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 linearen 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 $(n-l)$–parametrigen Schar der Lösungen des homogenen Gleichungssystems bzw. von der k–parametrigen 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, ... 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:

$$u+v=v+u,\quad (u+v)+w=u+(v+w),\quad v+0=v,\quad u+v=0 \Longleftrightarrow v=(-1)u,$$
$$1v=v,\quad 0v=0,\quad r(sv)=(rs)v,\quad (r+s)v=rv+sv,\quad r(u+v)=ru+rv.$$

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 $V, W, \ldots$ einfach Zahlenräume $\mathbb{R}^n, \mathbb{R}^m, \ldots$ mit den aus 3.3 bekannten "komponentenweisen" Rechenoperationen vorstellen.

Zu den Vektorräumen $V, W, \ldots$ gehören die Abbildungen $T\colon V \to W$ 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 $\sum_{j=1}^{k} v_j = v_1+v_2+\ldots+v_k \in V$ 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 $v_1, \ldots, v_k \in V$ mit Koeffizienten $r_1, \ldots, r_k \in \mathbb{R}$, also der Summe der Vielfachen $r_j v_j$ der Vektoren v_j,

$$\sum_{j=1}^{k} r_j v_j = r_1 v_1 + r_2 v_2 + \ldots + r_k v_k \in V \qquad (v_j \in V,\ r_j \in \mathbb{R} \text{ für } j = 1 \ldots k).$$

Lineare Abbildungen $T: V \to W$ zwischen Vektorräumen sind dann automatisch auch mit dieser abgeleiteten Rechenoperation verträglich, d.h. man kann die Linearkombinationsbildung "herausziehen" im Sinne von

$$T\Big(\sum_{j=1}^{k} r_j v_j\Big) = \sum_{j=1}^{k} r_j T(v_j)\,.$$

Das intuitive Konzept, dass gewisse Vektoren $v \in V$ durch gegebene Vektoren $v_1, \ldots, v_k$ "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 $v_1, \ldots, v_k$ des Vektorraums V **linear ausgedrückt** werden kann (oder dass v *von den Vektoren* $v_1, \ldots, v_k$ *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 $v_1, \ldots, v_k \in V$ 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 $v_1, \ldots, v_k$ im Vektorraum V:

- *Keiner der Vektoren v_i ist Linearkombination der vorangehenden $v_1, \ldots, v_{i-1}$.*
- *Der Nullvektor ist aus $v_1, \ldots, v_k$ nur auf triviale Weise linear kombinierbar, d.h.*
  $r_1 v_1 + r_2 v_2 + \ldots + r_k v_k = 0$ *gilt nur, wenn* $r_1 = r_2 = \ldots = r_k = 0$.
- *Die Koeffizienten von Linearkombinationen der v_j sind eindeutig bestimmt, d.h. aus*
  $r_1 v_1 + r_2 v_2 + \ldots + r_k v_k = s_1 v_1 + s_2 v_2 + \ldots + s_k v_k$ *folgt* $r_1 = s_1,\ r_2 = s_2,\ \ldots, r_k = s_k$.

Aus der linearen Unabhängigkeit von $v_1, \ldots, v_k$ 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 $v_1, \ldots, v_{i-1}$. 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 $r_{i+1} = \ldots = r_k = 0 \neq r_i$, 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 $v_1, \ldots, v_{i-1}$ 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 $v_1, \ldots, v_k$ 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 $v_1, \ldots, v_k \in \mathbb{R}^n$ sind linear unabhängig, genau wenn das homogene lineare Gleichungssystem $A\mathbf{x} = 0$ mit der aus den Spalten v_j zusammengestellten $n \times k$–Matrix $A = (v_1\, v_2\, \ldots\, v_k)$ nur die triviale Lösung $\mathbf{x} = 0 \in \mathbb{R}^k$ hat.*

Das ist mit (1) klar, weil $A\mathbf{x}$ nichts anderes ist als die Linearkombination $x_1v_1 + \ldots + x_kv_k$ der Vektoren v_j mit den Komponenten x_j des Vektors $\mathbf{x} \in \mathbb{R}^k$ als Koeffizienten. Die Überprüfung der linearen Unabbhängigkeit 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 $\mathbf{x} = 0$, 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 nichttriviale 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 $e_1, \ldots, e_n$ 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 $A = (v_1\, v_2\, \ldots\, v_k)$ darauf hinaus, dass man zwei der Spaltenvektoren v_i, v_j $(i \neq j)$ vertauscht, oder einen Spaltenvektor v_j ersetzt durch rv_j mit $r \in \mathbb{R}_{\neq 0}$, oder v_i ersetzt durch $v_i + v_j$ (bzw. durch $v_i + sv_j$ bei einer Kombination der beiden letzten Operationen). Solche Operationen ändern aber offenbar nichts an der linearen Abhängigkeit oder Unabhängigkeit der Vektoren $v_1, \ldots, v_k$. 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 $v_1, v_2, \ldots, v_k$ $(k \leq n)$ sind genau dann linear unabhängig in $\mathbb{R}^n$, wenn man die Matrix $A = (v_1 \, v_2 \, \ldots, \, v_k)$ 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 $v_1, v_2, \ldots, v_k$ sind genau dann linear abhängig in $\mathbb{R}^n$, wenn man aus der Matrix $A = (v_1 \, v_2 \, \ldots, \, v_k)$ 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 $(j = 1 \ldots k)$ linear abhängig in $\mathbb{R}^n$, so gilt das auch für die Vektoren $\widetilde{v}_j$ in $\mathbb{R}^{n-m}$, die aus den v_j durch Streichen von m Komponenten mit vorgegebenen Positionen entstehen.*
- *Sind die Vektoren v_j $(j = 1 \ldots k)$ linear unabhängig in $\mathbb{R}^n$, so gilt das auch für die Vektoren $\widehat{v}_j$ in $\mathbb{R}^{n+m}$, die aus den v_j durch Hinzufügen von m Komponenten mit vorgegebenen Positionen und beliebigen Einträgen (bei jedem $\widehat{v}_j$) 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 $A = (v_1 \, \ldots \, v_k)$ eine $k \times k$–Untermatrix $\widetilde{A}$ 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 $\widetilde{A}\mathbf{x} = 0$ genau dann nur trivial lösbar ist, wenn die $k \times k$–Matrix $\widetilde{A}$ invertierbar ist, wenn also $\det(\widetilde{A}) \neq 0$ ist, können wir aus (3) und (5) folgendes **Determinantenkriterium für lineare (Un–)Abhängigkeit** von Vektoren in $\mathbb{R}^n$ folgern:

- *Die Vektoren $v_1, v_2, \ldots, v_k$ sind linear unabhängig in $\mathbb{R}^n$, wenn die Matrix $A = (v_1 \, v_2 \, \ldots \, v_k)$ mit den Spalten v_j eine (durch Streichen von $n-k$ beliebigen Zeilen in A entstandene) $k \times k$–Untermatrix $\widetilde{A}$ mit $\det(\widetilde{A}) \neq 0$ besitzt.*
- *Haben aber alle derartigen $k \times k$–Untermatrizen von A Determinante 0, so sind die Vektoren $v_1, v_2, \ldots, v_k$ sind linear abhängig in $\mathbb{R}^n$.*

Die erste Aussage ist klar, weil eben $\widetilde{A}\mathbf{x} = 0$ nur die triviale Lösung hat, wenn $\det(\widetilde{A}) \neq 0$ ist, und weil aus der linearen Unabhängigkeit der Spalten von $\widetilde{A}$ die der durch Komponentenhinzufügung entstehenden Spalten von A folgt. Die zweite Aussage gilt, weil man im Fall der linearen Abhängigkeit von $v_1, \ldots, v_k$ 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 $v_1, \ldots, v_k$ 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 Sörung 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 $\widetilde{A}$ die Untermatrix der ersten k Zeilen von A, so ist dann $\widetilde{A}+s\mathbb{I}_k$ die entsprechende Untermatrix zu den Vektoren v_j+se_j, und es gilt $\det(\widetilde{A}+s\mathbb{I}_k) \neq 0$ 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 $T : V \to W$ zwischen Vektorräumen liest man ab, dass die Bildvektoren $T(v_1), \ldots, T(v_k)$ 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 $T(v) = 0 \Longrightarrow v = 0$ äquivalent mit $u \neq v \Longrightarrow T(u) \neq T(v)$, 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 $v_1, \ldots, v_k$ 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 $j_1, \ldots, j_k$ 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.

$$\begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}0\\-4\end{pmatrix} \quad \text{linear unabhängig in } \mathbb{R}^2 \text{ und } \quad \begin{pmatrix}1\\0\\-1\end{pmatrix}, \begin{pmatrix}\frac{1}{2}\\\frac{2}{3}\\-\frac{1}{2}\end{pmatrix} \quad \text{linear unabhängig in } \mathbb{R}^3,$$

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

$$\begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}-2\\-4\end{pmatrix} \quad \text{linear abhängig in } \mathbb{R}^2 \text{ und } \quad \begin{pmatrix}1\\0\\-1\\\sqrt{2}\end{pmatrix}, \begin{pmatrix}\frac{1}{2}\\0\\-\frac{1}{2}\\\frac{1}{\sqrt{2}}\end{pmatrix} \quad \text{linear abhängig in } \mathbb{R}^4,$$

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

$$\begin{pmatrix}a\\b\end{pmatrix}, \begin{pmatrix}c\\d\end{pmatrix} \quad \text{linear unabhängig, genau wenn } ad - bc \neq 0 \text{ ist.}$$

Das besagt hier das Determinantenkriterium, weil $ad - bc = \det\left(\begin{smallmatrix}a & b\\c & d\end{smallmatrix}\right)$ ist. Und

$$v = \begin{pmatrix}r\\s\\t\end{pmatrix}, \; w = \begin{pmatrix}1\\2\\3\end{pmatrix} \quad \text{sind linear abhängig in } \mathbb{R}^3, \text{ genau wenn } r \in \mathbb{R},\ s = 2r,\ t = 3r\,;$$

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 $A\mathbf{x} = 0$ (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)

$$u = \begin{pmatrix} 1 \\ 1 \\ 2 \\ r \end{pmatrix}, \quad v = \begin{pmatrix} 2 \\ 0 \\ 6 \\ s \end{pmatrix}, \quad w = \begin{pmatrix} 3 \\ 1 \\ 7 \\ t \end{pmatrix} \quad \text{in } \mathbb{R}^4,$$

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 $\widetilde{w}$, so führt dieselbe Rechnung für $u, v, \widetilde{w}$ 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 $u, v, \widetilde{w}$ 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 $u, v, \widetilde{w}$ 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:

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ r & s & t-r-s \end{pmatrix}.$$

Diese Dreiecksmatrix hat keinen Nulleintrag auf der Diagonalen, wenn $t \neq r{+}s$ ist; in diesem Fall sind also $u, v, \widetilde{w}$ 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 $u, v, \widetilde{w}$ linear abhängig. Dasselbe 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 $u, v, \widetilde{w}$ genau im Fall $t = r{+}s$ linear abhängig sind. (Dann ist übrigens $\widetilde{w} = u{+}v$, 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 $A\mathbf{x} = 0$ 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* $e_1, \ldots, e_n$ 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 $e_1, \ldots, e_n$ 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 abhängigen 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 $\mathbb{R}^\infty$ aller unendlichen reellen Zahlenfolgen $(r_1, r_2, \ldots)$ 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 $(r_1, \ldots, r_n) \in \mathbb{R}^n$ macht. In $\mathbb{R}^\infty$ 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 $p(x) = c_n x^n + c_{n-1}x^{n-1} + \ldots + c_1 x + c_0$ einer reellen Variablen $x \in \mathbb{R}$ mit beliebigem Grad n und beliebigen Koeffizienten $c_0, \ldots, c_n \in \mathbb{R}$, $c_n \neq 0$ (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 $p_0, \ldots, p_{n-1}$ linear ausdrücken, d.h. $p_n = s_{n-1}p_{n-1} + \ldots + s_1 p_1 + s_0 p_0$ gälte mit gewissen Koeffizienten $s_0, s_1, \ldots, s_{n-1} \in \mathbb{R}$. Dies wäre aber eine für alle $x \in \mathbb{R}$ gültige algebraische Gleichung $x^n = s_{n-1}x^{n-1} + \ldots + s_1 x + s_0$, 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–parametrigen linearen Schar $v(r_1, \ldots, r_k)$ von Vektoren in einem Vektorraum V präzise fassen. Die Linearität der Schar bedeutet, dass $\mathbb{R}^k \ni \mathbf{r} = (r_1, \ldots, r_k) \mapsto v(\mathbf{r}) = v(r_1, \ldots, r_k) \in V$ 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 $v(r_1, \ldots, r_k) = \sum_{j=1}^k r_j v_j$ mit den Vektoren $v_j := v(e_j) \in V$. Eine k–parametrige lineare Schar in V ist also nichts anderes als die Bildung aller möglichen Linearkombinationen von k gegebenen Vektoren $v_1, \ldots, v_k \in V$. Dabei muss man die lineare Parametrisierung, das ist die Abbildung $(r_1, \ldots, r_k) \mapsto \sum_{j=1}^k r_j v_j$, 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 $(v_1, \ldots, v_k)$, 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 $u, \widetilde{u}$ auch die Summe $u + \widetilde{u}$ sowie alle Vielfachen ru $(r \in \mathbb{R})$.

(ii) Sind $v_1, \ldots, v_k$ gegebene Vektoren im Vektorraum V, so heißt die Menge aller Linearkombinationen $\sum_{j=1}^k r_j v_j$ mit Koeffizienten $r_1, \ldots, r_k \in \mathbb{R}$ der von den Vektoren $v_1, \ldots, v_k$ **aufgespannte Unterraum** U und wird $\mathrm{Span}(v_1, \ldots, v_k)$ oder $\mathbb{R}v_1 + \ldots + \mathbb{R}v_k$ notiert. Man sagt auch, dass die Vektoren $v_1, \ldots, v_k$ ein **Erzeugendensystem** des Unterraums U bilden und dass die Abbildung $\mathbb{R}^k \ni (r_1, \ldots, r_k) \mapsto \sum_{j=1}^k r_j v_j \in U$ eine ***k*–parametrige 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 $v_1, \ldots, v_k$ in (ii) ist tatsächlich ein Unterraum von V; denn mit $u = \sum_{j=1}^k s_j v_j \in U$ und $\widetilde{u} = \sum_{j=1}^k t_j v_j \in U$ sind auch $u + \widetilde{u} = \sum_{j=1}^k (s_j + t_j) v_j$ 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 $v_1, \ldots, v_k$ liegen. In der Literatur sind neben den in (ii) angegebenen Benennungen und Notationen noch andere verbreitet. So wird U auch als der von $v_1, \ldots, v_k$ *erzeugte Unterraum* oder als *lineare Hülle* der Vektoren $v_1, \ldots, v_k$ bezeichnet und $\mathrm{Lin}\{v_1, \ldots, v_k\}$ oder $\langle v_1, \ldots, v_k \rangle$ 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 $e_1, \ldots, e_n$; denn jeder Vektor $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ 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 $\mathbb{R}v_1 + \ldots + \mathbb{R}v_k$ beschreiben mit geeigneten Wahlen von $k \in \mathbb{N}_0$ und von $v_1, \ldots, v_k \in V$. 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 $\mathbf{a}_1, \ldots, \mathbf{a}_n$, so gilt $A\mathbf{x} = x_1\mathbf{a}_1 + \ldots + x_n\mathbf{a}_n$, 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 $\mathbf{x} = (x_1, \ldots, x_n)$; bei einer **parametrischen Darstellung des Unterraums** gibt man eine lineare Parametrisierung von U an oder, was auf dasselbe hinausläuft, ein Erzeugendensystem $\mathbf{a}_1, \ldots, \mathbf{a}_k$ 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 $\ell_h(\mathbf{x}) = c_{h1}x_1 + \ldots + c_{hn}x_n$ 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 $h = l{+}1 \ldots n$ äquivalent damit, dass $\mathbf{x}$ eine zu A konsistente rechte Seite ist, also ein Vektor im Unterraum $U = \mathbb{R}\mathbf{a}_1 + \ldots + \mathbb{R}\mathbf{a}_k$. Damit ist U beschrieben als Lösungsmenge des homogenen linearen Gleichungssystems $\sum_{i=1}^n c_{hi}x_i = 0$, $h = l{+}1 \ldots n$. (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 $T : V \to W$ 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 $T : V \to W$ wieder ein Unterraum von W, und für jeden Unterraum $\widetilde{U}$ von W ist das Urbild $T^{-1}\widetilde{U} := \{v \in V : T(v) \in \widetilde{U}\}$ ein Unterraum von V. Das folgt aus der Unterraum–Definition und der Definition der Linearität von Abbildungen auch sofort. Wir halten fest:

- *Für $T : V \to W$ linear ist das Bild $T(U)$ jedes Unterraums U von V ein Unterraum von W und das Urbild $T^{-1}\widetilde{U}$ jedes Unterraums $\widetilde{U}$ 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 $T : V \to W$ 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 $v_1, \ldots, v_k \in V$, so bilden die Bildvektoren $T(v_1), \ldots, T(v_k)$ 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 $T : V \to W$ führt ein gegebenes Erzeugendensystem von V in ein Erzeugendensysteme von W über, genau wenn sie sujektiv ist.*

(8) Die *Lösungsmenge eines inhomogenen linearen Gleichungssystems* $A\mathbf{x} = \mathbf{b}$ für n Unbekannte $\mathbf{x} = (x_1, \ldots, x_n)$ 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 $Ax = 0$ 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 $H_1 \cap \ldots \cap H_{i-1}$ 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 $T : V \to W$ 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 $w \notin \mathrm{Bild}(T)$ ist) oder ein zu $\mathrm{Kern}(T)$ paralleler affiner Unterraum von V (nämlich $T^{-1}\{w\} = v + \mathrm{Kern}(T)$ 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 $\mathbb{R}^\infty$ 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 $p_1, \ldots, p_k \in \mathcal{P}$ Polynome von einem Grad, der höchstens so groß ist wie das Maximum m der Grade von $p_1, \ldots, p_k$; also kann man kein Polynom $p \in \mathcal{P}$ vom Grad $> m$ als Linearkombination von $p_1, \ldots, p_k$ darstellen. Bei unendlichdimensionalen Vektorräumen V lassen sich nur die endlich erzeugten Unterräume in der Form $\mathbb{R}v_1 + \ldots + \mathbb{R}v_k$ 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 $2v_1, v_2, \ldots, v_k$ und $v_1+sv_2, v_2, \ldots, v_k$ offenbar denselben Unterraum wie $v_1, v_2, \ldots, v_k$. 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 $v_1, \ldots, v_k$ von Vektoren aus U äquivalent:*

(i) $v_1, \ldots, v_k$ *ist ein* **minimales Erzeugendensystem** *für U;*

(ii) $v_1, \ldots, v_k$ *ist ein* **maximales System linear unabhängiger Vektoren** *in U;*

(iii) $v_1, \ldots, v_k$ *ist ein* **linear unabhängiges Erzeugendensystem** *für U;*

(iv) *jeder Vektor $u \in U$ ist* **eindeutige Linearkombination** *von $v_1, \ldots, v_k$.*

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 $v_1, \ldots, v_k$ nicht mehr jedes $u \in U$ als Linearkombination der restlichen Vektoren darstellen kann. (ii) bedeutet, dass Hinzufügen eines einzigen Vektors aus U zu $v_1, \ldots, v_k$ 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 $(r_1, \ldots, r_k) \in \mathbb{R}^k$ ihrer Koeffizienten bzgl. der Basis bestimmen sich daher gegenseitig. Das bedeutet letztlich, dass man in einem k–dimensonalen 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)$\Longleftrightarrow$(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 $v_1, \ldots, v_k$ 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 $v_1, \ldots, v_k$ 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 $v_1, \ldots, v_k$ 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 $u_1, \ldots, u_l$ besitzt, so können wir darin ein linear unabhängiges Erzeugendensystem $v_j = u_{i_j}$, $j = 1 \ldots k$, 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 $v_1, \ldots, v_k$. Sind nun $u_1, \ldots, u_l$ 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 $R\mathbf{s} = 0$ nichttriviale Lösungen $\mathbf{s} = (s_1, \ldots, s_l)$, also $\sum_{h=1}^{l} r_{jh} s_h = 0$ für $j = 1 \ldots k$. 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 $u_1, \ldots, u_l$ 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 $v_1, \ldots, v_k$ von U den linearen Parametrisierungen $\mathbb{R}^k \ni (r_1, \ldots, r_k) \mapsto \sum_{j=1}^{k} r_j v_j$ 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 eines maximalen linear unabhängigen Systems $v_1, \ldots, v_k$ in U gibt eine zweite Interpretation der Dimension: Jedes solche System definiert eine sog. *Fahne* in U, d.h. eine aufsteigende Folge von Unterräumen $\{0\} = U_0 \subset U_1 \subset U_2 \subset \ldots \subset U_k = U$ von denen jeder echter Unterraum des folgenden ist, nämlich $U_j := \mathbb{R}v_1 + \ldots + \mathbb{R}v_j$. 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 eine 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 $1, 2, 3, \ldots$

(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 $\mathbb{R}^\infty$ 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 $v_1, \ldots, v_n$, so kann man zu jedem Vektor $v \in V$ seinen eindeutig bestimmten **Koeffizientenvektor** $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ 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 seine Koeffizientenvektor zuordnet, ist dann linear und hat die lineare Umkehrabbildung $\mathbb{R}^n \ni (x_1, \ldots, x_n) \mapsto \sum_{j=1}^n x_j v_j \in V$, die jedem gegebenen Koeffizientenvektor die zugehörige Linearkombination der Basisvektoren zuordnet.

Man nennt eine lineare Abbildung $T : V \to W$ 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 $T^{-1} : W \to V$ existiert dann und ist automatisch auch linear.) Hat man einen Isomorphismus T zwischen zwei Vektorräumen V, W, so kann man damit alle Aufgabestellungen 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 endlichdimesionalen 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 $v_1, \ldots, v_n$ ab! Wählt man eine andere Basis $\widetilde{v}_1, \ldots, \widetilde{v}_n$, so hat derselbe Vektor v bzgl. dieser zweiten Basis im Allgemeinen einen *anderen* Koeffizientenvektor $\widetilde{\mathbf{x}}$. Um den Unterschied zu berechnen, stellen wir die Vektoren der zweiten Basis als Linearkombinationen der ersten Basis dar, $\widetilde{v}_j = \sum_{i=1}^n c_{ij}v_i$ für $j = 1\ldots n$, und erhalten so die sog. **Übergangsmatrix** $C = (c_{ij}) \in \mathbb{R}^{n\times n}$ von der ersten zur zweiten Basis. Aus $\sum_{i=1}^n x_i v_i = v = \sum_{j=1}^n \widetilde{x}_j \widetilde{v}_j = \sum_{i=1}^n (\sum_{j=1}^n \widetilde{x}_j c_{ij}) v_i$ und der linearen Unabhängigkeit der v_i erhalten wir dann $x_i = \sum_{j=1}^n c_{ij}\widetilde{x}_j$ für $i = 1 \ldots n$, also in Matrix-Schreibweise $\mathbf{x} = C\widetilde{\mathbf{x}}$. Da hier $\mathbf{x}$ und $\widetilde{\mathbf{x}}$ beliebige Vektoren in $\mathbb{R}^n$ sein können und sich gegenseitig eindeutig bestimmen, muss C invertierbar sein, und wir sehen:

- *Den Koeffizientvektor $\widetilde{\mathbf{x}}$ 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,* $\widetilde{\mathbf{x}} = C^{-1}\mathbf{x}$.

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 $\widetilde{v}_j = \sum_{i=1}^n c_{ij}v_i$ und *nicht* $\widetilde{v}_j = \sum_{i=1}^n c_{ji}v_i$)! Die Argumentation zeigt übrigens auch, dass ein System von n beliebigen Vektoren $\widetilde{v}_j = \sum_{i=1}^n c_{ij}v_i$ 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 $\widehat{v}_1, \ldots, \widehat{v}_n$ und die Übergangsmatrix D von den $\widetilde{v}_j$ zu den $\widehat{v}_h$ 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 Paradebespiel einer Basis ist natürlich die *kanonische Basis* $e_1, \ldots, e_n$ des n-dimensionalen Zahlenraums $\mathbb{R}^n$. Hier ist die fundamentale Basiseigenschaft unmittelbar klar: Jeder Vektor $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ ist Linearkombination der kanonischen Basisvektoren, und der Koeffizient bei e_j muss dabei die entsprechende Komponente x_j von $\mathbf{x}$ sein:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \ldots + x_n \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = \sum_{j=1}^{n} x_j e_j \,.$$

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 $\widetilde{v} = cv$ ($c \in \mathbb{R}_{\neq 0}$) 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 lineare 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 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 $C = \left(\begin{smallmatrix} 0 & 1 \\ -1 & -1 \end{smallmatrix}\right)$, 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 $\widetilde{\mathbf{x}} = C^{-1}\left(\begin{smallmatrix} 2 \\ -3 \end{smallmatrix}\right) = \left(\begin{smallmatrix} -1 & -1 \\ 1 & 0 \end{smallmatrix}\right)\left(\begin{smallmatrix} 2 \\ -3 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right)$. (Das kann man natürlich direkt sehen.)

Die affine Ebene $M := \{\mathbf{x} \in \mathbb{R}^3 : x_1 + x_2 + x_3 = 1\}$ 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 $u_1, \ldots, u_l$. Man geht dann diese Vektoren in aufsteigender Nummerierung durch und streicht diejenigen, die Null sind oder Linearkombination der vorhergehenden, und behält eine Basis $v_1, \ldots, v_k$ 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 $A\mathbf{x} = 0$, 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 $a_1x_1 + \ldots + a_nx_n = 0$ (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 $(0, \ldots, 0, -a_j/a_i, 0, \ldots, 0, 1, 0, \ldots, 0) \in \mathbb{R}^n$ 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 $\mathbb{R}^3$:

$$\left.\begin{array}{rrr} x_1 & & -x_3 = 0 \\ x_1 & +2x_2 & -x_3 = 0 \end{array}\right\} \text{ hat Koeff.–Matrix } \begin{pmatrix} 1 & 0 & -1 \\ 1 & 2 & 1 \end{pmatrix} \text{ mit Z.–St.–Form } \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 2 \end{pmatrix}.$$

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 $\mathbb{R}^4$:

$$\left.\begin{array}{rrrr} x_1 & & +x_3 & +x_4 = 0 \\ & x_2 & & +x_4 = 0 \end{array}\right\}, \text{ Koeff.–Matrix } \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix} \text{ hat schon Z.–St.–Form.}$$

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. ■

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 $\widetilde{A}$ 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 $\widetilde{A}$, so dass k auch die maximale Zahl der unabhängigen Spalten von $\widetilde{A}$ 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 Zeilen–Operationen 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:

$$\textbf{Zeilenrang} = \textbf{Spaltenrang}.$$

Diese Zahl wird infolgedessen einfach der **Rang der Matrix** A genannt und $\mathrm{Rang}(A)$ 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:

$$\text{Rang}(A) \leq \max\{\text{Zeilenzahl}(A), \text{Spaltenzahl}(A)\} = \max\{m, n\} \quad \text{für} \quad A \in \mathbb{R}^{m \times n},$$
$$\text{Rang}(BA) \leq \min\{\text{Rang}(B), \text{Rang}(A)\} \quad \text{für} \quad A \in \mathbb{R}^{m \times n},\ B \in \mathbb{R}^{l \times m}.$$

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 $A\mathbf{x} = 0$ aus $n-k$ Vektoren (wenn $k < n$). Der Lösungsraum hat somit die Dimension $\dim L = n-k$, und dieses Ergebnis ist der sog. **Rangsatz:**

$$\mathbf{Rang}(\boldsymbol{A}) + \mathbf{Defekt}(\boldsymbol{A}) = \mathbf{Spaltenzahl}(\boldsymbol{A})$$

oder

$$\mathbf{dim}\,\boldsymbol{K} + \mathbf{dim}\,\boldsymbol{L} = \boldsymbol{n} \quad \textbf{für} \quad \boldsymbol{A} \in \mathbb{R}^{m \times n},$$

wobei die Dimension des Lösungsraums L der **Defekt** der Matrix A genannt wird und K den Raum der zu A konsistenten rechten Seiten bezeichet. 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 $T: V \to W$ 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:

$$\mathbf{Rang}(\boldsymbol{T}) + \mathbf{Defekt}(\boldsymbol{T}) = \mathbf{dim\,Def}(\boldsymbol{T}),$$

bzw. ausführlicher

$$\mathbf{dim\,Bild}(\boldsymbol{T}) + \mathbf{dim\,Kern}(\boldsymbol{T}) = \mathbf{dim}\,\boldsymbol{V} \quad \text{für lineare } T: V \to W.$$

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 soviele Dimensionen kleiner wie die Dimension des Anullierungsraums von T im Definitionsbereich V.

Zum Beweis wählen wir eine Basis $w_1, \ldots, w_k$ 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 $v_{k+1}, \ldots, v_n$ 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 $v_1, \ldots, v_n$ 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 $j = 1 \ldots k$ und $T(v_j) = 0$ für $j > k$ zunächst $r_1 = \ldots = r_k = 0$ wegen der Basiseigenschaft von $w_1, \ldots, w_k$ und sodann auch $r_{k+1} = \ldots = r_n = 0$ wegen der Basiseigenschaft von $v_{k+1}, \ldots, v_n$. Außerdem erzeugen $v_1, \ldots, v_n$ 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 $v_{k+1}, \ldots, v_n$. Somit ist $v_1, \ldots, v_n$ 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 $\{v \in U : T(u) = 0\} = U \cap T^{-1}\{0\}$, 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 $v_1, \ldots, v_n$ bzw. $w_1, \ldots, w_m$ fixiert, so kann man jeder lineare Abbildung $T: V \to W$ 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 $T: V \to W$ ist eindeutig bestimmt durch ihre Werte auf den Vektoren einer gegebenen Basis $v_1, \ldots, v_n$ 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 $T: V \to W$ die Matrix $A \in \mathbb{R}^{m \times n}$ zu, deren Spalten die Koeffizientenvektoren der Bilder $T(v_j)$ bzgl. einer gegebenen Basis $w_1, \ldots, w_m$ 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. $w_1, \ldots, w_m$, wenn $\mathbf{x}$ der Koeffizientenvektor von $v \in V$ bzgl. $v_1, \ldots, v_n$ 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 $w_1, \ldots, w_m$ 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 $\widetilde{v}_j$ in V und $\widetilde{w}_i$ in W übergeht. Für die Matrix $\widetilde{A}$ bzgl. der neuen Basen gilt $\widetilde{A}C^{-1}\mathbf{x} = D^{-1}A\mathbf{x}$, wenn C bzw. D die Übergangsmatrizen zwischen den Basen in V bzw. in W sind, weil ja $\widetilde{\mathbf{x}} = C^{-1}\mathbf{x}$ die Änderung der Koeffizientenvektoren von Elementen in V beschreibt und $\widetilde{\mathbf{y}} = D^{-1}\mathbf{y}$ die Änderung der Koeffizientenvektoren von Elementen in W. Somit ist

$$\widetilde{A} = D^{-1}AC \quad \text{die geänderte Matrixdarstellung nach dem Basiswechsel.}$$

Da die Matrizen C und D invertierbar sind, folgt insbesondere, dass A und $\widetilde{A}$ 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 $\widetilde{v}_j$ geänderte Matrixdarstellung $\widetilde{A} = C^{-1}AC$, wobei C die Übergangsmatrix zur neuen Basis ist. $\widetilde{A}$ 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** $T\colon V \to V$ 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. $T(v) = \lambda v_j$ 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 $C = \left(\begin{smallmatrix} 1 & -1 \\ 1 & 1 \end{smallmatrix}\right)$ mit der Inversen $C^{-1} = \frac{1}{2}\left(\begin{smallmatrix} 1 & 1 \\ -1 & 1 \end{smallmatrix}\right)$. Die Matrixdarstellung von T bzgl. der neuen Basis ist dann

$$C^{-1}AC = \tfrac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 5 & 3 \\ 3 & 5 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 8 & 0 \\ 0 & 2 \end{pmatrix}.$$

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 $v_1, \ldots, v_k$ 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 $u_1, \ldots, u_k$ 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:

$$u_i \bullet u_j = d_{ij} := \begin{cases} 0 & \text{für } i \neq j \\ 1 & \text{für } i = j \end{cases} \quad (1 \leq i, j \leq k). \qquad \blacksquare$$

Hierbei ist $\mathbf{x} \bullet \mathbf{y} = x_1 y_1 + \ldots + x_n y_n$ das in 3.3 eingeführte (Euklidische) **Skalarprodukt** von Vektoren $\mathbf{x} = (x_1, \ldots, x_n)$ und $\mathbf{y} = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$. Die Bedingung $\mathbf{x} \bullet \mathbf{y} = 0$ 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 $\sqrt{\mathbf{x} \bullet \mathbf{x}}$, 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 $u_i \bullet u_j = 0$ ist für $i \neq j$, also verschiedene Vektoren des Systems orthogonal zueinander, und zweitens $u_i \bullet u_i = 1$, 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 $v \bullet w$ zuordnet mit folgenden Eigenschaften:

(*Positivität*) $\quad v \bullet v > 0$ wenn $v \neq 0$ (und $0 \bullet 0 = 0$),

(*Symmetrie*) $\quad v \bullet w = w \bullet v$,

(*Bilinearität*) $\quad (rv) \bullet w = r(v \bullet w)$ für $r \in \mathbb{R}$, $\quad (v + \tilde{v}) \bullet w = v \bullet w + \tilde{v} \bullet w$.

Die letzten Gleichungen gelten wegen der Symmetrie natürlich analog für den zweiten Faktor, d.h. $v \bullet w$ 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 $|v| := \sqrt{v \bullet v}$ zu und hat folgende Grundeigenschaften:

(*Positivität*) $\quad |v| > 0$ wenn $v \neq 0$ (und $|0| = 0$),

(*absolute Homogenität*) $\quad |rv| = |r||v|$ für $r \in \mathbb{R}$,

(*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 $|v+w|^2 = (v+w)\bullet(v+w) = v\bullet v + w\bullet v + v\bullet w + w\bullet w = |v|^2 + 2v\bullet w + |w|^2$ berechnet, auf die sog. **Cauchy–Schwarz–Ungleichung** hinaus:

$$|v \bullet w| \leq |v||w|.$$

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 $|v+tw|^2 = |v|^2+2tv\bullet w+t^2|w|^2$ der Variablen $t \in \mathbb{R}$ nur nichtnegative Werte hat. Daher kann die Diskriminante nicht positiv sein, und das heißt $(v\bullet w)^2 - |v|^2|w|^2 \leq 0$. 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 $v \bullet w = 0$ 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 $u_1, \ldots, u_k$ 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 $v = \sum_{j=1}^k (u_j \bullet v) u_j$;*

(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*

$$v \bullet w = \sum_{j=1}^k (u_j \bullet v)(u_j \bullet w) \qquad \text{und} \qquad |v|^2 = \sum_{j=1}^k (u_j \bullet v)^2;$$

(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 $u_i \bullet v = \sum_{j=1}^k r_j u_i \bullet u_j$, und weil $u_i \bullet u_j = 0$ ist für $j \neq i$ sowie $u_i \bullet u_i = 1$, folgt $u_i \bullet v = r_i$. 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) $v \bullet w = \sum_{j=1}^k (u_j \bullet v) u_j \bullet w = \sum_{j=1}^k (u_j \bullet v) \sum_{i=1}^k (u_i \bullet w) u_j \bullet u_i$. Berücksichtigt man wieder $u_j \bullet u_i = 0$ für $j \neq i$ und $u_j \bullet u_j = 1$, 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 $v_1, \ldots, v_k$ von U rekursiv eine Orthonormalbasis $u_1, \ldots, u_k$ berechnet und für praktische Anwendungen wichtig ist. Deshalb beschreiben wir nun dieses sog. **Gram–Schmidt–Orthonormalisierungsverfahren**:

$$u_1 = \frac{v_1}{|v_1|},\ u_2 = \frac{v_2 - (u_1 \bullet v_2) u_1}{\text{Norm des Zählers}}, \ldots,\ u_j = \frac{v_j - (u_1 \bullet v_j) u_1 - \ldots - (u_{j-1} \bullet v_j) u_{j-1}}{\text{Norm des Zählers}}, \ldots$$

Man erhält also u_1 durch Normierung des ersten Basisvektors v_1. Vor dem j–ten Schritt des Verfahrens hat man bereits die Vektoren $u_1, \ldots, u_{j-1}$ des Orthonormalsystems berechnet, die denselben Unterraum aufspannen wie $v_1, \ldots v_{j-1}$. Beim j–ten Schritt subtrahiert man zunächst von v_j die Linearkombination $\sum_{i=1}^{j-1} (u_i \bullet v_j) u_i$ der bereits berechneten Vektoren, also den Vektor aus $\mathbb{R}u_1 + \ldots + \mathbb{R}u_{j-1}$, der dieselben Skalarprodukte mit $u_1, \ldots, u_{j-1}$ hat wie v. Dies bedeutet, dass der Differenzvektor orthogonal zu $u_1 \ldots, u_{j-1}$ ist (er ist gewissermaßen die Komponente von v_j senkrecht zu $u_1, \ldots, u_{j-1}$). Er ist auch nicht der Nullvektor; denn sonst wäre v_j eine Element von von $\mathbb{R}u_1 + \ldots + \mathbb{R}u_{j-1} = \mathbb{R}v_1 + \ldots + \mathbb{R}v_{j-1}$, was der linearen Unabhängigkeit von $v_1, \ldots, v_j$ widerspräche. Daher kann man den Differenzvektor noch normieren durch Multiplikation mit dem Reziproken seiner Norm und erhält so den zu $u_1, \ldots, u_{j-1}$ orthogonalen Einheitsvektor u_j, der zusammen mit $u_1, \ldots, u_{j-1}$ denselben Unterraum aufspannt wie $v_1, \ldots, v_j$. Das Verfahren kann bis $j = k$ fortgesetzt werden und liefert die gesuchte Orthonormalbasis in $U = \mathbb{R}v_1 + \ldots + \mathbb{R}v_k$.

BEISPIELE: **(1)** Die *kanonische Basis* $e_1, \ldots, e_n$ 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

$$\widetilde{u}_2 = \begin{pmatrix}-3\\0\\1\end{pmatrix} - \left[u_1 \bullet \begin{pmatrix}-3\\0\\1\end{pmatrix}\right] u_1 = \begin{pmatrix}-3\\0\\1\end{pmatrix} - \frac{-6}{5}\begin{pmatrix}2\\1\\0\end{pmatrix} = \frac{1}{5}\begin{pmatrix}-3\\6\\5\end{pmatrix}.$$

Dieser Vektor muss jetzt noch normiert, also mit dem Reziproken seiner Norm multipliziert werden. Das ergibt

$$u_2 = \frac{\widetilde{u}_2}{|\widetilde{u}_2|} = \frac{1}{\sqrt{(-3^2+6^2+5^2}}\begin{pmatrix}-3\\6\\5\end{pmatrix} = \frac{1}{\sqrt{70}}\begin{pmatrix}-3\\6\\5\end{pmatrix},$$

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 $\widetilde{u}_3 = v - (u_1 \bullet v)u_1 - (u_2 \bullet v)u_2$ 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

$$u_3 = \frac{1}{\sqrt{1^2+(-2)^2+3^2}}\begin{pmatrix}1\\-2\\3\end{pmatrix} = \frac{1}{\sqrt{14}}\begin{pmatrix}1\\-2\\3\end{pmatrix}.$$

■

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 $U^\perp := \{v \in V : v \bullet u = 0 \text{ für alle } u \in U\}$ 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**

$$v = u + u^\perp \quad \textit{mit} \quad u \in U \quad \textit{und} \quad u^\perp \in U^\perp.$$

Hat V endliche Dimension, so gilt

$$\dim U + \dim U^\perp = \dim V \quad \textit{und} \quad (U^\perp)^\perp = U.$$

Beweis: Mit der Bilinearität des Skalarprodukts sieht man sofort, dass mit v, $\widetilde{v}$ auch rv und $v+\widetilde{v}$ 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 $u_1, \ldots, u_k$ 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 $v \bullet u_j = u \bullet u_j$ gelten muss für $j = 1 \ldots k$. Da die Koeffizienten von u bzgl. der Orthonormalbasis die Skalarprodukte mit den Basisvektoren sind, folgt $u = \sum_{j=1}^k (u_j \bullet v)u_j$. 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 $u := \sum_{j=1}^k (u_j \bullet v)u_j$. 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 $u_1, \ldots, u_k$ durch eine Orthonormalbasis $u_{k+1}, \ldots, u_n$ 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 $u_1, \ldots, u_n$ 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 $u_{k+1}, \ldots, u_n$ orthogonal ist, wenn $s_{k+1} = \ldots = s_n = 0$ gilt, wenn also die Linearkombination in U liegt.

Allgemein heißt ein Unterraum $\widetilde{U}$ ein **Komplement** des Unterraums U eines Vektorraums V, wenn jeder Vektor $v \in V$ eindeutig zerlegbar ist $v = u + \widetilde{u}$ mit $u \in U$ und $\widetilde{u} \in \widetilde{U}$. Man schreibt symbolisch $V = U \oplus \widetilde{U}$ für diesen Sachverhalt und nennt u bzw. $\widetilde{u}$ die **Komponenten** von v **bzgl. der Zerlegung**. Im endlichdimensionalen Fall ist äquivalent, dass sich Basen von U und $\widetilde{U}$ 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 $(u-u') \bullet u^\perp = 0$ 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 $u_1, \ldots, u_k$ von U:

$$\sum_{j=1}^{k} (u_j \bullet v) u_j \quad \text{ist die } \textbf{Orthogonalprojektion von } v \textbf{ auf } U = \mathbb{R}u_1 + \ldots + \mathbb{R}u_k.$$

BEISPIELE (Anwendungen auf lineare Gleichungssysteme):

(1) Ist $A\mathbf{x} = 0$ ein homogenes lineares Gleichungssystem und sind $\mathbf{a}_i = (a_{i1}, \ldots, a_{in})$ die Zeilenvektoren der Koeffizientenmatrix, so kann man die i-te Gleichung schreiben $\mathbf{a}_i \bullet \mathbf{x} = a_{i1}x_1 + \ldots + a_{in}x_n = 0$. 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 $G = (\mathbb{R}\binom{a}{b})^\perp$ 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 $G = \mathbb{R}\binom{-b}{a}$ und $G^\perp = \mathbb{R}\binom{a}{b}$ 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 orthonormalen 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 $A\mathbf{x} = 0$ 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 $B\mathbf{x}= 0$ 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 Grade das Gleichungssystem.

$$\begin{array}{rrrl} x_1 & & +x_3 & = 0 \\ -2x_1 & +x_2 & & = 0 \end{array}$$

(6) Zur der durch
$$\begin{cases} x_1 \quad +x_3 \quad = 0 \\ \quad x_2 + x_3 + x_4 = 0 \end{cases}$$
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 $\left(\begin{smallmatrix}1&0&1&0&|&0\\0&1&1&1&|&0\end{smallmatrix}\right)$ 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
$$\begin{cases} -x_1 - x_2 + x_3 \quad = 0 \\ \quad -x_2 \quad + x_4 = 0 \end{cases}$$
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
$$\begin{cases} \quad x_1 \quad + x_3 \quad = 0 \\ -x_1 + x_2 \quad + x_4 = 0 \end{cases}$$
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 $v_i \bullet v_j$. 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 $A, \widetilde{A}, \ldots$ bilden bzgl. der Matrix–Multiplikation eine Untergruppe der Gruppe aller invertierbaren $n \times n$–Matrizen (d.h. A^{-1} und $A\widetilde{A}$ sind wieder orthogonal), genannt **orthogonale Gruppe** von $\mathbb{R}^n$. Unter Beachtung der Symmetrie von $A^\mathsf{T} A$ und von $\mathbb{I}_n$ enthält die Matrix–Gleichung $A^\mathsf{T} A = \mathbb{I}_n$ 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 Spaltenvektorpaare 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 $D_{\alpha_1}, \ldots, D_{\alpha_m}$ 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 $\alpha_1, \ldots, \alpha_m$ 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 $T : \mathbb{R}^n \to \mathbb{R}^n$, $T(\mathbf{x}) = A\mathbf{x}$, die durch eine orthogonale Matrix A definiert wird, führt die kanonische Basis in eine Orthonormalbasis $v_1, \ldots, v_n$ 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 $T(\mathbf{x}) \bullet T(\mathbf{y}) = \mathbf{x} \bullet \mathbf{y}$ 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 $(A\mathbf{x}) \bullet (A\mathbf{y}) = (A\mathbf{x})^\mathsf{T} A\mathbf{y} = \mathbf{x}^\mathsf{T} A^\mathsf{T} A\mathbf{y} = \mathbf{x}^\mathsf{T} \mathbb{I}\mathbf{y} = \mathbf{x}^\mathsf{T} \mathbf{y} = \mathbf{x} \bullet \mathbf{y}$.)

- *Eine $n \times n$–Matrix ist orthogonal, genau wenn die lineare Abbildung $T(\mathbf{x}) = A\mathbf{x}$ Skalarprodukte erhält, d.h. $(A\mathbf{x}) \bullet (A\mathbf{y}) = \mathbf{x} \bullet \mathbf{y}$ 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 $|\mathbf{x}| = \sqrt{\mathbf{x} \bullet \mathbf{x}}$), Orthogonalität $\mathbf{x} \perp \mathbf{y} \iff T(\mathbf{x}) \perp T(\mathbf{y})$ (da $\mathbf{x} \perp \mathbf{y} \iff \mathbf{x} \bullet \mathbf{y} = 0$) und die Winkel $\sphericalangle(\mathbf{x}, \mathbf{y})$ zwischen Einheitsvektoren (die definiert sind durch die Bedingung, dass ihr Kosinus gleich $\mathbf{x} \bullet \mathbf{y}$ ist). Man nennt solche Abbildungen $T : \mathbb{R}^n \to \mathbb{R}^n$ 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 $T : V \to W$ zwischen Vektorräumen mit Skalarprodukt heißt **orthogonale Abbildung**, wenn sie Skalarprodukte erhält, wenn also gilt $T(v) \bullet T(\widetilde{v}) = v \bullet \widetilde{v}$ für alle $v, \widetilde{v} \in V$. 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 $T : V \to W$, 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 $T : V \to W$ zwischen Vektorräumen mit Skalarprodukt folgende Aussagen äquivalent:

(i) *T ist orthogonale Abbildung, erhält also Skalarprodukte, $T(v) \bullet T(w) = v \bullet w$;*

(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*:

$$|v + w|^2 - |v - w|^2 = |v|^2 + 2v \bullet w + |w|^2 - (|v|^2 - 2v \bullet w + |w|^2) = 4\, v \bullet w.$$

Wenn man überlegt, welche linearen Abbildungen $T : V \to W$ *Orthogonalitäts–erhaltend* sind, also $v \perp w \Longrightarrow T(v) \perp T(w)$ erfüllen, so findet man, dass dies genau die Vielfachen $T = rS$ von orthogonalen Abbildungen $S : V \to W$ 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 $V \to W$ 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 $u_1, \ldots, u_n$ 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 $e_1, \ldots, e_n$ 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 $u_1, \ldots, u_n$ besteht aus Eigenvektoren von A und die Zahlen λ_i sind die zugehörigen Eigenwerte (siehe 3.4). Wenn dies gilt und wenn die Basis $u_1, \ldots, u_n$ orthonormal ist, so haben wir (wegen $\delta_{ij} = 0$ für $i \neq j$)

$$u_i \bullet (Au_j) = u_i \bullet (\lambda_j u_j) = \lambda_j \delta_{ij} = \lambda_i \delta_{ij} = (\lambda_i u_i) \bullet u_j = (Au_i) \bullet u_j,$$

und für Linearkombinationen $v = \sum_{i=1}^n r_i u_i$ und $w = \sum_{i=1}^n s_j u_j$ daher:

$$v \bullet (Aw) = \sum_{i=1}^n \sum_{j=1}^n r_i s_j \; u_i \bullet (Au_j) = \sum_{i=1}^n \sum_{j=1}^n r_i s_j (Au_i) \bullet u_j = (Av) \bullet w.$$

Für die Einträge der Matrix A folgt insbesondere

$$a_{ij} = e_i \bullet (Ae_j) = (Ae_i) \bullet e_j = a_{ji},$$

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 $u_1, \ldots, u_n$ von $\mathbb{R}^n$ aus Eigenvektoren von A, also $Au_i = \lambda_i u_i$ für gewisse Eigenwerte $\lambda_i \in \mathbb{R}$ und $i = 1 \ldots n$.

Ist B die orthogonale Matrix mit den u_i als Spalten, so ist also dann $B^{-1}AB = B^{\mathsf{T}}AB$ 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 $u\bullet(Av) = (Au)\bullet v = \lambda u\bullet v$ 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 $\det(t\mathbb{I}_n - A) = 0$, 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 $\mathbf{x}+\mathring{\imath}\mathbf{y} \neq 0$ hat mit $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ (und der imaginären Einheit $\mathring{\imath}$, also $\mathring{\imath}^2 = -1$). Die mit der Symmetrie der Matrix A äquivalente Gleichung $v\bullet(Aw) = (Av)\bullet w$ gilt nun aber auch für Vektoren mit komplexen Einträgen. Wählen wir insbesondere $v := \mathbf{x}+\mathring{\imath}\mathbf{y}$ und $w := \overline{v} = \mathbf{x}-\mathring{\imath}\mathbf{y}$, so folgt: $\lambda v\bullet\overline{v} = (Av)\bullet\overline{v} = v\bullet(A\overline{v}) = v\bullet(\overline{Av}) = v\bullet(\overline{\lambda v}) = \overline{\lambda} v\bullet\overline{v}$. (Hier bezeichnet $\overline{z} = x - \mathring{\imath}y$ die konjugiert komplexe Zahl zu $z = x + \mathring{\imath}y$ für $x, y \in \mathbb{R}$.) Da nun $v\bullet\overline{v} = |\mathbf{x}|^2 + |\mathbf{y}|^2$ reell und positiv ist, folgt $\lambda = \overline{\lambda}$, 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 $\mathbf{x}\bullet(A\mathbf{x})$ einen kleinstmöglichen Wert hat. Eine solche Stelle existiert, weil die quadratische Funktion $\mathbf{x} \mapsto \mathbf{x}\bullet(A\mathbf{x}) = \sum_{i,j=1}^n a_{ij}x_ix_j$ 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 $u\bullet(Au) \leq \frac{1}{1+t^2}(u+tv)\bullet(Au+tAv) = \frac{1}{1+t^2}[u\bullet(Au)+tu\bullet(Av)+tv\bullet(Au)+t^2v\bullet(Av)]$, weil $\frac{1}{\sqrt{1+t^2}}(u+tv)$ auch ein Einheitsvektor ist. Für kleine $t > 0$ kann das nur sein, wenn $u\bullet(Av) + v\bullet(Au) = 0$ ist; denn die Terme mit t^2 sind vernachlässigbar, wenn $0 < t \ll 1$. Wegen der Symmetrie $u\bullet(Av) = (Au)\bullet v = v\bullet(Au)$ folgt dann $v\bullet(Au) = 0$ 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 $T: V \to W$ eine lineare Abbildung, so gibt es eine eindeutig bestimmte Abbildung $T^*: W \to V$ mit

$$w\bullet T(v) = T^*(w)\bullet v \quad \text{für alle} \quad v \in V,\ w \in W.$$

T^* heißt die zu T **adjungierte lineare Abbildung.** Sind Orthonormalbasen $v_1, \ldots, v_n$ bzw. $w_1, \ldots, w_m$ in V bzw. in W gegeben und hat T diesbezüglich die Matrixdarstellung A, also $a_{ij} = w_i\bullet T(v_j)$, so hat T^* bzgl. derselben Basen die transponierte Matrix $A^\top$ als Matrixdarstellung; denn es ist $v_j\bullet T^*(w_i) = T(v_j)\bullet w_i = a_{ij}$. Damit kann man auch die Existenz und Eindeutigkeit von T^* einsehen: Es muss $T^*(w) = \sum_{j=1}^n (T(v_j)\bullet w)v_j$ gelten, und durch diese Formel kann man T^* definieren und $T^*(w)\bullet v = w\bullet T(v)$ nachweisen.

Ist nun $T: V \to V$ 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 Übetragung durch einen orthogonalen Automorphismus von $\mathbb{R}^n$ auf V in folgender Form aussprechen (für endlichdimensionale V):

- *Ist $T: V \to V$ eine selbstadjungierte lineare Abbildung, so hat V eine Orthonormalbasis aus Eigenvektoren zu T.*

(2) Eine Funktion auf $\mathbb{R}^n$ der Form

$$Q(\mathbf{x}) = \sum_{i,j=1}^{n} a_{ij} x_i x_j \quad \text{für} \quad \mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$$

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.

$$S(\mathbf{x}, \mathbf{y}) := \mathbf{x} \bullet (A\mathbf{y}) = \sum_{i,j=1}^{n} a_{ij} x_i y_j \quad \text{für} \quad \mathbf{x},\, \mathbf{y} \in \mathbb{R}^n$$

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 jeweil 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 $Q(\mathbf{x}) = \mathbf{x} \bullet (A\mathbf{x})$ folgt, wenn man $\mathbf{x} = \sum_{i=1}^{n} (u_i \bullet x) u_i$ als Linearkombination einer Orthonormalbasis von Eigenvektoren u_i von A schreibt und $u_j \bullet (Au_i) = u_j \bullet (\lambda_i u_i) = \lambda_i \delta_{ij}$ berücksichtigt:

$$Q(\mathbf{x}) = \lambda_1 (u_1 \bullet \mathbf{x})^2 + \lambda_2 (u_2 \bullet \mathbf{x})^2 + \ldots + \lambda_n (u_n \bullet \mathbf{x})^2.$$

Dies bedeutet, dass Q darstellbar ist als Linearkombination von "vollständigen Quadraten", genauer von Quadraten linearer Funktionen $\ell_i(\mathbf{x}) = u_{1i} x_1 + \ldots + u_{ni} x_n$ 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 $Q(\mathbf{x}) = \mathbf{x} \bullet (A\mathbf{x})$ 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 $\mathbf{x} \bullet (A\mathbf{x}) > 0$ 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 $a_{ii} = e_i \bullet (Ae_i)$ 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 ($k = 1 \ldots n$). 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 $B'^{\mathsf{T}} A' B'$ Diagonalmatrix ist mit positiven Diagonaleinträgen $\lambda'_1, \ldots, \lambda'_{n-1}$. Es folgt:

$$\begin{pmatrix} B'^{\mathsf{T}} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ 0 \ldots 0 & 1 \end{pmatrix} A \underbrace{\begin{pmatrix} B' & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ 0 \ldots 0 & 1 \end{pmatrix}}_{=:\,B} = \begin{pmatrix} \lambda'_1 & 0 & \ldots & 0 & c_1 \\ 0 & & & & \\ \vdots & & \ddots & & \vdots \\ 0 & & & \lambda'_{n-1} & c_{n-1} \\ c_1 & & \ldots & c_{n-1} & a_{nn} \end{pmatrix}.$$

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 $\delta := a_{nn} - c_1^2/\lambda'_1 - \ldots - c_{n-1}^2/\lambda'_{n-1}$ berechnet werden. Dieser Ausdruck ist also positiv. Außerdem erhält man

$$C^{-1}\mathbf{x} \bullet (AC^{-1}\mathbf{x}) = \sum_{i=1}^{n-1} (\lambda'_i x_i^2 + 2c_i x_i x_n) + a_{nn} x_n^2 = \sum_{i=1}^{n-1} \lambda'_i (x_i + c_i x_n/\lambda'_i)^2 + \delta x_n^2,$$

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 $k = 1 \ldots n$. ■

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 $Q(\mathbf{x}) = \mathbf{x} \bullet (A\mathbf{x})$. An den Koeffizienten von Q liest man ab (beachte, dass $a_{ij} = a_{ji}$ die *Hälfte* des Koeffizienten von x_ix_j ist für $i < j$, wenn nur Produkte x_ix_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

$$\widetilde{u}_1 = \begin{pmatrix} 3 \\ 1 \\ -1 \end{pmatrix}, \quad \widetilde{u}_2 = \begin{pmatrix} 1 \\ -2-\sqrt{3} \\ 1-\sqrt{3} \end{pmatrix}, \quad \widetilde{u}_3 = \begin{pmatrix} 1 \\ -2+\sqrt{3} \\ 1+\sqrt{3} \end{pmatrix}$$

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 $\lambda(u \bullet v) = (Au) \bullet v = u \bullet (Av) = \mu(u \bullet v)$, 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 $\widetilde{u}_i$ also nur noch zu normieren. Resultat ist die folgende Orthonormalbasis des $\mathbb{R}^3$ aus Eigenvektoren von A:

$$\widetilde{u}_1 = \frac{1}{\sqrt{11}} \begin{pmatrix} 3 \\ 1 \\ -1 \end{pmatrix}, \quad \widetilde{u}_2 = \frac{1}{\sqrt{12+2\sqrt{3}}} \begin{pmatrix} 1 \\ -2-\sqrt{3} \\ 1-\sqrt{3} \end{pmatrix}, \quad \widetilde{u}_3 = \frac{1}{\sqrt{12-2\sqrt{3}}} \begin{pmatrix} 1 \\ -2+\sqrt{3} \\ 1+\sqrt{3} \end{pmatrix}.$$

Damit ergibt sich nun die folgende Darstellung der quadratischen Form als Summe "vollständiger Quadrate":

$$\begin{aligned}Q(\mathbf{x}) &= \lambda_1(u_1 \bullet \mathbf{x})^2 + \lambda_2(u_2 \bullet \mathbf{x})^2 + \lambda_3(u_3 \bullet \mathbf{x})^2 \\ &= \frac{5-\sqrt{3}}{\sqrt{12+2\sqrt{3}}}[x_1-(2+\sqrt{3})x_2+(1-\sqrt{3})x_3]^2 + \frac{5+\sqrt{3}}{\sqrt{12-2\sqrt{3}}}[x_1-(2-\sqrt{3})x_2+(1+\sqrt{3})x_3]^2.\end{aligned}$$

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. ∎