein Eigenvektor von D und d_i der zugehörige Eigenwert. Allgemeiner sieht man, dass $Dx = \lambda x$ genau dann gilt, wenn $d_j = \lambda$ gilt für alle j mit $x_j \neq 0$. Falls also alle Diagonaleinträge verschieden sind, so sind nur die kanonischen Basisvektoren und Vielfache davon ($\neq 0$) Eigenvektoren zu D . Sind aber mehrere Diagonaleinträge gleich, so sind alle Linearkombinationen ($\neq 0$) der zugehörigen kanonischen Basisvektoren Eigenvektoren.

4) $A \in \mathbb{R}^{n \times n}$ hat den *Eigenwert Null*, genau wenn $Ax = 0$ eine nichttriviale Lösung besitzt, also genau wenn $\det(A) = 0$ ist. Die nichttrivialen Lösungen sind dann die Eigenvektoren zum Eigenwert $\lambda = 0$.

5) *Potenzen* A^k von A mit Exponenten $k \in \mathbb{N}$ haben die Potenz λ^k als Eigenwert und x als zugehörigen Eigenvektor, wenn x Eigenvektor von A zum Eigenwert λ ist; denn aus $Ax = \lambda x$ folgt $A^k x = A^{k-1}(Ax) = A^{k-1}(\lambda x) = \lambda A^{k-1} x = \lambda^2 A^{k-2} x = \dots = \lambda^k x$. Wenn A^{-1} existiert, so gilt das auch für negative ganze Exponenten; denn $A^k x = \lambda^k x \implies \lambda^{-k} x = \lambda^{-k} A^{-k}(A^k x) = \lambda^{-k} A^{-k}(\lambda^k x) = A^{-k} x$. ■

FRAGE: Wie findet man die Eigenwerte einer $n \times n$ -Matrix A und wie die zugehörigen Eigenvektoren $x \in \mathbb{R}^n$?

Hierzu schreiben wir die Gleichung $Ax = \lambda x$ um in der Form $Ax = \lambda \mathbb{I}x \iff (\lambda \mathbb{I} - A)x = 0$ und sehen, dass λ genau dann ein Eigenwert ist, wenn das homogene Gleichungssystem $(\lambda \mathbb{I} - A)x = 0$ nichttriviale Lösungen hat (die dann die zu λ gehörenden Eigenvektoren sind). Dies ist nun genau dann der Fall, wenn A *nicht* invertierbar ist, also gemäß dem Determinantenkriterium für (Nicht-)Invertierbarkeit genau dann, wenn $\det(\lambda \mathbb{I} - A) = 0$ ist. Nun wissen wir aber aus der Leibnizschen Formel, dass die Determinante einer $n \times n$ -Matrix ein vom Grad n homogenes Polynom in den Matrixeinträgen ist. Bei der Matrix $t\mathbb{I} - A$ sind diese Einträge $-a_{ij}$ für $i \neq j$ bzw. $t - a_{ii}$ für $i = j$. Daher ist $\det(t\mathbb{I} - A)$ als Funktion der skalaren Variablen t (bei fester Matrix A) eine Polynomfunktion vom Grad höchstens n . (Der Grad ist tatsächlich genau gleich n .) Und die Eigenwerte λ sind gerade die Nullstellen dieser Polynomfunktion!

DEFINITION: Das **charakteristische Polynom** einer $n \times n$ -Matrix A ist die Polynomfunktion vom Grad n

$$p_A(t) := \det(t\mathbb{I} - A) = t^n + c_{n-1}t^{n-1} + \dots + c_1t + c_0.$$

SATZ: Die Eigenwerte von A sind genau die Nullstellen λ des charakteristischen Polynoms von A . (Es gibt also höchstens n Eigenwerte.) Die zu einem Eigenwert λ gehörenden Eigenvektoren sind genau die nichttrivialen Lösungen $x \in \mathbb{R}^n$ des homogenen linearen Gleichungssystems $(\lambda \mathbb{I} - A)x = 0$. ■

DISKUSSION: 1) Dass das charakteristische Polynom den Grad n und den führenden Koeffizienten $c_n = 1$ hat, kann man aus der Leibnizschen Formel sehen. $\det(t\mathbb{I} - A)$ ist danach eine Linearkombination (mit Koeffizienten $+1$ und -1) von Produkten von n aus verschiedenen Zeilen und Spalten der Matrix herausgegriffenen Einträgen $t\delta_{ij} - a_{ij}$. Nur ein solches Produkt, nämlich das Produkt der Diagonaleinträge $(t - a_{11})(t - a_{22}) \dots (t - a_{nn}) = t^n \pm \dots$, gibt einen Beitrag zur führenden Potenz t^n .

2) Alle anderen auftretenden Produkte haben sogar zwei nichtdiagonale Einträge, tragen als höchstens zu Potenzen t^k mit $0 \leq k \leq n-2$ bei. Der Koeffizient c_{n-1} des charakteristischen Polynoms vor t^{n-1} ist daher derselbe wie bei $(t - a_{11})(t - a_{22}) \dots (t - a_{nn}) = t^n - (a_{11} + a_{22} + \dots + a_{nn})t^{n-1} \pm \dots$, und das ist gerade das Negative der Spur von A (Summe der Diagonaleinträge). Auch den Koeffizienten c_0 des charakteristischen Polynoms können wir sofort angeben, er ist $c_0 = p_A(0) = \det(0\mathbb{I} - A) = \det(-A) = (-1)^n \det(A)$. (Im letzten Schritt haben wir aus n Spalten den Faktor -1 herausgezogen!) Wir haben damit folgende genauere Information über die *Koeffizienten des charakteristischen Polynoms*:

$$p_A(t) = t^n - \text{Spur}(A)t^{n-1} + c_{n-2}t^{n-2} + \dots + c_1t + (-1)^n \det(A).$$

Speziell bei 2×2 -Matrizen $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ ist

$$p_A(t) = t^2 - \text{Spur}(A)t + \det(A) = t^2 - (a+d)t + (ad-bc),$$

was man natürlich auch direkt durch Berechnung der Determinante $\det(t\mathbb{I} - A) = \begin{vmatrix} t-a & -b \\ -c & t-d \end{vmatrix} = (t-a)(t-d) - bc$ erhält.

3) Es ist auch bekannt, wie man die *anderen Koeffizienten des charakteristischen Polynoms* $p_A(t)$ im Fall $n > 2$ berechnet: $(-1)^k c_{n-k}$ ist die Summe aller sog. k -Hauptminoren von A , d.h. der Determinanten aller $k \times k$ -Matrizen, die man aus A durch Streichen von $n-k$ Zeilen und $n-k$ Spalten mit denselben Nummern wie die Zeilen erhalten kann. (Diese $k \times k$ -Untermatrizen liegen also symmetrisch zur Diagonalen von A ; wenn die Nummern der gestrichenen Spalten nicht alle mit denen der gestrichenen Zeilen übereinstimmen, so nennt man die Determinante der $k \times k$ -Restmatrix einen k -Nebenminor von A .) Der Beweis kann geführt werden, indem man $\det(t\mathbb{I} - A)$ als Funktion der Spalten $te_j - a_j$ von $t\mathbb{I} - A$ auffasst und die Differenz sowie den Faktor t aus jeder Zeile herauszieht. Man findet so, dass der Koeffizient von t^{n-k} die Summe der Determinanten aller Matrizen ist, die aus $-A$ mit Ersetzen von $n-k$ Spalten durch die entsprechenden Spalten der Einheitsmatrix hervorgehen, und diese Determinanten sind — bis auf den Faktor $(-1)^k$ — gerade die k -Hauptminoren. Letzteres sieht man nach geeigneten Zeilen- und Spaltenvertauschungen mit dem Kästchensatz für die Determinantenberechnung.

4) *Nicht jede $n \times n$ -Matrix hat einen reellen Eigenwert*, weil nicht jede Polynomfunktion eine reelle Nullstelle hat. Zum Beispiel ist für $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ das charakteristische Polynom $p_A(t) = t^2 + 1$ ohne Nullstelle in \mathbb{R} . Geometrisch beschreibt A eine 90° -Drehung in der Ebene \mathbb{R}^2 , und es ist auch anschaulich klar, dass dabei kein Vektor $\neq 0$ auf ein Vielfaches von sich selbst abgebildet wird. Da aber jede Polynomfunktion von ungeradem Grad eine Nullstelle auf \mathbb{R} hat (z.B. wegen des Zwischenwertsatzes), gilt andererseits:

- *Ist n ungerade, so hat jede $n \times n$ -Matrix A mindestens einen reellen Eigenwert.*

Dies ist übrigens die algebraische Erklärung dafür, warum Drehungen im \mathbb{R}^3 (anders als in \mathbb{R}^2 oder auch \mathbb{R}^4) stets eine Achse festlassen.

5) Dem Fehlen von Eigenwerten kann man abhelfen, indem man *komplexe Eigenwerte* $\lambda \in \mathbb{C}$ zulässt, d.h. komplexe Nullstellen des charakteristischen Polynoms $p_A(t)$. Im Komplexen gilt der *Fundamentalsatz der Algebra*, wonach jedes Polynom vom Grad n genau n Nullstellen hat, wenn man jede mit ihrer Vielfachheit zählt. Somit hat jede $n \times n$ -Matrix auch genau n Eigenwerte λ im Bereich $\mathbb{C} \supset \mathbb{R}$ der komplexen Zahlen, wenn man jeden mit seiner Nullstellenvielfachheit beim charakteristischen Polynom zählt. Um zu λ gehörende Eigenvektoren zu finden, also nichttriviale Lösungen x zu $(\lambda\mathbb{I} - A)x = 0$, muss man dann natürlich auch n -gliedrige Vektoren x mit komplexen Zahlen als Komponenten zulassen (wenn λ nicht reell ist).

Sind $\lambda_1, \dots, \lambda_n$ die komplexen Eigenwerte, jeder mit seiner Vielfachheit aufgeführt, so sieht man aus der Linearfaktorzerlegung des charakteristischen Polynoms $p_A(t) = (t - \lambda_1) \cdot \dots \cdot (t - \lambda_n)$ durch Vergleich mit den in 2) angegebenen Koeffizienten:

- *Die Summe aller (mit ihrer Vielfachheit aufgeführten) komplexen Eigenwerte der $n \times n$ -Matrix A ist die Spur von A , ihr Produkt ist die Determinante von A .*

Das gilt erst recht, wenn $A \in \mathbb{R}^{n \times n}$ (zufällig) n reelle Eigenwerte hat (mit Vielfachheit). ■

BEISPIELE: 1) Die **Eigenwerte von Dreiecksmatrizen** $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ sind gerade ihre Diagonaleinträge; denn hier ist ja auch $t\mathbb{I} - A$ Dreiecksmatrix, also $p_A(t) = \det(t\mathbb{I} - A)$ das Produkt der Diagonaleinträge $(t - a_{11}) \cdot \dots \cdot (t - a_{nn})$, und dieses Polynom hat genau die Nullstellen a_{ii} . Die zum Eigenwert $\lambda = a_{ii}$ gehörenden Eigenvektoren bekommt man durch Auflösen des homogenen linearen Gleichungssystems $(a_{ii}\mathbb{I} - A)x = 0$ (die nichttrivialen Lösungen sind die Eigenvektoren). Das ist hier auch einfach, weil dieses Gleichungssystem ebenfalls Dreiecksform hat. Konkretes Beispiel:

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & -1 \end{pmatrix} \text{ hat die Eigenwerte } 1, 2 \text{ und } -1;$$

Das Gleichungssystem für die Eigenvektoren zum Eigenwert 1 ist:

$$\begin{pmatrix} 0 & 0 & -1 \\ 0 & -1 & -1 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \iff \begin{matrix} x_1 = r \in \mathbb{R} \\ x_2 = 0 \\ x_3 = 0 \end{matrix}, \text{ Eigenvektoren: } \begin{pmatrix} r \\ 0 \\ 0 \end{pmatrix} (r \neq 0);$$

das Gleichungssystem für die Eigenvektoren zum Eigenwert 2 lautet:

$$\begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & -1 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \iff \begin{matrix} x_1 = 0 \\ x_2 = s \in \mathbb{R} \\ x_3 = 0 \end{matrix}, \text{ Eigenvektoren: } \begin{pmatrix} 0 \\ s \\ 0 \end{pmatrix} (s \neq 0);$$

das Gleichungssystem für die Eigenvektoren zum Eigenwert -1 ist schließlich:

$$\begin{pmatrix} -2 & 0 & -1 \\ 0 & -3 & -1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \iff \begin{matrix} x_1 = -\frac{1}{2}t \\ x_2 = -\frac{1}{3}t \\ x_3 = t \in \mathbb{R} \end{matrix}, \text{ Eigenvektoren: } \begin{pmatrix} -\frac{1}{2}t \\ -\frac{1}{3}t \\ t \end{pmatrix} (t \neq 0).$$

2) Die *Eigenwerte einer Potenz* A^k von $A \in \mathbb{R}^{n \times n}$ sind genau die Potenzen λ^k aller komplexen Eigenwerte λ von A . (Diese Potenzen können durchaus reell sein, auch wenn λ nicht reell ist. Z.B. ist $i = \sqrt{-1}$ eine nichtreelle komplexe Zahl, aber $i^2 = -1$ ist reell.) Der Beweis kann am einfachsten geführt werden, indem man das Problem über dem Zahlbereich \mathbb{C} betrachtet, wo es auf den Fall von Dreiecksmatrizen rückgeführt werden kann. Ist aber A Dreiecksmatrix, so auch A^k , und die Diagonaleinträge, also die Eigenwerte, von A^k sind gerade die Potenzen $(a_{ii})^k$ der Diagonaleinträge von A .

3) *Nilpotente Matrizen* $N \in \mathbb{R}^{n \times n}$ haben nur den Eigenwert 0 (mit Vielfachheit n); denn aus $Nx = \lambda x$ folgt $N^k x = \lambda^k x$ und wegen der Nilpotenz ist $N^k = 0$ für genügend große k . Man kann zeigen, dass auch umgekehrt jede $n \times n$ -Matrix mit n -fachem Eigenwert 0 nilpotent ist.

4) *Involutorische Matrizen* A , d.h. $A^2 = \mathbb{I}$, haben nur die Eigenwerte $\lambda = 1$ oder $\lambda = -1$, da $\lambda^2 = 1$ ist nach 3). Geometrisch gedeutet sind involutorische lineare Abbildungen *Spiegelungen*. Die Eigenvektoren zum Eigenwert 1 (und der Ursprung 0) entsprechen dabei den Punkten des "Spiegels", die unter der Abbildung auf sich selbst abgebildet werden.

5) Die *transponierte Matrix* hat dasselbe charakteristische Polynom wie A , also auch dieselben Eigenwerte; denn es ist ja $\det(t\mathbb{I} - A^T) = \det((t\mathbb{I} - A)^T) = \det(t\mathbb{I} - A)$. Die Eigenvektoren bei A^T sind aber im Allgemeinen andere als bei A .

6) Ein *geschlossenes Leontief-Modell* ist ein Input-Output-Modell $(\mathbb{I} - A)x = y$ (siehe 3.2 und 3.4 oben), in dem es zum Absatzvektor $y = 0$ eine Lösung $x \neq 0$ mit lauter nichtnegativen Komponenten gibt. Dies bedeutet, dass die gesamte Produktion für den endogenen Input verwendet wird. Bei einem Unternehmen würde das nicht lange gut gehen, aber bei einer Volkswirtschaft ist es denkbar. Wegen $(\mathbb{I} - A)x = 0 \iff Ax = x$ ist ein Leontief-Modell genau dann geschlossen, wenn die Produktionsmatrix A einen Eigenvektor zum Eigenwert 1 mit lauter nichtnegativen Komponenten besitzt. Ist A positiv wie hier (d.h. alle $a_{ij} \geq 0$ und A ohne Nullzeile), so gibt es übrigens stets einen positiven Eigenwert und einen zugehörigen Eigenvektor mit lauter positiven Komponenten. Dies garantiert der sog. *Satz von Perron & Frobenius*, der in der Mathematik bewiesen wird. ■

3.5 Lineare Unabhängigkeit, Unterräume und Basen

In diesem Abschnitt präzisieren und vertiefen wir einige fundamentale Konzepte der Linearen Algebra. Diese Begriffsbildungen sind für die mathematische Theorie zentral und sie sind auch nützlich für das Verständnis von einigen ökonomischen Problemstellungen, insbesondere für die Lineare Optimierung (siehe 3.6), die mit Hilfe der Linearen Algebra mathematisch modelliert werden. Es handelt sich dabei um die lineare Unabhängigkeit von Vektoren und um parametrisierte lineare Scharen von Vektoren. Beide Konzepte sind in einer speziellen Situation im Zusammenhang mit linearen Systemen von m Gleichungen für n Unbekannte in 3.2 schon eingeführt worden, wo wir von der maximalen Zahl l der (linear) unabhängigen Gleichungen des Gleichungssystems gesprochen hatten und von der $(n-l)$ -parametrischen Schar der Lösungen des homogenen Gleichungssystems bzw. von der k -parametrischen linearen Schar der konsistenten rechten Seiten.

Die grundlegende und einfache Natur dieser Konzepte wird aber – wie oft in der Mathematik – erst richtig deutlich, wenn man sich von der speziellen Situation löst. Wir betrachten daher statt des Zahlenraums \mathbb{R}^n im Folgenden allgemeiner einen beliebigen **Vektorraum** V über dem Zahlbereich \mathbb{R} . (Man spricht hier auch von einem “reellen Vektorraum”. In der Mathematik werden auch Vektorräume über anderen Zahlbereichen behandelt, z.B. komplexe Vektorräume, bei denen der Zahlbereich \mathbb{C} zugrunde liegt. Für Anwendungen in der Ökonomie ist diese größere Allgemeinheit aber nicht nötig.) Darunter versteht man eine Menge V mit einem ausgezeichneten Element 0 , für deren Elemente u, v, w, \dots eine Summe $u+v \in V$ und die Vervielfachung $rv \in V$ mit Skalaren $r \in \mathbb{R}$ erklärt sind, derart dass die grundlegenden Rechenregeln (Axiome) gelten, die wir in 3.3 für das Rechnen mit (Spalten-)Vektoren angegeben haben; das sind die Folgenden:

$$\begin{aligned} u+v &= v+u, & (u+v)+w &= u+(v+w), & v+0 &= v, & u+v = 0 &\iff v = (-1)u, \\ 1v &= v, & 0v &= 0, & r(sv) &= (rs)v, & (r+s)v &= rv+sv, & r(u+v) &= ru+rv. \end{aligned}$$

Die Elemente von V nennt man dann “Vektoren”, das ausgezeichnete Element 0 den “Nullvektor”. Es ist hierbei irrelevant, wie die Elemente von V genau definiert sind, sondern es kommt nur darauf an, nach welchen Regeln man mit ihnen rechnet. Wer aber eine konkrete Situation leichter verständlich findet, kann sich im Folgenden unter den Vektorräumen V, W, \dots einfach Zahlenräume $\mathbb{R}^n, \mathbb{R}^m, \dots$ mit den aus 3.3 bekannten “komponentenweisen” Rechenoperationen vorstellen.

Zu den Vektorräumen V, W, \dots gehören die Abbildungen $T:V \rightarrow W$ dazwischen, die mit den Rechenoperationen verträglich sind, d.h. die Gleichungen $T(u+v) = T(u)+T(v)$ und $T(rv) = rT(v)$ gelten für alle $u, v \in V$ und $r \in \mathbb{R}$ (wobei auf der linken Seite der Gleichung jeweils die Rechenoperation in V ausgeführt wird und auf der rechten Seite die Operation in W). Eine solche Abbildung T eines Vektorraums V in einen (anderen oder denselben) Vektorraum W nennt man **lineare Abbildung** oder **linearer Operator**. Eine unmittelbare Folgerung ist $T(0) = 0$ (links der Nullvektor in V , rechts der in W). In der speziellen Situation $V = \mathbb{R}^n, W = \mathbb{R}^m$ haben wir lineare Operatoren schon in 3.3 betrachtet und gezeigt, dass sie sich durch $m \times n$ -Matrizen A beschreiben lassen in der Form $T(\mathbf{x}) = A\mathbf{x}$ für (Spalten-)Vektoren $\mathbf{x} \in \mathbb{R}^n$. Ein linearer Operator ist also in dieser Situation einfach eine Matrix, bzw. genauer gesagt, die Multiplikation einer (festen) Matrix von links an einen (variablen) Spaltenvektor. In der allgemeinen Situation von beliebigen Vektorräumen sind die linearen Abbildungen die Objekte, die den Matrizen entsprechen.

Ein Vorteil, den die Entwicklung der linearen Algebra im Rahmen von allgemeinen Vektorräumen und linearen Abbildungen bietet, ist die Einfachheit und bessere Verständlichkeit der grundlegenden Konzepte. Ein weiterer Vorteil ist, dass alle abgeleiteten Rechenoperationen und Folgerungen, die man in diesem allgemeinen Rahmen herleiten kann, automatisch auch in jeder konkreten Situation anwendbar sind, in der die grundlegenden Rechenregeln gelten. Diese konkreten Situationen können ganz unterschiedlich aussehen und betreffen z.B. das Rechnen mit Vektoren in \mathbb{R}^n , oder mit Matrizen eines festen Formats $m \times n$, oder mit reellen Funktionen auf einem gemeinsamen Definitionsbereich, ...

Eine abgeleitete Rechenoperation ist z.B. die Bildung der Differenz $u-v := u+(-1)v \in V$ von Vektoren u, v in einem Vektorraum V . Eine andere ist die Bildung der Summe $\sum_{j=1}^k v_j = v_1+v_2+\dots+v_k \in V$ von beliebig (endlich) vielen Vektoren $v_j \in V$, die nicht von der Reihenfolge und Klammerung der Summanden abhängt. Für das Thema dieses Abschnitts besonders wichtig ist die Möglichkeit der Bildung der **Linearkombination** von endlich vielen Vektoren $v_1, \dots, v_k \in V$ mit Koeffizienten $r_1, \dots, r_k \in \mathbb{R}$, also der Summe der Vielfachen $r_j v_j$ der Vektoren v_j ,

$$\sum_{j=1}^k r_j v_j = r_1 v_1 + r_2 v_2 + \dots + r_k v_k \in V \quad (v_j \in V, r_j \in \mathbb{R} \text{ für } j = 1 \dots k).$$

Lineare Abbildungen $T:V \rightarrow W$ zwischen Vektorräumen sind dann automatisch auch mit dieser abgeleiteten Rechenoperation verträglich, d.h. man kann die Linearkombinationsbildung "herausziehen" im Sinne von

$$T\left(\sum_{j=1}^k r_j v_j\right) = \sum_{j=1}^k r_j T(v_j).$$

Das intuitive Konzept, dass gewisse Vektoren $v \in V$ durch gegebene Vektoren v_1, \dots, v_k "ausgedrückt" werden können (oder auch nicht), kann man mit dem Begriff der Linearkombination nun präzise fassen wie folgt:

DEFINITION: (i) Wir sagen, dass ein Vektor v durch gegebene Vektoren v_1, \dots, v_k des Vektorraums V **linear ausgedrückt** werden kann (oder dass v von den Vektoren v_1, \dots, v_k *linear abhängig* ist), wenn er sich als Linearkombination $v = \sum_{j=1}^k r_j v_j$ darstellen lässt (mit geeigneter Wahl der Koeffizienten r_j).

(ii) Die Vektoren $v_1, \dots, v_k \in V$ heißen **linear unabhängig** (voneinander), wenn keiner von ihnen durch die anderen linear ausgedrückt werden kann. (Im Fall $k=1$ heißt der eine Vektor v_1 linear unabhängig in V , wenn er nicht der Nullvektor ist.) Andernfalls heißen die Vektoren **linear abhängig**. ■

DISKUSSION: (1) Es gibt verschiedene *äquivalente Beschreibungen der linearen Unabhängigkeit* von Vektoren v_1, \dots, v_k im Vektorraum V :

- Keiner der Vektoren v_i ist Linearkombination der vorangehenden v_1, \dots, v_{i-1} .
- Der Nullvektor ist aus v_1, \dots, v_k nur auf triviale Weise linear kombinierbar, d.h. $r_1 v_1 + r_2 v_2 + \dots + r_k v_k = 0$ gilt nur, wenn $r_1 = r_2 = \dots = r_k = 0$.
- Die Koeffizienten von Linearkombinationen der v_j sind eindeutig bestimmt, d.h. aus $r_1 v_1 + r_2 v_2 + \dots + r_k v_k = s_1 v_1 + s_2 v_2 + \dots + s_k v_k$ folgt $r_1 = s_1, r_2 = s_2, \dots, r_k = s_k$.

Aus der linearen Unabhängigkeit von v_1, \dots, v_k gemäß der Definition folgt offenbar die erste Aussage; denn kann man keinen der Vektoren v_i durch die anderen v_j , $j \neq i$, ausdrücken, so erst recht nicht durch die vorangehenden v_1, \dots, v_{i-1} . Aus der ersten Aussage folgt die zweite; denn hätte man eine nichttriviale Linearkombination $0 = \sum_{j=1}^k r_j v_j$, also $r_i \neq 0$ für einen größten Index i und somit $r_{i+1} = \dots = r_k = 0 \neq r_i$, so könnte man durch Division mit r_i und Auflösung nach v_i den Vektor $v_i = \sum_{j=1}^{i-1} (-\frac{r_j}{r_i}) v_j$ durch die vorangehenden Vektoren v_1, \dots, v_{i-1} linear ausdrücken. Die dritte Aussage folgt aus der zweiten durch Differenzbildung $0 = \sum_{j=1}^k (s_j - r_j) v_j$. Schließlich folgt aus der dritten Aussage die lineare Unabhängigkeit von v_1, \dots, v_k wie definiert, weil insbesondere eine Gleichung $\sum_{j \neq i} r_j v_j = v_i$ (also $r_i = 0$, $s_i = 1$ und $s_j = 0$ für alle $j \neq i$) ausgeschlossen ist.

(2) Eine triviale Beobachtung, die sich unmittelbar aus der Definition ergibt, ist folgende:

- *Entfernt man aus einem System linear unabhängiger Vektoren einen oder mehrere Vektoren, so ist das verkleinerte System erst recht linear unabhängig.*
- *Nimmt man zu einem System linear abhängiger Vektoren in V weitere Vektoren aus V hinzu, so ist das vergrößerte System erst recht linear abhängig.*

(3) **Rechenkriterium für lineare (Un-)Abhängigkeit in \mathbb{R}^n :**

- *Die Vektoren $v_1, \dots, v_k \in \mathbb{R}^n$ sind linear unabhängig, genau wenn das homogene lineare Gleichungssystem $Ax = 0$ mit der aus den Spalten v_j zusammengestellten $n \times k$ -Matrix $A = (v_1 \ v_2 \ \dots \ v_k)$ nur die triviale Lösung $x = 0 \in \mathbb{R}^k$ hat.*

Das ist mit (1) klar, weil Ax nichts anderes ist als die Linearkombination $x_1 v_1 + \dots + x_k v_k$ der Vektoren v_j mit den Komponenten x_j des Vektors $x \in \mathbb{R}^k$ als Koeffizienten. Die Überprüfung der linearen Unabhängigkeit bzw. Abhängigkeit von gegebenen Vektoren in \mathbb{R}^n ist damit zurückgeführt auf die Lösung eines homogenen linearen Gleichungssystems: Hat es nur eine Lösung, nämlich die triviale Lösung $x = 0$, so sind die Vektoren linear unabhängig; gibt es aber mehrere Lösungen, nämlich eine unendliche lineare Lösungsschar, so sind die Vektoren linear abhängig.

Da weniger als k homogene lineare Gleichungen für k Unbekannte gemäß 3.2 immer nicht-triviale Lösungen haben, ist eine unmittelbare Konsequenz:

- *Mehr als n Vektoren aus \mathbb{R}^n sind immer linear abhängig.*

Für $k \leq n$ dagegen gibt es viele Systeme von k linear unabhängigen Vektoren in \mathbb{R}^n . Zum Beispiel können wir aus den kanonischen Basisvektoren e_1, \dots, e_n einfach k verschiedene auswählen (siehe die folgenden Beispiele).

(4) Die Lösung des linearen Gleichungssystems aus (3) bewerkstelligt man gemäß 3.2 durch Herstellung der Zeilen-Stufen-Form der Koeffizientenmatrix $A \in \mathbb{R}^{n \times k}$ mit Zeilenoperationen. Solche Operationen erhalten die Lösungsmenge von $Ax = 0$, also die lineare (Un-)abhängigkeit der Spalten von A . Letzteres gilt aber auch für Spaltenoperationen mit A ; denn diese laufen bei der Matrix $A = (v_1 \ v_2 \ \dots \ v_k)$ darauf hinaus, dass man zwei der Spaltenvektoren v_i, v_j ($i \neq j$) vertauscht, oder einen Spaltenvektor v_j ersetzt durch $r v_j$ mit $r \in \mathbb{R}_{\neq 0}$, oder v_i ersetzt durch $v_i + v_j$ (bzw. durch $v_i + s v_j$ bei einer Kombination der beiden letzten Operationen). Solche Operationen ändern aber offenbar nichts an der linearen Abhängigkeit oder Unabhängigkeit der Vektoren v_1, \dots, v_k . Daher ergibt sich folgendes

effektiveres Rechenkriterium für lineare (Un-)Abhängigkeit in \mathbb{R}^n :

- Zeilen- und Spaltenoperationen bei einer Matrix ändern nichts an der linearen Abhängigkeit bzw. Unabhängigkeit ihrer Spaltenvektoren.
- Die Vektoren v_1, v_2, \dots, v_k ($k \leq n$) sind genau dann linear unabhängig in \mathbb{R}^n , wenn man die Matrix $A = (v_1 v_2 \dots v_k)$ mit den Spalten v_j durch Zeilenoperationen und/oder Spaltenoperationen auf eine Form bringen kann, deren erste k Zeilen eine Dreiecksmatrix ohne Nullen auf der Diagonalen bilden.
- Die Vektoren v_1, v_2, \dots, v_k sind genau dann linear abhängig in \mathbb{R}^n , wenn man aus der Matrix $A = (v_1 v_2 \dots v_k)$ durch Zeilenoperationen und/oder Spaltenoperationen eine Matrix mit einer Nullspalte bzw. mit mehr als $n-k$ Nullzeilen erzeugen kann.

Um Missverständnissen vorzubeugen, betonen wir, dass man aus der Erzeugung einer einzigen Nullzeile nichts schließen kann, wenn $n > k$, und dass hier nur von Umformungen der Koeffizientenmatrix A selbst durch Zeilen- oder Spaltenoperationen die Rede ist, nicht etwa von der Umformung einer erweiterten Koeffizientenmatrix $(A|\mathbf{b})$.

(5) Nützlich für die Untersuchung von gegebener Vektoren in \mathbb{R}^n auf lineare (Un-)Abhängigkeit ist oft folgende Beobachtung:

- Sind die Vektoren v_j ($j = 1 \dots k$) linear abhängig in \mathbb{R}^n , so gilt das auch für die Vektoren \tilde{v}_j in \mathbb{R}^{n-m} , die aus den v_j durch Streichen von m Komponenten mit vorgegebenen Positionen entstehen.
- Sind die Vektoren v_j ($j = 1 \dots k$) linear unabhängig in \mathbb{R}^n , so gilt das auch für die Vektoren \hat{v}_j in \mathbb{R}^{n+m} , die aus den v_j durch Hinzufügen von m Komponenten mit vorgegebenen Positionen und beliebigen Einträgen (bei jedem \hat{v}_j) entstehen.

Die erste Aussage ist klar, weil eine Linearkombination des Nullvektors in \mathbb{R}^n nach Streichen von m Komponenten (in denselben Positionen bei allen Vektoren) eine nichttriviale Linearkombination des Nullvektors in \mathbb{R}^{n-m} ist. Die zweite folgt aus der ersten, indem man die hinzugefügten Komponenten wieder streicht. Kann man z.B. in der $n \times k$ -Matrix $A = (v_1 \dots v_k)$ eine $k \times k$ -Untermatrix \tilde{A} finden, die linear unabhängige Spalten in \mathbb{R}^k hat, so sind auch die Spalten v_j der Matrix A selbst linear unabhängig in \mathbb{R}^n (weil sie durch Hinzufügen von Komponenten zu den Spalten der Untermatrix entstehen).

(6) Da gemäß 3.4 das homogene lineare Gleichungssystem $\tilde{A}\mathbf{x} = 0$ genau dann nur trivial lösbar ist, wenn die $k \times k$ -Matrix \tilde{A} invertierbar ist, wenn also $\det(\tilde{A}) \neq 0$ ist, können wir aus (3) und (5) folgendes **Determinantenkriterium für lineare (Un-)Abhängigkeit** von Vektoren in \mathbb{R}^n folgern:

- Die Vektoren v_1, v_2, \dots, v_k sind linear unabhängig in \mathbb{R}^n , wenn die Matrix $A = (v_1 v_2 \dots v_k)$ mit den Spalten v_j eine (durch Streichen von $n-k$ beliebigen Zeilen in A entstandene) $k \times k$ -Untermatrix \tilde{A} mit $\det(\tilde{A}) \neq 0$ besitzt.
- Haben aber alle derartigen $k \times k$ -Untermatrizen von A Determinante 0, so sind die Vektoren v_1, v_2, \dots, v_k sind linear abhängig in \mathbb{R}^n .

Die erste Aussage ist klar, weil eben $\tilde{A}\mathbf{x} = 0$ nur die triviale Lösung hat, wenn $\det(\tilde{A}) \neq 0$ ist, und weil aus der linearen Unabhängigkeit der Spalten von \tilde{A} die der durch Komponentenanzufügung entstehenden Spalten von A folgt. Die zweite Aussage gilt, weil man im Fall der linearen Abhängigkeit von v_1, \dots, v_k durch Spaltenoperationen bei A eine Matrix B mit Nullspalte herstellen kann; dann haben auch alle $k \times k$ -Untermatrizen von B eine Nullspalte also Determinante 0, und da Spaltenoperationen die Determinante nicht ändern, gilt das auch für alle $k \times k$ -Untermatrizen der ursprünglichen Matrix A .

Das Determinantenkriterium für lineare (Un-)abhängigkeit ist eher von theoretischem Interesse und für praktische Rechnungen im Allgemeinen zu aufwendig. Schließlich muss man bei seiner Anwendung die Determinanten von allen $\binom{n}{k}$ Untermatrizen des Formats $k \times k$ berechnen, wenn man die lineare Abhängigkeit von v_1, \dots, v_k verifizieren will. (Allerdings genügt eine einzige derartige Determinante zum Nachweis der linearen Unabhängigkeit – wenn sie nämlich $\neq 0$ ausfällt!) Selbst im Fall $k = n$, wo nur $\det(A)$ selbst auszurechnen ist, sind die Rechenverfahren in (3) und (4) im Allgemeinen weniger aufwendig.

(7) Eine theoretische Folgerung, die man aus dem Determinantenkriterium ziehen kann, ist zum Beispiel: *Lineare Unabhängigkeit ist bei $k \leq n$ Vektoren in \mathbb{R}^n der Normalfall* in dem Sinne, dass eine beliebig kleine Störung der Vektoren zu einem linear unabhängigen System führt. Dazu addiert man zu den Spalten v_j von A Vielfache se_j der kanonischen Basisvektoren e_j von \mathbb{R}^n . Ist \tilde{A} die Untermatrix der ersten k Zeilen von A , so ist dann $\tilde{A} + s\mathbb{I}_k$ die entsprechende Untermatrix zu den Vektoren $v_j + se_j$, und es gilt $\det(\tilde{A} + s\mathbb{I}_k) \neq 0$ für alle Zahlen $s \in \mathbb{R}$ bis auf endlich viele (weil die Determinante eine Polynomfunktion des Parameters $s \in \mathbb{R}$ ist, die ja nur endlich viele Nullstellen hat), insbesondere also für alle Zahlen $s \in \mathbb{R}_{\neq 0}$ von hinreichend kleinem Betrag. Man kann den “Normalfall” in einem Sinne, der mathematisch präzisiert werden kann, auch so beschreiben: *Wählt man $k \leq n$ Vektoren zufällig in \mathbb{R}^n , so sind sie mit Wahrscheinlichkeit 1 linear unabhängig.*

(8) Aus $T(\sum_{j=1}^k r_j v_j) = \sum_{j=1}^k r_j T(v_j)$ für lineare Abbildungen $T: V \rightarrow W$ zwischen Vektorräumen liest man ab, dass die Bildvektoren $T(v_1), \dots, T(v_k)$ genau dann linear unabhängig in W sind, wenn es keine nichttriviale Linearkombination $v = \sum_{j=1}^k r_j v_j$ gibt mit $T(v) = 0$. Gilt $T(v) = 0$ nur für $v = 0$, so folgt also aus der linearen Unabhängigkeit der v_j auch die der Bildvektoren. Für lineare Abbildungen ist die Eigenschaft $T(v) = 0 \implies v = 0$ äquivalent mit $u \neq v \implies T(u) \neq T(v)$, d.h. verschiedene Vektoren haben auch verschiedene Bilder. (Die Äquivalenz sieht man an $T(u) - T(v) = T(u - v)$.) Solche Abbildungen nennt man **injektiv**, also können wir sagen:

- *Lineare Abbildungen T erhalten lineare Abhängigkeit von Vektoren, aber im Allgemeinen nicht lineare Unabhängigkeit.*
- *Ist aber T injektive lineare Abbildung (also $T(v) = 0$ nur für $v = 0$), so erhält T auch lineare Unabhängigkeit.*

(9) Die Definition der linearen Unabhängigkeit haben wir für endliche Folgen v_1, \dots, v_k von Vektoren in einem Vektorraum V formuliert. Sie lässt sich aber ganz genau so für unendliche Folgen oder für beliebige Familien $(v_j)_{j \in J}$ von Vektoren $v_j \in V$ aussprechen, die mit einer endlichen oder unendlichen Indexmenge J nummeriert sind. Dabei hat man unter einer Linearkombination der v_j immer eine Summe $\sum_{i=1}^k r_i v_{j_i}$ mit beliebig (endlich) vielen Summanden zu verstehen, wobei j_1, \dots, j_k aus der Indexmenge J ausgewählt sind. (Summen von unendlich vielen Vektoren sind im Rahmen der Linearen Algebra nicht definierbar.) Also ist $(v_j)_{j \in J}$ linear unabhängig, genau wenn man kein v_i durch endlich viele v_j mit $j \in J_{\neq i}$ linear ausdrücken kann. Äquivalent ist, dass keine nichttriviale Linearkombination (von endlich vielen) der v_j den Nullvektor ergibt.

Im Zahlenraum \mathbb{R}^n und allgemeiner in einem Vektorraum V endlicher Dimension n (s.u.), ist die Betrachtung unendlicher Familien von Vektoren in der Definition der linearen Unabhängigkeit nicht sinnvoll, weil mehr als n Vektoren dann stets linear abhängig sind (vgl. (3) oben). Es gibt aber auch “größere” Vektorräume, die in der Mathematik und auch in einigen Anwendungen auf ökonomische Fragestellungen relevant sind und unendliche Systeme von linear unabhängigen Vektoren besitzen (siehe (5) der folgenden Beispiele). ■

BEISPIELE: (1) Ein einziger Vektor v in \mathbb{R}^n (oder in einem beliebigen Vektorraum) ist linear unabhängig, genau wenn er nicht der Nullvektor ist. Das ist die Definition der linearen Unabhängigkeit in diesem Fall, und mehr gibt es dazu nicht zu sagen.

(2) Zwei Vektoren v, w in \mathbb{R}^n (oder in einem beliebigen Vektorraum) sind genau dann linear abhängig, wenn einer der Vektoren ein Vielfaches des anderen ist, also $v = rw$ oder $w = sv$ für gewisse $r, s \in \mathbb{R}^n$. Das ist immer der Fall, wenn (mindestens) einer der beiden Vektoren der Nullvektor ist (wegen $0v = 0w = 0$), und im anderen Fall ist bei linear abhängigen Vektoren v, w jeder ein Vielfaches des anderen (wegen $v = rw \iff w = \frac{1}{r}v$ für $r \in \mathbb{R}_{\neq 0}$). Geometrisch bedeutet die lineare Unabhängigkeit von v, w , dass nicht beide Vektoren auf derselben Geraden durch den Nullpunkt des \mathbb{R}^n liegen.

Eine nützliche Beobachtung ist, dass zwei linear abhängige Vektoren $u \neq 0 \neq v$ in \mathbb{R}^n Nulleinträge in denselben Positionen haben müssen; damit kann man oft schon die lineare Unabhängigkeit mit einem Blick auf die Nulleinträge der beiden Vektoren ablesen. Auch für zwei beliebige Vektoren v, w in \mathbb{R}^n lässt sich die lineare Abhängigkeit oft ohne Rechnung "durch Hinsehen" feststellen: Man betrachtet eine Komponente $v_i \neq 0$ von v (gibt es die nicht, so ist $v = 0$, also sind v, w linear abhängig) und findet damit den einzig möglichen Faktor $s = w_i/v_i$, für den $w = sv$ gelten kann. Dann hat man noch nachzuprüfen, ob auch $w_j = sv_j$ für alle Komponenten mit Nummern $j \neq i$ gilt mit demselben Faktor $s = w_i/v_i$; falls ja, so gilt $w = sv$ und v, w sind linear abhängig, andernfalls sind die beiden Vektoren linear unabhängig. (Das Determinantenkriterium für lineare Abhängigkeit lautet hier übrigens $v_i w_j - v_j w_i \neq 0$ für $1 \leq i < j \leq n$ und läuft im Wesentlichen auf dieselbe Vorgehensweise hinaus.) So sind z.B.

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ -4 \end{pmatrix} \text{ linear unabhängig in } \mathbb{R}^2 \text{ und } \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ \frac{2}{3} \\ -\frac{1}{2} \end{pmatrix} \text{ linear unabhängig in } \mathbb{R}^3,$$

weil in einer bestimmten Position jeweils ein Vektor einen Nulleintrag hat, der andere aber nicht (und weil keiner der Vektoren der Nullvektor ist). Dagegen sind

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -2 \\ -4 \end{pmatrix} \text{ linear abhängig in } \mathbb{R}^2 \text{ und } \begin{pmatrix} 1 \\ 0 \\ -1 \\ \sqrt{2} \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ 0 \\ -\frac{1}{2} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \text{ linear abhängig in } \mathbb{R}^4,$$

weil der zweite Vektor ein Vielfaches des ersten ist (bzw. der erste ein Vielfaches des zweiten, wie man unmittelbar sieht. (Man findet den Vervielfachungsfaktor durch Blick auf die obersten Einträge und prüft für die anderen Komponenten denselben Vervielfachungsfaktor nach.) Bei zwei Vektoren, deren Einträge Parameter enthalten, kann man mit denselben Methoden feststellen, für welche Wahlen der Parameter sie linear abhängig sind oder nicht. So sind in \mathbb{R}^2 zwei allgemeine Vektoren

$$\begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} c \\ d \end{pmatrix} \text{ linear unabhängig, genau wenn } ad - bc \neq 0 \text{ ist.}$$

Das besagt hier das Determinantenkriterium, weil $ad - bc = \det \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ ist. Und

$$v = \begin{pmatrix} r \\ s \\ t \end{pmatrix}, w = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \text{ sind linear abhängig in } \mathbb{R}^3, \text{ genau wenn } r \in \mathbb{R}, s = 2r, t = 3r;$$

denn an den obersten Komponenten $v_1 = r, w_1 = 1$ sieht man, dass $v = rw$ gelten muss, wenn v überhaupt ein Vielfaches von w ist, und die Gleichungen $s = 2r, t = 3r$ bedeuten, dass dann auch $v_j = rw_j$ für die Komponenten mit Nummern $j = 2, 3$ gilt.

(3) Bei 3 Vektoren in \mathbb{R}^n kann man die lineare (Un-)Abhängigkeit im Allgemeinen nicht mehr "durch Hinsehen" erkennen, sondern muss zu einem Rechenverfahren greifen. Natürlich muss $n \geq 3$ sein, damit überhaupt 3 Vektoren in \mathbb{R}^n linear unabhängig sein können! Man stellt also die Vektoren spaltenweise zu einer $n \times 3$ -Matrix A zusammen und berechnet die Lösungen des homogenen Gleichungssystems $Ax = 0$ (nur triviale Lösung \iff lineare Unabhängigkeit). Oder man bringt A durch Zeilen- und/oder Spaltenoperationen in eine Form, an der man die lineare Abhängigkeit bzw. Unabhängigkeit der Spalten direkt ablesen kann, indem man z.B. eine 3×3 -Untermatrix von Dreiecksgestalt und ohne Nulleinträge auf der Diagonalen erzeugt (\iff lineare Unabhängigkeit) oder eine Nullspalte (\iff lineare Abhängigkeit). Geometrisch bedeutet lineare Unabhängigkeit von $u, v, w \in \mathbb{R}^n$, dass $u \neq 0$ ist, v nicht auf der durch u bestimmten Ursprungsgeraden liegt und w nicht auf der durch u, v bestimmten Ebene durch den Ursprung $0 \in \mathbb{R}^n$.

Für die drei Vektoren (mit beliebigen Einträgen r, s, t in der letzten Komponente)

$$u = \begin{pmatrix} 1 \\ 1 \\ 2 \\ r \end{pmatrix}, v = \begin{pmatrix} 2 \\ 0 \\ 6 \\ s \end{pmatrix}, w = \begin{pmatrix} 3 \\ 1 \\ 7 \\ t \end{pmatrix} \quad \text{in } \mathbb{R}^4,$$

ergibt sich z.B. die folgende Matrix $A = (u \ v \ w)$ und ihre Umformung durch zwei Zeilenoperationen und eine Spaltenvertauschung:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 2 & 6 & 7 \\ r & s & t \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 0 & 0 & -1 \\ r & s & t \end{pmatrix} \longrightarrow \begin{pmatrix} 2 & 1 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \\ s & r & t \end{pmatrix}.$$

Zuletzt hat die Untermatrix der drei obersten Zeilen Dreiecksgestalt ohne Nulleinträge auf der Diagonalen, also sind u, v, w linear unabhängig. Ändern wir bei w den Eintrag 7 ab zu 8 und bilden so den Vektor \tilde{w} , so führt dieselbe Rechnung für u, v, \tilde{w} auf die Matrix

$$B = \begin{pmatrix} 2 & 1 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ s & r & t \end{pmatrix} \quad \text{mit } 3 \times 3\text{-Untermatrizen } B' := \begin{pmatrix} 2 & 1 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}, B'' := \begin{pmatrix} 2 & 1 & 3 \\ 0 & 1 & 1 \\ s & r & t \end{pmatrix}.$$

Hier ist die Untermatrix B' der obersten 3 Zeilen von Dreiecksform mit einem Nulleintrag auf der Diagonalen. Das heißt freilich noch nicht, dass die Vektoren u, v, \tilde{w} in \mathbb{R}^4 linear abhängig sind, sondern eben nur die 3 Spalten dieser Untermatrix. Eine andere 3×3 -Untermatrix könnte aber unabhängige Spalten haben, und dann wären auch u, v, \tilde{w} linear unabhängig. Wegen der Nullzeile kommt als derartige Untermatrix einzig die aus den Zeilen Nr. 1, 2 und 4 zusammengestellte in Frage, die oben als Untermatrix B'' angegeben ist. (Die drei anderen 3×3 -Untermatrizen haben eine Nullzeile und daher linear abhängige Spalten.) Mit Spaltenoperationen erzeugen wir bei dieser Untermatrix $(0, 0, t-r-s)$ als letzte Spalte und mit Zeilen- und Spaltenvertauschung dann Folgendes:

Diese Dreiecksmatrix hat keinen Nulleintrag auf der Diagonalen, wenn $t \neq r+s$ ist; in diesem Fall sind also u, v, \tilde{w} linear unabhängig. Im Fall $t = r+s$ dagegen hat diese Matrix und auch die 4×3 -Matrix, die durch entsprechende Umformung aus B hervorgegangen ist, eine Nullspalte, also sind dann u, v, \tilde{w} linear abhängig. Das-

selbe Ergebnis hätte man durch Berechnung von $\det(B'') = 2(t-r-s)$ erhalten. Da die anderen 3×3 -Untermatrizen von B eine Nullzeile und damit Determinante 0 haben, zeigt das Determinantenkriterium, dass u, v, \tilde{w} genau im Fall $t = r+s$ linear abhängig sind. (Dann ist übrigens $\tilde{w} = u+v$, was man auch direkt sehen kann.)

(4) Analog überprüft man die lineare Unabhängigkeit von $k \geq 3$ Vektoren in \mathbb{R}^n . Mit wachsenden k, n wächst auch der Rechenaufwand. ($k \leq n$, sonst lineare Abhängigkeit!)

Relativ übersichtlich ist noch der Fall von n Vektoren in \mathbb{R}^n . Hier besteht lineare Unabhängigkeit, genau wenn die aus diesen Vektoren spaltenweise zusammengestellte $n \times n$ -Matrix invertierbar A ist; denn genau dann ist das homogene System $Ax = 0$ von n linearen Gleichungen für n Unbekannte ja nur trivial lösbar. Der Tatsache, dass sich an der linearen (Un-)Abhängigkeit der Spalten dieser Matrix durch Zeilen- und Spaltenoperationen nichts ändert, entspricht, dass die Determinante bei solchen Operationen höchstens um einen Faktor $\neq 0$ geändert werden kann, also entweder vor und nach den Operationen von Null verschieden ist oder vorher und nachher Null. Zur Überprüfung der linearen (Un-)Abhängigkeit von n Vektoren in \mathbb{R}^n stehen damit alle Verfahren aus 3.4 zur Verfügung, mit dem man die Invertierbarkeit von quadratischen Matrizen bzw. das Nichtverschwinden ihrer Determinante feststellen kann.

Die n kanonischen Basisvektoren e_1, \dots, e_n in \mathbb{R}^n bilden das einfachste Beispiel eines Systems von n linear unabhängigen Vektoren in \mathbb{R}^n . Da die Vektoren e_j mit $j \neq i$ allesamt Nulleinträge in i -ter Position haben, gilt das auch für jede Linearkombination dieser e_j , und deshalb kann der Vektor e_i , der ja den Eintrag 1 in i -ter Position hat, nicht linear ausgedrückt werden durch die e_j mit $j \neq i$. Die Matrix mit den Spalten e_1, \dots, e_n ist nichts anderes als die $n \times n$ -Einheitsmatrix \mathbb{I}_n , und die Invertierbarkeit dieser Matrix ($\mathbb{I}_n^{-1} = \mathbb{I}_n$ bzw. $\det(\mathbb{I}_n) = 1 \neq 0$) ist ein weiterer Nachweis der linearen Unabhängigkeit der kanonischen Basisvektoren. Natürlich gibt es viele andere Systeme von n linear unabhängigen Vektoren in \mathbb{R}^n , z.B. die Spalten einer beliebigen $n \times n$ -Dreiecksmatrix mit Diagonaleinträgen $\neq 0$; denn die n Vektoren sind ja linear unabhängig in \mathbb{R}^n , genau wenn sie die Spalten einer invertierbaren $n \times n$ -Matrix sind. Alle Systeme von n linear unabhängigen Vektoren in \mathbb{R}^n erhält man analog, indem man die Spalten von nicht invertierbaren $n \times n$ -Matrizen nimmt. Beispiele dafür sind natürlich weniger interessant; man kann ja einfach alle n Vektoren als Nullvektor wählen, oder als Vielfache eines festen Vektors, oder den dritten als Summe der beiden ersten etc.

(5) Unendlich viele linear unabhängige Vektoren gibt es z.B. im Vektorraum \mathbb{R}^∞ aller unendlichen reellen Zahlenfolgen (r_1, r_2, \dots) mit nur endlich (aber beliebig) vielen Einträgen $r_j \neq 0$. Diese Folgen addiert und vervielfacht man genau so komponentenweise, wie man es bei den n -gliedrigen Folgen $(r_1, \dots, r_n) \in \mathbb{R}^n$ macht. In \mathbb{R}^∞ hat man die kanonischen Basisfolgen e_i , die einen Eintrag 1 in i -ter Position haben und sonst nur (unendlich viele) Nulleinträge. Mit derselben Argumentation wie in (4) sieht man, dass keine Basisfolge e_i als Linearkombination von (endlich vielen) Basisfolgen e_j mit $j \neq i$ dargestellt werden kann, weil solche Linearkombinationen den Eintrag 0 in der i -ten Position haben.

Ein äquivalentes, aber vielleicht weniger "konstruiert" erscheinendes, Beispiel ist der Raum \mathcal{P} aller Polynomfunktionen $p(x) = c_n x^n + c_{n-1} x^{n-1} + \dots + c_1 x + c_0$ einer reellen Variablen $x \in \mathbb{R}$ mit beliebigem Grad n und beliebigen Koeffizienten $c_0, \dots, c_n \in \mathbb{R}$, $c_n \neq 0$ (außer im Fall $n = 0$, wo auch $c_0 = 0$ erlaubt ist, so dass p die Nullfunktion ist). Die Polynomfunktionen addiert und vervielfacht man wie üblich "koeffizientenweise" und erhält so einen Vektorraum \mathcal{P} . Darin haben wir als Vektoren insbesondere die Potenzfunktionen $p_n(x) = x^n$ für $n \in \mathbb{N}_0$ (mit $p_0(x) := 1$ für alle x), und diese sind linear unabhängig. Andernfalls könnte man nämlich p_n für ein n durch p_0, \dots, p_{n-1} linear ausdrücken, d.h. $p_n = s_{n-1} p_{n-1} + \dots + s_1 p_1 + s_0 p_0$ gälte mit gewissen Koeffizienten $s_0, s_1, \dots, s_{n-1} \in \mathbb{R}$. Dies wäre aber eine für alle $x \in \mathbb{R}$ gültige algebraische Gleichung $x^n = s_{n-1} x^{n-1} + \dots + s_1 x + s_0$, und eine derartige Gleichung hat bekanntlich (siehe 1.2) höchstens n Lösungen. ■

Wir kommen nun zum zweiten fundamentalen Begriff dieses Abschnitts, mit dem wir das Konzept einer k -parametrischen linearen Schar $v(r_1, \dots, r_k)$ von Vektoren in einem Vektorraum V präzise fassen. Die Linearität der Schar bedeutet, dass $\mathbb{R}^k \ni \mathbf{r} = (r_1, \dots, r_k) \mapsto v(\mathbf{r}) = v(r_1, \dots, r_k) \in V$ eine lineare Abbildung ist, also mit der Bildung von Linearkombinationen vertauscht. Da $\mathbf{r} = \sum_{j=1}^k r_j e_j$ als Linearkombination der kanonischen Basisvektoren e_j des \mathbb{R}^k darstellbar ist, bedeutet das $v(r_1, \dots, r_k) = \sum_{j=1}^k r_j v_j$ mit den Vektoren $v_j := v(e_j) \in V$. Eine k -parametrische lineare Schar in V ist also nichts anderes als die Bildung aller möglichen Linearkombinationen von k gegebenen Vektoren $v_1, \dots, v_k \in V$. Dabei muss man die lineare Parametrisierung, das ist die Abbildung $(r_1, \dots, r_k) \mapsto \sum_{j=1}^k r_j v_j$, unterscheiden von dem hierdurch parametrisierten Unterraum von V , das ist die Menge aller durch die Parametrisierung erfassten Vektoren aus V . Diese Unterscheidung ist nötig, weil verschiedene Parametrisierungen, d.h. verschiedene Wahlen der Vektorfolge (v_1, \dots, v_k) , denselben Unterraum parametrisieren können. Diese Überlegungen führen letztlich zu folgender

DEFINITION: (i) Eine nichtleere Teilmenge $U \neq \emptyset$ des Vektorraums V heißt ein (**linearer**) **Unterraum** von V , wenn U den Nullvektor enthält und mit je zwei Vektoren u, \tilde{u} auch die Summe $u + \tilde{u}$ sowie alle Vielfachen ru ($r \in \mathbb{R}$).

(ii) Sind v_1, \dots, v_k gegebene Vektoren im Vektorraum V , so heißt die Menge aller Linearkombinationen $\sum_{j=1}^k r_j v_j$ mit Koeffizienten $r_1, \dots, r_k \in \mathbb{R}$ der von den Vektoren v_1, \dots, v_k **aufgespannte Unterraum** U und wird $\text{Span}(v_1, \dots, v_k)$ oder $\mathbb{R}v_1 + \dots + \mathbb{R}v_k$ notiert. Man sagt auch, dass die Vektoren v_1, \dots, v_k ein **Erzeugendensystem** des Unterraums U bilden und dass die Abbildung $\mathbb{R}^k \ni (r_1, \dots, r_k) \mapsto \sum_{j=1}^k r_j v_j \in U$ eine **k -parametrische lineare Parametrisierung** des Unterraums U ist. ■

DISKUSSION und BEISPIELE: (1) Unterräume von V sind insbesondere V selbst und der **Nullraum** $\{0\}$, der nur aus dem Nullvektor von V besteht. Die von V verschiedenen Unterräume nennt man auch **echte Unterräume**. Mit den Rechenoperationen des Vektorraums V ist jeder Unterraum U wieder ein Vektorraum; denn Summen und Vielfache von Vektoren aus U liegen ja per Definition wieder im Unterraum U und die grundlegenden Rechenregeln gelten für das Rechnen mit Vektoren aus U natürlich auch, da sie ja sogar für das Rechnen in V Gültigkeit haben. Mit Summen und Vielfachen enthält ein Unterraum auch beliebige Linearkombinationen seiner Vektoren.

(2) Die Menge U aller Linearkombinationen der Vektoren v_1, \dots, v_k in (ii) ist tatsächlich ein Unterraum von V ; denn mit $u = \sum_{j=1}^k s_j v_j \in U$ und $\tilde{u} = \sum_{j=1}^k t_j v_j \in U$ sind auch $u + \tilde{u} = \sum_{j=1}^k (s_j + t_j) v_j$ und $ru = \sum_{j=1}^k (rs_j) v_j$ für alle $r \in \mathbb{R}$ Linearkombinationen der Vektoren v_j . Da ein Unterraum alle Linearkombinationen seiner Vektoren enthält, ist U offenbar der kleinste Unterraum von V , in dem alle Vektoren v_1, \dots, v_k liegen. In der Literatur sind neben den in (ii) angegebenen Benennungen und Notationen noch andere verbreitet. So wird U auch als der von v_1, \dots, v_k *erzeugte Unterraum* oder als *lineare Hülle* der Vektoren v_1, \dots, v_k bezeichnet und $\text{Lin}\{v_1, \dots, v_k\}$ oder $\langle v_1, \dots, v_k \rangle$ notiert.

(3) Der Nullvektor $0 \in V$ erzeugt den Nullraum $\{0\}$ (dem man auch das leere Erzeugendensystem formal zuordnet). Der von einem Vektor $v \neq 0$ aufgespannte Unterraum $\mathbb{R}v$ von V besteht aus allen Vielfachen rv ($r \in \mathbb{R}$) und wird eine *Ursprungsgerade* in V genannt, weil er im Fall der "Zeichenebene" $V = \mathbb{R}^2$ oder des "Anschauungsraumes" eine Gerade beschreibt, die durch den Ursprung (Nullvektor) verläuft. Zwei Vektoren $v, w \in V$ spannen dieselbe Ursprungsgerade $\mathbb{R}v$ auf, wenn w ein Vielfaches von v ist und $v \neq 0$, und eine sog. *Ursprungsebene* $\mathbb{R}v + \mathbb{R}w$ in V , wenn v, w linear unabhängig sind.

(4) Ein Erzeugendensystem des n -dimensionalen Zahlenraums \mathbb{R}^n ist z.B. die kanonische Basis e_1, \dots, e_n ; denn jeder Vektor $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ ist ja Linearkombination $\mathbf{x} = \sum_{j=1}^n x_j e_j$ der kanonischen Basisvektoren. Ein Vektorraum V heißt **endlich erzeugt**, wenn er – wie \mathbb{R}^n – ein endliches Erzeugendensystem besitzt. Man kann dann zeigen (s.u.), dass auch jeder Unterraum endlich erzeugt ist (von höchstens so vielen Vektoren, wie man zur Erzeugung von V braucht). Daher lassen sich in diesem Falle *alle* Unterräume von V in der Form $\mathbb{R}v_1 + \dots + \mathbb{R}v_k$ beschreiben mit geeigneten Wahlen von $k \in \mathbb{N}_0$ und von $v_1, \dots, v_k \in V$. Das gilt insbesondere für alle Unterräume von \mathbb{R}^n .

(5) Mit jeder $m \times n$ -Matrix A sind zwei Unterräume assoziiert, nämlich die **Lösungsmenge** $L \subset \mathbb{R}^n$ des **homogenen Gleichungssystems** $A\mathbf{x} = 0$ und der **Raum** $K \subset \mathbb{R}^m$ **der konsistenten rechten Seiten**, also die Menge der Vektoren $\mathbf{b} \in \mathbb{R}^m$, für die $A\mathbf{x} = \mathbf{b}$ lösbar ist. Hat A die Spaltenvektoren $\mathbf{a}_1, \dots, \mathbf{a}_n$, so gilt $A\mathbf{x} = x_1\mathbf{a}_1 + \dots + x_n\mathbf{a}_n$, daher wird K von den Spaltenvektoren von A aufgespannt.

- *Der Raum der konsistenten rechten Seiten zu $A \in \mathbb{R}^{m \times n}$ ist der von den Spaltenvektoren der Matrix A aufgespannte Unterraum von \mathbb{R}^m .*

Die Lösungsmenge L hatten wir in 3.2 linear parametrisiert, indem wir die Nicht-Basisvariablen einer Zeilen-Stufen-Form als Parameter gewählt hatten. Dies kann man so beschreiben:

- *Der Lösungsraum eines homogenen linearen Gleichungssystems $A\mathbf{x} = 0$ wird aufgespannt von den Lösungsvektoren, die man erhält, wenn man in einer Zeilen-Stufen-Form von A jeweils einer der Nicht-Basisvariablen den Wert 1 zuweist und den anderen Nicht-Basisvariablen den Wert 0.*

(6) Wir haben damit zwei verschiedene Möglichkeiten, Unterräume U von \mathbb{R}^n zu beschreiben: Bei einer **impliziten Darstellung des Unterraums** wird U angegeben als Lösungsmenge eines homogenen linearen Gleichungssystems $A\mathbf{x} = 0$ für n Unbekannte $\mathbf{x} = (x_1, \dots, x_n)$; bei einer **parametrischen Darstellung des Unterraums** gibt man eine lineare Parametrisierung von U an oder, was auf dasselbe hinausläuft, ein Erzeugendensystem $\mathbf{a}_1, \dots, \mathbf{a}_k$ von U . Mit Hilfe der Zeilen-Stufen-Form von A kann man, wie gerade in (5) beschrieben, den Übergang von einer impliziten zu einer parametrischen Darstellung vollziehen.

Auch der umgekehrte Übergang, also die Aufstellung eines homogenen linearen Gleichungssystems, das einen parametrisch gegebenen Unterraum als Lösungsmenge hat, lässt sich mit der Zeilen-Stufen-Form erledigen. Dazu bildet man die $n \times k$ -Matrix A mit den Vektoren \mathbf{a}_j des Erzeugendensystems von U als Spalten, erweitert sie zu $(A|\mathbf{x})$ durch eine Spalte \mathbf{x} von Unbekannten und bringt sie durch Zeilenoperationen auf Zeilen-Stufen-Form. Die Spalte der rechten Seiten der Zeilen-Stufen-Form hat dann Einträge $\ell_h(\mathbf{x}) = c_{h1}x_1 + \dots + c_{hn}x_n$ mit linearen Funktionen ℓ_h , deren Koeffizienten c_{hi} sich aus den durchgeführten Zeilenoperationen ergeben. Hat die Koeffizientenmatrix der Zeilen-Stufen-Form l Stufen und darunter $n-l$ Nullzeilen, so sind die Gleichungen $\ell_h(\mathbf{x}) = 0$ für $h = l+1 \dots n$ äquivalent damit, dass \mathbf{x} eine zu A konsistente rechte Seite ist, also ein Vektor im Unterraum $U = \mathbb{R}\mathbf{a}_1 + \dots + \mathbb{R}\mathbf{a}_k$. Damit ist U beschrieben als Lösungsmenge des homogenen linearen Gleichungssystems $\sum_{i=1}^n c_{hi}x_i = 0$, $h = l+1 \dots n$. (Im Fall $l = n$ ist jeder Vektor $\mathbf{x} \in \mathbb{R}^n$ konsistent mit A , d.h. U ist dann der ganze Raum \mathbb{R}^n selbst. Diesen kann man als Lösungsmenge einer homogenen linearen Gleichung mit lauter Nullkoeffizienten darstellen.)

Eine einfachere Methode, zu einem parametrisch beschriebenen Unterraum U von \mathbb{R}^n ein homogenes lineares Gleichungssystem zu konstruieren, das U als Lösungsmenge hat, werden wir später in diesem Abschnitt beschreiben. Es sei noch darauf hingewiesen, dass weder die linearen Gleichungssysteme, noch die linearen Parametrisierungen / Erzeugendensysteme, die einen gegebenen Unterraum beschreiben, eindeutig bestimmt sind: Verschiedene homogene lineare Gleichungssysteme können dieselbe Lösungsmenge haben und verschiedene Systeme von endlich vielen Vektoren denselben Unterraum aufspannen!

(7) Analog zu (5) sind mit jeder linearen Abbildung $T: V \rightarrow W$ zwischen Vektorräumen zwei Unterräume assoziiert. Einer ist der sog. **Kern** oder **Anullierungsraum** von T , notiert $\text{Kern}(T)$ oder $T^{-1}\{0\}$; das ist die Menge aller Vektoren $v \in V$ mit $T(v) = 0$. Der zweite ist das **Bild** der linearen Abbildung T , notiert $\text{Bild}(T)$ oder $T(V)$, das ist die Menge aller Werte $T(v) \in W$, die T annimmt, wenn v den Definitionsbereich V durchläuft. Mit Hilfe der Linearität der Abbildung T prüft man leicht nach, dass $\text{Kern}(T)$ ein Unterraum des Ausgangsraums V ist und $\text{Bild}(T)$ ein Unterraum des Zielraums W . Die Kerne und Bilder linearer Abbildungen verallgemeinern die in (5) mit Matrizen assoziierten Unterräume. Im Fall $V = \mathbb{R}^n$, $W = \mathbb{R}^m$ und $T(\mathbf{x}) = A\mathbf{x}$ mit einer $m \times n$ -Matrix A ist nämlich $\text{Kern}(T)$ der Lösungsraum des homogenen Gleichungssystems $A\mathbf{x} = 0$ und $\text{Bild}(T)$ der Raum der mit A konsistenten rechten Seiten.

Allgemeiner ist für jeden Unterraum U von V das Bild $T(U) := \{T(u) : u \in U\}$ des Unterraums unter einer linearen Abbildung $T: V \rightarrow W$ wieder ein Unterraum von W , und für jeden Unterraum \tilde{U} von W ist das Urbild $T^{-1}\tilde{U} := \{v \in V : T(v) \in \tilde{U}\}$ ein Unterraum von V . Das folgt aus der Unterraum-Definition und der Definition der Linearität von Abbildungen auch sofort. Wir halten fest:

- Für $T: V \rightarrow W$ linear ist das Bild $T(U)$ jedes Unterraums U von V ein Unterraum von W und das Urbild $T^{-1}\tilde{U}$ jedes Unterraums \tilde{U} von W ein Unterraum von V .
- Das gilt insbesondere für $\text{Bild}(T) = T(V) \subset W$ und für $\text{Kern}(T) = T^{-1}\{0\} \subset V$.

Bezüglich der Wirkung von linearen Abbildungen $T: V \rightarrow W$ auf Erzeugendensysteme können wir anhand der Gleichung $T(\sum_{j=1}^k r_j v_j) = \sum_{j=1}^k r_j T(v_j)$ folgende einfache Beobachtung machen: Wird der Unterraum U aufgespannt von den Vektoren $v_1, \dots, v_k \in V$, so bilden die Bildvektoren $T(v_1), \dots, T(v_k)$ ein Erzeugendensystem im Bild $T(U) \subset W$. Ein Erzeugendensystem des Zielraums W selbst bekommt man auf diese Weise aus einem Erzeugendensystem von V aber nur, wenn $\text{Bild}(T) = W$ ist, d.h. wenn es zu jedem $w \in W$ ein $v \in V$ gibt mit $w = T(v)$. Abbildungen mit dieser Eigenschaft nennt man **surjektiv** (bzgl. des Zielraums W), so dass wir festhalten können:

- Eine lineare Abbildungen $T: V \rightarrow W$ führt ein gegebenes Erzeugendensystem von V in ein Erzeugendensysteme von W über, genau wenn sie surjektiv ist.

(8) Die Lösungsmenge eines inhomogenen linearen Gleichungssystems $A\mathbf{x} = \mathbf{b}$ für n Unbekannte $\mathbf{x} = (x_1, \dots, x_n)$ ist kein Unterraum von \mathbb{R}^n , wenn wirkliche Inhomogenität vorliegt, wenn also der Vektor \mathbf{b} der rechten Seiten nicht der Nullvektor ist. Dann ist nämlich die Lösungsmenge leer (der Inkonsistenzfall – damit muss man immer rechnen, wenn das Gleichungssystem inhomogen ist!) oder sie enthält jedenfalls nicht den Nullvektor des \mathbb{R}^n . Da aber die Lösungsmenge M des inhomogenen Systems im Konsistenzfall in der Form $M = \mathbf{x}_* + L := \{\mathbf{x}_* + \mathbf{x} : \mathbf{x} \in L\}$ beschrieben werden kann, wo L die Lösungsmenge des homogenen Systems $A\mathbf{x} = 0$ ist und \mathbf{x}_* irgendeine (“spezielle”) Lösung des inhomogenen Systems, unterscheidet M sich von dem Unterraum L nur durch die Verschiebung um einen festen Vektor \mathbf{x}_* . Geometrisch bedeutet dies, dass M parallel zu dem Unterraum L ist (aber nicht durch den Nullpunkt geht, wenn $\mathbf{b} \neq 0$ bzw. $\mathbf{x}_* \notin L$).

Eine Teilmenge M eines Vektorraums V , die durch Parallelverschiebung $u \mapsto v+u$ eines Unterraums U von V um einen festen Vektor $v \in V$ entsteht, d.h. es ist $M = v + U := \{v+u : u \in U\}$, heißt **affiner Unterraum** (oder *affin-linearer Unterraum*) von V . Dies ist kein linearer Unterraum, abgesehen von dem Fall $v \in U$, in dem $M = U$ ist. Mit dieser Terminologie können wir nun sagen:

- Die Lösungsmenge eines beliebigen linearen Gleichungssystems für n Unbekannte (homogen oder nicht) ist entweder leer oder ein affiner Unterraum von \mathbb{R}^n , der zur Lösungsmenge des zugehörigen homogenen Gleichungssystems parallel ist.
- Nur bei einem homogenen linearen Gleichungssystem ist die Lösungsmenge ein linearer Unterraum von \mathbb{R}^n ; andernfalls ist sie leer oder ein nichtleerer affiner Unterraum, der nicht den Nullvektor enthält.

Geometrisch sind die affinen Unterräume $v + U$ beliebige, d.h. nicht unbedingt durch den "Ursprung" $0 \in V$ gehende Punkte, Geraden bzw. Ebenen, ... in V , wenn U der Nullraum, eine Ursprungsgerade bzw. eine Ursprungsebene ... in V ist. Man kann auch den Inkonsistenzfall geometrisch verstehen: Eine einzelne lineare Gleichung $\ell(\mathbf{x}) = b$ für n Unbekannte beschreibt, wenn sie nicht schon in sich inkonsistent ist (d.h. die linke Seite ist Null für alle \mathbf{x} , aber die rechte Seite ist nicht Null), eine sog. *affine Hyperebene* H in \mathbb{R}^n , d.h. eine Gerade in \mathbb{R}^2 , eine Ebene in \mathbb{R}^3 , einen 3-dimensionalen affinen Unterraum in \mathbb{R}^4 usw. (den Dimensionsbegriff erklären wir gleich noch genauer). Die Lösungsmenge eines Systems von konsistenten linearen Gleichungen $\ell_i(\mathbf{x}) = b_i$, $i = 1 \dots m$, ist nun der Durchschnitt der entsprechenden Hyperebenen H_i . Wenn z.B. zwei dieser Hyperebenen parallel sind, aber nicht identisch, so schneiden sie sich gar nicht, und die Lösungsmenge des Gleichungssystems ist leer. Dies ist die Situation, in der schon zwei der Gleichungen im Widerspruch zueinander stehen. Allgemeiner liegt der Inkonsistenzfall vor, wenn für ein i der Durchschnitt $H_1 \cap \dots \cap H_{i-1}$ parallel aber nicht identisch ist mit einem affinen Unterraum von H_i , so wie z.B. der Durchschnitt von zwei Ebenen in \mathbb{R}^3 eine Gerade sein kann, die eine dritte Ebene nicht schneidet.

Der obigen Feststellung, dass die Lösungsmenge eines linearen Gleichungssystems leer ist oder ein affiner Unterraum, entspricht bei linearen Abbildungen $T:V \rightarrow W$ zwischen allgemeinen Vektorräumen die Aussage, dass das Urbild $T^{-1}\{w\} = \{v \in V : T(v) = w\}$ eines beliebigen Punktes $w \in W$ entweder leer ist (wenn nämlich $w \notin \text{Bild}(T)$ ist) oder ein zu $\text{Kern}(T)$ paralleler affiner Unterraum von V (nämlich $T^{-1}\{w\} = v + \text{Kern}(T)$ für jeden Vektor $v \in V$ mit $T(v) = w$). Allgemeiner gilt, dass die T -Urbilder affiner Unterräume von W affine Unterräume von V sind und die T -Bilder affiner Unterräume von V affine Unterräume von W . Dieser geometrischen Eigenschaft, dass sie "lineare Mengen" wie Punkte, Geraden, Ebenen, ... in lineare Mengen (nicht unbedingt derselben Dimension) überführen, verdanken die "linearen" Abbildungen letztlich ihren Namen!

(9) Die Räume \mathbb{R}^∞ und \mathcal{P} aus (5) der letzten Beispielserie sind *nicht endlich erzeugte Vektorräume* (man sagt auch *Vektorräume unendlicher Dimension*), d.h. die Linearkombinationen von endlich vielen gegebenen Vektoren erfassen nie alle Vektoren des Raumes. Z.B. sind die Linearkombinationen von endlich vielen Polynomen $p_1, \dots, p_k \in \mathcal{P}$ Polynome von einem Grad, der höchstens so groß ist wie das Maximum m der Grade von p_1, \dots, p_k ; also kann man kein Polynom $p \in \mathcal{P}$ vom Grad $> m$ als Linearkombination von p_1, \dots, p_k darstellen. Bei unendlichdimensionalen Vektorräumen V lassen sich nur die endlich erzeugten Unterräume in der Form $\mathbb{R}v_1 + \dots + \mathbb{R}v_k$ wie in (4) darstellen; es gibt aber (sehr viel mehr) nicht endlich erzeugte Unterräume, die man (von wenigen Ausnahmen abgesehen) nicht explizit durch lineare Parametrisierungen oder als Lösungsräume zu homogenen linearen Gleichungssystemen beschreiben kann. ■

Ein Vektorraum oder Unterraum hat, wenn er endlich erzeugt und nicht der Nullraum ist, viele verschiedene Erzeugendensysteme und Parametrisierungen. Beispielsweise erzeugen $2v_1, v_2, \dots, v_k$ und $v_1 + sv_2, v_2, \dots, v_k$ offenbar denselben Unterraum wie v_1, v_2, \dots, v_k . Natürlich gibt es für jeden Raum auch Erzeugendensysteme mit unterschiedlich vielen Vektoren bzw. lineare Parametrisierungen mit unterschiedlicher Anzahl von reellen Parametern. Fügt man nämlich zu einem Erzeugendensystem eines Unterraums beliebige Vektoren aus diesem Unterraum hinzu, so erhält man offenbar wieder ein Erzeugendensystem dieses Unterraums. Daher stellt sich die Frage nach Parametrisierungen mit der kleinstmöglichen Zahl von Parametern. Bei linearen Gleichungssystemen wollen wir schließlich die allgemeine Lösung auch nicht mit mehr Parametern beschreiben als nötig. Die Antwort liegt in der Verbindung von Erzeugendensystemen mit linearer Unabhängigkeit und führt zur dritten grundlegenden Begriffsbildung dieses Abschnitts:

SATZ und DEFINITION: Für jeden Vektorraum U (und insbesondere für jeden Unterraum U eines Vektorraums V wie z.B. $V = \mathbb{R}^n$) sind folgende Aussagen über eine endliche Folge v_1, \dots, v_k von Vektoren aus U äquivalent:

- (i) v_1, \dots, v_k ist ein minimales Erzeugendensystem für U ;
- (ii) v_1, \dots, v_k ist ein maximales System linear unabhängiger Vektoren in U ;
- (iii) v_1, \dots, v_k ist ein linear unabhängiges Erzeugendensystem für U ;
- (iv) jeder Vektor $u \in U$ ist eindeutige Linearkombination von v_1, \dots, v_k .

Hat die Folge diese Eigenschaften, so heißt sie eine (endliche) **Basis** des Vektorraums U . Eine Basis von $U \neq \{0\}$ existiert genau dann, wenn U ein endliches Erzeugendensystem besitzt. Alle Basen von U haben dann dieselbe Anzahl $k \in \mathbb{N}$ von Vektoren, und diese Zahl wird die **Dimension** des Vektorraums U genannt und $\dim U$ notiert. (Für den Nullraum vereinbart man noch $\dim\{0\} = 0$.) Ist $U \neq \{0\}$ Unterraum eines endlich erzeugten Vektorraums V (wie z.B. \mathbb{R}^n), so hat U eine Basis und es gilt $\dim U \leq \dim V$.

Einige Worte zur Erklärung der hier verwendeten Terminologie: (i) besagt natürlich, dass man nach Entfernung eines Vektors aus v_1, \dots, v_k nicht mehr jedes $u \in U$ als Linearkombination der restlichen Vektoren darstellen kann. (ii) bedeutet, dass Hinzufügen eines einzigen Vektors aus U zu v_1, \dots, v_k schon die lineare Unabhängigkeit zerstört. (iii) ist selbsterklärend. (iv) besagt erstens, dass jeder Vektor $u \in U$ als Linearkombination $u = \sum_{j=1}^k r_j v_j$ darstellbar ist, und zweitens, dass dies nur auf eine Weise möglich ist, d.h. mit eindeutig bestimmten Koeffizienten r_j . Die Vektoren $u \in U$ und die k -tupel $(r_1, \dots, r_k) \in \mathbb{R}^k$ ihrer Koeffizienten bzgl. der Basis bestimmen sich daher gegenseitig. Das bedeutet letztlich, dass man in einem k -dimensionalen Vektorraum mit Basis genau so rechnen kann wie mit Vektoren im Zahlenraum \mathbb{R}^k , und darin liegt die Bedeutung des Basisbegriffs! (Wir präzisieren diese Bemerkung weiter unten.)

Obwohl es für das Verständnis und die Anwendung des Satzes nicht nötig ist, wollen wir auch einige Worte zum *Beweis* sagen, weil er einfach ist und letztlich auf Aussagen über lineare Gleichungssysteme zurückgeführt werden kann, die wir schon in 3.2 behandelt haben. Zunächst ist (iii) \iff (iv) klar; denn die Erzeugungseigenschaft ist äquivalent mit der Darstellbarkeit $u = \sum_{j=1}^k r_j v_j$ eines jeden Vektors $u \in U$ und die lineare Unabhängigkeit ist äquivalent mit der Eindeutigkeit der Koeffizienten in jeder solchen Darstellung. Klar ist auch, dass man ein System v_1, \dots, v_k mit der Eigenschaft (iii) nicht verkleinern kann, ohne die Erzeugungseigenschaft zu verlieren (der weggelassene Vektor kann ja nicht linear durch die anderen ausgedrückt werden), und auch nicht vergrößern kann, ohne die

lineare Unabhängigkeit zu zerstören (jeder aus U hinzugenommene Vektor kann ja durch die v_j linear ausgedrückt werden). Daher folgen (i) und (ii) aus (iii). Schließlich muss ein minimales Erzeugendensystem v_1, \dots, v_k von U linear unabhängig sein, weil man sonst einen der Vektoren, den man durch die anderen linear ausdrücken kann, für die Erzeugung von U gar nicht bräuchte, und ein maximales linear unabhängiges System v_1, \dots, v_k in U muss U erzeugen, weil man sonst einen Vektor aus U , der sich nicht linear durch die v_j ausdrücken lässt, hinzunehmen könnte, ohne die lineare Unabhängigkeit zu zerstören. Damit ist die Äquivalenz der vier Aussagen schon gezeigt.

Wenn $U \neq \{0\}$ ein endliches Erzeugendensystem u_1, \dots, u_l besitzt, so können wir darin ein linear unabhängiges Erzeugendensystem $v_j = u_{i_j}$, $j = 1 \dots k$, auswählen, indem wir die Vektoren u_i in aufsteigender Nummerierung durchgehen und diejenigen wegstreichen, die Null sind oder sich durch die vorangehenden Vektoren linear ausdrücken lassen. Daher hat dann U eine Basis v_1, \dots, v_k . Sind nun u_1, \dots, u_l irgendwelche Vektoren aus U , so können wir sie darstellen $u_h = \sum_{j=1}^k r_{jh} v_j$ mit einer eindeutigen $k \times l$ -Matrix $R = (r_{jh})$. Ist $k < l$, so hat gemäß 3.2 das homogene Gleichungssystem $Rs = 0$ nichttriviale Lösungen $s = (s_1, \dots, s_l)$, also $\sum_{h=1}^l r_{jh} s_h = 0$ für $j = 1 \dots k$. Dann ist $\sum_{h=1}^l s_h u_h = \sum_{j=1}^k (\sum_{h=1}^l s_h r_{jh}) v_j = 0$ eine nichttriviale Linearkombination, die den Nullvektor darstellt, also sind u_1, \dots, u_l linear abhängig in U . Dieses Argument zeigt, dass mehr als k Vektoren in U stets linear abhängig sind, d.h. dass keine Basen mit mehr als k Elementen existieren. Dasselbe Argument mit vertauschten Rollen der v_j und u_h zeigt aber, dass auch keine Basis mit weniger als k Elementen existiert, also haben alle Basen gleich viele Elemente, und die Dimension $\dim U = k$ ist eindeutig definiert. Schließlich zeigt das Argument noch, dass jedes linear unabhängige System in einem Unterraum von U höchstens k Elemente hat, und das gilt dann auch für ein maximales derartiges System, also eine Basis des Unterraums, weshalb seine Dimension $\leq k = \dim U$ sein muss. Damit ist der Satz vollständig bewiesen (wobei wir die letzte Aussage im Satz für Unterräume von V formuliert haben statt von U).

DISKUSSION: (1) Die oben definierte Dimension von Vektorräumen bzw. Unterräumen stimmt überein mit unseren intuitiven Vorstellungen und mit dem in 3.2 schon verwendeten Dimensionsbegriff. In \mathbb{R}^n hat also der Nullraum $\{0\}$ die Dimension 0, jede Ursprungsgerade $\mathbb{R}v$ ($v \neq 0$) die Dimension 1, jede Ursprungsebene $\mathbb{R}u + \mathbb{R}v$ (u, v linear unabhängig) die Dimension 2, ... und der ganze Raum natürlich die Dimension n .

Mit (i) und (ii) des vorigen Satzes haben wir zwei verschiedene Interpretationen der Dimension eines Vektorraums bzw. Unterraums U . Da die Erzeugendensysteme v_1, \dots, v_k von U den linearen Parametrisierungen $\mathbb{R}^k \ni (r_1, \dots, r_k) \mapsto \sum_{j=1}^k r_j v_j$ von U entsprechen, ist die kleinstmögliche Anzahl der Parameter einer solchen Parametrisierung gleich der Anzahl der Vektoren in den minimalen Erzeugendensystemen, also gleich $\dim U$. Dies ist eine Interpretation der Dimension, die wir schon in 3.2 verwendet haben. Die Beschreibung von $\dim U = k$ als Anzahl der Vektoren eines maximalen linear unabhängigen Systems v_1, \dots, v_k in U gibt eine zweite Interpretation der Dimension: Jedes solche System definiert eine sog. *Fahne* in U , d.h. eine aufsteigende Folge von Unterräumen $\{0\} = U_0 \subset U_1 \subset U_2 \subset \dots \subset U_k = U$ von denen jeder echter Unterraum des folgenden ist, nämlich $U_j := \mathbb{R}v_1 + \dots + \mathbb{R}v_j$. Und die Länge k dieser Fahne ist maximal, d.h. es gibt keine längere Folge von Unterräumen U_j von U mit derselben Eigenschaft (sonst erhielte man durch Wahl von v_j in $U_j \setminus U_{j-1}$ eine Folge von mehr als k linear unabhängigen Vektoren in U).

- Die Dimension eines Vektor(-Unter-)Raums U ist die minimale Zahl der Parameter, die man für eine lineare Parametrisierung von U benötigt.
- Die Dimension von U ist auch die Länge jeder maximalen Fahne in U , also die Anzahl der Glieder einer längstmöglichen Folge von echten Unterräumen von U , von denen jeder ein echter Unterraum des folgenden ist.

Die zweite Interpretation entspricht der geometrischen Anschauung noch direkter als die erste. Z.B. hat der "Anschauungsraum" \mathbb{R}^3 die Dimension 3, weil es darin eine Folge Ursprung \subset Ursprungsgerade \subset Ursprungsebene der Länge 3 gibt, aber keine längere Folge von echt ineinander enthaltenen echten Unterräumen. Als *Dimension eines affinen Unterraums* $M = v + U$ eines Vektorraums V definiert man natürlich die Dimension des zu M parallelen Unterraums U von V , der durch den Nullpunkt verläuft. Somit haben beliebige Punkte, Geraden, Ebenen, ... in V die Dimensionen 1, 2, 3, ...

(2) In (1) haben wir stillschweigend vorausgesetzt, dass U endlich erzeugt ist. Sonst hat ja U gar keine endliche Basis, und in diesem Fall setzt man formal $\dim U := \infty$, weil es dann unendliche Folgen von linear unabhängigen Vektoren in U gibt (wie man zeigen kann). In der Mathematik wird auch für Vektorräume V unendlicher Dimension der Begriff einer Basis definiert als (unendliches) System $(v_j)_{j \in J}$ von Vektoren in V , derart dass sich jeder Vektor $v \in V$ als eindeutige Linearkombination von endlich vielen der Vektoren v_j darstellen lässt. (Eindeutigkeit bedeutet hier, dass die in der Linearkombination auftretenden Koeffizienten $r_j \neq 0$ eindeutig bestimmt sind.) Es kann bewiesen werden, dass auch jeder unendlichdimensionale Vektorraum V eine Basis in diesem Sinne besitzt; jedoch ist dies eine reine Existenzaussage ohne praktischen Nutzen, und man kann nur in seltenen Ausnahmefällen eine unendliche Basis konkret angeben.

Im Raum \mathbb{R}^∞ der unendlichen Zahlenfolgen mit endlich (aber beliebig) vielen von Null verschiedenen Gliedern bilden z.B. die kanonischen Basisfolgen e_j (mit dem Eintrag 1 in j -ter Position und unendlich vielen Nulleinträgen sonst) eine Basis. Und im Raum \mathcal{P} der Polynomfunktionen von einer reellen Variablen x sind die Potenzfunktionen $p_j(x) = x^j$ ($j \in \mathbb{N}_0$) eine Basis. Im Raum aller unendlichen reellen Zahlenfolgen (ohne die Bedingung, dass nur endlich viele Glieder von Null verschieden sind) und im Raum aller Funktionen einer reellen Variablen (die z.B. stetig sind oder differenzierbar) kann dagegen niemand eine unendliche Basis konkret angeben (obwohl man weiß, dass solche Basen existieren).

(3) Ist V ein Vektorraum (z.B. Unterraum eines höherdimensionalen Zahlenraumes) mit Basis v_1, \dots, v_n , so kann man zu jedem Vektor $v \in V$ seinen eindeutig bestimmten **Koeffizientenvektor** $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ bilden, d.h. es gilt $v = \sum_{j=1}^n x_j v_j$ in der Linearkombinationsdarstellung bzgl. der gegebenen Basis. Die Abbildung $V \ni v \mapsto \mathbf{x} \in \mathbb{R}^n$, die jedem Vektor seinen Koeffizientenvektor zuordnet, ist dann linear und hat die lineare Umkehrabbildung $\mathbb{R}^n \ni (x_1, \dots, x_n) \mapsto \sum_{j=1}^n x_j v_j \in V$, die jedem gegebenen Koeffizientenvektor die zugehörige Linearkombination der Basisvektoren zuordnet.

Man nennt eine lineare Abbildung $T: V \rightarrow W$ zwischen Vektorräumen einen (**linearen**) **Isomorphismus** von V auf W , wenn sie eine lineare Umkehrabbildung besitzt. Äquivalent ist, dass T linear, injektiv und surjektiv ist. (Die Umkehrabbildung $T^{-1}: W \rightarrow V$ existiert dann und ist automatisch auch linear.) Hat man einen Isomorphismus T zwischen zwei Vektorräumen V, W , so kann man damit alle Aufgabenstellungen und Aussagen der Linearen Algebra in V übertragen in äquivalente Aufgabenstellungen und Aussagen in W . Zum Beispiel ist ein System von Vektoren in V genau dann linear unabhängig bzw. Erzeugendensystem bzw. Basis in V , wenn Entsprechendes für das System der T -Bilder dieser Vektoren in W gilt.

Für Vektorräume V mit gegebener Basis wie oben bedeuten diese Feststellungen konkret das Folgendes:

- Die Abbildung, die jedem Vektor aus V seinen Koeffizientenvektor bzgl. einer gegebenen Basis von V zuordnet, ist ein linearer Isomorphismus von V auf den Zahlenraum \mathbb{R}^n der Dimension $n = \dim V$.
- Mit diesem Isomorphismus kann man alle Aufgabenstellungen der Linearen Algebra in V übertragen in äquivalente Aufgaben im Zahlenraum \mathbb{R}^n und alle Aussagen der Linearen Algebra in \mathbb{R}^n in äquivalente Aussagen in V .

Das präzisiert die im Anschluss an die Basisdefinition gemachte Bemerkung, dass man in einem Vektorraum mit Basis genau so rechnen kann wie in einem Zahlenraum gleicher Dimension. Durch diese "Identifikation" eines Vektorraum V mit gegebener Basis mit einem Zahlenraum \mathbb{R}^n kann man insbesondere den in 3.2 – 3.4 entwickelten Vektor- und Matrix-Kalkül auch für Aufgabenstellungen in beliebigen endlichdimensionalen Vektorräumen einsetzen.

Es ist aber nicht immer sinnvoll, diese Identifikation wirklich vorzunehmen, weil dabei zusätzliche Struktur von V verloren gehen kann. Z.B. kann man den Vektorraum $V = \mathbb{R}^{k \times k}$ der $k \times k$ -Matrizen mit dem Zahlenraum \mathbb{R}^{k^2} identifizieren, indem man die Spalten einer $k \times k$ -Matrix untereinander schreibt als ein Spaltenvektor der Länge k^2 oder die Zeilen nebeneinander als ein Zeilenvektor der Länge k^2 . Beides aber ist nicht zweckmäßig, weil damit die durch die Matrix-Multiplikation gegebene Zusatzstruktur in $\mathbb{R}^{k \times k}$ nicht mehr erkennbar wäre.

(4) Eine *Warnung* ist angebracht: Die Identifikation von Vektoren $v \in V$ mit ihrem Koeffizientenvektor \mathbf{x} im Zahlenraum \mathbb{R}^n hängt wesentlich von der gegebenen Basis v_1, \dots, v_n ab! Wählt man eine andere Basis $\tilde{v}_1, \dots, \tilde{v}_n$, so hat derselbe Vektor v bzgl. dieser zweiten Basis im Allgemeinen einen *anderen* Koeffizientenvektor $\tilde{\mathbf{x}}$. Um den Unterschied zu berechnen, stellen wir die Vektoren der zweiten Basis als Linearkombinationen der ersten Basis dar, $\tilde{v}_j = \sum_{i=1}^n c_{ij} v_i$ für $j = 1 \dots n$, und erhalten so die sog. **Übergangsmatrix** $C = (c_{ij}) \in \mathbb{R}^{n \times n}$ von der ersten zur zweiten Basis. Aus $\sum_{i=1}^n x_i v_i = v = \sum_{j=1}^n \tilde{x}_j \tilde{v}_j = \sum_{i=1}^n (\sum_{j=1}^n \tilde{x}_j c_{ij}) v_i$ und der linearen Unabhängigkeit der v_i erhalten wir dann $x_i = \sum_{j=1}^n c_{ij} \tilde{x}_j$ für $i = 1 \dots n$, also in Matrix-Schreibweise $\mathbf{x} = C\tilde{\mathbf{x}}$. Da hier \mathbf{x} und $\tilde{\mathbf{x}}$ beliebige Vektoren in \mathbb{R}^n sein können und sich gegenseitig eindeutig bestimmen, muss C invertierbar sein, und wir sehen:

- Den Koeffizientenvektor $\tilde{\mathbf{x}}$ von $v \in V$ bzgl. der zweiten Basis erhält man aus dem Koeffizientenvektor \mathbf{x} bzgl. der ersten Basis durch Multiplikation mit der Inversen der Übergangsmatrix, $\tilde{\mathbf{x}} = C^{-1}\mathbf{x}$.

Hier sind die Koeffizientenvektoren natürlich als Spalten zu schreiben, und man muss sich genau an die Anordnung der Indizes bei der Definition der Übergangsmatrix halten (also $\tilde{v}_j = \sum_{i=1}^n c_{ij} v_i$ und *nicht* $\tilde{v}_j = \sum_{i=1}^n c_{ji} v_i$)! Die Argumentation zeigt übrigens auch, dass ein System von n beliebigen Vektoren $\tilde{v}_j = \sum_{i=1}^n c_{ij} v_i$ in V genau dann Basis ist, wenn die Matrix $C = (c_{ij})$ invertierbar ist und dass C^{-1} dann die Übergangsmatrix von der zweiten Basis zur ersten ist. Für eine dritte Basis $\hat{v}_1, \dots, \hat{v}_n$ und die Übergangsmatrix D von den \tilde{v}_j zu den \hat{v}_h folgt wegen $D^{-1}C^{-1} = (CD)^{-1}$, dass CD die Übergangsmatrix für den direkten Übergang von der ersten zur dritten Basis ist. ■

BEISPIELE: (1) Das Paradebeispiel einer Basis ist natürlich die *kanonische Basis* e_1, \dots, e_n des n -dimensionalen Zahlenraums \mathbb{R}^n . Hier ist die fundamentale Basiseigenschaft unmittelbar klar: Jeder Vektor $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ ist Linearkombination der kanonischen Basisvektoren, und der Koeffizient bei e_j muss dabei die entsprechende Komponente x_j von \mathbf{x} sein:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \dots + x_n \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = \sum_{j=1}^n x_j e_j.$$

Ein anderes System von n Vektoren in \mathbb{R}^n ist Basis, genau wenn die Zusammenstellung dieser Spaltenvektoren zu einer $n \times n$ -Matrix invertierbar ist. Diese Matrix ist dann gerade die Übergangsmatrix von der kanonischen Basis zu dieser zweiten Basis.

(2) In \mathbb{R}^3 bilden z.B. die Spaltentripel folgender (invertierbaren) Matrizen Basen:

$$\mathbb{I}_3 := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad C := \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad D := \begin{pmatrix} 0 & 3 & -1 \\ 2 & 1 & 1 \\ -1 & 0 & 0 \end{pmatrix}.$$

Die erste Basis ist die kanonische, zur zweiten bzw. dritten ist C bzw. D die Übergangsmatrix. Der (Spalten-)Vektor $\mathbf{x} := (2, 2, 1)$ hat als Koeffizientenvektor bzgl. der kanonischen Basis natürlich sich selbst, sein Koeffizientenvektor bzgl. der zweiten Basis ist $C^{-1}\mathbf{x} = (0, 1, 1)$ und sein Koeffizientenvektor bzgl. der dritten Basis ist $D^{-1}\mathbf{x} = (0, 1, 1)$, was "zufällig" derselbe Koeffizientenvektor wie bzgl. der zweiten Basis ist (das ist ja nicht ausgeschlossen, auch wenn die Basen verschieden sind). Man braucht dazu die Inversen der Übergangsmatrizen nicht zu berechnen, weil man hier direkt sieht, wie \mathbf{x} aus den Spalten der Matrizen C, D linear zu kombinieren ist: Jeweils als Summe der beiden letzten Spalten. Die Übergangsmatrix von der zweiten zur dritten Basis ist $C^{-1}D$ und muss $(0, 1, 1)$ auf sich abbilden, weil dies der Koeffizientenvektor von \mathbf{x} bzgl. beider Basen ist.

(3) *Basis in einem 1-dimensionalen Vektorraum V* ist jeder Vektor $v \neq 0$ (und der Nullvektor natürlich nicht). Die Übergangsmatrix von v zu einer andern Basis $\tilde{v} = cv$ ($c \in \mathbb{R}_{\neq 0}$) ist die 1×1 -Matrix (c) . In einem 2-dimensionalen Vektorraum wie \mathbb{R}^2 gibt es außer $\{0\}$ und V nur 1-dimensionale Unterräume.

(4) In \mathbb{R}^3 haben wir außer $\{0\}$ und den Ursprungsgeraden noch die Ursprungsebenen E als echte Unterräume. Diese kann man beschreiben als lineäre Hülle $\mathbb{R}v_1 + \mathbb{R}v_2$ von zwei linear unabhängigen Vektoren, die dann eine Basis bilden, oder als Lösungsmenge einer homogenen linearen Gleichung $a_1x_1 + a_2x_2 + a_3x_3 = 0$, deren Koeffizienten nicht alle Null sind. Z.B. werden die drei *Koordinatenebenen* aufgespannt durch Paare e_2, e_3 bzw. e_1, e_3 bzw. e_1, e_2 von kanonischen Basisvektoren oder beschrieben durch die lineare Gleichung $x_1 = 0$ bzw. $x_2 = 0$ bzw. $x_3 = 0$. Dies sind die drei Ursprungsebenen in \mathbb{R}^3 , die jeweils zwei der Kartesischen Koordinatenachsen enthalten.

Die durch $x_1 + x_2 + x_3 = 0$ beschriebene Ebene E der Vektoren mit Komponentensumme Null in \mathbb{R}^3 hat als Basis z.B. $(-1, 1, 0)$, $(-1, 0, 1)$ und auch $(1, 0, -1)$, $(0, 1, -1)$. Um nun z.B. den Koeffizientenvektor $\mathbf{x} = (x_1, x_2)$ von $v := (1, 2, -3) \in E$ bzgl. der ersten Basis zu bestimmen, müssen wir das folgende homogene Gleichungssystem lösen:

$$x_1 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}$$

Durch Betrachtung der zweiten und dritten Komponenten findet man hier sofort $x_1 = 2$, $x_2 = -3$. (Wir wissen ja, dass das Gleichungssystem eindeutig lösbar ist, weil die Spalten der Koeffizientenmatrix eine Basis von E sind und v in E liegt; also brauchen wir die Gleichung für die ersten Komponente gar nicht mehr nachzurechnen – oder allenfalls zur Probe!) Die Übergangsmatrix von der ersten zur zweiten Basis ist hier $C = \begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix}$, sie hat als Spalten die Koeffizientenvektoren der zweiten Basisvektoren bzgl. der ersten Basis. Der Koeffizientenvektor von v bzgl. der zweiten Basis ist dann gegeben durch $\tilde{x} = C^{-1} \begin{pmatrix} 2 \\ -3 \end{pmatrix} = \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 2 \\ -3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. (Das kann man natürlich direkt sehen.)

Die affine Ebene $M := \{x \in \mathbb{R}^3 : x_1 + x_2 + x_3 = 1\}$ aller Punkte in \mathbb{R}^3 mit Komponentensumme 1 kann man nun durch Verschiebung von E um einen beliebigen Vektor $u \in \mathbb{R}^3$ mit Komponentensumme 1 beschreiben, also z.B. $u := (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Dann ist $M = u + E$.

(5) Die Bestimmung einer Basis in einem Unterraum U von \mathbb{R}^n ist im Prinzip einfach, wenn U parametrisch gegeben ist, also mit einem Erzeugendensystem u_1, \dots, u_l . Man geht dann diese Vektoren in aufsteigender Nummerierung durch und streicht diejenigen, die Null sind oder Linearkombination der vorhergehenden, und behält eine Basis v_1, \dots, v_k von U übrig. (Wenn l groß ist, so kann das natürlich einen erheblichen Rechenaufwand bedeuten.) Ist dagegen U implizit beschrieben als Lösung eines homogenen linearen Gleichungssystems $Ax = 0$, so führt die Umformung in Zeilen–Stufen–Form zum Ziel. Man bestimmt die Lösungsvektoren, die sich ergeben, wenn man eine der Nicht–Basisvariablen 1 setzt und die anderen 0, und erhält so ein Erzeugendensystem für den Lösungsraum U . Diese Lösungsvektoren sind aber auch linear unabhängig, bilden also eine Basis in U . Das Verfahren ist um so weniger rechenaufwendig, je kleiner die Kodimension $n - \dim U$ des Unterraums U in \mathbb{R}^n ist, weil man ja $n - k$ unabhängige Gleichungen braucht, um einen k -dimensionalen Unterraum zu beschreiben.

Für eine Ursprungshyperebene $H \subset \mathbb{R}^n$, das ist ein durch eine einzige Gleichung $a_1x_1 + \dots + a_nx_n = 0$ (deren Koeffizienten nicht alle Null sind) beschriebener Unterraum, kann man als Basisvariable x_i wählen, wenn $a_i \neq 0$ ist. Das Verfahren liefert für H dann als Basis die $n-1$ Vektoren $(0, \dots, 0, -a_j/a_i, 0, \dots, 0, 1, 0, \dots, 0) \in \mathbb{R}^n$ mit dem Eintrag $-a_j/a_i$ in i -ter Position, dem Eintrag 1 in einer Position $j \neq i$ und Nulleinträgen sonst. Für eine Ursprungsgerade $G \subset \mathbb{R}^n$ andererseits, die durch $n-1$ unabhängige homogene lineare Gleichungen beschrieben wird, hat man nur eine einzige Basisvariable und erhält dementsprechend einen einzigen Lösungsvektor $\neq 0$, der Basis ist.

Konkret betrachten wir z.B. eine Ursprungsgerade in \mathbb{R}^3 :

$$\left. \begin{array}{l} x_1 \quad -x_3 = 0 \\ x_1 + 2x_2 - x_3 = 0 \end{array} \right\} \text{ hat Koeff.-Matrix } \begin{pmatrix} 1 & 0 & -1 \\ 1 & 2 & -1 \end{pmatrix} \text{ mit Z.-St.-Form } \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 2 \end{pmatrix}.$$

Hier ist x_3 Nicht–Basisvariable, und wenn man $x_3 = 1$ setzt, ergibt sich der Lösungsvektor $(1, -1, 1)$ als Basisvektor der Geraden. (Man hätte mit Spaltenvertauschungen hier auch x_1 oder x_2 als Nicht–Basisvariable wählen können. Wenn aber der Basisvektor der Geraden einen Nulleintrag hat, so kann man die zugehörige Variable natürlich niemals Nicht–Basisvariable sein.)

Als letztes Beispiel betrachten wir noch eine Ursprungsebene in \mathbb{R}^4 :

$$\left. \begin{array}{l} x_1 \quad +x_3 +x_4 = 0 \\ x_2 \quad \quad +x_4 = 0 \end{array} \right\}, \text{ Koeff.-Matrix } \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix} \text{ hat schon Z.-St.-Form.}$$

Setzt man für die Nicht–Basisvariablen (x_3, x_4) die kanonischen Basisvektoren des \mathbb{R}^2 ein, so erhält man $(-1, 0, 1, 0)$, $(-1, -1, 0, 1)$ als Basis der Ebene. ■

Mathematik für Wirtschaftswissenschaftler

(K. Steffen, Heinrich-Heine-Universität Düsseldorf, WS 2006/07)

Mit den nun eingeführten Konzepten der Basen und der Dimension können wir die in 3.2 gefundenen Ergebnisse über den Rang von Matrizen systematischer herleiten und besser verstehen:

DISKUSSION (Rang von Matrizen und von linearen Abbildungen):

(1) Der **Spaltenrang** einer $m \times n$ -Matrix A ist die maximale Zahl k von linear unabhängigen Spalten der Matrix. Da die Spalten von A den zugehörigen Unterraum $K \subset \mathbb{R}^m$ der zu A konsistenten rechten Seiten erzeugen, ist der Spaltenrang die Dimension von K . Spaltenoperationen mit der Matrix A ändern offenbar diesen Unterraum nicht, daher ist k auch die Anzahl der Stufen (also der Spalten, die nicht lauter Nulleinträge haben) in einer Spalten-Stufen-Form von A . Somit stimmt der hier definierte Spaltenrang mit dem in 3.2 definierten überein. Entsprechend ist der **Zeilenrang** von A die maximale Zahl l von linear unabhängigen Zeilen und wird durch Zeilenoperationen nicht geändert.

Nun ändern aber auch Zeilenoperationen den Spaltenrang k nicht. Eine Untermatrix aus k linear unabhängigen Spalten von A behält nämlich bei Zeilenoperationen linear unabhängige Spalten, wie wir schon gesehen haben, daher hat auch jede aus A durch Zeilenoperationen entstehende Matrix \tilde{A} mindestens k unabhängige Spalten. Andererseits hat jede aus mehr als k Spalten von A gebildete Untermatrix von A linear abhängige Spalten, und das gilt dann auch für die entsprechenden Untermatrizen von \tilde{A} , so dass k auch die maximale Zahl der unabhängigen Spalten von \tilde{A} ist. Entsprechend ändern Spaltenoperationen auch den Zeilenrang nicht. (Das folgt, indem man zu transponierten Matrizen übergeht.) Nun kann man A (wenn das nicht die Nullmatrix ist) durch Zeilenoperationen auf Zeilen-Stufen-Form bringen und danach durch Spaltenoperationen auf die einfache Form $\left(\begin{array}{c|c} \mathbb{I} & 0 \\ \hline 0 & 0 \end{array}\right)$, wo \mathbb{I} eine Einheitsmatrix ist und 0 für Nullmatrizen von passenden Formaten steht (evtl. auch für leere Matrizen). Für diese einfach gebaute Matrix sind nun aber offenbar Zeilenrang und Spaltenrang gleich, und das gilt folglich auch für die beliebige Ausgangsmatrix A , so dass wir das in 3.2 schon formulierte (aber dort nur plausibel gemachte) Ergebnis bewiesen haben:

$$\text{Zeilenrang} = \text{Spaltenrang}.$$

Diese Zahl wird infolgedessen einfach der **Rang der Matrix** A genannt und $\text{Rang}(A)$ notiert. Das ist also sowohl die maximale Zahl linear unabhängiger Zeilen von A als auch die maximale Zahl linear unabhängiger Spalten.

(2) Eine Konsequenz der Überlegungen aus (1) ist, dass man die Matrix A beliebig mit Zeilenoperationen und Spaltenoperationen umformen darf, wenn man den Rang bestimmen will. Man bringt so die Matrix – unter bestmöglicher Ausnutzung ihrer Nulleinträge – auf eine einfache Form, z.B. eine Zeilen- oder Spalten-Stufen-Form, an der man den (Zeilen- oder Spalten-)Rang unmittelbar ablesen kann. Für praktische Rangberechnungen weniger geeignet, aber theoretisch interessant, ist folgendes **Minoren-Kriterium**:

- *Der Rang einer Matrix ist k , genau wenn sie eine (durch Streichen von Zeilen und Spalten entstehende) $k \times k$ -Untermatrix mit Determinante $\neq 0$ hat, während alle $(k+1) \times (k+1)$ -Untermatrizen Determinante $= 0$ haben.*

Dies heißt “Minoren-Kriterium”, weil man die Determinante einer quadratischen Untermatrix als einen “Minor” der Matrix bezeichnet. Für $k = 0$ ist das Kriterium natürlich so zu verstehen, dass nur die Nullmatrizen Rang Null haben, und für $k+1$ größer als die Zeilen- oder Spaltenzahl der Matrix so, dass die zweite Bedingung entfällt. Der Beweis des Kriteriums ergibt sich daraus, dass im Fall $\text{Rang}(A) = k > 0$ einerseits k linear unabhängige Spalten existieren und daher, wie früher gesehen, eine invertierbare $k \times k$ -Untermatrix dieser k Spalten, während andererseits $k+1$ beliebig ausgewählte Spalten stets linear abhängig sind und daher alle $(k+1) \times (k+1)$ -Untermatrizen dieser $k+1$ Spalten singulär.

(3) Zwei einfache Ungleichungen für den Rang von Matrizen sind:

$$\text{Rang}(A) \leq \max\{\text{Zeilenzahl}(A), \text{Spaltenzahl}(A)\} = \max\{m, n\} \quad \text{für } A \in \mathbb{R}^{m \times n},$$

$$\text{Rang}(BA) \leq \min\{\text{Rang}(B), \text{Rang}(A)\} \quad \text{für } A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{l \times m}.$$

Die erste Ungleichung besagt einfach, dass die Anzahl der linear unabhängigen Zeilen von A nicht größer sein kann als die Zahl aller Zeilen von A und die Zahl der unabhängigen Spalten nicht größer als die Zahl aller Spalten. Wenn in der Ungleichung Gleichheit eintritt, wenn also A den größten Rang hat, der bei dem vorliegenden Format der Matrix überhaupt möglich ist, so nennt man A eine **Matrix von maximalem Rang**. Für quadratische Matrizen ist diese Bedingung gleichbedeutend mit Invertierbarkeit. Die zweite Ungleichung gilt, weil erstens die Spalten von BA Linearkombinationen der Spalten von B sind und zweitens die Zeilen von BA Linearkombinationen der Zeilen von A . Ist B eine invertierbare $m \times m$ -Matrix, so gilt auch $\text{Rang}(A) = \text{Rang}(B^{-1}BA) \leq \text{Rang}(BA)$, und analog $\text{Rang}(B) \leq \text{Rang}(BA)$, wenn A invertierbar ist. Daher kann man hinzufügen:

- *Die Multiplikation einer invertierbaren Matrix von links oder rechts an eine zweite passende Matrix ändert den Rang dieser zweiten Matrix nicht.*

(4) Hat $A \in \mathbb{R}^{m \times n}$ den Rang k , so hat jede Zeilen-Stufen-Form von A genau k Stufen, d.h. k Zeilen, die keine Nullzeilen sind (in der nicht durch rechte Seiten erweiterten Matrix). Als gibt es dann $n-k$ Nicht-Basisvariable und dazu, wie oben gesehen, eine Basis des Lösungsraums L zu $Ax = 0$ aus $n-k$ Vektoren (wenn $k < n$). Der Lösungsraum hat somit die Dimension $\dim L = n-k$, und dieses Ergebnis ist der sog. **Rangsatz**:

$$\text{Rang}(A) + \text{Defekt}(A) = \text{Spaltenzahl}(A)$$

oder

$$\dim K + \dim L = n \quad \text{für } A \in \mathbb{R}^{m \times n},$$

wobei die Dimension des Lösungsraums L der **Defekt** der Matrix A genannt wird und K den Raum der zu A konsistenten rechten Seiten bezeichnet. Anschaulich gesprochen ist der Raum der konsistenten rechten Seiten um so viele Dimensionen kleiner als der Raum \mathbb{R}^n , wieviel bei der Multiplikation von A an Spaltenvektoren aus \mathbb{R}^n “anulliert” wird.

(5) Für lineare Abbildungen $T(\mathbf{x}) = A\mathbf{x}$ von \mathbb{R}^n nach \mathbb{R}^m ist $\text{Bild}(T) = K$ der Raum der zur Matrix $A \in \mathbb{R}^{m \times n}$ konsistenten rechten Seiten, also $\text{Rang}(A) = \dim \text{Bild}(T)$. Daher definiert man sinnvollerweise für lineare Abbildungen $T: V \rightarrow W$ zwischen beliebigen Vektorräumen den **Rang der linearen Abbildung** durch

$$\text{Rang}(T) := \dim \text{Bild}(T)$$

(sofern diese Dimension endlich ist). Dem Lösungsraum L zu $A\mathbf{x} = 0$ entspricht in der allgemeineren Situation $\text{Kern}(T) = \{v \in V : T(v) = 0\}$ und dessen Dimension wird wiederum als **Defekt der linearen Abbildung** bezeichnet. Der Rangsatz gilt nun auch in dieser allgemeineren Situation, sofern das Bild und der Kern endlichdimensional sind:

$$\text{Rang}(T) + \text{Defekt}(T) = \dim \text{Def}(T),$$

bzw. ausführlicher

$$\dim \text{Bild}(T) + \dim \text{Kern}(T) = \dim V \quad \text{für lineare } T: V \rightarrow W.$$

wobei $\text{Def}(T)$ den Definitionsbereich der linearen Abbildung bezeichnet, also den Vektorraum V , in dem der Kern gebildet wird. Die Dimension des Bildes ist somit exakt um so viele Dimensionen kleiner wie die Dimension des Anullierungsraums von T im Definitionsbereich V .

Zum Beweis wählen wir eine Basis w_1, \dots, w_k im Unterraum $\text{Bild}(T)$ von W , wobei k dessen Dimension ist, also der Rang von T . Zu jedem w_j wählen wir ein T -Urbild v_j in V und ergänzen diese Vektoren durch eine Basis v_{k+1}, \dots, v_n in $\text{Kern}(T)$. (Wir können annehmen, dass T nicht die Nullabbildung ist. Im Fall $\text{Kern}(T) = \{0\}$ ist die Ergänzung nicht nötig.) Dann sind v_1, \dots, v_n linear unabhängig; denn aus $\sum_{j=1}^n r_j v_j = 0$ folgt mit Anwendung von T unter Ausnutzung von $T(v_j) = w_j$ für $j = 1 \dots k$ und $T(v_j) = 0$ für $j > k$ zunächst $r_1 = \dots = r_k = 0$ wegen der Basiseigenschaft von w_1, \dots, w_k und sodann auch $r_{k+1} = \dots = r_n = 0$ wegen der Basiseigenschaft von v_{k+1}, \dots, v_n . Außerdem erzeugen v_1, \dots, v_n auch V ; denn für $v \in V$ und Koeffizienten s_j mit $T(v) = \sum_{j=1}^k s_j w_j$ liegt $v - \sum_{j=1}^k s_j v_j$ im Kern von T , ist also eine Linearkombination von v_{k+1}, \dots, v_n . Somit ist v_1, \dots, v_n eine Basis in V , und die Behauptung läuft einfach auf die Gleichung $k + (n-k) = n$ hinaus.

Der Rangsatz in der hier bewiesenen Form ist allgemeiner als (5), weil er sich z.B. auch auf Unterräume U von V anwenden lässt. Der Kern von T im Ausgangsraum U ist dann der Durchschnitt $\{v \in U : T(v) = 0\} = U \cap T^{-1}\{0\}$, und der Rangsatz lautet $\dim T(U) + \dim(U \cap T^{-1}\{0\}) = \dim U$. Für ein lineares Gleichungssystem $A\mathbf{x} = 0$ mit $A \in \mathbb{R}^{m \times n}$ und Lösungsmenge L besagt das $\dim AU + \dim L \cap U = \dim U$ für jeden Unterraum U von \mathbb{R}^n . Eine Konsequenz, die auch einfach aus der Tatsache folgt, dass T jede Basis in U auf ein Erzeugendensystem von $T(U)$ abbildet, ist:

- *Lineare Abbildungen T können Dimensionen nicht vergrößern, d.h. für Unterräume U des Definitionsbereichs gilt stets $\dim T(U) \leq \dim U$.*

Der Rangsatz sagt aber genau, wann tatsächlich eine Dimensionsverkleinerung stattfindet, nämlich wenn $U \cap \text{Kern}(T) \neq \{0\}$ ist, und um wieviel die Dimension dann verkleinert wird, nämlich um die Dimension von $U \cap \text{Kern}(T) \neq \{0\}$.

(6) Sind in den Vektorräumen V bzw. W Basen v_1, \dots, v_n bzw. w_1, \dots, w_m fixiert, so kann man jeder lineare Abbildung $T: V \rightarrow W$ eine $m \times n$ -Matrix $A = (a_{ij})$ zuordnen, indem man die Bilder der Basisvektoren v_j durch die Basisvektoren w_i ausdrückt, $T(v_j) = \sum_{i=1}^m a_{ij} w_i$ (die Reihenfolge der Indizes bei a_{ij} ist zu beachten). Diese Matrix A heißt die **Matrixdarstellung der linearen Abbildung bzgl. der gegebenen Basen**. Sie hat als Spalten die Koeffizientenvektoren der Bilder $T(v_j)$ bzgl. der Basis in W . Im Fall der Zahlenräume \mathbb{R}^n und \mathbb{R}^m mit kanonischen Basen ist das gerade die Matrix mit $T(\mathbf{x}) = A\mathbf{x}$ für alle (Spalten-)Vektoren $\mathbf{x} \in \mathbb{R}^n$. Umgekehrt kann man mit einer gegebenen Matrix $A \in \mathbb{R}^{m \times n}$ eine lineare Abbildung T mit Matrixdarstellung A definieren, indem man zunächst für Basisvektoren $T(v_j) := \sum_{i=1}^m a_{ij} w_i$ setzt und für allgemeine $v = \sum_{j=1}^n x_j v_j$ dann $T(v) := \sum_{j=1}^n x_j T(v_j) = \sum_{i=1}^m (\sum_{j=1}^n a_{ij} x_j) w_i$ definiert. (Das ist möglich, weil die Koeffizienten x_j eindeutig bestimmt sind.) Die Argumentation zeigt insgesamt Folgendes:

- Eine lineare Abbildung $T: V \rightarrow W$ ist eindeutig bestimmt durch ihre Werte auf den Vektoren einer gegebenen Basis v_1, \dots, v_n von V ; diese Werte können beliebig in W vorgegeben werden, um eine lineare Abbildung von V nach W zu definieren.
- Ordnet man jeder linearen Abbildung $T: V \rightarrow W$ die Matrix $A \in \mathbb{R}^{m \times n}$ zu, deren Spalten die Koeffizientenvektoren der Bilder $T(v_j)$ bzgl. einer gegebenen Basis w_1, \dots, w_m von W sind, so erhält man einen Isomorphismus zwischen dem Raum der linearen Abbildungen von V nach W und dem Raum der $m \times n$ -Matrizen.
- Dabei ist $A\mathbf{x} \in \mathbb{R}^m$ der Koeffizientenvektor von $T(v) \in W$ bzgl. w_1, \dots, w_m , wenn \mathbf{x} der Koeffizientenvektor von $v \in V$ bzgl. v_1, \dots, v_n ist.

(7) Das Bild eines k -dimensionalen Unterraums von W , unter dem Isomorphismus, der den Elementen von W ihren Koeffizientenvektor in \mathbb{R}^m bzgl. der Basis w_1, \dots, w_m zuordnet, ist eine k -dimensionaler Unterraum von \mathbb{R}^m . Das ist klar, weil injektive lineare Abbildungen lineare Unabhängigkeit erhalten. Andererseits ist dieses Bild nach der letzten Bemerkung gleich der Menge aller $A\mathbf{x}$ zu $\mathbf{x} \in \mathbb{R}^n$, also gleich dem Raum der zu A konsistenten rechten Seiten, und dessen Dimension ist der Rang von A . Also gilt:

- Der Rang einer linearen Abbildung T ist gleich dem Rang der Matrix, die T bzgl. gegebener Basen im Ausgangsraum und im Zielraum darstellt.

Eine Folgerung ist, dass alle Matrizen, die dieselbe lineare Abbildung T bzgl. verschiedener Basen darstellen, denselben Rang haben.

(8) Mit der letzten Bemerkung in (6) kann man genau sehen, wie sich die Matrixdarstellung von T ändert, wenn man zu anderen Basen \tilde{v}_j in V und \tilde{w}_i in W übergeht. Für die Matrix \tilde{A} bzgl. der neuen Basen gilt $\tilde{A}C^{-1}\mathbf{x} = D^{-1}A\mathbf{x}$, wenn C bzw. D die Übergangsmatrizen zwischen den Basen in V bzw. in W sind, weil ja $\tilde{\mathbf{x}} = C^{-1}\mathbf{x}$ die Änderung der Koeffizientenvektoren von Elementen in V beschreibt und $\tilde{\mathbf{y}} = D^{-1}\mathbf{y}$ die Änderung der Koeffizientenvektoren von Elementen in W . Somit ist

$$\tilde{A} = D^{-1}AC \quad \text{die geänderte Matrixdarstellung nach dem Basiswechsel.}$$

Da die Matrizen C und D invertierbar sind, folgt insbesondere, dass A und \tilde{A} denselben Rang haben.

(9) Speziell im Fall $V = W$ arbeitet man normalerweise mit derselben Basis $v_j = w_j$ in V als Ausgangsraum und als Zielraum. Dann lautet die durch Wechsel zu einer Basis \tilde{v}_j geänderte Matrixdarstellung $\tilde{A} = C^{-1}AC$, wobei C die Übergangsmatrix zur neuen Basis ist. \tilde{A} ist also eine zur Matrix $A \in \mathbb{R}^{n \times n}$ ähnliche Matrix. Hieran erkennt man die eigentliche Bedeutung des Ähnlichkeitsbegriffs für quadratische Matrizen: *Ähnliche Matrizen beschreiben dieselbe lineare Abbildung*, nur eben bzgl. verschiedener Basen. Wegen des Multiplikationssatzes für die Determinante haben ähnliche Matrizen dieselbe Determinante. Dies ermöglicht, die **Determinante einer linearen Selbstabbildung** $T: V \rightarrow V$ eines endlichdimensionalen Vektorraums in sich zu definieren durch $\det(T) := \det(A)$ für die Matrixdarstellung A von T bzgl. einer beliebigen Basis von V ; diese Zahl hängt dann nur von T ab und nicht von der Wahl der Basis.

Man kann versuchen, sich einen Überblick über die Struktur einer linearen Selbstabbildung T von V zu verschaffen, indem man eine Basis bestimmt, bzgl. der T eine möglichst einfache Matrixdarstellung hat. Das läuft darauf hinaus, dass man in einer Klasse von zueinander ähnlichen Matrizen eine Matrix von möglichst einfacher Form sucht. Dies ist ein Problem, das in der Mathematik mit sehr befriedigendem Ergebnis behandelt werden kann, allerdings nur über dem Zahlbereich \mathbb{C} der komplexen Zahlen. Man kann dann stets erreichen, dass die gesuchten "einfachen" Matrizen *Jordansche Normalform* haben, das sind spezielle Dreiecksmatrizen, in denen außer auf der Diagonalen und evtl. der ersten Oberdiagonalen nur Nulleinträge stehen. (Über dem Zahlbereich \mathbb{R} ist die Antwort komplizierter; es lässt sich dann nicht in jedem Fall eine Basis finden, bzgl. der die Matrixdarstellung von T Dreiecksgestalt hat. Das kann man schon daran sehen, dass bei Darstellung durch eine Dreiecksmatrix mindestens ein Basisvektor v ein reeller Eigenvektor wäre, d.h. $T(v) = \lambda v$, gälte mit einem Eigenwert $\lambda \in \mathbb{R}$. Aber nicht jede lineare Abbildung eines reellen Vektorraums in sich hat einen reellen Eigenwert, z.B. nicht eine Drehung um 0 in \mathbb{R}^2 mit einem Drehwinkel echt zwischen 0 Grad und 180 Grad.)

Statt weiter auf diese schon recht fortgeschrittene mathematische Theorie einzugehen, begnügen wir uns mit einem Beispiel, das zeigt, wie ein Basiswechsel die einfache Struktur einer linearen Abbildung aufdecken kann (oder auch verschleiern, wenn man ihn in umgekehrter Richtung vornimmt). Dazu betrachten wir die lineare Abbildung von \mathbb{R}^2 in sich, die gegeben ist durch

$$T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5x_1 + 3x_2 \\ 3x_1 + 5x_2 \end{pmatrix} \quad \text{mit Matrixdarstellung } A = \begin{pmatrix} 5 & 3 \\ 3 & 5 \end{pmatrix}$$

bezüglich der kanonischen Basis von \mathbb{R}^2 . Gehen wir nun von der kanonischen Basis zur Basis $(1, 1), (-1, 1)$ über, so ist die Übergangsmatrix $C = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ mit der Inversen $C^{-1} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$. Die Matrixdarstellung von T bzgl. der neuen Basis ist dann

$$C^{-1}AC = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 5 & 3 \\ 3 & 5 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 8 & 0 \\ 0 & 2 \end{pmatrix}.$$

Dies bedeutet, dass die Abbildung bzgl. der neuen Basis (die abgesehen von einer Maßstabänderung einer Drehung der Koordinatenachsen um 45 Grad entspricht) einfach eine Achsenstreckung ist, wobei die erste Achse um den Faktor 8, die zweite um den Faktor 2 gestreckt wird. Diese einfache Struktur der linearen Abbildung war an der ursprünglichen Matrixdarstellung bzgl. der kanonischen Basis nicht unmittelbar zu erkennen.

Wir erklären am Ende dieses Abschnitts, wie man systematisch eine Basis bestimmen kann, bzgl. der die obige lineare Abbildung durch eine Diagonalmatrix dargestellt wird. Die Symmetrie der Matrix A ist dabei wesentlich. ■

Die Berechnung des Koeffizientenvektors \mathbf{x} eines Vektors v in einem Unterraum U von \mathbb{R}^n bezüglich einer allgemeinen Basis v_1, \dots, v_k von U erfordert einigen Aufwand: Man muss ein System $A\mathbf{x} = v$ von n linearen Gleichungen für k Unbekannte lösen, wobei A die Matrix mit den Spalten v_j ist. Es gibt aber eine Klasse von speziellen Basen, bei denen diese Aufgabe viel einfacher zu erledigen ist; man kann die Koeffizienten eines Vektors bzgl. einer solchen Basis in vielen Fällen direkt ablesen und jedenfalls sehr einfach berechnen. Diese für viele Rechnungen sehr praktische Basen sind der letzte Gegenstand dieses Abschnitts.

DEFINITION: Eine endliche Folge von Vektoren u_1, \dots, u_k in \mathbb{R}^n heißt **Orthonormalsystem** in \mathbb{R}^n oder **Orthonormalbasis** des davon aufgespannten Unterraums, wenn die Vektoren paarweise senkrecht zueinander sind und die Länge 1 haben, wenn also gilt:

$$u_i \cdot u_j = d_{ij} := \begin{cases} 0 & \text{für } i \neq j \\ 1 & \text{für } i = j \end{cases} \quad (1 \leq i, j \leq k). \quad \blacksquare$$

Hierbei ist $\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + \dots + x_n y_n$ das in 3.3 eingeführte (Euklidische) **Skalarprodukt** von Vektoren $\mathbf{x} = (x_1, \dots, x_n)$ und $\mathbf{y} = (y_1, \dots, y_n)$ in \mathbb{R}^n . Die Bedingung $\mathbf{x} \cdot \mathbf{y} = 0$ bedeutet geometrisch, dass die Vektoren (bzw. die dadurch definierten Ursprungsgeraden) **orthogonal** (senkrecht) zueinander sind, wofür man auch $\mathbf{x} \perp \mathbf{y}$ schreibt. Das haben wir ebenfalls schon in 3.3 erklärt und ebenso, dass $\sqrt{\mathbf{x} \cdot \mathbf{x}}$, also die Wurzel aus der Quadratsumme der Komponenten von \mathbf{x} , als Länge des Vektors \mathbf{x} zu interpretieren ist und auch $|\mathbf{x}|$ notiert und (Euklidische) **Norm** des Vektors genannt wird. Ein Vektor \mathbf{x} mit Norm 1 heißt **Einheitsvektor** oder **normiert**. Ist $\mathbf{x} \neq 0$, so kann man den Vektor **normieren**, d.h. auf die Länge 1 bringen, indem man ihn mit dem Reziproken seiner Norm multipliziert; $\frac{1}{|\mathbf{x}|} \mathbf{x}$ ist also ein Einheitsvektor. Die kanonischen Basisvektoren e_j des \mathbb{R}^n sind Beispiele von Einheitsvektoren, aber auch $-e_j$ und viele weitere wie $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ in \mathbb{R}^2 , $(\frac{2}{3}, \frac{1}{3}, \frac{2}{3})$ in \mathbb{R}^3 , ... und eben alle Vektoren in \mathbb{R}^n , die durch Pfeile im Ursprung repräsentiert sind, deren Spitze auf dem Rand der Kugel vom Radius 1 um den Ursprung liegt. Die Bedingungen der Definition besagen, dass erstens $u_i \cdot u_j = 0$ ist für $i \neq j$, also verschiedene Vektoren des Systems orthogonal zueinander, und zweitens $u_i \cdot u_i = 1$, also alle Vektoren des Systems normiert. Das erklärt die Wortbildung "orthonormal".

Da wir sie im Folgenden benötigen, wiederholen wir die grundlegenden Eigenschaften von Skalarprodukt und zugehöriger Norm. Unter einem **Skalarprodukt** oder **innerem Produkt** auf einem Vektorraum V versteht man eine Vorschrift, die je zwei Vektoren $v, w \in V$ einen Skalar, also eine reelle Zahl $v \cdot w$ zuordnet mit folgenden Eigenschaften:

$$(\text{Positivität}) \quad v \cdot v > 0 \text{ wenn } v \neq 0 \quad (\text{und } 0 \cdot 0 = 0),$$

$$(\text{Symmetrie}) \quad v \cdot w = w \cdot v,$$

$$(\text{Bilinearität}) \quad (rv) \cdot w = r(v \cdot w) \text{ für } r \in \mathbb{R}, \quad (v + \tilde{v}) \cdot w = v \cdot w + \tilde{v} \cdot w.$$

Die letzten Gleichungen gelten wegen der Symmetrie natürlich analog für den zweiten Faktor, d.h. $v \cdot w$ ist eine lineare Funktion von v , wenn man w fixiert, und von w , wenn man v festhält. (Das ist mit "Bilinearität" gemeint.) Die mit dem Skalarprodukt **assoziierte Norm** ordnet jedem $v \in V$ die nichtnegative Zahl $|v| := \sqrt{v \cdot v}$ zu und hat folgende Grundeigenschaften:

$$(\text{Positivität}) \quad |v| > 0 \text{ wenn } v \neq 0 \quad (\text{und } |0| = 0),$$

$$(\text{absolute Homogenität}) \quad |rv| = |r||v| \text{ für } r \in \mathbb{R},$$

$$(\text{Dreiecksungleichung}) \quad |v + w| \leq |v| + |w|.$$

Diese Normeigenschaften folgen aus denen Grundeigenschaften des Skalarprodukts. Die erste ist klar. Die zweite besagt, dass man einen skalaren Faktor aus der Norm mit seinem Betrag herausziehen kann, und ist ebenfalls leicht zu sehen. Die Dreiecksungleichung läuft, wenn man quadriert und mit Bilinearität und Symmetrie $|v+w|^2 = (v+w) \cdot (v+w) = v \cdot v + w \cdot w + v \cdot w + w \cdot v = |v|^2 + 2v \cdot w + |w|^2$ berechnet, auf die sog. **Cauchy-Schwarz-Ungleichung** hinaus:

$$|v \cdot w| \leq |v||w|.$$

Der Betrag des Skalarproduktes ist also höchstens so groß wie das Produkt der Normen seiner Faktoren. Diese Ungleichung ergibt sich aus der Beobachtung, dass die quadratische Funktion $|v+tw|^2 = |v|^2 + 2tv \cdot w + t^2|w|^2$ der Variablen $t \in \mathbb{R}$ nur nichtnegative Werte hat. Daher kann die Diskriminante nicht positiv sein, und das heißt $(v \cdot w)^2 - |v|^2|w|^2 \leq 0$. Die geometrische Bedeutung der Dreiecksungleichung ist, dass die Norm der Differenzvektoren $|v-w|$ eine sinnvolle Definition für den **Abstand** zwischen den Punkten v, w in V ist; denn es gilt dann $|u-w| \leq |u-v| + |v-w|$, d.h. der Abstand von u nach w ist höchstens so groß wie die Summe der Abstände, die sich beim "Umweg" über v ergeben.

Für das oben angegebene Euklidische Skalarprodukt auf \mathbb{R}^n sind die Grundeigenschaften leicht nachprüfbar, und der Euklidische Abstand $|\mathbf{x} - \mathbf{y}|$ zwischen Vektoren in \mathbb{R}^n , also die Wurzel aus der Quadratsumme der Komponentendifferenzen der Vektoren, ist der übliche geometrische Abstand in der Zeichenebene \mathbb{R}^2 bzw. im Anschauungsraum \mathbb{R}^3 , der durch den Satz des Pythagoras motiviert wird. Die für das Euklidische Skalarprodukt oben eingeführte Notationen und Sprechweisen verwendet man ganz analog für allgemeine Skalarprodukte, also bedeutet Orthogonalität $v \perp w$ von Vektoren, dass $v \cdot w = 0$ ist und Vektoren v mit $|v| = 1$ heißen Einheitsvektoren oder normiert. Die Definition der Orthonormalsysteme ist damit sinnvoll für jeden Vektorraum V , der mit einem Skalarprodukt versehen ist, und diese Situation setzen wir im Folgenden stets voraus. Das wichtigste Beispiel, das man sich immer vorstellen sollte, ist dabei der Zahlenraum \mathbb{R}^n mit dem kanonischen (Euklidischen) Skalarprodukt.

Wir kommen nun zu den angekündigten besonders schönen Eigenschaften der Orthonormalbasen. Die erste Aussage des folgenden Satzes rechtfertigt auch, dass wir überhaupt von "Basen" sprechen.

SATZ: (i) Jedes Orthonormalsystem u_1, \dots, u_k in V ist linear unabhängig, also Basis des davon aufgespannten Unterraums U ;

(ii) Die Koeffizienten eines Vektors $v \in U$ bzgl. der Orthonormalbasis sind die Skalarprodukte von v mit den Basisvektoren, also $v = \sum_{j=1}^k (u_j \cdot v) u_j$;

(iii) Skalarprodukt und Norm von Vektoren $v, w \in U$ lassen sich berechnen als Euklidisches Skalarprodukt bzw. Euklidische Norm ihrer Koeffizientenvektoren in \mathbb{R}^k , also

$$v \cdot w = \sum_{j=1}^k (u_j \cdot v)(u_j \cdot w) \quad \text{und} \quad |v|^2 = \sum_{j=1}^k (u_j \cdot v)^2;$$

(iv) Jeder endlichdimensionale Unterraum $U \neq \{0\}$ von V hat eine Orthonormalbasis.

Die Aussage (ii) bedeutet, dass man für die Berechnung der Koeffizienten eines Vektors bzgl. der Basis kein lineares Gleichungssystem mehr lösen muss, sondern nur Skalarprodukte auszurechnen braucht – das ist natürlich viel einfacher.

Zum *Beweis* von (i) und (ii) bilden wir das Skalarprodukt von u_i mit einer Linearkombination $v = \sum_{j=1}^k r_j u_j$. Wegen der Bilinearität gilt $u_i \cdot v = \sum_{j=1}^k r_j u_i \cdot u_j$, und weil $u_i \cdot u_j = 0$ ist für $j \neq i$ sowie $u_i \cdot u_i = 1$, folgt $u_i \cdot v = r_i$. Das zeigt, dass die Koeffizienten der Linearkombination eindeutig bestimmt sind, also die behauptete lineare Unabhängigkeit der u_j , und gleichzeitig die Behauptung (ii). Mit der Bilinearität folgt aus (ii) $v \cdot w = \sum_{j=1}^k (u_j \cdot v) u_j \cdot w = \sum_{j=1}^k (u_j \cdot v) \sum_{i=1}^k (u_i \cdot w) u_j \cdot u_i$. Berücksichtigt man wieder $u_j \cdot u_i = 0$ für $j \neq i$ und $u_j \cdot u_j = 1$, so erhält man die erste Gleichung von (iii); die zweite ist der Spezialfall $v = w$ davon.

Eine wesentliche Aussage, die z.B. erst die Anwendung des Satzes auf beliebige Unterräume von \mathbb{R}^n möglich macht, ist (iv). Dies beweist man mit einem nützlichen Verfahren, das ausgehend von einer beliebigen Basis v_1, \dots, v_k von U rekursiv eine Orthonormalbasis u_1, \dots, u_k berechnet und für praktische Anwendungen wichtig ist. Deshalb beschreiben wir nun dieses sog. **Gram-Schmidt-Orthonormalisierungsverfahren**:

$$u_1 = \frac{v_1}{|v_1|}, \quad u_2 = \frac{v_2 - (u_1 \cdot v_2)u_1}{\text{Norm des Zählers}}, \quad \dots, \quad u_j = \frac{v_j - (u_1 \cdot v_j)u_1 - \dots - (u_{j-1} \cdot v_j)u_{j-1}}{\text{Norm des Zählers}}, \quad \dots$$

Man erhält also u_1 durch Normierung des ersten Basisvektors v_1 . Vor dem j -ten Schritt des Verfahrens hat man bereits die Vektoren u_1, \dots, u_{j-1} des Orthonormalsystems berechnet, die denselben Unterraum aufspannen wie v_1, \dots, v_{j-1} . Beim j -ten Schritt subtrahiert man zunächst von v_j die Linearkombination $\sum_{i=1}^{j-1} (u_i \cdot v_j) u_i$ der bereits berechneten Vektoren, also den Vektor aus $\mathbb{R}u_1 + \dots + \mathbb{R}u_{j-1}$, der dieselben Skalarprodukte mit u_1, \dots, u_{j-1} hat wie v_j . Dies bedeutet, dass der Differenzvektor orthogonal zu u_1, \dots, u_{j-1} ist (er ist gewissermaßen die Komponente von v_j senkrecht zu u_1, \dots, u_{j-1}). Er ist auch nicht der Nullvektor; denn sonst wäre v_j ein Element von $\mathbb{R}u_1 + \dots + \mathbb{R}u_{j-1} = \mathbb{R}v_1 + \dots + \mathbb{R}v_{j-1}$, was der linearen Unabhängigkeit von v_1, \dots, v_j widerspräche. Daher kann man den Differenzvektor noch normieren durch Multiplikation mit dem Reziproken seiner Norm und erhält so den zu u_1, \dots, u_{j-1} orthogonalen Einheitsvektor u_j , der zusammen mit u_1, \dots, u_{j-1} denselben Unterraum aufspannt wie v_1, \dots, v_j . Das Verfahren kann bis $j = k$ fortgesetzt werden und liefert die gesuchte Orthonormalbasis in $U = \mathbb{R}v_1 + \dots + \mathbb{R}v_k$.

BEISPIELE: (1) Die *kanonische Basis* e_1, \dots, e_n im Zahlenraum \mathbb{R}^n mit dem kanonischen (Euklidischen) Skalarprodukt ist das Paradebeispiel einer Orthonormalbasis. Das Skalarprodukt verschiedener kanonischer Basisvektoren ist offensichtlich 0 und die Euklidische Norm jedes e_i ist offensichtlich 1. Wenn man einige kanonischen Basisvektoren e_i durch ihre Negativen ersetzt, so hat man natürlich ebenfalls eine Orthonormalbasis. Natürlich gibt es noch sehr viel mehr Orthonormalbasen in \mathbb{R}^n . Anschaulich erhält man sie alle durch (verallgemeinerte) Drehung der in (1) beschriebenen Basen. Eine Diskussion dieses Sachverhalts geben wir etwas später.

(2) Gesucht ist eine Orthonormalbasis in der Ursprungsebene des \mathbb{R}^3 , die beschrieben wird durch die Gleichung

$$x_1 - 2x_2 + 3x_3 = 0.$$

Eine Basis des Lösungsraums erhalten wir, indem wir z.B. x_2, x_3 als Basisvariable wählen. Das gibt die Basis $(2, 1, 0)$ und $(-3, 0, 1)$ der Ebene. Darauf wenden wir das Orthonormalisierungsverfahren an:

$$u_1 = \frac{1}{\sqrt{2^2+1^2+0^2}} \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}.$$

Dieser Einheitsvektor ist der erste der gesuchten Orthonormalbasis. Die Komponente von $(-3, 0, 1)$ senkrecht dazu ist

$$\tilde{u}_2 = \begin{pmatrix} -3 \\ 0 \\ 1 \end{pmatrix} - \left[u_1 \cdot \begin{pmatrix} -3 \\ 0 \\ 1 \end{pmatrix} \right] u_1 = \begin{pmatrix} -3 \\ 0 \\ 1 \end{pmatrix} - \frac{-6}{5} \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} = \frac{1}{5} \begin{pmatrix} -3 \\ 6 \\ 5 \end{pmatrix}.$$

Dieser Vektor muss jetzt noch normiert, also mit dem Reziproken seiner Norm multipliziert werden. Das ergibt

$$u_2 = \frac{\tilde{u}_2}{|\tilde{u}_2|} = \frac{1}{\sqrt{(-3)^2 + 6^2 + 5^2}} \begin{pmatrix} -3 \\ 6 \\ 5 \end{pmatrix} = \frac{1}{\sqrt{70}} \begin{pmatrix} -3 \\ 6 \\ 5 \end{pmatrix},$$

und u_1, u_2 sind eine Orthonormalbasis der Ursprungsebene. Soll diese durch u_3 ergänzt werden zu einer Orthonormalbasis des \mathbb{R}^3 , so können wir einen weiteren Schritt des Orthonormalisierungsverfahrens anwenden auf irgendeinen von u_1, u_2 linear unabhängigen Vektor v . Man bildet also $\tilde{u}_3 = v - (u_1 \cdot v)u_1 - (u_2 \cdot v)u_2$ und normiert dann diesen Vektor noch, um u_3 zu erhalten. Einfacher ist folgende Beobachtung: Die homogene lineare Ausgangsgleichung, die u_1, u_2 erfüllen, bedeutet, dass der Koeffizientenvektor dieser Gleichung $(1, -2, 3)$ (als Spaltenvektor aufgefasst) orthogonal zu u_1, u_2 ist. Wir brauchen also nur noch diesen Koeffizientenvektor zu normieren und erhalten

$$u_3 = \frac{1}{\sqrt{1^2 + (-2)^2 + 3^2}} \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix} = \frac{1}{\sqrt{14}} \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}. \quad \blacksquare$$

Die am Ende des letzten Beispiels gemachte Beobachtung, dass der Koeffizientenvektor einer homogenen linearen Gleichung orthogonal zu deren Lösungsvektoren ist, können wir systematischer beschreiben und anwenden mit Hilfe des folgenden Satzes, wobei wieder ein allgemeiner Vektorraum V mit Skalarprodukt zugrund liegt:

SATZ und DEFINITION: Ist U ein endlichdimensionaler Unterraum von V , so ist $U^\perp := \{v \in V : v \cdot u = 0 \text{ für alle } u \in U\}$ wieder ein Unterraum von V , genannt der zu U **orthogonale Unterraum** in V oder das **orthogonale Komplement** von U in V . Jeder Vektor $v \in V$ besitzt eine eindeutige **orthogonale Zerlegung**

$$v = u + u^\perp \quad \text{mit } u \in U \quad \text{und } u^\perp \in U^\perp.$$

Hat V endliche Dimension, so gilt

$$\dim U + \dim U^\perp = \dim V \quad \text{und} \quad (U^\perp)^\perp = U.$$

Beweis: Mit der Bilinearität des Skalarprodukts sieht man sofort, dass mit v, \tilde{v} auch rv und $v + \tilde{v}$ orthogonal zu allen $u \in U$ sind, also ist U^\perp ein Unterraum von V . Um die orthogonale Zerlegung zu beweisen, wählen wir eine Orthonormalbasis u_1, \dots, u_k in U . (Wenn $U = \{0\}$ ist, so gilt $U^\perp = V$ und die Behauptungen sind klar.) Ist $v = u + u^\perp$ zerlegt mit $u \in U$ und $u^\perp \in U^\perp$, so zeigt Skalarproduktbildung mit u_j , dass $v \cdot u_j = u \cdot u_j$ gelten muss für $j = 1 \dots k$. Da die Koeffizienten von u bzgl. der Orthonormalbasis die Skalarprodukte mit den Basisvektoren sind, folgt $u = \sum_{j=1}^k (u_j \cdot v) u_j$. Das demonstriert, dass u eindeutig durch v bestimmt ist und $u^\perp = v - u$ dann natürlich auch. Um die Existenz der Zerlegung zu zeigen, definieren wir für gegebenes $v \in V$ nun $u := \sum_{j=1}^k (u_j \cdot v) u_j$. Dann hat u mit allen u_j dasselbe Skalarprodukt wie v , und wegen der Bilinearität des

Skalarprodukts bedeutet dies, dass $v-u$ orthogonal ist zu allen Linearkombinationen der u_j , also zu allen Vektoren aus U . Mit $u^\perp := v-u \in U^\perp$ haben wir also die gesuchte Zerlegung $v = u + u^\perp$. Im Fall $\dim V < \infty$ können wir u_1, \dots, u_k durch eine Orthonormalbasis u_{k+1}, \dots, u_n von U^\perp ergänzen. (Wenn $U^\perp = \{0\}$ ist, so gilt aufgrund der bewiesenen orthogonalen Zerlegung $U = V$ und die Behauptungen sind klar.) Dann ist u_1, \dots, u_n Orthonormalbasis von V , da diese Vektoren wegen der bewiesenen Zerlegung V erzeugen und als Vektoren eines Orthonormalsystems auch linear unabhängig sind. Daraus folgt nun die Dimensionsgleichung und auch die letzte Aussage, weil eine Linearkombination $\sum_{j=1}^n s_j u_j$ genau dann zu u_{k+1}, \dots, u_n orthogonal ist, wenn $s_{k+1} = \dots = s_n = 0$ gilt, wenn also die Linearkombination in U liegt.

Allgemein heißt ein Unterraum \tilde{U} ein **Komplement** des Unterraums U eines Vektorraums V , wenn jeder Vektor $v \in V$ eindeutig zerlegbar ist $v = u + \tilde{u}$ mit $u \in U$ und $\tilde{u} \in \tilde{U}$. Man schreibt symbolisch $V = U \oplus \tilde{U}$ für diesen Sachverhalt und nennt u bzw. \tilde{u} die **Komponenten** von v **bzgl. der Zerlegung**. Im endlichdimensionalen Fall ist äquivalent, dass sich Basen von U und \tilde{U} zu Basen von V ergänzen. Eine besondere geometrische Eigenschaft der orthogonalen Zerlegung $v = u + u^\perp$ ist, dass hier u der Punkt in U mit kleinstem Abstand zu v ist, also der Fußpunkt des Lotes von v auf den Unterraum U . Das sieht man daran, dass das Quadrat des Abstands von v zu $u' \in U$ gleich $|v - u'|^2 = |(u - u') + u^\perp|^2 = |u - u'|^2 + |u^\perp|^2$ ist (die letzte Gleichung gilt wegen $(u - u') \cdot u^\perp = 0$ und ist nichts anderes als der Satz des Pythagoras) und am kleinsten wird, genau wenn $u' = u$ ist. Analog sieht man, dass u^\perp der zu v nächste Punkt in U^\perp ist. Man nennt daher u bzw. u^\perp die **Orthogonalprojektionen** von v auf U bzw. auf U^\perp . Der Beweis oben liefert folgende Formel für die Orthogonalprojektion von $v \in V$ auf U in Termen einer Orthonormalbasis u_1, \dots, u_k von U :

$$\sum_{j=1}^k (u_j \cdot v) u_j \quad \text{ist die Orthogonalprojektion von } v \text{ auf } U = \mathbb{R}u_1 + \dots + \mathbb{R}u_k.$$

BEISPIELE (Anwendungen auf lineare Gleichungssysteme):

(1) Ist $Ax = 0$ ein homogenes lineares Gleichungssystem und sind $\mathbf{a}_i = (a_{i1}, \dots, a_{in})$ die Zeilenvektoren der Koeffizientenmatrix, so kann man die i -te Gleichung schreiben $\mathbf{a}_i \cdot \mathbf{x} = a_{i1}x_1 + \dots + a_{in}x_n = 0$. Lösungsvektoren sind also genau die $\mathbf{x} \in \mathbb{R}^n$, die orthogonal zu allen Koeffizientenzeilen sind. Mit anderen Worten:

- *Der Lösungsraum eines homogenen linearen Gleichungssystems für n Unbekannte ist das orthogonale Komplement des Unterraums von \mathbb{R}^n , der von den Zeilenvektoren der Koeffizientenmatrix aufgespannt wird.*

(Wenn man die Vektoren des \mathbb{R}^n als Spalten schreibt, so muss man diese Zeilenvektoren natürlich transponieren, also als Spaltenvektoren umschreiben.)

Das eröffnet folgende prinzipielle Möglichkeit, eine Orthonormalbasis des Lösungsraums zu berechnen: Man wählt eine maximale Anzahl k von linear unabhängigen Zeilenvektoren der Koeffizientenmatrix aus und ergänzt sie zu einer Basis von \mathbb{R}^n . Auf diese Basis wendet man dann das Orthogonalisierungsverfahren an und lässt die ersten k Vektoren der erhaltenen Orthonormalbasis weg, die ja denselben Raum aufspannen wie die Zeilenvektoren der Koeffizientenmatrix. Die Ergänzung kann man immer so vornehmen, dass man kanonische Basisvektoren hinzunimmt, das ist rechnerisch günstig.

(2) Eine Ursprungsgerade G in \mathbb{R}^2 beschrieben durch eine Gleichung $ax_1 + bx_2 = 0$ (mit $a \neq 0$ oder $b \neq 0$) ist das orthogonale Komplement $G = (\mathbb{R} \begin{pmatrix} a \\ b \end{pmatrix})^\perp$ des vom Koeffizientenvektor erzeugten 1-dimensionalen Unterraums von \mathbb{R}^2 . Einen zum Koeffizientenvektor (a, b) orthogonalen Vektor $\neq 0$ kann man direkt angeben, nämlich $(-b, a)$, und dieser Vektor ist dann Basis von G . Die Lösungsmenge ist also $G = \mathbb{R} \begin{pmatrix} -b \\ a \end{pmatrix}$ und $G^\perp = \mathbb{R} \begin{pmatrix} a \\ b \end{pmatrix}$ ist die zu G orthogonale Ursprungsgerade, beschrieben durch die Gleichung $-bx_1 + ax_2 = 0$.

(3) In der Ebene $x_1 - 2x_2 + 3x_3 = 0$ in \mathbb{R}^3 soll eine Orthonormalbasis bestimmt werden. Hierzu ergänzen wir den Koeffizientenvektor $v_1 = (1, -2, 3)$ durch $v_2 := (2, 1, 0)$ und $v_3 := (0, 0, 1)$ zu einer Basis von \mathbb{R}^3 . (v_2 ist ein schon zu v_1 senkrechter Vektor, den man direkt "erraten" kann, und der kanonische Basisvektor v_3 ist günstig, weil v_2 einen Nulleintrag in dritter Position hat.) Da Orthonormierungsverfahren liefert dann die Vektoren

$$u_1 = \frac{1}{\sqrt{14}} \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}, \quad u_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \quad u_3 = \frac{1}{\sqrt{70}} \begin{pmatrix} -3 \\ 6 \\ 5 \end{pmatrix},$$

und die beiden Vektoren u_2, u_3 sind eine Orthonormalbasis der Ebene. Ein solche Orthonormalbasis hätte man auch erhalten können, indem man eine Basis des Lösungsraums direkt bestimmt, z.B. $(2, 1, 0), (-3, 0, 1)$ durch Wahl von x_2, x_3 als Nicht-Basisvariable, und diese dann orthonormiert. Hier ergibt sich so "zufällig" dieselbe Orthonormalbasis. Die zweite Vorgehensweise führt schneller zum Ziel, aber wir wollten hier die in (1) beschriebene Methode veranschaulichen.

Man erkennt, dass das in (1) beschriebene Verfahren, das die Berechnung von n orthormalen Vektoren in \mathbb{R}^n erfordert, im Allgemeinen aufwendiger ist als die Berechnung einer Basis des Lösungsraums mit Hilfe der Zeilen-Stufen-Form und anschließender Orthonormierung dieser k Vektoren.

(4) Will man einen (z.B. parametrisch, also durch ein Erzeugendensystem) gegebenen Unterraum U von \mathbb{R}^n als Lösungsmenge eines Gleichungssystems darstellen, so kann man sich Folgende zu Nutze machen:

- Nimmt man die Vektoren einer Basis des orthogonalen Komplements U^\perp in \mathbb{R}^n als Zeilen einer Koeffizientenmatrix $A \in \mathbb{R}^{n \times m}$, so hat das Gleichungssystem $Ax = 0$ den Unterraum U als Lösungsmenge.

Die Lösungsmenge ist nämlich nach (1) gleich $(U^\perp)^\perp$, und das ist nach dem letzten Satz gleich U . Eine Basis von U^\perp kann man bestimmen, indem man wie in (1) eine Basis von U zu einer von \mathbb{R}^n ergänzt und darauf das Orthonormierungsverfahren anwendet, oder indem man die Vektoren einer Basis (oder eines Erzeugendensystems) von U als Zeilen einer Matrix B nimmt und mit Transformation auf Zeilen-Stufen-Form eine Basis der Lösungsmenge des Gleichungssystem $Bx = 0$ berechnet; denn die Lösungsmenge ist ja U^\perp .

(5) Zu der von $u := (1, 2, -1)$ erzeugten Geraden in \mathbb{R}^3 soll ein beschreibendes Gleichungssystem aufgestellt werden. Dazu müssen wir gemäß (4) nur zwei linear unabhängige zu u orthogonale Vektoren finden. Das kann man "im Kopf" erledigen, z.B. sind $(1, 0, 1)$ und $(-2, 1, 0)$ solche Vektoren. Diese nimmt man nun als Koeffizientenzeilen eines homogenen linearen Gleichungssystems und erhält für die gegebene Gerade das Gleichungssystem.

$$\begin{array}{rcl} x_1 & & +x_3 = 0 \\ -2x_1 + x_2 & & = 0 \end{array}$$

(6) Zur der durch

$$\begin{cases} x_1 + x_3 = 0 \\ x_2 + x_3 + x_4 = 0 \end{cases}$$

definierten Ursprungsebene E in \mathbb{R}^4 soll eine Basis und für die dazu orthogonale Ursprungsebene eine Basis und ein Gleichungssystem bestimmt werden. Das gegebene Gleichungssystem hat schon Zeilen–Stufen–Form $\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$ und mit der Wahl von x_3, x_4 als Nicht–Basis–Variable erhält man $(-1, -1, 1, 0), (0, -1, 0, 1)$ als Basis von E . Ein Gleichungssystem für E^\perp ist somit

$$\begin{cases} -x_1 - x_2 + x_3 = 0 \\ -x_2 + x_4 = 0 \end{cases}$$

Eine Basis von E^\perp ist gegeben durch die Koeffizientenzeilen $(1, 0, 1, 0)$ und $(0, 1, 1, 1)$ des ursprünglichen Gleichungssystems (da diese linear unabhängig sind). Wenn man andererseits in dem für E^\perp aufgestellten Gleichungssystem x_3, x_4 als Basis–Variable wählt, so erhält man die Basis $(1, 0, 1, 0), (-1, 1, 0, 1)$ der Lösungsmenge E^\perp . Das Gleichungssystem mit diesen beiden Vektoren als Koeffizientenzeilen

$$\begin{cases} x_1 + x_3 = 0 \\ -x_1 + x_2 + x_4 = 0 \end{cases}$$

hat dann wiederum E als Lösungsmenge. (Es entsteht auch durch eine leicht erkennbare Zeilenoperation aus dem ursprünglichen Gleichungssystem für E .) ■

Die folgende Diskussion soll die frühere Ankündigung präzisieren, dass man alle Orthonormalbasen in \mathbb{R}^N durch Drehung der kanonischen Basis erhält und einen Eindruck von der Gesamtheit aller Orthonormalbasen vermitteln

DISKUSSION (orthogonale Matrizen und orthogonale Transformationen):

(1) Ist $A \in \mathbb{R}^{n \times n}$ die Übergangsmatrix von der kanonischen Basis zu einer anderen Basis v_1, \dots, v_n , sind also v_1, \dots, v_n die Spalten von A , so ist das Skalarprodukt der i -ten Zeile der transponierten Matrix A^\top mit der j -ten Spalte von A gleich $v_i \cdot v_j$. Dieses Skalarprodukt ist aber nichts anderes als der Eintrag in Position (i, j) beim Matrixprodukt $A^\top A$. Also ist v_1, \dots, v_n genau dann eine Orthonormalbasis in \mathbb{R}^n , wenn gilt

$$A^\top A = \mathbb{I}_n = AA^\top \quad \text{bzw. äquivalent} \quad A^\top = A^{-1},$$

d.h. die zu A transponierte Matrix ist die Inverse zu A . Eine quadratische Matrix mit dieser Eigenschaft heißt **orthogonale Matrix**. Äquivalent ist, wie die Überlegung gezeigt hat, dass die Spaltenvektoren Euklidische Länge 1 haben und paarweise orthogonal zueinander sind. Äquivalent ist auch dieselbe Aussage über die Zeilenvektoren, weil A offenbar genau dann orthogonale Matrix ist, wenn dies auch auf A^\top zutrifft (wegen $(A^\top)^\top = A$). (Die Benennung “orthonormale Matrix” statt “orthogonale Matrix” wäre daher präziser, ist aber unüblich.) Eine Konsequenz aus dem Multiplikationssatz für die Determinante ist $1 = \det(\mathbb{I}) = \det(A^\top A) = (\det A^\top)(\det A) = (\det A)(\det A) = (\det A)^2$, also gilt

$$\det(A) = \pm 1 \quad \text{für orthogonale Matrizen } A.$$

(2) Wir betrachten allgemeiner einen n -dimensionaler Vektorraum V mit Skalarprodukt und mit Orthonormalbasis u_1, \dots, u_n und eine zweite Basis v_1, \dots, v_n mit Übergangsmatrix $A = (a_{ij})$ von den u_i zu den v_j , also $v_j = \sum_{i=1}^n a_{ij} u_i$. Dann ist nach dem früheren Satz über Orthonormalsysteme das Skalarprodukt zwischen zwei der Vektoren v_j dasselbe wie das Skalarprodukt ihrer Koeffizientenvektoren bzgl. der Orthonormalbasis u_1, \dots, u_n , also zwischen den entsprechenden Spaltenvektoren von A . Daran sieht man,

dass die Übergangsmatrix genau dann eine orthogonale Matrix ist, wenn auch die zweite Basis orthonormal ist. Insbesondere ist die Übergangsmatrix zwischen zwei beliebigen Orthonormalbasen eine orthogonale Matrix.

(3) Die orthogonalen $n \times n$ -Matrizen A, \tilde{A}, \dots bilden bzgl. der Matrix-Multiplikation eine Untergruppe der Gruppe aller invertierbaren $n \times n$ -Matrizen (d.h. A^{-1} und $A\tilde{A}$ sind wieder orthogonal), genannt **orthogonale Gruppe** von \mathbb{R}^n . Unter Beachtung der Symmetrie von $A^T A$ und von \mathbb{I}_n enthält die Matrix-Gleichung $A^T A = \mathbb{I}_n$ formal $\frac{1}{2}n(n+1)$ verschiedene quadratische Gleichungen für die n^2 Matrixeinträge a_{ij} von A . Man wird also erwarten, dass bei einer orthogonalen Matrix in gewisser Weise $n^2 - \frac{1}{2}n(n+1) = \frac{1}{2}(n-1)n$ Einträge frei gewählt werden können. Dies trifft tatsächlich zu und kann mathematisch präzisiert werden. Die orthogonale Gruppe und damit auch die Menge aller Orthonormalbasen in \mathbb{R}^n ist eine "Mannigfaltigkeit" der Dimension $\frac{1}{2}(n-1)n$. Das erwähnen wir, damit man einen Eindruck davon gewinnen kann, "wieviele Orthonormalbasen es gibt". Die Gruppe aller invertierbaren $n \times n$ -Matrizen, bzw. die Mannigfaltigkeit aller (nicht unbedingt orthonormalen) Basen in \mathbb{R}^n , hat dagegen Dimension n^2 ; denn man kann in einer invertierbaren $n \times n$ -Matrix A alle Einträge im Wesentlichen frei wählen. (Man muss nur den Ausnahmefall $\det(A) = 0$ vermeiden.)

(4) Die orthogonalen 2×2 -Matrizen sind genau die der folgenden Form:

$$D_\alpha = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \quad \text{und} \quad S_\alpha = \begin{pmatrix} \cos \alpha & \sin \alpha \\ \sin \alpha & -\cos \alpha \end{pmatrix}.$$

D_α beschreibt eine Drehung um 0 in \mathbb{R}^2 mit dem Winkel α und S_α die Spiegelung in \mathbb{R}^2 an der Ursprungsgeraden mit Winkel $\frac{1}{2}\alpha$ zur ersten Koordinatenachse. Die Spaltenvektorenpaare dieser Matrizen sind also die Orthonormalbasen in \mathbb{R}^2 . (Das läuft auf die bekannte Gleichung $(\sin \alpha)^2 + (\cos \alpha)^2 = 1$ hinaus.)

Die Gruppe der orthogonalen 2×2 -Matrizen hat die Dimension $\frac{1}{2}(2-1)2 = 1$. In solchen Matrizen kann man daher einen reellen Parameter frei wählen, und das ist der Winkel α . Die Gruppe der orthogonalen 3×3 -Matrizen hat Dimension $\frac{1}{2}(3-1)3 = 3$, also kann man in einer orthogonalen 3×3 -Matrix 3 reelle Parameter frei wählen. Entsprechende Matrizen, die von 3 Winkeln abhängen, lassen sich explizit angeben. Eine bessere Einsicht in die Struktur der Abbildungen $T(\mathbf{x}) = A\mathbf{x}$ von \mathbb{R}^n in sich, die durch orthogonale Matrizen gegeben sind, gewinnt man aber für $n \geq 3$, wenn man statt der kanonischen Basis eine günstigere, zu T "passende" Orthonormalbasis wählt. Wie sich zeigen lässt, kann man dies stets so machen, dass die Matrixdarstellung von T bzgl. der neuen Orthonormalbasis eine einfache "Diagonal-Blockgestalt" besitzt, wobei auf der Diagonalen 2×2 -Blöcke $D_{\alpha_1}, \dots, D_{\alpha_m}$ mit Winkeln α_i zwischen 0 und 180 Grad stehen und außerdem evtl. noch Einträge $+1$ und, wenn $\det(A) = -1$ ist, noch ein Eintrag -1 (alle anderen Einträge sind Null). Dies bedeutet geometrisch, dass die Abbildung in m paarweise zueinander senkrechten Ursprungsebenen eine Drehung um die Winkel $\alpha_1, \dots, \alpha_m$ bewirkt und in dem zu diesen Ebenen orthogonalen Unterraum (wenn er nicht $\{0\}$ ist), alle Punkte festlässt bzw. als Spiegelung wirkt.

(5) Die lineare Abbildung $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$, $T(\mathbf{x}) = A\mathbf{x}$, die durch eine orthogonale Matrix A definiert wird, führt die kanonische Basis in eine Orthonormalbasis v_1, \dots, v_n über (die Spaltenvektoren $v_j = T(e_j) = Ae_j$ von A). Der Vektor $T(\mathbf{x})$ hat bzgl. dieser Basis den Koeffizientenvektor \mathbf{x} . Da man die Skalarprodukte von Vektoren als Skalarprodukte ihrer Koeffizientenvektoren berechnen kann, folgt $T(\mathbf{x}) \cdot T(\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$ für alle $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, d.h. die lineare Abbildung T erhält Skalarprodukte zwischen Vektoren. (Das kann man auch direkt nachrechnen mit $(A\mathbf{x}) \cdot (A\mathbf{y}) = (A\mathbf{x})^T A\mathbf{y} = \mathbf{x}^T A^T A\mathbf{y} = \mathbf{x}^T \mathbb{I}_n \mathbf{y} = \mathbf{x}^T \mathbf{y} = \mathbf{x} \cdot \mathbf{y}$.)

- Eine $n \times n$ -Matrix ist orthogonal, genau wenn die lineare Abbildung $T(\mathbf{x}) = A\mathbf{x}$ Skalarprodukte erhält, d.h. $(A\mathbf{x}) \cdot (A\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$ für alle $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Daran erkennt man die geometrische Bedeutung orthogonaler Matrizen A . Mit den Skalarprodukten erhält $T(\mathbf{x}) = A\mathbf{x}$ nämlich alle geometrischen Größen und Relationen, die mit dem Skalarprodukt definiert werden können, also z.B. Längen $|T(\mathbf{x})| = |\mathbf{x}|$ (da $|\mathbf{x}| = \sqrt{\mathbf{x} \cdot \mathbf{x}}$), Orthogonalität $\mathbf{x} \perp \mathbf{y} \iff T(\mathbf{x}) \perp T(\mathbf{y})$ (da $\mathbf{x} \perp \mathbf{y} \iff \mathbf{x} \cdot \mathbf{y} = 0$) und die Winkel $\sphericalangle(\mathbf{x}, \mathbf{y})$ zwischen Einheitsvektoren (die definiert sind durch die Bedingung, dass ihr Kosinus gleich $\mathbf{x} \cdot \mathbf{y}$ ist). Man nennt solche Abbildungen $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$ daher auch **verallgemeinerte Drehungen** des n -dimensionalen Zahlenraumes, genauer (eigentliche) **Drehungen**, wenn $\det(A) = 1$ ist, und **Drehspiegelungen** im Fall $\det(A) = -1$.

(6) Eine lineare Abbildung $T: V \rightarrow W$ zwischen Vektorräumen mit Skalarprodukt heißt **orthogonale Abbildung**, wenn sie Skalarprodukte erhält, wenn also gilt $T(v) \cdot T(\tilde{v}) = v \cdot \tilde{v}$ für alle $v, \tilde{v} \in V$. Nach (5) sind die orthogonalen Automorphismen des \mathbb{R}^n (d.h. die orthogonalen Abbildungen des Raums in sich) genau die der Form $\mathbf{x} \mapsto A\mathbf{x}$ mit einer orthogonalen $n \times n$ -Matrix A . Wie in (5) sieht man auch für allgemeine orthogonale Abbildungen $T: V \rightarrow W$, dass T die Länge von Vektoren erhält. Insbesondere gilt $T(v) = 0$ nur für $v = 0$, d.h. T ist injektiv; daher muss $\dim W \geq \dim V$ sein, wenn eine orthogonale Abbildung von V in W existiert. Außerdem erhält T auch Orthogonalität (und allgemeiner die Winkel) zwischen Vektoren, also transformiert T Orthonormalsysteme in V in Orthonormalsysteme in W und, wenn V und W dieselbe endliche Dimension haben, Orthonormalbasen von V in Orthonormalbasen von W . Genauer sind für lineare $T: V \rightarrow W$ zwischen Vektorräumen mit Skalarprodukt folgende Aussagen äquivalent:

- (i) T ist orthogonale Abbildung, erhält also Skalarprodukte, $T(v) \cdot T(w) = v \cdot w$;
- (ii) T erhält die Norm von Vektoren, $|T(v)| = |v|$;
- (iii) T erhält Abstände, $|T(v) - T(w)| = |v - w|$;
- (iv) T führt Orthonormalsysteme in Orthonormalsysteme über;

und wenn $\dim V < \infty$ und eine Orthonormalbasis in V spezifiziert ist, auch noch:

- (v) T führt die gegebene Orthonormalbasis in ein Orthonormalsystem von W über.

In den Gleichungen ist dabei auf der linken Seite natürlich stets das Skalarprodukt bzw. die assoziierte Norm in W zu nehmen und rechts im Raum V . Dass aus (i) die anderen Aussagen folgen, haben wir schon überlegt. Mit dem Argument aus (5) sieht man, dass aus (iv) und daher erst recht aus (v) auch (i) folgt, weil zwei linear unabhängige Vektoren in V stets in einer von zwei orthonormalen Vektoren aufgespannten Ursprungsebene liegen. Die Äquivalenz von (ii) mit (iii) ist klar wegen der Linearität $T(v) - T(w) = T(v - w)$. Es muss also nur noch überlegt werden, dass aus der Längenerhaltung (ii) die Skalarprodukterhaltung (i) folgt. Dies liegt daran, dass man Skalarprodukte durch Längen ausdrücken kann mit folgender sog. *Polarisationsformel*:

$$|v + w|^2 - |v - w|^2 = |v|^2 + 2v \cdot w + |w|^2 - (|v|^2 - 2v \cdot w + |w|^2) = 4v \cdot w.$$

Wenn man überlegt, welche linearen Abbildungen $T: V \rightarrow W$ *Orthogonalitäts-erhaltend* sind, also $v \perp w \implies T(v) \perp T(w)$ erfüllen, so findet man, dass dies genau die Vielfachen $T = rS$ von orthogonalen Abbildungen $S: V \rightarrow W$ sind ($r \in \mathbb{R}$). Solche Abbildungen nennt man, wenn $r \neq 0$, also T nicht die Nullabbildung ist, **Ähnlichkeitsabbildungen**, weil sie Abstandsverhältnisse erhalten (wenn auch nicht unbedingt die Abstände selbst) und weil man in der elementaren Geometrie Figuren mit gleichen Abstandsverhältnissen "ähnlich" zueinander nennt.

(7) Abstandserhaltende Abbildungen zwischen Räumen, in denen eine Abstandsmessung zwischen Punkten definiert ist, heißen **isometrische Abbildungen** oder **Isometrien**. Für Isometrien $V \rightarrow W$ zwischen Räumen mit Skalarprodukt kann man zeigen, dass sie automatisch linear sind, wenn man noch eine Verschiebung in W so vorgenommen hat, dass der Nullpunkt von V in den Nullpunkt von W abgebildet wird. Gemäß (6) haben solche Isometrien also die Form $V \ni v \mapsto T(v) + w \in W$ mit einer orthogonalen Abbildung T von V nach W und einem festen Vektor $w \in W$. Umgekehrt sind alle Abbildungen dieser Gestalt auch isometrisch. Im Fall $V = W = \mathbb{R}^n$ sind also die isometrischen Selbstabbildungen genau die der Form $\mathbf{x} \mapsto T(\mathbf{x}) + \mathbf{b} = A\mathbf{x} + \mathbf{b}$ mit einem orthogonalen Automorphismus T des \mathbb{R}^n bzw. einer orthogonalen $n \times n$ -Matrix A und einem Vektor $\mathbf{b} \in \mathbb{R}^n$. Diese Abbildungen nennt man auch **Bewegungen** oder **Kongruenzen** des n -dimensionalen Euklidischen Raumes. ■

Wir schließen diesen Abschnitt, der schon recht weit in die lineare Algebra geführt hat, mit einem Hauptsatz der linearen Algebra, der später bei der Optimierung (Maximum- oder Minimum-Bestimmung) von quadratischen Funktionen oder von allgemeineren differenzierbaren Funktionen mehrerer Veränderlicher nützlich sein wird. Der Satz hängt zusammen mit der früheren Beobachtung, dass eine lineare Selbstabbildung $T(\mathbf{x}) = A\mathbf{x}$ des \mathbb{R}^n bzgl. einer "günstigen" Basis u_1, \dots, u_n unter Umständen eine Matrixdarstellung hat, die viel einfacher ist als die Matrix $A \in \mathbb{R}^{n \times n}$, welche T bzgl. der kanonischen Basis e_1, \dots, e_n darstellt. "Einfach" wäre in diesem Zusammenhang z.B. eine Diagonalmatrixdarstellung, d.h. es gilt $T(u_i) = \lambda_i u_i$ mit gewissen reellen Zahlen λ_i (den Diagonaleinträgen). Anders gesagt: Die Basis u_1, \dots, u_n besteht aus Eigenvektoren von A und die Zahlen λ_i sind die zugehörigen Eigenwerte (siehe 3.4). Wenn dies gilt und wenn die Basis u_1, \dots, u_n orthonormal ist, so haben wir (wegen $\delta_{ij} = 0$ für $i \neq j$)

$$u_i \cdot (Au_j) = u_i \cdot (\lambda_j u_j) = \lambda_j \delta_{ij} = \lambda_i \delta_{ij} = (\lambda_i u_i) \cdot u_j = (Au_i) \cdot u_j,$$

und für Linearkombinationen $v = \sum_{i=1}^n r_i u_i$ und $w = \sum_{i=1}^n s_i u_i$ daher:

$$v \cdot (Aw) = \sum_{i=1}^n \sum_{j=1}^n r_i s_j u_i \cdot (Au_j) = \sum_{i=1}^n \sum_{j=1}^n r_i s_j (Au_i) \cdot u_j = (Av) \cdot w.$$

Für die Einträge der Matrix A folgt insbesondere

$$a_{ij} = e_i \cdot (Ae_j) = (Ae_i) \cdot e_j = a_{ji},$$

d.h. die Matrix A ist symmetrisch. Wenn wir also eine Chance haben wollen, zur linearen Abbildung $T(\mathbf{x}) = A\mathbf{x}$ eine Orthonormalbasis zu finden, bzgl. der die Abbildung durch eine Diagonalmatrix dargestellt wird, so muss A symmetrisch sein! Der folgende Satz sagt, dass man allein unter dieser – notwendigen – Voraussetzung dann auch tatsächlich schon die gesuchte Orthonormalbasis wie gewünscht finden kann.

Satz (von der Hauptachsentransformation):

Ist A eine symmetrische $n \times n$ -Matrix, so gibt es eine Orthonormalbasis u_1, \dots, u_n von \mathbb{R}^n aus Eigenvektoren von A , also $Au_i = \lambda_i u_i$ für gewisse Eigenwerte $\lambda_i \in \mathbb{R}$ und $i = 1 \dots n$.

Ist B die orthogonale Matrix mit den u_i als Spalten, so ist also dann $B^{-1}AB = B^T A B$ die Diagonalmatrix mit den Eigenwerten $\lambda_i \in \mathbb{R}$ als Diagonaleinträgen.

Zum *Beweis* genügt es im Wesentlichen, einen einzigen Eigenwert λ und zugehörigen normierten Eigenvektor u von A zu finden. Wegen der Symmetrie $u \cdot (Av) = (Au) \cdot v = \lambda u \cdot v$ bildet A dann das orthogonale Komplement $(\mathbb{R}u)^\perp$ in sich ab. Dieses Komplement ist ein $(n-1)$ -dimensionaler Unterraum und dafür können wir den Satz bereits als bewiesen annehmen (Induktionsbeweis), so dass es eine Orthonormalbasis von $(\mathbb{R}u)^\perp$ aus Eigenvektoren von A gibt, die u zu einer Orthonormalbasis von \mathbb{R}^n ergänzt. (Genauer bildet man $(\mathbb{R}u)^\perp$ mittels Wahl einer Orthonormalbasis durch einen Isomorphismus auf \mathbb{R}^{n-1} ab, wobei Skalarprodukte erhalten werden und die Wirkung von A auf $(\mathbb{R}u)^\perp$ daher im Bild \mathbb{R}^{n-1} durch eine symmetrische $(n-1) \times (n-1)$ -Matrix A' dargestellt werden kann. In \mathbb{R}^{n-1} gibt es dann nach Induktionsannahme eine orthonormale Basis aus Eigenvektoren von A' , und unter dem Isomorphismus entspricht diese einer Orthonormalbasis aus Eigenvektoren von A in $(\mathbb{R}u)^\perp$.)

Um den Eigenwert λ und Eigenvektor u zu finden gibt es zwei Wege. Ein algebraisches Argument ist, dass es jedenfalls einen komplexen Eigenwert $\lambda \in \mathbb{C}$ gibt, weil die algebraische Gleichung $\det(tI_n - A) = 0$, deren Lösungen ja die Eigenwerte sind, gemäß dem für \mathbb{C} gültigen Fundamentalsatz der Algebra stets eine Lösung hat. Zu λ gibt es dann auch einen Eigenvektor, der allerdings komplexe Einträge haben kann, also die Form $\mathbf{x} + i\mathbf{y} \neq 0$ hat mit $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ (und der imaginären Einheit i , also $i^2 = -1$). Die mit der Symmetrie der Matrix A äquivalente Gleichung $v \cdot (Aw) = (Av) \cdot w$ gilt nun aber auch für Vektoren mit komplexen Einträgen. Wählen wir insbesondere $v := \mathbf{x} + i\mathbf{y}$ und $w := \bar{v} = \mathbf{x} - i\mathbf{y}$, so folgt: $\lambda v \cdot \bar{v} = (Av) \cdot \bar{v} = v \cdot (A\bar{v}) = v \cdot (\overline{Av}) = v \cdot (\overline{\lambda v}) = \bar{\lambda} v \cdot \bar{v}$. (Hier bezeichnet $\bar{z} = x - iy$ die konjugiert komplexe Zahl zu $z = x + iy$ für $x, y \in \mathbb{R}$.) Da nun $v \cdot \bar{v} = |\mathbf{x}|^2 + |\mathbf{y}|^2$ reell und positiv ist, folgt $\lambda = \bar{\lambda}$, d.h. der Eigenwert λ ist in Wahrheit reell. Dann gibt es dazu natürlich auch einen reellen normierten Eigenvektor u .

Das zweite Argument ist analytisch und hat mehr Bezug zu den Optimierungsaufgaben, auf die der Satz später angewendet wird. Hierbei wählt man u als einen Einheitsvektor, für den $\mathbf{x} \cdot (A\mathbf{x})$ einen kleinstmöglichen Wert hat. Eine solche Stelle existiert, weil die quadratische Funktion $\mathbf{x} \mapsto \mathbf{x} \cdot (A\mathbf{x}) = \sum_{i,j=1}^n a_{ij} x_i x_j$ auf der Menge der Einheitsvektoren, die eine beschränkte und abgeschlossene Menge in \mathbb{R}^n ist, ein Minimum annimmt. (Das wird in der Analysis bewiesen.) Für jeden zu u orthogonalen Einheitsvektor und alle $t \in \mathbb{R}$ gilt dann $u \cdot (Au) \leq \frac{1}{1+t^2} (u+tv) \cdot (Au+tAv) = \frac{1}{1+t^2} [u \cdot (Au) + tu \cdot (Av) + tv \cdot (Au) + t^2 v \cdot (Av)]$, weil $\frac{1}{\sqrt{1+t^2}} (u+tv)$ auch ein Einheitsvektor ist. Für kleine $t > 0$ kann das nur sein, wenn $u \cdot (Av) + v \cdot (Au) = 0$ ist; denn die Terme mit t^2 sind vernachlässigbar, wenn $0 < t \ll 1$. Wegen der Symmetrie $u \cdot (Av) = (Au) \cdot v = v \cdot (Au)$ folgt dann $v \cdot (Au) = 0$ für alle zu u senkrechten Einheitsvektoren v . Dies bedeutet aber, dass Au in dem zu diesen Vektoren v orthogonalen Unterraum liegt und das ist $\mathbb{R}u$. Also gilt $Au = \lambda u$ mit einer reellen Zahl λ .

DISKUSSION: (1) Sind V, W endlichdimensionale Vektorräume mit Skalarprodukt und $T: V \rightarrow W$ eine lineare Abbildung, so gibt es eine eindeutig bestimmte Abbildung $T^*: W \rightarrow V$ mit

$$w \cdot T(v) = T^*(w) \cdot v \quad \text{für alle } v \in V, w \in W.$$

T^* heißt die zu T **adjungierte lineare Abbildung**. Sind Orthonormalbasen v_1, \dots, v_n bzw. w_1, \dots, w_m in V bzw. in W gegeben und hat T diesbezüglich die Matrixdarstellung A , also $a_{ij} = w_i \cdot T(v_j)$, so hat T^* bzgl. derselben Basen die transponierte Matrix A^T als Matrixdarstellung; denn es ist $v_j \cdot T^*(w_i) = T(v_j) \cdot w_i = a_{ij}$. Damit kann man auch die Existenz und Eindeutigkeit von T^* einsehen: Es muss $T^*(w) = \sum_{j=1}^n (T(v_j) \cdot w) v_j$ gelten, und durch diese Formel kann man T^* definieren und $T^*(w) \cdot v = w \cdot T(v)$ nachweisen.

Ist nun $T:V \rightarrow V$ eine Abbildung von V in sich, so heißt T **selbstadjungiert**, wenn $T^* = T$ gilt. Äquivalent ist, dass die Matrixdarstellung von T bzgl. einer (und jeder) Orthonormalbasis in V symmetrisch ist. Die selbstadjungierten Abbildungen entsprechen also hier den symmetrischen $n \times n$ -Matrizen im vorigen Satz. Und man kann den Satz mit gleichem Beweis oder mittels Übertragung durch einen orthogonalen Automorphismus von \mathbb{R}^n auf V in folgender Form aussprechen (für endlichdimensionale V):

- Ist $T:V \rightarrow V$ eine selbstadjungierte lineare Abbildung, so hat V eine Orthonormalbasis aus Eigenvektoren zu T .

(2) Eine Funktion auf \mathbb{R}^n der Form

$$Q(\mathbf{x}) = \sum_{i,j=1}^n a_{ij}x_i x_j \quad \text{für } \mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$$

nennt man eine **quadratische Form** oder eine **homogene quadratische Polynomfunktion** von n Variablen. Man kann hier stets Symmetrie der Matrix $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ annehmen; nötigenfalls ersetzt man a_{ij} durch $\frac{1}{2}(a_{ij} + a_{ji})$, ohne den Wert der Summe zu ändern.

$$S(\mathbf{x}, \mathbf{y}) := \mathbf{x} \cdot (A\mathbf{y}) = \sum_{i,j=1}^n a_{ij}x_i y_j \quad \text{für } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$

heißt die zu Q (oder zur symmetrischen Matrix A) gehörende **symmetrische Bilinearform** auf \mathbb{R}^n . Diese hat alle Eigenschaften eines Skalarproduktes bis auf möglicherweise die Positivität, d.h. man hat Symmetrie $S(\mathbf{x}, \mathbf{y}) = S(\mathbf{y}, \mathbf{x})$ und Linearität in jeder der Variablen \mathbf{x}, \mathbf{y} , wenn man die andere jeweils festhält. Die symmetrische Bilinearform S bestimmt Q durch $Q(\mathbf{x}) = S(\mathbf{x}, \mathbf{x})$ und die symmetrische Matrix A durch $a_{ij} = S(e_i, e_j)$. Umgekehrt bestimmt Q auch S durch die Polarisierungsformel $4S(\mathbf{x}, \mathbf{y}) = Q(\mathbf{x} + \mathbf{y}) - Q(\mathbf{x} - \mathbf{y})$. Quadratische Formen, symmetrische Bilinearformen und symmetrische Matrizen sind also äquivalente Begriffsbildungen.

(3) Für quadratische Formen $Q(\mathbf{x}) = \mathbf{x} \cdot (A\mathbf{x})$ folgt, wenn man $\mathbf{x} = \sum_{i=1}^n (u_i \cdot \mathbf{x})u_i$ als Linearkombination einer Orthonormalbasis von Eigenvektoren u_i von A schreibt und $u_j \cdot (A u_i) = u_j \cdot (\lambda_i u_i) = \lambda_i \delta_{ij}$ berücksichtigt:

$$Q(\mathbf{x}) = \lambda_1(u_1 \cdot \mathbf{x})^2 + \lambda_2(u_2 \cdot \mathbf{x})^2 + \dots + \lambda_n(u_n \cdot \mathbf{x})^2.$$

Dies bedeutet, dass Q darstellbar ist als Linearkombination von "vollständigen Quadraten", genauer von Quadraten linearer Funktionen $\ell_i(\mathbf{x}) = u_{i1}x_1 + \dots + u_{in}x_n$ mit linear unabhängigen Koeffizientenvektoren. Das ist gewissermaßen ein allgemeiner Satz über quadratische Ergänzung.

Sind alle Eigenwerte λ_i positiv, so beschreibt die Gleichung $Q(\mathbf{x}) = 1$ ein im Allgemeinen "schief" in \mathbb{R}^n gelegenes Ellipsoid (Ellipse, wenn $n = 2$; sind einige $\lambda_i \leq 0$, so können auch andere Kegelschnitte wie Hyperboloide, Paraboloide, ... auftreten). Die durch die Eigenvektoren definierten Geraden $\mathbb{R}u_1, \dots, \mathbb{R}u_n$ sind dann die geometrischen Achsen dieses Ellipsoids. Daher rührt der Namen "Hauptachsentransformation".

(4) Im Hinblick auf Anwendungen in der Optimierung ist man an Kriterien dafür interessiert, dass eine quadratische Form $Q(\mathbf{x}) = \mathbf{x} \cdot (A\mathbf{x})$ auf \mathbb{R}^n die eindeutige Minimumstelle 0 hat, dass also $Q(\mathbf{x}) > 0$ gilt für alle $\mathbf{x} \neq 0$. Man nennt in diesem Fall die quadratische Form Q bzw. die symmetrische $n \times n$ -Matrix A **positiv definit**. An dem Ausdruck $\sum_{i,j=1}^n a_{ij}x_i x_j$ kann man das oft nicht direkt ablesen, selbst wenn alle $a_{ij} \geq 0$ sind, weil die gemischten Produkte $x_i x_j, i \neq j$, ja auch negative Werte annehmen können.

Aus (3) sehen wir aber sofort:

- Die symmetrische Matrix A ist positiv definit, also $\mathbf{x} \cdot (A\mathbf{x}) > 0$ für alle $0 \neq \mathbf{x} \in \mathbb{R}^n$, genau wenn alle Eigenwerte der symmetrischen Matrix A positiv sind.

Die quadratische Form ist dann eine Summe von "vollständigen Quadraten". Da die Determinante von A gleich der Determinante der Diagonalmatrix $B^{-1}AB$ ist, also das Produkt der Eigenwerte (mit ihrer Vielfachheit aufgeführt), ist dann folglich auch $\det(A) > 0$. Allgemeiner sind die Eigenwerte und die Hauptminoren, d.h. die Determinanten der zur Diagonalen symmetrisch gelegenen $k \times k$ -Untermatrizen, positiv; das sieht man, indem man in die quadratische Form Vektoren einsetzt, die Nulleinträge in den Positionen haben, die den gestrichenen Zeilen und Spalten entsprechen, und die entstehende quadratische Form auf \mathbb{R}^k betrachtet. Insbesondere sind die Diagonaleinträge $a_{ii} = \mathbf{e}_i \cdot (A\mathbf{e}_i)$ einer positiv definiten Matrix allesamt positiv. An diesem notwendigen Kriterium kann man oft direkt erkennen, dass *keine* Definitheit vorliegt. Interessant ist, dass man umgekehrt an der Positivität der Determinante von wenigen speziellen Untermatrizen schon die positive Definitheit von A erkennen kann. Man braucht dazu nur die sog. führenden Hauptminoren zu betrachten. Unter dem k -ten *führenden Hauptminor* von $A \in \mathbb{R}^{n \times n}$ versteht man dabei die Determinante der Untermatrix, die aus A durch Streichen der letzten $n-k$ Zeilen und der letzten $n-k$ Spalten entsteht ($k = 1 \dots n$). Es gilt dann folgendes **Hauptminoren-Kriterium**:

- Eine symmetrische Matrix A ist positiv definit, genau wenn all ihre führenden Hauptminoren positiv sind.

Beim Beweis mit Induktion kann man wieder annehmen, dass das Kriterium für symmetrische $(n-1) \times (n-1)$ -Matrizen schon bewiesen ist. Die Untermatrix A' , die aus A durch Streichen der letzten Zeile und Spalte entsteht, ist also positiv definit. Daher gibt es eine orthogonale $(n-1) \times (n-1)$ -Matrix B' , so dass $B'^T A' B'$ Diagonalmatrix ist mit positiven Diagonaleinträgen $\lambda'_1, \dots, \lambda'_{n-1}$. Es folgt:

$$\begin{pmatrix} B'^T & 0 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} A \underbrace{\begin{pmatrix} B' & 0 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix}}_{=: B} = \begin{pmatrix} \lambda'_1 & 0 & \dots & 0 & c_1 \\ 0 & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots \\ 0 & & & \lambda'_{n-1} & c_{n-1} \\ c_1 & \dots & c_{n-1} & a_{nn} & \end{pmatrix}.$$

Die Determinante der linken Seite unterscheidet sich nur um einen positiven Faktor (das Quadrat von $\det(B)$) von $\det(A)$, und die der Matrix rechts kann mit Zeilenoperationen leicht zu $\delta := a_{nn} - c_1^2/\lambda'_1 - \dots - c_{n-1}^2/\lambda'_{n-1}$ berechnet werden. Dieser Ausdruck ist also positiv. Außerdem erhält man

$$C^{-1}\mathbf{x} \cdot (AC^{-1}\mathbf{x}) = \sum_{i=1}^{n-1} (\lambda'_i x_i^2 + 2c_i x_i x_n) + a_{nn} x_n^2 = \sum_{i=1}^{n-1} \lambda'_i (x_i + c_i x_n / \lambda'_i)^2 + \delta x_n^2,$$

wobei die rechte Seite > 0 ist, wenn $0 \neq \mathbf{x} \in \mathbb{R}^n$. Da mit \mathbf{x} auch $C^{-1}\mathbf{x}$ alle Vektoren in $\mathbb{R}^n \setminus \{0\}$ durchläuft, ist damit die positive Definitheit von A bewiesen.

Eine *Warnung* noch: Die symmetrische $n \times n$ -Matrix A ist **negativ definit**, d.h. $-A$ ist positiv definit, genau wenn all ihre Eigenwerte negativ sind. Für die Hauptminoren gilt das aber nicht! Weil sich die Determinante einer $k \times k$ -Matrix um den Faktor $(-1)^k$ ändert, wenn man sie mit -1 multipliziert, ist negative Definitheit vielmehr äquivalent damit, dass der k -te führende Hauptminor das Vorzeichen $(-1)^k$ hat für $k = 1 \dots n$. ■

BEISPIEL und DISKUSSION: Die quadratische Form

$$Q(\mathbf{x}) = x_1^2 + 3x_2^2 + 6x_3^2 - 2x_1x_2 + 4x_1x_3$$

aus \mathbb{R}^3 soll auf Hauptachsenform gebracht und auf positive Definitheit untersucht werden. Zunächst bestimmen wir die symmetrische 3×3 -Matrix A mit $Q(\mathbf{x}) = \mathbf{x} \cdot (A\mathbf{x})$. An den Koeffizienten von Q liest man ab (beachte, dass $a_{ij} = a_{ji}$ die Hälfte des Koeffizienten von $x_i x_j$ ist für $i < j$, wenn nur Produkte $x_i x_j$ mit $i < j$ aufgeführt sind):

$$A = \begin{pmatrix} 1 & -1 & 2 \\ -1 & 3 & 0 \\ 2 & 0 & 6 \end{pmatrix}.$$

Um die Eigenwerte zu bestimmen, berechnen wir das charakteristische Polynom:

$$p_A(t) = \det(t\mathbb{I}_3 - A) = \det \begin{pmatrix} t-1 & 1 & -2 \\ 1 & t-3 & 0 \\ -2 & 0 & t-6 \end{pmatrix} = t^3 - 10t^2 + 22t.$$

Die Eigenwerte sind die Nullstellen. Ein Eigenwert ist also $\lambda_1 = 0$, zwei weitere ergeben sich dann durch Lösen einer quadratischen Gleichung zu $\lambda_2 = 5 - \sqrt{3}$, $\lambda_3 = 5 + \sqrt{3}$. Zugehörige Eigenvektoren berechnet man nun als nichttriviale Lösungen der homogenen Gleichungssysteme $(\lambda_i \mathbb{I}_3 - A)\mathbf{x} = 0$. Man findet damit die Eigenvektoren

$$\tilde{u}_1 = \begin{pmatrix} 3 \\ 1 \\ -1 \end{pmatrix}, \quad \tilde{u}_2 = \begin{pmatrix} 1 \\ -2 - \sqrt{3} \\ 1 - \sqrt{3} \end{pmatrix}, \quad \tilde{u}_3 = \begin{pmatrix} 1 \\ -2 + \sqrt{3} \\ 1 + \sqrt{3} \end{pmatrix}$$

oder Vielfache davon. Diese Vektoren sind schon paarweise orthogonal zueinander, und das muss auch so sein; denn es gilt allgemein:

- *Eigenvektoren zu verschiedenen Eigenwerten einer symmetrischen Matrix sind orthogonal zueinander.*

Das sieht man aus $\lambda(u \cdot v) = (Au) \cdot v = u \cdot (Av) = \mu(u \cdot v)$, wenn $Au = \lambda u$ ist und $Av = \mu v$. Wenn Eigenwerte λ als Nullstellen des charakteristischen Polynoms mit einer Vielfachheit $k > 1$ auftreten, so hat der entsprechende *Eigenraum*, das ist die Lösungsmenge des Gleichungssystems $(\lambda \mathbb{I} - A)\mathbf{x} = 0$, die Dimension k ; das folgt aus dem Satz über die Hauptachsentransformation. Bei der Berechnung von k linear unabhängigen Lösungen wird man dann zunächst im Allgemeinen nicht schon k zueinander orthogonale Eigenvektoren zu diesem Eigenwert λ erhalten, aber man kann k linear unabhängige Eigenvektoren ja immer orthonormieren.

Um die Hauptachsentransformation für das obige Beispiel abzuschließen, brauchen wir die Eigenvektoren \tilde{u}_i also nur noch zu normieren. Resultat ist die folgende Orthonormalbasis des \mathbb{R}^3 aus Eigenvektoren von A :

$$\tilde{u}_1 = \frac{1}{\sqrt{11}} \begin{pmatrix} 3 \\ 1 \\ -1 \end{pmatrix}, \quad \tilde{u}_2 = \frac{1}{\sqrt{12+2\sqrt{3}}} \begin{pmatrix} 1 \\ -2-\sqrt{3} \\ 1-\sqrt{3} \end{pmatrix}, \quad \tilde{u}_3 = \frac{1}{\sqrt{12-2\sqrt{3}}} \begin{pmatrix} 1 \\ -2+\sqrt{3} \\ 1+\sqrt{3} \end{pmatrix}.$$

Damit ergibt sich nun die folgende Darstellung der quadratischen Form als Summe "vollständiger Quadrate":

$$Q(\mathbf{x}) = \lambda_1(u_1 \cdot \mathbf{x})^2 + \lambda_2(u_2 \cdot \mathbf{x})^2 + \lambda_3(u_3 \cdot \mathbf{x})^2 \\ = \frac{5-\sqrt{3}}{\sqrt{12+2\sqrt{3}}}[x_1-(2+\sqrt{3})x_2+(1-\sqrt{3})x_3]^2 + \frac{5+\sqrt{3}}{\sqrt{12-2\sqrt{3}}}[x_1-(2-\sqrt{3})x_2+(1+\sqrt{3})x_3]^2.$$

Auf diese Darstellung von $Q(\mathbf{x})$ als Summe von Quadraten linearer Ausdrücke in \mathbf{x} wäre man durch Probieren wohl kaum gekommen. Die Berechnung einer Hauptachsentransformation ist, wie man sieht, zwar aufwendig, erfordert aber nur einfache Standardalgorithmen. Im ersten Schritt muss man die Eigenwerte bestimmen als Nullstellen des charakteristischen Polynoms. Das ist bei Grad $n \geq 3$ nur "mit Glück" exakt durchführbar. Wenn man die Eigenwerte alle gefunden hat, so sind die weiteren Schritte elementare lineare Algebra; man muss nur noch homogene lineare Gleichungssysteme lösen und die Lösungen normieren, bzw. bei Dimension des Lösungsraums > 1 eine Orthonormalbasis von Lösungen darin bestimmen.

Da der Eigenwert $\lambda_1 = 0$ auftritt, ist übrigens Q nicht positiv definit; vielmehr nimmt $Q(\mathbf{x})$ den Wert Null auf gewissen Vektoren $\mathbf{x} \neq 0$ an, genauer gesagt: $Q(\mathbf{x}) = 0 \iff \mathbf{x} \in \mathbb{R}u_1$. Immerhin sieht man, dass $Q(\mathbf{x}) \geq 0$ gilt für alle \mathbf{x} . Dann nennt man die quadratische Form bzw. die zugehörige symmetrische Matrix A **positiv semidefinit**. Äquivalent ist, dass A keine negativen Eigenwerte hat. Bei positiv semidefiniten Matrizen sind alle Hauptminoren nichtnegativ, insbesondere also die Diagonaleinträge und die Determinante.

Das Hauptminoren-Kriterium für positive Definitheit lässt sich nur teilweise zu einem Kriterium für Semidefinitheit erweitern. Der Beweis des Hauptminorenkriteriums oben zeigt immerhin, dass A jedenfalls dann positiv semidefinit ist, wenn alle führenden Hauptminoren bis auf den zum größten Format $n \times n$ positiv sind und wenn nur der Hauptminor zum größten Format, also die Determinante der $n \times n$ -Matrix A selbst, Null ist. Es gibt dann einen Null-Eigenwert mit Vielfachheit 1, und alle anderen sind positiv. Wenn A aber den Eigenwert Null mit Vielfachheit > 1 hat, so ist außer $\det(A)$ auch mindestens noch ein führender Hauptminor zu kleinerem Format Null. Und man kann leider nicht umgekehrt auf positive Semidefinitheit schließen, wenn etwa die führenden Hauptminoren zu den beiden größten Formaten $n \times n$ und $(n-1) \times (n-1)$ Null sind und die anderen positiv. Das zeigt z.B. die Matrix

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix},$$

die zur quadratischen Form

$$x_1^2 + x_2^2 + 2x_1x_2 + 2x_1x_3 + 2x_2x_3$$

gehört, welche **indefinit** ist, d.h. positive und negative Werte annimmt. Äquivalent ist, dass die symmetrische Matrix positive und negative Eigenwerte hat. ■