Vorwort

Dieses Buch ist aus einer Vorlesung hervorgegangen, die der ältere der beiden Autoren im Wintersemester 1987/88 an der Universität Göttingen gehalten hat, die vom jüngeren Autor ausgearbeitet und zunächst vom Mathematischen Institut der Universität Göttingen herausgegeben worden ist. Für die Veröffentlichung als Lehrbuch ist dieses Manuskript noch einmal grundlegend überarbeitet und an zahlreichen Stellen ergänzt worden. Wir hoffen, daß die Darstellung so ausführlich, interessant und klar geraten ist, daß sich das Buch nicht nur als Begleitlektüre zu einer entsprechenden Vorlesung, sondern auch zum Selbststudium, zum Nachlesen oder zur Vorbereitung auf Prüfungen eignet.

Auf die Stoffauswahl gehen wir in der Einleitung ausführlich ein. Hier sei nur bemerkt, daß einerseits der Geometrie und vor allem der Nahtstelle zwischen geometrischer Anschauung und mathematisch-logischer Formulierung ein großes Gewicht eingeräumt wird. Andererseits wird ein großer Teil der grundlegenden Theorie, wie er etwa auch von Studenten der Wirtschaftsmathematik oder Physik benötigt wird, in diesem Buch behandelt, auch wenn es nur den Inhalt des ersten Teils der zweisemestrigen Vorlesung „Analytische Geometrie und lineare Algebra“ wiedergibt. Breiten Raum nehmen die Behandlung linearer Gleichungssysteme (Gaußscher Algorithmus), die Determinantentheorie und die Eigenwerttheorie linearer Abbildungen in „euklidischen“ Vektorräumen (Hauptachsentransformation) ein.

Wir sind bemüht, schnell zu den Kernpunkten der jeweiligen Themenbereiche vorzustoßen und verzichten dabei bewußt auf größtmögliche Allgemeinheit. Zahlreiche Beispiele, die teilweise den Charakter von gelösten Übungsaufgaben haben, sollen das Verständnis neuer Begriffe oder Methoden unmittelbar fördern.

Wir haben dem Verlag und insbesondere den Herren Andreas Türk und Martin Reck zu danken für die jederzeit angenehme Zusammenarbeit. Herrn Peter Lerner danken wir für seine tatkräftige Mitwirkung bei der Erstellung der LaTeX-Dateien und Herrn Christian Meister für seine Hilfe bei der Durchsicht des Manuskripts.

Göttingen und Bayreuth, im März 1999

Hans Grauert
Hans-Christoph Grunau

Zu den Autoren

Hans Grauert wurde am 8. Februar 1930 in Haren (Ems) geboren. Er studierte Mathematik und Physik in Mainz, Münster und an der ETH Zürich. Er wurde 1958 an die Universität Göttingen als Nachfolger von C. L. Siegel berufen. Er verbrachte insgesamt drei Jahre in Amerika (u. a. am Institute for Advanced Study in Princeton), war aber auch in anderen Ländern wie vor allem in Frankreich (Institut des Hautes Études Scientifiques) und u. a. in Japan, China, der UdSSR und Indien. Seine wissenschaftlichen Arbeiten befassen sich mit Fragen der komplexen Analysis.

Hans-Christoph Grunau wurde am 23. März 1961 in Wuppertal geboren. Er studierte Mathematik in Marburg und Göttingen und hat sich 1996 an der Universität Bayreuth habilitiert. Sein wissenschaftliches Interesse gilt der reellen Analysis und insbesondere der Theorie der partiellen Differentialgleichungen. Auf diesem Gebiet arbeitet er mit Wissenschaftlern u. a. aus den Niederlanden, Spanien und Italien zusammen.

Inhaltsverzeichnis

Einleitung

In der Anfängervorlesung „Analytische Geometrie und lineare Algebra" ist es heute vielfach üblich, einfach mit der abstrakten Theorie der Vektorräume und evtl. sogar der Moduln über Ringen zu beginnen. Man hat das Bestreben, sofort von den mathematischen Grundbegriffen auszugehen und aus ihnen alles zusammenzubauen. Das soll zu einer logischen Übersichtlichkeit führen. Aber der Anfänger wird staunen, was es so alles gibt. Er kann sich bei manchem nur wenig denken.

In der Technik, aber auch in der Physik wird im allgemeinen erst nur der 3-dimensionale anschauliche Raum auftreten. Er hat noch keine Koordinaten. Man hat noch keine Tripel von reellen Zahlen, dafür aber gibt es eine Vorstellung von Geraden, Winkeln, Translationen, Drehungen. Vektoren gibt es. Sie sind Pfeile einer bestimmten Länge. Man kann mit ihnen rechnen. Warum geht man dann nicht von diesem Raum aus? Man kann die Vektorraumeigenschaften ableiten und so den Studenten zu den in der Mathematik natürlich notwendigen Abstraktionen hinführen. Die Abstraktionen werden dann als natürlich empfunden. Man kann dann auch die Existenz von rechtwinkligen Koordinaten zeigen und die Isomorphie mit dem Zahlenraum herstellen. Die Existenz einer solchen Isomorphie bedeutet, daß der anschauliche Raum die gleiche Gestalt wie der Zahlenraum hat.
Wer seinen Baum beschneiden will, wird nicht erst Koordinaten in seinen Garten legen.
Das hier vorgestellte Axiomensystem ist auch ein vollständiges Axiomensystem der euklidischen Geometrie. Anders als bei Hilbert werden jedoch keine Geraden, sondern Pfeilvektoren verwendet. Das ist praktischer. Die Pfeilvektoren finden ja überall Anwendung. Natürlich wird nicht die alte hyperbolische Geometrie mit eingeschlossen. Diese dürfte aber auch heute ziemlich ohne Bedeutung sein. Wenn man sie machen will, ist es ohnehin besser, gleich die Quotienten von Liegruppen zu untersuchen! Die nicht-euklidischen Geometrien werden hier gleich dadurch ausgeschlossen, daß man fordert, daß die Relation „parallel" eine Äquivalenzrelation ist und daß das Parallelogramm der Kräfte gilt.
Es sei noch erwähnt, daß die Griechen keinen Beweis des Satzes von Pythagoras geliefert haben, der den heutigen Anforderungen entspricht. Sie haben nämlich den Flächeninhalt benutzt, ohne ihn zu definieren. Auch hätten sie zeigen müssen, daß er invariant gegen Drehungen ist. An sich hat der Pythagoräische Satz mit dem Flächeninhalt nichts zu tun. Dieses Buch enthält einen exakten Beweis.

In dem ersten Kapitel dieses Buches wird zwar zunächst der n-dimensionale reelle Zahlenraum $\mathbb{R}^n$ als Menge der n-tupel von reellen Zahlen und der Begriff des abstrakten Vektorraums eingeführt, damit man die späteren Begriffe einordnen kann. Dann geht man jedoch im zweiten Kapitel sofort zur Anschaulichkeit über. Insbesondere wird die anschauliche Ebene betrachtet. Hier wird unser Axiomensystem der anschaulichen Geometrie soweit entwickelt, daß die anschauliche Ebene als zweidimensionaler reeller Vektorraum erkannt werden kann.

Hierauf aufbauend entwickeln wir im 3. Kapitel die Grundlagen der Theorie endlichdimensionaler (zunächst ausschließlich reeller) Vektorräume sowie der linearen Abbildungen zwischen diesen. Wir werden zeigen, daß durch Einführung geeigneter „Isomorphismen " (Koordinatensysteme) die Betrachtung linearer Abbildungen zwischen endlichdimensionalen Vektorräumen zurückgeführt werden kann auf die Betrachtung linearer Abbildungen $\mathbb{R}^n \rightarrow \mathbb{R}^m$. Diese sind dem ebenso einfachen wie wirkungsvollen Matrizenkalkül (dabei handelt es sich lediglich um die Manipulation von Zahlenschemata) zugänglich, den wir simultan und gleichberechtigt zur Theorie linearer Abbildungen entwickeln und mit dessen Hilfe wir zahlreiche theoretische Resultate herleiten werden. Breiten Raum nimmt die Behandlung linearer Gleichungssysteme bzw., was, wie wir in Kapitel 3.7 darlegen werden, dazu gleichwertig ist, allgemeiner s-dimensionaler Ebenen ein. In diesem Zusammenhang wird der Gaußsche Algorithmus vorgestellt, aus dem wir neben rechnerischem auch stets großen theoretischen Nutzen ziehen werden.

In Kapitel 4 sprechen wir speziellere Fragen der Vektorraumtheorie an. Im Mittelpunkt stehen dort duale Vektorräume und duale Abbildungen sowie in diesem Zusammenhang einige Besonderheiten unendlichdimensionaler Vektorräume. Die Idee, zusammen mit einem Vektorraum V die Gesamtheit aller linearen Abbildungen $V \rightarrow \mathbb{R}$ zu betrachten, ist grundlegend in vielen Bereichen der modernen Mathematik, führt zum einen von einer abstrakten Seite hin auf den Begriff der Orthogonalität und bietet zum anderen in allgemeinen Situationen einen Ersatz für bzw. eine Verallgemeinerung von Orthogonalität. Ausführlichere Erklärungen versuchen wir zu Beginn und im Verlauf von Kapitel 4.

Kapitel 5 ist der klassischen Determinantentheorie und dabei wegen ihrer enormen Wichtigkeit unter anderem der Cramerschen Regel gewidmet. Auch dieses Kapitel führt wegen der geometrischen Bedeutung der Determinante hin zu Kapitel 6, in dem unser in Kapitel 2 begonnenes Axiomensystem der anschaulichen Geometrie um diejenigen Axiome ergänzt wird, die Orthogonalität, Länge, Winkel, Drehung, u. ä betreffen.

Damit erhalten die anschauliche Ebene, $\mathbb{R}^2$ und allgemeiner $\mathbb{R}^n$ die volle Struktur „euklidischer Vektorräume". Das Vorhandensein eines Skalarprodukts erlaubt es uns, in Kapitel 7 etwa „othogonale", „selbstadjungierte" oder „normale" Abbildungen zu betrachten. Die starken Eigenschaften solcher speziellen Abbildungen erlauben die Herleitung sehr weitreichender Resultate („Hauptachsentransformation"), deren volle

Eleganz und Allgemeinheit sich aber nur in komplexen Vektorräumen erreichen läßt. Deshalb beginnt dieses Kapitel mit einer Einführung in komplexe Zahlen und komplexe Vektorräume; in der Hoffnung, daß vieles schon aus der gleichzeitig erlernten Analysis vertraut ist, fassen wir uns hier recht kurz. Für die angesprochenen Klassen von Matrizen werden wir orthogonale Koordinatensysteme konstruieren, so daß das Rechnen mit linearen Abbildungen in diesen Koordinaten so einfach wird wie das Rechnen mit reellen oder komplexen Zahlen. Als Beispiele werden wir Wurzeln und Exponentialfunktionen geeigneter linearer Abbildungen bestimmen. Ganz wesentliches Standbein dieses Kapitels ist die „Eigenwerttheorie linearer Abbildungen", die mit all ihren Weiterentwicklungen und Verallgemeinerungen grundlegend für viele Bereiche der modernen Mathematik und Physik bis hin zur Quantenmechanik ist.

Mit dieser Stoffauswahl hoffen wir:

- die geometrischen Wurzeln der linearen Algebra mit logischer Strenge freizulegen,
- die wichtigsten Elemente der Theorie reeller und komplexer Vektorräume und linearer Abbildungen zu entwickeln und
- stets den Blick schon ein wenig zu öffnen in Richtung auf Weiterentwicklungen im Bereich der Geometrie (allgemeine Ebenen etwa als einfachste Prototypen allgemeiner Mannigfaltigkeiten), der Analysis (unendlichdimensionale Vektorräume werden gestreift, um Appetit zu machen auf die Theorie linearer Abbildungen zwischen vollständigen normierten Vektorräumen, die „Funktionalanalysis") und der mathematischen Physik (etwa Lorentztransformationen oder Eigenwerttheorie).

Technische Hinweise zum Lesen des Buches erübrigen sich weitestgehend. Das Kästchen $\square$ kennzeichnet das Ende von Beweisen. Die Numerierung sollte selbsterklärend und vollkommen unzweideutig sein: Sätze, Definitionen, Beispiele, usw. werden unterschiedslos dreistellig numeriert, die erste Stelle gehört zum Kapitel und die zweite zum Teilkapitel (Abschnitt). Am Ende des Buches befinden sich ein ausführliches Symbolverzeichnis und ein Sachverzeichnis, so daß (hoffentlich!) jedes interessierende Thema im Text mühelos aufgefunden werden kann.

Um eventuellen Mißverständnissen vorzubeugen, sei schon hier erwähnt, daß wir n-tupel aus dem $\mathbb{R}^n$ bzw. $\mathbb{C}^n$ zwar grundsätzlich als Zeilen $\vec{b} = (b_1, \ldots, b_n)$ schreiben, um die Lesbarkeit vor allem des fortlaufenden Textes zu gewährleisten, daß wir im Rahmen des Matrizenkalküls darunter jedoch (zumeist stillschweigend) die entsprechenden Spaltenvektoren

$$\vec{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

verstehen.

1 Der Vektorraumbegriff

In diesem Kapitel werden einige Grundbegriffe der linearen Algebra eingeführt, die gleichzeitig fundamental für viele Bereiche der modernen Mathematik sind. Der Vektorraumbegriff ist maßgeschneidert für alle „linearen“ Modelle und ist selbst im „Nichtlinearen“ häufig der geeignete Rahmen.

Unser unmittelbares Interesse gilt aber der Untersuchung der anschaulichen Ebene bzw. des anschaulichen Raumes, der die Kapitel 2 und 6 gewidmet sein werden. Um die dort aufzuzeigenden mathematischen Strukturen einordnen zu können, stellen wir in den folgenden Abschnitten die entsprechenden Begriffe bereit.

1.1 Der $\mathbb{R}^n$

In diesem Buch werden die reellen Zahlen als bekannt vorausgesetzt. Eine axiomatische Beschreibung der reellen Zahlen wird meist in Büchern wie z. B. dem von H. Grauert und I. Lieb über Differential- und Integralrechnung I gegeben (s. Eintrag [6] im Literaturverzeichnis). Man kann auch in dem Buch von H.-D. Ebbinghaus et al. [2], das allein dem Thema „Zahlen“ gewidmet ist, nachschlagen. Dort wird beschrieben, wie man, ausgehend von einigen unbeweisbaren, aber „evidenten“ Axiomen, sukzessive die natürlichen, ganzen, rationalen und reellen Zahlen konstruiert.

Wir können die reellen Zahlen als unendliche Dezimalbrüche auffassen, z. B. $10{,}23457\ldots$ Jeder endliche Dezimalbruch ist auch ein unendlicher Dezimalbruch: $10{,}23 = 10{,}2300\ldots$, Brüche wie $10{,}22999\ldots$ und $10{,}2300\ldots$ sind gleich. Es gibt die Null $0 = 0{,}00\ldots$ und negative reelle Zahlen, z. B. $-10{,}2300\ldots$

Reelle Zahlen können addiert, multipliziert, dividiert werden. Die Eigenschaften der reellen Zahlen werden durch die Körperaxiome, die Anordnungsaxiome, das Archimedische Axiom und das Vollständigkeitsaxiom beschrieben.

Das Wort „Axiom“ wird in verschiedenen Bedeutungen benutzt: die erwähnten Axiome der reellen Zahlen oder die weiter unten einzuführenden Gruppen- und Vektorraumaxiome haben den Charakter von grundlegenden Definitionen. Dagegen sind Axiome, mit deren Hilfe man etwa die natürlichen Zahlen gewinnt (Unendlichkeitsaxiom) oder in Kapitel 2 die anschauliche Ebene beschreibt, unbeweisbare, aber „anschaulich evidente“ Aussagen.

Die Menge der reellen Zahlen wird mit $\mathbb{R}$ bezeichnet:

$$\mathbb{R} = \{a : a \text{ ist eine reelle Zahl}\}.$$

Der Doppelpunkt in der Mengenklammer bedeutet „mit“ oder „wobei gilt“.

Eine wichtige Teilmenge von $\mathbb{R}$ ist die Menge der natürlichen Zahlen $\mathbb{N} = \{1, 2, 3, \ldots\}$. Vor 3000 v. Chr. kannten die Ägypter schon sehr große natürliche Zahlen. Um 2000 v. Chr. entwickelten die Sumerer mit dem Sexagesimalsystem erstmals eine Positionsschreibweise, der allerdings die Ziffer 0 noch fehlte. Deren Rolle übernahmen vorerst noch Leerzeichen, wodurch es natürlich häufig zu Verwechslungen kam. Um 500 v. Chr. führten die Babylonier die *Ziffer* 0 als Zwischennull ein, die auch von den Griechen übernommen wurde. Das Wissen von der griechischen Astronomie und Physik gelangte über Persien nach Indien. Dort entstand bis 800 n. Chr. unser heute vertrautes „arabisches“ Zahlensystem. Dieses enthielt auch schon die 0 als selbständige *Zahl*, mit der man richtig rechnen konnte. Über die Anfänge der Mathematik kann man sich genauer in dem Buch „Erwachende Wissenschaft“ [14] von B. L. van der Waerden informieren.

Das Enthaltensein von $\mathbb{N}$ in $\mathbb{R}$ wird in der mathematischen Symbolsprache so beschrieben: $\mathbb{N} \subset \mathbb{R}$. Das bedeutet, daß jedes Element n von $\mathbb{N}$ (kurz $n \in \mathbb{N}$) auch Element von $\mathbb{R}$ ist. Fügen wir noch die 0 zu den natürlichen Zahlen hinzu, so schreiben wir $\mathbb{N}_0 := \mathbb{N} \cup \{0\} = \{0, 1, 2, 3, \ldots\}$.

Der n-dimensionale Zahlenraum

Im Folgenden sei $n \in \mathbb{N}$ eine *feste* natürliche Zahl. Wir betrachten (geordnete) n-tupel reeller Zahlen $\vec{a} = (a_1, \ldots, a_n)$, dabei sind die Komponenten a_i für $i = 1, \ldots, n$ reelle Zahlen. Die $\vec{a} \in \mathbb{R}^n$ heißen *Vektoren*.

1.1.1 Definition

Wir definieren den **n-dimensionalen reellen Zahlenraum** *durch*

$$\mathbb{R}^n := \{\vec{a} = (a_1, \ldots, a_n) : \ a_i \in \mathbb{R}, \ i = 1, \ldots, n\}.$$

Das Zeichen „:=“ bedeutet „definitionsgemäß gleich“, der Doppelpunkt steht auf der Seite des zu definierenden Objekts. Zwei n-tupel $\vec{a} \in \mathbb{R}^n$ und $\vec{b} \in \mathbb{R}^n$ sind gleich, wenn $a_i = b_i$ gilt für $i = 1, \ldots, n$. Beispielsweise gilt im Fall $n = 2$: $(1{,}1;\ 2{,}3) \neq (2{,}3;\ 1{,}1)$.

Je zwei Vektoren aus $\mathbb{R}^n$ können *addiert* werden, das Ergebnis ist wieder ein Vektor aus $\mathbb{R}^n$:

$$\vec{a} + \vec{b} := (a_1 + b_1, \ldots, a_n + b_n), \qquad \vec{a}, \vec{b} \in \mathbb{R}^n.$$

Der *Nullvektor* $\underbrace{(0, \ldots, 0)}_{n\text{-mal}}$ wird mit dem Symbol $\vec{0}$ bezeichnet. Zu jedem Vektor

$\vec{a} \in \mathbb{R}^n$ wird ein *negativer Vektor*

$$-\vec{a} := (-a_1, \ldots, -a_n)$$

definiert.

Schließlich kann man eine *Skalarmultiplikation* einführen. Für $d \in \mathbb{R}$ und $\vec{a} \in \mathbb{R}^n$ setzt man:

$$d \cdot \vec{a} := (d \cdot a_1, \ldots, d \cdot a_n) \in \mathbb{R}^n.$$

1.1.2 Beispiel

Weiter unten wird ausführlich dargelegt werden, daß $\mathbb{R}^2$ bzw. $\mathbb{R}^3$ geeignete Modelle der anschaulichen Ebene bzw. des anschaulichen Raumes sind. Allerdings tritt der $\mathbb{R}^n$ auch für beliebig große n im täglichen Leben in sehr naheliegender Weise als geeignetes Modell überall dort auf, wo man mit großen Datenmengen umgehen muß. Man kann zum Beispiel an eine Sparkasse denken, n ist dann die Anzahl der Konten. Guthaben bzw. Unterdeckung des i-ten Kontos wird durch die reelle Zahl a_i beschrieben. Der Vektor $\vec{a} = (a_1, \ldots, a_n)$ beschreibt die Gesamtheit aller Kontostände. Buchungen werden durch Vektoraddition und Verzinsungen durch Skalarmultiplikation modelliert.

Der $\mathbb{R}^n$ zusammen mit den oben erklärten Operationen hat die folgenden grundlegenden Eigenschaften, die weiter unten abstrakt als Gruppen- bzw. Vektorraumaxiome formuliert werden:

(1) Für alle $\vec{a}, \vec{b}, \vec{c} \in \mathbb{R}^n$ ist $\vec{a} + (\vec{b} + \vec{c}) = (\vec{a} + \vec{b}) + \vec{c}$, d. h. das *Assoziativgesetz* gilt.

Beweis

Es wird die entsprechende Eigenschaft reeller Zahlen benutzt. Seien $\vec{a}, \vec{b}, \vec{c} \in \mathbb{R}^n$ gegeben, dann gilt:

$$\begin{aligned}
\vec{a} + \left(\vec{b} + \vec{c}\right) &= \vec{a} + \Big((b_1, \ldots, b_n) + (c_1, \ldots c_n)\Big) \\
&= (a_1, \ldots, a_n) + (b_1 + c_1, \ldots, b_n + c_n) \\
&= \Big(a_1 + (b_1 + c_1), \ldots, a_n + (b_n + c_n)\Big) \\
&= \Big((a_1 + b_1) + c_1, \ldots, (a_n + b_n) + c_n\Big) \\
&= (a_1 + b_1, \ldots, a_n + b_n) + (c_1, \ldots, c_n) \\
&= \Big((a_1, \ldots, a_n) + (b_1, \ldots, b_n)\Big) + \vec{c} \\
&= \left(\vec{a} + \vec{b}\right) + \vec{c}.
\end{aligned}$$

$\square$

(2) Stets gilt $\vec{a} + \vec{0} = \vec{a}$. („Stets“ kann hier die genaue Formulierung „für alle $\vec{a} \in \mathbb{R}^n$“ ersetzen, weil sich diese Bedeutung direkt aus dem Zusammenhang ergibt.)

Beweis

$\vec{a} + \vec{0} = (a_1, \ldots, a_n) + (0, \ldots, 0) = (a_1 + 0, \ldots, a_n + 0) = (a_1, \ldots, a_n) = \vec{a}$. $\square$

(3) Stets gilt $\vec{a} + (-\vec{a}) = \vec{0}$.

Beweis

$$\begin{aligned}\vec{a} + (-\vec{a}) &= (a_1, \ldots, a_n) + (-a_1, \ldots, -a_n) \\ &= \Big(a_1 + (-a_1), \ldots, a_n + (-a_n)\Big) = (0, \ldots, 0) = \vec{0}.\end{aligned}$$

$\square$

(4) Das *Kommutativgesetz* gilt, d. h. stets ist $\vec{a} + \vec{b} = \vec{b} + \vec{a}$.

Beweis

$$\begin{aligned}\vec{a} + \vec{b} &= (a_1, \ldots, a_n) + (b_1, \ldots, b_n) = (a_1 + b_1, \ldots, a_n + b_n) \\ &= (b_1 + a_1, \ldots, b_n + a_n) = \vec{b} + \vec{a}.\end{aligned}$$

$\square$

Dieselben mathematischen Strukturen treten in vielen, scheinbar ganz verschiedenen Situationen auf. Deshalb abstrahieren wir nun von unserem konkreten Beispiel $\mathbb{R}^n$ und führen den allgemeinen Begriff einer „Gruppe“ ein.

Gruppenaxiome

1.1.3 Definition

Es sei V eine nichtleere Menge mit einer „binären“ Operation $+ : V \times V \to V, \quad (a, b) \mapsto a + b$. Es gelte:

(1) Stets ist $a + (b + c) = (a + b) + c$. (Assoziativgesetz)

(2) Es gibt ein (festes) Element $0 \in V$, so daß stets $a + 0 = a$ ist. (Existenz eines (rechts-)neutralen Elements)

(3) Zu jedem $a \in V$ existiert ein Element $-a \in V$, so daß $a + (-a) = 0$ gilt. (Existenz eines (rechts-)inversen Elements)

Dann heißt V eine **Gruppe**. *Ist V eine Gruppe und gilt zusätzlich*

(4) für alle $a, b \in V$: $\quad a + b = b + a$, (Kommutativgesetz)

so heißt V eine **abelsche Gruppe**.

Genauer sollte man $(V, +)$ schreiben. Da die Eindeutigkeit von neutralem und inversem Element erst unten in Bemerkung 1.1.5 (b) geklärt wird, müßte man zunächst die noch genauere Bezeichnung $(V, +, 0, -)$ verwenden. Die charakteristischen Eigenschaften (1)-(3) bzw. (1)-(4) aus Definition 1.1.3 nennt man *Gruppenaxiome*.

Mit Hilfe dieser Definition lassen sich die oben erzielten Resultate zusammenfassen.

1.1.4 Satz

$(\mathbb{R}^n, +, \vec{0}, -)$ *ist eine abelsche Gruppe.*

Die folgenden Bemerkungen sollen die Definition von Gruppen erläutern und ergänzen.

1.1.5 Bemerkungen

(a) Bei abelschen Gruppen bevorzugt man meist die oben verwendete additive Schreibweise, man hat die „Null“ 0 und „negative Elemente“ $-a$. Dagegen wird insbesondere bei nicht abelschen Gruppen die Operation häufig als „Multiplikation“ mit „$\circ$“ (Kringel) statt mit „+“ bezeichnet. Man schreibt dann e oder mitunter auch 1 statt 0 und a^{-1} statt $-a$.

(b) Sei $(V, +, 0, -)$ eine Gruppe. Dann gilt:

(i) Für alle $a \in V$ ist $(-a) + a = 0$, d. h. das in Axiom (3) garantierte *rechtsinverse* Element von a ist auch *linksinvers* zu a.

(ii) Stets gilt $0 + a = a$, d. h. die *Rechts-Null* ist auch eine *Links-Null*.

(iii) Stets gilt $a = -(-a)$.

(iv) Sei $0' \in V$. Für ein $a \in V$ gelte $a + 0' = a$. Dann ist $0 = 0'$.

(v) Sei $a \in V$ gegeben. Gilt für ein $b \in V$: $a + b = 0$, so ist $b = -a$.

Beweis

Man beachte, daß das Kommutativgesetz tatsächlich nicht benötigt wird.

(i) Das Assoziativgesetz wird wiederholt verwendet:

$$\begin{aligned}
(-a) + a &= ((-a) + a) + 0 \\
&= \Big((-a) + a\Big) + \Big((-a) + (-(-a))\Big) \\
&= \Big(((-a) + a) + (-a)\Big) + (-(-a)) \\
&= \Big((-a) + (a + (-a))\Big) + (-(-a)) \\
&= ((-a) + 0) + (-(-a)) = (-a) + (-(-a)) = 0.
\end{aligned}$$

(ii)

$$\begin{aligned}
0 + a &= \Big(a + (-a)\Big) + a = a + \Big((-a) + a\Big) \\
&= a + 0 \qquad \text{unter Verwendung von (i)} \\
&= a.
\end{aligned}$$

(iii)

$$\begin{aligned}
(-(-a)) &= 0 + (-(-a)) \qquad \text{unter Verwendung von (ii)} \\
&= \Big(a + (-a)\Big) + (-(-a)) \\
&= a + \Big((-a) + (-(-a))\Big) = a + 0 = a.
\end{aligned}$$

(iv)
$$\begin{aligned} 0' &= 0 + 0' && \text{auf Grund von (ii)} \\ &= ((-a) + a) + 0' && \text{auf Grund von (i)} \\ &= (-a) + (a + 0') = (-a) + a = 0. \end{aligned}$$

(v)
$$\begin{aligned} b &= 0 + b = ((-a) + a) + b \\ &= (-a) + (a + b) = -a + 0 = -a. \end{aligned}$$
□

(c) Die Bemerkungen (iv) und (v) zeigen, daß das Nullelement und das inverse Element jeweils eindeutig bestimmt sind. Daher reicht es aus, eine Gruppe $(V, +, 0, -)$ einfach mit $(V, +)$ zu bezeichnen.

(d) Bei den Gruppenaxiomen ist es wesentlich, einheitlich die Existenz von *Rechts*-Null und *Rechts*-Inversen zu fordern.

Zur Illustration der letzten Bemerkung dient das folgende Beispiel:

1.1.6 Beispiel

Es sei V eine mindestens zweielementige Menge mit der Addition $a + b := b$. Es gibt eine Links-Null und stets ein Rechts-Inverses, nämlich die vorher ausgewählte Links-Null. Die Addition ist zudem assoziativ. Jedoch ist $(V, +)$ keine Gruppe, denn die Null ist nicht eindeutig bestimmt: jedes Element kann als Links-Null dienen.

Zum vorläufigen Abschluß unserer gruppentheoretischen Betrachtungen führen wir Ausdrücke der Form $a_1 + \cdots + a_n$ ein; die a_i sind Elemente einer Gruppe $(V, +)$. Dieses geschieht induktiv.

Der Fall $n = 2$ ist klar: Ausdrücke der Form $a_1 + a_2$ sind in einer Gruppe stets erklärt.

Sei also $n > 2$. Ausdrücke der Form $a_1 + \ldots + a_{n-1}$ seien bereits erklärt. Dann definiert man
$$a_1 + \ldots + a_n := (a_1 + \ldots + a_{n-1}) + a_n.$$

Es gilt das *allgemeine Assoziativgesetz*: in dem Ausdruck $a_1 + \ldots + a_n$ können nach Belieben Klammern gesetzt werden, ohne daß sich der Wert des Ausdrucks ändert.

Einen Beweis dieser Aussage findet man z. B. bei H. Grauert, I. Lieb [6, S. 13f].

Wir wenden uns wieder dem $\mathbb{R}^n$ zu und stellen einige Eigenschaften der Skalarmultiplikation zusammen. Deren Nachweis ist sehr leicht und wird daher nicht ausgeführt.

(5) Es gilt das *unitäre Gesetz*: Für alle $\vec{a} \in \mathbb{R}^n$ ist $1 \cdot \vec{a} = \vec{a}$.

(6) Es gilt das *Assoziativgesetz* der Skalarmultiplikation: Für alle $c, d \in \mathbb{R}$, $\vec{a} \in \mathbb{R}^n$ ist
$$c \cdot (d \cdot \vec{a}) = (c \cdot d) \cdot \vec{a}.$$

(7) Skalarmultiplikation und Vektoraddition werden durch *Distributivgesetze* verknüpft: Für alle $c, d \in \mathbb{R}$ und $\vec{a}, \vec{b} \in \mathbb{R}^n$ ist
$$(c + d) \cdot \vec{a} = c \cdot \vec{a} + d \cdot \vec{a}, \qquad c \cdot (\vec{a} + \vec{b}) = c \cdot \vec{a} + c \cdot \vec{b}.$$

Man bedenke, daß sowohl die Addition in $\mathbb{R}$ als auch die Addition in $\mathbb{R}^n$ durch dasselbe Symbol „+“ bezeichnet werden. Entsprechendes gilt für die Skalarmultiplikation und die Multiplikation in $\mathbb{R}$.

1.2 Reelle Vektorräume

In diesem Abschnitt lösen wir uns von der Betrachtung des konkreten Vektorraums $\mathbb{R}^n$ und gelangen zur abstrakten Definition eines reellen Vektorraums.

Das Symbol $\emptyset$ bezeichnet die leere Menge.

1.2.1 Definition

Es sei $V \neq \emptyset$ eine nichtleere Menge, auf der zwei binäre Operationen gegeben sind, eine Addition

$$+ : V \times V \to V, \qquad (x, y) \mapsto x + y$$

und eine Skalarmultiplikation

$$\cdot : \mathbb{R} \times V \to V, \qquad (c, x) \mapsto c \cdot x.$$

Es sei $(V, +)$ eine abelsche Gruppe. Für die Skalarmultiplikation gelte ferner:

(5) *Für alle $x \in V$ ist* $\quad 1 \cdot x = x.$ (Unitäres Gesetz)

(6) *Für alle $c, d \in \mathbb{R}$ und $x \in V$ ist:*
$(c \cdot d) \cdot x = c \cdot (d \cdot x).$ (Assoziativgesetz)

(7) *Für alle $c, d \in \mathbb{R}$ und $x, y \in V$ ist:*
$c \cdot (x + y) = c \cdot x + c \cdot y$ *und* $(c + d) \cdot x = c \cdot x + d \cdot x.$ (Distributivgesetze)

Dann heißt $(V, +, \cdot)$ ein **reeller Vektorraum** *oder* **$\mathbb{R}$-Vektorraum**.

Die charakteristischen Eigenschaften (1)-(7) aus den Definitionen 1.1.3 und 1.2.1 nennt man *Vektorraumaxiome*. Die Elemente $x, y, \ldots$ aus V heißen *Vektoren*. Das neutrale Element $0 \in V$ bzgl. $+$ heißt *Nullvektor*. Meist wird eine Verwechslung mit $0 \in \mathbb{R}$ durch den Zusammenhang ausgeschlossen, in Zweifelsfällen wird der Nullvektor als solcher gekennzeichnet: $0 = 0_V$. Hier ist $\mathbb{R}$ der *Skalarenkörper*, seine Elemente heißen *Skalare*. Man schreibt statt $(V, +, \cdot)$ auch kurz V, wenn hinsichtlich der Operationen auf V kein Mißverständnis möglich ist. Falls klar ist, daß $\mathbb{R}$ der Skalarenkörper ist, nennt man V auch einfach *Vektorraum*. Für *komplexe* Vektorräume dagegen verweisen wir auf Definition 7.1.3.

Die Ergebnisse des vorhergehenden Abschnitts über den $\mathbb{R}^n$ können durch Verwendung des neuen Begriffs zusammengefaßt werden.

1.2.2 Satz

$\mathbb{R}^n$ *ist ein reeller Vektorraum.*

Es folgen einige Eigenschaften reeller Vektorräume.

1.2.3 Satz

Sei V ein reeller Vektorraum. Für alle $c \in \mathbb{R}$ und $x \in V$ gilt:

(a) $0 \cdot x = 0$.

(b) *Sei* $0 = 0_V$ *der Nullvektor. Dann gilt* $c \cdot 0 = 0$.

(c) *Stets gilt* $c \cdot (-x) = (-c) \cdot x = -(c \cdot x)$.

(d) *Aus* $c \cdot x = 0$ *folgt:* $c = 0$ *oder* $x = 0$.

Beweis

(a) Aus $x = 1 \cdot x = (1 + 0) \cdot x = 1 \cdot x + 0 \cdot x = x + 0 \cdot x$ folgt $0 = 0 \cdot x$ auf Grund von Bemerkung 1.1.5.

(b) Ebenso ergibt sich $c \cdot 0 = 0$ aus $c \cdot 0 = c \cdot (0 + 0) = c \cdot 0 + c \cdot 0$.

(c) Das Distributivgesetz ergibt $0 = c \cdot 0 = c \cdot (x + (-x)) = c \cdot x + c \cdot (-x)$ und $0 = 0 \cdot x = (c + (-c)) \cdot x = c \cdot x + (-c) \cdot x$. Wegen der Eindeutigkeit des negativen Elements folgt $-(c \cdot x) = c \cdot (-x) = (-c) \cdot x$.

(d) Im Fall $c = 0$ ist nichts zu zeigen. Ist dagegen $c \neq 0$, so existiert der Kehrwert $\frac{1}{c} \in \mathbb{R}$. Mit Hilfe von Behauptung (b) folgt nun: $0 = \frac{1}{c} \cdot (c \cdot x) = (\frac{1}{c} \cdot c) \cdot x = 1 \cdot x = x$. In jedem Fall ist $c = 0$ oder $x = 0$. □

1.2.4 Bemerkung

(a) Damit die Formulierung von Mathematik nicht unnötig schwerfällig wird, überläßt man häufig die genaue Klärung der Bedeutung von Symbolen und Begriffen dem Zusammenhang. In Teil (a) des vorhergehenden Satzes wird sowohl die reelle Zahl Null als auch der Nullvektor mit demselben Symbol 0 bezeichnet. Da aber x voraussetzungsgemäß ein Vektor ist, muß mit 0 auf der linken Seite die reelle Zahl 0 gemeint sein. Da die Skalarmultiplikation Ergebnisse in V liefert, ist die 0 auf der rechten Seite der Nullvektor.

(b) Im Gegensatz zur Umgangssprache ist in der Mathematik das „oder“ ein nicht auschließendes „oder“. In Teil (d) des vorhergehenden Satzes bedeutet das, daß auch der Fall eintreten *kann*, in dem sowohl $c = 0$ als auch $x = 0$ gilt.

1.2.5 Definition

Sei V ein reeller Vektorraum. Eine Teilmenge $V' \subset V$ heißt **Untervektorraum** *oder* **linearer Teilraum** *von V, falls gilt:*

(1) $V' \neq \emptyset$.

(2) *Aus* $x_1, x_2 \in V'$ *folgt* $x_1 + x_2 \in V'$.

(3) *Aus* $c \in \mathbb{R}$, $x \in V'$ *folgt* $c \cdot x \in V'$.

Ein Untervektorraum von V ist also eine nichtleere Teilmenge von V, die gegenüber der Addition „+“ von Vektoren und der Skalarmultiplikation „·“ abgeschlossen ist.

1.2.6 Beispiel

Die Menge $V' := \left\{\vec{a} \in \mathbb{R}^3 : \ a_1 = a_3\right\}$ ist ein Untervektorraum des $\mathbb{R}^3$:

(1) $V' \neq \emptyset$, denn $\vec{0} \in V'$.

(2) Seien $\vec{a}, \vec{b} \in V'$. Dann gilt:

$$a_1 = a_3 \text{ und } b_1 = b_3 \ \Rightarrow \ (a_1 + b_1) = (a_3 + b_3) \ \Rightarrow \ \vec{a} + \vec{b} \in V'.$$

(3) Seien $c \in \mathbb{R}$, $\vec{a} \in V'$. Dann gilt:

$$a_1 = a_3 \ \Rightarrow \ c \cdot a_1 = c \cdot a_3 \ \Rightarrow \ c \cdot \vec{a} \in V'.$$

1.2.7 Folgerungen

Sei V ein reeller Vektorraum und $V' \subset V$ ein Untervektorraum.

(a) Dann ist $0 \in V'$. Denn $V' \neq \emptyset$, also existiert ein $x \in V'$. Eigenschaft (3) liefert zusammen mit Satz 1.2.3 (a):

$$0 = 0 \cdot x \in V'.$$

(b) Ist $x \in V'$, so ist auch $-x \in V'$. Eigenschaft (3) liefert nämlich zusammen mit Satz 1.2.3 (c):

$$-x = -(1 \cdot x) = (-1) \cdot x \in V'.$$

1.2.8 Satz

Es sei V ein reeller Vektorraum, $V' \subset V$ ein Untervektorraum. Dann ist V' mit denselben Operationen wie auf V ein reeller Vektorraum.

Noch genauer sollte es heißen: „V' mit den Einschränkungen auf V' der auf V erklärten Operationen.“

Beweis

Die Bedingungen (2) und (3) aus Definition 1.2.5 liefern

$$+ : V' \times V' \to V',$$

$$\cdot : \mathbb{R} \times V' \to V'.$$

Die Folgerungen 1.2.7 zeigen die Gültigkeit der Vektorraumaxiome (Gruppenaxiome für $(V', +)$) (2) und (3). Weil V' Teilmenge von V und V ein Vektorraum ist, ist die Gültigkeit der übrigen Vektorraumaxiome offensichtlich. $\square$

1.2.9 Satz

Jeder reelle Vektorraum V besitzt $\{0\}$ und V als „triviale" Untervektorräume.

Beweis

Klar! □

1.2.10 Satz

Seien V_1, V_2 Untervektorräume des reellen Vektorraums V. Dann ist auch $V_1 \cap V_2$ ein Untervektorraum von V.

Beweis

Wir haben sowohl $0 \in V_1$ als auch $0 \in V_2$. Also liegt das Nullelement auch im Durchschnitt $V_1 \cap V_2$, und dieser ist nicht leer. Seien nun $x, y \in V_1 \cap V_2$ und $c \in \mathbb{R}$ gegeben. Es gilt also sowohl $x, y \in V_1$ als auch $x, y \in V_2$. Da V_1 und V_2 Untervektorräume sind, erhalten wir $x + y \in V_1$, $x + y \in V_2$, $c \cdot x \in V_1$ und $c \cdot x \in V_2$. Damit folgt schließlich $x + y \in V_1 \cap V_2$ und $c \cdot x \in V_1 \cap V_2$. □

Es sollen nun einige Vereinfachungen der Schreibweise eingeführt werden. Mit V bezeichnen wir im Folgenden stets einen reellen Vektorraum. Multiplikationspunkte werden häufig weggelassen. Das Assoziativgesetz erlaubt es, auch bei der Skalarmultiplikation auf Klammern zu verzichten. Treten in einem Ausdruck Multiplikationen und Additionen auf, so binden die Multiplikationen stärker. Für $x, y \in V$ ist die Differenz durch $x - y := x + (-y)$ erklärt. Mit einem großen Sigma (Summenzeichen) werden Summen abgekürzt:

$$\sum_{i=1}^{n} x_i := x_1 + \cdots + x_n.$$

Die indizierten Vektoren x_i sind nicht mit den Komponenten eines n-tupels aus Abschnitt 1.1 zu verwechseln.

In der folgenden Definition wird beschrieben, welche Vektoren man mit Hilfe der Vektorraumoperationen aus einem endlichen System von „erzeugenden" Vektoren konstruieren kann.

1.2.11 Definition

Seien $n \in \mathbb{N}$, $x_1, \ldots, x_n \in V$, $c_1, \ldots, c_n \in \mathbb{R}$. Dann heißt

$$x = c_1 x_1 + \cdots + c_n x_n = \sum_{i=1}^{n} c_i x_i$$

Linearkombination *von $x_1, \ldots, x_n$.*

1.2.12 Satz

Seien $x_1, \dots, x_n \in V$ gegeben. Es sei V' die Menge aller Linearkombinationen von $x_1, \dots, x_n$. Dann ist V' ein Untervektorraum von V, der $x_1, \dots, x_n$ enthält.

Beweis

Zunächst gilt offensichtlich $x_1, \dots, x_n \in V'$ und damit auch $V' \neq \emptyset$. Seien nun

$$x = \sum_{i=1}^{n} c_i x_i\,, \qquad y = \sum_{i=1}^{n} d_i x_i\,, \qquad c_i,\ d_i \in \mathbb{R},\ i = 1, \dots, n$$

Linearkombinationen von $x_1, \dots, x_n$. Weiter sei $d \in \mathbb{R}$. Unter intensiver Verwendung der Kommutativ- und Distributivgesetze erhält man:

$$x + y = \sum_{i=1}^{n} (c_i + d_i) x_i, \qquad d \cdot y = \sum_{i=1}^{n} (d \cdot d_i) x_i.$$

Damit sind auch $x + y$ und $d \cdot y$ in V'. □

1.2.13 Definition

Die Menge V' aller Linearkombinationen der Vektoren $x_1, \dots, x_n \in V$ heißt der von $x_1, \dots, x_n$ **aufgespannte Untervektorraum** *und wird mit $\mathbb{R} \cdot (x_1, \dots, x_n) := V'$ bezeichnet. Hierbei heißen $x_1, \dots, x_n$* **Erzeugende** *oder* **Erzeuger** *von V'.*

Weitere, in der Literatur verbreitete Bezeichnungen für V' sind $<x_1, \dots, x_n>$, $[x_1, \dots, x_n]$ oder Span$(x_1, \dots, x_n)$.

1.3 Basis und Dimension

Am Ende des vorhergehenden Abschnitts ist ein Verfahren beschrieben worden, mit dem man den kleinstmöglichen Vektorraum findet, der ein vorgegebenes System von Vektoren $x_1, \dots, x_n$ enthält. Hier wollen wir nun umgekehrt vorgehen. Gegeben ist ein Vektorraum V; wir suchen Systeme von Vektoren, die den ganzen Vektorraum aufspannen, die also bereits die gesamte Information über den Vektorraum V enthalten. Unter allen solchen erzeugenden Systemen interessieren wir uns besonders für „möglichst kleine“, bei denen kein Vektor mehr entfernt werden kann, ohne die Eigenschaft des Erzeugens zu verlieren. Allein das Wissen, daß V ein Vektorraum (linearer Raum) ist, soll es uns ermöglichen, den ganzen Raum durch ein relativ extrem kleines Teilsystem, die „Basis“, zu beschreiben.

„Kleinstmögliche Systeme“ werden durch den nun zu definierenden Begriff der „linearen Unahhängigkeit“ charakterisiert.

Lineare Unabhängigkeit

Auch hier bezeichne V wieder stets einen reellen Vektorraum.

1.3.1 Definition

Die Vektoren $x_1, \ldots, x_n \in V$ *heißen* **linear abhängig**, *wenn es Zahlen* $c_1, \ldots, c_n \in \mathbb{R}$ *gibt, von denen wenigstens eine ungleich* 0 *ist, so daß*

$$0 = \sum_{i=1}^{n} c_i x_i$$

gilt. Umgekehrt heißen $x_1, \ldots, x_n$ **linear unabhängig**, *wenn* $x_1, \ldots, x_n$ *nicht linear abhängig sind.*

1.3.2 Beispiel

Die Vektoren $(1,2),\ (-2,0),\ (3,1) \in \mathbb{R}^2$ sind linear abhängig, denn es ist

$$(-1) \cdot (1,2) + \frac{5}{2} \cdot (-2,0) + 2 \cdot (3,1) = \vec{0}.$$

Die folgende Bemerkung ergibt sich direkt aus der Definition linearer Unabhängigkeit, indem man einige Grundregeln logischen Schließens anwendet: Die Verneinung von „es gibt ..., so daß " ist „für alle ... gilt nicht". Die Verneinung von „(nicht A) und B" ist „A oder (nicht B)" und damit äquivalent zu „$B \Rightarrow A$".

1.3.3 Bemerkung

Die Vektoren $x_1, \ldots, x_n \in V$ *sind genau dann linear unabhängig, wenn stets aus*

$$c_1, \ldots, c_n \in \mathbb{R}, \qquad \sum_{i=1}^{n} c_i x_i = 0$$

folgt:

$$c_1 = \ldots = c_n = 0.$$

1.3.4 Beispiel

Wir greifen das Beispiel 1.2.6 aus dem vorhergehenden Abschnitt wieder auf. Dort wurde bewiesen, daß $V' = \left\{ \vec{a} \in \mathbb{R}^3 :\ a_1 = a_3 \right\}$ ein Untervektorraum des $\mathbb{R}^3$ ist. Hier soll gezeigt werden, daß $\vec{b}_1 = (1,0,1)$ und $\vec{b}_2 = (1,1,1)$ linear unabhängige Vektoren aus V' sind:

Offensichtlich gilt $\vec{b}_1, \vec{b}_2 \in V'$, es bleibt der Nachweis der linearen Unabhängigkeit. Dazu sei der Nullvektor eine Linearkombination von $\vec{b}_1$ und $\vec{b}_2$:

$$c_1 \vec{b}_1 + c_2 \vec{b}_2 = \vec{0}, \qquad c_1, c_2 \in \mathbb{R}.$$

Es folgt:

$$(c_1 + c_2, c_2, c_1 + c_2) = (0, 0, 0).$$

Vektoren in $\mathbb{R}^n$ sind genau dann gleich, wenn alle ihre Komponenten übereinstimmen:

$$\Rightarrow \begin{cases} c_1 + c_2 = 0 \\ \quad\; c_2 = 0 \\ c_1 + c_2 = 0 \end{cases} \Rightarrow \begin{cases} c_1 + c_2 = 0 \\ \quad\; c_2 = 0 \end{cases} \Rightarrow c_2 = 0 \text{ und } c_1 = 0.$$

1.3.5 Satz

Ein Vektor $x \in V$ ist linear abhängig genau dann, wenn $x = 0$ ist.

Die Behauptung ist eine „genau dann, wenn"-Behauptung und besteht daher aus den zwei Aussagen:

„$\Rightarrow$": x linear abhängig $\Rightarrow x = 0$,

„$\Leftarrow$": $x = 0 \Rightarrow x$ linear abhängig.

Daher besteht auch der Beweis aus zwei Teilen.

Beweis

„$\Rightarrow$": Sei x linear abhängig. Dann existiert $c \in \mathbb{R}$, $c \neq 0$ mit $c \cdot x = 0$. Es folgt $0 = \frac{1}{c} \cdot 0 = \frac{1}{c} \cdot c \cdot x = 1 \cdot x = x$.

„$\Leftarrow$": Sei $x = 0$. Das unitäre Gesetz liefert $1 \cdot x = x = 0$. Also ist x linear abhängig.

$\square$

1.3.6 Satz

Die Vektoren $x_1, \ldots, x_n \in V$ seien linear abhängig. Es sei $x \in V$. Dann sind $x, x_1, \ldots, x_n$ ebenfalls linear abhängig.

Jedes „Obersystem" eines linear abhängigen Systems ist also wieder linear abhängig.

Beweis

Es gibt $c_1, \ldots, c_n \in \mathbb{R}$, wobei ein $c_i \neq 0$ ist, so daß $0 = \sum_{i=1}^{n} c_i x_i$ gilt. Mit der Setzung $c := 0$ ist dann auch $0 = cx + c_1 x_1 + \cdots + c_n x_n$. $\square$

1.3.7 Satz

Seien $x_1, \ldots, x_n \in V$ linear unabhängig. Es seien ℓ ($\ell \leq n$) Zahlen $1 \leq i_1 < i_2 < \ldots < i_\ell \leq n$ gegeben. Dann sind $x_{i_1}, \ldots, x_{i_\ell}$ linear unabhängig.

Jedes Teilsystem eines Systems linear unabhängiger Vektoren ist also wieder linear unabhängig.

Beweis

Der Nullvektor sei die Linearkombination $\sum_{k=1}^{\ell} c_{i_k} x_{i_k} = 0$, $c_{i_k} \in \mathbb{R}$. Für $j = 1, \dots, n$ definieren wir

$$d_j := \begin{cases} 0 \quad , & \text{falls } j \notin \{i_1, \dots, i_\ell\}, \\ c_{i_k} \ , & \text{falls } j = i_k \text{ für ein } k \in \{1, \dots, \ell\}. \end{cases}$$

Dann gilt

$$\sum_{j=1}^{n} d_j x_j = \sum_{k=1}^{\ell} c_{i_k} x_{i_k} = 0.$$

Die lineare Unabhängigkeit von $x_1, \dots, x_n$ ergibt $d_j = 0$, $j = 1, \dots, n$. Folglich ist auch $c_{i_k} = 0$, $k = 1, \dots, \ell$. □

1.3.8 Satz

Die Vektoren $x_1, \dots, x_n \in V$ seien linear unabhängig. Sei $x \in V$. Dann gilt: $x, x_1, \dots, x_n$ sind linear abhängig genau dann, wenn x Linearkombination von $x_1, \dots, x_n$ ist.

Beweis

„$\Rightarrow$“: Seien $x, x_1, \dots, x_n$ linear abhängig. Dann existieren reelle Zahlen $c, c_1, \dots, c_n$, die nicht alle gleich 0 sind, mit $0 = cx + c_1 x_1 + \dots + c_n x_n$. Wir wollen $c \neq 0$ zeigen und nehmen dazu an, c wäre gleich 0. Dann wäre $\sum_{i=1}^{n} c_i x_i = 0$, ein c_i wäre ungleich 0 im Widerspruch zur Voraussetzung. Also ist $c \neq 0$. Es folgt

$$0 = \frac{1}{c} \cdot 0 = \frac{1}{c} \left(cx + \sum_{i=1}^{n} c_i x_i \right) = x + \sum_{i=1}^{n} \frac{c_i}{c} \cdot x_i,$$

also

$$x = \sum_{i=1}^{n} \left(-\frac{c_i}{c} x_i \right).$$

„$\Leftarrow$“: Es sei x Linearkombination von $x_1, \dots, x_n$. Dann exisitieren $c_1, \dots, c_n \in \mathbb{R}$ mit $x = \sum_{i=1}^{n} c_i x_i$. Setzt man $c = -1$, so gilt $0 = cx + c_1 x_1 + \dots + c_n x_n$. Folglich sind $x, x_1, \dots, x_n$ linear abhängig. □

Man kann den vorhergehenden Satz auch so lesen, daß in einem linear unabhängigen System kein Vektor Linearkombination der übrigen Vektoren ist. Der von einem linear unabhängigen System aufgespannte Untervektorraum wird also durch Weglassen eines Erzeugers *echt* verkleinert. In diesem Sinne sind linear unabhängige Systeme „kleinstmöglich“. Einen anderen Aspekt dieser Minimal-Eigenschaft beleuchtet der folgende Satz:

1.3.9 Satz (Eindeutigkeit der Darstellung)

Es seien $x_1, \dots, x_n \in V$ *linear unabhängig und* c_i, d_i $(i = 1, \dots, n)$ *reelle Zahlen. Es gelte*

$$\sum_{i=1}^{n} c_i x_i = \sum_{i=1}^{n} d_i x_i.$$

Dann ist $c_i = d_i$ *für* $i = 1, \dots, n.$

Beweis

Die Voraussetzung ergibt $\sum_{i=1}^{n}(c_i - d_i)x_i = 0$. Wegen der linearen Unabhängigkeit von $x_1, \dots, x_n$ folgt $c_i - d_i = 0$ und somit $c_i = d_i$ für $i = 1, \dots, n$. □

Dimension

Für das Folgende ist es zweckmäßig, das Symbol „∞“ (unendlich) einzuführen. Bei „∞“ handelt es sich um eine Zahl, die definitionsgemäß folgende Eigenschaften besitzt: $\infty \notin \mathbb{R}$. Für alle $c \in \mathbb{R}$ gilt $c < \infty,\ c+\infty := \infty,\ \infty+c := \infty, \infty+\infty = \infty$. Man erklärt dagegen nicht „$\infty - \infty$“.

Für die ausführliche Definition des Supremums $\sup$ als kleinste obere Schranke einer Teilmenge von $\mathbb{R}$ und des Infimums $\inf$ als größte untere Schranke sei wieder auf das Buch von H. Grauert und I. Lieb [6, S. 30f] verwiesen. Hat eine Teilmenge M von $\mathbb{R}$ keine obere Schranke in $\mathbb{R}$, so ist $\sup M = \infty$.

Jedem reellen Vektorraum V soll nun eine Zahl $n = \dim V \in \mathbb{N} \cup \{0, \infty\}$, die Dimension von V, zugeordnet werden. Die Dimension soll ein Maß für die Größe eines Vektorraums sein und insbesondere so definiert werden, daß man $\dim \mathbb{R}^n = n$ beweisen kann.

1.3.10 Definition

Die **Dimension** *eines reellen Vektorraums* V *wird durch*

$$\dim V := \sup \{k \in \mathbb{N}_0 : \text{ es gibt } k \text{ linear unabhängige Vektoren in } V\}$$

definiert. Ist $\dim V < \infty$, *so heißt* V **endlichdimensional**, *andernfalls nennen wir* V **unendlichdimensional**.

Demnach ist $\dim V$ die maximale Anzahl linear unabhängiger Vektoren aus V. Dabei soll ein unendliches System von Vektoren linear unabhängig heißen, wenn jedes endliche Teilsystem linear unabhängig ist.

1.3.11 Beispiele

(a) In $V = \mathbb{R}^n$ zeichnen wir n Vektoren aus, die *Einheitsvektoren*: $\vec{e}_1 := (1, 0, \dots, 0), \dots, \vec{e}_n := (0, \dots, 0, 1)$. Die Vektoren $\vec{e}_1, \dots, \vec{e}_n$ sind linear unabhängig, denn aus $c_1, \ldots, c_n \in \mathbb{R}$, $\sum_{i=1}^{n} c_i \vec{e}_i = 0$ folgt:

$$\begin{aligned}(0, \ldots, 0) &= \sum_{i=1}^{n} c_i \cdot (0, \ldots, \underbrace{1}_{i\text{-te Stelle}}, \ldots, 0) \\ &= \sum_{i=1}^{n} (0, \ldots, \underbrace{c_i}_{i\text{-te Stelle}}, \ldots, 0) = (c_1, \ldots, c_n).\end{aligned}$$

Also ist $c_1 = \ldots = c_n = 0$, der Nullvektor also nur als triviale Linearkombination von $\vec{e}_1, \ldots, \vec{e}_n$ darstellbar. Daher ist $\dim \mathbb{R}^n \geq n$. Der Nachweis der Gleichheit kann erst später in Satz 3.3.14 erfolgen.

(b) Unter einem *(reellen) Polynom* versteht man einen *formalen Ausdruck* $c_0 + c_1 z + c_2 z^2 + \cdots + c_n z^n$ mit $n \in \mathbb{N}_0$ und $c_i \in \mathbb{R}$ für $i = 0, \ldots, n$. Dabei heißt z *Variable (Veränderliche, Unbestimmte)*. Man vereinbart, daß man beliebig viele Glieder $0 \cdot z^i \quad (i > n)$ zum Polynom hinzufügen darf, ohne daß sich das Polynom ändert. So kann man stets erreichen, daß bei je zwei Polynomen der höchste auftretende Exponent n gleich ist. Zwei Polynome $c_0 + c_1 z + \cdots + c_n z^n$ und $d_0 + d_1 z + \cdots + d_n z^n$ heißen *gleich*, wenn $c_i = d_i$ für $i = 0, \ldots, n$ gilt. Es handelt sich hier um die *algebraische Definition* eines Polynoms, im Gegensatz zur *analytischen Definition*: Ein Polynom ist danach eine für $z \in \mathbb{R}$ erklärte *Funktion* f der Form $f(z) = c_0 + c_1 z + c_2 z^2 + \ldots + c_n z^n$. Zwei Polynome f und g sind in diesem Sinne gleich, wenn für alle $z \in \mathbb{R}$ gilt: $f(z) = g(z)$.

Im reellen Fall sind beide Definitionen gleichwertig, denn man kann zeigen, daß ein Polynom $c_0 + c_1 z + c_2 z^2 + \cdots + c_n z^n$, welches nicht das Nullpolynom ist, höchstens n Nullstellen hat. Ersetzt man aber $\mathbb{R}$ durch endliche Körper, so ergeben sich grundlegende Unterschiede, insbesondere beim Gleichheitsbegriff. Wir legen hier die algebraische Definition zugrunde. Dann wird ein Polynom völlig gleichwertig durch seine Koeffizienten $(c_0, \ldots, c_n)$ beschrieben.

Es bezeichne V die Menge aller (reellen) Polynome. Auf V können Addition und Skalarmultiplikation erklärt werden:

$$\begin{gathered}(c_0 + c_1 z + \cdots + c_n z^n) + (d_0 + d_1 z + \cdots + d_n z^n) \\ := (c_0 + d_0) + (c_1 + d_1) z + \cdots + (c_n + d_n) z^n,\end{gathered}$$

$$c \cdot (c_0 + c_1 z + \cdots + c_n z^n) := (c c_0) + (c c_1) z + \cdots + (c c_n) z^n.$$

Das Ergebnis ist jeweils wieder ein Polynom. Man kann sich schnell davon überzeugen, daß V mit diesen Operationen ein reeller Vektorraum ist. Auch hier können „Einheitsvektoren" ausgezeichnet werden: $e_0 = 1$, $e_1 = z$, $e_2 = z^2, \ldots$, allgemein $e_n = z^n$ für $n \in \mathbb{N}_0$. Dieser Vektorraum V aller reellen Polynome ist das erste Beispiel eines *unendlichdimensionalen Vektorraums*. Für den Beweis reicht es zu zeigen, daß für jedes $n \in \mathbb{N}_0$ die Einheitsvektoren $e_0, \ldots, e_n$ linear unabhängig sind. Sei also $c_0 e_0 + \cdots + c_n e_n = 0$ das Nullpolynom mit

$c_0, \ldots, c_n \in \mathbb{R}$, dann ist $c_0 + c_1 z + \cdots + c_n z^n = 0$. Die Definition der Gleichheit liefert $c_0 = \ldots = c_n = 0$. Die Einheitsvektoren $e_i \quad (i = 0, 1, 2, \ldots)$ sind nicht nur linear unabhängig, sondern sie erzeugen auch V, denn jedes Polynom aus V läßt sich in offensichtlicher Weise als Linearkombination von endlich vielen Vektoren e_i darstellen.

Basis

Wir führen zunächst *endliche* Basen ein.

1.3.12 Definition

Eine endliche Teilmenge $\{x_1, \ldots, x_n\} \subset V$ *heißt* **Basis** *des reellen Vektorraums* V*, falls gilt:*

(1) $x_1, \ldots, x_n$ *sind linear unabhängig;*

(2) $x_1, \ldots, x_n$ *erzeugen V, d. h.* $V = \mathbb{R} \cdot (x_1, \ldots, x_n)$.

Für $V = \{0\}$ *vereinbart man die leere Menge* $\emptyset$ *als Basis.*

Bei der aufzählenden Beschreibung von Mengen wie $\{x_1, \ldots, x_n\}$ gehen wir stets davon aus, daß die Elemente paarweise verschieden sind, daß für $i \neq j$ also $x_i \neq x_j$ gilt.

1.3.13 Beispiel

(a) Eine Basis des $\mathbb{R}^n$ ist durch $\{\vec{e}_1, \ldots, \vec{e}_n\}$ gegeben. In Abschnitt 3.3 wird gezeigt: Jede Basis des $\mathbb{R}^n$ enthält genau n Elemente.

(b) Wir entwickeln das oben auf den Seiten 9 und 12 behandelte Beispiel weiter. Demnach ist $V' = \left\{\vec{a} \in \mathbb{R}^3 : a_1 = a_3\right\}$ ein Untervektorraum des $\mathbb{R}^3$, und $\vec{b}_1 = (1, 0, 1)$, $\vec{b}_2 = (1, 1, 1)$ sind linear unabhängige Vektoren aus V'. Zudem wird V' von $\vec{b}_1$ und $\vec{b}_2$ erzeugt: Ist $\vec{a} = (a_1, a_2, a_3) \in V'$ beliebig, so gilt definitionsgemäß $a_1 = a_3$ und folglich

$$\begin{aligned} \vec{a} &= (a_1, a_2, a_1) = \Big(a_2 + (a_1 - a_2), a_2, a_2 + (a_1 - a_2)\Big) \\ &= (a_1 - a_2) \cdot (1, 0, 1) + a_2 \cdot (1, 1, 1) = (a_1 - a_2) \cdot \vec{b}_1 + a_2 \cdot \vec{b}_2. \end{aligned}$$

Also ist $\left\{\vec{b}_1, \vec{b}_2\right\}$ eine Basis von V'.

Das Beispiel des Vektorraums der reellen Polynome legt nahe, auch unendliche Basen einzuführen:

1.3.14 Definition

Sei V ein reeller Vektorraum. Eine Teilmenge $\{x_i : i \in I\} \subset V$ (mit beliebiger Indexmenge $I \neq \emptyset$) heißt **Basis** *von V, falls gilt:*

(1) *Jedes endliche Teilsystem $x_{i_1}, \dots, x_{i_n}$ mit $n \in \mathbb{N}$ und paarweise verschiedenen Indizes $i_1, \dots, i_n \in I$ ist linear unabhängig.*

(2) *Jeder Vektor aus V ist eine (endliche!) Linearkombination von Elementen $x_{i_1}, \dots, x_{i_n}$, $n \in \mathbb{N}, i_1, \dots, i_n \in I$.*

1.3.15 Beispiel

Die Menge aller *Monome* $\{e_n = z^n : n \in \mathbb{N}_0\}$ ist eine Basis des Vektorraums aller reellen Polynome.

Dieser Abschnitt schließt mit einigen Bemerkungen über endlichdimensionale Vektorräume.

1.3.16 Satz

Sei V ein reeller Vektorraum mit $\dim V = n < \infty$. Die Vektoren $x_1, \dots, x_n \in V$ seien linear unabhängig. Dann ist $\{x_1, \dots, x_n\}$ eine Basis von V.

Jedes linear unabhängige System von maximaler Länge ist eine Basis.

Beweis

Es ist zu zeigen: $x_1, \dots, x_n$ erzeugen V. Sei $x \in V$ gegeben. Nun sind $x, x_1, \dots, x_n$ linear abhängig, andernfalls wäre $\dim V > n$. Nach Satz 1.3.8 ist x Linearkombination von $x_1, \dots, x_n$. □

In Abschnitt 3.3 werden wir sehen, daß alle Basen dieselbe Anzahl von Elementen haben und daß deshalb umgekehrt *jede* Basis ein linear unabhängiges System von maximaler Länge ist.

Schließlich klären wir die Existenzfrage für Basen:

1.3.17 Satz

Jeder endlichdimensionale reelle Vektorraum besitzt eine Basis.

Beweis

Sei $\dim V = n$. Nach Definition 1.3.10 der Dimension existieren n linear unabhängige Vektoren $x_1, \dots, x_n \in V$, welche nach Satz 1.3.16 eine Basis bilden. □

Ein entsprechender Satz gilt auch in beliebigen Vektorräumen, spielt dort aber eine nicht so zentrale Rolle und erfordert einen tiefliegenden Beweis. Unter anderem muß das „Zornsche Lemma“ verwendet werden. Deshalb verweisen wir hier auf die Literatur, z. B. auf H.-J. Kowalsky [9, S. 37f].

2 Der anschauliche Raum

2.1 Gibt es eine geometrische Anschauung?

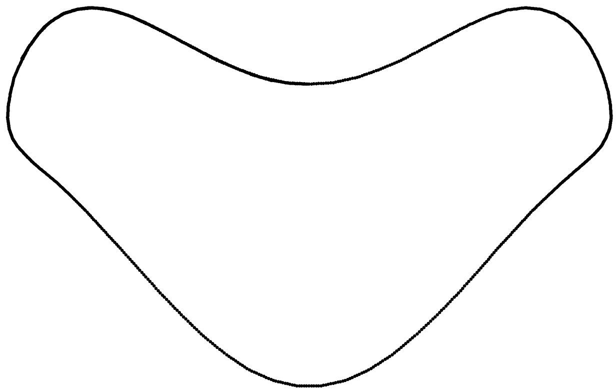
Die historische Entwicklung des Geometriebegriffes ist sehr wichtig für die Art unseres Denkens im allgemeinen. Kant meinte, daß die Vernunft vorschreibt, daß die Geometrie euklidisch ist. Später entdeckten die Mathematiker C. F. Gauß (1777-1855), N. I. Lobačevskij (1793-1856) und J. Bolyai (1802-1860), daß es noch andere Geometrien gibt, die genauso widerspruchsfrei aufgebaut werden können wie die euklidische. Man muß formale logische Konstruktionen durchführen! Man muß viel rechnen!

Wir meinen jedoch, daß dem natürlichen Menschen eine Anschauung gegeben ist, die ihn für sein tägliches Leben mit einer Geometrie ausrüstet. Statt sprachlich darlegen kann man einfach schauen. Die geometrische Anschauung ist also sehr praktisch! Schon Goethes Faust verlangte nach einer alles umfassenden Schau: „Daß ich nicht mehr mit saurem Schweiß zu sagen brauche, was ich nicht weiß... Schau alle Wirkenskraft und Samen und tu nicht mehr in Worten kramen.“ – Wie anders ist es bei der Sprache! Wenn man z. B. etwas in der mathematischen Textsprache $\TeX$ zeichnen will, muß man alles ausrechnen. Manchmal dauert das sehr lange. Man stelle sich etwa vor, Rembrandt hätte seine Bilder auf solche Weise malen müssen!

Räumliche Anschauung und Logik

Gewiß gibt es Menschen, die nicht zwischen Anschauung und Logik unterscheiden können. Das erinnert an die Zeit von Goethe, wo niemand sehen konnte, daß Licht und Farbe etwas Verschiedenes sind. Und doch ist das sehr einfach. Die gleiche Farbe kann ja durch ganz verschiedene Mischungen von Licht verschiedener Wellenlänge hervorgebracht werden!

Es gibt jedoch sehr wichtige Argumente dafür, sorgfältig zwischen Anschauung und Logik zu unterscheiden. In der ebenen Geometrie gilt z.B. der Jordansche Kurvensatz. Anschaulich sind nur stückweise glatte Kurven. Aber schon in diesen Fall ist es mühsam, den Satz logisch zu beweisen. In der Anschauung ist er dagegen völlig trivial. Man sieht direkt: Die Ebene wird durch jede einfach geschlossene Kurve in zwei disjunkte Teilgebiete zerlegt.



Der Jordansche Kurvensatz

A. Einstein (1879-1955), der 1896-1900 in Zürich Physik studiert hat und auch Vorlesungen bei H. Minkowski (1864-1909) gehört hat, sah, daß in der Physik Raum und Zeit vom gleichen Wesen sind: Das Einsteinsche Relativitätsprinzip besagt: *Zwei (mathematische) Inertialsysteme sind völlig gleichberechtigt. Wenn man aber von einem auf das andere transformiert, werden Raum- und Zeitgrößen durcheinander gemischt.* Diese müssen also in der Logik von gleicher Art sein. In der Anschauung sind dagegen Raum und Zeit etwas völlig Verschiedenes. Eine Einsteinsche Folgerung aus dem Relativitätsprinzip war z. B. die Zeitdilatation. Sie ist anschaulich gesehen etwas völlig Unsinniges.

Was ist nun eigentlich Anschauung? Viele sagen, sie sei Gewohnheit. Aber das ist keine Beschreibung. Wenn man ein räumliches Bild sieht, so ist zwar der Blick auf einen Punkt gerichtet. Aber die ganze Figur steht doch gleichzeitig zur Verfügung, sie ist gegenwärtig. Man kann all ihre Teile ausnutzen, um zu neuen Erkenntnissen zu kommen. Das ist, als ob man an der Figur Experimente macht. Die Anschauung (rein als Verarbeitung von Realität gesehen) ist also dem Werk eines Parallelcomputers oder eines Analogcomputers ähnlich. – Bei der Logik ist immer nur ein Ergebnis da. Sie ist wie ein einfacher Digitalcomputer.

Kant und Einstein

Der Philosoph Immanuel Kant (1724-1804) schuf eine Theorie der Erkenntnis. Die Erkenntnis war nicht einfach eine Spiegelung der realen Welt. Wesentlich benutzte sie das in der Vernunft vorgegebene Transzendentale. Dazu gehörte auch die Anschauungsform „Raum". Dieser Raum war mit der euklidischen Geometrie ausgerüstet. Man konnte alles nur in dieser Geometrie sehen.

Die Vernunft konnte sich nur auf (sinnliche) Dinge richten, die „gegenständlich" waren. Sonst kam man zu Widersprüchen, sog. Antinomien. Unendliches schied somit aus. Im Falle der Gegenständlichkeit war die Erkenntnis aber wahr, also real und unabhängig vom Menschen richtig. Die Dinge der Welt waren in der euklidischen Geometrie.

Genau diese Geometrie war vernünftig. Es gab dabei auch Urteile „a priori“. Sie benutzten die Welt außerhalb unserer Vernunft nicht, waren aber gültig.

Die Aufklärung hatte gelehrt, daß die Vernunft der Gott der Welt ist. Wir wissen heute, daß es eine menschliche Vernunft, die ohne Einschränkung vernünftig ist, gar nicht gibt. – Im Gegensatz zu Kant führte bei Einstein die Anschauung auch im gegenständlichen Bereich zu einem Widerspruch mit den Experimenten. Kant war damit widerlegt.

Man erfährt auch leicht durch reines Nachdenken, daß die Vernunft keine spannungslose Einheit ist. Sie besteht aus verschiedenen Organen, die durchaus miteinander in Konflikt geraten können. Man hat die Anschauung und die Logik. Harmonie besteht nur in solchen Teilen, die von der biologischen Evolution geformt wurden.

Die Mathematiker

Was aber taten die Mathematiker? Sie verzichteten ganz auf Anschaulichkeit und machten logische Konstruktionen. Aber sicherlich meinte Kant die anschauliche Geometrie und nicht ein logisches Konstrukt. Er konnte allerdings wohl kaum effektiv unterscheiden. Die Mathematiker sahen, daß auch andere Geometrien aufgebaut werden können, so daß sie genauso widerspruchsfrei sind wie die euklidische. So erhielt man z. B. die hyperbolische Geometrie. Gauß, Lobačevskij und Bolyai zeigten das zu Anfang des 19. Jahrhunderts. In der hyperbolischen Geometrie sind alle Axiome gültig mit Ausnahme des Parallelenaxioms: Wenn in der hyperbolischen Ebene g eine Gerade und P ein Punkt außerhalb von g ist, so gibt es nicht gerade eine Gerade durch P, die g nicht schneidet, sondern unendlich viele. Alle geistigen Menschen sahen irrtümlich einen Widerspruch zu Kant, und Kant war unfehlbar. Gauß fürchtete sich, er fürchtete das Geschrei der „Böotier“ (im Altertum: Bauern um Theben in Griechenland) so sehr, daß er seine Ideen nicht publizierte. Das tat dann Lobačevskij, der etwas weiter weg lebte.

Unsere Anschauung

Die Objekte der Anschauung sind ganz anders als die der Mathematik. Es seien einige Beispiele angegeben. Wie wir gesehen haben, hat der $\mathbb{R}^n$ mindestens die mathematische Dimension n, und n kann beliebig groß sein. In der Anschauung können wir uns jedoch nicht mehr als 3 Geraden vorstellen, die in einem Punkte paarweise aufeinander senkrecht stehen. Wir haben ferner Punkte, Kurven, Flächen. Als nächst Höheres kommt das Gebiet, das schon lokal den ganzen Raum ausfüllt. Darüber hinaus geht es nicht! Ist der anschauliche Raum isomorph zu einem $\mathbb{R}^n$, so muß daher n gleich 3 sein.

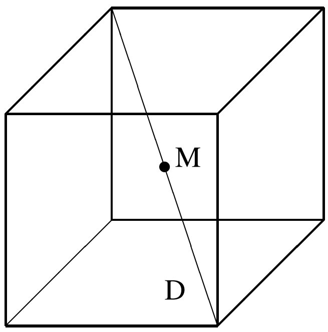
Interessant ist auch die folgende Betrachtung. Wenn man eine Schrift in einem Spiegel betrachtet, so erscheint die Spiegelschrift. Die Schrift sei etwa auf einem senkrechtstehenden Blatt Papier, der Spiegel stehe aufrecht. Dann werden durch

Spiegelung der Schrift „links“ und „rechts“ vertauscht, aber nicht „oben“ und „unten“. Wodurch kommt diese Bevorzugung zustande? Nach einigem Denken wird klar: Das kommt daher, daß die letztere Richtung durch die Schwerkraft definiert wird und daß diese deshalb in unserem Denken eine ausgezeichnete Rolle spielt. Beim Betrachten der Spiegelschrift wird diese nicht geändert. Da die Spiegelung die Orientierung umdreht, bleibt nur die Vertauschung von „rechts“ und „links“ (vom Betrachter aus gesehen). So hat der Raum, wie die Alten glaubten, eine Richtung. Ein Stein fällt, indem er dieser Richtung folgt. Erst später hat Newton entdeckt, daß der Stein von der Schwerkraft bewegt wird. Zwei Massen ziehen sich immer radial an.

Die Anschauung wird durch eine besondere Aktivität unseres neurophysiologischen Systems hervorgerufen. Wir können das bei dem folgenden Geschehen beobachten: Wenn wir des Morgens auf eine schwarz-weiße Karte in einer Zeitung schauen, dann halten wir manchmal zunächst die Meere für Länder und die Länder für Meere. Erst nach einer Weile springt die Sicht um. Das sollte bedeuten, daß in unserem Nervensystem etwas geschehen ist. Es gibt dafür noch bessere Beispiele. In speziellen Zeichnungen sieht man manchmal 6 Würfel und manchmal nur 5. Auch hier findet ein Umspringen statt: es ist ein psychologisches Ereignis. Man hat nun versucht, das anatomische Korrelat dazu zu finden. Es wurden dazu Experimente durchgeführt (Kernspintomograph). Beide Sichtweisen korrespondieren mit Zuständen im Gehirn, wo zwei Mengen von Milliarden von Neuronen synchron oszillieren. Beim Umspringen werden die Mengen geändert. Man vergleiche etwa den interessanten Vortrag, den Wolf Singer (Frankfurt) im April 1995 vor der Leopoldina in Halle gehalten hat: W. Singer, Funktionelle Organisation der Großhirnrinde, Nova Acta Leopoldina, Neue Serie, Bd. **72 (294)** (1996), S. 5ff.

Der Sieg der Logik

Die Anschauung bringt allerdings nur wenig Möglichkeiten, tiefere mathematische Verhältnisse zu sehen. Das gilt schon in der 3-dimensionalen Geometrie. Man betrachte etwa folgende Zeichnung:



Was schneidet die Ebene, die im Punkt M senkrecht zur Diagonalen D ist, aus dem Würfel heraus?

Die Antwort ist: ein regelmäßiges Sechseck. Aber das ist nicht sichtbar. Wir haben es zu beweisen. Mit den Methoden der Logik!

Einstein zeigte, daß man in der Physik mit dem anschaulichen Raum nicht auskommt. Man muß mindestens das 4-dimensionale Raum-Zeit-Kontinuum konstruieren.

Man muß also Logik statt Anschauung verwenden! Wir ersetzen deshalb fortan den anschaulichen Raum durch den $\mathbb{R}^3$ mit euklidischer Norm. Wir werden zeigen, daß beide zueinander isomorph sind. Um das zu tun, benutzen wir die Axiome der Geometrie. Das sind Aussagen, die sich nicht mehr logisch herleiten lassen. Sie müssen „evident“ sein. Evident bedeutet, daß es eine Anschauung gibt, in der man das Axiom sieht. Der Grund für seine Gültigkeit ist also nicht mathematischer, sondern psychologischer Natur!

Man merkt bald, daß man nicht ins Unendliche schauen kann. Wichtig, um die Axiome zu bekommen, ist aber, daß der anschauliche Raum im Großen wie im Kleinen aussieht. Das folgt nun, weil man zunächst nur den Raum als begrenzte Kugel hat und den weiteren Raum dann nach einer festen Vorschrift konstruieren muß. Er ist dann nie aktual ganz da, aber man kann mit der Konstruktion beliebig fortfahren. Aristoteles nannte das *das Potentialunendliche.* Aktual ist der Raum nur eine endliche Kugel. Er hat deshalb auch einen Mittelpunkt. Aristoteles glaubte, daß in diesem Mittelpunkt die Erde steht.

2.2 Addition von Vektoren in der anschaulichen Ebene

Wir sollten eigentlich den anschaulichen Raum behandeln. Es ist aber einfacher, eine Dimension herunterzugehen. Die *Ebene* besteht aus Punkten. Wir wählen einen festen Basispunkt O und nehmen alle Pfeile, die von einem Punkt zu einem weiteren führen. Wir stellen dann Axiome auf.

In den Axiomen der Geometrie, die von Euklid aufgestellt wurden, wird der ganze unendliche Raum benutzt. Aber seine Unendlichkeit ist nicht anschaulich. Sie wird aber bei dem Parallelenaxiom benutzt. Es ist deshalb nicht verwunderlich, daß es zweifelhaft wurde. In dem Axiomensystem, das hier entwickelt werden soll, leben alle Axiome in beschränkten Teilbereichen, die natürlich beliebig groß sein können. Sie sind immer anschaulich, und die Axiome sind evident. Hätte man immer dieses Axiomensystem gehabt, so wäre man (aus psychologischen Gründen) wohl nie zu Zweifeln an der euklidischen Geometrie gekommen. Diese Geometrie wäre als die einzig wahre angesehen worden!

Es sei E die anschauliche Ebene.

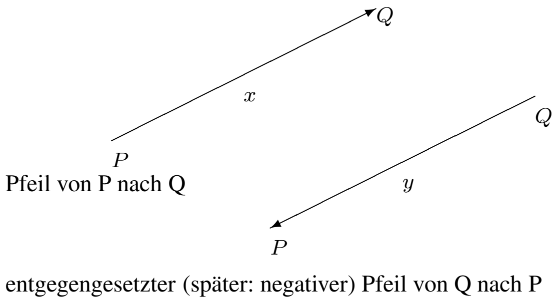
Axiom I

E ist eine nicht-leere Menge. Die Elemente $P \in E$ heißen **Punkte**.

2.2.1 Definition

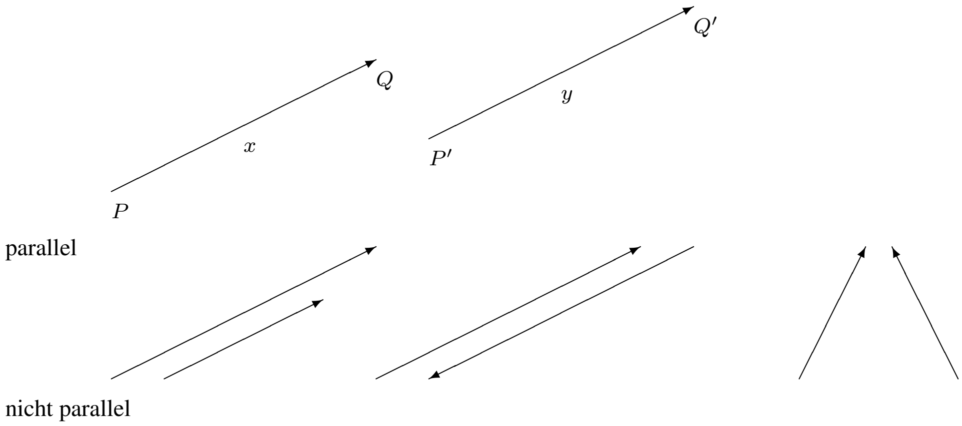
Ein **Pfeil** $x = (P, Q)$ *ist ein geordnetes Paar von Punkten* $P, Q \in E$.

Wir zeichnen x als einen wirklichen Pfeil, der sich von P nach Q erstreckt.



Wenn $y = (Q, P)$ der Pfeil in entgegengesetzter Richtung ist und wenn $P \neq Q$, so gilt $x \neq y$. Wir werden mit X die Menge der Pfeile x bezeichnen.

Der Begriff „voll parallel“ ist evident, wenigstens in kleinen Bereichen und dann auch in großen, weil diese zu kleinen isomorph sind. Die Bedeutung dieses Wortes für Pfeile ist: gleiche Länge und gleiche Richtung. Die Evidenz der folgenden Axiome folgt also aus der Evidenz in kleinen Bereichen.



Wir müssen die wichtigsten Eigenschaften dieses neuen Begriffs in Axiomen fassen. Dazu ziehen wir heran:

2.2.2 Definition

Eine **(binäre) Relation** $\sim$ *auf einer Menge* M *liegt vor, wenn für je zwei Elemente* $x,\ y$ *von* M *stets festgelegt ist, ob* $x \sim y$ *gilt oder nicht.*

Aus der Anschauung folgt:

Axiom II

In X ist eine binäre Relation $\|$ (in Worten: **voll parallel***) gegeben.*

Axiom III (Parallelen-Axiom)

Es sei $x = (P, Q) \in X$ und $P_0 \in E$. Dann gibt es genau einen Punkt $Q_0 \in E$ so daß $y = (P_0, Q_0) \parallel x$ gilt.

Die Mathematik befaßt sich nicht mit der Frage, wie die Anschauung entsteht. Wichtig ist nur die Gültigkeit der Axiome. Deshalb können wir fortfahren:

Axiom IV

Die binäre Relation $\|$ ist eine Äquivalenzrelation.

Dabei heißt eine binäre Relation „$\sim$" eine *Äquivalenzrelation*, wenn folgende drei Gesetze gelten:

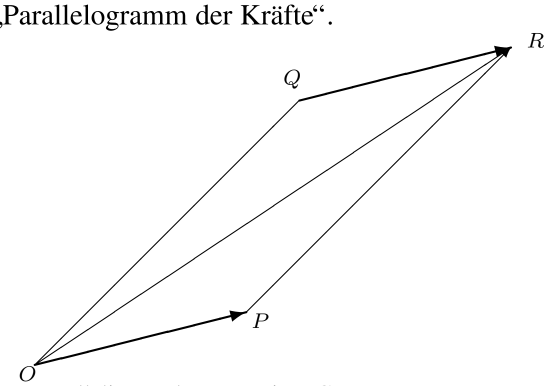
(1) Es ist immer $x \sim x$. (Reflexivität)

(2) Wenn $x \sim y$, dann gilt auch $y \sim x$. (Symmetrie)

(3) Wenn $x \sim y$ und $y \sim z$, dann folgt $x \sim z$. (Transitives Gesetz)

Ein anderes einfaches Beispiel einer Äquivalenzrelation ist die Gleichheit. In der Analysis wird ferner die Ungleichung $<$ eingeführt. Sie ist keine Äquivalenzrelation.

Das nächste Gesetz ist allen Physikern und Ingenieuren wohlbekannt. Sie sprechen vom „Parallelogramm der Kräfte".



Parallelismus, kommutatives Gesetz

Axiom V (Parallelogramm-Axiom)

Es seien für $O, P, Q, R \in E$ die Pfeile $(O, P) \parallel (Q, R)$. Dann folgt $(O, Q) \parallel (P, R)$.

Dieses Axiom ist wegen der Translationen evident:

Sind $P_0, P_1 \in E$ zwei Punkte, so gibt es immer genau eine Translation in E, die P_0 auf P_1 abbildet. Dabei geht jeder Vektor in einen zu ihm voll parallelen über. Es sei nun F

die Translation mit $F(O) = Q$. Dann haben wir wegen $(O, P) \parallel (Q, R)$ die Gleichheit $F(P) = R$. Aber bei Translationen sind auch immer die Verbindungsvektoren der Punkte mit den Bildpunkten zueinander voll parallel. Also folgt $(O, Q) \parallel (P, R)$. $\square$

Wie bei Aristoteles denken wir uns E mit einem Mittelpunkt O versehen. An sich sind die Axiome nur einsehbar, wenn man sich in der Nähe von O befindet. Wie schon gesagt, sieht der 3-dimensionale unendliche Raum im Großen so aus wie im Kleinen. Man vgl. den Aufsatz [5] von H. Grauert. Das Gleiche gilt aber auch für die Ebene E. Denn man erhält sie als Fixpunktmenge einer Spiegelung des Raumes und die Spiegelung ist natürlich auch im Großen so wie im Kleinen.

Vektorraumaxiome

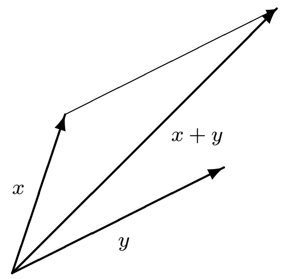
Wir wollen zeigen, daß die anschauliche Ebene ein Vektorraum im Sinne des ersten Kapitels ist. Dazu müssen wir beweisen, daß die grundlegenden Eigenschaften aus der dortigen Definition 1.2.1 erfüllt sind. Wir nennen diese Eigenschaften *Vektorraumaxiome*. Man muß jedoch sagen, daß sie keine Axiome im eigentlichen Sinne sind. Sie sind aus den eigentlichen, anschaulichen Axiomen hergeleitet und sind die Grundbegriffe des Vektorraums. Unsere ursprünglichen Axiome sind evident aus der Anschauung heraus. Das ist so ähnlich wie in der Mengenlehre. Auch die Mengenlehre ist auf unmittelbar einleuchtenden Axiomen (Zermelo) aufgebaut.

Wir müssen uns allerdings zunächst auf die Addition beschränken. Wir haben dafür inzwischen ein vollständiges Axiomensystem erhalten. Wir nennen unsere Pfeile fortan *Pfeilvektoren* oder auch einfach Vektoren. Es folgen dann aus den erkannten Axiomen einige mathematische Sätze, nämlich die Vektorraumaxiome.

Fortan sei $O \in E$ ein fester Punkt. Wir nennen ihn *Basispunkt* oder auch *Aufpunkt*. Die Menge $V = \{(O, P) : P \in E\} \subset X$ heißt die *Menge der Vektoren in O*. Wir numerieren die Vektorraumaxiome wie in Kapitel 1.

Es kann jetzt die *Vektoraddition* eingeführt werden. Es seien $x = (O, P),\ y = (O, Q)$ Vektoren aus V. Dann gibt es einen wohlbestimmten Punkt $R \in E$ mit $(O, Q) \parallel (P, R)$. Wir setzen $x + y = (O, P) + (O, Q) := (O, P) + (P, R) := (O, R) \in V$. Dadurch ist die Addition definiert, die man anschaulich so sehen kann:

Vektoraddition



Das kommutative Gesetz läßt sich nun ableiten:

Vektorraumaxiom (4)

Es gilt stets $x + y = y + x$.

Beweis

Wir schreiben $x = (O, P)$, $y = (O, Q)$. Es sei $(O, Q) \parallel (P, R)$. Wir erhalten $(O, P) \parallel (Q, R)$ nach dem Parallelogramm-Axiom V. Deswegen ist $y + x = (O, Q) + (O, P) = (O, Q) + (Q, R) = (O, R) = x + y$. □

2.2.3 Definition

Wir nennen den Vektor (P, P) *den* **Nullvektor in** P**.**

2.2.4 Satz

Es ist immer $(P, P) \parallel (Q, Q)$.

Beweis

Wir haben auf Grund der Reflexivität aus Axiom IV $(P, Q) \parallel (P, Q)$. Also ergibt das Parallelogramm-Axiom $(P, P) \parallel (Q, Q)$. □

2.2.5 Definition

Es heißt $\mathcal{O} := (O, O)$ *der* **Nullvektor** *von* V.

Vektorraumaxiom (2)

In V *gilt immer* $x + \mathcal{O} = x$.

Beweis

Wir wissen, daß $\mathcal{O} \parallel (P, P)$. Also erhalten wir $(O, P) + \mathcal{O} = (O, P) + (P, P) = (O, P)$. □

Bei gegebenem (O, P) gibt es gemäß dem Parallelen-Axiom III genau einen Punkt $Q \in E$ mit $(O, Q) \parallel (P, O)$.

2.2.6 Definition

Der Vektor $-(O, P) := (O, Q) \in V$ *heißt das* **Negative** *von* (O, P).

Vektorraumaxiom (3)

In V *gilt stets die Gleichung* $x + (-x) = \mathcal{O}$.

Beweis

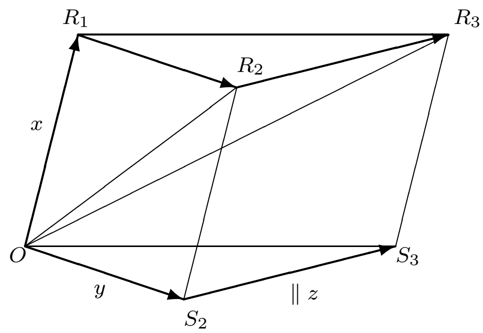
Man hat $(O, P) + (-(O, P)) = (O, P) + (O, Q) = (O, P) + (P, O) = (O, O) = \mathcal{O}$. □

Wir werden jetzt zeigen, daß für die Addition auch das assoziative Gesetz gilt.

Vektorraumaxiom (1)

In V gilt stets die Gleichung $(x + y) + z = x + (y + z)$.

Beweis



Das assoziative Gesetz

Es sei $x = (O, R_1)$, $y = (O, S_2)$, $z = (O, Z)$. Dann gibt es ein R_2 mit $(O, S_2) \parallel (R_1, R_2)$ und ein R_3 mit $(O, Z) \parallel (R_2, R_3)$, und weiter ein S_3 mit $(O, Z) \parallel (S_2, S_3)$. Nach dem Parallelogrammaxiom erhalten wir sukzessive $(O, R_1) \parallel (S_2, R_2)$, $(S_2, R_2) \parallel (S_3, R_3)$, $(O, R_1) \parallel (S_3, R_3)$ und schließlich $(O, S_3) \parallel (R_1, R_3)$. Also haben wir $(x + y) + z = \Big((O, R_1) + (O, S_2)\Big) + (O, Z) = \Big((O, R_1) + (R_1, R_2)\Big) + (R_2, R_3) = (O, R_3)$ und andererseits $x + (y + z) = (O, R_1) + \Big((O, S_2) + (O, Z)\Big) = (O, R_1) + \Big((O, S_2) + (S_2, S_3)\Big) = (O, R_1) + (O, S_3) = (O, R_1) + (R_1, R_3) = (O, R_3)$. □

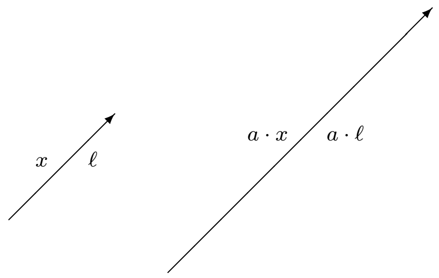
Damit ist bewiesen:

2.2.7 Satz

$(V, +)$ ist eine abelsche Gruppe.

2.3 Multiplikation mit Skalaren

Wir wollen zunächst wieder die durch die räumliche Anschauung gegebenen Axiome betrachten. Bei den Alten Griechen waren die reellen Zahlen Streckenlängen, also sehr anschaulich. Wir nehmen deshalb diese Zahlen als Skalare, um zu einer Vektorraumstruktur zu gelangen. Wir bezeichnen mit $\mathbb{R}_0^+ = \{a \in \mathbb{R} : a \geq 0\}$ die Menge der nicht-negativen reellen Zahlen. Es ist evident: *Ist $x = (O, P) \in V$ ein Vektor, so hat er eine Länge $\ell \in \mathbb{R}_0^+$ und eine Richtung (wenn er von $\mathcal{O}$ verschieden ist). Zu $a \in \mathbb{R}_0^+$ gibt es den wohlbestimmten Vektor $a \cdot x$ mit der gleichen Richtung und der Länge $a \cdot \ell$.*



Skalarenmultiplikation

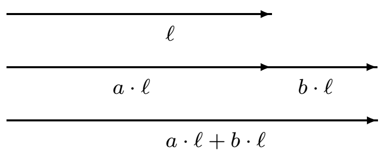
Axiom VI

Wir haben eine binäre Operation $\mathbb{R}_0^+ \times V \mapsto V : (a, x) \to a \cdot x \in V$.

Wenn $a, b \in \mathbb{R}_0^+$ sind, so haben $(a + b) \cdot x$ und $ax + bx$ die gleiche Richtung wie x. Aber die Längen sind $(a+b) \cdot \ell$ und $a\ell + b\ell$. Das ist auf Grund des Distributivgesetzes für die reellen Zahlen gleich.

Axiom VII

Es gilt immer $(a + b) \cdot x = ax + bx$, *wenn* $a, b \in \mathbb{R}_0^+$.



eingeschränktes Distributivgesetz

Auch das assoziative Gesetz ist richtig:

Axiom VIII

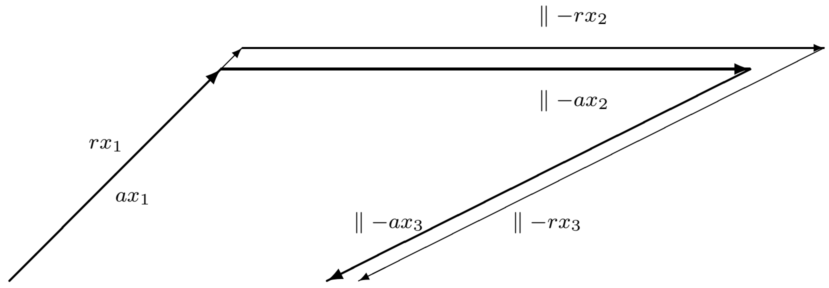
Es ist immer $(a \cdot b) \cdot x = a \cdot (b \cdot x)$, *wenn* $a, b \in \mathbb{R}_0^+$.

Unmittelbar einsichtig ist:

Axiom IX

Es gilt das unitäre Gesetz $1 \cdot x = x$.

Um weitere Gesetze für die Skalarmultiplikation zu erhalten, benötigen wir die Stetigkeit der Vektoroperationen. Auch das ist evident. Man hat nämlich: Wenn $x = ax_1 - ax_2 - ax_3$ mit $x_1, x_2, x_3 \in V, a \in \mathbb{R}_0^+$ ist, dann muß bei $r \approx a,\ r \in \mathbb{R}_0^+$, der Vektor $y = rx_1 - rx_2 - rx_3$ sehr nahe bei x sein.



$ax_1 - ax_2 - ax_3 \approx rx_1 - rx_2 - rx_3;\ r/a = 1{,}1$

Axiom X (Stetigkeit)

Wenn es bei $x = ax_1 - ax_2 - ax_3$ *in beliebiger Nähe von* a *Zahlen* $r \in \mathbb{R}_0^+$ *mit* $rx_1 - rx_2 - rx_3 = \mathcal{O}$ *gibt, so muß auch* $x = \mathcal{O}$ *gelten.*

Die Formulierung „in beliebiger Nähe von a gibt es ein $r \in \mathbb{R}_0^+$" ist nur scheinbar nicht exakt. Sie bedeutet nämlich: „zu jedem $\varepsilon > 0$ gibt es ein $r \in \mathbb{R}_0^+$ mit $|a - r| < \varepsilon$".

Folgerungen

2.3.1 Satz

Es gilt $0 \cdot x = \mathcal{O}$.

Beweis

Man hat nach Axiom VII: $ax = (a + 0) \cdot x = ax + 0 \cdot x$. Daraus folgt $0 \cdot x = \mathcal{O}$ wie in jeder Gruppe (man vgl. die Bemerkung 1.1.5). □

2.3.2 Definition

Ist $a \le 0$, *so setzen wir* $a \cdot x = -\Big((-a) \cdot x\Big)$.

Durch diese Festsetzung wird $a \cdot x$ für alle $a \in \mathbb{R}$ ohne Zweideutigkeit definiert. Sie könnte nur auftreten, wenn $a = 0$ ist. Gilt aber das, so folgt aus dem Axiom und nach der Festsetzung $a \cdot x = \mathcal{O}$. Liegt keine Zweideutigkeit vor, so verwenden wir fortan das Wort *wohldefiniert*.

2.3.3 Satz

Man hat für alle $a \in \mathbb{R}$ *die Gleichung* $(-a) \cdot x = -ax$.

Beweis

Ist $a \ge 0$, so ist $-a \le 0$. Nach unserer Festsetzung ist dann $(-a) \cdot x = -\Big((-(-a)) \cdot x\Big) = -a \cdot x$. Wenn $a \le 0$, so erhalten wir aus Definition 2.3.2 $-\Big((-a) \cdot x\Big) = a \cdot x$ und daraus wie in jeder Gruppe $-a \cdot x = (-a) \cdot x$ (vgl. Bemerkung 1.1.5). □

Es wird nun schrittweise das allgemeine Distributivgesetz $a \cdot (x+y) = a \cdot x + a \cdot y$ hergeleitet.

2.3.4 Satz

Wenn $n \in \mathbb{N}$ *ist, dann ist* $n \cdot x = \underbrace{x + \ldots + x}_{n\text{-mal}}$.

Beweis

Wir führen *vollständige Induktion* nach n. Die Aussage hängt von $n \in \mathbb{N}$ ab. Der Fall $n = 1$ ist richtig nach Axiom IX. Wir nennen das die *Induktionsbasis*. Es sei nun vorausgesetzt, daß $n > 1$ und der Fall $n-1$ schon bewiesen ist. Das ist dann die *Induktionsannahme*. Nun haben wir nach Axiom VII: $n \cdot x = (n-1) \cdot x + x = \underbrace{x + \ldots + x}_{(n-1)\text{-mal}} + x = \underbrace{x + \ldots + x}_{n\text{-mal}}$. Aus der Richtigkeit der Fälle für kleineres n folgt also die Aussage für n. Wir sagen, daß der *Induktionsschluß* durchführbar ist. Gilt für eine Folge von Aussagen die Induktionsbasis und immer der Induktionsschluß, so heißt die Induktion *vollständig*. Die Aussagen sind dann für alle n gültig. Unser Satz ist also bewiesen. □

2.3.5 Satz

Es gilt stets $n \cdot (x+y) = n \cdot x + n \cdot y$.

Beweis

Neben dem letzten Satz werden das allgemeine Assoziativgesetz und das Kommutativgesetz für die Addition in V benuzt.

$$n \cdot (x+y) = \underbrace{(x+y) + \ldots + (x+y)}_{n\text{-mal}} = \underbrace{(x + \ldots + x)}_{n\text{-mal}} + \underbrace{(y + \ldots + y)}_{n\text{-mal}} = n \cdot x + n \cdot y.$$

□

2.3.6 Satz

Es gilt für alle $n, m \in \mathbb{N}$ *die Gleichung:*

$$\frac{n}{m} \cdot (x+y) = \frac{n}{m} \cdot x + \frac{n}{m} \cdot y.$$

Beweis

Wir erhalten $m \cdot (\frac{n}{m} \cdot x + \frac{n}{m} \cdot y) = m\frac{n}{m}x + m\frac{n}{m}y = nx + ny = n \cdot (x+y)$. Die Multiplikation mit $\frac{1}{m}$ ergibt: $\frac{n}{m} \cdot x + \frac{n}{m} \cdot y = \frac{n}{m} \cdot (x+y)$. □

Wir nutzen nun das Stetigkeitsaxiom aus:

2.3.7 Satz

Für alle $a \in \mathbb{R}_0^+$ *ist* $a \cdot (x+y) = a \cdot x + a \cdot y$.

Beweis

Die Zahl a ist ein unendlicher Dezimalbruch. Es gibt in beliebiger Nähe von a endliche Dezimalbrüche, also rationale Zahlen $r = \frac{n}{m}$ mit $n, m \in \mathbb{N}$. Definieren wir $z = a \cdot (x+y) - ax - ay$, so gilt immer $r \cdot (x+y) - rx - ry = \mathcal{O}$. Wegen des Stetigkeitsaxioms folgt dann $z = \mathcal{O}$. □

Vektorraumaxiome

Wir werden jetzt die weiteren Axiome des Vektorraums ableiten.

Vektorraumaxiom (7a)

Für alle reellen Zahlen a *gilt die Gleichung* $a \cdot (x+y) = ax + ay$ *(distributives Gesetz).*

Beweis

Wir müssen nur noch den Fall $a < 0$ betrachten. Man hat

$$\begin{aligned} a \cdot (x+y) &= -\Big((-a) \cdot (x+y)\Big) = -\Big((-a) \cdot x + (-a) \cdot y\Big) \\ &= -\Big((-a \cdot x) + (-a \cdot y)\Big) = -\Big(-(ax+ay)\Big) = ax + ay, \end{aligned}$$

wie es in jeder abelschen Gruppe richtig ist. □

Vektorraumaxiom (7b)

Es gilt das distributive Gesetz $(a+b) \cdot x = ax + bx$ *für alle reellen Zahlen* a, b.

Beweis

Das gilt nach Axiom VII im Falle $a \geq 0,\ b \geq 0$. Es sei nun $a < 0,\ b \geq 0,\ a+b \geq 0$. Dann folgt $(a+b) \cdot x + (-a) \cdot x = (a+b-a)x = bx$. Das bedeutet: wir haben $(a+b) \cdot x - ax = bx$. Also ergibt sich $(a+b)x = ax + bx$. Alle anderen Fälle können analog gelöst werden. □

Um das assoziative Gesetz für die Skalarmultiplikation zu erhalten, beweisen wir:

2.3.8 Satz

Es gilt für alle $a \in \mathbb{R}$ *die Gleichung* $a \cdot \mathcal{O} = \mathcal{O}$.

Beweis

Wir haben $a \cdot x = a \cdot (x + \mathcal{O}) = ax + a \cdot \mathcal{O}$ und dann $a \cdot \mathcal{O} = \mathcal{O}$, wie es in jeder Gruppe richtig ist. – Hier wurde das Vektorraumaxiom (7a) benutzt. □

2.3.9 Satz

Für alle $a \in \mathbb{R}$ gilt: $a \cdot (-x) = -a \cdot x$.

Beweis

Wir haben $ax + a(-x) = a \cdot \Big(x + (-x)\Big) = a \cdot \mathcal{O} = \mathcal{O}$. Also folgt wie in jeder Gruppe $a \cdot (-x) = -ax$. □

Schließlich erhalten wir:

Vektorraumaxiom (6)

Für alle reellen Zahlen a, b gilt $a \cdot (b \cdot x) = (ab) \cdot x$.

Beweis

Ist $a \geq 0,\ b \geq 0$, so folgt dieses aus Axiom VIII. Es sei nun $a < 0$ und $b \geq 0$. Dann erhalten wir $a \cdot (b \cdot x) = -\Big((-a) \cdot (bx)\Big) = -\Big((-ab) \cdot x\Big) = -(-ab)x = (ab)x$. Die Fälle $a \geq 0, b < 0$ und $a < 0, b < 0$ können in ähnlicher Weise behandelt werden. □

Nachdem das Vektorraumaxiom (5) bereits in unserem Axiom IX enthalten ist, haben wir damit bewiesen, daß V den Axiomen eines reellen Vektorraums genügt.

2.3.10 Satz

V ist ein reeller Vektorraum.

Dimension

Wir werden sehen, daß V 2-dimensional ist. Dafür gebrauchen wir den Begriff der Geraden, den wir hier aus den Axiomen definieren müssen. Von $\mathcal{O}$ verschiedene Vektoren gibt es nach dem folgenden Axiom XI. Es seien $x_0, x_1 \in V$ mit $x_0 \neq \mathcal{O}$.

2.3.11 Definition

Die Menge $\mathcal{G} = \{x = x_1 + t \cdot x_0 : t \in \mathbb{R}\}$ heißt die **Gerade** *durch x_1 mit Richtungsvektor x_0.*

Um $\mathcal{G}$ zu erhalten, muß man t also alle reellen Zahlen durchlaufen lassen. Wir werden im Folgenden den Vektor $x = (O, P)$ mit P und entsprechend $x_1 = (O, P_1)$ mit P_1 identifizieren. Dann ist $\mathcal{G}$ eine Gerade in E durch P_1.

Axiom XI

Es gibt eine Gerade $\mathcal{G}$ durch O, so daß die Differenzmenge $V \setminus \mathcal{G} \neq \emptyset$ ist.

Dabei bezeichnet die *Differenzmenge* $M \setminus N$ die Menge der Elemente von M, die nicht in N enthalten sind.

Axiom XII

Es seien $\mathcal{G}_1, \mathcal{G}_2$ *zwei verschiedene Geraden durch* $\mathcal{O} \in V$. *Dann gibt es zu jedem Vektor* $x \in V$ *zwei Vektoren* $x_1 \in \mathcal{G}_1,\ x_2 \in \mathcal{G}_2$ *mit* $x = x_1 + x_2$.

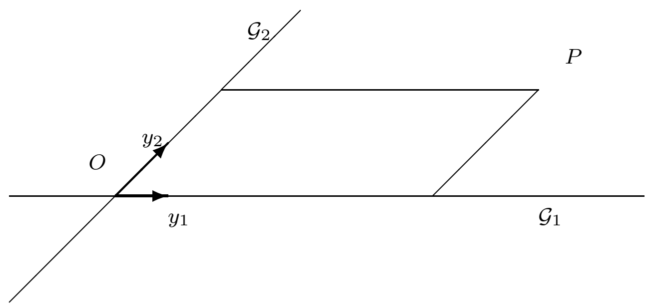
2.3.12 Satz

Es gilt $\dim V = 2$.

Beweis

Nach Axiom XI gibt es eine Gerade $\mathcal{G}_1$ durch O mit $V \setminus \mathcal{G}_1 \neq \emptyset$. Wir bezeichnen ihren Richtungsvektor mit y_1. Dann wählen wir einen Vektor $y_2 \in V \setminus \mathcal{G}_1$. Wir bezeichnen durch $\mathcal{G}_2$ die Gerade $\{x = t \cdot y_2 : t \in \mathbb{R}\}$. Wir erhalten $\mathcal{G}_1 \neq \mathcal{G}_2$. Die Vektoren y_1, y_2 sind linear unabhängig. Ist etwa die Gleichung $t_1 \cdot y_1 + t_2 \cdot y_2 = \mathcal{O}$ nicht trivial, so bekommen wir zuerst $t_1 \neq 0, t_2 \neq 0$ und dann, daß y_2 ein Vielfaches von y_1 ist und umgekehrt. Das ist ein Widerspruch, weil $\mathcal{G}_1$ und $\mathcal{G}_2$ verschieden sind. Andererseits erzeugen y_1, y_2 den ganzen Vektorraum V. Also sind sie eine Basis. Gilt $(O, P) = t_1 y_1 + t_2 y_2$, so kann man das Paar (t_1, t_2) als *Koordinaten* von P ansehen.

Gibt es mehr als 2 linear unabhängige Vektoren in V, so haben wir linear unabhängige Vektoren y_1, y_2, y_3. Die Geraden $\mathcal{G}_1, \mathcal{G}_2$ zu y_1, y_2 sind dann verschieden. Nach Axiom XII läßt sich y_3 als Linearkombination von y_1, y_2 schreiben. Das ist ein Widerspruch zur linearen Unabhängigkeit. Also ist die Höchstzahl der linear unabhängigen Vektoren zwei. Nach Definition 1.3.10 haben wir $\dim V = 2$. □

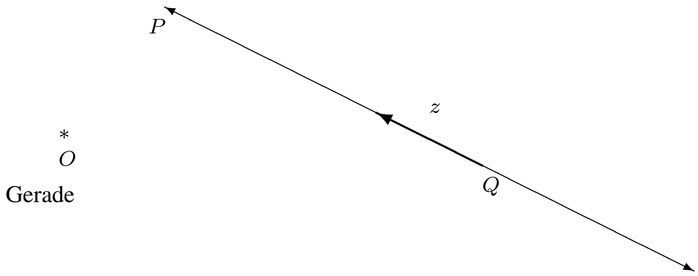


Koordinaten

2.4 Geometrie

Es sollen hier einige klassische geometrische Resultate abgeleitet werden. Wir betrachten zunächst Geraden in der Ebene.

Es sei $\mathcal{G} = \{P : (O,P) = (O,Q) + t \cdot z \text{ mit } t \in \mathbb{R}\}$ eine Gerade durch Q mit Richtungsvektor $z \neq \mathcal{O}$.



Es folgt:

2.4.1 Bemerkung

Ersetzt man Q durch ein $Q' \in \mathcal{G}$, so erhält man die gleiche Gerade.

Beweis

Es ist ja $x = (O,Q') + (t - t')z = (O,Q) + tz$, wenn $(O,Q') = (O,Q) + t'z$. □

2.4.2 Bemerkung

Man hat genau dann die gleiche Gerade, wenn man den Richtungsvektor z durch einen linear abhängigen Richtungsvektor z' ersetzt.

Beweis

Wir dürfen $Q = Q'$ annehmen. Die beiden Geraden sind genau dann gleich, wenn für alle Geradenpunkte x Gleichungen $x - (O,Q) = tz = \tau z'$ gelten. Das ist gerade bei linearer Abhängigkeit der Fall. □

2.4.3 Definition

Zwei Geraden heißen **parallel**, *wenn ihre Richtungsvektoren linear abhängig sind.*

2.4.4 Satz

Zwei verschiedene parallele Geraden schneiden sich nicht: Sie haben keinen Punkt gemeinsam.

Beweis

Ist Q ein gemeinsamer Punkt, so sind beides Geraden durch Q mit gleichem Richtungsvektor. Sie sind dann gleich. □

Ist eine Gerade gegeben, so existiert offensichtlich durch jeden Punkt eine dazu parallele Gerade.

2.4.5 Satz

Zwei nichtparallele Geraden schneiden sich in genau einem Punkt.

Beweis

Wir identifizieren E mit V. Es sei $\mathcal{G}$ die Gerade $\{x = y + tz\}$ und $\mathcal{G}'$ die Gerade $\{x = y' + \tau z'\}$. Wir haben zu zeigen, daß die Gleichung $y - y' = \tau z' - tz$ genau eine Lösung hat. Das ist der Fall, weil die beiden Geraden durch O mit den Richtungsvektoren z', z verschieden sind, die Richtungsvektoren sind linear unabhängig. $\square$

2.4.6 Satz

Durch je zwei verschiedene Punkte $Q, Q' \in E$ gibt es genau eine Gerade.

Beweis

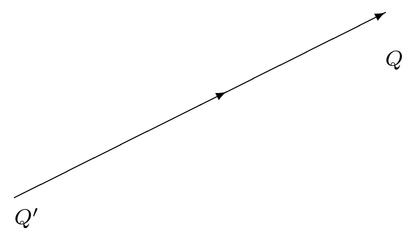
Wir identifizieren wieder E mit V. Es sei also $(O,Q) = y, (O,Q') = y'$. Wir nehmen die Gerade $\mathcal{G} = \{x = y + t\,(y' - y)\}$. Dann hat man für $t = 0: \;\; y \in \mathcal{G}$ und für $t = 1: \;\; y' \in \mathcal{G}$. Also geht $\mathcal{G}$ durch die beiden Punkte. Ist $\mathcal{G}'$ eine weitere solche Gerade, so geht $\mathcal{G}'$ durch y und hat einen Richtungsvektor z. Dann muß $y' - y = \tau z$ gelten, denn $\mathcal{G}'$ geht auch durch y'. Die Richtungsvektoren beider Geraden sind also linear abhängig, sie selbst also gleich. $\square$

Wir können $\mathcal{G}$ auch in der folgenden Form schreiben $\mathcal{G} = \{x : \; x = (1-t)y + ty', \; t \in \mathbb{R}\} = \{x = \mu y + \mu' y' : \; \mu + \mu' = 1\}$.

2.4.7 Definition

Es seien $Q \neq Q' \in E$. Dann heißt die Menge $\overline{QQ'} = \{P : (O,P) = (O,Q') + t \cdot ((O,Q) - (O,Q')), 0 \leq t \leq 1\}$ die **Verbindungsstrecke** *von Q und Q'.*

Für $t = 0$ ergibt sich $P = Q'$, für $t = 1$ folgt $P = Q$:



Verbindungsstrecke

Man sieht sofort:

2.4.8 Satz

Es ist stets $\overline{QQ'} = \{P : (O,P) = \mu(O,Q) + \mu'(O,Q'), \mu + \mu' = 1, \mu, \mu' \geq 0\}$.

2.4.9 Definition

Es seien $Q, Q' \in E$. *Dann heißt der Punkt* $P \in E$ *mit* $(O,P) = \frac{1}{2}((O,Q) + (O,Q'))$ *der* **Mittelpunkt** *von* $\overline{QQ'}$.

Es ist klar, warum man von Mitte spricht. Man hat nämlich:

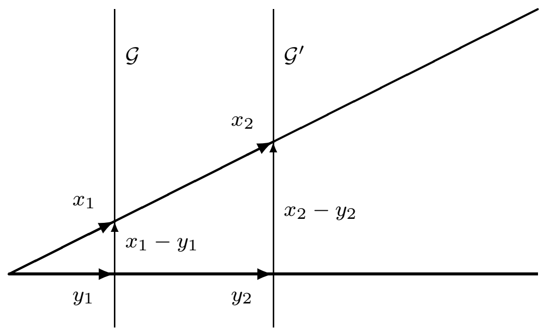
$$(O,P) = (O,Q') + \frac{1}{2}((O,Q) - (O,Q')) = (O,Q) + \frac{1}{2}((O,Q') - (O,Q)).$$

Wir wollen nun die Strahlensätze betrachten. Dabei sind *Strahlen* durch O Halbgeraden, die von O berandet werden. In den Sätzen brauchen wir nicht die noch nicht definierten Längen von Strecken. Wegen der Skalarenmultiplikation haben wir aber die Längenverhältnisse. Die genügen!

2.4.10 Satz (beide Strahlensätze)

Es seien zwei von $\mathcal{O}$ *verschiedene Vektoren* $x, y \in V$ *gegeben. Es seien* $\mathcal{G}, \mathcal{G}'$ *zwei verschiedene parallele Geraden, die nicht durch O gehen. Sie seien weder zu* x *noch zu* y *parallel. Wir setzen voraus, daß die Strahlen zu* x, y *von* $\mathcal{G}, \mathcal{G}'$ *geschnitten werden. Es sei* x_1 *der Vektor auf dem Strahl zu* x*, der von* O *bis zu dem Schnitt mit* $\mathcal{G}$ *reicht,* x_2 *der Vektor auf diesem Strahl von* O *bis zu dem Schnitt mit* $\mathcal{G}'$ *und* y_1, y_2 *die entsprechenden Vektoren auf dem Strahl zu* y*. Dann gilt für das Verhältnis der Längen:*

$$L(x_2)/L(x_1) = L(y_2)/L(y_1) = L(x_2 - y_2)/L(x_1 - y_1).$$



Strahlen

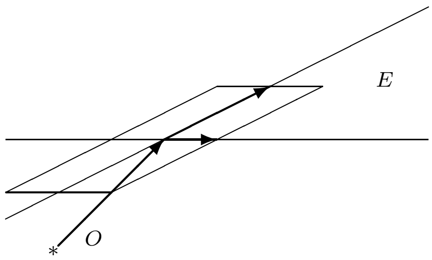
Beweis

Wir wählen eine reelle Zahl $r \geq 0$ so, daß $r \cdot x_1 = x_2$. Bei der Multiplikation mit r gehen Geraden stets in parallele Geraden über, $\mathcal{G}$ wird zu $\mathcal{G}'$. Also wird y_1 in y_2 abgebildet und damit auch $x_1 - y_1$ in $r(x_1 - y_1) = rx_1 - ry_1 = x_2 - y_2$. Die

Multiplikation mit r bedeutet nach der Anschauung aber eine Längung um den Faktor r. □

Der 3-dimensionale Raum

Es werde nun die Ebene durch den *3-dimensionalen anschaulichen Raum* ersetzt.



Ebenen E

Natürlich braucht man statt der Ebene E den ganzen 3-dimensionalen anschaulichen Raum R. Es sei wieder O ein Basispunkt und $V^3 = \{(O, P) : P \in R\}$. Es gelten dann die Axiome I - X unverändert. Es läßt sich aber besser begründen, daß der Raum im Großen wie im Kleinen aussieht.

2.4.11 Definition

Es sei E eine Teilmenge von R. Dann ist E eine **Ebene**, *wenn es einen Punkt $Q \in R$ und linear unabhängige Vektoren $x_1, x_2 \in V^3$ gibt, so daß*

$$E = \{P : (O, P) = (O, Q) + t_1 \cdot x_1 + t_2 \cdot x_2,\ t_1, t_2 \in \mathbb{R}\}.$$

Man erhält die gleiche Ebene genau dann, wenn man Q durch einen anderen Punkt aus E und x_1, x_2 durch von den gegebenen linear abhängige Vektoren ersetzt, die aber selbst linear unabhängig sind.

Das Axiom XI muß jetzt heißen: Es gibt eine Ebene E durch O, so daß $R \setminus E$ eine nicht leere Menge ist. Das Axiom XII besagt schließlich: Sind $\mathcal{G}_1, \mathcal{G}_2$ zwei verschiedene Geraden in E durch O und ist $\mathcal{G}_3$ eine nicht in E enthaltene Gerade durch O, und sind x_1, x_2, x_3 entsprechende Richtungsvektoren, so spannen sie V^3 auf. Das bedeutet, daß V^3 ein 3-dimensionaler Vektorraum ist.

3 Lineare Abbildungen

Abbildungen sind jedermann, zumindest implizit, aus dem täglichen Leben vertraut. Zwei einfache Beispiele sollen das illustrieren:

Das Einwohnermeldeamt bildet jedes Individuum ab auf einen Satz von Daten wie Name, Vorname, Geburtstag, usw. Dabei wird die Behörde bemüht sein, diese Abbildung so zu gestalten, daß jedes Individuum durch die erhobenen Daten zweifelsfrei beschrieben wird, daß die Abbildung also in dem unten zu erklärenden Sinne „injektiv" ist.

Ein Automobilkonzern berechnet aus den Anzahlen bestellter Fahrzeugtypen die Menge der benötigten Rohteile: Bleche, Schrauben, Kabel, … Dieses Beispiel hat einen anderen Charakter als das erste: Anzahlen können addiert und multipliziert werden; die Produktion doppelt so vieler Autos erfordert doppelt so viele Rohteile. Gehen mehrere Bestellungen ein, so addiert man die entsprechenden Anzahlen der Rohteile.

Somit ist es nicht verwunderlich, daß Abbildungen gerade in der Mathematik eine zentrale Rolle spielen. Bereits in den ersten beiden Kapiteln haben wir davon regen Gebrauch gemacht:

Jeder Vektor $\vec{a} \in \mathbb{R}^n$ ist eine Abbildung, die den Indizes $i = 1, \dots, n$ die Komponenten $a_1, \dots, a_n$ zuordnet. Das Urbild, der Index i, wird dabei implizit, durch die Position i des Bildes a_i, angegeben.

Bei der Addition werden je zwei Vektoren aus $\mathbb{R}^n$ bzw. in der anschaulichen Ebene auf einen Vektor aus $\mathbb{R}^n$ bzw. in der anschaulichen Ebene, den Summenvektor, abgebildet.

Im folgenden Abschnitt wird daher zunächst der allgemeine Abbildungsbegriff präzisiert.

Im übrigen ist dieses ganze Kapitel sehr speziellen Abbildungen, den *linearen* Abbildungen gewidmet, die gleichzeitig entsprechend starke Eigenschaften haben. Lineare Abbildungen wirken zwischen Vektorräumen und sind mit den Vektorraumoperationen „$+$" und „$\cdot$" verträglich. Etwa die Abbildung aus dem Automobilkonzern-Beispiel ist linear. Mit Hilfe von linearen „Koordinatenabbildungen" werden wir sehen, daß $\mathbb{R}^2$ bzw. $\mathbb{R}^3$ geeignete Modelle für die anschauliche Ebene bzw. den anschaulichen Raum sind. Allgemeiner werden wir jeden n-dimensionalen Vektorraum durch Einführung von Basen linear unter Erhaltung der Gestalt (isomorph) auf den $\mathbb{R}^n$ abbilden. Mit Hilfe dieser Koordinaten werden lineare Abbildungen zwischen endlichdimensionalen

Vektorräumen zu linearen Abbildungen $\mathbb{R}^n \to \mathbb{R}^m$. Deshalb werden letztere besondere Aufmerksamkeit genießen, zudem sie einem sehr einfachen und wirkungsvollen Kalkül, dem Matrizenkalkül, zugänglich sind. Die Untersuchung *linearer* Abbildungen zwischen endlichdimensionalen Vektorräumen wird auf Manipulationen von Zahlenschemata reduziert.

3.1 Abbildungen

Mit $M, M_1, M_2, \ldots$ werden im Folgenden stets Mengen bezeichnet.

3.1.1 Definition

F heißt **Abbildung** *von M_1 in M_2, wenn durch F* jedem *Element $P \in M_1$* genau ein *Element $Q \in M_2$ als Bild zugeordnet wird:*

$$F : M_1 \to M_2, \qquad P \mapsto Q = F(P).$$

Der Punkt $Q = F(P)$ heißt *Bild* von P oder *Wert* von F in P. Die Menge M_1 heißt *Urbildbereich* und M_2 *Wertevorrat*.

Zwei Abbildungen $F, F' : M_1 \to M_2$ sind *gleich* genau dann, wenn für alle $P \in M_1$ gilt: $F(P) = F'(P)$. Es wird nicht verlangt, daß die Zuordnungsvorschrift explizit angegeben werden kann; die Zuordnung kann sehr abstrakt sein. Dies gilt beispielsweise für Abbildungen, die mit Hilfe des „Zornschen Lemmas“ bzw. des „Auswahlaxioms“ konstruiert werden.

Auf jeder Menge M gibt es eine besonders einfache Abbildung, die *identische Abbildung*, $id : M \to M$, $P \mapsto P$. Werden gleichzeitig mehrere Mengen betrachtet, so schreibt man zur Vermeidung von Mißverständnissen auch id_M.

3.1.2 Beispiel

Geraden $\mathcal{G}$ im anschaulichen Raum R können als Bilder von Abbildungen beschrieben werden. Sei $x_0 \in V^3$, $x_0 \neq \mathcal{O}$, $P \in R$, $\mathcal{G} := \{Q \in R : \ (O,Q) = (O,P) + tx_0$ für ein $t \in \mathbb{R}\}$. Dann wird durch $t \mapsto Q$ eine Abbildung $G : \mathbb{R} \to \mathcal{G}$ erklärt.

3.1.3 Definition

Eine Abbildung $F : M_1 \to M_2$ heißt

(a) **surjektiv**, *wenn für alle $Q \in M_2$ (mindestens) ein $P \in M_1$ existiert mit $F(P) = Q$.*
(Jedes Element von M_2 kommt als Wert vor.)

(b) **injektiv**, *wenn aus $P_1, P_2 \in M_1$, $P_1 \neq P_2$ schon $F(P_1) \neq F(P_2)$ folgt.*
(Jedes Element von M_2 kommt höchstens einmal als Wert vor.)

(c) **bijektiv**, *wenn F surjektiv und injektiv ist.*
(Jedes Element von M_2 kommt genau einmal als Wert vor.)

Für surjektive Abbildungen schreibt man $F : M_1 \twoheadrightarrow M_2$ und sagt, daß F *auf* M_2 abbildet.

Injektive Abbildungen heißen auch *eineindeutig*, man schreibt $F : M_1 \hookrightarrow M_2$. Eine Abbildung F ist genau dann injektiv, wenn aus $P_1, P_2 \in M_1$, $F(P_1) = F(P_2)$ stets $P_1 = P_2$ folgt.

Für bijektive Abbildungen schreibt man auch $F : M_1 \leftrightarrow M_2$.

3.1.4 Beispiel

Die Geradenabbildung aus Beispiel 3.1.2 ist injektiv, denn der Vektor x_0 ist nicht der Nullvektor und deshalb linear unabhängig. Auf Grund der Wahl des Wertevorrats $\mathcal{G}$ ist diese Abbildung auch surjektiv und damit bijektiv. Betrachtet man dagegen die Abbildung $\mathbb{R} \to R$ mit derselben Abbildungsvorschrift, so ist diese Abbildung nicht mehr surjektiv.

3.1.5 Definition

Sei $F : M_1 \to M_2$ eine Abbildung. Dann heißt

$$\mathrm{im}(F) := F(M_1) := \{Q \in M_2 : \text{ es gibt ein } P \in M_1 \text{ mit } Q = F(P)\}$$

Bildmenge *von M_1 unter F.*

3.1.6 Beispiel

Sei $G : \mathbb{R} \to \mathcal{G} \subset R$ die Geraden-Abbildung aus Beispiel 3.1.2. Nach Definition dieser Geraden ist $\mathcal{G} = G(\mathbb{R}) = \mathrm{im}(G)$.

3.1.7 Definition

Es seien $F_1 : M_1 \to M_2$, $F_2 : M_2 \to M_3$ Abbildungen. Die **Zusammensetzung** *(***Hintereinanderschaltung**, **Komposition***) von F_1 und F_2 wird durch*

$$F_2 \circ F_1 : M_1 \to M_3, \qquad (F_2 \circ F_1)(P) := F_2(F_1(P)) \qquad (P \in M_1)$$

erklärt.

Solche Sachverhalte werden durch *kommutative Diagramme* beschrieben.

$$\begin{tikzcd}
M_1 \arrow[rr, "F_2 \circ F_1"'] \arrow[dr, "F_1"'] & & M_3 \\
& M_2 \arrow[ur, "F_2"'] &
\end{tikzcd}$$

Komposition, Zusammensetzung von Abbildungen

Definitionsgemäß liefern beide Wege von M_1 nach M_3 stets dasselbe Ergebnis; das Diagramm *kommutiert.*

3.1.8 Beispiel

Wir betrachten

$$F_1 : \mathbb{R}^2 \to \mathbb{R}^3, \quad F_1\,(\vec{a}) = (a_1 + a_2, a_1, -a_2) \text{ und}$$

$$F_2 : \mathbb{R}^3 \to \mathbb{R}^2, \quad F_2\left(\vec{b}\right) = \left(\frac{1}{2}(b_1 + b_2 + b_3), \frac{1}{2}(b_1 - b_2 - b_3)\right).$$

Wir erhalten für die Komposition $F_2 \circ F_1 : \mathbb{R}^2 \to \mathbb{R}^2$:

$$F_2 \circ F_1\,(\vec{a}) = \left(\frac{1}{2}(a_1 + a_2 + a_1 - a_2), \frac{1}{2}(a_1 + a_2 - a_1 + a_2)\right) = \vec{a}.$$

□

Wir stellen einige einfache Gesetze für die Komposition von Abbildungen zusammen.

3.1.9 Satz

Es seien $F_1 : M_1 \to M_2,\ F_2 : M_2 \to M_3,\ F_3 : M_3 \to M_4$ *Abbildungen. Dann gilt:*

(a) $F_1 \circ id_{M_1} = F_1$.

(b) $id_{M_2} \circ F_1 = F_1$.

(c) $(F_3 \circ F_2) \circ F_1 = F_3 \circ (F_2 \circ F_1)$. (Assoziativgesetz)

Das Diagramm kommutiert:

$$\begin{tikzcd}
M_1 \arrow[rr, "F_2 \circ F_1"] \arrow[dr, "F_1"'] & & M_3 \arrow[dr, "F_3"] & \\
& M_2 \arrow[ur, "F_2"] \arrow[rr, "F_3 \circ F_2"] & & M_4
\end{tikzcd}$$

Assoziativgesetz bei der Komposition von Abbildungen

Beweis

Nur (c) ist nicht völlig trivial. Für alle $P \in M_1$ gilt: $((F_3 \circ F_2) \circ F_1)(P) = (F_3 \circ F_2)(F_1(P)) = F_3(F_2(F_1(P))) = F_3((F_2 \circ F_1)(P)) = (F_3 \circ (F_2 \circ F_1))(P)$. □

Umkehrabbildungen

Ist $F : M_1 \to M_2$ bijektiv, so kann die *Umkehrabbildung* oder *inverse Abbildung* $F^{-1} : M_2 \to M_1$ erklärt werden. Für jedes $Q \in M_2$ existiert genau ein $P \in M_1$ mit $F(P) = Q$. Man setzt $F^{-1}(Q) = P$. Wir können die Kompositionen $F^{-1} \circ F : M_1 \to M_1, \quad F \circ F^{-1} : M_2 \to M_2$ von F und F^{-1} bilden und stellen fest, daß F^{-1} tatsächlich zu F invers ist:

3.1.10 Satz

Sei $F : M_1 \to M_2$ bijektiv mit Umkehrabbildung $F^{-1} : M_2 \to M_1$. Dann gilt:

(a) $F^{-1} \circ F = id_{M_1}$,

(b) $F \circ F^{-1} = id_{M_2}$.

Beweis

(a) Für beliebiges $P \in M_1$ ist $F^{-1}(F(P))$ das eindeutig bestimmte Element $P' \in M_1$ mit $F(P') = F(P)$. Die Injektivität von F ergibt $P = P'$. Also ist stets $F^{-1}(F(P)) = P$.

(b) Sei $Q \in M_2$. Nach Definition ist $F^{-1}(Q)$ dasjenige Element aus M_1 mit $F(F^{-1}(Q)) = Q$. □

Ist für eine bijektive Abbildung $F : M_1 \leftrightarrow M_2$ allein durch Angabe von M_1 und M_2 klar, welche Abbildung gemeint ist, so heißt F *kanonisch* oder *natürlich*. Häufig werden M_1 und M_2 dann identifiziert, M_1 und M_2 werden als gleich aufgefaßt: $M_1 = M_2$.

Wir zeigen nun, daß die Eigenschaften (a), (b) aus Satz 3.1.10 die (Existenz genau einer) Umkehrabbildung charakterisieren.

3.1.11 Satz

Die Abbildung $F : M_1 \to M_2$ ist genau dann bijektiv, wenn ein $G : M_2 \to M_1$ mit $G \circ F = id_{M_1}$ und $F \circ G = id_{M_2}$ existiert. Für ein solches G gilt $G = F^{-1}$.

Beweis

„$\Rightarrow$“: Siehe Satz 3.1.10.

„$\Leftarrow$“: Die Identität $G \circ F = id_{M_1}$ liefert die Injektivität von F: Seien $P_1, P_2 \in M_1$ gegeben mit $F(P_1) = F(P_2)$. Es folgt $G(F(P_1)) = G(F(P_2))$, also $P_1 = P_2$. Andererseits ergibt $F \circ G = id_{M_2}$ die Surjektivität von F: Sei $Q \in M_2$, setze $P := G(Q)$. Es gilt $F(P) = F(G(Q)) = F \circ G(Q) = Q$.

Sei $Q \in M_2$ gegeben. Für $P = G(Q)$ gilt $F(P) = F \circ G(Q) = Q$. Nach Definition von F^{-1} gilt $F^{-1}(Q) = P$. Also ist $G = F^{-1}$. □

3.1.12 Folgerung

Vertauscht man in Satz 3.1.11 die Rollen von G und F, so sieht man: $G = F^{-1}$ ist dort auch bijektiv, es gilt $F = G^{-1}$.

Also hat man für jede bijektive Abbildung F:

$$F = (F^{-1})^{-1}.$$

In Abschnitt 1.1 hatten wir gesehen: in *Gruppen* ist jedes Rechtsinverse auch Linksinverses. Für Abbildungen haben wir ohne zusätzliche Bedingungen im allgemeinen *keine* Gruppenstruktur. Um die Existenz von Umkehrabbildungen zu zeigen, sind *beide* Identitäten aus Satz 3.1.11 nachzuweisen, wie das folgende Beispiel zeigt:

3.1.13 Beispiel

Für die in Beispiel 3.1.8 definierten Abbildungen hatten wir gesehen: $F_2 \circ F_1 = id_{\mathbb{R}^2}$. Wäre F_1 bijektiv, so hätte man $F_1^{-1} : \mathbb{R}^3 \to \mathbb{R}^2$ und damit $F_1^{-1} = id_{\mathbb{R}^2} \circ F_1^{-1} = (F_2 \circ F_1) \circ F_1^{-1} = F_2 \circ \left(F_1 \circ F_1^{-1}\right) = F_2 \circ id_{\mathbb{R}^3} = F_2$. Es wäre also $F_1 \circ F_2 = F_1 \circ F_1^{-1} = id_{\mathbb{R}^3}$. Jedoch ist $F_1 \circ F_2(\vec{b}) = \left(b_1, \frac{1}{2}(b_1 + b_2 + b_3), -\frac{1}{2}(b_1 - b_2 - b_3)\right)$, also z. B. $(F_1 \circ F_2)(0,0,1) = \left(0, \frac{1}{2}, \frac{1}{2}\right)$. Demnach ist $F_1 \circ F_2 \neq id_{\mathbb{R}^3}$ und folglich F_1 nicht bijektiv.

3.1.14 Satz

Es seien $F_1 : M_1 \to M_2$, $F_2 : M_2 \to M_3$ Abbildungen. Für die Zusammensetzung $G := F_2 \circ F_1$ gilt:

(a) Sind F_1 und F_2 injektiv, so ist G injektiv.

(b) Sind F_1 und F_2 surjektiv, so ist G surjektiv.

(c) Sind F_1 und F_2 bijektiv, so ist G bijektiv.

Beweis

(a) Seien $P_1, P_2 \in M_1$ mit $G(P_1) = G(P_2)$, d. h. $F_2(F_1(P_1)) = F_2(F_1(P_2))$. Es folgt $F_1(P_1) = F_1(P_2)$ aus der Injektivität von F_2 und entsprechend $P_1 = P_2$ aus der Injektivität von F_1.

(b) Sei $Q_3 \in M_3$ gegeben. Weil F_2 surjektiv ist, existiert $Q_2 \in M_2$ mit $Q_3 = F_2(Q_2)$, die Surjektivität von F_1 liefert ein $Q_1 \in M_1$ mit $Q_2 = F_1(Q_1)$. Es folgt $Q_3 = F_2 \circ F_1(Q_1)$.

(c) Folgt aus (a) und (b). □

3.1.15 Satz

Die Abbildungen $F_1 : M_1 \to M_2$, $F_2 : M_2 \to M_3$ seien bijektiv. Dann gilt $(F_2 \circ F_1)^{-1} = {F_1}^{-1} \circ {F_2}^{-1}$.

$$\begin{tikzcd}
M_1 \arrow[rr, shift left, "F_2 \circ F_1"] \arrow[dr, shift left, "F_1"] & & M_3 \arrow[ll, shift left, "(F_2 \circ F_1)^{-1}"] \arrow[dl, shift left, "F_2^{-1}"] \\
& M_2 \arrow[ul, shift left, "F_1^{-1}"] \arrow[ur, shift left, "F_2"] &
\end{tikzcd}$$

Es gibt keinen anderen Weg zurück.

Beweis

Für $G := {F_1}^{-1} \circ {F_2}^{-1}$ gilt auf Grund der Assoziativität der Zusammensetzung von Abbildungen:

$$\begin{aligned}(F_2 \circ F_1) \circ G &= F_2 \circ (F_1 \circ {F_1}^{-1}) \circ {F_2}^{-1} \\ &= F_2 \circ id_{M_2} \circ {F_2}^{-1} = F_2 \circ {F_2}^{-1} = id_{M_3}\end{aligned}$$

und entsprechend

$$G \circ (F_2 \circ F_1) = {F_1}^{-1} \circ {F_2}^{-1} \circ F_2 \circ F_1 = id_{M_1}.$$

Satz 3.1.11 liefert $(F_2 \circ F_1)^{-1} = G = {F_1}^{-1} \circ {F_2}^{-1}$. □

Mit Hilfe von bijektiven Abbildungen können Koordinaten in der anschaulichen Ebene E und im anschaulichen Raum R eingeführt werden. Wir verwenden im Folgenden die Bezeichnungen aus Kapitel 2.

Koordinaten in der anschaulichen Ebene *E*

Nachdem der Basispunkt $O \in E$ einmal fixiert worden ist, existieren zwischen der anschaulichen Ebene E und dem Vektorraum V der Pfeile in O kanonische Abbildungen:

$$\begin{aligned} F &: V \to E, \qquad (O, P) \mapsto P, \\ G &: E \to V, \qquad P \mapsto (O, P). \end{aligned}$$

Nach Satz 3.1.11 sind F und G bijektiv, $F = G^{-1}$, $G = F^{-1}$. Wir können E und V identifizieren: $E = V$. Nun soll eine bijektive Abbildung zwischen V und $\mathbb{R}^2$ konstruiert werden. Gemäß Satz 2.3.12 ist $\dim V = 2$. Also existieren zwei linear unabhängige Vektoren $x_1, x_2 \in V$, die nach Satz 1.3.16 eine Basis von V bilden. Wir definieren nun eine Abbildung, die *nicht* kanonisch ist, weil sie von der Wahl der Basis $\{x_1, x_2\}$ abhängt:

$$H : \mathbb{R}^2 \to V, \qquad (a_1, a_2) \mapsto a_1 x_1 + a_2 x_2.$$

Da x_1, x_2 linear unabhängig sind, ist die Darstellung von Vektoren aus V bzgl. der Basis $\{x_1, x_2\}$ nach Satz 1.3.9 eindeutig, also ist H injektiv. Nachdem V von x_1 und x_2 erzeugt wird, ist H aber auch surjektiv und somit bijektiv.

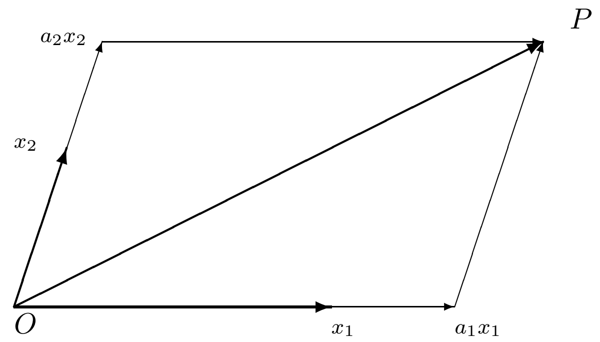
Wir betrachten nun die bijektive Abbildung

$$F \circ H : \mathbb{R}^2 \to E$$

und die Umkehrabbildung

$$(F \circ H)^{-1} : E \to \mathbb{R}^2.$$

Die Abbildung $F \circ H$ heißt *Karte*, weil jedem Paar aus $\mathbb{R}^2$ ($\leftrightarrow$ Kartenblatt) ein Punkt der Ebene ($\leftrightarrow$ Erdoberfläche) zugeordnet wird. Die Umkehrabbildung $(F \circ H)^{-1}$ heißt *Koordinatensystem*, die Zahlenpaare $\vec{a} = (a_1, a_2) = (F \circ H)^{-1}(P)$ heißen *Koordinaten* von P bzgl. der Basis $\{x_1, x_2\}$.



Anschauliche Ebene isomorph zu $\mathbb{R}^2$

Koordinaten im anschaulichen Raum R

Man geht völlig analog wie oben vor; $F : V^3 \to R, \quad (O, P) \mapsto P$ sei die kanonische Abbildung zwischen V^3 und R. Nachdem V^3 ein dreidimensionaler Vektorraum ist, existiert eine dreielementige Basis $\{x_1, x_2, x_3\}$ von V^3. Wir setzen

$$H : \mathbb{R}^3 \to V^3, \quad (a_1, a_2, a_3) \mapsto a_1x_1 + a_2x_2 + a_3x_3,$$

dann ist $F \circ H : \mathbb{R}^3 \to R$ wieder bijektiv und heißt Karte. Jeder Punkt $P \in R$ hat drei Koordinaten $(a_1, a_2, a_3) = (F \circ H)^{-1}(P)$.

Diese so weitreichende Idee, mittels Koordinaten die in $\mathbb{R}^2$ bzw. $\mathbb{R}^3$ zur Verfügung stehenden algebraischen und analytischen Methoden der Geometrie zu Nutze zu machen, geht auf den französischen Philosophen René Descartes (1596-1650) zurück.

3.2 Lineare Abbildungen

Lineare Abbildungen sind spezielle Abbildungen zwischen Vektorräumen. Sie „beachten“ Addition und Skalarmultiplikation. Jeder Vektorraum ist bezüglich der Vektoraddition eine abelsche Gruppe, daher betrachten wir zunächst solche Abbildungen zwischen Gruppen, die die Gruppenoperation beachten.

Homomorphismen

Im Folgenden bezeichnen G_1, G_2 Gruppen; da die Verwechslungsgefahr gering ist, bezeichnet „+“ die Operation sowohl auf G_1 wie auf G_2. Ebenso verwenden wir für die jeweiligen Nullelemente dasselbe Symbol 0.

3.2.1 Definition

Eine Abbildung $F : G_1 \to G_2$ *heißt* **Gruppenhomomorphismus**, *wenn für alle* $x, y \in G_1$ *gilt:* $\quad F(x+y) = F(x) + F(y).$

Man sagt, daß F die Gruppenoperation „+“ beachtet. „Homomorph“ ist eine Ableitung aus dem Griechischen und bedeutet etwa „von gleicher Gestalt“.

3.2.2 Beispiel

$F : (\mathbb{R}^+, \cdot) \to (\mathbb{R}, +)$, $F(x) = \ln x$ ist wegen $\ln(x \cdot y) = \ln x + \ln y$ ein Homomorphismus. Dagegen ist $G : (\mathbb{R}, +) \to (\mathbb{R}, +)$, $F(x) = 1 + x$ kein Homomorphismus.

3.2.3 Satz

Ist $F : G_1 \to G_2$ *ein Gruppenhomomorphismus, so gilt:*

(a) $F(0) = 0$.

(b) $F(-x) = -F(x)$ *für alle* $x \in G_1$.

Beweis

(a) Aus $F(0) = F(0+0) = F(0) + F(0)$ folgt $F(0) = 0$.

(b) Wegen (a) gilt: $0 = F(0) = F(x + (-x)) = F(x) + F(-x)$ und damit $-F(x) = F(-x)$. □

Im Folgenden bezeichnen $V, V_1, V_2, \ldots$ stets reelle Vektorräume.

3.2.4 Definition

Eine Abbildung $F : V_1 \to V_2$ *heißt* **lineare Abbildung** *oder* **Vektorraumhomomorphismus**, *falls gilt:*

(1) *Für alle* $x, y \in V_1$ *ist* $F(x+y) = F(x) + F(y)$. (Additivität)

(2) *Für alle* $a \in \mathbb{R}, x \in V_1$ *ist* $F(a \cdot x) = a \cdot F(x)$. (Homogenität)

Eine lineare Abbildung beachtet also die Vektorraumoperationen. Eigenschaft (1) bedeutet, daß F insbesondere ein Gruppenhomomorphismus $(V_1, +) \to (V_2, +)$ ist.

Offensichtlich ist die identische Abbildung $id : V \to V$ stets eine lineare Abbildung.

3.2.5 Beispiel

Sei V der Vektorraum der reellen Polynome, vgl. Beispiel 1.3.11. Zwei Polynome $P = \sum_{i=0}^n a_i z^i$ und $Q = \sum_{i=0}^n b_i z^i$ können miteinander multipliziert werden. Indem man formal ausmultipliziert und $a_{n+1} = \ldots = a_{2n} = b_{n+1} = \ldots = b_{2n} = 0$ ergänzt, gelangt man zu der Setzung:

$$P \cdot Q := \sum_{i=0}^{2n} \left(\sum_{\ell=0}^{i} a_\ell b_{i-\ell} \right) z^i.$$

Für diese Multiplikation läßt sich leicht ein Distributivgesetz zeigen. Sei $Q' = \sum_{i=0}^n {b'}_i z^i$, dann ist auf Grund des reellen Distributivgesetzes:

$$\begin{aligned} P \cdot (Q + Q') &= \sum_{i=0}^{2n} \left(\sum_{\ell=0}^{i} a_\ell \left(b_{i-\ell} + {b'}_{i-\ell} \right) \right) z^i \\ &= \sum_{i=0}^{2n} \left(\left(\sum_{\ell=0}^{i} a_\ell b_{i-\ell} \right) + \left(\sum_{\ell=0}^{i} a_\ell {b'}_{i-\ell} \right) \right) z^i \\ &= \sum_{i=0}^{2n} \left(\sum_{\ell=0}^{i} a_\ell b_{i-\ell} \right) z^i + \sum_{i=0}^{2n} \left(\sum_{\ell=0}^{i} a_\ell {b'}_{i-\ell} \right) z^i \\ &= P \cdot Q + P \cdot Q'. \end{aligned}$$

Wir betrachten ein *festes* Polynom $P_0 \in V$ und definieren eine Abbildung

$$F : V \to V, \qquad F(P) := P_0 \cdot P.$$

Diese Abbildung ist *linear*, denn für alle $a \in \mathbb{R}$, $P, Q \in V$ gilt:

(1) $F(P + Q) = P_0 \cdot (P + Q) = P_0 \cdot P + P_0 \cdot Q = F(P) + F(Q)$.

(2) $F(a \cdot P) = P_0 \cdot (a \cdot P) = a \cdot (P_0 \cdot P) = a \cdot F(P)$.

Dagegen ist $G : V \to V$, $G(P) := P \cdot P$ *nicht linear*, denn es gilt beispielsweise $G(2 \cdot z) = 4 \cdot z^2 \neq 2 \cdot z^2 = 2 \cdot G(z)$.

3.2.6 Folgerungen

Sei $F : V_1 \to V_2$ *eine lineare Abbildung, ferner seien* $x_1, \ldots, x_n \in V_1$, $a_1, \ldots, a_n \in \mathbb{R}$. *Dann gilt:*

(a) $F\left(\sum_{i=1}^n a_i x_i \right) = \sum_{i=1}^n a_i F(x_i)$.

(b) *Sind* $x_1, \ldots, x_n$ *linear abhängig, so auch* $F(x_1), \ldots, F(x_n)$.

(c) *Sind* $F(x_1), \ldots, F(x_n)$ *linear unabhängig, so auch* $x_1, \ldots, x_n$.

Beweis

(a) Folgt durch einen leichten Induktionsbeweis nach n aus den Eigenschaften einer linearen Abbildung.

(b) Sind $x_1, \ldots, x_n$ linear abhängig, so existieren $a_1, \ldots, a_n \in \mathbb{R}$ mit $\sum_{i=1}^n a_i x_i = 0_{V_1}$, wobei ein $a_i \neq 0$ ist. Mit Satz 3.2.3 und (a) folgt: $0_{V_2} = F(0_{V_1}) = F\left(\sum_{i=1}^n a_i x_i\right) = \sum_{i=1}^n a_i F(x_i)$. Also ist 0_{V_2} eine nichttriviale Linearkombination von $F(x_1), \ldots, F(x_n)$.

(c) Da die Aussagen „$A \Rightarrow B$“ und „nicht $B \Rightarrow$ nicht A“ äquivalent sind, ist (c) lediglich eine Umformulierung von (b). □

3.2.7 Satz

Sei $F : V_1 \to V_2$ *eine lineare Abbildung. Dann ist die Bildmenge* $\operatorname{im}(F) = F(V_1) = \{F(x) : x \in V_1\}$ *ein Untervektorraum von* V_2*, und es gilt* $\dim F(V_1) \leq \dim V_1$.

Beweis

Die erste Behauptung ist offensichtlich, die zweite ergibt sich aus der vorhergehenden Folgerung 3.2.6 (c). □

Lineare Abbildungen können demnach die Dimension nicht erhöhen, wogegen man in der Differential- und Integralrechnung oder Topologie lernt, daß es durchaus stetige surjektive Abbildungen $\mathbb{R} \to \mathbb{R}^2$, sogenannte Peano-Kurven gibt. Dazu s. z. B. J. Cigler, H.-C. Reichel, *Topologie*, Bibliographisches Institut: Mannheim 1978.

3.2.8 Satz

Es seien $F_1 : V_1 \to V_2$, $F_2 : V_2 \to V_3$ *lineare Abbildungen. Dann ist auch* $F_2 \circ F_1 : V_1 \to V_3$ *eine lineare Abbildung.*

Beweis

Wir müssen die Eigenschaften (1) und (2) aus Definition 3.2.4 prüfen. Für beliebige $x, y \in V_1$, $a \in \mathbb{R}$ gilt: $(F_2 \circ F_1)(x + y) = F_2(F_1(x + y)) = F_2(F_1(x) + F_1(y)) = F_2(F_1(x)) + F_2(F_1(y)) = (F_2 \circ F_1)(x) + (F_2 \circ F_1)(y)$ und $(F_2 \circ F_1)(a \cdot x) = F_2(F_1(a \cdot x)) = F_2(a \cdot F_1(x)) = a \cdot F_2(F_1(x)) = a \cdot (F_2 \circ F_1)(x)$. □

Der nächste Satz bekräftigt, daß die Linearität einer Abbildung F eine sehr starke Eigenschaft ist: Schon durch die Werte auf einem System von Erzeugern ist F vollkommen bestimmt.

3.2.9 Satz

Seien $x_1, \ldots, x_n \in V_1$ *Erzeugende von* V_1*. Weiter seien* $F, G : V_1 \to V_2$ *lineare Abbildungen mit* $F(x_i) = G(x_i)$ *für* $i = 1, \ldots, n$*. Dann ist* $F = G$.

Beweis

Es ist zu zeigen, daß $F(x) = G(x)$ für alle $x \in V_1$ gilt. Sei also $x \in V_1$ beliebig. Es existieren $a_1, \ldots, a_n \in \mathbb{R}$ mit $x = \sum_{i=1}^n a_i x_i$. Folgerung 3.2.6 (a) liefert:

$$F(x) = F\left(\sum_{i=1}^n a_i x_i\right) = \sum_{i=1}^n a_i F(x_i) = \sum_{i=1}^n a_i G(x_i) = G\left(\sum_{i=1}^n a_i x_i\right) = G(x).$$

□

Konstruktion von linearen Abbildungen

Der folgende fundamentale Satz liefert die Existenz von nichttrivialen linearen Abbildungen zwischen endlichdimensionalen Vektorräumen. Der Beweis gibt sogar ein *Konstruktionsverfahren* an, die *lineare Fortsetzung*. Wir erinnern an Satz 1.3.17, wonach jeder endlichdimensionale Vektorraum eine Basis besitzt.

3.2.10 Satz

Sei $\{x_1, \ldots, x_n\} \subset V_1$ eine Basis von V_1. Die Vektoren $y_1, \ldots, y_n \in V_2$ seien beliebig. Dann existiert genau eine lineare Abbildung $F : V_1 \to V_2$ mit $F(x_i) = y_i$ für $i = 1, \ldots, n$.

Beweis

Die Eindeutigkeitsaussage ist auf Grund von Satz 3.2.9 klar. Für den Nachweis der Existenz ist die *eindeutige* Darstellbarkeit beliebiger Vektoren $x \in V_1$ durch die Basis $\{x_1, \ldots, x_n\}$ entscheidend. Zu $x \in V_1$ existieren eindeutig bestimmte Zahlen $a_1, \ldots, a_n \in \mathbb{R}$ mit $x = \sum_{i=1}^n a_i x_i$. Wir definieren $F(x) := \sum_{i=1}^n a_i y_i \in V_2$ und erhalten eine Abbildung $F : V_1 \to V_2$ mit $F(x_i) = y_i$, $i = 1, \ldots, n$. Die Abbildung F ist wohldefiniert wegen der Eindeutigkeit der Darstellung von x.

Es bleibt die Linearität von F zu zeigen. Für $a \in \mathbb{R}$, $x = \sum_{i=1}^n a_i x_i$, $x' = \sum_{i=1}^n a'_i x_i$ ergibt die Definition von F:

$$\begin{aligned} F(x + x') &= F\left(\sum_{i=1}^n a_i x_i + \sum_{i=1}^n a'_i x_i\right) = F\left(\sum_{i=1}^n (a_i + a'_i)\, x_i\right) \\ &= \sum_{i=1}^n (a_i + a'_i)\, y_i = \sum_{i=1}^n a_i y_i + \sum_{i=1}^n a'_i y_i \\ &= F\left(\sum_{i=1}^n a_i x_i\right) + F\left(\sum_{i=1}^n a'_i x_i\right) = F(x) + F(y); \end{aligned}$$

$$
\begin{aligned}
F(a \cdot x) &= F\left(a \cdot \left(\sum_{i=1}^{n} a_i x_i\right)\right) = F\left(\sum_{i=1}^{n} (a a_i) x_i\right) = \sum_{i=1}^{n} (a a_i) y_i \\
&= a \cdot \sum_{i=1}^{n} a_i y_i = a \cdot F\left(\sum_{i=1}^{n} a_i x_i\right) = a \cdot F(x) .
\end{aligned}
$$

□

Legt man die Existenz von Basen auch in unendlichdimensionalen Vektorräumen zugrunde (dazu vgl. die Bemerkung am Ende von Abschnitt 1.3), kann man dort auch einen entsprechenden Satz zeigen.

Isomorphismen

3.2.11 Definition

Eine bijektive lineare Abbildung $F : V_1 \to V_2$ heißt **Vektorraumisomorphismus** *oder kurz* **Isomorphismus**.

Bei Isomorphismen schreibt man auch $F : V_1 \xrightarrow{\sim} V_2$.

3.2.12 Beispiel

(a) Die lineare Abbildung $F : \mathbb{R}^2 \to \mathbb{R}^2$, $F(c_1, c_2) = (c_1 + 2c_2, c_1 - c_2)$ ist ein Isomorphismus mit Umkehrabbildung $F^{-1} : \mathbb{R}^2 \to \mathbb{R}^2$, $F^{-1}(d_1, d_2) = \left(\frac{1}{3}(d_1 + 2d_2), \frac{1}{3}(d_1 - d_2)\right)$.

(b) Sei $V_1 = \left\{\sum_{j=0}^{n} a_j z^j : n \in \mathbb{N}_0,\ a_0, \ldots, a_n \in \mathbb{R}\right\}$ der schon wiederholt betrachtete Vektorraum aller reellen Polynome und $V_2 = \left\{\sum_{j=0}^{n} a_j z^{2j} : n \in \mathbb{N}_0,\ a_0, \ldots, a_n \in \mathbb{R}\right\}$ der Vektorraum aller reellen Polynome in z^2. Die Polynome in V_2 enthalten also nur gerade Potenzen von z. Durch $F\left(\sum_{j=0}^{n} a_j z^j\right) := \left(\sum_{j=0}^{n} a_j z^{2j}\right)$ wird ein kanonischer Isomorphismus $V_1 \to V_2$ vermittelt. Gleichzeitig kann man V_2 als echten Untervektorraum von V_1 auffassen.

3.2.13 Satz

Sei $F : V_1 \to V_2$ ein Isomorphismus. Dann ist die Umkehrabbildung $F^{-1} : V_2 \to V_1$ ebenfalls ein Isomorphismus.

Beweis

Wegen der Folgerung 3.1.12 ist F^{-1} auch bijektiv, und es reicht, die Linearität von F^{-1} zu zeigen:

(1) Zu $y_1, y_2 \in V_2$ definieren wir $x_i := F^{-1}(y_i)$, damit ist auch $y_i = F(x_i)$. Weil F linear ist, gilt $F(x_1 + x_2) = F(x_1) + F(x_2)$. Anwendung von F^{-1} auf diese Gleichung liefert: $F^{-1}(y_1) + F^{-1}(y_2) = x_1 + x_2 = F^{-1}(F(x_1 + x_2)) = F^{-1}(F(x_1) + F(x_2)) = F^{-1}(y_1 + y_2)$.

(2) Seien $a \in \mathbb{R}$ und $y \in V_2$ gegeben, wir setzen $x := F^{-1}(y)$. Analog zu (1) gilt: $a \cdot F^{-1}(y) = a \cdot x = F^{-1}(F(a \cdot x)) = F^{-1}(a \cdot F(x)) = F^{-1}(a \cdot y)$. □

3.2.14 Folgerungen

Sei $F : V_1 \to V_2$ ein Isomorphismus. Dann gilt:

(a) Sind $x_1, \ldots, x_n \in V_1$ linear unabhängig, so auch $F(x_1), \ldots, F(x_n)$.

(b) Wird V_1 von $x_1, \ldots, x_n \in V_1$ erzeugt, so wird V_2 von $F(x_1), \ldots, F(x_n)$ erzeugt.

(c) Ist $\{x_1, \ldots, x_n\}$ eine Basis von V_1, so ist $\{F(x_1), \ldots, F(x_n)\}$ eine Basis von V_2.

(d) $\dim V_1 = \dim V_2$.

Beweis

(a) Gemäß Voraussetzung sind $x_1 = F^{-1}(F(x_1)), \ldots, x_n = F^{-1}(F(x_n))$ linear unabhängig. Anwendung der Folgerung 3.2.6 auf F^{-1} ergibt die lineare Unabhängigkeit von $F(x_1), \ldots, F(x_n)$.

(b) Sei $y \in V_2$ gegeben. Zu $x = F^{-1}(y)$ existieren $a_1, \ldots, a_n \in \mathbb{R}$ mit $x = \sum_{i=1}^n a_i x_i$. Es folgt $y = F(x) = \sum_{i=1}^n a_i F(x_i)$.

(c) Folgt aus (a) und (b).

(d) Folgt aus (a) und der entsprechenden Aussage für F^{-1}. □

Diese Folgerungen zeigen, daß Vektorräume, zwischen denen ein Isomorphismus existiert, sehr eng verwandt sind. Die spezielle Gestalt der Abbildung ist dabei von untergeordnetem Interesse, weshalb ein weiterer Begriff eingeführt wird.

3.2.15 Definition

Zwei Vektorräume V_1, V_2 heißen **isomorph**, *wenn ein Isomorphismus $F : V_1 \xrightarrow{\sim} V_2$ existiert. In diesem Falle schreibt man $V_1 \approx V_2$. Ist F kanonisch, so nennt man V_1, V_2* **kanonisch isomorph** *und schreibt $V_1 \cong V_2$ oder $V_1 = V_2$.*

3.2.16 Satz

Sei V ein Vektorraum mit der Basis $\{x_1, \ldots, x_n\}$. Dann gilt $V \approx \mathbb{R}^n$.

Beweis

Wir versehen $\mathbb{R}^n$ mit der kanonischen Basis $\{\vec{e}_1, \ldots, \vec{e}_n\}$. Nach Satz 3.2.10 existieren eindeutig bestimmte lineare Abbildungen $G : \mathbb{R}^n \to V$ und $\hat{G} : V \to \mathbb{R}^n$ mit $G(\vec{e}_i) =$

x_i bzw. $\hat{G}(x_i) = \vec{e}_i$ für $i = 1, \ldots, n$. In Anlehnung an die geometrisch naheliegenden Fälle zweier und dreier Dimensionen heißt G *Karte* und $\hat{G}$ *Koordinatensystem*. Aus $(\hat{G} \circ G)(\vec{e}_i) = \vec{e}_i$ und $(G \circ \hat{G})(x_i) = x_i$ $(i = 1, \ldots, n)$ folgt $\hat{G} \circ G = id_{\mathbb{R}^n}$ und $G \circ \hat{G} = id_V$, denn lineare Abbildungen sind durch die Werte auf einer Basis schon festgelegt. Nach Satz 3.1.11 sind G und $\hat{G}$ Isomorphismen und zueinander invers. □

Dieser Satz ermöglicht es, aus linearen Abbildungen zwischen endlichdimensionalen Vektorräumen in naheliegender Weise lineare Abbildungen zwischen $\mathbb{R}^n$ und $\mathbb{R}^m$ (n, m geeignet) zu konstruieren. Letztere sind besonders einfach zu behandeln, weil sie dem Matrizenkalkül zugänglich sind.

Seien also V_1, V_2 Vektorräume mit $\dim V_1 = n$, $\dim V_2 = m$ und $H : V_1 \to V_2$ eine lineare Abbildung. Es existieren Basen $\{x_1, \ldots, x_n\}$ von V_1 und $\{y_1, \ldots, y_m\}$ von V_2, siehe Abschnitt 1.3. (Schon im nächsten Abschnitt werden wir sehen, daß *jede* Basis von V_1 bzw. V_2 genau n bzw. m Elemente hat.)

Es seien $G_1 : \mathbb{R}^n \to V_1$ und $G_2 : \mathbb{R}^m \to V_2$ die im vorhergehenden Beweis konstruierten Karten. Wir haben folgende Situation:

$$\begin{tikzcd}
V_1 \arrow[r, "H"] & V_2 \arrow[d, "G_2^{-1}"] \\
\mathbb{R}^n \arrow[u, "G_1"] \arrow[r, "F = \text{„}H \text{ in den Koordinaten“}"'] & \mathbb{R}^m
\end{tikzcd}$$

Durch $F := G_2^{-1} \circ H \circ G_1$ wird eine lineare Abbildung $\mathbb{R}^n \to \mathbb{R}^m$ gegeben, wir nennen sie H *in den Koordinaten*. Sie bildet nämlich die Koordinaten $G_1^{-1}(x)$ von $x \in V_1$ ab auf $\left(G_2^{-1} \circ H \circ G_1\right)\left(G_1^{-1}(x)\right) = G_2^{-1}\left(H(x)\right)$, die Koordinaten von $H(x)$.

3.3 Lineare Abbildungen $\mathbb{R}^n \to \mathbb{R}^m$

Die vorhergehenden Erläuterungen erlauben, uns ganz auf die Untersuchung dieser speziellen linearen Abbildungen zu konzentrieren.

In diesem Abschnitt wird zunächst eine bijektive Zuordnung zwischen den linearen Abbildungen $\mathbb{R}^n \to \mathbb{R}^m$ und bestimmten Zahlenschemata, den „Matrizen“, hergestellt. Wesentlich ist dabei, daß lineare Abbildungen $\mathbb{R}^n \to \mathbb{R}^m$ gemäß Satz 3.2.10 durch die Angabe der Werte auf einer (etwa der kanonischen) Basis des $\mathbb{R}^n$ vollkommen bestimmt sind. Für die Matrizen wird ein einfacher Kalkül entwickelt, der sich sogar leicht auf dem Computer programmieren läßt.

Gleichzeitig können wir mit Hilfe dieses Kalküls wichtige theoretische Resultate herleiten. Als Folgerung aus dem „Gaußschen Algorithmus“ erhalten wir in Satz 3.3.14, daß die Dimension von $\mathbb{R}^n$ tatsächlich n ist, und in Satz 3.3.15, daß alle Basen eines Vektorraums dieselbe Anzahl von Elementen besitzen und diese Anzahl die Dimension des Vektorraums ist.

Matrizen

3.3.1 Definition

*Eine $m \times n$-**Matrix** A ist ein rechteckiges Schema reeller Zahlen*

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} = (a_{ij})_{\substack{i=1,\dots,m \\ j=1,\dots,n}}$$

*mit m Zeilen und n Spalten. Die $a_{ij} \in \mathbb{R}$ heißen **Komponenten** oder **Koeffizienten** von A.*

Der erste Index heißt Zeilenindex, er bleibt in jeder Zeile fest. Der zweite Index heißt Spaltenindex, er bleibt in jeder Spalte fest. Will man das Format einer Matrix angeben, so schreibt man $A = A^{m,n}$ oder auch $A \in \mathbb{R}^{m \times n}$.

Im Rahmen des Matrizenkalküls fassen wir n-tupel $\vec{a} = (a_1, \dots, a_n) \in \mathbb{R}^n$ als Spalten, d. h. als $n \times 1$-Matrizen auf:

$$\vec{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = (a_i)_{i=1,\dots,n} .$$

Zusammenhang zwischen linearen Abbildungen und Matrizen

Sei zunächst eine beliebige lineare Abbildung $F : \mathbb{R}^n \to \mathbb{R}^m$ gegeben. Wir wollen dieser eine $m \times n$-Matrix A zuordnen. Wir betrachten dazu die Einheitsvektoren

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, \vec{e}_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \in \mathbb{R}^n$$

und deren Bilder

$$\vec{a}_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix} := F\left(\vec{e}_j\right) \in \mathbb{R}^m.$$

Die Matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} = (\vec{a}_1, \dots, \vec{a}_n)$$

ist die *zu F gehörige Matrix*. Wir haben A gemäß der Merkregel gebildet:

Die Spalten sind die Bilder der Einheitsvektoren.

Umgekehrt kann man gemäß den Ausführungen aus Abschnitt 3.2 zu jeder $m \times n$-Matrix $A = (\vec{a}_1, \dots, \vec{a}_n)$ eine lineare Abbildung $F : \mathbb{R}^n \to \mathbb{R}^m$ konstruieren, indem man $F\left(\vec{e}_j\right) := \vec{a}_j$ setzt. Die beschriebenen Zuordnungen $F \mapsto A$ bzw. $A \mapsto F$ von linearen Abbildungen $F : \mathbb{R}^n \to \mathbb{R}^m$ und $m \times n$-Matrizen A sind zueinander invers und deshalb insbesondere bijektiv.

Es soll nun ein Produkt von Matrizen und Spaltenvektoren eingeführt werden, das der Auswertung der zugehörigen linearen Abbildung in den jeweiligen Punkten entspricht.

3.3.2 Definition

Das **Produkt** *einer $m \times n$-Matrix $A = (a_{ij})_{\substack{i=1,\dots,m \\ j=1,\dots,n}}$ mit einem Spaltenvektor $\vec{c} \in \mathbb{R}^n$ wird definiert durch*

$$A \circ \vec{c} := \begin{pmatrix} \sum_{j=1}^n a_{1j} c_j \\ \sum_{j=1}^n a_{2j} c_j \\ \vdots \\ \sum_{j=1}^n a_{mj} c_j \end{pmatrix} = \left(\sum_{j=1}^n a_{ij} c_j \right)_{i=1,\dots,m} \in \mathbb{R}^m.$$

Mitunter verwendet man auch den gewöhnlichen Multiplikationspunkt: $A \cdot \vec{c}$.

Man kann sich diese Definition folgendermaßen merken: Man kippt die Spalte $\vec{c}$ und legt sie über jede Zeile von A. Übereinanderliegende Zahlen werden multipliziert und die Komponenten einer Zeile dann jeweils addiert.

3.3.3 Beispiel

(a)

$$\begin{pmatrix} 1 & -2 & 3 \\ 2 & 0 & 1 \end{pmatrix} \circ \begin{pmatrix} -1 \\ \frac{5}{2} \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \vec{0}.$$

(b) Dagegen ist

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

nicht erklärt.

Mit dem so eingeführten Produkt reduziert sich die Auswertung linearer Abbildungen tatsächlich auf die Multiplikation von Zahlenschemata:

3.3.4 Satz

Sei $F : \mathbb{R}^n \to \mathbb{R}^m$ eine lineare Abbildung und A die zu F gehörige $m \times n$-Matrix. Dann gilt für alle $\vec{c} \in \mathbb{R}^n$:

$$F(\vec{c}) = A \circ \vec{c}.$$

Ist B eine weitere $m \times n$-Matrix derart, daß $F(\vec{c}) = B \circ \vec{c}$ für alle $\vec{c} \in \mathbb{R}^n$ gilt, so ist $B = A$.

Beweis

Gemäß der Zuordnungsvorschrift $F \mapsto A$ ist $A = (a_{ij})_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}} = (\vec{a}_1, \ldots, \vec{a}_n)$ mit

$$\vec{a}_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix} := F\left(\vec{e}_j\right).$$

Auf Grund der Linearität von F folgt:

$$\begin{aligned} F\left(\vec{c}\right) &= F\left(\sum_{j=1}^{n} c_j \vec{e}_j\right) = \sum_{j=1}^{n} c_j F\left(\vec{e}_j\right) \\ &= \sum_{j=1}^{n} c_j \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix} = \sum_{j=1}^{n} \begin{pmatrix} a_{1j}c_j \\ a_{2j}c_j \\ \vdots \\ a_{mj}c_j \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{n} a_{1j}c_j \\ \sum_{j=1}^{n} a_{2j}c_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj}c_j \end{pmatrix} \\ &\overset{\text{def}}{=} A \circ \vec{c}. \end{aligned}$$

Für die j-te Spalte der Matrix B gilt nun:

$$\vec{b}_j = \begin{pmatrix} b_{1j} \\ \vdots \\ b_{mj} \end{pmatrix} = B \circ \vec{e}_j = F\left(\vec{e}_j\right).$$

Also stimmen alle Spalten von A und B und damit die beiden Matrizen überein. □

Die Linearität der zur $m \times n$-Matrix gehörigen Abbildung schlägt sich auf Grund des vorhergehenden Satzes unmittelbar in den folgenden Regeln nieder:

3.3.5 Folgerung

Sei A eine $m \times n$-Matrix. Dann gilt für alle $\vec{b}$, $\vec{c} \in \mathbb{R}^n$ und $d \in \mathbb{R}$:

$$A \circ \left(\vec{b} + \vec{c}\right) = A \circ \vec{b} + A \circ \vec{c}, \qquad \text{(Distributivgesetz)}$$

$$A \circ \left(d \cdot \vec{b}\right) = d \cdot \left(A \circ \vec{b}\right).$$

3.3.6 Beispiele

(a) Die linearen Abbildungen $F : \mathbb{R} \to \mathbb{R}$ werden durch 1×1-Matrizen, also reelle Zahlen gegeben. Sie sind also stets von der Form $F(c) = a \cdot c$ mit einer festen reellen Zahl a.

(b) Wir betrachten den Isomorphismus $F : \mathbb{R}^2 \to \mathbb{R}^2$ aus Beispiel 3.2.12:

$$F(\vec{c}) = \begin{pmatrix} c_1 + 2c_2 \\ c_1 - c_2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix} \circ \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix} \circ \vec{c},$$

also gehört zu F die 2×2-Matrix $A = \begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix}$.

(c) Sei $n = m$ und $F = id_{\mathbb{R}^n}$. Wegen $F(\vec{e}_j) = \vec{e}_j$ gehört zur Identität die **Einheitsmatrix**

$$E := E^{n,n} := (\vec{e}_1, \dots, \vec{e}_n) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

Die Diagonale besteht aus Einsen, sonst findet man in der Matrix nur Nullen.

In der englischsprachigen Literatur findet man häufig die Bezeichnung I für „Identity".

Zur Beschreibung der Komponenten der Einheitsmatrix führen wir das Kroneckersymbol ein:

3.3.7 Definition

Für je zwei Indizes $i, j \in \mathbb{N}$ wird das **Kroneckersymbol** δ_{ij} *durch*

$$\delta_{ij} := \begin{cases} 1, & \text{falls } i = j, \\ 0, & \text{falls } i \neq j, \end{cases}$$

definiert.

Damit haben wir für die Einheitsmatrix:

$$E = E^{n,n} = (\delta_{ij})_{\substack{i=1,\ldots,n \\ j=1,\ldots,n}} .$$

Lineare Gleichungssysteme

Lineare Gleichungssysteme treten in unzähligen Beispielen auf. Wir werden gleich sehen, daß sich deren Lösungsmengen mit Hilfe von Matrizen sehr einfach beschreiben lassen. Zwischen Matrizen und linearen Abbildungen besteht wiederum der soeben beschriebene unmittelbare Zusammenhang. Das bedeutet, daß man die Theorie linearer Gleichungssysteme, linearer Abbildungen und der Matrizen simultan entwickeln kann.

Wir betrachten ein (homogenes) lineares Gleichungssystem aus m Gleichungen in n Veränderlichen

$$\begin{array}{ccccccc}
a_{11}x_1 & + & a_{12}x_2 & + \ldots + & a_{1n}x_n & = 0 \\
a_{21}x_1 & + & a_{22}x_2 & + \ldots + & a_{2n}x_n & = 0 \\
\vdots & & \vdots & \ddots & \vdots & \vdots \\
a_{m1}x_1 & + & a_{m2}x_2 & + \ldots + & a_{mn}x_n & = 0
\end{array}$$

mit Koeffizientenmatrix $A = (a_{ij})_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}}$. Sei $F : \mathbb{R}^n \to \mathbb{R}^m$ die zu dieser Matrix gehörige lineare Abbildung. Wegen Definition 3.3.2 und Satz 3.3.4 gilt für die Lösungsmenge des Gleichungssystems:

$$\begin{aligned}
&\left\{ \vec{x} \in \mathbb{R}^n : \begin{array}{ccccccc}
a_{11}x_1 & + & a_{12}x_2 & + \ldots + & a_{1n}x_n & = 0 \\
a_{21}x_1 & + & a_{22}x_2 & + \ldots + & a_{2n}x_n & = 0 \\
\vdots & & \vdots & \ddots & \vdots & \vdots \\
a_{m1}x_1 & + & a_{m2}x_2 & + \ldots + & a_{mn}x_n & = 0
\end{array} \right\} \\
&= \left\{ \vec{x} \in \mathbb{R}^n : \sum_{j=1}^{n} a_{ij}x_j = 0,\ i = 1, \ldots, m \right\} \\
&= \left\{ \vec{x} \in \mathbb{R}^n : A \circ \vec{x} = \vec{0}_{\mathbb{R}^m} \right\} = \left\{ \vec{x} \in \mathbb{R}^n : F(\vec{x}) = \vec{0}_{\mathbb{R}^m} \right\} .
\end{aligned}$$

Das Lösen von linearen Gleichungssystemen ist also gleichwertig mit der „Nullstellensuche“ bei linearen Abbildungen. Daher schieben wir jetzt wieder einige Betrachtungen über lineare Abbildungen ein.

3.3.8 Definition

Es seien M_1, M_2 Mengen und $F : M_1 \to M_2$ eine Abbildung, ferner sei $M_2' \subset M_2$. Dann heißt $F^{-1}(M_2') := \{x \in M_1 :\ F(x) \in M_2'\}$ **Urbild** *von M_2' unter F.*

Die Umkehrabbildung F^{-1} von F muß nicht existieren. In der Definition wird jeder Teilmenge von M_2 eine (möglicherweise leere) Teilmenge von M_1 zugeordnet.

Der Begriff des Urbildes unter linearen Abbildungen „beachtet“ ebenfalls die Vektorraumstruktur:

3.3.9 Satz

Es seien V_1, V_2 reelle Vektorräume und $F : V_1 \to V_2$ eine lineare Abbildung. Ist $V_2' \subset V_2$ ein Untervektorraum von V_2, dann ist $V_1' := F^{-1}(V_2')$ ebenfalls ein Untervektorraum von V_1.

Beweis

(1) $V_1' \neq \emptyset$, denn wegen $F(0) = 0$ und $0 \in V_2'$ gilt $0 \in V_1'$.

(2) Seien $x, y \in V_1'$. Dann sind $F(x), F(y) \in V_2'$, also ist $F(x+y) = F(x)+F(y) \in V_2'$. Es folgt $x + y \in V_1'$.

(3) Sei $x \in V_1'$ und $a \in \mathbb{R}$. Dann ist $F(x) \in V_2'$, weiter gilt $F(a\cdot x) = a\cdot F(x) \in V_2'$. Es folgt $a \cdot x \in V_1'$. $\square$

Besonders wichtig sind die Urbilder des einfachsten Untervektorraums $\{0\}$ von V_2:

3.3.10 Definition

Es seien V_1, V_2 reelle Vektorräume und $F : V_1 \to V_2$ eine lineare Abbildung. Dann heißt

$$\ker(F) := F^{-1}(\{0\}) = \{x \in V_1 : \ F(x) = 0 \in V_2\}$$

Kern *(engl. kernel) oder* **Nullraum** *von F.*

Der Kern einer linearen Abbildung ist ein Untervektorraum.

Die Lösungsmenge eines homogenen linearen Gleichungssystems ist der Kern der zur Koeffizientenmatrix gehörigen linearen Abbildung $\mathbb{R}^n \to \mathbb{R}^m$ und folglich ein Untervektorraum des $\mathbb{R}^n$.

Wir stellen im Folgenden einen Algorithmus zur effektiven Bestimmung der Lösungsmengen (homogener) linearer Gleichungssysteme bzw. der Kerne linearer Abbildungen vor.

Gaußscher Algorithmus

Die Koeffizientenmatrix A soll auf einfache Weise so umgeformt werden, daß

- sich die Lösungsmenge des Gleichungssystems $A \circ \vec{x} = \vec{0}$ nicht ändert und
- die Matrix A eine Gestalt erhält, die ein direktes Ablesen von Vektoren ermöglicht, die $\ker(F)$ bzw. die Lösungsmenge des Gleichungssystems erzeugen.

Diesen Anforderungen werden die folgenden *elementaren Zeilenumformungen* gerecht:

(1) Die i-te Zeile wird mit einer Zahl $c \in \mathbb{R} \setminus \{0\}$ multipliziert. Die Zeile

$$a_{i1} \, \dots \, a_{in}$$

geht also über in

$$ca_{i1} \, \dots \, ca_{in},$$

alle anderen Zeilen bleiben unverändert. Wegen $c \neq 0$ gilt

$$a_{i1}x_1 + \ldots + a_{in}x_n = 0$$

genau dann, wenn

$$ca_{i1}x_1 + \ldots + ca_{in}x_n = 0$$

erfüllt ist. Die Lösungsmenge ändert sich also nicht.

(2) Für Indizes $\ell \neq i$ und eine reelle Zahl c wird das c-fache der i-ten Zeile zur ℓ-ten Zeile addiert, alle anderen Zeilen bleiben unverändert. (Insbesondere die i-te Zeile.) Die Zeile

$$a_{\ell 1} \, \dots \, a_{\ell n}$$

geht also über in

$$a_{\ell 1} + ca_{i1} \, \dots \, a_{\ell n} + ca_{in}.$$

Das gleichzeitige Erfülltsein von

$$a_{i1}x_1 + \ldots + a_{in}x_n = 0 \quad \textit{und} \quad a_{\ell 1}x_1 + \ldots + a_{\ell n}x_n = 0$$

ist äquivalent zur gleichzeitigen Gültigkeit von

$$a_{i1}x_1 + \ldots + a_{in}x_n = 0 \quad \textit{und} \quad (a_{\ell 1} + ca_{i1})x_1 + \ldots + (a_{\ell n} + ca_{in})x_n = 0.$$

(3) Zwei Zeilen werden miteinander vertauscht. Offensichtlich ändert auch diese Umformung die Lösungsmenge nicht.

3.3.11 Satz

Sei A eine $m \times n$-Matrix. Durch elementare Zeilenumformungen kann erreicht werden, daß A die folgende Gestalt annimmt:

$$\begin{array}{cc}
\begin{smallmatrix}\text{Spalte/}\\ \text{Zeile}\end{smallmatrix} & \begin{array}{cccccccccccccccc} 1 & & & j_1 & & & & j_2 & & \cdots & & j_s & & & & n \end{array} \\
\begin{array}{c} 1 \\ \\ \\ \\ s \\ s{+}1 \\ \\ m \end{array} &
\begin{pmatrix}
0 & \dots & 0 & 1 & * & \dots & * & 0 & * & \dots & * & 0 & * & \dots & * \\
 & & & 0 & & & & 1 & & & & \vdots & & & \\
\vdots & \ddots & \vdots & 0 & \vdots & \ddots & \vdots & 0 & \vdots & \ddots & \vdots & 0 & \vdots & \ddots & \vdots \\
 & & & \vdots & & & & \vdots & & & & 0 & & & \\
0 & \dots & 0 & 0 & * & \dots & * & 0 & * & \dots & * & 1 & * & \dots & * \\
0 & \dots & & 0 & & \dots & & \dots & & \dots & & 0 & & \dots & 0 \\
\vdots & & & & & & & \ddots & & & & & & & \vdots \\
0 & \dots & & 0 & & \dots & & \dots & & \dots & & 0 & & \dots & 0
\end{pmatrix}
\end{array}.$$

In den mit „$*$“ gekennzeichneten Bereichen befinden sich Komponenten, über die keine Aussage gemacht wird.

Beweis

Ohne Einschränkung können wir annehmen, daß mindestens eine Komponente der Matrix ungleich 0 ist. Es sei $j_1 \in \{1, \ldots, n\}$ der kleinste Spaltenindex mit $\vec{a}_{j_1} \neq \vec{0}$. Also gibt es ein $i_1 \in \{1, \ldots, m\}$ mit $a_{i_1 j_1} \neq 0$. Man multipliziere die i_1-te Zeile mit $\frac{1}{a_{i_1 j_1}}$ und addiere für alle $\ell \neq i_1$, $\ell \in \{1, \ldots, m\}$ das $-a_{\ell j_1}$-fache der i_1-ten Zeile zur ℓ-ten Zeile. So erhält die j_1-te Spalte die Gestalt

$$\begin{array}{r} \\ \\ i_1\text{-te Zeile} \\ \\ \\ \end{array} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}.$$

Dieses Verfahren wird nun wiederholt, allerdings sind $j_2 > j_1$ kleinstmöglich und $i_2 \in \{1, \ldots, m\} \setminus \{i_1\}$ so zu wählen, daß $a_{i_2 j_2} \neq 0$ gilt. Dadurch bleibt die j_1-te Spalte unverändert.

Im k-ten Schritt sind $j_k > j_{k-1}$ und $i_k \in \{1, \ldots, m\} \setminus \{i_1, \ldots, i_{k-1}\}$ zu wählen. Das Verfahren wird nach s Schritten abgebrochen, wenn $s = m$ ist, oder wenn für die verbleibenden Zeilenindizes $i \in \{1, \ldots, m\} \setminus \{i_1, \ldots, i_s\}$ die zugehörigen Zeilen nur noch aus Nullen bestehen. Die Spalten j_k haben für $k = 1, \ldots, s$ die Gestalt

$$\begin{array}{r} \\ \\ i_k\text{-te Zeile} \\ \\ \\ \end{array} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}.$$

Durch Zeilenvertauschungen wird die im Satz behauptete Gestalt erreicht: Die Zeile i_1 wird in die erste Zeile geschrieben, die Zeile i_2 in die zweite Zeile, usw. An das untere Ende der Matrix schreibt man die Null-Zeilen. $\square$

3.3.12 Beispiel

$$A = \begin{pmatrix} 2 & 2 & 2 & 4 & 0 \\ \boxed{1} & 1 & 2 & 3 & 4 \\ 1 & 1 & 2 & 3 & 3 \\ -1 & -1 & -3 & -4 & -5 \end{pmatrix} \to \begin{pmatrix} 0 & 0 & -2 & -2 & -8 \\ 1 & 1 & 2 & 3 & 4 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & \boxed{\text{-1}} & -1 & -1 \end{pmatrix}$$

$$\to \begin{pmatrix} 0 & 0 & -2 & -2 & -8 \\ 1 & 1 & 2 & 3 & 4 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & \boxed{1} & 1 & 1 \end{pmatrix} \to \begin{pmatrix} 0 & 0 & 0 & 0 & -6 \\ 1 & 1 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & \boxed{\text{-1}} \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

$$\to \begin{pmatrix} 0 & 0 & 0 & 0 & -6 \\ 1 & 1 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix} \qquad \to \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

$$\to \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Es bleibt anzugeben, wie die Lösungen von $A \circ \vec{x} = \vec{0} \in \mathbb{R}^m$ zu bestimmen sind, wenn die $m \times n$-Matrix A die in Satz 3.3.11 beschriebene Gestalt hat. Der Spezialfall $n = s$ ist einfach: $x_1 = x_2 = \ldots = x_n = 0$ ist die einzige Lösung.

Sei also nun $s < n$. Wir bezeichnen mit $1 \leq k_1 < k_2 < \ldots < k_{n-s} \leq n$ die Spaltenindizes aus $\{1, \ldots, n\} \backslash \{j_1, \ldots, j_s\}$. Dann wird durch A das folgende Gleichungssystem gegeben:

$$\begin{array}{ccccccc} x_{j_1} & = & -a_{1k_1}x_{k_1} & - & a_{1k_2}x_{k_2} & - \ldots - & a_{1k_{n-s}}x_{k_{n-s}} \\ x_{j_2} & = & -a_{2k_1}x_{k_1} & - & a_{2k_2}x_{k_2} & - \ldots - & a_{2k_{n-s}}x_{k_{n-s}} \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ x_{j_s} & = & -a_{sk_1}x_{k_1} & - & a_{sk_2}x_{k_2} & - \ldots - & a_{sk_{n-s}}x_{k_{n-s}}. \end{array}$$

Man erhält alle Lösungen dieses Gleichungssystems, indem $(x_{k_1}, \ldots, x_{k_{n-s}})$ alle Werte aus $\mathbb{R}^{n-s}$ durchläuft und die zugehörigen $x_{j_1}, \ldots, x_{j_s}$ berechnet werden. Wählt man für $(x_{k_1}, \ldots, x_{k_{n-s}})$ die $(n-s)$ linear unabhängigen Vektoren $\vec{e}_1, \ldots, \vec{e}_{n-s} \in \mathbb{R}^{n-s}$, so erhält man $(n-s)$ linear unabhängige Lösungsvektoren von $A \circ \vec{x} = \vec{0}$; es ist also $\dim(\text{Lösungsmenge}) \geq n - s$.

Diese Vektoren erzeugen sogar die Lösungsmenge, den Beweis dieser Tatsache verschieben wir ebenso wie den Beweis von $\dim(\text{Lösungsmenge}) = n - s$ auf den Abschnitt 3.5. Dabei werden wir von der zwischenzeitlich weiter entwickelten Theorie profitieren. Dort werden wir ebenfalls inhomogene lineare Gleichungssysteme $A \circ \vec{x} = \vec{b}$ mit gegebenem $\vec{b} \in \mathbb{R}^m$ behandeln.

3.3.13 Folgerung

Sei $F : \mathbb{R}^n \to \mathbb{R}^m$ *eine lineare Abbildung mit* $m < n$. *Dann ist* $\dim \ker(F) > 0$.

Beweis

Sei A die zugehörige Matrix. Für die Zahl s der linear unabhängigen Zeilen aus Satz 3.3.11 ist $s \leq m$, daher liefern die obigen Überlegungen: $\dim \ker(F) \geq n - s \geq n - m > 0$. □

Dimension und Basislänge

Mit Hilfe der vorhergehenden Resultate gelingt nun der Nachweis, daß die Dimension als Maximalzahl linear unabhängiger Vektoren gleich der Länge einer *jeden* Basis ist.

Wir untersuchen zunächst $\mathbb{R}^m$:

3.3.14 Satz

Für alle $m \in \mathbb{N}$ *gilt* $\dim \mathbb{R}^m = m$.

Beweis

Durch Angabe der kanonischen Basis aus Einheitsvektoren hatten wir bereits $\dim \mathbb{R}^m \geq m$ gezeigt. Es bleibt also nachzuweisen, daß es *höchstens* m linear unabhängige Vektoren in $\mathbb{R}^m$ gibt. Seien also $\vec{a}_1, \dots, \vec{a}_n \in \mathbb{R}^m$ linear unabhängige Vektoren. Wir schreiben die $\vec{a}_i$ als Spalten der $m \times n$-Matrix $A = (\vec{a}_1, \dots, \vec{a}_n)$. Sei F die zu A gehörige lineare Abbildung $\mathbb{R}^n \to \mathbb{R}^m$, dann ist $F(\vec{x}) = \vec{0}$ äquivalent zu

$$\begin{array}{ccccccc} a_{11}x_1 & + & a_{12}x_2 & + \dots + & a_{1n}x_n & = 0 \\ a_{21}x_1 & + & a_{22}x_2 & + \dots + & a_{2n}x_n & = 0 \\ \vdots & & \vdots & \ddots & \vdots & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + \dots + & a_{mn}x_n & = 0, \end{array}$$

d. h. zu $x_1\vec{a}_1 + \cdots + x_n\vec{a}_n = \vec{0} \in \mathbb{R}^m$. Aus der linearen Unabhängigkeit von $\vec{a}_1, \dots, \vec{a}_n$ folgt $\vec{x} = (x_1, \dots, x_n) = \vec{0}$. Also ist $\ker(F) = \left\{\vec{0}\right\}$. Wie vorher folgt $n - m \leq \dim \ker(F) = 0$, damit ist $n \leq m$. □

Mit Hilfe von Karten bzw. Koordinatensystemen können wir dieses Resultat auf beliebige Vektorräume ausdehnen („liften“).

3.3.15 Satz

Sei V *ein Vektorraum mit Basis* $\{x_1, \dots, x_n\}$*. Dann gilt:*

(a) *Sind* $y_1, \dots, y_m$ *linear unabhängig, so gilt* $m \leq n$.

(b) *Ist* $\{y_1, \dots, y_m\}$ *eine Basis von* V*, so gilt* $m = n$.

(c) $\dim V = n$.

Insbesondere ist *jede* Basis ein linear unabhängiges System von größtmöglicher Länge.

Beweis

(a) Wir betrachten den durch $G(x_i) = \vec{e}_i$ gegebenen Isomorphismus $G : V \to \mathbb{R}^n$. Wegen der Folgerungen 3.2.14 sind $G(y_1), \dots, G(y_m) \in \mathbb{R}^n$ linear unabhängig, es folgt $m \leq n$.

(b) Da sowohl $\{y_1, \ldots, y_m\}$ als auch $\{x_1, \ldots, x_n\}$ Basen von V sind, gilt $V \approx \mathbb{R}^n$ und $V \approx \mathbb{R}^m$, also $\mathbb{R}^n \approx \mathbb{R}^m$. Die Folgerungen 3.2.14 ergeben $n = \dim \mathbb{R}^n = \dim \mathbb{R}^m = m$.

(c) Folgt aus (a). □

In der Literatur wird dieser Satz meist abstrakter bewiesen, etwa durch den *Austauschsatz von Steinitz*. Dessen Rolle hat hier der Gaußsche Algorithmus übernommen.

3.3.16 Satz

Sei V sei ein reeller Vektorraum und $V' \subset V$ ein Untervektorraum.

(a) Es gilt $\dim V' \leq \dim V$.

(b) Ist $\dim V' < \infty$, *so gilt:* $V' \neq V \Leftrightarrow \dim V' < \dim V$.

Beweis

(a) Folgt aus der Definition der Dimension: Kapitel 1.3.

(b) „$\Rightarrow$“: Es sei $V' \subset V$, $V' \neq V$. Nach Satz 1.3.17 exisitiert eine Basis $\{x_1, \ldots, x_m\}$ von V' mit $m = \dim V'$. Es existiert ein $x \in V \setminus V'$, wegen Satz 1.3.8 sind $x, x_1, \ldots, x_m$ linear unabhängig. Also gilt $\dim V \geq m + 1$.
„$\Leftarrow$“: Wäre $V' = V$, so wäre auch $\dim V' = \dim V$. □

3.3.17 Bemerkung

Wir haben wiederholt den Vektorraum der reellen Polynome betrachtet, zuletzt in Beispiel 3.2.12. Dort hatten wir gesehen, daß dieser echte Untervektorräume besitzt, die auch von unendlicher Dimension sind. Das bedeutet: Die Endlichkeitsvoraussetzung in Teil (b) des vorhergehenden Satzes ist notwendig.

3.3.18 Satz

Sei V ein reeller Vektorraum mit $\dim V = m < \infty$. *Für die Vektoren* $x_1, \ldots, x_m \in V$ *gelte* eine *der folgenden Bedingungen:*

(a) $x_1, \ldots, x_m$ *sind linear unabhängig.*

(b) $x_1, \ldots, x_m$ *erzeugen* V.

Dann ist $\{x_1, \ldots, x_m\}$ *eine Basis von* V.

Beweis

(a) Siehe Satz 1.3.16.

(b) Es ist zu zeigen: $x_1, \ldots, x_m$ sind linear unabhängig. Wir führen einen Widerspruchsbeweis und nehmen dazu an, dieses sei nicht der Fall. Es sei s die maximale Anzahl linear unabhängiger Vektoren von $x_1, \ldots, x_m$; es ist $s < m$. Gegebenenfalls nach Umnumerierung kann man erreichen: $x_1, \ldots, x_s$ sind linear unabhängig, $x_1, \ldots, x_s, x_{s+i}$ sind linear abhängig für $i = 1, \ldots, m - s$. Nach

Satz 1.3.8 ist jedes x_{s+i} für $i = 1, \dots, m-s$ Linearkombination von $x_1, \dots, x_s$. Folglich erzeugen $x_1, \dots, x_s$ schon V und bilden deshalb eine Basis. Also ist $\dim V = s < m$ nach Satz 3.3.15. Wir erhalten einen Widerspruch! Deshalb ist die Annahme falsch und die Behauptung damit bewiesen. □

Dasselbe Argument zeigt:

3.3.19 Folgerung

Es seien $x_1, \dots, x_n$ Erzeuger eines m-dimensionalen Vektorraums V. Dann gilt $n \geq m$, und es ist $n = m$ genau dann, wenn $\{x_1, \dots, x_n\}$ eine Basis von V ist.

Jedes System von Erzeugern von kleinstmöglicher Länge ist also eine Basis.

3.4 Rang und Corang

In diesem Abschnitt werden wir eines der wichtigsten Resultate aus der Theorie der linearen Abbildungen herleiten: In der „Dimensionsformel" wird ein einfacher Zusammenhang zwischen den Dimensionen von Urbildraum, Kern und Bild hergestellt.

Rang und Corang einer linearen Abbildung

Zunächst wollen wir Injektivität linearer Abbildungen genauer charakterisieren. Injektivität einer *beliebigen* Abbildung bedeutet, daß jeder Punkt im Wertebereich höchstens ein Urbild hat. Im folgenden Satz werden wir sehen, daß das bei linearen Abbildungen gleichwertig damit ist, daß dieses für den speziellen Wert 0 richtig ist. Damit wird erneut deutlich, eine wie starke Eigenschaft die Linearität einer Abbildung ist.

Es bezeichnen $V, V_1, V_2, \dots$ wieder stets reelle Vektorräume.

3.4.1 Satz

Eine lineare Abbildung $F : V_1 \to V_2$ ist genau dann injektiv, wenn $\ker(F) = \{0\}$ ist.

Letzteres ist gleichbedeutend mit $\dim \ker(F) = 0$.

Beweis

„$\Rightarrow$": Wegen $F(0) = 0$ folgt aus der Injektivität $\ker(F) = \{0\}$.

„$\Leftarrow$": Es sei $\ker(F) = \{0\}$. Für $x, y \in V_1$ mit $F(x) = F(y)$ gilt $0 = F(x) - F(y) = F(x-y)$, d. h. $(x-y) \in \ker(F)$. Man erhält $x - y = 0$, d. h. $x = y$. □

Wir haben nur benutzt, daß $(V_i, +)$, $i = 1, 2,$ Gruppen sind. Ein entsprechender Satz gilt also auch für Gruppenhomomorphismen.

Die Surjektivität einer linearen Abbildung $F : V_1 \to V_2$ wird durch den Bildraum $\mathrm{im}(F) = F(V_1)$ bzw. dessen Dimension bestimmt. Bereits in Satz 3.2.7 hatten wir gesehen, daß das Bild $\mathrm{im}(F)$ tatsächlich ein Untervektorraum von V_2 ist.

3.4.2 Satz

Es sei $\dim V_1 = n < \infty$ *und* $F : V_1 \to V_2$ *eine lineare Abbildung. Dann ist* F *surjektiv genau dann, wenn* $\dim \operatorname{im}(F) = \dim V_2$ *gilt.*

Beweis

„$\Rightarrow$“: Ist F surjektiv, so ist $\operatorname{im}(F) = V_2$.

„$\Leftarrow$“: Sicher ist $\dim \operatorname{im}(F) \le n < \infty$. Deshalb ist $\operatorname{im}(F) = V_2$ gemäß Satz 3.3.16.

Diese beiden Sätze deuten schon an, daß die Zahlen $\dim \ker(F)$ und $\dim \operatorname{im}(F)$ eine große Rolle spielen werden.

3.4.3 Definition

Sei $F : V_1 \to V_2$ *eine lineare Abbildung. Man erklärt den* **Rang** *von* F *durch* $\operatorname{rg}(F) := \dim \operatorname{im}(F) = \dim F(V_1)$ *und den* **Corang** *oder* **Defekt** *von* F *durch* $\operatorname{crg}(F) := \dim \ker(F)$.

3.4.4 Satz (Dimensionsformel)

Es sei $\dim V_1 = n < \infty$ *und* $F : V_1 \to V_2$ *eine lineare Abbildung. Dann gilt:*

$$\dim V_1 = \operatorname{rg}(F) + \operatorname{crg}(F).$$

Beweis

Wegen $\ell := \dim \ker(F) \le \dim V_1 < \infty$ existiert eine Basis $\{x_1, \ldots, x_\ell\}$ von $\ker(F)$. Wir zeigen nun, daß diese zu einer Basis von V_1 ergänzt werden kann (Basisergänzungssatz). Dazu wähle man sukzessive Vektoren $x_{\ell+i} \in V_1$, $(i = 1, 2, \ldots)$ derart, daß $x_1, \ldots, x_{\ell+i}$ linear unabhängig sind. Das Verfahren wird abgebrochen, wenn kein derartiger Vektor mehr gefunden werden kann. Dieses ist möglicherweise schon nach dem „nullten“ Schritt und wegen $\dim V_1 = n$ spätestens nach $i = n - \ell$ Schritten der Fall.

Auf diese Weise ist ein System $x_1, \ldots, x_k$, $(\ell \le k \le n)$ linear unabhängiger Vektoren entstanden, wobei für alle $x \in V_1$ das System $x, x_1, \ldots, x_k$ linear abhängig ist. Nach Satz 1.3.8 wird V_1 von $x_1, \ldots, x_k$ erzeugt. Also ist $\{x_1, \ldots, x_k\}$ eine Basis von V_1 und deswegen $k = \dim V_1 = n$.

Mit Hilfe der Basisvektoren $x_{\ell+1}, \ldots, x_n$, die nicht zur Basis von $\ker(F)$ gehören, wird eine Basis von $F(V_1)$ konstruiert. Für $y_{\ell+1} := F(x_{\ell+1}), \ldots, y_n := F(x_n)$ werden wir zeigen:

(1) $y_{\ell+1}, \ldots, y_n$ sind linear unabhängig;

(2) $y_{\ell+1}, \ldots, y_n$ erzeugen $F(V_1)$;

folglich ist $\{y_{\ell+1}, \ldots, y_n\}$ eine Basis von $F(V_1)$.

Zum Nachweis von (1) betrachte man den Nullvektor als Linearkombination von $y_{\ell+1},\ldots,y_n$:

$$a_{\ell+1}y_{\ell+1}+\ldots+a_ny_n=0\in V_2 \quad \text{mit} \quad a_{\ell+1},\ldots,a_n\in\mathbb{R}.$$

Wegen $0=a_{\ell+1}F(x_{\ell+1})+\ldots+a_nF(x_n)=F(a_{\ell+1}x_{\ell+1}+\ldots+a_nx_n)$ ist $a_{\ell+1}x_{\ell+1}+\ldots+a_nx_n\in\ker(F)$ und kann daher als Linearkombination von $x_1,\ldots,x_\ell$ dargestellt werden: Es existieren $a_1,\ldots,a_\ell\in\mathbb{R}$ mit: $a_{\ell+1}x_{\ell+1}+\ldots+a_nx_n=a_1x_1+\ldots+a_\ell x_\ell$, d. h. $-a_1x_1-\ldots-a_\ell x_\ell+a_{\ell+1}x_{\ell+1}+\ldots+a_nx_n=0\in V_1$. Die lineare Unabhängigkeit von $x_1,\ldots,x_n$ ergibt insbesondere

$$a_{\ell+1}=\ldots=a_n=0.$$

Zum Nachweis von (2) sei ein beliebiges $y\in F(V_1)$ gegeben. Es existiert ein Urbild $x\in V_1$ mit $F(x)=y$, dieses besitzt die Darstellung:

$$x=a_1x_1+\ldots+a_nx_n, \qquad a_1,\ldots,a_n\in\mathbb{R}.$$

Es folgt

$$\begin{aligned}y=F(x)&=a_1F(x_1)+\ldots+a_\ell F(x_\ell)+a_{\ell+1}F(x_{\ell+1})+\ldots+a_nF(x_n)\\&=a_{\ell+1}y_{\ell+1}+\ldots+a_ny_n\in\mathbb{R}\cdot(y_{\ell+1},\ldots,y_n)\,.\end{aligned}$$

Bisher wurde also gezeigt: $\{x_1,\ldots,x_\ell\}$ ist eine Basis von $\ker(F)$, $\{x_1,\ldots,x_n\}$ von V_1 und $\{y_{\ell+1},\ldots,y_n\}$ von $F(V_1)$. Mit Satz 3.3.15 folgt $\ell=\dim\ker(F)=\text{crg}(F)$ und $n=\dim V_1$ und schließlich

$$\text{rg}(F)=\dim F(V_1)=n-\ell=\dim V_1-\text{crg}(F).$$

Man beachte, daß im Fall $\ell=0$ bzw. $\ell=n$ die Basen $\{x_1,\ldots,x_\ell\}$ bzw. $\{y_{\ell+1},\ldots,y_n\}$ durch die leere Menge $\emptyset$ zu ersetzen sind. □

Stimmen Urbildbereich und Wertevorrat einer linearen Abbildung überein, so ergibt sich durch Kombination der vorhergehenden drei Sätze ein bemerkenswerter Zusammenhang zwischen Injektivität und Surjektivität. Die Voraussetzung endlicher Dimension ist dabei ganz wesentlich, wie das Beispiel 3.2.12 (b) zeigt.

3.4.5 Folgerung

Es sei $\dim V<\infty$ und $F:V\to V$ eine lineare Abbildung. Dann ist F surjektiv genau dann, wenn F injektiv ist.

3.4.6 Beispiel

Es soll der Rang der linearen Abbildung

$$F:\mathbb{R}^3\to\mathbb{R}^3\,, \qquad F(\vec{c}):=(c_1-2c_2,c_2+c_3,2c_1+3c_2+7c_3)$$

bestimmt werden. Wegen $F(\vec{c}) = c_1(1,0,2) + c_2(-2,1,3,) + c_3(0,1,7)$ wird $\mathrm{im}(F)$ von $\vec{b}_1 := (1,0,2)$, $\vec{b}_2 := (-2,1,3)$ und $\vec{b}_3 = (0,1,7)$ erzeugt. Allerdings sind $\vec{b}_1, \vec{b}_2, \vec{b}_3$ linear abhängig: $2\vec{b}_1 + \vec{b}_2 - \vec{b}_3 = \vec{0}$.

Dagegen sind $\vec{b}_1$ und $\vec{b}_2$ linear unabhängig:

$$d_1\vec{b}_1 + d_2\vec{b}_2 = \vec{0} \;\Rightarrow\; \begin{cases} d_1 - 2d_2 = 0 \\ \phantom{d_1 -{}} d_2 = 0 \\ 2d_1 + 3d_2 = 0 \end{cases} \;\Rightarrow\; d_2 = 0 \text{ und } d_1 = 0.$$

Wegen $\vec{b}_3 = 2\vec{b}_1 + \vec{b}_2$ ist stets $F(\vec{c}) = c_1\vec{b}_1 + c_2\vec{b}_2 + c_3\vec{b}_3 = (c_1 + 2c_3)\vec{b}_1 + (c_2 + c_3)\vec{b}_2$, also wird $\mathrm{im}(F)$ von $\vec{b}_1$, $\vec{b}_2$ auch erzeugt, und $\left\{\vec{b}_1,\ \vec{b}_2\right\}$ ist eine Basis von $\mathrm{im}(F)$. Wir erhalten $\mathrm{rg}(F) = 2$.

Rang einer Matrix

Die Bestimmung des Ranges einer linearen Abbildung gemäß seiner Definition ist eine langwierige Aufgabe: Man muß Systeme von Erzeugern von $\mathrm{im}(F)$ bestimmen und durch Überprüfen der linearen Unabhängigkeit bzw. Abhängigkeit zu einer Basis übergehen. Deren Länge ist dann $\mathrm{rg}(F)$.

Wir wollen uns wieder den Matrizenkalkül zur effektiven Behandlung dieser Aufgabe zu Nutze machen, indem wir einen Zusammenhang zwischen dem Rang von F und der maximalen Anzahl linear unabhängiger Spalten bzw. Zeilen von A herleiten.

Für den Rest des Abschnitts bezeichne $F : \mathbb{R}^n \to \mathbb{R}^m$ eine lineare Abbildung und A die zugehörige $m \times n$-Matrix.

3.4.7 Definition

Die $n \times m$-Matrix

$$A^t = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{pmatrix} = (b_{ij})_{\substack{i=1,\dots,n \\ j=1,\dots,m}}$$

mit $b_{ij} = a_{ji}$ heißt die zur $m \times n$-Matrix $A = (a_{ij})_{\substack{i=1,\dots,m \\ j=1,\dots,n}}$ **transponierte Matrix**. *Die zu A^t gehörige lineare Abbildung $F^* : \mathbb{R}^m \to \mathbb{R}^n$ heißt die zu F* **duale Abbildung**.

Ist $n = m$ und stimmt A mit der transponierten Matrix A^t überein, so daß $a_{ij} = a_{ji}$ für alle $i, j = 1, \dots, n$ gilt, dann nennt man A **symmetrisch**.

Die transponierte Matrix A^t entsteht, indem man die Zeilen von A als Spalten schreibt.

Auf die Bedeutung der dualen Abbildung werden wir in den Abschnitten 4.3 und 4.4 über Dualräume eingehen. Im Zusammenhang mit der Betrachtung reeller Vektorräume, die mit einem Skalarprodukt (Orthogonalitätsbegriff) versehen sind, spricht man von *adjungierten Abbildungen.* Dieses Thema werden wir in Kapitel 7 ausführlich behandeln.

3.4.8 Definition

Wir definieren den **Spaltenrang** *einer Matrix A als Rang der zugehörigen linearen Abbildung F und den* **Zeilenrang** *von A als Spaltenrang der transponierten Matrix A^t:*

$$\text{Spaltenrang}(A) := \text{rg}(F).$$

$$\text{Zeilenrang}(A) := \text{Spaltenrang}(A^t).$$

Die folgende Bemerkung ist grundlegend für die weiteren Überlegungen.

3.4.9 Bemerkung

Die Spalten $\vec{a}_1, \ldots, \vec{a}_n$ von A erzeugen $F(\mathbb{R}^n)$.

Beweis

Folgt aus $A \circ \vec{x} = \sum_{j=1}^{n} x_j \vec{a}_j$ für $\vec{x} \in \mathbb{R}^n$. □

Man beachte die Dualität der folgenden Resultate zum Gaußschen Algorithmus. Zur Bestimmung von $\text{im}(F) = F(\mathbb{R}^n)$ und $\text{rg}(F)$ eignen sich die folgenden *elementaren Spaltenumformungen* für die Matrix A, da diese $\text{im}(F)$ *nicht* ändern (wohl aber $\ker(F)$ und F selbst):

(1) Für ein $j \in \{1, \ldots, n\}$ und ein reelles $c \neq 0$ wird die j-te Spalte mit c multipliziert, die übrigen Spalten werden nicht geändert.

Die Spalte $\vec{a}_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}$ geht über in $c \cdot \vec{a}_j = \begin{pmatrix} ca_{1j} \\ ca_{2j} \\ \vdots \\ ca_{mj} \end{pmatrix}$.

Ein Vektor $\vec{y} \in \mathbb{R}^m$ ist genau dann Linearkombination von $\vec{a}_1, \ldots, \vec{a}_n$, wenn $\vec{y}$ Linearkombination von $\vec{a}_1, \ldots, \vec{a}_{j-1}, c\vec{a}_j, \vec{a}_{j+1}, \ldots, \vec{a}_n$ ist; denn $\vec{y} = x_1\vec{a}_1 + \ldots + x_n\vec{a}_n = x_1\vec{a}_1 + \ldots + \frac{x_j}{c}(c\vec{a}_j) + \ldots + x_n\vec{a}_n$. Die Bildmenge bleibt also unverändert.

(2) Für zwei Indizes $k \neq j$ und ein $c \in \mathbb{R}$ wird das c-fache der j-ten Spalte zur k-ten Spalte addiert. Die Spalte $\vec{a}_k$ wird durch $\vec{a}_k + c\vec{a}_j$ ersetzt, die übrigen Spalten bleiben unverändert. Wegen $x_1\vec{a}_1 + \ldots + x_n\vec{a}_n = x_1\vec{a}_1 + \ldots + x_k(\vec{a}_k + c\vec{a}_j) + \ldots + (x_j - cx_k)\vec{a}_j + \ldots + x_n\vec{a}_n$ ändert auch diese Umformung die Bildmenge nicht.

(3) Die Vertauschung zweier Spalten ändert ebenfalls die Bildmenge nicht.

Entsprechend Satz 3.3.11 gilt:

3.4.10 Satz

Durch elementare Spaltenumformungen kann erreicht werden, daß A die folgende Gestalt annimmt:

$$\begin{array}{cc}
\begin{array}{c}\text{Spalte /}\\ \text{Zeile}\end{array} & \begin{array}{cccccccc} 1 & & & & s & s{+}1 & & n \end{array} \\
\begin{array}{c} 1 \\ \\ \\ i_1 \\ \\ \\ \\ i_2 \\ \\ \vdots \\ \\ i_s \\ \\ \\ m \end{array} &
\begin{pmatrix}
0 & & \cdots & & 0 & 0 & \cdots & 0 \\
\vdots & & \ddots & & \vdots & \vdots & & \vdots \\
0 & & \cdots & & 0 & & & \\
1 & 0 & 0 & \cdots & 0 & 0 & & 0 \\
* & & \cdots & & * & & & \\
\vdots & & \ddots & & \vdots & \vdots & & \vdots \\
* & & \cdots & & * & & & \\
0 & 1 & 0 & \cdots & 0 & \vdots & \ddots & \vdots \\
* & & \cdots & & * & & & \\
\vdots & & \ddots & & \vdots & \vdots & & \vdots \\
* & & \cdots & & * & & & \\
0 & 0 & 0 & \cdots & 1 & 0 & & 0 \\
* & & \cdots & & * & & & \\
\vdots & & \ddots & & \vdots & \vdots & & \vdots \\
* & & \cdots & & * & 0 & \cdots & 0
\end{pmatrix}
\end{array}.$$

Beweis

Wir wenden Satz 3.3.11 auf A^t an; Spaltenumformungen bei A entsprechen Zeilenumformungen bei A^t. $\square$

Zur Bestimmung von $F(\mathbb{R}^n)$ und $\mathrm{rg}(F)$ können wir ohne Einschränkung annehmen, daß A die gerade beschriebene Gestalt hat. Die restlichen Zeilenindizes aus $\{1,\ldots,m\} \setminus \{i_1,\ldots,i_s\}$ werden mit $1 \leq \ell_1 < \ldots < \ell_{m-s} \leq m$ bezeichnet. Sei $\vec{x} \in \mathbb{R}^n$, $\quad \vec{y} = F(\vec{x}) = A \circ \vec{x}$, dann erhält man $\vec{y}$ aus den folgenden Gleichungen:

$$\begin{aligned}
y_{i_1} &= x_1 \\
\vdots\ & \ \ \vdots \\
y_{i_s} &= x_s \\
y_{\ell_1} &= a_{\ell_1 1}x_1 + \ldots + a_{\ell_1 s}x_s \\
\vdots\ & \ \ \vdots \\
y_{\ell_{m-s}} &= a_{\ell_{m-s}1}x_1 + \ldots + a_{\ell_{m-s}s}x_s.
\end{aligned}$$

Die Bildmenge $F(\mathbb{R}^n)$ wird durch ein Gleichungssystem aus $(m-s)$ Gleichungen in

m Veränderlichen gegeben:

$$\begin{aligned} y_{\ell_1} &= a_{\ell_1 1} y_{i_1} + \ldots + a_{\ell_1 s} y_{i_s} \\ \vdots \quad & \vdots \\ y_{\ell_{m-s}} &= a_{\ell_{m-s} 1} y_{i_1} + \ldots + a_{\ell_{m-s} s} y_{i_s}. \end{aligned}$$

Man erhält also $F(\mathbb{R}^n)$, indem $(x_1, \ldots, x_s)$ den $\mathbb{R}^s$ durchläuft und man $y_1, \ldots, y_m$ jeweils gemäß dem obigen Gleichungssystem bestimmt.

Dadurch wird eine lineare Abbildung

$$G : \mathbb{R}^s \to \mathbb{R}^m, \qquad (x_1, \ldots x_s) \mapsto (y_1, \ldots, y_m)$$

mit $G(\mathbb{R}^s) = F(\mathbb{R}^n)$ gegeben. Diese ist injektiv, denn aus $y_1 = \ldots = y_m = 0$ folgt $x_1 = y_{i_1} = 0, \ldots, x_s = y_{i_s} = 0$. Also ist G ein Isomorphismus von $\mathbb{R}^s$ auf $G(\mathbb{R}^s)$. Da Isomorphismen die Dimension erhalten, haben wir schließlich

$$\text{Spaltenrang}(A) \overset{\text{def}}{=} \text{rg}(F) = \dim F(\mathbb{R}^n) = \dim G(\mathbb{R}^s) = \dim \mathbb{R}^s = s.$$

Der Spaltenrang(A) kann also unmittelbar von der Matrix aus Satz 3.4.10 abgelesen werden.

Den s-dimensionalen Untervektorraum $F(\mathbb{R}^n)$ von $\mathbb{R}^m$ nennt man auch *s-dimensionale Ebene durch* $\vec{0} \in \mathbb{R}^m$.

3.4.11 Satz

Der Spaltenrang(A) *ist die maximale Anzahl linear unabhängiger Spalten der Matrix* A.

Beweis

Es bezeichne s^* die maximale Zahl linear unabhängiger Spalten von A. Wir finden also linear unabhängige Spalten $\vec{a}_{j_1}, \ldots, \vec{a}_{j_{s^*}}$ von A. Dann ist jede andere Spalte Linearkombination von $\vec{a}_{j_1}, \ldots, \vec{a}_{j_{s^*}}$. Da $\text{im}(F)$ von $\vec{a}_1, \ldots, \vec{a}_n$ erzeugt wird, ist $\{\vec{a}_{j_1}, \ldots, \vec{a}_{j_{s^*}}\}$ eine Basis von $\text{im}(F)$. Es folgt $s^* = \dim \text{im}(F) = \text{rg}(F) \overset{\text{def}}{=}$ Spaltenrang(A). □

3.4.12 Folgerung

Der Zeilenrang(A) *ist die maximale Anzahl linear unabhängiger Zeilen von* A.

Beweis

Folgt sofort aus Satz 3.4.11 und Definition 3.4.8, weil die Zeilen von A die Spalten von A^t sind. □

Bei der Bestimmung von $\ker(F)$ dürfen nur elementare Zeilenumformungen durchgeführt werden, bei der Bestimmung von $\mathrm{im}(F)$ dagegen nur elementare Spaltenumformungen. Die Dimensionsformel gestattet den deshalb bemerkenswerten Satz:

3.4.13 Satz

Zeilenrang(A) *und* Spaltenrang(A) *ändern sich bei elementaren (Zeilen- oder Spalten-) Umformungen nicht.*

Beweis

Es reicht, die Behauptung für Spaltenrang(A) zu zeigen; das Resultat für den Zeilenrang erhält man durch Betrachtung der transponierten Matrix. Die Matrix $\tilde{A}$ gehe durch eine elementare Umformung aus A hervor, es sei $\tilde{F}$ die zu $\tilde{A}$ gehörige Abbildung. Handelt es sich um eine Spaltenumformung, so gilt sogar $\mathrm{im}(F) = \mathrm{im}(\tilde{F})$. Handelt es sich dagegen um eine Zeilenumformung, so gilt nach Abschnitt 3.3: $\ker(F) = \ker(\tilde{F})$. In diesem Fall ergibt die Dimensionsformel: Spaltenrang$(A) =$ $\mathrm{rg}(F) = \dim(\mathbb{R}^n) - \mathrm{crg}(F) = \dim(\mathbb{R}^n) - \mathrm{crg}(\tilde{F}) = \mathrm{rg}(\tilde{F}) =$ Spaltenrang$(\tilde{A})$. □

3.4.14 Satz

Zeilenrang$(A) =$ Spaltenrang(A).

Beweis

Die Matrix A werde durch Zeilenumformungen auf die Gestalt von Satz 3.3.11 gebracht, durch Spaltenumformungen kann man schließlich folgende Form erreichen:

$$\tilde{A} = \begin{array}{cc} \begin{array}{c}\text{Spalte/}\\ \text{Zeile}\end{array} & \begin{array}{cccccccc} 1 & 2 & 3 & & s & s+1 & & n \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ \\ s \\ s+1 \\ \\ m \end{array} & \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 & & & \\ 0 & 0 & 1 & \dots & 0 & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & & & \\ 0 & 0 & 0 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \end{pmatrix} \end{array}.$$

Wie wir oben hergeleitet haben, ist Spaltenrang$(\tilde{A}) =$ Spaltenrang(A), Zeilenrang$(\tilde{A})$ $=$ Zeilenrang(A). Man liest unmittelbar ab, daß s sowohl die maximale Anzahl linear unabhängiger Spalten als auch Zeilen von $\tilde{A}$ ist:

$$\text{Spaltenrang}(\tilde{A}) = s = \text{Zeilenrang}(\tilde{A}).$$ □

Für die zugehörigen linearen Abbildungen bedeutet dieser Satz:

3.4.15 Folgerung

Jede lineare Abbildung $F : \mathbb{R}^n \to \mathbb{R}^m$ hat denselben Rang wie ihre duale Abbildung $F^* : \mathbb{R}^m \to \mathbb{R}^n$. Ist $n = m$, so haben beide Abbildungen auch denselben Corang.

Satz 3.4.14 erlaubt es, die Unterscheidung von Zeilen- und Spaltenrang aufzugeben und einfach vom *Rang einer Matrix* A zu sprechen, der wie im obigen Beweis bestimmt wird:

$$\text{Rang}\,(A) := \text{Spaltenrang}(A) = \text{Zeilenrang}(A).$$

3.4.16 Beispiel

Wie in Beispiel 3.4.6 bestimmen wir den Rang der linearen Abbildung $F : \mathbb{R}^3 \to \mathbb{R}^3$, $\quad F(\vec{c}) := (c_1 - 2c_2, c_2 + c_3, 2c_1 + 3c_2 + 7c_3)$, allerdings hier mit Hilfe des Matrizenkalküls. Die zugehörige Matrix ist

$$A = \begin{pmatrix} 1 & -2 & 0 \\ 0 & 1 & 1 \\ 2 & 3 & 7 \end{pmatrix}.$$

Es ist $\text{rg}(F) = \text{Rang}\,(A)$. Indem man das (-2)-fache der ersten Zeile zur dritten Zeile addiert, erhält man:

$$\text{rg}(F) = \text{Rang} \begin{pmatrix} 1 & -2 & 0 \\ 0 & 1 & 1 \\ 0 & 7 & 7 \end{pmatrix}$$

subtrahiere 7-faches der 2. Zeile von der 3. Zeile und addiere doppeltes der 2. Zeile zur 1. Zeile

$$= \text{Rang} \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

subtrahiere doppeltes der 1. Spalte von der 3. Spalte

$$= \text{Rang} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

subtrahiere 2. Spalte von der 3. Spalte

$$= \text{Rang} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} = 2.$$

3.5 Allgemeine lineare Gleichungssysteme

Auch in diesem Abschnitt sei stets A eine $m \times n$-Matrix und $F : \mathbb{R}^n \to \mathbb{R}^m$ die zugehörige lineare Abbildung.

Homogene lineare Gleichungssysteme

Nun stehen alle Hilfsmittel zur Verfügung, um die Lösungstheorie für homogene lineare Gleichungssysteme $A \circ \vec{x} = \vec{0}$ abzuschließen. Auf S. 62 hatten wir bereits beschrieben, wie man $n - s$ linear unabhängige Lösungsvektoren, also Elemente von $\left\{\vec{x} : A \circ \vec{x} = \vec{0}\right\} = \ker(F)$ konstruiert, nachdem man das Gleichungssystem mit Hilfe des Gaußschen Algorithmus' in eine Standardform gebracht hat. Dabei ist s die maximale Anzahl linear unabhängiger Zeilen von A. Die Resultate des vorhergehenden Abschnitts zeigen $s = \text{Zeilenrang}(A) = \text{Spaltenrang}(A) = \text{rg}(F)$, auf Grund der Dimensionsformel haben wir dann

$$\dim(\text{Lösungsmenge}) = \dim\ker(F) \overset{\text{def}}{=} \text{crg}(F) = n - \text{rg}(F) = n - s.$$

Bereits in Satz 3.3.18 hatten wir gesehen, daß ein System linear unabhängiger Vektoren, deren Anzahl gleich der Dimension des Vektorraums ist, eine Basis desselben bildet. Also ist durch die auf S. 62 konstruierten $(n - s)$ Vektoren tatsächlich eine Basis der Lösungsmenge $\left\{\vec{x} \in \mathbb{R}^n : A \circ \vec{x} = \vec{0}\right\}$ gegeben.

3.5.1 Beispiel

Gesucht ist die Lösungsmenge des Gleichungssystems $A \circ \vec{x} = \vec{0}$ mit der 4×5-Koeffizientenmatrix

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

In Beispiel 3.3.12 hatten wir diese Matrix durch Zeilenumformungen so vereinfacht, daß $A \circ \vec{x} = \vec{0}$ äquivalent ist zu

$$\begin{aligned} x_1 &= -x_2 - x_4 \\ x_3 &= - x_4 \\ x_5 &= 0. \end{aligned}$$

Es ist $n = 5$, $s = 3$, der Lösungsraum ist also zweidimensional. Wir setzen nacheinander $x_2 = 1, x_4 = 0$ und dann $x_2 = 0, x_4 = 1$, bestimmen dazu jeweils x_1, x_3, x_5 und erhalten mit $\{(-1, 1, 0, 0, 0)^t, (-1, 0, -1, 1, 0)^t\}$ eine Basis der Lösungsmenge $\left\{\vec{x} \in \mathbb{R}^5 : A \circ \vec{x} = \vec{0}_{\mathbb{R}^4}\right\}$.

Inhomogene lineare Gleichungssysteme

Hierunter versteht man Gleichungssysteme der Form $A \circ \vec{x} = \vec{b}$, wobei nun auch der Vektor $\vec{b} \in \mathbb{R}^m$ gegeben ist. Gesucht sind wieder Lösungsvektoren $\vec{x} \in \mathbb{R}^n$. Das

Beispiel

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad \vec{b} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

zeigt, daß hier im Gegensatz zum homogenen Fall i. a. nicht stets eine Lösung existiert.

Neben der Koeffizientenmatrix A ist auch die *erweiterte Koeffizientenmatrix* $\left(A, \vec{b}\right) \in \mathbb{R}^{m \times (n+1)}$ von Bedeutung. Denn es gilt das folgende Lösbarkeitskriterium:

3.5.2 Satz

Das Gleichungssystem $A \circ \vec{x} = \vec{b}$ *besitzt eine Lösung* $\vec{x} \in \mathbb{R}^n$ *genau dann, wenn der Rang der Koeffizientenmatrix und der Rang der erweiterten Koeffizientenmatrix übereinstimmen:* $\mathrm{Rang}\,(A) = \mathrm{Rang}\,\left(A, \vec{b}\right)$.

Beweis

Seien $\vec{a}_j$, $j = 1, \dots, n$ die Spalten von A. Wie oben erwähnt, bezeichnet F die zugehörige lineare Abbildung.

Offensichtlich besitzt das Gleichungssystem $A \circ \vec{x} = \vec{b}$ eine Lösung genau dann, wenn $\vec{b} \in \mathrm{im}(F)$ gilt. Nach Bemerkung 3.4.9 ist $\mathrm{im}(F) = \mathbb{R} \cdot (\vec{a}_1, \dots, \vec{a}_n)$. Im vorhergehenden Abschnitt hatten wir auch gesehen, daß $s := \mathrm{Rang}\,(A) = \mathrm{rg}(F) \leq n$ die Maximalzahl linear unabhängiger Spalten von A ist. Also existieren Indizes $1 \leq j_1 < \ldots < j_s \leq n$, so daß $\{\vec{a}_{j_1}, \dots, \vec{a}_{j_s}\}$ eine Basis von $\mathrm{im}(F)$ ist. Indem wir noch verwenden, daß der Rang einer Matrix die Dimension des von ihren Spalten erzeugten Untervektorraums ist, erhalten wir:

$$\begin{aligned} & A \circ \vec{x} = \vec{b} \text{ ist lösbar} \quad \Leftrightarrow \quad \vec{b} \in \mathrm{im}(F) \quad \Leftrightarrow \quad \vec{b} \in \mathbb{R} \cdot (\vec{a}_1, \dots, \vec{a}_n) \\ & \Leftrightarrow \{\vec{a}_{j_1}, \dots, \vec{a}_{j_s}\} \text{ ist eine Basis von } \mathbb{R} \cdot \left(\vec{a}_1, \dots, \vec{a}_n, \vec{b}\right) \\ & \Leftrightarrow \mathrm{Rang}\,\left(A, \vec{b}\right) = s = \mathrm{Rang}\,(A). \end{aligned}$$

□

3.5.3 Folgerung

Ist $\mathrm{Rang}\,(A^{m,n}) = m$, *so ist das Gleichungssystem* $A \circ \vec{x} = \vec{b}$ *für jedes* $\vec{b} \in \mathbb{R}^m$ *lösbar.*

Beweis

Folgt aus dem vorhergehenden Satz und $m = \mathrm{Rang}\,(A) \leq \mathrm{Rang}\,\left(A, \vec{b}\right) \leq$ Zeilenzahl $\left(A, \vec{b}\right) \leq m$. □

Im Falle der Lösbarkeit ergibt sich die Lösungsmenge des inhomogenen Gleichungssystems leicht aus der Lösungsmenge des zugehörigen homogenen Systems:

3.5.4 Satz

Sei $\vec{x}_0 \in \mathbb{R}^n$ eine Lösung des linearen Gleichungssystems $A \circ \vec{x} = \vec{b}$, wobei $A = A^{m,n}$ und $\vec{b} \in \mathbb{R}^m$ gegeben sind. Dann gilt:

$$\left\{\vec{x} \in \mathbb{R}^n : \ A \circ \vec{x} = \vec{b}\right\} = \left\{\vec{x} \in \mathbb{R}^n : \ \vec{x} = \vec{x}_0 + \vec{y} \text{ mit } A \circ \vec{y} = \vec{0}\right\}.$$

Die allgemeine Lösung des inhomogenen Systems ist die Summe aus einer speziellen Lösung des inhomogenen Systems und der allgemeinen Lösung des homogenen Systems. Ganz ähnlich wird man beispielsweise auch bei linearen Differentialgleichungen vorgehen.

Beweis

Die Behauptung folgt direkt aus dem Distributivgesetz in Folgerung 3.3.5:

$$A \circ \vec{x} = \vec{b} \Leftrightarrow A \circ \vec{x} = A \circ \vec{x}_0 \Leftrightarrow A \circ \vec{x} - A \circ \vec{x}_0 = \vec{0} \Leftrightarrow A \circ (\vec{x} - \vec{x}_0) = \vec{0}. \qquad \square$$

Wie trifft man nun die Lösbarkeitsentscheidung und bestimmt gegebenenfalls eine spezielle Lösung des inhomogenen Systems?

Wie in Abschnitt 3.3 kann man auch bei der erweiterten Koeffizientenmatrix $\left(A, \vec{b}\right)$ elementare Zeilenumformungen durchführen, diese ändern die Lösungsmenge nicht. Nach Anwendung des Gaußschen Algorithmus' (Satz 3.3.11) können wir davon ausgehen, daß das Gleichungssystem die folgende Gestalt hat:

$$\begin{array}{rcllcll}
x_{j_1} = & b_1 & -a_{1k_1}x_{k_1} & - \; a_{1k_2}x_{k_2} & - \ldots - & a_{1k_{n-s}}x_{k_{n-s}} \\
x_{j_2} = & b_2 & -a_{2k_1}x_{k_1} & - \; a_{2k_2}x_{k_2} & - \ldots - & a_{2k_{n-s}}x_{k_{n-s}} \\
\vdots \quad & \vdots & \quad\vdots & \quad\vdots & \ddots & \quad\vdots \\
x_{j_s} = & b_s & -a_{sk_1}x_{k_1} & - \; a_{sk_2}x_{k_2} & - \ldots - & a_{sk_{n-s}}x_{k_{n-s}} \\
0 = & b_{s+1} & & & & \\
\vdots \quad & \vdots & & & & \\
0 = & b_m. & & & &
\end{array}$$

Dabei sind $1 \leq j_1 < \ldots < j_s \leq n$ geeignete Spaltenindizes und s der Rang der Koeffizientenmatrix. Mit $1 \leq k_1 < \ldots < k_{n-s} \leq n$ werden die verbliebenen Spaltenindizes aus $\{1, \ldots, n\} \setminus \{j_1, \ldots, j_s\}$ bezeichnet.

Dieses Gleichungssystem ist lösbar genau dann, wenn $b_{s+1} = \ldots = b_m = 0$ gilt. In diesem Fall erhält man eine spezielle Lösung, indem man $x_{j_1} = b_1, \ldots, x_{j_s} = b_s$ und $x_{k_1} = \ldots = x_{k_{n-s}} = 0$ setzt.

3.5.5 Beispiel

Wir betrachten die 4×5-Koeffizientenmatrix A aus Beispiel 3.5.1.

(a) Für $\vec{b} = (1, -2, 0, 7)^t$ erhalten wir:

$$A = \left(\begin{array}{ccccc|c} 2 & 2 & 2 & 4 & 0 & 1 \\ \boxed{1} & 1 & 2 & 3 & 4 & -2 \\ 1 & 1 & 2 & 3 & 3 & 0 \\ -1 & -1 & -3 & -4 & -5 & 7 \end{array}\right) \to \left(\begin{array}{ccccc|c} 0 & 0 & -2 & -2 & -8 & 5 \\ 1 & 1 & 2 & 3 & 4 & -2 \\ 0 & 0 & 0 & 0 & -1 & 2 \\ 0 & 0 & \boxed{-1} & -1 & -1 & 5 \end{array}\right)$$

$$\to \left(\begin{array}{ccccc|c} 0 & 0 & -2 & -2 & -8 & 5 \\ 1 & 1 & 2 & 3 & 4 & -2 \\ 0 & 0 & 0 & 0 & -1 & 2 \\ 0 & 0 & \boxed{1} & 1 & 1 & -5 \end{array}\right) \to \left(\begin{array}{ccccc|c} 0 & 0 & 0 & 0 & -6 & -5 \\ 1 & 1 & 0 & 1 & 2 & 8 \\ 0 & 0 & 0 & 0 & \boxed{-1} & 2 \\ 0 & 0 & 1 & 1 & 1 & -5 \end{array}\right)$$

$$\to \left(\begin{array}{ccccc|c} 0 & 0 & 0 & 0 & -6 & -5 \\ 1 & 1 & 0 & 1 & 2 & 8 \\ 0 & 0 & 0 & 0 & \boxed{1} & -2 \\ 0 & 0 & 1 & 1 & 1 & -5 \end{array}\right) \to \left(\begin{array}{ccccc|c} 0 & 0 & 0 & 0 & 0 & -17 \\ 1 & 1 & 0 & 1 & 0 & 12 \\ 0 & 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 1 & 1 & 0 & -3 \end{array}\right)$$

$$\to \left(\begin{array}{ccccc|c} 1 & 1 & 0 & 1 & 0 & 12 \\ 0 & 0 & 1 & 1 & 0 & -3 \\ 0 & 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 & -17 \end{array}\right),$$

also das Gleichungssystem

$$\begin{aligned} x_1 &= 12 \quad -x_2 - x_4 \\ x_3 &= -3 \qquad\quad -x_4 \\ x_5 &= -2 \\ 0 &= -17, \end{aligned}$$

dieses besitzt keine Lösung.

(b) Für $\vec{b} = (2, 12, 9, -14)^t$ ist $A \circ \vec{x}$ dagegen äquivalent zu

$$\begin{aligned} x_1 &= 2 \quad -x_2 - x_4 \\ x_3 &= -1 \qquad -x_4 \\ x_5 &= 3. \end{aligned}$$

Dieses Gleichungssystem ist lösbar, $\vec{x}_0 = (2, 0, -1, 0, 3)^t$ ist eine spezielle Lösung. Mit den Basisvektoren $\vec{b}_1 = (-1, 1, 0, 0, 0)^t$, $\vec{b}_2 = (-1, 0, -1, 1, 0)^t$ der Lösungsmenge des homogenen Systems $\left\{\vec{y} \in \mathbb{R}^5 : A \circ \vec{y} = \vec{0}\right\}$ aus Beispiel 3.5.1 erhalten wir:

$$\left\{\vec{x} \in \mathbb{R}^5 : A \circ \vec{x} = \vec{b}\right\} = \left\{\vec{x}_0 + a_1\vec{b}_1 + a_2\vec{b}_2 : a_1, a_2 \in \mathbb{R}\right\}.$$

3.6 Matrizenoperationen

Aus dem Zusammenhang von linearen Abbildungen und Matrizen haben wir schon großen Nutzen gezogen. Wir wollen diesen Zusammenhang vertiefen, indem wir Matrizenoperationen einführen, die der gleich zu definierenden Summe und dem skalaren Vielfachen von Abbildungen und vor allem der bereits ausführlich behandelten Komposition linearer Abbildungen entsprechen.

Matrizenaddition und Skalarmultiplikation

Seien zunächst V, V_1, $V_2, \ldots$ beliebige reelle Vektorräume.

3.6.1 Definition

Die **Summe** *der Abbildungen* $F : V_1 \to V_2$ *und* $G : V_1 \to V_2$ *wird durch*

$$F + G : V_1 \to V_2, \qquad (F+G)(x) := F(x) + G(x)$$

erklärt. Für eine reelle Zahl c *ist*

$$c \cdot F : V_1 \to V_2, \qquad (c \cdot F)(x) := c \cdot F(x)$$

das **skalare Vielfache** *der Abbildung* F **(Skalarmultiplikation)**.

Bezüglich dieser *Addition von Abbildungen* hat man ein *Nullelement*

$$\mathcal{O} : V_1 \to V_2, \qquad \mathcal{O}(x) := 0_{V_2}$$

und zu jeder Abbildung $F : V_1 \to V_2$ ein *negatives Element*

$$-F : V_1 \to V_2, \qquad (-F)(x) := -F(x).$$

Aus den Vektorraumeigenschaften von V_2 – der Definitionsbereich V_1 hätte bis hier auch lediglich eine beliebige Menge sein können – folgt unmittelbar:

3.6.2 Satz

Mit dieser Addition und Skalarmultiplikation ist $\{F : \ F \text{ ist eine Abbildung } V_1 \to V_2\}$ *ein reeller Vektorraum.*

Im Folgenden spezialisieren wir uns auf *lineare* Abbildungen F, $G : V_1 \to V_2$.

Zunächst halten wir fest, daß Addition und Skalarmultiplikation die Linearität von Abbildungen erhalten:

3.6.3 Satz

Seien $F, G : V_1 \to V_2$ *lineare Abbildungen und* $c \in \mathbb{R}$*. Dann sind* $F + G : V_1 \to V_2$ *und* $c \cdot F : V_1 \to V_2$ *ebenfalls linear. Insbesondere ist* $\{F : \ F \text{ ist eine lineare Abbildung } V_1 \to V_2\}$ *ein Untervektorraum des Vektorraums aller Abbildungen* $V_1 \to V_2$.

Beweis

Für alle $a \in \mathbb{R}$ und $x, y \in V_1$ gilt:

$$\begin{aligned}
(F+G)\,(x+y) &\overset{\text{def}}{=} F(x+y) + G(x+y) = F(x) + F(y) + G(x) + G(y) \\
&\overset{\text{def}}{=} (F+G)(x) + (F+G)(y); \\
(F+G)\,(ax) &\overset{\text{def}}{=} F(ax) + G(ax) = aF(x) + aG(x) = a(F(x) + G(x)) \\
&\overset{\text{def}}{=} a(F+G)(x); \\
(c \cdot F)\,(x+y) &\overset{\text{def}}{=} c\,F(x+y) = c\,(F(x) + F(y)) = c\,F(x) + c\,F(y) \\
&\overset{\text{def}}{=} (c \cdot F)(x) + (c \cdot F)(y); \\
(c \cdot F)\,(ax) &\overset{\text{def}}{=} c\,F(ax) = c\,(a\,(F(x)) = a\,(c\,(F(x)) \overset{\text{def}}{=} a\,(c \cdot F)(x).
\end{aligned}$$

□

Indem wir nun zusätzlich $V_1 = \mathbb{R}^n$ und $V_2 = \mathbb{R}^m$ annehmen, können wir entsprechende Matrizen einführen.

3.6.4 Definition

Es seien A, $B \in \mathbb{R}^{m \times n}$ Matrizen und F, $G : \mathbb{R}^n \to \mathbb{R}^m$ die zugehörigen linearen Abbildungen. Ferner sei eine reelle Zahl $c \in \mathbb{R}$ gegeben. Dann ist die **Summe** *$A + B$ die zur linearen Abbildung $F + G$ und das* **skalare Vielfache** *$c \cdot A$ die zur linearen Abbildung $c \cdot F$ gehörige $m \times n$-Matrix.*

Diese *Matrixaddition* und *Skalarmultiplikation* lassen sich auch durch eine einfache Rechenvorschrift beschreiben:

3.6.5 Satz

Für $A = (a_{ij})_{\substack{i=1,\dots,m\\ j=1,\dots,n}}$, $B = (b_{ij})_{\substack{i=1,\dots,m\\ j=1,\dots,n}}$, $c \in \mathbb{R}$ gilt:

$$A + B = \begin{pmatrix} a_{11} + b_{11} & \dots & a_{1n} + b_{1n} \\ \vdots & & \vdots \\ a_{m1} + b_{m1} & \dots & a_{mn} + b_{mn} \end{pmatrix} = (a_{ij} + b_{ij})_{\substack{i=1,\dots,m\\ j=1,\dots,n}}\,,$$

$$c \cdot A := \begin{pmatrix} c \cdot a_{11} & \dots & c \cdot a_{1n} \\ \vdots & & \vdots \\ c \cdot a_{m1} & \dots & c \cdot a_{mn} \end{pmatrix} = (c \cdot a_{ij})_{\substack{i=1,\dots,m\\ j=1,\dots,n}}\,.$$

Beweis

Für alle $\vec{x} \in \mathbb{R}^n$ gilt:

$$(F+G)(\vec{x}) \overset{\text{def}}{=} F\,(\vec{x}) + G\,(\vec{x}) = A \circ \vec{x} + B \circ \vec{x}$$

$$= \left(\sum_{j=1}^{n} a_{ij} x_j \right)_{i=1,\ldots,m} + \left(\sum_{j=1}^{n} b_{ij} x_j \right)_{i=1,\ldots,m}$$

$$= \left(\sum_{j=1}^{n} (a_{ij} + b_{ij})\, x_j \right)_{i=1,\ldots,m} = (a_{ij} + b_{ij})_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}} \circ \vec{x};$$

$$(c \cdot F)(\vec{x}) \overset{\text{def}}{=} c \cdot F\,(\vec{x}) = c \cdot (A \circ \vec{x}) = c \cdot \left(\sum_{j=1}^{n} a_{ij} x_j \right)_{i=1,\ldots,m}$$

$$= \left(\sum_{j=1}^{n} (c \cdot a_{ij})\, x_j \right)_{i=1,\ldots,m} = (c \cdot a_{ij})_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}} \circ \vec{x}.$$

Die Behauptung folgt nun aus Satz 3.3.4. □

Der Zusammenhang von $A + B$ und $F + G$ schlägt sich in dem *Distributivgesetz*

$$(A + B) \circ \vec{x} = (F + G)\,(\vec{x}) \overset{\text{def}}{=} F\,(\vec{x}) + G\,(\vec{x}) = A \circ \vec{x} + B \circ \vec{x}$$

nieder, während das schon früher bewiesene Gesetz

$$A \circ (\vec{x} + \vec{y}) = A \circ \vec{x} + A \circ \vec{y}$$

die Linearität (genauer die Additivität) der Abbildung F widerspiegelt.

Ganz analog wie bei den Abbildungen hat man eine *Nullmatrix*

$$\mathcal{O} = \mathcal{O}^{m,n} = \begin{pmatrix} 0 & \ldots & 0 \\ \vdots & & \vdots \\ 0 & \ldots & 0 \end{pmatrix}$$

und zu jeder Matrix $A = A^{m,n}$ eine *negative Matrix*

$$-A := \begin{pmatrix} -a_{11} & \ldots & -a_{1n} \\ \vdots & & \vdots \\ -a_{m1} & \ldots & -a_{mn} \end{pmatrix} = (-a_{ij})_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}},$$

und es gilt in vollkommener Analogie zu dem obigen Satz 3.6.3:

3.6.6 Satz

Mit der Matrixaddition und Skalarmultiplikation ist $\{A : A \text{ ist } m \times n\text{-Matrix}\}$ ein reeller Vektorraum.

Matrizenmultiplikation

Gemäß Satz 3.2.8 ist die Zusammensetzung linearer Abbildungen ebenfalls linear.

3.6.7 Definition

Es sei A eine $m \times n$-Matrix und B eine $\ell \times m$-Matrix. Weiter seien $F : \mathbb{R}^n \to \mathbb{R}^m$ und $G : \mathbb{R}^m \to \mathbb{R}^\ell$ die zugehörigen linearen Abbildungen. Die $\ell \times n$-Matrix C gehöre zur Zusammensetzung $G \circ F : \mathbb{R}^n \to \mathbb{R}^\ell$. Das **Matrixprodukt** *von A und B wird durch $B \circ A := C$ erklärt.*

Häufig verwendet man auch den gewöhnlichen Multiplikationspunkt $C = A \cdot B$ oder läßt diesen ganz weg: $C = A\,B$. Man bemerke, daß nur Matrizen geeigneten Formats multipliziert werden können:

$$(\ell \times \underbrace{m) \circ (m}\times n).$$

Auch hier lassen sich die Komponenten der Produktmatrix leicht aus den Komponenten von A und B berechnen.

3.6.8 Satz

Es seien $A = (a_{ij})_{\substack{i=1,\dots,m\\ j=1,\dots,n}} \in \mathbb{R}^{m\times n}$, $B = (b_{ki})_{\substack{k=1,\dots,\ell\\ i=1,\dots,m}} \in \mathbb{R}^{\ell\times m}$ und $C = B \circ A = (c_{kj})_{\substack{k=1,\dots,\ell\\ j=1,\dots,n}} \in \mathbb{R}^{\ell\times n}$ die Produktmatrix. Dann gilt:

$$c_{kj} = \sum_{i=1}^{m} b_{ki} \cdot a_{ij}; \qquad k = 1,\dots,\ell, \quad j = 1,\dots,n.$$

Beweis

Mit F und G werden wie in der Definition die zu A und B gehörigen linearen Abbildungen bezeichnet. Wir erinnern wieder an Satz 3.3.4.

Für jeden Vektor $\vec{x} \in \mathbb{R}^n$ gilt:

$$\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} := F(\vec{x}) = A \circ \vec{x} = \left(\sum_{j=1}^{n} a_{ij} x_j \right)_{i=1,\dots,m},$$

$$(G \circ F)(\vec{x}) = G(\vec{y}) = B \circ \vec{y} = \left(\sum_{i=1}^{m} b_{ki} y_i \right)_{k=1,\dots,\ell}$$

$$= \left(\sum_{i=1}^{m} b_{ki} \left(\sum_{j=1}^{n} a_{ij} x_j \right) \right)_{k=1,\dots,\ell} = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} (b_{ki} a_{ij} x_j) \right)_{k=1,\dots,\ell}$$

$$= \left(\sum_{j=1}^{n} \left(\sum_{i=1}^{m} b_{ki} a_{ij} \right) x_j \right)_{k=1,\dots,\ell} = \left(\left(\sum_{i=1}^{m} b_{ki} a_{ij} \right) \right)_{\substack{k=1,\dots,\ell\\ j=1,\dots,n}} \circ \vec{x}.$$

$\square$

Dieser Satz zeigt auch, daß das allgemeine Matrizenprodukt direkt mit dem bereits bekannten Produkt von einer Matrix und einem Spaltenvektor zusammenhängt und im Spezialfall $n = 1$ mit diesem zusammenfällt:

Sind $\vec{a}_1, \dots, \vec{a}_n \in \mathbb{R}^m$ die Spalten von A und $\vec{c}_1, \dots, \vec{c}_n \in \mathbb{R}^\ell$ die Spalten von C, so folgt:

$$\vec{c}_j = B \circ \vec{a}_j.$$

Wir stellen einige Gesetze für die Matrizenmultiplikation zusammen.

Zunächst sei an Beispiel 3.3.6 (c) erinnert, wo wir die zur Identität gehörige (quadratische) Einheitsmatrix E eingeführt haben. Somit gilt ganz offensichtlich:

3.6.9 Satz

Sei A eine $m \times n$-Matrix. Dann gilt

$$A \circ E^{n,n} = A, \qquad E^{m,m} \circ A = A.$$

Für Matrizen A und B wie in der Definition ist das Produkt $A \circ B$ auf Grund ihres Formats im allgemeinen gar nicht erklärt. Nur wenn $\ell = n$ gilt, kann man auch $A \circ B$ bilden. Doch auch in diesem Fall ist die Matrizenmultiplikation *nicht kommutativ*:

3.6.10 Beispiel

Für

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \qquad \text{und} \qquad B = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$$

gilt nämlich

$$A \circ B = \begin{pmatrix} 1 & 2+2 \\ 3 & 6+4 \end{pmatrix} = \begin{pmatrix} 1 & 4 \\ 3 & 10 \end{pmatrix},$$

aber

$$B \circ A = \begin{pmatrix} 1+6 & 2+8 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 7 & 10 \\ 3 & 4 \end{pmatrix}.$$

Die Matrizenmultiplikation genügt aber dem *Assoziativgesetz*:

3.6.11 Satz

Sei A eine $m \times n$-Matrix, B eine $\ell \times m$-Matrix und C eine $k \times \ell$-Matrix. Dann gilt:

$$(C \circ B) \circ A = C \circ (B \circ A).$$

Beweis

Es sei $F : \mathbb{R}^n \to \mathbb{R}^m$ bzw. $G : \mathbb{R}^m \to \mathbb{R}^\ell$ bzw. $H : \mathbb{R}^\ell \to \mathbb{R}^k$ die zu A bzw. B bzw. C gehörige lineare Abbildung. Gemäß Definition 3.6.7 gehört

$$\begin{array}{llllll} H \circ G & \text{zu} & C \circ B, & (H \circ G) \circ F & \text{zu} & (C \circ B) \circ A, \\ G \circ F & \text{zu} & B \circ A, & H \circ (G \circ F) & \text{zu} & C \circ (B \circ A). \end{array}$$

Die Behauptung folgt nun aus dem Assoziativgesetz für die Komposition von Abbildungen: siehe Satz 3.1.9. □

Eine besondere Rolle spielen zu Isomorphismen gehörige Matrizen. Da dann Dimension von Urbildraum und Wertevorrat notwendigerweise übereinstimmen, reicht es, im Folgenden nur quadratische Matrizen (d. h. Spaltenzahl = Zeilenzahl) zu betrachten.

3.6.12 Definition

Zu der $n \times n$-Matrix A existiere eine $n \times n$-Matrix B mit $A \circ B = B \circ A = E = E^{n,n}$. Dann heißt A **nichtsingulär** *oder auch* **regulär**. *Man nennt B die zu A* **inverse Matrix** *oder* **Umkehrmatrix** *und schreibt $B = A^{-1}$.*

In Satz 3.1.11 hatten wir entsprechend die Existenz von Umkehrabbildungen charakterisiert. Deshalb folgt unmittelbar:

3.6.13 Folgerung

Die $n \times n$-Matrix A ist nichtsingulär genau dann, wenn die zugehörige lineare Abbildung $F : \mathbb{R}^n \to \mathbb{R}^n$ ein Isomorphismus, d. h. bijektiv ist. In diesem Fall existiert die inverse Matrix A^{-1}, ist die zu F^{-1} gehörige Matrix und ist deshalb auch eindeutig bestimmt.

Ganz analog folgern wir aus Satz 3.1.15 für das Inverse von Produktmatrizen:

3.6.14 Folgerung

Die $n \times n$-Matrizen A, B seien nichtsingulär. Dann ist auch die Produktmatrix nichtsingulär, und für die inverse Matrix gilt:

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

Auf die Frage nach der Existenz und vor allem auf die Berechnung der inversen Matrix gehen wir in Abschnitt 3.8 noch einmal ausführlich ein. An dieser Stelle sollen die folgenden Beispiele genügen:

3.6.15 Beispiele

(a) Es soll untersucht werden, ob die 2×2-Matrix

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

regulär (d. h. nichtsingulär) ist. Auf Grund von Folgerung 3.4.5 reicht es zu zeigen, daß die zugehörige lineare Abbildung injektiv ist. Dazu prüfen wir mittels des Gaußschen Algorithmus', ob das lineare Gleichungssystem $A \circ \vec{x} = \vec{0}$ nur die triviale Lösung $\vec{x} = \vec{0} \in \mathbb{R}^2$ besitzt. Wir gehen dabei gemäß dem Beweis von Satz 3.3.11 vor.

$$\begin{pmatrix} \boxed{1} & 2 \\ 3 & 4 \end{pmatrix} \to \begin{pmatrix} 1 & 2 \\ 0 & \boxed{-2} \end{pmatrix} \to \begin{pmatrix} 1 & 2 \\ 0 & \boxed{1} \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Also gilt:

$$\left\{\vec{x} \in \mathbb{R}^2 : A \circ \vec{x} = \vec{0}\right\} = \left\{\vec{x} \in \mathbb{R}^2 : \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \circ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \vec{0}\right\}$$
$$= \left\{\vec{x} \in \mathbb{R}^2 : x_1 = 0,\ x_2 = 0\right\} = \left\{\vec{0}\right\}.$$

Somit ist A regulär, und A^{-1} existiert. Um die inverse Matrix zu bestimmen, ist also $B = \left(\vec{b}_1, \vec{b}_2\right) = A^{-1}$ derart gesucht, daß insbesondere gilt:

$$A \circ B = E = E^{2,2} \qquad \Leftrightarrow \qquad A \circ \vec{b}_1 = \vec{e}_1 \text{ und } A \circ \vec{b}_2 = \vec{e}_2.$$

Wir lösen beide Gleichungssysteme mit Hilfe des Gaußschen Algorithmus'.

$$\left(\begin{array}{cc|c} \boxed{1} & 2 & 1 \\ 3 & 4 & 0 \end{array}\right) \to \left(\begin{array}{cc|c} 1 & 2 & 1 \\ 0 & \boxed{-2} & -3 \end{array}\right) \to \left(\begin{array}{cc|c} 1 & 2 & 1 \\ 0 & \boxed{1} & \frac{3}{2} \end{array}\right)$$
$$\to \left(\begin{array}{cc|c} 1 & 0 & -2 \\ 0 & 1 & \frac{3}{2} \end{array}\right), \text{ also } \vec{b}_1 = \begin{pmatrix} -2 \\ \frac{3}{2} \end{pmatrix};$$
$$\left(\begin{array}{cc|c} \boxed{1} & 2 & 0 \\ 3 & 4 & 1 \end{array}\right) \to \left(\begin{array}{cc|c} 1 & 2 & 0 \\ 0 & \boxed{-2} & 1 \end{array}\right) \to \left(\begin{array}{cc|c} 1 & 2 & 0 \\ 0 & \boxed{1} & -\frac{1}{2} \end{array}\right)$$
$$\to \left(\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & -\frac{1}{2} \end{array}\right), \text{ also } \vec{b}_2 = \begin{pmatrix} 1 \\ -\frac{1}{2} \end{pmatrix}.$$

Mit

$$B = \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix}$$

gilt neben

$$A \circ B = \begin{pmatrix} -2+3 & 1-1 \\ -6+6 & 3-2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = E$$

bemerkenswerterweise auch

$$B \circ A = \begin{pmatrix} -2+3 & -4+4 \\ \frac{3}{2} - \frac{3}{2} & 3-2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = E,$$

folglich ist $A^{-1} = B$.

Die vorab geklärte Existenz von A^{-1} ist der Grund dafür, daß die Gültigkeit von $A \circ B = E$ schon die Gültigkeit von $B \circ A = E$ impliziert.

(b) Ebenso wollen wir untersuchen, ob

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

regulär ist. Die Lösungsmenge $\left\{\vec{x} \in \mathbb{R}^3 : A \circ \vec{x} = \vec{0}\right\}$, d. h. der Kern der zugehörigen linearen Abbildung, wird durch elementare Umformungen bestimmt.

$$\begin{pmatrix} \boxed{1} & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 \\ 0 & \boxed{-3} & -6 \\ 0 & -6 & -12 \end{pmatrix}$$
$$\to \begin{pmatrix} 1 & 2 & 3 \\ 0 & \boxed{1} & 2 \\ 0 & -6 & -12 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Gemäß Abschnitt 3.5 wird $\left\{\vec{x} \in \mathbb{R}^3 : A \circ \vec{x} = \vec{0}\right\}$ von $(1, -2, 1)$ erzeugt, ist also ein eindimensionaler Untervektorraum von $\mathbb{R}^3$. Die zugehörige Abbildung ist nicht injektiv und insbesondere nicht bijektiv, die Matrix A ist daher nicht nichtsingulär, man nennt A dann *singulär*.

Für quadratische Matrizen A kann man auch Potenzen definieren. Für $p \in \mathbb{N}$ setzt man

$$A^p := \underbrace{A \circ \ldots \circ A}_{p\text{-mal}}.$$

Dabei tritt eine weitere Besonderheit des Matrixprodukts auf. Im Gegensatz zur Multiplikation reeller Zahlen gibt es *Nullteiler*; ein Produkt kann die Null ergeben, obwohl beide Faktoren von Null verschieden sind:

3.6.16 Beispiel

Für $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ gilt $A^2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \circ \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \mathcal{O}^{2,2}$; man erhält die Nullmatrix.

3.6.17 Definition

Man nennt eine $n \times n$-Matrix A **nilpotent**, *wenn eine Zahl $p \in \mathbb{N}$ mit $A^p = \mathcal{O}^{n,n}$ existiert.*

Skalarmultiplikation, Matrizenaddition und -multiplikation

3.6.18 Satz (Distributivgesetze)

Es seien $m \times \ell$-Matrizen A_1, A_2 und $\ell \times n$-Matrizen B_1, B_2 gegeben. Dann gilt:

(a) $(A_1 + A_2) \circ B_1 = A_1 \circ B_1 + A_2 \circ B_1$,

(b) $A_1 \circ (B_1 + B_2) = A_1 \circ B_1 + A_1 \circ B_2$.

Beweis

Seien $F_1, F_2 : \mathbb{R}^\ell \to \mathbb{R}^m$ und $G_1, G_2 : \mathbb{R}^n \to \mathbb{R}^\ell$ die zu A_1, A_2 und B_1, B_2 gehörigen linearen Abbildungen. Zum Nachweis von (a) verwenden wir nur die Definitionen von Summe und Komposition von Abbildungen. Für alle $\vec{x} \in \mathbb{R}^n$ gilt:

$$\begin{aligned}
&((F_1 + F_2) \circ G_1)(\vec{x}) = (F_1 + F_2)(G_1(\vec{x})) \\
&= F_1(G_1(\vec{x})) + F_2(G_1(\vec{x})) = (F_1 \circ G_1)(\vec{x}) + (F_2 \circ G_1)(\vec{x}) \\
&= (F_1 \circ G_1 + F_2 \circ G_1)(\vec{x}).
\end{aligned}$$

Zum Nachweis von (b) verwenden wir dagegen zusätzlich die Linearität der Abbildung F_1. Für alle $\vec{x} \in \mathbb{R}^n$ gilt:

$$\begin{aligned}
&(F_1 \circ (G_1 + G_2))(\vec{x}) = F_1((G_1 + G_2)(\vec{x})) = F_1(G_1(\vec{x}) + G_2(\vec{x})) \\
&= F_1(G_1(\vec{x})) + F_1(G_2(\vec{x})) = (F_1 \circ G_1)(\vec{x}) + (F_1 \circ G_2)(\vec{x}) \\
&= (F_1 \circ G_1 + F_1 \circ G_2)(\vec{x}).
\end{aligned}$$

$\square$

3.6.19 Satz

Sei A eine $m \times \ell$-Matrix und B eine $\ell \times n$-Matrix. Dann gilt für alle reellen Zahlen c:

$$c \cdot (A \circ B) = (c \cdot A) \circ B = A \circ (c \cdot B).$$

Beweis

Seien $F : \mathbb{R}^\ell \to \mathbb{R}^m$ und $G : \mathbb{R}^n \to \mathbb{R}^\ell$ die zugehörigen linearen Abbildungen. Im Folgenden verwenden wir die Linearität der Abbildung F. Für alle $\vec{x} \in \mathbb{R}^n$ gilt:

$$\begin{aligned}
&(c \cdot (F \circ G))(\vec{x}) = c \cdot (F \circ G)(\vec{x}) = c \cdot F(G(\vec{x})) \\
&= (c \cdot F)(G(\vec{x})) = ((c \cdot F) \circ G)(\vec{x}) \\
&= F(c \cdot G(\vec{x})) = F((c \cdot G)(\vec{x})) = (F \circ (c \cdot G))(\vec{x}).
\end{aligned}$$

$\square$

Koordinatenwechsel

Wir erinnern an die Überlegungen am Ende von Abschnitt 3.2: Es seien V_1, V_2 endlichdimensionale Vektorräume, $\dim V_1 = n$, $\dim V_2 = m$, und $H : V_1 \to V_2$ eine lineare Abbildung. Durch Auswahl von Basen $\{x_1, \ldots, x_n\}$ von V_1 und $\{y_1, \ldots, y_m\}$ von V_2 konnte die Abbildung H *in den Koordinaten* konstruiert werden.

Es seien $G_1 : \mathbb{R}^n \to V_1$ und $G_2 : \mathbb{R}^m \to V_2$ die durch $G_1(\vec{e}_j) = x_j$ bzw. $G_2(\vec{e}_k) = y_k$ definierten Isomorphismen (Karten), dann erhalten wir mit

$$F : \mathbb{R}^n \to \mathbb{R}^m, \qquad F := G_2^{-1} \circ H \circ G_1$$

die Abbildung H in den Koordinaten. Die dieser Abbildung F zugehörige $m \times n$-Matrix werde mit A bezeichnet.

$$\begin{tikzcd}
V_1 \arrow[r, "H"] & V_2 \arrow[d, "G_2^{-1}"] \\
\mathbb{R}^n \arrow[u, "G_1"] \arrow[r, "F = H \text{ in den Koordinaten}"'] & \mathbb{R}^m
\end{tikzcd}$$

Genauso gut hätte man andere Basen $\{\hat{x}_1, \ldots, \hat{x}_n\}$ von V_1 und $\{\hat{y}_1, \ldots, \hat{y}_m\}$ von V_2 wählen können, hätte entsprechend durch $\hat{G}_1(\vec{e}_j) = \hat{x}_j$ bzw. $\hat{G}_2(\vec{e}_k) = \hat{y}_k$ andere Karten (Isomorphismen) $\hat{G}_1 : \mathbb{R}^n \to V_1$ und $\hat{G}_2 : \mathbb{R}^m \to V_2$ und mit

$$\hat{F} : \mathbb{R}^n \to \mathbb{R}^m, \qquad \hat{F} := \hat{G}_2^{-1} \circ H \circ \hat{G}_1$$

die Abbildung H bezüglich dieser Koordinaten erhalten. Es bezeichne $\hat{A}$ die zu dieser Abbildung $\hat{F}$ gehörige $m \times n$-Matrix.

$$\begin{tikzcd}
V_1 \arrow[rrr, "H"] & & & V_2 \arrow[dl, "G_2^{-1}"'] \arrow[dd, "\hat{G}_2^{-1}"] \\
& \mathbb{R}^n \arrow[ul, "G_1"'] \arrow[r, "F"] & \mathbb{R}^m \arrow[dr, "\hat{G}_2^{-1} \circ G_2"'] & \\
\mathbb{R}^n \arrow[uu, "\hat{G}_1"] \arrow[ur, "G_1^{-1} \circ \hat{G}_1"'] \arrow[rrr, "\hat{F}"'] & & & \mathbb{R}^m
\end{tikzcd}$$

Koordinatenwechsel

Offensichtlich besteht zwischen F und $\hat{F}$ der Zusammenhang:

$$\begin{aligned}\hat{F} &= \hat{G}_2^{-1} \circ H \circ \hat{G}_1 = \hat{G}_2^{-1} \circ G_2 \circ G_2^{-1} \circ H \circ G_1 \circ G_1^{-1} \circ \hat{G}_1 \\ &= \left(\hat{G}_2^{-1} \circ G_2\right) \circ F \circ \left(G_1^{-1} \circ \hat{G}_1\right).\end{aligned}$$

Wir wollen nun herausfinden, ob die Kenntnis eines Zusammenhangs zwischen den Basen $\{x_1, \dots, x_n\}$ und $\{\hat{x}_1, \dots, \hat{x}_n\}$ auf einfache Weise die Berechnung der zur *Koordinatenwechsel-Abbildung (Koordinatentransformation)* $G_1^{-1} \circ \hat{G}_1$ gehörigen $n \times n$-Matrix gestattet.

Dazu gehen wir davon aus, daß die Darstellung der „Dach"-Basisvektoren in den „ungedachten" Basisvektoren gegeben sei:

$$\hat{x}_j = \sum_{i=1}^{n} b_{ij} x_i, \qquad j = 1, \dots, n,$$

und entsprechend

$$\hat{y}_\ell = \sum_{k=1}^{m} c_{k\ell} y_k, \qquad \ell = 1, \dots, m.$$

Die möglicherweise im Vergleich zur Matrizenmultiplikation zunächst verwirrend wirkende Wahl der *Transformationsmatrizen* $B = (b_{ij})_{i,j=1,\dots,n}$ und $C = (c_{k\ell})_{k,\ell=1,\dots,m}$ ist dadurch begründet, daß man die Komponenten des j-ten „Dach"-vektors bezüglich seiner Darstellung in der „ungedachten" Basis in der j-ten *Spalte* findet.

Für einen beliebigen Vektor $x \in V_1$ seien

$$\begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix} = G_1^{-1}(x) \text{ bzw. } \begin{pmatrix} \hat{p}_1 \\ \vdots \\ \hat{p}_n \end{pmatrix} = \hat{G}_1^{-1}(x)$$

die Koordinaten bezüglich $\{x_1, \dots, x_n\}$ bzw. $\{\hat{x}_1, \dots, \hat{x}_n\}$. Es ist also einerseits $x = \sum_{i=1}^{n} p_i x_i$ und andererseits

$$x = \sum_{j=1}^{n} \hat{p}_j \hat{x}_j = \sum_{j=1}^{n} \hat{p}_j \left(\sum_{i=1}^{n} b_{ij} x_i\right) = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} b_{ij} \hat{p}_j\right) x_i.$$

Da die Darstellung des Vektors x in der Basis $\{x_1, \dots, x_n\}$ eindeutig ist, erhalten wir

$$\left(p_i\right)_{i=1,\dots,n} = \left(\sum_{j=1}^{n} b_{ij} \hat{p}_j\right)_{i=1,\dots,n} = B \circ \left(\hat{p}_i\right)_{i=1,\dots,n}.$$

Wegen

$$\left(p_i\right)_{i=1,\dots,n} = \left(G_1^{-1} \circ \hat{G}_1\right) \left(\hat{p}_i\right)_{i=1,\dots,n}$$

folgt also, daß B die zum Koordinatenwechsel $G_1^{-1} \circ \hat{G}_1$ gehörige Matrix ist.

Entsprechend gehört die Matrix C zum Isomorphismus $G_2^{-1} \circ \hat{G}_2$. Auf Grund von Folgerung 3.6.13 existiert C^{-1} und gehört zum inversen Koordinatenwechsel $\left(G_2^{-1} \circ \hat{G}_2\right)^{-1} = \hat{G}_2^{-1} \circ \left(G_2^{-1}\right)^{-1} = \hat{G}_2^{-1} \circ G_2$.

Damit erhalten wir als Ergebnis für die zu F und $\hat{F}$ gehörigen Matrizen:

$$\hat{A} = C^{-1} \circ A \circ B$$

und durch Multiplikation von links mit C und von rechts mit B^{-1}:

$$A = C \circ \hat{A} \circ B^{-1}.$$

Für die Frage nach der *Berechnung* von inversen Matrizen verweisen wir nochmals auf Abschnitt 3.8.

In Kapitel 7 werden wir für den Fall $V_1 = V_2$ eingehend die Frage studieren, ob man bei gegebener Matrix A eine Transformationsmatrix $B = C$ und damit eine geeignete neue Basis finden kann, so daß die Matrix $\hat{A}$ eine möglichst einfache Gestalt, etwa „Diagonalgestalt“ erhält.

3.7 Ebenen

Wir knüpfen hier an die Überlegungen aus Abschnitt 3.4 an, in dem die Bilder linearer Abbildungen untersucht wurden. Durch „Verschiebung“ dieser Bildräume in einen vorgegebenen Punkt gelangen wir zum allgemeinen Begriff der *Ebene*.

Wie schon teilweise in Kapitel 2 unterscheiden wir hier nicht zwischen Punkten und Vektoren. Wir wollen uns auch nicht wie dort auf eindimensionale Geraden und zweidimensionale Ebenen im dreidimensionalen Raum R beschränken, sondern allgemein die „Geometrie“ s-dimensionaler Ebenen in $\mathbb{R}^m$ studieren. Diese allgemeinen Ebenen sind der einfachste Spezialfall von (im allgemeinen „gekrümmten“) Untermannigfaltigkeiten des $\mathbb{R}^m$, die in der Regel im dritten Teil des Analysis-Kurses behandelt werden.

Parameterdarstellung

3.7.1 Definition

Es sei

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

eine $m \times s$*-Matrix mit* $s \leq m$ *und maximalem Rang; d. h.* $\mathrm{Rang}\,(A) = s$*. Ferner sei* $\vec{c} \in \mathbb{R}^m$ *gegeben. Dann heißt*

$$M := \{\vec{x} \in \mathbb{R}^m : \ \vec{x} = \vec{c} + A \circ \vec{t} \text{ für ein } \vec{t} \in \mathbb{R}^s\}$$

s**-dimensionale Ebene** *durch* $\vec{c}$*.*

Eine **nulldimensionale Ebene** *ist eine einelementige Menge* $\{\vec{c}\}$ *mit* $\vec{c} \in \mathbb{R}^m$*.*

Die Bedingung $\mathrm{Rang}\,(A) = s$ stellt sicher, daß die Spalten von A linear unabhängig sind, sie erzeugen also einen s-dimensionalen Untervektorraum. Die in der Definition gewählte Darstellung heißt *Parameterdarstellung* der Ebene. Ebenen sind also Bilder eines Vektorraums unter einer affin linearen Abbildung, eine affin lineare Abbildung ist die Summe aus einer linearen Abbildung und einem konstanten Vektor.

Wir wollen im Folgenden der Frage nachgehen, ob Ebenen auch auf andere Weise dargestellt werden können.

Ebenen als Lösungsmengen linearer Gleichungssysteme

Sei M eine s-dimensionale Ebene ($0 < s < m$) in $\mathbb{R}^m$ durch $\vec{c}$ wie in Definition 3.7.1. Da Spaltenumformungen die Bildmenge der zu A gehörigen Abbildung nicht ändern, können wir nach geeigneter Umnumerierung der Komponenten von $\vec{c}, \vec{x} \in \mathbb{R}^m$ (entspricht Zeilenvertauschungen von A) annehmen, daß A die folgende Gestalt hat (siehe Abschnitt 3.4):

$$A = \begin{array}{c} \text{Zeile/} \\ \text{Spalte} \\ 1 \\ 2 \\ 3 \\ \\ s \\ s{+}1 \\ \\ m \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & & s \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ * & * & * & \dots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ * & * & * & \dots & * \end{pmatrix} \end{array} .$$

Dann ist die Gültigkeit der Darstellung $\vec{x} = \vec{c} + A \circ \vec{t}$ für ein $\vec{t} \in \mathbb{R}^s$ und damit $\vec{x} \in M$

äquivalent zu

$$\begin{cases} x_1 - c_1 &= t_1 \\ x_2 - c_2 &= t_2 \\ \quad\vdots & \quad\vdots \\ x_s - c_s &= t_s \\ x_{s+1} - c_{s+1} &= a_{s+1,1}t_1 + \ldots + a_{s+1,s}t_s \\ x_{s+2} - c_{s+2} &= a_{s+2,1}t_1 + \ldots + a_{s+2,s}t_s \\ \quad\vdots & \quad\vdots \\ x_m - c_m &= a_{m,1}t_1 + \ldots + a_{m,s}t_s \end{cases} \text{ für ein } \vec{t} \in \mathbb{R}^s,$$

$$\Leftrightarrow \begin{cases} x_{s+1} - c_{s+1} &= a_{s+1,1}(x_1 - c_1) + \ldots + a_{s+1,s}(x_s - c_s) \\ x_{s+2} - c_{s+2} &= a_{s+2,1}(x_1 - c_1) + \ldots + a_{s+2,s}(x_s - c_s) \\ \quad\vdots & \quad\vdots \\ x_m - c_m &= a_{m,1}(x_1 - c_1) + \ldots + a_{m,s}(x_s - c_s). \end{cases}$$

Setzt man nun

$$d_i = -c_i + \sum_{j=1}^{s} a_{ij}c_j \text{ für } i = s+1, \ldots, m,$$

so werden die Punkte $\vec{x}$ der Ebene M genau durch das Gleichungssystem

$$\begin{cases} d_{s+1} &= a_{s+1,1}x_1 + \ldots + a_{s+1,s}x_s - x_{s+1} \\ d_{s+2} &= a_{s+2,1}x_1 + \ldots + a_{s+2,s}x_s - x_{s+2} \\ \vdots & \quad\vdots \\ d_m &= a_{m,1}x_1 + \ldots + a_{m,s}x_s - x_m \end{cases}$$

beschrieben. Die s-dimensionale Ebene M ist also Lösungsmenge eines inhomogenen linearen Gleichungssystems aus $(m-s)$ Gleichungen in m Veränderlichen. Die Koeffizientenmatrix lautet

$$\begin{pmatrix} a_{s+1,1} & \ldots & a_{s+1,s} & -1 & 0 & \ldots & 0 \\ a_{s+2,1} & \ldots & a_{s+2,s} & 0 & -1 & \ldots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & \ldots & a_{m,s} & 0 & 0 & \ldots & -1 \end{pmatrix}$$

und hat den größtmöglichen Rang $m - s$.

An dieser Stelle ordnen sich nun auch nulldimensionale Ebenen $\{\vec{c}\}$ ohne weiteres ein; sie sind genau die eindeutig bestimmten Lösungen des Gleichungssystems

$$-E^{m,m} \circ \vec{x} = -\vec{c}.$$

Zusammenfassend gilt:

3.7.2 Satz

Sei $M \subset \mathbb{R}^m$ eine s-dimensionale Ebene, $s < m$. Dann gibt es eine $(m-s) \times m$-Matrix B mit $\operatorname{Rang}(B) = m - s$ und einen Vektor $\vec{d} \in \mathbb{R}^{m-s}$, so daß $M = \left\{\vec{x} \in \mathbb{R}^m : B \circ \vec{x} = \vec{d}\right\}$ gilt. Die Ebene M ist also die Lösungsmenge eines linearen Gleichungssystems aus $m - s$ Gleichungen in m Veränderlichen, dessen Koeffizientenmatrix maximalen Rang hat.

Ist $M \subset \mathbb{R}^m$ eine m-dimensionale Ebene, so gilt $M = \mathbb{R}^m$.

3.7.3 Beispiele

(a) Im Fall $s = 1$ sprechen wir von *Geraden*, vgl. auch Abschnitt 2.3. Es ist $A = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} =: \vec{a} \neq 0$ und $M = \{\vec{x} \in \mathbb{R}^m : \vec{x} = \vec{c} + t \cdot \vec{a}$ für ein $t \in \mathbb{R}\}$. Geraden werden bei geeigneter Numerierung der Variablen durch die $m - 1$ Gleichungen

$$\begin{aligned} d_2 &= \tilde{a}_2 x_1 - x_2 \\ d_3 &= \tilde{a}_3 x_1 - x_3 \\ \vdots &\qquad \vdots \\ d_m &= \tilde{a}_m x_1 - x_m \end{aligned}$$

beschrieben. Ist zusätzlich $m = 2$, so erhält man die vertraute Darstellung von Geraden in $\mathbb{R}^2$:

$$x_2 = bx_1 + d \qquad \text{mit geeigneten Zahlen} \quad b, d \in \mathbb{R}.$$

(b) Ebenen der Dimension $m - 1$ heißen auch *Hyperebenen*. Sie werden durch eine Gleichung beschrieben:

$$d = b_1 x_1 + \ldots + b_{m-1} x_{m-1} - x_m.$$

Es soll nun umgekehrt untersucht werden, ob durch inhomogene Gleichungssysteme aus $m - r$ Gleichungen in m Veränderlichen stets eine r-dimensionale Ebene gegeben wird.

Wir betrachten das Gleichungssystem $B \circ \vec{x} = \vec{d}$ mit der $(m-r) \times m$-Matrix $B \neq \mathcal{O}$ und $\vec{d} \in \mathbb{R}^{m-r}$ und untersuchen mit Hilfe der in Abschnitt 3.5 bereitgestellten Theorie die Lösungsmenge

$$L := \left\{\vec{x} \in \mathbb{R}^m : \ B \circ \vec{x} = \vec{d}\right\}.$$

Diese Lösungsmenge ist leer genau dann, wenn $\operatorname{Rang}(B) < \operatorname{Rang}\left(B, \vec{d}\right)$ ist, wenn also $\vec{d}$ nicht im Bild der zu B gehörigen linearen Abbildung $\mathbb{R}^m \to \mathbb{R}^{m-r}$ liegt.

Die leere Menge fassen wir nicht als Ebene auf. Diese Möglichkeit läßt sich gemäß Folgerung 3.5.3 jedoch ausschließen, indem man $\mathrm{Rang}\,(B) = m - r$ voraussetzt:

Hat B maximalen Rang, so ist die Lösungsmenge stets nichtleer: $L \neq \emptyset$.

Für das Folgende setzen wir $L \neq \emptyset$ voraus. Dann existiert ein $\vec{c} \in \mathbb{R}^m$ mit $B \circ \vec{c} = \vec{d}$, und die Bestimmung von L reduziert sich auf das Lösen eines homogenen Gleichungssystems:

$$L = \left\{ \vec{x} \in \mathbb{R}^m : \; \vec{x} = \vec{c} + \vec{y} \text{ mit } B \circ \vec{y} = \vec{0} \right\}.$$

Es sei $\mathrm{Rang}\,(B) =: m - s \leq m - r$. Zunächst sei ferner $s \neq 0$, auf den Fall $s = 0$ gehen wir unten kurz ein. Durch elementare Zeilenumformungen und geeigntete Numerierung der Komponenten von $\vec{y}, \vec{x}, \vec{c}$ (entspricht Spaltenvertauschungen von B) gelangt man für die Lösungen von $B \circ \vec{y} = \vec{0}$ zu folgendem Gleichungssystem (siehe Satz 3.3.11, hier stehen die Nullzeilen oben und die Einheitsmatrix rechts in der Matrix, der Rang $m - s$ von B ersetzt dort die Zahl s):

$$\begin{array}{rcl} y_{s+1} &=& -b_{s+1,1}y_1 - \ldots - b_{s+1,s}y_s \\ y_{s+2} &=& -b_{s+2,1}y_1 - \ldots - b_{s+2,s}y_s \\ \vdots &=& \vdots \\ y_m &=& -b_{m,1}y_1 - \ldots - b_{m,s}y_s. \end{array} \tag{$*$}$$

Die Gesamtheit der Lösungen dieses Gleichungssystems erhält man, indem man den Parametervektor $\vec{t} \in \mathbb{R}^s$ einführt und damit den ganzen $\mathbb{R}^s$ durchläuft, $y_1 = t_1, \ldots, y_s = t_s$ setzt und $y_{s+1}, \ldots, y_m$ aus dem Gleichungssystem $(*)$ bestimmt. Mit der $m \times s$-Matrix:

$$A = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \\ -b_{s+1,1} & -b_{s+1,2} & \ldots & -b_{s+1,s} \\ \vdots & \vdots & \ddots & \vdots \\ -b_{m,1} & -b_{m,2} & \ldots & -b_{m,s} \end{pmatrix}$$

ist also $L = \left\{ \vec{x} = \vec{c} + A \circ \vec{t} : \; \vec{t} \in \mathbb{R}^s \right\}$. Die Matrix A hat schließlich s linear unabhängige Spalten, es ist also $\mathrm{Rang}\,(A) = s$. Somit ist L eine s-dimensionale Ebene, dabei ist $s = m - \mathrm{Rang}\,(B) \geq r$.

Nulldimensionale Ebenen erhält man im Fall $s = 0$ und $\mathrm{Rang}\,(B) = m - s = m$, d. h. als Lösungsmenge eines quadratischen Gleichungssystems $B \circ \vec{x} = \vec{d}$ mit der *regulären* Koeffizientenmatrix $B = B^{m,m}$. Den ganzen Raum, d. h. $s = m$, erhält man mittels der Nullmatrix $B = \mathcal{O}$ und $\vec{d} = \vec{0}$, bzw. indem man keine Gleichungen an $\vec{x}$ verlangt. $\square$

Ist der Rang von B maximal, so lassen sich unsere Ergebnisse wie folgt zusammenfassen:

3.7.4 Satz

Sei B *eine* $(m-s)\times m$*-Matrix mit* $\operatorname{Rang}(B)=m-s$. *Dann ist die Lösungsmenge* $L := \left\{\vec{x}\in\mathbb{R}^m : B\circ\vec{x}=\vec{d}\right\}$ *für jedes* $\vec{d}\in\mathbb{R}^{m-s}$ *nichtleer und beschreibt eine* s*-dimensionale Ebene.*

Ebenen lassen sich also gleichwertig mit Hilfe von Parameterdarstellungen und als Lösungsmengen linearer Gleichungssysteme beschreiben.

3.7.5 Beispiele

(a) Wir betrachten das Gleichungssystem

$$\begin{aligned} x_1+2x_2&=1\\ x_1+x_2&=2, \end{aligned} \qquad \text{d. h. } B\circ\vec{x}=\vec{d} \text{ mit } B=\begin{pmatrix}1&2\\1&1\end{pmatrix} \text{ und } \vec{d}=\begin{pmatrix}1\\2\end{pmatrix}.$$

Zur Bestimmung der Lösungsmenge formen wir die erweiterte Koeffizientenmatrix $\left(B,\vec{d}\right)$ gemäß Abschnitt 3.5 um, hier allerdings mit dem Ziel, den Einheitsblock rechts unten in der Matrix zu erhalten:

$$\left(\begin{array}{cc|c}1&2&1\\1&\boxed{1}&2\end{array}\right)\to\left(\begin{array}{cc|c}\boxed{-1}&0&-3\\1&1&2\end{array}\right)\to\left(\begin{array}{cc|c}\boxed{1}&0&3\\1&1&2\end{array}\right)\to\left(\begin{array}{cc|c}1&0&3\\0&1&-1\end{array}\right).$$

Dieses Gleichungssystem hat die eindeutig bestimmte Lösung $\vec{x}=(x_1,x_2)=(3,-1)$ und beschreibt einen Punkt in $\mathbb{R}^2$; es ist $\operatorname{Rang}(B)=2$.

(b) Sei nun

$$B=\begin{pmatrix}1&2&0&-1\\2&-1&0&1\\4&3&0&-1\end{pmatrix} \text{ und } \vec{d}=\begin{pmatrix}3\\1\\7\end{pmatrix}.$$

Mit den Bezeichnungen von S. 93 haben wir also $m=4$, $r=1$. Wir betrachten wieder die erweiterte Koeffizientenmatrix und erhalten auf Grund von

$$\left(\begin{array}{cccc|c}1&2&0&-1&3\\2&-1&0&1&1\\4&3&0&\boxed{-1}&7\end{array}\right)\to\left(\begin{array}{cccc|c}1&2&0&-1&3\\2&-1&0&1&1\\-4&-3&0&\boxed{1}&-7\end{array}\right)$$

$$\to\left(\begin{array}{cccc|c}-3&-1&0&0&-4\\6&\boxed{2}&0&0&8\\-4&-3&0&1&-7\end{array}\right)\to\left(\begin{array}{cccc|c}-3&-1&0&0&-4\\3&\boxed{1}&0&0&4\\-4&-3&0&1&-7\end{array}\right)$$

$$\to\left(\begin{array}{cccc|c}0&0&0&0&0\\3&1&0&0&4\\5&0&0&1&5\end{array}\right),$$

daß $\operatorname{Rang}(B) = 2 = m - 2$ ist und daß das Gleichungssystem $B \circ \vec{x} = \vec{d}$ äquivalent zu

$$\begin{aligned} x_2 &= 4 - 3x_1, \\ x_4 &= 5 - 5x_1 \end{aligned}$$

ist. Indem man $x_1 = t_1$, $x_3 = t_2$ setzt, erhält man alle Lösungen durch

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} t_1 \\ 4 - 3t_1 \\ t_2 \\ 5 - 5t_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 4 \\ 0 \\ 5 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ -3 & 0 \\ 0 & 1 \\ -5 & 0 \end{pmatrix} \circ \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}.$$

Also ist $\left\{\vec{x} : B \circ \vec{x} = \vec{d}\right\}$ eine zweidimensionale Ebene in $\mathbb{R}^4$.

Lineare Ausdehnung von Ebenen

Im folgenden Satz wird gezeigt, daß jede (allgemeine) Ebene in eine Ebene vorgegebener höherer Dimension eingebettet werden kann.

3.7.6 Satz

Es sei $M \subset \mathbb{R}^m$ eine s-dimensionale Ebene ($0 \leq s \leq m$). Weiter sei eine natürliche Zahl r mit $s \leq r \leq m$ gegeben. Dann gibt es eine r-dimensionale Ebene $N \subset \mathbb{R}^m$ mit $M \subset N$. Gilt weiter $s < r < m$, so ist N nicht *eindeutig bestimmt.*

Beweis

Ist $r = s$, so setzen wir $N = M$, während im Falle $r = m$ notwendigerweise $N = \mathbb{R}^m$ gilt. Wir können uns also auf den Fall $s < r < m$ beschränken.

Wir zeigen zunächst die Existenz einer derartigen Ebene N:

Die gegebene Ebene M wird definitionsgemäß mit Hilfe einer $m \times s$-Matrix A mit $\operatorname{Rang}(A) = s$ und eines „Aufpunktes" $\vec{c} \in \mathbb{R}^m$ beschrieben:

$$M = \left\{\vec{x} = \vec{c} + A \circ \vec{t} : \vec{t} \in \mathbb{R}^s\right\}.$$

Wie im Anschluß an Definition 3.7.1 können wir nach Umnumerierung der Komponenten von $\vec{c}, \vec{x} \in \mathbb{R}^m$ annehmen, daß A die folgende Gestalt hat:

$$A = \begin{pmatrix} E^{s,s} \\ \tilde{A}^{m-s,s} \end{pmatrix}.$$

Zur Konstruktion der Ebene N ergänzen wir die $m \times s$-Matrix A in der folgenden

Weise zu einer $m \times r$-Matrix B:

$$B = \left(\begin{array}{c|c} E^{s,s} & \mathcal{O}^{s,r-s} \\ \hline \tilde{A}^{m-s,s} & \begin{array}{c} E^{r-s,r-s} \\ \hline \tilde{B}^{m-r,r-s} \end{array} \end{array}\right),$$

die Teilmatrix $\tilde{B}$ kann dabei beliebig gewählt werden. Die Betrachtung der ersten r Zeilen liefert $\operatorname{Rang}(B) = r$, also ist

$$N := \{\vec{x} = \vec{c} + B \circ \vec{\tau} : \ \vec{\tau} \in \mathbb{R}^r\}$$

eine r-dimensionale Ebene in $\mathbb{R}^m$.

Wir zeigen noch, daß diese Ebene auch tatsächlich M umfaßt: Sei $\vec{x} \in M$, dann gibt es ein $\vec{t} \in \mathbb{R}^s$ mit $\vec{x} = \vec{c} + A \circ \vec{t}$. Für $\vec{\tau} = (t_1, \ldots, t_s, 0, \ldots, 0)^t$ (transponiert man eine $1 \times r$-Matrix, d. h. eine Zeile, so erhält man eine $r \times 1$-Matrix, d. h. eine Spalte) gilt:

$$B \circ \vec{\tau} = t_1 \cdot \vec{b}_1 + \ldots + t_s \cdot \vec{b}_s = A \circ \vec{t},$$

wobei $\vec{b}_j$ die j-te Spalte von B bezeichnet. Also ist $\vec{x} \in N$.

Wir zeigen nun, daß für $s < r < m$ die soeben konstruierte Ebene N *nicht* eindeutig bestimmt ist:

Die $(m-r) \times (r-s)$-Teilmatrix $\tilde{B}$ von B tritt tatsächlich auf. Wir betrachten eine einparametrige Familie derartiger Teilmatrizen

$$\tilde{B}(a) = \begin{pmatrix} 0 & \ldots & 0 \\ \vdots & & \vdots \\ 0 & \ldots & a \end{pmatrix}, \quad a \in \mathbb{R},$$

und die entsprechenden Familien $B(a)$ von $m \times r$-Matrizen und von Ebenen $N(a)$. Wir wollen beweisen, daß verschiedene Werte des Parameters a auch verschiedene Ebenen $N(a)$ ergeben.

Spaltenumformungen von $B(a)$ ändern $N(a)$ als Bildmenge einer affin-linearen Abbildung nicht; wir können also annehmen, daß

$$B = B(a) = \left(\begin{array}{c} E^{r,r} \\ \hline * \ \big| \tilde{B}(a) \end{array}\right)$$

gilt. Man bemerke, daß die nötigen Spaltenumformungen die $(r-s)$ Spalten rechts in der Matrix, also insbesondere $\tilde{B}(a)$, nicht ändern. Wir konstruieren ein Element aus $N(a)$, das in keiner weiteren Ebene $N(a')$ enthalten ist. Für $\vec{x}(a) := \vec{c} + B(a) \circ \vec{e}_r \in$

$N(a)$ gilt

$$x_j(a) = \begin{cases} c_j & \text{, falls } j \neq r,\ j \neq m, \\ c_j + 1 & \text{, falls } j = r, \\ c_j + a & \text{, falls } j = m. \end{cases}$$

Tatsächlich gilt $\vec{x}(a) \notin N(a')$, falls $a \neq a'$. Sonst gäbe es nämlich ein $\vec{\tau} \in \mathbb{R}^r$ mit $\vec{x}(a) = \vec{c} + B(a') \circ \vec{\tau}$. Ein Vergleich der beiden Darstellungen von $\vec{x}(a)$ ergäbe $c_j + \tau_j = c_j$ für $j = 1, \ldots, r-1$ und $c_r + \tau_r = c_r + 1$, also $\vec{\tau} = \vec{e}_r$. Man erhielte $x_m(a) = c_m + a = c_m + a'$, also $a = a'$ und damit einen Widerspruch!

Somit gilt $N(a) \neq N(a')$ für $a \neq a'$; es gibt also eine ganze Schar von Ebenen mit der gewünschten Eigenschaft. □

Dieser Beweis läßt sich in dem Fall veranschaulichen, daß M eine (eindimensionale) Gerade und N eine (zweidimensionale) Ebene in $\mathbb{R}^3$ sind. Wir haben angenommen, daß der Richtungsvektor von M nicht in der x_2-x_3-Ebene liegt und ergänzen M durch einen zweiten Richtungsvektor $(0, 1, a)^t$ aus der x_2-x_3-Ebene zur Ebene $N = N(a)$.

Ebenen durch gegebene Punkte

In dem der anschaulichen Geometrie gewidmeten Abschnitt 2.4 haben wir uns mit der Existenz von Geraden in der anschaulichen Ebene durch vorgegebene Punkte beschäftigt. Wir übersetzen diese Frage in die Sprache der Vektorraumtheorie und betrachten deren Verallgemeinerung auf beliebige Dimensionen.

3.7.7 Satz

Es seien $(r+1)$ Punkte $\vec{x}_0, \ldots, \vec{x}_r \in \mathbb{R}^m$ gegeben, dabei sei $r \leq m$. Dann existiert eine r-dimensionale Ebene M durch $\vec{x}_0, \ldots, \vec{x}_r$.

Beweis

Im Fall $r = 0$ ist die Lösung durch $M = \{\vec{x}_0\}$ gegeben. Für das Folgende können wir also $r > 0$ annehmen.

Sei $A := (\vec{x}_1 - \vec{x}_0, \vec{x}_2 - \vec{x}_0, \ldots, \vec{x}_r - \vec{x}_0)$ die $m \times r$-Matrix mit den Spalten $\vec{x}_j - \vec{x}_0$. Sei $s := \mathrm{Rang}\,(A)$, es ist $s \leq r$. Man erkennt, daß die Menge

$$\tilde{M} := \left\{\vec{x} = \vec{x}_0 + A \circ \vec{t} :\ \vec{t} \in \mathbb{R}^r\right\}$$

die vorgegebenen Punkte enthält, indem man nacheinander $\vec{t} = \vec{0}, \vec{e}_1, \ldots, \vec{e}_r$ betrachtet.

Wir zeigen nun: $\tilde{M}$ ist eine s-dimensionale Ebene.

Beweis dazu: Da $\tilde{M}$ durch Spaltenumformungen von A nicht geändert wird, kann man nach Umnumerieren der Komponenten der Elemente von $\mathbb{R}^m$ (Zeilenvertauschungen von A) annehmen, daß A die folgende Gestalt hat:

$$A = \left(\begin{array}{c|c} E^{s,s} & \\ \hline & \mathcal{O}^{m,r-s} \\ \tilde{A}^{m-s,s} & \end{array} \right) = (B|\,\mathcal{O})\,.$$

Dabei ist B die $m \times s$-Matrix, die aus den s linken Spalten von A besteht. Es ist $\mathrm{Rang}\,(B) = \mathrm{Rang}\,(A) = s$. Für $\vec{t} = (t_1, \ldots, t_r)^t \in \mathbb{R}^r$ bezeichne $\vec{\tau} = (t_1, \ldots, t_s)^t \in \mathbb{R}^s$, es gilt

$$A \circ \vec{t} = (B|\,\mathcal{O}) \circ \vec{t} = B \circ \vec{\tau}$$

und damit $\tilde{M} = \{\vec{x} = \vec{x}_0 + B \circ \vec{\tau} :\ \vec{\tau} \in \mathbb{R}^s\}$.

Damit haben wir gezeigt, daß $\tilde{M}$ tatsächlich eine s-dimensionale Ebene ist. Im allgemeinen ist jedoch $s \leq r$. Nun stellt allerdings Satz 3.7.6 sicher, daß eine r-dimensionale Ebene $M \subset \mathbb{R}^m$ mit $M \supset \tilde{M}$ existiert. Diese enthält insbesondere die vorgegebenen Punkte $\vec{x}_0, \ldots, \vec{x}_r$. □

Der Beweis von Satz 3.7.6 zeigt, daß unter Umständen eine ganze Schar von Ebenen das in Satz 3.7.7 formulierte Problem löst. Liegen z. B. drei Punkte auf einer Geraden, so läßt sich der zweite Richtungsvektor der gesuchten Ebene frei wählen.

Einer ganz ähnlichen Situation sind wir in Abschnitt 3.3 begegnet: Ein System aus s Vektoren erzeugt einen s-dimensionalen Untervektorraum dann und nur dann, wenn diese Vektoren linear unabhängig sind.

Entsprechend soll ein Begriff eingeführt werden, der gewährleistet, daß eine s-dimensionale Ebene durch Vorgabe von $(s+1)$ geeigneten Punkten eindeutig festgelegt ist.

3.7.8 Definition

Wir sagen, daß $(s+1)$ Punkte $\vec{x}_0, \ldots, \vec{x}_s \in \mathbb{R}^m \ (s \leq m)$, in **allgemeiner Lage** *sind, wenn die Vektoren*

$$\begin{pmatrix} \vec{x}_0 \\ 1 \end{pmatrix}, \ldots, \begin{pmatrix} \vec{x}_s \\ 1 \end{pmatrix} \in \mathbb{R}^{m+1}$$

linear unabhängig sind.

Es sei noch einmal darin erinnert, daß wir zwischen Punkten und Vektoren nicht mehr unterscheiden.

3.7.9 Bemerkung

Der Begriff „allgemeine Lage“ erfährt seine Berechtigung durch folgende Überlegung, dabei sei $s \leq m$: Mit Hilfe von topologischen und maßtheoretischen Begriffen und Methoden, die in der Regel im dritten Semester besprochen werden, kann man zeigen, daß die Menge aller $(s+1)$-tupel von Punkten, die sich in allgemeiner Lage befinden, den Raum sämtlicher $(s+1)$-tupel von Punkten „fast“ vollständig ausfüllt. Könnte man wirklich „zufällig“ jeweils $(s+1)$ Punkte aus $\mathbb{R}^m$ auswählen, so wäre die „Wahrscheinlichkeit“, daß diese in allgemeiner Lage sind, gleich 1.

3.7.10 Satz

Die Punkte $\vec{x}_0, \dots, \vec{x}_s \in \mathbb{R}^m$ $(s \le m)$ sind genau dann in allgemeiner Lage, wenn die Vektoren $\vec{x}_1 - \vec{x}_0, \dots, \vec{x}_s - \vec{x}_0 \in \mathbb{R}^m$ linear unabhängig sind.

Beweis

Definitionsgemäß sind $\vec{x}_0, \dots, \vec{x}_s \in \mathbb{R}^m$ genau dann in allgemeiner Lage, wenn die $(m+1) \times (s+1)$-Matrix

$$\begin{pmatrix} \vec{x}_0 & \dots & \vec{x}_s \\ 1 & \dots & 1 \end{pmatrix}$$

maximalen Rang, d. h. Rang $s+1$ hat. Da elementare Umformungen den Rang nicht ändern, ist das äquivalent zu

$$\operatorname{Rang} \begin{pmatrix} \vec{x}_0 & \vec{x}_1 - \vec{x}_0 & \dots & \vec{x}_s - \vec{x}_0 \\ 1 & 0 & \dots & 0 \end{pmatrix} = s+1.$$

(Die erste Spalte wurde von den übrigen subtrahiert.)

$$\Leftrightarrow \operatorname{Rang} \begin{pmatrix} \vec{0} & \vec{x}_1 - \vec{x}_0 & \dots & \vec{x}_s - \vec{x}_0 \\ 1 & 0 & \dots & 0 \end{pmatrix} = s+1.$$

(Das $-x_{0,j}$-fache der letzten Zeile wurde zur j-ten Zeile addiert.)

$$\Leftrightarrow (\vec{x}_1 - \vec{x}_0), \dots, (\vec{x}_s - \vec{x}_0) \quad \text{sind linear unabhängig.}$$

Wir zeigen noch die Gültigkeit der letzten Äquivalenz:

„$\Rightarrow$“: Insbesondere sind $\begin{pmatrix} \vec{x}_1 - \vec{x}_0 \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} \vec{x}_s - \vec{x}_0 \\ 0 \end{pmatrix}$ linear unabhängig. Wegen

$$\sum_{j=1}^{s} a_j \begin{pmatrix} \vec{x}_j - \vec{x}_0 \\ 0 \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{s} a_j (\vec{x}_j - \vec{x}_0) \\ 0 \end{pmatrix}$$

$(a_j \in \mathbb{R}$ beliebig) sind $(\vec{x}_1 - \vec{x}_0), \dots, (\vec{x}_s - \vec{x}_0)$ linear unabhängig.

„$\Leftarrow$“: Seien $a_0, \dots, a_s \in \mathbb{R}$ mit

$$\vec{0} = a_0 \begin{pmatrix} \vec{0} \\ 1 \end{pmatrix} + \sum_{j=1}^{s} a_j \begin{pmatrix} \vec{x}_j - \vec{x}_0 \\ 0 \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{s} a_j (\vec{x}_j - \vec{x}_0) \\ a_0 \end{pmatrix}.$$

Die Betrachtung der letzten Komponente liefert $a_0 = 0$, aus der linearen Unabhängigkeit von $(\vec{x}_1 - \vec{x}_0), \dots, (\vec{x}_s - \vec{x}_0)$ folgt $a_1 = \dots = a_s = 0$. □

Man hätte auch Satz 3.7.10 zur Definition erheben können. Die hier gewählte Definition hat aber den Vorteil, daß im Gegensatz dazu die Formulierung in $\vec{x}_0, \dots, \vec{x}_s$ symmetrisch ist; kein Vektor ist ausgezeichnet.

3.7.11 Satz

Für $s \leq m$ seien $(s+1)$ Punkte $\vec{x}_0, \dots, \vec{x}_s \in \mathbb{R}^m$ in allgemeiner Lage gegeben. Dann existiert genau eine *s-dimensionale Ebene M durch diese Punkte.*

Die Ebene M ist dabei in dem folgenden Sinne minimal: Jede Ebene $\tilde{M} \subset \mathbb{R}^m$, die $\vec{x}_0, \dots, \vec{x}_s$ enthält, umfaßt auch die Ebene M: $\tilde{M} \supset M$.

Insbesondere können sich in einer s-dimensionalen Ebene jeweils höchstens $s+1$ Punkte in allgemeiner Lage befinden. Zweidimensionale Ebenen (Tischplatten) können nicht vier Punkte (Tischbeine) in allgemeiner Lage enthalten. Zusammen mit der Bemerkung über den Hintergrund des Begriffs „allgemeine Lage“ bestätigt das die Erfahrung, daß Tische „im allgemeinen“ wackeln.

Beweis

Wegen Satz 3.7.7 reicht es für den Beweis des ersten Teils zu zeigen, daß *höchstens* eine derartige Ebene existiert. Dazu nimmt man ein derartiges M als gegeben an und zeigt, daß M eine durch $\vec{x}_0, \dots, \vec{x}_s$ bestimmte Gestalt hat.

Sei also $M = \{\vec{x} = \vec{c} + A \circ \vec{t} : \ \vec{t} \in \mathbb{R}^s\}$ eine s-dimensionale Ebene durch $\vec{x}_0, \dots, \vec{x}_s$, es ist also $\operatorname{Rang}(A) = s$. Da $\vec{x}_0$ in M enthalten ist, gibt es ein $\vec{t}_0 \in \mathbb{R}^s$ mit $\vec{x}_0 = \vec{c} + A \circ \vec{t}_0$. Für beliebiges $\vec{t}$ folgt $\vec{c} + A \circ \vec{t} = \vec{c} + A \circ (\vec{t} - \vec{t}_0) + A \circ \vec{t}_0 = \vec{x}_0 + A \circ (\vec{t} - \vec{t}_0)$. Wir können also ohne Einschränkung annehmen, daß $M = \{\vec{x} = \vec{x}_0 + A \circ \vec{t} : \ \vec{t} \in \mathbb{R}^s\}$ gilt. Wir setzen $\vec{y}_i = \vec{x}_i - \vec{x}_0$, $(i = 1, \dots, s)$ und $V := \mathbb{R} \cdot (\vec{y}_1, \dots, \vec{y}_s)$. Da $\vec{x}_0, \dots, \vec{x}_s$ in allgemeiner Lage sind, ist $\dim V = s$.

Es gilt: $M = \{\vec{x} \in \mathbb{R}^m : \ \vec{x} - \vec{x}_0 \in V\}$.

Beweis dazu: Es sei $F : \mathbb{R}^s \to \mathbb{R}^m$ die zu A gehörige lineare Abbildung, wegen $s = \operatorname{Rang}(A) = \operatorname{rg}(F)$ ist $\operatorname{im}(F) = \{A \circ \vec{t} : \ \vec{t} \in \mathbb{R}^s\}$ ein s-dimensionaler Untervektorraum des $\mathbb{R}^m$. Wegen $\vec{x}_i \in M$ gilt $\vec{y}_i \in \operatorname{im}(F)$, also $V \subset \operatorname{im}(F)$. Da $\operatorname{im}(F)$ und V dieselbe Dimension haben, folgt $\operatorname{im}(F) = V$. Offensichtlich ist aber $M = \{\vec{x} \in \mathbb{R}^m : \ \vec{x} - \vec{x}_0 \in \operatorname{im}(F)\}$.

Wir zeigen noch den Zusatz über die Minimalität der Ebene M:

Sei $\tilde{M}$ eine Ebene in $\mathbb{R}^m$ mit $\vec{x}_0, \dots, \vec{x}_s \in \tilde{M}$. Wie vorher zeigt man, daß $\tilde{V} := \left\{\vec{x} - \vec{x}_0 : \ \vec{x} \in \tilde{M}\right\}$ ein Untervektorraum von $\mathbb{R}^m$ ist. Gemäß Voraussetzung enthält $\tilde{V}$ die Vektoren $\vec{y}_i = \vec{x}_i - \vec{x}_0$, $i = 1, \dots, s$. Damit ist $\tilde{V} \supset V$ und $\tilde{M} \supset M$. □

3.7.12 Beispiel

Wir wollen prüfen, ob die Punkte $\vec{x}_0 = (1, 0, 2)$, $\vec{x}_1 = (0, -3, 2)$, $\vec{x}_2 = (-1, 1, 0) \in \mathbb{R}^3$ in allgemeiner Lage sind. Gemäß Definition 3.7.8 bestimmen wir den Rang der folgenden Matrix:

$$\operatorname{Rang}\begin{pmatrix} \boxed{1} & 0 & -1 \\ 0 & -3 & 1 \\ 2 & 2 & 0 \\ 1 & 1 & 1 \end{pmatrix} = \operatorname{Rang}\begin{pmatrix} 1 & 0 & -1 \\ 0 & -3 & 1 \\ 0 & 2 & 2 \\ 0 & \boxed{1} & 2 \end{pmatrix} = \operatorname{Rang}\begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 7 \\ 0 & 0 & \boxed{-2} \\ 0 & 1 & 2 \end{pmatrix}$$
$$= \operatorname{Rang}\begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 7 \\ 0 & 0 & \boxed{1} \\ 0 & 1 & 2 \end{pmatrix} = \operatorname{Rang}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} = \operatorname{Rang}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} = 3.$$

Diese Matrix hat den größtmöglichen Rang, nämlich 3. Damit sind $\vec{x}_0$, $\vec{x}_1$, $\vec{x}_2$ in allgemeiner Lage, und es gibt genau eine (zweidimensionale) Ebene durch diese Punkte.

Schnitt zweier Ebenen

Für die als nicht leer angenommene Schnittebene zweier (allgemeiner) Ebenen wird sich herausstellen, daß deren Dimension auch von der Dimension des umgebenden Raumes abhängt: Zwei sich schneidende zweidimensionale Ebenen haben in $\mathbb{R}^3$ mindestens eine Gerade, in $\mathbb{R}^4$ jedoch nur mindestens einen Punkt gemeinsam.

Deshalb führen wir mit der Codimension eine Maßzahl dafür ein, wieviele freie Richtungen außerhalb einer Ebene im einbettenden Raum existieren.

3.7.13 Definition

Für eine s-dimensionale Ebene $M \subset \mathbb{R}^m$ heißt codim $M := m - s$ *die* **Codimension** *von M.*

3.7.14 Satz

Ist für Ebenen M_1 und M_2 in $\mathbb{R}^m$ der Durchschnitt nicht leer, dann schneiden sie sich in einer Ebene $M := M_1 \cap M_2$, deren Codimension der Abschätzung

$$\operatorname{codim} M \leq \operatorname{codim} M_1 + \operatorname{codim} M_2$$

genügt.

Beweis

Bei der Untersuchung von Schnittgebilden von Ebenen ist es naheliegend, deren Beschreibung als Lösungsmengen von Gleichungssystemen zugrundezulegen. Es seien s_1 und s_2 die Dimensionen von M_1 und M_2. Wir können ohne Einschränkung $0 < s_1, s_2 < m$ annehmen. Dann existieren $(m - s_i) \times m$-Matrizen B_i mit $\operatorname{Rang}(B_i) = m - s_i$ und Vektoren $\vec{d_i} \in \mathbb{R}^{m-s_i}$, so daß gilt:

$$M_i = \left\{ \vec{x} \in \mathbb{R}^m : \, B_i \circ \vec{x} = \vec{d_i} \right\}, \qquad i = 1, 2.$$

Wir führen die $(m-s_1+m-s_2)\times m$-Matrix

$$B=\begin{pmatrix}B_1\\B_2\end{pmatrix}\quad\text{und den Vektor}\quad\vec{d}=\begin{pmatrix}\vec{d}_1\\\vec{d}_2\end{pmatrix}\in\mathbb{R}^{2m-s_1-s_2}$$

ein, es sei $m-s$ der Rang von B. Wegen $\vec{x}\in M_1\cap M_2\Leftrightarrow(\vec{x}\in M_1$ und $\vec{x}\in M_2)\Leftrightarrow$ $(B_1\circ\vec{x}=\vec{d}_1$ und $B_2\circ\vec{x}=\vec{d}_2)$ und wegen

$$\begin{pmatrix}B_1\\B_2\end{pmatrix}\circ\vec{x}=\begin{pmatrix}B_1\circ\vec{x}\\B_2\circ\vec{x}\end{pmatrix}\in\mathbb{R}^{2m-s_1-s_2}$$

gilt:

$$M=\left\{\vec{x}\in\mathbb{R}^m:\ B\circ\vec{x}=\vec{d}\right\}.$$

Da voraussetzungsgemäß $M\neq\emptyset$ ist, ist folglich M eine s-dimensionale Ebene. Die offensichtliche Ungleichung Zeilenrang$(B)\leq$ Zeilenrang$(B_1)+$ Zeilenrang(B_2) ergibt die Abschätzung für die Codimensionen

$$m-s\leq(m-s_1)+(m-s_2).$$ □

Dem Beweis entnimmt man auch, daß $M_1\cap M_2\neq\emptyset$ folgt und daß man deshalb auf diese Voraussetzung verzichten kann, falls der $\mathrm{Rang}\,(B)$ gleich der Zeilenzahl von B ist, d. h. falls $\mathrm{Rang}\,(B)=(m-s_1)+(m-s_2)$. Wir widmen die folgende Definition dieser Situation, die sich als „typisch" herausstellen wird, solange die Gesamtzahl $2m-(s_1+s_2)$ der Gleichungen die Zahl m der Veränderlichen nicht übersteigt.

3.7.15 Definition

Die s_i-dimensionalen Ebenen $M_i=\left\{\vec{x}\in\mathbb{R}^m:\ B_i\circ\vec{x}=\vec{d}_i\right\}$ seien mittels der $(m-s_i)\times m$-Matrizen B_i gegeben, $i=1,2$.

Man sagt, daß die Ebenen M_1 und M_2 in **allgemeiner Lage** *sind, wenn der Rang $m-s$ der $(m-s_1+m-s_2)\times m$-Matrix*

$$B=\begin{pmatrix}B_1\\B_2\end{pmatrix}$$

gleich der Zeilenzahl von B ist, d. h., wenn die Gleichung

$$m-s=\operatorname{codim}M_1+\operatorname{codim}M_2$$

gilt.

Die allgemeine Lage von Ebenen ist nicht mit der vorher definierten allgemeinen Lage von Punkten zu verwechseln.

3.7.16 Bemerkung

Sind M_1, M_2 in allgemeiner Lage, so gilt insbesondere $(m - s_1) + (m - s_2) = \operatorname{codim} M_1 + \operatorname{codim} M_2 = m - s \leq m$, also

$$s_1 + s_2 \geq m.$$

So sind etwa zwei Geraden in $\mathbb{R}^3$ niemals in allgemeiner Lage.

Unsere bisherigen Überlegungen lassen sich unter Verwendung des neuen Begriffs zusammenfassen:

3.7.17 Satz

Genau dann sind die Ebenen M_1 und M_2 in allgemeiner Lage, wenn deren Durchschnitt $M := M_1 \cap M_2$ eine Ebene mit $\operatorname{codim} M = \operatorname{codim} M_1 + \operatorname{codim} M_2$ ist.

Das Schnittgebilde von Ebenen in allgemeiner Lage hat größtmögliche Codimension und damit kleinstmögliche Dimension $s = s_1 + s_2 - m$.

3.7.18 Bemerkung

Betrachtet man nun Ebenen M_1, M_2 in $\mathbb{R}^m$ mit $\operatorname{codim} M_1 + \operatorname{codim} M_2 \leq m$, d. h. $m \leq s_1 + s_2$, dann kann man entsprechend der Bemerkung 3.7.9 mit den Methoden der Analysis zeigen, daß „fast alle“ Paare von Ebenen in allgemeiner Lage sind. Insbesondere kann man allgemeine Lage bei Ebenen, die selbst nicht in allgemeiner Lage sind, durch beliebig kleine „Drehungen“ einer Ebene erreichen. Für die präzise Definition und Untersuchung von Drehungen müssen wir auf die Kapitel 6 und 7 verweisen.

3.7.19 Beispiel

Wir betrachten $M_i := \{(x_1, x_2) \in \mathbb{R}^2 : x_2 = a_i \cdot x_1\}$ mit $a_i \in \mathbb{R}$, $i = 1, 2$. Mit den 1×2-Matrizen $B_i = (\,-a_i \quad 1\,)$, $\operatorname{Rang}(B_i) = 1$, gilt

$$M_i = \left\{\vec{x} \in \mathbb{R}^2 : B_i \circ \vec{x} = 0 \in \mathbb{R}\right\}.$$

Also sind die M_i eindimensionale Ebenen, d. h. Geraden in $\mathbb{R}^2$. Es soll der Durchschnitt

$$M := M_1 \cap M_2 = \left\{(x_1, x_2) \in \mathbb{R}^2 : x_2 = a_1 x_1 = a_2 x_1\right\}$$

untersucht werden.

Sicher ist $\vec{0} \in M$, M ist nach Satz 3.7.14 eine (allgemeine) Ebene. Zur Bestimmung der Dimension von M ist die zusammengesetzte Matrix

$$B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} = \begin{pmatrix} -a_1 & 1 \\ -a_2 & 1 \end{pmatrix}$$

zu untersuchen. Offensichtlich gibt es zwei verschiedene Fälle:

Fall $a_1 \neq a_2$: Dann sind die Zeilen von B linear unabhängig, es ist $\mathrm{Rang}\,(B) = 2 = 2 - 0$. Also ist M eine nulldimensionale Ebene, ein Punkt, und es gilt $\mathrm{codim}\, M = 2 = 1 + 1 = \mathrm{codim}\, M_1 + \mathrm{codim}\, M_2$, M_1 und M_2 sind in allgemeiner Lage.

Fall $a_1 = a_2$: Dann sind die Zeilen von B gleich, es ist $\mathrm{Rang}\,(B) = 1$. Also ist M eine Gerade, was man auch direkt aus $M_1 = M_2$ folgern kann. Wegen $\mathrm{codim}\, M = 1 < 1 + 1 = \mathrm{codim}\, M_1 + \mathrm{codim}\, M_2 = 2$ sind M_1 und M_2 nicht in allgemeiner Lage. Allerdings kann man durch eine beliebig kleine „Drehung“, etwa von M_1 erreichen, daß $a_1 \neq a_2$ wird und damit M_1 und M_2 in allgemeiner Lage sind. Diese Bemerkung stimmt mit der Anschauung überein.

3.7.20 Beispiel

Wir betrachten $M_1 := \{(x_1, \ldots, x_{2n}) \in \mathbb{R}^{2n} : x_1 = \ldots = x_n = 0\}$. Das Gleichungssystem wird durch die $n \times 2n$-Matrix

$$B_1 = B_1^{n,2n} = (\, E^{n,n}, \quad \mathcal{O}^{n,n}\,)$$

beschrieben, es ist $M_1 = \left\{\vec{x} \in \mathbb{R}^{2n} : B_1 \circ \vec{x} = \vec{0}\right\}$. Da B_1 genau n linear unabhängige Spalten besitzt, ist $\mathrm{Rang}\,(B_1) = n = 2n - n$, also ist M_1 eine n-dimensionale Ebene. Wir betrachten ferner die n-dimensionale Ebene $M_2 := \left\{\vec{x} \in \mathbb{R}^{2n} : x_{n+1} = \ldots = x_{2n} = 0\right\} = \left\{\vec{x} \in \mathbb{R}^{2n} : B_2 \circ \vec{x} = \vec{0}\right\}$, wobei

$$B_2 = B_2^{n,2n} = (\, \mathcal{O}^{n,n}, \quad E^{n,n}\,) \quad \text{und} \quad \mathrm{Rang}\,(B_2) = n$$

ist. Für den Durchschnitt gilt $M := M_1 \cap M_2 = \{\vec{x} \in \mathbb{R}^{2n} : x_1 = \ldots = x_n = x_{n+1} = \ldots = x_{2n} = 0\} = \left\{\vec{0}\right\}$, M ist ein Punkt, d. h. $\mathrm{codim}\, M = 2n = \mathrm{codim}\, M_1 + \mathrm{codim}\, M_2$. Folglich sind M_1 und M_2 in allgemeiner Lage.

Durch eine scheinbar kleine Änderung erhalten wir eine ganz andere Situation. Statt M_2 betrachten wir die n-dimensionale Ebene

$$M_2' := \left\{\vec{x} \in \mathbb{R}^{2n} : x_n = \ldots = x_{2n-1} = 0\right\} = \left\{\vec{x} \in \mathbb{R}^{2n} : B_2' \circ \vec{x} = \vec{0}\right\}$$

mit $B_2' = (\, \mathcal{O}^{n,n-1}, \quad E^{n,n}, \quad \mathcal{O}^{n,1}\,)$ und $\mathrm{Rang}\,(B_2') = n$. Eine $M' := M_1 \cap M_2' = \{\vec{x} \in \mathbb{R}^{2n} : x_1 = \ldots = x_{2n-1} = 0\}$ beschreibende $(2n-1) \times 2n$-Matrix ist

$$B' = (\, E^{2n-1,2n-1}, \quad \mathcal{O}^{2n-1,1}\,)\,,$$

offensichtlich ist $\mathrm{Rang}\,(B') = 2n - 1$. Also ist M' eine eindimensionale Ebene, eine Gerade. Es gilt $2n - 1 = \mathrm{codim}\, M' < \mathrm{codim}\, M_1 + \mathrm{codim}\, M_2' = 2n$, M_1 und M_2' sind nicht in allgemeiner Lage, die Dimension des Durchschnitts ist zu groß.

Wir „drehen“ nun M_2' um einen geeigneten, beliebig kleinen „Winkel“, um wieder allgemeine Lage herzustellen. Für $\varepsilon \neq 0$ sei

$$\begin{aligned} M'_{2,\varepsilon} &:= \left\{\vec{x} \in \mathbb{R}^{2n} : x_{n+1} = \ldots = x_{2n-1} = 0, x_n = \varepsilon x_{2n}\right\} \\ &= \left\{\vec{x} \in \mathbb{R}^{2n} : B'_{2,\varepsilon} \circ \vec{x} = \vec{0}\right\} \end{aligned}$$

mit $B'_{2,\varepsilon} = (\,\mathcal{O}^{n,n-1},\quad E^{n,n},\quad -\varepsilon\,\vec{e}_1\,)$ und Rang $\left(B'_{2,\varepsilon}\right) = n$. Nun folgt wieder aus $\vec{x} \in M'_\varepsilon := M_1 \cap M'_{2,\varepsilon}$, daß $x_1 = \ldots x_{2n-1} = 0$, $x_{2n} = \frac{1}{\varepsilon}x_n = 0$ und damit $\vec{x} = \vec{0}$ gilt. Es ist $M'_\varepsilon = \left\{\vec{0}\right\}$, und M_1 und $M'_{2,\varepsilon}$ sind in allgemeiner Lage.

3.8 Spezielle lineare Abbildungen

Wir wollen dieses Kapitel 3 beschließen, indem wir einige Beispiele linearer Abbildungen $\mathbb{R}^n \to \mathbb{R}^m$ zusammenstellen, die in zwei und drei Dimensionen sämtlich der anschaulichen Geometrie entstammen. Im ersten Teil dieses Abschnitts stellen wir Projektionsabbildungen als typische Beispiele für nicht-injektive lineare Abbildungen vor. Im zweiten Teil behandeln wir bijektive lineare Abbildungen (Isomorphismen) des $\mathbb{R}^n$ und stellen zunächst noch einmal die hierzu bislang erzielten theoretischen Resultate zusammen. An Beispielen diskutieren wir insbesondere Homothetien (Streckungen bzw. Stauchungen des Raumes), Dehnungen und Scherungen.

Projektionen

3.8.1 Definition

Es sei $m \leq n$. Eine lineare Abbildung $F : \mathbb{R}^n \to \mathbb{R}^m$ heißt **Projektion** *von $\mathbb{R}^n$ auf $\mathbb{R}^m$, falls m Indizes $1 \leq i_1 < \ldots < i_m \leq n$ existieren, so daß für alle $\vec{x} \in \mathbb{R}^n$ gilt:*

$$F(\vec{x}) = F(x_1, \ldots, x_n) = (x_{i_1}, \ldots, x_{i_m}).$$

Im Spezialfall $m = 1$ nennt man

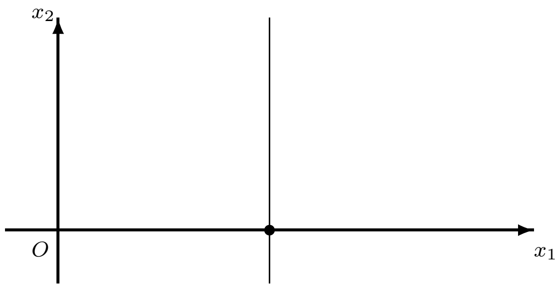
$$P_i : \mathbb{R}^n \to \mathbb{R}, \qquad \vec{x} = (x_1, \ldots, x_n) \mapsto x_i$$

die i-te Projektion von $\mathbb{R}^n \to \mathbb{R}$.

Offensichtlich sind Projektionen surjektiv und für $m < n$ nicht injektiv.

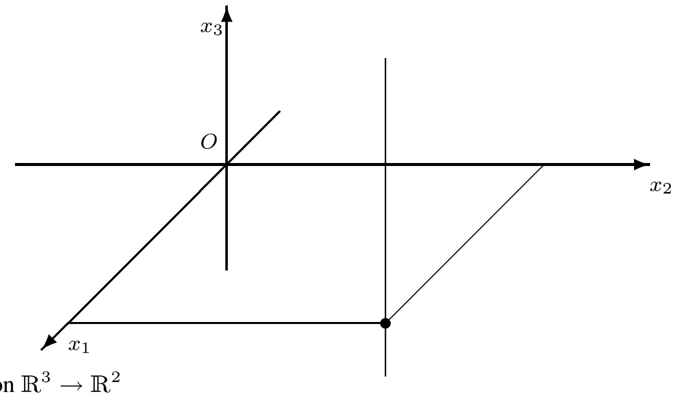
3.8.2 Beispiele

(a) Wir betrachten $P_1 : \mathbb{R}^2 \to \mathbb{R}, \quad (x_1, x_2) \mapsto x_1$. Die zugehörige Matrix ist $(\,1,\quad 0\,)$. Durch P_1 werden anschaulich Parallelen zur x_2-Achse auf einen Punkt abgebildet.



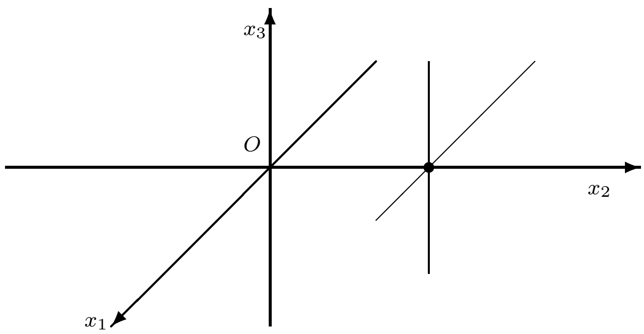
Projektion auf die x_1-Achse

(b) Auch durch $F : \mathbb{R}^3 \to \mathbb{R}^2, \quad (x_1, x_2, x_3) \mapsto (x_1, x_2)$ werden Geraden auf einen Punkt „projiziert".



Projektion $\mathbb{R}^3 \to \mathbb{R}^2$

(c) Durch die zweite Koordinatenprojektion $P_2 : \mathbb{R}^3 \to \mathbb{R}, \quad (x_1, x_2, x_3) \mapsto x_2$ werden Ebenen auf Punkte abgebildet.



Projektion $\mathbb{R}^3 \to \mathbb{R}$

Umkehrbar lineare Abbildungen

Da bijektive lineare Abbildungen nur zwischen Vektorräumen derselben Dimension existieren, betrachten wir im Rest dieses Abschnitts nur noch Abbildungen $F : \mathbb{R}^n \to \mathbb{R}^n$. Um die Tatsache hervorzuheben, daß Urbild- und Wertebereich übereinstimmen, definieren wir:

3.8.3 Definition

Unter einer **umkehrbar linearen** *oder* **invertierbaren linearen Abbildung** *des* $\mathbb{R}^n$ *verstehen wir eine bijektive lineare Abbildung* $F : \mathbb{R}^n \to \mathbb{R}^n$*, also einen Isomorphismus* $F : \mathbb{R}^n \to \mathbb{R}^n$*.*

Im folgenden Satz stellen wir eine Vielzahl von Kriterien dafür zusammen, ob eine lineare Abbildung $F : \mathbb{R}^n \to \mathbb{R}^n$ umkehrbar linear ist. Auch hier wollen wir wieder Nutzen aus dem Matrizenkalkül ziehen.

3.8.4 Satz

Es sei $F : \mathbb{R}^n \to \mathbb{R}^n$ *eine lineare Abbildung und* A *die zugehörige* $n \times n$*-Matrix. Dann sind folgende Aussagen äquivalent:*

(a) F *ist umkehrbar linear.*

(b) A *ist nichtsingulär.*

(c) Es gibt eine $n \times n$*-Matrix* B *mit* $B \circ A = E$*.*

(d) Es gibt eine $n \times n$*-Matrix* B *mit* $A \circ B = E$*.*

(e) $\mathrm{Rang}\,(A) = n$*.*

(f) Alle Spalten von A *sind linear unabhängig.*

(g) Alle Zeilen von A *sind linear unabhängig.*

(h) Die Spalten von A *erzeugen* $\mathbb{R}^n$*.*

(i) Die Zeilen von A *erzeugen* $\mathbb{R}^n$*.*

(j) F *ist injektiv.*

(k) F *ist surjektiv.*

Beweis

Wir brauchen nur soviele Äquivalenzen zu beweisen, daß sich die übrigen mit Hilfe der Transitivität von „$\Leftrightarrow$“ ergeben.

„(a) $\Leftrightarrow$ (b)“: Diese Äquivalenz haben wir bereits in Folgerung 3.6.13 im Zusammenhang mit der Definition nichtsingulärer Matrizen erläutert.

„(j) $\Leftrightarrow$ (k)“: Siehe Folgerung 3.4.5.

„(a) $\Leftrightarrow$ (j)“: Folgt dann unmittelbar.

„(b) $\Rightarrow$ (c)“: Offensichtlich.

„(c) $\Rightarrow$ (j)“: Sei $F(\vec{x}) = \vec{0}$. Nach (c) folgt $\vec{x} = E \circ \vec{x} = (B \circ A) \circ \vec{x} = B \circ (A \circ \vec{x}) = B \circ \vec{0} = \vec{0}$.

Wir wissen bereits „(j) $\Leftrightarrow$ (b)“, und damit haben wir „(b) $\Leftrightarrow$ (c)“ und „(c) $\Leftrightarrow$ (j)“.

„(b) $\Rightarrow$ (d)“: Offensichtlich.

„(d) $\Rightarrow$ (k)“: Sei $\vec{y} \in \mathbb{R}^n$ gegeben. Mit B gemäß (d) setze $\vec{x} := B \circ \vec{y}$; es folgt $\vec{y} = E \circ \vec{y} = (A \circ B) \circ \vec{y} = A \circ (B \circ \vec{y}) = A \circ \vec{x}$, also ist $\vec{y} = F(\vec{x})$.

Indem wir hier „(k) $\Leftrightarrow$ (b)“ verwenden, folgt damit wie oben „(b) $\Leftrightarrow$ (d)“ und „(d) $\Leftrightarrow$ (k)“.

„(h) $\Leftrightarrow$ (k)“: Die Spalten von A erzeugen $\mathrm{im}(F)$.

Die übrigen Äquivalenzen folgen aus Zeilenrang(A) = Spaltenrang$(A) = \mathrm{Rang}\,(A)$ und Satz 3.3.18, wonach ein System aus n Vektoren in $\mathbb{R}^n$, das linear unabängig ist *oder* den $\mathbb{R}^n$ erzeugt, bereits eine Basis ist. □

Es liegt nahe, auf der mit der Matrizenmultiplikation versehenen Menge der nichtsingulären Matrizen eine Gruppenstruktur zu vermuten.

3.8.5 Definition

Die Menge $GL(n, \mathbb{R}) := \{A : A$ ist nichtsinguläre (reelle) $n \times n$-Matrix$\}$ heißt **allgemeine lineare Gruppe** *vom Rang n über $\mathbb{R}$. Die Elemente von $GL(n, \mathbb{R})$ heißen auch* **reguläre** *oder* **invertierbare** *Matrizen.*

Die Abkürzung GL erklärt sich durch die englische Bezeichnung „general linear group“.

3.8.6 Satz

Bezüglich der Matrizenmultiplikation ist $GL(n, \mathbb{R})$ eine Gruppe. Für $n \geq 2$ ist diese Gruppe nicht kommutativ.

Beweis

Die Einheitsmatrix ist nichtsingulär, also ist $GL(n, \mathbb{R}) \neq \emptyset$. In Folgerung 3.6.14 hatten wir festgestellt, daß das Produkt nichtsingulärer Matrizen ebenfalls nichtsingulär ist. Also ist $GL(n, \mathbb{R})$ gegenüber der Matrizenmultiplikation abgeschlossen.

(1) Die Assoziativität der Matrizenmultiplikation wurde in Satz 3.6.11 gezeigt.

(2) Die Einheitsmatrix E ist nichtsingulär, und für alle $A \in GL(n, \mathbb{R})$ gilt $A \circ E = A$.

(3) Die Existenz jeweils eines (Rechts-)Inversen zu jeder Matrix aus $GL(n, \mathbb{R})$ wird durch die Definition 3.6.12 der nichtsingulären Matrizen gesichert.

Zum Nachweis der Nicht-Kommutativität der Matrizenmultiplikation bauen wir auf Beispiel 3.6.10 auf: Dort hatten wir reguläre 2×2-Matrizen A, B mit $A \circ B \neq B \circ A$

gefunden. Für $n \geq 2$ setzen wir

$$\tilde{A} = \left(\begin{array}{c|c} A^{2,2} & \mathcal{O}^{2,n-2} \\ \hline \mathcal{O}^{n-2,2} & E^{n-2,n-2} \end{array}\right)$$

und entsprechend $\tilde{B}$. Die so entstehenden Matrizen haben maximalen Rang n und sind folglich ebenfalls regulär. Man erhält

$$\tilde{A} \circ \tilde{B} = \left(\begin{array}{c|c} A \circ B & \mathcal{O} \\ \hline \mathcal{O} & E \end{array}\right), \qquad \tilde{B} \circ \tilde{A} = \left(\begin{array}{c|c} B \circ A & \mathcal{O} \\ \hline \mathcal{O} & E \end{array}\right)$$

und damit $\tilde{A} \circ \tilde{B} \neq \tilde{B} \circ \tilde{A}$. □

Man beachte, daß die Summe invertierbarer Matrizen im allgemeinen *nicht* invertierbar ist.

Um die Invertierbarkeit einer gegebenen $n \times n$-Matrix A zu prüfen, verwendet man in der Regel Kriterium (d) aus Satz 3.8.4. Die Existenz einer $n \times n$-Matrix B mit $A \circ B = E$ ist gleichbedeutend damit, daß die Gleichungssysteme $A \circ \vec{b}_j = \vec{e}_j$ für $j = 1, \ldots, n$ eine Lösung $\vec{b}_j$ besitzen. Dabei ist $\vec{b}_j$ die j-te Spalte der gesuchten Matrix B. Anstatt diese n Gleichungssysteme nacheinander zu behandeln und immer wieder dieselben Umformungen durchzuführen, werden diese Gleichungssysteme simultan betrachtet. Man gelangt so zu der folgenden Variante des *Gaußschen Algorithmus*':

Man schreibt gleichzeitig alle rechten Seiten in die erweiterte Koeffizientenmatrix, man beginnt also mit der $n \times 2n$-Matrix (A, E). Durch *Zeilenumformungen* gelangt man zu einer Matrix $\left(\tilde{A}, B\right)$.

Ist $\operatorname{Rang}(A) = n$, so läßt sich $\tilde{A} = E$ erreichen, die Matrix A ist invertierbar, $A \in GL(n, \mathbb{R})$, und es ist $B = A^{-1}$.

Ist dagegen $\operatorname{Rang}(A) < n$, so erhält man eine Matrix $\tilde{A}$, die mindestens eine Nullzeile enthält. In diesem Fall besitzt A keine inverse Matrix, $A \notin GL(n, \mathbb{R})$.

3.8.7 Beispiele

(a) Wir wollen untersuchen, ob die Matrix

$$A = \begin{pmatrix} -1 & 1 & 1 \\ -3 & 4 & 3 \\ -4 & 6 & 5 \end{pmatrix}$$

nichtsingulär ist und eine inverse Matrix besitzt.

Wir erhalten durch Zeilenumformungen:

$$
\begin{aligned}
(A,\ E) &= \left(\begin{array}{ccc|ccc} \boxed{\text{-1}} & 1 & 1 & 1 & 0 & 0 \\ -3 & 4 & 3 & 0 & 1 & 0 \\ -4 & 6 & 5 & 0 & 0 & 1 \end{array}\right) \quad \to \left(\begin{array}{ccc|ccc} 1 & -1 & -1 & -1 & 0 & 0 \\ 0 & \boxed{1} & 0 & -3 & 1 & 0 \\ 0 & 2 & 1 & -4 & 0 & 1 \end{array}\right) \\
&\to \left(\begin{array}{ccc|ccc} 1 & 0 & -1 & -4 & 1 & 0 \\ 0 & 1 & 0 & -3 & 1 & 0 \\ 0 & 0 & \boxed{1} & 2 & -2 & 1 \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -2 & -1 & 1 \\ 0 & 1 & 0 & -3 & 1 & 0 \\ 0 & 0 & 1 & 2 & -2 & 1 \end{array}\right) \\
&= (E,\ B).
\end{aligned}
$$

Also ist $\operatorname{Rang}(A) = 3$ und A invertierbar, $A \in GL(3, \mathbb{R})$, es gilt

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

Es mag zunächst überraschen, daß sowohl A als auch A^{-1} nur ganzzahlige Komponenten haben. Dieses „Rätsel“ werden wir im Abschnitt über die Cramersche Regel, Beispiel 5.5.7, auflösen.

(b) Entsprechend wollen wir

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

auf Invertierbarkeit untersuchen:

$$
\begin{aligned}
(A,\ E) &= \left(\begin{array}{ccc|ccc} \boxed{1} & 1 & 2 & 1 & 0 & 0 \\ -2 & 3 & 0 & 0 & 1 & 0 \\ -1 & 4 & 2 & 0 & 0 & 1 \end{array}\right) \quad \to \left(\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & \boxed{5} & 4 & 2 & 1 & 0 \\ 0 & 5 & 4 & 1 & 0 & 1 \end{array}\right) \\
&\to \left(\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 5 & 4 & 2 & 1 & 0 \\ 0 & 0 & 0 & -1 & -1 & 1 \end{array}\right) = \left(\tilde{A},\ B\right).
\end{aligned}
$$

Es ist $\operatorname{Rang}(A) = \operatorname{Rang}\left(\tilde{A}\right) = 2 < 3$, also besitzt A keine inverse Matrix, $A \notin GL(3, \mathbb{R})$.

3.8.8 Beispiel

Wir betrachten die $n \times n$-Matrix

$$A = (a_{ij})_{i,j=1,\ldots,n} = \begin{pmatrix} 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & \ldots & 1 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 1 & \ldots & 0 & 0 \\ 1 & 0 & \ldots & 0 & 0 \end{pmatrix}$$

mit

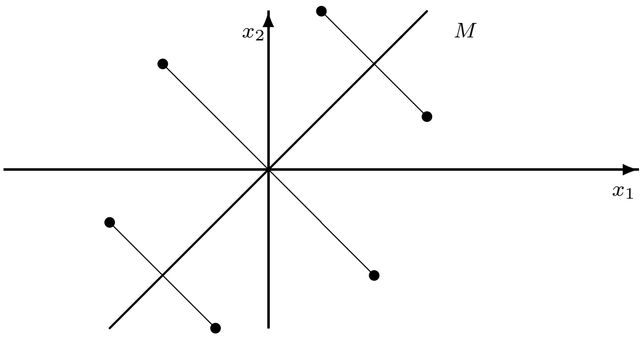
$$a_{ij} = \begin{cases} 1, & \text{falls } i + j = n + 1, \\ 0, & \text{sonst.} \end{cases}$$

Durch A wird die lineare Abbildung

$$F : \mathbb{R}^n \to \mathbb{R}^n,\ (x_1, \ldots, x_n) \mapsto (x_n, \ldots, x_1)$$

gegeben. Man berechnet $A \circ A = E$, entsprechend gilt $F \circ F = id$. Nach Satz 3.8.4 ist F umkehrbar linear.

Zur anschaulichen Interpretation sei nun speziell $n = 2$. Es ist $F(x_1, x_2) = (x_2, x_1)$. Obwohl der Begriff „Spiegelung an einer beliebigen Achse“ und die zugrundeliegende Orthogonalität erst in Kapitel 6 streng eingeführt werden, sei erwähnt, daß F eine Spiegelung an der Winkelhalbierenden $M = \{(x_1, x_2) \in \mathbb{R}^2 :\ x_1 = x_2\}$ ist.



Spiegelung an der Winkelhalbierenden

Zugleich ist M die Menge der *Fixpunkte* $\left\{\vec{x} \in \mathbb{R}^2 :\ F(\vec{x}) = \vec{x}\right\}$ von F, denn $F(\vec{x}) = \vec{x}$ ist äquivalent zu $x_1 = x_2$.

Spezielle umkehrbar lineare Abbildungen

Im Folgenden wollen wir einige Klassen spezieller linearer Abbildungen bzw. spezieller Matrizen als Untergruppen von $GL(n, \mathbb{R})$ betrachten. Dabei ist eine Untergruppe ganz analog dem Untervektorraum definiert, vgl. Definition 1.2.5.

3.8.9 Definition

Es sei $(G, \circ)$ eine Gruppe. Eine Teilmenge $H \subset G$ heißt **Untergruppe**, *falls gilt:*

(1) $H \neq \emptyset$.

(2) Für alle $x, y \in H$ ist auch $x \circ y \in H$.

(3) Für alle $x \in H$ ist $x^{-1} \in H$.

Es gilt wieder: Eine Untergruppe H von G ist mit derselben Operation wie auf G eine Gruppe.

Homothetien

3.8.10 Definition

Eine lineare Abbildung $F : \mathbb{R}^n \to \mathbb{R}^n$ heißt **Homothetie**, *falls eine reelle Zahl $a \neq 0$ existiert, so daß für alle $\vec{x} \in \mathbb{R}^n$ gilt:*

$$F(\vec{x}) = a \cdot \vec{x}.$$

Wegen $F(\vec{e}_i) = a\vec{e}_i$ ist $A = \begin{pmatrix} a & 0 & \dots & 0 \\ 0 & a & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a \end{pmatrix} = a \cdot E$ die zu F gehörige Matrix.

3.8.11 Satz

Die Menge aller eine Homothetie beschreibenden $n \times n$-Matrizen

$$H := \{A : \ A = a \cdot E^{n,n}, \ a \in \mathbb{R} \setminus \{0\}\}$$

ist eine Untergruppe von $GL(n, \mathbb{R})$.

Beweis

Wir verwenden die Rechenregeln aus Abschnitt 3.6 für Skalarmultiplikation und Matrizenmultiplikation.

Sei $A = a \cdot E \in H$. Wegen $\left(\frac{1}{a} \cdot E\right) \circ (a \cdot E) = (a \cdot E) \circ \left(\frac{1}{a} \cdot E\right) = \left(a \cdot \frac{1}{a}\right) \cdot (E \circ E) = E$ ist A nichtsingulär. Damit gilt $H \subset GL(n, \mathbb{R})$ und H enthält mit jedem Element auch sein Inverses. Offenbar ist $H \neq \emptyset$, schließlich gilt für $A = a \cdot E, \ B = b \cdot E \in H$: $A \circ B = (a \cdot E) \circ (b \cdot E) = (a \cdot b) \cdot (E \circ E) = (ab) \cdot E$, also $A \circ B \in H$. Die Menge H ist also bezüglich der Gruppenoperation „$\circ$“ abgeschlossen. □

Wir präzisieren noch die offensichtliche Ähnlichkeit der Multiplikation in H und der Multiplikation reeller Zahlen.

3.8.12 Definition

Ein bijektiver Gruppenhomomorphismus heißt **Gruppenisomorphismus**.

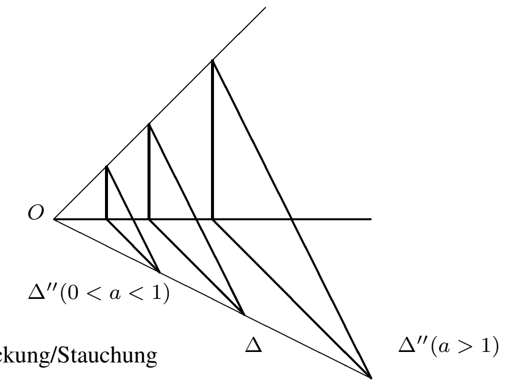
3.8.13 Bemerkung

Die Menge der von 0 verschiedenen Zahlen $\mathbb{R}^* := \mathbb{R} \setminus \{0\}$ ist mit der gewöhnlichen Multiplikation eine abelsche Gruppe. Durch $h : \mathbb{R}^* \to H,\ a \mapsto a \cdot E$ wird offensichtlich ein Gruppenisomorphismus $\mathbb{R}^* \to H$ definiert. Daher ist H eine *abelsche Untergruppe* der *nichtabelschen Gruppe* $GL(n, \mathbb{R})$.

3.8.14 Beispiel

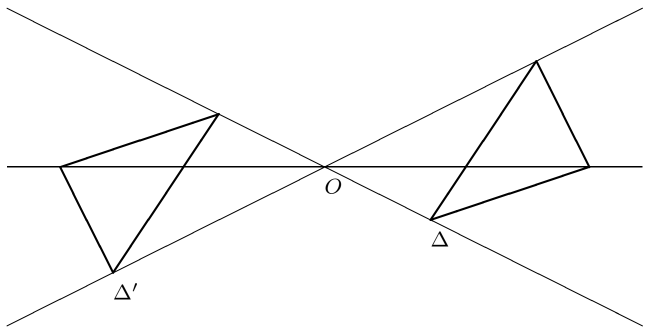
Zur Veranschaulichung betrachten wir für $a \in \mathbb{R}^*$ die Homothetie $F : \mathbb{R}^2 \to \mathbb{R}^2,\ F(\vec{x}) = a\vec{x}$.

Fall 1: $a > 0$. Die Strahlensätze zeigen, daß ein Dreieck durch die Streckung unter H in ein ähnliches Dreieck (mit denselben Längenverhältnissen) überführt wird.



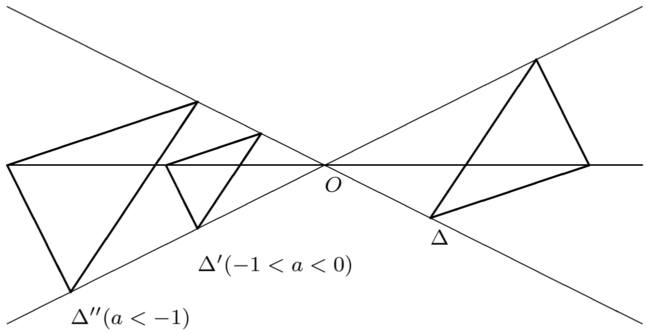
Streckung/Stauchung

Fall 2: $a = -1$. Die Abbildung F ist eine Spiegelung am Nullpunkt. In Kapitel 6 werden wir sehen, daß (nur) im $\mathbb{R}^2$ diese Spiegelung auch als Drehung um $180°$ aufgefaßt werden kann.



Spiegelung am Nullpunkt

Fall 3: $a < 0$. Dann ist $-a > 0$ und $F(\vec{x}) = (-1) \cdot ((-a) \cdot \vec{x}) = H \circ G(\vec{x})$ mit $G(\vec{x}) = (-a) \cdot \vec{x}$ und $H(\vec{x}) = (-1) \cdot \vec{x}$. In diesem Falle ist F Zusammensetzung von einer Streckung und einer Spiegelung am Nullpunkt.



Zusammensetzung einer Spiegelung und einer Streckung

In jedem Fall bleiben die Längenverhältnisse in Dreiecken erhalten, Homothetien sind also spezielle Ähnlichkeitstransformationen. Zudem sind hier die Seiten der Bilddreiecke und der Urbilddreiecke parallel.

Dehnungen

3.8.15 Definition

Eine lineare Abbildung $F : \mathbb{R}^n \to \mathbb{R}^n$ heißt **Dehnung**, *falls n reelle Zahlen $a_1, \ldots, a_n \in \mathbb{R}^*$ existieren, so daß $F(\vec{e}_i) = a_i \cdot \vec{e}_i$ für $i = 1, \ldots, n$ gilt.*

Zu F gehört also die Matrix $\begin{pmatrix} a_1 & 0 & \ldots & 0 \\ 0 & a_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_n \end{pmatrix}$.

3.8.16 Satz

Die Menge $D := \{A \in \mathbb{R}^{n \times n} : A$ beschreibt eine Dehnung des $\mathbb{R}^n\}$ ist eine Untergruppe von $GL(n, \mathbb{R})$.

Beweis

Jede Homothetie ist insbesondere eine Dehnung, $H \subset D$, also ist D nicht leer. Für $A = \begin{pmatrix} a_1 & 0 & \ldots & 0 \\ 0 & a_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_n \end{pmatrix} \in D$ existiert wegen $a_i \neq 0$, $i = 1, \ldots, n$, stets die inverse Matrix $A^{-1} = \begin{pmatrix} \frac{1}{a_1} & 0 & \dots & 0 \\ 0 & \frac{1}{a_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{a_n} \end{pmatrix}$, es folgt $D \subset GL(n, \mathbb{R})$.

Außerdem beschreibt A^{-1} ebenfalls eine Dehnung. Schließlich ist D gegenüber der Matrizenmultiplikation abgeschlossen, denn mit

$$A = \begin{pmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_n \end{pmatrix} \in D, \quad B = \begin{pmatrix} b_1 & 0 & \dots & 0 \\ 0 & b_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & b_n \end{pmatrix} \in D$$

ist auch stets

$$A \circ B = \begin{pmatrix} a_1 b_1 & 0 & \dots & 0 \\ 0 & a_2 b_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_n b_n \end{pmatrix} \in D.$$

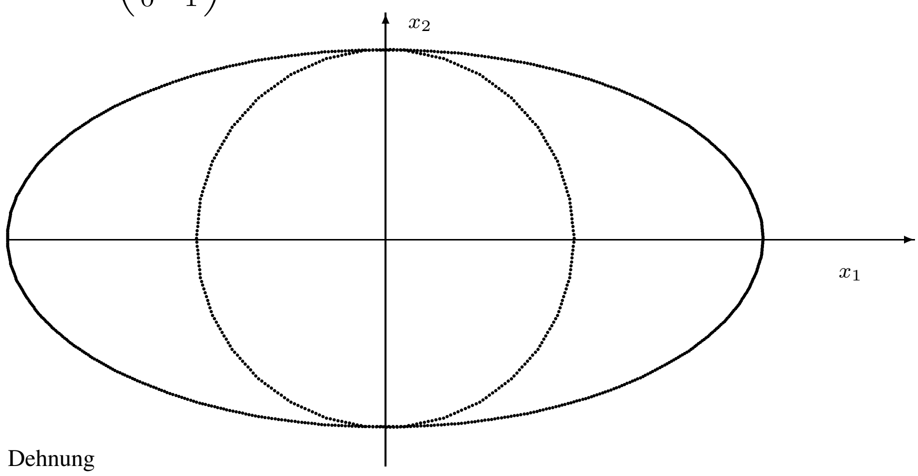
□

3.8.17 Bemerkung

Die Menge der Homothetien H ist auch eine Untergruppe der Gruppe D der Dehnungen.

3.8.18 Beispiel

Dieses Beispiel soll den Begriff „Dehnung“ erläutern. Die Abbildung $F : \mathbb{R}^2 \to \mathbb{R}^2$ werde durch $\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ gegeben, sie führt den Einheitskreis in eine Ellipse über.



Dehnung

3.8.19 Beispiel

Durch $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ wird eine Spiegelung an der x_1-Achse $(x_1, x_2) \mapsto (x_1, -x_2)$ beschrieben.

Scherungen

3.8.20 Definition

Eine lineare Abbildung $F : \mathbb{R}^n \to \mathbb{R}^n$ heißt **Scherung**, *falls Zahlen $a_2, \dots, a_n \in \mathbb{R}$ existieren, so daß für die zu F gehörige $n \times n$-Matrix A gilt:*

$$A = A_{a_2,\dots,a_n} := \begin{pmatrix} 1 & a_2 & \dots & a_n \\ 0 & & & \\ \vdots & & E^{n-1,n-1} & \\ 0 & & & \end{pmatrix}.$$

Für das Produkt derartiger „Schermatrizen“ gilt:

$$\begin{aligned} A_{a_2,\dots,a_n} \circ A_{a_2',\dots,a_n'} &= \begin{pmatrix} 1 & a_2 & \dots & a_n \\ 0 & & & \\ \vdots & & E^{n-1,n-1} & \\ 0 & & & \end{pmatrix} \circ \begin{pmatrix} 1 & a_2' & \dots & a_n' \\ 0 & & & \\ \vdots & & E^{n-1,n-1} & \\ 0 & & & \end{pmatrix} \\ &= \begin{pmatrix} 1 & a_2 + a_2' & \dots & a_n + a_n' \\ 0 & & & \\ \vdots & & E^{n-1,n-1} & \\ 0 & & & \end{pmatrix}, \end{aligned}$$

also

$$A_{a_2,\dots,a_n} \circ A_{a_2',\dots,a_n'} = A_{a_2+a_2',\dots,a_n+a_n'}. \qquad (*)$$

Die Zusammensetzung zweier Scherungen ist also wieder eine Scherung. Diese Formel zeigt ferner $A_{a_2,\dots,a_n} \circ A_{-a_2,\dots,-a_n} = E$. Schermatrizen sind also invertierbar, und die inverse Matrix beschreibt ebenfalls eine Scherung. Also haben wir das folgende Resultat:

3.8.21 Satz

Die Menge der zu Scherungen des $\mathbb{R}^n$ gehörigen Matrizen

$$S := \left\{ A_{a_2,\dots,a_n} \in \mathbb{R}^{n \times n} : \ a_2, \dots, a_n \in \mathbb{R} \right\}$$

ist eine Untergruppe von $GL(n, \mathbb{R})$.

Auch diese Untergruppe von $GL(n,\mathbb{R})$ können wir mit einer bereits bekannten Gruppe identifizieren:

3.8.22 Satz

Die Abbildung $f : \mathbb{R}^{n-1} \to S$, $(a_2,\dots,a_n) \mapsto A_{a_2,\dots,a_n}$ *ist ein Gruppenisomorphismus von* $(\mathbb{R}^{n-1},+)$ *auf* $(S,\circ)$. *Insbesondere ist* $(S,\circ)$ *abelsch.*

Beweis

Die Bijektivität von f ist offensichtlich. Es ist also nur die Gruppenhomomorphismuseigenschaft

$$f\left((a_2,\dots,a_n)+(a'_2,\dots,a'_n)\right) = f(a_2,\dots,a_n)\circ f(a_2,\dots,a_n)$$

zu zeigen. Diese ist aber gleichbedeutend mit der oben bewiesenen Identität $(*)$. □

Sei $F : \mathbb{R}^n \to \mathbb{R}^n$ eine Scherung und $A = A_{a_2,\dots,a_n} \in S$ die zugehörige Matrix. Für $\vec{x} \in \mathbb{R}^n$ gilt $A\circ\vec{x} = (x_1 + a_2x_2 + \cdots + a_nx_n, x_2,\dots,x_n)$. Die Abbildung F läßt die letzten $(n-1)$ Komponenten unverändert. Hält man $(x_2,\dots,x_n)$ fest, so wird x_1 durch F auf einer zur x_1-Achse parallelen Geraden durch $(0,x_2,\dots,x_n)$ um $a_2x_2+\cdots+a_nx_n$ verschoben.

Für die Menge der Fixpunkte $M := \{\vec{x}\in\mathbb{R}^n : \ A\circ\vec{x}=\vec{x}\}$ gilt:

$$M = \left\{\vec{x}\in\mathbb{R}^n : \ A\circ\vec{x} - E\circ\vec{x} = \vec{0}\right\} = \left\{\vec{x}\in\mathbb{R}^n : (A-E)\circ\vec{x} = \vec{0}\right\}.$$

Die Matrix

$$A - E = \begin{pmatrix} 0 & a_2 & \dots & a_n \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

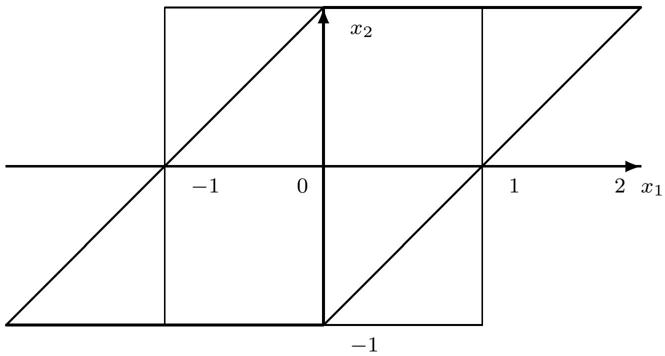
hat

$$\text{Rang}\,(A-E) = \begin{cases} 0, & \text{falls } a_2 = \ldots = a_n = 0; \\ 1, & \text{falls ein } a_i \neq 0. \end{cases}$$

Daher ist M eine $(n-1)$-dimensionale Ebene, d. h. eine Hyperebene, sofern wenigstens ein $a_i \neq 0$ ist, und eine n-dimensionale Ebene, d. h. $M = \mathbb{R}^n$, falls $a_2 = \ldots = a_n = 0$.

3.8.23 Beispiel

Sei $n = 2$ und $A = A_1$, d. h. $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Die Menge der Fixpunkte ist $M = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 = 0\}$, also die x_1-Achse. Das Einheitsquadrat geht durch die entsprechende Abbildung in ein Parallelogramm über, indem zwei Seiten parallel zur x_1-Achse verschoben und die anderen zwei Seiten um -45 Grad gedreht und entsprechend verlängert werden.



Scherung

4 Vektorraumkonstruktionen

Häufig ist mit der Untersuchung von Vektorräumen und linearen Abbildungen in ganz naheliegender Weise die Betrachtung daraus hervorgehender Vektorräume und Abbildungen verbunden.

In der Mechanik wird die Position eines Massepunktes durch Vektoren aus $\mathbb{R}^3$ ($\approx$ physikalischer Raum) beschrieben, für die Beschreibung der Zeit verwendet man $\mathbb{R}$. Oft sind aber Raum und Zeit gleichzeitig zu betrachten und zu einem Vektorraum, der „direkten Summe $\mathbb{R}^3 \oplus \mathbb{R}$“, zusammenzufügen. Für die Geschwindigkeit eines Massepunktes verwendet man ebenfalls Vektoren aus $\mathbb{R}^3$. Da in der Mechanik aber stets Position und Geschwindigkeit (Impuls) und im allgemeinen eine Anzahl n von Massepunkten gleichzeitig betrachtet werden, möchte man „Orts-“ und „Impulsraum“ zu „$\mathbb{R}^3 \oplus \mathbb{R}^3$“ und für das Gesamtsystem n Kopien derselben zusammensetzen. Mit diesem „Zusammensetzen“ und umgekehrt auch dem „Zerlegen“ von Vektorräumen befassen wir uns in Abschnitt 4.2.

In Abschnitt 3.8 hatten wir u. a. Projektionsabbildungen und insbesondere Projektionen auf jeweils eine Koordinatenrichtung vorgestellt. Das ganze Architektur- und Ingenieurwesen basiert auf der intensiven Verwendung von Projektionen, und das schon bei der Beschreibung dreidimensionaler Objekte. Um so mehr sind Projektionsabbildungen geeignet, um höher- oder unendlichdimensionale Vektorräume zu studieren. Wie wichtig solche unendlichdimensionalen Vektorräume sind, wird in Abschnitt 4.1 an Hand einiger Beispiele (Folgen- und Funktionenräume) erläutert werden. In solchen Räumen wird beispielsweise Gleichheit zweier Elemente häufig so gezeigt, daß man die Gleichheit deren Werte unter allen bzw. hinreichend vielen Projektionsabbildungen zeigt.

Da wir hier über den Begriff der Orthogonalität noch nicht verfügen und dieser in vielen wichtigen Beispielen unendlichdimensionaler (normierter) Vektorräume *nicht eingeführt werden kann*, lösen wir uns von der damit verbundenen speziellen Vorstellung von Projektionsabbildungen und gelangen zum allgemeineren Begriff der „Linearform“, deren Gesamtheit den zum ursprünglichen Vektorraum V „dualen Vektorraum“ V^* bildet. Diesen „Dualräumen“ widmen wir den zweiten Teil dieses Kapitels.

4.1 Beispiele unendlichdimensionaler Vektorräume

Gerade die in den folgenden Abschnitten einzuführenden Zerlegungen von Vektorräumen in direkte Summen sowie die Dual- und Bidualräume spielen in der Theorie der „unendlichdimensionalen normierten Vektorräume“ (Funktionalanalysis) eine wesentliche Rolle. Um schon hier den Blickwinkel ein wenig in diese Richtung öffnen zu können, diskutieren wir detailliert zwei Beispiele unendlichdimensionaler Vektorräume und dabei insbesondere die Begriffe „Basis“ und „Dimension“.

Wir erinnern zunächst an Definition 1.3.14.

4.1.1 Definition

Sei V ein reeller Vektorraum und $M \subset V$ eine Teilmenge.

(a) *Die Menge M heißt* **linear unabhängig**, *wenn jedes endliche System $x_1, \ldots, x_n$ $(n \in \mathbb{N})$ paarweise verschiedener Vektoren aus M linear unabhängig ist.*

(b) *Man sagt, daß V von M* **erzeugt** *wird, wenn zu jedem $x \in V$ eine endliche Anzahl $n \in \mathbb{N}$ von Vektoren $x_1, \ldots, x_n \in M$ und von reellen Zahlen $a_1, \ldots, a_n \in \mathbb{R}$ mit $x = a_1x_1 + \ldots + a_nx_n$ existiert.*

(c) *Man nennt M eine* **Basis** *von V, wenn M linear unabhängig ist und V erzeugt.*

In der Funktionalanalysis tritt ein anderer Basisbegriff (Schauderbasis, Hilbertbasis) in den Vordergrund. Man läßt dort auch die Darstellung mittels konvergenter Reihen statt nur endlicher Linearkombinationen zu. Um dann Verwechslungen zu vermeiden, nennt man Basen im oben erklärten Sinne auch *Basen im Sinne der (linearen) Algebra* oder *Hamelbasen*.

Wie bereits nach Satz 1.3.17 erläutert, kann man allgemein beweisen:

4.1.2 Satz

Jeder Vektorraum hat eine Basis.

Aus Satz 3.3.15 folgt auch: Ein unendlichdimensionaler Vektorraum kann keine endliche Basis besitzen. Auch wenn man „∞“ als Maßzahl für die Länge einer Basis hinzunimmt, bleibt also richtig: Die Dimension ist gleich der Basislänge.

Wie in der Differential- und Integralrechnung gezeigt wird, gibt es jedoch „verschiedene Stufen von Unendlichkeit“. So gibt es abzählbar unendliche Mengen wie $\mathbb{N}$ und überabzählbare Mengen wie etwa $\mathbb{R}$ oder Intervalle der Form

$$[a, b] := \{x \in \mathbb{R} : \ a \leq x \leq b\} \text{ mit } a < b.$$

Die folgenden Beispiele zeigen, daß bei verschiedenen Vektorräumen linear unabhängige Mengen verschiedener „Mächtigkeit“ auftreten können. Bei der Dimensionszahl ∞ könnte also weiter differenziert werden. Wegen des großen Aufwandes wollen wir hier jedoch nicht in die Theorie der „Kardinalzahlen“ einsteigen.

4.1.3 Beispiel

Wir erinnern an den in Kapitel 1.3 eingeführten Vektorraum V der *reellen Polynome*. Dort wurden die Monome als Einheitsvektoren $e_0 = 1$, $e_1 = z$, $e_2 = z^2, \dots$ ausgezeichnet. Diese Monome erzeugen V, und je endlich viele (paarweise verschiedene) von ihnen sind linear unabhängig. Also ist $M := \{e_i : i \in \mathbb{N}_0\}$ eine Basis von V. Die Elemente von M können mit Indizes aus $\mathbb{N}$ durchnumeriert werden, M ist eine *abzählbar unendliche* Basis von V.

4.1.4 Beispiel

Für zwei reelle Zahlen $a < b$ betrachten wir das Intervall $I := [a, b]$ und den Vektorraum $V := \{f : I \to \mathbb{R}\}$ aller *reellen Funktionen*. Auf V ist ganz analog wie zu Beginn von Abschnitt 3.6 eine Addition durch $(f + g)(x) := f(x) + g(x)$ und eine Skalarmultiplikation durch $(c \cdot f)(x) := c \cdot f(x)$, $x \in I$, gegeben. Zwei Funktionen $f, g \in V$ sind gleich, wenn $f(x) = g(x)$ für alle $x \in I$ gilt. Die Vektorraumeigenschaften von V folgen dann direkt aus den entsprechenden Gesetzen für die reellen Zahlen. Das Nullelement $\mathcal{O} \in V$ ist die Nullfunktion: $x \mapsto 0$. Zu jedem $f \in V$ existiert mit $(-f)(x) := -(f(x))$, $x \in I$, ein negatives Element.

Gemäß dem obigen Satz besitzt V eine Basis, die man allerdings nicht explizit konstruieren kann. Es kann aber eine sehr große linear unabhängige Menge angegeben werden. Für beliebiges, aber festes $y \in I$ definieren wir $f_y \in V$ durch

$$f_y(x) = \begin{cases} 1, & \text{falls } x = y, \\ 0, & \text{sonst.} \end{cases}$$

Die Menge $M = \{f_y : y \in I\}$ enthält wie I *überabzählbar viele* Elemente und ist linear unabhängig. Seien nämlich $f_{y_1}, \dots, f_{y_n} \in M$ paarweise verschieden, d. h. $y_i \neq y_j$ für $i \neq j$; dann folgt aus

$$\sum_{i=1}^{n} a_i f_{y_i} = \mathcal{O} \quad \text{(Nullfunktion!)}, \quad a_1, \dots, a_n \in \mathbb{R},$$

daß für alle $x \in I$ gilt:

$$0 = \mathcal{O}(x) = \sum_{i=1}^{n} a_i f_{y_i}(x) = \begin{cases} a_j, & \text{falls } x = y_j, \\ 0, & \text{falls } x \notin \{y_1, \dots, y_n\}. \end{cases}$$

Also ist $a_1 = \ldots = a_n = 0$.

Gleichzeitig wird V aber nicht von M erzeugt: Sei dazu $A \subset [a, b]$ eine unendliche Menge und

$$1_A : [a, b] \to \mathbb{R}, \quad 1_A(x) = \begin{cases} 1, & \text{falls } x \in A, \\ 0, & \text{falls } x \notin A, \end{cases}$$

die *charakteristische Funktion* der Menge A. Diese Funktion ist nicht (endliche!)

Linearkombination der Elemente von M. Durch Hinzunahme dieser Funktion erhält man mit $M \cup \{1_A\}$ eine weitere linear unabhängige Menge in V.

Der Vektorraum der reellen Funktionen ist also „höherdimensional“ (in dem vor Beispiel 4.1.3 angedeuteten Sinne) als der Vektorraum der reellen Polynome.

In Verallgemeinerung zu Satz 4.1.2 gilt auch in dieser Situation der Basisergänzungssatz: es existiert eine Basis von V, die die Menge M enthält (ebenfalls ohne Beweis). So erhält man die Existenz einer *überabzählbaren Basis* von V.

4.2 Direkte Summen

Mit $V, V_1, V_2, \ldots, W, W_1, W_2, \ldots$ werden reelle Vektorräume bezeichnet.

Zuammensetzen

Vermutlich haben die einleitenden Bemerkungen zu Beginn dieses Kapitels schon den Gedanken nahegelegt, die Elemente der jeweiligen Vektorräume als ℓ-tupel zu einem Element der „direkten Summe“ zusammenzufügen.

4.2.1 Definition

Es heißt $V := V_1 \oplus \ldots \oplus V_\ell := \oplus_{i=1}^{\ell} V_i := \{(x_1, \ldots, x_\ell) : \ x_i \in V_i$ *für* $i = 1, \ldots, \ell\}$ *die* **direkte Summe** *von* $V_1, \ldots, V_\ell$*. Auf* V *wird durch*

$$(x_1, \ldots, x_\ell) + (y_1, \ldots, y_\ell) := (x_1 + y_1, \ldots, x_\ell + y_\ell)$$

eine **Addition** *und durch*

$$c \cdot (x_1, \ldots, x_\ell) := (c \cdot x_1, \ldots, c \cdot x_\ell)$$

eine **Skalarmultiplikation** *erklärt. Zwei Elemente* $(x_1, \ldots, x_\ell), (y_1, \ldots, y_\ell) \in V$ *sind* **gleich***, wenn* $x_i = y_i$ *für alle* $i = 1, \ldots, \ell$ *gilt.*

Die Definition von $V_1 \oplus \ldots \oplus V_\ell$ entspricht der Definition von $\mathbb{R}^\ell$ in Kapitel 1. Ganz analog wie dort zeigt man, daß $V_1 \oplus \ldots \oplus V_\ell$ ein Vektorraum ist. Das Nullelement ist durch

$$0_{V_1 \oplus \ldots \oplus V_\ell} = (0_{V_1}, \ldots, 0_{V_\ell})$$

und das zu $(x_1, \ldots, x_\ell) \in \oplus_{i=1}^{\ell} V_i$ negative Element durch

$$-(x_1, \ldots, x_\ell) = (-x_1, \ldots, -x_\ell)$$

gegeben. Offensichtlich gilt

$$\mathbb{R}^\ell = \underbrace{\mathbb{R} \oplus \ldots \oplus \mathbb{R}}_{\ell\text{-mal}} \qquad \text{und} \qquad \mathbb{R}^{\ell+n} = \mathbb{R}^\ell \oplus \mathbb{R}^n.$$

Die Konstruktion der kanonischen Basis in $\mathbb{R}^\ell$ gibt die Idee, wie man aus Basen der (nun als endlichdimensional angenommenen) Vektorräume $V_1, \ldots, V_\ell$ zu einer Basis der direkten Summe gelangt.

4.2.2 Satz
Es seien $\left\{x_i^{(1)}, \ldots, x_i^{(n_i)}\right\} \subset V_i$ *Basen von* V_i, $n_i \in \mathbb{N}$, $i = 1, \ldots, \ell$. *Dann ist*

$$\begin{aligned} M &:= \left\{\left(0, \ldots, \underbrace{x_i^{(j)}}_{i\text{-te Stelle}}, \ldots, 0\right) : j = 1, \ldots, n_i,\, i = 1, \ldots, \ell\right\} \\ &= \left\{\left(x_1^{(1)}, 0, \ldots, 0\right), \ldots, \left(x_1^{(n_1)}, \ldots, 0\right), \ldots, \right. \\ &\qquad \left. \left(0, \ldots, 0, x_\ell^{(1)}\right), \ldots, \left(0, \ldots, 0, x_\ell^{(n_\ell)}\right)\right\} \end{aligned}$$

eine Basis von $V = V_1 \oplus \ldots \oplus V_\ell$. *Insbesondere hat* V *die Dimension*

$$\dim(V_1 \oplus \ldots \oplus V_\ell) = n_1 + \ldots + n_\ell.$$

Beweis

(1) Die Elemente von M sind linear unabhängig: Sei nämlich das Nullelement der direkten Summe eine Linearkombination aus Elementen von M:

$$\begin{aligned} \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} a_i^{(j)} \left(0, \ldots, x_i^{(j)}, \ldots, 0\right) &= \sum_{i=1}^{\ell} \left(0, \ldots, \sum_{j=1}^{n_i} a_i^{(j)} x_i^{(j)}, \ldots, 0\right) \\ &= (0, \ldots, 0) = 0_V \end{aligned}$$

mit $a_i^{(j)} \in \mathbb{R}$. Die Betrachtung der i-ten Komponente liefert

$$\sum_{j=1}^{n_i} a_i^{(j)} x_i^{(j)} = 0 \in V_i, \quad i = 1, \ldots, \ell.$$

Aus der linearen Unabhängigkeit von $x_i^{(1)}, \ldots, x_i^{(n_i)}$ in V_i folgt schließlich

$$a_i^{(1)} = \ldots = a_i^{(n_i)} = 0, \quad i = 1, \ldots, \ell.$$

(2) Die Elemente von M erzeugen V: Sei nämlich ein beliebiges Element der direkten Summe $(x_1, \ldots, x_\ell) \in V$ gegeben. Da V_i von $x_i^{(1)}, \ldots, x_i^{(n_i)}$ erzeugt wird, gibt es reelle Zahlen $a_i^{(1)}, \ldots, a_i^{(n_i)}$ mit

$$x_i = \sum_{j=1}^{n_i} a_i^{(j)} x_i^{(j)}, \quad i = 1, \ldots, \ell.$$

Es folgt die Darstellbarkeit mittels der Elemente aus M:

$$(x_1, \ldots, x_\ell) = \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} a_i^{(j)} \left(0, \ldots, x_i^{(j)}, \ldots, 0\right).$$

Die Dimension von V ergibt sich nun als die Länge der Basis M, vgl. Satz 3.3.15. □

Auch lineare Abbildungen von den direkten Summanden in einen gemeinsamen Wertevorrat lassen sich zu einer auf der direkten Summe definierten linearen Abbildung zusammenfügen:

4.2.3 Definition

Seien $F_i : V_i \to W$ $(i = 1, \ldots, \ell)$ lineare Abbildungen, es sei $V = V_1 \oplus \ldots \oplus V_\ell$. Die Abbildung $F : V \to W$, $F(x_1, \ldots, x_\ell) := F_1(x_1) + \ldots + F_\ell(x_\ell)$ heißt von $F_1, \ldots, F_\ell$ **erzeugte lineare Abbildung** *oder* **direkte Summe** *von $F_1, \ldots, F_\ell$. Man schreibt $F =: F_1 \oplus \ldots \oplus F_\ell =: \oplus_{i=1}^{\ell} F_i$.*

4.2.4 Bemerkung

Auf Grund der Linearität der einzelnen Abbildungen F_i ist F tatsächlich linear, denn für $c \in \mathbb{R}$, $(x_1, \ldots, x_\ell), (y_1, \ldots, y_\ell) \in V$ gilt:

$$\begin{aligned}
&F((x_1, \ldots, x_\ell) + (y_1, \ldots, y_\ell)) = F(x_1 + y_1, \ldots, x_\ell + y_\ell)\\
&= \sum_{i=1}^{\ell} F_i(x_i + y_i) = \sum_{i=1}^{\ell} (F_i(x_i) + F_i(y_i))\\
&= \sum_{i=1}^{\ell} F_i(x_i) + \sum_{i=1}^{\ell} F_i(y_i) = F(x_1, \ldots, x_\ell) + F(y_1, \ldots, y_\ell);
\end{aligned}$$

$$\begin{aligned}
&F(c \cdot (x_1, \ldots, x_\ell)) = F(c \cdot x_1, \ldots, c \cdot x_\ell) = \sum_{i=1}^{\ell} F_i(c \cdot x_i)\\
&= \sum_{i=1}^{\ell} c \cdot F_i(x_i) = c \cdot \left(\sum_{i=1}^{\ell} F_i(x_i)\right) = c \cdot F(x_1, \ldots, x_\ell).
\end{aligned}$$

□

Seien nun speziell $F_i : \mathbb{R}^{n_i} \to \mathbb{R}^m$, $i = 1, \ldots, \ell$, lineare Abbildungen mit zugehörigen $m \times n_i$-Matrizen A_i. Im Rahmen des Matrizenkalküls schreiben wir für die Elemente der direkten Summe $\oplus_{i=1}^{\ell} \mathbb{R}^{n_i}$ alle „Komponenten" aus $\mathbb{R}^{n_i}$ untereinander, d. h. in eine Spalte der Länge $\sum_{i=1}^{\ell} n_i$. Dann gehört zu der direkten Summe

$F = \oplus_{i=1}^{\ell} F_i$ die $m \times \left(\sum_{i=1}^{\ell} n_i\right)$-Matrix

$$A := \Big(A_1, \dots, A_\ell\Big),$$

diese nennt man auch die *direkte Summe* der Matrizen $A_1, \dots, A_\ell$.

In dem Fall, daß bei den linearen Abbildungen $F_i : V_i \to W_i$, $i = 1, \dots, \ell$ die Wertevorräte nicht notwendig übereinstimmen, lassen sich diese in ganz ähnlicher Weise zu einer Abbildung zwischen den entsprechenden direkten Summen der Vektorräume zusammensetzen:

$$\hat{F} : \oplus_{i=1}^{\ell} V_i \to \oplus_{i=1}^{\ell} W_i, \quad \hat{F}(x_1, \dots, x_\ell) := (F_1(x_1), \dots, F_\ell(x_\ell)) .$$

Sind speziell $V_i = \mathbb{R}^{n_i}$, $W_i = \mathbb{R}^{m_i}$, $i = 1, \dots, \ell$ und A_i die zu F_i gehörigen $m_i \times n_i$-Matrizen, so gehört zu $\hat{F}$ die $\left(\sum_{i=1}^{\ell} m_i\right) \times \left(\sum_{i=1}^{\ell} n_i\right)$-Matrix

$$\hat{A} = \begin{pmatrix} A_1 & \mathcal{O} & \dots & \mathcal{O} \\ \mathcal{O} & A_2 & \dots & \mathcal{O} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{O} & \mathcal{O} & \dots & A_\ell \end{pmatrix} .$$

Zerlegen

Wir kehren nun unseren Standpunkt um: Gegeben sei ein Vektorraum W, der auf „geeignete“ Weise als direkte Summe zweier – oder, was bis auf eine Iteration auf dasselbe hinausläuft, endlich vieler – Untervektorräume V_1, V_2 dargestellt werden soll. Dabei wird noch zu erklären sein, unter welchen Umständen man $V_1 \oplus V_2$ als Untervektorraum von W selbst auffassen kann. Da die Darstellung von Elementen $x \in V_1 \oplus V_2$ mittels ihrer Komponenten $x_i \in V_i$, $x = (x_1, x_2)$, *eindeutig* ist, läßt sich die Darstellung als direkte Summe $W = V_1 \oplus V_2$ als Weiterentwicklung bzw. als eine (im jeweiligen Kontext erwünschte) Vergröberung des Basisbegriffs ansehen, vgl. Satz 4.2.7. Gerade bei unendlichdimensionalen Vektorräumen, für die man häufig keine Basis explizit angeben kann (dazu vgl. den vorhergehenden Abschnitt 4.1), kommt solchen Zerlegungen eine große Bedeutung zu.

Häufig ist neben W bereits ein Untervektorraum V_1 gegeben, und es wird ein geeignetes „Komplement“ V_2 mit $W = V_1 \oplus V_2$ gesucht. So waren wir z. B. beim Beweis der Dimensionsformel in endlichdimensionalen Vektorräumen vorgegangen, wo wir V_1 als Kern einer linearen Abbildung F gewählt haben und dazu ein Komplement konstruiert haben, auf dem (die Einschränkung von) F injektiv geworden ist.

4.2.5 Satz

Seien V_1, V_2 Untervektorräume von W mit $V_1 \cap V_2 = \{0\}$. Wir betrachten die kanonischen Injektionsabbildungen $F_i :\ V_i \to W, F_i(x) = x$. Dann ist $F := F_1 \oplus F_2 : V_1 \oplus V_2 \to W$, $F(x_1, x_2) = x_1 + x_2$ injektiv.

Beweis

Weil F linear ist, reicht es zu zeigen: $\ker(F) = \{0_{V_1 \oplus V_2}\} = \{(0,0)\}$. Sei also $(x_1, x_2) \in \ker(F) \subset V_1 \oplus V_2$. Es ist $0 = F(x_1 + x_2) = F_1(x_1) + F_2(x_2) = x_1 + x_2$. Folglich ist $x_1 = -x_2 \in V_1 \cap V_2 = \{0\}$, d. h. $(x_1, x_2) = (0, 0)$. □

Unter der Voraussetzung $V_1 \cap V_2 = \{0\}$ können wir also $V_1 \oplus V_2$ und das Bild unter der kanonischen Abbildung $F : (x_1, x_2) \mapsto x_1 + x_2$ identifizieren und somit $V_1 \oplus V_2$ als Untervektorraum von W auffassen. Ist diese Abbildung F zusätzlich surjektiv, so erhält man einen kanonischen Isomorphismus $V_1 \oplus V_2 \leftrightarrow W$ und $V_1 \oplus V_2$ und W werden identifiziert. Dieses fassen wir in der folgenden Definition zusammen:

4.2.6 Definition

Seien V_1, V_2 Untervektorräume von W mit $V_1 \cap V_2 = \{0\}$. Die kanonische Abbildung $F : V_1 \oplus V_2 \to W,\ F(x_1, x_2) = x_1 + x_2$ sei surjektiv. Dann heißt W **direkte Summe** *von V_1 und V_2, und man schreibt $W = V_1 \oplus V_2$.*

4.2.7 Satz

Es sei $W = V_1 \oplus V_2$. Zu jedem $x \in V$ existieren eindeutig bestimmte Vektoren $x_1 \in V_1,\ x_2 \in V_2$ mit $x = x_1 + x_2$.

Beweis

Sei $x \in W$ beliebig. Da F bijektiv ist, existiert genau ein Paar $(x_1, x_2) \in V_1 \oplus V_2$ mit $x = F(x_1, x_2) = F_1(x_1) + F_2(x_2) = x_1 + x_2$. □

4.2.8 Beispiele

(a) Seien $V_1 = \left\{\vec{a} \in \mathbb{R}^3 :\ a_2 = a_3 = 0\right\}$ und $V_2 = \{\vec{a} :\ a_1 + 2a_2 + 3a_3 = 0\}$. Dann ist ganz offensichtlich $V_1 \cap V_2 = \left\{\vec{0}\right\}$. Zu beliebigem $\vec{x} \in \mathbb{R}^3$ bestimmt man ferner $\vec{a}_1 = (x_1 + 2x_2 + 3x_3, 0, 0) \in V_1$ und $\vec{a}_2 = (-2x_2 - 3x_3, x_2, x_3) \in V_2$ mit $\vec{x} = \vec{a}_1 + \vec{a}_2 = F(\vec{a}_1, \vec{a}_2)$. Also ist $\mathbb{R}^3 = V_1 \oplus V_2$.

(b) Sei $V = \{f : [-1, 1] \to \mathbb{R}\}$ der Vektorraum aller reellen Funktionen auf dem Intervall $[-1, 1]$, vgl. Beispiel 4.1.4. Wir betrachten die Untervektorräume

$$V_1 = \{f \in V :\ \text{für alle } x \text{ ist } f(x) = f(-x)\}$$

der geraden und

$$V_2 = \{f \in V :\ \text{für alle } x \text{ ist } f(x) = -f(-x)\}$$

der ungeraden Funktionen.

Ist $f \in V_1 \cap V_2$, also gleichzeitig gerade und ungerade, so ist für alle $x \in [-1, 1]$: $f(x) = f(-x) = -f(x)$, folglich $f(x) = 0$. Wir haben $V_1 \cap V_2 = \{\mathcal{O}\}$.

Eine beliebige Funktion $f \in V$ läßt sich in einen geraden Anteil $f_1(x) := \frac{1}{2}(f(x) + f(-x))$ und einen ungeraden Anteil $f_2(x) := \frac{1}{2}(f(x) - f(-x))$ zerlegen: Es ist $f = f_1 + f_2$, $f_i \in V_i$.

Insgesamt folgt $V = V_1 \oplus V_2$.

Wir wenden uns nun der schon oben angeschnittenen Frage nach der Existenz von „Komplementen" zu: Gegeben sei ein Vektorraum W und ein Untervektorraum V_1, gesucht ist ein weiterer Untervektorraum V_2 derart, daß W direkte Summe von V_1 und V_2 wird: $W = V_1 \oplus V_2$. Es liegt auf der Hand, daß bei der Lösung dieses Problems geeignete Basen hilfreich sein werden.

Damit die folgenden Beweise elementar bleiben und wir auf das Zornsche Lemma als transzendentes Hilfsmittel verzichten können, beschränken wir uns im Folgenden auf endlichdimensionale Vektorräume. Die Resultate gelten jedoch entsprechend auch im Unendlichdimensionalen.

Wir erinnern an den Basisergänzungssatz, den wir bereits beim Beweis der Dimensionsformel (s. Satz 3.4.4) gezeigt haben:

4.2.9 Satz (Basisergänzungssatz)

Sei W ein endlichdimensionaler Vektorraum, $\dim W = n$, und es seien linear unabhängige Vektoren $x_1, \ldots, x_s \in W$ gegeben. Dann existieren Vektoren $x_{s+1}, \ldots, x_n \in W$ derart, daß $\{x_1, \ldots, x_n\}$ eine Basis von W ist.

4.2.10 Satz (Existenz eines Komplements)

Es sei W ein endlichdimensionaler Vektorraum und $V_1 \subset W$ ein Untervektorraum. Dann gibt es einen Untervektorraum $V_2 \subset W$ derart, daß $W = V_1 \oplus V_2$ gilt.

Beweis

Da mit W auch V_1 endlichdimensional ist, existiert nach Satz 1.3.17 eine Basis $\{x_1, \ldots, x_s\}$ von V_1. Diese werde gemäß Satz 4.2.9 zu einer Basis $\{x_1, \ldots, x_s, x_{s+1}, \ldots, x_n\}$ von W ergänzt.

Wir setzen $V_2 := \mathbb{R} \cdot (x_{s+1}, \ldots, x_n)$ und zeigen zunächst $V_1 \cap V_2 = \{0\}$: Sei dazu $x \in V_1 \cap V_2$. Wegen $x \in V_1$ bzw. $x \in V_2$ existieren $a_1, \ldots, a_s \in \mathbb{R}$ bzw. $a_{s+1}, \ldots, a_n \in \mathbb{R}$ mit $x = a_1x_1 + \ldots + a_sx_s = a_{s+1}x_{s+1} + \ldots + a_nx_n$. Also ist $0 = a_1x_1 + \ldots + a_sx_s - a_{s+1}x_{s+1} - \ldots - a_nx_n$. Die lineare Unabhängigkeit von $x_1, \ldots, x_n$ liefert $a_1 = \ldots = a_n = 0$ und damit $x = 0$. Umgekehrt gilt offensichtlich $0 \in V_1 \cap V_2$.

Sei nun F wie in Satz 4.2.5. Wir weisen die Gültigkeit von $F(V_1 \oplus V_2) = W$ nach: Sei ein beliebiges $x \in W$ gegeben, es existieren $a_1, \ldots, a_n \in \mathbb{R}$ mit $x = a_1x_1 + \ldots + a_nx_n$. Setze $y_1 := a_1x_1 + \ldots + a_sx_s \in V_1$, $y_2 := a_{s+1}x_{s+1} + \ldots + a_nx_n \in V_2$. Es ist $x = y_1 + y_2 = F(y_1, y_2)$; also ist F surjektiv. $\square$

Das folgende Beispiel zeigt, daß das Komplement V_2 von V_1 (außer im Fall $V_1 = W$ oder $V_1 = \{0\}$) im allgemeinen nicht eindeutig bestimmt ist:

4.2.11 Beispiel

Wir kommen auf das Beispiel 4.2.8 (a) zurück. Seien $W = \mathbb{R}^3$ und $V_1 = \left\{\vec{a} \in \mathbb{R}^3 : a_2 = a_3 = 0\right\}$. Dann sind $V_2 = \left\{\vec{a} \in \mathbb{R}^3 : a_1 + 2a_2 + 3a_3 = 0\right\}$ und auch $\hat{V}_2 = \left\{\vec{a} \in \mathbb{R}^3 : a_1 = 0\right\}$ Komplemente von V_1 im Sinne des vorhergehenden Satzes: $W = V_1 \oplus V_2 = V_1 \oplus \hat{V}_2$.

Beim Verzicht auf die Voraussetzung $V_1 \cap V_2 = \{0\}$ können wir $V_1 \oplus V_2$ zwar nicht mehr als Untervektorraum von W ansehen, es ergibt sich aber dennoch eine wichtige Begriffsbildung:

4.2.12 Definition

Seien V_1, V_2 Untervektorräume von W und $F : V_1 \oplus V_2 \to W$ die durch $F(x_1, x_2) = x_1 + x_2$ gegebene kanonische Abbildung. Dann heißt $V_1 + V_2 := F(V_1 \oplus V_2) = \{x = x_1 + x_2 : x_1 \in V_1, x_2 \in V_2\}$ der von V_1 und V_2 **aufgespannte Untervektorraum** *von W.*

Weil F linear ist, ist $V_1 + V_2$ tatsächlich ein Untervektorraum. Im Gegensatz zur Situation von Satz 4.2.7 ist die Zerlegung der Elemente von $V_1 + V_2$ im allgemeinen nicht mehr eindeutig. Die Eindeutigkeit ist eine Folge der Bedingung $V_1 \cap V_2 = \{0\}$.

Diese Bildung $V_1 + V_2$ tritt an die Stelle der in der Vektorraumtheorie wenig sinnvollen Vereinigungsmenge, im allgemeinen ist nämlich $V_1 \cup V_2$ kein Untervektorraum. Der kleinste Untervektorraum, der $V_1 \cup V_2$ enthält, ist gerade der von V_1 und V_2 aufgespannte Untervektorraum $V_1 + V_2$.

4.3 Linearformen und Dualräume

Auch hier bezeichnen $V, V_1, V_2, \ldots$ wieder stets reelle Vektorräume.

Duale Vektorräume

4.3.1 Definition

(a) *Eine* **Linearform** *ist eine lineare reelle Funktion $L : V \to \mathbb{R}$.*

(b) *Es heißt $V^* := \{L : V \to \mathbb{R} : L \text{ ist Linearform}\}$ der zu V* **duale Vektorraum** *oder* **Dualraum** *von V.*

4.3.2 Bemerkung

Wir erinnern an den ersten Teil von Abschnitt 3.6. Danach erhält man für $L_1, L_2 \in V^*$, $c \in \mathbb{R}$ durch $(L_1 + L_2)(x) \overset{\text{def}}{=} L_1(x) + L_2(x)$, $(c \cdot L_1)(x) \overset{\text{def}}{=} c \cdot L_1(x)$ lineare reelle

Funktionen $V \to \mathbb{R}$, also Elemente des Dualraums: $L_1 + L_2 \in V^*$, $c \cdot L_1 \in V^*$, und V^* ist tatsächlich ein Vektorraum. Das Nullelement von V^* ist die Nullform $\mathcal{O} : V \to \mathbb{R},\ x \mapsto 0$. Das zu $L \in V^*$ negative Element ist durch $(-L)(x) = -(L(x))$ gegeben.

4.3.3 Beispiel

Sei V der Vektorraum der reellen Polynome (siehe Beispiel 4.1.3) mit der kanonischen Basis $\{e_0 = 1, e_1 = z, \ldots\}$. Daher ist V vermöge der Abbildung $(a_0, \ldots, a_n) \mapsto \sum_{i=0}^n a_i z^i$ kanonisch isomorph zum Vektorraum aller *endlichen* reellen Folgen $\{(a_0, \ldots, a_n) :\ n \in \mathbb{N}_0,\ a_i \in \mathbb{R}\}$.

Um den Dualraum V^* zu charakterisieren, erinnern wir an Satz 3.2.10, der entsprechend auch für unendlichdimensionale Vektorräume gilt. Demnach ist jede Linearform $L \in V^*$ durch die Werte $a_i = L(e_i),\ i \in \mathbb{N}_0$, eindeutig bestimmt.

Also ist V^* isomorph zum Vektorraum aller *unendlichen* reellen Folgen $W := \{(a_i)_{i \in \mathbb{N}_0} :\ a_i \in \mathbb{R}\}$ und damit wesentlich „größer“ als V. Genauer gilt, daß V eine abzählbar unendliche Basis hat, daß aber jede Basis von W und damit von V^* *überabzählbar* ist.

Wir orientieren uns am Beweis der Überabzählbarkeit von $\mathbb{R}$ und nehmen dazu an, es gäbe eine abzählbare Basis

$$\left\{ \left(b_i^{(j)} \right)_{i \in \mathbb{N}_0} :\ j \in \mathbb{N}_0 \right\}$$

von W.

Zunächst wählen wir Zahlen $(\nu_k)_{k \in \mathbb{N}_0}$, $\nu_k < \nu_{k+1}$, so daß für $k \in \mathbb{N}$ die Betrachtung der ersten $\nu_{k-1} + 1$ Komponenten bereits die lineare Unabhängigkeit der ersten k Basiselemente zeigt:

$$\dim \left(\mathbb{R} \cdot \left(\left(b_0^{(j)}, \ldots, b_{\nu_{k-1}}^{(j)} \right),\ j = 0, \ldots, k-1 \right) \right) = k.$$

Es ist $\nu_k \geq k$. Wir begründen am Ende dieses Beweises, daß man derartige Indizes ν_k tatsächlich finden kann.

Wir konstruieren induktiv die Glieder einer Folge $(c_i)_{i \in \mathbb{N}_0} \in W$, die nicht Linearkombination der Basiselemente ist und erhalten so einen Widerspruch. Im ersten Schritt wählen wir $c_0, \ldots, c_{\nu_1}$ und im k-ten $(k \geq 2)$ Schritt $c_{\nu_{k-1}+1}, \ldots, c_{\nu_k}$, so daß für alle $k \in \mathbb{N}$ gilt:

$$(c_0, \ldots, c_{\nu_k}) \notin \mathbb{R} \cdot \left(\left(b_0^{(j)}, \ldots, b_{\nu_k}^{(j)} \right),\ j = 0, \ldots, k-1 \right). \qquad (*)$$

Die Wahl von $c_0, c_1, \ldots, c_{\nu_1}$ ist trivial, da man mit mindestens zwei freien Parametern lediglich einen eindimensionalen Raum meiden muß. Seien also bereits $(c_0, \ldots, c_{\nu_k})$

gemäß $(*)$ gewählt. Wir nehmen nun $\left(b_i^{(k)}\right)_{i\in\mathbb{N}_0}$ als Erzeuger hinzu. Gilt auch

$$(c_0,\ldots,c_{\nu_k}) \notin \mathbb{R}\cdot\left(\left(b_0^{(j)},\ldots,b_{\nu_k}^{(j)}\right),\ j=0,\ldots,k-1,k\right),$$

so können wir $c_{\nu_k+1},\ldots,c_{\nu_{k+1}}$ beliebig wählen. Ist dagegen

$$(c_0,\ldots,c_{\nu_k}) \in \mathbb{R}\cdot\left(\left(b_0^{(j)},\ldots,b_{\nu_k}^{(j)}\right),\ j=0,\ldots,k-1,k\right),$$

so existieren *eindeutig* bestimmte $a_0,\ldots,a_k\in\mathbb{R}$ mit

$$c_i=\sum_{j=0}^{k} a_j b_i^{(j)},\quad i=0,\ldots,\nu_k.$$

Hier benutzen wir, daß schon die ersten ν_k+1 Komponenten die lineare Unabhängigkeit der ersten $(k+1)$-Basisvektoren zeigen. In diesem Fall wählen wir

$$c_{\nu_k+1}\neq\sum_{j=0}^{k} a_j b_{\nu_k+1}^{(j)}$$

und ggfs. $c_{\nu_k+2},\ldots,c_{\nu_{k+1}}$ beliebig. In jedem Fall ist dann $(*)$ erfüllt, wobei k durch $k+1$ ersetzt ist.

Für jedes $k\in\mathbb{N}$ ergibt sich damit

$$(c_i)_{i\in\mathbb{N}_0}\notin\mathbb{R}\cdot\left(\left(b_i^{(j)}\right)_{i\in\mathbb{N}_0},\ j=0,\ldots,k-1\right)$$

allein schon aus der Betrachtung der ersten (ν_k+1) Folgenglieder. Wir haben also ein Element von W gefunden, das nicht als Linearkombination der Elemente einer als abzählbar angenommenen Basis dargestellt werden kann. Wäre nämlich $(c_i)_{i\in\mathbb{N}_0}$ (endliche!) Linearkombination der Elemente $\left(b_i^{(j)}\right)_{i\in\mathbb{N}_0}$, $j\in\mathbb{N}_0$, so wäre $(c_i)_{i\in\mathbb{N}_0}$ für eine hinreichend große Zahl k auch Linearkombination der Elemente $\left(b_i^{(j)}\right)_{i\in\mathbb{N}_0}$, $j=0,\ldots,k-1$. Dieser Widerspruch zeigt die Überabzählbarkeit jeder Basis von W und damit von V^*.

Wir haben noch die Existenz geeigneter Zahlen ν_k zu zeigen, und nehmen dazu bei festgehaltenem k an, daß

$$\left(b_0^{(j)},\ldots,b_\nu^{(j)}\right),\quad j=0,\ldots,k-1,$$

für jedes $\nu\in\mathbb{N}$ linear abhängig sind. Dann existieren also Zahlen $a_j^{(\nu)}$, $j=0,\ldots,k-1$, die wir durch $\max_{j=0,\ldots,k-1}\left|a_j^{(\nu)}\right|=1$ normiert annehmen können, so daß für $i\leq\nu$ gilt: $\sum_{j=0}^{k-1}a_j^{(\nu)}b_i^{(j)}=0$. Indem man für die k Folgen $(a_j^{(\nu)})_{\nu\in\mathbb{N}}$

gemäß dem Satz von Bolzano-Weierstraß sukzessive k-mal Teilfolgen auswählt, kann man simultan Konvergenz erreichen: $\lim_{\ell\to\infty} a_j^{(\nu_\ell)} = a_j$ für eine geeignete Teilfolge ν_ℓ. Für wenigstens ein j gilt unendlich oft $\left|a_j^{(\nu_\ell)}\right| = 1$, also sind nicht alle $a_j = 0$. Wir erhalten $\sum_{j=0}^{k-1} a_j b_i^{(j)} = 0$ für alle i und damit einen Widerspruch zur linearen Unabhängigkeit von $\left(b_i^{(j)}\right)_{i\in\mathbb{N}_0}$, $j = 0, \ldots, k-1$. □

Direkte Summen in Dualräumen

Am Ende von Abschnitt 4.2 hatten wir ausführlich die Zerlegung von Vektorräumen in direkte Summen $V = V_1 \oplus V_2$ diskutiert. Wir zeigen nun, daß sich die Existenz derartiger Zerlegungen unmittelbar in entsprechenden Zerlegungen des Dualraums V^* niederschlägt.

4.3.4 Satz

Es seien V_1, V_2 Untervektorräume von V mit $V = V_1 \oplus V_2$. Dann gilt $V^* = V_1^* \oplus V_2^*$.

Beweis

Zunächst ist zu klären, in welchem Sinne V_i^*, $i = 1, 2$, als Untervektorräume von V^* aufgefaßt werden können. Es wird V_1^* behandelt, für V_2^* gilt Entsprechendes. Wir betrachten den Untervektorraum

$$W := \{L \in V^* : \ L(x_2) = 0 \text{ für alle } x_2 \in V_2\}$$

und die lineare Abbildung $G : W \to V_1^*, L \mapsto L|V_1$. Dabei ist $L|V_1$ die Einschränkung von L auf V_1. Diese Abbildung ist injektiv, denn für $L \in W$ folgt aus $L|V_1 = \mathcal{O}$, daß für alle $x \in V$, $x = x_1 + x_2$ mit $x_i \in V_i$, gilt:

$$L(x) = L(x_1) + L(x_2) = 0 + 0 = 0.$$

Folglich ist $L = \mathcal{O}$.

Außerdem ist G auch surjektiv, denn jedes $L \in V_1^*$ kann zu einem $\hat{L} \in W$ mit $G(\hat{L}) = L$ fortgesetzt werden: Jedes $x \in V$ gestattet eine *eindeutige* Zerlegung $x = x_1 + x_2$, $x_i \in V_i$ (siehe Satz 4.2.7). Wir setzen $\hat{L}(x) = L(x_1)$, $\hat{L}$ ist wegen der Eindeutigkeit der Zerlegung wohldefiniert. Es ist $\hat{L} \in V^*$, $\hat{L}|V_2 = 0$ und damit $\hat{L} \in W$. Weiter ist $G(\hat{L}) = \hat{L}|V_1 = L$. Also sind V_1^* und W kanonisch isomorph und werden identifiziert.

Die Gültigkeit von $V_1^* \cap V_2^* = \{\mathcal{O}\}$ läßt sich nun leicht einsehen: Ist $L \in V_1^* \cap V_2^*$ und damit sowohl $L|V_2 = \mathcal{O}$ als auch $L|V_1 = \mathcal{O}$, so ergibt sich für alle $x \in V$:

$$x = x_1 + x_2, \ x_i \in V_i \Rightarrow L(x) = L(x_1) + L(x_2) = 0 + 0 = 0.$$

Zu jedem $L \in V^*$ existieren schließlich $L_1 \in V_1^*$, $L_2 \in V_2^*$ mit $L = L_1 + L_2$: Jedes $x \in V$ gestattet nämlich eine *eindeutige* Zerlegung $x = x_1 + x_2$, $x_i \in V_i$. Durch die Setzung $L_1(x) := L(x_1)$, $L_2(x) := L(x_2)$ erhalten wir das Gewünschte. □

Duale Abbildungen

Wir betrachten nun auch lineare Abbildungen $F : V_1 \to V_2$ und wollen untersuchen, ob diese auf kanonische Weise Abbildungen zwischen den Dualräumen V_1^* und V_2^* erzeugen („induzieren"): Das Diagramm

$$\begin{tikzcd} V_1 \arrow[r, "F"] \arrow[dr, "F^*(L_2) := L_1 := L_2 \circ F"'] & V_2 \arrow[d, "L_2"] \\ & \mathbb{R} \end{tikzcd}$$

Duale Abbildung

zeigt, daß mit Hilfe der *linearen* Abbildung F jede Linearform $L_2 \in V_2^*$ zu einer Linearform $L_1 \in V_1^*$ zusammengesetzt werden kann.

4.3.5 Definition

Sei $F : V_1 \to V_2$ eine lineare Abbildung. Dann heißt

$$F^* : V_2^* \to V_1^*, \quad F^*(L_2) := L_2 \circ F \quad (L_2 \in V_2^*),$$

die zu F **duale Abbildung**.

Die duale Abbildung F^* wirkt in „umgekehrter Richtung" wie F, sie „holt Linearformen auf V_2 nach V_1 zurück". Im Spezialfall $V_1 = \mathbb{R}^n$, $V_2 = \mathbb{R}^m$ steht diese Definition der dualen Abbildung neben Definition 3.4.7. Deren Übereinstimmung zeigen wir unten in Satz 4.4.3.

4.3.6 Satz

Sei $F : V_1 \to V_2$ eine lineare Abbildung. Dann ist die duale Abbildung $F^* : V_2^* \to V_1^*$ ebenfalls linear.

Beweis

Seien $L_2, \hat{L}_2 \in V_2^*$, $c \in \mathbb{R}$ gegeben. Wir gehen ganz analog wie beim Beweis der Sätze 3.6.18 und 3.6.19 vor. Für alle $x \in V_1$ gilt:

$$\begin{aligned} F^*\left(L_2 + \hat{L}_2\right)(x) &= \left(\left(L_2 + \hat{L}_2\right) \circ F\right)(x) = \left(L_2 + \hat{L}_2\right)(F(x)) \\ &= L_2(F(x)) + \hat{L}_2(F(x)) = L_2 \circ F(x) + \hat{L}_2 \circ F(x) \\ &= \left(L_2 \circ F + \hat{L}_2 \circ F\right)(x) = \left(F^*\left(L_2\right) + F^*\left(\hat{L}_2\right)\right)(x) \end{aligned}$$

und

$$\begin{aligned} F^*(c \cdot L_2)(x) &= ((c \cdot L_2) \circ F)(x) = (c \cdot L_2)(F(x)) = c \cdot L_2(F(x)) \\ &= c \cdot ((L_2 \circ F)(x)) = (c \cdot (L_2 \circ F))(x) = (c \cdot F^*(L_2))(x). \end{aligned}$$

Also gilt

$$F^*\left(L_2 + \hat{L}_2\right) = F^*\left(L_2\right) + F^*\left(\hat{L}_2\right)$$

und

$$F^*(c \cdot L_2) = c \cdot F^*(L_2). \qquad \square$$

4.3.7 Satz

(a) $(id_V)^* = id_{V^*}$.

(b) *Es seien $F_1 : V_1 \to V_2$, $F_2 : V_2 \to V_3$ lineare Abbildungen. Dann gilt für $(F_2 \circ F_1)^* : V_3^* \to V_1^*$:*

$$(F_2 \circ F_1)^* = F_1^* \circ F_2^*. \qquad \text{(Transitives Gesetz)}$$

(c) *Ist $F : V_1 \to V_2$ bijektiv, so ist auch $F^* : V_2^* \to V_1^*$ bijektiv, und es gilt:*

$$(F^*)^{-1} = \left(F^{-1}\right)^*.$$

Beweis

(a) Für $L \in V^*$ gilt $L \circ id = L$.

(b) Auf Grund der Definition der dualen Abbildung gilt für alle $L \in V_3^*$:

$$\begin{aligned}(F_2 \circ F_1)^*(L) &= L \circ (F_2 \circ F_1) = (L \circ F_2) \circ F_1 = F_2^*(L) \circ F_1 \\ &= F_1^*(F_2^*(L)) = (F_1^* \circ F_2^*)\,(L).\end{aligned}$$

$$\begin{tikzcd} V_1 \arrow[r, "F_1"] \arrow[drr, "F_1^*(F_2^*(L))"'] & V_2 \arrow[r, "F_2"] \arrow[dr, "F_2^*(L)"] & V_3 \arrow[d, "L"] \\ & & \mathbb{R} \end{tikzcd}$$

Transitives Gesetz

(c) Gemäß (a) und (b) ist

$$id_{V_2^*} = (id_{V_2})^* = \left(F \circ F^{-1}\right)^* = \left(F^{-1}\right)^* \circ F^*$$

und

$$id_{V_1^*} = (id_{V_1})^* = \left(F^{-1} \circ F\right)^* = F^* \circ \left(F^{-1}\right)^*,$$

woraus sich zusammen die Behauptung ergibt. □

Der Bidualraum

Das Verfahren der Dualraumbildung läßt sich iterieren.

4.3.8 Definition

Der Dualraum des Dualraums $V^{**} := (V^*)^*$ heißt **Bidualraum** *des Vektorraums V.*

Diese Begriffsbildung mag zunächst sehr abstrakt und willkürlich erscheinen. Der nächste Satz wird allerdings zeigen, daß jeder Vektorraum auf kanonische Weise als Untervektorraum seines Bidualraums aufgefaßt werden kann. Ist diese kanonische Injektion auch surjektiv, so kann man $V = V^{**}$ identifizieren: V ist dann ebenso Dualraum von V^* wie umgekehrt, die Dualraumbildung verliert ihre ursprüngliche Asymmetrie. Um wieviel der Bidualraum größer ist als der Vektorraum selbst (bzw. sein Bild unter der kanonischen Injektion) und man damit von dem symmetrischen Idealbild entfernt ist, ist ein Maß für die „Kompliziertheit" des Vektorraums V. Im nächsten Abschnitt werden sich die endlichdimensionalen Vektorräume als die in diesem Sinne idealen Vektorräume herausstellen. Entsprechende Überlegungen, dann unter Einbeziehung von Norm und Stetigkeit, sind z. B. in der Funktionalanalysis von zentraler Wichtigkeit.

4.3.9 Satz und Definition

Durch $\Phi(x)(L) := L(x)$, $x \in V$, $L \in V^*$ *wird eine injektive lineare Abbildung* $\Phi : V \hookrightarrow V^{**}$, *die* **kanonische Injektion** *erklärt.*

Beweis

(a) Wir zeigen zunächst, daß Φ in V^{**} abbildet, daß also $\Phi(x)$ für beliebiges $x \in V$ eine Linearform auf V^* ist. Tatsächlich gilt für $L_1, L_2 \in V^*$, $c \in \mathbb{R}$ auf Grund der Definition der Summe und der Skalarmultiplikation von Abbildungen:

$$\Phi(x)(L_1 + L_2) = (L_1 + L_2)(x) = L_1(x) + L_2(x) = \Phi(x)(L_1) + \Phi(x)(L_2);$$
$$\Phi(x)(c \cdot L_1) = (c \cdot L_1)(x) = c \cdot L_1(x) = c \cdot \Phi(x)(L_1).$$

(b) Wir zeigen nun die Linearität der Abbildung Φ, daß also für alle $x, y \in V$, $c \in \mathbb{R}$ gilt:

$$\Phi(x + y) = \Phi(x) + \Phi(y), \quad \Phi(c \cdot x) = c \cdot \Phi(x).$$

Die Gültigkeit dieser Gleichungen wird für alle Argumente $L \in V^*$ gezeigt:

$$\begin{aligned}
\Phi(x + y)(L) &\overset{\text{def}}{=} L(x + y) \overset{L \text{ linear}}{=} L(x) + L(y) \\
&\overset{\text{def}}{=} \Phi(x)(L) + \Phi(y)(L) \overset{\text{def}}{=} (\Phi(x) + \Phi(y))(L); \\
\Phi(c \cdot x)(L) &\overset{\text{def}}{=} L(c \cdot x) \overset{L \text{ linear}}{=} c \cdot L(x) \\
&\overset{\text{def}}{=} c \cdot (\Phi(x)(L)) \overset{\text{def}}{=} (c \cdot \Phi(x))(L).
\end{aligned}$$

(c) Wegen der Linearität von Φ erhalten wir die Injektivität aus dem Nachweis von $x \neq 0 \Rightarrow x \notin \ker(\Phi)$.

Sei also $x \in V \setminus \{0\}$. Um $\Phi(x) \neq \mathcal{O}$ (Nullabbildung!) zu zeigen, konstruieren wir ein $L \in V^*$ mit $0 \neq \Phi(x)(L) = L(x)$. Der Basisergänzungssatz 4.2.9 gilt (ohne daß wir das hier beweisen) auch in unendlichdimensionalen Vektorräumen. Wir ergänzen also hier das linear unabhängige Element x zu einer Basis M von V. Mit Hilfe dieser Basis können wir genauso wie in Satz 3.2.10 eine lineare reelle Funktion $L : V \to \mathbb{R}$ konstruieren, indem wir die Werte von L auf der Basis M vorgeben und die Funktion linear von M nach V fortsetzen. Also gibt es eine Linearform $L \in V^*$ mit $L(x) = 1$ und $L(y) = 0$ für $y \in M \setminus \{x\}$. Es folgt

$$\Phi(x)(L) = L(x) = 1 \neq 0. \qquad \square$$

Die Beweisteile sind sehr verschieden. Der Nachweis von $\Phi(V) \subset V^{**}$ und der Linearität von Φ erfordert nur geschicktes Anwenden einiger Definitionen, kann also als „leicht“ erachtet werden. Für den Nachweis der Injektivität muß dagegen ein tieferliegender Satz über die Existenz geeigneter Basen verwendet werden, der hier so nicht bewiesen wurde.

4.3.10 Beispiel

Seien V, V^* wie in Beispiel 4.3.3, V ist isomorph zum Vektorraum aller endlichen und V^* isomorph zum Vektorraum aller unendlichen reellen Folgen. Außerdem haben wir gezeigt, daß V^* eine überabzählbare Basis besitzt, also ist V^{**} mit Hilfe „überabzählbar langer Folgen“ zu beschreiben. In einem hier nicht weiter zu präzisierenden Sinne ist also V^{**} noch einmal wesentlich „größer“ als V^* und damit insbesondere auch als V.

Im Falle endlichdimensionaler Vektorräume ist die Situation jedoch einfacher.

4.4 Dualräume endlichdimensionaler Vektorräume

Bei endlichdimensionalen Vektorräumen V wird es auf naheliegende Weise gelingen, aus einer Basis von V eine entsprechende Basis des Dualraums V^* zu konstruieren. Daraus werden wir den auf dem Hintergrund der dieses Kapitel einleitenden Bemerkungen wohl nicht sonderlich überraschenden Schluß ziehen, daß endlichdimensionale Vektorräume und ihre Dualräume und damit auch ihre Bidualräume isomorph sind.

Als Folgerung werden wir den Matrizenkalkül auf die dualen Abbildungen ausdehnen können.

In diesem Abschnitt bezeichnen $V, V_1, V_2, \ldots$ stets endlichdimensionale Vektorräume, für die Dimension von V schreiben wir $\dim V =: n < \infty$.

Duale Basen

Wir erinnern an die Definition des Kroneckersymbols

$$\delta_{ij} = \begin{cases} 1, & \text{falls } i = j, \\ 0, & \text{falls } i \neq j. \end{cases}$$

4.4.1 Satz und Definition

Es sei $\{x_1, \ldots, x_n\} \subset V$ *eine Basis von* V. *Dann gibt es Linearformen* $x_1^*, \ldots, x_n^* : V \to \mathbb{R}$ *mit*

$$x_i^*(x_j) = \delta_{ij}, \quad i, j = 1, \ldots, n,$$

und $\{x_1^*, \ldots, x_n^*\}$ *ist eine Basis von* V^*. *Sie heißt die zu* $\{x_1, \ldots, x_n\}$ **duale Basis**.

Beweis

Gemäß Satz 3.2.10 existieren derartige Linearformen, deren Werte auf einer Basis von V vorgegeben werden.

(1) Wir zeigen zunächst die lineare Unabhängigkeit. Ist $a_1x_1^* + \ldots + a_nx_n^* = \mathcal{O}$ (Nullform!) mit $a_1, \ldots, a_n \in \mathbb{R}$, so folgt für alle $x \in V$:

$$0 = \mathcal{O}(x) = a_1x_1^*(x) + \ldots + a_1x_1^*(x) = \sum_{i=1}^{n} a_ix_i^*(x).$$

Wählt man spziell $x = x_j$, $j = 1, \ldots, n$, so erhält man auf Grund der Definition der x_i^*:

$$0 = \sum_{i=1}^{n} a_ix_i^*(x_j) = \sum_{i=1}^{n} a_i\delta_{ij} = a_j, \text{ also } a_1 = \ldots = a_n = 0.$$

(2) Die Linearformen $x_1^*, \ldots, x_n^*$ erzeugen auch V^*: Sei dazu $L \in V^*$ beliebig. Mit der Setzung $a_i := L(x_i) \in \mathbb{R}$, $i = 1, \ldots, n$ folgt auf Grund der Definition von Summe und skalaren Vielfachen von Linearformen für $j = 1, \ldots, n$:

$$L(x_j) = a_j = \sum_{i=1}^{n} a_i\delta_{ij} = \sum_{i=1}^{n} a_ix_i^*(x_j) = \left(\sum_{i=1}^{n} a_ix_i^*\right)(x_j).$$

Die Übereinstimmung beider Linearformen auf einer Basis von V ergibt:

$$L = \sum_{i=1}^{n} a_ix_i^*.$$

□

4.4.2 Folgerungen

(a) $\dim V = \dim V^* = \dim V^{**} = n$.

(b) Die kanonische Injektion $\Phi : V \to V^{**}$ ist sogar ein Isomorphismus, V und V^{**} können identifiziert werden.

Beweis

(a) Klar wegen Satz 4.4.1 und $V^{**} = (V^*)^*$.

(b) Die Surjektivität von Φ folgt unmittelbar aus der Dimensionsformel Satz 3.4.4.

□

Wegen $\dim V = \dim V^*$ sind V und V^* isomorph, im allgemeinen aber nicht kanonisch isomorph. Der durch Satz 4.4.1 nahegelegte Isomorphismus mit $x_i \mapsto x_i^*$ hängt von der Wahl der Basis ab, ist also im allgemeinen nicht kanonisch.

Nur wenn V mit einer kanonischen Basis $\{e_1, \ldots, e_n\}$ versehen ist (wie etwa der $\mathbb{R}^n$), so wird durch $e_i \mapsto e_i^*$ ein kanonischer Isomorphismus gegeben, und man kann $V = V^*$ schreiben.

Duale Abbildungen in den Koordinaten

Um lineare Abbildungen $H : V_1 \to V_2$ zwischen beliebigen endlichdimensionalen Vektorräumen dem Matrizenkalkül zugänglich zu machen, haben wir am Ende von Abschnitt 3.2 erläutert, wie man zu den Abbildungen in den Koordinaten gelangt. Sei dazu $\{x_1, \ldots, x_n\}$ eine Basis von V_1 und $\{y_1, \ldots, y_m\}$ eine Basis von V_2, $G_1 : \mathbb{R}^n \to V_1$ der durch $G_1(\vec{e}_i) = x_i$ und $G_2 : \mathbb{R}^m \to V_2$ der durch $G_2(\vec{e}_j) = y_j$ gegebene Isomorphismus (Karte, mit Umkehrabbildung als Koordinatensystem). Dann ist $F := G_2^{-1} \circ H \circ G_1$ die lineare Abbildung H in den Koordinaten.

$$\begin{tikzcd}
V_1 \arrow[r, "H"] & V_2 \arrow[d, "G_2^{-1}"] \\
\mathbb{R}^n \arrow[u, "G_1"] \arrow[r, "F = H \text{ in den Koordinaten}"'] & \mathbb{R}^m
\end{tikzcd}$$

Mit Hilfe von Satz 4.3.7 über die duale Abbildung von Kompositionen und Umkehrabbildungen erhält man daraus unmittelbar das dazu duale kommutative Diagramm:

$$\begin{tikzcd}
V_2^* \arrow[r, "H^*"] & V_1^* \arrow[d, "G_1^*"] \\
(\mathbb{R}^m)^* \arrow[u, "(G_2^*)^{-1}"] \arrow[r, "F^*"'] & (\mathbb{R}^n)^*
\end{tikzcd}$$

Wir wollen noch den Zusammenhang des dualen Koordinatensystems $G_1^* : V_1^* \to (\mathbb{R}^n)^*$ mit den dualen Basen $\{x_1^*, \ldots, x_n^*\}$ von V_1^* und $\{\vec{e}_1^*, \ldots, \vec{e}_n^*\}$ von $(\mathbb{R}^n)^* \cong \mathbb{R}^n$ studieren, für $G_2^* : V_2^* \to (\mathbb{R}^m)^* \cong \mathbb{R}^m$ gilt Entsprechendes.

Für $i, j = 1, \ldots, n$ ist

$$G_1^*(x_i^*)(\vec{e}_j) = (x_i^* \circ G_1)(\vec{e}_j) = x_i^*(G_1(\vec{e}_j)) = x_i^*(x_j) = \delta_{ij} = \vec{e}_i^*(\vec{e}_j),$$

also

$$G_1^*(x_i^*) = \vec{e}_i^*,\ i = 1, \ldots, n; \quad G_1^*(y_j^*) = \vec{e}_j^*,\ j = 1, \ldots, m.$$

Bis auf die Umkehr der Richtung gehen also die *dualen Koordinatensysteme* G_1^*, G_2^* auf dieselbe Weise aus den dualen Basen hervor wie die ursprünglichen Koordinatensysteme aus den Basen der ursprünglichen Vektorräume. Wegen $(\mathbb{R}^m)^* \cong \mathbb{R}^m$, $(\mathbb{R}^n)^* \cong \mathbb{R}^n$ ist also F^* die duale Abbildung H^* in den dualen Koordinaten.

In diesem Sinne spezialisieren wir nun auf lineare Abbildungen $F : \mathbb{R}^n \to \mathbb{R}^m$ und deren duale Abbildungen $F^* : \mathbb{R}^m \to \mathbb{R}^n$.

Duale Abbildungen $\mathbb{R}^m \to \mathbb{R}^n$

Wir betrachten $\mathbb{R}^n$ bzw. $(\mathbb{R}^n)^*$ mit der kanonischen Basis $\{\vec{e}_1, \ldots, \vec{e}_n\}$ bzw. $\{\vec{e}_1^*, \ldots, \vec{e}_n^*\}$ und beschreiben die durch $\vec{e}_i \mapsto \vec{e}_i^*$, $i = 1, \ldots, n$, gegebenen kanonischen Isomorphismen $G : \mathbb{R}^n \to (\mathbb{R}^n)^*$ und $G^{-1} : (\mathbb{R}^n)^* \to \mathbb{R}^n$ explizit. Ist $L \in (\mathbb{R}^n)^*$, so gilt gemäß dem Beweis von Satz 4.4.1:

$$L = \sum_{i=1}^{n} b_i \vec{e}_i^* \quad \text{mit } b_i = L\left(\vec{e}_i\right),\ i = 1, \ldots, n.$$

Also gilt

$$G : (b_1, \ldots, b_n) \mapsto \sum_{i=1}^{n} b_i \vec{e}_i^* \quad \text{und} \quad G^{-1} : L \mapsto \left(L\left(\vec{e}_1\right), \ldots, L\left(\vec{e}_n\right)\right).$$

Im Folgenden sind unter Linearformen auf $\mathbb{R}^n$ immer n-tupel $(b_1, \ldots, b_n)$ zu verstehen. Die Anwendung von $(b_1, \ldots, b_n)$ auf ein Argument $\vec{x} \in \mathbb{R}^n$ ergibt:

$$(b_1, \ldots, b_n)(\vec{x}) \stackrel{\text{def}}{=} \sum_{i=1}^{n} b_i \vec{e}_i^*(\vec{x}) = \sum_{i,j=1}^{n} b_i x_j \vec{e}_i^*(\vec{e}_j) = \sum_{i=1}^{n} b_i x_i.$$

Sei nun $F : \mathbb{R}^n \to \mathbb{R}^m$ eine lineare Abbildung, $A = (a_{ij})_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}}$ die zugehörige $m \times n$-Matrix. Wir wollen jetzt zeigen, daß zu der dualen Abbildung $F^* : \mathbb{R}^m \to \mathbb{R}^n$ (im Sinne von Definition 4.3.5) die transponierte Matrix A^t gehört und sich so eine Übereinstimmung mit der bereits für den Spezialfall getroffenen Definition 3.4.7 ergibt.

Sei $\vec{b} = (b_1, \ldots, b_m) \in \mathbb{R}^m = (\mathbb{R}^m)^*$ gegeben. Für die i-te Komponente $(F^*(b_1, \ldots, b_m))_i$ von $F^*(b_1, \ldots, b_m) \in \mathbb{R}^n$ gilt:

$$\begin{aligned}(F^*(b_1, \ldots, b_m))_i &= F^*(b_1, \ldots, b_m)\left(\vec{e}_i\right) = (b_1, \ldots, b_m)\left(F\left(\vec{e}_i\right)\right) \\ &= \sum_{j=1}^{m} b_j \cdot F(\vec{e}_i)_j = \sum_{j=1}^{m} b_j \cdot a_{ji}.\end{aligned}$$

Also ist

$$F^*\left(\vec{b}\right) = \left(a_{ji}\right)_{\substack{i=1,\ldots,n \\ j=1,\ldots m}} \circ \vec{b} = A^t \circ \vec{b}.$$

4.4.3 Satz

Sei $F : \mathbb{R}^n \to \mathbb{R}^m$ eine lineare Abbildung und A die zugehörige $m \times n$-Matrix. Dann gehört zur (im Sinne von Definition 4.3.5) dualen Abbildung $F^* : \mathbb{R}^m \to \mathbb{R}^n$ die transponierte $n \times m$-Matrix A^t.

In Kapitel 7 werden wir den Abbildungen F besondere Aufmerksamkeit widmen, bei denen Urbildraum und Wertevorrat gleich sind und die mit ihrer dualen Abbildung F^* übereinstimmen, die also durch symmetrische Matrizen gegeben werden.

4.4.4 Folgerungen

(a) *Sei A eine $m \times n$-Matrix, B eine $\ell \times m$-Matrix. Dann gilt $(B \circ A)^t = A^t \circ B^t$.*

(b) *Stets gilt $\mathrm{Rang}\,(A) = \mathrm{Rang}\,(A^t)$.*

Beweis

(a) Sei $F : \mathbb{R}^n \to \mathbb{R}^m$ bzw. $G : \mathbb{R}^m \to \mathbb{R}^\ell$ die zu A bzw. B gehörige lineare Abbildung. Dann gehört $B \circ A$ zu $G \circ F$, $(B \circ A)^t$ zu $(G \circ F)^*$, $A^t \circ B^t$ zu $F^* \circ G^*$. Die Behauptung folgt nun aus $(G \circ F)^* = F^* \circ G^*$, siehe Satz 4.3.7.

(b) Siehe Satz 3.4.14. □

In Verallgemeinerung hiervon und von Folgerung 3.4.15 erhält man durch Betrachtung von Koordinatensystemen:

4.4.5 Folgerung

Es seien V_1, V_2 endlichdimensionale Vektorräume, $F : V_1 \to V_2$ eine lineare Abbildung und $F^* : V_2^* \to V_1^*$ die dazu duale Abbildung. Dann gilt:

$$\mathrm{rg}(F) = \mathrm{rg}(F^*).$$

Für die in Abschnitt 3.6 eingeführte Matrizenaddition und Skalarmultiplikation haben wir schließlich:

4.4.6 Satz

Seien $m \times n$-Matrizen A, B und eine reelle Zahl a gegeben. Dann gilt $(A + B)^t = A^t + B^t$, $(a \cdot A)^t = a \cdot A^t$.

Beweis

Der Beweis folgt sofort aus der Definition der transponierten Matrix. Entsprechendes gilt gemäß Satz 4.4.3 für die zugehörigen linearen Abbildungen, dabei beachte man: Gehört $F : \mathbb{R}^n \to \mathbb{R}^m$ zu A und $G : \mathbb{R}^n \to \mathbb{R}^m$ zu B, so gehört $F + G$ zu $A + B$ und $a \cdot F$ zu $a \cdot A$. Es folgt:

$$(F + G)^* = F^* + G^*, \qquad (a \cdot F)^* = a \cdot F^*.$$ □

Die Abbildung $F \mapsto F^*$ ist also ebenfalls linear.

4.5 Weitere spezielle lineare Abbildungen

In diesem Abschnitt werden die Ausführungen von Kapitel 3.8 um solche Beispiele ergänzt, deren Behandlung intensiven Gebrauch transponierter Matrizen und der in Abschnitt 4.4 hergeleiteten Gesetze erfordert.

Die orthogonale Gruppe

Die Anschauung von Länge, Winkel und insbesondere Orthogonalität behandeln wir erst in Kapitel 6. Mit der Betrachtung der „dualen Paare“ $(\mathbb{R}^n, (\mathbb{R}^n)^*) \cong (\mathbb{R}^n, \mathbb{R}^n)$ und vor allem der Auszeichnung der kanonischen Basen $\{\vec{e_1}, \ldots, \vec{e_n}\}$ bzw. $\{\vec{e_1}^*, \ldots, \vec{e}_n^*\}$ haben wir implizit bereits einen abstrakten Orthogonalitätsbegriff eingeführt. Dazu vgl. man unten die Definition 4.5.5 des kanonischen Skalarprodukts und die Wirkung einer Linearform $(b_1, \ldots, b_n) \in (\mathbb{R}^n)^* \cong \mathbb{R}^n$ auf einen Vektor $\vec{x} \in \mathbb{R}^n$: $(b_1, \ldots, b_n)(\vec{x}) = \sum_{i=1}^n b_i x_i$. In abstrakten Situationen kann man häufig kein Skalarprodukt einführen. Dann ist tatsächlich oft die Betrachtung „dualer Paare“ (V, V^*) und „dualer Paarungen“ $\langle x, L\rangle := L(x)$, $(x \in V,\ L \in V^*)$ ein hinreichender Ersatz für die fehlende Orthogonalität.

4.5.1 Definition

Eine $n \times n$-Matrix A heißt **orthogonal**, *wenn $A^t \circ A = E$ gilt. Die Menge aller orthogonalen Matrizen $O(n) := O(n, \mathbb{R}) := \{A : A \text{ ist orthogonale } n \times n\text{-Matrix}\}$ heißt* **orthogonale Gruppe** *vom Rang n über $\mathbb{R}$.*

Letztere Bezeichnung ist gerechtfertigt durch:

4.5.2 Satz

$O(n, \mathbb{R})$ ist eine Untergruppe von $GL(n, \mathbb{R})$.

Beweis

Zunächst ist jedes $A \in O(n, \mathbb{R})$ nichtsingulär mit $A^{-1} = A^t$, dazu beachte man Satz 3.8.4. Folglich gilt $O(n, \mathbb{R}) \subset GL(n, \mathbb{R})$.

Offensichtlich ist $E \in O(n, \mathbb{R})$ und deshalb $\emptyset \neq O(n, \mathbb{R})$.

Die orthogonale Gruppe ist bezüglich der Matrizenmultiplikation abgeschlossen, denn für $A, B \in O(n, \mathbb{R})$ gilt nach Folgerung 4.4.4:

$$(A \circ B)^t \circ (A \circ B) = B^t \circ (A^t \circ A) \circ B = B^t \circ E \circ B = B^t \circ B = E.$$

Schließlich ist auch das Inverse jeder orthogonalen Matrix A wieder orthogonal. Denn für A^{-1} gilt wegen $A^{-1} = A^t$:

$$(A^{-1})^t \circ A^{-1} = (A^t)^t \circ A^t = A \circ A^t = A \circ A^{-1} = E. \qquad \square$$

Die orthogonalen 2×2-Matrizen lassen sich auf elementare Weise explizit beschreiben:

4.5.3 Satz

Sei

$$M_1 := \left\{ \begin{pmatrix} a & b \\ -b & a \end{pmatrix} : \ a, b \in \mathbb{R},\ a^2 + b^2 = 1 \right\}$$

und

$$M_2 := \left\{ \begin{pmatrix} a & b \\ b & -a \end{pmatrix} : \ a, b \in \mathbb{R},\ a^2 + b^2 = 1 \right\}.$$

Dann gilt:

(a) $O(2, \mathbb{R}) = M_1 \cup M_2$.

(b) M_1 *ist eine Untergruppe von* $O(2, \mathbb{R})$.

Wegen $E \notin M_2$ ist M_2 dagegen keine Untergruppe von $O(2, \mathbb{R})$.

Beweis

(a) Für $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{R}^{2\times 2}$ gilt $A^t \circ A = \begin{pmatrix} a^2 + c^2 & ab + cd \\ ab + cd & b^2 + d^2 \end{pmatrix}$. Also ist $A \in O(2, \mathbb{R})$ genau dann, wenn die *Orthogonalitätsbedingungen*

$$a^2 + c^2 = 1, \tag{1}$$

$$b^2 + d^2 = 1, \tag{2}$$

$$ab + cd = 0 \tag{3}$$

gelten.

Die Inklusion „$M_1 \cup M_2 \subset O(2, \mathbb{R})$“ gilt dann ganz offensichtlich, und es bleibt der Nachweis von „$M_1 \cup M_2 \supset O(2, \mathbb{R})$“. Dazu untersuchen wir für $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in O(2, \mathbb{R})$ das *nichtlineare* Gleichungssytem (1)-(3).

1. Fall: $a \neq 0$. Aus (3) folgt $b = -\frac{cd}{a}$. Einsetzen in (2) liefert unter Verwendung von (1):

$$1 = b^2 + d^2 = \frac{d^2(a^2 + c^2)}{a^2} = \frac{d^2}{a^2}.$$

Ist $a = d$, so folgt aus (3) $c = -b$ und damit $A = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ und $a^2 + b^2 = a^2 + c^2 = 1$, also $A \in M_1$. Ist hingegen $a = -d$, so gilt $b = c$ und folglich $A = \begin{pmatrix} a & b \\ b & -a \end{pmatrix}$, also $A \in M_2$.

2. Fall: $a = 0$. Es folgt $c^2 = 1$, $c \cdot d = 0$, d. h. $d = 0$ und $1 = b^2 + d^2 = b^2$. Also gilt $c \in \{-1, 1\}$ und $b \in \{-1, 1\}$. Es gibt vier Möglichkeiten, diese Werte für b und c zu kombinieren:

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in M_1, \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \in M_2.$$

(b) Offenbar enthält M_1 die Matrix E und mit A auch $A^{-1} = A^t$. Für $A = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \in M_1$, $A' = \begin{pmatrix} a' & b' \\ -b' & a' \end{pmatrix} \in M_1$ gilt

$$A \circ A' = \begin{pmatrix} aa' - bb' & ab' + a'b \\ -ab' - a'b & aa' - bb' \end{pmatrix} \tag{$*$}$$

und

$$\begin{aligned} &(aa' - bb')^2 + (ab' + a'b)^2 \\ &= a^2 a'^2 - 2aa'bb' + b^2 b'^2 + a^2 b'^2 + 2aa'bb' + a'^2 b^2 \\ &= a^2(a'^2 + b'^2) + b^2(a'^2 + b'^2) = a^2 + b^2 = 1, \end{aligned}$$

also $A \circ A' \in M_1$. □

4.5.4 Bemerkungen

(a) $O(2, \mathbb{R})$ ist nicht abelsch, denn es ist

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \circ \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \circ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

An der Gleichung $(*)$ aus dem obigen Beweis erkennt man, daß dagegen die Untergruppe M_1 von $O(2, \mathbb{R})$ abelsch ist.

(b) Die Elemente der Untergruppe M_1 sind genau von der Form:

$$A = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}. \tag{$+$}$$

In Kapitel 6 werden wir sehen, daß diese Matrizen genau den *Drehungen* der anschaulichen Ebene entsprechen.

Beweis

Matrizen von der Form $(+)$ sind offenbar Elemente von M_1.

Sei umgekehrt $A = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \in M_1$, insbesondere gilt $|a| \leq 1$. Also gibt es ein $\tilde{\alpha} \in \mathbb{R}$ mit $\cos\tilde{\alpha} = a$; es folgt $1 = \cos^2\tilde{\alpha} + b^2$, also $b^2 = \sin^2\tilde{\alpha}$. Im Fall $b = -\sin\tilde{\alpha}$ wählen wir $\alpha := \tilde{\alpha}$. Im Fall $b = \sin\tilde{\alpha}$ setzen wir $\alpha := -\tilde{\alpha}$ und erhalten wegen $\cos\alpha = \cos(-\alpha)$, $\sin(-\alpha) = -\sin\alpha$ die gewünschte Form $(+)$ für A. □

(c) In Abschnitt 3.8 wurde erwähnt, daß durch $C := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ eine Spiegelung an der Winkelhalbierenden $\{(x_1, x_2) \in \mathbb{R}^2 : x_1 = x_2\}$ gegeben ist. Für $A = \begin{pmatrix} a & b \\ b & -a \end{pmatrix} \in M_2$ gilt $C \circ A = \begin{pmatrix} b & -a \\ a & b \end{pmatrix}$, also $C \circ A \in M_1$. Wegen $C^2 = E$ folgt $A = C \circ (C \circ A)$ mit $C \circ A \in M_1$. Jede Matrix aus M_2 beschreibt also die Zusammensetzung einer Drehung und einer Spiegelung, also eine *Drehspiegelung*.

Zum Abschluß wollen wir die für den zweidimensionalen Fall formulierten „Orthogonalitätsbedingungen" aus dem Beweis von Satz 4.5.3 auf beliebige Dimensionen ausdehnen. Dazu definieren wir (ganz formal) das kanonische Skalarprodukt.

4.5.5 Definition

Durch $\vec{x} \cdot \vec{y} := \sum_{i=1}^n x_i y_i,\ \vec{x}, \vec{y} \in \mathbb{R}^n$, *wird das* **kanonische Skalarprodukt** $\mathbb{R}^n \oplus \mathbb{R}^n \to \mathbb{R}$ *gegeben.*

Da wir die Anschauung von Orthogonalität noch nicht besprochen haben, interpretieren wir das Skalarprodukt hier als Auswertung der Linearform $\vec{x}$ im Punkte $\vec{y}$.

4.5.6 Bemerkung

Für jede orthogonale $n \times n$-Matrix $A = (\vec{a}_1, \dots, \vec{a}_n)$ hat man

$$E = A^t \circ A = \begin{pmatrix} \vec{a}_1^{\,t} \\ \vdots \\ \vec{a}_n^{\,t} \end{pmatrix} \circ (\vec{a}_1, \dots, \vec{a}_n) = (\vec{a}_i \cdot \vec{a}_j)_{\substack{i=1,\dots,n \\ j=1,\dots,n}},$$

das heißt, es gelten die *Orthogonalitätsbedingungen*

$$\vec{a}_i \cdot \vec{a}_j = \delta_{ij} = \begin{cases} 1, & \text{falls } i = j, \\ 0, & \text{falls } i \neq j. \end{cases}$$

Die Spalten von A bilden ein sogenanntes „Orthonormalsystem". Entsprechendes gilt für die Zeilen, da $A^{-1} = A^t$ auch eine orthogonale Matrix ist.

Die Lorentzgruppe

4.5.7 Definition

Sei

$$H := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Dann heißt $L := \{A \in \mathbb{R}^{4\times 4} : A^t \circ H \circ A = H\}$ **Lorentzgruppe**. *Die linearen Abbildungen* $\vec{x} \mapsto A \circ \vec{x}$, $A \in L$, *heißen* **Lorentztransformationen**.

4.5.8 Satz

Die Lorentzgruppe L *ist eine Untergruppe von* $GL(4, \mathbb{R})$.

Beweis

Sicher ist $E \in L$, also ist L nichtleer.

Wir verwenden im Folgenden wiederholt $H = H^t$ und $H^2 = E$, d. h. $H^{-1} = H$.

Für $A \in L$ ist $A^t \circ H \circ A = H$ und folglich

$$(H \circ A^t \circ H) \circ A = H \circ (A^t \circ H \circ A) = H \circ H = E.$$

Somit ist A auf Grund von Satz 3.8.4 nichtsingulär mit

$$A^{-1} = H \circ A^t \circ H.$$

Insbesondere ist die Lorentzgruppe Teilmenge von $GL(4, \mathbb{R})$.

Diese ist auch bezüglich der Matrizenmultiplikation abgeschlossen, denn für $A, B \in L$ gilt $(A \circ B)^t \circ H \circ (A \circ B) = B^t \circ A^t \circ H \circ A \circ B = B^t \circ H \circ B = H$, das heißt $A \circ B \in L$.

Schließlich ist mit $A \in L$ auch $A^{-1} = H \circ A^t \circ H \in L$, denn man hat

$$\begin{aligned}(A^{-1})^t \circ H \circ A^{-1} &= (H \circ A^t \circ H)^t \circ H \circ (H \circ A^t \circ H) \\ &= H \circ A \circ H \circ H \circ H \circ A^t \circ H = H \circ A \circ (H \circ A^t \circ H) \\ &= H \circ A \circ A^{-1} = H.\end{aligned}$$

□

Die Lorentzgruppe ist in der Relativitätstheorie von Bedeutung. Im physikalischen Raum sei ein „ruhendes“ Bezugssystem (Inertialsystem) K gegeben. Jedem Punkt P des physikalischen Raums sind dann bezüglich K Koordinaten t, x_1, x_2, x_3 zugeordnet, dabei ist t die Zeitkoordinate und x_1, x_2, x_3 sind die Ortskoordinaten. In der Relativitätstheorie erweist es sich als nützlich, statt t das Vielfache $x_0 = ct$ als „Zeitkoordinate“ zu nehmen, wobei $c \approx 3 \cdot 10^8$m/sec die Lichtgeschwindigkeit bedeutet. Vermöge K haben wir eine bijektive Abbildung

$$\text{physikalischer Raum} \to \mathbb{R}^4, \qquad P \mapsto (x_0, x_1, x_2, x_3).$$

Sei jetzt K' ein anderes Koordinatensystem, das sich relativ zu K mit konstanter Geschwindigkeit v bewegt. Einem Punkt P sind dann bzgl. K' Koordinaten (y_0, y_1, y_2, y_3) zugeordnet.

In der speziellen Relativitätstheorie gilt nun, anders als in der klassischen Mechanik:

$$\vec{y} = A \circ \vec{x} \text{ mit einer geeigneten Matrix } A \in L,$$

wobei $\vec{x} = (x_0, \dots, x_3)$ die Koordinaten von P bzgl. K und $\vec{y} = (y_0, \dots, y_3)$ die Koordinaten von P bzgl. K' sind.

Man betrachte etwa den Fall, daß sich K' mit der Geschwindigkeit v relativ zu K in x_1-Richtung bewegt und die y_i-Achse jeweils parallel zur x_i-Achse ist $(i = 1, 2, 3)$. Dann hat A die Gestalt

$$A = \left(\begin{array}{c|c} A'^{2,2} & \mathcal{O}^{2,2} \\ \hline \mathcal{O}^{2,2} & E^{2,2} \end{array} \right),$$

wobei

$$A' = \frac{1}{\sqrt{1-\beta^2}} \begin{pmatrix} 1 & -\beta \\ -\beta & 1 \end{pmatrix} \text{ mit } \beta = \frac{v}{c}$$

ist. Wegen $|v| < c$ ist $\beta^2 < 1$, und $\sqrt{1-\beta^2}$ existiert als reelle Zahl.

Eine Herleitung dieses Ergebnisses findet man etwa in M. v. Laue, *Die Relativitätstheorie*, Erster Band, 7. Auflage, Kapitel III, § 7.

5 Determinantentheorie

Mit der Determinante werden wir eine Abbildung von dem Raum der n-reihigen quadratischen Matrizen $\mathbb{R}^{n\times n}$ in die reellen Zahlen $\mathbb{R}$ einführen, mit deren Hilfe wir entscheiden können, ob gegebene Matrizen singulär oder nichtsingulär (regulär, invertierbar) sind. Zwar sind wir zu Beginn des Abschnitts „Umkehrbar lineare Abbildungen" in Kapitel 3.8 bereits ausführlich auf die Frage nach der Invertierbarkeit quadratischer Matrizen eingegangen und haben dort mit dem Gaußschen Algorithmus ein sehr effizientes Verfahren angegeben, wie man die Invertierbarkeit einer *fest vorgegebenen* Matrix prüfen und ggfs. deren inverse Matrix berechnen kann. Für subtilere Fragen vor allem auch analytischer Natur wie z. B.

- Bleibt die Invertierbarkeit einer Matrix bei „kleinen" Änderungen der Komponenten der Matrix erhalten?
- Hängt im Falle der Existenz die inverse Matrix A^{-1} stetig von den Komponenten von A ab?
- Wie sieht die „Mannigfaltigkeit der singulären Matrizen" in der Menge aller quadratischen Matrizen $\mathbb{R}^{n\times n}$ aus und welchen Raum nimmt sie dort ein?

ist der Gaußsche Algorithmus nicht (ohne weiteres) geeignet. Dazu werden wir mit der Determinante als Polynom in den n^2 Matrixkomponenten ein adäquates Instrument vorstellen.

Wir werden in Kapitel 6 zeigen, daß der Determinante neben dieser analytischen auch eine große anschaulich geometrische Bedeutung zukommt (im Zusammenhang mit Flächeninhalt bzw. allgemein Volumen, Kreuzprodukt, Orientierung, . . .).

Mit Hilfe von Karten und des überaus wichtigen Multiplikationssatzes lassen sich Determinanten schließlich leicht für lineare Abbildungen $F : V \to V$ von endlichdimensionalen Vektorräumen in sich erklären.

5.1 Allgemeines Kroneckersymbol

Wie bereits einleitend angedeutet, werden wir die Determinante als Polynom in den Matrixkomponenten erklären. In diesem Abschnitt führen wir eine Vorzeichenfunktion für die einzelnen Summanden dieses Determinanten-Polynoms ein.

Dazu betrachten wir das folgende reelle Polynom in den n Veränderlichen $x_1, \ldots, x_n$:

$$\begin{aligned} D(x_1, \ldots, x_n) &:= \prod_{1 \le i < j \le n} (x_j - x_i), && \text{falls } n \ge 2, \\ D(x_1) &:= 1, && \text{falls } n = 1. \end{aligned}$$

Dabei bedeutet das Produktzeichen (großes Pi) $\prod_{1 \le i < j \le n}$, daß das Produkt über alle Indexpaare $i, j \in \{1, \ldots, n\}$ zu erstrecken ist, für die $i < j$ gilt.

Wie in Beispiel 1.3.13 sind $x_1, \ldots, x_n$ Veränderliche (Unbestimmte, Variable), und wir können die algebraische Definition dieses Polynoms gleichwertig neben der analytischen verwenden. Indem man für $x_1, \ldots, x_n$ reelle Zahlen einsetzt, erhält man eine Funktion $D : \mathbb{R}^n \to \mathbb{R}$.

Für $n = 2$ lautet beispielsweise

$$D(x_1, x_2) = x_2 - x_1$$

und für $n = 3$:

$$D(x_1, x_2, x_3) = (x_3 - x_2) \cdot (x_3 - x_1) \cdot (x_2 - x_1).$$

5.1.1 Satz

Sei $k \in \{1, \ldots, n-1\}$*, dann gilt:*

$$D(x_1, \ldots, x_n) = -D(x_1, \ldots, x_{k-1}, x_{k+1}, x_k, x_{k+2}, \ldots, x_n).$$

Bei Vertauschung zweier benachbarter Veränderlicher ändert D das Vorzeichen.

Beweis

Die beiden Polynome bestehen aus derselben Anzahl von Faktoren. Wir werden jedem Faktor von $D(x_1, \ldots, x_n)$ einen Faktor von $D(x_1, \ldots, x_{k+1}, x_k, \ldots, x_n)$ zuordnen. Mit $P = \prod_{\substack{i<j \\ (i,j) \ne (k,k+1)}} (x_j - x_i)$ gilt $D(x_1, \ldots, x_n) = (x_{k+1} - x_k) \cdot P$. Wir betrachten sämtliche Faktoren $(x_j - x_i)$ von $D(x_1, \ldots, x_n)$:

Fall 1: $i \notin \{k, k+1\}$, $j \notin \{k, k+1\}$. Die Veränderlichen x_i, x_j sind von der Vertauschung nicht betroffen.

Fall 2: $i < k,\ j = k$ oder $i < k,\ j = k+1$. Der Faktor $(x_k - x_i)$ geht durch Vertauschung über in $(x_{k+1} - x_i)$, der Faktor $(x_{k+1} - x_i)$ in $(x_k - x_i)$. Beide Faktoren treten jeweils in beiden Produkten auf.

Fall 3: $i = k,\ j = k+1$. Der Faktor $(x_{k+1} - x_k)$ geht durch Vertauschung über in $(x_k - x_{k+1})$, es findet ein *Vorzeichenwechsel* statt.

Fall 4: $i = k,\ j > k+1$ oder $i = k+1,\ j > k+1$. Der Faktor $x_j - x_k$ geht über in $x_j - x_{k+1}$, $x_j - x_{k+1}$ in $x_j - x_k$. Beide Faktoren treten jeweils in beiden Produkten auf.

Insgesamt gilt:

$$D(x_1, \ldots, x_{k+1}, x_k, \ldots, x_n) = (x_k - x_{k+1}) \cdot P = -D(x_1, \ldots, x_n). \qquad \square$$

Ein entsprechender Satz gilt auch für allgemeine Vertauschungen zweier Variabler.

5.1.2 Satz

Für alle Indizes $k, \ell \in \{1, \ldots, n\}$ *mit* $k < \ell$ *gilt:*

$$D(x_1, \ldots, x_{k-1}, x_\ell, x_{k+1}, \ldots, x_{\ell-1}, x_k, x_{\ell+1}, \ldots, x_n) = -D(x_1, \ldots, x_n).$$

Beweis

Wir erhalten dieses Ergebnis durch eine hinreichend häufige Anwendung des vorhergehenden Satzes. Ausgehend von $D(x_1, \ldots, x_n)$ wird x_k $(\ell - k)$-mal mit seinem jeweiligen Nachfolger vertauscht:

$$D(x_1, \ldots, x_n) = (-1)^{\ell-k} D(x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_\ell, x_k, x_{\ell+1}, \ldots, x_n).$$

Nun wird x_ℓ $((\ell - 2) - (k - 1))$-mal mit seinem jeweiligen Vorgänger vertauscht:

$$\begin{aligned} &D(x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_\ell, x_k, x_{\ell+1}, \ldots, x_n) \\ &= (-1)^{\ell-k-1} D(x_1, \ldots, x_{k-1}, x_\ell, x_{k+1}, \ldots, x_{\ell-1}, x_k, x_{\ell+1}, \ldots, x_n). \end{aligned}$$

Wegen $(-1)^{\ell-k+\ell-k-1} = (-1)^{2(\ell-k)-1} = -1$ folgt die Behauptung. $\square$

5.1.3 Definition

Die **Signum**-*Funktion* $\operatorname{sgn} : \mathbb{R} \to \mathbb{R}$ *wird durch*

$$\operatorname{sgn}(x) = \begin{cases} 1, & \text{falls } x > 0, \\ 0, & \text{falls } x = 0, \\ -1, & \text{falls } x < 0, \end{cases}$$

erklärt.

5.1.4 Definition

Für jeweils n *reelle Zahlen* $r_1, \ldots, r_n$ *definieren wir das* **allgemeine Kroneckersymbol** *durch* $\delta(r_1, \ldots, r_n) := \operatorname{sgn}(D(r_1, \ldots, r_n))$.

Ist speziell $n = 2$, so erhalten wir für $(i, j) \in \mathbb{N} \times \mathbb{N}$:

$$\delta(i, j) = \operatorname{sgn}(j - i) = \begin{cases} 1, & \text{falls } j > i, \\ 0, & \text{falls } j = i, \\ -1, & \text{falls } j < i, \end{cases}$$

und damit für das vorher eingeführte (spezielle) Kroneckersymbol

$$\delta_{ij} = 1 - |\delta(i, j)|.$$

Wir wollen den Zusammenhang zwischen der Anordnung der Zahlen $r_1, \ldots, r_n$ und dem allgemeinen Kroneckersymbol $\delta(r_1, \ldots, r_n)$ studieren. Dazu zeichnen wir eine spezielle Anordnung dieser Zahlen aus:

5.1.5 Definition

Es seien $r_1, \ldots, r_n \in \mathbb{R}$ paarweise verschiedene reelle Zahlen, d. h. es gelte $r_i \neq r_j$ für $i \neq j$. Wir sagen, daß $r_1, \ldots, r_n$ in **natürlicher Anordnung** *sind, wenn $r_1 < r_2 < \ldots < r_n$ gilt.*

5.1.6 Bemerkung

(a) Genau dann ist $\delta(r_1, \ldots, r_n) = 0$, wenn für mindestens ein Indexpaar $i \neq j$ gilt: $r_i = r_j$.

(b) Sind $r_1, \ldots, r_n \in \mathbb{R}$ in natürlicher Anordnung, so gilt für $j > i$ stets $r_j - r_i > 0$, also $\delta(r_1, \ldots, r_n) = 1$.

5.1.7 Satz

Die reellen Zahlen $r_1, \ldots, r_n$ seien paarweise verschieden. Dann gilt:

(a) Durch eine endliche Anzahl ($\leq n-1$) von Vertauschungen je zweier Zahlen kann man für $r_1, \ldots, r_n$ natürliche Anordnung erreichen.

(b) Genau dann wird die natürliche Anordnung durch eine gerade bzw. ungerade Anzahl von Vertauschungen erreicht, wenn

$$\delta(r_1, \ldots, r_n) = 1 \quad \textit{bzw.} \quad \delta(r_1, \ldots, r_n) = -1$$

gilt.

Beweis

(a) Zunächst wird r_1 mit der kleinsten Zahl vertauscht. In der neuen Anordnung wird r_2 mit der zweitkleinsten Zahl vertauscht, usw. Nach maximal $(n-1)$ Schritten hat man erreicht, daß r_{n-1} die zweitgrößte und damit r_n die größte Zahl geworden ist.

(b) Folgt aus Satz 5.1.2 und Bemerkung 5.1.6 (b). □

Insbesondere ist es unmöglich, daß natürliche Anordnung sowohl durch eine gerade als auch eine ungerade Anzahl von Vertauschungen erreicht werden kann.

5.1.8 Beispiel

(a) $(3, 1, 2) \to (2, 1, 3) \to (1, 2, 3)$, zwei Vertauschungen.

$(3, 1, 2) \to (3, 2, 1) \to (2, 3, 1) \to (1, 3, 2) \to (1, 2, 3)$, vier Vertauschungen.

(b) $(2,1,3) \to (1,2,3)$, eine Vertauschung.

$(2,1,3) \to (2,3,1) \to (3,2,1) \to (1,2,3)$, drei Vertauschungen.

5.1.9 Definition

Es seien $r_1, \ldots, r_n \in \mathbb{R}$ paarweise verschiedene reelle Zahlen in natürlicher Anordnung. Sind $i_1, \ldots, i_n \in \{1, \ldots, n\}$ paarweise verschiedene Indizes, so heißt die Anordnung $r_{i_1}, \ldots, r_{i_n}$ eine **Permutation** *von $r_1, \ldots, r_n$. Die Permutation heißt* **gerade**, *wenn $\delta(r_{i_1}, \ldots, r_{i_n}) = 1$ ist, und* **ungerade**, *wenn $\delta(r_{i_1}, \ldots, r_{i_n}) = -1$ ist. Man nennt $\delta(r_{i_1}, \ldots, r_{i_n})$ auch das* **Vorzeichen** *der Permutation.*

Die Wahl der Bezeichnungen „gerade“ und „ungerade“ erklärt sich unmittelbar durch Satz 5.1.7.

Permutationen können auch als bijektive Abbildungen $\pi : \{r_1, \ldots, r_n\} \to \{r_1, \ldots, r_n\}$ erklärt werden. Durch $\pi(r_1), \ldots, \pi(r_n)$ wird eine Anordnung gegeben. Liegt umgekehrt eine Anordnung $r_{i_1}, \ldots, r_{i_n}$ vor, kann π durch $\pi(r_j) = r_{i_j}$ gewonnen werden.

5.1.10 Definition

Die **Fakultätsfunktion** $\mathbb{N} \to \mathbb{N}$, $n \mapsto n!$ *wird induktiv durch* $1! := 1$, $(n+1)! := n! \cdot (n+1)$ *erklärt.*

Offensichtlich gilt auch $n! = \prod_{j=1}^{n} j$.

5.1.11 Satz

Es gibt genau $n!$ Permutationen von $1, \ldots, n$ oder allgemeiner von n reellen Zahlen $r_1 < \ldots < r_n$.

Beweis

Wir führen den Beweis durch vollständige Induktion nach n. Der Induktionsanfang $n = 1$ ist trivial, denn die Zahl 1 kann auf genau eine Art angeordnet werden.

Um den Induktionsschritt durchzuführen, wählen wir $n \in \mathbb{N}$ beliebig und nehmen an, die Behauptung sei bereits für $1, \ldots, n$ bewiesen. Zu jeder der folgenden $(n+1)$ Anordnungen gibt es durch Permutation von $1, \ldots, n$ je $n!$ verschiedene Anordnungen, die die Zahl $n+1$ fest lassen:

$$\begin{gathered} n+1, 1, 2, \ldots, n, \\ 1, n+1, 2, \ldots, n, \\ \vdots \\ 1, 2, \ldots, n, n+1. \end{gathered}$$

Insgesamt gibt es $n! \cdot (n+1) = (n+1)!$ verschiedene Anordnungen. □

Die Fakultätsfunktion wächst außerordentlich schnell: $1! = 1$, $2! = 2$, $3! = 6$, $4! = 24, \ldots$ Für große n ist daher zur Berechnung von $n!$ eine Näherungsformel zweckmäßig. Gemäß der *Stirlingschen Formel*, die in vielen Lehrbüchern der Analysis hergeleitet wird, ist

$$n! \approx \sqrt{\frac{2\pi}{n+1}} \left(\frac{n+1}{e}\right)^{n+1}.$$

Die Zahlen $\pi = 3{,}14\ldots$ und $e = 2{,}718\ldots$ werden in der Analysis eingeführt. Für $n = 100$ ergibt sich beispielsweise

$$100! \approx 9{,}3 \cdot 10^{157}.$$

5.2 Determinanten

5.2.1 Definition

Die **Determinante** *der quadratischen Matrix* $A = (a_{ij})_{i,j=1,\ldots,n}$ *wird durch*

$$\det(A) := \sum_{i_1,\ldots,i_n=1}^{n} \delta(i_1, \ldots, i_n) \cdot a_{i_1 1} \cdot \ldots \cdot a_{i_n n}$$

definiert.

Das n-fach indizierte Summenzeichen $\sum\limits_{i_1,\ldots,i_n=1}^{n}$ steht abkürzend für n Summenzeichen $\sum\limits_{i_1=1}^{n} \cdots \sum\limits_{i_n=1}^{n}$, auf deren Reihenfolge es wegen des allgemeinen Kommutativgesetzes nicht ankommt. In der Formel für $\det(A)$ treten n^n Summanden auf. Wenn allerdings $i_1, \ldots, i_n$ nicht paarweise verschieden sind, so ist $\delta(i_1, \ldots, i_n) = 0$; die Summe ist also letztlich „nur“ über die $n!$ Permutationen von $1, \ldots, n$ zu erstrecken.

Die Determinante ist ein Polynom in den Matrixkomponenten im folgenden Sinne:

Ein reelles Polynom in m Veränderlichen ist ein Ausdruck der Form

$$p(x_1, \ldots, x_m) = \sum_{\substack{i_1=0,\ldots,k_1 \\ \vdots \\ i_m=0,\ldots,k_m}} c_{i_1,\ldots,i_m} x_1^{i_1} \cdot \ldots \cdot x_m^{i_m}$$

mit $c_{i_1,\ldots,i_m} \in \mathbb{R}$, $k_1, \ldots, k_m \in \mathbb{N}_0$; dabei wird $x_i^0 = 1$ vereinbart. Für $x_1, \ldots, x_m$ können reelle Zahlen eingesetzt werden, p ist dann eine Funktion $\mathbb{R}^m \to \mathbb{R}$.

Sei nun $A = (a_{ij})_{i,j=1,\ldots,n}$ eine $n \times n$-Matrix, dann fassen wir die a_{ij} als Veränderliche auf:

$$(x_1, \ldots, x_m) := (a_{11}, \ldots, a_{1n}, a_{21}, \ldots, a_{2n}, \ldots, a_{n1}, \ldots, a_{nn})$$

mit $m = n^2$. Die Determinante

$$\det(A) = \sum_{i_1,\ldots,i_n=1}^{n} \delta(i_1, \ldots, i_n) \cdot a_{i_1 1} \cdot \ldots \cdot a_{i_n n}$$

ist also ein Polynom in $m = n^2$ Veränderlichen mit ganzzahligen Koeffizienten.

5.2.2 Beispiel

Im Fall $n = 1$ hat die Matrix A eine Komponente, $A = (a_{11})$, und es ist

$$\det(A) = \sum_{i_1=1}^{1} \delta(i_1) \cdot a_{i_1 1} = \delta(1) \cdot a_{11} = a_{11}.$$

Für die 2-reihige Matrix $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ ergibt die Definition

$$\det(A) = \sum_{i_1,i_2=1}^{2} \delta(i_1, i_2) a_{i_1 1} a_{i_2 2} = \delta(1,2) a_{11} a_{22} + \delta(2,1) a_{21} a_{12}.$$

Die Zahlen $1, 2$ sind in natürlicher Anordnung; $2, 1$ gehen durch eine Vertauschung in natürliche Anordnung über, also ist $\delta(1,2) = 1$, $\delta(2,1) = -1$. Folglich ist

$$\det(A) = a_{11} a_{22} - a_{21} a_{12}.$$

Im Fall $n = 3$ kann man die schon etwas unübersichtliche Formel von Sarrus

$$\det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23}$$
$$- a_{31}a_{22}a_{13} - a_{21}a_{12}a_{33} - a_{11}a_{32}a_{23}$$

herleiten.

Für größere n wird Definition 5.2.1 für die *Berechnung* von Determinanten schnell zu aufwendig. Im Folgenden werden deshalb Eigenschaften der Determinante hergeleitet, auf Grund derer dafür der Gaußsche Algorithmus eingesetzt werden kann.

Wir fassen die Determinante nun als Funktion $\underbrace{\mathbb{R}^n \oplus \ldots \oplus \mathbb{R}^n}_{n\text{-mal}} \to \mathbb{R}$ der Spalten der Matrix $A = (\vec{a}_1, \ldots, \vec{a}_n)$ auf und stellen deren *charakteristische* Eigenschaften zusammen.

5.2.3 Definition

Eine Funktion $f : \underbrace{\mathbb{R}^n \oplus \ldots \oplus \mathbb{R}^n}_{m\text{-mal}} \to \mathbb{R}$ *heißt* **multilinear** *oder genauer* m*-fach multilinear, falls sie in jeder Komponente linear ist, d. h. falls für jedes* $k = 1, \ldots, m$ *stets gilt:*

(1) $f(\vec{a}_1, \ldots, \vec{a}_k + \vec{a}_k{}', \ldots, \vec{a}_m) = f(\vec{a}_1, \ldots, \vec{a}_k, \ldots, \vec{a}_m) + f(\vec{a}_1, \ldots, \vec{a}_k{}', \ldots, \vec{a}_m)$,

(2) $f(\vec{a}_1, \ldots, c \cdot \vec{a}_k, \ldots, \vec{a}_m) = c \cdot f(\vec{a}_1, \ldots, \vec{a}_k, \ldots, \vec{a}_m)$.

Eine multilineare Funktion heißt **Multilinearform**, *im Falle* $m = 2$ *spricht man von* **Bilinearformen**.

Im Fall $m = 1$ erhält man die Definition einer Linearform, siehe Kapitel 4.3.

5.2.4 Satz

Die Determinantenfunktion

$$f : \underbrace{\mathbb{R}^n \oplus \ldots \oplus \mathbb{R}^n}_{n\text{-mal}} \to \mathbb{R}, \quad f(\vec{a}_1, \ldots, \vec{a}_n) := \det(\vec{a}_1, \ldots, \vec{a}_n)$$

ist eine n*-fache Multilinearform.*

Beweis

Sei $k \in \{1, \ldots, n\}$ beliebig. Für $\vec{a}_k{}', \vec{a}_1, \ldots, \vec{a}_n \in \mathbb{R}^n$, $c \in \mathbb{R}$ gilt:

$$\begin{aligned}
&\det(\vec{a}_1, \ldots, \vec{a}_k + \vec{a}_k{}', \ldots, \vec{a}_n) \\
&= \sum_{i_1, \ldots, i_n = 1}^{n} \delta(i_1, \ldots, i_n) \cdot a_{i_1 1} \cdot \ldots \cdot (a_{i_k k} + a'_{i_k k}) \cdot \ldots \cdot a_{i_n n} \\
&= \sum_{i_1, \ldots, i_n = 1}^{n} \delta(i_1, \ldots, i_n) \cdot a_{i_1 1} \cdot \ldots \cdot a_{i_k k} \cdot \ldots \cdot a_{i_n n} \\
&\quad + \sum_{i_1, \ldots, i_n = 1}^{n} \delta(i_1, \ldots, i_n) \cdot a_{i_1 1} \cdot \ldots \cdot a'_{i_k k} \cdot \ldots \cdot a_{i_n n} \\
&= \det(\vec{a}_1, \ldots, \vec{a}_k, \ldots, \vec{a}_n) + \det(\vec{a}_1, \ldots, \vec{a}_k{}', \ldots, \vec{a}_n); \\
&\det(\vec{a}_1, \ldots, c \cdot \vec{a}_k, \ldots, \vec{a}_n) \\
&= \sum_{i_1, \ldots, i_n = 1}^{n} \delta(i_1, \ldots, i_n) \cdot a_{i_1 1} \cdot \ldots \cdot (c \cdot a_{i_k k}) \cdot \ldots \cdot a_{i_n n} \\
&= c \cdot \sum_{i_1, \ldots, i_n = 1}^{n} \delta(i_1, \ldots, i_n) \cdot a_{i_1 1} \cdot \ldots \cdot a_{i_k k} \cdot \ldots \cdot a_{i_n n} \\
&= c \cdot \det(\vec{a}_1, \ldots, \vec{a}_k, \ldots, \vec{a}_n).
\end{aligned}$$

□

5.2.5 Satz

Sei $f : \underbrace{\mathbb{R}^n \oplus \ldots \oplus \mathbb{R}^n}_{m\text{-mal}} \to \mathbb{R}$ *eine* m*-fache Multilinearform. Gilt für ein Argument* $\vec{a}_j = \vec{0}$*, so ist* $f(\vec{a}_1, \ldots, \vec{a}_n) = 0$.

Beweis

Wir halten die Vektoren $\vec{a}_1, \ldots, \vec{a}_{j-1}$ und $\vec{a}_{j+1}, \ldots, \vec{a}_n$ fest und betrachten f als *lineare* Abbildung $\mathbb{R}^n \to \mathbb{R}$ allein im j-ten Spaltenvektor. Jede solche lineare Abbildung bildet aber $\vec{0}$ auf 0 ab. □

5.2.6 Definition

Eine m*-fache Multilinearform* $f : \underbrace{\mathbb{R}^n \oplus \ldots \oplus \mathbb{R}^n}_{m\text{-mal}} \to \mathbb{R}$ *heißt* **alterniernd**, *wenn für je zwei Indizes* $1 \leq k < \ell \leq m$ *stets gilt:*

$$f(\vec{a}_1, \ldots, \vec{a}_k, \ldots \vec{a}_\ell, \ldots, \vec{a}_n) = -f(\vec{a}_1, \ldots \underbrace{, \vec{a}_\ell,}_{k\text{-te Stelle}} \ldots \underbrace{, \vec{a}_k,}_{\ell\text{-te Stelle}} \ldots, \vec{a}_n).$$

5.2.7 Satz

Die Determinantenfunktion $\det : \underbrace{\mathbb{R}^n \oplus \ldots \oplus \mathbb{R}^n}_{n\text{-mal}} \to \mathbb{R}$ *ist alternierend.*

Beweis

Seien $1 \leq k < \ell \leq n$ zwei Indizes. Der Beweis beruht auf der entsprechenden Eigenschaft des allgemeinen Kroneckersymbols, vgl. Satz 5.1.2.

$$\begin{aligned}
&\det(\vec{a}_1, \ldots \underbrace{, \vec{a}_\ell,}_{k\text{-te Stelle}} \ldots \underbrace{, \vec{a}_k,}_{\ell\text{-te Stelle}} \ldots, \vec{a}_n) \\
&= \sum_{i_1, \ldots, i_n = 1}^{n} \delta(i_1, \ldots, i_n) \cdot a_{i_1 1} \cdot \ldots \cdot a_{i_k \ell} \cdot \ldots \cdot a_{i_\ell k} \cdot \ldots \cdot a_{i_n n} \\
&\quad \text{(Umbenennung: der Index } i_k \text{ heißt nun } i_\ell \text{ und } i_\ell \text{ heißt nun } i_k\text{)} \\
&= \sum_{i_1, \ldots, i_n = 1}^{n} \delta(i_1, \ldots, i_\ell, \ldots, i_k, \ldots, i_n) \cdot a_{i_1 1} \cdot \ldots \cdot a_{i_\ell \ell} \cdot \ldots \cdot a_{i_k k} \cdot \ldots \cdot a_{i_n n} \\
&= -\sum_{i_1, \ldots, i_n = 1}^{n} \delta(i_1, \ldots, i_k, \ldots, i_\ell, \ldots, i_n) \cdot a_{i_1 1} \cdot \ldots \cdot a_{i_n n} \\
&= -\det(\vec{a}_1, \ldots, \vec{a}_n).
\end{aligned}$$

□

5.2.8 Satz

Für die Einheitsmatrix ist $\det(E) = \det(\vec{e}_1, \dots, \vec{e}_n) = 1$.

Beweis

In diesem Falle besteht die Summe aus Definition 5.2.1 aus einem nicht verschwindenden Summanden, für diesen ist $i_1 = 1, \dots, i_n = n$:

$$\det(E) = \sum_{i_1,\dots,i_n=1}^{n} \delta(i_1, \dots, i_n)\delta_{i_1 1} \cdot \ldots \cdot \delta_{i_n n} = \delta(1, \dots, n) = 1. \qquad \square$$

Gemäß Satz 5.2.4 bis Satz 5.2.8 ist det eine alternierende, normierte[1] n-fache Multilinearform. Es soll nun gezeigt werden, daß die Determinante durch diese Eigenschaften bereits vollkommen bestimmt ist. Man hätte die Determinante auch gleich auf diese Weise definieren können, dann hätte man allerdings die vorhergehenden Überlegungen zum Nachweis der Existenz einer solchen Funktion nachtragen müssen.

Zunächst benötigen wir allerdings noch einen Hilfssatz:

5.2.9 Satz

Es sei $f : \underbrace{\mathbb{R}^n \oplus \ldots \oplus \mathbb{R}^n}_{m\text{-mal}} \to \mathbb{R}$ *eine alternierende* m*-fache Multilinearform.*

(a) Sind $\vec{a}_1, \dots, \vec{a}_m \in \mathbb{R}^n$ *nicht paarweise verschieden, so gilt* $f(\vec{a}_1, \dots, \vec{a}_m) = 0$.

(b) Für je m *Indizes* $1 \leq i_1, \dots, i_m \leq m$ *und* m *Vektoren* $\vec{a}_1, \dots, \vec{a}_m \in \mathbb{R}^n$ *gilt:*

$$f(\vec{a}_{i_1}, \dots, \vec{a}_{i_m}) = \delta(i_1, \dots, i_m) \cdot f(\vec{a}_1, \dots, \vec{a}_m).$$

Beweis

(a) Sind $\vec{a}_1, \dots, \vec{a}_m \in \mathbb{R}^n$ nicht paarweise verschieden, so gibt es zwei Indizes $1 \leq k < \ell \leq m$ mit $\vec{a}_k = \vec{a}_\ell$. Da f alternierend ist, folgt

$$f(\vec{a}_1, \dots, \vec{a}_m) = -f(\vec{a}_1, \dots \underbrace{, \vec{a}_\ell,}_{k\text{-te Stelle}} \dots \underbrace{, \vec{a}_k,}_{\ell\text{-te Stelle}} \dots, \vec{a}_n) = -f(\vec{a}_1, \dots, \vec{a}_m),$$

also $2 \cdot f(\vec{a}_1, \dots, \vec{a}_m) = 0$.

(b) Wir können ohne Einschränkung annehmen, daß $\vec{a}_1, \dots, \vec{a}_m$ paarweise verschieden sind; andernfalls reduziert sich die Behauptung wegen (a) auf $0 = 0$. Weiter seien $i_1, \dots, i_m$ ohne Einschränkung paarweise verschieden, $i_1, \dots, i_m$ ist also eine Permutation von $1, \dots, m$. Damit ist $\delta(i_1, \dots, i_m) = \pm 1$. Wegen $\delta(i_1, \dots, i_m) = \frac{1}{\delta(i_1,\dots,i_m)}$ reicht es zu zeigen:

$$\delta(i_1, \dots, i_m) \cdot f(\vec{a}_{i_1}, \dots, \vec{a}_{i_m}) = f(\vec{a}_1, \dots, \vec{a}_m).$$

[1] D. h. $\det(\vec{e}_1, \dots, \vec{e}_n) = 1$.

Bei Vertauschung zweier Indizes ändert sich die linke Seite nicht, δ und f ändern nämlich jeweils ihr Vorzeichen. Durch endlich viele Vertauschungen erreicht man natürliche Reihenfolge von $i_1, \ldots, i_m$, d. h. $1, \ldots, m$. Dabei ist $\delta(1, \ldots, m) = 1$.
$\square$

5.2.10 Satz

Sei $f : \underbrace{\mathbb{R}^n \oplus \ldots \oplus \mathbb{R}^n}_{n\text{-mal}} \to \mathbb{R}$ *eine alternierende* n*-fache Multilinearform. Dann gilt für alle* $\vec{a}_1, \ldots, \vec{a}_n \in \mathbb{R}^n$*:*

$$f(\vec{a}_1, \ldots, \vec{a}_n) = f(\vec{e}_1, \ldots, \vec{e}_n) \cdot \det(\vec{a}_1, \ldots, \vec{a}_n).$$

Hat man zusätzlich $f(\vec{e}_1, \ldots, \vec{e}_n) = 1$*, so ist* $f = \det$.

Beweis

Wegen $\vec{a}_k = \sum_{i=1}^n a_{ik}\vec{e}_i$ ergibt sich unter Benutzung von Satz 5.2.9:

$$\begin{aligned}
f(\vec{a}_1, \ldots, \vec{a}_n) &= \sum_{i_1=1}^n a_{i_1 1} f(\vec{e}_{i_1}, \vec{a}_2, \ldots, \vec{a}_n) \\
&= \sum_{i_1=1}^n a_{i_1 1} \sum_{i_2=1}^n a_{i_2 2} f(\vec{e}_{i_1}, \vec{e}_{i_2}, \vec{a}_3, \ldots, \vec{a}_n) \\
&\quad \vdots \\
&= \sum_{i_1, \ldots, i_n = 1}^n a_{i_1 1} \cdot \ldots \cdot a_{i_n n} \cdot f(\vec{e}_{i_1}, \ldots, \vec{e}_{i_n}) \\
&= \sum_{i_1, \ldots, i_n = 1}^n a_{i_1 1} \cdot \ldots \cdot a_{i_n n} \cdot \delta(i_1, \ldots, i_n) \cdot f(\vec{e}_1, \ldots, \vec{e}_n) \\
&= \det(\vec{a}_1, \ldots, \vec{a}_n) \cdot f(\vec{e}_1, \ldots, \vec{e}_n).
\end{aligned}$$

$\square$

5.3 Weitere Eigenschaften der Determinante

Nachdem wir im vorherigen Abschnitt die Determinante eingeführt und als alternierende n-fache Multilinearform charakterisiert haben, wollen wir hier daraus zwei der wichtigsten Sätze der Determinantentheorie herleiten: die Gleichheit der Determinanten von A und A^t und den Multiplikationssatz für Determinanten.

Determinanten transponierter Matrizen

5.3.1 Satz

Für jede $n \times n$*-Matrix* A *gilt* $\det(A) = \det(A^t)$.

Beweis

Definitionsgemäß gilt

$$\det(A^t) = \sum_{\substack{\underbrace{j_1,\ldots,j_n}_{\text{paarweise verschieden}}=1}}^{n} \delta(j_1,\ldots,j_n) \cdot a_{1j_1} \cdot \ldots \cdot a_{nj_n}, \tag{$*$}$$

denn das allgemeine Kroneckersymbol verschwindet, sobald zwei Indizes übereinstimmen.

Wir betrachten zunächst einen festen Summanden, der durch eine Anordnung $j_1,\ldots,j_n$ von $1,\ldots,n$ gegeben ist. Jeder Faktor ist durch ein Indexpaar (i,j_i) gegeben, das Produkt kann durch die Abfolge $(1,j_1),\ldots,(n,j_n)$ beschrieben werden. Die Vertauschung zweier Faktoren entspricht der Vertauschung der zugehörigen Indexpaare. Man kann durch eine Anzahl von Vertauschungen erreichen, daß die jeweils zweiten Indizes $j_1,\ldots,j_n$ natürliche Anordnung einnehmen, die jeweils ersten Indizes $1,\ldots,n$ treten nun in der Anordnung $i_1,\ldots,i_n$ auf. Aufgrund des Kommutativgesetzes in $\mathbb{R}$ gilt

$$a_{1j_1} \cdot \ldots \cdot a_{nj_n} = a_{i_1 1} \cdot \ldots \cdot a_{i_n n}.$$

Da nach Konstruktion $j_1,\ldots,j_n$ und $i_1,\ldots,i_n$ aus $1,\ldots,n$ durch dieselbe Anzahl von Vertauschungen hervorgehen, gilt auch

$$\delta(i_1,\ldots,i_n) = \delta(j_1,\ldots,j_n).$$

Damit erhalten wir eine Abbildung $\Phi : (j_1,\ldots,j_n) \mapsto (i_1,\ldots,i_n)$ von M in M, wobei M die Menge aller Permutationen von $1,\ldots,n$ bezeichnet. Diese Abbildung ist wohldefiniert, denn jedes k taucht in den Indexpaaren $(1,j_1),\ldots,(n,j_n)$ genau einmal als zweiter Index auf, der entsprechende erste Index ist dann i_k. Um die Summation in $(*)$ statt über $j_1,\ldots,j_n$ über $i_1,\ldots,i_n$ erstrecken zu können, ist zu zeigen, daß die Zuordnung Φ bijektiv ist.

Zum Nachweis der Injektivität seien zwei Permutationen $(j_1,\ldots,j_n)$ und $(\hat{j}_1,\ldots,\hat{j}_n)$ von $1,\ldots,n$ mit

$$\Phi(j_1,\ldots,j_n) =: (i_1,\ldots,i_n) = (\hat{\imath}_1,\ldots,\hat{\imath}_n) := \Phi(\hat{j}_1,\ldots,\hat{j}_n) \tag{$**$}$$

gegeben. Die Abfolge der zugehörigen Indexpaare ist

$$(i_1,1),\ldots,(i_n,n) \quad \text{bzw.} \quad (\hat{\imath}_1,1),\ldots,(\hat{\imath}_n,n),$$

gemäß unserer Annahme (∗∗) stimmen diese Abfolgen überein. Durch umgekehrte Anwendung der Vertauschungen, die

$$(1, j_1), \ldots, (n, j_n) \quad \text{bzw.} \quad (1, \hat{\jmath}_1), \ldots, (n, \hat{\jmath}_n)$$

in

$$(i_1, 1), \ldots, (i_n, n) \quad \text{bzw.} \quad (\hat{\imath}_1, 1), \ldots, (\hat{\imath}_n, n)$$

überführt haben, gelangt man zu

$$(1, j_1), \ldots, (n, j_n) \quad \text{bzw.} \quad (1, \hat{\jmath}_1), \ldots, (n, \hat{\jmath}_n).$$

Beide Abfolgen bestehen aus denselben Indexpaaren, also gilt

$$j_1 = \hat{\jmath}_1, \ldots, j_n = \hat{\jmath}_n.$$

Folglich ist Φ eine injektive Abbildung der endlichen Menge M in sich und damit auch surjektiv.

Da nun $\Phi : M \to M$ bijektiv ist, können wir in (∗) über die Permutationen $i_1, \ldots, i_n$ summieren und erhalten:

$$\begin{aligned} \det(A^t) &= \sum_{\substack{\underbrace{j_1, \ldots, j_n}_{\text{paarweise verschieden}}=1}}^{n} \delta(j_1, \ldots, j_n) \cdot a_{1j_1} \cdot \ldots \cdot a_{nj_n} \\ &= \sum_{\substack{\underbrace{i_1, \ldots, i_n}_{\text{paarweise verschieden}}=1}}^{n} \delta(i_1, \ldots, i_n) \cdot a_{i_1 1} \cdot \ldots \cdot a_{i_n n} = \det(A). \end{aligned}$$

□

5.3.2 Bemerkung

Folglich gelten alle Aussagen über Determinanten, die mit Spalten formuliert werden, auch für Zeilen. Faßt man etwa $\det$ als Funktion von den je n Zeilen der Matrizen auf, so ist $\det$ eine alternierende n-fache Multilinearform in den Zeilen.

Mit dieser Bemerkung erhalten wir beispielsweise aus Satz 5.2.5:

5.3.3 Satz

Hat eine Matrix A eine Nullzeile oder eine Nullspalte, so ist $\det(A) = 0$.

Ebenso folgt aus Satz 5.2.9:

5.3.4 Satz

Sei A eine $n \times n$-Matrix. Sind in A zwei Spalten gleich, so gilt $\det(A) = 0$. *Entsprechendes gilt auch für Zeilen.*

Es ist nun leicht, die Änderungen der Determinante unter elementaren Matrixumformungen zu beschreiben. Multiplikation einer Zeile bzw. einer Spalte mit dem Faktor $c \in \mathbb{R}$ ver-c-facht den Wert der Determinante, denn die Determinante ist n-fach linear und damit insbesondere n-fach homogen.

Bei Vertauschungen zweier Spalten bzw. Zeilen ändert die (alternierende!) Determinante ihr Vorzeichen.

Schließlich gilt noch:

5.3.5 Satz

Es sei A eine $n \times n$-Matrix, und es seien $j, k \in \{1, \ldots, n\}$ zwei verschiedene Indizes: $j \neq k$. Addiert man ein Vielfaches der j-ten Spalte von A zur k-ten Spalte, so ändert sich die Determinante nicht. Entsprechendes gilt auch für Zeilen.

Beweis

Sind $\vec{a}_1, \ldots, \vec{a}_n$ die Spalten von A, so lautet die k-te Spalte der umgeformten Matrix: $\vec{a}_k + c \cdot \vec{a}_j$ mit einem $c \in \mathbb{R}$, die übrigen Spalten werden nicht geändert. Da die Determinante multilinear und alternierend ist, gilt:

$$\begin{aligned} &\det(\vec{a}_1, \ldots, \vec{a}_k + c \cdot \vec{a}_j, \ldots, \vec{a}_n) \\ &= \det(\vec{a}_1, \ldots, \vec{a}_n) + c \cdot \det(\vec{a}_1, \ldots, \underbrace{\vec{a}_j,}_{k\text{-te Stelle}} \ldots, \vec{a}_n) = \det(A) + 0, \end{aligned}$$

denn in der zweiten Matrix stimmen zwei Spalten überein. □

Multiplikationssatz

5.3.6 Satz (Multiplikationssatz)

Für alle $n \times n$-Matrizen A, B gilt $\det(B \circ A) = \det(B) \cdot \det(A)$.

Beweis

Wir führen den Beweis mit Hilfe unserer axiomatischen Charakterisierung der Determinante als alternierende n-fache Multilinearform.

Sei B eine beliebige, aber im Folgenden feste $n \times n$-Matrix. Wir definieren eine Funktion

$$f : \underbrace{\mathbb{R}^n \oplus \ldots \oplus \mathbb{R}^n}_{n\text{-mal}} \to \mathbb{R},$$

$$f(\vec{a}_1, \ldots, \vec{a}_n) := \det(B \circ \vec{a}_1, \ldots, B \circ \vec{a}_n) = \det(B \circ A),$$

wobei wir $A = (\vec{a}_1, \ldots, \vec{a}_n)$ schreiben.

Diese Funktion f ist multilinear, denn die Linearität der zu B gehörigen Abbildung (man vgl. auch Folgerung 3.3.5) und die Multilinearität von $\det$ ergeben

$$\begin{aligned} f(\vec{a}_1, \ldots, \vec{a}_k + \vec{a}_k{}', \ldots, \vec{a}_n) &= \det(B \circ \vec{a}_1, \ldots, B \circ (\vec{a}_k + \vec{a}_k{}'), \ldots, B \circ \vec{a}_n) \\ &= \det(B \circ \vec{a}_1, \ldots, B \circ \vec{a}_k + B \circ \vec{a}_k{}', \ldots, B \circ \vec{a}_n) \\ &= \det(B \circ \vec{a}_1, \ldots, B \circ \vec{a}_k, \ldots, B \circ \vec{a}_n) \\ &\quad + \det(B \circ \vec{a}_1, \ldots, B \circ \vec{a}_k{}', \ldots, B \circ \vec{a}_n) \\ &= f(\vec{a}_1, \ldots, \vec{a}_k, \ldots, \vec{a}_n) + f(\vec{a}_1, \ldots, \vec{a}_k{}', \ldots, \vec{a}_n) \end{aligned}$$

und entsprechend

$$f(\vec{a}_1, \ldots, c \cdot \vec{a}_k, \ldots, \vec{a}_n) = c \cdot f(\vec{a}_1, \ldots, \vec{a}_k, \ldots, \vec{a}_n).$$

Außerdem ist f alternierend, denn $\det$ ist alternierend.

Die Anwendung von Satz 5.2.10 liefert nun für alle $n \times n$-Matrizen $A = (\vec{a}_1, \ldots, \vec{a}_n)$:

$$\begin{aligned} \det(B \circ A) &= f(\vec{a}_1, \ldots, \vec{a}_n) = \det(\vec{a}_1, \ldots, \vec{a}_n) \cdot f(\vec{e}_1, \ldots, \vec{e}_n) \\ &= \det(A) \cdot \det(B \circ \vec{e}_1, \ldots, B \circ \vec{e}_n) = \det(A) \cdot \det(B). \end{aligned}$$

□

5.3.7 Satz

Für eine $n \times n$-Matrix A gilt $\det(A) \neq 0$ genau dann, wenn A nichtsingulär ist. In diesem Fall gilt $\det(A^{-1}) = \frac{1}{\det(A)}$.

Die Determinante legt fest (determiniert), ob eine Matrix singulär oder nichtsingulär (regulär, invertierbar) ist.

Beweis

„$\Leftarrow$“: Sei A nichtsingulär. Dann existiert die Umkehrmatrix A^{-1}, und es gilt

$$1 = \det(E) = \det(A \circ A^{-1}) = \det(A) \cdot \det(A^{-1}).$$

Es folgt $\det(A) \neq 0$ und $\det(A^{-1}) = \frac{1}{\det(A)}$.

„$\Rightarrow$“: Sei nun A singulär, nach Satz 3.8.4 sind die Spalten linear abhängig. Gegebenenfalls nach einer Spaltenvertauschung, die nur das Vorzeichen von $\det(A)$ ändert, kann man annehmen, daß die erste Spalte Linearkombination der übrigen ist:

$$\vec{a}_1 = \sum_{j=2}^{n} c_j \vec{a}_j, \qquad c_j \in \mathbb{R},$$

und wir erhalten aus Satz 5.3.4:

$$\begin{aligned}\det(A) &= \det\left(\sum_{j=2}^{n} c_j \cdot \vec{a}_j,\, \vec{a}_2,\, \ldots,\, \vec{a}_n\right)\\ &= \sum_{j=2}^{n} c_j \cdot \det(\vec{a}_j, \vec{a}_2, \ldots \underbrace{, \vec{a}_j,}_{j\text{-te Stelle}} \ldots, \vec{a}_n) = 0.\end{aligned}$$

Wir haben also gezeigt:

$$A \,\text{singulär} \Rightarrow \det(A) = 0$$

und damit

$$\det(A) \neq 0 \Rightarrow A \,\text{nichtsingulär}. \qquad \square$$

Matrizen mit Null-Blöcken

Der Definition der Determinante hatten wir für 2×2-Matrizen $\det\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22}\end{pmatrix} = a_{11}a_{22} - a_{21}a_{12} = \det(a_{11})\det(a_{22}) - \det(a_{21})\det(a_{12})$ entnommen. Wir wollen nun der Frage nachgehen, ob eine entsprechende Gleichung auch für größere Matrizen gilt, wenn $a_{11}, \ldots, a_{22}$ durch Untermatrizen geeigneter Formate ersetzt werden. Das folgende Beispiel zeigt, daß die Frage in dieser Allgemeinheit nicht positiv beantwortet werden kann:

5.3.8 Beispiel

Es ist

$$\det\left(\begin{array}{cc|cc} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ \hline 1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 2 \end{array}\right) = 0,$$

da zwei Spalten übereinstimmen, und dagegen

$$\det\begin{pmatrix} 1 & 0 \\ 0 & 1\end{pmatrix} \cdot \det\begin{pmatrix} 1 & 0 \\ 0 & 2\end{pmatrix} - \det\begin{pmatrix} 1 & 0 \\ 0 & -1\end{pmatrix} \cdot \det\begin{pmatrix} 1 & 0 \\ 0 & 1\end{pmatrix} = 3.$$

Auch der naheliegende Gedanke, aus Gründen des Matrizenformats das Minuszeichen durch ein Pluszeichen zu ersetzen, führt auf $2 + (-1)$ und damit auf ein falsches Ergebnis.

Unter geeigneten, recht einschränkenden Voraussetzungen haben wir aber tatsächlich derartige Gesetze.

5.3.9 Satz

Es seien Matrizen $E = E^{k,k}$, $A = A^{\ell,\ell}$, $B = B^{\ell,k}$ und $\mathcal{O} = \mathcal{O}^{k,\ell}$ gegeben. Dann gilt

$$\det \begin{pmatrix} E & \mathcal{O} \\ B & A \end{pmatrix} = \det(A).$$

Beweis

Wir verwenden wieder die Charakterisierung der Determinante als alternierende Multilinearform aus Satz 5.2.10. Sei $A = (\vec{a}_1, \dots, \vec{a}_\ell)$ und $f : \underbrace{\mathbb{R}^\ell \oplus \ldots \oplus \mathbb{R}^\ell}_{\ell\text{-mal}} \to \mathbb{R}$ durch $f(\vec{a}_1, \dots, \vec{a}_\ell) := \det \begin{pmatrix} E & 0 \\ B & A \end{pmatrix}$ gegeben. Die Spalten von A werden nach oben durch Nullen zu Spalten aus $\mathbb{R}^{k+\ell}$ ergänzt. Daher haben die entsprechenden Eigenschaften der Determinante zur Folge, daß f ℓ-fach multilinear und alternierend ist. Der oben zitierte Satz liefert

$$f(\vec{a}_1, \dots, \vec{a}_\ell) = f(\vec{e}_1, \dots, \vec{e}_\ell) \cdot \det(\vec{a}_1, \dots, \vec{a}_\ell).$$

Es ist also noch zu zeigen: $f(\vec{e}_1, \dots, \vec{e}_\ell) = 1$. Tatsächlich gilt

$$f(\vec{e}_1, \dots, \vec{e}_\ell) = \det \begin{pmatrix} 1 & & 0 & 0 & \dots & 0 \\ & \ddots & & \vdots & & \vdots \\ 0 & & 1 & 0 & \dots & 0 \\ b_{11} & \dots & b_{1k} & 1 & & 0 \\ \vdots & & \vdots & & \ddots & \\ b_{\ell 1} & \dots & b_{\ell k} & 0 & & 1 \end{pmatrix} = \det(E^{k+\ell,k+\ell}).$$

Letzteres folgt auf Grund von Satz 5.3.5, denn b_{ij} kann durch Addition des $-b_{ij}$-fachen der $(k+i)$-ten Spalte zur j-ten Spalte durch 0 ersetzt werden. Der Rest der Matrix bleibt jeweils unverändert. Wegen $\det(E) = 1$ ist der Satz bewiesen. □

Durch Anwendung von $\det(A) = \det(A^t)$ auf Satz 5.3.9 gelangt man zu:

5.3.10 Satz

Es seien Matrizen $E = E^{k,k}$, $A = A^{\ell,\ell}$, $B = B^{k,\ell}$, $\mathcal{O} = \mathcal{O}^{\ell,k}$ gegeben. Dann gilt

$$\det \begin{pmatrix} E & B \\ \mathcal{O} & A \end{pmatrix} = \det(A).$$

5.3.11 Satz

Es seien Matrizen $A = A^{k,k}$, $B = B^{\ell,\ell}$, $C = C^{k,\ell}$, $D = D^{\ell,k}$ gegeben. Dann gilt

$$\det \begin{pmatrix} A & C \\ \mathcal{O} & B \end{pmatrix} = \det \begin{pmatrix} A & \mathcal{O} \\ D & B \end{pmatrix} = \det(A) \cdot \det(B).$$

Beweis

Wir gehen wie im Beweis von Satz 5.3.9 vor. Es sei wieder $A = (\vec{a}_1, \ldots, \vec{a}_k)$ und $f : \underbrace{\mathbb{R}^k \oplus \ldots \oplus \mathbb{R}^k}_{k\text{-mal}} \to \mathbb{R}$ durch

$$f(\vec{a}_1, \ldots, \vec{a}_k) := \det \begin{pmatrix} A & C \\ \mathcal{O} & B \end{pmatrix}$$

gegeben. Diese Funktion f ist k-fach multilinear und alternierend, es folgt

$$\begin{aligned} f(\vec{a}_1, \ldots, \vec{a}_k) &= \det(\vec{a}_1, \ldots, \vec{a}_k) \cdot f(\vec{e}_1, \ldots, \vec{e}_k) \\ &= \det(A) \cdot \det \begin{pmatrix} E & C \\ \mathcal{O} & B \end{pmatrix} = \det(A) \cdot \det(B) \end{aligned}$$

wegen Satz 5.3.10.

Die zweite Gleichung folgt aus der ersten durch Betrachtung der transponierten Matrizen. □

Man kann für $n \times n$-Matrizen A, B, C, D unter geeigneten Voraussetzungen wie etwa, daß D regulär ist und die Vertauschbarkeitsrelation $C \circ D = D \circ C$ gilt, zwar Identitäten der Art $\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(A \circ D - B \circ C)$ zeigen, weitere Aussagen sind aber im allgemeinen nicht möglich, da $\det$ als Funktion in den Matrizen $\mathbb{R}^{n \times n} \to \mathbb{R}$ nicht linear ist.

5.4 Berechnung von Determinanten

In diesem Abschnitt werden einander zwei Verfahren zur Berechnung von Determinanten gegenübergestellt. Zunächst werden wir sehen, daß wir wie schon bei der Bestimmung der Invertierbarkeit, der *Berechnung* der inversen Matrizen und wie bei der Auflösung linearer Gleichungssysteme den Gaußschen Algorithmus einsetzen können, die Grundlage dafür ist Satz 5.3.5. Das zweite Verfahren gründet sich auf den Laplaceschen Entwicklungssatz und ist von größerem *theoretischen* Interesse.

Gaußsches Verfahren zur Berechnung von Determinanten

5.4.1 Definition

Eine $n \times n$-Matrix $A = (a_{ij})_{i,j=1,\ldots,n}$ heißt **(obere) Dreiecksmatrix**, *wenn in A unterhalb der Hauptdiagonalen nur Nullen stehen (d. h. $a_{ij} = 0$ für $i > j$), und* **(untere) Dreiecksmatrix**, *wenn in A oberhalb der Hauptdiagonalen nur Nullen stehen (d. h. $a_{ij} = 0$ für $i < j$).*

Der folgende Satz zeigt, daß wir in diesem Zusammenhang die gegebene Matrix A mittels des Gaußschen Algorithmus' nicht in die Einheitsmatrix unformen müssen, sondern daß es reicht, Dreiecksgestalt anzustreben.

5.4.2 Satz

Es sei $A = (a_{ij})_{i,j=1,\ldots,n}$ eine Dreiecksmatrix. Dann gilt

$$\det(A) = \prod_{i=1}^{n} a_{ii}.$$

Die Determinante von Dreiecksmatrizen ist das Produkt der Diagonalelemente.

Beweis

Es reicht, wegen $\det(A) = \det(A^t)$ untere Dreiecksmatrizen zu betrachten, für die also $a_{ij} = 0$ ist, falls $i < j$. Der Beweis wird durch vollständige Induktion nach n geführt.

Die Induktionsbasis $n = 1$ ist trivial. Für den Induktionsschritt nehmen wir an, die Behauptung sei für $1, \ldots, n$ bereits gezeigt. Sei $A = (a_{ij})_{i,j=1,\ldots,n+1}$ eine $(n+1) \times (n+1)$-Dreiecksmatrix. Durch Zuhilfenahme von Satz 5.3.11 folgt aus der Induktionsannahme

$$\det(A) = \det \left(\begin{array}{cccc|c} a_{11} & 0 & \ldots & 0 & 0 \\ \vdots & \ddots & \ddots & 0 & \vdots \\ \vdots & & \ddots & 0 & \vdots \\ a_{n1} & \ldots & \ldots & a_{nn} & 0 \\ \hline a_{n+1,1} & \ldots & \ldots & a_{n+1,n} & a_{n+1,n+1} \end{array} \right)$$

$$= \det(a_{n+1,n+1}) \cdot \det \begin{pmatrix} a_{11} & 0 & \ldots & 0 \\ \vdots & \ddots & 0 & \vdots \\ \vdots & & \ddots & 0 \\ a_{n1} & \ldots & \ldots & a_{nn} \end{pmatrix}$$

$$= a_{n+1,n+1} \cdot (a_{11} \cdot \ldots \cdot a_{nn}) = \prod_{i=1}^{n+1} a_{ii}.$$

□

Damit stehen alle Hilfsmittel bereit, um den Gaußschen Algorithmus zur Berechnung von Determinanten darzustellen:

5.4.3 Satz

Es sei A eine beliebige $n \times n$-Matrix. Durch die elementaren Umformungen

(1) Addition des Vielfachen einer Zeile zu einer anderen *Zeile,*

(2) Spaltenvertauschungen

kann man eine der beiden folgenden Situationen herstellen:

(a) Eine Zeile von A geht in die Nullzeile über, in diesem Fall gilt $\det(A) = 0$.

(b) Die Matrix A geht in eine Dreiecksmatrix mit den Diagonalelementen $d_1, \ldots, d_n$ über. Ist k die Anzahl der durchgeführten Spaltenvertauschungen, dann gilt $\det(A) = (-1)^k \cdot d_1 \cdot \ldots \cdot d_n$.

Beweis

Hat die Matrix A selbst eine Nullzeile, so ist $\det(A) = 0$, wie wir schon in Satz 5.3.3 festgehalten hatten.

Sei nun A eine $n \times n$-Matrix ohne Nullzeilen. Gegebenenfalls nach einer Spaltenvertauschung kann man $a_{11} \neq 0$ erreichen. Durch Addition geeigneter Vielfacher der ersten Zeile zu den darunterliegenden Zeilen geht A in eine Matrix der Form

$$\left(\begin{array}{c|c} a_{11} & a_{12} \ \ldots \ a_{1n} \\ \hline 0 & \\ \vdots & A' \\ 0 & \end{array}\right)$$

über. Enthält A' eine Nullzeile, so auch die gesamte Matrix; das Verfahren wird mit dem Ergebnis $\det(A) = 0$ beendet.

Enthält auch A' keine Nullzeile, so wird das oben beschriebene Verfahren auf A' angewandt. Umformungen des Typs (1) werden innerhalb von A' durchgeführt, wegen der Nullen links in der Matrix sind es automatisch Umformungen der gesamten Matrix. Umformungen vom Typ (2) sind stets auf die gesamte Matrix zu erstrecken.

Ist im Laufe des Verfahrens keine Nullzeile aufgetreten, so hat man nach höchstens $(n - 1)$ Wiederholungen eine Dreiecksmatrix erhalten, deren Wert wir gemäß Satz 5.4.2 berechnen. Die Behauptung für $\det(A)$ folgt, weil gemäß Satz 5.3.5 Umformungen vom Typ (1) den Wert der Determinante nicht ändern. Umformungen vom Typ (2) ändern das Vorzeichen der Determinante. □

Laplacescher Entwicklungssatz

Hier werden Determinanten von „Streichungsmatrizen" eine zentrale Rolle spielen.

5.4.4 Definition

Es seien $A = (a_{ij})_{i,j=1,\ldots,n}$ eine $n \times n$-Matrix und $k, \ell \in \{1, \ldots, n\}$ zwei Indizes. Durch Streichen der k-ten Zeile und ℓ-ten Spalte entsteht eine $(n-1) \times (n-1)$-Matrix, die mit $A_{k\ell}$ bezeichnet wird. Deren Determinante $d_{k\ell} := \det(A_{k\ell})$ heißt (k, ℓ)-ter **Cofaktor** *von A.*

5.4.5 Satz (Laplacescher Entwicklungssatz)

Für die Determinante jeder $n \times n$-Matrix A gilt die

(1) Entwicklung nach einer beliebigen Spalte: Für jeden Index $\ell \in \{1, \dots, n\}$ ist

$$\det(A) = \sum_{k=1}^{n} (-1)^{k+\ell} a_{k\ell} d_{k\ell}.$$

(2) und die Entwicklung nach einer beliebigen Zeile: Für jeden Index $k \in \{1, \dots, n\}$ ist

$$\det(A) = \sum_{\ell=1}^{n} (-1)^{k+\ell} a_{k\ell} d_{k\ell}.$$

Beweis

Wir betrachten zunächst die Entwicklung nach der ersten Spalte, sei also $\ell = 1$. Nach Definition gilt

$$\begin{aligned}\det(A) &= \sum_{\substack{i_1,\dots,i_n=1 \\ \text{paarweise verschieden}}}^{n} \delta(i_1, \dots, i_n) \cdot a_{i_1 1} \cdot \ldots \cdot a_{i_n n} \\ &= \sum_{k=1}^{n} a_{k1} \cdot \Big(\sum_{\substack{i_2,\dots,i_n=1 \\ \text{paarweise verschieden} \\ \text{ungleich } k}}^{n} \delta(k, i_2, \dots, i_n) \cdot a_{i_2 2} \cdot \ldots \cdot a_{i_n n} \Big).\end{aligned}$$

Zur Berechung von $\delta(k, i_2, \dots, i_n)$: Die Anordnung $i_2, \dots, i_n$ geht durch eine Anzahl a einfacher Vertauschungen in natürliche Anordnung $1, \dots, k-1, k+1, \dots, n$ über. Die Anordnung $k, 1, \dots, k-1, k+1, \dots, n$ kann durch $(k-1)$-maliges Vertauschen von k mit seinem jeweiligen Nachfolger in natürliche Anordnung $1, \dots, n$ überführt werden. Für $k, i_2, \dots, i_n$ kann natürliche Anordnung durch $a + k - 1$ einfache Vertauschungen erreicht werden. Zusammenfassend gilt:

$$\begin{aligned}\delta(k, i_2, \dots, i_n) &= (-1)^{a+k-1} = (-1)^{k-1} \cdot (-1)^a \\ &= (-1)^{k+1} \cdot (-1)^a = (-1)^{k+1} \cdot \delta(i_2, \dots, i_n),\end{aligned}$$

und folglich

$$\begin{aligned}\det(A) &= \sum_{k=1}^{n} (-1)^{k+1} a_{k1} \cdot \Big(\sum_{\substack{i_2,\dots,i_n=1 \\ \text{alle } i_j \neq k}}^{n} \delta(i_2, \dots, i_n) a_{i_2 2} \cdot \ldots \cdot a_{i_n n} \Big) \\ &= \sum_{k=1}^{n} (-1)^{k+1} a_{k1} d_{k1},\end{aligned}$$

denn es ist

$$d_{k1} = \det(A_{k1}) = \det(a_{ij})_{\substack{i=1,\ldots,k-1,k+1,\ldots,n \\ j=2,\ldots,n}}.$$

Sei nun $\ell \in \{1, \ldots, n\}$ beliebig, $A = (\vec{a}_1, \ldots, \vec{a}_n)$. Durch $(\ell-1)$-maliges Vertauschen von $\vec{a}_\ell$ mit seinem jeweiligen Vorgänger folgt

$$\det(A) = (-1)^{\ell-1} \cdot \det(\vec{a}_\ell, \vec{a}_1, \ldots, \vec{a}_{\ell-1}, \vec{a}_{\ell+1}, \ldots, \vec{a}_n).$$

Sei $\tilde{d}_{k1}$ der $(k,1)$-te Cofaktor von $\tilde{A} = (\vec{a}_\ell, \vec{a}_1, \ldots, \vec{a}_{\ell-1}, \vec{a}_{\ell+1}, \ldots, \vec{a}_n)$. Diese Matrix geht durch Streichen der k-ten Zeile und der ersten Spalte in $A_{k\ell}$ über, also ist $\tilde{d}_{k1} = d_{k\ell}$. Durch Entwicklung von $\det(\tilde{A})$ nach der ersten Spalte folgt:

$$\det(A) = (-1)^{\ell-1} \det(\tilde{A}) = (-1)^{\ell-1} \left(\sum_{k=1}^{n} (-1)^{k+1} a_{k\ell} \tilde{d}_{k1} \right) = \sum_{k=1}^{n} (-1)^{k+\ell} a_{k\ell} d_{k\ell}.$$

Da transponierte Matrizen jeweils dieselbe Determinante haben, läßt sich unser Ergebnis über die Entwicklung nach Spalten schnell auf die Entwicklung nach Zeilen ausdehnen. Sei $d^t_{\ell k}$ der (ℓ,k)-te Cofaktor von A^t, wegen $(A^t)_{\ell k} = (A_{k\ell})^t$ gilt $d^t_{\ell k} = d_{k\ell}$. Durch Entwicklung von $\det(A^t)$ nach der k-ten Spalte erhält man:

$$\det(A) = \det(A^t) = \sum_{\ell=1}^{n} (-1)^{k+\ell} a_{k\ell} d^t_{\ell k} = \sum_{\ell=1}^{n} (-1)^{k+\ell} a_{k\ell} d_{k\ell}. \qquad \square$$

5.4.6 Beispiel

Wir berechnen die Determinante der 3×3-Matrix

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{pmatrix}$$

mittels

(a) des Gaußschen Verfahrens:

$$A \to \begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & -5 \\ 0 & -5 & -7 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & -5 \\ 0 & 0 & 18 \end{pmatrix}$$

Da keine Spaltenvertauschungen durchgeführt wurden, gilt:

$$\det(A) = 1 \cdot (-1) \cdot 18 = -18.$$

(b) und des Entwicklungssatzes: Wir entwickeln nach der ersten Spalte.

$$\begin{aligned}\det\begin{pmatrix}1&2&3\\2&3&1\\3&1&2\end{pmatrix} &= 1\cdot\det\begin{pmatrix}3&1\\1&2\end{pmatrix} - 2\cdot\det\begin{pmatrix}2&3\\1&2\end{pmatrix} + 3\cdot\det\begin{pmatrix}2&3\\3&1\end{pmatrix}\\ &= (3\cdot 2 - 1\cdot 1) - 2\cdot(2\cdot 2 - 1\cdot 3) + 3\cdot(2\cdot 1 - 3\cdot 3)\\ &= 5 - 2 - 21 = -18.\end{aligned}$$

Zur Berechnung konkret angegebener Determinanten ist der Gaußsche Algorithmus wesentlich einfacher und schneller. Bei $(n-1)$-maliger Anwendung des Entwicklungssatzes auf die jeweils entstehenden quadratischen Untermatrizen von A mit n, $n-1,\ldots,2$ Zeilen gelangt man letztlich wieder zu der Definition 5.2.1 der Determinante.

Für das nächste Beispiel erinnern wir an die Definition des Polynoms

$$D(r_1,\ldots,r_n) = \prod_{1\le i<j\le n}(r_j - r_i)$$

zu Beginn von Abschnitt 5.1.

5.4.7 Satz (Vandermondesche Determinante)

$$\det\begin{pmatrix}1 & r_1 & r_1^2 & \ldots & r_1^{n-1}\\ \vdots & \vdots & \vdots & & \vdots\\ 1 & r_n & r_n^2 & \ldots & r_n^{n-1}\end{pmatrix} = D(r_1,\ldots,r_n).$$

Beweis

Durch Induktion nach $n \ge 2$. Für $n = 2$ ist

$$\det\begin{pmatrix}1 & r_1\\ 1 & r_2\end{pmatrix} = r_2 - r_1 = D(r_1, r_2).$$

Sei nun $n \ge 3$ und die Behauptung für $n-1$ bewiesen. Wir ziehen in

$$A = \begin{pmatrix}1 & r_1 & r_1^2 & \ldots & r_1^{n-1}\\ \vdots & \vdots & \vdots & & \vdots\\ 1 & r_n & r_n^2 & \ldots & r_n^{n-1}\end{pmatrix}$$

das r_1-fache der $(n-1)$-ten Spalte von der n-ten Spalte ab. In dieser neuen Matrix subtrahieren wir das r_1-fache der $(n-2)$-ten Spalte von der $(n-1)$-ten. So fahren wir fort und erhalten schließlich:

$$A \to B = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & r_2 - r_1 & \dots & r_2^{n-1} - r_1 r_2^{n-2} \\ \vdots & \vdots & & \vdots \\ 1 & r_n - r_1 & \dots & r_n^{n-1} - r_1 r_n^{n-2} \end{pmatrix}$$

$$= \left(\begin{array}{c|ccc} 1 & 0 & \dots & 0 \\ \hline 1 & r_2 - r_1 & \dots & (r_2 - r_1) r_2^{n-2} \\ \vdots & \vdots & & \vdots \\ 1 & r_n - r_1 & \dots & (r_n - r_1) r_n^{n-2} \end{array}\right).$$

Folglich gilt auf Grund von Satz 5.3.9 über Determinanten mit einem $\mathcal{O}$-Block, der Multilinearität von $\det$ und der Induktionsannahme

$$\begin{aligned} \det(A) &= \det(B) \\ &= (r_2 - r_1) \cdot \ldots \cdot (r_n - r_1) \cdot \det \begin{pmatrix} 1 & r_2 & r_2^2 & \dots & r_2^{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & r_n & r_n^2 & \dots & r_n^{n-2} \end{pmatrix} \\ &= (r_2 - r_1) \cdot \ldots \cdot (r_n - r_1) \cdot \prod_{n \geq j > i \geq 2} (r_j - r_i) \\ &= \prod_{n \geq j > i \geq 1} (r_j - r_i) = D(r_1, \dots, r_n). \end{aligned}$$

□

5.5 Cramersche Regel

In diesem Paragraphen werden wir aus dem Laplaceschen Entwicklungssatz einige insbesondere für die reine Mathematik wichtige Sätze herleiten. So wird etwa in Satz 5.5.4 eine explizite Formel zur Bestimmung von A^{-1} angegeben, aus der u. a. hervorgeht, daß im Falle der Existenz die Komponenten von A^{-1} rationale Funktionen der Komponenten von A sind. Damit kann man beispielsweise quantitativ angeben, inwiefern „kleine" Änderungen bei A „kleine" Änderungen bei A^{-1} nach sich ziehen.

Die Adjunkte

5.5.1 Definition

Für $n \times n$-Matrizen A ist die **Adjunkte** $\operatorname{ad} A$ *die transponierte, mit alternierendem Vorzeichen versehene Cofaktormatrix*

$$\operatorname{ad} A := \left(\left((-1)^{k+\ell} d_{k\ell} \right)_{k,\ell=1,\dots,n} \right)^t.$$

5.5.2 Beispiel

Für

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

lautet

$$\text{ad}\, A = \begin{pmatrix} (-3) & -(-6) & (-3) \\ -(-6) & (-12) & -(-6) \\ (-3) & -(-6) & (-3) \end{pmatrix}^t = \begin{pmatrix} -3 & 6 & -3 \\ 6 & -12 & 6 \\ -3 & 6 & -3 \end{pmatrix}.$$

5.5.3 Satz

Für alle $n \times n$-Matrizen A gilt:

$$\text{ad}\, A \circ A = A \circ \text{ad}\, A = \det(A) \cdot E.$$

Beweis

Wir zeigen zunächst $A \circ \text{ad}\, A = \det(A) \cdot E$, d. h., daß für alle $i, k = 1, \dots, n$ gilt:

$$\sum_{\ell=1}^{n} a_{i\ell}((-1)^{k+\ell} d_{k\ell}) = \delta_{ik} \det(A).$$

Für $i = k$ liefert die Entwicklung nach der k-ten Zeile die Behauptung:

$$\det(A) = \sum_{\ell=1}^{n} (-1)^{k+\ell} a_{k\ell} d_{k\ell}.$$

Sei nun $i \neq k$. Wir betrachten die Matrix A', die in der k-ten Zeile die i-te Zeile von A und sonst dieselben Zeilen wie A enthält:

$$A' = \begin{array}{r} \\ \\ i\text{-te Zeile} \\ \\ k\text{-te Zeile} \\ \\ \\ \end{array} \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{i1} & \dots & a_{in} \\ \vdots & & \vdots \\ a_{i1} & \dots & a_{in} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}$$

Da A' zwei gleiche Zeilen enthält, ist $\det(A') = 0 = \delta_{ik} \det(A)$. Andererseits liefert die Entwicklung von $\det(A')$ nach der k-ten Zeile:

$$\det(A') = \sum_{\ell=1}^{n} (-1)^{k+\ell} a_{i\ell} \cdot d_{k\ell}.$$

Damit haben wir die Gültigkeit von $A \circ \operatorname{ad} A = \det(A) \cdot E$ gezeigt. Durch Transponieren folgt daraus $(\operatorname{ad} A)^t \circ A^t = \det(A) \cdot E$ und, indem man A durch A^t ersetzt, $(\operatorname{ad} A^t)^t \circ A = \det(A) \cdot E$. Im Beweis des Laplaceschen Entwicklungssatzes hatten wir gezeigt, daß der (k, ℓ)-te Cofaktor von A^t der (ℓ, k)-te Cofaktor von A ist, daß also $\operatorname{ad} A^t = (\operatorname{ad} A)^t$ gilt. Es folgt $\operatorname{ad} A \circ A = \det(A) \cdot E$. □

5.5.4 Satz

Für alle nichtsingulären $n \times n$-Matrizen A gilt:

$$A^{-1} = \frac{1}{\det(A)} \cdot \operatorname{ad} A.$$

Beweis

Da voraussetzungsgemäß A^{-1} existiert, kann man in Satz 5.5.3 mit A^{-1} und mit $\frac{1}{\det(A)}$ multiplizieren. □

Im Anschluß an die Definition 5.2.1 hatten wir erläutert, daß die Determinante als Polynom mit den n^2-Matrixkomponenten als Veränderlichen angesehen werden kann. Ebenso sind die Cofaktoren d_{ij} (Unter-) Determinanten und damit Polynome. Wegen Satz 5.5.4 ist also jede Komponente von A^{-1} eine rationale Funktion in den $m = n^2$ Veränderlichen $(x_1, \dots, x_m) = (a_{11}, \dots, a_{1n}, \dots, a_{n1}, \dots, a_{nn})$ gemäß der folgenden Definition.

5.5.5 Definition

Seien P, Q Polynome in m Veränderlichen, Q sei nicht das Nullpolynom. Der Quotient $\frac{P}{Q}$ heißt **rationale Funktion***.*

5.5.6 Beispiel

Die zweireihige Matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} =: \begin{pmatrix} x_1 & x_2 \\ x_3 & x_4 \end{pmatrix}$ sei nichtsingulär, es ist also $\det(A) = x_1 x_4 - x_2 x_3 \neq 0$. Für die Cofaktoren berechnet man $d_{11} = x_4$, $d_{12} = x_3$, $d_{21} = x_2$, $d_{22} = x_1$, die Adjunkte lautet

$$\operatorname{ad} A = \begin{pmatrix} x_4 & -x_3 \\ -x_2 & x_1 \end{pmatrix}^t = \begin{pmatrix} x_4 & -x_2 \\ -x_3 & x_1 \end{pmatrix}.$$

Für

$$A^{-1} =: \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}$$

erhält man also aus Satz 5.5.4:

$$a' = \frac{x_4}{x_1x_4 - x_2x_3},\ b' = \frac{-x_2}{x_1x_4 - x_2x_3},\ c' = \frac{-x_3}{x_1x_4 - x_2x_3},\ d' = \frac{x_1}{x_1x_4 - x_2x_3}.$$

5.5.7 Beispiel

In Beispiel 3.8.7 hatten wir für

$$A = \begin{pmatrix} -1 & 1 & 1 \\ -3 & 4 & 3 \\ -4 & 6 & 5 \end{pmatrix}$$

die inverse Matrix mit

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

berechnet. Daß sowohl A als auch A^{-1} nur ganzzahlige Komponenten haben, erklärt sich aus Satz 5.5.4 und

$$\det(A) = \det \begin{pmatrix} -1 & 0 & 0 \\ -3 & 1 & 0 \\ -4 & 2 & 1 \end{pmatrix} = -1.$$

Hier haben wir die erste Spalte von A zur zweiten und zur dritten Spalte addiert.

Allgemein gilt sogar umgekehrt: Haben A und A^{-1} ganzzahlige Komponenten, so sind $\det(A)$, $\det(A^{-1}) \in \mathbb{Z}$, da $\det$ ein Polynom in den Matrixkomponenten mit ganzzahligen Koeffizienten ist. Es folgt $1 = \det(A) \cdot \det(A^{-1})$, und diese Gleichung hat in $\mathbb{Z}$ nur die Lösungen $\det(A), \det(A^{-1}) = \pm 1$.

Auflösung von Gleichungssystemen

Der Zusammenhang zwischen Adjunkten und inversen Matrizen führt zu einem weiteren Verfahren zur Auflösung von linearen Gleichungssystemen $A \circ \vec{x} = \vec{b}$, wobei $A = A^{n,n}$ eine $n \times n$-Matrix und $\vec{b} \in \mathbb{R}^n$ ein (Spalten-) Vektor ist. Für das Folgende setzen wir stets voraus, daß die Matrix A nichtsingulär (invertierbar) ist. Dann wird durch A ein Isomorphismus $F : \mathbb{R}^n \to \mathbb{R}^n$ gegeben, zu jedem $\vec{b} \in \mathbb{R}^n$ existiert also genau eine Lösung $\vec{x} \in \mathbb{R}^n$ von $A \circ \vec{x} = \vec{b}$, für diese gilt $\vec{x} = A^{-1} \circ \vec{b}$. Durch Verwendung von Satz 5.5.4 ergibt sich:

5.5.8 Satz (Cramersche Regel)

Es sei A eine nichtsinguläre $n \times n$-Matrix mit den Spalten $\vec{a}_1, \dots, \vec{a}_n$, ferner sei $\vec{b} \in \mathbb{R}^n$. Für $j = 1, \dots, n$ setze man

$$B_j := (\vec{a}_1, \dots, \vec{a}_{j-1}, \vec{b}, \vec{a}_{j+1}, \dots, \vec{a}_n).$$

Dann gilt für die eindeutig bestimmte Lösung $\vec{x} \in \mathbb{R}^n$ des linearen Gleichungssystems $A \circ \vec{x} = \vec{b}$:

$$x_j = \frac{\det(B_j)}{\det(A)}, \qquad j = 1, \dots, n.$$

Beweis

Es bezeichne $(c_{ji})_{\substack{j=1,\dots,n \\ i=1,\dots,n}} := A^{-1}$, gemäß Satz 5.5.4 gilt

$$c_{ji} = (-1)^{i+j} \frac{d_{ij}}{\det(A)}.$$

Die Lösung des Gleichungssystems ist durch $\vec{x} = A^{-1} \circ \vec{b}$ gegeben, also folgt:

$$x_j = \sum_{i=1}^{n} c_{ji} b_i = \frac{1}{\det(A)} \cdot \sum_{i=1}^{n} (-1)^{i+j} b_i d_{ij} = \frac{\det(B_j)}{\det(A)},$$

denn die Entwicklung von $\det(B_j)$ nach der j-ten Spalte ergibt:

$$\det(B_j) = \sum_{i=1}^{n} (-1)^{i+j} b_i d_{ij}. \qquad \square$$

5.5.9 Beispiel

Sei

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad \text{und} \quad \vec{b} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}.$$

Es gilt $\det(A) = 4 - 6 = -2$, die Cramersche Regel kann also verwendet werden. Es ist $B_1 = \begin{pmatrix} 2 & 2 \\ 3 & 4 \end{pmatrix}$, $B_2 = \begin{pmatrix} 1 & 2 \\ 3 & 3 \end{pmatrix}$, $\det(B_1) = 2$, $\det(B_2) = -3$, wir erhalten

$$x_1 = \frac{2}{-2} = -1 \quad \text{und} \quad x_2 = \frac{-3}{-2} = \frac{3}{2}.$$

Somit löst $(-1, \frac{3}{2})$ das Gleichungssystem

$$\begin{aligned} x_1 + 2x_2 &= 2 \\ 3x_1 + 4x_2 &= 3. \end{aligned}$$

Unterdeterminanten und Rang einer Matrix

Es wird ein Zusammenhang zwischen dem Rang einer nicht notwendig quadratischen Matrix und deren größtmöglichen quadratischen nichtsingulären Untermatrizen hergestellt.

5.5.10 Definition

Sei $A = (a_{ij})_{\substack{i=1,\ldots,m\\ j=1,\ldots,n}}$ eine $m \times n$-Matrix. Eine $d \times d$-Matrix C mit $d \leq \min(m,n)$ heißt d-dimensionale **Untermatrix** *von A, wenn d Zeilenindizes $1 \leq i_1 < i_2 < \ldots < i_d \leq m$ und d Spaltenindizes $1 \leq j_1 < \ldots < j_d \leq n$ existieren, so daß gilt:*

$$C = \begin{pmatrix} a_{i_1j_1} & \ldots & a_{i_1j_d} \\ \vdots & & \vdots \\ a_{i_dj_1} & \ldots & a_{i_dj_d} \end{pmatrix}.$$

Unter einer d-dimensionalen **Unterdeterminante** *von A versteht man die Determinante einer d-dimensionalen Untermatrix von A.*

Die Untermatrix C geht aus A hervor, indem man die $m-d$ Zeilen von A mit Index $i \neq \{i_1, \ldots, i_d\}$ und die $n-d$ Spalten von A mit Index $j \notin \{j_1, \ldots, j_d\}$ streicht.

Die in Definition 5.4.4 für n-reihige quadratische Matrizen A erklärten „Sreichungs"-Matrizen A_{ij} sind $(n-1)$-dimensionale Untermatrizen von A, die Cofaktoren d_{ij} sind $(n-1)$-dimensionale Unterdeterminanten von A.

5.5.11 Definition

Es sei A eine $m \times n$-Matrix. Ist A nicht die Nullmatrix, so bezeichnen wir mit $d(A)$ die größte Zahl $d \in \mathbb{N}$, für die eine von 0 verschiedene d-dimensionale Unterdeterminante von A existiert. Für die Nullmatrix setzen wir $d(\mathcal{O}^{m,n}) := 0$.

5.5.12 Satz

Für jede $m \times n$-Matrix A gilt $\mathrm{Rang}\,(A) = d(A)$.

Beweis

Es bezeichnen $\vec{a}_1, \ldots, \vec{a}_n$ die Spalten von A, $d := d(A)$ und $r := \mathrm{Rang}\,(A)$. Wegen der Äquivalenz $r = 0 \Leftrightarrow A = \mathcal{O} \Leftrightarrow d = 0$ können wir ohne Einschränkung $r > 0$ und $d > 0$ annehmen.

„$d \leq r$": Dieser Beweisteil ist einfach und beruht darauf, daß die lineare Unabhängigkeit von Vektoren aus $\mathbb{R}^d$ durch Hinzufügen weiterer Komponenten nicht verlorengeht. Durch Zeilen- und Spaltenvertauschungen werden r und d nicht geändert wir können also ohne Einschränkung

$$\det \begin{pmatrix} a_{11} & \ldots & a_{1d} \\ \vdots & & \vdots \\ a_{d1} & \ldots & a_{dd} \end{pmatrix} \neq 0$$

annehmen. Die Spalten dieser Matrix sind linear unabhängig und damit $\vec{a}_1, \dots, \vec{a}_d \in \mathbb{R}^m$ ebenfalls. Folglich ist $r \geq d$.

„$r \leq d$": Da Spaltenvertauschungen den Rang einer Matrix nicht ändern, können wir annehmen, daß $\vec{a}_1, \dots, \vec{a}_r$ linear unabhängig sind.

Wir bezeichnen $A' := (\vec{a}_1, \dots, \vec{a}_r)$ und $d' := d(A')$, offenbar ist $d' \leq r \leq m$. Da jede Unterdeterminante von A' auch Unterdeterminante von A ist, gilt $d' \leq d$. Es reicht also zu zeigen: $r \leq d'$. Dazu wird ein Widerspruchsbeweis geführt.

Angenommen, es wäre $r > d'$. Da wir nur den Fall $r \geq 1$ betrachten, ist offensichtlich auch $d' \geq 1$. Wir können, gegebenenfalls wieder nach Zeilen- und Spaltenvertauschungen, ohne Einschränkung von

$$\det \begin{pmatrix} a_{11} & \dots & a_{1d'} \\ \vdots & & \vdots \\ a_{d'1} & \dots & a_{d'd'} \end{pmatrix} \neq 0 \qquad (*)$$

ausgehen. Für $i = 1, \dots, m$ setzen wir

$$B_i := B_i^{d'+1,d'+1} := \begin{pmatrix} a_{11} & \dots & a_{1d'} & a_{1,d'+1} \\ \vdots & & \vdots & \vdots \\ a_{d'1} & \dots & a_{d'd'} & a_{d',d'+1} \\ a_{i1} & \dots & a_{id'} & a_{i,d'+1} \end{pmatrix}.$$

Wegen der Annahme $d' < r$ ist diese Setzung möglich, und es gilt stets $\det(B_i) = 0$. Die Entwicklung von $\det(B_i)$ nach der $(d'+1)$-ten Zeile ergibt:

$$0 = \det(B_i) = \sum_{j=1}^{d'+1} (-1)^{d'+1+j} a_{ij} \cdot d_j, \quad i = 1, \dots, m,$$

mit

$$d_j = \det \begin{pmatrix} a_{11} & \dots & a_{1,j-1} & a_{1,j+1} & \dots & a_{1,d'+1} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{d'1} & \dots & a_{d',j-1} & a_{d',j+1} & \dots & a_{d',d'+1} \end{pmatrix}.$$

Diese m Gleichungen werden zu einer Vektorgleichung zusammengefaßt:

$$\sum_{j=1}^{d'+1} (-1)^{d'+1+j} d_j \vec{a}_j = \vec{0}_{\mathbb{R}^m}.$$

Gemäß $(*)$ ist $d_{d'+1} \neq 0$; die Spalten $\vec{a}_1, \dots, \vec{a}_{d'+1}$ sind linear abhängig. Wir erhalten einen Widerspruch zur linearen Unabhängigkeit von $\vec{a}_1, \dots, \vec{a}_r$.

Also gilt $r \leq d'$ und folglich $r \leq d$. □

5.5.13 Beispiel

Mit Hilfe dieses Satzes soll der Rang der Matrix

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

bestimmt werden, vgl. Beispiel 3.6.15. Mit einem der in Abschnitt 5.4 entwickelten Verfahren berechnet man $\det(A) = 0$, es ist also $d(A) \leq 2$. Für die zweidimensionale Untermatrix $A' = \begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix}$ gilt: $\det(A') = -3 \neq 0$, folglich ist $d(A) \geq 2$. Man erhält $2 = d(A) = \mathrm{Rang}\,(A)$.

5.6 Die Determinante einer linearen Abbildung $F : V \to V$

In den Kapiteln 3 und 4 haben wir stets die Theorie linearer Abbildungen und den Matrizenkalkül simultan entwickelt. In diesem Sinne soll nun die Determinantendefinition auf lineare Abbildungen endlichdimensionaler Vektorräume in sich ausgedehnt werden.

Sei V ein n-dimensionaler Vektorraum und $F : V \to V$ eine lineare Abbildung. Wir betrachten eine Karte $H : \mathbb{R}^n \xrightarrow{\sim} V$, das entsprechende Koordinatensystem $H^{-1} : V \xrightarrow{\sim} \mathbb{R}^n$ und die lineare Abbildung F in den Koordinaten: $H^{-1} \circ F \circ H : \mathbb{R}^n \to \mathbb{R}^n$, (vgl. Kapitel 3.2).

Es liegt nahe, die Determinante von F als Determinante der zur Abbildung in den Koordinaten $H^{-1} \circ F \circ H$ gehörigen $n \times n$-Matrix A zu erklären: $\det(F) := \det(A)$. Zuvor ist allerdings zu zeigen, daß diese Setzung nicht von der Wahl der Karte H abhängt.

Zur Klärung der Wohldefiniertheit betrachten wir eine weitere Karte $\hat{H} : \mathbb{R}^n \xrightarrow{\sim} V$. In diesen Koordinaten gehört zu $\hat{H}^{-1} \circ F \circ \hat{H}$ eine Matrix $\hat{A}$. Die zur Koordinatenwechselabbildung $H^{-1} \circ \hat{H} : \mathbb{R}^n \to \mathbb{R}^n$ gehörige reguläre $n \times n$-Matrix werde mit B bezeichnet. Wie im Abschnitt „Koordinatenwechsel" in Kapitel 3.6 berechnet man:

$$\begin{aligned} \hat{H}^{-1} \circ F \circ \hat{H} &= \hat{H}^{-1} \circ H \circ H^{-1} \circ F \circ H \circ H^{-1} \circ \hat{H} = \\ &= (H^{-1} \circ \hat{H})^{-1} \circ (H^{-1} \circ F \circ H) \circ (H^{-1} \circ \hat{H}), \end{aligned}$$

also $\hat{A} = B^{-1} \circ A \circ B$. Der Determinanten-Multiplikationssatz zeigt schließlich:

$$\det(\hat{A}) = \det(B^{-1}) \cdot \det(A) \cdot \det(B) = \frac{\det(B)}{\det(B)} \cdot \det(A) = \det(A).$$

Die folgende Definition ist also sinnvoll:

5.6.1 Definition

Es sei $F : V \to V$ eine lineare Abbildung des n-dimensionalen Vektorraums V in sich. Sei $H : \mathbb{R}^n \xrightarrow{\sim} V$ eine beliebige Karte und A die zu $H^{-1} \circ F \circ H$ gehörige $n \times n$-Matrix. Die **Determinante** *von F wird durch $\det(F) := \det(A)$ erklärt.*

Mit dieser Definition finden alle Gesetze für Determinanten von Matrizen ihre Entsprechung bei Determinanten von linearen Abbildungen. Im folgenden Satz wird davon eine Auswahl zusammengestellt:

5.6.2 Satz

Es seien $F, G : V \to V$ lineare Abbildungen des n-dimensionalen Vektorraums V in sich. Dann gilt:

(a) $\det(id_V) = 1$.

(b) $\det(F \circ G) = \det(F) \cdot \det(G)$.

(c) *Genau dann ist F ein Isomorphismus, wenn $\det(F) \neq 0$ gilt.*

Beweis

Sei $H : \mathbb{R}^n \xrightarrow{\sim} V$ eine Karte und A bzw. B die zu $H^{-1} \circ F \circ H$ bzw. $H^{-1} \circ G \circ H$ gehörige Matrix.

(a) Folgt aus $H^{-1} \circ id_V \circ H = id_{\mathbb{R}^n}$.

(b) Die Matrix $A \circ B$ gehört zu $H^{-1} \circ F \circ H \circ H^{-1} \circ G \circ H = H^{-1} \circ (F \circ G) \circ H$, d. h. zur Abbildung $F \circ G$ in den Koordinaten. Der Determinanten-Multiplikationssatz 5.3.6 ergibt $\det(F \circ G) \stackrel{\text{def}}{=} \det(A \circ B) = \det(A) \cdot \det(B) \stackrel{\text{def}}{=} \det(F) \cdot \det(G)$.

(c) Unter Verwendung von Satz 5.3.7 erhalten wir die folgende Kette äquivalenter Aussagen: F ist ein Isomorphismus $\Leftrightarrow$ $H^{-1} \circ F \circ H$ ist ein Isomorphismus $\Leftrightarrow$ A ist nichtsingulär $\Leftrightarrow$ $\det(A) \neq 0 \Leftrightarrow \det(F) \neq 0$. □

6 Orthogonalität

6.1 Die euklidische Ebene

Wir knüpfen hier an Kapitel 2 an. Es sei wieder E die anschauliche Ebene und $O \in E$ der Basispunkt. Wir haben den Vektorraum der von O ausgehenden Pfeilvektoren in $E:\ V = \{(O,P) : P \in E\}$. Bislang sind in Kapitel 2 zwölf Axiome für V aufgestellt worden. Wir übernehmen sie. Aus ihnen wurde abgeleitet:

V ist ein 2-dimensionaler Vektorraum.

Bislang hat V nur die *lineare Geometrie*. Es fehlt der Abstandsbegriff, der Winkelbegriff, das Senkrechtstehen. Wir müssen deshalb das Axiomensystem ergänzen.

Wir werden V nicht sogleich mit einem Skalarprodukt versehen, sondern den Begriff des Winkels zugrundelegen. Dann kann man auf V auch ein Skalarprodukt festlegen. Es ist bis auf eine Eichtransformation bestimmt. Wir gehen vom anschaulichen Begriff der *Drehung* aus.

Axiome der Drehungen

In der räumlichen Anschauung ist evident, daß es Drehungen um $O \in E$ gibt. Wir werden für diese sechs weitere Axiome aufstellen.

Axiom XIII

Zu jedem $\Theta \in \mathbb{R}$ ist eine Abbildung $D_\Theta : E \to E$ mit $O \mapsto O$ definiert.

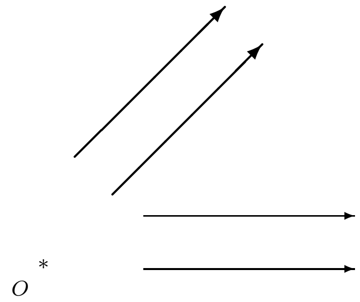
In der räumlichen Anschauung bedeutet die Abbildung D_Θ die Drehung von E um O mit dem Winkel Θ. Wir wollen den Winkel in gewöhnlichen Grad messen. Die Drehung um 360 Grad ist also gerade wieder die Identität. Ist $\Theta > 0$ so erfolgt die Drehung im mathematisch positiven Sinn, ist $\Theta < 0$, so geschieht sie negativ, also wie bei einer Uhr. Natürlich ist durch die Drehung Θ nicht eindeutig festgelegt. Man kann ein Vielfaches von 360 dazu addieren oder davon abziehen.

Die Abbildung D_Θ ergibt auch immer eine Abbbildung der Vektoren. Wir setzen einfach $\hat{D}_\Theta(P,Q) = \Big(D_\Theta(P), D_\Theta(Q)\Big)$ für $P,Q \in E$. Wir erhalten $\hat{D}_\Theta(O,P) = \Big(O, D_\Theta(P)\Big) \in V$. Also wirft die Abbildung $\hat{D}_\Theta$ den Vektorraum V in sich selbst.

Es ist auch evident, daß $\hat{D}_\Theta$ „parallel“ invariant läßt:

Axiom XIV

Ist $(P,Q) \parallel (P_0,Q_0)$*, so folgt* $\hat{D}_\Theta(P,Q) \parallel \hat{D}_\Theta(P_0,Q_0)$.

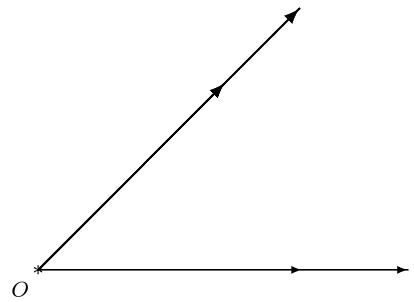


Drehung paralleler Vektoren um 45°

Die Vergrößerung mit einem Faktor a vor der Drehung ergibt eine Vergrößerung mit a nach der Drehung.

Axiom XV

Wenn $a \in \mathbb{R}_0^+$ *ist, folgt* $\hat{D}_\Theta\Big(a \cdot (O,P)\Big) = a \cdot \hat{D}_\Theta(O,P)$.



Verlängerung eines Vektors, Drehung

Wir beweisen nun, daß alle Drehungen lineare Abbildungen $V \to V$ sind.

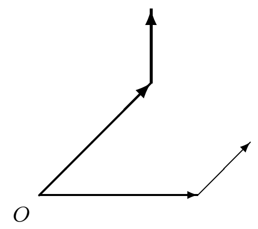
6.1.1 Satz

Jede Drehung $\hat{D}_\Theta : V \to V$ *ist eine lineare Abbildung.*

Beweis

(1) Wir müssen zeigen, daß $\hat{D}_\Theta\Big((O,P) + (O,Q)\Big) = \hat{D}_\Theta(O,P) + \hat{D}_\Theta(O,Q)$ ist. Es gibt genau einen Punkt $R \in E$ mit $(O,Q) \parallel (P,R)$. Das folgt aus Axiom III. Wir erhalten $(O,P) + (O,Q) = (O,P) + (P,R) = (O,R)$. Weiter

haben wir $\Big(O, D(Q)\Big) \parallel \Big(D(P), D(R)\Big) = \hat{D}(P,R)$. Das bedeutet $\hat{D}(O,P) + \hat{D}(O,Q) = \Big(O, D(P)\Big) + \Big(D(P), D(R)\Big) = \Big(O, D(R)\Big) = \hat{D}(O,R)$. Also ist $\hat{D} : (V,+) \to (V,+)$ ein Gruppenhomomorphismus.



Vektoraddition, Drehung um $45°$

(2) Wir müssen weiter beweisen: Für $a \in \mathbb{R}$ ist $a \cdot \hat{D}(O,P) = \hat{D}\Big(a \cdot (O,P)\Big)$. Wenn $a \geq 0$ ist, ist das gerade Axiom XV. Also setzen wir $a < 0$ voraus. Dann folgt: $\hat{D}\Big(a \cdot (O,P)\Big) = \hat{D}\Big(-((-a) \cdot (O,P))\Big) = -\hat{D}\Big((-a) \cdot (O,P)\Big)$, da $\hat{D}$ ein Gruppenhomomorphismus ist. Das heißt $\hat{D}\Big(a \cdot (0,P)\Big) = -\Big((-a) \cdot \hat{D}(O,P)\Big) = a \cdot \hat{D}(O,P)$. $\square$

In der räumlichen Anschauung ist evident:

(a) Wenn wir eine Drehung um den Winkel Θ_2 ausführen und dann eine Drehung um den Winkel Θ_1, so erhalten wir eine Drehung um den Winkel $\Theta_1 + \Theta_2$.

(b) Zwei Drehungen D_{Θ_1} und D_{Θ_2} sind genau dann gleich, wenn $\Theta_2 - \Theta_1 = k \cdot 360$ mit $k \in \mathbb{Z}$ ist.

Wir haben also noch die folgenden Axiome:

Axiom XVI

$D_{\Theta_1} \circ D_{\Theta_2} = D_{\Theta_1 + \Theta_2}$.

Axiom XVII

Die Abbildung D_0 ist die Identität. Es gilt $D_{\Theta_1} = D_{\Theta_2}$ dann und nur dann, wenn $\Theta_2 = \Theta_1 + k \cdot 360$ ($k \in \mathbb{Z}$).

Es gelten die gleichen Aussagen für die Abbildungen $\hat{D}_\Theta$. Man hat: $\hat{D}_{\Theta_1} \circ \hat{D}_{\Theta_2} = \hat{D}_{\Theta_1+\Theta_2}$, $\hat{D}_0 = id : V \to V$, $\hat{D}_{\Theta_1} = \hat{D}_{\Theta_2}$ dann und nur dann, wenn $\Theta_2 = \Theta_1 + k \cdot 360$.

6.1.2 Satz

Für jedes $\Theta \in \mathbb{R}$ ist $\hat{D}_\Theta : V \to V$ ein Isomorphismus.

Beweis

Es muß gezeigt werden: $\hat{D}_\Theta$ hat eine lineare Umkehrabbildung. Es ist aber $\hat{D}_{-\Theta} \circ \hat{D}_\Theta = \hat{D}_\Theta \circ \hat{D}_{-\Theta} = id : V \to V$. □

Da V ein 2-dimensionaler Vektorraum ist, gibt es Isomorphismen $\alpha : \mathbb{R}^2 \xrightarrow{\sim} V$. Wir nennen wie in Kapitel 3.2 ein solches α stets eine *Karte* von V und die Punkte von $\mathbb{R}^2$ *Koordinaten* in V. Jede Drehung $\hat{D} : V \to V$ wird in $\mathbb{R}^2$ zu einer linearen Abbildung $\varphi = \alpha^{-1} \circ \hat{D} \circ \alpha : \mathbb{R}^2 \to \mathbb{R}^2$. Wir nennen sie die Drehung $\hat{D}$ in den Koordinaten von V. Diese Abbildung φ wird durch eine Matrix gegeben:

$$A_\alpha(\Theta) = \begin{pmatrix} a_\alpha(\Theta) & c_\alpha(\Theta) \\ b_\alpha(\Theta) & d_\alpha(\Theta) \end{pmatrix}.$$

Es soll nun der aus der Differentialrechnung für Funktionen $\mathbb{R} \to \mathbb{R}$ bekannte Stetigkeitsbegriff verwendet werden:

Axiom XVIII

Es gibt eine Karte $\alpha : \mathbb{R}^2 \xrightarrow{\sim} V$, so daß $a_\alpha(\Theta), \ldots, d_\alpha(\Theta)$ stetig in Θ sind.

Natürlich ist die Stetigkeit unabhängig von der Karte:

6.1.3 Satz

Ist $\beta : \mathbb{R}^2 \xrightarrow{\sim} V$ eine andere Karte von V, so sind auch $a_\beta(\Theta), \ldots, d_\beta(\Theta)$ stetig in Θ.

Beweis

Es gilt $\varphi_\beta = \beta^{-1} \circ \hat{D} \circ \beta = \beta^{-1} \circ \alpha \circ \alpha^{-1} \circ \hat{D} \circ \alpha \circ \alpha^{-1} \circ \beta = \gamma^{-1} \circ \varphi_\alpha \circ \gamma$ mit $\gamma = \alpha^{-1} \circ \beta : \mathbb{R}^2 \xrightarrow{\sim} \mathbb{R}^2$. Die Abbildung γ ist die Koordinatentransformation. Sie transformiert die neuen Koordinaten, die durch β gegebenen, in die alten. Sie wird durch eine Matrix C gegeben, die der folgenden Gleichung genügt (man vgl. dazu auch Abschnitt 3.6):

$$A_\beta = C^{-1} \circ A_\alpha \circ C.$$

Ist A_α stetig in Θ, so ist auch A_β stetig in Θ. □

Die Determinante

6.1.4 Satz

Für alle $\Theta \in \mathbb{R}$ gilt $\det(\hat{D}_\Theta) \stackrel{\text{def}}{=} \det(\alpha^{-1} \circ \hat{D}_\Theta \circ \alpha) = 1$.

Beweis

Wir betrachten die Funktion $f(\Theta) = \det(\hat{D}_\Theta)$. Sie hat folgende drei Eigenschaften:

(1) f ist stetig in Θ;
denn es ist $f(\Theta) = \det(A_\alpha(\Theta)) = a_\alpha(\Theta)\, d_\alpha(\Theta) - b_\alpha(\Theta)\, c_\alpha(\Theta)$.

(2) Es ist $f(\Theta_1 + \Theta_2) = f(\Theta_1) \cdot f(\Theta_2)$:
wegen $\det\left(\hat{D}_{\Theta_1+\Theta_2}\right) = \det\left(\hat{D}_{\Theta_1} \circ \hat{D}_{\Theta_2}\right) = \det\left(\hat{D}_{\Theta_1}\right) \cdot \det\left(\hat{D}_{\Theta_2}\right)$, auf Grund von Axiom XVI.

(3) Für ein $a \neq 0$ ist $f(a) = 1$,
wegen $\det(360) = \det(id) = 1$ nach Axiom XVII.

Unser Satz ergibt sich also aus der folgenden Behauptung:

6.1.5 Satz

Die Funktion $f(\Theta) : \mathbb{R} \to \mathbb{R}$ habe die Eigenschaften (1)-(3). Dann ist $f(\Theta) \equiv 1$.

Beweis

Aus $1 = f(a) = f(a + 0) = f(a) \cdot f(0) = 1 \cdot f(0)$ folgt $f(0) = 1$.

Wir zeigen mit vollständiger Induktion nach $n = 0, 1, 2, \ldots$, daß $f(a/2^n) = 1$.
Die Induktionsbasis $n = 0$ folgt wegen $2^0 = 1$ aus Eigenschaft (3). Ist nun die Behauptung für $n - 1$ bewiesen, so folgt mit Eigenschaft (2):

$$1 = f(a/2^{n-1}) = f(a/2^n + a/2^n) = (f(a/2^n))^2.$$

Also ist $f(a/2^n) = \pm 1$. Es ist auch $f(a/2^n) = (f(a/2^{n+1}))^2 \geq 0$. Folglich ist $f(a/2^n) = 1$.

Durch eine einfache Induktion folgt aus Eigenschaft (2), daß $f(m \cdot \Theta) = (f(\Theta))^m$ für $m = 0, 1, 2, \ldots$. Ferner hat man $f(-\Theta) \cdot f(\Theta) = f(0) = 1$, also $f(-\Theta) = 1/f(\Theta)$. Das ergibt $f(\pm(m/2^n) \cdot a) = 1$ für $n, m \in \mathbb{N}_0$. Die Menge $\{\pm(m/2^n)a\}$ ist aber „dicht“ in $\mathbb{R}$: Sind $\delta > 0$, Θ reelle Zahlen, so gibt es immer ein $\pm(m/2^n)a$ mit $|\pm(m/2^n)a - \Theta| < \delta$. Wegen der Stetigkeit von f kann man zu beliebig kleinem $\epsilon > 0$ ein $\delta > 0$ finden, so daß in dem ganzen Intervall um Θ mit Radius δ die Ungleichung $|f(x) - f(\Theta)| < \epsilon$ gilt. Das Intervall enthält aber eine Einsstelle von f. Der Abstand zweier aufeinanderfolgender solcher Stellen ist ja höchstens $|a|/2^n$ und n kann beliebig groß sein. Es muß daher $|f(\Theta) - 1| < \epsilon$ für beliebig kleines ϵ sein. Dann kann nur $f(\Theta) = 1$ sein. Unser Satz ist bewiesen. $\square$

Spezielle lineare Gruppen

Im Folgenden müssen Eigenschaften von einigen Gruppen von reellen Matrizen dargestellt werden.

6.1.6 Definition

Die **spezielle lineare Gruppe** $SL(n, \mathbb{R})$ *ist die mit der Matrizenmultiplikation versehene Menge* $\{C \in \mathbb{R}^{n \times n} : \det(C) = 1\}$.

Es gilt:

(a) $SL(n,\mathbb{R}) \subset GL(n,\mathbb{R})$ ist eine Untergruppe.

(b) $A_\alpha \in SL(2,\mathbb{R})$ (nach Satz 6.1.4).

Ist $O(n,\mathbb{R})$ die in Kapitel 4.5 eingeführte orthogonale Gruppe, so ist für $A \in O(n,\mathbb{R})$:

$$1 = \det(A^t \circ A) = \det(A^t) \cdot \det(A) = (\det(A))^2, \text{ also } \det(A) = \pm 1.$$

6.1.7 Definition

Die **spezielle orthogonale Gruppe** $SO(n,\mathbb{R})$ *ist die Menge* $SL(n,\mathbb{R}) \cap O(n,\mathbb{R})$ $= \{C \in O(n,\mathbb{R}) : \det(C) = 1\}$.

Wir werden später sehen, daß $SO(2,\mathbb{R})$ gerade aus den Drehungen der anschaulichen Ebene besteht. $SO(n,\mathbb{R})$ ist eine Untergruppe von $O(n,\mathbb{R})$.

6.1.8 Satz

Es gilt

$$SO(2,\mathbb{R}) = \left\{ \begin{pmatrix} a & -b \\ b & a \end{pmatrix} : a^2 + b^2 = 1 \right\}, \qquad (*)$$

und $SO(2,\mathbb{R})$ *ist eine abelsche Gruppe.*

Beweis

Nach Satz 4.5.3 ist jede orthogonale 2×2-Matrix in der Form $(*)$ oder als

$$\begin{pmatrix} a & b \\ b & -a \end{pmatrix}$$

darstellbar. Die Determinante ist im zweiten Fall aber nicht positiv. Es wurde in Bemerkung 4.5.4 auch gezeigt, daß die Multiplikation in $SO(2,\mathbb{R})$ abelsch ist. □

Die Aussage dieses Satzes gilt hinsichtlich der Kommutativität von $SO(n,\mathbb{R})$ nicht mehr, wenn $n \geq 3$. Häufig lassen wir das $\mathbb{R}$ in der Bezeichnung dieser Gruppen fort, da wir auch in diesem Abschnitt weiterhin nur reelle Matrizen betrachten.

Der folgende Satz behandelt „Quadratwurzeln“ orthogonaler Matrizen und dient der genaueren Untersuchung der Matrizen $A_\alpha(\Theta)$.

6.1.9 Satz

(a) Ist $A \in SO(2), A \neq -E$, *dann gibt es genau zwei Matrizen* $B_1, B_2 \in SL(2)$ *mit* $B_1^2 = B_2^2 = A$. *Dabei ist* $B_1 = -B_2,\ B_1, B_2 \in SO(2)$.

(b) Für $B \in SL(2)$ *gilt* $B^2 = -E$ *genau dann, wenn*

$$B = \begin{pmatrix} r & u \\ s & -r \end{pmatrix}, \ \det(B) = -r^2 - us = 1.$$

Beweis

Es sei $A \in SO(2),\ B \in SL(2)$, also

$$A = \begin{pmatrix} a & -b \\ b & a \end{pmatrix},\ a^2 + b^2 = 1, \quad B = \begin{pmatrix} r & u \\ s & t \end{pmatrix},\ rt - us = 1.$$

Unter Benutzung von $us = rt - 1$ errechnet man:

$$B^2 = \begin{pmatrix} r(r+t) - 1 & u(r+t) \\ s(r+t) & t(r+t) - 1 \end{pmatrix}.$$

Daher gilt $B^2 = A$ genau dann, wenn die folgenden Gleichungen erfüllt sind:

$$a = r(r+t) - 1 = t(r+t) - 1; \tag{$*$}$$

$$b = s(r+t) = -u(r+t). \tag{$**$}$$

(a) Es sei nun $A \neq -E,\ B^2 = A$. Dann ist $a+1 \neq 0$, also wegen $(*)$ auch $r+t \neq 0$. Es folgt $r = t,\ s = -u$ und $r^2 + s^2 = 1$ wegen $\det(B) = 1$, mithin $B \in SO(2)$. Weiter folgt mit $(*)$ und $(**)$: $2r^2 = a + 1, \quad s = b/2r$. Wegen $a \neq -1$ ist $r^2 > 0$, und es existieren zwei verschiedene Lösungen r_1, r_2 von $r^2 = \frac{a+1}{2}$. Dagegen ist s durch r festgelegt. Also sind

$$B_1 = \begin{pmatrix} \sqrt{\frac{a+1}{2}} & -\frac{b}{2} \cdot \sqrt{\frac{2}{a+1}} \\ \frac{b}{2} \cdot \sqrt{\frac{2}{a+1}} & \sqrt{\frac{a+1}{2}} \end{pmatrix}, \quad B_2 = -B_1$$

die Lösungen (Man kann leicht nachrechnen, daß diese Bedingungen auch hinreichend für $B^2 = A$ sind).

(b) Ist $A = -E$, so lauten $(*)$ und $(**)$:

$$r(r+t) = t(r+t) = 0,$$

$$s(r+t) = u(r+t) = 0.$$

Wäre $r + t \neq 0$, so wäre im Widerspruch dazu $r = u = s = t = 0$. Also gilt $r = -t$ und

$$B = \begin{pmatrix} r & u \\ s & -r \end{pmatrix},\ \det(B) = -r^2 - us = 1.$$

Alle diese B erfüllen die Gleichung $B^2 = -E$. □

Orthogonale Karten

Es sei $\alpha : \mathbb{R}^2 \to V$ eine Karte in unserem Vektorraum der Pfeilvektoren. Die Drehungen $\hat{D}_\Theta$ werden in den Koordinaten durch die Matrizen $A_\alpha(\Theta) \in SL(2)$

beschrieben. Auf Grund von Axiom XVI besitzt die Abbildung $\Theta \mapsto A_\alpha(\Theta)$ die multiplikative Eigenschaft:

$$A_\alpha(\Theta_1 + \Theta_2) = A_\alpha(\Theta_1) \circ A_\alpha(\Theta_2) \text{ für } \Theta_1, \Theta_2 \in \mathbb{R}.$$

Insbesondere gilt nach Axiom XVII:

$$E = A_\alpha(360) = A_\alpha(180) \circ A_\alpha(180).$$

Gerade die Matrizen $B = E, -E$ lösen die Gleichung $B^2 = E$ in $SL(2)$. Folglich ist $A_\alpha(180) = \pm E$. Aber $+E$ kommt wegen Axiom XVII nicht in Frage. Also ist $A_\alpha(180) = -E$ unabhängig von α. Jedoch kann man nicht $A_\alpha(90)$ bestimmen. Wir erhalten nur:

$$A_\alpha(90) = \begin{pmatrix} r & u \\ s & -r \end{pmatrix} \text{ mit } -r^2 - us = 1.$$

Das ist aber auch selbstverständlich, da die aufspannenden Vektoren

$$x_1 := \alpha(\vec{e}_1), \qquad x_2 := \alpha(\vec{e}_2)$$

in beliebigem Winkel aufeinander stehen können.

6.1.10 Definition

Eine Karte heißt **orthogonal,** *wenn* $x_2 = \hat{D}_{90}(x_1)$.

Die Existenz von orthogonalen Karten ist beinahe trivial. Wir wählen einfach einen beliebigen Vektor $x_1 \neq \mathcal{O}$ und setzen dann $x_2 = \hat{D}_{90}(x_1)$. Dann sind beide Vektoren linear unabhängig. Aus $x_2 = a \cdot x_1$ folgt nämlich

$$-x_1 = \hat{D}_{180}(x_1) = \hat{D}_{90}(x_2) = \hat{D}_{90}(ax_1) = a\hat{D}_{90}(x_1) = ax_2 = a^2 x_1.$$

Also ist $a^2 = -1$. Das ist nicht möglich. Man hat in x_1, x_2 eine Basis von V. Die zugehörige Abbildung $\alpha : \mathbb{R}^2 \to V$ ist eine Karte.

6.1.11 Satz

Bei einer orthogonalen Karte gilt:

$$A_\alpha(0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad A_\alpha(90) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

$$A_\alpha(180) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad A_\alpha(270) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Beweis

Nach Axiom XVII gilt $A_\alpha(0) = E$, $A_\alpha(180) = -E$ wurde soeben gezeigt. Es ist

$$A_\alpha(90) = \begin{pmatrix} r & u \\ s & -r \end{pmatrix}$$

und

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} = A_\alpha(90) \circ \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

mithin $r = 0, s = 1$. Man hat $\det(A_\alpha(90)) = 1$. Also muß $u = -1$ sein. Es ist $A_\alpha(270) = A_\alpha(90 + 180) = A_\alpha(90) \circ (-E) = -A_\alpha(90)$. Damit ist alles gezeigt. □

6.1.12 Satz

Bei einer orthogonalen Karte ist immer $A_\alpha(\Theta) \in SO(2)$.

Wir schreiben immer $A_\alpha(\Theta) = \begin{pmatrix} a_\alpha(\Theta) & c_\alpha(\Theta) \\ b_\alpha(\Theta) & d_\alpha(\Theta) \end{pmatrix}$.

Beweis

Weil die Determinante der Abbildung immer gleich 1 ist, muß man nur die Orthogonalität beweisen.

(a) Durch vollständige Induktion über $n = 0, 1, 2, \ldots$ zeigen wir: Es ist stets $A_\alpha(90/2^n) \neq \pm E$ und orthogonal. Der Induktionsanfang $n = 0$ ist klar. Sei nun die Behauptung für $n - 1$ bewiesen. Dann folgt mit Satz 6.1.9 aus $(A_\alpha(90/2^n))^2 = A_\alpha(90/2^{n-1})$, daß $A_\alpha(90/2^n) \neq \pm E$ und wieder orthogonal ist.

(b) Für $\Theta = q \cdot 90$, $q = m/2^n \in [0, 1]$, $m, n \in \mathbb{N}_0$, gelten die Gleichungen

$$a_\alpha(\Theta) = d_\alpha(\Theta),\ b_\alpha(\Theta) = -c_\alpha(\Theta), \qquad (*)$$

denn Potenzen orthogonaler Matrizen sind wieder orthogonal. Diese Zahlen $q \cdot 90$ liegen wieder „dicht" im Intervall zwischen 0 und 90. Weil die Funktionen stetig sind, folgt dann wie beim Beweis von Satz 6.1.5, daß die Orthogonalitätsbedingungen $(*)$ im ganzen Intervall gelten. Man kann nun jedes $\hat{D}_\Theta$ durch Multiplikation einer Drehung $\hat{D}_{\Theta_1}$ zu einem Θ_1 aus dem Intervall $[0, 90]$ mit $\hat{D}_{90}, \hat{D}_{180}, \hat{D}_{270}$ erreichen. Durch Multiplikation der zugehörigen Matrizen sieht man, daß immer orthogonale Matrizen entstehen. □

6.1.13 Satz

Es sei α eine orthogonale Karte. Dann sind im Intervall $\{\Theta : 0 \leq \Theta < 360\}$ die Nullstellen von $a_\alpha(\Theta)$ die Zahlen 90 und 270, die von $b_\alpha(\Theta)$ die Zahlen 0 und 180. (Echt) zwischen 0 und 90 sowie zwischen 270 und 360 gilt $a_\alpha(\Theta) > 0$, zwischen 90 und 270 ist $a_\alpha(\Theta) < 0$, die Funktion $b_\alpha(\Theta)$ ist positiv zwischen 0 und 180 und negativ zwischen 180 und 360.

Beweis

Daß die angegebenen Zahlen Nullstellen sind, ersieht man aus Satz 6.1.11. Ist umgekehrt Θ eine Nullstelle von a_α, so folgt $1 = (a_\alpha(\Theta))^2 + (b_\alpha(\Theta))^2 = (b_\alpha(\Theta))^2$, also

$b_\alpha(\Theta) = \pm 1$ und folglich $A_\alpha(\Theta) = A_\alpha(90)$ oder $A_\alpha(\Theta) = A_\alpha(270)$. Axiom XVII liefert dann $\Theta = 90$ oder 270. Entsprechend untersucht man $b_\alpha(\Theta)$.

Die Vorzeichenaussagen folgen wegen $a_\alpha(0) = -a_\alpha(180) = a_\alpha(360) = 1$ und $b_\alpha(90) = -b_\alpha(270) = 1$ und der Nullstellenaussagen aus dem Zwischenwertsatz für stetige Funktionen. Wäre etwa zwischen 0 und 90 ein $a_\alpha \leq 0$, so müßte in diesem Intervall eine weitere Nullstelle von a_α vorliegen, was nicht der Fall ist. $\square$

Es folgt jetzt, daß $A_\alpha(\Theta)$ unabhängig von α ist.

6.1.14 Satz

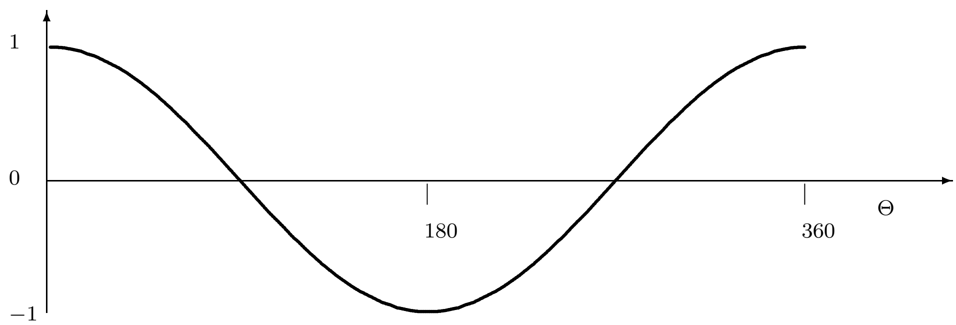
Bei orthogonalen Karten α, β gilt für alle $\Theta \in \mathbb{R}: \; A_\alpha(\Theta) = A_\beta(\Theta)$.

Wir schreiben deshalb fortan

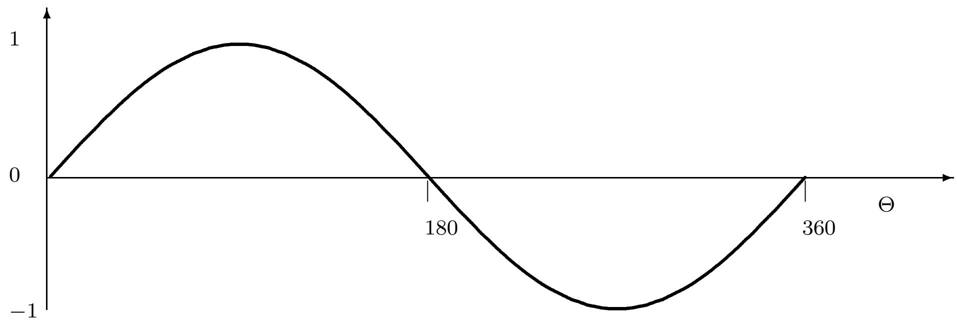
$$A_\alpha(\Theta) = A(\Theta) = \begin{pmatrix} a(\Theta) & -b(\Theta) \\ b(\Theta) & a(\Theta) \end{pmatrix}.$$

Beweis

Wir führen ihn wie bei Satz 6.1.12 durch vollständige Induktion über $n = 0, 1, 2, \ldots$, indem wir die Werte der beiden Matrixfunktionen für alle $\Theta = 90/2^n$ ausrechnen. Beim Übergang von $\Theta = 90/2^{n-1}$ zu $\Theta = 90/2^n$ haben wir zwei Möglichkeiten, die sich jedoch nur durch das Vorzeichen unterscheiden. Es muß aber $a_\alpha(\Theta) > 0$ und $a_\beta(\Theta) > 0$ gelten. Dadurch ist der Übergang festgelegt. Wir erhalten die Gleichheit von $a_\alpha(90/2^n)$ und $a_\beta(90/2^n)$ für alle n und durch anschließendes Potenzieren die Gleichheit von $a_\alpha(q \cdot 90) = a_\beta(q \cdot 90)$, $q = m/2^n \in [0, 1]$. Wegen der Stetigkeit folgt dann $a_\alpha(\Theta) = a_\beta(\Theta)$ und ganz entsprechend $b_\alpha(\Theta) = b_\beta(\Theta)$ für das ganze Intervall zwischen 0 und 90. Wieder durch Multiplikation mit $A(90)$, $A(180)$, $A(270)$ ergibt sich die Gleichheit dann überall. $\square$



Funktion a(Θ)



Funktion b(Θ)

In Kapitel 4.5 wurden bereits die *trigonometrischen Funktionen* sin, cos benutzt. Es gibt für sie eine analytische Definition durch unendliche Potenzreihen. Diese wird meist in der Differential- und Integralrechnung gebracht. Wir geben jetzt eine exakte geometrische Definition. Die Gleichheit beider Definitionen (bis auf Umrechnungen der Winkel von Grad in Bogenmaß und umgekehrt) folgt dann später, weil der $\mathbb{R}^2$ mit der analytischen Definition von sin, cos auch zu einem euklidischen Vektorraum gemacht werden kann, dazu s. man unten S. 198.

6.1.15 Definition

Für $\Theta \in \mathbb{R}$ *sei* $\cos(\Theta) := a(\Theta), \sin(\Theta) := b(\Theta)$.

Die trigonometrischen Funktionen

(a) *Die Funktionen* sin *und* cos *sind stetig.*

Der Beweis folgt aus Satz 6.1.3. □

(b) *Für* $k \in \mathbb{Z}$, $\Theta \in \mathbb{R}$ *gilt* $\cos(\Theta + k \cdot 360) = \cos(\Theta)$, $\sin(\Theta + k \cdot 360) = \sin(\Theta)$, *d. h.* sin *und* cos *haben die Periode* 360.

Der Beweis folgt aus Axiom XVII. □

(c) *Im abgeschlossenen Intervall zwischen* 0 *und* 360 *sind* 0, 180, 360 *die Nullstellen von* sin *und* 90, 270 *die Nullstellen von* cos .

Der Beweis dazu ist Satz 6.1.13. Er ist es auch für die folgenden beiden Aussagen:

(d) $\sin(90) = -\sin(270) = 1$, $\cos(0) = -\cos(180) = 1$.

(e) $\cos(\Theta) > 0$ *in den offenen Intervallen zwischen* 0, 90 *und zwischen* 270 *und* 360.
$\cos(\Theta) < 0$ *im offenen Intervall zwischen* 90, 270.
$\sin(\Theta) > 0$ *im offenen Intervall zwischen* 0, 180.
$\sin(\Theta) < 0$ *im offenen Intervall zwischen* 180, 360.

(f) *Für alle* $\Theta \in \mathbb{R}$ *gilt* $\cos^2(\Theta) + \sin^2(\Theta) = 1$.

Der Satz folgt, weil $1 = \det(A(\Theta)) = a^2(\Theta) + b^2(\Theta)$. □

(g) *Für alle* $\Theta \in \mathbb{R}$ *gilt* $\cos(-\Theta) = \cos(\Theta), \sin(-\Theta) = -\sin(\Theta)$.

Beweis

Es ist $A(\Theta) \circ A(-\Theta) = E$ und damit $A(-\Theta)$ die Umkehrmatrix von $A(\Theta)$. Wegen $A(\Theta) \in SO(2)$ bedeutet das:

$$\begin{pmatrix} \cos(\Theta) & \sin(\Theta) \\ -\sin(\Theta) & \cos(\Theta) \end{pmatrix} = A(\Theta)^t = A(-\Theta) = \begin{pmatrix} \cos(-\Theta) & -\sin(-\Theta) \\ \sin(-\Theta) & \cos(-\Theta) \end{pmatrix}.$$

Das ist die Behauptung. □

(h) *Zu jedem* $\vec{x} = (x_1, x_2) \in \mathbb{R}^2$ *mit* $x_1^2 + x_2^2 = 1$ *existiert genau ein* Θ *mit* $0 \leq \Theta < 360$ *und* $\vec{x} = (\cos(\Theta), \sin(\Theta))$.

Beweis

Es gilt $|x_1| \leq 1$. Also gibt es ein $\Theta_1 : 0 \leq \Theta_1 \leq 90$, so daß $|x_1| = \cos(\Theta_1)$. Es ist dann $x_2 = \pm\sin(\Theta_1)$. Es ist auch $x_1 = \pm\cos(\Theta_1)$. Durch Übergang zu $\Theta = \Theta_1 + 180$, $\Theta = 360 - \Theta_1$, $\Theta = 180 - \Theta_1$ folgt die Existenz einer Lösung.

Gilt aber $(\cos(\Theta_1), \sin(\Theta_1)) = (\cos(\Theta_2), \sin(\Theta_2))$, so folgt $A(\Theta_1) = A(\Theta_2)$ und nach Axiom XVII auch $\Theta_1 = \Theta_2$, weil $0 \leq \Theta_1 < 360$ und $0 \leq \Theta_2 < 360$. □

(i) *Es gelten die Additiontheoreme:*

$$\sin(\Theta_1 + \Theta_2) = \sin(\Theta_1) \cdot \cos(\Theta_2) + \cos(\Theta_1) \cdot \sin(\Theta_2),$$
$$\cos(\Theta_1 + \Theta_2) = \cos(\Theta_1) \cdot \cos(\Theta_2) - \sin(\Theta_1) \cdot \sin(\Theta_2)$$

für $\Theta_i \in \mathbb{R}$.

Beweis

Auf Grund von Axiom XVI gilt:

$$\begin{aligned} A(\Theta_1 + \Theta_2) &= A(\Theta_1) \circ A(\Theta_2) \\ &= \begin{pmatrix} \cos(\Theta_1) \cdot \cos(\Theta_2) - \sin(\Theta_1) \cdot \sin(\Theta_2) & -\sin(\Theta_1) \cdot \cos(\Theta_2) - \cos(\Theta_1) \cdot \sin(\Theta_2) \\ \sin(\Theta_1) \cdot \cos(\Theta_2) + \cos(\Theta_1) \cdot \sin(\Theta_2) & \cos(\Theta_1) \cdot \cos(\Theta_2) - \sin(\Theta_1) \cdot \sin(\Theta_2) \end{pmatrix}. \end{aligned}$$

Das ist die Behauptung. □

6.2 Die Länge eines Vektors

Zunächst wird im n-dimensionalen reellen Zahlenraum $\mathbb{R}^n$ abstrakt mathematisch die euklidische Norm eingeführt. Im anschaulichen Vektorraum V läßt sich aber die Länge auch anschaulich definieren. Bei Wahl einer geeigneten rechtwinkligen Karte im $\mathbb{R}^2$ stimmen beide Definitionen überein. Etwas Analoges gilt für den Winkelbegriff. Im Folgenden können deshalb alle Probleme der anschaulichen Geometrie in V mit den üblichen mathematischen Methoden des $\mathbb{R}^2$ behandelt werden.

Im Kapitel 4.5 wurde das sogenannte kanonische Skalarprodukt im $\mathbb{R}^n$ abstrakt mathematisch eingeführt. Es gilt für $\vec{x}, \vec{y} \in \mathbb{R}^n$, wenn wir sie als Spalten schreiben:

$$\vec{x} \cdot \vec{y} \stackrel{\text{def}}{=} \sum_{i=1}^{n} x_i y_i = \vec{x}^{\,t} \circ \vec{y}$$

als Matrixprodukt. Mitunter läßt man den Skalarproduktpunkt auch weg. Man hat die folgenden typischen Eigenschaften (Axiome) des Skalarproduktes:

6.2.1 Satz

Für alle $a, b \in \mathbb{R}$, $\vec{x}$, $\vec{y}$, $\vec{x}\,'$, $\vec{y}\,' \in \mathbb{R}^n$ gilt:

(1) $(a \cdot \vec{x} + b \cdot \vec{x}\,') \cdot \vec{y} = a \cdot \vec{x} \cdot \vec{y} + b \cdot \vec{x}\,' \cdot \vec{y}$,
$\vec{x} \cdot (a \cdot \vec{y} + b \cdot \vec{y}\,') = a \cdot \vec{x} \cdot \vec{y} + b \cdot \vec{x} \cdot \vec{y}\,'$; (Bilinearität)

(2) $\vec{x} \cdot \vec{y} = \vec{y} \cdot \vec{x}$; (Kommutativität, Symmetrie)

(3) $\vec{x} \cdot \vec{x} > 0$, *wenn* $\vec{x} \neq 0$. (Positive Definitheit)

Von besonderer Wichtigkeit ist:

6.2.2 Satz (Cauchy-Schwarzsche Ungleichung)

Für alle $\vec{x}, \vec{y} \in \mathbb{R}^n$ gilt $|\vec{x} \cdot \vec{y}\,| \leq \sqrt{(\vec{x} \cdot \vec{x})} \cdot \sqrt{(\vec{y} \cdot \vec{y})}$.

Beweis

Es werden allein die Ergebnisse von Satz 6.2.1 benutzt. Ferner sei $\vec{x} \neq \vec{0}$, $\vec{y} \neq \vec{0}$. Sonst wäre die Behauptung trivial. Für alle $t \in \mathbb{R}$ gilt:

$$0 \leq (\vec{x} + t\vec{y}\,) \cdot (\vec{x} + t\vec{y}\,) = \vec{x} \cdot \vec{x} + 2t\,\vec{x} \cdot \vec{y} + t^2\,\vec{y} \cdot \vec{y}.$$

Setzt man $t = -(\vec{x} \cdot \vec{y}\,)/(\vec{y} \cdot \vec{y}\,)$, so folgt

$$0 \leq \vec{x}\,\vec{x} - 2\frac{(\vec{x}\,\vec{y}\,)^2}{\vec{y}\,\vec{y}} + \frac{(\vec{x}\,\vec{y}\,)^2}{\vec{y}\,\vec{y}} = \vec{x}\,\vec{x} - \frac{(\vec{x}\,\vec{y}\,)^2}{\vec{y}\,\vec{y}}.$$

Multiplikation mit $\vec{y}\,\vec{y}$ und Wurzelziehen liefert die Behauptung, denn die Wurzel ist eine monotone Funktion. □

6.2.3 Definition

Die (mathematische) **euklidische Norm** *eines Vektors $\vec{x} \in \mathbb{R}^n$ ist definiert durch:*

$$\| \vec{x} \| = \sqrt{\vec{x} \cdot \vec{x}}$$

Wir zeigen, daß die typischen Eigenschaften einer *Norm* erfüllt sind:

6.2.4 Satz

Für alle $a \in \mathbb{R},\ \vec{x}, \vec{y} \in \mathbb{R}^n$ gilt:

(1) $\| \vec{x} \| \geq 0$ *und* $\| \vec{x} \| > 0$, *falls* $\vec{x} \neq 0$; (Positive Definitheit)

(2) $\| a \cdot \vec{x} \| = |a| \cdot \| \vec{x} \|$; (Multiplikatives Gesetz)

(3) $\| \vec{x} + \vec{y} \| \leq \| \vec{x} \| + \| \vec{y} \|$. (Dreiecksungleichung)

Beweis

Die Eigenschaften (1) und (2) folgen unmittelbar aus Satz 6.2.1. Die Dreiecksungleichung (3) folgt aus der Cauchy-Schwarzschen Ungleichung:

$$\begin{aligned} \| \vec{x} + \vec{y} \|^2 &= (\vec{x} + \vec{y}) \cdot (\vec{x} + \vec{y}) = \vec{x}\,\vec{x} + 2\vec{x}\,\vec{y} + \vec{y}\,\vec{y} \\ &\leq \| \vec{x} \|^2 + 2\| \vec{x} \| \, \| \vec{y} \| + \| \vec{y} \|^2 = (\| \vec{x} \| + \| \vec{y} \|)^2. \end{aligned}$$

Wurzelziehen liefert die Behauptung. □

6.2.5 Satz

Multiplikation mit orthogonalen Matrizen ändert die Norm eines Vektors nicht.

Beweis

In Matrizenschreibweise gilt:

$$\| A \circ \vec{x} \|^2 = (A \circ \vec{x})^t \circ (A \circ \vec{x}) = \vec{x}^{\,t} \circ (A^t \circ A) \circ \vec{x} = \vec{x}^{\,t} \circ E \circ \vec{x} = \| \vec{x} \|^2$$

wegen $A^t \circ A = E$. □

Wie schon angedeutet, ist in allgemeinen Vektorräumen W der Normbegriff definiert. Wir sagen es möglichst einfach:

6.2.6 Definition

Eine Abbildung $\| \cdot \| : W \to \mathbb{R}$ heißt eine **Norm**, *wenn gilt:*

(1) *für $x \in W$, $x \neq 0$ ist immer $\| x \| \neq 0$;*

(2) *es ist stets $\| a \cdot x \| = |a| \cdot \| x \|$;*

(3) *man hat die Dreiecksungleichung: $\| x + y \| \leq \| x \| + \| y \|$.*

Ein Vektorraum, der eine Norm trägt, heißt ein **normierter Vektorraum**.

Wegen (2) ist klar, daß $\| 0 \| = 0$ gilt, wegen $0 = \| x - x \| \leq 2\| x \|$ ist auch stets $\| x \| \geq 0$.

Der Längenbegriff

Man kann nun anschaulich den Begriff der Länge der Vektoren in der anschaulichen Ebene E einführen. Parallele Vektoren sollen gleiche Länge haben. Es sei $O \in E$ der Basispunkt. Es genügt dann, die Länge von Vektoren aus $V = \{(O, P) : P \in E\}$ zu definieren. Man nimmt einen *Maßvektor* $z \in V \setminus \{\mathcal{O}\}$ und gibt ihm die Länge 1.

Von dem Längenbegriff wird man fordern, daß er invariant unter Drehungen ist. Wir haben in V die orthogonale Karte $\alpha : \mathbb{R}^2 \to V$ mit $\alpha(\vec{e}_1) = z$, $\alpha(\vec{e}_2) = \hat{D}_{90}(z)$. Die nach $\mathbb{R}^2$ übertragenen Drehungen $\alpha^{-1} \circ \hat{D}_\Theta \circ \alpha$ werden durch Matrizen $A(\Theta) = \begin{pmatrix} \cos(\Theta) & -\sin(\Theta) \\ \sin(\Theta) & \cos(\Theta) \end{pmatrix}$, $\Theta \in \mathbb{R}$, gegeben. Es sei $x \in V \setminus \{\mathcal{O}\}$ und $\vec{x} = (x_1, x_2) = \alpha^{-1}(x)$. Wir wählen $t \in \mathbb{R}^+$ so, daß $t^{-2}(x_1^2 + x_2^2) = 1$. Die Zahl t ist eindeutig bestimmt. Nach (h) am Ende von Kapitel 6.1 gibt es dann ein Θ, $0 \leq \Theta < 360$, mit $t^{-1}x_1 = \cos(\Theta)$, $t^{-1}x_2 = \sin(\Theta)$. Die Drehung $A(\Theta)$ wirft dann $\vec{e}_1$ auf $t^{-1}\vec{x}$, und mithin dreht $\hat{D}_\Theta$ den Vektor z auf $t^{-1}x$. Die Drehung $\hat{D}_\Theta$ ist eindeutig bestimmt.

6.2.7 Definition

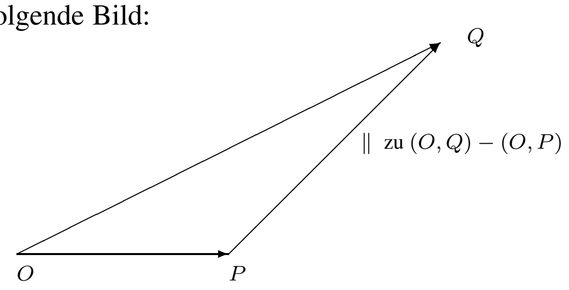
Es sei $\| x \| := t$, $\| \mathcal{O} \| := 0$.

Bei $x = z$ ist die Drehung die Identität. Also ist auch nach der Definition $\| z \| = 1$. Es gilt $\| x \| = t = \sqrt{x_1^2 + x_2^2}$. Also haben wir eine Norm wie in $\mathbb{R}^2$. Wir nennen $\| x \|$ auch die Länge des Vektors x. Diese Länge hängt wesentlich von der Auswahl des Maßvektors z ab. Wir brauchen ein „Urmeter“. Ersetzen wir z durch ein $\hat{D}_{\Theta_1}(z)$, so ändert sich die Norm nicht. Wird dagegen z durch az mit $a > 0$ ersetzt, so multipliziert sich die Norm mit a^{-1}.

6.2.8 Definition

Der Abstand zweier Punkte $P, Q \in E$ *ist* $\operatorname{dist}(P, Q) := \| (O, Q) - (O, P) \|$.

Wir haben das folgende Bild:



Entfernung von P und Q

Die Dreiecksungleichung besagt, daß $\| (O,Q) - (O,P) \| \le \| (O,Q) \| + \| (O,P) \|$, also $\mathrm{dist}(P,Q) \le \mathrm{dist}(O,P) + \mathrm{dist}(O,Q)$. Das heißt: *Bei einem Dreieck ist die Summe zweier Seitenlängen nicht kleiner als die dritte Seitenlänge.* Das erklärt den Ausdruck!

Der Winkelbegriff

Wir haben gesehen: *Sind $y, x \in V \setminus \{\mathcal{O}\}$ Vektoren, so gibt es genau ein $\Theta : 0 \le \Theta < 360$ mit $\| x \|^{-1} x = \hat{D}_\Theta \left(\| y \|^{-1} y \right)$*. Dabei haben wir $\| y \|^{-1} y$ als Maßvektor anstelle des ursprünglichen Maßvektors z betrachtet; die Norm ändert sich dadurch nicht. Allerdings brauchen wir die Norm bzgl. des Maßvektors auch gar nicht zu verwenden: Äquivalent ist, daß x positives Vielfaches von $\hat{D}_\Theta(y)$ ist. Dadurch ist Θ schon bestimmt.

6.2.9 Definition

$\Theta = \angle(y, x)$ *heißt der* **Winkel** *zwischen y und x.*

Wenn wir y oder x mit einer positiven reellen Zahl multiplizieren, so ändert sich der Winkel nicht. Es gilt $\angle(y,x) = -\angle(x,y) \bmod (360)$, d. h. man muß evtl. ein ganzzahliges Vielfaches von 360 addieren.

Es sei V mit einer Norm versehen, die von einem Maßvektor z herkommt. Wir haben dann eine orthogonale Karte $\alpha : \mathbb{R}^2 \to V$ und $\alpha^{-1} : V \to \mathbb{R}^2$ mit $\alpha(\vec{e}_1) = z$. Sind dann $x, y \in V$ Vektoren und ist $\vec{x} = \alpha^{-1}(x)$, $\vec{y} = \alpha^{-1}(y)$, so definieren wir als *Skalarprodukt*: $(x,y) = \vec{x} \cdot \vec{y}$. Es verschwindet, wenn einer der Vektoren gleich $\mathcal{O}$ ist. Im anderen Falle sei $\Theta = \angle(x,y)$. Es gilt dann $\frac{1}{\| y \|} y = \frac{1}{\| x \|} \hat{D}_\Theta(x)$ und damit $\frac{1}{\| \vec{y} \|} \vec{y} = A \circ \left(\frac{1}{\| \vec{x} \|} \vec{x} \right)$ mit $A = \begin{pmatrix} \cos(\Theta) & -\sin(\Theta) \\ \sin(\Theta) & \cos(\Theta) \end{pmatrix}$. Man hat dann $(\frac{1}{\| \vec{x} \|} \vec{x}) \cdot (\frac{1}{\| \vec{y} \|} \vec{y}) = \left((x_1^2 + x_2^2) / \| \vec{x} \|^2 \right) \cos(\Theta) = \cos(\Theta)$. Das Skalarprodukt ist also insofern unabhängig von z definiert, als nur der durch den Maßvektor gegebene Längenbegriff eingeht.

6.2.10 Satz

Es seien $x, y \in V$ beliebige Vektoren. Dann gilt $(x,y) = \| x \| \cdot \| y \| \cdot \cos(\Theta)$. Dabei ist Θ der eingeschlossene Winkel. Im Falle $x = \mathcal{O}$ oder $y = \mathcal{O}$ kann er beliebig sein. Es folgt $(x,x) = \| x \|^2$.

Wir betrachten jetzt das Senkrechtstehen.

6.2.11 Definition

Zwei Vektoren $x, y \in V$ stehen **senkrecht** *aufeinander (sind* **orthogonal**, *$x \perp y$), wenn einer von ihnen $\mathcal{O}$ ist oder wenn der eingeschlossene Winkel 90 oder 270 ist.*

6.2.12 Satz

Die Vektoren x, y sind genau dann orthogonal, wenn das Skalarprodukt $(x, y) = 0$ ist.

Beweis

Der Satz ist trivial, wenn einer der Vektoren $\mathcal{O}$ ist. Das sei nun nicht der Fall. Dann gilt $(x, y) = \| x \| \| y \| \cos(\Theta)$, wobei Θ der eingeschlossene Winkel ist. Es ist aber $\cos(\Theta)$ genau dann gleich 0, wenn $\Theta = 90$ oder $\Theta = 270$. Das beweist alles. □

Diese Eigenschaft benutzen wir, um im $\mathbb{R}^n$ die Orthogonalität zu definieren.

6.2.13 Definition

Zwei Vektoren $\vec{x}, \vec{y} \in \mathbb{R}^n$ heißen **orthogonal** *($\vec{x} \perp \vec{y}$), wenn $\vec{x} \cdot \vec{y} = 0$.*

Klassische Sätze

Die bekannten klassischen Sätze der euklidischen Geometrie werden meistens nicht exakt bewiesen. Das soll hier geschehen.

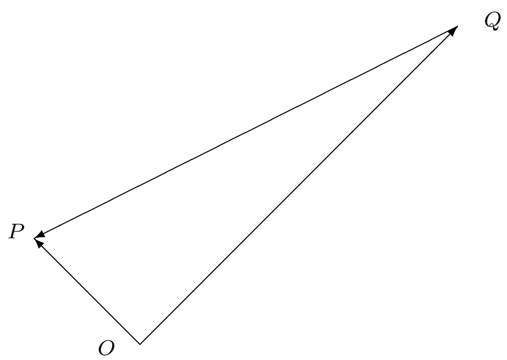
6.2.14 Satz (von Pythagoras)

Es seien $x = (O, P)$, $y = (O, Q) \in V \setminus \{\mathcal{O}\}$ orthogonale Vektoren. Dann gilt

$$\text{dist}^2(Q, P) = \text{dist}^2(O, P) + \text{dist}^2(O, Q).$$

Beweis

Es ist $\text{dist}^2(Q, P) = (x - y, x - y) = (x, x) - 2(x, y) + (y, y) = (x, x) + (y, y) = \text{dist}^2(O, P) + \text{dist}^2(O, Q)$. Dabei haben wir Satz 6.2.12 verwendet. □



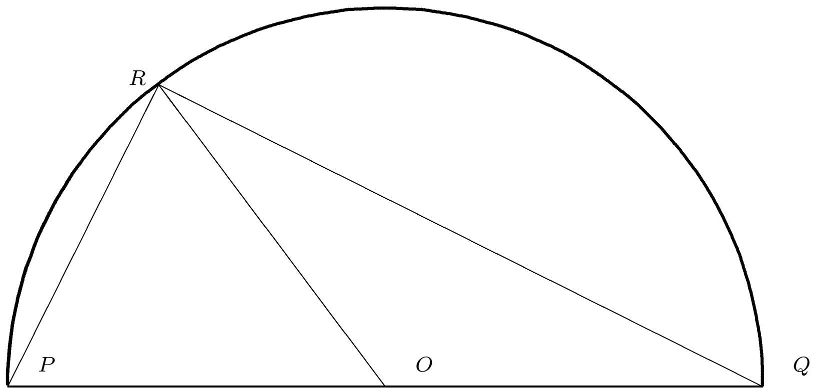
Hypotenusenquadrat = Summe der beiden Kathetenquadrate

6.2.15 Satz (von Thales)

Es seien P, Q zwei verschiedene Punkte auf einer Geraden durch O, R ein weiterer Punkt, der nicht auf der Geraden liegt. Es gelte $\text{dist}(O, P) = \text{dist}(O, Q) = \text{dist}(O, R)$. Dann sind (P, R) und (Q, R) zueinander orthogonal.

Beweis

Es sei $x = (O, Q)$, $y = (O, R)$. Dann gilt $(O, P) = -x$, $(P, R) = y + x$, $(Q, R) = y - x$. Das Skalarprodukt $(y - x, y + x) = (y, y) - (x, x)$ ist aber 0. Also stehen (P, R), (Q, R) senkrecht aufeinander. □



Der Winkel im Halbkreis ist ein rechter

Das Fällen des Lotes

Es seien $\mathcal{G} = \{x = x_1 + t \cdot x_0 :\ t \in \mathbb{R}\}$ und $\mathcal{G}' = \{x = x_1' + t \cdot x_0' :\ t \in \mathbb{R}\}$ Geraden in V mit $x_0 \neq \mathcal{O}$, $x_0' \neq \mathcal{O}$ wie in Kapitel 2.3.

6.2.16 Definition

Die Geraden $\mathcal{G}, \mathcal{G}'$ sind **orthogonal** *(stehen* **senkrecht** *aufeinander), wenn das Skalarprodukt $(x_0', x_0) = 0$ ist.*

Offenbar ist diese Definition unabhängig von der Auswahl der Richtungsvektoren.

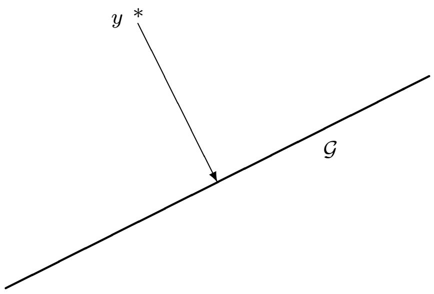
6.2.17 Satz und Definition

Es sei $y \in V \setminus \mathcal{G}$ ein Punkt. Dann gibt es genau eine Gerade $\mathcal{G}'$ durch y, die auf $\mathcal{G}$ senkrecht steht. Wir nennen $\mathcal{G}'$ das **Lot** *von y auf $\mathcal{G}$.*

Beweis

Wir betrachten die Abstandsfunktion $t \mapsto (y - x_1 - t \cdot x_0, y - x_1 - t \cdot x_0) = (y - x_1, y - x_1) - 2t(y - x_1, x_0) + t^2(x_0, x_0)$. Diese positive Parabel hat ein Minimum in einem t_0. Hier verschwindet die Ableitung nach t. Es ist also $-2(y - x_1, x_0) + 2t_0(x_0, x_0) = 0$. Das Minimum wird also in $x_1' = x_1 + ((y - x_1, x_0)/(x_0, x_0))x_0$ angenommen. Der Richtungsvektor der Geraden durch y, x_1' ist dann $x_0' := y - x_1' = y - x_1 - ((y - x_1, x_0)/(x_0, x_0))x_0$. Das Skalarprodukt dieses Vektors mit x_0 verschwindet. Die Verbindungsgerade von y und x_1' steht also senkrecht auf $\mathcal{G}$. Wir nehmen diese für $\mathcal{G}'$. Wählt man t_0 irgendwie anders, ist die Gleichung

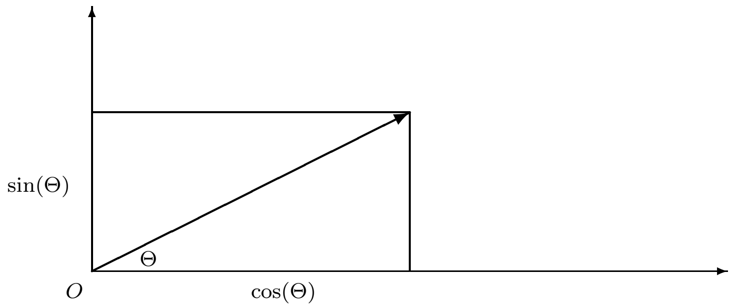
$y - x_1' = y - x_1 - ((y - x_1, x_0)/(x_0, x_0))x_0$ nicht mehr erfüllt, es ist dann $y - x_1' = a_1 (y - x_1 - ((y - x_1, x_0)/(x_0, x_0))x_0) + a_0 x_0$ mit reellen Zahlen $a_0 \neq 0$ und a_1. Das Skalarprodukt mit x_0 ist dann $\neq 0$. Also ist $\mathcal{G}'$ die einzige senkrechte Gerade durch y. Das Lot ist eindeutig bestimmt. $\square$



Lot von y auf $\mathcal{G}$

Deutung von cos und sin

Wir haben (orthogonale) Isomorphismen $\alpha^{-1} : V \to \mathbb{R}^2$, so daß die Drehungen $\hat{D}_\Theta$ in die Multiplikationen der Vektoren mit den Matrizen $A(\Theta) = \begin{pmatrix} \cos(\Theta) & -\sin(\Theta) \\ \sin(\Theta) & \cos(\Theta) \end{pmatrix}$ übergehen. Dann ist $A(\Theta) \circ \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ gleich $\begin{pmatrix} \cos(\Theta) \\ \sin(\Theta) \end{pmatrix}$, also $\cos(\Theta)$ die Länge der Ankathete des Dreiecks zur x-Achse unter dem Bildvektor und $\sin(\Theta)$ die Länge der Gegenkathete.



Die Definitionen nach der Schulmathematik

Der $\mathbb{R}^2$ als euklidische Ebene

Die Mathematik ist formal. Es kommt nie darauf an, was die behandelten Objekte eigentlich sind, sondern nur auf die Relationen, die zwischen ihnen bestehen. In unserem Falle spielt es keine Rolle, daß die Elemente der Menge E anschauliche

Punkte sind.

Wir wählen jetzt für E den $\mathbb{R}^2$ und nehmen in E den Basispunkt $\vec{0}$. Wir identifizieren stets $\left(\vec{0}, \vec{x}\right)$ mit $\vec{x}$ und haben dann auch $V = \mathbb{R}^2$. In E nennen wir zwei Vektoren $(\vec{x}, \vec{y}\,), (\vec{x}\,', \vec{y}\,')$ *(voll) parallel*, wenn $\vec{y}\,' - \vec{x}\,' = \vec{y} - \vec{x}$. Gilt $(\vec{x}, \vec{y}\,) \parallel (\vec{x}\,', \vec{y}\,')$, so folgt auch $(\vec{x}, \vec{x}\,') \parallel (\vec{y}, \vec{y}\,')$. Es sind deshalb die Axiome I bis V erfüllt. Als Multiplikation mit reellen Zahlen übernehmen wir die Skalarenmultiplikation in $\mathbb{R}^2$. Bekanntlich sind Addition und Multiplikation stetig im Sinne der Differentialrechnung. Es gilt deshalb das Axiom X. Die übrigen Axiome bis XII einschließlich sind trivial.

Wir nehmen nun irgendeine Definition von $\cos(\Theta)$, $\sin(\Theta)$, etwa die mit den Potenzreihen, so daß folgende Eigenschaften gelten:

(1) $\cos(\Theta)$ und $\sin(\Theta)$ sind stetig in Θ;

(2) das Paar $(\cos(\Theta), \sin(\Theta))$ hat genau die Periode 360;

(3) es gelten die Additionstheoreme:

$$\cos(\Theta_1 + \Theta_2) = \cos(\Theta_1)\cos(\Theta_2) - \sin(\Theta_1)\sin(\Theta_2),$$

$$\sin(\Theta_1 + \Theta_2) = \sin(\Theta_1)\cos(\Theta_2) + \cos(\Theta_1)\sin(\Theta_2).$$

Wir setzen $A(\Theta) = \begin{pmatrix} \cos(\Theta) & -\sin(\Theta) \\ \sin(\Theta) & \cos(\Theta) \end{pmatrix}$. Es sei $D_\Theta(\vec{x}, \vec{y}\,) = (A(\Theta) \circ \vec{x}, A(\Theta) \circ \vec{y}\,)$. Es folgt dann $\hat{D}_\Theta(\vec{x}) = A(\Theta) \circ \vec{x}$. Da es sich um lineare Abbildungen handelt, sind die Axiome bis XV erfüllt. Axiom XVI und XVII folgen aus den Eigenschaften (2) und (3), Axiom XVIII ist die Eigenschaft (1). Somit trägt $\mathbb{R}^2$ die euklidische Geometrie. Die Karte $\alpha = id : \mathbb{R}^2 \to V = \mathbb{R}^2$ ist orthogonal. Die Koeffizienten a, b der Matrix $A(\Theta)$ sind eindeutig bestimmt. Somit gilt (wie bei Satz 6.1.14):

$$\cos(\Theta) = a(\Theta), \quad \sin(\Theta) = b(\Theta).$$

Das heißt, daß die neue Definition der trigonometrischen Funktionen mit unserer geometrischen übereinstimmt.

6.3 Orthogonale k-Beine

Die Ausführungen von Kapitel 4.5 sollen ergänzt werden. Es gibt zwei Typen orthogonaler Matrizen. Für $A \in O(n)$ kann $\det(A) = 1$ oder gleich -1 sein. Gilt $\det(A) = 1$, so heißt A, motiviert durch die Betrachtungen der beiden vorhergehenden Paragraphen, eine *Drehung*. Gilt dagegen $\det(A) = -1$, so nennen wir A eine

Drehspiegelung. Dann ist A Produkt einer Drehung und der Spiegelung:

$$S = \begin{pmatrix} -1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Dabei ist S eine Spiegelung an der Hyperebene $\{\vec{x} \in \mathbb{R}^n : x_1 = 0\}$. Es gilt: $S = S^t$, $S^t \circ S = E$, $\det(S) = -1$, $S \in O(n)$. Für $B := S \circ A$ ist: $\det(B) = (-1) \cdot (-1) = 1$. Also ist B eine Drehung. Man hat $A = E \circ A = S^2 \circ A = S \circ (S \circ A) = S \circ B$.

Unter Verwendung der folgenden Definition kann eine Anzahl von äquivalenten Kriterien für die Orthogonalität von Matrizen hergeleitet werden.

6.3.1 Definition

Ein k-tupel $(\vec{x}_1, \ldots, \vec{x}_k)$ von Vektoren $\vec{x}_i \in \mathbb{R}^n$ heißt ein **orthogonales k-Bein**, *wenn für $i, j = 1, \ldots, k$ gilt:*

$$\vec{x}_i \cdot \vec{x}_j = \delta_{ij}.$$

6.3.2 Satz

Es sei $A = (\vec{a}_1, \ldots, \vec{a}_n)$ eine $n \times n$-Matrix. Dann sind die folgenden Aussagen äquivalent:

(a) *A ist orthogonal;*

(b) *für alle $\vec{x}, \vec{y} \in \mathbb{R}^n$ gilt $(A \circ \vec{x}) \cdot (A \circ \vec{y}) = \vec{x} \cdot \vec{y}$;*

(c) *für alle $\vec{x} \in \mathbb{R}^n$ gilt $\| A \circ \vec{x} \| = \| \vec{x} \|$;*

(d) *$(\vec{a}_1, \ldots, \vec{a}_n)$ ist ein orthogonales n-Bein.*

Die orthogonale Matrix A bildet also das orthogonale n-Bein $(\vec{e}_1, \ldots, \vec{e}_n)$, die kanonische Basis des $\mathbb{R}^n$, auf das orthogonale n-Bein $(\vec{a}_1, \ldots, \vec{a}_n)$ ab.

Beweis

„(a) $\Rightarrow$ (c)“: Siehe Satz 6.2.5.

„(c) $\Rightarrow$ (b)“: Es sei $C = A^t \circ A$. Dann ist $C^t = A^t \circ (A^t)^t = C$. Nach Voraussetzung gilt für alle $\vec{x} \in \mathbb{R}^n$ die Gleichung $\vec{x} \cdot \vec{x} = (\vec{x}^{\,t} \circ A^t) \circ A \circ \vec{x} = \vec{x}^{\,t} \circ C \circ \vec{x}$. Für $\vec{x}, \vec{y} \in \mathbb{R}^n$ folgt:

$$\begin{aligned} \vec{x} \cdot \vec{x} + 2\vec{x} \cdot \vec{y} + \vec{y} \cdot \vec{y} &= (\vec{x} + \vec{y}) \cdot (\vec{x} + \vec{y}) = (\vec{x}^{\,t} + \vec{y}^{\,t}) \circ C \circ (\vec{x} + \vec{y}) \\ &= \vec{x}^{\,t} \circ C \circ \vec{x} + \vec{x}^{\,t} \circ C \circ \vec{y} + \vec{y}^{\,t} \circ C \circ \vec{x} + \vec{y}^{\,t} \circ C \circ \vec{y} \\ &= \vec{x} \cdot \vec{x} + \vec{x}^{\,t} \circ C \circ \vec{y} + \vec{y}^{\,t} \circ C \circ \vec{x} + \vec{y} \cdot \vec{y} \end{aligned}$$

und damit

$$2\vec{x} \cdot \vec{y} = \vec{x}^{\,t} \circ C \circ \vec{y} + \vec{y}^{\,t} \circ C \circ \vec{x}.$$

Wegen der Symmetrie von C ist die reelle Zahl $\vec{y}^{\,t} \circ C \circ \vec{x} = (\vec{y}^{\,t} \circ C \circ \vec{x}\,)^t = \vec{x}^{\,t} \circ C \circ \vec{y}$. Also folgt:

$$\vec{x} \cdot \vec{y} = \vec{x}^{\,t} \circ C \circ \vec{y} = (A \circ \vec{x}\,) \cdot (A \circ \vec{y}\,).$$

„(b) $\Rightarrow$ (d)“: Es seien wieder $\vec{e}_i \in \mathbb{R}^n$, $i = 1, \ldots, n$, die Einheitsvektoren. Für $i, j = 1, \ldots, n$ gilt dann:

$$\delta_{ij} = \vec{e}_i \cdot \vec{e}_j = (A \circ \vec{e}_i) \cdot (A \circ \vec{e}_j) = \vec{a}_i \cdot \vec{a}_j.$$

„(d) $\Rightarrow$ (a)“: Die Gleichung $\vec{a}_i \cdot \vec{a}_j = \delta_{ij}$ ergibt, daß :

$$A^t \circ A = \begin{pmatrix} \vec{a}_1^{\,t} \\ \vdots \\ \vec{a}_n^{\,t} \end{pmatrix} \circ (\vec{a}_1, \ldots, \vec{a}_n) = (\vec{a}_i \cdot \vec{a}_j)_{i,j=1,\ldots,n} = (\delta_{ij})_{i,j=1,\ldots,n} = E. \qquad \square$$

Dieser Satz enthält auch, daß jedes orthogonale n-Bein linear unabhängig ist. Allgemeiner gilt:

6.3.3 Satz

Es sei $(\vec{x}_1, \ldots, \vec{x}_k)$ ein orthogonales k-Bein im $\mathbb{R}^n$. Dann sind die Vektoren $\vec{x}_1, \ldots, \vec{x}_k$ linear unabhängig. Insbesondere ist $k \leq n$.

Beweis

Gilt etwa $\sum_{i=1}^k a_i \cdot \vec{x}_i = \vec{0}$ mit $a_i \in \mathbb{R}$, so folgt:

$$0 = \left(\sum_{i=1}^k a_i \cdot \vec{x}_i \right) \cdot \vec{x}_j = \sum_{i=1}^k a_i \cdot (\vec{x}_i \cdot \vec{x}_j) = \sum_{i=1}^k a_i \cdot \delta_{ij} = a_j. \qquad \square$$

Schmidtsches Orthonormalisierungsverfahren

6.3.4 Satz

Es sei $V \subset \mathbb{R}^n$ ein Untervektorraum und $(\vec{x}_1, \ldots, \vec{x}_k)$ ein orthogonales k-Bein in V mit $k < \dim V$. Dann gibt es ein $\vec{x}_{k+1} \in V$, so daß $(\vec{x}_1, \ldots, \vec{x}_{k+1})$ ein orthogonales $(k+1)$-Bein ist.

Beweis

Der von $\vec{x}_1, \ldots, \vec{x}_k$ aufgespannte Vektorraum $W \subset V$ hat kleinere Dimension als V. Es gibt daher einen Vektor $\vec{y} \in V \setminus W$. Wir setzen $\vec{y}\,' = \vec{y} - \sum_{i=1}^k (\vec{y} \cdot \vec{x}_i)\vec{x}_i$ und erhalten $\vec{y}\,' \neq \vec{0}$. Sonst wäre ja $\vec{y}$ Linearkombination von $\vec{x}_1, \ldots, \vec{x}_k$. Der Vektor

$\vec{x}_{k+1} = \frac{1}{\|\vec{y}'\|} \cdot \vec{y}'$ ist dann ein Einheitsvektor. Es gilt für $j = 1, \ldots, k$:

$$\vec{x}_{k+1} \cdot \vec{x}_j = \frac{1}{\|\vec{y}'\|} \left(\vec{y} \cdot \vec{x}_j - \sum_{i=1}^{k} (\vec{y} \cdot \vec{x}_i)(\vec{x}_i \cdot \vec{x}_j) \right) = \frac{1}{\|\vec{y}'\|} (\vec{y} \cdot \vec{x}_j - \vec{y} \cdot \vec{x}_j) = 0.$$

□

6.3.5 Satz

Es sei $V \subset \mathbb{R}^n$ ein ℓ-dimensionaler Untervektorraum und $(\vec{x}_1, \ldots, \vec{x}_k)$ ein orthogonales k-Bein in V mit $k < \ell$. Dann gibt es Vektoren $\vec{x}_{k+1}, \ldots, \vec{x}_\ell \in V$, so daß $(\vec{x}_1, \ldots, \vec{x}_\ell)$ ein orthogonales ℓ-Bein ist. Dieses ist dann eine Basis von V. Im Falle $V = \mathbb{R}^n$ kann man $\det(\vec{x}_1, \ldots, \vec{x}_n) = 1$ erreichen.

Beweis

Wiederholte Anwendung von Satz 6.3.4 liefert die Existenz. Die Vektoren $\vec{x}_1, \ldots, \vec{x}_\ell$ sind dann linear unabhängig. Da aber $\ell = \dim V$ ist, erzeugen sie auch V. Im Falle $\ell = n$ ist $\det = \pm 1$. Ersetzt man ggfs. einen Vektor durch sein Negatives, erhält man $\det = +1$. □

In den Beweisen dieser beiden Sätze wird ein explizites Konstruktionsverfahren gegeben. Es ist in der höheren Analysis von großer Bedeutung.

Das orthogonale Komplement

Es sei V ein k-dimensionaler Untervektorraum des $\mathbb{R}^n$. Wir setzen

$$V^\perp = \{\vec{y} \in \mathbb{R}^n : \ \vec{y} \cdot \vec{x} = 0 \text{ für alle } \vec{x} \in V\}.$$

Die Vektorenmenge $V^\perp$ ist dann gegenüber Addition und Skalarenmultiplikation abgeschlossen und ein Untervektorraum von $\mathbb{R}^n$. Wir nennen ihn das *orthogonale Komplement von V*.

6.3.6 Satz

Es ist $\dim V^\perp = n - k$. Nimmt man ein orthogonales k-Bein in V und ein orthogonales $(n-k)$-Bein in $V^\perp$, so ist das aus diesen Vektoren zusammengesetzte n-tupel ein orthogonales n-Bein im $\mathbb{R}^n$.

Beweis

Wir beginnen mit einem orthogonalen k-Bein $(\vec{x}_1, \ldots, \vec{x}_k)$ in V. Wir bilden die Matrix

$$A = \begin{pmatrix} \vec{x}_1^{\,t} \\ \vdots \\ \vec{x}_k^{\,t} \end{pmatrix}.$$

Es ist $\vec{y} \cdot \vec{x} = 0$ für alle $\vec{x} \in V$, wenn das für alle $\vec{x} = \vec{x}_i$, $i = 1, \ldots, k$ gilt, da alle $\vec{x}$ Linearkombination der $\vec{x}_i$ sind. Genau dann ist $\vec{y} \in V^\perp$. Also ist $V^\perp = \left\{\vec{y} : A \circ \vec{y} = \vec{0}\right\}$. Die Matrix A hat k linear unabhängige Zeilen und damit den Rang k. Dieser Rang ist gleich dem Spaltenrang nach Kapitel 3.4. Das Bild unter der linearen Abbildung $\vec{y} \mapsto A \circ \vec{y}$ ist also der ganze $\mathbb{R}^k$. Der Kern dieser linearen Abbildung ist aber $V^\perp$. Er hat damit die Dimension $n - k$.

Wählt man nun irgendein orthogonales $(n - k)$-Bein $(\vec{y}_1, \ldots, \vec{y}_{n-k})$ in $V^\perp$, so gilt immer $\vec{x}_i \cdot \vec{y}_j = 0$ für $i = 1, \ldots, k$ und $j = 1, \ldots, n - k$. Also sind diese Vektoren paarweise zueinander orthogonal. Wir haben ein orthogonales n-Bein im $\mathbb{R}^n$. $\square$

Man hat nun zwei unmittelbar einsichtige Folgerungen:

(a) Es ist $V \cap V^\perp = \left\{\vec{0}\right\}$;

(b) jeder Vektor aus $\mathbb{R}^n$ ist eindeutig bestimmte Summe eines Vektors aus V und eines Vektors aus $V^\perp$.

Das heißt: $\mathbb{R}^n = V \oplus V^\perp$.

6.4 Das Kreuzprodukt

Wir wollen in diesem Abschnitt das Kreuzprodukt von $n - 1$ Vektoren des $\mathbb{R}^n$ definieren. Im Falle $n = 3$ hängt es anschaulich eng mit dem Flächeninhalt des Parallelogramms zusammen, das von den beiden Vektoren aufgespannt wird. Wir machen folgende Betrachtung:

Es seien $\vec{x}_2, \ldots, \vec{x}_n$ linear unabhängige Vektoren des $\mathbb{R}^n$. Dann ist der von diesen $n - 1$ Vektoren aufgespannte Vektorraum V $(n - 1)$-dimensional und damit $V^\perp$ eindimensional. Es gibt zwei Einheitsvektoren $\pm\vec{y}$, die auf V senkrecht stehen. Wir setzen

$$\vec{d} := \det(\pm\vec{y}, \vec{x}_2, \ldots, \vec{x}_n) \cdot (\pm\vec{y}).$$

Dieser Vektor ist unabhängig von der Auswahl des Vorzeichens von $\vec{y}$ bestimmt.

6.4.1 Definition

Das **Kreuzprodukt** $\vec{x}_2 \times \ldots \times \vec{x}_n$ *ist ein Vektor des* $\mathbb{R}^n$. *Sind die Vektoren* $\vec{x}_2, \ldots, \vec{x}_n$ *linear abhängig, so ist es* $\vec{0}$. *Im anderen Falle ist es* $\vec{d}$.

Wir definieren nun für $i = 1, \ldots, n$:

$$z_i = -(-1)^i \cdot \det \begin{pmatrix} x_{12} & \ldots & x_{1n} \\ \vdots & & \vdots \\ x_{i2} & \text{nicht} & x_{in} \\ \vdots & & \vdots \\ x_{n2} & \ldots & x_{nn} \end{pmatrix}.$$

Bei der Determinante wird also die i-te Zeile ausgelassen. Es gilt:

6.4.2 Satz

Für das Kreuzprodukt ist: $d_i = z_i$.

Beweis

Nach dem Entwicklungssatz aus Kapitel 5.4 hat man für alle $\vec{x} \in \mathbb{R}^n$

$$\vec{x} \cdot \vec{z} = \sum_{i=1}^{n} x_i z_i = \det(\vec{x}, \vec{x}_2, \ldots, \vec{x}_n) \qquad (*)$$

und insbesondere $\vec{z} \cdot \vec{x}_i = \det(\vec{x}_i, \vec{x}_2, \ldots, \vec{x}_n) = 0$ für $i = 2, \ldots, n$. Der Vektor $\vec{z}$ ist also in $V^\perp$. Es folgt $\det(\vec{z}, \vec{x}_2, \ldots, x_n) = \sum_{i=1}^{n} z_i^2 = \|\vec{z}\|^2$. Wir zeigen, daß $\vec{z} \neq \vec{0}$ ist. Dazu ergänzen wir das linear unabhängige System $\vec{x}_2, \ldots, \vec{x}_n$ durch einen geeigneten Vektor $\vec{x}_1 \in \mathbb{R}^n$ zu einer Basis. Auf Grund von $(*)$ gilt $\vec{x}_1 \cdot \vec{z} = \det(\vec{x}_1, \vec{x}_2, \ldots \vec{x}_n) \neq 0$, also $\vec{z} \neq \vec{0}$. Das bedeutet für den Einheitsvektor $\vec{y} = (1/\|\vec{z}\|) \cdot \vec{z}$ die Gleichung $\|\vec{z}\| = \det(\vec{y}, \vec{x}_2, \ldots, \vec{x}_n)$ und $\vec{z} = \|\vec{z}\| \cdot \vec{y} = \det(\vec{y}, \vec{x}_2, \ldots, \vec{x}_n) \cdot \vec{y}$. Also ist $\vec{z}$ das Kreuzprodukt. □

6.4.3 Satz

Die Abbildung $\times : \underbrace{\mathbb{R}^n \oplus \ldots \oplus \mathbb{R}^n}_{(n-1)\text{-mal}} \to \mathbb{R}^n$ *hat folgende Eigenschaften:*

(a) *es gilt* $\vec{x}_i \perp (\vec{x}_2 \times \ldots \times \vec{x}_n)$ *für* $i = 2, \ldots, n$*;*

(b) $\times$ *ist multilinear, d. h. man hat für alle* i*:*

$$\begin{aligned}&\vec{x}_2 \times \ldots \times (\vec{x}_i + \vec{x}_i{}') \times \ldots \times \vec{x}_n \\ &= (\vec{x}_2 \times \ldots \times \vec{x}_i \times \ldots \times \vec{x}_n) + (\vec{x}_2 \times \ldots \times \vec{x}_i{}' \times \ldots \times \vec{x}_n)\end{aligned}$$

und für alle $a \in \mathbb{R}$*:*

$$\vec{x}_2 \times \ldots \times (a \cdot \vec{x}_i) \times \ldots \times \vec{x}_n = a \cdot (\vec{x}_2 \times \ldots \times \vec{x}_i \times \ldots \times \vec{x}_n);$$

(c) $\times$ *ist alternierend, d. h. das Kreuzprodukt ändert beim Vertauschen zweier Vektoren das Vorzeichen;*

(d) *es gilt* $\vec{x}_2 \times \ldots \times \vec{x}_n = \vec{0}$ *genau dann, wenn* $\vec{x}_2, \ldots, \vec{x}_n$ *linear abhängig sind.*

Beweis

(a) Gilt nach Definition.

(b) Wir verwenden Satz 6.4.2. Determinanten sind multilinear!

(c) Ebenso folgt „alternierend“.

(d) Nach Definition ist das Kreuzprodukt gleich $\vec{0}$, wenn die Vektoren linear abhängig sind. Sei nun umgekehrt das Kreuzprodukt gleich $\vec{0}$. Wir nehmen an, daß die Vektoren nicht linear abhängig sind. Wir hatten dann den orthogonalen Einheitsvektor $\vec{y}$ konstruiert. Er ist $\neq \vec{0}$. Es muß jedoch $\det(\vec{y}, \vec{x}_2, \ldots, \vec{x}_n) = \vec{y} \cdot (\vec{x}_2 \times \ldots \times \vec{x}_n) = 0$ sein. Dann sind aber die Vektoren $\vec{y}, \vec{x}_2, \ldots, \vec{x}_n$ linear abhängig. Da aber $\vec{y}$ senkrecht auf $\vec{x}_2, \ldots, \vec{x}_n$ steht, kann $\vec{y}$ nicht Linearkombination von $\vec{x}_2, \ldots, \vec{x}_n$ sein. In der Kombination des Nullvektors durch $\vec{y}, \vec{x}_2, \ldots, \vec{x}_n$ muß also der Koeffizient bei $\vec{y}$ gleich 0 sein. Es sind also $\vec{x}_2, \ldots, \vec{x}_n$ doch linear abhängig. Das ist ein Widerspruch! □

Das Kreuzprodukt ist invariant gegenüber Drehungen:

6.4.4 Satz

Für $A \in SO(n)$ gilt stets:

$$(A \circ \vec{x}_2) \times \ldots \times (A \circ \vec{x}_n) = A \circ (\vec{x}_2 \times \ldots \times \vec{x}_n).$$

Beweis

Das Senkrechtstehen und die Länge eines Vektors sind invariant gegenüber Drehungen und sogar gegenüber allgemeinen orthogonalen Transformationen. Also wird $\vec{y}$ auf $A \circ \vec{y}$ abgebildet. Nach der Determinantentheorie ist $\det(A \circ \vec{y}, A \circ \vec{x}_2, \ldots, A \circ \vec{x}_n) = \det(A \circ (\vec{y}, \vec{x}_2, \ldots, \vec{x}_n)) = \det(A) \cdot \det(\vec{y}, \vec{x}_2, \ldots, \vec{x}_n) = \det(\vec{y}, \vec{x}_2, \ldots, \vec{x}_n)$, wenn A eine Drehung ist. □

Ist dagegen A eine Drehspiegelung, so ist die Determinante gleich -1. Das neue Kreuzprodukt ist dann $-A \circ (\vec{x}_2 \times \ldots \times \vec{x}_n)$.

Im anschaulichen $\mathbb{R}^3$

Wir betrachten nun nur noch den Fall $n = 3$. Es ist hier $\times : \mathbb{R}^3 \oplus \mathbb{R}^3 \to \mathbb{R}^3$ ein Produkt im $\mathbb{R}^3$. Man hat folgende Rechenregeln:

(a) *Distributivgesetze:*

$$(\vec{x} + \vec{y}\,) \times \vec{z} = \vec{x} \times \vec{z} + \vec{y} \times \vec{z}, \quad \vec{x} \times (\vec{y} + \vec{z}\,) = \vec{x} \times \vec{y} + \vec{x} \times \vec{z}.$$

(b) *Für $a \in \mathbb{R}$ gilt:* $(a \cdot \vec{x}\,) \times \vec{y} = \vec{x} \times (a \cdot \vec{y}\,) = a \cdot (\vec{x} \times \vec{y}\,)$.

(c) *Das Kreuzprodukt ist antikommutativ:* $\vec{x} \times \vec{y} = -(\vec{y} \times \vec{x}\,)$.

(d) $\vec{x} \cdot (\vec{y} \times \vec{z}\,) = \det(\vec{x}, \vec{y}, \vec{z}\,)$.

Beweis

Es sei $\vec{d} = \vec{y} \times \vec{z}$ wieder dargestellt als $\pm$Unterdeterminanten der Matrix $(\vec{y}, \vec{z}\,)$. Dann ist nach dem Entwicklungssatz $\vec{x} \cdot (\vec{y} \times \vec{z}\,) = \sum_{i=1}^{3} x_i \cdot d_i = \det(\vec{x}, \vec{y}, \vec{z}\,)$. □

(e) *Ist* $A \in SO(3)$ *eine Drehung, so gilt:* $A \circ (\vec{x} \times \vec{y}\,) = (A \circ \vec{x}\,) \times (A \circ \vec{y}\,)$.

Es sei hier auf eine Besonderheit des $\times$-Produktes hingewiesen. Es ist *nicht assoziativ*, wie das folgende Beispiel lehrt:

$$\vec{e}_3 \times (\vec{e}_1 \times (1,1,0)) = \vec{e}_3 \times \vec{e}_3 = \vec{0}, \quad (\vec{e}_3 \times \vec{e}_1) \times (1,1,0) = -\vec{e}_3.$$

Es gibt ein komplizierteres Gesetz, das hier ohne Beweis angegeben wird. Es gilt die *Jacobi-Identität*:

$$(\vec{x} \times \vec{y}\,) \times \vec{z} + (\vec{y} \times \vec{z}\,) \times \vec{x} + (\vec{z} \times \vec{x}\,) \times \vec{y} = \vec{0}.$$

Um eine anschauliche Deutung des Kreuzproduktes zu geben, gehen wir wie folgt vor.

6.4.5 Satz

Es seien $\vec{x}, \vec{y} \in \mathbb{R}^3$ *linear unabhängig,* V *die von* $\vec{e}_1, \vec{e}_2$ *aufgespannte Ebene. Dann gibt es eine Drehung* $A \in SO(3)$*, so daß gilt:*

(a) $A \circ \vec{x} = a \cdot \vec{e}_1$ *mit einem* $a > 0$*;*

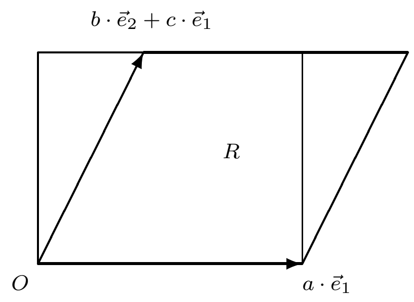
(b) $A \circ \vec{y} \in V$*;*

(c) die zweite Komponente von $A \circ \vec{y}$ *ist positiv.*

Beweis

Es sei W die von $\vec{x}, \vec{y}$ aufgespannte Ebene. Wir setzen $\vec{x}_1 = (1/\|\,\vec{x}\,\|) \cdot \vec{x}$. Wegen Satz 6.3.4 gibt es ein $\vec{x}_2 \in W$, so daß $(\vec{x}_1, \vec{x}_2)$ ein orthogonales 2-Bein in W ist. Nach Satz 6.3.5 existiert dann ein $\vec{x}_3 \in \mathbb{R}^3$, so daß $(\vec{x}_1, \vec{x}_2, \vec{x}_3)$ ein orthogonales 3-Bein mit $\det(\vec{x}_1, \vec{x}_2, \vec{x}_3) = 1$ ist. Es ist also $B = (\vec{x}_1, \vec{x}_2, \vec{x}_3) \in SO(3)$. Es sei $A = B^{-1} \in SO(3)$. Man hat $B \circ \vec{e}_1 = \vec{x}_1$, also $A \circ \vec{x}_1 = \vec{e}_1$ und damit $A \circ \vec{x} = \|\,\vec{x}\,\| \cdot \vec{e}_1$. Genauso folgt $A \circ \vec{x}_2 = \vec{e}_2$, $A \circ \vec{x}_3 = \vec{e}_3$. Bei der Linearkombination $\vec{y}$ von $\vec{x}_1, \vec{x}_2$ liegt $A \circ \vec{y}$ in V. Ist nun die zweite Komponente dieses Vektors negativ, so ersetzen wir $\vec{x}_2 \to -\vec{x}_2$, $\vec{x}_3 \to -\vec{x}_3$. Die Determinante ist dann 1 geblieben, aber die zweite Komponente von $A \circ \vec{y}$ ist nun positiv, da eine Spiegelung in V an der ersten Koordinatenachse stattgefunden hat. $\square$

Wir benutzen hier, ohne ausführlich auf eine präzise Definition einzugehen, einen Begriff des Flächeninhaltes, der insbesondere bei Drehungen und Translationen nicht verändert wird. Ist $P \subset W$ das Parallelogramm $\{\vec{z} = t_1\vec{x} + t_2\vec{y} : \; 0 \le t_1 \le 1, \; 0 \le t_2 \le 1\}$ und $Q \subset V$ das Bildparallelogramm von P unter der Drehung A gemäß Satz 6.4.5, so haben beide gleichen Flächeninhalt. Es sei $A \circ \vec{x} = \vec{v}_1 = a\vec{e}_1$, $A \circ \vec{y} = \vec{v}_2 = b\vec{e}_2 + c\vec{e}_1$. Dann gilt $a > 0$, $b > 0$ und $(A \circ \vec{x}\,) \times (A \circ \vec{y}\,) = (a\vec{e}_1) \times (b\vec{e}_2 + c\vec{e}_1) = ab\vec{e}_3$. Andererseits ist der Flächeninhalt von Q auch gleich ab, wie man an der folgenden Zeichnung sieht.



Parallelogramm und Rechteck

Das Rechteck R hat den Flächeninhalt ab. Die beiden Dreiecke links und rechts in der Zeichnung gehen durch Parallelverschiebung auseinander hervor. Sie haben deswegen den gleichen Inhalt. Nimmt man von R das linke Dreieck und fügt das rechte hinzu, so erhält man aus R das Parallelogramm Q. Also hat auch Q den Inhalt ab. – Das Parallelogramm P hat also den Flächeninhalt ab. Es gilt auch $ab = \| \vec{x} \times \vec{y} \|$.

Die Orientierung

Wir betrachten nun linear unabhängige Vektoren $\vec{v}_1, \vec{v}_2, \vec{v}_3$. Diese bilden ein Rechtssystem, wenn gilt: Legt man den Daumen der rechten Hand in Richtung $\vec{v}_1$, den Zeigefinger in Richtung $\vec{v}_2$, so zeigt der Mittelfinger in Richtung $\vec{v}_3$. Würde man die linke Hand verwenden, spräche man von einem Linkssystem. Es gibt Rechts- und Linkssysteme.

6.4.6 Satz

Bei einer Drehung um eine Achse geht ein Rechtssystem in ein Rechtssystem über.

Beweis

Die Achse werde durch einen Einheitsvektor $\vec{v}$ bezeichnet. Die Drehungen um $\vec{v}$ erhält man auf folgendem Wege. Man ergänzt $\vec{v}$ zu einem orthogonalen 3-Bein mit Determinante 1 und faßt dann dieses Dreibein als Drehung B auf, die $(\vec{e}_1, \vec{e}_2, \vec{e}_3)$ auf dieses 3-Bein abbildet. Die Drehungen um $\vec{v}$ sind dann die Abbildungen $B \circ C \circ B^{-1}$, wobei C die Drehungen um $\vec{e}_1$, also die konstant in Richtung $\vec{e}_1$ fortgesetzten Drehungen der $(\vec{e}_2, \vec{e}_3)$-Ebene um einen Winkel Θ sind. Man kann diesen Winkel stetig auf 0 verkleinern und dadurch die Drehung um $\vec{v}$ stetig auf die Identität deformieren. Durch die inverse Deformation erhält man eine Schar von 3-Beinen, wenn man anfangs ein (nicht notwendig orthogonales) 3-Bein vorgegeben hat. Ist dieses ein Rechtssystem, so besteht die ganze Schar aus Rechtssystemen, da diese wegen der Stetigkeit nicht plötzlich auf Linkssysteme springen können. Ein Rechtssystem geht also durch eine Drehung um $\vec{v}$ in eine Rechtssystem über. $\square$

Wir werden auch benutzen:

6.4.7 Satz

Jede Drehung ist das Produkt von zwei Drehungen um jeweils eine Achse.

Beweis

Die Drehung möge das orthogonale 3-Bein $(\vec{e}_1, \vec{e}_2, \vec{e}_3)$ auf das orthogonale 3-Bein $(\vec{x}_1, \vec{x}_2, \vec{x}_3)$ abbilden. Wir drehen zunächst um einen Einheitsvektor $\vec{v}$, der zu $\vec{e}_1, \vec{x}_1$ orthogonal ist, den Vektor $\vec{e}_1$ auf $\vec{x}_1$ und danach weiter den mitgedrehten Vektor $\vec{e}_2$ um $\vec{x}_1$ auf $\vec{x}_2$. Nach den beiden Drehungen gilt $\vec{x}_3 = \pm\,\vec{x}_3{}'$, wobei $\vec{x}_3{}'$ der Vektor $\vec{e}_3$ nach den beiden Drehungen ist. Es muß dann aber die Gleichheit gelten, da $\det(\vec{x}_1, \vec{x}_2, \vec{x}_3) = \det(\vec{x}_1, \vec{x}_2, \vec{x}_3{}') = 1$ ist. □

Man kann allerdings auch zeigen, z. B. mit Hilfe der Eigenwerttheorie aus dem nächsten Kapitel, daß jede Drehung (in $\mathbb{R}^3$!) eine Drehung um eine einzige Achse ist. Dazu vgl. man Satz 7.3.21 und die Erläuterungen auf S. 257.

Durch eine beliebige Drehung geht also ein Rechtssystem in ein Rechtssystem über. Man hat also, indem man die Sätze 6.4.5–6.4.7 kombiniert und beachtet, daß für $A \circ \vec{x}$, $A \circ \vec{y}$ aus Satz 6.4.5 das Kreuzprodukt $(A \circ \vec{x}) \times (A \circ \vec{y})$ in $\vec{e}_3$-Richtung zeigt:

6.4.8 Satz

Es ist $\vec{x}, \vec{y}, \vec{x} \times \vec{y}$ ein Rechtssystem. Die Länge des Produktvektors ist der Inhalt des von $\vec{x}$, $\vec{y}$ aufgespannten Parallelogramms P. Der Produktvektor steht senkrecht auf den beiden Vektoren $\vec{x}$, $\vec{y}$.

Ebenso sind auch $\vec{x} \times \vec{y}, \vec{x}, \vec{y}$ und $\vec{y}, \vec{x} \times \vec{y}, \vec{x}$ Rechtssysteme.

Wir nennen die Gesamtheit der Rechtssysteme im $\mathbb{R}^3$ eine *Orientierung* des $\mathbb{R}^3$. Spiegelungen verwandeln Rechtssysteme in Linkssysteme. Das $\times$-Produkt ist also nicht gegen Spiegelungen (an Ebenen) invariant. Es ist im $\mathbb{R}^3$ nur dann koordinatenunabhängig definiert, wenn dieser mit einem Längenbegriff und einer Orientierung versehen ist.

Wir erhalten die Orientierung durch den Körper des Menschen: linke Hand – rechte Hand. Seit den fünfziger Jahren ist aber bekannt, daß die Physik nicht invariant gegen Spiegelungen ist. 1957 erhielten Yang und Lee dafür den Nobelpreis (Kobalt-60-Experiment). Später wurde gezeigt, daß die Invarianz auch nicht nach nachfolgender Vertauschung von Teilchen und Antiteilchen gilt. Der physikalisch wirkliche Raum ist also auch ohne menschliche Festlegung orientiert. (Wir kennen nun drei Arten von Räumen: den realen physikalischen Raum, den anschaulichen Raum und die vielen mathematischen Räume.)

7 Eigenwerte und Hauptachsentransformation

In diesem Kapitel werden wir eingehend die am Ende von Abschnitt 3.6 aufgeworfene Frage studieren: Sei $F : V \to V$ eine lineare Abbildung des n-dimensionalen Vektorraums V in sich, $G : \mathbb{R}^n \to V$ eine Karte, $G^{-1} : V \to \mathbb{R}^n$ das entsprechende Koordinatensystem und A die zur Abbildung in den Koordinaten $G^{-1} \circ F \circ G$ gehörige $n \times n$-Matrix.

Für welche Matrizen A läßt sich eine Koordinatenwechselabbildung $G^{-1} \circ \hat{G} : \mathbb{R}^n \to \mathbb{R}^n$ mit zugehöriger $n \times n$-Matrix B finden, so daß $\hat{A} := B^{-1} \circ A \circ B$ *Diagonalgestalt*

$$\hat{A} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \dots & \dots & \lambda_n \end{pmatrix}$$

hat? Wir wollen uns hier auf den Fall beschränken, daß die Karten G und $\hat{G}$ auseinander durch Drehungen hervorgehen, daß die Transformationsmatrix B also orthogonal ist. In diesem Fall sagt man, daß B die *Hauptachsentransformation* für die Matrix A vermittelt.

Man ist deshalb so an einer Diagonalform interessiert, weil diese einerseits eine besonders naheliegende Interpretation der entsprechenden Abbildungen gestattet und weil so andererseits Matrizen auf einfache Weise auch komplexen Berechnungen wie Potenzieren oder unendlichen Reihenbildungen zugänglich gemacht werden.

Hat $\hat{A}$ Diagonalgestalt, so gilt $\hat{A} \circ \vec{e}_j = \lambda_j \vec{e}_j$ und deshalb für die Basiselemente $\hat{x}_j := \hat{G}(\vec{e}_j)$ von V:

$$F(\hat{x}_j) = \lambda_j \hat{x}_j.$$

Derartige Vektoren $\hat{x}_j$ heißen *Eigenvektoren* zum jeweiligen *Eigenwert* λ_j, diese werden unter der Anwendung von F lediglich um den Faktor λ_j gestreckt. Solche Eigenvektoren werden beim Aufsuchen der Diagonalform $\hat{A}$ von A eine grundlegende Rolle spielen, dazu vgl. man den Abschnitt 7.3.

Es wird sich herausstellen, daß eine *geschlossene* Behandlung der angesprochenen Fragen für reelle Vektorräume im allgemeinen nicht möglich ist. Deshalb werden wir im folgenden Abschnitt zunächst den Körper der komplexen Zahlen und komplexe Vektorräume einführen und in Abschnitt 7.2 die unitären Matrizen als Verallgemeinerung der orthogonalen Matrizen behandeln.

In Abschnitt 7.4 werden wir, zunächst im *Komplexen*, für die „größtmögliche“ Klasse der „normalen“ Matrizen Diagonalisierungsresultate herleiten. Ein entsprechendes *reelles* Resultat können wir daraus im allgemeinen nur für die wesentlich speziellere Klasse der „symmetrischen (selbstadjungierten)“ Matrizen folgern, da nur in diesem Fall stets die Existenz hinreichend vieler *reeller* Eigenwerte gesichert ist.

7.1 Der allgemeine Körperbegriff, komplexe Vektorräume

Allgemeine Körper

7.1.1 Definition

Eine Menge K, die mit einer Addition

$$+ : K \times K \to K$$

und einer Multiplikation

$$\cdot : K \times K \to K$$

versehen ist, heißt **Körper**, *falls gilt:*

(1) $(K, +)$ *ist eine abelsche Gruppe.*

(2) *Es bezeichne* 0 *das Nullelement von* $(K, +)$ *und* $K^* := K \backslash \{0\}$. *Dann ist* $(K^*, \cdot)$ *ebenfalls eine abelsche Gruppe.*

(3) *Die Distributivgesetze gelten, d. h. für alle* $a, b, c, d \in K$ *ist*

$$(a + b) \cdot c = a \cdot c + b \cdot c, \qquad a \cdot (c + d) = a \cdot c + a \cdot d.$$

Das neutrale Element bzgl. der Multiplikation wird mit 1 bezeichnet und die inversen Elemente bzgl. der Addition mit $(-a)$ und bzgl. der Multiplikation mit a^{-1} oder $\frac{1}{a}$. In jedem Körper gelten insbesondere sowohl für die Addition als auch für die Multiplikation die Regeln, wie wir sie in Abschnitt 1.1 für (abelsche) Gruppen zusammengestellt haben. Darüber hinaus haben wir für die Verknüpfung additiver und multiplikativer Eigenschaften folgende Gesetze:

7.1.2 Satz

Sei K ein Körper, dann gilt für alle $a, b \in K$:

(a) $a \cdot 0 = 0 \cdot a = 0,$

(b) aus $a \cdot b = 0$ folgt $a = 0$ oder $b = 0$, (Nullteilerfreiheit)

(c) $(-a) \cdot b = a \cdot (-b) = -(a \cdot b), \qquad (-a) \cdot (-b) = a \cdot b.$

Für den Beweis sowie für weitere Folgerungen aus den Körperaxiomen verweisen wir auf die Lehrbücher der Analysis etwa von O. Forster [4, § 2] und von H. Grauert, I. Lieb [6, § 3].

Der Körper der reellen Zahlen verfügt zudem über einen Abstandsbegriff (Betrag), bzgl. dessen er vollständig ist, und über eine archimedische Anordnung ($\leq, \geq$).

Der Körper der komplexen Zahlen

Der für das Folgende außerordentlich wichtige Körper $\mathbb{C}$ der komplexen Zahlen wird im allgemeinen in der Differentialrechnung definiert. Daher werden hier manche Eigenschaften nur skizziert und einige Beweise übergangen. Zur weiteren Information sei wieder auf O. Forster [4, § 13] verwiesen.

Wir führen $\mathbb{C}$ als den zweidimensionalen reellen Vektorraum

$$\mathbb{C} := \mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\}$$

ein. Damit ist $\mathbb{C}$ bereits eine additive abelsche Gruppe. Eine Multiplikation

$$\cdot : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$$

wird durch

$$(x, y) \cdot (u, v) := (xu - yv,\ xv + yu)$$

erklärt. Damit wird $\mathbb{C}$ zum Körper: das neutrale Element ist bzgl. der Addition $0 = (0, 0)$ und bzgl. der Multiplikation $1 = (1, 0)$. Das Negative von (x, y) ist $(-x, -y)$, für $(x, y) \neq 0$ gilt $(x, y)^{-1} = \frac{1}{x^2+y^2}(x, -y)$.

Die reellen Zahlen werden folgendermaßen in den komplexen Zahlkörper eingebettet: Die Abbildung

$$\Phi : \mathbb{R} \to \mathbb{C}, \quad x \mapsto (x, 0)$$

ist ein injektiver Körperhomomorphismus, d. h. stets ist

$$\Phi(x + y) = \Phi(x) + \Phi(y) \quad \text{und} \quad \Phi(x \cdot y) = \Phi(x) \cdot \Phi(y).$$

Infolgedessen ist $\Phi(\mathbb{R})$ ein Teilkörper von $\mathbb{C}$, d. h. eine Teilmenge von $\mathbb{C}$ und, mit denselben Verknüpfungen wie $\mathbb{C}$ versehen, ein Körper. Man identifiziert

$$\mathbb{R} = \Phi(\mathbb{R}) = \{(x, 0) : x \in \mathbb{R}\}.$$

Für $a \in \mathbb{R}$, $(x, y) \in \mathbb{C}$ gilt $a \cdot (x, y) = (a, 0) \cdot (x, y) = (a \cdot x, a \cdot y)$, die Skalarmultiplikation auf $\mathbb{R}^2$ ist also ein Spezialfall der Multiplikation auf $\mathbb{C}$.

Durch Einführung der *imaginären Einheit* $i := (0,1)$ mit $i^2 = -1$ gelangt man zur üblichen Darstellung komplexer Zahlen

$$(x,y) = (x,0) + (0,y) = x \cdot 1 + y \cdot i = x + iy.$$

Da jetzt i stets die imaginäre Einheit bezeichnet, steht es als Summationsindex nicht mehr zur Verfügung. Deshalb ersetzen wir fortan die bisherigen Standard-Summationsindizes i, j durch μ, ν.

Für $z = x + iy \in \mathbb{C}$ ist $\mathrm{Re}(z) := x$ der *Realteil* und $\mathrm{Im}(z) := y$ der *Imaginärteil* von z. Zu jeder komplexen Zahl $z = x + iy$ wird die *konjugiert komplexe Zahl* $\bar{z} = x - iy$ erklärt. Wegen $\bar{\bar{z}} = z$ ist die Konjugation eine bijektive Abbildung. Ferner gilt $\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2$ und $\overline{z_1 \cdot z_2} = \bar{z}_1 \cdot \bar{z}_2$, die Konjugation ist also ein Körperisomorphismus.

Da reelle Zahlen durch die Konjugation nicht verändert werden, ist diese insbesondere eine bijektive lineare Abbildung des $\mathbb{R}$-Vektorraums $\mathbb{C}$ in sich, eine *umkehrbar $\mathbb{R}$-lineare Abbildung*. Sie ist anschaulich eine Spiegelung an der x-Achse. Die Elemente $iy, y \in \mathbb{R}$, der y-Achse heißen *rein imaginär*; es gilt: $\mathbb{R} = \{(x,0) : x \in \mathbb{R}\} = \{z : z = \bar{z}\}$ und $i\mathbb{R} := \{iy : y \in \mathbb{R}\} = \{(0,y) : y \in \mathbb{R}\} = \{z : z = -\bar{z}\}$.

Mittels der Konjugation erhält man geschlossene Formeln für Real- und Imaginärteil:

$$\mathrm{Re}(z) = \frac{1}{2}(z + \bar{z}), \quad \mathrm{Im}(z) = \frac{1}{2i}(z - \bar{z}).$$

Betrag einer komplexen Zahl

Für $z = x + iy \in \mathbb{C}$ wird entsprechend der euklidischen Länge in $\mathbb{R}^2$ durch

$$|z| := \sqrt{x^2 + y^2}$$

der Betrag erklärt, offensichtlich ist $|z|^2 = z \cdot \bar{z}$. Damit erhält man für $z \neq 0$ eine einfache Formel zur Bestimmung der komplexen Kehrwerte:

$$\frac{1}{z} = \frac{\bar{z}}{|z|^2}.$$

Es gelten die typischen Eigenschaften eines Betrages, für alle $z, z_1, z_2 \in \mathbb{C}$ ist:

(1) $|z| = 0 \Leftrightarrow z = 0$,

(2) $|z_1 \cdot z_2| = |z_1| \cdot |z_2|$,

(3) $|z_1 + z_2| \leq |z_1| + |z_2|$. (Dreiecksungleichung)

Die erste und die dritte Eigenschaft folgen unmittelbar daraus, daß der Betrag der komplexen Zahl $z = x + iy$ mit der euklidischen Länge des Vektors $(x,y) \in \mathbb{R}^2$ übereinstimmt. Für die multiplikative Eigenschaft berechnet man

$$|z_1 \cdot z_2| = |(x_1 + iy_1) \cdot (x_2 + iy_2)| = |(x_1x_2 - y_1y_2) + i(x_1y_2 + x_2y_1)|$$

$$= \sqrt{x_1^2 x_2^2 + y_1^2 y_2^2 + x_1^2 y_2^2 + x_2^2 y_1^2} = \sqrt{(x_1^2 + y_1^2) \cdot (x_2^2 + y_2^2)} = |z_1| \cdot |z_2|.$$

Da Konvergenz einer komplexen Zahlenfolge $(z_k)_{k\in\mathbb{N}} = (x_k + iy_k)_{k\in\mathbb{N}}$ äquivalent zur Konvergenz der beiden reellen Zahlenfolgen $(x_k)_{k\in\mathbb{N}}$ und $(y_k)_{k\in\mathbb{N}}$ ist, erhält man aus der Vollständigkeit von $\mathbb{R}$:

Der Körper $\mathbb{C}$ der komplexen Zahlen ist vollständig.

Im Gegensatz zu $\mathbb{R}$ ist jedoch $\mathbb{C}$ nicht angeordnet, man kann also nicht von „größeren“ und „kleineren“ komplexen Zahlen sprechen.

Komplexe Vektorräume

7.1.3 Definition

Auf einer nichtleeren Menge V sei eine Addition

$$+ : V \times V \to V$$

und eine Skalarmultiplikation

$$\cdot : \mathbb{C} \times V \to V$$

gegeben. Es sei $(V, +)$ eine abelsche Gruppe, es gelten also die Eigenschaften (1)-(4) aus Definition 1.1.3. Ferner gelte für alle $c, d \in \mathbb{C}$ und $x, y \in V$:

(5) $1 \cdot x = x,$ (Unitäres Gesetz)

(6) $(c \cdot d) \cdot x = c \cdot (d \cdot x),$ (Assoziativgesetz)

(7) $(c + d) \cdot x = c \cdot x + d \cdot x$ *und* $c \cdot (x + y) = c \cdot x + c \cdot y.$ (Distributivgesetze)

Dann heißt V ein **komplexer Vektorraum** *oder* **$\mathbb{C}$-Vektorraum**.

Man beachte die nahezu wörtliche Übereinstimmung mit der Definition 1.2.1, es wird nur der Skalarenkörper $\mathbb{R}$ durch $\mathbb{C}$ ersetzt. Mit Ausnahme der Abschnitte, in denen Orthogonalität eine Rolle spielt (d. h. Kapitel 4.5 und Kapitel 6), kann im gesamten Buch $\mathbb{R}$ ohne weiteres durch $\mathbb{C}$ ersetzt werden. Hinsichtlich der Orthogonalität werden wir unten zeigen, daß sich die Definition des Skalarproduktes in geeigneter Weise modifizieren läßt. Zur Vermeidung von Mißverständnissen bezüglich des zugrundeliegenden Skalarenkörpers (man vgl. Satz 7.1.6) spricht man von linearer (Un-) Abhängigkeit *über* $\mathbb{C}$, $\mathbb{C}$*-Dimension* (man schreibt $\dim_{\mathbb{C}}$), usw.

Der n-dimensionale komplexe Zahlenraum

$$\mathbb{C}^n := \{(z_1, \ldots, z_n) : z_\nu \in \mathbb{C}\}$$

ist mit den Verknüpfungen

$$\vec{z} + \vec{w} := (z_1 + w_1, \ldots, z_n + w_n) \quad \text{und} \quad c \cdot \vec{z} := (c \cdot z_1, \ldots, c \cdot z_n)$$

ein komplexer Vektorraum mit der $\mathbb{C}$-Basis

$$\{\vec{e}_1 = (1, 0, \ldots, 0),\ \ldots, \vec{e}_n = (0, \ldots, 0, 1)\}.$$

In derselben Weise, wie $\mathbb{R}$ als Teilmenge von $\mathbb{C}$ aufgefaßt werden kann, ist auch der n-dimensionale *reelle* Zahlenraum $\mathbb{R}^n$ in $\mathbb{C}^n$ enthalten; für $\vec{z} \in \mathbb{C}^n$ schreiben wir $\vec{z} = \vec{x} + i\vec{y}$ mit $\vec{x}, \vec{y} \in \mathbb{R}^n$.

Es gilt wörtlich wie in Satz 3.3.14:

7.1.4 Satz

$\dim_{\mathbb{C}} \mathbb{C}^n = n.$

Da $\mathbb{R}$ ein Teilkörper von $\mathbb{C}$ ist, wird jeder komplexe Vektorraum V durch Einschränkung der Skalarmultiplikation von $\mathbb{C} \times V$ auf $\mathbb{R} \times V$ zu einem reellen Vektorraum. Im Folgenden stellen wir dar, wie die Eigenschaften von V als $\mathbb{R}$-Vektorraum und als $\mathbb{C}$-Vektorraum zusammenhängen.

7.1.5 Satz

Es sei V *ein* $\mathbb{C}$*-Vektorraum und* $\{z_1, \ldots, z_n\}$ *eine Basis von* V *über* $\mathbb{C}$*. Dann ist* $\{z_1, i \cdot z_1, \ldots, z_n, i \cdot z_n\}$ *eine* $\mathbb{R}$*-Basis von* V*.*

Beweis

Wir zeigen zunächst die lineare Unabhängigkeit über $\mathbb{R}$. Seien $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{R}$ mit

$$\sum_{\nu=1}^{n} a_\nu z_\nu + \sum_{\nu=1}^{n} b_\nu (i \cdot z_\nu) = 0.$$

Für $c_\nu := a_\nu + ib_\nu$ folgt $\sum_{\nu=1}^{n} c_\nu z_\nu = 0$. Die lineare Unabhängigkeit über $\mathbb{C}$ liefert $c_1 = \cdots = c_n = 0$, also

$$a_1 = \ldots = a_n = b_1 = \ldots = b_n = 0.$$

Andererseits wird V auch von dem angegebenen System über $\mathbb{R}$ erzeugt. Da $\{z_1, \ldots, z_n\}$ über $\mathbb{C}$ ein Erzeugendensystem von V ist, existieren zu beliebigem $z \in V$ komplexe Zahlen $c_\nu = a_\nu + ib_\nu \in \mathbb{C}$, $a_\nu, b_\nu \in \mathbb{R}$, $\nu = 1, \ldots, n$, mit

$$z = \sum_{\nu=1}^{n} c_\nu z_\nu = \sum_{\nu=1}^{n} (a_\nu + ib_\nu) z_\nu = \sum_{\nu=1}^{n} a_\nu z_\nu + \sum_{\nu=1}^{n} b_\nu (iz_\nu). \qquad \square$$

Eine unmittelbare Folgerung ist

7.1.6 Satz

Sei V *ein komplexer Vektorraum mit* $\dim_{\mathbb{C}} V = n$*. Dann gilt* $\dim_{\mathbb{R}} V = 2n$*.*

Die Isomorphie von $\mathbb{C}^n$ und $\mathbb{R}^{2n}$ über $\mathbb{R}$ sieht man am deutlichsten, indem man im n-tupel $\vec{z} = (z_1, \dots, z_n)$, $z_\nu = x_\nu + iy_\nu$ jede Komponente als Zahlenpaar (x_ν, y_ν) schreibt: $((x_1, y_1), \dots, (x_n, y_n))$. Gemäß Satz 7.1.5 erhält man eine $\mathbb{R}$-Basis von $\mathbb{C}^n$ durch

$$\begin{aligned}&\{\vec{e}_1, i \cdot \vec{e}_1, \dots, \vec{e}_n, i \cdot \vec{e}_n\}\\ &= \Big\{ \Big((1,0),(0,0),\dots,(0,0)\Big),\ \Big((0,1),(0,0),\dots,(0,0)\Big), \dots\\ &\quad \dots, \Big((0,0),\dots,(0,0),(1,0)\Big),\ \Big((0,0),\dots,(0,0),(0,1)\Big)\Big\}.\end{aligned}$$

Komplex lineare Abbildungen

7.1.7 Definition

Unter einer **komplex linearen** *oder* **$\mathbb{C}$-linearen Abbildung** $F : V \to W$ *zwischen den komplexen Vektorräumen* V *und* W *versteht man eine Abbildung, die mit den Vektorraumoperationen verträglich ist, so daß für alle* $x, y \in V$, $c \in \mathbb{C}$ *gilt:*

$$F(x + y) = F(x) + F(y), \qquad F(c \cdot x) = c \cdot F(x).$$

Sei nun $F : \mathbb{C}^n \to \mathbb{C}^n$ eine $\mathbb{C}$-lineare Abbildung, die durch die komplexe $n \times n$-Matrix

$$C = \begin{pmatrix} c_{11} & \dots & c_{1n} \\ \vdots & & \vdots \\ c_{n1} & \dots & c_{nn} \end{pmatrix}, c_{\mu\nu} = a_{\mu\nu} + ib_{\mu\nu},\ a_{\mu\nu}, b_{\mu\nu} \in \mathbb{R},$$

gegeben wird.

Man kann F insbesondere auch als $\mathbb{R}$-lineare Abbildung $\mathbb{R}^{2n} \to \mathbb{R}^{2n}$ auffassen. Wir wollen untersuchen, wie die zugehörige reelle $2n \times 2n$-Matrix A aus C hervorgeht. Dazu sind die Bilder der Einheitsvektoren $\vec{e}_1, i \cdot \vec{e}_1, \dots, \vec{e}_n, i \cdot \vec{e}_n$ zu bestimmen. Für $\nu = 1, \dots, n$ gilt:

$$\begin{aligned} F(\vec{e}_\nu) &= (c_{1\nu}, \dots, c_{n\nu}) = \Big((a_{1\nu}, b_{1\nu}), \dots, (a_{n\nu}, b_{n\nu})\Big), \\ F(i\vec{e}_\nu) &= i \cdot (c_{i\nu}, \dots, c_{n\nu}) = (i \cdot c_{1\nu}, \dots, i \cdot c_{n\nu}) \\ &= \Big((-b_{1\nu}, a_{1\nu}), \dots, (-b_{n\nu}, a_{n\nu})\Big), \end{aligned}$$

und man erhält:

$$A = \begin{pmatrix} a_{11} & -b_{11} & \dots & a_{1n} & -b_{1n} \\ b_{11} & a_{11} & \dots & b_{1n} & a_{1n} \\ \vdots & \vdots & & \vdots & \vdots \\ a_{n1} & -b_{n1} & \dots & a_{nn} & -b_{nn} \\ b_{n1} & a_{n1} & \dots & b_{nn} & a_{nn} \end{pmatrix}.$$

Also geht die reelle Matrix A aus der komplexen Matrix C hervor, indem man jede komplexe Komponente $c_{\mu\nu} = a_{\mu\nu} + ib_{\mu\nu}$ von C durch das reelle 2×2-Kästchen

$$\begin{matrix} a_{\mu\nu} & -b_{\mu\nu} \\ b_{\mu\nu} & a_{\mu\nu} \end{matrix}$$

ersetzt.

Auch die Determinantentheorie für komplexe Matrizen läßt sich genauso wie im reellen Fall durchführen. Insbesondere gilt:

7.1.8 Satz

Es sei $F : \mathbb{C}^n \to \mathbb{C}^n$ eine $\mathbb{C}$-lineare Abbildung mit zugehöriger Matrix $C = C^{n,n}$. Dann sind äquivalent:

(a) Die Abbildung F ist bijektiv.

(b) Die Matrix C ist nichtsingulär.

(c) $\det(C) \neq 0$.

Nach diesen Vorbereitungen definieren wir genauso wie in 3.8.5:

7.1.9 Definition

Die Menge aller nichtsingulären komplexen $n \times n$-Matrizen

$$GL(n, \mathbb{C}) := \{C : \ C \text{ ist komplexe } n \times n\text{-Matrix mit } \det(C) \neq 0\}$$

heißt **allgemeine lineare Gruppe** *vom Rang n über $\mathbb{C}$.*

7.1.10 Bemerkung

Im Zusammenhang mit der Determinantentheorie $\mathbb{C}$-linearer *Abbildungen* ist Vorsicht geboten, da sich hier der Wert der Determinante ändern kann, wenn man zu der Interpretation derselben Abbildung als $\mathbb{R}$-linearer Abbildung zwischen $\mathbb{R}$-Vektorräumen mit doppelt so großer $\mathbb{R}$-Dimension wechselt. Als Beispiel betrachten wir $F : \mathbb{C} \to \mathbb{C}$, $F(z) = -z$; die entsprechende komplexe 1×1-Matrix lautet $C = (-1)$, $\det(C) = -1$. Zu $F : \mathbb{R}^2 \to \mathbb{R}^2$, $F(x, y) = -(x, y)$ gehört dagegen die reelle 2×2-Matrix $A = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ mit $\det(A) = 1$.

Ein komplexes Skalarprodukt

Um auch $\mathbb{C}^n$ mit einem Skalarprodukt zu versehen, geht man von der Idee aus, daß $\mathbb{C}^n$ die euklidische Norm $\| \vec{z} \| = \sqrt{|z_1|^2 + \ldots + |z_n|^2}$ tragen soll und daß diese von dem gesuchten Skalarprodukt erzeugt werden soll: $(\vec{z}, \vec{z}) \stackrel{!}{=} \| \vec{z} \|^2$. Indem man noch das bereits erwähnte Gesetz $|z_\nu|^2 = \bar{z}_\nu \cdot z_\nu$ beachtet, gelangt man zu der Setzung

$$(\vec{z}, \vec{w}) := \sum_{\nu=1}^{n} \bar{z}_\nu \cdot w_\nu.$$

Es stellt sich heraus, daß die axiomatische Beschreibung des Skalarproduktes in $\mathbb{C}^n$ oder allgemeiner in komplexen Vektorräumen gegenüber der Charakterisierung in Satz 6.2.1 für den $\mathbb{R}^n$ bzw. für reelle Vektorräume hinsichtlich der Linearitäts- und Symmetrieeigenschaften zu modifizieren ist.

7.1.11 Satz

Für das durch $(\vec{z}, \vec{w}) = \sum_{\nu=1}^{n} \bar{z}_\nu \cdot w_\nu$ *auf* $\mathbb{C}^n$ *gegebene* **kanonische komplexe Skalarprodukt** *gilt:*

(1) Es ist **konjugiert linear** *in der ersten Komponente, d. h. für* $c \in \mathbb{C}$, $\vec{z}, \vec{z}\,', \vec{w} \in \mathbb{C}^n$ *gilt:*

$$(\vec{z} + \vec{z}\,', \vec{w}) = (\vec{z}, \vec{w}) + (\vec{z}\,', \vec{w}), \quad (c \cdot \vec{z}, \vec{w}) = \bar{c} \cdot (\vec{z}, \vec{w}).$$

(2) Es ist linear in der zweiten Komponente, d. h. für $c \in \mathbb{C}$, $\vec{z}, \vec{w}, \vec{w}\,' \in \mathbb{C}^n$ *gilt:*

$$(\vec{z}, \vec{w} + \vec{w}\,') = (\vec{z}, \vec{w}) + (\vec{z}, \vec{w}\,'), \quad (\vec{z}, c \cdot \vec{w}) = c \cdot (\vec{z}, \vec{w}).$$

(3) Es ist **Hermitesch** *(konjugiert symmetrisch), d. h. für* $\vec{z}, \vec{w} \in \mathbb{C}^n$ *gilt:*

$$(\vec{z}, \vec{w}) = \overline{(\vec{w}, \vec{z})}.$$

(4) Es ist **positiv definit**, *d. h. für* $\vec{z} \in \mathbb{C}^n$ *ist* $(\vec{z}, \vec{z})$ *reell, und es gilt:*

$$(\vec{z}, \vec{z}) \geq 0 \quad \textit{und} \quad (\vec{z}, \vec{z}) > 0, \ \textit{falls } \vec{z} \neq \vec{0}.$$

Beweis

Der Beweis folgt durch elementares Nachrechnen aus der Definition des Skalarprodukts. □

Abbildungen $\mathbb{C}^n \oplus \mathbb{C}^n \to \mathbb{C}$ mit den Eigenschaften (1) und (2) nennt man *sesquilinear*. Das kanonische komplexe Skalarprodukt ist also eine Hermitesche, positiv definite *Sesquilinearform*.

Es wird in der Literatur nicht einheitlich gehandhabt, ob das Skalarprodukt in der ersten oder in der zweiten Komponente als konjugiert linear definiert wird. Häufig wird z. B. das Skalarprodukt in $\mathbb{C}^n$ auch durch $\sum_{\nu=1}^{n} z_\nu \bar{w}_\nu$ eingeführt.

7.1.12 Satz

Durch die Setzung

$$\| \vec{z} \| := \sqrt{(\vec{z}, \vec{z})} = \sqrt{\sum_{\nu=1}^{n} \bar{z}_\nu \cdot z_\nu}$$

wird $\mathbb{C}^n$ *zu einem* **normierten komplexen Vektorraum**. *Die Abbildung* $\| \,.\, \| : \mathbb{C}^n \to \mathbb{R}$ *hat also die typischen Eigenschaften einer Norm; für alle* $c \in \mathbb{C}$, $\vec{z}, \vec{w} \in \mathbb{C}^n$ *gilt:*

(1) $\| \vec{z} \| = 0 \Leftrightarrow \vec{z} = \vec{0}$,

(2) $\| c \cdot \vec{z} \| = |c| \cdot \| \vec{z} \|$,

(3) $\| \vec{z} + \vec{w} \| \le \| \vec{z} \| + \| \vec{w} \|$.

Beweis

Man könnte zunächst die auch in $\mathbb{C}^n$ gültige Cauchy-Schwarzsche Ungleichung

$$|(\vec{z}, \vec{w})| \le \| \vec{z} \| \cdot \| \vec{w} \|$$

herleiten und anschließend genauso wie im Beweis von Satz 6.2.4 verfahren.

Hier wollen wir allerdings direkt auf das dort erzielte Resultat zurückgreifen und verwenden, daß $\mathbb{C}^n$ und $\mathbb{R}^{2n}$ als $\mathbb{R}$-Vektorräume kanonisch isomorph sind. Wie bereits vorher bemerkt, stimmen für

$$\vec{z} = (z_1, \ldots, z_n) = (x_1 + iy_1, \ldots, x_n + iy_n) = ((x_1, y_1), \ldots, (x_n, y_n))$$

die euklidische Norm in $\mathbb{R}^{2n}$ und die euklidische Norm in $\mathbb{C}^n$ überein:

$$\| (x_1, y_1, \ldots, x_n, y_n) \|^2 = \sum_{\nu=1}^{n} \left(x_\nu^2 + y_\nu^2 \right) = \sum_{\nu=1}^{n} \bar{z}_\nu \cdot z_\nu = \| \vec{z} \|^2.$$

Die Eigenschaften (1) und (3) folgen also direkt aus dem reellen Resultat. Für $c \in \mathbb{C}$, $\vec{z} \in \mathbb{C}^n$ gilt weiter:

$$\| c \cdot \vec{z} \|^2 = (c \cdot \vec{z}, c \cdot \vec{z}) = \bar{c} \cdot (\vec{z}, c \cdot \vec{z}) = \bar{c} \cdot c \cdot (\vec{z}, \vec{z}) = |c|^2 \cdot \| \vec{z} \|^2. \qquad \square$$

7.2 Unitäre und adjungierte Abbildungen

Die unitäre Gruppe

Die unitären Abbildungen sind die komplexen Verallgemeinerungen der orthogonalen Abbildungen. Zwar könnte man die folgende Theorie auch allgemein in endlichdimensionalen Vektorräumen mit Skalarprodukt (diese nennt man *unitär* bzw. im reellen Fall *euklidisch*) durchführen, wir beschränken uns aber der Einfachheit halber auf Abbildungen $\mathbb{C}^n \to \mathbb{C}^n$. Zudem kann man sich, ganz ähnlich wie bei allgemeinen linearen Abbildungen in Kapitel 3, durch die Einführung *unitärer* Karten stets auf diesen Fall zurückziehen.

Zur Definition wählen wir eine der Eigenschaften aus Satz 6.3.2.

7.2.1 Definition

Eine komplexe $n \times n$-Matrix C heißt **unitär**, *falls $\| C \circ \vec{z} \| = \| \vec{z} \|$ für alle $\vec{z} \in \mathbb{C}^n$ gilt. Eine lineare Abbildung $F : \mathbb{C}^n \to \mathbb{C}^n$ heißt* **unitär**, *falls die zugehörige Matrix unitär ist. Wir nennen*

$$U(n) := \{C : C \text{ ist unitäre } n \times n\text{-Matrix}\}$$

die **unitäre Gruppe** *vom Rang n.*

Tatsächlich ist die Bezeichnung „unitäre Gruppe“ gerechtfertigt:

7.2.2 Satz

$U(n)$ ist eine Untergruppe von $GL(n, \mathbb{C})$.

Beweis

Offenbar ist die zu $C \in U(n)$ gehörige lineare Abbildung injektiv. Wie in Kapitel 3 folgt daraus, daß C nichtsingulär ist und $C^{-1} \in GL(n, \mathbb{C})$ existiert. Es folgt $U(n) \subset GL(n, \mathbb{C})$ und auf Grund von $\| \vec{z} \| = \| C \circ C^{-1} \circ \vec{z} \| = \| C^{-1} \circ \vec{z} \|$ auch $C^{-1} \in U(n)$. Wegen $E^{n,n} \in U(n)$ ist $U(n) \neq \emptyset$. Schließlich ist $U(n)$ auch gegenüber der Matrizenmultiplikation abgeschlossen, denn mit $C_1, C_2 \in U(n)$ ist wegen $\| C_1 \circ C_2 \circ \vec{z} \| = \| C_2 \circ \vec{z} \| = \| \vec{z} \|$ auch $C_1 \circ C_2 \in U(n)$. □

Wir studieren zunächst die Verbindungen mit dem reellen Begriff der orthogonalen Matrix. Eine reelle Matrix $C \in \mathbb{R}^{n \times n}$ ist natürlich auch eine komplexe Matrix $C \in \mathbb{C}^{n \times n}$ und vermittelt damit sowohl Abbildungen $F_{\mathbb{R}} : \mathbb{R}^n \to \mathbb{R}^n$ als auch $F_{\mathbb{C}} : \mathbb{C}^n \to \mathbb{C}^n$. Die komplexe Abbildung $F_{\mathbb{C}}$ geht dabei aus $F_{\mathbb{R}}$ hervor, indem man für $\vec{z} = \vec{x} + i\vec{y} \in \mathbb{C}^n$, $\vec{x}, \vec{y} \in \mathbb{R}^n$ beachtet:

$$F_{\mathbb{C}}(\vec{z}\,) = F_{\mathbb{C}}(\vec{x} + i\vec{y}\,) = F_{\mathbb{C}}(\vec{x}\,) + iF_{\mathbb{C}}(\vec{y}\,) = C \circ \vec{x} + i\, C \circ \vec{y} = F_{\mathbb{R}}(\vec{x}\,) + iF_{\mathbb{R}}(\vec{y}\,).$$

Dieses Verfahren, das man auch unabhängig vom Matrizenkalkül durchführen kann, heißt *Komplexifizierung*.

7.2.3 Satz

$O(n) \subset U(n)$.

Beweis

Sei $C \in O(n)$ und $\vec{z} = \vec{x} + i\vec{y}$, $\vec{x}, \vec{y} \in \mathbb{R}^n$. Die Eigenschaften der euklidischen Norm in $\mathbb{C}^n$ ergeben:

$$\begin{aligned} \| \vec{z} \|^2 &= \| \vec{x} \|^2 + \| \vec{y} \|^2 = \| C \circ \vec{x} \|^2 + \| C \circ \vec{y} \|^2 \\ &= \| C \circ \vec{x} + i\, C \circ \vec{y} \|^2 = \| C \circ (\vec{x} + i\vec{y}\,) \|^2 = \| C \circ \vec{z} \|^2. \end{aligned}$$

□

Wie bereits im vorhergehenden Abschnitt erläutert, können wir $\mathbb{C}^n$ und $\mathbb{R}^{2n}$ als $\mathbb{R}$-Vektorräume identifizieren und entsprechend $\mathbb{C}$-lineare Abbildungen $\mathbb{C}^n \to \mathbb{C}^n$ als $\mathbb{R}$-lineare Abbildungen $\mathbb{R}^{2n} \to \mathbb{R}^{2n}$ auffassen. In diesem Sinne ist der folgende Satz zu verstehen.

7.2.4 Satz

$U(n)$ ist eine Untergruppe von $O(2n)$.

Beweis

Es sei $\Phi : U(n) \to \mathbb{R}^{2n\times 2n}$ die in Abschnitt 7.1 beschriebene Abbildung: Für $C = (c_{\mu\nu})_{\mu,\nu=1,\dots,n}$ erhält man $A = \Phi(C)$, indem man jede komplexe Komponente $c_{\mu\nu} = a_{\mu\nu} + ib_{\mu\nu}$ durch das reelle 2×2-Kästchen $\begin{pmatrix} a_{\mu\nu} & -b_{\mu\nu} \\ b_{\mu\nu} & a_{\mu\nu} \end{pmatrix}$ ersetzt.

Wir haben also zu zeigen, daß $\Phi : U(n) \to O(2n)$ ein injektiver Gruppenhomomorphismus ist.

Da C und $A = \Phi(C)$ dieselbe Abbildung beschreiben und die euklidische Norm in $\mathbb{C}^n$ mit der in $\mathbb{R}^{2n}$ übereinstimmt, folgt $\Phi(U(n)) \subset O(2n)$ aus Satz 6.3.2.

Wir zeigen nun, daß Φ ein Gruppenhomomorphismus ist. Seien $C_1, C_2 \in U(n)$. Die durch $C_1 \circ C_2$ gegebene Abbildung $\mathbb{C}^n \to \mathbb{C}^n$ stimmt mit der durch $\Phi(C_1) \circ \Phi(C_2)$ gegebenen Abbildung $\mathbb{R}^{2n} \to \mathbb{R}^{2n}$ überein, denn die Komposition der Abbildungen ist unabhängig davon, ob man $\mathbb{C}^n$ als komplexen oder reellen Vektorraum interpretiert. Also gilt $\Phi(C_1 \circ C_2) = \Phi(C_1) \circ \Phi(C_2)$.

Die Injektivität von Φ ergibt sich schließlich aus

$$\Phi(C) = E^{2n,2n} \quad \Rightarrow \quad C = E^{n,n}. \qquad \square$$

Im Spezialfall $n = 1$ läßt sich Genaueres aussagen:

7.2.5 Satz

$U(1) \cong SO(2)$.

Beweis

Für die im vorherigen Beweis betrachtete Abbildung $\Phi : U(1) \to O(2)$ bleibt zu zeigen:

$$\Phi(U(1)) \subset SO(2), \quad \Phi : U(1) \to SO(2) \text{ ist surjektiv.}$$

Sei $c \in U(1)$, es ist also insbesondere $|c|^2 = |c \cdot \bar{c}| = |\bar{c}| = |c|$ und damit $|c| = 1$. Mit $c = a + ib$ gilt $a^2 + b^2 = 1$ und $\Phi(c) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$, also $\Phi(c) \in SO(2)$.

Um noch die Surjektivität von Φ zu zeigen, sei $A \in SO(2)$ gegeben. Gemäß Satz 6.1.8 existieren dann $a, b \in \mathbb{R}$ mit $a^2 + b^2 = 1$ und $A = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$. Für $c := a + ib$ gilt

$|c| = 1$ und damit für alle $d \in \mathbb{C}$: $|c \cdot d| = |c| \cdot |d| = |d|$, d. h. $c \in U(1)$. Ferner ist $\Phi(c) = A$ und folglich $\mathrm{im}(\Phi) = SO(2)$. □

Eine Drehung des $\mathbb{R}^2$ ist also die Multiplikation mit einer komplexen Zahl c, deren Betrag $|c| = 1$ ist.

Das nächste Ziel ist die komplexe Entsprechung des Satzes 6.3.2, es sollen also einige äquivalente und für das Folgende grundlegende Charakterisierungen unitärer Matrizen zusammengestellt werden. Dazu sind vorbereitend einige aus Kapitel 6 bekannte Begriffe auf den $\mathbb{C}^n$ auszudehnen und einige Hilfsresultate zu beweisen.

Komplexe Orthogonalität

7.2.6 Definition

Die Vektoren $\vec{z}, \vec{w} \in \mathbb{C}^n$ heißen **(komplex) orthogonal**, *wenn $(\vec{z}, \vec{w}) = 0$ gilt. Unter dem* **orthogonalen Komplement** *des Untervektorraums $V \subset \mathbb{C}^n$ verstehen wir*

$$V^\perp := \{\vec{z} \in \mathbb{C}^n : \ (\vec{z}, \vec{w}) = 0 \text{ für alle } \vec{w} \in V\}\,.$$

7.2.7 Satz

Die Vektoren $\vec{z}, \vec{w} \in \mathbb{C}^n$ seien (komplex) orthogonal, $\vec{z} = \vec{x} + i\vec{y}$, $\vec{w} = \vec{u} + i\vec{v}$ mit $\vec{x}$, $\vec{y}$, $\vec{u}$, $\vec{v} \in \mathbb{R}^n$. Dann sind die entsprechenden Vektoren $(x_1, y_1, \ldots, x_n, y_n)$, $(u_1, v_1, \ldots, u_n, v_n) \in \mathbb{R}^{2n}$ (reell) orthogonal.

Beweis

Aus der komplexen Orthogonalität folgt

$$\begin{aligned} 0 = (\vec{z}, \vec{w}) &= \sum_{\nu=1}^{n} (x_\nu - i\, y_\nu)(u_\nu + i\, v_\nu) \\ &= \sum_{\nu=1}^{n} (x_\nu u_\nu + y_\nu v_\nu) + i \cdot \sum_{\nu=1}^{n} (x_\nu v_\nu - y_\nu u_\nu) \end{aligned}$$

und damit durch Betrachtung des Realteils die Behauptung. □

Der Beweis zeigt, daß komplexe Orthogonalität in $\mathbb{C}^n$ eine wesentlich stärkere Eigenschaft als reelle Orthogonalität in $\mathbb{R}^{2n}$ ist, denn das entsprechende reelle Skalarprodukt ist lediglich der Realteil des komplexen Skalarprodukts. Dazu geben wir folgendes einfache Beispiel:

7.2.8 Beispiel

Die komplexen Zahlen 1 und i sind in $\mathbb{C}$ natürlich nicht orthogonal, die entsprechenden reellen Vektoren $(1, 0)$ und $(0, 1)$ in $\mathbb{R}^2$ hingegen schon.

7.2.9 Satz

Für jeden Untervektorraum $V \subset \mathbb{C}^n$ gilt

$$V^{\perp\perp} := \left(V^{\perp}\right)^{\perp} = V.$$

Beweis

Offensichtlich ist $\left(V^{\perp}\right)^{\perp} = \{\vec{z} \in \mathbb{C}^n : (\vec{z}, \vec{w}) = 0 \text{ für alle } \vec{w} \in V^{\perp}\} \supset V$.

Die umgekehrte Inklusion beweist man mit Hilfe eines Dimensionsarguments. Sei $k = \dim_{\mathbb{C}} V$, wie in Satz 6.3.6 zeigt man $\dim_{\mathbb{C}} V^{\perp} = n - k$. Eine Wiederholung dieses Schlusses ergibt $\dim_{\mathbb{C}} \left(V^{\perp}\right)^{\perp} = n - (n-k) = k$ und damit wegen $V \subset V^{\perp\perp}$ die behauptete Gleichheit. □

In unendlichdimensionalen (unitären) Vektorräumen kann das Dimensionsargument nicht verwendet werden. Tatsächlich behält der obige Satz dort nur unter zusätzlichen Voraussetzungen seine Gültigkeit.

7.2.10 Definition

Man nennt ein k-tupel $(\vec{z}_1, \ldots, \vec{z}_k)$ von Vektoren aus $\mathbb{C}^n$ ein **(komplexes) orthogonales k-Bein**, *wenn für $\mu, \nu = 1, \ldots, k$ gilt:*

$$(\vec{z}_\mu, \vec{z}_\nu) = \delta_{\mu\nu}.$$

Ist ein orthogonales k-Bein gleichzeitig auch Erzeugendensystem eines Untervektorraums V von $\mathbb{C}^n$, so spricht man von einer **Orthonormalbasis** *von V.*

Eine Orthonormalbasis von V ist also eine Basis aus paarweise orthogonalen, normierten (d. h. mit Länge 1) Vektoren.

Wie im Reellen stehen nun das Schmidtsche Orthonormalisierungsverfahren, Projektionen auf orthogonale k-Beine etc. zur Verfügung.

Wir bezeichnen mit $\bar{C}^t$ die zur $m \times n$-Matrix C konjugiert transponierte $n \times m$-Matrix, die also aus $C = (c_{\mu\nu})_{\substack{\mu=1,\ldots,m \\ \nu=1,\ldots,n}}$ durch Transponieren und komponentenweises Konjugieren hervorgeht:

$$\bar{C}^t = \begin{pmatrix} \bar{c}_{11} & \bar{c}_{21} & \ldots & \bar{c}_{m1} \\ \bar{c}_{12} & \bar{c}_{22} & \ldots & \bar{c}_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{c}_{1n} & \bar{c}_{2n} & \ldots & \bar{c}_{mn} \end{pmatrix}.$$

Wir betrachten nun wieder ausschließlich quadratische Matrizen.

7.2.11 Satz

Für jede $n \times n$-Matrix C und alle Vektoren $\vec{z}, \vec{w} \in \mathbb{C}^n$ gilt:

$$(\vec{w}, C \circ \vec{z}) = (\bar{C}^t \circ \vec{w}, \vec{z}).$$

Beweis

Das komplexe Skalarprodukt lautet in Matrizenschreibweise:

$$(\vec{z}, \vec{w}) = \bar{\vec{z}}^t \circ \vec{w},$$

daraus erhält man:

$$(\bar{C}^t \circ \vec{w}, \vec{z}\,) = \overline{(\bar{C}^t \circ \vec{w})}^{\,t} \circ \vec{z} = (\bar{\bar{C}}^t \circ \bar{\vec{w}})^t \circ \vec{z} = \bar{\vec{w}}^{\,t} \circ (C \circ \vec{z}\,) = (\vec{w}, C \circ \vec{z}\,). \quad \square$$

7.2.12 Satz

Ist D eine komplexe $n \times n$-Matrix derart, daß $\bar{\vec{z}}^t \circ D \circ \vec{z} = 0$ für alle $\vec{z} \in \mathbb{C}^n$ gilt, dann folgt: $D = \mathcal{O}^{n,n}$.

Beweis

Wir zeigen zunächst, daß die Sesquilinearform, die je zwei Vektoren $\vec{z}, \vec{w} \in \mathbb{C}^n$ die komplexe Zahl $(\vec{z}, D \circ \vec{w}) = \bar{\vec{z}}^t \circ D \circ \vec{w}$ zuordnet, bereits durch die Werte der quadratischen Form $\vec{z} \mapsto \bar{\vec{z}}^t \circ D \circ \vec{z}$ bestimmt ist. Bei dem folgenden Verfahren der *Polarisierung* setzt man geeignete Kombinationen der beliebig vorgegebenen Vektoren $\vec{z}, \vec{w} \in \mathbb{C}^n$ ein:

$$\begin{aligned}
\overline{(\vec{z}+\vec{w})}^{\,t} \circ D \circ (\vec{z}+\vec{w}) &= \bar{\vec{z}}^t \circ D \circ \vec{z} + \bar{\vec{z}}^t \circ D \circ \vec{w} + \bar{\vec{w}}^{\,t} \circ D \circ \vec{z} + \bar{\vec{w}}^{\,t} \circ D \circ \vec{w}, \\
\overline{(\vec{z}-\vec{w})}^{\,t} \circ D \circ (\vec{z}-\vec{w}) &= \bar{\vec{z}}^t \circ D \circ \vec{z} - \bar{\vec{z}}^t \circ D \circ \vec{w} - \bar{\vec{w}}^{\,t} \circ D \circ \vec{z} + \bar{\vec{w}}^{\,t} \circ D \circ \vec{w}, \\
\overline{(\vec{z}+i\vec{w})}^t \circ D \circ (\vec{z}+i\vec{w}) &= \bar{\vec{z}}^t \circ D \circ \vec{z} + i\bar{\vec{z}}^t \circ D \circ \vec{w} - i\bar{\vec{w}}^{\,t} \circ D \circ \vec{z} + \bar{\vec{w}}^{\,t} \circ D \circ \vec{w}, \\
\overline{(\vec{z}-i\vec{w})}^t \circ D \circ (\vec{z}-i\vec{w}) &= \bar{\vec{z}}^t \circ D \circ \vec{z} - i\bar{\vec{z}}^t \circ D \circ \vec{w} + i\bar{\vec{w}}^{\,t} \circ D \circ \vec{z} + \bar{\vec{w}}^{\,t} \circ D \circ \vec{w}.
\end{aligned}$$

Indem man jeweils die ersten und letzten beiden Gleichungen voneinander subtrahiert, erhält man:

$$\begin{aligned}
&\bar{\vec{z}}^t \circ D \circ \vec{w} + \bar{\vec{w}}^{\,t} \circ D \circ \vec{z} \\
&= \frac{1}{2}\left(\overline{(\vec{z}+\vec{w})}^{\,t} \circ D \circ (\vec{z}+\vec{w}) - \overline{(\vec{z}-\vec{w})}^{\,t} \circ D \circ (\vec{z}-\vec{w})\right); \\
&\bar{\vec{z}}^t \circ D \circ \vec{w} - \bar{\vec{w}}^{\,t} \circ D \circ \vec{z} \\
&= \frac{1}{2i}\left(\overline{(\vec{z}+i\vec{w})}^{\,t} \circ D \circ (\vec{z}+i\vec{w}) - \overline{(\vec{z}-i\vec{w})}^{\,t} \circ D \circ (\vec{z}-i\vec{w})\right).
\end{aligned}$$

Diese Gleichungen werden zueinander addiert, und es folgt:

$$\begin{aligned}
\bar{\vec{z}}^t \circ D \circ \vec{w} = &\frac{1}{4}\left(\overline{(\vec{z}+\vec{w})}^{\,t} \circ D \circ (\vec{z}+\vec{w}) - \overline{(\vec{z}-\vec{w})}^{\,t} \circ D \circ (\vec{z}-\vec{w})\right) \\
&+\frac{1}{4i}\left(\overline{(\vec{z}+i\vec{w})}^{\,t} \circ D \circ (\vec{z}+i\vec{w}) - \overline{(\vec{z}-i\vec{w})}^{\,t} \circ D \circ (\vec{z}-i\vec{w})\right).
\end{aligned}$$

Wir setzen nun in dieses auch an sich interessante Zwischenergebnis unsere Voraussetzung ein, daß stets $\bar{\vec{z}}^t \circ D \circ \vec{z} = 0$ ist. Dadurch erhalten wir für alle $\vec{z}, \vec{w} \in \mathbb{C}^n$:

$$\bar{\vec{z}}^t \circ D \circ \vec{w} = 0.$$

Indem wir für $\vec{z}$ und $\vec{w}$ die Elemente $\vec{e}_\mu$ und $\vec{e}_\nu$ der kanonischen Basis des $\mathbb{C}^n$ einsetzen, folgt schließlich für $D = (d_{\mu\nu})_{\mu,\nu=1,\ldots,n}$:

$$d_{\mu\nu} = \vec{e}_\mu^{\,t} \circ D \circ \vec{e}_\nu = 0. \qquad \square$$

7.2.13 Bemerkung

Die komplexe $n \times n$-Matrix D und damit die entsprechende Sesquilinearform $S : \mathbb{C}^n \oplus \mathbb{C}^n \to \mathbb{C},\ S(\vec{z}, \vec{w}) = \bar{\vec{z}}^t \circ D \circ \vec{w} = (\vec{z}, D \circ \vec{w})$ sind durch die quadratische Form $\vec{z} \mapsto \bar{\vec{z}}^t \circ D \circ \vec{z}$, d. h. durch die Werte der Sesquilinearform auf der „Diagonalen" $\vec{z} = \vec{w}$ eindeutig bestimmt.

Ein entsprechendes Resultat gilt im Reellen nur unter Zusatzvoraussetzungen, wie etwa der Symmetrie (man vgl. den Beweis von Satz 6.3.2), wie das Beispiel $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ zeigt.

Unitäre Matrizen

7.2.14 Satz

Für alle komplexen $n \times n$-Matrizen $A = (\vec{a}_1, \ldots, \vec{a}_n)$ sind äquivalent:

(a) $\bar{A}^t \circ A = E$;

(b) für alle $\vec{z}, \vec{w} \in \mathbb{C}^n$ gilt $(A \circ \vec{z}, A \circ \vec{w}) = (\vec{z}, \vec{w})$;

(c) A ist unitär;

(d) $(\vec{a}_1, \ldots, \vec{a}_n)$ ist ein komplexes orthogonales n-Bein.

Beweis

„(a) $\Rightarrow$ (c)": Für alle $\vec{z} \in \mathbb{C}^n$ gilt in Matrizenschreibweise:

$$\begin{aligned} \|\, \vec{z}\,\|^2 &= \bar{\vec{z}}^t \circ \vec{z} = \bar{\vec{z}}^t \circ E \circ \vec{z} = (\bar{\vec{z}}^t \circ \bar{A}^t) \circ (A \circ \vec{z}\,) \\ &= \overline{(A \circ \vec{z}\,)}^{\,t} \circ (A \circ \vec{z}\,) = \|\, A \circ \vec{z}\,\|^2. \end{aligned}$$

„(c) $\Rightarrow$ (b)": In diesem Beweisteil können wir von den in Satz 7.2.12 geleisteten Vorarbeiten profitieren. Wir wollen zeigen, daß die Sesquilinearform

$$S : \mathbb{C}^n \oplus \mathbb{C}^n \to \mathbb{C},\ S(\vec{z}, \vec{w}) = (A \circ \vec{z}, A \circ \vec{w}) - (\vec{z}, \vec{w}) \qquad (*)$$

stets den Wert 0 ergibt. Gemäß Voraussetzung (c) ist der Wert dieser Form auf der „Diagonalen" $\vec{z} = \vec{w}$ stets 0. Indem man noch beachtet, daß für alle $\vec{z}, \vec{w} \in \mathbb{C}^n$

$$\begin{aligned} (A \circ \vec{z}, A \circ \vec{w}) - (\vec{z}, \vec{w}) &= \bar{\vec{z}}^t \circ \big(\bar{A}^t \circ A\big) \circ \vec{w} - \bar{\vec{z}}^t \circ \vec{w} \\ &= \bar{\vec{z}}^t \circ \big(\bar{A}^t \circ A - E\big) \circ \vec{w} \end{aligned}$$

gilt, daß also die Sesquilinearform S in $(*)$ durch die Matrix $\bar{A}^t \circ A - E$ gegeben wird, ergibt Satz 7.2.12 tatsächlich, daß es sich bei dieser Matrix um die Nullmatrix handelt. Mithin verschwindet die Sesquilinearform S in $(*)$ identisch, und wir haben gleichzeitig auch „(c) $\Rightarrow$ (a)" mitbewiesen.

„(b) $\Rightarrow$ (d)": Gemäß Voraussetzung (b) ist die Matrix A „längen-" und „winkeltreu", das gilt insbesondere auch für die Elemente der kanonischen Basis des $\mathbb{C}^n$:

$$\delta_{\mu\nu} = (\vec{e}_\mu, \vec{e}_\nu) = (A \circ \vec{e}_\mu, A \circ \vec{e}_\nu) = (\vec{a}_\mu, \vec{a}_\nu).$$

„(d) $\Rightarrow$ (a)": Dieser Schluß basiert allein auf der Definition der Matrizenmultiplikation:

$$\begin{aligned} \bar{A}^t \circ A &= \begin{pmatrix} \bar{\vec{a}}_1{}^t \\ \vdots \\ \bar{\vec{a}}_n{}^t \end{pmatrix} \circ (\vec{a}_1, \ldots, \vec{a}_n) = \left(\bar{\vec{a}}_\mu{}^t \circ \vec{a}_\nu\right)_{\mu,\nu=1,\ldots,n} \\ &= \Big((\vec{a}_\mu, \vec{a}_\nu)\Big)_{\mu,\nu=1,\ldots,n} = (\delta_{\mu\nu})_{\mu,\nu=1,\ldots,n} = E. \end{aligned}$$

□

Adjungierte Abbildungen

Wiederholt hat die zur komplexen $n \times n$-Matrix A konjugiert transponierte Matrix $\bar{A}^t$ eine Rolle gespielt, nicht zuletzt als Umkehrmatrix unitärer Matrizen. Deshalb werden wir nun die entsprechenden Abbildungen eingehender studieren.

7.2.15 Definition

Sei A eine komplexe $n \times n$-Matrix und $F : \mathbb{C}^n \to \mathbb{C}^n$ die zugehörige $\mathbb{C}$-lineare Abbildung.

(a) *Die zur konjugiert transponierten Matrix $\bar{A}^t$ gehörige $\mathbb{C}$-lineare Abbildung $\mathbb{C}^n \to \mathbb{C}^n$ heißt die zu F* **adjungierte Abbildung**.

(b) *Gilt $\bar{A}^t = A$, so heißt A* **Hermitesch** *und F* **selbstadjungiert**.

Ein Spezialfall hiervon ist die entsprechende reelle Definition:

7.2.16 Definition

Sei A eine reelle $n \times n$-Matrix und $F : \mathbb{R}^n \to \mathbb{R}^n$ die zugehörige $\mathbb{R}$-lineare Abbildung.

(a) *Die zur transponierten Matrix A^t gehörige $\mathbb{R}$-lineare Abbildung $\mathbb{R}^n \to \mathbb{R}^n$ heißt die zu F* **adjungierte Abbildung**.

(b) *Ist A symmetrisch, d. h. $A = A^t$, so nennen wir F* **selbstadjungiert**.

7.2.17 Bemerkung

Bei reellen Matrizen $A \in \mathbb{R}^{n\times n}$ und zugehörigen linearen Abbildungen $F : \mathbb{R}^n \to \mathbb{R}^n$ fallen die adjungierte und die duale Abbildung zusammen. Im Komplexen dagegen wird die duale Abbildung F^* unverändert durch A^t, die adjungierte Abbildung $F^{\text{adjungiert}}$ dagegen durch $\bar{A}^t$ gegeben. Hier besteht bei den Abbildungen der Zusammenhang: $F^{\text{adjungiert}}(\vec{z}\,) = \overline{F^*\left(\bar{\vec{z}}\right)}$.

Um im Folgenden nicht unnötig komplizierte Bezeichnungen verwenden zu müssen, werden wir mitunter die zur Matrix A gehörige lineare Abbildung ebenfalls mit A bezeichnen und beispielsweise von $\bar{A}^t$ als der zu A adjungierten Abbildung sprechen.

7.2.18 Satz

Eine komplexe $n \times n$-Matrix A ist genau dann Hermitesch, wenn für alle $\vec{z} \in \mathbb{C}^n$ die Zahl $(\vec{z}, A \circ \vec{z}\,)$ reell ist.

Beweis

Grundlegend ist die folgende, auf den Eigenschaften konjugiert transponierter Matrizen beruhende Gleichheit:

$$(\vec{z}, \bar{A}^t \circ \vec{z}\,) = (A \circ \vec{z}, \vec{z}\,) = \overline{(\vec{z}, A \circ \vec{z}\,)}.$$

Ist A Hermitesch, so gilt für alle $\vec{z} \in \mathbb{C}^n$ also $(\vec{z}, A \circ \vec{z}\,) = \overline{(\vec{z}, A \circ \vec{z}\,)}$, demgemäß sind diese Zahlen stets reell.

Ist dagegen $(\vec{z}, A \circ \vec{z}\,)$ stets reell, d. h. $(\vec{z}, A \circ \vec{z}\,) = \overline{(\vec{z}, A \circ \vec{z}\,)}$, so verschwindet die von $\bar{A}^t - A$ gegebene quadratische Form identisch, es ist also stets $(\vec{z}, (\bar{A}^t - A) \circ \vec{z}\,) = 0$. Der schon wiederholt zitierte Satz 7.2.12 zeigt $\bar{A}^t - A = \mathcal{O}$ und damit die Hermitizität von A. □

Auch um die Symmetrie einer *reellen* $n \times n$-Matrix zu prüfen, ist die quadratische Form $(\vec{z}, A \circ \vec{z}\,)$ für alle *komplexen* Vektoren $\vec{z} \in \mathbb{C}^n$ zu betrachten.

7.2.19 Satz

Für die zur komplexen $n \times n$-Matrix A gehörige lineare Abbildung gilt:

(a) $\ker(A) = \operatorname{im}(\bar{A}^t)^\perp$,

(b) $\operatorname{im}(A) = \ker(\bar{A}^t)^\perp$.

Beweis

Sei $\vec{z} \in \ker(A)$, d. h. $A \circ \vec{z} = \vec{0}$. Für alle $\bar{A}^t \circ \vec{w} \in \operatorname{im}(\bar{A}^t)$, $\vec{w} \in \mathbb{C}^n$, folgt mit Satz 7.2.11

$$(\vec{z}, \bar{A}^t \circ \vec{w}) = (A \circ \vec{z}, \vec{w}) = (\vec{0}, \vec{w}) = 0.$$

Somit ist $\ker(A) \subset \operatorname{im}(\bar{A}^t)^\perp$.

Sei nun umgekehrt $\vec{z} \in \mathrm{im}(\bar{A}^t)^\perp$, für alle $\vec{w} \in \mathbb{C}^n$ haben wir also

$$0 = (\vec{z}, \bar{A}^t \circ \vec{w}) = (A \circ \vec{z}, \vec{w}).$$

Indem man speziell $\vec{w} = A \circ \vec{z}$ wählt, ergibt sich $A \circ \vec{z} = \vec{0}$, $\vec{z} \in \ker(A)$, $\mathrm{im}(\bar{A}^t)^\perp \subset \ker(A)$ und damit insgesamt Behauptung (a).

Indem man in (a) zu den orthogonalen Komplementen übergeht, die Matrix A durch $\bar{A}^t$ ersetzt und $\overline{\bar{A}^t}^t = A$ beachtet, erhält man

$$\ker(\bar{A}^t)^\perp = \mathrm{im}(A)^{\perp\perp}.$$

Vorher hatten wir schon für beliebige Untervektorräume $V \subset \mathbb{C}^n$ die Gültigkeit von $V^{\perp\perp} = V$ gezeigt, und folglich ist auch Behauptung (b) bewiesen. □

Bei selbstadjungierten Abbildungen können wir mit Hilfe dieses Satzes durch Abspaltung des Kerns zu Isomorphismen übergehen.

7.2.20 Satz

Die lineare Abbildung $F : \mathbb{C}^n \to \mathbb{C}^n$ sei selbstadjungiert, es bezeichne $V = \ker(F)^\perp$. Dann ist die Einschränkung von F auf V

$$F|V : V \to V$$

ein Isomorphismus.

Beweis

Gemäß dem vorhergehenden Satz ist $\mathrm{im}(F|V) \subset \mathrm{im}(F) = \ker(F)^\perp = V$. Die Abbildung $F|V$ ist offensichtlich injektiv und damit auf Grund der Dimensionsformel auch surjektiv. □

Normale Abbildungen

Unitäre und selbstadjungierte Abbildungen scheinen zunächst wenig miteinander zu tun zu haben: Unitäre Abbildungen sind längentreu, mithin stets Isomorphismen. Selbstadjungierte Abbildungen können durchaus einen nichttrivialen Kern haben und die eingesetzten Vektoren strecken oder stauchen. Dennoch teilen diese Abbildungen eine Eigenschaft, die sich in Abschnitt 7.4 als entscheidend für die Diagonalisierbarkeit erweisen wird.

7.2.21 Definition

Wir nennen eine $n \times n$-Matrix A und die zugehörige lineare Abbildung $F : \mathbb{C}^n \to \mathbb{C}^n$ **normal**, *wenn die Abbildung und ihre adjungierte kommutieren, falls also die Bedingung*

$$\bar{A}^t \circ A = A \circ \bar{A}^t$$

erfüllt ist.

7.2.22 Satz

Eine $n \times n$-Matrix A ist normal genau dann, wenn für alle $\vec{z}, \vec{w} \in \mathbb{C}^n$ gilt:

$$(\bar{A}^t \circ \vec{z}, \bar{A}^t \circ \vec{w}) = (A \circ \vec{z}, A \circ \vec{w}). \qquad (*)$$

Beweis

Ist A normal, so folgt für alle $\vec{z}, \vec{w} \in \mathbb{C}^n$

$$\begin{aligned}(A \circ \vec{z}, A \circ \vec{w}) &= (\vec{z}, \bar{A}^t \circ A \circ \vec{w}) = (\vec{z}, A \circ \bar{A}^t \circ \vec{w}) \\ &= (\bar{A}^t \circ \vec{z}, \bar{A}^t \circ \vec{w}).\end{aligned}$$

Gilt nun umgekehrt $(*)$ für alle $\vec{z}, \vec{w} \in \mathbb{C}^n$, so folgt

$$\begin{aligned}\left(\vec{z}, \left(\bar{A}^t \circ A\right) \circ \vec{w}\right) &= (A \circ \vec{z}, A \circ \vec{w}) = \left(\bar{A}^t \circ \vec{z}, \bar{A}^t \circ \vec{w}\right) \\ &= \left(\vec{z}, \left(A \circ \bar{A}^t\right) \circ \vec{w}\right), \\ 0 &= \left(\vec{z}, \left(\bar{A}^t \circ A - A \circ \bar{A}^t\right) \circ \vec{w}\right).\end{aligned}$$

Indem man bei beliebigem $\vec{w}$ für $\vec{z} = \left(\bar{A}^t \circ A - A \circ \bar{A}^t\right) \circ \vec{w}$ einsetzt, erhält man $\left(\bar{A}^t \circ A - A \circ \bar{A}^t\right) \circ \vec{w} = \vec{0}$ und damit

$$\bar{A}^t \circ A - A \circ \bar{A}^t = \mathcal{O};$$

die Matrix A ist normal. □

7.3 Eigenwerte

Wir haben stets die Theorie linearer Abbildungen und den Matrizenkalkül simultan und gleichberechtigt entwickelt und haben häufig mit Gewinn zwischen diesen beiden Standpunkten gewechselt. In diesem Abschnitt stellen wir den „Abbildungsstandpunkt“ in den Vordergrund.

Es bezeichne V stets einen n-dimensionalen komplexen Vektorraum und $F : V \to V$ eine $\mathbb{C}$-lineare Abbildung.

Eigenräume

7.3.1 Definition

Ein Vektor $z \in V \setminus \{0\}$ heißt **Eigenvektor** *von F zum* **Eigenwert** *$\lambda \in \mathbb{C}$, wenn die Eigenwertgleichung*

$$F(z) = \lambda \cdot z$$

erfüllt ist.

Da gleichzeitig Eigenwert λ und Eigenvektor z gesucht werden und die Abbildung $\mathbb{C} \times V \to V$, $(\lambda, z) \mapsto \lambda \cdot z$ nicht linear ist, handelt es sich bei der Eigenwertbestimmung um ein nichtlineares Problem.

Diese Begriffe stehen ebenso wie die im Folgenden noch zu definierenden auch für beliebige $n \times n$-Matrizen zur Verfügung, indem man die entsprechende $\mathbb{C}$-lineare Abbildung betrachtet. So ist $\vec{z} \in \mathbb{C}^n \setminus \{\vec{0}\}$ Eigenvektor von A zum Eigenwert $\lambda \in \mathbb{C}$, falls $A \circ \vec{z} = \lambda \cdot \vec{z}$ gilt. Das Bestehen dieser Eigenwertgleichung ist äquivalent zu $(A - \lambda E) \circ \vec{z} = \vec{0}$, bzw. auf Abbildungsebene zu $(F - \lambda \cdot id)(z) = 0$. Also gilt

$$\begin{aligned}&\{z \in V:\ z \text{ ist Eigenvektor von } F: V \to V \text{ zu } \ \lambda \in \mathbb{C}\} \cup \{0\}\\ &= \{z:\ (F - \lambda \cdot id)(z) = 0\} = \ker(F - \lambda \cdot id).\end{aligned}$$

7.3.2 Definition

Für $\lambda \in \mathbb{C}$ heißt die Menge aller Eigenvektoren zu λ, ergänzt um den Nullvektor,

$$V_\lambda := V_\lambda(F) := \ker(F - \lambda \cdot id)$$

Eigenraum *von F zum Eigenwert $\lambda \in \mathbb{C}$.*

7.3.3 Bemerkung

Als Kern einer linearen Abbildung ist V_λ ein $\mathbb{C}$-Untervektorraum von V. Genau dann ist λ Eigenwert von F, wenn $\dim V_\lambda \neq 0$ gilt. Ist z Eigenvektor zu λ von F, so auch jedes komplexe Vielfache $c \cdot z$, $c \in \mathbb{C}$.

7.3.4 Satz

Es seien $z_1, \ldots, z_k$ Eigenvektoren von F zu den paarweise verschiedenen Eigenwerten $\lambda_1, \ldots, \lambda_k \in \mathbb{C}$:

$$F(z_\nu) = \lambda_\nu z_\nu, \quad z_\nu \neq 0, \quad \lambda_\mu \neq \lambda_\nu \text{ für } \mu \neq \nu.$$

Dann sind $z_1, \ldots, z_k$ linear unabhängig.

Insbesondere hat F höchstens n paarweise verschiedene Eigenwerte.

Beweis

Wir führen den Beweis durch Induktion nach der Anzahl k der Eigenvektoren. Der Fall $k = 1$ ist offensichtlich, denn Eigenvektoren sind definitionsgemäß vom Nullvektor verschieden. Sei nun die Behauptung für $k - 1$ gezeigt, und es seien $z_1, \ldots, z_k$ Eigenvektoren zu den paarweise verschiedenen Eigenwerten $\lambda_1, \ldots, \lambda_k$. Wir betrachten Linearkombinationen des Nullvektors aus $z_1, \ldots, z_k$:

$$\sum_{\nu=1}^{k} a_\nu z_\nu = 0, \qquad a_\nu \in \mathbb{C}.$$

Man erhält durch Multiplikation mit λ_k:

$$\sum_{\nu=1}^{k} a_\nu \lambda_k z_\nu = 0$$

und durch die Anwendung von F:

$$0 = F(0) = F\left(\sum_{\nu=1}^{k} a_\nu z_\nu\right) = \sum_{\nu=1}^{k} a_\nu F(z_\nu) = \sum_{\nu=1}^{k} a_\nu \lambda_\nu z_\nu.$$

Durch Subtraktion der beiden Gleichungen folgt:

$$\sum_{\nu=1}^{k} a_\nu(\lambda_k - \lambda_\nu) z_\nu = \sum_{\nu=1}^{k-1} a_\nu(\lambda_k - \lambda_\nu) z_\nu = 0.$$

Nach Induktionsvoraussetzung sind $z_1, \dots, z_{k-1}$ linear unabhängig und somit $a_\nu(\lambda_k - \lambda_\nu) = 0$. Da die Eigenwerte als paarweise verschieden vorausgesetzt sind, erhalten wir $a_1 = \ldots = a_{k-1} = 0$. Aus der ursprünglichen Linearkombination des Nullvektors wird $0 = a_k z_k$, hieraus folgt schließlich auch $a_k = 0$. □

Das charakteristische Polynom

Die Bestimmung von Eigenwerten soll auf die Bestimmung der Nullstellen eines komplexen Polynoms reduziert werden. Wir fassen hier komplexe Polynome im Sinne der analytischen Definition als Funktionen $\mathbb{C} \to \mathbb{C}$ auf.

Sei $G : \mathbb{C}^n \to V$ eine Karte, d. h. ein $\mathbb{C}$-Vektorraumisomorphismus. Die Abbildung F in den Koordinaten, $G^{-1} \circ F \circ G$, wird durch eine $n \times n$-Matrix A gegeben. Wegen der Linearität des Koordinatensystems G^{-1} lautet die Abbildung $F - \lambda \cdot id$ in den Koordinaten: $G^{-1} \circ (F - \lambda \cdot id) \circ G = G^{-1} \circ (F \circ G - \lambda \cdot G) = G^{-1} \circ F \circ G - \lambda \cdot G^{-1} \circ G = G^{-1} \circ F \circ G - \lambda \cdot id_{\mathbb{C}^n}$, sie wird also durch die Matrix $A - \lambda \cdot E$ beschrieben. Da die komplexe Zahl λ Eigenwert von A bzw. F genau dann ist, wenn $A - \lambda E$ singulär (nicht invertierbar) ist, liegt es nahe, die entsprechende Determinante zu betrachten:

$$\begin{aligned} \det(A - \lambda E) &= \sum_{\nu_1, \dots, \nu_n = 1}^{n} \delta(\nu_1, \dots, \nu_n)(a_{\nu_1 1} - \lambda \delta_{\nu_1 1}) \cdot \ldots \cdot (a_{\nu_n n} - \lambda \delta_{\nu_n n}) \\ &= (-1)^n \cdot (\lambda^n + Q(\lambda)), \end{aligned}$$

dabei ist $Q(\lambda)$ ein Polynom in λ vom Grade $\leq n - 1$. Es handelt sich bei $\det(A - \lambda E)$ also um ein Polynom n-ten Grades in λ mit höchtem Koeffizienten $(-1)^n$. Auf Grund der Überlegungen aus Kapitel 5.6 ist $\det(A - \lambda E) = \det(F - \lambda \cdot id)$ unabhängig von der Wahl der Karte G.

7.3.5 Definition

Wir nennen $P(\lambda) := P_F(\lambda) := (-1)^n \cdot \det(F - \lambda \cdot id)$ das **charakteristische Polynom** *der linearen Abbildung F.*

Das charakteristische Polynom ist normiert, besitzt also höchsten Koeffizienten 1.

7.3.6 Satz

Sei $P(\lambda)$ das charakteristische Polynom einer $\mathbb{C}$-linearen Abbildung $F : V \to V$. Dann ist $\lambda \in \mathbb{C}$ Eigenwert von F genau dann, wenn $P(\lambda) = 0$ gilt.

Beweis

Die komplexe Zahl $\lambda \in \mathbb{C}$ ist Eigenwert von F genau dann, wenn die Abbildung $F - \lambda \cdot id$ nicht injektiv und damit nicht invertierbar ist. Letzteres ist jedoch äquivalent zum Verschwinden der Determinante: $P(\lambda) = \det(F - \lambda\, id) = 0$. □

Das charakteristische Polynom ist invariant gegenüber $\mathbb{C}$-Isomorphismen:

7.3.7 Satz

Sei $P(\lambda)$ das charakteristische Polynom einer $\mathbb{C}$-linearen Abbildung $F : V \to V$. Sei W ein weiterer n-dimensionaler $\mathbb{C}$-Vektorraum und $H : W \xrightarrow{\sim} V$ ein $\mathbb{C}$-Vektorraumisomorphismus. Für das charakteristische Polynom $\hat{P}(\lambda)$ der Abbildung $\hat{F} := H^{-1} \circ F \circ H$ gilt $\hat{P}(\lambda) = P(\lambda)$.

Beweis

Sei $G : \mathbb{C}^n \to V$ eine Karte. Dann ist $H^{-1} \circ G$ eine Karte von W und $\hat{F}$ lautet in diesen Koordinaten

$$(H^{-1} \circ G)^{-1} \circ \hat{F} \circ (H^{-1} \circ G) = G^{-1} \circ H \circ H^{-1} \circ F \circ H \circ H^{-1} \circ G = G^{-1} \circ F \circ G,$$

das ist auch die Abbildung F in den Koordinaten G. Das charakteristische Polynom ist aber von der Wahl der Karte unabhängig. □

Als Folgerung erhalten wir die Existenz einer weiteren Matrixinvariante:

7.3.8 Definition

Unter der **Spur** *einer quadratischen Matrix $A = (a_{\mu\nu})_{\mu,\nu=1,\dots,n}$ verstehen wir die Summe der Diagonalelemente:*

$$\mathrm{Sp}(A) := \sum_{\nu=1}^{n} a_{\nu\nu}.$$

7.3.9 Satz

Sei $A \in \mathbb{C}^{n\times n}$ beliebig, $B \in \mathbb{C}^{n\times n}$ invertierbar. Dann haben A und $B^{-1} \circ A \circ B$ dieselbe Spur.

Beweis

Eine genaue Inspektion der Definition des charakteristischen Polynoms ergibt

$$\begin{aligned} P_A(\lambda) &= (-1)^n \det(A - \lambda E) \\ &= (-1)^n \sum_{\nu_1,\ldots,\nu_n=1}^{n} \delta(\nu_1,\ldots,\nu_n)(a_{\nu_1 1} - \lambda\delta_{\nu_1 1}) \cdot \ldots \cdot (a_{\nu_n n} - \lambda\delta_{\nu_n n}) \\ &= \lambda^n - \mathrm{Sp}(A)\lambda^{n-1} + \tilde{Q}(\lambda), \end{aligned}$$

das Polynom $\tilde{Q}(\lambda)$ ist vom Grade $\leq n-2$. Die Koeffizienten dieses Polynoms ändern sich unter dem Koordinatenwechsel B nicht. □

7.3.10 Bemerkung (Existenz reeller Eigenvektoren)

Sei $\lambda \in \mathbb{R}$ ein Eigenwert der reellen $n \times n$-Matrix A. Dann ist $\det(A - \lambda E) = 0$, und es existiert ein $\vec{x} \in \mathbb{R}^n \setminus \{\vec{0}\}$ mit $(A - \lambda E) \circ \vec{x} = \vec{0}$. Also besitzt A einen reellen Eigenvektor.

Der folgende, für unsere Zwecke zentrale Satz verliert in den reellen Zahlen seine Gültigkeit und ist der Grund dafür, daß die Eigenwerttheorie in *komplexen* Vektorräumen durchgeführt wird.

7.3.11 Satz (Fundamentalsatz der Algebra)

Jedes komplexe Polynom zerfällt über $\mathbb{C}$ in Linearfaktoren. Genauer gilt: Sei $P(\lambda)$ ein Polynom vom Grade n über $\mathbb{C}$. Dann gibt es Zahlen $s \in \mathbb{N}$, $s \leq n$, $r_1, \ldots, r_s \in \mathbb{N}$ mit $r_1 + \cdots + r_s = n$, $a \in \mathbb{C} \setminus \{0\}$, $c_1, \ldots, c_s \in \mathbb{C}$ mit $c_\mu \neq c_\nu$ für $\mu \neq \nu$, so daß gilt:

$$P(\lambda) = a \cdot (\lambda - c_1)^{r_1} \cdot \ldots \cdot (\lambda - c_s)^{r_s}.$$

Man sagt, daß $P(\lambda)$ in c_ν eine Nullstelle *r_ν-ter Ordnung* besitzt.

Der Beweis kann mit den hier zur Verfügung stehenden Mitteln nicht geführt werden und wird im Rahmen der Funktionentheorie oder der Algebra gegeben.

Als Folgerung erhalten wir insbesondere ein Existenzresultat für *komplexe* Eigenwerte:

7.3.12 Satz

Jede $\mathbb{C}$-lineare Abbildung $F : V \to V$ besitzt mindestens einen Eigenwert.

7.3.13 Definition

Sei $P_F(\lambda) = (\lambda - \lambda_1)^{r_1} \cdot \ldots \cdot (\lambda - \lambda_s)^{r_s}$ mit $\lambda_\nu \neq \lambda_\mu$ für $\mu \neq \nu$, $r_1, \ldots, r_s \in \mathbb{N}$ und $r_1 + \ldots + r_s = n$ das charakteristische Polynom einer $\mathbb{C}$-linearen Abbildung F. Man nennt die Nullstellenordnung r_ν die **algebraische** *und die Dimension des Eigenraumes $\dim V_{\lambda_\nu}$ die* **geometrische Vielfachheit** *des Eigenwertes λ_ν.*

7.3.14 Beispiele

(a) Sei

$$A := \begin{pmatrix} a_1 & \dots & * \\ & \ddots & \vdots \\ \mathcal{O} & & a_n \end{pmatrix}$$

eine Dreiecksmatrix. Gemäß Satz 5.4.2 gilt:

$$P(\lambda) = (-1)^n \cdot \det(A - \lambda E) = (-1)^n \cdot \prod_{\nu=1}^{n} (a_\nu - \lambda) = \prod_{\nu=1}^{n} (\lambda - a_\nu).$$

Die Diagonalelemente $a_1, \dots, a_n$ sind die Eigenwerte von A.

(b) Sei

$$A := \begin{pmatrix} a_1 & & \mathcal{O} \\ & \ddots & \\ \mathcal{O} & & a_n \end{pmatrix}$$

eine Diagonalmatrix. Es gilt $A \circ \vec{e}_\nu = a_\nu \cdot \vec{e}_\nu$, d. h. $\vec{e}_\nu$ ist Eigenvektor von A zum Eigenwert a_ν. Die kanonische Basis ist eine Orthonormalbasis aus Eigenvektoren von A. Die algebraischen und geometrischen Vielfachheiten der Eigenwerte stimmen überein.

(c) Auch reelle Matrizen haben i. a. nur komplexe Eigenwerte. Dazu berechnen wir die Eigenwerte von $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Das charakteristische Polynom lautet $P(\lambda) = \det \begin{pmatrix} -\lambda & 1 \\ -1 & -\lambda \end{pmatrix} = \lambda^2 + 1$, die Eigenwerte sind i und $-i$.

(d) Nicht für jede Matrix existiert eine Basis aus Eigenvektoren. Man betrachte dazu $A = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$ mit $a \in \mathbb{C} \setminus \{0\}$. Es ist $P(\lambda) = (-1)^2(1-\lambda)^2 = (1-\lambda)^2$, 1 ist also einziger Eigenwert von A. Zur Bestimmung der Eigenvektoren ist folgendes Gleichungssystem zu lösen:

$$\vec{0} = (A - E) \circ \vec{z} = \begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} a \cdot z_2 \\ 0 \end{pmatrix}.$$

Der Eigenraum zum einzigen Eigenwert 1 ist also

$$V_1 = \{(z_1, z_2) \in \mathbb{C}^2 : \; z_2 = 0\} = \{(z_1, 0) : z_1 \in \mathbb{C}\}.$$

Insbesondere existieren nicht zwei linear unabhängige Eigenvektoren von A. Die algebraische Vielfachheit des Eigenwertes 1 ist 2, die geometrische 1.

7.3.15 Folgerung

Die Determinante ist das Produkt der Eigenwerte (entsprechend ihren algebraischen Vielfachheiten). Genauer gilt: Sei $P_F(\lambda) = (\lambda - \lambda_1)^{r_1} \cdot \ldots \cdot (\lambda - \lambda_s)^{r_s}$ mit $\lambda_\nu \neq \lambda_\mu$ für $\mu \neq \nu$, $r_1, \ldots, r_s \in \mathbb{N}$ und $r_1 + \ldots + r_s = n$ das charakteristische Polynom der $\mathbb{C}$-linearen Abbildung F. Dann gilt: $\det(F) = \prod_{\nu=1}^{s} \lambda_\nu^{r_\nu}$.

Beweis

Wir haben

$$P_F(\lambda) = (-1)^n \det(F - \lambda\, id) = \prod_{\nu=1}^{s} (\lambda - \lambda_\nu)^{r_\nu}\,.$$

Durch Einsetzen von $\lambda = 0$ folgt

$$\det(F) = \prod_{\nu=1}^{s} \lambda_\nu^{r_\nu}. \qquad \square$$

Eigenvektoren normaler Matrizen

Am Ende von Abschnitt 7.2 haben wir die normalen Matrizen auf zunächst wenig anschauliche Weise als Oberbegriff für die Hermiteschen und unitären Matrizen eingeführt. Die nächsten Sätze verdeutlichen, wie wichtig dieser neue Begriff tatsächlich ist.

7.3.16 Satz

Die $n \times n$-Matrix A sei normal. Dann ist λ Eigenwert von A genau dann, wenn $\bar\lambda$ Eigenwert der adjungierten Abbildung $\bar A^t$ ist; die Eigenräume stimmen überein:

$$V_\lambda(A) = V_{\bar\lambda}(\bar A^t).$$

Beweis

Zunächst sei bemerkt, daß mit A auch $A - \lambda\, E$ normal ist.

Sei $\vec z \in V_\lambda(A)$, also $A \circ \vec z - \lambda\, \vec z = \vec 0$. Mit Hilfe von Satz 7.2.22 folgt

$$\begin{aligned} 0 &= \|\, A \circ \vec z - \lambda \vec z\,\|^2 = ((A - \lambda E) \circ \vec z, (A - \lambda E) \circ \vec z\,) \\ &= \left(\overline{(A - \lambda E)}^t \circ \vec z, \overline{(A - \lambda E)}^t \circ \vec z\right) \\ &= \left((\bar A^t - \bar\lambda \cdot E) \circ \vec z, (\bar A^t - \bar\lambda \cdot E) \circ \vec z\right) \\ &= \|\, \bar A^t \circ \vec z - \bar\lambda \cdot \vec z\,\|^2; \\ \bar A^t \circ \vec z &= \bar\lambda \cdot \vec z. \end{aligned}$$

Somit ist $V_\lambda(A) \subset V_{\bar\lambda}(\bar A^t)$. Derselbe Schluß, nun angewandt auf $\bar\lambda$ und $\bar A^t$, zeigt $V_{\bar\lambda}(\bar A^t) \subset V_{\bar{\bar\lambda}}\left(\overline{\bar A^t}^t\right) = V_\lambda(A)$ und damit auch $V_\lambda(A) = V_{\bar\lambda}(\bar A^t)$. $\square$

7.3.17 Satz

Je zwei Eigenvektoren $\vec{z}_1, \vec{z}_2$ einer normalen Matrix A zu verschiedenen Eigenwerten $\lambda_1 \neq \lambda_2$ sind orthogonal.

Beweis

Gemäß Voraussetzung ist $A \circ \vec{z}_1 = \lambda_1 \vec{z}_1$, $A \circ \vec{z}_2 = \lambda_2 \vec{z}_2$; der vorhergehende Satz impliziert $\bar{A}^t \circ \vec{z}_1 = \bar{\lambda}_1 \vec{z}_1$, $\bar{A}^t \circ \vec{z}_2 = \bar{\lambda}_2 \vec{z}_2$. Damit folgt:

$$\begin{aligned}\lambda_2(\vec{z}_1, \vec{z}_2) &= (\vec{z}_1, \lambda_2 \vec{z}_2) = (\vec{z}_1, A \circ \vec{z}_2)\\ &= (\bar{A}^t \circ \vec{z}_1, \vec{z}_2) = (\bar{\lambda}_1 \circ \vec{z}_1, \vec{z}_2) = \lambda_1(\vec{z}_1, \vec{z}_2),\\ 0 &= (\lambda_1 - \lambda_2)(\vec{z}_1, \vec{z}_2).\end{aligned}$$

Aus der Verschiedenheit der Eigenwerte erhalten wir die Orthogonalität der Eigenvektoren: $(\vec{z}_1, \vec{z}_2) = 0$. □

Eigenwerte Hermitescher Matrizen

7.3.18 Satz

Alle Eigenwerte Hermitescher Matrizen sind reell.

Beweis

Sei $A = A^{n,n}$ Hermitesch mit Eigenvektor $\vec{z} \in \mathbb{C}^n \setminus \{\vec{0}\}$ zum Eigenwert $\lambda \in \mathbb{C}$. Satz 7.2.18 zeigt:

$$\lambda \| \vec{z}\, \|^2 = \lambda(\vec{z}, \vec{z}\,) = (\vec{z}, \lambda \cdot \vec{z}\,) = (\vec{z}, A \circ \vec{z}\,) \in \mathbb{R}.$$

Wegen $\| \vec{z}\, \|^2 \in \mathbb{R} \setminus \{0\}$ folgt, daß λ reell ist. □

7.3.19 Beispiele

(a) Für die Hermitesche 2×2-Matrix

$$A = \begin{pmatrix} 1 & -3i \\ 3i & 1 \end{pmatrix}$$

lautet das charakteristische Polynom

$$\begin{aligned}P(\lambda) = \det(A - \lambda\, E) = \det \begin{pmatrix} 1-\lambda & -3i \\ 3i & 1-\lambda \end{pmatrix} &= (1-\lambda)^2 + 9i^2\\ = (\lambda - 1)^2 - 9 = (\lambda - 4)(\lambda + 2);\end{aligned}$$

die Eigenwerte sind $\lambda_1 = 4$ und $\lambda_2 = -2$.

(b) Wir betrachten nun die reelle symmetrische 3×3-Matrix

$$A = \begin{pmatrix} \frac{7}{6} & -\frac{2}{3} & \frac{1}{6} \\ -\frac{2}{3} & \frac{2}{3} & -\frac{2}{3} \\ \frac{1}{6} & -\frac{2}{3} & \frac{7}{6} \end{pmatrix}.$$

Das charakteristische Polynom lautet

$$\begin{aligned}
P(\lambda) = (-1)^3 \det(A - \lambda\, E) &= -\frac{1}{216} \det \begin{pmatrix} (7-6\lambda) & -4 & 1 \\ -4 & (4-6\lambda) & -4 \\ 1 & -4 & (7-6\lambda) \end{pmatrix} \\
&= -\frac{1}{216} \det \begin{pmatrix} (7-6\lambda) & -4 & 1 \\ 24(1-\lambda) & -6(2+\lambda) & 0 \\ 12(-3\lambda^2+7\lambda-4) & 24(1-\lambda) & 0 \end{pmatrix} \\
&= -\frac{72}{216} \left\{ 8(1-\lambda)^2 + (2+\lambda)(-3\lambda^2+7\lambda-4) \right\} \\
&= -\frac{1}{3} \left\{ -3\lambda^3 + 9\lambda^2 - 6\lambda \right\} = \lambda(\lambda-1)(\lambda-2),
\end{aligned}$$

wir haben also die drei Eigenwerte $\lambda_1 = 0$, $\lambda_2 = 1$, $\lambda_3 = 2$.

Eigenwerte unitärer Matrizen

7.3.20 Satz

Für jeden Eigenwert λ einer unitären Matrix A gilt $|\lambda| = 1$.

Beweis

Sei λ Eigenwert von $A \in U(n)$ zum Eigenvektor $\vec{z} \in \mathbb{C}^n \setminus \{\vec{0}\}$. Die Längentreue der unitären Matrix impliziert

$$\|\, \vec{z}\, \|^2 = \|\, A \circ \vec{z}\, \|^2 = (A \circ \vec{z}, A \circ \vec{z}\,) = (\lambda \cdot \vec{z}, \lambda \cdot \vec{z}\,) = \lambda\bar{\lambda}\, \|\, \vec{z}\, \|^2.$$

Da $\|\, \vec{z}\, \|^2 \neq 0$ ist, folgt $|\lambda| = 1$. □

7.3.21 Satz

Sei $A \in O(n, \mathbb{R})$ eine reelle orthogonale Matrix. Die Raumdimension n sei ungerade. Dann ist $+1$ oder -1 Eigenwert von A. Ist A sogar eine Drehung, d. h. $A \in SO(n, \mathbb{R})$, dann ist $+1$ Eigenwert von A.

Wir werden auf S. 257 mit Hilfe der Diagonalisierungsresultate erläutern, daß in $\mathbb{R}^3$ der zu $+1$ gehörige Eigenvektor als Drehachse interpretiert werden kann.

Beweis

Das charakteristische Polynom $P_A(\lambda)$ von A hat reelle Koeffizienten und wird hier als Funktion $\mathbb{R} \to \mathbb{R}$ aufgefaßt. Der Term höchster Ordnung ist λ^n mit der ungeraden Potenz n. Aus der Analysis weiß man $\lim_{\lambda\to\infty} P_A(\lambda) = +\infty$ und $\lim_{\lambda\to-\infty} P_A(\lambda) = -\infty$. Der Zwischenwertsatz ergibt die Existenz wenigstens einer reellen Nullstelle λ_0 mit $|\lambda_0| = 1$, denn A ist orthogonal und insbesondere unitär.

Die nichtreellen Eigenwerte von A treten immer in zueinander konjugierten Paaren auf: Sei λ_1 Eigenwert von A, d. h. Nullstelle des charakteristischen Polynoms

$$0 = P_A(\lambda_1) = \lambda_1^n + \sum_{\nu=0}^{n-1} a_\nu \lambda_1^\nu, \qquad a_\nu \in \mathbb{R}.$$

Durch Konjugation erhält man

$$0 = \overline{P_A(\lambda_1)} = \overline{\lambda_1^n} + \overline{\sum_{\nu=0}^{n-1} a_\nu \lambda_1^\nu} = \bar\lambda_1^n + \sum_{\nu=0}^{n-1} a_\nu \bar\lambda_1^\nu,$$

$\bar\lambda_1$ ist also ebenfalls Eigenwert. Indem man den quadratischen Faktor $(\lambda-\lambda_1)(\lambda-\bar\lambda_1)$ abspaltet, der nur reelle Koeffizienten besitzt, und das Verfahren iteriert, erhält man: Die Matrix A hat $0 \le k \le (n-1)/2$ (nicht notwendigerweise verschiedene) Paare zueinander konjugiert komplexer Eigenwerte $\lambda_1, \bar\lambda_1, \ldots, \lambda_k, \bar\lambda_k$ und eine ungerade Anzahl reeller Eigenwerte $\lambda_{2k+1}, \ldots, \lambda_n \in \{+1, -1\}$; die Eigenwerte werden jeweils entsprechend ihrer algebraischen Vielfachheit aufgeführt. Die Determinante ist das Produkt der Eigenwerte:

$$\det(A) = |\lambda_1|^2 \cdot \ldots \cdot |\lambda_k|^2 \cdot \lambda_{2k+1} \cdot \ldots \cdot \lambda_n = \lambda_{2k+1} \cdot \ldots \cdot \lambda_n.$$

Ist nun $\det(A) = +1$, so können nicht alle reellen Eigenwerte $= -1$ sein, denn n wurde als ungerade vorausgesetzt. In diesem Fall ist $+1$ (zumindest einfacher) Eigenwert von A. □

7.3.22 Beispiel

Durch direktes Nachrechnen sieht man, daß die reelle 3×3-Matrix

$$A = \begin{pmatrix} \frac12 + \frac{\sqrt2}{4} & -\frac12 + \frac{\sqrt2}{4} & -\frac12 \\ -\frac12 + \frac{\sqrt2}{4} & \frac12 + \frac{\sqrt2}{4} & -\frac12 \\ \frac12 & \frac12 & \frac{\sqrt2}{2} \end{pmatrix}$$

orthogonal ist, daß also $A^t \circ A = E$ gilt. Wir berechnen das charakteristische Polynom:

$$P_A(\lambda) = -\det(A - \lambda E) = -\det\begin{pmatrix} \frac12 + \frac{\sqrt2}{4} - \lambda & -\frac12 + \frac{\sqrt2}{4} & -\frac12 \\ -\frac12 + \frac{\sqrt2}{4} & \frac12 + \frac{\sqrt2}{4} - \lambda & -\frac12 \\ \frac12 & \frac12 & \frac{\sqrt2}{2} - \lambda \end{pmatrix}$$

$$
\begin{aligned}
&= -\det\begin{pmatrix} \frac{1}{2}+\frac{\sqrt{2}}{4}-\lambda & -1+\lambda & -\frac{1}{2} \\ -\frac{1}{2}+\frac{\sqrt{2}}{4} & 1-\lambda & -\frac{1}{2} \\ \frac{1}{2} & 0 & \frac{\sqrt{2}}{2}-\lambda \end{pmatrix} \\
&= -\det\begin{pmatrix} \frac{1}{2}+\frac{\sqrt{2}}{4}-\lambda & -1+\lambda & -\frac{1}{2} \\ \frac{\sqrt{2}}{2}-\lambda & 0 & -1 \\ \frac{1}{2} & 0 & \frac{\sqrt{2}}{2}-\lambda \end{pmatrix} \\
&= (\lambda-1)\cdot\det\begin{pmatrix} \frac{\sqrt{2}}{2}-\lambda & -1 \\ \frac{1}{2} & \frac{\sqrt{2}}{2}-\lambda \end{pmatrix} = (\lambda-1)\cdot\left(\left(\frac{\sqrt{2}}{2}-\lambda\right)^2+\frac{1}{2}\right) \\
&= (\lambda-1)\cdot\left(\lambda-\frac{\sqrt{2}}{2}-i\frac{\sqrt{2}}{2}\right)\cdot\left(\lambda-\frac{\sqrt{2}}{2}+i\frac{\sqrt{2}}{2}\right),
\end{aligned}
$$

die Eigenwerte von A lauten also $\lambda_1 = \frac{\sqrt{2}}{2}(1+i)$, $\lambda_2 = \frac{\sqrt{2}}{2}(1-i)$ und $\lambda_3 = 1$. Daraus folgt auch $\det(A) = \frac{\sqrt{2}}{2}(1+i)\cdot\frac{\sqrt{2}}{2}(1-i)\cdot 1 = 1$, also ist A eine Drehung.

Wir bestimmen die Drehachse (sie sei vorbehaltlich der nachzutragenden Erläuterungen von S. 257 bereits so bezeichnet), d. h. den Eigenvektor zum Eigenwert 1, als nichttriviale Lösung des linearen Gleichungssystems

$$
\vec{0} = (A-E)\circ\vec{x} = \begin{pmatrix} -\frac{1}{2}+\frac{\sqrt{2}}{4} & -\frac{1}{2}+\frac{\sqrt{2}}{4} & -\frac{1}{2} \\ -\frac{1}{2}+\frac{\sqrt{2}}{4} & -\frac{1}{2}+\frac{\sqrt{2}}{4} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{\sqrt{2}}{2}-1 \end{pmatrix}\circ\vec{x}.
$$

Etwa mittels des Gaußschen Algorithmus' findet man, daß dieses Gleichungssystem eine linear unabhängige Lösung hat. Der Lösungsraum (die Drehachse) wird von dem Einheitsvektor $\vec{x} = \left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, 0\right)^t$ aufgespannt.

7.4 Die Hauptachsentransformation

In diesem Abschnitt betrachten wir nur $n\times n$-Matrizen $A \in \mathbb{C}^{n\times n}$ bzw. $A \in \mathbb{R}^{n\times n}$ und die entsprechenden linearen Abbildungen $F : \mathbb{C}^n \to \mathbb{C}^n$ bzw. $F : \mathbb{R}^n \to \mathbb{R}^n$. Lineare Abbildungen allgemeiner endlichdimensionaler Vektorräume in sich könnten, wie schon erwähnt, durch die Einführung *unitärer* Koordinaten auf diesen Fall zurückgeführt werden, die relevanten Eigenschaften wie „normal", „selbstadjungiert" oder „unitär" blieben dabei erhalten.

Sei also $F : \mathbb{C}^n \to \mathbb{C}^n$ eine lineare Abbildung mit zugehöriger $n\times n$-Matrix A. Wir können uns nun der zu Beginn dieses Kapitels wieder aufgeworfenen Frage zuwenden: Unter welchen Bedingungen an die Matrix A läßt sich ein orthonormiertes Koordinatensystem $(\vec{b}_1,\ldots,\vec{b}_n)$ in $\mathbb{C}^n$ so geschickt „einzeichnen", daß die Abbildung

F in den durch die unitäre Matrix $B = (\vec{b}_1, \ldots, \vec{b}_n)$ bzw. die entsprechende unitäre Karte $G : \mathbb{C}^n \to \mathbb{C}^n$ gegebenen Koordinaten

$$\hat{F} = G^{-1} \circ F \circ G \quad \text{mit Matrix} \quad \hat{A} = \bar{B}^t \circ A \circ B$$

Diagonalgestalt hat?

Wir werden zeigen, daß ein solches Diagonalisierungsresultat genau in der Klasse der normalen Matrizen Gültigkeit besitzt. Insbesondere folgt damit die Diagonalisierbarkeit unitärer und Hermitescher Matrizen. Da die Eigenwerte reeller symmetrischer Matrizen sämtlich reell sind, erhalten wir daraus unmittelbar ein Diagonalisierungsresultat auch im Reellen, während der Übergang vom unitären zum reell orthogonalen Fall größere Modifikationen erfordert.

Transformation auf Dreiecksgestalt, Trigonalisierung

Wir werden zunächst beweisen, daß jede komplexe $n \times n$-Matrix durch eine unitäre Transformationsmatrix auf Dreiecksgestalt gebracht werden kann, und daraus später unsere Diagonalisierungsresultate herleiten.

Zur Vorbereitung dienen die folgenden beiden Hilfssätze:

7.4.1 Satz

Sei $m < n$ und $B'' \in \mathbb{C}^{m \times m}$ eine unitäre $m \times m$-Matrix. Dann ist

$$B := \begin{pmatrix} E^{n-m,n-m} & \mathcal{O}^{n-m,m} \\ \mathcal{O}^{m,n-m} & B'' \end{pmatrix} \in \mathbb{C}^{n \times n}$$

ebenfalls unitär.

Beweis

Sei $\vec{z} \in \mathbb{C}^n$ beliebig. Wir bezeichnen $\vec{z}\,' := (z_1, \ldots, z_{n-m})^t \in \mathbb{C}^{n-m}$ und $\vec{z}\,'' := (z_{n-m+1}, \ldots, z_n)^t \in \mathbb{C}^m$. Auf Grund der Nullblöcke in B ist $B \circ \vec{z} = (\vec{z}\,', B'' \circ \vec{z}\,'')^t \in \mathbb{C}^{n-m} \oplus \mathbb{C}^m$. Wir greifen noch auf die Definition der euklidischen Norm in $\mathbb{C}^n$ zurück und erhalten

$$\| B \circ \vec{z} \|^2 = \| \vec{z}\,' \|^2 + \| B'' \circ \vec{z}\,'' \|^2 = \| \vec{z}\,' \|^2 + \| \vec{z}\,'' \|^2 = \| \vec{z} \|^2. \qquad \square$$

7.4.2 Satz

Zu je zwei Einheitsvektoren $\vec{z}, \vec{w} \in \mathbb{C}^n$ (d. h. $\| \vec{z} \| = \| \vec{w} \| = 1$) existiert eine unitäre $n \times n$-Matrix $B \in U(n)$ mit $\vec{w} = B \circ \vec{z}$.

Beweis

Genauso wie in Satz 6.3.5 beweist man: es gibt Vektoren $\vec{z}_2, \ldots, \vec{z}_n$ derart, daß $(\vec{z}, \vec{z}_2, \ldots, \vec{z}_n)$ ein orthogonales n-Bein ist, d. h. nach Satz 7.2.14 ist $B_1 :=$

$(\vec{z}, \vec{z}_2, \ldots, \vec{z}_n) \in U(n)$. Es gilt

$$B_1 \circ \vec{e}_1 = \vec{z}, \text{ also } \vec{e}_1 = B_1^{-1} \circ \vec{z} = \bar{B}_1^t \circ \vec{z}.$$

Entsprechend konstruiert man ein $B_2 \in U(n)$ mit $B_2 \circ \vec{e}_1 = \vec{w}$. Für $B := B_2 \circ \bar{B}_1^t \in U(n)$ ist dann in der Tat $B \circ \vec{z} = B_2 \circ \vec{e}_1 = \vec{w}$. □

7.4.3 Satz

Zu jeder komplexen $n \times n$-Matrix A existiert eine unitäre $n \times n$-Matrix $B \in U(n)$ derart, daß $\hat{A} := \bar{B}^t \circ A \circ B$ eine (obere) Dreiecksmatrix ist, daß also $\hat{a}_{\mu\nu} = 0$ gilt für $\mu > \nu$.

Beweis

Wir führen den Beweis durch vollständige Induktion nach n. Bei der Induktionsbasis $n = 1$ ist nichts zu zeigen, denn jede 1×1-„Matrix" ist eine Dreiecks-„Matrix".

Sei nun $n > 1$, und wir nehmen an, die Behauptung sei für $n - 1$ bewiesen. Das charakteristische Polynom $P(\lambda) = P_A(\lambda)$ der $n \times n$-Matrix A zerfällt über $\mathbb{C}$ in Linearfaktoren, es gibt also einen Eigenwert $\lambda_1 \in \mathbb{C}$ mit zugehörigem Eigenvektor $\vec{z}_1 \neq \vec{0}$. Wir können den Eigenvektor normieren und ohne Einschränkung $\| \vec{z}_1 \| = 1$ annehmen. Gemäß Satz 7.4.2 existiert eine unitäre Abbildung $B_1 \in U(n)$, die den ersten Einheitsvektor auf diesen Eigenvektor abbildet: $B_1 \circ \vec{e}_1 = \vec{z}_1$.

Für

$$A_1 := \bar{B}_1^t \circ A \circ B_1$$

gilt

$$A_1 \circ \vec{e}_1 = \bar{B}_1^t \circ A \circ (B_1 \circ \vec{e}_1) = \bar{B}_1^t \circ (A \circ \vec{z}_1) = \bar{B}_1^t \circ (\lambda_1 \vec{z}_1) = \lambda_1 \vec{e}_1,$$

also

$$A_1 = \left(\begin{array}{c|c} \lambda_1 & * \\ \hline 0 & \\ \vdots & A' \\ 0 & \end{array} \right)$$

mit einer $(n-1) \times (n-1)$-Matrix A'. Nach Induktionsvoraussetzung existiert ein $B' \in U(n-1)$, so daß in $\bar{B'}^t \circ A' \circ B'$ unterhalb der Diagonalen ausschließlich Nullen stehen. Auf die in Satz 7.4.1 beschriebene Weise erhalten wir aus B' eine n-reihige unitäre Matrix:

$$B_2 := \begin{pmatrix} 1 & \mathcal{O}^{1,n-1} \\ \mathcal{O}^{n-1,1} & B' \end{pmatrix} \in U(n).$$

Wir betrachten die Hintereinanderschaltung

$$B := B_1 \circ B_2 \in U(n)$$

der unitären Transformationen B_1 und B_2. Für

$$\hat{A} := \bar{B}_2^t \circ A_1 \circ B_2 = \bar{B}_2^t \circ \bar{B}_1^t \circ A \circ B_1 \circ B_2 = \overline{(B_1 \circ B_2)}^t \circ A \circ (B_1 \circ B_2) = \bar{B}^t \circ A \circ B$$

erhalten wir nun Dreiecksgestalt:

$$\begin{aligned} \hat{A} &= \begin{pmatrix} 1 & \mathcal{O} \\ \mathcal{O} & \bar{B}'^t \end{pmatrix} \circ \begin{pmatrix} \lambda_1 & * \\ \mathcal{O} & A' \end{pmatrix} \circ \begin{pmatrix} 1 & \mathcal{O} \\ \mathcal{O} & B' \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 & * \\ \mathcal{O} & \bar{B}'^t \circ A' \end{pmatrix} \circ \begin{pmatrix} 1 & \mathcal{O} \\ \mathcal{O} & B' \end{pmatrix} = \begin{pmatrix} \lambda_1 & * \\ \mathcal{O} & \bar{B}'^t \circ A' \circ B' \end{pmatrix}. \end{aligned}$$ □

Im Beweis haben wir wesentlichen Gebrauch vom Fundamentalsatz der Algebra gemacht. Betrachtet man eine reelle Matrix A, so hat deren charakteristisches Polynom $P_A(\lambda)$ zwar nur reelle Koeffizienten, der Fundamentalsatz der Algebra besitzt im Reellen jedoch keine Gültigkeit. Tatsächlich wird der soeben bewiesene Satz in dieser Situation im allgemeinen falsch:

7.4.4 Beispiel

Wir betrachten die reelle normale 2×2-Matrix $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Für beliebiges $B = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in O(2, \mathbb{R})$ berechnet man

$$B^t \circ A \circ B = \begin{pmatrix} 0 & -ad + bc \\ -bc + ad & 0 \end{pmatrix} = \begin{pmatrix} 0 & -\det(B) \\ \det(B) & 0 \end{pmatrix};$$

Dreiecksgestalt läßt sich nicht erreichen. Mit der komplexen unitären Matrix $C = \frac{\sqrt{2}}{2}\begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}$ ist dagegen $\bar{C}^t \circ A \circ C = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}$.

Um das reelle Analogon von Satz 7.4.3 zu beweisen, haben wir also die Existenz von reellen Eigenwerten vorauszusetzen.

7.4.5 Satz

Die reelle $n \times n$-Matrix A habe nur reelle Eigenwerte, d. h. das charakteristische Polynom $P_A(\lambda)$ zerfalle über $\mathbb{R}$ *in Linearfaktoren. Dann existiert eine (reelle) Matrix $B \in SO(n)$ derart, daß $\hat{A} = B^{-1} \circ A \circ B = B^t \circ A \circ B$ eine Dreiecksmatrix ist.*

Beweis

Weil wir hier weitgehend parallel zum Beweis von Satz 7.4.3 vorgehen, wird nur der Induktionsschritt skizziert.

Gemäß Voraussetzung existiert ein Eigenwert $\lambda_1 \in \mathbb{R}$ von A, wegen Bemerkung 7.3.10 gibt es zu λ_1 einen reellen Eigenvektor:

$$A \circ \vec{x}_1 = \lambda_1 \cdot \vec{x}_1 \text{ mit } \vec{x}_1 \in \mathbb{R}^n, \ \|\vec{x}_1\| = 1.$$

In Satz 6.3.5 haben wir eine Drehung $B_1 \in SO(n)$ konstruiert mit $\vec{x}_1 = B_1 \circ \vec{e}_1$. Die transformierte Matrix $A_1 := B_1^t \circ A \circ B_1$ hat dasselbe Aussehen wie oben und gemäß Satz 7.3.7 dasselbe charakteristische Polynom wie A. Ist $Q(\lambda)$ das charakteristische Polynom von A', so liefert die Entwicklung von $\det(A_1 - \lambda E)$ nach der ersten Spalte

$$P(\lambda) = (\lambda - \lambda_1) \cdot Q(\lambda).$$

Also zerfällt auch $Q(\lambda)$ über $\mathbb{R}$ in Linearfaktoren. Die Induktionsvoraussetzung liefert ein $B' \in SO(n-1)$, so daß $B'^t \circ A' \circ B'$ Dreiecksmatrix ist. Der Rest des Beweises verläuft wie oben, wobei die Konjugationsstriche entfallen und $U(n)$ durch $SO(n)$ zu ersetzen ist. $\square$

7.4.6 Folgerung

Die geometrische Vielfachheit eines Eigenwertes kann seine algebraische Vielfachheit nicht übersteigen.

Beweis

Wir können uns auf den komplexen Fall beschränken: Die geometrische Vielfachheit im Reellen kann die im Komplexen nicht übersteigen, denn *reelle* Vektoren, also solche aus $\mathbb{R}^n$, die über $\mathbb{R}$ linear unabhängig sind, sind auch über $\mathbb{C}$ linear unabhängig. Die algebraischen Vielfachheiten stimmen im Reellen und Komplexen überein.

Sei λ_0 Eigenwert der $n \times n$-Matrix A mit geometrischer Vielfachheit $s = \dim V_{\lambda_0}$ und algebraischer Vielfachheit r. Ähnlich wie im Beweis von Satz 7.4.3 wollen wir den Eigenraum V_{λ_0} von $\mathbb{C}^n$ abspalten. Sei dazu $\vec{z}_1, \dots, \vec{z}_s$ eine Orthonormalbasis von V_{λ_0}, diese können wir durch Verwendung des Schmidtschen Orthonormalisierungsverfahrens zu einem orthogonalen n-Bein $\vec{z}_1, \dots, \vec{z}_n$ ergänzen. Die Matrix $B_1 := (\vec{z}_1, \dots, \vec{z}_n)$ ist also unitär. Mit der Wahl von $A_1 := \bar{B}_1^t \circ A \circ B_1$ folgt für $\nu = 1, \dots, s$:

$$A_1 \circ \vec{e}_\nu = \bar{B}_1^t \circ A \circ (B_1 \circ \vec{e}_\nu) = \bar{B}_1^t \circ (A \circ \vec{z}_\nu) = \lambda_0 \left(\bar{B}_1^t \circ \vec{z}_\nu\right) = \lambda_0 \vec{e}_\nu,$$

also

$$A_1 = \begin{array}{cc}
\begin{array}{c}\text{Zeile/}\\ \text{Spalte}\end{array} & \begin{array}{cccccc} 1 & \cdots & s & s+1 & \cdots & n \end{array} \\
\begin{array}{c} 1 \\ \vdots \\ s \\ s+1 \\ \vdots \\ n \end{array} & \begin{pmatrix} \lambda_0 & & \mathcal{O} & & & \\ & \ddots & & & * & \\ \mathcal{O} & & \lambda_0 & & & \\ & & & & & \\ & \mathcal{O} & & & A' & \\ & & & & & \end{pmatrix}
\end{array}.$$

Gemäß Satz 7.4.3 existiert eine unitäre Matrix $B' \in U(n-s)$, so daß $\bar{B}'^t \circ A' \circ B'$ Dreiecksgestalt hat. Mit

$$B_2 := \begin{pmatrix} E^{s,s} & \mathcal{O} \\ \mathcal{O} & B' \end{pmatrix} \in U(n)$$

erhalten wir wie in Satz 7.4.3

$$\hat{A} := \bar{B}_2^t \circ A_1 \circ B_2 = \begin{array}{cc} \text{Zeile/ Spalte} & \begin{matrix} 1 & \cdots & s & s{+}1 & \cdots & n \end{matrix} \\ \begin{matrix} 1 \\ \vdots \\ s \\ s{+}1 \\ \vdots \\ n \end{matrix} & \begin{pmatrix} \lambda_0 & & \mathcal{O} & & & \\ & \ddots & & & * & \\ \mathcal{O} & & \lambda_0 & & & \\ & & & & & \\ & \mathcal{O} & & & \bar{B'}^t \circ A' \circ B' & \\ & & & & & \end{pmatrix} \end{array}$$

$$= \begin{array}{cc} \text{Zeile/ Spalte} & \begin{matrix} 1 & \cdots & s & s{+}1 & \cdots & n \end{matrix} \\ \begin{matrix} 1 \\ \vdots \\ s \\ s{+}1 \\ \vdots \\ n \end{matrix} & \begin{pmatrix} \lambda_0 & & \mathcal{O} & & & \\ & \ddots & & & * & \\ \mathcal{O} & & \lambda_0 & & & \\ & & & \lambda_{s+1} & & * \\ & \mathcal{O} & & & \ddots & \\ & & & \mathcal{O} & & \lambda_n \end{pmatrix} \end{array}.$$

Die Matrizen A und $\hat{A}$ haben dasselbe charakteristische Polynom, es lautet $P(\lambda) = (\lambda - \lambda_0)^s \cdot \prod_{\nu=s+1}^{n} (\lambda - \lambda_\nu)$. Die Nullstellen λ_ν sind i. a. *nicht* paarweise und auch nicht von λ_0 verschieden, es folgt $s \leq r$. □

Diagonalisierbarkeit normaler Matrizen

Wir wollen zeigen, daß die Dreiecksmatrix aus Satz 7.4.3 sogar eine Diagonalmatrix ist, sofern eine normale Matrix vorgegeben wurde. Hier liegt die Beobachtung zugrunde, daß Normalität unter unitären Koordinatentransformationen erhalten bleibt:

7.4.7 Satz

Die Matrix $A \in \mathbb{C}^{n \times n}$ sei normal. Dann ist $\bar{B}^t \circ A \circ B$ für jede unitäre Matrix $B \in U(n)$ ebenfalls normal.

Beweis

$$\begin{aligned} &\overline{(\bar{B}^t \circ A \circ B)}^t \circ \bar{B}^t \circ A \circ B - \bar{B}^t \circ A \circ B \circ \overline{(\bar{B}^t \circ A \circ B)}^t \\ &= \bar{B}^t \circ \bar{A}^t \circ B \circ \bar{B}^t \circ A \circ B - \bar{B}^t \circ A \circ B \circ \bar{B}^t \circ \bar{A}^t \circ B \\ &= \bar{B}^t \circ \bar{A}^t \circ A \circ B - \bar{B}^t \circ A \circ \bar{A}^t \circ B \\ &= \bar{B}^t \circ \left(\bar{A}^t \circ A - A \circ \bar{A}^t\right) \circ B = \mathcal{O}. \end{aligned}$$

□

7.4.8 Satz

Zu jeder normalen Matrix $A \in \mathbb{C}^{n\times n}$ existiert eine unitäre Matrix $B \in U(n)$ derart, daß $\hat{A} := \bar{B}^t \circ A \circ B$ Diagonalgestalt hat. Die Spalten von $B = \left(\vec{b}_1, \ldots, \vec{b}_n\right)$ bilden eine Orthonormalbasis von $\mathbb{C}^n$ aus Eigenvektoren von A. Die Diagonalelemente von $\hat{A}$ sind genau die Eigenwerte von A entsprechend ihren geometrischen Vielfachheiten.

Beweis

Gemäß Satz 7.4.3 finden wir eine unitäre Matrix $B \in U(n)$, so daß $\hat{A} := \bar{B}^t \circ A \circ B$ Dreiecksgestalt hat; $\hat{A}$ ist ebenfalls normal. Es bleibt zu zeigen, daß normale Dreiecksmatrizen schon Diagonalmatrizen sind.

Wir haben $\hat{a}_{\mu\nu} = 0$ für $\mu > \nu$. Die Normalität ergibt für $\mu = 1, \ldots, n$:

$$\begin{aligned} 0 &= \sum_{\sigma=1}^{n} \left(\bar{\hat{a}}_{\sigma\mu}\hat{a}_{\sigma\mu} - \hat{a}_{\mu\sigma}\bar{\hat{a}}_{\mu\sigma}\right) = \sum_{\sigma=1}^{\mu} |\hat{a}_{\sigma\mu}|^2 - \sum_{\sigma=\mu}^{n} |\hat{a}_{\mu\sigma}|^2 \\ &= \sum_{\sigma=1}^{\mu-1} |\hat{a}_{\sigma\mu}|^2 - \sum_{\sigma=\mu+1}^{n} |\hat{a}_{\mu\sigma}|^2. \end{aligned}$$

Die Betrachtung von $\mu = 1$ ergibt

$$0 = \sum_{\sigma=2}^{n} |\hat{a}_{1\sigma}|^2, \qquad \hat{a}_{1\sigma} = 0 \text{ für } \sigma = 2, \ldots, n.$$

Wir fahren induktiv fort und nehmen an, wir hätten im k-ten Schritt $\hat{a}_{\mu\sigma} = 0$ für $\sigma = \mu + 1, \ldots, n$, $\mu = 1, \ldots, k$ erreicht. Durch Betrachtung von $\mu = k + 1$ folgt:

$$0 = \sum_{\sigma=1}^{k} |\hat{a}_{\sigma,k+1}|^2 - \sum_{\sigma=k+2}^{n} |\hat{a}_{k+1,\sigma}|^2 = - \sum_{\sigma=k+2}^{n} |\hat{a}_{k+1,\sigma}|^2,$$

also ist auch $\hat{a}_{k+1,\sigma} = 0$ für $\sigma = k + 2, \ldots, n$.

Nach $(n-1)$ Schritten erhalten wir, daß $\hat{a}_{\mu\nu} = 0$ auch für alle μ, ν mit $\mu < \nu$ ist. Auch oberhalb der Diagonalen enthält $\hat{A}$ nur Nullen und ist folglich eine Diagonalmatrix.

Da die Matrix B unitär ist, bilden deren Spalten ein orthogonales n-Bein, also eine Orthonormalbasis von $\mathbb{C}^n$. Wir haben zu zeigen, daß es sich dabei um Eigenvektoren von A handelt. Aus $\hat{A} = \bar{B}^t \circ A \circ B$ folgt $B \circ \hat{A} = A \circ B$. Wir betrachten jeweils die ν-te Spalte:

$$A \circ \vec{b}_\nu = B \circ \vec{\hat{a}}_\nu = B \circ (\hat{a}_{\nu\nu}\vec{e}_\nu) = \hat{a}_{\nu\nu}\,(B \circ \vec{e}_\nu) = \hat{a}_{\nu\nu} \circ \vec{b}_\nu,$$

in der Tat ist $\vec{b}_\nu$ Eigenvektor von A zum Eigenwert $\lambda_\nu := \hat{a}_{\nu\nu}$. Jeder Eigenwert tritt gemäß seiner geometrischen Vielfachheit auf, denn die Anzahl der entsprechenden Eigenvektoren in B ist gerade die Dimension des Eigenraums. Somit sind $\hat{a}_{11}, \ldots, \hat{a}_{nn}$

genau die Eigenwerte von A, denn A und $\hat{A}$ haben dieselben Eigenwerte mit denselben geometrischen Vielfachheiten. □

Mit den normalen Matrizen haben wir die größtmögliche Klasse derjenigen Matrizen bestimmt, die mit Hilfe unitärer Koordinatenwechselabbildungen diagonalisiert werden können:

7.4.9 Satz

Für die $n \times n$-Matrix A existiere eine unitäre $n \times n$-Matrix $B \in U(n)$ derart, daß $\bar{B}^t \circ A \circ B$ Diagonalgestalt hat. Dann ist A normal.

Beweis

Es bezeichne

$$D = \begin{pmatrix} d_1 & & \mathcal{O} \\ & \ddots & \\ \mathcal{O} & & d_n \end{pmatrix} = \bar{B}^t \circ A \circ B;$$

diese Diagonalmatrix ist normal:

$$\begin{aligned} \bar{D}^t \circ D &= \begin{pmatrix} \bar{d}_1 & & \mathcal{O} \\ & \ddots & \\ \mathcal{O} & & \bar{d}_n \end{pmatrix} \circ \begin{pmatrix} d_1 & & \mathcal{O} \\ & \ddots & \\ \mathcal{O} & & d_n \end{pmatrix} \\ &= \begin{pmatrix} \bar{d}_1 \cdot d_1 & & \mathcal{O} \\ & \ddots & \\ \mathcal{O} & & \bar{d}_n \cdot d_n \end{pmatrix} = D \circ \bar{D}^t. \end{aligned}$$

Nun ist $A = B \circ D \circ \bar{B}^t = \overline{\bar{B}^t}^t \circ D \circ \bar{B}^t$ mit $\bar{B}^t \in U(n)$, und A ist gemäß Satz 7.4.7 normal. □

Das reelle Analogon zu Satz 7.4.8 leitet man mit wörtlich demselben Beweis aus Satz 7.4.5 her.

7.4.10 Satz

Die reelle $n \times n$-Matrix A sei normal und habe nur reelle Eigenwerte. Dann existiert eine reelle Drehung $B \in SO(n)$ derart, daß $\hat{A} := B^t \circ A \circ B$ eine reelle Diagonalmatrix ist.

Bezüglich jeder normalen Matrix können wir mit Hilfe des Diagonalisierungssatzes den $\mathbb{C}^n$ in eine orthogonale Summe über deren Eigenräume zerlegen:

7.4.11 Satz

Es seien $\lambda_1, \ldots, \lambda_s$ sämtliche paarweise verschiedenen Eigenwerte einer normalen $n \times n$-Matrix A. Dann besitzt $\mathbb{C}^n$ eine orthogonale Zerlegung in die Eigenräume von A:

$$\mathbb{C}^n = \oplus_{\nu=1}^{s} V_{\lambda_\nu}.$$

Ist $P(\lambda) = \prod_{\nu=1}^{s}(\lambda - \lambda_\nu)^{r_\nu}$ das charakteristische Polynom von A, so ist $r_\nu = \dim V_{\lambda_\nu}$, die geometrischen und algebraischen Vielfachheiten der Eigenwerte stimmen also überein.

Beweis

Die Orthogonalität der Eigenräume wurde bereits in Satz 7.3.17 nachgewiesen.

Sei $B \in U(n)$ derart, daß $\hat{A} := \bar{B}^t \circ A \circ B$ Diagonalgestalt hat, die Spalten $B = \left(\vec{b}_1, \ldots, \vec{b}_n\right)$ bilden eine Orthonormalbasis von Eigenvektoren von A, in der Diagonalen von $\hat{A}$ treten die Eigenwerte von $\hat{A}$ und damit die von A entsprechend ihrer geometrischen Vielfachheit $\dim V_{\lambda_\nu}$ auf; man beachte, daß sich geometrische Vielfachheiten bei Koordinatenwechseln nicht ändern. In $\hat{A}$ und folglich auch in A stimmen somit offensichtlich geometrische und algebraische Vielfachheiten r_ν überein: $r_\nu = \dim V_{\lambda_\nu}$. Indem man ggfs. die Spalten von B vertauscht, kann man erreichen:

$$\begin{array}{ll}
\vec{b}_1, \ldots, \vec{b}_{r_1} & \text{sind Eigenvektoren zu } \lambda_1, \\
\vec{b}_{r_1+1}, \ldots, \vec{b}_{r_1+r_2} & \text{sind Eigenvektoren zu } \lambda_2, \\
\vdots & \\
\vec{b}_{r_1+\ldots+r_{s-1}+1}, \ldots, \vec{b}_n & \text{sind Eigenvektoren zu } \lambda_s.
\end{array}$$

Da jedes Element $\vec{z} \in \mathbb{C}^n$ eine Darstellung in der Basis $\{\vec{b}_1, \ldots, \vec{b}_n\}$ besitzt, folgt durch entsprechende Klammerung die Zerlegung von $\vec{z}$ in Komponenten in den jeweiligen Eigenräumen:

$$\begin{aligned}
\vec{z} &= \sum_{\nu=1}^{n} c_\nu \vec{b}_\nu = \left(\sum_{\nu=1}^{r_1} c_\nu \vec{b}_\nu\right) + \left(\sum_{\nu=r_1+1}^{r_1+r_2} c_\nu \vec{b}_\nu\right) + \ldots + \left(\sum_{\nu=r_1+\ldots+r_{s-1}+1}^{n} c_\nu \vec{b}_\nu\right) \\
&=: \vec{z}_1 + \ldots + \vec{z}_s, \qquad \text{mit } \vec{z}_\nu \in V_{\lambda_\nu}.
\end{aligned}$$

Die Eindeutigkeit dieser Darstellung folgt aus der Orthogonalität der Eigenräume. □

Die Hauptachsentransformation für Hermitesche und reell symmetrische Matrizen

Wir erhalten den folgenden Satz direkt aus dem Diagonalisierungsresultat für normale Matrizen Satz 7.4.8 zusammen mit der Beobachtung aus Satz 7.3.18, daß Eigenwerte Hermitescher Matrizen stets reell sind:

7.4.12 Satz (Hauptachsentransformation)

Zu jeder Hermiteschen $n \times n$-Matrix A existiert eine unitäre Matrix $B \in U(n)$ derart, daß $\hat{A} := \bar{B}^t \circ A \circ B$ eine reelle Diagonalmatrix ist.

Man erhält das reelle Analogon, indem man auf Satz 7.4.10 anstatt auf Satz 7.4.8 verweist.

7.4.13 Satz (Hauptachsentransformation)

Sei A eine symmetrische reelle $n \times n$-Matrix. Dann existiert eine Drehung $B \in SO(n, \mathbb{R})$ derart, daß $\hat{A} := B^t \circ A \circ B$ eine (reelle) Diagonalmatrix ist.

7.4.14 Bemerkung

Um den Begriff „Hauptachsentransformation“ zu erläutern, betrachten wir zunächst eine reelle $n \times n$-Diagonalmatrix $\hat{A}$ mit den (nicht notwendigerweise verschiedenen) Diagonalelementen $\lambda_1, \ldots, \lambda_n$. Unter dieser Abbildung wird die Einheitskugel

$$K = \left\{ \vec{x} \in \mathbb{R}^n : \ \| \vec{x} \|^2 \leq 1 \right\}$$

auf ein Ellipsoid

$$E = \left\{ \vec{x} \in \mathbb{R}^n : \ \sum_{\nu=1}^{n} \frac{x_\nu^2}{\lambda_\nu^2} \leq 1 \right\}$$

abgebildet. Die Basisvektoren $\lambda_\nu \vec{e}_\nu$ sind die Hauptachsen dieses Ellipsoids.

Sei nun $A \in \mathbb{R}^{n \times n}$ eine symmetrische Matrix und $B \in SO(n, \mathbb{R})$ eine Transformationsmatrix derart, daß die Abbildung in diesen Koordinaten Diagonalgestalt $\hat{A} := B^t \circ A \circ B$ wie oben hat. Man kann also in $\mathbb{R}^n$ eine Orthonormalbasis $B = \left(\vec{b}_1, \ldots, \vec{b}_n \right)$ „einzeichnen“, so daß unter der Abbildung A die Einheitskugel K in ein Ellipsoid mit den Basisvektoren $\lambda_\nu \vec{b}_\nu$ als Hauptachsen übergeht.

7.4.15 Beispiel

Eine wichtige Anwendung des Diagonalisierungssatzes für reell symmetrische Matrizen findet man in der Theorie der Drehbewegung starrer Körper. Für eine ausführliche Darstellung verweisen wir auf die Lehrbücher der Mechanik wie etwa das von Á. Budó [1, §§ 51 ff].

Wir stellen uns einen solchen Körper vor und legen den Ursprung und ein orthogonales Koordinatensystem in dessen Schwerpunkt. Wir betrachten Drehbewegungen nur um Achsen, die durch diesen Ursprung gehen, um insgesamt resultierende *Kräfte* auf diese Achsen zu vermeiden. Der Zusammenhang zwischen dem Drehimpulsvektor $\vec{L}$ und der Winkelgeschwindigkeit $\vec{\omega}$ wird durch den „Trägheitstensor“

$$J = \begin{pmatrix} j_{11} & -j_{12} & -j_{13} \\ -j_{21} & j_{22} & -j_{23} \\ -j_{31} & -j_{32} & j_{33} \end{pmatrix} \in \mathbb{R}^{3 \times 3}$$

vermittelt:

$$\vec{L} = J \circ \vec{\omega}.$$

Dabei sind die Diagonalelemente $j_{\nu\nu}$ die „Trägheitsmomente“ bzgl. der Rotation um die x_ν-Achse und $j_{\mu\nu}$ für $\mu \neq \nu$ „Deviationsmomente“. Die Deviationsmomente sind ein Maß dafür, wie sehr bei der Rotation um die Koordinatenachsen *Drehmomente* auf diese Achsen auftreten, die deren Lage zu ändern versuchen. Bei realen Bewegungen von Maschinen oder Fahrzeugen müssen diese Momente von den Lagern der Achse kompensiert werden und führen zu einem „unrunden“ Lauf und erhöhtem Verschleiß. Deshalb ist man bemüht, die Drehachse und damit $\vec{\omega}$ so einzurichten, daß Drehimpuls und Winkelgeschwindigkeit parallel sind. Es sind also die Eigenvektoren des Trägheitstensors J zu bestimmen.

Den definierenden Gleichungen für dessen Komponenten entnimmt man (vgl. [1, § 51]), daß der Trägheitstensor J symmetrisch ist. Er besitzt somit eine Orthonormalbasis $\left\{\vec{b}_1, \vec{b}_2, \vec{b}_3\right\}$ aus Eigenvektoren, den „Hauptträgheitsachsen“ des Körpers im Schwerpunkt. Führt man die durch $B = \left(\vec{b}_1, \vec{b}_2, \vec{b}_3\right)$ gegebenen Koordinaten ein, so lautet der Trägheitstensor in diesen Koordinaten:

$$\hat{J} := B^t \circ J \circ B = \begin{pmatrix} \hat{j}_{11} & 0 & 0 \\ 0 & \hat{j}_{22} & 0 \\ 0 & 0 & \hat{j}_{33} \end{pmatrix}.$$

Bezüglich der Hauptträgheitsachsen $\vec{b}_1, \vec{b}_2, \vec{b}_3$ treten keine Deviationsmomente auf.

> Unter „Auswuchten“ eines Körpers versteht man dessen Veränderung derart, daß die (fest vorgegebene) Drehachse zu einer Hauptträgheitsachse im Schwerpunkt wird.

7.4.16 Beispiele

Wir setzen hier die Behandlung der Beispiele 7.3.19 fort.

(a) Für $A = \begin{pmatrix} 1 & -3i \\ 3i & 1 \end{pmatrix}$ hatten wir dort $\lambda_1 = 4$ und $\lambda_2 = -2$ als Eigenwerte bestimmt. Zugehörige Eigenvektoren findet man als nichttriviale Lösungen der Gleichungssysteme

$$\vec{0} = (A - \lambda_1 E) \circ \vec{x}_1 = \begin{pmatrix} -3 & -3i \\ 3i & -3 \end{pmatrix} \circ \vec{x}_1$$

und

$$\vec{0} = (A - \lambda_2 E) \circ \vec{x}_2 = \begin{pmatrix} 3 & -3i \\ 3i & 3 \end{pmatrix} \circ \vec{x}_2.$$

Man berechnet sofort $\vec{x}_1 = (-i, 1)^t$ und $\vec{x}_2 = (i, 1)^t$ als Lösungsvektoren. Durch Normieren erhält man die Transformationsmatrix

$$B = \begin{pmatrix} -i\frac{\sqrt{2}}{2} & i\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} = \frac{\sqrt{2}}{2} \begin{pmatrix} -i & i \\ 1 & 1 \end{pmatrix}.$$

Hiermit gilt

$$\begin{aligned} \hat{A} = \bar{B}^t \circ A \circ B &= \begin{pmatrix} i\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -i\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \circ \begin{pmatrix} 1 & -3i \\ 3i & 1 \end{pmatrix} \circ \begin{pmatrix} -i\frac{\sqrt{2}}{2} & i\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \\ &= \begin{pmatrix} 4 & 0 \\ 0 & -2 \end{pmatrix}. \end{aligned}$$

(b) Die Eigenwerte von

$$A = \begin{pmatrix} \frac{7}{6} & -\frac{2}{3} & \frac{1}{6} \\ -\frac{2}{3} & \frac{2}{3} & -\frac{2}{3} \\ \frac{1}{6} & -\frac{2}{3} & \frac{7}{6} \end{pmatrix}$$

lauten $\lambda_1 = 0$, $\lambda_2 = 1$, $\lambda_3 = 2$. Als zugehörige Eigenvektoren bestimmt man mit dem Gaußschen Algorithmus beispielsweise $\vec{x}_1 = (1, 2, 1)^t$, $\vec{x}_2 = (-1, 0, 1)^t$ und $\vec{x}_3 = (1, -1, 1)^t$. Durch Normieren gelangt man zu der Transformationsmatrix

$$B = \begin{pmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{pmatrix} = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 & -\sqrt{3} & \sqrt{2} \\ 2 & 0 & -\sqrt{2} \\ 1 & \sqrt{3} & \sqrt{2} \end{pmatrix},$$

es ist

$$\det(B) = \frac{1}{6^{3/2}} \det \begin{pmatrix} 1 & -\sqrt{3} & \sqrt{2} \\ 0 & 2\sqrt{3} & -3\sqrt{2} \\ 0 & 2\sqrt{3} & 0 \end{pmatrix} = \frac{1}{6^{3/2}} \cdot 6\sqrt{6} = 1.$$

Ohne daß eine Spalte von B durch ihr Negatives ersetzt werden müßte, haben wir bereits $B \in SO(3)$, so daß also B eine Drehung des $\mathbb{R}^3$ ist. In den durch B gegebenen Koordinaten lautet nun A:

$$\begin{aligned} \hat{A} &= B^t \circ A \circ B \\ &= \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} \circ \begin{pmatrix} \frac{7}{6} & -\frac{2}{3} & \frac{1}{6} \\ -\frac{2}{3} & \frac{2}{3} & -\frac{2}{3} \\ \frac{1}{6} & -\frac{2}{3} & \frac{7}{6} \end{pmatrix} \circ \begin{pmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}. \end{aligned}$$

7.4.17 Beispiel

In der Analysis benötigt man beispielsweise bei der Behandlung gewöhnlicher Differentialgleichungssysteme die „Exponentialfunktion“ quadratischer n-reihiger Matrizen A:

$$\exp(A) := \sum_{k=0}^{\infty} \frac{1}{k!} A^k. \qquad (*)$$

Die Partialsummen $S_N := \left(s_{\mu\nu}^{(N)}\right)_{\mu,\nu=1,\ldots,n} := \sum_{k=0}^{N} \frac{1}{k!} A^k$ sind jeweils $n \times n$-Matrizen, von deren Komponenten $s_{\mu\nu}^{(N)}$ man in der Analysis Konvergenz nachweist. Man findet Zahlen $s_{\mu\nu} := \lim_{N\to\infty} s_{\mu\nu}^{(N)} \in \mathbb{C}$ und setzt $\exp(A) := (s_{\mu\nu})_{\mu,\nu=1,\ldots,n}$, die Definition $(*)$ ist also im Sinne komponentenweiser Konvergenz zu verstehen. Wir wollen uns hier aber nicht mit Konvergenz, sondern mit der Frage beschäftigen, ob man die Komponenten von $\exp(A)$ auf einfache Weise aus den Komponenten von A bestimmen kann. Im Falle normaler und damit insbesondere auch Hermitescher Matrizen ist das tatsächlich möglich.

Sei A eine normale $n \times n$-Matrix und $B \in U(n)$ derart, daß wir Diagonalgestalt

$$\hat{A} := \bar{B}^t \circ A \circ B = \begin{pmatrix} \lambda_1 & & \mathcal{O} \\ & \ddots & \\ \mathcal{O} & & \lambda_n \end{pmatrix}, \qquad \lambda_1, \ldots, \lambda_n \in \mathbb{C},$$

erreichen. Es folgt:

$$\begin{aligned}
A &= B \circ \hat{A} \circ \bar{B}^t, \\
A^k &= \left(B \circ \hat{A} \circ \bar{B}^t\right)^k \\
&= B \circ \hat{A} \circ \left(\bar{B}^t \circ B\right) \circ \hat{A} \circ \bar{B}^t \circ \ldots \circ B \circ \hat{A} \circ \bar{B}^t \\
&= B \circ \hat{A}^k \circ \bar{B}^t = B \circ \begin{pmatrix} \lambda_1^k & & \mathcal{O} \\ & \ddots & \\ \mathcal{O} & & \lambda_n^k \end{pmatrix} \circ \bar{B}^t, \\
\exp(A) &= \sum_{k=0}^{\infty} \frac{1}{k!} B \circ \begin{pmatrix} \lambda_1^k & & \mathcal{O} \\ & \ddots & \\ \mathcal{O} & & \lambda_n^k \end{pmatrix} \circ \bar{B}^t \\
&= B \circ \left(\sum_{k=0}^{\infty} \frac{1}{k!} \begin{pmatrix} \lambda_1^k & & \mathcal{O} \\ & \ddots & \\ \mathcal{O} & & \lambda_n^k \end{pmatrix} \right) \circ \bar{B}^t \\
&= B \circ \begin{pmatrix} \sum_{k=0}^{\infty} \frac{1}{k!} \lambda_1^k & & \mathcal{O} \\ & \ddots & \\ \mathcal{O} & & \sum_{k=0}^{\infty} \frac{1}{k!} \lambda_n^k \end{pmatrix} \circ \bar{B}^t
\end{aligned}$$

$$= B \circ \begin{pmatrix} \exp(\lambda_1) & & \mathcal{O} \\ & \ddots & \\ \mathcal{O} & & \exp(\lambda_n) \end{pmatrix} \circ \bar{B}^t.$$

Man erhält die Exponentialfunktion normaler Matrizen, indem man die gewöhnliche komplexe Exponentialfunktion der Eigenwerte berechnet und diese Diagonalmatrix mittels B und $\bar{B}^t$ „zurück"-transformiert. Die Eigenwerte von $\exp(A)$ sind ebenfalls $\exp(\lambda_1), \ldots, \exp(\lambda_n)$. Die Matrizen A und $\exp(A)$ haben mit den Spalten von B dieselben Eigenvektoren.

Wir berechnen $\exp(A)$ für die Hermitesche 2×2-Matrix $A = \begin{pmatrix} 1 & -3i \\ 3i & 1 \end{pmatrix}$ aus dem vorhergehenden Beispiel. Diese hatten wir mittels $B = \frac{\sqrt{2}}{2} \begin{pmatrix} -i & i \\ 1 & 1 \end{pmatrix} \in U(2)$ diagonalisiert: $\hat{A} = \bar{B}^t \circ A \circ B = \begin{pmatrix} 4 & 0 \\ 0 & -2 \end{pmatrix}$. Daraus erhalten wir:

$$\begin{aligned} \exp(A) &= B \circ \exp(\hat{A}) \circ \bar{B}^t \\ &= \frac{1}{2} \begin{pmatrix} -i & i \\ 1 & 1 \end{pmatrix} \circ \begin{pmatrix} e^4 & 0 \\ 0 & e^{-2} \end{pmatrix} \circ \begin{pmatrix} i & 1 \\ -i & 1 \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} e^4 + e^{-2} & -i\left(e^4 - e^{-2}\right) \\ i\left(e^4 - e^{-2}\right) & e^4 + e^{-2} \end{pmatrix}. \end{aligned}$$

Sesquilinearformen

Die Abbildung, die jedem Paar von Vektoren $\vec{z}, \vec{w} \in \mathbb{C}^n$ die komplexe Zahl $(\vec{z}, A \circ \vec{w})$ zuordnet, hat für Hermitesche Matrizen A ähnliche Eigenschaften wie das komplexe Skalarprodukt. Sie ist sesquilinear und Hermitesch:

$$(\vec{z}, A \circ \vec{w}) = (\bar{A}^t \circ \vec{z}, \vec{w}) = (A \circ \vec{z}, \vec{w}) = \overline{(\vec{w}, A \circ \vec{z})},$$

und wegen $(\vec{z}, A \circ \vec{z}) = \overline{(\vec{z}, A \circ \vec{z})}$ ist die entsprechende quadratische Form stets reellwertig. Die Positivität dieser Terme kann aber im allgemeinen nicht gezeigt werden, sondern ist nur von einer Unterklassse der Hermiteschen Matrizen erfüllt:

7.4.18 Definition

(a) *Eine Hermitesche Matrix $A \in \mathbb{C}^{n \times n}$ heißt* **positiv definit**, *falls $(\vec{z}, A \circ \vec{z}) > 0$ für alle $\vec{z} \in \mathbb{C}^n \setminus \{\vec{0}\}$ gilt.*

(b) *Eine symmetrische Matrix $A \in \mathbb{R}^{n \times n}$ heißt* **positiv definit**, *falls $\vec{x} \cdot (A \circ \vec{x}) > 0$ für alle $\vec{x} \in \mathbb{R}^n \setminus \{\vec{0}\}$ gilt.*

Mit Hilfe der Hauptachsentransformation kann man die positive Definitheit einer Matrix leicht erkennen.

7.4.19 Satz

Eine komplexe Hermitesche (bzw. reelle symmetrische) $n \times n$-Matrix A ist genau dann positiv definit, wenn alle Eigenwerte von A positiv sind.

Beweis

Hier werden nur komplexe Hermitesche Matrizen betrachtet; der reelle Fall ist genauso zu behandeln.

Es gibt eine unitäre Matrix $B \in U(n)$ derart, daß $\hat{A} = \bar{B}^t \circ A \circ B$ eine Diagonalmatrix ist:

$$\hat{A} = \begin{pmatrix} \lambda_1 & & \mathcal{O} \\ & \ddots & \\ \mathcal{O} & & \lambda_n \end{pmatrix}.$$

Die (nicht notwendigerweise verschiedenen) reellen Zahlen λ_ν $(\nu = 1, \ldots, n)$ sind die Eigenwerte von $\hat{A}$ und daher auch von A. Für alle $\vec{z} \in \mathbb{C}^n$ gilt mit der Setzung $\vec{w} := \bar{B}^t \circ \vec{z}$:

$$(\vec{z}, A \circ \vec{z}\,) = \bar{\vec{z}}^{\,t} \circ B \circ \hat{A} \circ \bar{B}^t \circ \vec{z} = \bar{\vec{w}}^{\,t} \circ \hat{A} \circ \vec{w} = \sum_{\nu=1}^{n} \lambda_\nu |w_\nu|^2.$$

Da $\vec{z} = \vec{0} \Leftrightarrow \vec{w} = B \circ \vec{z} = \vec{0}$, ist die Positivität der von A erzeugten quadratischen Form

$$(\vec{z}, A \circ \vec{z}\,) > 0 \text{ für alle } \vec{z} \in \mathbb{C}^n \setminus \{\vec{0}\}$$

äquivalent zur Positivität sämtlicher Eigenwerte

$$\lambda_\nu > 0, \qquad \nu = 1, \ldots, n. \qquad \Box$$

7.4.20 Folgerung

Die reelle symmetrische $n \times n$-Matrix A sei positiv definit. Dann besitzt A genau eine symmetrische positiv definite Quadratwurzel $C \in \mathbb{R}^{n \times n}$, für die also $C^2 = A$ gilt. Man schreibt auch $C =: A^{1/2}$.

Beweis

Wir diagonalisieren A mittels einer Drehung $B \in SO(n, \mathbb{R})$:

$$\hat{A} = B^t \circ A \circ B = \begin{pmatrix} \lambda_1 & & & & & & \\ & \ddots & & & & \mathcal{O} & \\ & & \lambda_1 & & & & \\ & & & \ddots & & & \\ & & & & \lambda_s & & \\ & \mathcal{O} & & & & \ddots & \\ & & & & & & \lambda_s \end{pmatrix}.$$

Dabei sind $\lambda_1, \ldots, \lambda_s$ sämtliche paarweise verschiedenen Eigenwerte von A und von $\hat{A}$; jeder Eigenwert λ_ν tritt in $\hat{A}$ entsprechend seiner Vielfachheit r_ν-mal auf; $r_1 + \ldots + r_s = n$. Wie wir soeben gezeigt haben, sind wegen der positiven Definitheit von A sämtliche Eigenwerte positiv: $\lambda_\nu > 0$. Wir setzen

$$\hat{C} := \begin{pmatrix} \sqrt{\lambda_1} & & & & & & \\ & \ddots & & & & \mathcal{O} & \\ & & \sqrt{\lambda_1} & & & & \\ & & & \ddots & & & \\ & & & & \sqrt{\lambda_s} & & \\ & \mathcal{O} & & & & \ddots & \\ & & & & & & \sqrt{\lambda_s} \end{pmatrix} \text{ und } C := B \circ \hat{C} \circ B^t;$$

es ist

$$C^2 = B \circ \hat{C} \circ B^t \circ B \circ \hat{C} \circ B^t = B \circ \hat{C}^2 \circ B^t = B \circ \hat{A} \circ B^t = A.$$

Die einander entsprechenden Eigenräume von C und von A werden jeweils von denselben Spalten von B erzeugt und stimmen deshalb auf Grund unserer Konstruktion überein:

$$V_{\lambda_\nu}(A) = V_{\sqrt{\lambda_\nu}}(C). \qquad (*)$$

Es bleibt die Eindeutigkeit der Quadratwurzel in der angegebenen Matrizenklasse nachzuweisen. Sei also $\tilde{C}$ irgendeine positiv definite symmetrische $n \times n$-Matrix mit $\tilde{C}^2 = A$. Wir wollen zeigen, daß die Gleichheit $(*)$ auch für die Eigenräume von $\tilde{C}$ erfüllt ist. Sei dazu $\vec{x} \in \mathbb{R}^n$ Eigenvektor von $\tilde{C}$ zum Eigenwert μ: $\tilde{C} \circ \vec{x} = \mu\,\vec{x}$. Da $\tilde{C}$ als positiv definit vorausgesetzt ist, haben wir $\mu > 0$. Aus $A \circ \vec{x} = \tilde{C} \circ \tilde{C} \circ \vec{x} = \tilde{C} \circ (\mu\,\vec{x}) = \mu^2\,\vec{x}$ folgt, daß μ^2 mit einem Eigenwert λ_ν von A übereinstimmen muß, wegen $\mu > 0$ ist $\mu = \sqrt{\lambda_\nu}$. Folglich sind unter $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_s}$ sämtliche paarweise verschiedenen Eigenwerte von $\tilde{C}$, und das gerade benutzte Argument zeigt auch $V_{\sqrt{\lambda_\nu}}(\tilde{C}) \subset V_{\lambda_\nu}(A)$. Da wegen Satz 7.4.11 bzw. dessen reeller Entsprechung aber die orthogonale Zerlegung

$$\mathbb{R}^n = \oplus_{\nu=1}^{s} V_{\sqrt{\lambda_\nu}}(\tilde{C}) = \oplus_{\nu=1}^{s} V_{\lambda_\nu}(A)$$

gilt, erhalten wir die Gleichheit der Eigenräume $V_{\sqrt{\lambda_\nu}}(\tilde{C}) = V_{\lambda_\nu}(A)$. Insbesondere sind genau $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_s}$ die Eigenwerte von $\tilde{C}$, und in Verbindung mit $(*)$ folgt:

$$V_{\sqrt{\lambda_\nu}}(\tilde{C}) = V_{\sqrt{\lambda_\nu}}(C).$$

Für alle

$$\vec{x} = \sum_{\nu=1}^{s} \vec{x}_\nu \in \mathbb{R}^n = \oplus_{\nu=1}^{s} V_{\sqrt{\lambda_\nu}}(\tilde{C}) = \oplus_{\nu=1}^{s} V_{\sqrt{\lambda_\nu}}(C)$$

folgt dann

$$\tilde{C} \circ \vec{x} = \sum_{\nu=1}^{s} \tilde{C} \circ \vec{x}_\nu = \sum_{\nu=1}^{s} \sqrt{\lambda_\nu}\, \vec{x}_\nu = \sum_{\nu=1}^{s} C \circ \vec{x}_\nu = C \circ \vec{x},$$

die positiv definiten Quadratwurzeln C und $\tilde{C}$ von A stimmen überein. □

Diagonalisierbarkeit unitärer Matrizen, Normalform für reell orthogonale Matrizen

Hier erhalten wir das Diagonalisierungsresultat ebenfalls durch direkten Verweis auf den entsprechenden Satz 7.4.8 für normale Matrizen und auf Satz 7.3.20.

7.4.21 Satz

Zu jeder unitären Matrix $A \in U(n)$ existiert eine unitäre Transformationsmatrix $B \in U(n)$ derart, daß $\hat{A} := \bar{B}^t \circ A \circ B$ eine Diagonalmatrix ist. Die Diagonalelemente von $\hat{A}$ haben sämtlich den Betrag 1.

Wir wollen im Folgenden noch reelle orthogonale Matrizen untersuchen. Der Fall zweier Raumdimensionen läßt sich elementar diskutieren und wird die Richtung für unser angestrebtes allgemeines reelles Resultat anzeigen.

7.4.22 Beispiele

(a) Seien $\sin_a, \cos_a$ die aus der Analysis bekannten im Bogenmaß erklärten trigonometrischen Funktionen. Für eine Drehung des $\mathbb{R}^2$

$$A = \begin{pmatrix} \cos_a(\varphi) & -\sin_a(\varphi) \\ \sin_a(\varphi) & \cos_a(\varphi) \end{pmatrix}$$

gilt

$$P(\lambda) = (\cos_a(\varphi) - \lambda)^2 + \sin_a^2(\varphi) = 1 - 2\lambda \cos_a(\varphi) + \lambda^2.$$

Die Eigenwerte von A sind $\lambda_1 = e^{i\varphi}$, $\lambda_2 = e^{-i\varphi}$; diese sind für $\varphi \in]0, 2\pi[\setminus \{\pi\}$ nicht reell und verschieden. Im Reellen können wir A also nicht diagonalisieren. Es existiert jedoch eine komplexe unitäre 2×2-Matrix B derart, daß $\hat{A} = \bar{B}^t \circ A \circ B = \begin{pmatrix} e^{i\varphi} & 0 \\ 0 & e^{-i\varphi} \end{pmatrix}$ ist.

(b) Eine Drehspiegelung des $\mathbb{R}^2$ wird durch eine Matrix

$$A = \begin{pmatrix} a & b \\ b & -a \end{pmatrix}, \quad a^2 + b^2 = 1,$$

gegeben. Es ist

$$P(\lambda) = (a - \lambda)(-a - \lambda) - b^2 = \lambda^2 - (a^2 + b^2) = \lambda^2 - 1 = (\lambda - 1)(\lambda + 1).$$

Weil die Eigenwerte von A mit $\lambda_1 = 1$ und $\lambda_2 = -1$ sämtlich reell sind, kann man aus Satz 7.4.10 die Existenz einer orthogonalen Matrix $B \in SO(2)$ folgern, so daß $\hat{A} := B^t \circ A \circ B = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ gilt. Bezüglich eines geeigneten Koordinatensystems ist die Drehspiegelung A nur eine Spiegelung an einer Koordinatenachse.

Um Satz 7.4.21 auch für den Fall reeller orthogonaler Matrizen $A \in O(n)$ nutzen zu können, gehen wir von der im Beweis von Satz 7.3.21 gemachten Beobachtung aus, daß die nichtreellen Eigenwerte in Paaren zueinander konjugierter Eigenwerte derselben (geometrischen und algebraischen) Vielfachheit auftreten: $\lambda_1, \bar{\lambda}_1, \ldots, \lambda_k, \bar{\lambda}_k \in \mathbb{C} \setminus \mathbb{R}$ mit einer geeigneten Zahl $k \leq n/2$; die Eigenwerte werden entsprechend ihrer Vielfachheit aufgeführt. Dabei können wir erreichen, daß unter $\lambda_1, \lambda_2, \ldots, \lambda_k$ nicht zwei Eigenwerte zueinander konjugiert sind: $\lambda_\nu \neq \bar{\lambda}_\mu$, $\mu, \nu = 1, \ldots, k$. An das Ende dieser Eigenwertliste schreiben wir die reellen Eigenwerte $\lambda_{2k+1}, \ldots, \lambda_n \in \{\pm 1\} \subset \mathbb{R}$. Bei den Eigenvektoren können wir eine ganz ähnliche Beobachtung machen: Sind $\vec{b}_\nu \in \mathbb{C}^n$ $(1 \leq \nu \leq k)$ paarweise orthogonale normierte Eigenvektoren zu den Eigenwerten λ_ν $(1 \leq \nu \leq k)$, so folgt aus der Reellwertigkeit der Komponenten von A:

$$(A - \lambda_\nu E) \circ \vec{b}_\nu = \vec{0} \;\; \Rightarrow \;\; \overline{(A - \lambda_\nu E)} \circ \bar{\vec{b}}_\nu = \vec{0} \;\; \Rightarrow \;\; (A - \bar{\lambda}_\nu E) \circ \bar{\vec{b}}_\nu = \vec{0};$$

die konjugierten Vektoren $\bar{\vec{b}}_\nu$ $(1 \leq \nu \leq k)$ sind also normierte Eigenvektoren zu den Eigenwerten $\bar{\lambda}_\nu$. Wir wählen nun die Transformationsmatrix zur Diagonalisierung von A gemäß

$$B = \left(\vec{b}_1, \bar{\vec{b}}_1, \ldots, \vec{b}_k, \bar{\vec{b}}_k, \vec{b}_{2k+1}, \ldots, \vec{b}_n\right) \in U(n),$$

dabei sind $\vec{b}_{2k+1}, \ldots, \vec{b}_n$ paarweise orthogonale normierte *reelle* Eigenvektoren zu den reellen Eigenwerten $\lambda_{2k+1}, \ldots, \lambda_n$. Wir haben noch zu beachten, daß aus der Orthogonalität von $\vec{b}_\nu, \vec{b}_\mu$, $1 \leq \nu \neq \mu \leq k$ auch die Orthogonalität der konjugierten Vektoren folgt: $\left(\bar{\vec{b}}_\nu, \bar{\vec{b}}_\mu\right) = \overline{\left(\vec{b}_\nu, \vec{b}_\mu\right)} = 0$. Wegen der Verschiedenheit $\lambda_\nu \neq \bar{\lambda}_\mu$ gilt für $\mu, \nu = 1, \ldots, k$ ohnehin: $\left(\vec{b}_\nu, \bar{\vec{b}}_\mu\right) = 0$. Bezüglich dieser Transformation B lautet die Diagonalmatrix

$$\hat{A} = \bar{B}^t \circ A \circ B = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & & & \cdots & 0 \\ 0 & \bar{\lambda}_1 & 0 & \cdots & & & \cdots & 0 \\ \vdots & & \ddots & & & & & \vdots \\ & \cdots & 0 & \lambda_k & 0 & 0 & \cdots & \\ & \cdots & 0 & 0 & \bar{\lambda}_k & 0 & \cdots & \\ & \cdots & 0 & 0 & 0 & \lambda_{2k+1} & \cdots & \\ \vdots & & & & & & \ddots & \vdots \\ 0 & \cdots & & & & & \cdots & \lambda_n \end{pmatrix}.$$

Wir gelangen nun zu einer reellen orthogonalen Transformationsmatrix C, indem wir jeweils bei Paaren zueinander konjugierter Eigenvektoren $\vec{b}_\nu, \bar{\vec{b}}_\nu$, $\nu = 1, \ldots, k$ zu deren normierten Real- und Imaginärteilen übergehen:

$$\vec{c}_{2\nu-1} := \frac{\sqrt{2}}{2}\left(\vec{b}_\nu + \bar{\vec{b}}_\nu\right), \qquad \vec{c}_{2\nu} := \frac{\sqrt{2}}{2i}\left(\vec{b}_\nu - \bar{\vec{b}}_\nu\right) \in \mathbb{R}^n,$$

wegen der Orthogonalität von $\vec{b}_\nu$ und $\bar{\vec{b}}_\nu$ handelt es sich dabei nicht um die Nullvektoren. Ferner ist in $\mathbb{C}^n$ auf Grund dieser Orthogonalität:

$$\begin{aligned}\left(\vec{b}_\nu + \bar{\vec{b}}_\nu, \vec{b}_\nu - \bar{\vec{b}}_\nu\right) &= \left(\vec{b}_\nu, \vec{b}_\nu\right) - \left(\vec{b}_\nu, \bar{\vec{b}}_\nu\right) + \left(\bar{\vec{b}}_\nu, \vec{b}_\nu\right) - \left(\bar{\vec{b}}_\nu, \bar{\vec{b}}_\nu\right)\\ &= \| \vec{b}_\nu \|^2 - \overline{\left(\vec{b}_\nu, \vec{b}_\nu\right)} = 0,\end{aligned}$$

also sind $\vec{c}_{2\nu-1}$ und $\vec{c}_{2\nu}$ in $\mathbb{R}^n$ zueinander orthogonal. Indem man die reellen Eigenvektoren unverändert läßt,

$$\vec{c}_{2k+1} := \vec{b}_{2k+1}, \ldots, \vec{c}_n := \vec{b}_n,$$

und nur ggfs. einen Eigenvektor durch sein Negatives ersetzt, erhält man mit

$$C := (\vec{c}_1, \ldots, \vec{c}_{2k}, \vec{c}_{2k+1}, \ldots, \vec{c}_n) \in SO(n, \mathbb{R})$$

eine reelle Transformationsmatrix. Um die Matrix der Abbildung A in den durch C gegebenen Koordinaten zu ermitteln, berechnen wir für $\nu = 1, \ldots, k$:

$$\begin{aligned}A \circ \vec{c}_{2\nu-1} &= A \circ \frac{\sqrt{2}}{2}\left(\vec{b}_\nu + \bar{\vec{b}}_\nu\right) = \frac{\sqrt{2}}{2}\left(\lambda_\nu \vec{b}_\nu + \bar{\lambda}_\nu \bar{\vec{b}}_\nu\right)\\ &= \frac{\sqrt{2}}{4}\left(\left(\lambda_\nu + \bar{\lambda}_\nu\right)\left(\vec{b}_\nu + \bar{\vec{b}}_\nu\right) + \left(\lambda_\nu - \bar{\lambda}_\nu\right)\left(\vec{b}_\nu - \bar{\vec{b}}_\nu\right)\right)\\ &= \mathrm{Re}(\lambda_\nu) \cdot \vec{c}_{2\nu-1} - \mathrm{Im}(\lambda_\nu) \cdot \vec{c}_{2\nu},\\ A \circ \vec{c}_{2\nu} &= A \circ \frac{\sqrt{2}}{2i}\left(\vec{b}_\nu - \bar{\vec{b}}_\nu\right) = \frac{\sqrt{2}}{2i}\left(\lambda_\nu \vec{b}_\nu - \bar{\lambda}_\nu \bar{\vec{b}}_\nu\right)\\ &= \frac{\sqrt{2}}{4i}\left(\left(\lambda_\nu - \bar{\lambda}_\nu\right)\left(\vec{b}_\nu + \bar{\vec{b}}_\nu\right) + \left(\lambda_\nu + \bar{\lambda}_\nu\right)\left(\vec{b}_\nu - \bar{\vec{b}}_\nu\right)\right)\\ &= \mathrm{Im}(\lambda_\nu) \cdot \vec{c}_{2\nu-1} + \mathrm{Re}(\lambda_\nu) \cdot \vec{c}_{2\nu}.\end{aligned}$$

Für die Matrix

$$\tilde{A} := C^t \circ A \circ C,$$

die die Abbildung $A : \mathbb{R}^n \to \mathbb{R}^n$ in den durch die orthogonale Matrix C gegebenen

Koordinaten beschreibt, haben wir also gefunden:

$$\tilde{A} = \begin{pmatrix} \mathrm{Re}(\lambda_1) & \mathrm{Im}(\lambda_1) & & & & & & \\ -\mathrm{Im}(\lambda_1) & \mathrm{Re}(\lambda_1) & & & & \mathcal{O} & & \\ & & \ddots & & & & & \\ & & & \mathrm{Re}(\lambda_k) & \mathrm{Im}(\lambda_k) & & & \\ & & & -\mathrm{Im}(\lambda_k) & \mathrm{Re}(\lambda_k) & & & \\ & & & & & \lambda_{2k+1} & & \\ & & \mathcal{O} & & & & \ddots & \\ & & & & & & & \lambda_n \end{pmatrix} \in \mathbb{R}^{n\times n}.$$

Diese Normalform zeigt, daß für Matrizen $A \in SO(3,\mathbb{R})$ der Eigenvektor zum Eigenwert $+1$ als Drehachse interpretiert werden kann: Indem man ein geeignetes Koordinatensystem einführt, erhält die Abbildung A in diesen Koordinaten die folgende Darstellung mit geeigneten reellen Zahlen a, b, $a^2 + b^2 = 1$:

$$\tilde{A} = \begin{pmatrix} a & -b & 0 \\ b & a & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \cos\Theta & -\sin\Theta & 0 \\ \sin\Theta & \cos\Theta & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

die Zahl $\Theta \in [0, 360[$ ist geeignet zu wählen. Die Abbildung $\tilde{A}$ läßt die dritte Komponente fest, die ersten beiden Komponenten werden durch die 2×2-Untermatrix $\begin{pmatrix} \cos\Theta & -\sin\Theta \\ \sin\Theta & \cos\Theta \end{pmatrix}$ abgebildet. Nach unseren Überlegungen am Ende von Kapitel 6.1 (insbesondere Satz 6.1.12 und Definition 6.1.15) bedeutet das eine Drehung der x_1-x_2-Ebene um den Winkel Θ, die x_3-Achse (in den ursprünglichen Koordinaten dem Eigenvektor zum Eigenwert 1 entsprechend) ist die Drehachse.

7.4.23 Beispiel

Wir bestimmen sowohl die komplexe als auch die reelle Normalform für die schon in Beispiel 7.3.22 betrachtete Matrix

$$A = \begin{pmatrix} \frac{1}{2} + \frac{\sqrt{2}}{4} & -\frac{1}{2} + \frac{\sqrt{2}}{4} & -\frac{1}{2} \\ -\frac{1}{2} + \frac{\sqrt{2}}{4} & \frac{1}{2} + \frac{\sqrt{2}}{4} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \in SO(3,\mathbb{R}).$$

Die Eigenwerte hatten wir als $\lambda_1 = \frac{\sqrt{2}}{2}(1+i)$, $\lambda_2 = \frac{\sqrt{2}}{2}(1-i)$ und $\lambda_3 = 1$ bestimmt, wir hatten auch schon die Drehachse als normierten Eigenvektor $\vec{b}_3 = \left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, 0\right)^t$ zum Eigenwert 1 berechnet. Mittels des Gaußschen Algorithmus' findet man weiter, daß $\vec{b}_1 = \left(\frac{i}{2}, \frac{i}{2}, \frac{\sqrt{2}}{2}\right)^t$ und $\vec{b}_2 = \left(-\frac{i}{2}, -\frac{i}{2}, \frac{\sqrt{2}}{2}\right)^t$ normierte Eigenvektoren zu λ_1 und

λ_2 sind. Mit
$$B := \begin{pmatrix} \frac{i}{2} & -\frac{i}{2} & \frac{\sqrt{2}}{2} \\ \frac{i}{2} & -\frac{i}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \end{pmatrix}$$
erhalten wir die komplexe Diagonalform
$$\hat{A} = \bar{B}^t \circ A \circ B = \begin{pmatrix} \frac{\sqrt{2}}{2}(1+i) & 0 & 0 \\ 0 & \frac{\sqrt{2}}{2}(1-i) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$
Um die reelle Normalform zu erhalten, ersetzen wir, wie oben erläutert, die ersten beiden Spalten von B durch deren normierte Real- und Imaginärteile:
$$C := \begin{pmatrix} 0 & \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ 1 & 0 & 0 \end{pmatrix} \in SO(3).$$
Um mit C eine Drehung, d. h. eine orthogonale Abbildung mit Determinante 1, zu erhalten, haben wir gleichzeitig die letzte Spalte durch ihr Negatives ersetzt. Damit berechnen wir
$$\tilde{A} := C^t \circ A \circ C = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$
Bei der Abbildung A handelt es sich also um eine Drehung um 315 Grad um die Drehachse $\left(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, 0\right)^t$.

Literaturverzeichnis

[1] Budó, Ágoston, *Theoretische Mechanik*, 10. Auflage, Deutscher Verlag der Wissenschaften: Berlin 1980.

[2] Ebbinghaus, Heinz-Dieter, H. Hermes, F. Hirzebruch, M. Koecher, K. Mainzer, A. Prestel, R. Remmert, *Zahlen*, Springer-Verlag: Berlin etc. 1983.

[3] Fischer, Gerd, *Lineare Algebra*, 6. Auflage, Vieweg: Braunschweig, Wiesbaden 1980.

[4] Forster, Otto, *Analysis 1*, 3. Auflage, Vieweg: Braunschweig, Wiesbaden 1980.

[5] Grauert, Hans, Die Axiome der Elementargeometrie und die griechischen Theoreme, *Math. Semesterber.* **44** (1997), 19-36.

[6] Grauert, Hans, I. Lieb, *Differential- und Integralrechnung I*, 4. Auflage, Springer-Verlag: Berlin etc. 1976.

[7] Klingenberg, Wilhelm, *Lineare Algebra und Geometrie*, Springer-Verlag: Berlin etc. 1984.

[8] Koecher, Max, *Lineare Algebra und analytische Geometrie*, Springer-Verlag: Berlin etc. 1983.

[9] Kowalsky, Hans-Joachim, *Lineare Algebra*, 9. Auflage, de Gruyter: Berlin, New York 1979.

[10] Pickert, Günter, *Analytische Geometrie*, 4. Auflage, Akademische Verlagsgesellschaft Geest & Portig: Leipzig 1961.

[11] Rudin, Walter, *Analysis*, Oldenbourg Verlag: München, Wien 1998.

[12] Sperner, Emanuel, *Einführung in die Analytische Geometrie und Algebra*, 7. Auflage, Vandenhoeck & Ruprecht: Göttingen 1969.

[13] Tietz, Horst, *Lineare Geometrie*, Verlag Aschendorff: Münster 1967.

[14] van der Waerden, B. L., *Erwachende Wissenschaft*, 2. Auflage, Birkhäuser Verlag: Basel, Stuttgart 1966.

Symbolverzeichnis

$:=$	Die linke Seite der Gleichung wird durch die rechte Seite definiert.
$\stackrel{\text{def}}{=}$	Gleichheit auf Grund einer Definition
$\equiv$	identisch gleich; $f(x) \equiv 0$ bedeutet, daß $f(x) = 0$ für *alle* x gilt
$\Rightarrow$	Folgerung: Wenn die linke Aussage gilt, dann auch die rechte.
$\Leftrightarrow$	Äquivalenz zweier Aussagen
$\emptyset$	leere Menge
$x \in M$	x ist Element der Menge M
$x \notin M$	x ist nicht Element der Menge M
$M \subset N$	Teilmenge: $x \in M \Rightarrow x \in N$.
$M \supset N$	Obermenge: $N \subset M$
$M = N$	Gleichheit von Mengen: $M \subset N$ und $N \subset M$
$M \cap N$	$= \{x :\ x \in M$ und $x \in N\}$, Durchschnitt
$M \cup N$	$= \{x :\ x \in M$ oder $x \in N\}$, Vereinigung
$M \setminus N$	$= \{x \in M :\ x \notin N\}$, Differenzmenge
$M \times N$	$= \{(x, y) :\ x \in M, y \in N\}$, kartesisches Produkt
$\mathbb{N}$	Menge der natürlichen Zahlen $\{1, 2, 3, \ldots\}$
$\mathbb{N}_0$	$\mathbb{N} \cup \{0\} = \{0, 1, 2, 3, \ldots\}$
$\mathbb{Z}$	Menge der ganzen Zahlen
$\mathbb{R}$	Menge der reellen Zahlen
$\mathbb{R}^+$	Menge der positiven reellen Zahlen $\{a \in \mathbb{R} :\ a > 0\}$
$\mathbb{R}_0^+$	Menge der nichtnegativen reellen Zahlen $\{a \in \mathbb{R} :\ a \geq 0\}$
$\mathbb{R}^*$	$= \mathbb{R} \setminus \{0\}$
$\max(a, b)$	die größere der reellen Zahlen a, b
$\min(a, b)$	die kleinere der reellen Zahlen a, b

$\sup$	Supremum
$\inf$	Infimum
$\sum$	Summenzeichen
$\prod$	Produktzeichen
$\mathbb{R}^n$	n-dimensionaler reeller Zahlenraum
$\vec{a}$	$= (a_1, \dots, a_n) \in \mathbb{R}^n$, n-tupel reeller Zahlen
$\vec{0}$	$= (0, \dots, 0) \in \mathbb{R}^n$
$\vec{e}_j$	Einheitsvektoren in $\mathbb{R}^n$
$V, V_1, V_2, \dots$	allgemeine reelle Vektorräume
x, y	Elemente von V
$\left.\begin{array}{l} \text{Span}(x_1, \dots, x_n) \\ \mathbb{R} \cdot (x_1, \dots, x_n) \\ <x_1, \dots, x_n> \\ [x_1, \dots, x_n] \end{array}\right\}$	von $x_1, \dots, x_n$ aufgespannter Untervektorraum
$\dim$	Dimension eines Vektorraums
E	anschauliche Ebene
R	anschaulicher Raum
O	Basispunkt, Aufpunkt in E bzw. R
$x = (P, Q)$	Pfeile, Pfeilvektoren in E bzw. R
(P, P)	Nullvektor in P
$(P, Q) \parallel (P', Q')$	Die Pfeile (P, Q) und (P', Q') sind voll parallel.
X	Menge aller Pfeile in E bzw. R
V	$= \{x = (O, P) : P \in E\}$, Vektorraum der Pfeile in O
V^3	$= \{x = (O, P) : P \in R\}$, Vektorraum der Pfeile in O
$\mathcal{O} = (O, O)$	Nullvektor in V bzw. V^3
$F : M_1 \to M_2$	Abbildung
$id_M, id : M \to M$	identische Abbildung
$F : M_1 \twoheadrightarrow M_2$	surjektive Abbildung
$F : M_1 \hookrightarrow M_2$	injektive Abbildung
$F : M_1 \leftrightarrow M_2$	bijektive Abbildung
F^{-1}	Umkehrabbildung einer bijektiven Abbildung
$F \circ G$	Hintereinanderschaltung, Komposition, Zusammensetzung von Abbildungen

$F^{-1}(N)$	$= \{x \in M : F(x) \in N\}$, Urbildmenge
$\mathrm{im}(F) = F(M)$	Bildmenge
$\mathrm{rg}(F)$	$= \dim \mathrm{im}(F)$, Rang einer linearen Abbildung
$\ker(F)$	Kern einer linearen Abbildung
$\mathrm{crg}(F)$	$= \dim \ker(F)$, Corang (Defekt) einer linearen Abbildung
$V_1 \approx V_2$	V_1 und V_2 sind isomorph.
$V_1 \cong V_2, V_1 = V_2$	V_1 und V_2 sind kanonisch isomorph.
$F : V_1 \xrightarrow{\sim} V_2$	F ist ein Isomorphismus.
A, B, C	Matrizen
$A^{m,n} = (a_{ij})_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}}$	$m \times n$-Matrix
$\mathbb{R}^{m \times n}$	Menge aller $m \times n$-Matrizen
$E = E^{n,n}$; I	Einheitsmatrix
δ_{ij}	Kroneckersymbol, Komponenten der Einheitsmatrix
$\mathcal{O} = \mathcal{O}^{m,n}$	Nullmatrix
A^t	zu A transponierte Matrix
$\mathrm{Rang}\,(A)$	Rang einer Matrix
$A \circ \vec{c}$	Produkt einer $m \times n$-Matrix und einem Spaltenvektor aus $\mathbb{R}^n$
$A + B$	Matrixaddition
$c \cdot A$	Skalarmultiplikation für Matrizen
$A \circ B$	Matrixprodukt
$GL(n, \mathbb{R})$	allgemeine lineare Gruppe vom Rang n über $\mathbb{R}$, Gruppe der invertierbaren $n \times n$-Matrizen
$SL(n, \mathbb{R})$	spezielle lineare Gruppe
$O(n, \mathbb{R})$	orthogonale Gruppe
$SO(n, \mathbb{R})$	spezielle orthogonale Gruppe
$M,\ N$	in Abschnitt 3.7: allgemeine Ebenen
codim	Codimension von allgemeinen Ebenen
$V_1 \oplus V_2$	direkte Summe der Vektorräume V_1 und V_2
$V_1 + V_2$	von V_1, V_2 aufgespannter Untervektorraum
$F_1 \oplus F_2$	direkte Summe der linearen Abbildungen F_1 und F_2
V^*	Dualraum von V

V^{**}	Bidualraum
$F^* : V_2^* \to V_1^*$	duale Abbildung zu $F : V_1 \to V_2$,
$\delta(i_1, \ldots, i_n)$	allgemeines Kroneckersymbol
$\det(A), \det(F)$	Determinante
$A_{k\ell}$	Streichungsmatrix
$d_{k\ell}$	(k, ℓ)-ter Cofaktor
ad (A)	Adjunkte zu A
D_Θ, $\hat{D}_\Theta$	Drehungen der anschaulichen Ebene bzw. der Pfeile
$A_\alpha(\Theta)$	Drehungen in den Koordinaten bzgl. der Karte α
$\vec{x} \cdot \vec{y}$	kanonisches Skalarprodukt in $\mathbb{R}^n$
(x, y)	Skalarprodukt in der anschaulichen Ebene
$\| \,.\, \|$	Länge bzw. Norm eines Vektors
$\angle(y, x)$	Winkel zwischen $y, x \in V$
$x \perp y$	x, y sind orthogonal
$V^\perp$	orthogonales Komplement des Untervektorraums V
$\mathbb{C}$	Menge der komplexen Zahlen
$\mathbb{C}^n$	n-dimensionaler komplexer Zahlenraum
$\dim_\mathbb{C} V$	Dimension eines Vektorraums über $\mathbb{C}$
$\dim_\mathbb{R} V$	Dimension eines Vektorraums über $\mathbb{R}$
Re(z)	Realteil der komplexen Zahl z
Im(z)	Imaginärteil der komplexen Zahl z
$\bar{z}$	konjugiert komplexe Zahl
$(\vec{z}, \vec{w})$	kanonisches Skalarprodukt in $\mathbb{C}^n$
$\bar{A}^t$	konjugiert transponierte Matrix
$GL(n, \mathbb{C})$	allgemeine lineare Gruppe vom Rang n über $\mathbb{C}$
$U(n)$	unitäre Gruppe
Sp(A)	Spur der Matrix A
$P_A(\lambda), P_F(\lambda)$	charakteristisches Polynom
λ_ν	Eigenwerte
V_λ	Eigenraum
$\exp(A)$	Exponentialfunktion für Matrizen
$A^{1/2}$	Quadratwurzel positiv definiter Matrizen

Sachverzeichnis