



Lineare Algebra 1 & 2

Hans Joachim Pflug (pflug@itc.rwth-aachen.de)

grundlegend überarbeitet von

Matthias Grajewski (grajewski@fh-aachen.de)

Inge Gutheil (i.gutheil@fz-juelich.de)

Benno Wienke (b.wienke@itc.rwth-aachen.de)

01. August 2019

Inhaltsverzeichnis

1	Motivation und Vorbereitung	7
1.1	Inhalt und Anwendungen der Linearen Algebra	7
1.2	Mathematisches Arbeiten und Problemlösen	10
1.3	Vektoren im \mathbb{R}^n	12
1.4	Addition und Multiplikation im \mathbb{R}^n	14
1.5	Lineare Gleichungssysteme und Matrizen	16
1.6	Lösung eines linearen Gleichungssystems	19
1.6.1	Das Gaußsche Eliminationsverfahren	19
1.6.2	Unter- und überbestimmte Gleichungssysteme	23
2	Analytische Geometrie	27
2.1	Skalarprodukt und Norm	27
2.1.1	Einheitsvektoren	32
2.1.2	Orthogonale (senkrechte) Vektoren	33
2.1.3	Winkel zwischen Vektoren	35
2.1.4	Das Vektorprodukt (Kreuzprodukt)	39
2.2	Geraden und Ebenen	42
2.2.1	Definition und grundlegende Eigenschaften	42
2.2.2	Umrechnen zwischen verschiedenen Darstellungsformen	49
2.2.3	Lagebeziehungen von Geraden und Ebenen	51
2.2.4	Schnittmengen zwischen Geraden und Ebenen in \mathbb{R}^2 und \mathbb{R}^3	54
2.2.5	Abstandsbestimmung in \mathbb{R}^2 und \mathbb{R}^3	57
2.3	Die Determinante im \mathbb{R}^2 und \mathbb{R}^3	61
2.3.1	Berechnung und geometrische Deutung	61
2.3.2	Lineare 3×3 -Gleichungssysteme	64
3	Algebraische Strukturen	67
3.1	Gruppen	67
3.1.1	Grundlagen	67
3.1.2	Vertiefung: Endliche Gruppen und Restklassen	70
3.2	Körper	72

3.3	Vektorräume	74
3.4	Lineare Unabhängigkeit, Basis, Dimension	78
3.4.1	Lineare Unabhängigkeit	78
3.4.2	Nachweis linearer Unabhängigkeit	81
3.4.3	Basis und Dimension	84
3.4.4	Exkurs: Nicht endlich erzeugte Vektorräume	89
3.4.5	Exkurs: Hyperebenen im \mathbb{R}^n	91
3.5	Polynome	95
3.6	Skalarprodukt, euklidische und unitäre Räume	102
3.7	Orthogonalität in unitären Vektorräumen	108
3.8	Das Verfahren von Gram-Schmidt und Anwendungen	115
4	Lineare Abbildungen	123
4.1	Vorbereitung	123
4.2	Grundlegende Eigenschaften linearer Abbildungen	126
4.3	Matrizen und lineare Abbildungen	135
4.4	Abbildungsverkettung und Matrizenmultiplikation	145
4.5	Koordinatentransformationen	153
5	Determinanten	161
5.1	Motivation und Einführung	161
5.2	Vorbereitung: Elementarmatrizen	163
5.3	Eigenschaften der Determinante	166
5.4	Verfahren zur Berechnung der Determinante	169
5.4.1	Leibniz-Formel	169
5.4.2	Laplacescher Entwicklungssatz	171
5.4.3	Gauß-Algorithmus	172
5.4.4	Determinantenberechnung in der Praxis	173
5.5	Exkurs: Invertierung von Matrizen mittels Unterdeterminanten	175
6	Lineare Gleichungssysteme	179
6.1	Lösbarkeit eines linearen Gleichungssystems	179
6.2	Das Gaußsche Eliminationsverfahren reloaded	184
6.3	Die Cramersche Regel	188
6.4	Überbestimmte lineare Gleichungssysteme	190
6.5	Unterbestimmte lineare Gleichungssysteme	196
7	Geometrie linearer Abbildungen	199
7.1	Orthogonale Abbildungen und Matrizen	199
7.2	Exkurs: QR-Zerlegung und Anwendungen	204
7.3	Eigenwerte und Eigenvektoren	206
7.4	Diagonalisierung linearer Abbildungen	213
7.5	Definitheit und Skalarprodukte	220

<i>INHALTSVERZEICHNIS</i>	5
Index	226
Literatur	231

Kapitel 1

Motivation und Vorbereitung

1.1 Inhalt und Anwendungen der Linearen Algebra

Die Lineare Algebra ist keine eigenständige mathematische Disziplin wie etwa die Analysis oder die Geometrie und ist doch ein unverzichtbarer Bestandteil jeden Studiums, in dem mathematische Inhalte eine wesentliche Rolle spielen. Dies liegt daran, dass sie grundlegende mathematische Techniken und Begriffe bereitstellt, die die gesamte Mathematik durchsetzen und auf die man deswegen im weiteren Studium immer wieder zurückgreifen wird. Drei wesentliche Themenbereiche prägen die Lineare Algebra:

1. lineare Gleichungssysteme
2. analytische Geometrie
3. algebraischen Strukturen

Wir werden zur Motivation drei Beispiele skizzieren, ohne dabei zu viel Wert auf mathematische Strenge zu legen.

Beispiel 1.1: Zur Eisenerzeugung wird im Hochofen Schrott eingeschmolzen, um die Kosten der Verhüttung zu senken und vorhandenes Material wiederzuverwenden. Bei Schrott handelt es sich nicht um chemisch reines Eisen, weil u. a. Metalle wie Kupfer oder Zink enthalten sind. Zudem wird in der Praxis selten reines Eisenerz erzeugt, sondern gleich eine Legierung hergestellt. Beim Betrieb einer Eisenhütte ergibt sich also die Frage, welche Schrottsorten man in welcher Menge beimischen darf, damit am Ende die gewünschte Legierung entsteht. Wir gehen vereinfachend von drei Schrottsorten S_1 , S_2 und S_3 aus, die nach Massenanteil wie folgt zusammengesetzt sind:

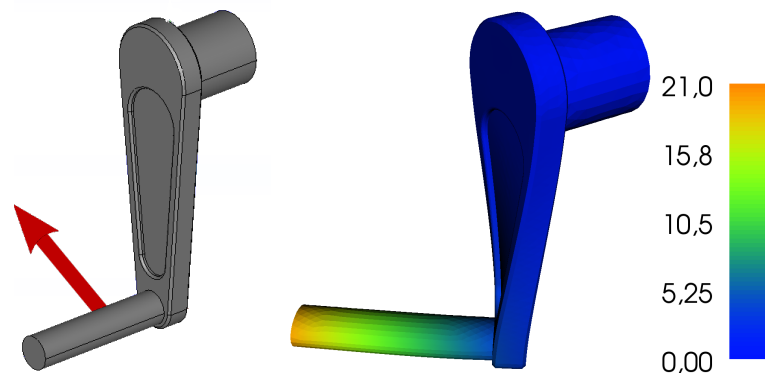


Abbildung 1.1: Handkurbel mit angreifender Kraft (links; als Pfeil dargestellt); resultierende Verformung (rechts; Verformung zur Verdeutlichung überhöht dargestellt)

	S_1	S_2	S_3
Eisen	0,8	0,8	0,8
Kupfer	0	0,2	0,1
Zink	0,2	0	0,1

Das Endprodukt soll 80% Eisen, 12% Kupfer und 8% Zink enthalten. Wir bezeichnen die Massenanteile der drei Legierungen mit a, b und c und stellen die Massenbilanzen für alle drei Elemente getrennt auf. Für Eisen erhält man $0,8a + 0,8b + 0,8c = 0,8$ und analog für Kupfer $0a + 0,2b + 0,1c = 0,12$ bzw. für Zink $0,2a + 0b + 0,1c = 0,08$. Da alle drei Gleichungen zugleich gelten müssen, werden wir auf ein lineares Gleichungssystem der Form

$$\begin{aligned}
 0,8a + 0,8b + 0,8c &= 0,8 \\
 0a + 0,2b + 0,1c &= 0,12 \\
 0,2a + 0b + 0,1c &= 0,08
 \end{aligned} \tag{1.1}$$

geführt. Man überprüft leicht, dass die Werte $a = 0, b = 0,2$ und $c = 0,8$ das System lösen, indem man sie in die Formel (1.1) einsetzt und sich von der Gleichheit der beiden Seiten überzeugt.

Beispiel 1.2: Ein wesentliches Anwendungsgebiet der Mathematik ist die Numerische Simulation, d. h. die Verhaltensvorhersage eines physikalischen Systems mithilfe mathematischer Methoden. Ein physikalisches System kann ein Auto, die Erde, ein Atom oder ein mechanisches Bauteil wie die in Abb. 1.1 gezeigte Kurbel sein. Soll in einem Industrieunternehmen eine solche Kurbel produziert werden, wird man nach den Regeln und Verfahrensweisen des Maschinenbaus die Kurbel entwerfen, dann einen Prototypen bauen und diesen testen. Bei der Kurbel wird es wahrscheinlich

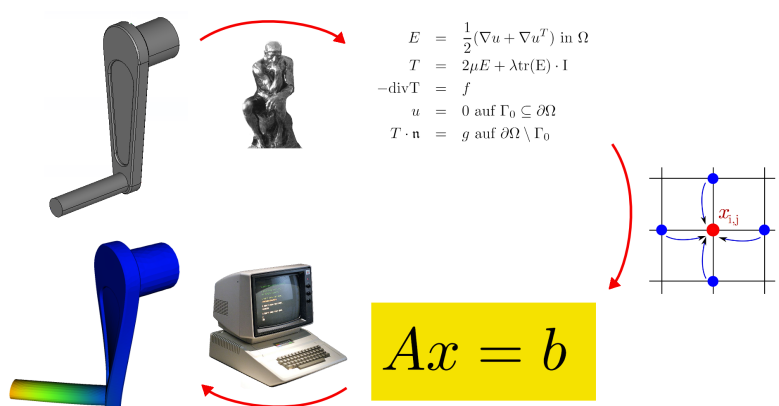


Abbildung 1.2: vereinfachtes Schema einer Numerischen Simulation

um die Frage gehen, ob sie allen Kräften standhält, die im Betrieb erwartet werden. Besteht sie alle Tests, kann sie in Serie gefertigt werden, ansonsten wird der Entwurf solange modifiziert und erneut getestet, bis er alle Anforderungen erfüllt. Mithilfe der Numerischen Simulation lassen sich die Tests der Prototypen im Entwicklungsprozess durch Berechnungen ersetzen. Dies spart Zeit und Geld. Diese Ersparnis mag bei der Kurbel unbedeutend sein, bei der Entwicklung eines Verkehrsflugzeuges ist sie es nicht. Numerische Simulation hat sich als unverzichtbares Werkzeug in der Produktentwicklung erwiesen.

Eine numerische Simulation, z. B. mit der Methode der finiten Elemente, ist das Ergebnis des Zusammenwirkens verschiedenster mathematischer Disziplinen wie Analysis, Numerik, Funktionalanalysis und Bereichen der Informatik sowie von Erkenntnissen aus Physik und Maschinenbau. Am Ende aller Überlegungen steht sehr häufig ein lineares Gleichungssystem der Form $Ax = b$, das es zu lösen gilt (Abb. 1.2). Umfasste das Gleichungssystem in Beispiel 1.1 3 Bedingungen an 3 Unbekannte, sind bei Numerischer Simulation 1000000 Bedingungen an 1000000 Unbekannte nicht unüblich. Eine Handrechnung zur Lösung derartiger Systeme verbietet sich. Mit der effizienten Lösung mithilfe von Computern befasst sich die Numerische Lineare Algebra, die auf den Ergebnissen der Linearen Algebra aufbaut. Numerische Simulation setzt also ein tiefes Verständnis linearer Gleichungssysteme notwendig voraus.

Beispiel 1.3: Der Begriff der Gruppe als einfache algebraische Struktur ist aus der Vorlesung „Mathematische Grundlagen“ bekannt. Es handelt sich dabei um das Paar einer Menge M zusammen mit einer auf M definierten

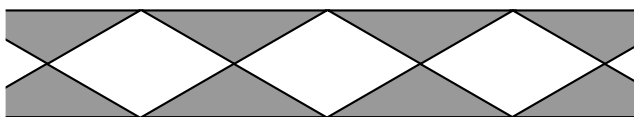


Abbildung 1.3: Beispiel eines Frieses

Verknüpfung \circ , das gewisse algebraische Bedingungen wie z. B. Assoziativität erfüllen muss. Wir betrachten ein Fries, also eine in einer Raumrichtung unendlich ausgedehnte ebene Struktur, die sich dieser Richtung wiederholt (Abb. 1.3). Es sei M die Menge aller Starrkörperbewegungen (alle längen- und winkelerhaltenden bijektiven Abbildungen der Ebene in sich), die das Fries wieder in sich überführen, als Verknüpfung \circ auf M wählen wir die Hintereinanderausführung von Abbildungen. Dann bildet (M, \circ) eine Gruppe, die sog. *Automorphismengruppe* des Frieses, kurz Friesgruppe. Man kann zeigen, dass zu jedem Fries-Typ genau eine Friesgruppe existiert und umgekehrt und dass es im Wesentlichen 7 Friesgruppen gibt. Es gibt also (bis auf Designmerkmale wie Farbe, Größe etc.) genau 7 verschiedene Friesse.

Analoge Betrachtungen lassen sich für unendliche periodische räumliche Objekte anstellen, die als mathematische Modelle für den atomaren Aufbau von Kristallen dienen. Fedorov¹ konnte 1891 zeigen, dass genau 230 sogenannter Raumgruppen existieren und damit den Aufbau von Kristallen klassifizieren. Dies war eine der ersten bedeutenden außermathematischen Anwendungen der Gruppentheorie und gleichzeitig ein großer Durchbruch in der Kristallographie.

1.2 Mathematisches Arbeiten und Problemlösen

Kern allen mathematischen Arbeitens ist die systematische Beschäftigung mit mathematischen Ideen, Problemen und ihrer Lösung. Konkrete Rechnungen sind dagegen häufig zweitrangig. Ein mathematisches Problem gilt als gelöst, wenn man eine entsprechende Aussage formulieren und streng beweisen kann bzw. ein geeignetes Gegenbeispiel angibt. Folgerichtig wird in der Vorlesung Lineare Algebra (und nicht nur da!) der Formulierung mathematischer Aussagen und ihrem Beweis breiter Raum gegeben. Dies unterscheidet sich bisweilen stark vom mathematischen Vorgehen außerhalb der Hochschulen. Dieser Paradigmenwechsel zu Beginn des Studiums ist der Erfahrung nach eine der wesentlichen Hürden für die Studierenden. Es ist normal, dass die Lösung eines mathematischen Problems nicht auf der Hand liegt und sollte nicht entmutigen. Wir skizzieren im Folgenden

¹Evgraf S. Fedorov (1853–1919); russischer Kristallograph und Mineraloge

sehr kurz einige Techniken zur Lösung mathematischer Probleme. Für weitergehende Beschäftigung mit diesem Thema verweisen wir auf das sehr lesenswerte Buch von Grieser [4] und die klassische Abhandlung von Pólya [6].

Wir gehen davon aus, dass Ihnen eine mathematische Aussage vorgelegt wird und Sie diese beweisen oder widerlegen sollen. Dies ist eine Standardanforderung im Studium und zugleich eine große Vereinfachung, denn in der mathematischen Praxis ist es häufig viel schwieriger, die richtigen (d. h. zielführenden) Fragen zu stellen als die richtigen Antworten zu finden. Die folgenden Hinweise können Ihnen bei Ihrer Beschäftigung mit mathematischen Problemen helfen.

1. Versuchen Sie, das Problem zu verstehen!

- Sind Ihnen alle verwendeten mathematischen Begriffe in der Formulierung klar? Könnten Sie einem Kommilitonen alle verwendeten mathematischen Begriffe erklären? Wenn nicht, wiederholen Sie zunächst die entsprechenden Inhalte der Vorlesung.
- Sollte die Aussage eine Formel sein, versuchen Sie, dieselbe Aussage für Sie selbst als Text zu formulieren und umgekehrt.
- Betrachten Sie Beispiele und Spezialfälle. Kann man an Beispielen bereits erkennen, warum die Aussage wahr oder falsch sein sollte?
- Fertigen Sie, wann immer das Problem es ermöglicht, eine Skizze an!

2. Kennen Sie ähnliche Probleme?

- Kennen Sie bei vergleichbaren Aussagen sogar Beweis oder Gegenbeispiel?
- Kann man die gegebene Aussage auf Bekanntes zurückführen?
- Wenn nicht, was genau ist anders?

3. Vorwärtsarbeiten

- Jeder Beweisversuch lebt von den Voraussetzungen. Was lässt sich mit den gegebenen Voraussetzungen anfangen? Welche Aussagen lassen sich damit zeigen?

4. Rückwärtsarbeiten

- Unter welchen zusätzlichen Voraussetzungen könnten Sie denn die gewünschte Aussage beweisen?
- Wie kann man sich im zweiten Schritt von diesen zusätzlichen Voraussetzungen befreien?

5. Zwischenziele formulieren

- Bei komplexeren Sachverhalten kann es helfen, dass man das Gesamtproblem in Teilprobleme zerlegt, die man dann getrennt bearbeitet.
- Für jede Etappe lassen sich die oben skizzierten Problemlösestrategien verwenden.

6. Problemlösestrategien kombinieren und ausprobieren

- In vielen Fällen bringt erst eine Kombination der obigen Strategien den Erfolg.
- Wenn eine Strategie nicht weiterführt, muss man eine andere ausprobieren. Mathematik bedeutet manchmal auch hartnäckiges Herumprobieren!

7. Zum Schluss: richtig Aufschreiben!

- Schreiben Sie ihre Argumentation detailliert und nachvollziehbar auf. Oftmals wird umgekehrt die Argumentation erst bei ihrer Formulierung wirklich klar.
- Vermeiden Sie umständliche Prosa, sondern bedienen Sie sich mathematischer Formeln und Formulierungen.
- Kontrollieren Sie zum Schluss: Ist Ihre Beweisführung lückenlos? Ist das Gegenbeispiel wirklich eins?
- Für weiterführende Literatur zur Formulierung mathematischer Gedanken verweisen wir auf Beutelspacher [1].

1.3 Vektoren im \mathbb{R}^n

Ein epochaler Fortschritt in der Geometrie wurde mit der Einführung von Koordinaten durch René Descartes² erzielt (ihm zu Ehren spricht man vom „kartesischen Koordinatensystem“), weil hierdurch die seit alters her bestehende Kluft zwischen Geometrie und Arithmetik beseitigt werden konnte. Nun konnte man Aussagen der Geometrie mit Methoden der Arithmetik untersuchen und umgekehrt.

Definition 1.4: 1. Für $n \in \mathbb{N}$ sei x das n -Tupel

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = (x_i)_{i=1}^n$$

²1596–1650; franz. Mathematiker, Philosoph, Naturwissenschaftler

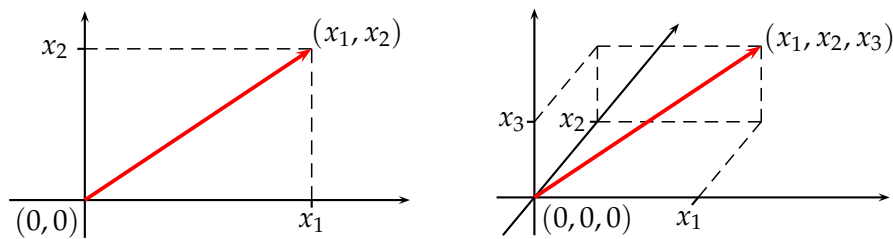


Abbildung 1.4: Beschreibung von Punkten in der Ebene (links) und im Raum (rechts) mithilfe eines Koordinatensystems

mit $x_i \in \mathbb{R}$ für $1 \leq i \leq n$. Dann heißt x *Vektor*, die Zahl x_i die *i-te Koordinate*.

2. Zu x wie oben sei

$$x^T := (x_1, \dots, x_n)$$

der *transponierte Vektor*. Weiter definieren wir $(x^T)^T := x$.

3. Zwei Vektoren x, y sind *gleich*, wenn alle ihre Koordinaten gleich sind.

4. Der Zahlraum

$$\mathbb{R}^n = \{(x_1, \dots, x_n)^T \mid x_1, \dots, x_n \in \mathbb{R}\}$$

sei die Menge aller geordneten n -Tupel reeller Zahlen.

Wir betrachten den Spezialfall $n = 2$. Durch Wahl eines Koordinatensystems gelingt es, jeden Punkt der Ebene durch ein Zahlenpaar (x_1, x_2) zu parametrisieren (Abb. 1.4 links). Umgekehrt lässt sich jedes Zahlenpaar (x_1, x_2) als Punkt in der Ebene veranschaulichen. Daher kann man die Ebene mit dem Zahlraum \mathbb{R}^2 identifizieren. Analog lässt sich durch die Wahl eines Koordinatensystems jeder Punkt des Raumes mit einem Zahlentripel $(x_1, x_2, x_3) \in \mathbb{R}^3$ identifizieren (Abb. 1.4 rechts).

Der allgemeine Zahlraum \mathbb{R}^n eignet sich auch zur Beschreibung nicht-geometrischer Zusammenhänge.

Beispiel 1.5: In den Wirtschaftswissenschaften verwendet man Vektoren, um die Gesamtproduktion von Waren oder Dienstleistungen zu repräsentieren. Als Maßstab wird dann der Preis (z. B. in 10^9 \$) verwendet. Wir betrachten 5 Wirtschaftszweige:

1.	Schwerindustrie		2.	Elektronik und Computer
3.	Automobilindustrie		4.	Landwirtschaft
5.	Finanzsektor			

und modellieren die wirtschaftlichen Vorgänge in \mathbb{R}^5 . Ordnen wir die 5 Koordinaten in der obigen Reihenfolge an, bedeutet der Punkt

$$(44, 182, 404, 36, 230)^T,$$

dass die Schwerindustrie Waren in einem Wert von 44 Milliarden Dollar produziert und der Finanzsektor 230 Milliarden Dollar umgesetzt hat.

Wir kehren nun zur geometrischen Deutung von Vektoren im \mathbb{R}^2 bzw. \mathbb{R}^3 zurück. Die Interpretation eines Vektors als Beschreibung eines Punktes wie oben erfordert, dass man einen Koordinatenursprung festgelegt hat, weil die Koordinaten eines Punktes relativ zu diesem angegeben werden. Da aber in der Ebene bzw. im Raum alle Punkte gleichberechtigt sind, ist die Festlegung eines Koordinatenursprungs ein Akt der Willkür und daher geometrisch fragwürdig. Wir deuten vor diesem Hintergrund ein n -Tupel alternativ als *Verschiebung*. Bei einer Verschiebung handelt es sich zunächst um eine Abbildung in der Ebene bzw. im Raum. Sie ist dann eindeutig festgelegt, wenn man von irgendeinem Punkt P den Bildpunkt Q kennt. Man kann also eine solche Verschiebung durch die gerichtete Strecke \overrightarrow{PQ} eindeutig repräsentieren. Allerdings ist die Wahl von P nicht eindeutig; ein anderer Punkt P' repräsentiert mit dessen Bildpunkt Q' dieselbe Verschiebung ebenso. Man kann zeigen, dass alle gerichteten Strecken, die dieselbe Verschiebung repräsentieren, eine Äquivalenzklasse bilden; diese nennt man in der euklidischen Geometrie *Vektor*.³ Ein solcher (Verschiebungs-)Vektor in diesem geometrischen Sinn hat eine bestimmte Richtung und Länge, aber keine bestimmte Lage, benötigt also auch nicht die Festlegung eines Koordinatenursprungs. Geht man nun von der Existenz eines Koordinatenursprungs 0 aus und wählt $P = 0$, so wählt man damit aus der Äquivalenzklasse „Vektor“ die entsprechende gerichtete Strecke aus, die auf einen festen Punkt Q zeigt. Dann spricht man vom *Ortsvektor* und schreibt $\overrightarrow{0Q}$ oder einfach q . Nimmt man jetzt noch die Existenz ausgezeichneter Richtungen an, die den Koordinatenachsen entsprechen, wie man es ja durch die Beschreibung der Ebene durch \mathbb{R}^2 und des Raumes durch \mathbb{R}^3 implizit tut, dann wird Q über seine Koordinaten durch ein n -Tupel repräsentiert, also durch einen Vektor im Sinn der Definition 1.4. Dieses repräsentiert dann aber zugleich die Verschiebung und damit auch den Verschiebungsvektor. Ob ein n -Tupel als Ortsvektor oder Verschiebungsvektor gedeutet werden soll, erschließt sich nur aus dem Zusammenhang.

1.4 Addition und Multiplikation im \mathbb{R}^n

Wir betrachten die Vektoren $a = (2, 1)^T$ und $b = (-1, 1)^T$ und deuten sie als Verschiebungen in der Ebene. Verschiebt man zunächst längs a und

³Der Terminus „Vektor“ ist vom lateinischen „vehere“ = transportieren abgeleitet

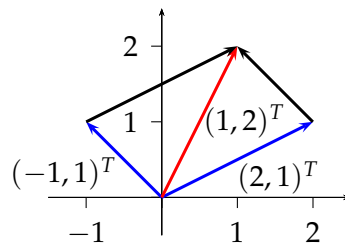


Abbildung 1.5: Addition von Vektoren

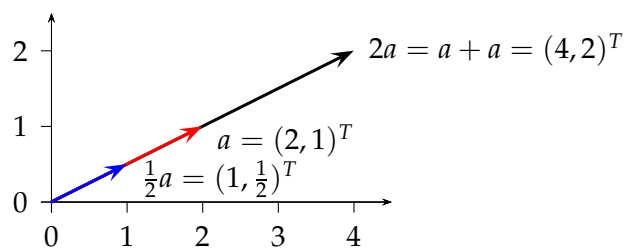


Abbildung 1.6: Multiplikation eines Vektors mit einem Skalar

dann längs b , entspricht das einer Verschiebung um den Vektor $c = (1, 2)^T$ (Abb. 1.5), der komponentenweise der Summe der Komponenten von a und b entspricht. Ein analoger Zusammenhang gilt auch im \mathbb{R}^3 . Weil man also die Verschiebung durch den Vektor b der Verschiebung um den Vektor a hinzufügt, liegt es nahe, folgende Vektoraddition zu definieren.

Definition 1.6: Seien $a = (a_1, \dots, a_n)^T, b = (b_1, \dots, b_n)^T \in \mathbb{R}^n$. Dann ist

$$a + b := \begin{pmatrix} a_1 + b_1 \\ \vdots \\ a_n + b_n \end{pmatrix}.$$

Wird der Vektor $a = (2, 1)^T$ in der Ebene um den Faktor 2 verlängert, erhält man den Vektor $(4, 2)^T$ (Abb. 1.6). Dies entspricht genau einer Skalierung der Komponenten von a mit dem Faktor 2 und motiviert die folgende Definition.

Definition 1.7: Für $\lambda \in \mathbb{R}$ und $a = (a_1, \dots, a_n)^T \in \mathbb{R}^n$ sei

$$\lambda a := \begin{pmatrix} \lambda a_1 \\ \vdots \\ \lambda a_n \end{pmatrix}.$$

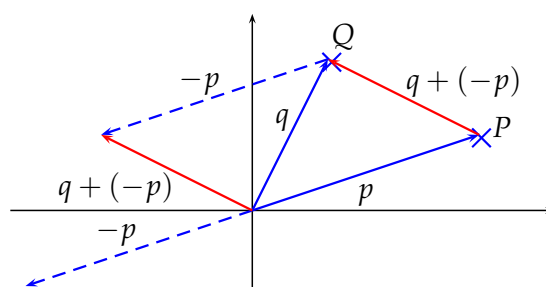


Abbildung 1.7: Differenz zweier Vektoren

Für $0 < \lambda \in \mathbb{R}$ entspricht also λa einem Punkt mit derselben Richtung wie a zum Ursprung, jedoch mit einem λ -fachen Abstand. Anhand einer Skizze erkennt man, dass die Multiplikation von a mit einer negativen Zahl die Richtung von a umkehrt.

Definition 1.8: Für $a, b \in \mathbb{R}^n$ sei

$$a - b := a + (-1)b$$

der *Differenzvektor* von a und b .

Abb. 1.7 liefert eine geometrische Deutung der Subtraktion zweier Vektoren: Interpretiert man die Vektoren p und q als Punkte, so entspricht $q - p$ dem Verschiebungsvektor von P nach Q , $p - q$ dagegen dem Verschiebungsvektor von Q nach P .

Definition 1.9: Zwei Vektoren $a, b \neq 0$ heißen *parallel* (Schreibweise $a \parallel b$) : $\Leftrightarrow \exists \alpha \in \mathbb{R} : a = \alpha b$. Gilt $\alpha > 0$, haben parallele Vektoren die *gleiche Richtung*, im Fall von $\alpha < 0$ *gegensätzliche Richtung*. Der Nullvektor ist zu jedem Vektor parallel.

Beispiel 1.10: Die Vektoren $a = (1, 2)^T$ und $b = (2, 4)^T$ sind parallel und haben die gleiche Richtung; der Vektor $c = (1, 3)^T$ ist weder zu a noch zu b parallel.

1.5 Lineare Gleichungssysteme und Matrizen

In der Einleitung wurde die zentrale Rolle linearer Gleichungssysteme betont. Daher beginnen wir hier mit ihrer systematischen Behandlung. Wir betrachten erneut das lineare Gleichungssystem aus Beispiel 1.1,

$$\begin{aligned} 0,8a + 0,8b + 0,8c &= 0,8 \\ 0a + 0,2b + 0,1c &= 0,12 \\ 0,2a + 0b + 0,1c &= 0,08 \end{aligned} \tag{1.2}$$

Offenbar hängt der Wert der Lösung nicht von der Bezeichnung der Variablen (hier: a, b und c) ab, sondern vielmehr von den Vorfaktoren vor den Variablen im linearen Gleichungssystem, den *Koeffizienten*. Zusammen mit der rechten Seite enthalten sie alle relevanten Informationen. Man kann somit ein Gleichungssystem kompakt schreiben, indem man nur die Koeffizienten als *Koeffizientenmatrix* angibt. Zum linearen Gleichungssystem (1.2) gehört die Koeffizientenmatrix

$$\begin{pmatrix} 0,8 & 0,8 & 0,8 \\ 0 & 0,2 & 0,1 \\ 0,2 & 0 & 0,1 \end{pmatrix}.$$

Wir werden also bei der Untersuchung von linearen Gleichungssystemen auf Matrizen geführt.

Definition 1.11: Eine $m \times n$ -Matrix A ist ein rechteckiges Schema von reellen oder komplexen Zahlen a_{ij} (den *Elementen* der Matrix) mit m Zeilen und n Spalten

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}. \quad (1.3)$$

Es heißt i *Zeilenindex* und j *Spaltenindex* von a_{ij} . Eine Matrix wird alternativ in der Form (a_{ij}) bzw. $(a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$ notiert. Die Menge aller $m \times n$ -Matrizen mit reellen bzw. komplexen Elementen wird mit $\mathbb{R}^{m \times n}$ bzw. $\mathbb{C}^{m \times n}$ bezeichnet. Soll nicht zwischen \mathbb{R} und \mathbb{C} unterschieden werden, schreiben wir $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$.

Definition 1.12:

1. Seien $A = (a_{ij}), B = (b_{ij}) \in \mathbb{K}^{m \times n}$. Man setzt $A = B := \Leftrightarrow a_{ij} = b_{ij} \forall 1 \leq i \leq m, 1 \leq j \leq n$.
2. Eine Matrix, deren Elemente alle den Wert 0 annehmen, heißt *Nullmatrix*.
3. Zu A wie oben sei

$$A^T := (a_{ji}) \in \mathbb{K}^{n \times m}$$

die *transponierte Matrix*. Man erhält also A^T , indem man die Spalten von A als Zeilen von A^T verwendet.

Die Koeffizientenmatrix des allgemeinen linearen Gleichungssystems

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \cdots + \cdots + \cdots + \cdots &= \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \quad (1.4)$$

von m linearen Gleichungen und n Unbekannten entspricht der Matrix A aus Formel (1.3). Sie allein beschreibt ein lineares Gleichungssystem aber nicht vollständig, weil die rechte Seite nicht vorkommt. Wird A rechts um eine Spalte b ergänzt, die die Werte der rechten Seite enthält, gelangt man zur erweiterten Koeffizientenmatrix (A, b) . Um hierin die Sonderrolle der letzten Spalte zu betonen, wird sie oftmals durch einen senkrechten Strich von A getrennt.

Beispiel 1.13: Die erweiterte Koeffizientenmatrix des linearen Gleichungssystems aus Beispiel 1.1 lautet

$$\begin{pmatrix} 0,8 & 0,8 & 0,8 & 0,8 \\ 0 & 0,2 & 0,1 & 0,12 \\ 0,2 & 0 & 0,1 & 0,08 \end{pmatrix} \quad \text{bzw.} \quad \left(\begin{array}{ccc|c} 0,8 & 0,8 & 0,8 & 0,8 \\ 0 & 0,2 & 0,1 & 0,12 \\ 0,2 & 0 & 0,1 & 0,08 \end{array} \right).$$

Als besonders interessant werden sich Matrizen mit gleich vielen Zeilen wie Spalten erweisen.

Definition 1.14:

1. Eine $n \times n$ -Matrix heißt *quadratisch*.
2. Eine quadratische Matrix A mit Elementen a_{ij} wird häufig in der Form $A = (a_{ij})_{i,j=1}^n$ geschrieben.
3. Sei $A = (a_{ij})_{i,j=1}^n$ quadratisch. Die Elemente von A mit $i = j$ bilden die *Hauptdiagonale* von A .
4. Eine quadratische Matrix, bei der alle Elemente oberhalb der Hauptdiagonalen gleich 0 sind, heißt *untere Dreiecksmatrix*.
5. Eine quadratische Matrix, bei der alle Elemente unterhalb der Hauptdiagonalen gleich 0 sind, heißt *obere Dreiecksmatrix*.

Beispiel 1.15: Die Matrix

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 8 & 4 \end{pmatrix}$$

ist eine untere Dreiecksmatrix.

Definition 1.16:

1. Sei $A \in \mathbb{K}^{n \times n}$. Gilt $a_{ij} = 0$ für alle $i \neq j$, heißt A *Diagonalmatrix*.

2. Die Diagonalmatrix

$$E_n := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} \in \mathbb{K}^{n \times n}$$

heißt *Einheitsmatrix*.

Bemerkung 1.17: Die m Werte der rechten Seite eines reellen linearen Gleichungssystems (vgl. Formel (1.4)) lassen sich zu einem Vektor

$$b = (b_1, \dots, b_m)^T \in \mathbb{R}^m$$

zusammenfassen und als Punkt im \mathbb{R}^m deuten, ebenso bilden die n Komponenten der Lösung einen Vektor $x \in \mathbb{R}^n$.

1.6 Lösung eines linearen Gleichungssystems

1.6.1 Das Gaußsche Eliminationsverfahren

Wir stellen nachfolgend ein Standard-Verfahren zum Lösen von linearen Gleichungssystemen vor, das *Gaußsche Eliminationsverfahren*, das ein gegebenes Gleichungssystem in ein anderes überführt, dessen Lösung unmittelbar ablesbar ist. Diese Lösung ist zugleich die Lösung des ursprünglich gegebenen linearen Gleichungssystems.

Definition 1.18:

1. Zwei Gleichungssysteme heißen *äquivalent*, falls sie die gleiche Lösungsmenge haben.
2. Zwei Matrizen A und B heißen *äquivalent* ($A \sim B$), falls die entsprechenden Gleichungssysteme die gleiche Lösungsmenge besitzen.

Bemerkung 1.19: Äquivalenz im Sinne von Definition 1.18 ist eine *Äquivalenzrelation* (vgl. die Vorlesung „Mathematische Grundlagen“).

Beispiel 1.20: Das Gleichungssystem

$$\begin{aligned} 2x + 3y - 4z &= 8 \\ -4x + 2y + 3z &= -5 \\ 3x + y + 2z &= 13 \end{aligned} \quad (a)$$

ist äquivalent zu den Gleichungssystemen

$$\begin{aligned} 2x + 3y - 4z &= 8 \\ 8y - 5z &= 11 \\ z &= 1 \end{aligned} \quad (b)$$

und

$$\begin{aligned} x &= 3 \\ y &= 2 \quad (c) \\ z &= 1 \end{aligned}$$

Auch die drei zugehörigen erweiterten Matrizen A, B, C sind dann äquivalent:

$$A = \begin{pmatrix} 2 & 3 & -4 & 8 \\ -4 & 2 & 3 & -5 \\ 3 & 1 & 2 & 13 \end{pmatrix}, B = \begin{pmatrix} 2 & 3 & -4 & 8 \\ 0 & 8 & -5 & 11 \\ 0 & 0 & 1 & 1 \end{pmatrix}, C = \begin{pmatrix} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

Man sagt, dass lineare Gleichungssystem (b) liegt in *Stufenform* vor, das lineare Gleichungssystem (c) in *reduzierter Stufenform*. Die zu Stufenform gehörige Matrix B weist unterhalb der Hauptdiagonalen nur Nullen auf (da eine erweiterte Koeffizientenmatrix nicht notwendig quadratisch ist, ist sie i. A. keine Dreiecksmatrix). Es enthält C Einsen auf der Hauptdiagonalen und Nullen sonst mit Ausnahme der letzten Spalte. Hier sind beliebige Werte möglich.

Offenbar ist die Stufenform (Ergebnis des Gauß-Algorithmus) und insbesondere die reduzierte Stufenform (Ergebnis des Gauß-Jordan-Algorithmus) übersichtlicher und die Bestimmung einer Lösung deutlich einfacher. Bei der reduzierten Stufenform kann man eine Lösung sogar einfach ablesen. Darum werden wir versuchen, durch *Äquivalenzumformungen der Zeilen* ein äquivalentes Gleichungssystem in reduzierter Stufenform zu finden. Wir verwenden drei Arten solcher Umformungen:

Z1. Addition eines Vielfachen einer Zeile zu einer anderen.

Z2. Vertauschen zweier Zeilen.

Z3. Multiplikation⁴ einer Zeile mit einem Skalar $\lambda \neq 0, \lambda \in \mathbb{K}$.

Um auf die reduzierte Stufenform zu gelangen, errechnet man zuerst eine Stufenform. Dies gelingt, indem man hauptsächlich durch Z1 die Elemente unterhalb der Hauptdiagonalen zu 0 macht. Man beginnt in der ersten Spalte und geht spaltenweise vor:

$$\begin{array}{ccc|c} 2 & 3 & -4 & 8 \\ -4 & 2 & 3 & -5 & | + 2 \cdot (\text{I}) \\ 3 & 1 & 2 & 13 & | - 3/2 \cdot (\text{I}) \\ \hline 2 & 3 & -4 & 8 \\ 0 & 8 & -5 & 11 \\ 0 & -7/2 & 8 & 1 & | + 7/16 \cdot (\text{II}) \\ \hline 2 & 3 & -4 & 8 \\ 0 & 8 & -5 & 11 \\ 0 & 0 & 93/16 & 93/16 \end{array}$$

⁴Z3 ist für den Gauß-Algorithmus nicht erforderlich, aber insbesondere für Handrechnungen vorteilhaft.

Die Stufenform ist nicht eindeutig, denn es gibt mehrere äquivalente Stufenformen. Hat man eine Stufenform erreicht, geht man zeilenweise von unten nach oben vor. Zunächst bringt man das Element auf der Hauptdiagonalen auf 1, anschließend alle Elemente darüber auf 0. Wir illustrieren das an Beispiel 1.20.

$$\begin{array}{ccc|c}
 2 & 3 & -4 & 8 \\
 0 & 8 & -5 & 11 \\
 0 & 0 & \frac{93}{16} & \frac{93}{16} & | \div \frac{93}{16} \\
 \hline
 2 & 3 & -4 & 8 & | + 4 \cdot (\text{III}) \\
 0 & 8 & -5 & 11 & | + 5 \cdot (\text{III}) \\
 0 & 0 & 1 & 1 \\
 \hline
 2 & 3 & 0 & 12 \\
 0 & 8 & 0 & 16 & | \div 8 \\
 0 & 0 & 1 & 1 \\
 \hline
 2 & 3 & 0 & 12 & | - 3 \cdot (\text{II}) \\
 0 & 1 & 0 & 2 \\
 0 & 0 & 1 & 1 \\
 \hline
 2 & 0 & 0 & 6 & | \div 2 \\
 0 & 1 & 0 & 2 \\
 0 & 0 & 1 & 1 \\
 \hline
 1 & 0 & 0 & 3 \\
 0 & 1 & 0 & 2 \\
 0 & 0 & 1 & 1
 \end{array}$$

Transferiert man das Schema wieder in das eigentliche Gleichungssystem zurück, erhält man die Lösung:

$$\begin{aligned}
 x &= 3 \\
 y &= 2 \\
 z &= 1
 \end{aligned}$$

Dass das Gaußsche Eliminationsverfahren tatsächlich zu äquivalenten linearen Gleichungssystemen führt, wird (mit etwas mehr Theorie) später in Kapitel 6 bewiesen werden.

In vielen Fällen werden die Zahlen im Laufe der Berechnung der Lösung eines linearen Gleichungssystems schnell unhandlich. Dies macht eine Berechnung per Hand aufwändig und fehleranfällig. Daher werden in der Praxis lineare Gleichungssysteme häufig nicht durch Handrechnung gelöst, sondern mithilfe des Computers. Dazu werden oft spezielle Programmpakete verwendet, die primär auf Berechnungen mit Hilfe von Matrizen ausgelegt ist, oftmals aber einen weitaus größeren Funktionsumfang bieten. In der Industrie weit verbreitet ist das kommerzielle Programmpaket MATLAB (www.mathworks.com), als Alternativen zu MATLAB existieren z. B. die Open-Source-Programme python (<http://www.python.org>) oder

Octave (<http://www.gnu.org/software/octave/>). Die drei genannten Programmpakete gleichen sich hinsichtlich ihrer Bedienung weitgehend, so dass wir uns hier prototypisch auf MATLAB beschränken.

In MATLAB kann man interaktiv Anweisungen ausführen lassen. Dazu schreibt man hinter das Prompt-Zeichen `>>` die Eingabe. MATLAB gibt dann darunter das Ergebnis aus. Die Ausgabe des Ergebnisses kann man unterdrücken, indem man die Eingabezeile mit einem Semikolon abschließt. Wir legen die Matrix zu Beispiel 1.20 in MATLAB an:

```
>> a=[1 2 3;4 5 6;7 8 9]
```

```
a =
```

```
1     2     3
4     5     6
7     8     9
```

Die oberste Zeile ist unsere Eingabe, alle unteren Zeilen entsprechen der Ergebnisausgabe von MATLAB. Matrixzeilen werden durch ein Semikolon getrennt. Wir verwenden den Befehl `rref` (`rref` steht für *reduced row echelon form*, englisch für *reduzierte Stufenform*) zum Lösen des linearen Gleichungssystems und betrachten das lineare Gleichungssystem aus Beispiel 1.20. Die Eingabe `format rat` veranlasst MATLAB, das Ergebnis als Bruch auszugeben, was bei unseren simplen Beispielrechnungen gut gelingt. MATLAB wie auch python und Octave rechnet intern aber immer mit Fließkommazahlen. In der zweiten Zeile geben wir die erweiterte Matrix zu Beispiel 1.20 ein, die in die reduzierte Stufenform umgerechnet wird.

```
>> format rat
```

```
>> a=[2 3 -4 8;-4 2 3 -5;3 1 2 13]
```

```
a =
```

```
2           3           -4           8
-4          2           3           -5
3           1           2           13
```

```
>> rref(a)
```

```
ans =
```

```
1           0           0           3
0           1           0           2
0           0           1           1
```

Die Rückübersetzung des Resultats des `rref`-Kommandos liefert die Lösung

$$x = 3$$

$$y = 2$$

$$z = 1$$

1.6.2 Unter- und überbestimmte Gleichungssysteme

Im letzten Kapitel hatten wir uns auf lineare Gleichungssysteme beschränkt, die genauso viele Unbekannte wie Gleichungen aufweisen. Unter gewissen Zusatzvoraussetzungen existiert dann genau eine Lösung, und in diesem Fall ist diese immer mit dem Gauß-Algorithmus berechenbar. Wir betrachten jetzt allgemeine lineare Gleichungssysteme. Dann sind drei Fälle möglich:

1. Es existiert eine eindeutige Lösung.
2. Es existiert keine Lösung.
3. Es existieren unendlich viele Lösungen.

Im zweiten Fall spricht man von *überbestimmten linearen Gleichungssystemen*, im dritten Fall von *unterbestimmten linearen Gleichungssystemen*. Wir werden dieses Phänomen hier nicht systematisch untersuchen, weil die reduzierte Stufenform als Ergebnis des Gaußschen Eliminationsverfahrens immer zeigt, welcher Fall vorliegt. Die folgende Definition vereinfacht die Analyse der reduzierten Stufenform.

Definition 1.21: Der erste Nicht-Null-Eintrag jeder Zeile einer Matrix heißt *Pivot-Element*. Eine Spalte, in der ein Pivot-Element vorkommt, heißt *Pivot-Spalte*.

In der reduzierten Stufenform ist das Pivot-Element immer 1. Die reduzierte Stufenform ist im allgemeinen Fall nicht eindeutig. Zur Eindeutigkeit fordern wir daher von der reduzierten Stufenform folgende Eigenschaften:

- Kommen Nullzeilen vor, also Zeilen, die nur Nullen enthalten, dann müssen diese die untersten Zeilen der Matrix sein. Nullzeilen repräsentieren die Gleichung $0 = 0$ und können ignoriert werden.
- Die Pivot-Spalte der Zeile $i + 1$ muss rechts der Pivot-Spalte der Zeile i liegen.
- Pivot-Spalten enthalten in der reduzierten Stufenform außer dem Pivot-Element nur Nullen. Diese Bedingung gilt nicht für die letzte Spalte der erweiterten Matrix.

Diese Forderungen lassen sich für jede Matrix durch die Äquivalenzumformungen Z1 – Z3 erfüllen.

Beispiel 1.22: Wir bezeichnen ein Pivot-Element symbolisch mit P und ein beliebiges Element mit x . Die Matrix

$$\begin{pmatrix} P & 0 & 0 & x \\ 0 & P & 0 & x \\ 0 & 0 & P & x \end{pmatrix}$$

liegt in reduzierter Stufenform vor. Alle Spalten mit Ausnahme der letzten sind Pivot-Spalten. Ebenso liegen folgende Matrizen in reduzierter Stufenform vor:

$$\begin{pmatrix} P & 0 & x & 0 & x \\ 0 & P & x & 0 & x \\ 0 & 0 & 0 & P & x \\ 0 & 0 & 0 & 0 & P \end{pmatrix} \begin{pmatrix} P & 0 & x & x & 0 & x & x & x \\ 0 & P & x & x & 0 & x & x & x \\ 0 & 0 & 0 & 0 & P & x & x & x \end{pmatrix} \begin{pmatrix} P & 0 & 0 & x \\ 0 & P & 0 & x \\ 0 & 0 & P & x \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Wir nennen im Folgenden die Spalten 1 bis $n - 1$ der erweiterten Matrix *Koeffizientenmatrix* und die n -te Spalte *letzte Spalte*. Hat das Gleichungssystem eine eindeutige Lösung, so erhält man immer die Stufenform bzw. die reduzierte Stufenform

$$\left(\begin{array}{ccc|c} P & x & x & x \\ 0 & P & x & x \\ 0 & 0 & P & x \end{array} \right) \quad \text{bzw.} \quad \left(\begin{array}{ccc|c} P & 0 & 0 & x \\ 0 & P & 0 & x \\ 0 & 0 & P & x \end{array} \right).$$

Falls die letzte Spalte der Stufenform eine Pivot-Spalte ist, ist das Gleichungssystem *nicht lösbar*. Man könnte also z. B. die folgende Stufenform bzw. reduzierte Stufenform erhalten:

$$\left(\begin{array}{cccc|c} P & x & x & x & x \\ 0 & P & x & x & x \\ 0 & 0 & 0 & 0 & P \end{array} \right) \quad \text{bzw.} \quad \left(\begin{array}{cccc|c} P & 0 & x & x & x \\ 0 & P & x & x & x \\ 0 & 0 & 0 & 0 & P \end{array} \right).$$

Beispiel 1.23: Die erweiterte Matrix sei

$$\left(\begin{array}{ccc|c} 2 & 0 & 3 & 0 \\ 0 & -1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right).$$

Die letzte Zeile lautet ausgeschrieben

$$0 \cdot x + 0 \cdot y + 0 \cdot z = 1.$$

Diese Gleichung ist unerfüllbar für beliebige Werte von x, y und z . Somit ist das lineare Gleichungssystem unlösbar.

Erhält man in der reduzierten Stufenform weniger Pivotspalten als das Gleichungssystem Unbekannte hat, hat das Gleichungssystem unendlich viele Lösungen. Mögliche erweiterte Matrizen sind dann z. B.

$$\left(\begin{array}{cccc|c} P & x & x & x & x \\ 0 & P & x & x & x \\ 0 & 0 & 0 & 0 & 0 \end{array} \right) \quad \text{bzw.} \quad \left(\begin{array}{cccc|c} P & 0 & x & x & x \\ 0 & P & x & x & x \\ 0 & 0 & 0 & 0 & 0 \end{array} \right).$$

Es gibt (mindestens) einen *freien Parameter*, d. h. eine der Variablen kann einen beliebigen Wert annehmen. Nach einer einfachen Lösungsregel gehört zu jeder Nicht-Pivot-Spalte der Koeffizientenmatrix eine freie Variable.

Beispiel 1.24: Zum Gleichungssystem

$$\begin{aligned} 2 \cdot x - 4 \cdot y + 2 \cdot z &= 8 \\ 1 \cdot x - 1 \cdot y - 7 \cdot z &= 6 \end{aligned}$$

berechnen wir die äquivalente reduzierte Stufenform in MATLAB:

```
>> format rat
>> a=[2 -4 2 8;1 -1 -7 6];
>> rref(a)
```

ans =

$$\begin{array}{cccc} 1 & 0 & -15 & 8 \\ 0 & 1 & -8 & 2 \end{array}$$

Das Ergebnis entspricht dem Gleichungssystem

$$\begin{aligned} 1 \cdot x + 0 \cdot y - 15 \cdot z &= 8 \\ 0 \cdot x + 1 \cdot y - 8 \cdot z &= 2 \end{aligned}$$

bzw.

$$\begin{aligned} x &= 8 + 15 \cdot z \\ y &= 2 + 8 \cdot z \end{aligned}$$

oder in Vektor-Schreibweise

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 8 \\ 2 \end{pmatrix} + z \begin{pmatrix} 15 \\ 8 \end{pmatrix}.$$

Es handelt sich dabei noch nicht um eine Lösung, denn diese besteht aus Tripeln (x, y, z) . Den Wert von z können wir aber wegen $z = 0 + 1 \cdot z$ leicht ergänzen. Hängt man diese Zeile an, erhält man das Endergebnis

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 8 \\ 2 \\ 0 \end{pmatrix} + z \begin{pmatrix} 15 \\ 8 \\ 1 \end{pmatrix}.$$

Zusammengefasst geht man bei nicht eindeutig lösbaeren Gleichungssystemen so vor:

- Die Nicht-Pivot-Spalten werden mit der zugehörigen Variablen versehen auf die rechte Seite gebracht, wobei sich alle Vorzeichen umkehren.
- Zeilen für fehlende Variablen werden in den Lösungsvektor nach dem Schema $x = x$ eingefügt.

Beispiel 1.25: Man löse das lineare Gleichungssystem

$$\begin{aligned} 2a + b - 2c &= -2 \\ -4a - 2b + c + 2d &= 2 \\ -2a - b + d &= 1 \end{aligned}$$

Man erhält die reduzierte Stufenform

```
>> a=[2 1 -2 0 -2;-4 -2 1 2 2;-2 -1 0 1 1];
>> rref(a)
```

ans =

$$\begin{array}{cccccc} 1 & & 1/2 & & 0 & & -1 \\ 0 & & 0 & & 1 & & 0 \\ 0 & & 0 & & 0 & & 1 & & -1 \end{array}$$

Die zweite Spalte der Matrix ist keine Pivot-Spalte. Dies führt zu

$$\begin{pmatrix} a \\ c \\ d \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ -1 \end{pmatrix} + b \begin{pmatrix} -\frac{1}{2} \\ 0 \\ 0 \end{pmatrix}$$

und dann zum Endergebnis

$$\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 0 \\ -1 \end{pmatrix} + b \begin{pmatrix} -\frac{1}{2} \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

Kapitel 2

Analytische Geometrie

2.1 Skalarprodukt und Norm

Wir werden nun geometrische Konzepte wie die Länge eines Vektors und den Winkel zwischen zwei Vektoren auf den abstrakten Raum \mathbb{R}^n übertragen. Das wird selbstverständlich so geschehen, dass sich in den Spezialfällen \mathbb{R}^2 und \mathbb{R}^3 die intuitiv bekannten Längen- und Winkelbegriffe ergeben. Entscheidend hierfür ist das Konzept des *Skalarprodukts*, dessen konkrete geometrische Bedeutung wir zunächst nicht betrachten. Vorbereitend definieren wir das *kartesische Produkt* von Mengen.

Definition 2.1: Seien X_1, \dots, X_n mit $n \in \mathbb{N}$ nichtleere Mengen. Dann versteht man unter dem *kartesischen Produkt* $X_1 \times \dots \times X_n$ die Menge aller geordneten n -Tupel:

$$X_1 \times \dots \times X_n := \{(x_1, \dots, x_n)^T \mid x_i \in X_i\}.$$

Für $X_1 = \dots = X_n = X$ schreibt man kurz X^n statt $X_1 \times \dots \times X_n$.

Beispiel 2.2: Offenbar gilt $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ und allgemeiner $\mathbb{R}^n = \mathbb{R} \times \dots \times \mathbb{R}$ durch die Wahl von $X_1 = \dots = X_n = \mathbb{R}$ in Definition 2.1.

Bemerkung 2.3:

1. Nach Beispiel 2.2 verallgemeinert das kartesische Produkt die Definition 1.4 des Zahlraums \mathbb{R}^n .
2. Lässt man statt der *Indexmenge* $\{1, \dots, n\}$ wie oben für i alle natürlichen Zahlen zu, wird man auf unendliche kartesische Produkte geführt. Elemente solcher Produktmengen werden *Folgen* genannt und in der Analysis eingehend untersucht.

Von nun an seien im ganzen Kapitel immer $a = (a_i)_{i=1}^n, b = (b_i)_{i=1}^n \in \mathbb{R}^n$.

Definition 2.4: Eine Abbildung $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ heißt *Skalarprodukt*, wenn folgende Bedingungen erfüllt sind:

SP1 (Symmetrie): $\forall a, b \in \mathbb{R}^n : \langle a, b \rangle = \langle b, a \rangle$.

SP2 $\forall a, b, c \in \mathbb{R}^n :$

$$\langle a, b + c \rangle = \langle a, b \rangle + \langle a, c \rangle$$

SP3 $\forall \alpha \in \mathbb{R}$ gilt

$$\langle \alpha a, b \rangle = \alpha \langle a, b \rangle = \langle a, \alpha b \rangle.$$

SP4 (positive Definitheit): $\forall a \in \mathbb{R}^n \setminus \{0\} : \langle a, a \rangle > 0$, und $\langle 0, 0 \rangle = 0$.

Das bei Weitem wichtigste Skalarprodukt ist das *Standardskalarprodukt* oder *euklidische Skalarprodukt*. Wird im Folgenden von *dem* Skalarprodukt gesprochen, ist das euklidische Skalarprodukt gemeint.

Definition 2.5: Für a, b sei ihr *euklidisches Skalar-* bzw. *Punktprodukt* $\langle a, b \rangle$ definiert als

$$\langle a, b \rangle := a_1 b_1 + \cdots + a_n b_n = \sum_{i=1}^n a_i b_i.$$

Bemerkung 2.6: Oft wird das euklidische Skalarprodukt als $a \cdot b$ geschrieben. In der Mathematik wird jedoch am häufigsten $\langle a, b \rangle$ verwendet. Das (Skalar)produkt zweier Vektoren entspricht in seinen Eigenschaften nicht dem Produkt zweier Zahlen, so z. B. sind höhere Potenzen von Vektoren wie a^3 nicht definiert.

Beispiel 2.7: Seien

$$a = \begin{pmatrix} 2 \\ -1 \\ -2 \end{pmatrix} \quad \text{und} \quad b = \begin{pmatrix} 1 \\ 3 \\ -4 \end{pmatrix}$$

gegeben. Dann gilt

$$\langle a, b \rangle = 2 \cdot 1 + (-1) \cdot 3 + (-2) \cdot (-4) = 7.$$

Aus Gründen der Widerspruchsfreiheit formulieren und beweisen wir folgenden Satz.

Satz 2.8: Das euklidische Skalarprodukt ist ein Skalarprodukt im Sinne der Definition 2.4.

BEWEIS Seien a, b beliebig. Man hat $\langle a, b \rangle = a_1 b_1 + \cdots + a_n b_n \in \mathbb{R}$, da Produkte und Summen reeller Zahlen wieder reelle Zahlen sind.

SP1:

$$\langle a, b \rangle = \sum_{i=1}^n a_i b_i = \sum_{i=1}^n b_i a_i = \langle b, a \rangle,$$

weil es bei der Multiplikation zweier reeller Zahlen nicht auf ihre Reihenfolge ankommt.

SP2: Sei nun $c = (c_i)_{i=1}^n \in \mathbb{R}^n$ beliebig. Dann gilt $b + c = (b_i + c_i)_{i=1}^n$ und deswegen mit dem Distributivgesetz der reellen Zahlen

$$\begin{aligned} \langle a, b + c \rangle &= \sum_{i=1}^n a_i (b_i + c_i) = \sum_{i=1}^n (a_i b_i + a_i c_i) \\ &= \sum_{i=1}^n a_i b_i + \sum_{i=1}^n a_i c_i = \langle a, b \rangle + \langle a, c \rangle \end{aligned}$$

SP3: Für $\alpha \in \mathbb{R}$ gilt

$$\langle \alpha a, b \rangle = \sum_{i=1}^n (\alpha a_i) b_i = \sum_{i=1}^n \alpha (a_i b_i) = \alpha \sum_{i=1}^n a_i b_i = \alpha \langle a, b \rangle$$

SP4: Offenbar gilt $\langle 0, 0 \rangle = 0$. Sei $a \neq 0 \in \mathbb{R}^n$. Dann ist mindestens eine Koordinate $a_j \neq 0$. Also gilt wegen $a_j^2 > 0$

$$\langle a, a \rangle = a_1^2 + \dots + a_n^2 \geq a_j^2 > 0. \quad \blacksquare$$

Wir betrachten nun einen Vektor $a = (a_1, a_2)^T \in \mathbb{R}^2$. Nach dem Satz des Pythagoras beträgt seine Länge $\sqrt{a_1^2 + a_2^2}$. Analog findet man durch zweimalige Anwendung des Satzes von Pythagoras für die Länge eines Vektors $a = (a_1, a_2, a_3)^T \in \mathbb{R}^3$ den Ausdruck $\sqrt{a_1^2 + a_2^2 + a_3^2}$ (Abb. 2.1). Es liegt daher nahe, diesen intuitiven Längenbegriff auf den \mathbb{R}^n zu verallgemeinern.

Definition 2.9: Einem Vektor a wird die *euklidische Norm* oder *Standardnorm* $\|a\|$ zugeordnet durch

$$\|a\| := \left(\sum_{i=1}^n a_i^2 \right)^{1/2}. \quad (2.1)$$

Satz 2.10: Die in Gleichung (2.1) definierte Norm hat folgende Eigenschaften:

N0 : $\|a\| \in \mathbb{R}$.

N1 : $\|a\| \geq 0$.

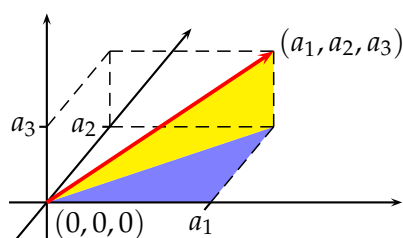


Abbildung 2.1: Berechnung der euklidischen Länge eines Vektors im \mathbb{R}^3 durch doppelte Anwendungen des Satzes von Pythagoras

$$\text{N2 : } \|a\| = 0 \Leftrightarrow a = 0.$$

$$\text{N3 : } \forall \lambda \in \mathbb{R} : \|\lambda a\| = |\lambda| \|a\|$$

$$\text{N4 : (Dreiecksungleichung) } \|a + b\| \leq \|a\| + \|b\|$$

BEWEIS

N0 : Wegen $\sum_{i=1}^n a_i^2 \geq 0$ ist die Wurzel dieses Ausdrucks reell.

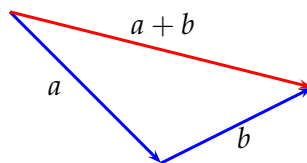
N1 : $\|a\|$ ist die positive Wurzel einer reellen Zahl.

N2 : $\|a\| = 0 \Leftrightarrow \|a\|^2 = a_1^2 + \dots + a_n^2 = 0 \Leftrightarrow a_i^2 = 0 \forall 1 \leq i \leq n \Leftrightarrow a = 0$, da Quadrate reeller Zahlen nicht negativ sind.

$$\text{N3 : } \|\lambda a\| = \sqrt{(\lambda^2 a_1^2) + \dots + (\lambda^2 a_n^2)} = \sqrt{\lambda^2} \|a\| = |\lambda| \|a\|.$$

N4 : später ■

Bemerkung 2.11: Die untenstehende Abbildung illustriert, warum die Eigenschaft **N4 Dreiecksungleichung** genannt wird: Die Länge einer Seite entspricht höchstens der Summe aus den Längen der beiden anderen Seiten.



Prägend für die Mathematik ist das Streben nach Abstraktion und Allgemeinheit. So haben wir bereits durch Verallgemeinerung auf \mathbb{R}^n einen Längenbegriff definiert. Wir werden nun das Konzept der Länge erneut verallgemeinern. Dazu deuten wir die Zuordnung einer Länge zu einem Vektor als Abbildung von \mathbb{R}^n nach \mathbb{R} .

Definition 2.12: Eine Abbildung $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}$ heißt *Norm*, wenn sie die Eigenschaften **N0** bis **N4** aus Satz 2.10 besitzt.

Man sieht sofort, dass sich die euklidische Norm nach Definition 2.9 als Wurzel des euklidischen Skalarprodukts eines Vektors mit sich selbst schreiben lässt. Man sagt, *das euklidische Skalarprodukt induziert die euklidische Norm*. Dieser Zusammenhang gilt für jedes Skalarprodukt.

Satz 2.13: Sei $\langle \cdot, \cdot \rangle$ ein beliebiges Skalarprodukt auf \mathbb{R}^n . Dann wird durch $\|\cdot\| := \sqrt{\langle \cdot, \cdot \rangle}$ eine Norm auf \mathbb{R}^n induziert.

BEWEIS Da $\langle \cdot, \cdot \rangle \geq 0$, existiert eine reelle positive Wurzel, und $\|\cdot\|$ ist in der Tat eine Abbildung von \mathbb{R}^n nach \mathbb{R} . **N1** gilt aus demselben Grund.

Offenbar gilt $\|0\| = \sqrt{\langle 0, 0 \rangle} = 0$ nach SP4. Sei also $\|a\| = 0$. Dann gilt auch $\|a\|^2 = 0 = \langle 0, 0 \rangle$ und damit nach SP4 $a = 0$, also gilt **N2**.

Wegen $\|\lambda a\| = \sqrt{\langle \lambda a, \lambda a \rangle} = \sqrt{\lambda^2 \langle a, a \rangle} = |\lambda| \|a\|$ aufgrund von SP3 gilt auch **N3**.

N4 beweisen wir später. ■

Es existieren viele Normen, d. h. Längenbegriffe auf dem \mathbb{R}^n . Wir werden künftig mit „der Norm“ immer die euklidische Norm meinen, mit „einer Norm“ jedoch eine unbestimmte Norm. Es folgen einige gebräuchliche Normen des \mathbb{R}^n .

- Auf \mathbb{R} ist der Absolutbetrag eine Norm.
- Für $p \geq 1$ definiert man die ℓ_p -Norm durch

$$\|a\|_p := \left(\sum_{i=1}^n |a_i|^p \right)^{1/p} \quad (2.2)$$

Die Dreiecksungleichung der ℓ_p -Norm ist in der Analysis als *diskrete Minkowskische Ungleichung* bekannt. Für ihren nichttrivialen Beweis verweisen wir auf einschlägige Literatur zur Analysis, z. B. [5]. Der Ausdruck (2.2) ist zwar auch für $0 < p < 1$ definiert, dann aber gilt die Dreiecksungleichung i. A. nicht mehr.

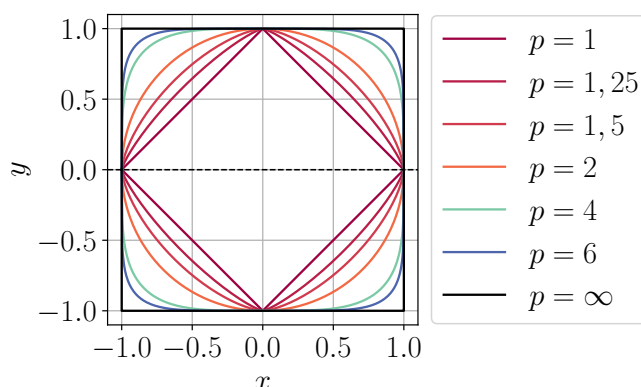
- Der Spezialfall $p = 1$ wird als *Betragssummennorm* oder *Einernorm* bezeichnet und ist definiert als

$$\|a\|_1 = |a_1| + \cdots + |a_n|.$$

- Der Spezialfall $p = 2$ entspricht der euklidischen Norm.
- Die *Maximumnorm* oder ℓ_∞ -Norm ist definiert als

$$\|a\|_\infty = \max\{|a_1|, \dots, |a_n|\};$$

der Bezeichnung erklärt sich durch $\lim_{p \rightarrow \infty} \|a\|_p = \|a\|_\infty$.

Abbildung 2.2: Einheitskugeln für verschiedene ℓ_p -Normen

Beispiel 2.14: Für $a = (2, -1, -2)^T$ gilt

$$\|a\| = \sqrt{4 + 1 + 4} = \sqrt{9} = 3$$

$$\|a\|_1 = 2 + 1 + 2 = 5$$

$$\|a\|_\infty = \max\{2, 1, 2\} = 2.$$

Bemerkung 2.15: Satz 2.13 besagt, dass jedes Skalarprodukt eine Norm induziert. Umgekehrt ist aber nicht jede Vektornorm von einem Skalarprodukt induziert. So z. B. ist *keine* der ℓ_p -Normen von irgendeinem Skalarprodukt induziert bis auf den Spezialfall $p = 2$.

2.1.1 Einheitsvektoren

Definition 2.16: Sei $\|\cdot\|$ eine Norm. Ein Vektor $e \in \mathbb{R}^n$ heißt *Einheitsvektor* (zur Norm $\|\cdot\|$), wenn $\|e\| = 1$ ist.

Bemerkung 2.17: Ob ein gegebener Vektor ein Einheitsvektor ist, hängt entscheidend von der verwendeten Norm ab. Wir betrachten die *Einheitskugel* S^1 im \mathbb{R}^2 ,

$$S^1 := \{x \in \mathbb{R}^2 \mid \|x\| = 1\}.$$

Im Fall der euklidischen Norm handelt es sich bei S^1 um den wohlbekannten Einheitskreis; für bestimmte p -Normen ergeben sich die in Abb. 2.2 gezeigten Mengen.

Besondere Bedeutung besitzen die *kanonischen Einheitsvektoren*

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, e_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix},$$

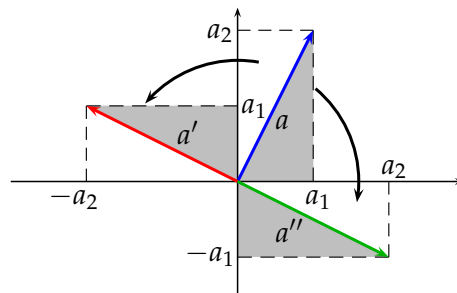


Abbildung 2.3: Drehung eines Vektors a um 90° im und entgegen dem Uhrzeigersinn

speziell in \mathbb{R}^3

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad (2.3)$$

die den Koordinatenrichtungen entsprechen und für die offensichtlich $\|e_i\|_p = 1$ für jede ℓ_p -Norm gilt. Jeder Vektor $a \neq 0$ kann *normiert* werden, d. h. man findet zu jedem Vektor a einen zu a parallelen und gleichgerichteten Einheitsvektor e_a durch

$$e_a = \frac{1}{\|a\|} a.$$

Beispiel 2.18: Sei $a = (2, -1, -2)^T$. Wegen $\|a\| = 3$ ist $e_a = 1/3(2, -1, -2)^T$.

Bemerkung 2.19: Die Vektoren v und λv ($\lambda > 0$) zeigen in die gleiche Richtung. Nach Normierung sind beide Vektoren gleich. Man kann beim Normieren also als ersten Schritt einen positiven Vorfaktor des Vektors streichen, um die Rechnung zu vereinfachen.

Beispiel 2.20: Möchte man den Vektor $b = \frac{11}{\sqrt{17}}(1, 0, 1)^T$ normieren, kann man statt dessen den Vektor $b' = (1, 0, 1)^T$ normieren und erhält bei einfacherer Rechnung das gleiche Ergebnis.

2.1.2 Orthogonale (senkrechte) Vektoren

Bemerkung 2.21: Wir gehen von der Ebene aus und betrachten den Vektor $a = (a_1, a_2)^T$. Anhand der Skizze in Abb. 2.3 erkennt man, dass für den um 90° gegen den Uhrzeigersinn gedrehte Vektor a' gilt: $a' = (-a_2, a_1)^T$,

und der um 90° im Uhrzeigersinn gedrehte Vektor a'' die Koordinaten $(a_2, -a_1)^T$ besitzt. Es gilt also mit dem euklidischen Skalarprodukt

$$\langle a, a' \rangle = \langle a'', a \rangle = 0.$$

Umgekehrt wird man durch die Bedingung $\langle a, b \rangle = 0$ über ein lineares Gleichungssystem auf das Ergebnis $b = \lambda(-a_2, a_1)^T$ mit einer beliebigen reellen Zahl λ geführt, so dass b offenbar parallel zu a' und a'' verläuft und damit senkrecht auf a steht. Also steht a genau dann auf b senkrecht, wenn $\langle a, b \rangle = 0$ gilt. Diese Betrachtungen erlauben es, den Begriff „senkrecht“ in den allgemeinen \mathbb{R}^n zu übertragen.

Definition 2.22: Seien $a, b \in \mathbb{R}^n$ und $\langle \cdot, \cdot \rangle$ ein beliebiges Skalarprodukt. Die Vektoren a und b stehen *senkrecht* (oder auch *orthogonal*) zueinander bzgl. $\langle \cdot, \cdot \rangle$, Schreibweise $a \perp b$, wenn $\langle a, b \rangle = 0$ ist.

Für den Spezialfall des euklidischen Skalarprodukts und $n = 2$ entspricht diese Definition der intuitiven Vorstellung des Begriffes „senkrecht“.

Bemerkung 2.23:

1. Die kanonischen Einheitsvektoren (2.3) stehen senkrecht zueinander im Sinne der Definition 2.22.
2. Ob zwei Vektoren aufeinander senkrecht stehen, hängt entscheidend von der Wahl des Skalarproduktes ab: Man kann zeigen, dass im \mathbb{R}^2 durch $\langle a, b \rangle' := a_1 b_1 + 2a_2 b_2$ ein Skalarprodukt definiert wird. Die Vektoren $(-1, 1)^T$ und $(1, 1)^T$ stehen bezogen auf das euklidische Skalarprodukt senkrecht aufeinander, nicht aber bezogen auf das Skalarprodukt $\langle \cdot, \cdot \rangle'$, da $\langle a, b \rangle' = 1$.

Bemerkung 2.24: Aus $a \perp b$ folgt $\alpha a \perp \beta b$ für beliebige reelle Zahlen $\alpha, \beta \in \mathbb{R}$ wegen

$$\langle \alpha a, \beta b \rangle = \alpha \beta \langle a, b \rangle = 0.$$

Wir beweisen nun den Satz des Pythagoras.

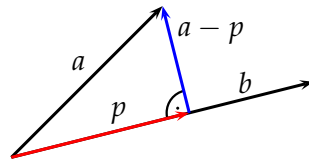
Satz 2.25: Seien $a, b \in \mathbb{R}^n$ mit $a \perp b$, $\langle \cdot, \cdot \rangle$ ein beliebiges Skalarprodukt und $\| \cdot \|$ die dadurch induzierte Norm. Dann gilt

$$\|a + b\|^2 = \|a\|^2 + \|b\|^2.$$

BEWEIS

$$\|a + b\|^2 = \langle a + b, a + b \rangle = \langle a, a \rangle + \langle a, b \rangle + \langle b, a \rangle + \langle b, b \rangle$$

Mit $\langle a, b \rangle = 0 = \langle b, a \rangle$ wegen $a \perp b$ und $\langle \cdot, \cdot \rangle = \| \cdot \|^2$ nach Definition folgt die Behauptung. ■

Abbildung 2.4: orthogonale Projektion von a auf b

Bemerkung 2.26: Dass der Satz des Pythagoras gilt, wird aus der Schule längst bekannt sein. Den Wert der obigen Aussage mögen folgende Aspekte verdeutlichen:

1. Der Satz des Pythagoras gilt nicht nur für das euklidische Skalarprodukt und die von ihm induzierte Norm, sondern für *jedes* Skalarprodukt mit induzierter Norm.
2. Er gilt nicht nur in der Ebene, sondern allgemeiner im \mathbb{R}^n .
3. Der Beweis mithilfe der Vektorrechnung ist kurz und elegant verglichen mit dem elementargeometrischen Beweis, der auf Flächenvergleichen basiert.
4. Dass der Satz des Pythagoras mit den von der Ebene abstrahierten geometrischen Begriffen und Definitionen immer noch gilt, rechtfertigt diese Abstraktionen.

2.1.3 Winkel zwischen Vektoren

Wie schon zuvor gehen wir immer davon aus, dass die Norm $\|\cdot\|$ durch das gegebene Skalarprodukt induziert wird. Seien a, b zwei Vektoren und $b \neq 0$. Wir möchten die orthogonale Projektion p von a in Richtung von b definieren (Abb. 2.4). Wir bestimmen einen Vektor p , so dass gilt:

$$\begin{aligned} 1. \quad p &\parallel b &\Rightarrow p &= \alpha b \\ 2. \quad (a - p) &\perp b &\Rightarrow \langle a - p, b \rangle &= 0 \end{aligned} \quad (2.4)$$

Daraus folgt:

$$\begin{aligned} \langle a - \alpha b, b \rangle &= 0 \\ \Leftrightarrow \langle a, b \rangle - \alpha \langle b, b \rangle &= 0 \\ \Leftrightarrow \alpha &= \frac{\langle a, b \rangle}{\langle b, b \rangle} \end{aligned} \quad (2.5)$$

$$\Rightarrow p = \frac{\langle a, b \rangle}{\langle b, b \rangle} \cdot b \quad (2.6)$$

Definition 2.27: Der für zwei Vektoren a und $b \neq 0$ durch (2.6) eindeutig bestimmte Vektor p heißt *orthogonale Projektion von a in Richtung b* , man schreibt auch $p_b(a)$.

Wir berechnen die Länge der orthogonale Projektion.

$$\|p_b(a)\| = \left\| \frac{\langle a, b \rangle}{\langle b, b \rangle} b \right\| = \left| \frac{\langle a, b \rangle}{\langle b, b \rangle} \right| \|b\| = \frac{|\langle a, b \rangle|}{\|b\|^2} \cdot \|b\| = \frac{|\langle a, b \rangle|}{\|b\|} \quad (2.7)$$

Bemerkung 2.28: Ist b ein Einheitsvektor, so gilt:

$$p_b(a) = \langle a, b \rangle \cdot b.$$

Beispiel 2.29: Seien a und b wie in Beispiel 2.7, also $a = (2, -1, -2)^T$ und $b = (1, 3, -4)^T$. Dann gilt

$$p_b(a) = \frac{\langle a, b \rangle}{\langle b, b \rangle} b = \frac{2 - 3 + 8}{1 + 9 + 16} \begin{pmatrix} 1 \\ 3 \\ -4 \end{pmatrix} = \frac{7}{26} \begin{pmatrix} 1 \\ 3 \\ -4 \end{pmatrix}.$$

Bemerkung 2.30: Die Idee der orthogonalen Projektion mag an dieser Stelle unscheinbar anmuten, es handelt sich aber um eines der zentralsten Konzepte der Mathematik, dessen Bedeutung nicht überschätzt werden kann. So basiert die Methode der Finiten Elemente, die zur Numerischen Simulation mechanischer Eigenschaften von Bauteilen weite Anwendung findet (vgl. Beispiel 1.2 aus Kapitel 1.1), auf orthogonalen Projektionen, wenn auch in einem erneut verallgemeinerten Sinn.

Ähnliche Bedeutung in der Mathematik hat die Cauchy¹-Schwarzsche² Ungleichung.

Satz 2.31 (Cauchy-Schwarzsche Ungleichung): Für $a, b \in \mathbb{R}^n$ gilt

$$|\langle a, b \rangle| \leq \|a\| \|b\|. \quad (2.8)$$

BEWEIS Für $b = 0$ oder $a = 0$ sind beide Seiten der Ungleichung gleich 0, und Formel (2.8) gilt. Es sei jetzt $b \neq 0$ und p die orthogonale Projektion von a auf b . Wegen $p \parallel b$ gilt nach Gleichung (2.4b) $p - a \perp p$ und daher mit dem Satz des Pythagoras

$$\|a\|^2 = \|p\|^2 + \|p - a\|^2 \geq \|p\|^2$$

¹Augustin-Louis Cauchy (1789–1857), franz. Mathematiker; bedeutende Beiträge zu den Grundlagen der Analysis, zur Funktionentheorie und Mechanik; Beweis der Cauchy-Schwarzschen Ungleichung für den Spezialfall des Standardskalarprodukts 1821.

²Hermann Amandus Schwarz (1843–1921), Beiträge zur Funktionentheorie und Analysis

aufgrund von $\|p - a\|^2 \geq 0$. Man erhält also $\|a\| \geq \|p\|$ aufgrund der Monotonie der Wurzelfunktion. Einsetzen dieser Ungleichung in (2.7) liefert

$$\|a\| \geq \|p\| = \frac{|\langle a, b \rangle|}{\|b\|}$$

und damit nach Multiplikation mit $\|b\|$ die Behauptung. ■

Wir holen nun den Beweis der Normeigenschaft **N4** für eine durch ein Skalarprodukt induzierte Norm nach und vollenden so den Beweis von Satz 2.13.

Satz 2.32 (Dreiecksungleichung): Für $a, b \in \mathbb{R}^n$ gilt

$$\|a + b\| \leq \|a\| + \|b\|.$$

BEWEIS Beide Seiten dieser Ungleichung sind nicht negativ. Daher genügt es, zu beweisen, dass ihre Quadrate die gewünschte Ungleichung erfüllen d. h.

$$\langle a + b, a + b \rangle \leq (\|a\| + \|b\|)^2.$$

Man hat

$$\langle a + b, a + b \rangle = \langle a, a \rangle + 2\langle a, b \rangle + \langle b, b \rangle \leq \|a\|^2 + 2\|a\| \|b\| + \|b\|^2$$

aufgrund der Cauchy-Schwarzschen Ungleichung (2.8). Mit der ersten binomischen Formel folgt die Behauptung. ■

Bemerkung 2.33: Dass bei einer beliebigen Norm, die eventuell nicht durch ein Skalarprodukt induziert ist, die Dreiecksungleichung gilt, folgt nicht aus Satz 2.32 und muss von Fall zu Fall bewiesen werden.

Wir kommen jetzt zur Definition von Winkeln zwischen zwei Vektoren. Dazu seien $a, b \neq 0$ Vektoren in der Ebene mit dem Standardskalarprodukt und der Standardnorm. Für den Winkel θ zwischen a und b gelte zunächst $0 \leq \theta \leq \pi/2$. Durch die orthogonale Projektion $p_a(b)$ ergänzen wir a und b zu einem rechtwinkligen Dreieck (Abb. 2.5 links). Es folgt

$$\cos \theta = \frac{\|p_b(a)\|}{\|a\|} \stackrel{(2.7)}{=} \frac{|\langle a, b \rangle|}{\|a\| \|b\|}.$$

Da $p_b(a)$ dieselbe Richtung wie b hat, gilt $p = \alpha b$ mit $\alpha \geq 0$. Andererseits gilt nach (2.5) auch $\alpha = \langle a, b \rangle / \|b\|^2$, also $\langle a, b \rangle \geq 0$. Man erhält

$$\cos \theta = \frac{\langle a, b \rangle}{\|a\| \|b\|}. \quad (2.9)$$

Analog diskutiert man den Fall $\pi/2 < \theta$ (Abb. 2.5 rechts). Hier ist $\langle a, b \rangle < 0$, weil $p_b(a)$ in Gegenrichtung von b zeigt, aber mit $\cos(\pi - \theta) = -\cos \theta$ erhält man erneut (2.9). Dies legt folgende Definition nahe.

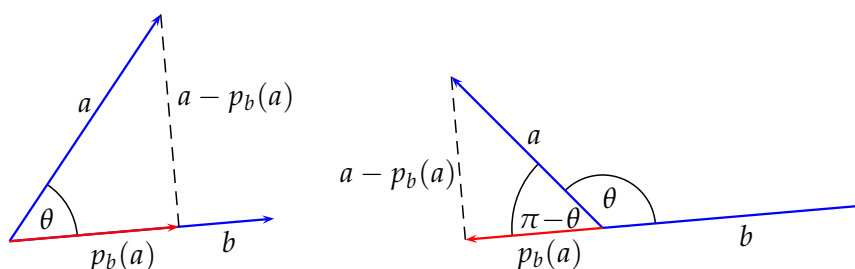


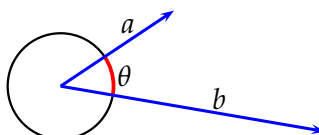
Abbildung 2.5: Winkel zwischen Vektoren in der Ebene

Definition 2.34: Seien $a, b \in \mathbb{R}^n \setminus \{0\}$. Der Winkel zwischen a und b , geschrieben $\angle(a, b)$, wird definiert als

$$\angle(a, b) := \arccos \frac{\langle a, b \rangle}{\|a\| \|b\|}. \quad (2.10)$$

Aufgrund der Cauchy-Schwarzschen Ungleichung ist der Betrag des Arguments des Arcuscosinus in Definition höchstens 1 und damit der Winkel durch (2.10) wohldefiniert. Es gilt $\angle(a, b) \in [0, \pi]$ aufgrund des Wertebereichs des Arcuscosinus. Weiterhin enthält nach Konstruktion Definition 2.34 den bislang verwendeten Winkelbegriff in der Ebene als Spezialfall.

Bemerkung 2.35: Wegen $\angle(a, b) \in [0, \pi]$ handelt es sich bei dem Winkel im Sinne von Definition 2.34 immer um den kleineren der Winkel zwischen den entsprechenden Ortsvektoren.



Beispiel 2.36: Wir bestimmen für $a = (1, 3, 2)^T$ und $b = (2, -1, 4)^T$ den Cosinus des Winkels θ zwischen a und b . Nach Definition 2.34 gilt

$$\cos \theta = \frac{\langle a, b \rangle}{\|a\| \|b\|} = \frac{2 - 3 + 8}{\sqrt{14} \sqrt{21}} = \frac{7}{\sqrt{2} \sqrt{7} \sqrt{3} \sqrt{7}} = \frac{1}{\sqrt{6}},$$

also $\theta \approx 65,91^\circ$.

Beispiel 2.37: Gegeben seien die Punkte $P = (3, -1, 2)$, $Q = (2, 1, 7)$, $R = (4, 0, 3)$. Wir bestimmen den Cosinus des Winkels θ zwischen den Vektoren \vec{PQ} und \vec{PR} . Man berechnet $\vec{PQ} = Q - P = (-1, 2, 5)^T$, $\vec{PR} = R - P = (1, 1, 1)^T$ und daraus

$$\cos \theta = \frac{\langle \vec{PQ}, \vec{PR} \rangle}{\|\vec{PQ}\| \|\vec{PR}\|} = \frac{-1 + 2 + 5}{\sqrt{30} \sqrt{3}} = \frac{6}{3\sqrt{10}} = \frac{1}{5} \sqrt{10}.$$

2.1.4 Das Vektorprodukt (Kreuzprodukt)

In diesem Abschnitt beziehen wir uns ausschließlich auf \mathbb{R}^3 . Wir wollen nun zu zwei gegebenen Vektoren $a = (a_1, a_2, a_3)^T$ und $b = (b_1, b_2, b_3)^T$, die nicht parallel sein sollen, einen Vektor konstruieren, der (bezogen auf das euklidische Skalarprodukt) senkrecht auf a und b steht. Die geometrische Intuition legt nahe, dass es einen solchen Vektor immer gibt. Um dies streng zeigen zu können, definieren wir zunächst das Vektorprodukt als algebraische Operation und werden dann nachweisen, dass das Ergebnis alle gewünschten geometrischen Eigenschaften besitzt.

Definition 2.38: Seien a, b wie oben. Dann heißt

$$a \times b := \begin{pmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{pmatrix}$$

das *Vektorprodukt* oder *Kreuzprodukt* von a und b .

Beispiel 2.39: Für $a = (2, -1, -2)^T$ und $b = (3, 2, -1)^T$ gilt

$$a \times b = \begin{pmatrix} (-1) \cdot (-1) - (-2) \cdot 2 \\ (-2) \cdot 3 - 2 \cdot (-1) \\ 2 \cdot 2 - (-1) \cdot 3 \end{pmatrix} = \begin{pmatrix} 5 \\ -4 \\ 7 \end{pmatrix}.$$

Wir stellen zwei Schemata zur Berechnung des Kreuzproduktes per Hand vor. Beim ersten Rechenschema multipliziert man über Kreuz entlang der Pfeile und subtrahiert anschließend die Produkte („blau minus rot“).

$$\begin{pmatrix} 2 \\ -1 \\ -2 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix} \quad \begin{pmatrix} 2 \\ -1 \\ -2 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix} \quad \begin{pmatrix} 2 \\ -1 \\ -2 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix}$$

$$\begin{array}{lll} \text{1. Komponente:} & \text{2. Komponente:} & \text{3. Komponente:} \\ 1 - (-4) = 5 & -6 - (-2) = -4 & 4 - (-3) = 7 \end{array}$$

Beim zweiten Rechenschema verlängert man die beiden Vektoren um ihre ersten beiden Komponenten multipliziert wie durch die Pfeile angedeutet. Man notiert die Ergebnisse am linken und rechten Rand und zieht abschließend die Produkte am linken Rand von denen am rechten Rand ab:

$$\begin{array}{ccc} & \begin{pmatrix} 2 \\ -1 \\ -2 \end{pmatrix} & \begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix} \\ -4 & & 1 \\ -2 & \leftarrow 2 & \rightarrow -6 \\ -3 & -1 & 4 \end{array} \Rightarrow a \times b = \begin{pmatrix} 1 - (-4) \\ -6 - (-2) \\ 4 - (-3) \end{pmatrix} = \begin{pmatrix} 5 \\ -4 \\ 7 \end{pmatrix}$$

Es gelten folgende Rechenregeln:

Satz 2.40: Für a, b wie oben gilt

1. $a \times b = -b \times a$.
2. $a \times a = 0 \in \mathbb{R}^3$.
3. Sei $c \in \mathbb{R}^3$. Dann gilt $(a + b) \times c = a \times c + b \times c$.
4. Für $\alpha \in \mathbb{R}$ gilt $\alpha a \times b = \alpha(a \times b) = a \times (\alpha b)$.
5. $\langle a \times b, a \rangle = \langle a \times b, b \rangle = 0$.

BEWEIS 1.) Es gilt

$$b \times a = \begin{pmatrix} b_2 a_3 - b_3 a_2 \\ b_3 a_1 - b_1 a_3 \\ b_1 a_2 - b_2 a_1 \end{pmatrix} = - \begin{pmatrix} b_3 a_2 - b_2 a_3 \\ b_1 a_3 - b_3 a_1 \\ b_2 a_1 - b_1 a_2 \end{pmatrix} = - \begin{pmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{pmatrix} = -(a \times b)$$

- 2.) Nach 1.) gilt $a \times a = -a \times a$.
- 3.), 4.) folgen durch elementare Rechnung wie in 1.)
- 5.) Es ist

$$\begin{aligned} \langle a \times b, a \rangle &= (a_2 b_3 - a_3 b_2)a_1 + (a_3 b_1 - a_1 b_3)a_2 + (a_1 b_2 - a_2 b_1)a_3 \\ &= a_1 a_2 b_3 - a_1 a_3 b_2 + a_2 a_3 b_1 - a_1 a_2 b_3 + a_1 a_3 b_2 - a_2 a_3 b_1 \\ &= 0 \end{aligned}$$

Analog folgt die zweite Aussage. ■

Mit Aussage 5.) des Satzes 2.40 ist die Frage nach der Existenz eines Vektors, der auf a und b senkrecht steht, noch nicht vollständig beantwortet. Es könnte ja sein, dass man durch $a \times b$ den Nullvektor erhält, der immer auf a und b senkrecht steht. Dies ist nicht der senkrechte Vektor, nach dem man sucht. Von jetzt an seien immer $a, b \in \mathbb{R}^3 \setminus \{0\}$.

Satz 2.41: Die Vektoren a, b sind genau dann nicht parallel, wenn $a \times b \neq 0$ gilt.

BEWEIS „ \Leftarrow “ Wir nehmen an, a und b seien parallel. Dann existiert $\alpha \in \mathbb{R} \setminus \{0\}$ mit $\alpha a = b$. Nach Satz 2.40 erhält man $a \times b = \alpha a \times a = 0$.
 „ \Rightarrow “ Sind a und b nicht parallel, dann ist $b + \alpha a \neq 0$ für alle $\alpha \in \mathbb{R}$: sonst gäbe es ja irgendein $\hat{\alpha}$ mit $b + \hat{\alpha} a = 0 \Leftrightarrow b = -\hat{\alpha} a$, und a und b wären parallel. Aufgrund von $a \neq 0$ besitzt a mindestens eine Komponente $a_i \neq 0$. Wir nehmen o. B. d. A. an, dass $a_1 \neq 0$, die anderen Fälle werden analog behandelt. Sei $\lambda = -b_1/a_1$. Für $c := b + \lambda a = (0, c_2, c_3)$ folgt $c \neq 0$ aufgrund obiger Überlegungen. Damit gilt $c_2 \neq 0$ oder auch $c_3 \neq 0$. Weiterhin gilt

$$a \times c = a \times (b + \lambda a) = a \times b + \lambda a \times a = a \times b$$

und damit

$$a \times b = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \times \begin{pmatrix} 0 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} a_2 c_3 - a_3 c_2 \\ -a_1 c_3 \\ a_1 c_2 \end{pmatrix}.$$

Aufgrund von $a_1 \neq 0$ folgt entweder $-a_1 c_3 \neq 0$ oder $a_1 c_2 \neq 0$ und damit $a \times b \neq 0$. ■

Wir kommen nun zur Frage nach der Länge von $a \times b$.

Satz 2.42: Für a, b gilt

$$\|a \times b\|^2 = \|a\|^2 \|b\|^2 - \langle a, b \rangle^2.$$

BEWEIS Ausrechnen und geschicktes Ergänzen mit 0 (Terme in blau dargestellt) ergibt

$$\begin{aligned} \|a \times b\|^2 &= \langle a \times b, a \times b \rangle \\ &= (a_2 b_3 - a_3 b_2)^2 + (a_3 b_1 - a_1 b_3)^2 + (a_1 b_2 - a_2 b_1)^2 \\ &= (a_2 b_3)^2 + (a_3 b_2)^2 + (a_3 b_1)^2 + (a_1 b_3)^2 + (a_1 b_2)^2 + (a_2 b_1)^2 \\ &\quad - 2a_2 a_3 b_2 b_3 - 2a_1 a_3 b_1 b_3 - 2a_1 a_2 b_1 b_2 \\ &\quad + (a_1 b_1)^2 + (a_2 b_2)^2 + (a_3 b_3)^2 - ((a_1 b_1)^2 + (a_2 b_2)^2 + (a_3 b_3)^2) \\ &= \underbrace{(a_1^2 + a_2^2 + a_3^2)}_{=\|a\|^2} \underbrace{(b_1^2 + b_2^2 + b_3^2)}_{=\|b\|^2} \\ &\quad - 2(a_2 a_3 b_2 b_3 + a_1 a_3 b_1 b_3 + a_1 a_2 b_1 b_2) - ((a_1 b_1)^2 + (a_2 b_2)^2 + (a_3 b_3)^2) \end{aligned}$$

Andererseits gilt nach der ersten binomischen Formel

$$\begin{aligned} \langle a, b \rangle^2 &= (a_1 b_1 + a_2 b_2 + a_3 b_3)^2 \\ &= (a_1 b_1 + (a_2 b_2 + a_3 b_3))^2 \\ &= a_1^2 b_1^2 + 2(a_1 a_2 b_1 b_2 + a_1 a_3 b_1 b_3) + (a_2 b_2 + a_3 b_3)^2 \\ &= a_1^2 b_1^2 + 2a_1 a_2 b_1 b_2 + 2a_1 a_3 b_1 b_3 + a_2^2 b_2^2 + a_3^2 b_3^2 + 2a_2 a_3 b_2 b_3. \end{aligned}$$

Dies sind genau die Terme in der letzten Zeile der ersten Rechnung. ■

Folgerung 2.43: Sei θ der Winkel zwischen a und b . Dann gilt

$$\|a \times b\| = \|a\| \|b\| \sin \theta.$$

BEWEIS Nach Definition 2.34 des Winkels gilt $\theta \in [0, \pi]$, also ist $\sin \theta \geq 0$. Wiederum mit Definition 2.34 und Satz 2.42 gilt

$$\begin{aligned} \|a \times b\|^2 &= \|a\|^2 \|b\|^2 - \langle a, b \rangle^2 \\ &= \|a\|^2 \|b\|^2 (1 - \cos^2 \theta) \\ &= \|a\|^2 \|b\|^2 \sin^2 \theta. \end{aligned}$$

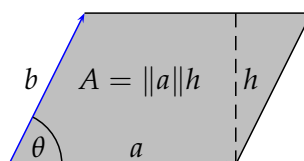


Abbildung 2.6: Fläche A des durch a und b aufgespannten Parallelogramms

Für die Höhe h eines von a und b aufgespannten Parallelogramms entnimmt man Abb. 2.6 $h = \sin \theta \|b\|$; für dessen Fläche A folgt $A = \|a\|h = \|a\| \|b\| \sin \theta$. Geometrisch lässt sich die Länge des Vektorprodukts daher als Flächeninhalt des von den beiden Vektoren aufgespannten Parallelogramms deuten.

Bemerkung 2.44: Das Vektorprodukt ist i. A. nicht assoziativ. Seien z. B. $a = (1, 0, 0)^T$, $b = (0, 1, -1)^T$ und $c = (0, 0, 1)^T$. Dann gilt $a \times b = (0, 1, 1)^T$ und $b \times c = (1, 0, 0)^T$. Daraus folgt $(a \times b) \times c = (1, 0, 0)^T$, aber $a \times (b \times c) = (0, 0, 0)^T$. Es gilt jedoch die *Graßmannsche Identität*

$$a \times (b \times c) = \langle a, c \rangle b - \langle a, b \rangle c,$$

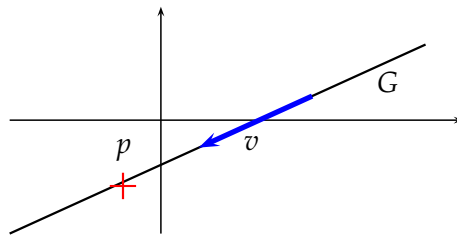
die man direkt nachrechnet.

Bemerkung 2.45: Alle bisher diskutierten geometrischen Eigenschaften von $a \times b$ gelten auch für $-(a \times b)$. Die Definition des Kreuzproduktes gewährleistet aber, dass die drei Vektoren $a, b, a \times b$ so gedreht werden können, dass sie in ihrer Lage zueinander den kanonischen Einheitsvektoren e_1, e_2, e_3 entsprechen. Man spricht von einem *Rechtssystem*: Versucht man, an der rechten Hand Daumen, Zeige- und Mittelfinger so zu halten, dass alle senkrecht zueinander stehen, und identifiziert man den Daumen mit e_1 , den Zeigefinger mit e_2 und den Mittelfinger mit e_3 , dann kann man so die Lage der kanonischen Einheitsvektoren zueinander nachempfinden („Rechte-Hand-Regel“). Für $a, b, -a \times b$ gilt das nicht: Bei jeder Drehung zeigt mindestens ein Vektor in Gegenrichtung eines kanonischen Einheitsvektors; $a, b, -a \times b$ bilden ein *Linkssystem*.

2.2 Geraden und Ebenen

2.2.1 Definition und grundlegende Eigenschaften

Bei der Definition einer Geraden in der Ebene gehen wir direkt von der Anschauung aus. Von jetzt an gelte immer $\alpha, \beta \in \mathbb{R}$ und $n, p, v \in \mathbb{R}^2$.

**Definition 2.46:**

1. Für einen Ortsvektor oder Aufpunkt p und einen Richtungsvektor $v \neq 0$ heie

$$G := \{x \in \mathbb{R}^2 \mid \exists \alpha \in \mathbb{R} : x = p + \alpha v\} \quad (2.11)$$

eine Gerade.

2. Die Gleichung

$$x = p + \alpha v$$

aus (2.11) heit *Punkt-Richtungsgleichung* von G , die reelle Zahl α aus Formel (2.11) nennt man *Parameter*.

Bemerkung 2.47: Wir identifizieren im Folgenden hufig eine Gerade, die ja als Teilmenge von \mathbb{R}^2 definiert ist, mit ihrer Geradengleichung, weil sie die Gerade eindeutig festlegt.

Nach Definition 2.46 legen Aufpunkt und Richtungsvektor eine Gerade eindeutig fest, jeder Punkt x auf der Geraden gengt der Geradengleichung $x = p + \alpha v$ fr ein bestimmtes α . Andererseits sind bei einer gegebenen Gerade Aufpunkt und Richtungsvektor nicht eindeutig bestimmt.

Satz 2.48: Sei G eine Gerade mit dem Richtungsvektor v und dem Aufpunkt p .

1. Jeder Vektor der Form $\tilde{v} = \beta v$ mit $\beta \neq 0$ ist ebenfalls Richtungsvektor.
2. Jeder Vektor der Form $\tilde{p} = p + \gamma v$ mit irgendeinem $\gamma \in \mathbb{R}$ ist ebenfalls Aufpunkt.

BEWEIS 1.) Sei G gegeben durch $x = p + \alpha v$ und \tilde{G} durch $x = p + \tilde{\alpha} \tilde{v}$. Wir zeigen $G = \tilde{G}$. Man hat

$$\begin{aligned} x \in G &\Leftrightarrow \exists \hat{\alpha} : x = p + \hat{\alpha} v \\ &\Leftrightarrow \exists \hat{\alpha} : x = p + \frac{\hat{\alpha}}{\beta} (\beta v) \\ &\Leftrightarrow \exists \tilde{\alpha} := \frac{\hat{\alpha}}{\beta} : x = p + \tilde{\alpha} \tilde{v} \Leftrightarrow x \in \tilde{G}. \end{aligned}$$

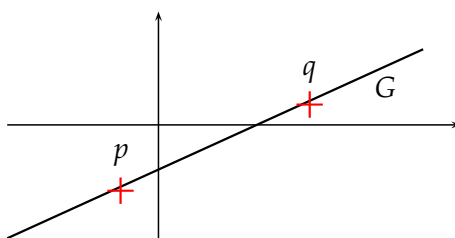
Also sind die beiden Mengen gleich, und es gilt $G = \tilde{G}$. Man beachte, dass man hier wegen $\beta \neq 0$ nicht durch Null dividiert.

2.) Sei G gegeben durch $x = p + \alpha v$ und \tilde{G} durch $x = \tilde{p} + \tilde{\alpha} v$. Dann gilt

$$\begin{aligned} x \in G &\Leftrightarrow \exists \hat{\alpha} : x = p + \hat{\alpha} v \\ &\Leftrightarrow \exists \hat{\alpha} : x = (p + \gamma v) + \hat{\alpha} v - \gamma v \\ &\Leftrightarrow \exists \tilde{\alpha} := \hat{\alpha} - \gamma : x = \tilde{p} + \tilde{\alpha} v \Leftrightarrow x \in \tilde{G} \end{aligned}$$

und daher $G = \tilde{G}$. ■

Wir gehen nun davon aus, dass anstelle eines Aufpunktes und eines Richtungsvektors von einer Geraden zwei verschiedene Punkte bekannt sind.



Satz 2.49: Seien $p, q \in \mathbb{R}^2$ mit $p \neq q$. Dann gibt es genau eine Gerade durch p und q .

BEWEIS Existenz Wir wählen als Aufpunkt p und als Richtungsvektor $v = q - p$, die hierdurch festgelegte Gerade sei G mit der Geradengleichung

$$x = p + \alpha(q - p).$$

Es gilt $p = p + 0v \in G$ und $q = p + 1(q - p) = p + 1v \in G$, also enthält G sowohl p als auch q .

Eindeutigkeit Sei \tilde{G} noch eine Gerade mit der Geradengleichung $x = \tilde{p} + \alpha \tilde{v}$, die p und q enthält. Weil nach Satz 2.48.2 sich jeder Punkt auf \tilde{G} , also auch p , als Aufpunkt eignet, wird \tilde{G} auch durch die Geradengleichung $x = p + \alpha \tilde{v}$ festgelegt. Wegen $q \in \tilde{G}$ gilt $q = p + \alpha \tilde{v}$ für ein bestimmtes $\alpha \in \mathbb{R}$. Aufgrund von $p \neq q$ erhält man $\alpha \neq 0$ und durch Subtraktion von p auf beiden Seiten:

$$v = q - p = \alpha \tilde{v}.$$

Beide Richtungsvektoren sind somit Vielfache voneinander; aus Satz 2.48.1 folgt $G = \tilde{G}$, also die Eindeutigkeit. ■

Definition 2.50:

1. Seien p, q zwei verschiedene Punkte auf der Geraden G . Dann heißt die Geradengleichung

$$x = p + \alpha(q - p)$$

Zweipunktform von G . Man sagt, „die Gerade liegt in der Zwei-Punkte-Form vor“.

2. Alle Gleichungen, die einen reellen Parameter enthalten und eine Gerade beschreiben, nennt man *Parametergleichungen* oder *Parameterformen* einer Geraden.

Bemerkung 2.51:

1. Nach Satz 2.48 ist man frei, statt eines vorliegenden Richtungsvektors ein Vielfaches davon zu wählen. Das kann man in der Praxis nutzen, um einen Richtungsvektor mit „möglichst einfachen Zahlen“ zu finden. Ersetzt man z. B. den Richtungsvektor $(\pi/3, \pi/6)^T$ durch $(2, 1)^T$, wird das in vielen Fällen Berechnungen wesentlich handhabbarer werden lassen. Analoges gilt für die Wahl des Aufpunkts.
2. Die Umrechnung von der Zwei-Punkte-Form in die Punkt-Richtungsform und umgekehrt ist einfach möglich: Liegt die Zwei-Punkte-Form vor, so findet man z. B. durch $q - p$ einen Richtungsvektor; als Aufpunkt kann man z. B. p oder q wählen. Umgekehrt findet man ausgehend vom Richtungsvektor v und dem Aufpunkt p z. B. durch $p + v$ oder $p - v$ einen weiteren Punkt auf der Geraden.
3. Dass man durch zwei verschiedene Punkte genau eine Gerade legen kann, hätte man sicherlich auch ohne höhere Mathematik anhand einiger Experimente mit Papier und Lineal einsehen können. Entscheidend ist aber, dass man diese Tatsache jetzt lückenlos bewiesen hat, ohne an eine wie auch immer geartete Einsicht appellieren zu müssen. Zudem ist es ein Indiz dafür, dass die Definition 2.46 des geometrischen Objekts „Gerade“ mittels einer algebraischen Gleichung zielführend ist, weil typische Eigenschaften einer Geraden in der Tat aus der Definition folgen.

Wir betrachten nun alternative Beschreibungen von Geraden in der Ebene, bei denen auf einen Parameter verzichtet wird. Vorbereitend zeigen wir folgendes Lemma.

Lemma 2.52: Seien $a, b, c \in \mathbb{R}^2 \setminus \{0\}$ beliebig, $a \perp b$ und $b \perp c$. Dann gilt $a \parallel c$.

BEWEIS Sei $a = (a_1, a_2)^T$. Dann muss nach Bemerkung 2.21 $b = (b_1, b_2)^T = \alpha(-a_2, a_1)^T$ gelten und abermals nach Bemerkung 2.21 $c = \beta(-b_2, b_1)^T = \alpha\beta(-a_1, -a_2)^T$. ■

Satz 2.53: Sei $G \subset \mathbb{R}^2$ eine Gerade mit Richtungsvektor v und Aufpunkt p . Für jedes $n \neq 0$ mit $n \perp v$ gilt

$$x \in G \Leftrightarrow \langle x, n \rangle = \langle p, n \rangle.$$

Der Vektor n ist bis auf Skalierung eindeutig bestimmt.

BEWEIS „ \Rightarrow “: Sei $x \in G$. Dann existiert $\alpha : x = p + \alpha v$. Es folgt mit den Rechenregeln des Skalarprodukts:

$$\langle x, n \rangle = \langle p, n \rangle + \underbrace{\alpha \langle v, n \rangle}_{=0}$$

„ \Leftarrow “: Offenbar ist $p \in G$. Erfülle $x \neq p$ die Gleichung $\langle x, n \rangle = \langle p, n \rangle$. Es folgt $\langle x - p, n \rangle = 0$, daher $0 \neq (x - p) \perp n$ und nach Lemma 2.52 $x - p \parallel v$, also nach Definition 1.9 $x - p = \alpha v$ für ein $\alpha \in \mathbb{R}$. Es folgt durch Addition von p auf beiden Seiten $x = p + \alpha v$.

Die Eindeutigkeit von n bis auf Skalierung folgt sofort aus Bemerkung 2.21.

■

Definition 2.54: Sei G eine Gerade in der Ebene und p der Aufpunkt von G .

1. Ein Vektor n wie in Satz 2.53 heißt *Normalenvektor* von G .
2. Die Gleichung

$$\langle x, n \rangle = \langle p, n \rangle$$

heißt *Normalform* von G .

Wir haben oben bereits gesehen, dass jeder Punkt auf der Geraden als Aufpunkt gewählt werden kann. Es bleibt noch zu zeigen, dass auch die Normalform einer Geraden nicht von der Wahl des Aufpunkts abhängt.

Bemerkung 2.55: Seien p, q zwei Punkte auf der Geraden G mit Normalenvektor n . Dann gilt

$$\langle p, n \rangle = \langle q, n \rangle.$$

BEWEIS Die Behauptung folgt mit $x = q$ sofort aus Satz 2.53. ■

Bemerkung 2.56:

1. Seien G, n und p wie zuvor. Mit $n = (a, b)^T \in \mathbb{R}^2$, $x = (x_1, x_2)^T$ und $c := \langle p, n \rangle \in \mathbb{R}$ folgt die *allgemeine Geradengleichung*

$$ax_1 + bx_2 = c.$$

2. Gilt $b \neq 0$, hat also der Richtungsvektor eine x_1 -Komponente, setzt man $m = -a/b$ und $n = c/b$ (dieses n ist an dieser Stelle *nicht* der Normalenvektor, sondern eine reelle Zahl) und erhält so die vielleicht aus der Schule bekannte Geradengleichung

$$y = mx + n,$$

wobei hier $x := x_1$ und $y := x_2$ gesetzt wird, um auch in der Notation den Bezug zu vielleicht Bekanntem herzustellen. Es handelt sich also um einen Spezialfall einer Normalform einer Geraden.

Beispiel 2.57: Seien $p = (1, 2)^T$, $n = (4, 3)^T$ und $x = (x_1, x_2)^T$. Dann gilt:

$$\begin{aligned} \langle x, n \rangle &= \langle p, n \rangle \\ \Leftrightarrow \left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} 4 \\ 3 \end{pmatrix} \right\rangle &= \left\langle \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 \\ 3 \end{pmatrix} \right\rangle \\ \Leftrightarrow 4x_1 + 3x_2 &= 10 \end{aligned}$$

Wir erhalten also eine allgemeine Geradengleichung zur Beschreibung der Geraden. Man beachte, dass die Vorfaktoren von x_1 und x_2 genau den Komponenten des Normalenvektors entsprechen.

Die Normalform ist nur bis auf einen Faktor genau bestimmt. Etwas eindeutiger wird die Darstellung in der Hesseschen Normalform.

Definition 2.58: Sei G eine Gerade mit Normalenvektor n . Gilt $\|n\| = 1$, so heißt die damit gebildete Normalform *Hessesche Normalform*.

Bemerkung 2.59: Man erhält die Hessesche Normalform aus einer beliebigen Normalform, indem man die Normalform durch $\|n\|$ teilt:

$$\frac{\langle x, n \rangle}{\|n\|} = \frac{\langle p, n \rangle}{\|n\|}.$$

Damit liegt der Normalvektor bis auf das Vorzeichen eindeutig fest.

Beispiel 2.60: Die Gerade

$$4x_1 + 3x_2 = 10$$

hat den Normalenvektor $n = (4, 3)^T$, was sich direkt aus den Vorfaktoren der linken Seite ablesen lässt. Damit ist

$$\|n\| = \sqrt{16 + 9} = 5,$$

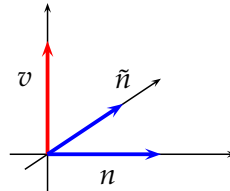
und die Hessesche Normalform lautet

$$\frac{1}{5}(4x_1 + 3x_2) = \frac{10}{5} = 2.$$

Wir kommen nun zu Geraden im Raum. Wir definieren sie völlig analog zu Geraden in der Ebene und verweisen dazu auf den Wortlaut von Definition 2.46, nur dass jetzt die Vektoren p, q, x Elemente von \mathbb{R}^3 sind statt von \mathbb{R}^2 wie zuvor. Die Sätze 2.48 und 2.49 gelten dann wortgleich mit wortgleichen Beweisen (es wurde dort niemals benutzt, dass Vektoren nur aus zwei Komponenten bestanden), ebenso lässt sich Definition 2.50 wortgleich auf den vorliegenden Fall übertragen.

Im Gegensatz zur Ebene existiert keine parameterlose Beschreibung einer Geraden im Raum. Dies liegt daran, dass die Richtung eines Normalenvektors zu einem Richtungsvektor nicht eindeutig festliegt.

Beispiel 2.61: Sei $v = (0, 0, 1)^T$ der Richtungsvektor einer Geraden. Dann ist $n = (1, 0, 0)^T$ ein Normalenvektor aufgrund von $\langle v, n \rangle = 0$, aber genauso auch $\tilde{n} = (0, 1, 0)^T$ und damit nach den Rechenregeln des Skalarprodukts auch alle Vektoren der Form $\alpha n + \beta \tilde{n}$, also jeder Vektor, der in der (x_1, x_2) -Ebene liegt.



Wir gehen nun noch kurz auf Ebenen im Raum ein. Den Vektorraum \mathbb{R}^2 , der ja sich geometrisch als Ebene deuten lässt, kann man mit der Teilmenge

$$\tilde{E} := \left\{ x \in \mathbb{R}^3 \mid x = (x_1, x_2, 0)^T; x_1, x_2 \in \mathbb{R} \right\}$$

identifizieren. Es gilt offensichtlich

$$\tilde{E} := \left\{ x \in \mathbb{R}^3 \mid x = (0, 0, 0)^T + x_1(1, 0, 0)^T + x_2(0, 1, 0)^T, x_1, x_2 \in \mathbb{R} \right\};$$

die Ebene \tilde{E} besteht daher genau aus allen Vektoren der Form $0 + \alpha(1, 0, 0)^T + \beta(0, 1, 0)^T$ mit den reellen Parametern α und β . Man beachte, dass die beiden Vektoren unterschiedliche Richtungen aufweisen. Verallgemeinernd definieren wir:

Definition 2.62: Seien $p, v, w \in \mathbb{R}^3$, $v \neq 0$ und $w \neq 0$, und seien v und w nicht parallel. Dann heißt

$$E := \left\{ x \in \mathbb{R}^3 \mid x = p + \alpha v + \beta w, \alpha, \beta \in \mathbb{R} \right\}$$

Ebene; die Vektoren v und w heißen *Richtungsvektoren*. Die Gleichung $x = p + \alpha v + \beta w$ wird *Punkt-Richtungsgleichung* genannt; man sagt, die „Ebene liegt in Punkt-Richtungsform vor“.

Bemerkung 2.63:

1. Ebenso wie bei einer Geraden in der Ebene in Parameterform eignet sich jeder Punkt der Ebene als Aufpunkt; beliebige zwei nichtparallele Vektoren in der Ebene ungleich dem Nullvektor eignen sich als Richtungsvektoren.
2. Drei paarweise verschiedene Punkte, die nicht auf einer Geraden liegen, legen eine Ebene eindeutig fest.

BEWEIS Die Beweise werden ähnlich zu den Beweisen der entsprechenden Aussagen für Geraden in der Ebene geführt, sind aber etwas komplizierter. Wir verzichten auf Details. ■

Ebenso ohne Beweis geben wir folgenden Satz an.

Satz 2.64: Sei E eine Ebene mit Aufpunkt p und den Richtungsvektoren v und w . Dann existiert ein Vektor $n \neq 0$ mit $v \perp n$, $w \perp n$ und

$$x \in E \Leftrightarrow \langle x, n \rangle = \langle p, n \rangle.$$

Der Vektor n ist bis auf Skalierung eindeutig bestimmt.

Definition 2.65: Der Vektor n aus Satz 2.64 heißt *Normalenvektor* der Ebene, die Gleichung

$$\langle x, n \rangle = \langle p, n \rangle \quad (2.12)$$

heißt *Normalform* oder *Normalgleichung* der Ebene.

Bemerkung 2.66:

1. Die Normalform hängt nicht von der Wahl des Aufpunkts p ab.
2. Sei $n = (a, b, c)^T$ und $d = \langle p, n \rangle \in \mathbb{R}$. Dann erhält man aus (2.12) sofort die *allgemeine Ebenengleichung*

$$ax_1 + bx_2 + cx_3 = d.$$

3. Man erklärt die Hessesche Normalform einer Ebene sinngemäß zu Definition 2.58; sie ist bis auf das Vorzeichen eindeutig.

Wir fassen unsere Ergebnisse in der folgenden Tabelle zusammen.

Raum	Objekt	Punkt-Richtungs-Gl.	Normalgleichung
\mathbb{R}^2	Gerade	$x = p + \alpha v$	$\langle x, n \rangle = \langle p, n \rangle$ $ax_1 + bx_2 = c$
\mathbb{R}^3	Gerade	$x = p + \alpha v$	Existiert nicht!
\mathbb{R}^3	Ebene	$x = p + \alpha v + \beta w$	$\langle x, n \rangle = \langle p, n \rangle$ $ax_1 + bx_2 + cx_3 = d$

2.2.2 Umrechnen zwischen verschiedenen Darstellungsformen

Wir diskutieren nun die Umrechnung zwischen der Parameterform einer Geraden und ihrer Normalform als parameterfreie Darstellung. Da es in \mathbb{R}^3 keine Normalform von Geraden gibt, beziehen wir uns auf Geraden in \mathbb{R}^2 . Zur sehr einfachen Umrechnung zwischen der Punkt-Richtungsform und der Zwei-Punkte-Form verweisen wir auf Bemerkung 2.51 und betrachten

nur die Punkt-Richtungsform.

Sei eine Gerade G gegeben durch $x = p + \alpha v$ mit $v = (v_1, v_2)^T$. Einen Normalenvektor n findet man durch

$$n := \begin{pmatrix} v_2 \\ -v_1 \end{pmatrix}.$$

Dies ist nach Bemerkung 2.21 bis auf Skalierung die einzig mögliche Wahl. Durch Ausrechnen von $\langle p, n \rangle$ erhält man die rechte Seite der Normalform.

Beispiel 2.67: Eine Gerade sei gegeben durch

$$x = \begin{pmatrix} -2 \\ -2 \end{pmatrix} + \alpha \begin{pmatrix} 2 \\ 3 \end{pmatrix}.$$

Wir lesen $p = (-2, -2)^T$ und $v = (2, 3)^T$ ab, erhalten $n = (3, -2)^T$ und daher mit $x = (x_1, x_2)^T$ eine Normalform

$$\langle x, n \rangle = \langle p, n \rangle \Leftrightarrow 3x_1 - 2x_2 = -2.$$

Liegt die Gerade in der Normalform $ax_1 + bx_2 = c$ vor, wird ein Richtungsvektor $v \perp n = (a, b)^T$ benötigt. Man kann hier $v = (b, -a)^T$ wählen. Einen Aufpunkt p erhält man, in dem man z. B. $x_1 = 0$ oder $x_2 = 0$ setzt und aus der parameterlosen Form die andere Komponente errechnet.

Beispiel 2.68: Eine Gerade in \mathbb{R}^2 liegt in ihrer Normalform $3x_1 - 2x_2 = -2$ vor. Man liest $n = (3, -2)^T$ ab und wählt $v = (2, 3)^T$. Für $x_1 = 0$, folgt aus der Normalform $x_2 = 1$ und damit $p = (0, 1)^T$. Wir erhalten

$$x = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \alpha \begin{pmatrix} 2 \\ 3 \end{pmatrix}.$$

Wir kommen nun zur Umrechnung von Ebenendarstellungen. Zur Umrechnung der Punkt-Richtungsform in die Normalform benötigt man einen Normalenvektor. In \mathbb{R}^3 kann man dazu das Kreuzprodukt der beiden Richtungsvektoren verwenden. Als Aufpunkt lässt sich jeder Punkt der Ebene verwenden, insbesondere der Vektor p aus der Punkt-Richtungsform.

Beispiel 2.69: Eine Ebene in \mathbb{R}^3 sei durch ihre Punkt-Richtungsform

$$x = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} + \alpha \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$$

definiert. Wir erhalten einen Normalenvektor n durch

$$n = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \times \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ 3 \end{pmatrix}$$

und mit $x = (x_1, x_2, x_3)^T$ die Normalform

$$\left\langle \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \\ 3 \end{pmatrix} \right\rangle = \left\langle \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \\ 3 \end{pmatrix} \right\rangle \Leftrightarrow -x_1 - x_2 + 3x_3 = 1.$$

Zur Umrechnung einer Normalform einer Ebene in die Punkt-Richtungsform benötigen wir zwei nicht parallele Richtungsvektoren, die beide senkrecht auf n stehen müssen. Aus der Normalgleichung $ax_1 + bx_2 + cx_3 = d$ liest man den Normalenvektor $n = (a, b, c)^T$ ab. Mindestens eine Komponente n_i von n ist ungleich 0. Wir vertauschen n_i mit einer anderen Komponente n_j und verändern das Vorzeichen von n_i im so erzeugten Vektor. Weil es zwei Möglichkeiten gibt, $j \neq i$ zu wählen, erhalten wir zwei Vektoren v und w . Man erkennt sofort $v \perp n$ und $w \perp n$, und weder v noch w sind der Nullvektor. Aufgrund der Position der Nullen können v und w nicht parallel sein. Ein Aufpunkt lässt sich errechnen, indem man zwei der drei Koordinaten von $x = (x_1, x_2, x_3)^T$ z. B. den Wert 0 zuweist und dann aus der Normalform den Wert der fehlenden Koordinate errechnet. Sollte dies nicht möglich sein, wähle man ein anderes Koordinatenpaar. Es gibt immer zwei Koordinaten in x , mit denen obige Rechnung möglich ist.

Beispiel 2.70: Aus der Normalform

$$-x_1 - x_2 + 3x_3 = 1$$

einer Ebene liest man den Normalenvektor $n = (-1, -1, 3)^T$ ab und erhält als Richtungsvektoren $v = (1, -1, 0)^T$ und $w = (3, 0, 1)^T$. Für $x_2 = x_3 = 0$ folgt aus der Normalform $x_1 = -1$. Das führt zu $p = (-1, 0, 0)^T$ und damit zur Punkt-Richtungsform

$$x = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}.$$

2.2.3 Lagebeziehungen von Geraden und Ebenen

Wir erklären zunächst Orthogonalität und Winkel zwischen Geraden und Ebenen.

Definition 2.71:

1. Zwei Geraden in \mathbb{R}^2 oder \mathbb{R}^3 heißen *parallel*, wenn ihre Richtungsvektoren parallel sind.
2. Zwei sich schneidende Geraden heißen *orthogonal*, wenn ihre Richtungsvektoren orthogonal sind.

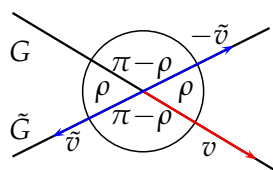


Abbildung 2.7: Winkel zwischen zwei Geraden

3. Seien G und \tilde{G} zwei sich schneidende Geraden mit Richtungsvektoren v bzw. \tilde{v} . Der Winkel $\angle(G, \tilde{G})$ zwischen den Geraden wird definiert durch

$$\angle(G, \tilde{G}) := \min\{\angle(v, \tilde{v}), \angle(v, -\tilde{v})\}$$

Bemerkung 2.72:

1. Es ist nicht möglich, den Winkel zwischen zwei Geraden einfach als den Winkel zwischen den beiden Richtungsvektoren der Geraden zu definieren, weil diese ja nicht eindeutig festliegen. Ersetzt man den Richtungsvektor \tilde{v} durch $-\tilde{v}$, dann wird aus dem Winkel ρ zwischen den Richtungsvektoren der Winkel $180^\circ - \rho$, obwohl die Geraden unverändert geblieben sind (vgl. Abb. 2.7). Skalieren des Richtungsvektors mit einer positiven Zahl dagegen erhält den Winkel.
2. Abb. 2.7 zeigt zudem, dass der Winkel zwischen zwei Geraden höchstens 90° beträgt.
3. Für den Winkel zwischen zwei Vektoren v, \tilde{v} gilt

$$\angle(v, \tilde{v}) = \arccos\left(\frac{\langle v, \tilde{v} \rangle}{\|v\| \|\tilde{v}\|}\right) \quad (2.13)$$

(vgl. Definition 2.34), so dass Winkel größer als 90° gerade einem negativen Wert des Skalarprodukts im Zähler entsprechen. Der Übergang von \tilde{v} auf $-\tilde{v}$ entspricht genau einer Vorzeichenumkehr des Bruches in Gleichung (2.13). Daher gilt für den Winkel zwischen zwei sich schneidenden Geraden G und \tilde{G} :

$$\angle(G, \tilde{G}) = \arccos\left(\frac{|\langle v, \tilde{v} \rangle|}{\|v\| \|\tilde{v}\|}\right). \quad (2.14)$$

Beispiel 2.73: Sei G durch

$$x = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} + \alpha \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$$

und \tilde{G} durch

$$x = \begin{pmatrix} 2 \\ -1 \\ 2 \end{pmatrix} + \alpha \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$$

gegeben. Dann gilt nach Formel (2.14):

$$\cos(\angle(G, \tilde{G})) = \frac{|-1|}{\sqrt{2}\sqrt{6}} = \frac{1}{\sqrt{12}}$$

Der gesuchte Winkel beträgt damit ca. $73,22^\circ$. Der Winkel zwischen den beiden Richtungsvektoren beträgt dagegen ca. $106,78^\circ$.

Definition 2.74:

1. Zwei Ebenen heißen *parallel*, wenn ihre Normalenvektoren parallel sind. Sie heißen *orthogonal*, wenn ihre Normalenvektoren orthogonal sind.
2. Sei E eine Ebene mit Normalenvektor n . Eine Gerade mit Richtungsvektor v heißt *parallel* zur Ebene E , falls $v \perp n$.
3. Seien n und \tilde{n} Normalenvektoren der beiden Ebenen E und \tilde{E} . Dann wird der Winkel $\angle(E, \tilde{E})$ zwischen den beiden Ebenen erklärt durch

$$\angle(E, \tilde{E}) := \min\{\angle(n, \tilde{n}), \angle(n, -\tilde{n})\}$$

Mit einer zum Fall der Geraden analogen Argumentation erkennt man

$$\angle(E, \tilde{E}) = \arccos\left(\frac{|\langle n, \tilde{n} \rangle|}{\|n\| \|\tilde{n}\|}\right). \quad (2.15)$$

Beispiel 2.75: Zwei Ebenen E und \tilde{E} seien gegeben durch

$$1x_1 + 2x_2 + 2x_3 = 0 \quad \text{bzw.} \quad -3x_1 + 4x_2 = 1.$$

Normalenvektoren lassen sich ablesen zu

$$n = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \quad \text{und} \quad \tilde{n} = \begin{pmatrix} -3 \\ 4 \\ 0 \end{pmatrix}.$$

Dadurch ergibt sich nach Formel (2.15)

$$\cos(\angle(E, \tilde{E})) = \frac{|\langle n, \tilde{n} \rangle|}{\|n\| \|\tilde{n}\|} = \frac{5}{3 \cdot 5} = \frac{1}{3} \Rightarrow \angle(E, \tilde{E}) \approx 70,52^\circ.$$

2.2.4 Schnittmengen zwischen Geraden und Ebenen in \mathbb{R}^2 und \mathbb{R}^3

Wir stellen nun einen Rechenweg vor, mit dem man die Schnittmenge zwischen Geraden und Ebenen bestimmen kann. Es existieren natürlich auch andere, gleichwertige Lösungswege. Wir werden Geraden in der Ebene und Ebenen im Raum gemeinsam behandeln, weil beide eine Normalform besitzen. In diesem Zusammenhang verwenden wir verkürzend den Begriff *Hyperebene*, wenn wir entweder eine Gerade in der Ebene oder eine Ebene im Raum meinen.

Sei im Folgenden zumindest eines der beiden Objekte, deren Schnittmenge bestimmt werden soll, eine Hyperebene. Die beschreibenden Gleichungen können parameterbehaftet oder parameterlos vorliegen. Folgender Rechenweg führt stets zur Schnittmenge.

Rechenweg 2.76 (Bestimmung der Schnittmenge):

1. Man sorgt eventuell durch Umrechnung dafür, dass das eine Objekt durch eine parameterlose und das andere durch eine parameterbehaftete Gleichung beschrieben wird.
2. Man setzt die Parametergleichung in die parameterlose Gleichung ein und erhält Ausdrücke für den oder die Parameter.
3. Diese setzt man in die Parametergleichung ein und erhält eine Parametrisierung der Schnittmenge.

Rechenweg 2.76 eignet sich nicht zur Bestimmung des Schnittpunkts zweier Geraden in \mathbb{R}^3 , weil diese keine parameterfreie Darstellung besitzen. Wir stellen diesen Fall zunächst zurück.

Beispiel 2.77: Gegeben seien $p = (1, -1, 2)^T$, $q = (1, 1, 1)^T$ und $n = (1, 2, 3)^T$. Gesucht wird der Schnittpunkt der Geraden G durch p in Richtung von n mit der Ebene E durch q senkrecht zu n . Man verwendet bei der Gleichsetzung für E eine Normal- und für G eine Parameterform, z. B.

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} + \alpha \cdot \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Komponentenweise liest man daraus die Gleichungen

$$x_1 = 1 + \alpha; \quad x_2 = -1 + 2\alpha; \quad x_3 = 2 + 3\alpha \quad (2.16)$$

ab. Eine Normalform von E lautet

$$\langle x, n \rangle = \langle q, n \rangle \Leftrightarrow x_1 + 2x_2 + 3x_3 = 6.$$

Einsetzen von (2.16) in die Normalgleichung liefert

$$\begin{aligned} 1 + \alpha + 2(-1 + 2\alpha) + 3(2 + 3\alpha) &= 6 \\ \Leftrightarrow 5 + 14\alpha &= 6 \\ \Leftrightarrow \alpha &= \frac{1}{14} \end{aligned}$$

Den Schnittpunkt s erhält man durch Einsetzen von α in die Parametergleichung:

$$s = p + \alpha n = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} + \frac{1}{14} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \left(\frac{15}{14}, -\frac{12}{14}, \frac{31}{14} \right)^T.$$

Es bleibt zu zeigen, dass Verfahren 2.76 wirklich immer zum Ziel führt und dass der gesuchte eindeutige Schnittpunkt tatsächlich existiert.

Satz 2.78: Sei E eine Hyperebene und G eine nichtparallele Gerade. Dann existiert ein eindeutiger Schnittpunkt von E und G , der sich mit obigem Verfahren berechnen lässt.

BEWEIS Eine Normaldarstellung von E sei gegeben durch $x \cdot n = c$, G werde durch $x = p + \alpha v$ beschrieben. Weil jeder Schnittpunkt beide Gleichungen simultan erfüllen muss, lassen sich alle Schnittpunkte durch Gleichsetzen errechnen. Sie werden, da auf G , durch ein $\hat{\alpha} \in \mathbb{R}$ eindeutig beschrieben. Man erhält

$$\begin{aligned} \langle p + \hat{\alpha}v, n \rangle &= c \\ \Leftrightarrow \hat{\alpha} \langle v, n \rangle &= c - \langle p, n \rangle. \end{aligned} \quad (2.17)$$

Weil E und G nicht parallel sind, gilt $\langle v, n \rangle \neq 0$, so dass man Gleichung (2.17) stets nach $\hat{\alpha}$ auflösen kann:

$$\hat{\alpha} = \frac{c - \langle p, n \rangle}{\langle v, n \rangle}.$$

Der zu diesem eindeutigen $\hat{\alpha}$ gehörende Punkt ist der gesuchte eindeutige Schnittpunkt. ■

Analog, aber mit etwas mehr Aufwand zeigt man die Existenz einer Schnittgeraden von zwei nicht parallelen Ebenen. Wir kommen nun zur Berechnung des Schnittpunkts zweier Geraden in \mathbb{R}^3 . Gegeben seien die Geraden G_1 und G_2 durch

$$G_1: \quad x = p_1 + \alpha v_1; \quad G_2: \quad y = p_2 + \alpha v_2.$$

Durch Gleichsetzen von G_1 und G_2 erhält man das lineare Gleichungssystem

$$p_1 + \alpha v_1 = p_2 + \beta v_2.$$

Man versucht, durch Lösen dieses linearen Gleichungssystems die Parameter des Schnittpunkts zu bestimmen. Drei Fälle können auftreten:

1. Die Geraden haben einen Schnittpunkt. Das Gleichungssystem hat dann genau eine Lösung.
2. Die Geraden sind parallel. Man erhält entweder mehrere Lösungen (bei identischen Geraden) oder keine Lösung.
3. Die Geraden haben weder einen Schnittpunkt, noch sind sie parallel. In diesem Fall spricht man von *windschiefen* Geraden. Das Gleichungssystem hat keine Lösung.

Beispiel 2.79: Gesucht ist der Schnittpunkt der Geraden

$$G_1 : x = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \alpha \cdot \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} \quad \text{und} \quad G_2 : y = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} + \beta \cdot \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix},$$

falls ein solcher existiert. Gleichsetzen von G_1 und G_2 ergibt

$$\begin{aligned} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \alpha \cdot \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} &= \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} + \beta \cdot \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} \\ \Leftrightarrow \alpha \cdot \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} + \beta \cdot \begin{pmatrix} -2 \\ 1 \\ -1 \end{pmatrix} &= \begin{pmatrix} -1 \\ 0 \\ -2 \end{pmatrix}. \end{aligned}$$

Wir lösen das lineare Gleichungssystem mit dem Gauß-Verfahren.

$$\begin{array}{ccc|c} -1 & -2 & -1 & \\ 1 & 1 & 0 & +(I) \\ 1 & -1 & -2 & +(I) \\ \hline -1 & -2 & -1 & \cdot(-1) \\ 0 & -1 & -1 & \cdot(-1) \\ 0 & -3 & -3 & -3 \cdot (II) \\ \hline 1 & 2 & 1 & -2 \cdot (II) \\ 0 & 1 & 1 & \\ 0 & 0 & 0 & \\ \hline 1 & 0 & -1 & \\ 0 & 1 & 1 & \end{array}$$

Es wird nur eine der beiden Variablen α oder β gebraucht. Mit $\beta = 1$ erhalten wir die Koordinaten s des Schnittpunktes durch

$$s = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} + \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}.$$

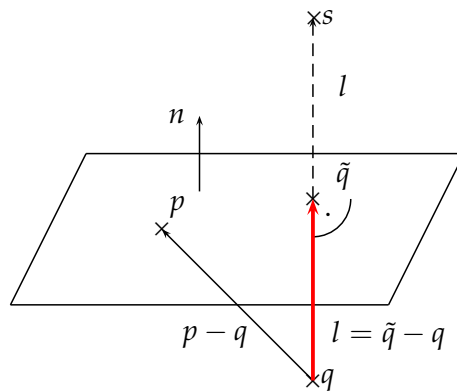
2.2.5 Abstandsbestimmung in \mathbb{R}^2 und \mathbb{R}^3

Wir betrachten zunächst eine Hyperebene E in \mathbb{R}^2 oder \mathbb{R}^3 mit einem Normalenvektor n und einen Punkt $q \notin E$ mit dem Ziel, einen Abstandsbegriff zwischen einem Punkt und einer Hyperebene zu definieren.

Definition 2.80:

1. Seien E und q wie oben. Der Schnittpunkt der Geraden $q + \alpha n$ mit E heißt *Lotfußpunkt*, $l := \tilde{q} - q$ heißt *Lot*, $s := q + 2l$ heißt *Spiegelpunkt* von q an E .
2. Der Abstand d von q zu E wird definiert als die Länge des Lotes, also

$$d := \|l\|.$$



Definition 2.80 ist gerechtfertigt, weil nach Satz 2.78 immer ein eindeutiger Lotfußpunkt \tilde{q} existiert. Warum man gerade die Länge des Lotes als Abstandsbegriff wählt, zeigt folgender Satz.

Satz 2.81: Seien E und $q \notin E$ wie oben. Dann gilt

$$\|q - \tilde{q}\| = \min_{p \in E} \|q - p\|,$$

und der Lotfußpunkt ist der einzige Punkt in E mit minimalem Abstand zu q .

BEWEIS Man hat $\tilde{q} - p \perp n$ und damit nach dem Satz des Pythagoras

$$\begin{aligned}\|p - q\|^2 &= \|(p - \tilde{q}) + (\tilde{q} - q)\|^2 \\ &= \|p - \tilde{q}\|^2 + \|\tilde{q} - q\|^2.\end{aligned}$$

Minimal wird $\|p - q\|$ genau dann, wenn $\|p - \tilde{q}\| = 0$, also nach N2 aus Satz 2.10 genau dann, wenn $p = \tilde{q}$. ■

Auf ähnliche Weise geht man bei der Bestimmung des Abstands zwischen einem Punkt und einer Geraden in \mathbb{R}^3 vor.

Definition 2.82: Sei G eine Gerade mit Richtungsvektor v und $q \notin G$. Ein Punkt $\tilde{q} \in G$ heißt *Lotfußpunkt*, wenn $l := q - \tilde{q} \perp G$ gilt, l heißt *Lot*, und der Abstand d zwischen einem Punkt und einer Geraden in \mathbb{R}^3 wird definiert durch $d := \|l\|$.

Wie im Fall der Hyperebene existiert immer ein eindeutiger Lotfußpunkt, der zudem stets der zu q nächste Punkt auf G ist. Wir betrachten kurz zwei nichtparallele Geraden in \mathbb{R}^3 .

Satz 2.83: Seien G_1 und G_2 zwei nicht parallele Geraden in \mathbb{R}^3 . Dann steht der kürzeste Verbindungsvektor zwischen G_1 und G_2 senkrecht sowohl auf G_1 als auch auf G_2 .

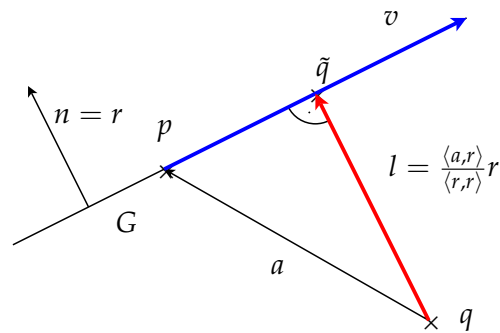
Abstandsberechnungen zwischen Punkten, Geraden und Hyperebenen lassen sich wie oben dargelegt auf die Berechnung des Lotfußpunkts zurückführen. Dies leistet folgender Rechenweg.

Rechenweg 2.84 (Abstandsberechnung):

1. Bestimme die Richtung r des Lots.
2. Bestimme jeweils einen Punkt auf den beiden Objekten und (durch Differenzbildung) den Abstandsvektor a zwischen den beiden Punkten.
3. Das Lot l ist die Projektion von a auf r . Der gesuchte Abstand ist $d = \|l\|$.

$$l = \frac{\langle a, r \rangle}{\langle r, r \rangle} r; \quad d = \frac{|\langle a, r \rangle|}{\langle r, r \rangle} \|r\| = \frac{|\langle a, r \rangle|}{\|r\|}$$

Folgende Grafik erläutert Rechenweg 2.84 anhand des Abstands eines Punktes zu einer Geraden in \mathbb{R}^2 .



Die einzelnen Fälle unterscheiden sich nur in der Wahl von r :

- Ist eine Hyperebene beteiligt, dann wählt man r als Normalenvektor n der Hyperebene.
- Bei zwei Geraden in \mathbb{R}^3 muss r senkrecht auf beiden Geraden stehen. Bei nicht parallelen Geraden wählt man r als Vektorprodukt der beiden Richtungsvektoren. Der Abstand zweier paralleler Geraden lässt sich auf den Abstand eines Punkts zu einer Geraden zurückführen.

Beispiel 2.85: Gegeben seien $q = (2, 0)^T$ und die Gerade G durch

$$x = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Gesucht ist der Abstand zwischen G und q . Aus dem Richtungsvektor $v = (2, 1)^T$ von G ermittelt man einen Normalenvektor $n = (1, -2)^T = r$. Ein möglicher Abstandsvektor ist

$$a = \begin{pmatrix} 2 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Der gesuchte Abstand beträgt also

$$d = \frac{|\langle a, r \rangle|}{\|r\|} = \frac{1}{\sqrt{5}}.$$

Beispiel 2.86: Gegeben seien $q = (1, 3, 5)^T \in \mathbb{R}^3$ und die Ebene E durch

$$-x + y - z = -5.$$

Gesucht wird der Abstand von q zu E . Man wählt $p \in E$ beliebig, z. B. $p = (5, 0, 0)^T$. Damit gilt $a = q - p = (-4, 3, 5)^T$. Das Lot zeigt in Richtung des Normalenvektors $r = n = (-1, 1, -1)^T$ der Ebene. Damit ist

$$d = \frac{|\langle a, r \rangle|}{\|r\|} = \frac{|4 + 3 - 5|}{\sqrt{3}} = \frac{2}{\sqrt{3}}.$$

Wenn der Abstand einer Ebene vom Nullpunkt gesucht ist, dann erhält man die einfachere Formel

$$d = \frac{|\langle p, n \rangle|}{\|n\|},$$

also die Länge der Projektion von p in Richtung des Normalenvektors n . Der Vergleich mit der Hesseschen Normalform

$$\frac{\langle x, n \rangle}{\|n\|} = \frac{\langle p, n \rangle}{\|n\|}$$

zeigt, dass der Betrag der rechten Seite der Hesseschen Normalform gerade der Abstand der Ebene zum Nullpunkt ist.

Wir kommen jetzt zur Abstandbestimmung eines Punktes zu einer Geraden in \mathbb{R}^3 und damit verbunden auch zur Abstandberechnung zweier paralleler Geraden in \mathbb{R}^3 .

Rechenweg 2.87 (Abstand Punkt-Gerade in \mathbb{R}^3):

1. $a = l + r$, also $l = a - r$.

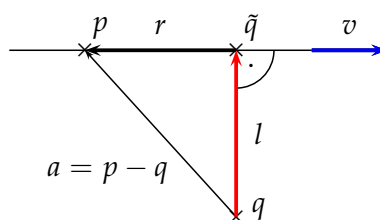
2. r ist die Projektion von a auf den Richtungsvektor der Geraden v , also

$$r = \frac{\langle a, v \rangle}{\langle v, v \rangle} v.$$

3. Zusammen ergibt sich:

$$l = a - \frac{\langle a, v \rangle}{\langle v, v \rangle} v$$

Die nachfolgende Grafik erläutert obigen Rechenweg.



Beispiel 2.88: Gegeben seien die Gerade G durch

$$x = \begin{pmatrix} 4 \\ 3 \\ 5 \end{pmatrix} + \alpha \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}$$

und $q = (1, -2, 3)^T$. Ein Richtungsvektor der Geraden lautet $v = (2, 2, 1)^T$, ein möglicher Abstandsvektor ist $a = (3, 5, 2)^T$. Das Lot l von q auf G errechnet sich zu

$$\begin{aligned} l &= a - \frac{\langle a, v \rangle}{\langle v, v \rangle} v \\ &= \begin{pmatrix} 3 \\ 5 \\ 2 \end{pmatrix} - \frac{6 + 10 + 2}{4 + 4 + 1} \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}. \end{aligned}$$

Daraus erhält man $d = \sqrt{2}$. Zudem gilt für den Lotfußpunkt

$$\tilde{q} = q + l = \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ 3 \end{pmatrix}.$$

2.3 Die Determinante im \mathbb{R}^2 und \mathbb{R}^3

2.3.1 Berechnung und geometrische Deutung

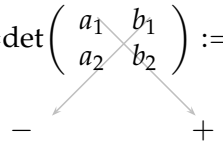
Wir stellen unsystematisch den Begriff der *Determinante* für die Spezialfälle von (1×1) , (2×2) - und (3×3) -Matrizen vor, weil es sich bei der Determinante auch um ein nützliches Rechenwerkzeug handelt, das die Berechnungen im Rahmen der analytischen Geometrie vereinfachen kann. Wir definieren zu einer quadratischen Matrix A die *Determinante* $\det(A)$ oder kürzer $|A|$ als reelle Zahl durch Angabe von Rechenvorschriften:

Definition 2.89: Sei $A \in \mathbb{R}^{n \times n}$.

n=1: $A = (a_1)$. Dann gilt

$$\det(A) := a_1$$

n=2: Sei $A = (a, b)$ mit den Spaltenvektoren $a, b \in \mathbb{R}^2$. Dann gilt

$$\det A = \det(a, b) = \det \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} := a_1 b_2 - b_1 a_2$$


Entlang der Diagonalen wird das Produkt gebildet und entsprechend dem angezeigten Vorzeichen aufsummiert.

n=3:

$$\begin{aligned} \det(a, b, c) &= \det \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} \\ &= a_1 b_2 c_3 + b_1 c_2 a_3 + c_1 a_2 b_3 - a_1 c_2 b_3 - b_1 a_2 c_3 - c_1 b_2 a_3 \end{aligned}$$

Die Formel für $n = 3$ kann durch die Berechnungsmethode nach *Sarrus*³ veranschaulicht werden:

$$\begin{array}{ccccccc}
 & a_1 & b_1 & c_1 & | & a_1 & b_1 \\
 & a_2 & b_2 & c_2 & | & a_2 & b_2 \\
 & a_3 & b_3 & c_3 & | & a_3 & b_3 \\
 \swarrow & & & & & \searrow & \\
 - & - & - & & & + & + & +
 \end{array}$$

Entlang der Diagonalen werden Produkte gebildet und diese mit den abgebildeten Vorzeichen versehen aufsummiert.

Wir deuten nun die Determinante zunächst für den Fall $n = 2$ geometrisch. Man hat

$$\left| \det \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} \right| = |a_1 b_2 - a_2 b_1| = \left\| \begin{pmatrix} 0 \\ 0 \\ a_1 b_2 - a_2 b_1 \end{pmatrix} \right\| = \left\| \begin{pmatrix} a_1 \\ a_2 \\ 0 \end{pmatrix} \times \begin{pmatrix} b_1 \\ b_2 \\ 0 \end{pmatrix} \right\|.$$

Nach Folgerung 2.43 gibt die euklidische Norm des Kreuzprodukts zweier Vektoren im \mathbb{R}^3 aber genau den Flächeninhalt des aufgespannten Parallelogramms an. Somit entspricht der Betrag der Determinante genau der Fläche des von a und b aufgespannten Parallelogramms.

Zur geometrischen Deutung der Determinante für $n = 3$ wird folgende Definition benötigt.

Definition 2.90: Für drei Vektoren $a, b, c \in \mathbb{R}^3$ nennt man

$$\langle a, b \times c \rangle \in \mathbb{R}$$

das *Spatprodukt* der drei Vektoren a, b, c .

Satz 2.91: Seien $a, b, c \in \mathbb{R}^3$, ϕ der Winkel zwischen a und b sowie ψ der Winkel zwischen den Vektoren $a \times b$ und c . Dann gilt:

$$\det(a, b, c) = \|a\| \cdot \|b\| \cdot \|c\| \sin \phi \cos \psi \quad (2.18)$$

BEWEIS Seien $a = (a_1, a_2, a_3)^T$, b und c entsprechend. Nach Definition 2.89 der Determinante gilt dann

$$\begin{aligned}
 \det(a, b, c) &= a_1 b_2 c_3 + b_1 c_2 a_3 + c_1 a_2 b_3 \\
 &\quad - a_3 b_2 c_1 - b_3 c_2 a_1 - c_3 a_2 b_1 \\
 &= c_1 (a_2 b_3 - a_3 b_2) + c_2 (a_3 b_1 - a_1 b_3) + c_3 (a_1 b_2 - a_2 b_1) \\
 &= \left\langle c, \begin{pmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{pmatrix} \right\rangle
 \end{aligned}$$

³Pierre Frédéric Sarrus (1798–1861), franz. Mathematiker

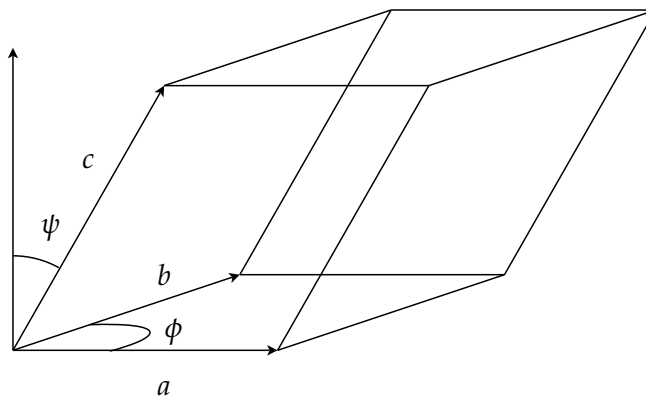


Abbildung 2.8: 3 Vektoren spannen einen Spat auf

Mit der Definition des Kreuzprodukts folgt daraus

$$\begin{aligned} \det(a, b, c) &= \langle c, a \times b \rangle \\ &= \|a \times b\| \cdot \|c\| \cos \psi \\ &= \|a\| \cdot \|b\| \cdot \|c\| \sin \phi \cos \psi \end{aligned}$$

Die Vektoren a, b, c bilden die Kanten eines Körpers im dreidimensionalen Raum, eines *Parallelepipeds* oder *Spats* (Abb. 2.8). Es entspricht also nach Formel (2.18) der Betrag der Determinante dem Volumen des durch die drei Spaltenvektoren aufgespannten Spats.

Bemerkung 2.92: Das Vorzeichen der Determinante zeigt die Orientiertheit der drei Spaltenvektoren: ist die Determinante positiv, handelt es sich um ein Rechtssystem, ist sie negativ, um ein Linkssystem. Daher spricht man in diesem Zusammenhang auch vom *orientierten Volumen* eines Spats.

Satz 2.93: Die Determinante hat folgende Eigenschaften.

D1 : $\det(a, b, c) = \det(c, a, b) = \det(b, c, a)$.

D2 : $\det(a, b, c) = -\det(b, a, c)$.

D3 : $\det(a, a, c) = 0$.

D4 : Für $\alpha \in \mathbb{R}$ gilt $\det(\alpha \cdot a, b, c) = \alpha \cdot \det(a, b, c)$.

D5 : $\det(a, b, c + d) = \det(a, b, c) + \det(a, b, d)$.

D6 : $\det(A) = \det(A^T)$

BEWEIS Im Fall $n = 2$: direktes Nachrechnen, im Fall $n = 3$ verwendet man Satz 2.91 zusammen mit den Eigenschaften von Kreuz- und Skalarprodukt. ■

Erläuterungen, Beispiele und Folgerungen:

- Beispiel zu D4:

$$\det \begin{pmatrix} 1 & 2 & 0 \\ 3 & 4 & -1 \\ 5 & 6 & 2 \end{pmatrix} = 2 \cdot \det \begin{pmatrix} 1 & 1 & 0 \\ 3 & 2 & -1 \\ 5 & 3 & 2 \end{pmatrix}$$

- Folgerung aus D4:

$$\det(0, b, c) = 0 \cdot \det(0, b, c) = 0$$

Ist in einer Matrix eine Spalte (oder eine Zeile, siehe D6) gleich 0, dann ist auch deren Determinante gleich 0.

- Beispiel zu D5:

$$\det \begin{pmatrix} 1 & 2 & 0 \\ 3 & 4 & -1 \\ 5 & 6 & 2 \end{pmatrix} = \det \begin{pmatrix} 1 & 1 & 0 \\ 3 & 1 & -1 \\ 5 & 1 & 2 \end{pmatrix} + \det \begin{pmatrix} 1 & 1 & 0 \\ 3 & 3 & -1 \\ 5 & 5 & 2 \end{pmatrix}$$

Bemerkung 2.94: Es gilt:

$$\begin{aligned} \det(a, b, c + \alpha a) &\stackrel{D5}{=} \det(a, b, c) + \det(a, b, \alpha a) \\ &\stackrel{D4}{=} \det(a, b, c) + \alpha \cdot \det(a, b, a) \\ &\stackrel{D1}{=} \det(a, b, c) + \alpha \cdot \det(a, a, b) \\ &\stackrel{D3}{=} \det(a, b, c) \end{aligned}$$

2.3.2 Lineare 3×3 -Gleichungssysteme

Mit Hilfe der Determinante lassen sich lineare Gleichungssystemen auf eindeutige Lösbarkeit untersuchen. Durch

$$(LG) \begin{cases} (E_1) & a_1 \cdot x_1 + a_2 \cdot x_2 + a_3 \cdot x_3 = d_1 \\ (E_2) & b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 = d_2 \\ (E_3) & c_1 \cdot x_1 + c_2 \cdot x_2 + c_3 \cdot x_3 = d_3 \end{cases}$$

ist ein *lineares Gleichungssystem* mit den Unbekannten (x_1, x_2, x_3) gegeben. Jede Gleichung beschreibt eine Ebene in \mathbb{R}^3 . Eindeutige Lösbarkeit bedeutet

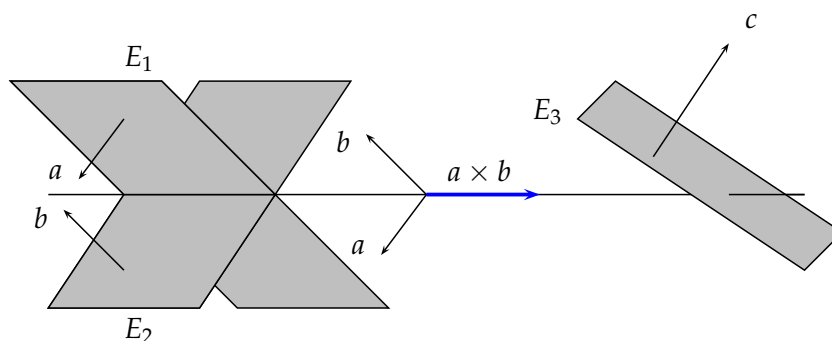
also die Existenz eines eindeutigen Schnittpunkts der drei Ebenen. Der Vektor

$$a = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$$

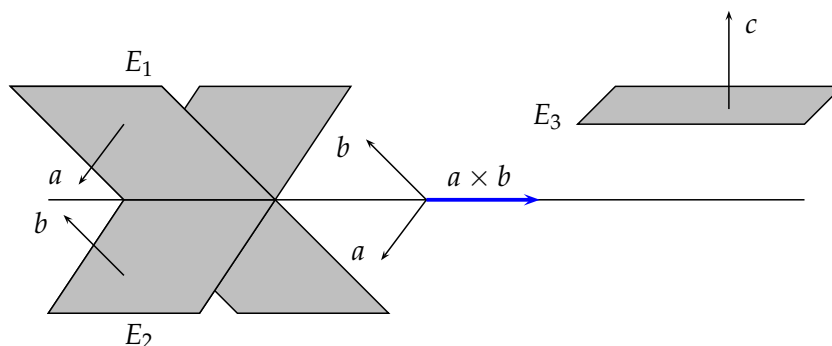
ist der Normalenvektor der Ebene E_1 . Entsprechend sind b und c die Normalenvektoren der Ebenen E_2 und E_3 . Im Fall eines eindeutigen Schnittpunkts dürfen E_1 und E_2 nicht parallel liegen, es muss somit $a \times b \neq 0$ gelten. Der Vektor $a \times b$ zeigt in Richtung der Schnittgeraden $E_1 \cap E_2$. Die Schnittgerade darf nicht parallel zu E_3 , d. h. nicht senkrecht zum Normalenvektor c , sein. Es muss also gelten

$$\langle (a \times b), c \rangle = \det(a, b, c) \neq 0.$$

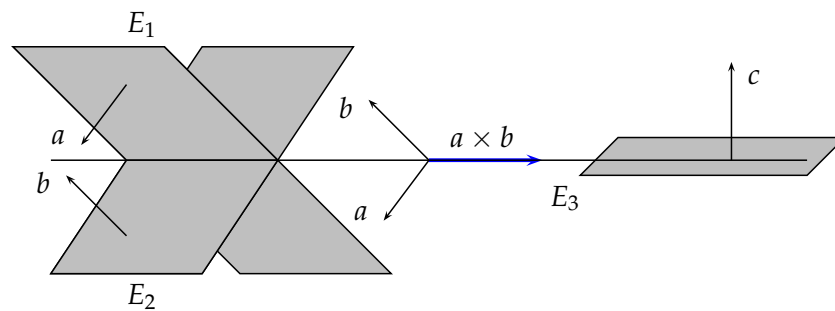
Damit ist die gewünschte Bedingung für die eindeutige Lösbarkeit des obigen Gleichungssystems gefunden.



Eindeutiger Schnittpunkt: $\langle a \times b, c \rangle \neq 0$



Kein Schnittpunkt: $\langle a \times b, c \rangle = 0$



Schnittgerade, kein eindeutiger Schnittpunkt: $\langle a \times b, c \rangle = 0$

Satz 2.95: Notwendig und hinreichend dafür, dass das lineare Gleichungssystem LG eine eindeutige Lösung besitzt, ist die Bedingung

$$\det(a, b, c) \neq 0.$$

Kapitel 3

Algebraische Strukturen

3.1 Gruppen

3.1.1 Grundlagen

Die Mathematik ist traditionell eine Strukturwissenschaft, d. h. sie ersinnt abstrakte Objekte und untersucht dann ihre Eigenschaften. Dieses Vorgehen prägt insbesondere die Algebra, die als Teil der reinen Mathematik algebraische Strukturen betrachtet. Die Resultate der Algebra haben aber, obwohl ursprünglich aus intellektuelle Neugier heraus erzielt, Eingang in die gesamte Mathematik gefunden und auch die angewandte Mathematik entscheidend beeinflusst. Daher gehen wir hier kurz auf einige wesentliche algebraische Strukturen ein. Eine der einfachsten algebraischen Strukturen ist die aus der Vorlesung Mathematische Grundlagen bekannte *Gruppe*.

Definition 3.1:

1. Sei M eine Menge, und $\circ : M \times M \rightarrow M$ eine Abbildung, $(x, y) \rightarrow \circ(x, y)$. Eine solche Abbildung heie *Verknpfung*; wir schreiben zur Abkrzung statt $\circ(x, y)$ einfacher $x \circ y$.
2. Fr M und \circ wie oben heit das Paar (M, \circ) eine *Gruppe*, wenn gilt:

G1 (Assoziativitt): Die Verknpfung ist assoziativ, d. h. es gilt:

$$\forall x, y, z \in M : (x \circ y) \circ z = x \circ (y \circ z)$$

G2 (Neutralelement): Es existiert ein neutrales Element $e \in M$ so dass

$$\forall x \in M : x \circ e = x$$

G3 (Inverses Element): $\forall x \in M \exists x'$ mit

$$x \circ x' = e.$$

3. Gilt für eine Gruppe $G = (M, \circ)$, dass $x \circ y = y \circ x \forall x, y \in M$, dann heißt G *abelsche Gruppe* oder *kommutative Gruppe*.

Bemerkung 3.2: Zum Nachweis, dass (M, \circ) eine Gruppe ist, gehört implizit auch der Nachweis, dass \circ tatsächlich eine Verknüpfung im Sinne der Definition 3.1 ist. Man muss also zeigen, dass

1. $x \circ y$ für alle x, y existiert und eindeutig festgelegt ist.
2. das Ergebnis $x \circ y$ in M liegt.

Ist dies der Fall, sagt man, die Abbildung $\circ : M \times M \rightarrow M$ sei *wohldefiniert*, also sinnvoll definiert.

Beispiel 3.3:

1. $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$ und $(\mathbb{R}, +)$ sind Gruppen.
2. Die Menge $M = \{-5, -4, \dots, 0, \dots, 4, 5\}$ bildet zusammen mit der gewöhnlichen Addition *keine* Gruppe. Natürlich erfüllt $(M, +)$ die Bedingungen **G1** – **G3** aus Definition 3.1, aber aufgrund z. B. $4 + 4 = 8 \notin M$ ist „+“ *keine* Verknüpfung auf M , und daher kann $(M, +)$ keine Gruppe bilden.

Satz 3.4: Sei $G = (M, \circ)$ eine Gruppe. Dann gilt

1. $M \neq \emptyset$
2. Das inverse Element kommutiert, d. h. $\forall x \in M : x' \circ x = x \circ x'$.
3. Das neutrale Element kommutiert, also $\forall x \in M : x \circ e = e \circ x$.
4. Das inverse Element ist eindeutig, d. h. $\forall x$ existiert genau ein $x' \in M : x' \circ x = e$.
5. Das neutrale Element e ist eindeutig.

BEWEIS

1. Nach **G2** gilt $e \in M$.

2. Es gilt: $x \circ x' = e$, d. h. x' ist ein inverses Element von x . Auch x' hat nach G3 ein inverses Element z : $x' \circ z = e$. Dann ist

$$\begin{aligned}
 e &= x' \circ z \\
 &= (x' \circ e) \circ z \\
 &= x' \circ (e \circ z) \\
 &= x' \circ ((x \circ x') \circ z) \\
 &= x' \circ (x \circ (x' \circ z)) \\
 &= x' \circ (x \circ e) \\
 &= x' \circ x
 \end{aligned}$$

3. Nach obiger Rechnung gilt

$$\begin{aligned}
 e \circ x &= (x \circ x') \circ x \\
 &= x \circ (x' \circ x) \\
 &= x \circ e \\
 &= x
 \end{aligned}$$

4. Seien y und y' zwei inverse Elemente zu x . Dann folgt:

$$y' = y' \circ e = y' \circ (x \circ y) = (y' \circ x) \circ y = e \circ y = y$$

5. Für ein weiteres neutrales Element e' folgt $e' = e' \circ e = e$ nach G2. ■

Bemerkung 3.5:

1. Da das inverse Element zu a eindeutig ist, spricht man von *dem* inversen Element und bezeichnet es mit a^{-1} .
2. Für eine Gruppe $G = (M, \circ)$ schreibt man oft kurz $a \in G$, wenn man eigentlich $a \in M$ meint. Ist die Verknüpfung aus dem Zusammenhang klar, identifiziert man häufig M und G .

Beispiel 3.6:

1. Die *triviale Gruppe* $(\{1\}, \cdot)$ ist die kleinste aller Gruppen.
2. $(\mathbb{N}, +)$ ist keine Gruppe, weil z. B. zur 1 das additiv Inverse -1 keine natürliche Zahl ist. Gäbe es ein weiteres Inverses in \mathbb{N} , wäre dies auch Inverses von 1 in \mathbb{Z} . Dies widerspräche der Eindeutigkeit des inversen Elements in \mathbb{Z} .
3. Der Zahlenraum \mathbb{R}^n bildet mit der Vektoraddition eine Gruppe (Übungsaufgabe!).

4. (\mathbb{R}, \cdot) mit der gewöhnlichen Multiplikation ist keine Gruppe, weil es zu 0 keine reelle Zahl a gibt mit $0a = 1$, also **G3** nicht erfüllt ist.
5. Vektoren im \mathbb{R}^3 bilden bzgl. des Vektorproduktes keine Gruppe, weil das Vektorprodukt nach Bemerkung 2.44 nicht assoziativ ist.

Bemerkung 3.7: $(\mathbb{R} \setminus \{0\}, \cdot)$ ist eine abelsche Gruppe.

BEWEIS Wir zeigen **G1** – **G3**. **G1** ist klar, ebenso **G2** mit $e = 1$. Mit $a^{-1} = 1/a$ gilt **G3**. Man beachte, dass $a \neq 0$ wegen $a \in \mathbb{R} \setminus \{0\}$ n. V. ■

3.1.2 Vertiefung: Endliche Gruppen und Restklassen

Wir wiederholen sehr kurz einige Inhalte der Vorlesung „Mathematische Grundlagen“, um sie im Kontext von Gruppen erneut zu betrachten. Für $m \in \mathbb{N}$ definiert man $\bar{k} = \{k + m\mathbb{Z}\}$, die sog. *Restklassen* als Äquivalenzklassen der Äquivalenzrelation $a \sim b :\Leftrightarrow (a - b) \bmod m = 0$ und erhält eine disjunkte Zerlegung von \mathbb{Z} : $\mathbb{Z} = \bigcup_{k=0}^{m-1} \bar{k}$. Auf der Menge $\mathbb{Z}/m\mathbb{Z} := \{\bar{0}, \dots, \overline{m-1}\}$ wird dann die Verknüpfung \oplus definiert durch $\bar{k} \oplus \bar{l} := \overline{k+l}$. Dann gilt:

Satz 3.8: Für $m \in \mathbb{N}$ ist $(\mathbb{Z}/m\mathbb{Z}, \oplus)$ eine abelsche Gruppe.

BEWEIS Dass \oplus wohldefiniert ist, ist aus der Vorlesung Mathematische Grundlagen bekannt. Wir zeigen die drei Bedingungen **G1** – **G3**.

G1 : Für $\bar{a}, \bar{b}, \bar{c} \in \mathbb{Z}/m\mathbb{Z}$ gilt

$$\begin{aligned} (\bar{a} \oplus \bar{b}) \oplus \bar{c} &= \overline{a+b} \oplus \bar{c} = \overline{(a+b)+c} = \overline{a+(b+c)} \\ &= \bar{a} \oplus \overline{b+c} = \bar{a} \oplus (\bar{b} \oplus \bar{c}). \end{aligned}$$

G2 : Offenbar gilt $\bar{a} \oplus \bar{0} = \overline{a+0} = \bar{a} \forall \bar{a} \in \mathbb{Z}/m\mathbb{Z}$, also ist $\bar{0}$ das neutrale Element.

G3 : Für \bar{a} gilt $\bar{a} \oplus \overline{m-a} = \overline{a+(m-a)} = \bar{m} = \bar{0}$, also ist $\overline{m-a}$ das gesuchte inverse Element.

Weiterhin gilt $\bar{a} \oplus \bar{b} = \overline{a+b} = \overline{b+a} = \bar{b} \oplus \bar{a}$, also ist die Gruppe abelsch. ■

Endliche Gruppen werden häufig über ihre *Verknüpfungstafel* definiert, in der man das Ergebnis jeder möglichen Verknüpfung explizit angibt (vgl. Abb. 3.1 für die Gruppen $\mathbb{Z}/2\mathbb{Z}$ und $\mathbb{Z}/4\mathbb{Z}$). Für abelsche Gruppen ist die Verknüpfungstafel stets spiegelsymmetrisch zur Hauptdiagonalen.

Man sieht leicht, dass die Menge $m\mathbb{Z} = \{\dots, -2m, -m, 0, m, 2m, \dots\}$ mit der gewöhnlichen Addition eine Gruppe bildet. Weil die Obermenge \mathbb{Z} mit derselben Addition auch eine Gruppe bildet, liegt hier eine Gruppe innerhalb einer Gruppe vor. Diese Beobachtung motiviert die folgende Definition.

\oplus	$\bar{0}$	$\bar{1}$
$\bar{0}$	$\bar{0}$	$\bar{1}$
$\bar{1}$	$\bar{1}$	$\bar{0}$

\oplus	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$
$\bar{0}$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$
$\bar{1}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{0}$
$\bar{2}$	$\bar{2}$	$\bar{3}$	$\bar{0}$	$\bar{1}$
$\bar{3}$	$\bar{3}$	$\bar{0}$	$\bar{1}$	$\bar{2}$

Abbildung 3.1: Verküpfungstabeln von $(\mathbb{Z}/2\mathbb{Z}, \oplus)$ (links) und $(\mathbb{Z}/4\mathbb{Z}, \oplus)$ (rechts)

\oplus	$\bar{0}$	$\bar{2}$
$\bar{0}$	$\bar{0}$	$\bar{2}$
$\bar{2}$	$\bar{2}$	$\bar{0}$

Abbildung 3.2: Verküpfungstafel von (\mathbb{Z}_2, \oplus)

Definition 3.9: Sei $G = (M, \circ)$ eine Gruppe und $M' \subseteq M$. Bildet $U = (M', \circ)$ eine Gruppe, so heißt U *Untergruppe* von G , Schreibweise $U \leq G$.

Folgerung 3.10: Für G und U wie oben gilt $U \leq G$ genau dann, wenn gilt:

1. $M' \neq \emptyset$
2. $\forall a, b \in M' : a \circ b \in M'$
3. $\forall a \in M' : a^{-1} \in M'$

BEWEIS „ \Rightarrow “: Ist U Gruppe, so gilt $e \in U$, also $U \neq \emptyset$. 2. folgt aus der Anforderung, dass \circ die Verknüpfung von U ist und deswegen von $M' \times M'$ nach M' abbildet. Nach **G3** gilt Eigenschaft 3.

„ \Leftarrow “: Aus 2. folgt analog zu oben die Wohldefiniiertheit der Verknüpfung; **G1** gilt, weil es bereits in der Obermenge M gilt. Wegen 1. ex. $a \in M'$; wegen 3. gilt dann $a^{-1} \in M'$ und damit nach 2. $e = a \circ a^{-1} \in M'$, also **G2**. **G3** folgt aus 3. ■

Beispiel 3.11: Offensichtlich gilt $(m\mathbb{Z}, +) \leq (\mathbb{Z}, +) \leq (\mathbb{Q}, +) \leq (\mathbb{R}, +)$.

Auf der Suche nach Beispielen für Untergruppen betrachten wir erneut die Verknüpfungstafel von $\mathbb{Z}/4\mathbb{Z}$ (Abb. 3.1). Es fällt auf, dass $\mathbb{Z}_2 := (\{\bar{0}, \bar{2}\}, \oplus)$ eine Untergruppe bildet. Ihre Verknüpfungstafel (Abb. 3.2) gleicht bis auf Bezeichnung der Verknüpfungstafel von $\mathbb{Z}/2\mathbb{Z}$ (Abb. 3.1 links). Die beiden Gruppen sind nicht gleich, weil es sich um verschiedene Restklassen handelt, aber besitzen doch die gleiche algebraische Struktur. Dieses Phänomen führt zu folgender Definition.

Definition 3.12: Zwei Gruppen $G_1 = (M_1, \circ)$ und $G_2 = (M_2, +)$ heißen *isomorph*, Schreibweise $G_1 \simeq G_2$, wenn es eine bijektive Abbildung $\phi : G_1 \rightarrow G_2$ gibt mit

$$\phi(a \circ b) = \phi(a) + \phi(b) \forall a, b \in G_1.$$

Bemerkung 3.13: Der Begriff „isomorph“ kommt aus dem Griechischen und bedeutet „gleichgestaltig“, ϕ wird Gruppenisomorphismus genannt. Die Abbildung „übersetzt“ die Elemente von G_1 in die von G_2 und respektiert dabei die Gruppenstruktur.

Beispiel 3.14:

1. Isomorphie ist eine Äquivalenzrelation.
2. Man hat $\mathbb{Z}/2\mathbb{Z} \simeq \mathbb{Z}_2$ durch den elementweise definierten Gruppenisomorphismus $\phi(\bar{0}_2) = \bar{0}_4, \phi(\bar{1}_2) = \bar{2}_4$. Die Indizes an den Restklassen bezeichnen jeweils den Modul m ; man beachte $\bar{0}_2 \neq \bar{0}_4$. Es gilt also $\mathbb{Z}/2\mathbb{Z} \simeq \mathbb{Z}_2 \leq \mathbb{Z}/4\mathbb{Z}$.
3. Sowohl $(\mathbb{R}^4, +)$ als auch $(\mathbb{R}^2 \times \mathbb{R}^2, +)$ sind offenbar Gruppen, aber dennoch formal nicht gleich. Durch den Gruppenisomorphismus $\varphi : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^4, \varphi((x_1, x_2), (x_3, x_4)) := (x_1, x_2, x_3, x_4)$ sind aber die beiden Gruppen isomorph. Daher werden wir künftig nicht zwischen \mathbb{R}^4 und $\mathbb{R}^2 \times \mathbb{R}^2$ o. ä. unterscheiden.
4. Es gilt $\mathbb{C} \neq \mathbb{R}^2$, wohl aber (als Gruppen) $(\mathbb{C}, +) \simeq (\mathbb{R}^2, +)$ durch den Isomorphismus $\varphi : \mathbb{C} \rightarrow \mathbb{R}^2, \varphi(a + ib) = (a, b)^T$. Dies rechtfertigt die Veranschaulichung der komplexen Zahlen als komplexe Ebene.

Bemerkung 3.15 (Produktgruppen): Seien $G_1 := (A, +)$ und $G_2 := (B, \circ)$ Gruppen. Dann lässt sich auf dem kartesischen Produkt $A \times B$ komponentenweise eine Verknüpfung \star definieren durch

$$\begin{pmatrix} a_1 \\ b_1 \end{pmatrix} \star \begin{pmatrix} a_2 \\ b_2 \end{pmatrix} := \begin{pmatrix} a_1 + a_2 \\ b_1 \circ b_2 \end{pmatrix}, \quad a_1, a_2 \in A, b_1, b_2 \in B.$$

Durch direktes Nachrechnen lässt sich zeigen, dass $(A \times B, \star)$ wiederum eine Gruppe bildet. Durch dieses Vorgehen lässt sich z. B. $(\mathbb{R}^2, +)$ aus $(\mathbb{R}, +)$ gewinnen. Enthält A m und B n Elemente, so enthält $A \times B$ mn Elemente. Damit enthält z. B. $(\mathbb{Z}/2\mathbb{Z})^2$ genau 4 Elemente, die so erhaltene Gruppe, die sog. *Kleinsche Vierergruppe*, ist jedoch nicht isomorph zu $\mathbb{Z}/4\mathbb{Z}$.

3.2 Körper

Wir wollen nun ausgehend von den reellen Zahlen eine allgemeinere algebraische Struktur definieren und untersuchen. Auf den reellen Zahlen ist neben der Addition auch die Multiplikation erklärt. Das Tripel $(\mathbb{R}, +, \cdot)$ hat folgende Eigenschaften:

1. $(\mathbb{R}, +)$ bildet eine abelsche Gruppe.
2. $(\mathbb{R} \setminus \{0\}, \cdot)$ bildet eine abelsche Gruppe (vgl. Bemerkung 3.7).
3. Es gelten die *Distributivgesetze*

$$\begin{aligned}\forall x, y, z \in \mathbb{R} : (x + y) \cdot z &= (x \cdot z) + (y \cdot z) \\ x \cdot (y + z) &= (x \cdot y) + (x \cdot z)\end{aligned}$$

Jedes Tripel (M, \oplus, \odot) , das die obigen drei Bedingungen erfüllt, wird *Körper* genannt:

Definition 3.16: Sei M eine nichtleere Menge. M zusammen mit zwei Verknüpfungen $\oplus, \odot : M \times M \rightarrow M$ heißt *Körper*, wenn gilt :

1. (M, \oplus) bildet eine kommutative Gruppe.
2. Nach Ausschluss des neutralen Elements der Verknüpfung \oplus bilden die restlichen Elemente von M eine kommutative Gruppe bzgl. \odot .
3. Für \oplus und \odot gelten die *Distributivgesetze*:
 - a) $\forall x, y, z \in M : (x \oplus y) \odot z = (x \odot z) \oplus (y \odot z)$
 - b) $\forall x, y, z \in M : z \odot (x \oplus y) = (z \odot x) \oplus (z \odot y)$

Das neutrale Element bzgl. der Addition \oplus wird häufig mit „0“ bezeichnet (Null-Element), das neutrale Element bzgl. \odot häufig mit „1“ (Eins-Element). Im Folgenden verwenden wir für die beiden neutralen Elemente die Notationen $\mathbb{0}$ und $\mathbb{1}$, um eine Verwechslung mit den Zahlen 0 und 1 zu vermeiden.

Dass $\mathbb{0}$ - das neutrale Element der Addition - in Punkt 2. der Definition 3.16 ausgeschlossen wird, ist notwendig aufgrund folgenden Satzes.

Satz 3.17: Sei K ein Körper und $a \in K$. Dann gilt $a \odot \mathbb{0} = \mathbb{0}$.

BEWEIS Es ist $a \odot (\mathbb{0} + \mathbb{0}) = a \odot \mathbb{0} + a \odot \mathbb{0}$ und $a \odot (\mathbb{0} + \mathbb{0}) = a \odot \mathbb{0}$, also $a \odot \mathbb{0} + a \odot \mathbb{0} = a \odot \mathbb{0}$ und daher $a \odot \mathbb{0} = \mathbb{0}$. ■

Existierte in einem Körper K $\mathbb{0}^{-1} = x$, folgt $x \odot \mathbb{0} = \mathbb{1}$. Es gilt aber $x \odot \mathbb{0} = \mathbb{0}$, also $\mathbb{0} = \mathbb{1}$. Für $a \in K$ folgt dann $a = \mathbb{1} \odot a = \mathbb{0} \odot a = \mathbb{0}$, also bestünde K aus einem einzigen Element $\mathbb{0}$ und den Verknüpfungen

$$\begin{aligned}\mathbb{0} \oplus \mathbb{0} &= \mathbb{0} \\ \mathbb{0} \odot \mathbb{0} &= \mathbb{0}.\end{aligned}$$

Jedoch impliziert Eigenschaft 2 eines Körpers, dass die neutralen Elemente bzgl. der Addition und der Multiplikation verschieden sind. Jeder Körper besitzt daher mindestens 2 Elemente. Somit ist $\mathbb{0} = \mathbb{1}$ ausgeschlossen, und $\mathbb{0}$ besitzt kein multiplikativ Inverses.

Beispiel 3.18:

1. Nach Konstruktion ist $(\mathbb{R}, +, \cdot)$ ein Körper.
2. $(\mathbb{Q}, +, \cdot)$ ist ein Körper, ebenso $(\mathbb{C}, +, \cdot)$ mit der komplexen Multiplikation.
3. $(\mathbb{Z}, +, \cdot)$ ist kein Körper, da $(\mathbb{Z} \setminus \{0\}, \cdot)$ keine Gruppe ist, weil z. B. $2^{-1} = \frac{1}{2} \notin \mathbb{Z}$, also **G3** nicht erfüllt ist.

Es liegt die Frage nahe, ob es analog zu endlichen Gruppen auch endliche Körper gibt. Der folgende Satz liefert eine partielle Antwort.

Satz 3.19: Wir definieren auf $\mathbb{Z}/m\mathbb{Z}$ die Verknüpfung \odot durch

$$\bar{a} \odot \bar{b} := \overline{a \cdot b} \quad \forall \bar{a}, \bar{b} \in \mathbb{Z}/m\mathbb{Z}.$$

Ist m eine Primzahl, dann ist $(\mathbb{Z}/m\mathbb{Z}, \oplus, \odot)$ ein Körper.

Für den Beweis verweisen wir auf Beutelspacher [2]. Endliche Körper haben für die Lineare Algebra nur geringe Bedeutung, daher werden wir sie nicht weiter betrachten.

3.3 Vektorräume

In Kapitel 1.3 haben wir den Zahlraum \mathbb{R}^n definiert und als Menge aller Vektoren charakterisiert. Wir werden nun den Zahlraum zum Vektorraum verallgemeinern, indem wir in Analogie zur Gruppe und zum Körper die wesentlichen algebraischen Eigenschaften des \mathbb{R}^n sammeln und *jede* Struktur, die diese Eigenschaften hat, *Vektorraum* nennen.

Definition 3.20: Sei K ein beliebiger Körper. Eine nicht-leere Menge V zusammen mit den beiden Abbildungen

$$\begin{aligned} \oplus : V \times V &\rightarrow V, & (x, y) &\mapsto x \oplus y \in V \\ \text{und } \odot : K \times V &\rightarrow V, & (\lambda, x) &\mapsto \lambda \odot x \in V \end{aligned}$$

heißt *Vektorraum über K* oder *K -Vektorraum*, wenn folgende Axiome gelten:

1. (V, \oplus) ist eine kommutative Gruppe.
2. $\forall \lambda, \mu \in K$ und $x \in V$ gilt $\lambda \odot (\mu \odot x) = (\lambda\mu) \odot x$, wobei mit $\lambda\mu$ die Multiplikation aus K gemeint ist.
3. $\forall x \in V$ gilt $\mathbb{1} \odot x = x$ ($\mathbb{1}$ ist das neutrale Element der Multiplikation aus K).
4. $\forall \lambda \in K, x, y \in V$ gilt: $\lambda \odot (x \oplus y) = (\lambda \odot x) \oplus (\lambda \odot y)$.

$$5. \forall \lambda, \mu \in K, x \in V \text{ gilt } (\lambda + \mu) \odot x = (\lambda \odot x) \oplus (\mu \odot x).$$

Üblicherweise geht man von \mathbb{R} oder \mathbb{C} als Grundkörper aus. Im Fall $K = \mathbb{R}$ heißt V *reeller Vektorraum*, für $K = \mathbb{C}$ heißt V *komplexer Vektorraum*. Wie bei Gruppen auch identifizieren wir (V, \oplus, \odot) mit V , wenn die Verknüpfungen und der Körper aus dem Zusammenhang heraus klar sind.

Definition 3.21: Die Elemente der Menge V eines Vektorraums heißen *Vektoren*, die Elemente aus K *Skalare*.

Bemerkung 3.22: Im Vektorraum muss die Multiplikation eines Skalars mit einem Vektor definiert sein, nicht aber die Multiplikation zweier Vektoren.

Beispiel 3.23:

1. \mathbb{R}^n ist nach Konstruktion ein reeller Vektorraum mit der Vektoraddition gemäß Definition 1.6 und der Multiplikation eines Vektors mit einem Skalar nach Definition 1.7.
2. $(\mathbb{C}^n, +, \cdot)$ ist ein komplexer Vektorraum mit der Addition von Vektoren und der Multiplikation mit einem (komplexen) Skalar wie oben.
3. Jeder Körper ist ein Vektorraum über sich selbst mit den Verknüpfungen des Körpers \oplus als Vektoraddition und \odot als Multiplikation eines Vektors mit einem Skalar.

Beispiel 3.24: Sei $K = \mathbb{R}$ und $V = \mathcal{C}[a, b]$ die Menge der auf dem Intervall $[a, b]$ definierten reellen stetigen Funktionen. Wie in der Analysis üblich werden Summe und skalares Vielfaches von $f, g \in \mathcal{C}[a, b]$ punktweise definiert:

$$\begin{aligned} (f + g)(x) &:= f(x) + g(x) \\ (\alpha f)(x) &:= \alpha f(x) \end{aligned}$$

Dann ist V ein reeller Vektorraum. Zum Nachweis nutzt man die aus der Analysis bekannte Tatsache, dass Summen und Vielfache stetiger Funktionen wieder stetig sind und führt alle anderen Axiome auf die entsprechenden Aussagen für reelle Zahlen zurück.

Beispiel 3.25: Auf der Menge der $m \times n$ -Matrizen mit Elementen im Körper K lässt sich analog zur Addition von Vektoren eine komponentenweise Addition definieren: Für $A = (a_{ij})$ und $B = (b_{ij})$ sei $A + B := (a_{ij} + b_{ij})$. Analog definiert man die Multiplikation eines Skalars mit einer Matrix A : $\lambda A := (\lambda a_{ij})$. Mit diesen Verknüpfungen wird die Menge der $m \times n$ -Matrizen zu einem K -Vektorraum und wird mit $K^{m \times n}$ bezeichnet. Der Vektorraum unterscheidet sich nur durch die Schreibweise seiner Elemente im Rechteck vom Vektorraum $K^{m \cdot n}$. Man rechnet die Vektorraumaxiome leicht nach.

Beispiel 3.26: Sei \mathcal{M} die Menge aller komplexwertigen Folgen (x_1, x_2, \dots) und für zwei Folgen $(x_n)_{n \in \mathbb{N}}, (y_n)_{n \in \mathbb{N}}$ eine Addition erklärt durch $(x_n)_{n \in \mathbb{N}} \oplus (y_n)_{n \in \mathbb{N}} := (x_n + y_n)_{n \in \mathbb{N}}$ sowie eine Multiplikation mit einem komplexen Skalar λ durch $\lambda \odot (x_n)_{n \in \mathbb{N}} := (\lambda x_n)_{n \in \mathbb{N}}$. Dann ist $(\mathcal{M}, \oplus, \odot)$ ein komplexer Vektorraum.

Analog zur Gruppe innerhalb einer Gruppe, also einer Untergruppe, kann es Vektorräume innerhalb von Vektorräumen geben. Man spricht von *Untervektorräumen*. Im restlichen Kapitel bezeichne V einen K -Vektorraum.

Definition 3.27: Eine Teilmenge $U \subseteq V$ heißt *Untervektorraum* oder *Unterraum* von V , wenn $U \neq \emptyset$ (leere Menge) und $\forall x, y \in U$ und alle $\lambda \in K$ gilt:

$$x \oplus y \in U, \lambda \odot x \in U.$$

Bemerkung 3.28: $U \subseteq V$ ist genau dann ein Vektorraum, wenn U ein Untervektorraum von V im Sinne von Definition 3.27 ist.

BEWEIS Sei U ein Vektorraum. Weil $(U, +)$ eine abelsche Gruppe bildet, gilt $U \neq \emptyset$. Die Abgeschlossenheit von U in V ist äquivalent zur Wohldefiniertheit der Vektoraddition und der Multiplikation mit einem Skalar im Vektorraum U .

Umgekehrt leiten sich aus $U \subset V$ alle Vektorraumaxiome für U mit Ausnahme der Abgeschlossenheit ab. ■

Beispiel 3.29: Sei $V = \mathbb{R}^2$ und $U = \{(x, 0); x \in \mathbb{R}\} \subset V$. Wir zeigen, dass U ein Unterraum von V ist und prüfen dazu die Kriterien der Definition 3.27. Sei $u_1 = (x_1, 0)$ und $u_2 = (x_2, 0)$. Dann gilt:

$$\begin{aligned} u_1 + u_2 &= (x_1 + x_2, 0) \in U \\ \lambda u_1 &= (\lambda x_1, 0) \in U, \end{aligned}$$

also ist U ein Unterraum von V und zugleich ist U nach Bemerkung 3.28 ein Vektorraum.

Bemerkung 3.30: Ist U ein Unterraum von V , dann liegt der Nullvektor 0 von V und zu jedem $x \in U$ auch $-x$ in U , weil $0 \cdot x = 0$ und $(-1) \cdot x = -x$ gilt gemäß der Definition eines Vektorraums.

Bemerkung 3.31: Es sind $\{0\}$ und V Unterräume von V .

Bemerkung 3.32: Sind U_1 und U_2 Unterräume von V , dann ist $U_1 \cap U_2$ ein Unterraum von V .

BEWEIS Wir prüfen die Kriterien der Definition. Seien $u, v \in U_1 \cap U_2$. Dann gilt nach Definition $u + v \in U_1$ und $u + v \in U_2$, da U_1 und U_2 Untervektorräume sind, also $u + v \in U_1 \cap U_2$. Weiter gilt für $u \in U_1 \cap U_2$ und $\lambda \in K$, dass $\lambda u \in U_1$ und $\lambda u \in U_2$ und damit $\lambda u \in U_1 \cap U_2$. ■

Im Gegensatz zu Schnitten sind Vereinigungen von Untervektorräumen i. A. keine Untervektorräume.

Beispiel 3.33: Es sind

$$U_1 = \{(x, 0); x \in \mathbb{R}\}$$

$$U_2 = \{(0, y); y \in \mathbb{R}\}$$

Unterräume des \mathbb{R}^2 . $U_v = U_1 \cup U_2$ ist die Teilmenge des \mathbb{R}^2 , bei der mindestens eine der beiden Koordinaten gleich 0 ist. Mit $x_1 = (1, 0)^T \in U_v$ und $x_2 = (0, 1)^T \in U_v$ gilt aber $x_1 + x_2 = (1, 1)^T \notin U_v$, also ist U_v kein Untervektorraum.

Im Gegensatz zur Vereinigung von Mengen erhält die Summe von Untervektorräumen die Vektorraumeigenschaft.

Definition 3.34:

1. Seien U_1, U_2 Unterräume von V . Die *Summe* von U_1 und U_2 ist dann definiert durch

$$U_1 + U_2 := \{v \in V \mid \exists u_1 \in U_1, u_2 \in U_2 : v = u_1 + u_2\}.$$

2. Ein Vektorraum W heißt *direkte Summe* von U_1 und U_2 , Schreibweise $W = U_1 \oplus U_2$, wenn gilt $W = U_1 + U_2$ und $U_1 \cap U_2 = \{0\}$.

Satz 3.35: Sind U_1 und U_2 Unterräume von V , ist $U_1 + U_2$ ein Unterraum von V .

BEWEIS Übungsaufgabe ■

Beispiel 3.36:

$$U_1 = \{(x, 0); x \in \mathbb{R}\}$$

$$U_2 = \{(0, y); y \in \mathbb{R}\}$$

sind Unterräume des \mathbb{R}^2 . Man hat

$$U_p = U_1 \oplus U_2 = \{(x + 0, 0 + y); x, y \in \mathbb{R}\} = \mathbb{R}^2.$$

Bemerkung 3.37: Gilt $V = U_1 \oplus U_2$, ist die Darstellung $v = u_1 + u_2$ mit $u_1 \in U_1$ und $u_2 \in U_2$ eindeutig für alle $v \in V$.

BEWEIS Sei $v \in U_1 \oplus U_2$ mit $v = u_1 + u_2 = w_1 + w_2$ mit $u_1, w_1 \in U_1$ und $u_2, w_2 \in U_2$. Dann gilt $0 = u_1 - w_1 + u_2 - w_2$, also $u_1 - w_1 = -(u_2 - w_2)$. Aus der Abgeschlossenheit von U_1 folgt $u_1 - w_1 \in U_1$ und deswegen $u_2 - w_2 \in U_1$, aber aus der Abgeschlossenheit von U_2 auch $u_2 - w_2 \in U_2$. Wegen $U_1 \cap U_2 = \{0\}$ n. V. erhält man $u_2 = w_2$ und damit dann $u_1 = w_1$. ■

3.4 Lineare Unabhängigkeit, Basis, Dimension

3.4.1 Lineare Unabhängigkeit

Man wird intuitiv die Ebene als zweidimensionales Objekt auffassen und damit dem Zahlraum \mathbb{R}^2 , den man ja mit der Ebene identifizieren kann, die Dimension 2 zuordnen wollen. Analog liegt es nahe, dem Zahlraum \mathbb{R}^3 die Dimension 3 zuzuordnen. Es stellt sich die Frage, ob und wie man einem allgemeinen Vektorraum eine Dimension sinnvoll zuordnen kann. Bis diese Frage beantwortet werden kann, ist einige Vorarbeit zu leisten. Wir beginnen mit der einfachen Beobachtung, dass im \mathbb{R}^3 jeder Vektor $v = (x, y, z)^T$ eindeutig als gewichtete Summe der drei kanonischen Einheitsvektoren geschrieben werden kann:

$$v = x \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + z \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Wir verallgemeinern dies in folgender Definition.

Definition 3.38: Seien $v_1, \dots, v_r \in V$ Vektoren eines K -Vektorraums V :

- (a) Lässt sich $v \in V$ als Summe dieser Vektoren mit Vorfaktoren darstellen,

$$v = \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_r v_r \quad \text{mit } \lambda_i \in K,$$

heißt v *Linearkombination* von v_1, \dots, v_r .

- (b) Die Menge

$$L(v_1, \dots, v_r) := \{\lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_r v_r \mid \lambda_i \in K\} \subseteq V$$

aller Linearkombinationen heißt *Lineare Hülle* von v_1, \dots, v_r .

Nach Definition ist jede Lineare Hülle ein Unterraum von V .

Beispiel 3.39: Gegeben seien die Vektoren

$$a = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \quad \text{und} \quad b = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

Dann ist

$$c = \begin{pmatrix} 0 \\ 2 \\ 3 \end{pmatrix} = 1 \cdot a + (-1) \cdot b \in L(a, b)$$

eine Linearkombination von a und b und liegt somit auch in

$$L(a, b) = \{\lambda_1 a + \lambda_2 b; \lambda_1, \lambda_2 \in \mathbb{R}\}.$$

Beispiel 3.40: Beispiele in \mathbb{R}^3 :

$$L(e_1, e_2, e_3) = \{(\lambda_1, \lambda_2, \lambda_3)^T \mid \lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}\} = \mathbb{R}^3$$

$$L(e_1, e_2) = \{(\lambda_1, \lambda_2, 0)^T \mid \lambda_1, \lambda_2 \in \mathbb{R}\} \quad (x_1\text{-}x_2\text{-Ebene})$$

$$L(e_1, e_2, e_1 + e_2) = \{(\lambda_1 + \lambda_3, \lambda_2 + \lambda_3, 0)^T \mid \lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}\} \quad (x_1\text{-}x_2\text{-Ebene})$$

Der englische Ausdruck für „Lineare Hülle“ lautet *span*. Auch im Deutschen schreibt man oft $\text{span}(e_1, e_2)$ statt $L(e_1, e_2)$. Der deutsche Ausdruck *Spann* für die lineare Hülle wird selten verwendet. Gängig ist aber eine Wendung wie „Die Vektoren e_1 und e_2 spannen die $x_1\text{-}x_2\text{-Ebene}$ auf.“

Beispiel 3.40 zeigt, dass man den gesamten Zahlraum \mathbb{R}^3 aus den drei Vektoren e_1, e_2 und e_3 durch Linearkombination erzeugen kann. Dies führt zu folgender Definition.

Definition 3.41:

1. Sei V ein K -Vektorraum und (v_1, \dots, v_n) ein n -Tupel von Vektoren in V . Gilt $L(v_1, \dots, v_n) = V$, nennt man (v_1, \dots, v_n) ein *Erzeugendensystem* von V .
2. V heißt *endlich erzeugt*, wenn es endlich viele Vektoren v_1, \dots, v_r gibt, so dass $L(v_1, \dots, v_r) = V$, ansonsten nicht endlich erzeugt.

Beispiel 3.42: Das Tripel (e_1, e_2, e_3) der kanonischen Einheitsvektoren ist ein Erzeugendensystem von \mathbb{R}^3 ; ebenso erzeugen die Vektoren $(0, 1)^T$ und $(1, 0)^T$ den Zahlraum \mathbb{R}^2 . Offenbar sind \mathbb{R}^2 und \mathbb{R}^3 endlich erzeugt. Ebenso ist der Zahlraum \mathbb{R}^n durch e_1, \dots, e_n endlich erzeugt.

Wir nehmen ab jetzt im ganzen Kapitel an, der Vektorraum V sei endlich erzeugt.

Man kann für einen Vektorraum viele Erzeugendensysteme angeben, so ist neben $E = (e_1, e_2, e_3)$ auch $\tilde{E} = (e_1, e_2, e_3, e_1 + e_2)$ ein Erzeugendensystem von \mathbb{R}^3 , allerdings ein unnötig großes, weil der zusätzliche Vektor $e_1 + e_2$ die Lineare Hülle nicht mehr vergrößert. Da jeder Vektorraum ein Nullelement enthält, betrachten wir dieses beispielhaft und schreiben die 0 als Linearkombination der Vektoren des Erzeugendensystems. In E lässt sich die Null nur darstellen als $0 = 0e_1 + 0e_2 + 0e_3$, in \tilde{E} gilt $0 = 0e_1 + 0e_2 + 0e_3 + 0(e_1 + e_2)$, aber auch $0 = (-1)e_1 + (-1)e_2 + 0e_3 + 1(e_1 + e_2)$. Die Redundanz in \tilde{E} hängt also mit der Möglichkeit zusammen, die 0 auf verschiedene Weise als Linearkombination zu schreiben.

Definition 3.43: Sei V ein K -Vektorraum. Ein r -Tupel (v_1, \dots, v_r) von Vektoren in V heißt *linear unabhängig*, wenn aus $\lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_r v_r = 0$ stets folgt, dass $\lambda_1 = \lambda_2 = \dots = \lambda_r = 0$ gilt.

Ist die Darstellung des Nullvektors als Linearkombination eindeutig, gilt dies sogar für *jeden* Vektor.

Bemerkung 3.44: Alle Vektoren v aus $L(v_1, \dots, v_n)$ sind genau dann eindeutig als Linearkombination der Vektoren v_1, \dots, v_n darstellbar, wenn das n -Tupel (v_1, \dots, v_n) linear unabhängig ist.

BEWEIS Seien

$$\begin{aligned} v &= \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n \\ &= \beta_1 v_1 + \beta_2 v_2 + \dots + \beta_n v_n \end{aligned}$$

zwei verschiedene Darstellungen von v . Es gilt

$$(\alpha_1 - \beta_1)v_1 + (\alpha_2 - \beta_2)v_2 + \dots + (\alpha_n - \beta_n)v_n = 0.$$

Die Vektoren v_1, \dots, v_n sind nach Definition genau dann linear unabhängig, wenn alle Differenzen $\alpha_n - \beta_n$ gleich 0 sind. ■

Bemerkung 3.45: Für linear unabhängige Vektoren gilt:

- Keiner der Vektoren ist eine Linearkombination der übrigen.
- Keiner der Vektoren ist der Nullvektor.

Beispiel 3.46 (\mathbb{R}^n): Zwei Vektoren v_1 und v_2 sind linear abhängig, wenn

$$\lambda_1 v_1 + \lambda_2 v_2 = 0$$

mit λ_1 oder $\lambda_2 \neq 0$ gilt. Sei o. B. d. A. $\lambda_1 \neq 0$ (für $\lambda_2 \neq 0$ ist die Rechnung ähnlich). Dann ist

$$v_1 = -\frac{\lambda_2}{\lambda_1} v_2.$$

Diese Gleichung ist erfüllt, wenn entweder $v_1 = 0$ und $\lambda_2 = 0$ (aber $\lambda_1 \neq 0$) ist oder v_1 und v_2 parallel sind. Zwei Vektoren des \mathbb{R}^n sind also genau dann linear abhängig, wenn sie entweder parallel sind oder einer der beiden Vektoren der Nullvektor ist.

Die n Vektoren v_1, \dots, v_n sind genau dann linear abhängig, wenn

$$\lambda_1 v_1 + \dots + \lambda_n v_n = 0$$

gilt, wobei mindestens einer der Vorfaktoren $\lambda_1, \dots, \lambda_n \neq 0$ ist. Sei $\lambda_1 \neq 0$ (für die anderen λ_i ist die Rechnung äquivalent). Dann gilt die Gleichung in zwei Fällen:

1. $v_1 = 0$. Dann können alle anderen $\lambda_i = 0$ sein.
2. $v_1 = -\frac{1}{\lambda_1}(\lambda_2 v_2 + \dots + \lambda_n v_n)$. Dann ist v_1 eine Linearkombination der anderen v_i .

Beispiel 3.47: Das n -Tupel der kanonischen Einheitsvektoren im \mathbb{R}^n ist linear unabhängig, das 4-Tupel $\tilde{E} = (e_1, e_2, e_3, e_1 + e_2)$ von oben nicht.

3.4.2 Nachweis linearer Unabhängigkeit

Wir betrachten zunächst den Vektorraum \mathbb{K}^n , der sich als prototypisch für endlich erzeugte reelle oder komplexe Vektorräume erweisen wird. Gegeben seien Vektoren v_1, \dots, v_r , $r \leq n$, die auf lineare Unabhängigkeit getestet werden sollen. Die Gleichung

$$\lambda_1 v_1 + \dots + \lambda_r v_r = 0$$

aus der Definition 3.43 ist ein lineares Gleichungssystem in den Koeffizienten λ_j . Die Vektoren sind genau dann linear unabhängig, wenn man als einzige Lösung den Nullvektor erhält.

Beispiel 3.48: Gegeben sind

$$v_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}, \quad v_3 = \begin{pmatrix} -1 \\ 2 \\ 5 \end{pmatrix}.$$

Sie führen zum Gleichungssystem

$$\begin{cases} \lambda_1 + 2\lambda_2 - \lambda_3 = 0 \\ 2\lambda_1 + 3\lambda_2 + 2\lambda_3 = 0 \\ \lambda_1 + \lambda_2 + 5\lambda_3 = 0 \end{cases}$$

Mit dem Gauß-Verfahren erhält man:

$$\begin{array}{ccc|c} 1 & 2 & -1 & 0 \\ 2 & 3 & 2 & 0 & | -2 \cdot (\text{I}) \\ 1 & 1 & 5 & 0 & | -(\text{I}) \\ \hline 1 & 2 & -1 & 0 \\ 0 & -1 & 4 & 0 \\ 0 & -1 & 6 & 0 & | -(\text{II}) \\ \hline 1 & 2 & -1 & 0 \\ 0 & -1 & 4 & 0 \\ 0 & 0 & 2 & 0 \end{array}$$

Daraus folgt $\lambda_3 = \lambda_2 = \lambda_1 = 0$. Also sind die drei Vektoren linear unabhängig.

Beispiel 3.49: Wir betrachten die Vektoren

$$v_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}, \quad v_3 = \begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix}$$

Wir lösen das korrespondierende lineare Gleichungssystem mit dem Gaußschen Eliminationsverfahren:

$$\begin{array}{ccc|c}
 1 & 2 & -1 & 0 \\
 2 & 3 & 2 & 0 & | -2 \cdot (\text{I}) \\
 1 & 1 & 3 & 0 & | -(\text{I}) \\
 \hline
 1 & 2 & -1 & 0 \\
 0 & -1 & 4 & 0 \\
 0 & -1 & 4 & 0 & | -(\text{II}) \\
 \hline
 1 & 2 & -1 & 0 \\
 0 & -1 & 4 & 0 \\
 0 & 0 & 0 & 0
 \end{array}$$

Es gibt nur noch zwei Gleichungen mit drei Unbekannten, so dass zur Lösung $\lambda_1 = \lambda_2 = \lambda_3 = 0$ (Trivialsolution) noch andere Lösungen hinzukommen. In diesem Fall sind dies:

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = \alpha \cdot \begin{pmatrix} -7 \\ 4 \\ 1 \end{pmatrix}.$$

Im Spezialfall $r = n$ erhält man ein $n \times n$ -Gleichungssystem. Wir haben gesehen, dass sich ein 3×3 -Gleichungssystem genau dann eindeutig lösen lässt, wenn die Determinante der Koeffizientenmatrix A ungleich 0 ist. Da $\lambda_1 = \dots = \lambda_3 = 0$ eine Lösung ist, folgt:

$$\begin{aligned}
 \det(A) \neq 0 &\Leftrightarrow \text{Lösung des Gleichungssystems ist eindeutig} \\
 &\Leftrightarrow \text{Nur } \lambda_1 = \dots = \lambda_3 = 0 \text{ ist Lösung} \\
 &\Leftrightarrow \text{Die Vektoren sind linear unabhängig.}
 \end{aligned}$$

Es würde in den vorigen Beispielen also schon ausreichen, die Determinanten der Koeffizientenmatrizen zu prüfen. Für das erste, linear unabhängige Beispiel gilt:

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

Für das zweite, linear abhängige Beispiel ist:

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

Wir betrachten nun den Vektorraum $\mathcal{C}[a, b]$. Vektoren dieses Vektorraums sind stetige Funktionen. Zum Überprüfen der Linearen Unabhängigkeit geht man hier üblicherweise von der Definition aus. Man muss also zeigen, dass die Nullfunktion nur trivial darstellbar ist bzw. für lineare Abhängigkeit, dass eine Funktion als Linearkombination der anderen dargestellt werden kann.

Beispiel 3.50:

1. Die Funktionen

$$f_1(x) = x; \quad f_2(x) = x^2; \quad f_3(x) = x^2 - 2x$$

sind auf einem beliebigen Intervall $[a, b]$ linear abhängig, denn

$$2f_1(x) - f_2(x) + f_3(x) = 2x - x^2 + x^2 - 2x = 0 \quad \forall x \in \mathbb{R}$$

2. Die Funktionen

$$f_1(x) = \sin^2(x); \quad f_2(x) = \cos^2(x); \quad f_3(x) = 1$$

sind auf einem beliebigen Intervall $[a, b]$ linear abhängig, denn

$$f_1(x) + f_2(x) - f_3(x) = \cos^2(x) + \sin^2(x) - 1 = 0 \quad \forall x \in \mathbb{R}$$

Beispiel 3.51: Die Funktionen

$$f_1(x) = x^2; \quad f_2(x) = x$$

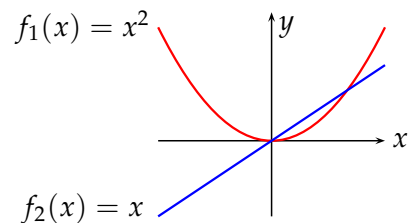
sind auf einem beliebigen Intervall $[a, b]$ linear unabhängig. Zum Nachweis dessen müssen wir alle Lösungen von

$$\lambda_1 f_1(x) + \lambda_2 f_2(x) = \lambda_1 x^2 + \lambda_2 x = 0$$

berechnen. Wir unterscheiden zwei Fälle.

1. Sei $\lambda_1 = 0$. Es folgt $\lambda_2 x = 0 \forall x$. Dies ist nur für $\lambda_2 = 0$ möglich.
2. Sei $\lambda_1 \neq 0$. Dann gilt $x^2 + \frac{\lambda_2}{\lambda_1} x = 0$. Diese Gleichung ist erfüllt für $x = 0$ und $x = -\frac{\lambda_2}{\lambda_1}$, aber nicht für alle x . Somit bleibt für alle x als einzige Lösung $\lambda_1 = \lambda_2 = 0$.

Damit sind $f_1(x)$ und $f_2(x)$ linear unabhängig.



Man kann in Gegensatz zum \mathbb{K}^n kein festes Rechenschema angeben, mit dessen Hilfe man in jedem Fall über Lineare Abhängigkeit oder Unabhängigkeit von Funktionen entscheiden kann. Folgende Aussage kann hier nützlich sein.

Bemerkung 3.52: Wir bilden zu den Funktionen f_1, \dots, f_n für paarweise verschiedene x_1, \dots, x_n die n Vektoren

$$(f_1(x_1), \dots, f_1(x_n))^T, \dots, (f_n(x_1), \dots, f_n(x_n))^T \in \mathbb{K}^n.$$

Sind diese Vektoren linear unabhängig, dann sind die Funktionen f_1, \dots, f_n selbst linear unabhängig.

BEWEIS Sei $0 = \sum_{i=1}^n \lambda_i f_i$. Diese Gleichung gilt punktweise im gesamten Definitionsbereich, also insbesondere für die x_j . Aus der linearen Unabhängigkeit der Vektoren $(f_i(x_j))_{i=1}^n$ folgt $\lambda_j = 0$ und daraus die Behauptung.

■

Beispiel 3.53: Wir betrachten die Funktionen $f_1(x) = 1, f_2(x) = \sin x, f_3(x) = \cos x$ auf $[-\pi, \pi]$ und wählen die Punkte $x_1 = 0, x_2 = \pi/2$ und $x_3 = \pi$. Man erhält die Vektoren

$$u = \begin{pmatrix} f_1(x_1) \\ f_1(x_2) \\ f_1(x_3) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad v = \begin{pmatrix} f_2(x_1) \\ f_2(x_2) \\ f_2(x_3) \end{pmatrix} = \begin{pmatrix} \sin 0 \\ \sin \pi/2 \\ \sin \pi \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

und noch

$$w = \begin{pmatrix} f_3(x_1) \\ f_3(x_2) \\ f_3(x_3) \end{pmatrix} = \begin{pmatrix} \cos 0 \\ \cos \pi/2 \\ \cos \pi \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

Wegen $\det(u, v, w) = -2 \neq 0$ sind u, v und w linear unabhängig und damit nach Bemerkung 3.52 auch die Funktionen f_1, f_2 und f_3 .

Der Nachweis der linearen Unabhängigkeit von Funktionen mittels Bemerkung 3.52 setzt aber eine geeignete Wahl der Auswertungspunkte x_i voraus. Wählt man statt x_1, x_2, x_3 die Punkte $x'_1 = -\pi, x'_2 = 0$ und $x'_3 = \pi$, erhält man analog zu oben die Vektoren $u' = u, v' = 0$ und $w' = (-1, 0, -1)^T$, die linear abhängig sind, weil sich unter ihnen der Nullvektor befindet. Trotzdem sind die Funktionen f_1, f_2, f_3 nach wie vor linear unabhängig. Man kann also von der linearen *Unabhängigkeit* der Vektoren auf die lineare *Unabhängigkeit* der Funktionen schließen, nicht aber von ihrer linearen *Abhängigkeit* auf die lineare *Abhängigkeit* der Funktionen.

Üblicherweise muss man zum Nachweis der linearen Abhängigkeit oder Unabhängigkeit die Besonderheiten der jeweiligen Funktionen ausnutzen und Hilfsmittel aus der Analysis anwenden.

3.4.3 Basis und Dimension

Definition 3.54: Sei $V \neq \{0\}$ ein endlich erzeugter K -Vektorraum und $v_1, \dots, v_n \in V$. Das n -Tupel $\mathcal{B} = (v_1, \dots, v_n)$ heißt *Basis* oder *minimales Erzeugendensystem* von V , wenn \mathcal{B} linear unabhängig ist und wenn gilt $L(\mathcal{B}) = V$. Weiter sei \emptyset die Basis des Nullvektorraums $\{0\}$.

Sei ab jetzt stets $V \neq \{0\}$.

Definition 3.55: Sei (v_1, \dots, v_n) eine Basis von V und $v \in V$ mit

$$v = \lambda_1 v_1 + \dots + \lambda_n v_n,$$

$\lambda_1, \dots, \lambda_n \in K$. Das n -Tupel der obigen Vorfaktoren $(\lambda_1, \dots, \lambda_n) \in K^n$ nennt man *Koordinaten* des Vektors v bzgl. der Basis (v_1, \dots, v_n) .

Bemerkung 3.56: Die Koordinaten eines Vektors bzgl. einer gegebenen Basis sind eindeutig.

BEWEIS Wegen $V = L(v_1, \dots, v_n)$ existiert zu jedem $v \in V$ mindestens ein $(\lambda_1, \dots, \lambda_n) \in K^n$ mit

$$v = \lambda_1 v_1 + \dots + \lambda_n v_n$$

Da (v_1, \dots, v_n) linear unabhängig ist, ist diese Darstellung eindeutig nach Bemerkung 3.44. ■

Es liegt nun nahe, die Dimension eines Vektorraums über die Anzahl der Vektoren zu definieren, die in einer Basis enthalten sind. Dies ist ja zugleich die Anzahl der Koordinaten eines Vektors im gegebenen Vektorraum. Der Dimensionsbegriff ist aber nur dann wohldefiniert, wenn alle Basen gleichviele Vektoren enthalten. Es ist nicht einmal klar, ob überhaupt jeder Vektorraum eine Basis besitzt. Wir beschränken uns jetzt erneut auf endlich erzeugte Räume. Um den Dimensionsbegriff wie oben skizziert definieren zu können, benötigen wir 2 Aussagen:

- Jeder endlich erzeugte Vektorraum besitzt eine Basis.
- Alle Basen bestehen aus gleich vielen Vektoren (Eindeutigkeit der Basislänge).

Wir beginnen mit der Existenz von Basen.

Satz 3.57 (Basisergänzungssatz): Sei V ein K -Vektorraum und seien

$$v_1, \dots, v_r, w_1, \dots, w_s$$

Vektoren in V . Ist (v_1, \dots, v_r) linear unabhängig und ist

$$L(v_1, \dots, v_r, w_1, \dots, w_s) = V,$$

dann kann man (v_1, \dots, v_r) durch evtl. Hinzunahme geeigneter Vektoren aus $\{w_1, \dots, w_s\}$ zu einer Basis von V ergänzen.

BEWEIS Im Fall $L(v_1, \dots, v_r) = V$ ist die Aussage offensichtlich. Sei also

$$L(v_1, \dots, v_r) \neq V.$$

Dann existiert mindestens ein w_i mit $w_i \notin L(v_1, \dots, v_r)$, denn wären alle $w_i \in L(v_1, \dots, v_r)$, dann müsste $L(v_1, \dots, v_r) = L(v_1, \dots, v_r, w_1, \dots, w_s) = V$ gelten, was der Annahme $L(v_1, \dots, v_r) \neq V$ widerspricht.

Das Tupel (w_i, v_1, \dots, v_r) ist linear unabhängig, denn aus $\sum_{j=1}^r \lambda_j v_j + \lambda w_i = 0$ folgt $\lambda = 0$, da $w_i \notin L(v_1, \dots, v_r)$, und weiter folgt dann $\lambda_j = 0 \forall j$, da alle v_j linear unabhängig sind.

Möglicherweise ist (w_i, v_1, \dots, v_r) aber noch keine Basis. Dann wird der vorige Schritt wiederholt und es werden solange weitere w_i dazugenommen, bis sie (v_1, \dots, v_r) zu einer Basis von V ergänzen. Dies ist nach endlich vielen Schritten möglich wegen $L(v_1, \dots, v_r, w_1, \dots, w_s) = V$. ■

Folgerung 3.58: Jeder endlich erzeugte Vektorraum V hat eine Basis.

BEWEIS Weil V endlich erzeugt ist, gibt es $v_1, \dots, v_n \in V$ mit $L(v_1, \dots, v_n) = V$. Streicht man alle Nullvektoren im n -Tupel (v_1, \dots, v_n) , so erhält man evtl. nach Umindizierung ein m -Tupel (v_1, \dots, v_m) mit $L(v_1, \dots, v_m) = V$. Aufgrund von $v_1 \neq 0$ ist (v_1) linear unabhängig. Nach dem Basisergänzungssatz 3.57 lässt sich (v_1) mit Vektoren aus v_2, \dots, v_m zu einer Basis von V ergänzen. ■

Beispiel 3.59: Sei $V = \mathbb{R}^3$. Weiterhin sei

$$v_1 = (1, 0, 0); \quad v_2 = (0, 1, 0).$$

Die Vektoren v_1 und v_2 sind linear unabhängig. Außerdem sei

$$w_1 = (1, 1, 0); \quad w_2 = (0, 0, 1).$$

Da $L(v_1, v_2, w_1, w_2) = \mathbb{R}^3$, sind die Voraussetzungen für den Basisergänzungssatz erfüllt. Dieser besagt: Da (v_1, v_2) keine Basis des \mathbb{R}^3 ist, ist mindestens eines der Tripel (v_1, v_2, w_1) , (v_1, v_2, w_2) oder (v_1, v_2, w_1, w_2) eine Basis des \mathbb{R}^3 , d. h. linear unabhängig und erzeugend. In diesem Beispiel ist dies das Tripel (v_1, v_2, w_2) . Alle vier Vektoren bilden keine Basis des \mathbb{R}^3 , da sie linear abhängig sind.

Satz 3.60 (Austauschlemma): Sind (v_1, \dots, v_n) und (w_1, \dots, w_m) Basen eines K -Vektorraums V , dann gibt es zu jedem v_i ein w_j , so dass aus (v_1, \dots, v_n) wieder eine Basis entsteht, wenn man in ihr v_i durch w_j ersetzt.

BEWEIS Seien (v_1, \dots, v_n) und (w_1, \dots, w_m) zwei Basen von V . Aus der ersten Basis wird der Vektor v_i gestrichen. Für das verkürzte Tupel $(v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n)$ gilt

$$L(v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n) \neq V,$$

denn wäre $L(v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n) = V$, dann wäre auch

$$v_i \in L(v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n),$$

ließe sich also durch eine Linearkombination der restlichen Vektoren darstellen. Damit wären (v_1, \dots, v_n) linear abhängig und keine Basis im Widerspruch zur Voraussetzung.

Damit existiert nach dem Basisergänzungssatz 3.57 ein w_j mit

$$w_j \notin L(v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n),$$

und das Tupel $(v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n, w_j)$ ist linear unabhängig.

Handelt es sich nicht um eine Basis, könnte man es durch einen der Vektoren v_1, \dots, v_n zu einer Basis ergänzen. Alle Vektoren außer v_i kommen dazu trivialerweise nicht in Frage. Aber auch v_i ist nicht möglich. In diesem Fall wäre die Basis nämlich (v_1, \dots, v_n, w_j) . Da aber bereits (v_1, \dots, v_n) eine Basis ist, ist w_j linear abhängig von diesen Vektoren und (v_1, \dots, v_n, w_j) ist keine Basis. Damit muss $(v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n, w_j)$ bereits eine Basis von V sein. ■

Beispiel 3.61: Wir betrachten die Vektoren

$$\begin{array}{lll} v_1 = (1, 0, 0); & v_2 = (0, 1, 0); & v_3 = (0, 0, 1) \\ w_1 = (2, 0, 0); & w_2 = (0, 2, 0); & w_3 = (0, 0, 3) \end{array}$$

Hierbei sind (v_1, v_2, v_3) und (w_1, w_2, w_3) Basen des \mathbb{R}^3 . Aus der ersten Basis wird v_3 entfernt. Nach dem Austauschlemma kann man dafür einen der drei Vektoren w_i einsetzen, so dass wieder eine Basis entsteht. Im Beispiel ist dies w_3 . Das Tripel (v_1, v_2, w_3) ist eine Basis des \mathbb{R}^3 .

Wir zeigen nun die Eindeutigkeit der Basislänge in endlich erzeugten Vektorräumen und legen damit die Grundlagen für den Dimensionsbegriff.

Satz 3.62 (Eindeutigkeit der Basislänge): Sind (v_1, \dots, v_n) und (w_1, \dots, w_m) Basen eines K -Vektorraums V , dann gilt $n = m$.

BEWEIS Nimmt man $n < m$ an (der Fall $n > m$ ist äquivalent), dann kann man mit Hilfe des Austauschlemmas aus der Basis (w_1, \dots, w_m) eine Basis $(v_1, \dots, v_n, w_{n+1}, \dots, w_m)$ erzeugen. Nun ist aber (v_1, \dots, v_n) bereits eine Basis, so dass $(v_1, \dots, v_n, w_{n+1}, \dots, w_m)$ linear abhängig sein müssen. Dies ist ein Widerspruch zur linearen Unabhängigkeit der Basiselemente. ■

Definition 3.63: Besitzt ein Vektorraum $V \neq \{0\}$ eine Basis (v_1, \dots, v_n) , so definieren wir die *Dimension* von V als $\dim(V) := n$; besitzt V keine endliche Basis, dann setzt man $\dim(V) := \infty$. Weiter sei $\dim(\{0\}) := 0$.

Die folgende Bemerkung liefert uns bei der Suche nach einer Basis für einen vorgegeben Vektorraum einen entscheidenden Hinweis:

Bemerkung 3.64: Ist $\dim(V) = n$ und sind v_1, \dots, v_n n linear unabhängige Vektoren in V , so ist (v_1, \dots, v_n) eine Basis von V .

BEWEIS Andernfalls könnte man (v_1, \dots, v_n) nach dem Basisergänzungssatz 3.57 um $k \geq 1$ Vektoren aus einer Basis (w_1, \dots, w_n) zu einer Basis ergänzen. Dann jedoch gäbe es eine Basis der Dimension $n + k$, was ein Widerspruch zur Eindeutigkeit der Basislänge ist. ■

Beispiel 3.65: Die Vektoren $v_1 = (1, 0)$ und $v_2 = (0, 1)$ bilden eine Basis des \mathbb{R}^2 .

$$\lambda_1 \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \lambda_2 \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0 \quad \Rightarrow \quad \lambda_1 = \lambda_2 = 0$$

Es sind somit v_1 und v_2 linear unabhängig und bilden wegen $\dim(\mathbb{R}^2) = 2$ eine Basis.

Als Anwendung der in diesem Kapitel entwickelten Theorie untersuchen wir abschließend die Dimension von Summen von Untervektorräumen.

Satz 3.66: Seien U, W Untervektorräume eines endlich erzeugten Vektorraums. Dann gilt

$$\dim(U + W) = \dim(U) + \dim(W) - \dim(U \cap W).$$

BEWEIS Sei $\dim(U) = n$, $\dim(W) = m$ und zunächst $U \cap W = \{0\}$. Wir wählen Basen (u_1, \dots, u_n) und (w_1, \dots, w_m) von U bzw. W . Die Vektoren $(u_1, \dots, u_n, w_1, \dots, w_m)$ erzeugen dann $U + W$: Für $v \in U + W$ existieren $u \in U$ und $w \in W$ mit $v = u + w$. Wegen $u = \sum_{i=1}^n \lambda_i u_i$ und $w = \sum_{i=1}^m \mu_i w_i$ für gewisse $\lambda_i, \mu_i \in K$ folgt

$$v = \sum_{i=1}^n \lambda_i u_i + \sum_{i=1}^m \mu_i w_i,$$

also $v \in L(u_1, \dots, u_n, w_1, \dots, w_m)$. Weiter sei

$$0 = \sum_{i=1}^n \lambda_i u_i + \sum_{i=1}^m \mu_i w_i$$

Es folgt

$$\underbrace{\sum_{i=1}^n \lambda_i u_i}_{\in U} = \underbrace{\sum_{i=1}^m -\mu_i w_i}_{\in W}$$

und damit $\sum_{i=1}^n \lambda_i u_i \in U \cap W$, also $\sum_{i=1}^n \lambda_i u_i = 0$. Aus der linearen Unabhängigkeit von (u_1, \dots, u_n) folgt $\lambda_i = 0, 1 \leq i \leq n$. Analog erschließt man

$\mu_i = 0, 1 \leq i \leq m$. Es ist also $(u_1, \dots, u_n, w_1, \dots, w_m)$ linear unabhängig, erzeugend und damit eine Basis von $U + W$.

Sei nun $\dim(U \cap W) = k > 0$ und (v_1, \dots, v_k) eine Basis von $U \cap W$. Nach Satz 3.57 ergänzen wir zu einer Basis $(v_1, \dots, v_k, u_1, \dots, u_{n-k})$ von U und andererseits zu einer Basis $(v_1, \dots, v_k, w_1, \dots, w_{m-k})$ von W . Sei

$$\mathcal{B} = (v_1, \dots, v_k, u_1, \dots, u_{n-k}, w_1, \dots, w_{m-k}).$$

Wir zeigen, dass \mathcal{B} eine Basis von $U + W$ bildet.

Lineare Unabhängigkeit: Man hat $u_i \notin U \cap W$, weil sonst $\dim(U \cap W) > k$ gälte aufgrund der linearen Unabhängigkeit von $v_1, \dots, v_k, u_1, \dots, u_{n-k}$. Analog erkennt man $w_i \notin U \cap W$. Weiter sind die Vektoren u_1, \dots, u_{n-k} und w_1, \dots, w_{m-k} linear unabhängig: Die Vektoren u_1, \dots, u_{n-k} und w_1, \dots, w_{m-k} sind für sich genommen linear unabhängig, weil sie aus einer Basis stammen. Wir nehmen jetzt, an, für irgendein u_j gälte $u_j \in L(w_1, \dots, w_{m-k})$. Dann gälte insbesondere $u_j \in U \cap W$, aber das ist nach obigen Überlegungen nicht möglich. Also liegt kein u_j in $L(w_1, \dots, w_{m-k})$, und daher sind die Vektoren linear unabhängig.

Erzeugend: Sei jetzt $v \in U + W$ beliebig. Dann existieren $u \in U$ und $w \in W$ mit $v = u + w$. Weil \mathcal{B} Basen von U und von W enthält, lassen sich u und w als Linearkombinationen von Vektoren aus \mathcal{B} schreiben und damit auch v (Linearkombinationen von Linearkombinationen sind wiederum Linearkombinationen). ■

Folgerung 3.67: Sei V ein Vektorraum mit $\dim(V) = n$, U_1, U_2 Unterräume und $V = U_1 \oplus U_2$. Dann gilt $\dim(U_1) + \dim(U_2) = n$.

3.4.4 Exkurs: Nicht endlich erzeugte Vektorräume

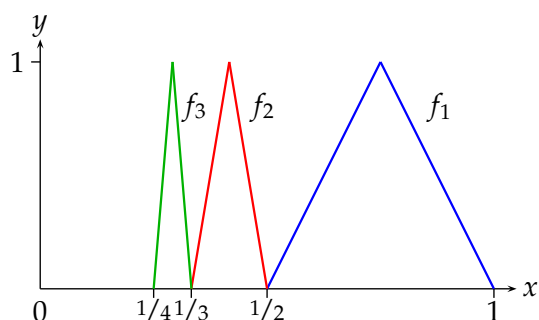
Wir betrachten den Vektorraum der stetigen Funktionen auf dem Intervall $[0, 1]$, also $\mathcal{C}[0, 1]$ und definieren stetige Funktionen f_n durch

$$f_n(x) = \left\{ \begin{array}{l} 0, \quad x \notin \left[\frac{1}{n+1}, \frac{1}{n} \right] \\ 1, \quad x = \left(\frac{1}{n+1} + \frac{1}{n} \right) / 2 \\ \text{linear auf den Zwischenintervallen} \end{array} \right\}, n \in \mathbb{N}.$$

Das n -Tupel (f_1, \dots, f_n) ist linear unabhängig für jedes $n \in \mathbb{N}$. Abbildung 3.3 zeigt die Funktionen f_1, f_2, f_3 , wobei die „Dreiecke“ von rechts nach links jeweils die Graphen der Funktionen f_1, f_2, f_3 darstellen. Außerhalb sind die Funktionswerte jeweils gleich Null.

Satz 3.68: Der Vektorraum $\mathcal{C}[0, 1]$ ist nicht endlich erzeugt.

BEWEIS Sonst sei \mathcal{B} eine Basis mit m Vektoren. Nach dem Austauschlemma lässt sich aus jeder endlichen Basis \mathcal{B} eine Basis \mathcal{B}' erzeugen, die die Funktionen f_1, \dots, f_{m+1} enthält. Dann aber enthielte \mathcal{B}' und damit auch \mathcal{B} mindestens $m + 1$ Vektoren. Dies ist ein Widerspruch. ■

Abbildung 3.3: Funktionen f_1, f_2 und f_3

Es erscheint wegen Satz 3.68 wünschenswert, den Basisbegriff auf nicht endlich erzeugte Vektorräume auszudehnen. Dazu erweitern wir zunächst den Begriff des n -Tupels.

Definition 3.69: Für eine beliebige geordnete Indexmenge I (z. B. die reellen Zahlen) und einer Menge $X \neq \emptyset$ bezeichnet man $(a_i)_{i \in I}$ mit $a_i \in X \forall i \in I$ als *Familie*.

Im Fall einer endlichen Indexmenge handelt es sich bei einer Familie um ein n -Tupel, im Fall einer abzählbar unendlichen Indexmenge I um eine Folge. Wir definieren nun durch Zurückführung auf den endlich erzeugten Fall:

Definition 3.70: Sei V ein Vektorraum und $(v_i)_{i \in I}$ eine Familie von Vektoren.

1. **(Lineare Hülle)** Sei $L(v_i)_{i \in I}$ die Menge aller Vektoren, die sich als Linearkombination von endlich vielen Vektoren $v_1, \dots, v_r \in (v_i)_{i \in I}$ darstellen lassen.
2. **(Erzeugendensystem)** Gilt $L(v_i)_{i \in I} = V$, so nennt man $(v_i)_{i \in I}$ ein *Erzeugendensystem*.
3. **(Lineare Unabhängigkeit)** Eine beliebige Familie heißt linear unabhängig, wenn jede endliche Teilfamilie linear unabhängig ist.

Bemerkung 3.71: Die eindeutige Darstellbarkeit eines Vektors als Linearkombination von Basisvektoren nach Bemerkung 3.44 gilt sofort in beliebigen Vektorräumen durch Auswahl einer endlichen Teilfamilie.

Beispiel 3.72: Die Funktionenfolge $(f_i)_{i \in \mathbb{N}}$ von oben ist offenbar linear unabhängig im Sinne von Definition 3.70, aber kein Erzeugendensystem von $\mathcal{C}[0, 1]$, weil z. B. die (stetige) konstante Funktion $g(x) = 1$ nicht dargestellt werden kann: Auf $[1/2, 1]$ ist nur die Funktion f_1 ungleich der Nullfunktion, zugleich aber ist kein Vielfaches von f_1 die Funktion g .

Jetzt stehen die Konzepte zur Verfügung, um den Basisbegriff auf beliebige Vektorräume ausdehnen zu können.

Definition 3.73: Sei V ein K -Vektorraum. Eine Familie $\mathcal{B} = (v_i)_{i \in I} \subseteq V$ heißt *Basis* oder *minimales Erzeugendensystem* von V , wenn \mathcal{B} linear unabhängig ist und wenn gilt $L(\mathcal{B}) = V$.

Bemerkung 3.74: Die Familie f_1, \dots ist kein Erzeugendensystem und daher auch keine Basis von $\mathcal{C}[0, 1]$.

Sei $\mathcal{C}^k[0, 1]$ die Menge aller k -fach stetig differenzierbaren Funktionen. Dann ist nach der Summen- und Faktorregel der Analysis $\mathcal{C}^k[0, 1]$ ein Vektorraum, und es ist $\mathcal{C}^{k+1}[0, 1]$ ein Unterraum $\mathcal{C}^k[0, 1]$ für alle k . Keiner dieser Räume hat endliche Dimension; man kann zum Nachweis analoge Funktionenfolgen zu f_1, \dots wie oben konstruieren, keine dieser Funktionenfolgen ist eine Basis. Es gilt aber:

Satz 3.75: Jeder Vektorraum hat eine Basis.

Zum Beweis dieses Satzes lassen sich die Techniken aus dem endlich erzeugten Fall nicht anwenden. Der Beweis verwendet stattdessen Hilfsmittel aus der Mengenlehre und wird hier ausgelassen. Zudem ist er nicht konstruktiv, er gibt also keinen Hinweis, wie man eine Basis eines eventuell unendlichdimensionalen Vektorraums konstruieren könnte.

Bemerkung 3.76: Man könnte dazu neigen, den recht abstrakten Begriff der Familie durch den eher vertrauten Begriff einer Folge zu ersetzen. Damit aber ließe man nur abzählbare Basen zu, was a priori nicht zu rechtfertigen wäre. In der Tat kann man zeigen, dass für alle $k \in \mathbb{N}$ die Vektorräume $\mathcal{C}^k[a, b]$ überabzählbare Basen besitzen. Konkret angeben lässt sich keine von ihnen.

3.4.5 Exkurs: Hyperebenen im \mathbb{R}^n

Mit der jetzt zur Verfügung stehenden Theorie können wir Geraden in \mathbb{R}^2 und Ebenen in \mathbb{R}^3 zur *Hyperebene* in allgemeinen endlich erzeugten Vektorräumen verallgemeinern. Wir gehen von der Beobachtung aus, dass man eine Gerade in der Ebene als verschobenen Untervektorraum deuten kann. Der Unterraum besitzt die Dimension $2 - 1 = 1$. Analog lässt sich eine Ebene in \mathbb{R}^3 als verschobenen Untervektorraum der Dimension $3 - 1 = 2$ deuten. Dies führt zur folgenden Definition.

Definition 3.77: Sei V ein K -Vektorraum, $\dim(V) = n$, U ein Untervektorraum mit $\dim(U) = n - 1$ und $p \in V$ beliebig.

1. Dann heißt

$$H(U; p) := p + U := \{x \in V \mid x = p + u, u \in U\}$$

Hyperebene, p heißt *Aufpunkt* der Hyperebene.

2. Existieren linear unabhängige Vektoren v_1, \dots, v_{n-1} , so dass

$$H = \{x \in V \mid x = p + \sum_{i=1}^{n-1} \alpha_i v_i, \alpha_i \in K\} \quad (3.1)$$

gilt, dann heißen v_1, \dots, v_{n-1} *Richtungsvektoren* von H , die Darstellung einer Hyperebene in Formel (3.1) heißt *Parameterform* der Hyperebene.

Bemerkung 3.78: Seien U, V, p wie oben.

1. Jede Hyperebene H besitzt eine Parameterdarstellung.
2. Richtungsvektoren sind nicht eindeutig.
3. Sei $H = p + U$ eine Hyperebene. Dann ist jeder Vektor $p' := p + u$ mit $u \in U$ ebenfalls Aufpunkt.

BEWEIS

1. Nach Folgerung 3.58 besitzt U eine Basis, deren Basisvektoren die Richtungsvektoren von U bilden.
2. Wir wählen eine Basis \mathcal{B} von U und irgendeinen Vektor $0 \neq u \in U \setminus \mathcal{B}$. Nach dem Austauschlemma 3.60 kann man dann eine neue Basis bilden, indem man dann einen geeignet gewählten Vektor in \mathcal{B} durch u ersetzt, erhält so eine neue Basis von U und damit andere Richtungsvektoren von H .
3. Offenbar gilt $u + U = U$ aufgrund der Abgeschlossenheit von U . Also hat man

$$H(U; p) = p + U = p + (u + U) = (p + u) + U = H(U; p'). \quad \blacksquare$$

Beispiel 3.79:

1. Sei $V = \mathbb{R}^2$. Dann wird jede Hyperebene in V durch einen Richtungsvektor v und einen Aufpunkt p erzeugt, so dass

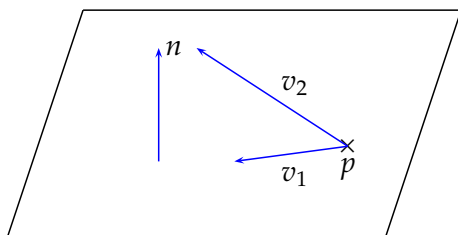
$$H = \{x \in \mathbb{R}^2 \mid x = p + \alpha v, \alpha \in \mathbb{R}\}$$

gilt und wir wie vorgesehen eine Gerade in der Ebene erhalten.

2. Eine Hyperebene in \mathbb{R}^3 ist gegeben durch

$$H = \{x \in \mathbb{R}^3 \mid x = p + \alpha_1 v_1 + \alpha_2 v_2, \alpha_1, \alpha_2 \in \mathbb{R}\}.$$

Ebenen in \mathbb{R}^3 sind also Spezialfälle von Hyperebenen; die lineare Unabhängigkeit der beiden Richtungsvektoren bedeutet, dass die beiden die Ebene aufspannenden Vektoren nicht in die gleiche Richtung zeigen (in dem Fall erhielte man ja eine Gerade und keine wirkliche Ebene).



Bemerkung 3.80: Hyperebenen sind genau dann Unterräume von V , wenn sie den Nullvektor enthalten, im Allgemeinen also nicht. Daher besitzen sie keine Dimension im Sinne unserer Definition 3.63.

Wir beschränken uns im folgenden auf den Fall $V = \mathbb{R}^n$.

Satz 3.81: Sei $V = \mathbb{R}^n$ und H eine Hyperebene. Dann existiert ein Normalenvektor $w \in \mathbb{R}^n \setminus \{0\}$ mit

$$H = \{x \in \mathbb{R}^n \mid \langle x, w \rangle = \langle p, w \rangle\}$$

Der Vektor w ist bis auf seine Länge eindeutig festgelegt.

BEWEIS Wir holen den Beweis nach, wenn die erforderliche Theorie entwickelt worden ist. ■

Definition 3.82: Die Gleichung $\langle x, w \rangle = \langle p, w \rangle$ heißt *Normalgleichung* oder auch *Normalform* der Hyperebene. Eine Normalform mit $\|w\| = 1$ heißt *Hessesche Normalform*.

Bemerkung 3.83:

1. Sei $w = (w_i)_{i=1}^n$ ein Normalenvektor der Hyperebene H und $p \in H$. Sei weiter $\langle p, w \rangle = c \in \mathbb{R}$. Dann lautet die Normalform von H :

$$\sum_{i=1}^n w_i x_i = c.$$

2. Nach 1. besteht ein lineares Gleichungssystem mit m Gleichungen und n Unbekannten aus m Hyperebenen in \mathbb{R}^n , die in Normalform vorliegend die Zeilen des linearen Gleichungssystems bilden. Da ein Lösungsvektor $x = (x_i)_{i=1}^n$ alle Gleichungen gleichzeitig erfüllen muss, erfüllt er alle Normalgleichungen dieser Hyperebenen zugleich, liegt also in deren Schnittmenge. Die Charakterisierung der Menge aller Lösungen eines linearen Gleichungssystems als Schnitt von m Hyperebenen in \mathbb{R}^n verallgemeinert die geometrische Deutung von 3×3 -Systemen aus Kapitel 2.3.2 auf beliebige lineare Gleichungssysteme.

Soll eine Normalform einer Hyperebene, die ja parameterfrei ist, aus einer Parameterform berechnet werden betrachtet man die Parametergleichung als lineares Gleichungssystem und versucht, alle Parameter zu eliminieren. Das Ergebnis ist eine parameterfreie Darstellung, also eine Normalform. Aufgrund von Satz 3.81 ist dies immer möglich. Zur vereinfachten Berechnung von w im Fall \mathbb{R}^2 und \mathbb{R}^3 verweisen wir auf Kapitel 2.2.2.

Beispiel 3.84: Wir demonstrieren das allgemeine Vorgehen anhand von \mathbb{R}^3 . Gegeben sei eine Ebene in Parameterform durch

$$x = \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix} + \alpha_1 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} -1 \\ 3 \\ -1 \end{pmatrix}.$$

Die Parametergleichungen führen zum linearen Gleichungssystem

$$\begin{array}{rcl} x_1 & = & 3 + \alpha_1 - \alpha_2 \\ x_2 & = & \alpha_1 + 3\alpha_2 \\ x_3 & = & 1 - \alpha_2 \\ \hline x_1 - x_2 & = & 3 - 4\alpha_2 \\ x_3 & = & 1 - \alpha_2 \\ \hline x_1 - x_2 - 4x_3 & = & -1 \end{array}$$

Zur Umrechnung einer Normalform in eine Parameterform werden $n - 1$ linear unabhängige Vektoren v_i benötigt mit $v_i \perp w$. Sie können z. B. mit dem folgenden Verfahren erzeugt werden.

Gegeben sei ein Normalenvektor $w = (w_i)_{i=1}^n$. Wegen $w \neq 0$ existiert mindestens ein $w_i \neq 0$. Für jeden auf w senkrechten Vektor führe man folgende Schritte durch:

1. Man vertausche w_i mit einer anderen Komponente w_j ; $j \neq i$.
2. Man ändere das Vorzeichen von w_i im so erzeugten Vektor.
3. Man setze alle Komponenten $w_k = 0$ für $k \neq i, j$.

Weil man bei einem Vektor der Länge n genau $n - 1$ verschiedene Möglichkeiten hat, $j \neq i$ zu wählen, entstehen so $n - 1$ Vektoren, die nach Konstruktion senkrecht auf w stehen. Für ein Beispiel hierzu verweisen wir auf Beispiel 2.70.

3.5 Polynome

Als eine einfache Anwendung der zuvor entwickelten Theorie betrachten wir *Polynome*, die zum einen in den verschiedensten Bereichen der Mathematik verwendet werden, und die zum anderen mit algebraischen Mitteln behandelt werden können.

Definition 3.85:

1. Ein *Polynom* oder ganzrationale Funktion $p : \mathbb{K} \rightarrow \mathbb{K}$ ist eine Funktion der Gestalt

$$p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n = \sum_{k=0}^n a_kx^k \quad (3.2)$$

mit den *Koeffizienten* $a_k \in \mathbb{K}, k = 0, 1, 2, \dots, n$. Der Koeffizient a_n heißt *Leitkoeffizient*. Im Fall $a_n = 1$ heißt p *normiert*.

2. Die Funktion $p(x) \equiv 0$ heißt *Nullpolynom*.
3. Ist $a_n \neq 0$, so heißt n *Grad* des Polynoms, Schreibweise $\deg(p) = n$. Weiter sei $\deg(0) = -\infty$.
4. Sei P_n die Menge aller Polynome mit einem Grad von höchstens n .

Bemerkung 3.86:

1. In der Algebra wird zwischen einem Polynom als einer endlichen Koeffizientenfolge mit Koeffizienten aus einem abstrakten Körper K und einer Polynomfunktion $p : K \rightarrow K$ im Sinne von Definition 3.85 unterschieden (vgl. z. B. [2]). Das ist unerlässlich, wenn man Polynome über endlichen Körpern betrachtet. Da wir uns auf die Körper \mathbb{R} oder \mathbb{C} beschränken, identifizieren wir im Folgenden Polynome und Polynomfunktionen.
2. Durch den Übergang des Laufindex von k nach $n - k$ in Formel (3.2) erhält man sofort

$$p(x) = \sum_{k=0}^n a_kx^k = \sum_{k=0}^n a_{n-k}x^{n-k}.$$

Wir werden nachfolgend zeigen, dass jedes Polynom vom Grad n höchstens n Nullstellen besitzt. Vorbereitend dafür betrachten wir das Abspalten von Linearfaktoren.

Satz 3.87: Sei $x_0 \in \mathbb{K}$ beliebig und p ein Polynom mit $\deg(p) = n \in \mathbb{N}$. Dann gilt für alle $x \in \mathbb{K}$:

$$p(x) = (x - x_0)p_{n-1}(x) + r$$

mit einem Polynom p_{n-1} vom Grad $n - 1$ und $r \in \mathbb{K}$.

BEWEIS Wir führen den Beweis mittels vollständiger Induktion. Für $n = 1$ gilt

$$\begin{aligned} p_1(x) &= a_1x + a_0 \\ &= a_1(x - x_0) + a_1x_0 + a_0 \\ &= a_1(x - x_0) + r \end{aligned}$$

Die Behauptung gelte nun für n . Dann errechnet man

$$\begin{aligned} p_{n+1}(x) &= p_n(x)x + a_0 \\ &= ((x - x_0)p_{n-1}(x) + r) \cdot x + a_0 \\ &= x(x - x_0)p_{n-1}(x) + rx + a_0 + \overbrace{rx_0 - rx_0}^{=0} \\ &= x(x - x_0)p_{n-1}(x) + r(x - x_0) + rx_0 + a_0 \\ &= (x - x_0) \overbrace{(xp_{n-1}(x) + r)}^{\tilde{p}_n(x)} + \overbrace{rx_0 + a_0}^s \\ &= (x - x_0)\tilde{p}_n(x) + s \end{aligned}$$

und erhält die Behauptung für $n + 1$. ■

Bemerkung 3.88: Seien x_0, p, n wie in Satz 3.87.

1. Es gilt $r = p(x_0)$.
2. Ist x_0 Nullstelle von p , dann gilt

$$p(x) = (x - x_0)p_{n-1}(x).$$

BEWEIS

1. Es gilt $p(x_0) = (x_0 - x_0)p_{n-1}(x_0) + r = r$.
2. Man hat dann $0 = p(x_0) = r$ nach 1.) ■

Eine wiederholte Anwendung von Bemerkung 3.88 auf p_{n-1} liefert folgenden Satz.

Satz 3.89: Hat ein Polynom p mit $\deg(p) = n \in \mathbb{N}$ genau n Nullstellen, dann gilt die *Faktorzerlegung*

$$p_n(x) = (x - x_1)(x - x_2) \cdots (x - x_n)a_n.$$

Daraus folgt direkt:

Satz 3.90: Ein Polynom vom Grad $n \in \mathbb{N}$ hat höchstens n Nullstellen.

Satz 3.91: Hat ein Polynom vom Grad $\leq n$ mehr als n Nullstellen, dann ist es das Nullpolynom, d. h. es gilt $a_0 = a_1 = \dots = a_n = 0$.

Bemerkung 3.92: Satz 3.90 besagt nicht etwa, dass ein reelles Polynom vom Grad n genau n reelle Nullstellen aufweisen müsste. So etwa besitzt das Polynom $p(x) = x^2 + 1$ überhaupt keine reelle Nullstelle. Eine solche Aussage gilt aber sehr wohl für $\mathbb{K} = \mathbb{C}$.

Satz 3.93 (Fundamentalsatz der Algebra): Jedes nichtkonstante Polynom besitzt mindestens eine komplexe Nullstelle.

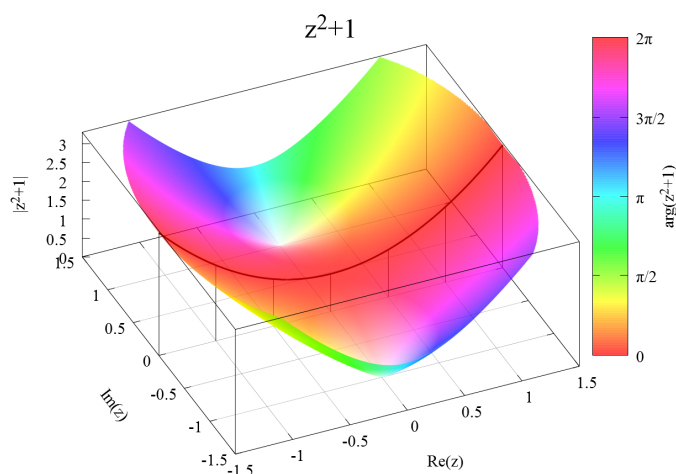
Ein vollständiger Beweis des Fundamentalsatzes wurde zuerst 1799 von C. F. Gauß in seiner Dissertation geführt. Wir verzichten hier aus Gründen des Aufwandes auf einen Beweis.

Folgerung 3.94: Jedes Polynom vom Grad $n \in \mathbb{N}$ weist genau n komplexe Nullstellen auf.

BEWEIS Sei p ein Polynom vom Grad n . Mit einer Nullstelle x_0 gilt dann nach Bemerkung 3.88 $p(x) = (x - x_0)p_{n-1}(x)$. Da nach dem Fundamentalsatz der Algebra p_{n-1} wieder eine Nullstelle besitzt, wendet man Bemerkung 3.88 auf p_{n-1} an und fährt solange fort, bis die Faktorzerlegung von p erreicht ist. ■

Beispiel 3.95: Das Polynom $p(z) = z^2 + 1$ besitzt die komplexen Nullstellen i und $-i$ (vgl. Abb. 3.4). Die Höhe der Fläche entspricht dem Absolutbetrag von $p(z)$, ihre Färbung dem Argument von $p(z)$, also dem Winkel in Polarkoordinatendarstellung. Eine komplexe Zahl mit einem Argument von $0, \pi$ und 2π entspricht somit einer reellen Zahl, bei einem Argument von $\pi/2$ oder $3/2\pi$ ist sie rein imaginär.

Bemerkung 3.96: Dass jedes Polynom (komplexe) Nullstellen besitzt, bedeutet nicht, dass man sie immer berechnen könnte. Die Nullstellen eines Polynoms vom Grad 2 lassen sich mit der bekannten pq -Formel ausrechnen, die schon in der Antike bekannt war. Für Polynome vom Grad 3 und 4 existieren ziemlich komplizierte Analogie, die sog. *Cardanischen Formeln*, die in der Renaissance von Cardano erstmalig veröffentlicht wurden. Die Suche

Abbildung 3.4: $z^2 + 1$ als komplexes Polynom

nach einer derartigen Formel für Polynome vom Grad mindestens 5 verlief über Jahrhunderte erfolglos, bis Abel 1824 mit algebraischen Methoden zeigte, dass es diese nicht geben kann. In der Praxis ist man daher in den allermeisten Fällen auf Näherungsrechnungen zur Nullstellenbestimmung angewiesen.

Wir betrachten nun Polynome als Elemente von Vektorräumen. Polynome werden als Funktionen punktweise addiert, ebenso erklärt man ein skalares Vielfaches eines Polynoms.

Satz 3.97: Seien $p, q \in P_n$. Dann sind für $\lambda \in \mathbb{K}$ sowohl λp als auch $(p + q)$ Polynome, und es gilt

$$\deg(\lambda p) \leq n, \quad \deg(p + q) \leq n$$

BEWEIS Die Polynome p und q besitzen die Darstellung $p(x) = \sum_{k=0}^n a_k x^k$ und $q(x) = \sum_{k=0}^n b_k x^k$, und deswegen gilt $(p + q)(x) = \sum_{k=0}^n (a_k x^k + b_k x^k) = \sum_{k=0}^n (a_k + b_k) x^k$. Es handelt sich also um ein Polynom mit Höchstgrad n . Die zweite Aussage folgt analog aus der Darstellung $(\lambda p)(x) = \sum_{k=0}^n \lambda a_k x^k$.

■

Folgerung 3.98: Sei $+$ die Addition von Funktionen und \cdot die Multiplikation einer Funktion mit einem Skalar. Dann bildet $(P_n, +, \cdot)$ einen \mathbb{K} -Vektorraum.

BEWEIS Da Polynome als Produkte und Summen stetiger Funktionen stetig sind, zeigen wir, dass P_n ein Untervektorraum von $\mathcal{C}(\mathbb{K})$ bildet. Die Abgeschlossenheit im Sinne der Definition 3.27 ist genau die Aussage von Satz 3.97. ■

Wir werden nun die Dimension von P_n bestimmen und eine Basis angeben.

Satz 3.99: Die Funktionen $1, x, x^2, \dots, x^n$ bilden eine Basis des Vektorraums P_n , und es gilt

$$\dim(P_n) = n + 1.$$

BEWEIS Nach Konstruktion von P_n bilden die Funktionen $1, x, x^2, \dots, x^n$ ein Erzeugendensystem von P_n . Zu zeigen bleibt die lineare Unabhängigkeit. Der Nullvektor in P_n ist das Nullpolynom, die Funktion $p(x) = 0 \forall x \in \mathbb{K}$. Offenbar ist diese als triviale Linearkombination der Funktionen x^k darstellbar. Wir müssen nun zeigen, dass dies die einzige Darstellung des Nullpolynoms ist. Wir führen den Beweis mittels vollständiger Induktion über den Polynomgrad n . Im Fall $n = 0$ folgt sofort $a_0 = 0$. Sei jetzt die Behauptung bewiesen für $n - 1$. Wir schreiben

$$p_n(x) = \underbrace{\sum_{k=0}^{n-1} a_k x^k}_{:=p_{n-1}(x)} + a_n x^n \equiv 0$$

mit $\deg(p_{n-1}) = n - 1$. Wäre $a_n \neq 0$, dann müsste $p_{n-1}(x) = -a_n x^n$ gelten, also $\deg(p_{n-1}) = n > n - 1$. Dies ist ein Widerspruch, also $a_n = 0$. Nach Voraussetzung folgt weiter $a_0 = \dots = a_{n-1} = 0$. ■

Bemerkung 3.100:

1. Die Basis $(1, x, \dots)$ aus Satz 3.99 wird *Monombasis* genannt.
2. Der Koeffizientenvektor (a_0, \dots, a_n) bildet die Koordinaten des Polynoms $p_n(x) = \sum_{k=0}^n a_k x^k$ bezüglich der Monombasis. Aufgrund der Eindeutigkeit der Koordinatendarstellung nach Bemerkung 3.56 sind die Koeffizienten eines Polynoms eindeutig festgelegt.
3. Stimmen zwei Polynome $p, q \in P_n$ in mindestens $(n + 1)$ Stellen überein, dann sind sie gleich, d. h. für

$$p_n(x) = \sum_{k=0}^n a_k x^k \quad \text{und} \quad q_n(x) = \sum_{k=0}^n b_k x^k$$

gilt dann

$$a_k = b_k, k = 0, 1, \dots, n.$$

BEWEIS Das *Differenzpolynom*

$$d_n(x) = p_n(x) - q_n(x) = \sum_{k=0}^n (a_k - b_k)x^k$$

hat nach Voraussetzung mindestens $n + 1$ Nullstellen. Also gilt $d_n(x) = 0$ nach Satz 3.91. ■

Wir kommen nun zu einer ersten konkreten Anwendung von Polynomen. Wir nehmen an, dass zu bestimmten Zeitpunkten $t_0 < \dots < t_n$ Messungen irgendeiner physikalischen Größe durchgeführt wurden. Möchte man nun nachträglich vielleicht zu Auswertungszwecken einen Messwert zwischen zwei Messzeitpunkten generieren, muss man aus den vorhandenen Daten eine Funktion f erstellen, die zu den Messzeitpunkten die gemessenen Werte annimmt. Wertet man dann f zu einem Zwischenzeitpunkt aus, hofft man so einen brauchbaren Ersatzwert für eine reale Messung, die ja dort nicht stattgefunden hat, zu erhalten. Man spricht von *Interpolation* von Daten. Es liegt nahe, f als Polynom anzusetzen. Es stellt sich die Frage, ob die Interpolationsaufgabe überhaupt immer eindeutig lösbar ist.

Satz 3.101: Gegeben seien die $n + 1$ Punkte $(x_k, y_k), 0 \leq k \leq n$ mit *paarweise verschiedenen* x_k . Dann existiert genau ein $p_n \in P_n$ mit $y_k = p_n(x_k) \forall 0 \leq k \leq n$. Dies ist das sogenannte *Interpolationspolynom*.

BEWEIS Die Eindeutigkeit folgt sofort aus Bemerkung 3.100. Die Existenz zeigen wir durch Induktion über n . Für $n = 0$ wähle man $p_n(x) = y_0$. Sei nun die Behauptung gezeigt für $n - 1$. Das Polynom p_{n-1} interpoliere $(x_0, y_0), \dots, (x_{n-1}, y_{n-1})$. Setze $p_n(x) = p_{n-1}(x) + q(x)$ mit

$$q(x) = \frac{(x - x_0) \cdots (x - x_{n-1})}{(x_n - x_0) \cdots (x_n - x_{n-1})} (y_n - p_{n-1}(x_n))$$

Man hat $q \in P_n$, nach Folgerung 3.98 also auch $p_n \in P_n$ und weiter $q(x_k) = 0$ für $k \leq n - 1$, weil dann immer ein Linearfaktor im Zähler den Wert 0 annimmt. Es gilt also $p_n(x_k) = y_k$ für $k \leq n - 1$. Weiterhin gilt $q(x_n) = y_n - p_{n-1}(x_n)$, also $p_n(x_n) = y_n$. ■

Beispiel 3.102: Wir betrachten die drei Punkte $(-2, 1), (-1, -1)$ und $(1, 1)$. Nach Satz 3.101 legen diese Punkte eine interpolierende Parabel p_2 eindeutig fest. Diese kann man mit der Definition von p_n aus dem Beweis des Satzes 3.101 bestimmen. Für eine Handrechnung und wenige zu interpolierende Punkte erweist sich folgender Ansatz als ebenfalls geeignet. Die allgemeine Form des Polynoms ist $p_2(x) = ax^2 + bx + c$. Einsetzen der drei Punkte ergibt die Gleichungen

$$\begin{aligned} (1/1) : \quad & 1 = a + b + c \\ (-1/-1) : \quad & -1 = a - b + c \\ (-2/1) : \quad & 1 = 4a - 2b + c \end{aligned}$$

und führt damit zum Gleichungssystem

$$\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 4 & -2 & 1 & 1 \end{array}$$

Man errechnet als Lösung $a = 1, b = 1, c = -1$ und erhält

$$p_2(x) = x^2 + x - 1.$$

Bemerkung 3.103: In vielen Fällen werden Polynome vom Grad n auf Teilmengen $M \subset \mathbb{K}$ eingeschränkt. Man bezeichnet solche Polynomräume mit $P_n(M)$. Enthält M mehr als $n + 1$ Elemente, dann gelten alle Aussagen dieses Kapitels unverändert, weil dann nicht alle Elemente von M Nullstellen sein können und man deswegen mit Satz 3.91 schließen kann wie oben geschehen. Insbesondere gelten alle Aussagen dieses Kapitels für sämtliche offenen Mengen (sie enthalten alle unendlich viele Elemente), alle echten reellen Intervalle $[a, b]$ mit $a < b$ und alle Kreisscheiben $|z - z_0| \leq r \subset \mathbb{C}$ mit $r > 0$.

Wir zeigen abschließend ein Verfahren, mit dem man die lineare Abhängigkeit von Polynomen einfach bestimmen kann. Seien $p_1(x), \dots, p_n(x) \in P_m$. Aus

$$\sum_{i=1}^n \lambda_i p_i(x) = 0 \quad (3.3)$$

muss für lineare Unabhängigkeit $\lambda_1 = \lambda_2 = \dots = 0$ folgen. Jedes Polynom $p_i(x)$ lässt sich schreiben als

$$p_i(x) = \sum_{k=0}^m a_{ki} x^k.$$

Einsetzen in Gleichung (3.3) und Umordnen der Summanden ergibt

$$0 = \sum_{i=1}^n \lambda_i \left(\sum_{k=0}^m a_{ki} x^k \right) = \sum_{k=0}^m \left(\sum_{i=1}^n \lambda_i a_{ki} \right) x^k.$$

Aus der linearen Unabhängigkeit der Monome folgt

$$\sum_{i=1}^n a_{ki} \lambda_i = 0 \quad \forall 0 \leq k \leq m$$

Dies ist ein lineares Gleichungssystem in $(\lambda_1, \dots, \lambda_n)$, dessen Spalten durch die Koeffizientenvektoren der einzelnen Polynome gebildet werden. Gibt es andere Lösungen als $\lambda_1 = \dots = \lambda_n = 0$, sind die Polynome linear abhängig, sonst linear unabhängig. Man überprüft also die lineare Unabhängigkeit von Polynomen anhand der linearen Unabhängigkeit ihrer Koeffizientenvektoren.

Beispiel 3.104: Die Polynome

$$p_1(x) = (1 - x)^2; \quad p_2(x) = (1 - x)x; \quad p_3(x) = x^2$$

sollen auf lineare Unabhängigkeit geprüft werden. Wir berechnen zunächst die Koeffizientenvektoren.

$$p_1(x) = (1 - x)^2 = 1 - 2x + 1x^2 \Rightarrow a_{01} = 1, a_{11} = -2, a_{21} = 1$$

$$p_2(x) = (1 - x)x = 0 + 1x - x^2 \Rightarrow a_{02} = 0, a_{12} = 1, a_{22} = -1$$

$$p_3(x) = x^2 \Rightarrow a_{03} = 0, a_{13} = 0, a_{23} = 1$$

Das zu lösende Gleichungssystem lautet somit:

$$\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \quad | + 2 \cdot (\text{I}) \\ 1 & -1 & 1 & 0 \quad | - (\text{I}) \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \quad | + (\text{II}) \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}$$

Daher ist allein $\lambda_1 = \lambda_2 = \lambda_3 = 0$ Lösung, und die Polynome sind linear unabhängig.

3.6 Skalarprodukt, euklidische und unitäre Räume

In Kapitel 2 hatte sich erwiesen, dass sich wesentliche geometrische Zusammenhänge in \mathbb{R}^n wie Lagebeziehungen von Geraden und Ebenen, Abstandsberechnungen, Winkelbestimmung u. s. w. unter Rückgriff auf Norm und Skalarprodukt untersuchen lassen. In Kapitel 3.3 wurde der Vektorraum \mathbb{R}^n zum allgemeinen K -Vektorraum erweitert, so dass es naheliegt, die geometrischen Betrachtungen in \mathbb{R}^n auf allgemeine Vektorräume auszudehnen. Wir gehen daher von $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ aus und erweitern zunächst die Definition 2.4 des Skalarprodukts von \mathbb{R}^n auf allgemeine \mathbb{K} -Vektorräume, indem wir wie schon in ähnlichen Fällen zuvor jede Abbildung $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{K}$ ein Skalarprodukt nennen, die die in Definition 2.4 geforderten Eigenschaften besitzt. Danach werden wir eine Norm, also einen Abstandsbegriff auf allgemeinen \mathbb{K} -Vektorräumen definieren. Dann lässt sich analog zu Kapitel 2 fortfahren.

Definition 3.105: Sei V ein \mathbb{K} -Vektorraum. Eine Abbildung $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{K}$ heißt *Skalarprodukt*, wenn folgende Bedingungen erfüllt sind:

SP1: $\forall a, b \in V$:

$$\langle a, b \rangle = \begin{cases} \langle b, a \rangle & \text{für } \mathbb{K} = \mathbb{R} \\ \overline{\langle b, a \rangle} & \text{für } \mathbb{K} = \mathbb{C}. \end{cases}$$

SP2 $\forall a, b, c \in V$:

$$\begin{aligned} \langle a, (b + c) \rangle &= \langle a, b \rangle + \langle a, c \rangle \\ \langle (a + b), c \rangle &= \langle a, c \rangle + \langle b, c \rangle \end{aligned}$$

SP3: $\forall \alpha \in \mathbb{K}$ gilt

$$\langle \alpha a, b \rangle = \alpha \langle a, b \rangle = \begin{cases} \langle a, \alpha b \rangle & \text{für } \mathbb{K} = \mathbb{R} \\ \langle a, \bar{\alpha} b \rangle & \text{für } \mathbb{K} = \mathbb{C}. \end{cases}$$

SP4: (positive Definitheit) $\forall a \in V \setminus \{0\} : \langle a, a \rangle > 0$, und $\langle 0, 0 \rangle = 0$.

Bemerkung 3.106: Um die positive Definitheit **SP4** formulieren zu können, muss $\langle a, a \rangle$ immer eine reelle Zahl sein, auch wenn a komplexe Anteile enthält, weil nur \mathbb{R} die Ordnungsrelation „ $>$ “ besitzt. Man erreicht dies für $\mathbb{K} = \mathbb{C}$, indem man in die Symmetriebedingung an das Skalarprodukt wie in **SP1** modifiziert. Dann gilt nämlich $\forall a \in V : \langle a, a \rangle = \overline{\langle a, a \rangle}$, und deswegen $\text{Im}(\langle a, a \rangle) = 0$.

Beispiel 3.107: Ein Skalarprodukt auf \mathbb{R}^n nach Definition 2.4 (Kapitel 2.1) ist auch Skalarprodukt im Sinn von Definition 3.105, und deswegen sind die beiden Definitionen 2.4 und 3.105 miteinander konsistent.

Wir werden zunächst in Analogie zum euklidischen Skalarprodukt das Standardskalarprodukt auf \mathbb{C}^n definieren.

Definition 3.108: Auf \mathbb{C}^n sei für $a = (a_i)_{i=1}^n, b = (b_i)_{i=1}^n$ das Standardskalarprodukt definiert durch

$$\langle a, b \rangle := \sum_{i=1}^n a_i \bar{b}_i. \quad (3.4)$$

Beispiel 3.109:

1. Seien $a = (1, -1)^T, b = (1 + i, 2 - i)^T \in \mathbb{C}^2$. Dann gilt

$$\langle a, b \rangle = 1 \cdot \overline{1 + i} + (-1) \cdot \overline{2 - i} = (1 - i) - (2 + i) = -1 - 2i \in \mathbb{C}$$

und weiter $\langle a, a \rangle = 2$ sowie $\langle b, b \rangle = (1 - i)(1 + i) + (2 - i)(2 + i) = 2 + 5 = 7 \in \mathbb{R}$.

2. Für Vektoren mit reellen Komponenten entspricht das Standardskalarprodukt auf \mathbb{C}^n genau dem euklidischen Skalarprodukt auf \mathbb{R}^n .

Wir definieren nun ein abstraktes Skalarprodukt auf dem Vektorraum der stetigen Funktionen, der sich ja wesentlich von \mathbb{K}^n unterscheidet.

Beispiel 3.110: Seien $f, g \in C[a, b]$. Auf $C[a, b]$ wird durch

$$\langle f, g \rangle := \int_a^b f(x) \cdot g(x) dx \quad (3.5)$$

ein Skalarprodukt definiert.

BEWEIS Auf einem beschränkten und abgeschlossenen Intervall nehmen stetige Funktionen nach einem Satz von Weierstraß ihr Minimum und Maximum an. Es existiert also $\alpha = \max_{x \in [a, b]} (f \cdot g)(x) \in \mathbb{R}$ und $\beta = \min_{x \in [a, b]} (f \cdot g)(x) \in \mathbb{R}$, weil $f \cdot g$ stetig ist. Damit gilt

$$\begin{aligned} \langle f, g \rangle &\leq \int_a^b \alpha dx = \alpha(b - a) < \infty \\ \langle f, g \rangle &\geq \int_a^b \beta dx = \beta(b - a) > -\infty, \end{aligned}$$

und daher ist die Abbildung $\langle \cdot, \cdot \rangle : V \rightarrow \mathbb{R}$ wohldefiniert, weil ja das Integral immer einen endlichen Wert annimmt.

SP1:

$$\begin{aligned} \langle f, g \rangle &= \int_a^b f(x) \cdot g(x) dx \\ &= \int_a^b g(x) \cdot f(x) dx = \langle g, f \rangle \end{aligned}$$

SP2:

$$\begin{aligned} \langle f, g + h \rangle &= \int_a^b f(x) \cdot (g(x) + h(x)) dx \\ &= \int_a^b (f(x) \cdot g(x)) + (f(x) \cdot h(x)) dx \\ &= \int_a^b (f(x) \cdot g(x)) dx + \int_a^b (f(x) \cdot h(x)) dx = \langle f, g \rangle + \langle f, h \rangle \end{aligned}$$

SP3:

$$\begin{aligned} \langle \lambda f, g \rangle &= \int_a^b \lambda f(x) \cdot g(x) dx \\ &= \lambda \cdot \int_a^b f(x) \cdot g(x) dx = \lambda \cdot \langle f, g \rangle \end{aligned}$$

SP4:

$$\langle f, f \rangle = \int_a^b f^2(x) dx$$

Im Fall $f = 0$ gilt $\langle f, f \rangle = \int_a^b 0 dx = 0$. Sei $f \neq 0$. Dann gilt $f^2 \geq 0$, und es existiert $\tilde{x} \in [a, b]$ mit $f(\tilde{x}) \neq 0$, also $f^2(\tilde{x}) > 0$. Wir nehmen zunächst $\tilde{x} \neq a$ und $\tilde{x} \neq b$ an. Sei $f^2(\tilde{x}) = \varepsilon > 0$. Aufgrund der Stetigkeit von f^2 existiert ein $\delta > 0 : f^2(y) > \varepsilon/2 \forall y$ im offenen Intervall $I = (\tilde{x} - \delta, \tilde{x} + \delta)$. Man beachte dazu, dass nach unserer Annahme $\tilde{x} \in (a, b)$ gilt und daher ein $\tilde{\delta} > 0$ existiert mit $(\tilde{x} - \tilde{\delta}, \tilde{x} + \tilde{\delta}) \subset [a, b]$, aus dem man eventuell durch abermaliges Verkleinern I gewinnen kann. Daher gilt

$$\langle f, f \rangle = \int_a^b f^2(x) dx \geq \int_{\tilde{x}-\tilde{\delta}}^{\tilde{x}+\tilde{\delta}} f^2(x) dx \geq \varepsilon/2 \cdot 2\tilde{\delta} > 0.$$

Im Fall $x = a$ gilt mit derselben Schlussweise $f^2(y) > \varepsilon/2$ auf $[a, a + \delta)$ und deswegen

$$\langle f, f \rangle = \int_a^b f^2(x) dx \geq \int_a^{a+\delta} f^2(x) dx \geq \varepsilon/2 \cdot \delta > 0.$$

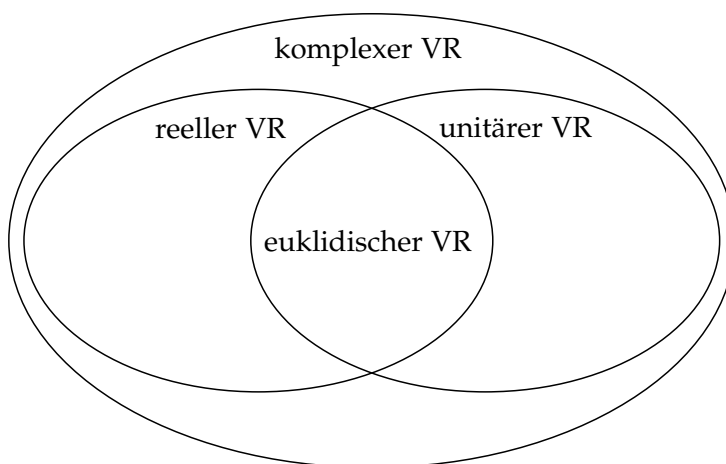
Analog argumentiert man im Fall $x = b$ mit dem Intervall $(b - \delta, b]$. Es folgt damit **SP4**. ■

Beispiel 3.111: Gegeben seien $f(x) = 1$ und $g(x) = x$ auf dem Intervall $[-1, 1]$. Sowohl f als auch g sind stetige Funktionen. Man errechnet

$$\langle 1, x \rangle = \int_{-1}^1 1 \cdot x dx = \frac{1}{2} x^2 \Big|_{-1}^1 = 0$$

Definition 3.112: Ein reeller Vektorraum gemeinsam mit einem Skalarprodukt heie *Euklidischer Vektorraum*, ein komplexer Vektorraum mit einem Skalarprodukt heie *Unitärer Vektorraum*.

Grafisch kann man die Zusammenhänge als Mengendiagramm wie folgt darstellen:



Beispiel 3.113: Die Vektorräume \mathbb{R}^n mit dem euklidischen Skalarprodukt und $\mathcal{C}[a, b]$ mit dem Skalarprodukt aus Beispiel 3.110 sind euklidische Vektorräume, ebenso natürlich jeder Unterraum dieser Vektorräume wie z. B. $P_n([a, b])$.

Zur Definition einer Norm auf einem allgemeinen Vektorraum V gehen wir in der bewährten Weise vor: Wir gehen von der Definition 2.12 einer Norm auf \mathbb{R}^n aus und nennen *jede* Abbildung $\|\cdot\| : V \rightarrow \mathbb{R}$, die die in Definition 2.12 geforderten Eigenschaften aufweist, eine Norm auf V . Die Konsistenz beider Definitionen ist damit nach Konstruktion gesichert.

Definition 3.114: Sei V ein \mathbb{K} -Vektorraum und $a, b \in V$. Eine Abbildung $\|\cdot\| : V \rightarrow \mathbb{R}$ heißt *Norm* genau dann, wenn

$$\mathbf{N0} : \|a\| \in \mathbb{R}.$$

$$\mathbf{N1} : \|a\| \geq 0.$$

$$\mathbf{N2} : \|a\| = 0 \Leftrightarrow a = 0.$$

$$\mathbf{N3} : \forall \lambda \in \mathbb{K} : \|\lambda a\| = |\lambda| \|a\|$$

$$\mathbf{N4} : (\text{Dreiecksungleichung}) \|a + b\| \leq \|a\| + \|b\|$$

Wie im Spezialfall $V = \mathbb{R}^n$ (vgl. Kapitel 2.1) induziert ein Skalarprodukt eine Norm.

Satz 3.115: In einem unitären (bzw. euklidischen) Raum induziert das Skalarprodukt eine (Standard-)Norm durch

$$\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}.$$

BEWEIS **N0** ist für euklidische Vektorräume klar und folgt für unitäre Vektorräume aus Bemerkung 3.106. Die Eigenschaften **N1** – **N4** wurden bereits in Satz 2.13 für \mathbb{R}^n gezeigt. Für **N1** – **N2** haben wir dazu nur die Gleichung $\|x\| = \sqrt{\langle x, x \rangle}$ verwendet und keine der speziellen Eigenschaften von \mathbb{R}^n benötigt. Daher sind mit dem dortigen Beweis **N1** – **N2** schon gezeigt. Es gilt für $a \in V$ und $\lambda \in \mathbb{C}$ nach **SP3** $\|\lambda a\| = \sqrt{\lambda \bar{\lambda}} \|a\|$. Sei jetzt $\lambda = x + iy$. Nach der dritten binomischen Formel folgt $\lambda \bar{\lambda} = x^2 + y^2 = |\lambda|^2$ und damit **N3**. Den Beweis von **N4** stellen wir kurz zurück. ■

In allen unitären Vektorräumen gilt die Cauchy-Schwarzsche Ungleichung, deren Beweis wir zunächst zurückstellen.

Satz 3.116: In allen unitären Vektorräumen V gilt die *Cauchy-Schwarzsche Ungleichung*

$$|\langle a, b \rangle| \leq \|a\| \|b\| \quad \forall a, b \in V. \quad (3.6)$$

Der Beweis der Dreiecksungleichung N4 auf \mathbb{R}^n in Satz 2.32 beruht auf der Cauchy-Schwarzschen Ungleichung und der Symmetrie des Skalarproduktes, die nur in euklidischen Räumen gilt. Daher lässt sich dieser Beweis unmittelbar nur auf euklidische Vektorräume übertragen. Wir zeigen also N4 für unitäre Vektorräume (und damit auch für euklidische Vektorräume) unter Verwendung der (hier noch unbewiesenen) Cauchy-Schwarzschen Ungleichung.

BEWEIS Beide Seiten der Dreiecksungleichung sind reell und insbesondere nicht negativ. Daher genügt es, zu beweisen, dass ihre Quadrate die gewünschte Ungleichung erfüllen d. h.

$$\langle a + b, a + b \rangle \leq (\|a\| + \|b\|)^2.$$

Man hat

$$\langle a + b, a + b \rangle = \langle a, a \rangle + \langle a, b \rangle + \langle b, a \rangle + \langle b, b \rangle$$

Es gilt $\langle b, a \rangle = \overline{\langle a, b \rangle}$ und deswegen $\langle a, b \rangle + \langle b, a \rangle = 2\operatorname{Re}\langle a, b \rangle \leq 2|\langle a, b \rangle|$, weil der Absolutbetrag einer komplexen Zahl mindestens so groß ist wie ihr Realteil. Aufgrund der Cauchy-Schwarzschen Ungleichung (3.6) gilt dann $\langle a, b \rangle + \langle b, a \rangle \leq 2\|a\| \|b\|$. Mit der ersten binomischen Formel folgt die Behauptung. ■

Beispiel 3.117:

1. Für $V = \mathbb{C}$ und $z = a + ib \in \mathbb{C}$ induziert das Standardskalarprodukt (3.4) die Norm

$$\|z\| = \sqrt{z \cdot \bar{z}} = \sqrt{(a + ib)(a - ib)} = \sqrt{a^2 + b^2} = |z|,$$

also den gewöhnlichen Betrag komplexer Zahlen.

2. Für \mathbb{R}^n erhält man aus dem euklidischen Skalarprodukt die euklidische Norm (vgl. Kapitel 2.1).
3. Auf $V = \mathcal{C}[a, b]$ und $f \in V$ induziert das Skalarprodukt aus Beispiel 3.111 die sog. L_2 -Norm

$$\|f\|_{L_2} := \left(\int_a^b f^2(x) dx \right)^{1/2}. \quad (3.7)$$

Sei z. B. $[a, b] = [-1, 1]$ und $f(x) = x$. Dann ist

$$\begin{aligned} \|f\|_{L_2}^2 &= \|x\|_{L_2}^2 = \int_{-1}^1 x \cdot x dx \\ &= \frac{1}{3} x^3 \Big|_{-1}^1 = \frac{2}{3}. \end{aligned}$$

4. Nicht jede Norm auf einem Vektorraum wird durch ein Skalarprodukt induziert (vgl. Bemerkung 2.15). So wird die auf $C[a, b]$ in der Analysis sehr gebräuchliche Maximumsnorm

$$\|f\|_\infty := \max_{x \in [a, b]} |f(x)|$$

von keinem Skalarprodukt induziert.

Bemerkung 3.118: Weil die Cauchy-Schwarzsche Ungleichung gilt, ist es auf euklidischen Vektorräumen problemlos möglich, einen Winkelbegriff wie in Kapitel 2.1.3 zwischen Vektoren einzuführen. In der Praxis haben aber Winkel in beliebigen euklidischen Räumen eine sehr geringe Bedeutung. Winkel in unitären Räumen sind dagegen nicht ohne Weiteres wie in Kapitel 2.1.3 zu definieren, weil das Skalarprodukt in Gleichung (2.10), die zur Winkeldefinition dienen müsste, komplexe Werte annehmen kann.

3.7 Orthogonalität in unitären Vektorräumen

Für die analytische Geometrie in \mathbb{R}^n (vgl. Kapitel 2) hat sich die *orthogonale Projektion* eines Vektors auf einen anderen als fundamental herausgestellt. Man darf also erwarten, dass auch in unitären Vektorräumen, die ja \mathbb{R}^n verallgemeinern, eine orthogonale Projektion eine wesentliche Rolle spielen wird. Wir definieren zunächst Orthogonalität in Analogie zur Definition 2.22 in \mathbb{R}^n auf allgemeinen unitären Vektorräumen. Der euklidische Vektorraum ist ein Spezialfall dessen, so dass wir nur unitäre Vektorräume betrachten. Sei im ganzen Kapitel von nun an V ein unitärer Vektorraum.

Definition 3.119: Seien $a, b \in V$. Es stehen a und b orthogonal zueinander, falls

$$\langle a, b \rangle = 0$$

gilt. Man schreibt $a \perp b$.

Bemerkung 3.120: Mit wortgleichem Beweis wie in Satz 2.25 gilt der Satz des Pythagoras $\|a + b\|^2 = \|a\|^2 + \|b\|^2$ für $a \perp b$ in allen unitären Vektorräumen.

Bei der Herleitung der orthogonalen Projektion (2.6) in \mathbb{R}^n wurden nur die Eigenschaften **SP2** und **SP3** des Skalarprodukts benutzt, wobei skalare Größen nur aus dem ersten Argument gezogen wurden. Weil also nur die Eigenschaften genutzt wurden, die im reellen wie komplexen Fall gleichermaßen gelten, erhält man sofort

Satz 3.121: Für die orthogonale Projektion $p_b(a)$ eines Vektors a auf b , $b \neq 0$, gilt in jedem unitären Vektorraum

$$p_b(a) = \frac{\langle a, b \rangle}{\langle b, b \rangle} b \tag{3.8}$$

Wir holen nun den Beweis der Cauchy-Schwarzschen Ungleichung nach und schließen damit die Lücke in unserer Theorie.

BEWEIS Im Beweis der Cauchy-Schwarzschen Ungleichung auf \mathbb{R}^n (vgl. Satz 2.31) werden der Satz des Pythagoras und die Formel für die orthogonale Projektion verwendet. Beides gilt unverändert in unitären Vektorräumen. Daher gilt der Beweis von Satz 2.31 wortgleich für den allgemeinen Fall.

Wir kehren zur orthogonalen Projektion zurück. Man erkennt an Gleichung (3.8), dass sich der Wert von $p_b(a)$ nicht ändert, wenn man b durch ein skalares Vielfaches (bis auf den Nullvektor natürlich) ersetzt. Alle Vektoren des von b aufgespannten Untervektorraums ohne den Nullvektor führen also zur gleichen Projektion, so dass es nahe liegt, $p_b(a)$ als eine orthogonale Projektion von a auf den Untervektorraum $U = L(b)$ anzusehen. Insbesondere gilt $a - p_b(a) \perp u \forall u \in U$. Wir untersuchen jetzt eine solche orthogonale Projektion auf beliebige Untervektorräume.

Definition 3.122: Sei U ein endlich erzeugter Untervektorraum von V und $a \in V$. Ein Vektor $p_U(a) \in U$ heißt *orthogonale Projektion* von a auf U , wenn

$$a - p_U(a) \perp u \quad \forall u \in U \quad (3.9)$$

gilt.

Es stellt sich die Frage nach der Wohldefiniertheit, d. h. ob es immer einen derartigen Vektor $p_U(a)$ gibt und ob er eindeutig ist. Zur Diskussion der Eindeutigkeit hilft folgender Begriff.

Definition 3.123: Für $M \subseteq V$ heißt

$$M^\perp = \{v \in V \mid v \perp u \forall u \in M\}$$

das *orthogonale Komplement* von M .

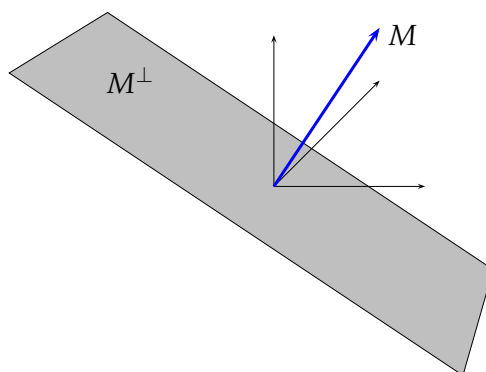
Bemerkung 3.124:

1. M^\perp ist ein Untervektorraum von V .
2. Sei U ein Untervektorraum von V . Dann gilt $U \cap U^\perp = \{0\}$.

BEWEIS 1.) Es ist die Abgeschlossenheit zu prüfen. Für $u \in M$; $x, y \in M^\perp$ und $\lambda \in \mathbb{R}$ gilt:

$$\begin{aligned} \langle x + y, u \rangle &= \langle x, u \rangle + \langle y, u \rangle = 0 \\ \langle \lambda x, u \rangle &= \lambda \langle x, u \rangle = 0 \end{aligned}$$

2.) Sei $a \in U \cap U^\perp$. Dann gilt $\langle a, u \rangle = 0 \forall u \in U$, da $a \in U^\perp$, also insbesondere $\langle a, a \rangle = 0$ wegen $a \in U$ und deswegen $a = 0$. ■

Abbildung 3.5: orthogonales Komplement von $M = \{v\}$

Beispiel 3.125: Sei $V = \mathbb{R}^3$, $v = (2, 0, 0)^T$ und $M = \{v\}$, eine Menge mit nur einem Vektor. Dann gilt $M^\perp = \{x \in \mathbb{R}^3 \mid \langle x, v \rangle = 0\}$, also für die konkrete Wahl von v genau die (x_2, x_3) -Ebene. Für beliebige Vektoren $v \neq 0$ entspricht M^\perp der Ebene mit dem Normalenvektor v durch den Nullpunkt (vgl. Kapitel 2.2 und Abb. 3.5).

Folgerung 3.126: Seien U, V und a wie in Definition 3.122. Die orthogonale Projektion von a auf U ist eindeutig.

BEWEIS Sei $q \in U$ mit $a - q \perp u \forall u \in U$. Dann gilt $q - p_U(a) \in U$, aber wegen

$$\begin{aligned} 0 &= \langle a - q, u \rangle = \langle a - p_U(a), u \rangle \quad \forall u \in U \\ &\Leftrightarrow \langle a, u \rangle - \langle q, u \rangle = \langle a, u \rangle - \langle p_U(a), u \rangle \quad \forall u \in U \\ &\Leftrightarrow \langle q - p_U(a), u \rangle = 0 \quad \forall u \in U \end{aligned}$$

auch $q - p_U(a) \in U^\perp$, also nach Bemerkung 3.124 $q = p_U(a)$. ■

Lemma 3.127: Sei U wie zuvor und (u_1, \dots, u_m) eine Basis von U . Für $v \in V$ gilt:

$$v \in U^\perp \Leftrightarrow \langle v, u_i \rangle = 0 \quad \forall 1 \leq i \leq m$$

BEWEIS „ \Rightarrow “ ist klar. „ \Leftarrow “: Sei $u \in U$ beliebig. Dann gilt $u = \sum_{i=1}^m \lambda_i u_i$ mit $\lambda_i \in \mathbb{K}$ und damit

$$\langle v, u \rangle = \left\langle v, \sum_{i=1}^m \lambda_i u_i \right\rangle = \sum_{i=1}^m \overline{\lambda_i} \langle v, u_i \rangle = 0. \quad \blacksquare$$

Sei (u_1, \dots, u_m) eine Basis von U . Existiert $p_U(a) \in U$, dann folgt

$$p_U(a) = \sum_{i=1}^m \alpha_i u_i \quad (3.10)$$

mit noch zu bestimmenden Koeffizienten α_i und aus Gleichung (3.9) zusammen mit Lemma 3.127:

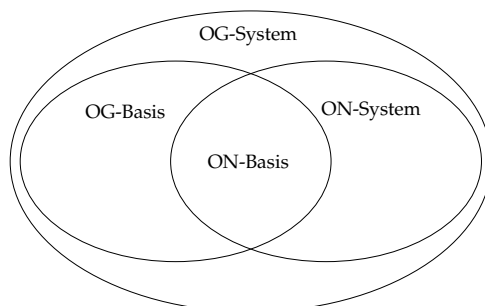
$$\begin{aligned} \left\langle a - \sum_{i=1}^m \alpha_i u_i, u_j \right\rangle &= 0 \quad \forall 1 \leq j \leq m \\ \Leftrightarrow \sum_{i=1}^m \langle u_i, u_j \rangle \alpha_i &= \langle a, u_j \rangle \quad \forall 1 \leq j \leq m \\ \Leftrightarrow G\alpha &= b \end{aligned} \quad (3.11)$$

mit dem Koeffizientenvektor $\alpha = (\alpha_1, \dots, \alpha_m)^T$ und der Matrix $G = (\langle u_i, u_j \rangle)_{i,j=1}^m$. Die Matrix G wird Gram¹ zu Ehren *Gram-Matrix* genannt. Man wird also auf ein lineares Gleichungssystem in den Koeffizienten geführt. Die orthogonale Projektion von a auf U existiert daher genau dann, wenn dieses Gleichungssystem eine Lösung besitzt. Wir stellen die Frage der Lösbarkeit zunächst zurück.

Wie gesehen erfordert die Berechnung der orthogonalen Projektion die eventuell aufwändige Lösung eines linearen Gleichungssystems. Andererseits hängt die Matrix G dieses Gleichungssystems unmittelbar von der Wahl der Basis von U ab. Am besten geeignet erscheint eine Basis von U , für die $\langle u_i, u_j \rangle = 0$ für $i \neq j$ und $\langle u_i, u_i \rangle = 1$ für $i = j$ gilt, also eine mit paarweise orthogonalen Basisvektoren der Länge eins. Dann nämlich gälte $A = E$, und das Lösen des linearen Gleichungssystems (3.11) entfiel.

Definition 3.128: Sei $\mathcal{B} = (v_1, \dots, v_m)$ ein m -Tupel mit Vektoren in $V \setminus \{0\}$.

1. \mathcal{B} heißt *Orthogonalsystem* in V , falls alle v_i paarweise orthogonal sind.
2. Ein Orthogonalsystem, für das zusätzlich $\|v_i\| = 1 \forall i = 1, \dots, m$ gilt, heißt *Orthonormalsystem*.
3. Ein Orthogonalsystem, das eine Basis von V bildet, heißt *Orthogonalbasis* von V .
4. Ein Orthonormalsystem, das eine Basis von V bildet, heißt *Orthonormalbasis* von V .



¹Jørgen Pederson Gram (1850–1916): dänischer Mathematiker; Beiträge zur Mathematik und Forstwirtschaft; Vorstand einer Versicherungsgesellschaft

Bemerkung 3.129: Mit dem *Kronecker-Symbol*²

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

gilt in jedem Orthonormalsystem $\langle v_i, v_j \rangle = \delta_{i,j}$.

Beispiel 3.130: Die Vektoren

$$a_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad a_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{und} \quad a_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

sind zwar eine Basis, aber nicht orthogonal, da $\langle a_1, a_3 \rangle = 1 \neq 0$.

Beispiel 3.131: Die Vektoren

$$a_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad a_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{und} \quad a_3 = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}$$

sind eine Basis und orthogonal, aber a_3 ist nicht normiert. Also bilden die drei Vektoren eine Orthogonalbasis.

Beispiel 3.132: Die Vektoren

$$a_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{und} \quad a_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

sind orthogonal und normiert, bilden aber keine Basis. Es handelt sich also um ein Orthonormalsystem.

Beispiel 3.133: Die Vektoren

$$a_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \quad a_2 = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} \quad \text{und} \quad a_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

sind eine Basis, orthogonal und normiert. Sie bilden somit eine Orthonormalbasis.

Satz 3.134: Ein Orthogonalsystem (v_1, \dots, v_n) ist linear unabhängig.

²Leopold Kronecker (1823–1891), dt. Mathematiker; Beiträge vornehmlich zur Algebra und Zahlentheorie

BEWEIS Sei $\sum_{i=1}^n \lambda_i v_i = 0$. Für v_j folgt

$$\left\langle \sum_{i=1}^n \lambda_i v_i, v_j \right\rangle = 0 \quad \Leftrightarrow \quad \sum_{i=1}^n \lambda_i \langle v_i, v_j \rangle = 0$$

und daraus aufgrund der Orthogonalität $\lambda_j \langle v_j, v_j \rangle = 0$. Wegen $v_j \neq 0$ in jedem Orthogonalsystem gilt auch $\langle v_j, v_j \rangle = \|v_j\|^2 > 0$ und daher $\lambda_j = 0$ für jedes $1 \leq j \leq n$. ■

Satz 3.135: Ist $\mathcal{B} = (v_1, \dots, v_n)$ eine Orthogonalbasis von V , dann gilt für jedes $v \in V$:

$$v = \sum_{k=1}^n \frac{\langle v, v_k \rangle}{\langle v_k, v_k \rangle} v_k,$$

d. h. v hat bzgl. \mathcal{B} die Koordinaten $(\frac{1}{\|v_k\|^2} \langle v, v_k \rangle, 1 \leq k \leq n)^T$.

BEWEIS Da \mathcal{B} eine Basis ist, existiert immer eine Darstellung $v = \sum_{k=1}^n \lambda_k v_k$. Bildet man auf beiden Seiten das Skalarprodukt mit v_ℓ , dann erhält man wegen $\langle v_k, v_\ell \rangle = \delta_{k,\ell} \|v_k\|^2$:

$$\langle v, v_\ell \rangle = \lambda_\ell \langle v_\ell, v_\ell \rangle = \lambda_\ell \|v_\ell\|^2 \quad \forall 1 \leq \ell \leq n. \quad \blacksquare$$

Bemerkung 3.136: Die Koordinaten eines Vektors bezogen auf eine Orthogonalbasis entsprechen genau den Vorfaktoren der Projektionen des Vektors auf den k -ten Basisvektor.

Ist \mathcal{B} wie in Satz 3.135 sogar eine Orthonormalbasis, besitzt v die Koordinaten $(\langle v, v_k \rangle, 1 \leq k \leq n)^T$ bezogen auf \mathcal{B} .

Beispiel 3.137: Die Vektoren

$$a_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \quad a_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} \quad \text{und} \quad a_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

bilden eine Orthonormalbasis von \mathbb{R}^3 . Der Vektor $v = (5, 3, 7)^T$ lässt sich also als $\lambda_1 a_1 + \lambda_2 a_2 + \lambda_3 a_3$ schreiben. Man erhält λ_1, λ_2 und λ_3 aus:

$$\begin{aligned} \lambda_1 &= \langle v, a_1 \rangle = \frac{5}{\sqrt{2}} + \frac{3}{\sqrt{2}} = 4 \cdot \sqrt{2} \\ \lambda_2 &= \langle v, a_2 \rangle = \frac{5}{\sqrt{2}} - \frac{3}{\sqrt{2}} = \sqrt{2} \\ \lambda_3 &= \langle v, a_3 \rangle = 7 \end{aligned}$$

Also ist $v = 4\sqrt{2} \cdot a_1 + \sqrt{2} \cdot a_2 + 7 \cdot a_3$.

Satz 3.138: Sei $\mathcal{B} = (v_1, \dots, v_m)$ ein Orthogonalsystem in V und $U = L(\mathcal{B})$ der von \mathcal{B} aufgespannte Unterraum.

1. Für jedes $v \in V$ gilt

$$p_U(v) = \sum_{i=1}^m \frac{\langle v, v_i \rangle}{\langle v_i, v_i \rangle} v_i.$$

2. Jedes $v \in V$ lässt sich eindeutig als Summe $v = p_U(v) + w$ mit $w \in U^\perp$ schreiben. Dabei gilt $w = v - p_U(v)$.

3. $V = U \oplus U^\perp$.

4. Sei $\dim(V) = n$. Dann gilt $\dim(U) + \dim(U^\perp) = n$ für jeden Untervektorraum U .

BEWEIS 1.) Die Koordinaten $(\alpha_1, \dots, \alpha_m)^T$ von $p_U(v)$ sind Lösung des Gleichungssystems (3.11). Im Fall eines Orthogonalsystems erhält man

$$\begin{pmatrix} \langle v_1, v_1 \rangle & 0 & \cdots & 0 \\ 0 & \langle v_2, v_2 \rangle & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \langle v_m, v_m \rangle \end{pmatrix} \cdot \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix} = \begin{pmatrix} \langle v, v_1 \rangle \\ \langle v, v_2 \rangle \\ \vdots \\ \langle v, v_m \rangle \end{pmatrix}$$

Es folgt $(\alpha_1, \dots, \alpha_m)^T = (\frac{1}{\|v_1\|^2} \langle v, v_1 \rangle, \dots, \frac{1}{\|v_m\|^2} \langle v, v_m \rangle)^T$ (man beachte dabei $\|v_i\| > 0$ wegen $v_i \neq 0$ nach Definition eines Orthogonalsystems). Einsetzen in Formel (3.10) liefert die Behauptung.

2.) Es gilt nach Definition 3.122 der orthogonalen Projektion $w \perp u \forall u \in U$, also $w \in U^\perp$. Die Eindeutigkeit der orthogonalen Projektion (Folgerung 3.126) impliziert die Eindeutigkeit der Zerlegung von v .

3.) Folgt aus der Definition 3.34 der direkten Summe und 2.).

4.) erhält man unmittelbar aus Folgerung 3.67. ■

Beispiel 3.139:

$$\mathcal{B} = \left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right)$$

ist sogar ein Orthonormalsystem. Seine lineare Hülle ist

$$U = L(\mathcal{B}) = \left\{ \begin{pmatrix} a \\ b \\ 0 \end{pmatrix}; a, b \in \mathbb{R} \right\}.$$

Das Orthogonale Komplement $U^\perp = \{v \in V; v \perp u \forall u \in U\}$ ist

$$U^\perp = \left\{ \begin{pmatrix} 0 \\ 0 \\ c \end{pmatrix}; c \in \mathbb{R} \right\}.$$

3.8 Das Verfahren von Gram-Schmidt und Anwendungen

Im letzten Kapitel wurde deutlich, dass eine orthogonale Projektion eines abstrakten Vektors immer und relativ einfach möglich ist, wenn von dem Unterraum, auf den projiziert werden soll, eine Orthogonal-, besser noch eine Orthonormalbasis bekannt ist. Es ist aber nicht sofort klar, ob jeder endlich erzeugte unitäre Vektorraum eine Orthonormalbasis besitzt und wie man, wenn ja, eine solche konstruieren kann. Beide Fragen werden mithilfe des *Orthonormalisierungsverfahrens nach Gram-Schmidt* positiv beantwortet. Das Verfahren wurde unabhängig voneinander von Schmidt³ und Gram (1879) veröffentlicht. Beide gelten als „Entdecker“ des Verfahrens, allerdings wurde das Verfahren schon 1836 von Cauchy benutzt.

Gegeben seien m linear unabhängige Vektoren v_1, \dots, v_m eines unitären Vektorraums. Das Verfahren von Gram-Schmidt erzeugt aus diesen Vektoren ein Orthonormalsystem w_1, \dots, w_m mit

$$L(v_1, \dots, v_m) = L(w_1, \dots, w_m),$$

also eine Orthonormalbasis des Untervektorraums $L(v_1, \dots, v_m)$. Zur Orthogonalisierung wird die orthogonale Projektion von Basisvektoren verwendet. Wir betrachten zur Veranschaulichung eine beliebige Basis (v_1, v_2, v_3) von \mathbb{R}^3 . Wir gehen so vor:

1. Man wählt einfach $w_1 = \frac{1}{\|v_1\|} v_1$, weil dann offenbar $\|w_1\| = 1$ gilt.
2. Wir konstruieren einen Vektor r_2 , der senkrecht auf w_1 steht. Dazu projizieren wir v_2 auf den von w_1 erzeugten Unterraum $L(w_1)$ und setzen $r_2 := v_2 - p_{L(w_1)}(v_2)$. Dann gilt nach Definition der orthogonalen Projektion $v_1 \perp r_2$. Nach Satz 3.138 erhält man

$$r_2 = v_2 - \langle v_2, w_1 \rangle w_1.$$

Aufgrund der linearen Unabhängigkeit von v_1 und v_2 gilt $r_2 \neq 0$ wegen $v_2 \notin L(w_1)$. Normierung von r_2 liefert w_2 . Weil r_2 und damit w_2 eine Linearkombination von v_1 und v_2 ist, gilt $L(w_1, w_2) \subseteq L(v_1, v_2)$. Es sind aber w_1 und w_2 orthonormal und damit nach Satz 3.134 linear unabhängig, also gilt auch $\dim(L(w_1, w_2)) = \dim(L(v_1, v_2))$ und daher

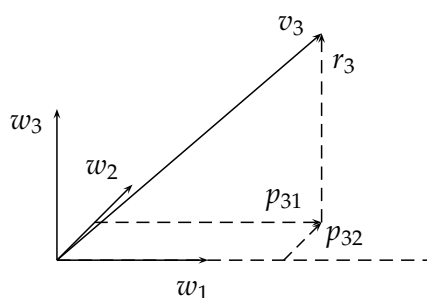
$$L(w_1, w_2) = L(v_1, v_2).$$

³Erhard Schmidt (1876–1959), dt. Mathematiker; bedeutende Beiträge zur Funktionalanalysis, Entwicklung des nach ihm benannte Verfahrens im Kontext nicht-endlich erzeugter unitärer Vektorräume (1907)

3. Der Vektor $r_3 := v_3 - p_{L(v_1, v_2)}(v_3)$ steht nach Konstruktion senkrecht auf $L(v_1, v_2)$ und daher gilt insbesondere $r_3 \perp v_1$ und $r_3 \perp v_2$. Aufgrund der linearen Unabhängigkeit der v_i liegt v_3 nicht in $L(v_1, v_2) = L(w_1, w_2)$, und daher $r_3 \neq 0$. Wir finden w_3 durch Normierung von r_3 . Weil (w_1, w_2) nach 2.) eine Orthonormalbasis von $L(v_1, v_2)$ bilden, gilt nach Satz 3.138:

$$r_3 = v_3 - \langle v_3, w_1 \rangle w_1 - \langle v_3, w_2 \rangle w_2$$

Das folgende Bild veranschaulicht die Konstruktion von w_3 .



Man kann die obige Konstruktion leicht verallgemeinern.

Satz 3.140 (Gram-Schmidtsches Orthogonalisierungsverfahren): Sei V ein unitärer Vektorraum und v_1, \dots, v_m linear unabhängig. Seien

$$w_1 := \frac{v_1}{\|v_1\|}$$

$$r_{k+1} := v_{k+1} - \sum_{i=1}^k \langle v_{k+1}, w_i \rangle w_i$$

$$w_{k+1} = \frac{r_{k+1}}{\|r_{k+1}\|}$$

Dann bilden (w_1, \dots, w_m) eine Orthonormalbasis von $L(v_1, \dots, v_m)$.

BEWEIS Man zeigt obigen Satz per vollständiger Induktion. Der Induktionsanfang entspricht Schritt 2.), der Induktionsschluss verläuft analog zu Schritt 3.) mit etwas allgemeineren Indizes. ■

Beispiel 3.141:

$$v_1 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 7 \\ 1 \\ 1 \end{pmatrix}$$

Berechnung von w_1 :

$$w_1 = \frac{v_1}{\|v_1\|} = \frac{1}{3} \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$$

Berechnung von w_2 :

$$\begin{aligned}
 r_2 &= v_2 - \langle v_2, w_1 \rangle w_1 \\
 &= \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix} - \frac{1}{3}(3 + 8 + 10) \cdot \frac{1}{3} \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \\
 &= \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix} - \frac{7}{3} \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \\
 &= \frac{1}{3} \left[\begin{pmatrix} 9 \\ 12 \\ 15 \end{pmatrix} - \begin{pmatrix} 7 \\ 14 \\ 14 \end{pmatrix} \right] = \frac{1}{3} \begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix} \\
 w_2 &= \frac{r_2}{\|r_2\|} = \frac{1}{3} \begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix}
 \end{aligned}$$

Berechnung von w_3 :

$$\begin{aligned}
 r_3 &= v_3 - \langle v_3, w_1 \rangle w_1 - \langle v_3, w_2 \rangle w_2 \\
 &= \begin{pmatrix} 7 \\ 1 \\ 1 \end{pmatrix} - \frac{1}{3} \cdot 11 \cdot \frac{1}{3} \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} - \frac{1}{3} \cdot 13 \cdot \frac{1}{3} \begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix} \\
 &= \frac{1}{9} \left[\begin{pmatrix} 63 \\ 9 \\ 9 \end{pmatrix} - \begin{pmatrix} 11 \\ 22 \\ 22 \end{pmatrix} - \begin{pmatrix} 26 \\ -26 \\ 13 \end{pmatrix} \right] \\
 &= \frac{1}{9} \begin{pmatrix} 26 \\ 13 \\ -26 \end{pmatrix} = \frac{13}{9} \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix} \\
 w_3 &= \frac{r_3}{\|r_3\|} = \frac{1}{3} \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix}
 \end{aligned}$$

Folgerung 3.142: Jeder endlich erzeugte unitäre Vektorraum besitzt eine Orthonormalbasis.

Bemerkung 3.143: Folgerung 3.142 ist in nicht endlich erzeugten Vektorräumen i. A. falsch.

Wir holen jetzt den Existenzbeweis der orthogonalen Projektion nach.

Folgerung 3.144: Sei V ein unitärer Vektorraum und U ein endlich erzeugter Untervektorraum. Dann existiert für jedes $v \in V$ die orthogonale Projektion $p_U(v)$ von v auf U .

BEWEIS Sei $U = L(v_1, \dots, v_m)$. Mithilfe des Gram-Schmidt-Verfahrens findet man eine Orthonormalbasis (w_1, \dots, w_m) von U . Anwendung von Satz 3.138 ergibt die Behauptung. ■

Folgerung 3.145: Sei V ein endlich erzeugter unitärer Vektorraum und U irgend ein Untervektorraum. Dann gilt $V = U \oplus U^\perp$, und $\dim(V) = \dim(U) + \dim(U^\perp)$.

BEWEIS Die Aussage wurde bereits in Satz 3.138 unter der Voraussetzung gezeigt, dass U durch ein Orthogonalsystem erzeugt wird. Mithilfe des Gram-Schmidt-Verfahrens lässt sich dies für jeden Unterraum U von V konstruieren, so dass wir jetzt auf die dort genannte Voraussetzung verzichten können. ■

Jetzt holen wir den Beweis dafür nach, dass jede Hyperebene in \mathbb{R}^n eine Normaldarstellung besitzt (vgl. Satz 3.81).

Folgerung 3.146: Jede Hyperebene in \mathbb{R}^n besitzt eine Normaldarstellung; der Normalenvektor ist bis auf Skalierung eindeutig.

BEWEIS Jede Hyperebene in \mathbb{R}^n besitzt die Darstellung $H = p + U$ mit $\dim(U) = n - 1$. Es gilt somit $\dim(U^\perp) = 1$, also existiert $v \in \mathbb{R}^n \setminus \{0\}$ mit $U^\perp = L(v)$. Dies ist der gesuchte Normalenvektor. Jeder andere Normalenvektor muss auch in U^\perp liegen und ist daher ein Vielfaches von v . ■

Bemerkung 3.147: Das Ergebnis des Gram-Schmidt-Verfahrens w_1, \dots, w_m hängt von der Reihenfolge der Ausgangsvektoren v_1, \dots, v_m ab. Vertauscht man diese, erhält man i. A. nicht etwa w_1, \dots, w_m in vertauschter Reihenfolge. Die Aussagen von Satz 3.140 bleiben natürlich unverändert gültig.

Satz 3.148: Sei V wie oben und $v_1, \dots, v_m \in V$. Gelingt es, aus diesen mithilfe des Gram-Schmidt-Verfahrens orthonormale Vektoren w_1, \dots, w_m zu erzeugen, dann sind (v_1, \dots, v_m) linear unabhängig.

BEWEIS Andernfalls existiert v_k mit $v_k \in L(v_1, \dots, v_{k-1})$. Dann gilt $v_k = p_{L(v_1, \dots, v_{k-1})}(v_k)$ und daher $r_k = 0$. Damit ist r_k nicht normierbar, und das Gram-Schmidt-Verfahren bricht ab. ■

Es ist also nicht erforderlich, die lineare Unabhängigkeit der Ausgangsvektoren bei der Orthonormalisierung nach Gram-Schmidt zu prüfen, weil das im Laufe der Rechnung ohnehin klar wird.

Beispiel 3.149: Wir betrachten $P_2([-1, 1])$ mit der bekannten Basis $v_1(x) = 1, v_2(x) = x, v_3(x) = x^2$ und wollen daraus eine Orthonormalbasis bezogen auf das Skalarprodukt (3.5) konstruieren. Wir verwenden dazu

das Verfahren von Gram-Schmidt und normieren zunächst v_1 . Man hat $\|1\|_{L_2}^2 = \int_{-1}^1 1^2 dx = 2$, und daher

$$w_1(x) = \frac{1}{\sqrt{2}}v_1(x).$$

Für $w_2(x)$ erhält man:

$$\begin{aligned} r_2(x) &= v_2(x) - \langle v_2(x), w_1(x) \rangle w_1(x) \\ &= x - \int_{-1}^1 x \cdot \frac{1}{\sqrt{2}} dx \cdot \frac{1}{\sqrt{2}} \\ &= x - \frac{1}{2} \left[\frac{1}{2} x^2 \right]_{-1}^1 = x \end{aligned}$$

und weiter $\|x\|_{L_2}^2 = \int_{-1}^1 x \cdot x dx = \frac{2}{3}$, also

$$w_2(x) = \frac{r_2(x)}{\|r_2(x)\|_{L_2}} = \sqrt{\frac{3}{2}}x$$

Mit einer analogen Rechnung erhält man

$$w_3(x) = \frac{\sqrt{10}}{4}(3x^2 - 1).$$

Orthonormale Funktionensysteme wie in Beispiel 3.149 konstruiert spielen in der angewandten Mathematik eine große Rolle.

In der Praxis kommt es häufig vor, dass man eine komplizierte Funktion durch eine einfache Funktion ersetzen möchte, weil diese vielleicht effizienter zu berechnen ist oder in der Praxis besser handhabbar ist (Wie würde man z. B. $\sin(35^\circ)$ ohne Hilfsmittel praktisch berechnen?). Dabei soll natürlich der Fehler durch diese Ersetzung möglichst klein sein. Man spricht hier von *Approximation*. Die mathematische Disziplin der Approximationstheorie beschäftigt sich mit der möglichst geschickten Konstruktion approximierender Funktionen und derartiger Fehlerschranken. Es liegt nahe, als approximierende Funktionen Polynome aufgrund ihrer Einfachheit zu verwenden. Wir wollen prototypisch die Funktion

$$f(x) = \frac{1}{1+x^2}$$

auf dem Intervall $[-5, 5]$ durch Polynome vom Höchstgrad n approximieren. Dazu werden wir drei Ansätze vergleichen:

1. Approximation durch Taylorpolynome mit Grad n (Polynome t_n)
2. Approximation durch Interpolationspolynome mit Grad n (Polynome i_n)

n	$\ f - t_n\ _{L_2}$	$\ f - i_n\ _{L_2}$	$\ f - p_n\ _{L_2}$
1	1,045	1,170	0,9007
2	1,108	1,438	0,9007
3	2,037	0,7712	0,6092
4	4,483	0,8829	0,6092
5	9,463	0,462	0,4103
6	19,02	0,8663	0,4103
10	218,4	1,835	0,1857
15	1074	-	0,0564

Tabelle 3.1: L_2 -Normen der Approximationsfehler an f

3. Approximation durch orthogonale Projektion auf P_n (Polynome p_n)

Wir führen die durchaus umfangreichen Berechnungen in MATLAB aus. Die $n + 1$ Interpolationspunkte werden auf dem Intervall gleichabständig verteilt, für die Taylorpolynome verwenden wir den Entwicklungspunkt $x_0 = -5$.

In Tabelle 3.1 werden die Approximationsfehler $\|f - t_n\|_{L_2}$, $\|f - i_n\|_{L_2}$ und $\|f - p_n\|_{L_2}$ für die drei Ansätze für einige Polynomgrade n gezeigt. Man erkennt, dass allein die orthogonale Projektion von f auf P_n brauchbare polynomielle Näherungen an f hervorbringt; insbesondere scheint der Approximationsfehler kleiner zu werden, je größer n gewählt wird. Es gilt in der Tat $\|f - p_n\|_{L_2} \rightarrow 0, n \rightarrow \infty$, ohne hier genau zu erklären, warum das so ist. In diesem Fall lässt sich der Approximationsfehler kleiner als jede vom Anwender gewählte Toleranz machen, wenn man nur den Polynomgrad hoch genug wählt. Solche Aussagen gelten für die anderen beiden Ansätze offenbar nicht. Der Vergleich der verschiedenen Polynome in den Abbildungen 3.6 bis 3.8 zeigt, dass die Taylor-Polynome zwar eine exzellente Approximation an f in der Nähe des Entwicklungspunktes liefern, aber eben nicht auf dem ganzen Intervall $[-5, 5]$. Die interpolierenden Polynome zeigen starke Oszillationen, so dass sie zwar den Wert von f in den Interpolationspunkten genau treffen, ansonsten aber den Verlauf von f nicht nachbilden. Dieses Oszillationsphänomen tritt in der Praxis eher selten auf und lässt sich durch eine günstigere Wahl von Interpolationspunkten wesentlich reduzieren, aber man wird von einem Verfahren, das nur meistens funktioniert, in der Praxis doch eher Abstand nehmen. Wir werden jetzt die Überlegenheit der orthogonalen Projektion erklären.

Definition 3.150: Sei V ein unitärer Vektorraum und $v \in V$ sowie $M \subset V$ eine beliebige nichtleere Menge. Ein Vektor $v^* \in M$ heißt *Bestapproximation* in M an v , falls

$$\|v^* - v\| = \inf_{x \in M} \|x - v\|.$$

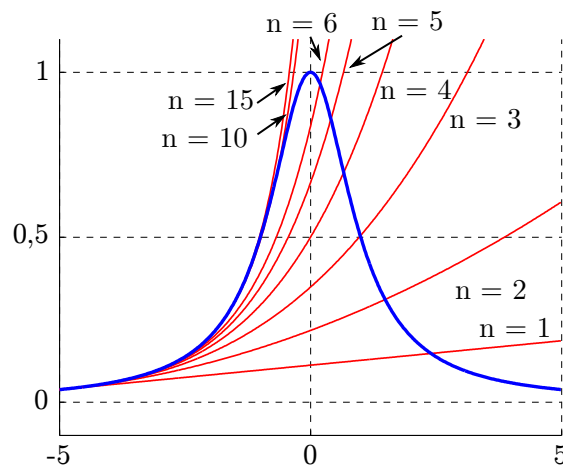


Abbildung 3.6: f (blau dargestellt) und Taylorpolynome t_n

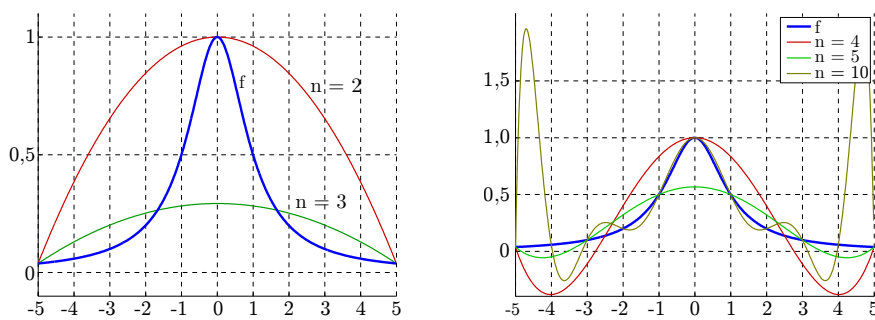


Abbildung 3.7: f (blau dargestellt) und Interpolationspolynome i_n

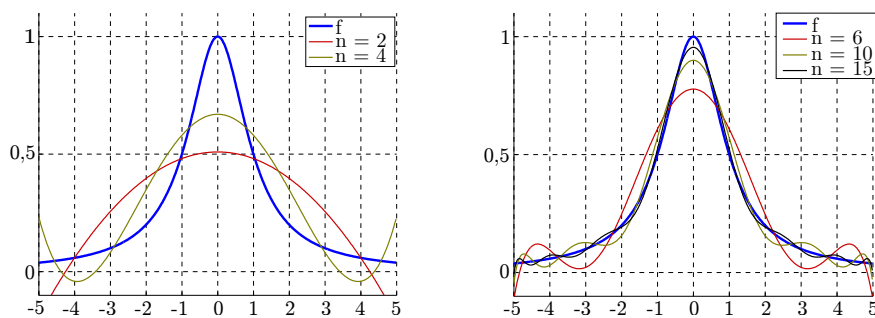


Abbildung 3.8: f (blau dargestellt) und p_n

Für allgemeine Mengen muss es eine Bestapproximation an v weder geben, noch muss sie eindeutig sein. Anders ist das, wenn M ein endlich erzeugter Untervektorraum ist.

Satz 3.151 (Bestapproximation): Sei V wie oben, $v \in V$ und U ein endlich erzeugter Untervektorraum von V . Dann gilt

$$\|v - p_U(v)\| = \min_{u \in U} \|u - v\|,$$

die orthogonale Projektion von v auf U ist also die einzige Bestapproximation an v in U .

BEWEIS Sei $u \in U$ beliebig. Dann ist $u - P_U(v) \in U$ und $v - P_U(v) \in U^\perp$, und daher $u - P_U(v) \perp v - P_U(v)$. Nach dem Satz des Pythagoras gilt dann

$$\|u - v\|^2 = \|u - p_U(v)\|^2 + \|v - p_U(v)\|^2 \geq \|v - p_U(v)\|^2. \quad (3.12)$$

Somit ist $p_U(v)$ eine Bestapproximation. Gleichheit in Formel (3.12) gilt nur, wenn $\|u - P_U(v)\| = 0$, also $u = p_U(v)$ gilt. Daher ist $p_U(v)$ die einzige Bestapproximation in U an v . ■

Nach Satz 3.151 ist der Approximationsfehler gemessen in der L_2 -Norm der p_n an f bei festem Grad nicht mehr zu unterbieten; die durch orthogonale Projektion gewonnenen p_n stellen die bestmögliche Wahl dar.

Folgerung 3.152: Aus $P_n \subset P_{n+1}$ folgt $\|f - p_{P_{n+1}}(f)\|_{L_2} \leq \|f - p_{P_n}(f)\|_{L_2}$ nach Satz 3.151; erhöht man den Polynomgrad, wird also der Approximationsfehler auf keinen Fall größer (aber auch nicht notwendig kleiner, siehe Tabelle 3.1).

Bemerkung 3.153: Man kann Bestapproximation bezüglich beliebiger Normen betrachten, so liegt es hier durchaus nahe, statt der L_2 -Norm die Maximumsnorm auf $\mathcal{C}[a, b]$ zu wählen. Das Fehlermaß ist dann die maximale punktweise Abweichung von p_n zu f . Die Maximumsnorm wird aber nach Bemerkung 3.118 von keinem Skalarprodukt induziert, so dass ein Zugang über Orthogonalität wie oben unmöglich ist. Tatsächlich sind dann sowohl praktische Rechnungen als auch die theoretischen Betrachtungen dazu ungleich komplizierter als bei einer Norm, die von einem Skalarprodukt induziert wird.

Kapitel 4

Lineare Abbildungen

4.1 Vorbereitung

In der Mathematik spielen die Begriffe „Menge“ und „Abbildung“ eine fundamentale Rolle. Wir wiederholen vorbereitend einige wichtige Begriffe zur Charakterisierung von Abbildungen. Intuitiv verbindet man mit einer Abbildung die Vorstellung, dass Elementen einer Menge andere Elemente einer anderen Menge eindeutig zugeordnet werden. Etwas formaler kann man dies wie folgt fassen.

Definition 4.1:

1. Unter einer *Abbildung* $f : X \rightarrow Y$ der *Definitionsmenge* X auf die *Zielmenge* Y versteht man eine Vorschrift, die jedem Element aus X ein eindeutiges Element aus Y zuordnet.
2. Für $x \in X$ heißt $f(x) \in Y$ der *Wert von f in x* oder auch *das Bild von x bezüglich f* .

Statt des Begriffs „Definitionsmenge“ wird auch der Begriff „Definitionsbereich“ verwendet, statt von „Zielmenge“ spricht man bisweilen von „Wertebereich“. Man kann eine Abbildung statt nur auf einzelne Elemente von X auf ganze Teilmengen anwenden.

Definition 4.2: Seien f, X, Y wie in Definition 4.1 und $\tilde{X} \subseteq X$.

1. Das *Bild* $f(\tilde{X})$ von \tilde{X} *bezüglich* f wird erklärt durch

$$f(\tilde{X}) := \{f(x) \mid x \in \tilde{X}\} \subseteq Y.$$

2. Für $\tilde{Y} \subseteq Y$ sei

$$f^{-1}(\tilde{Y}) := \{x \in X \mid f(x) \in \tilde{Y}\}$$

das *Urbild* von \tilde{Y} *bezüglich* f .

3. Im Spezialfall einer einelementigen Menge $\tilde{Y} = \{a\}$ sei

$$f^{-1}(a) := f^{-1}(\{a\})$$

4. Man definiert $\text{Bild}(f) := f(X)$.

Bemerkung 4.3: Der Ausdruck $f^{-1}(a)$ beschreibt eine *Menge*, die auch leer sein oder mehrere Elemente enthalten kann. So gilt für $f : \mathbb{R} \rightarrow \mathbb{R}, f(x) = x^2$: $f^{-1}(-1) = \emptyset$ und $f^{-1}(4) = \{-2, 2\}$.

Folgende Eigenschaften von Abbildungen haben sich als fundamental herausgestellt.

Definition 4.4: Seien f, X, Y wie in Definition 4.1. Die Abbildung f heißt

1. *injektiv*, wenn

$$f(x) = f(\tilde{x}) \Rightarrow x = \tilde{x} \quad \forall x, \tilde{x} \in X$$

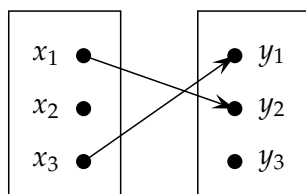
gilt.

2. *surjektiv*, falls $\text{Bild}(f) = Y$ gilt.

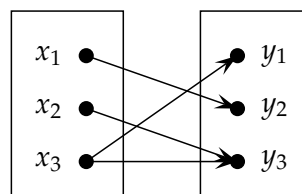
3. *bijektiv*, falls f injektiv und surjektiv ist.

4. *invertierbar*, falls es eine *Umkehrabbildung* oder *inverse Abbildung* $g : Y \rightarrow X$ gibt mit $f(g(y)) = y \forall y \in Y$ und $g(f(x)) = x \forall x \in X$.

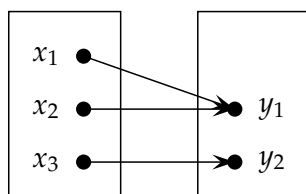
In den folgenden Diagrammen sind verschiedene Abbildungen zwischen Mengen skizziert, an denen die Begriffe aus Definition 4.4 illustriert werden.



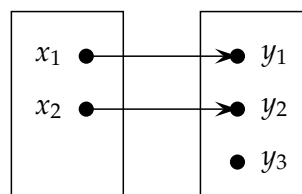
Keine Abbildung, kein Element ist $x_2 \in X$ zugeordnet.



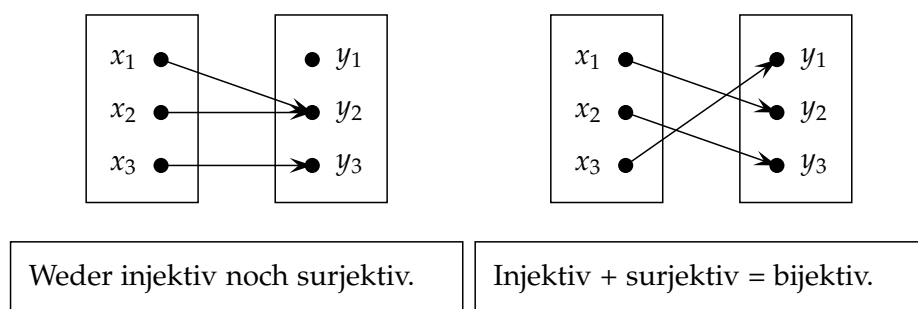
Keine Abbildung, zwei Elemente sind $x_3 \in X$ zugeordnet.



Surjektiv, aber nicht injektiv.



Injektiv, aber nicht surjektiv.



Bemerkung 4.5: Eine Abbildung ist genau dann invertierbar, wenn sie bijektiv ist.

BEWEIS „ \Leftarrow “: Sei f bijektiv, und $y \in Y$ beliebig. Aufgrund der Surjektivität gibt es mindestens ein $x \in X$ mit $f(x) = y$ und aufgrund der Injektivität von f gibt es höchstens ein $x \in X$ mit $f(x) = y$. Also existiert für jedes $y \in Y$ genau ein $x \in X$ mit $f(x) = y$, so dass eine Umkehrabbildung g punktweise durch $g(y) := x$ definiert werden kann. Dann gilt offenbar $f(g(y)) = f(x) = y \forall y \in Y$ und weiter $g(f(x)) = g(y) = x \forall x \in X$. Um letzteres einzusehen, definiert man für ein gegebenes x : $y := f(x)$ und kann dann wie oben argumentieren.

„ \Rightarrow “: Umgekehrt impliziert die Existenz einer Umkehrabbildung g , dass das Urbild $f^{-1}(y)$ für ein beliebiges $y \in Y$ genau ein Element enthält. Wg. $f^{-1}(y) \neq \emptyset \forall y \in Y$ ist f surjektiv. Wäre f nicht injektiv, fände man mindestens ein $y' \in Y$ mit $f(x) = y' = f(\tilde{x})$, aber $x \neq \tilde{x}$. Dann aber enthielte $f^{-1}(y')$ mindestens zwei Elemente. ■

Bemerkung 4.6: Die zu f inverse Abbildung ist eindeutig.

BEWEIS Seien g_1, g_2 zwei inverse Abbildungen zu f . Es gilt $f(g_1(y)) = f(g_2(y)) \forall y \in Y$ und deswegen $g_1(y) = g_2(y) \forall y \in Y$, weil f bijektiv und damit injektiv ist. Also erhält man $g_1 = g_2$. ■

Man bezeichnet die eindeutige Umkehrabbildung von f als f^{-1} . Dabei handelt es sich eigentlich um einen eleganten Missbrauch der Notation, weil ja f^{-1} eine Menge beschreibt. Weil aber diese Menge bei bijektivem f für jedes Element y aus $\text{Bild}(f)$ genau ein Element enthält, identifiziert man $f^{-1}(y)$ mit diesem einen Element und kann über die Zuordnung $y \rightarrow f^{-1}(y)$ eine Abbildung erklären. Dabei handelt es sich genau um die Umkehrabbildung.

Beispiel 4.7: Ob eine Abbildung injektiv oder surjektiv ist, hängt wesentlich von der Wahl des Definitions- und Wertebereichs ab. Die Funktion

$$f: \mathbb{R} \rightarrow \mathbb{R}, f(x) = x^2$$

ist wegen $f^{-1}(-1) = \emptyset$ nicht surjektiv und wegen $f^{-1}(4) = \{-2, 2\}$ nicht injektiv. Die Funktion

$$f : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, f(x) = x^2$$

ist surjektiv, weil jede reelle nichtnegative Zahl eine reelle Wurzel, also ein Urbild unter f , besitzt. Aus demselben Grund wie oben ist f aber nicht injektiv. Dagegen ist die Funktion

$$f : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}, f(x) = x^2$$

sogar bijektiv.

4.2 Grundlegende Eigenschaften linearer Abbildungen

Wir wenden uns jetzt erneut Vektorräumen zu, die ja in der Linearen Algebra eine zentrale Rolle spielen und beschränken uns darauf, *lineare* Abbildungen zwischen Vektorräumen zu untersuchen. Wir werden zunächst elementare Eigenschaften linearer Abbildungen zusammentragen. Im ganzen Kapitel seien V und W zwei K -Vektorräume über demselben Körper K .

Definition 4.8:

1. Eine Abbildung $f : V \rightarrow W$ heißt *linear* oder ein *Homomorphismus*, falls gilt:

$$\begin{aligned} f(x+y) &= f(x) + f(y) && \text{(Additivität)} \\ f(\lambda x) &= \lambda f(x) && \text{(Homogenität)} \end{aligned} \quad \forall x, y \in V, \forall \lambda \in K.$$

2. Es sei $\text{Hom}(V, W)$ die Menge aller Homomorphismen von V nach W .

Beispiel 4.9: Die Abbildung $f : \mathbb{R} \rightarrow \mathbb{R}, f(x) = \alpha x$ mit $\alpha \in \mathbb{R}$ ist linear wegen

$$\begin{aligned} \text{Additivität: } f(x+y) &= \alpha(x+y) \\ &= \alpha x + \alpha y \\ &= f(x) + f(y) && \forall x, y \in \mathbb{R} \\ \text{Homogenität: } f(\lambda x) &= \alpha(\lambda x) \\ &= \lambda \alpha x \\ &= \lambda f(x) && \forall x \in \mathbb{R}, \forall \lambda \in \mathbb{R} \end{aligned}$$

Beispiel 4.10: Die Abbildung $f : \mathbb{R} \rightarrow \mathbb{R}, f(x) = x^2$ ist nicht linear aufgrund von z. B. $f(1+1) = 4 \neq 2 = f(1) + f(1)$.

Dass es höchst sinnvoll ist, lineare Abbildungen zu untersuchen, wird sich vollständig erst am Ende des Kapitels erschließen. Dennoch wollen wir einen Versuch einer Motivation unternehmen.

Bemerkung 4.11:

1. Nach dem Taylorschen Satz aus der Analysis hat jede Funktion $f \in \mathcal{C}^2(\mathbb{R})$ eine Darstellung der Form

$$\begin{aligned} f(x) &= f(x_0) + f'(x_0)(x - x_0) + \mathcal{O}((x - x_0)^2) \\ &= \underbrace{f(x_0) - f'(x_0)x_0}_{:=c} + \underbrace{f'(x_0)x}_{:=\alpha} + \mathcal{O}((x - x_0)^2) \\ &= c + \alpha x + \mathcal{O}((x - x_0)^2), \end{aligned}$$

lässt sich also darstellen als eine Summe einer konstanten Funktion c , einer linearen Funktion αx und einem Restglied, das nahe des Entwicklungspunktes x_0 sehr klein ist. Jede hinreichend glatte Funktion lässt sich daher lokal beliebig gut durch eine lineare Funktion (und konstanten Anteil) approximieren. Analoge Resultate gelten für glatte Funktionen mit mehreren Veränderlichen.

2. Es ist die Frage offen, wie viele strukturell verschiedene Vektorräume es eigentlich gibt. Eine ähnliche Fragestellung in Bezug auf Gruppen wurde in Kapitel 3.1 mithilfe des Isomorphiebegriffs untersucht. Gruppenisomorphismen sind strukturerhaltende bijektive Abbildungen. Betrachtet man in Analogie dazu Abbildungen, die die algebraischen Strukturen von Vektorräumen erhalten, wird man genau auf die linearen Abbildungen aus Definition 4.8 geführt: Die Additivität erhält die Struktur der Gruppe (V, \oplus) , wohingegen die Homogenität die Multiplikation eines Skalars mit einem Vektor widerspiegelt.

Bemerkung 4.12: Seien $f, g : V \rightarrow W$ linear.

1. Für eine lineare Funktion f gilt $f(0) = 0$.
2. Die Funktion f ist genau dann linear, wenn

$$f(x + \lambda y) = f(x) + \lambda f(y) \quad \forall x, y \in V, \lambda \in K \quad (4.1)$$

gilt.

3. Summen, Vielfache linearer Abbildungen und vektorwertige Abbildungen, deren Komponenten aus linearen Abbildungen bestehen, sind wiederum linear.

BEWEIS 1.: Wegen $0 = 0 + 0 \in V$ gilt $f(0) = f(0 + 0) = f(0) + f(0)$. Abziehen von $f(0)$ auf beiden Seiten der Gleichung liefert die Behauptung.

2.: Sei f linear. Dann gilt $f(x + \lambda y) = f(x) + f(\lambda y) = f(x) + \lambda f(y)$. Gelte umgekehrt (4.1). Mit $x = 0 \in V$ und Aussage 1. folgt die Homogenität. Mit $\lambda = 1$ und $1y = y$ folgt die Additivität von f .

3.: Wir betrachten zunächst zwei lineare Abbildungen $f, g : V \rightarrow W$. Dann gilt für beliebige $x, y \in V, \lambda \in K$:

$$\begin{aligned} (f + g)(x + \lambda y) &= f(x + \lambda y) + g(x + \lambda y) \\ &= f(x) + \lambda f(y) + g(x) + \lambda g(y) \\ &= (f + g)(x) + \lambda(f + g)(y). \end{aligned}$$

Für eine Summe von k linearen Funktionen f_1, \dots, f_k zeigt man die Behauptung per vollständiger Induktion: Der Fall $k = 1$ ist klar; die Behauptung gelte für $k - 1$. Die Funktion $\tilde{f}_{k-1} = f_1 + \dots + f_{k-1}$ ist demnach linear und damit nach dem Argument oben auch $\tilde{f}_{k-1} + f_k = f_1 + \dots + f_k$.

Für $\alpha \in K$ gilt

$$(\alpha f)(x + \lambda y) = \alpha f(x + \lambda y) = \alpha (f(x) + \lambda f(y)) = (\alpha f)(x) + \lambda(\alpha f)(y),$$

also sind Vielfache linearer Funktionen wieder linear.

Sei $F(x) := (f(x), g(x))^T$ mit linearen Funktionen f und g . Man hat

$$\begin{aligned} F(x + \lambda y) &= \begin{pmatrix} f(x + \lambda y) \\ g(x + \lambda y) \end{pmatrix} \\ &= \begin{pmatrix} f(x) \\ g(x) \end{pmatrix} + \lambda \begin{pmatrix} f(y) \\ g(y) \end{pmatrix} = F(x) + \lambda F(y). \end{aligned}$$

Im Fall einer Funktion F mit k Komponenten argumentiert man wiederum induktiv. ■

Beispiel 4.13: Die Funktion $f(x) = x + 3$ ist *nicht* linear, weil nach Bemerkung 4.12.1 dann $f(0) = 0$ gelten müsste. Es gilt aber $f(0) = 3$. Aus demselben Grund sind Funktionen mit einem konstanten Anteil niemals linear.

Folgerung 4.14: $(\text{Hom}(V, W), +, \cdot)$ ist ein K -Vektorraum.

BEWEIS Die Menge aller Abbildungen von V nach W , $\text{Abb}(V, W)$, bildet bekanntlich einen Vektorraum. Es genügt also zu zeigen, dass $\text{Hom}(V, W)$ ein Untervektorraum ist. Die Abgeschlossenheit von $\text{Hom}(V, W)$ gegenüber der Addition und der Multiplikation mit einem Skalar folgt unmittelbar aus Bemerkung 4.12.3. ■

Wie umfassend der Begriff der linearen Abbildung ist, mag nachfolgendes Beispiel verdeutlichen.

Beispiel 4.15:

1. Sei V sogar euklidisch, $\langle \cdot, \cdot \rangle$ das darauf definierte Skalarprodukt und $a \in V$. Dann ist die Abbildung $f_a : V \rightarrow \mathbb{R}, f_a(x) := \langle a, x \rangle$ linear. Dies folgt sofort aus den Eigenschaften **SP2** und **SP3** in Definition 3.105:

$$f_a(x + \lambda y) = \langle a, x + \lambda y \rangle = \langle a, x \rangle + \langle a, \lambda y \rangle = f_a(x) + \lambda f_a(y).$$

2. Sei $D : \mathcal{C}^1(\mathbb{R}) \rightarrow \mathcal{C}(\mathbb{R}), D(f) := f'$ die Abbildung, die einer Funktion f ihre Ableitungsfunktion f' zuordnet. Nach der Summen- und Faktorregel der Differentiation gilt für $f, g \in \mathcal{C}^1(\mathbb{R})$:

$$D(f + \lambda g) = (f + \lambda g)' = f' + (\lambda g)' = f' + \lambda g' = D(f) + \lambda D(g)$$

Das „Ableiten an sich“ ist also eine lineare Abbildung.

3. Wir definieren eine Abbildung T , die einer Matrix ihre transponierte Matrix zuordnet, also sei $T : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{m \times n}, T(A) := A^T$. Für zwei Matrizen $A, B \in \mathbb{R}^{n \times m}$ erhält man

$$\begin{aligned} (A + B)^T &= (a_{ij} + b_{ij})_{i=1..n, j=1..m}^T \\ &= (a_{ji} + b_{ji})_{i=1..n, j=1..m} \\ &= (a_{ji})_{i=1..n, j=1..m} + (b_{ji})_{i=1..n, j=1..m} = A^T + B^T. \end{aligned}$$

Weiterhin gilt $(\lambda A)^T = (\lambda a_{ij})_{i=1..n, j=1..m} = \lambda A^T$ für irgendein $\lambda \in K$ und damit

$$T(A + \lambda B) = (A + \lambda B)^T = A^T + \lambda B^T = T(A) + \lambda T(B).$$

Somit lässt sich das Transponieren einer Matrix als lineare Abbildung deuten.

In Kapitel 4.1 wurden die Eigenschaften Injektivität, Surjektivität und Bijektivität von Abbildungen als fundamental herausgestellt. Wir untersuchen nun lineare Abbildungen auf diese Eigenschaften. Ein wesentliches Hilfsmittel dazu ist der *Kern* einer linearen Abbildung.

Definition 4.16: Der *Kern* einer linearen Abbildung $f : V \rightarrow W$ wird definiert durch

$$\ker(f) := f^{-1}(0).$$

Beispiel 4.17:

1. Für die lineare reelle Funktion $f(x) = \alpha x$ aus Beispiel 4.9 gilt $\ker(f) = \{0\}$ für $\alpha \neq 0$ und $\ker(f) = \mathbb{R}$ für $\alpha = 0$.

2. Für die Abbildung $f_a : V \rightarrow \mathbb{R}$ aus Beispiel 4.15 erhält man

$$\ker(f_a) = \{x \in V \mid f_a(x) = 0\} = \{x \in V \mid \langle a, x \rangle = 0\} = \{a\}^\perp$$

also genau das orthogonale Komplement von a (vgl. Definition 3.123). Nach Bemerkung 3.124 handelt es sich dabei um einen Untervektorraum von V .

Dass im obigen Beispiel der Kern immer einen Untervektorraum bildet, ist kein Zufall.

Satz 4.18: Sei $f : V \rightarrow W$ linear. Dann gilt:

1. $\text{Bild}(f)$ ist ein Untervektorraum von W .
2. $\ker(f)$ ist ein Untervektorraum von V .

BEWEIS

1. Seien $w, w' \in \text{Bild}(f)$ und $\lambda \in K$. Dann gibt es $v, v' \in V$ mit $f(v) = w$ und $f(v') = w'$. Folglich gilt

$$w + w' = f(v) + f(v') = f(v + v') \in \text{Bild}(f)$$

und

$$\lambda w = \lambda f(v) = f(\lambda v) \in \text{Bild}(f).$$

2. Gelte $f(v) = f(v') = 0$. Es folgt $f(v) + f(v') = f(v + v') = 0$, also $v + v' \in \ker(f)$ und weiter $\lambda f(v) = f(\lambda v) = 0$ und deswegen $\lambda v \in \ker(f)$. ■

Die Injektivität von linearen Abbildungen lässt sich jetzt einfach charakterisieren.

Satz 4.19: Eine lineare Abbildung f ist genau dann injektiv, wenn $\ker(f) = \{0\}$ gilt.

BEWEIS „ \Rightarrow “: Man hat nach Bemerkung 4.12 $f(0) = 0$. Aufgrund der Injektivität von f existiert kein weiteres $v \neq 0$ mit $f(v) = 0$, und somit enthält $\ker(f)$ als einziges Element $0 \in V$.

„ \Leftarrow “: Sonst gibt es $v \neq v' \in V$ mit $f(v) = f(v')$, also $0 = f(v) - f(v') = f(v - v')$ und damit $0 \neq v - v' \in \ker(f)$. ■

Beispiel 4.20: Nach Satz 4.19 und Bemerkung 4.17 ist $f(x) = \alpha x$ für $\alpha \neq 0$ injektiv.

Wir gehen jetzt davon aus, dass V und W endlich erzeugt seien. Sowohl das Bild als auch der Kern einer linearen Abbildungen sind nach Satz 4.18 Unterräume von V bzw. W mit endlicher Dimension. Wir suchen nach Zusammenhängen zwischen den Dimensionen dieser (Unter)vektorräume. Folgende Definition vereinfacht die Beschreibung der Situation.

Definition 4.21: Für $f : V \rightarrow W$ linear definiert man den *Rang* von f durch

$$\operatorname{rg}(f) := \dim(\operatorname{Bild}(f)).$$

Beispiel 4.22:

1. Für die Funktion $f(x) = \alpha x$ mit $\alpha \neq 0$ aus Beispiel 4.9 erhält man $\operatorname{rg}(f) = 1$, weil $\operatorname{Bild}(f) = \mathbb{R}$ gilt und daher $\operatorname{rg}(f) = \dim(\operatorname{Bild}(f)) = \dim(\mathbb{R}) = 1$. Im Fall $\alpha = 0$ hat man $\operatorname{Bild}(f) = \{0\}$ und deswegen $\operatorname{rg}(f) = 0$.
2. Für f_a mit $a \neq 0$ aus Beispiel 4.15 gilt $\operatorname{rg}(f_a) = 1$, weil z. B. $f_a(a) > 0$ gilt und deswegen $\dim(\operatorname{Bild}(f_a)) \geq 1$. Wegen $\dim(\mathbb{R}) = 1$ hat man aber auch $\operatorname{rg}(f_a) \leq 1$.

Satz 4.23 (Dimensionsformel für lineare Abbildungen): Es sei $f : V \rightarrow W$ linear und $\dim(V) = n$. Dann gilt

$$\dim(\ker(f)) + \operatorname{rg}(f) = n.$$

BEWEIS Nach Satz 4.18 bildet $\ker(f)$ einen Untervektorraum von V und deswegen $\dim(\ker(f)) := r \leq n$. Wir ergänzen eine beliebige Basis (v_1, \dots, v_r) von $\ker(f)$ zu einer Basis $(v_1, \dots, v_r, v_{r+1}, \dots, v_n)$ von V . Setzt man $w_{r+i} = f(v_{r+i})$ für $i = 1, \dots, n - r$, dann gilt $\forall v \in V$:

$$\begin{aligned} f(v) &= f(\lambda_1 v_1 + \dots + \lambda_r v_r + \lambda_{r+1} v_{r+1} + \dots + \lambda_n v_n) \\ &= \lambda_1 \underbrace{f(v_1)}_{=0} + \dots + \lambda_r \underbrace{f(v_r)}_{=0} + \lambda_{r+1} f(v_{r+1}) + \dots + \lambda_n f(v_n) \\ &= \lambda_{r+1} f(v_{r+1}) + \dots + \lambda_n f(v_n) \\ &= \lambda_{r+1} w_{r+1} + \dots + \lambda_n w_n, \end{aligned}$$

also $\operatorname{Bild}(f) = L(w_{r+1}, \dots, w_n)$. Wir zeigen nun, dass w_{r+1}, \dots, w_n linear unabhängig sind. Sei also

$$\lambda_{r+1} w_{r+1} + \dots + \lambda_n w_n = 0.$$

Aus

$$\lambda_{r+1} w_{r+1} + \dots + \lambda_n w_n = f(\lambda_{r+1} v_{r+1} + \dots + \lambda_n v_n) = 0$$

folgt

$$\lambda_{r+1} v_{r+1} + \dots + \lambda_n v_n \in \ker(f).$$

Also gilt

$$\lambda_{r+1}v_{r+1} + \cdots + \lambda_n v_n = \lambda_1 v_1 + \cdots + \lambda_r v_r$$

für gewisse $\lambda_1, \dots, \lambda_r$. Da aber v_1, \dots, v_n linear unabhängig ist, hat man $\lambda_1 = \cdots = \lambda_r = 0$. Somit sind die Vektoren w_{r+1}, \dots, w_n linear unabhängig. Es folgt $\dim(\text{Bild}(f)) = \text{rg}(f) = n - r$ und damit

$$\dim(\ker(f)) + \dim(\text{Bild}(f)) = \dim(\ker(f)) + \text{rg}(f) = r + n - r = n. \quad \blacksquare$$

Als erste Anwendung der Dimensionsformel betrachten wir nun lineare bijektive Abbildungen.

Definition 4.24: Sei $f : V \rightarrow W$ linear.

1. Ist f bijektiv, dann heißt f *Isomorphismus*.
2. Ist f bijektiv und gilt $V = W$, dann heißt f *Automorphismus*.

Bemerkung 4.25: Eine lineare Abbildung $f : V \rightarrow W$ ist genau dann ein Isomorphismus, wenn $\ker(f) = \{0\}$ und $\text{Bild}(f) = W$ gilt.

Folgerung 4.26:

1. Gelte $\dim(V) = \dim(W) = n$ und sei $f : V \rightarrow W$ linear. Dann gilt: f ist injektiv $\Leftrightarrow f$ ist surjektiv $\Leftrightarrow f$ ist bijektiv.
2. Sei $f : V \rightarrow W$ ein Isomorphismus. Dann gilt $\dim(V) = \dim(W)$.

BEWEIS 1. folgt direkt aus der Dimensionsformel 4.23 und Bemerkung 4.25. Zu 2.: Ist f ein Isomorphismus, dann ist f injektiv und deswegen gilt $\dim(\ker(f)) = 0$. Mit der Dimensionsformel 4.23 folgt $\text{rg}(f) = \dim(V)$. Zugleich ist f surjektiv, und deswegen gilt $\text{Bild}(f) = W$. Man erhält $\dim(V) = \text{rg}(f) = \dim(\text{Bild}(f)) = \dim(W)$. \blacksquare

Beispiel 4.27:

1. Nach Bemerkung 4.25 ist $f(x) = \alpha x$ aus Beispiel 4.9 für $\alpha \neq 0$ ein Automorphismus.
2. Soll geprüft werden, ob eine lineare Abbildung f ein Isomorphismus ist, kann man die Dimensionen von V und W vergleichen. Sind sie ungleich, kann f niemals ein Isomorphismus sein. Gilt $\dim(V) < \dim(W)$, dann kann f nicht surjektiv sein; gilt $\dim(V) > \dim(W)$, dann ist f niemals injektiv. So ist für V mit $\dim(V) \geq 2$ die Funktion f_a aus Beispiel 4.15 kein Isomorphismus. Dies erkennt man nach obiger Argumentation ohne nähere Betrachtung von f_a .

Satz 4.28: Sei $f : V \rightarrow W$ ein Isomorphismus. Dann ist $f^{-1} : W \rightarrow V$ ebenfalls ein Isomorphismus.

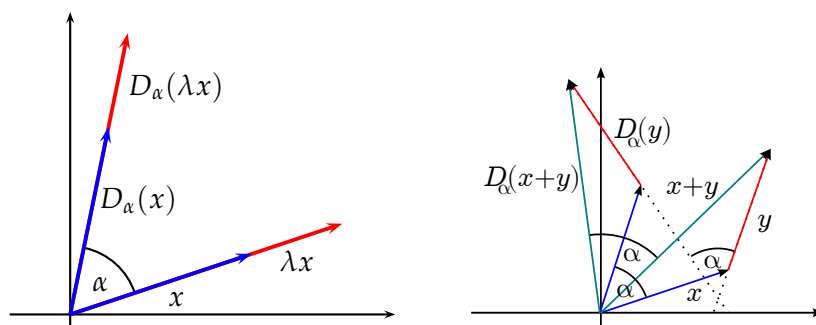


Abbildung 4.1: Die Drehung um den Nullpunkt ist linear

BEWEIS Seien $w, w' \in W, \lambda \in K$. Dann gibt es eindeutige $v, v' \in V$ mit $f^{-1}(w) = v$ und $f^{-1}(w') = v'$. Aus der Linearität von f folgt wegen

$$\begin{aligned} f^{-1}(w + \lambda w') &= f^{-1}(f(v) + \lambda f(v')) \\ &= f^{-1}(f(v + \lambda v')) \\ &= v + \lambda v' = f^{-1}(w) + \lambda f^{-1}(w') \end{aligned}$$

die Linearität von f^{-1} . Wegen $(f^{-1})^{-1} = f$ ist f^{-1} invertierbar und daher nach Bemerkung 4.5 bijektiv. ■

Beispiel 4.29: Wir betrachten die Drehung D_α eines Vektors $x = (x_1, x_2)^T$ in \mathbb{R}^2 um den Winkel α gegen den Uhrzeigersinn um den Nullpunkt. Abb. 4.1 zeigt die Linearität von D_α . Weil Drehungen die Länge eines Vektors erhalten, gilt $\|D_\alpha(x)\| = \|x\| \forall x \in \mathbb{R}^2$ in der euklidischen Norm und daher $\ker(D_\alpha) = \{\|x\| = 0\} = \{0\}$. Damit ist D_α injektiv und nach Bemerkung 4.26 ein Automorphismus von \mathbb{R}^2 . Nach Satz 4.28 existiert also eine ebenfalls lineare Umkehrabbildung D_α^{-1} . Dabei handelt es sich natürlich um die Drehung um den Nullpunkt um α im Uhrzeigersinn.

Der folgende Satz stellt ein einfaches Kriterium bereit, mit dem man prüfen kann, ob eine lineare Abbildung ein Isomorphismus ist oder nicht.

Satz 4.30: Sei $\dim(V) = \dim(W) = n$, (v_1, \dots, v_n) eine Basis von V und $f : V \rightarrow W$ linear. Die Abbildung f ist genau dann ein Isomorphismus, wenn die Bilder $f(v_1), \dots, f(v_n)$ der Basisvektoren von V eine Basis von W bilden.

BEWEIS Bilden die Vektoren $f(v_1), \dots, f(v_n)$ eine Basis von W , gilt $\text{rg}(f) = \dim(\text{Bild}(f)) = n$. Weil gemäß Satz 4.18 $\text{Bild}(f)$ ein Untervektorraum von W ist, folgt $\text{Bild}(f) = W$, also ist f surjektiv. Nach der Dimensionsformel 4.23 gilt dann $\dim(\ker(f)) = 0$ und daher $\ker(f) = \{0\}$. Nach Satz 4.19 ist daher f auch injektiv.

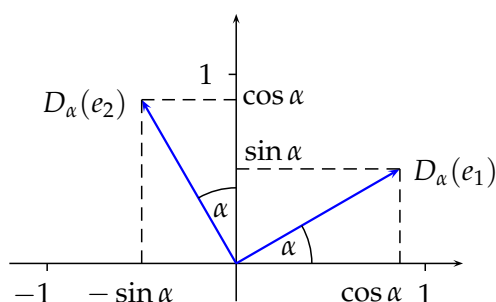


Abbildung 4.2: Drehung der (kanonischen) Basisvektoren um α gegen den Uhrzeigersinn

Umgekehrt sei f ein Isomorphismus. Dann ist f injektiv, also gilt $\ker(f) = \{0\}$. Nach der Dimensionsformel muss dann $\dim(\text{Bild}(f)) = n$ gelten und damit $\text{Bild}(f) = W$. Gleichzeitig wird $\text{Bild}(f) = W$ von den Vektoren $f(v_1), \dots, f(v_n)$ aufgespannt. Sie sind daher erzeugend und auch, da genau n Stück, linear unabhängig. Also bilden sie eine Basis von W . ■

Beispiel 4.31: Wir wollen erneut zeigen, dass D_α ein Isomorphismus ist, aber ohne die Umkehrabbildung explizit anzugeben oder Bild bzw. Kern zu berechnen. Dazu berechnen wir die Bilder der kanonischen Basisvektoren e_1 und e_2 . Elementare Trigonometrie liefert (vgl. Abbildung 4.2)

$$D_\alpha(e_1) = \begin{pmatrix} \cos \alpha \\ \sin \alpha \end{pmatrix} \quad \text{und} \quad D_\alpha(e_2) = \begin{pmatrix} -\sin \alpha \\ \cos \alpha \end{pmatrix}$$

Man errechnet $\langle D_\alpha(e_1), D_\alpha(e_2) \rangle = 0$. Die Bilder der Basisvektoren stehen also orthogonal zueinander, sind damit nach Satz 3.134 linear unabhängig und bilden so eine Basis von $W = \mathbb{R}^2$. Satz 4.30 zufolge ist D_α ein Isomorphismus.

Bemerkung 4.32: Für den Nachweis, dass eine lineare Abbildung f ein Isomorphismus ist, kann man natürlich versuchen, die Umkehrabbildung direkt auszurechnen. Das ist jedoch i. A. sehr aufwändig. Daher sollte man dies wenn möglich vermeiden.

Alternativ dazu ließen sich auch Kern und Bild explizit ausrechnen und mit der Dimensionsformel und ihren Folgerungen argumentieren. Zielführende Rechenverfahren hierzu werden im folgenden Kapitel thematisiert. Ist man aber an Kern und Bild nicht interessiert, sollte man von ihrer ebenfalls aufwändigen Berechnung absehen und das Problem mittels Satz 4.30 auf die einfachere Prüfung auf lineare Unabhängigkeit zurückführen.

4.3 Matrizen und lineare Abbildungen

In diesem Kapitel werden wir die Zusammenhänge zwischen Matrizen und linearen Abbildungen untersuchen. Als ein Nebenergebnis dieser Betrachtungen erhalten lineare Abbildungen, die ja als recht abstrakte Gebilde erscheinen, eine konkrete Gestalt.

Matrizen wurden in der Vorlesung eingeführt als eine effiziente Möglichkeit, lineare Gleichungssysteme zu notieren. Das allgemeine lineare Gleichungssystem (vgl. Formel (1.4))

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \cdots + \cdots + \cdots + \cdots &= \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \quad (4.2)$$

konnte mithilfe der Koeffizientenmatrix $A = (a_{ij})_{i=1\dots m, j=1\dots n}$ und dem Vektor $b = (b_1, \dots, b_m)^T$ beschrieben werden. Man definiert in diesem Zusammenhang die *Matrix-Vektor-Multiplikation*, um den Lösungsvektor $x = (x_1, \dots, x_n)^T$ in die Beschreibung des linearen Gleichungssystems durch A und b einzubeziehen.

Definition 4.33: Seien $A = (a_{ij})_{i=1\dots m, j=1\dots n} \in K^{m \times n}$ und $x = (x_i)_{i=1}^n \in K^n$. Dann sei

$$Ax := \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \cdots + \cdots + \cdots + \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix} \in K^m.$$

Beispiel 4.34: Für

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, \quad x = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$$

errechnet man

$$Ax = \begin{pmatrix} 1 \cdot 1 + 2 \cdot (-1) + 3 \cdot 2 \\ 4 \cdot 1 + 5 \cdot (-1) + 6 \cdot 2 \\ 7 \cdot 1 + 8 \cdot (-1) + 9 \cdot 2 \end{pmatrix} = \begin{pmatrix} 5 \\ 11 \\ 17 \end{pmatrix}.$$

Bemerkung 4.35: Ein Vektor $x \in K^n$ beinhaltet genau dann eine Lösung des linearen Gleichungssystems (4.2), wenn

$$Ax = b$$

gilt.

Wir deuten nun die Matrix-Vektor-Multiplikation als Abbildung.

Satz 4.36: Sei $A \in K^{m \times n}$. Die Abbildung $f_A : K^n \rightarrow K^m, f_A(x) := Ax$ ist linear.

BEWEIS Seien a_1, \dots, a_m die Zeilenvektoren der Matrix. Mit dem euklidischen Skalarprodukt gilt

$$f_A(x) = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} \cdot x = \begin{pmatrix} \langle a_1^T, x \rangle \\ \vdots \\ \langle a_m^T, x \rangle \end{pmatrix}.$$

Damit ist f_A nach Beispiel 4.15.1 und Bemerkung 4.12.3 als vektorwertige Funktion mit komponentenweise linearen Abbildungen linear. ■

Bemerkung 4.37: Für das Matrix-Vektor-Produkt gilt $(A+B)x = Ax + Bx$ sowie $(\lambda A)x = \lambda Ax$ mit $A, B \in K^{m \times n}$ und $\lambda \in K$. Diese durch einfache Rechnung nachvollziehbare Tatsache kann man auch sehr vornehm mit den neu eingeführten Begriffen formulieren: Für ein festes $x \in K^n$ ist die Abbildung $F_x : K^{m \times n} \rightarrow K^m, F_x(A) = Ax$ linear.

Es gilt die Umkehrung von Satz 4.36.

Satz 4.38: Sei $f : K^n \rightarrow K^m$ linear. Dann gibt es genau ein $A \in K^{m \times n}$ mit $f(x) = Ax \forall x \in K^n$, und es gilt

$$A = (f(e_1), \dots, f(e_n)).$$

BEWEIS Eindeutigkeit: Für eine beliebige Matrix $A \in K^{m \times n}$ sei $Ax = 0 \forall x \in K^n$. Mit dem j -ten kanonischen Einheitsvektor e_j gilt $Ae_j = (a_{1j}, \dots, a_{mj})^T = 0$, also $a_{ij} = 0$ und deswegen $A = 0$. Sei jetzt $f(x) = A_1x = A_2x \forall x \in K^n$. Dann gilt nach Satz 4.36, dass $0 = f(x) - f(x) = A_1x - A_2x = (A_1 - A_2)x \forall x \in K^n$ und deswegen $A_1 - A_2 = 0$, also $A_1 = A_2$.

Existenz: Wir setzen

$$A := (f(e_1), \dots, f(e_n)).$$

Die Spalten der Matrix A sind die Bilder der kanonischen Basisvektoren. Es bleibt zu zeigen, dass die lineare Abbildung $x \rightarrow Ax$ der gegebenen Abbildung f entspricht. Sei dazu $x \in K^n$ beliebig. Man hat

$$x = \sum_{j=1}^n x_j e_j$$

und deswegen mit der Linearität von f

$$f(x) = f\left(\sum_{j=1}^n x_j e_j\right) = \sum_{j=1}^n x_j f(e_j). \quad (4.3)$$

Andererseits gilt für die i -te Komponente von Ax :

$$(Ax)_i = \sum_{j=1}^n a_{ij}x_j = \sum_{j=1}^n x_j f_i(e_j),$$

wobei $f_i(e_j)$ die i -te Komponente des Vektors $f(e_j)$ sein soll. Betrachtet man alle Komponenten von Ax zusammen, folgt

$$Ax = \sum_{i=1}^n x_i f(e_i)$$

und mit Gleichung (4.3) die Behauptung. ■

Beispiel 4.39: Für beliebiges $a \in \mathbb{R}^3$ ist die Abbildung $f: \mathbb{R}^3 \rightarrow \mathbb{R}^3$,

$$f(x) = a \times x$$

linear. Dies folgt aus den Eigenschaften des Kreuzprodukts aus Satz 2.40. Damit existiert eine Matrix A , so dass $f(x) = Ax \forall x \in \mathbb{R}^3$ gilt. Nach Satz 4.38 enthalten die Spalten der Matrix die Bilder der kanonischen Basisvektoren. Man errechnet

$$\begin{aligned} f(e_1) &= \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \times \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ a_3 \\ -a_2 \end{pmatrix} \\ f(e_2) &= \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \times \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -a_3 \\ 0 \\ a_1 \end{pmatrix} \\ f(e_3) &= \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \times \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} a_2 \\ -a_1 \\ 0 \end{pmatrix}. \end{aligned}$$

Damit ergibt sich

$$A = \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix}.$$

Beispiel 4.40: Um eine Matrixdarstellung der Drehung D_α aus Beispiel 4.29 zu erhalten, benötigt man die Bilder der (kanonischen) Basisvektoren. Aus Beispiel 4.31 weiß man $D_\alpha(e_1) = (\cos \alpha, \sin \alpha)^T$ sowie $D_\alpha(e_2) = (-\sin \alpha, \cos \alpha)^T$. Also gilt

$$D_\alpha(x) = \underbrace{\begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}}_{:=A_\alpha} x.$$

Matrizen der Form A_α werden *Drehmatrizen* genannt und spielen in der Linearen Algebra eine bedeutende Rolle.

Bemerkung 4.41: Matrizen der Gestalt $A = (f(e_1), \dots, f(e_n))$ lassen sich mit jeder Abbildung f , sei sie linear oder nicht, erzeugen. Nur im linearen Fall aber gilt dann auch $f(x) = Ax \forall x$. Zur Verdeutlichung betrachten wir die offenkundig nichtlineare Funktion

$$f \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_2^2 \\ x_1^2 \end{pmatrix}.$$

Man hat

$$f(e_1) = (0, 1)^T, \quad f(e_2) = (1, 0)^T,$$

also

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

Für $x = (-1, -1)$ aber ist $f(x) = (1, 1)^T \neq (-1, -1)^T = Ax$. Man muss also prüfen, ob eine Funktion f überhaupt linear ist, bevor man eine Darstellung der Form $f(x) = Ax$ angibt.

Im Beweis des Satzes 4.38 wurde die Abbildungsmatrix aus f gewonnen, indem man die Bilder der kanonischen Basisvektoren e_i zu den Spalten von A macht. Weil die Matrix A wiederum f eindeutig festlegt, genügt also die Angabe der Bilder der kanonischen Basisvektoren, um die gesamte Abbildung f eindeutig festzulegen. Es stellt sich die Frage, ob allgemeiner eine beliebige lineare Abbildung zwischen beliebigen Vektorräumen durch die Vorgabe der Bilder von einzelnen Vektoren eindeutig festgelegt wird. Eine Antwort liefert der folgende Satz.

Satz 4.42: Gegeben seien Vektoren $v_1, \dots, v_n \in V$ und $w_1, \dots, w_n \in W$. Bildet (v_1, \dots, v_n) eine Basis von V , dann gibt es genau ein $f \in \text{Hom}(V, W)$ mit $f(v_i) = w_i, 1 \leq i \leq n$. Die Abbildung f hat folgende Eigenschaften:

1. $\text{Bild}(f) = L(f(v_1), \dots, f(v_n))$.
2. f ist injektiv $\Leftrightarrow w_1, \dots, w_n$ sind linear unabhängig.

BEWEIS Sei $v \in V$ beliebig. Dann existiert nach Bemerkung 3.56 eine eindeutige Darstellung der Form

$$v = \sum_{i=1}^n \lambda_i v_i,$$

und wir definieren f durch

$$f(v) = f \left(\sum_{i=1}^n \lambda_i v_i \right) := \sum_{i=1}^n \lambda_i w_i. \quad (4.4)$$

Damit ist f eindeutig festgelegt. Zu zeigen bleibt die Linearität von f . Seien also $v, v' \in V$ und $\alpha \in K$. Es existieren eindeutige $\lambda_i, \lambda'_i, 1 \leq i \leq n$ mit $v = \sum_{i=1}^n \lambda_i v_i$ und $v' = \sum_{i=1}^n \lambda'_i v_i$. Man erhält

$$\begin{aligned} f(v + \alpha v') &= f\left(\sum_{i=1}^n \lambda_i v_i + \alpha \sum_{i=1}^n \lambda'_i v_i\right) \\ &= f\left(\sum_{i=1}^n (\lambda_i + \alpha \lambda'_i) v_i\right) \end{aligned}$$

Aufgrund der Darstellung (4.4) von f folgt daraus

$$\begin{aligned} f(v + \alpha v') &= \sum_{i=1}^n (\lambda_i + \alpha \lambda'_i) w_i \\ &= \sum_{i=1}^n \lambda_i w_i + \alpha \sum_{i=1}^n \lambda'_i w_i \\ &= f(v) + \alpha f(v'). \end{aligned}$$

Offenbar gilt $\text{Bild}(f) \subseteq L(w_1, \dots, w_n)$ aufgrund der Darstellung (4.4) von f . Sei nun $w = \sum_{i=1}^n \beta_i w_i$ ein beliebiges Element aus $L(w_1, \dots, w_n)$. Mit $v := \sum_{i=1}^n \beta_i v_i$ erhält man

$$f(v) = f\left(\sum_{i=1}^n \beta_i v_i\right) = \sum_{i=1}^n \beta_i \underbrace{f(v_i)}_{=w_i} = w$$

und deswegen $L(w_1, \dots, w_n) \subseteq \text{Bild}(f)$. Damit gilt Aussage 1.

Aussage 2, „ \Rightarrow “: Nehmen wir an, w_1, \dots, w_n wären linear abhängig. Dann existiert ein Koeffizientenvektor $(\beta_1, \dots, \beta_n)^T \neq 0$ mit $\sum_{i=1}^n \beta_i w_i = 0$. Weil (v_1, \dots, v_n) eine Basis von V ist, sind die Vektoren v_1, \dots, v_n linear unabhängig. Daher gilt $v := \beta_1 v_1 + \dots + \beta_n v_n \neq 0$, aber $f(v) = 0$. Nach Bemerkung 4.12 gilt aber auch $f(0) = 0$, so dass f nicht injektiv ist.

„ \Leftarrow “: Seien w_1, \dots, w_n linear unabhängig und $v \in V$ beliebig mit $f(v) = 0$. Es gilt $v = \sum_{i=1}^n \alpha_i v_i$ mit gewissen α_i und daher $0 = f(v) = \sum_{i=1}^n \alpha_i f(v_i) = \sum_{i=1}^n \alpha_i w_i$. Aus der linearen Unabhängigkeit der w_i folgt $\alpha_1 = \dots = \alpha_n = 0$ und daher $v = 0$, also $\ker(f) = \{0\}$. Damit ist f nach Satz 4.19 injektiv. ■

Oben wurde gezeigt, dass im Fall von $V = K^n$ und $W = K^m$ jede lineare Abbildung in der Form $f(x) = Ax$ geschrieben werden kann und umgekehrt jede Matrix $A \in K^{m \times n}$ durch $f(x) = Ax$ eine lineare Abbildung f induziert. Dass dies für allgemeine (endlich erzeugte) Vektorräume gilt, besagt Satz 4.43.

Satz 4.43: Seien V und W zwei K -Vektorräume, $\mathcal{B}_V = (v_1, \dots, v_n)$ eine Basis von V und $\mathcal{B}_W = (w_1, \dots, w_m)$ eine Basis von W und weiter $f : V \rightarrow W$ linear. Dann existiert genau eine Matrix $M_{\mathcal{B}_W}^{\mathcal{B}_V}(f) = (a_{ij}) \in K^{m \times n}$ mit

$$f(v_j) = \sum_{i=1}^m a_{ij} w_i \quad \forall j = 1, \dots, n \quad (4.5)$$

BEWEIS Die lineare Abbildung f wird nach Satz 4.42 durch die Vorgabe der Bilder der Basisvektoren von V eindeutig festgelegt. Bezogen auf die Basis \mathcal{B}_W hat jeder Vektor $f(v_j)$ eindeutige Koordinaten $a_{ij} \in K$, $1 \leq i \leq m$, und es gilt $f(v_j) = \sum_{i=1}^m a_{ij} w_i$. Damit wird f eindeutig festgelegt durch die Angabe der Koordinaten der Bilder von v_i unter f . Schreibt man diese Koordinaten spaltenweise in eine Matrix $M_{\mathcal{B}_W}^{\mathcal{B}_V}(f)$, so legt diese mit der Rechenvorschrift (4.5) f eindeutig fest. ■

Bemerkung 4.44: Indem die Matrix $M_{\mathcal{B}_W}^{\mathcal{B}_V}(f)$ die Bilder der Basisvektoren von V kodiert, legt sie f eindeutig fest. Es ist also allgemein gerechtfertigt, statt linearer Abbildungen f Matrizen zu betrachten. Man sagt, $M_{\mathcal{B}_W}^{\mathcal{B}_V}(f)$ sei die *Darstellungsmatrix* von f bezüglich der Basen \mathcal{B}_V und \mathcal{B}_W .

Ein Ergebnis unserer Untersuchungen ist der folgende Merksatz:

In den Spalten von A stehen die Koordinaten der Bilder der Basisvektoren von V bzgl. der gewählten Basis von W .

Beispiel 4.45: Die Abbildung D mit $D(f) = f'$ aus Beispiel 4.15 besitzt überhaupt keine Darstellung durch eine Matrix, weil nach Bemerkung 3.76 weder $V = \mathcal{C}^1(\mathbb{R})$ noch $W = \mathcal{C}(\mathbb{R})$ endlich erzeugt sind. Man kann aber D auf einen endlich erzeugten Unterraum V' einschränken. Dann besitzt $D|_{V'} \rightarrow W'$ eine Matrixdarstellung. Wir wählen $V' = W' = P_4(\mathbb{R})$, und $\mathcal{B} = (1, x, \dots, x^4)$. Sei $f = \sum_{i=0}^4 \alpha_i x^i \in P_4$ beliebig. Ableiten ergibt $D(f) = f' = \sum_{i=1}^4 i \alpha_i x^{i-1}$. Das Polynom f hat bezogen auf \mathcal{B} die Koordinaten $(\alpha_0, \dots, \alpha_4)^T$ und entsprechend f' die Koordinaten $(\alpha_1, 2\alpha_2, 3\alpha_3, 4\alpha_4, 0)^T$. Andererseits stehen nach obigem Merksatz die Koordinaten der Bilder der Basisvektoren in den Spalten der Matrix. Wegen $(x^i)' = ix^{i-1}$ erhält man

$$M_{\mathcal{B}}^{\mathcal{B}}(D) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Damit ergibt sich

$$M_{\mathcal{B}}^{\mathcal{B}}(D) \cdot \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ 2\alpha_2 \\ 3\alpha_3 \\ 4\alpha_4 \\ 0 \end{pmatrix},$$

also genau die Koordinaten von f' bezogen auf \mathcal{B} .

Beispiel 4.46: Die Wahl anderer Basen von V und W führt zu einer anderen Darstellungsmatrix derselben Abbildung f . Sei $V = W = \mathbb{R}^2$ und (bezogen auf die Standardbasis $\mathcal{E} = (e_1, e_2)$) sei $f(x) = Ax$ mit

$$A = \begin{pmatrix} 1,5 & 0,5 \\ 0,5 & 1,5 \end{pmatrix},$$

also $A = M_{\mathcal{E}}^{\mathcal{E}}(f)$. Um die Abbildung f in ihrer Wirkung zu verstehen, betrachten wir das Bild des Einheitsquadrats $[0, 1]^2$ unter f , das in Abb. 4.3 blau dargestellt wird. Es handelt sich um eine bestimmte Kombination aus Drehung und Scherung. Wir wechseln jetzt von der Standardbasis repräsentiert durch (x, y) -Koordinaten in eine um 45° gedrehte Basis $\tilde{\mathcal{E}} = (\tilde{e}_1, \tilde{e}_2)$, die den Koordinaten (\tilde{x}, \tilde{y}) entspricht. Ein analoges Quadrat bezogen auf $\tilde{\mathcal{E}}$, in Abb. 4.3 grün dargestellt, wird durch f längs der \tilde{x} -Achse um den Faktor 2 gestreckt. Es gilt folglich $f(\tilde{e}_1) = 2\tilde{e}_1$ und $f(\tilde{e}_2) = \tilde{e}_2$ und daher

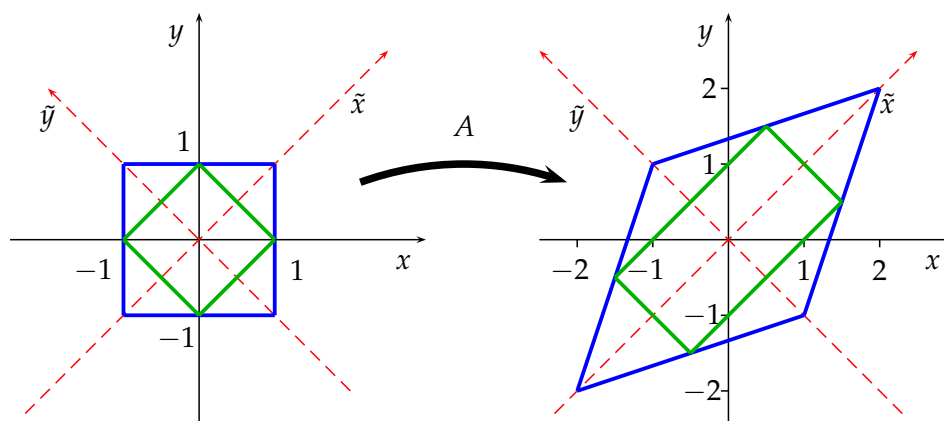
$$\tilde{A} = M_{\tilde{\mathcal{E}}}^{\tilde{\mathcal{E}}}(f) = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix},$$

weil in der Spalten der Matrix die Koordinaten der Bilder der Basisvektoren stehen. Sowohl A als auch \tilde{A} beschreiben dieselbe lineare Abbildung, aber bezogen auf andere Basen, also andere Koordinaten. Wir werden die Umrechnungen von Darstellungsmatrizen linearer Abbildungen bei Basiswechseln in Kapitel 4.5 eingehender untersuchen.

Die zu Beginn des Kapitels in Bemerkung 4.11 aufgeworfene Frage, wie viele strukturell gleiche Vektorräume über einem Körper es denn nun gibt, lässt sich jetzt einfach beantworten.

Definition 4.47: Seien V und W zwei Vektorräume über demselben Körper K . Dann heißen V und W *isomorph*, Schreibweise $V \simeq W$, falls ein Isomorphismus von V nach W existiert.

Satz 4.48: Seien V und W zwei K -Vektorräume mit $\dim(V) = \dim(W) = n$. Dann gilt $K^n \simeq V \simeq W$.

Abbildung 4.3: Quadrate unter der linearen Abbildung f aus Beispiel 4.46

BEWEIS Sei (v_1, \dots, v_n) eine beliebige Basis von V und (w_1, \dots, w_n) eine Basis von W . Nach Satz 4.42 existiert genau eine lineare Abbildung f mit $f(v_i) = w_i, 1 \leq i \leq n$. Nach Satz 4.30 handelt es sich um einen Isomorphismus. Der erste Teil der Aussage folgt sofort mutatis mutandis.

■

Es existiert also (bis auf Isomorphie) nur ein endlich erzeugter K -Vektorraum mit Dimension n , nämlich K^n .

Beispiel 4.49: Aus Satz 4.48 folgt, dass der Vektorraum $P_4(\mathbb{R})$ der reellwertigen Polynome mit Höchstgrad 4, der ja die Dimension 5 aufweist, zu \mathbb{R}^5 isomorph ist. Ein Isomorphismus f lässt sich nach dem Beweis des Satzes konkret angeben: Wir wählen als Basis von $P_4(\mathbb{R})$ die Monombasis $(1, x, \dots, x^4)$ und legen f durch die Bedingungen $f(x^i) = e_{i+1}, 0 \leq i \leq n$ eindeutig fest.

Wir kehren nun zum Spezialfall $V = K^n$ zurück und wollen der Theorie konkrete Berechnungen folgen lassen. Eine einfache Folgerung aus Satz 4.38 vereinfacht die Berechnung des Bildes einer linearen Abbildung wesentlich.

Folgerung 4.50: Sei $A \in K^{m \times n}$, und $f(x) = Ax$. Dann ist das Bild von f gleich der Linearen Hülle der Spaltenvektoren von A .

BEWEIS Sei $e_i \in K^n$ der i -te Vektor der Standardbasis. Dann entspricht $f(e_i)$ der i -ten Spalte von A . Also entspricht die lineare Hülle der Spaltenvektoren von A genau der linearen Hülle der Bilder der Basisvektoren unter f und damit nach Satz 4.38 dem Bild von f . ■

Wir wollen nun die bisher entwickelte Theorie nutzen, um bei zwei linearen Abbildungen beispielhaft Bild und Kern zu berechnen. Es sind viele Rechenwege möglich, aber folgende Reihenfolge der Berechnungen hat sich als besonders effizient herausgestellt:

1. Bestimmung des Kerns
2. Bestimmung der Dimension des Kerns
3. Bestimmung des Rangs (\Rightarrow Dimensionsformel)
4. Bestimmung des Bildes

Beispiel 4.51: Gegeben ist

$$f \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2x_1 + x_2 \\ x_1 - x_2 + x_3 \\ 4x_1 - x_2 + 2x_3 \end{pmatrix}.$$

Es soll gezeigt werden, dass f in der Tat linear ist, sowie $\ker(f)$, $\text{Bild}(f)$ und deren Dimensionen bestimmt werden.

Ein direkter Nachweis der Linearität oder mittels Bemerkung 4.12.2 ist ohne Weiteres möglich. Wir geben stattdessen die Abbildungsmatrix A an. Die Bilder der (kanonischen) Basisvektoren lauten

$$f \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 4 \end{pmatrix}, \quad f \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}, \quad f \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}.$$

Man erhält

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

Man muss jetzt aber noch zwingend nachweisen, dass in der Tat $f(x) = Ax \forall x \in \mathbb{R}^3$ gilt, indem man z. B. für ein allgemeines x sowohl Ax als auch $f(x)$ ausrechnet und Gleichheit zeigt: Es gilt hier mit $x = (x_1, x_2, x_3)^T$

$$A \cdot x = \begin{pmatrix} 2 & 1 & 0 \\ 1 & -1 & 1 \\ 4 & -1 & 2 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2x_1 + x_2 \\ x_1 - x_2 + x_3 \\ 4x_1 - x_2 + 2x_3 \end{pmatrix}.$$

Dies entspricht offenbar $f(x)$, so dass nach Satz 4.36 die Abbildung f linear ist. Zur Bestimmung des Kerns muss man das lineare Gleichungssystem

$$\begin{array}{ccc|c} 2 & 1 & 0 & 0 \\ 1 & -1 & 1 & 0 \\ 4 & -1 & 2 & 0 \end{array}$$

lösen, das ja gerade der Gleichung $Ax = f(x) = 0$ entspricht. Das Gauß-Verfahren liefert $x_3 = \lambda'$; $x_2 = \frac{2}{3} \cdot \lambda'$; $x_1 = -\frac{1}{3} \cdot \lambda'$, also mit $\lambda = \frac{1}{3}\lambda'$:

$$\ker(f) = \left\{ x = \lambda \begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix} \mid \lambda \in \mathbb{R} \right\}$$

Es folgt $\dim(\ker(f)) = 1$ und wegen $\dim(V) = 3$ aus der Dimensionsformel $\dim(\text{Bild}(f)) = 2$. Nach Folgerung 4.50 entspricht $\text{Bild}(f)$ der linearen Hülle der Spalten der Matrix. Man wählt folglich $\dim(\text{Bild}(f))$ Spaltenvektoren aus, z. B. die ersten, und testet, ob sie linear unabhängig sind. Im konkreten Fall ist dies offensichtlich, weil die zweite Spalte kein Vielfaches der ersten ist. Es folgt daher

$$\text{Bild}(f) = \left\{ x; x = \lambda \begin{pmatrix} 2 \\ 1 \\ 4 \end{pmatrix} + \mu \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix} \mid \lambda, \mu \in \mathbb{R} \right\}.$$

Beispiel 4.52: Wir möchten nun $\text{Bild}(D)$ und $\ker(D)$ der Abbildung $D : P_4(\mathbb{R}) \rightarrow P_4(\mathbb{R})$ bestimmen (vgl. Beispiel 4.45). Wir beginnen wiederum mit $\ker(f)$. Wir nutzen aus, dass ein Vektor genau dann der Nullvektor ist, wenn seine Koordinaten sämtlich den Wert 0 annehmen und können so die Aufgabe auf Berechnungen im K^n zurückführen. Die Abbildungsmatrix $M_B^B(D)$ liegt uns bereits aus Beispiel 4.45 vor. Somit lässt sich $\ker(D)$ als Menge aller Vektoren charakterisieren, für deren Koordinaten $x = (x_1, \dots, x_5)^T \in \mathbb{R}^5$ gilt: $M_B^B(D) \cdot x = 0$. Mit dem Gauß-Verfahren löst man das entsprechende lineare Gleichungssystem

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = 0$$

und erhält $x = (\lambda, 0, 0, 0, 0)^T$, $\lambda \in \mathbb{R}$. Dies entspricht allen Vektoren (d. h. Polynomen) der Gestalt $\lambda + 0x + 0x^2 + 0x^3 + 0x^4$, also allen konstanten Polynomen. (Dass deren Ableitungen die Nullfunktion ist, sollte auch unmittelbar klar sein.) Aufgrund der Dimensionsformel muss dann aber $\dim(\text{Bild}(D)) = 4$ gelten. Das Bild von D sind alle Polynome, deren Koordinaten von den Spalten von $M_B^B(D)$ aufgespannt werden, also hier alle Polynome von Höchstgrad 3.

Wesentlich ist, dass man auch bei abstrakteren linearen Abbildungen zwischen endlich erzeugten Räumen sich für alle relevanten Berechnungen auf Berechnungen an Matrizen zurückziehen kann.

4.4 Abbildungsverkettung und Matrizenmultiplikation

Wir werden nun die Hintereinanderausführung von linearen Abbildungen untersuchen.

Satz 4.53: Seien U, V, W K -Vektorräume und $f : V \rightarrow W$ sowie $g : U \rightarrow V$ linear. Dann ist auch $f \circ g : U \rightarrow W$ linear.

BEWEIS Für $u, u' \in U$ und $\lambda \in K$ gilt

$$\begin{aligned}(f \circ g)(u + \lambda u') &= f(g(u + \lambda u')) \\ &= f(g(u) + \lambda g(u')) \\ &= f(g(u)) + \lambda f(g(u')) \\ &= (f \circ g)(u) + \lambda (f \circ g)(u')\end{aligned}$$

Folgerung 4.54: Führt man Isomorphismen hintereinander aus, erhält man wiederum einen Isomorphismus.

Etwas allgemeiner gilt folgende Ranggleichung.

Satz 4.55: Seien in der Situation von Satz 4.53 $\dim(V) = \dim(W) = n$ und f ein Isomorphismus. Dann gilt:

$$\operatorname{rg}(f \circ g) = \operatorname{rg}(g).$$

BEWEIS Sei $r = \operatorname{rg}(g)$ und v_1, \dots, v_r linear unabhängig in $\operatorname{Bild}(g)$. Dann sind $f(v_1), \dots, f(v_r) \in \operatorname{Bild}(f \circ g)$ linear unabhängig: Sei $\sum_{i=1}^r \lambda_i f(v_i) = 0$. Mit der Linearität von f folgt $f(\sum_{i=1}^r \lambda_i v_i) = 0$, also $\sum_{i=1}^r \lambda_i v_i = 0$, da $\ker(f) = \{0\}$ aufgrund der Injektivität von f . Aus der linearen Unabhängigkeit der v_i selbst erhält man $\lambda_i = 0, 1 \leq i \leq r$ und deswegen $\operatorname{rg}(f \circ g) \geq r$.

Seien umgekehrt $w_1, \dots, w_k \in \operatorname{Bild}(f \circ g)$ linear unabhängig. Analog zu oben folgert man, dass $f^{-1}(w_1), \dots, f^{-1}(w_k) \in \operatorname{Bild}(g)$ linear unabhängig sind und gelangt so zu $k \leq r$ und deswegen zu $\operatorname{rg}(f \circ g) \leq r$. ■

Wenn zwei lineare Abbildungen $f : K^n \rightarrow K^m$ und $g : K^\ell \rightarrow K^n$ hintereinander ausgeführt werden, ist die Verkettung $f \circ g$ wiederum linear und besitzt deswegen nach Satz 4.38 eine Darstellung der Gestalt $(f \circ g)(x) = Cx$ mit einer Matrix $C \in K^{m \times \ell}$, die wir jetzt errechnen werden. Dazu gehen wir davon aus, dass $z = f(y) = Ay$ und $y = g(x) = Bx$ gilt. Das Matrix-Vektor-Produkt von B und x liefert $y_j = \sum_{k=1}^{\ell} b_{jk} x_k, 1 \leq j \leq n$. Dann erhält man in Analogie $z_i = \sum_{j=1}^n a_{ij} y_j, 1 \leq i \leq m$. Zugleich gilt aber $z = A(B(x)) = Cx = (f \circ g)(x)$, also auch $z_i = \sum_{j=1}^{\ell} c_{ij} x_j, 1 \leq i \leq m$. Wir

setzen die verschiedenen Darstellungen ineinander ein und sortieren die Summationsreihenfolge um.

$$\begin{aligned} z_i &= \sum_{j=1}^n a_{ij} y_j = \sum_{j=1}^n \left[a_{ij} \left(\sum_{k=1}^{\ell} b_{jk} x_k \right) \right] \\ &= \sum_{j=1}^n \left(\sum_{k=1}^{\ell} a_{ij} b_{jk} x_k \right) \\ &= \sum_{k=1}^{\ell} \left(\sum_{j=1}^n a_{ij} b_{jk} x_k \right) = \sum_{k=1}^{\ell} \underbrace{\left(\sum_{j=1}^n a_{ij} b_{jk} \right)}_{:=c_{ik}} x_k \end{aligned}$$

Damit ist C bestimmt. Ausgehend hiervon definieren wir nun die *Multiplikation von Matrizen*.

Definition 4.56: Es sei $A \in K^{m \times n}$ und $B \in K^{n \times \ell}$. Dann heißt $C = AB \in K^{m \times \ell}$ mit

$$c_{ik} = a_{i1}b_{1k} + \cdots + a_{in}b_{nk} = \sum_{j=1}^n a_{ij}b_{jk} \quad \forall i = 1, \dots, m; k = 1, \dots, \ell$$

das *Produkt* der Matrizen A und B .

$$\begin{array}{c} \begin{pmatrix} b_{11} & \cdots & b_{1k} & \cdots & b_{1\ell} \\ b_{21} & \cdots & b_{2k} & \cdots & b_{2\ell} \\ \vdots & & \vdots & & \vdots \\ b_{n1} & \cdots & b_{nk} & \cdots & b_{n\ell} \end{pmatrix} = B \\ \downarrow \\ \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \xrightarrow{\quad} \left(\begin{array}{c} \\ \\ \longrightarrow c_{ik} \\ \\ \end{array} \right) = C \end{array}$$

Beispiel 4.57: Gegeben sind die Matrizen

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}, \quad B = \begin{pmatrix} 4 & 3 & 2 & 1 \\ 0 & -1 & -2 & -3 \\ 2 & 5 & 4 & 3 \end{pmatrix}.$$

A ist eine 2×3 -Matrix und B ist eine 3×4 -Matrix, also ist das Produkt $C = AB$ eine 2×4 -Matrix. c_{23} ist das Produkt aus der 2. Zeile von A und der 3. Spalte von B .

$$c_{23} = (4 \cdot 2) + (5 \cdot (-2)) + (6 \cdot 4) = 22$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 4 & 3 & 2 & 1 \\ 0 & -1 & -2 & -3 \\ 2 & 5 & 4 & 3 \end{pmatrix} = \begin{pmatrix} \square & \square & \square & \square \\ \square & \square & 22 & \square \end{pmatrix}$$

Das Element c_{14} ist das Produkt aus der 1. Zeile von A und der 4. Spalte von B :

$$c_{14} = (1 \cdot 1) + (2 \cdot (-3)) + (3 \cdot 3) = 4$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 4 & 3 & 2 & 1 \\ 0 & -1 & -2 & -3 \\ 2 & 5 & 4 & 3 \end{pmatrix} = \begin{pmatrix} \square & \square & \square & 4 \\ \square & \square & \square & \square \end{pmatrix}$$

Die restlichen Elemente werden analog berechnet.

Bemerkung 4.58: Die Produktbildung ist nur dann möglich, wenn die Spaltenzahl von A mit der Zeilenzahl von B übereinstimmt. Ansonsten ist das Produkt von A und B nicht definiert.

Wir sammeln einige wichtige Eigenschaften der Matrixmultiplikation.

Satz 4.59: Seien A, B, C so, dass die nachfolgend vorkommenden Matrixmultiplikationen definiert sind. Dann gilt:

1. $A(BC) = (AB)C$ (Assoziativgesetz)
2. $A(B + C) = AB + AC$ und $(A + B)C = AC + BC$ (Distributivgesetz)
3. $(AB)^T = B^T A^T$
4. Sei $A \in K^{m \times n}$ und $E_k \in K^{k \times k}$ die $(k \times k)$ -Einheitsmatrix. Dann gilt $AE_n = E_m A = A$.
5. Es sei 0 eine Nullmatrix. Falls sich die Produkte $0 \cdot A$ und $A \cdot 0$ bilden lassen, sind sie ebenfalls eine Nullmatrix.

BEWEIS

1. Durch A, B und C seien die linearen Abbildungen f, g, h induziert. Dann entspricht $A(BC)$ der Abbildung $f \circ (g \circ h)$ und $(AB)C$ der Abbildung $(f \circ g) \circ h$. Weil aber die Verkettung von Abbildungen assoziativ ist, sind die beiden Abbildungen gleich und damit auch ihre Abbildungsmatrizen.
2. Seien $A \in K^{\ell \times m}$ und $B, C \in K^{m \times n}$. Das i, k -te Element der Matrix $A(B + C)$ wird nach Definition der Matrizenmultiplikation durch $\sum_{j=1}^m a_{ij}(b_{jk} + c_{jk}) = \sum_{j=1}^m a_{ij}b_{jk} + \sum_{j=1}^m a_{ij}c_{jk}$, das i, k -te Element von $AB + AC$ durch $\sum_{j=1}^m a_{ij}b_{jk} + \sum_{j=1}^m a_{ij}c_{jk}$. Damit folgt Gleichheit und damit die Behauptung.

3. Sei $A = (a_{ij}) \in K^{m \times n}$, $B = (b_{ij}) \in K^{n \times \ell}$ und $C = (c_{ij}) := AB \in K^{m \times \ell}$. Dann gilt $c_{ji} = \sum_{k=1}^n a_{jk} b_{ki}$. Sei weiter $D = (d_{ij}) := B^T A^T \in K^{\ell \times m}$. Nach Definition der Matrixmultiplikation erhält man

$$d_{ij} = \sum_{k=1}^n b_{ki} a_{jk} = \sum_{k=1}^n a_{jk} b_{ki} = c_{ji}$$

und wegen $C^T = D$ die Behauptung.

4. Sei $C = (c_{ij}) := AE_n$. Nach Definition der Matrixmultiplikation gilt

$$c_{ik} = a_{i1}e_{1k} + \cdots + a_{in}e_{nk} = a_{ik}e_{kk} = a_{ik}, \quad \forall i = 1, \dots, m; k = 1, \dots, n$$

also $C = A$. Analog zeigt man $EA = A$.

5. folgt durch einfache Rechnung. ■

Bemerkung 4.60: Das Matrixprodukt ist im Allgemeinen *nicht* kommutativ: Für

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

gilt

$$AB = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \text{aber} \quad BA = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Beispiel 4.61: Sei V endlich erzeugt und $id : V \rightarrow V$ die identische Abbildung, also $id(x) = x \forall x \in V$, und \mathcal{B} eine beliebige Basis von V . Dann gilt $M_{\mathcal{B}}^{\mathcal{B}}(id) = E$: Die k -te Spalte von $M_{\mathcal{B}}^{\mathcal{B}}(id)$ enthält die Koordinaten des Bildes des k -ten Basisvektors, also hier die Koordinaten des k -ten Basisvektors selbst. Diese bilden aber gerade den k -ten Einheitsvektor.

Bemerkung 4.62: Fasst man einen Vektor $x \in K^n$ als eine einspaltige Matrix mit n Zeilen auf, dann lässt sich das euklidische Skalarprodukt als Matrixmultiplikation deuten wegen

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i = x^T \cdot y,$$

wobei man die 1×1 -Matrix $x^T \cdot y$ mit ihrem einzigen Eintrag $\langle x, y \rangle$ identifiziert. In diesem Sinn ist also das (euklidische) Skalarprodukt ein Spezialfall der Matrixmultiplikation. In gleicher Weise erscheint jetzt auch die Umwandlung eines Spalten- in einen Zeilenvektor als Spezialfall der Transposition von Matrizen. Entsprechend handelt es sich bei der Matrixvektor-Multiplikation (Definition 4.33) um einen Spezialfall der Matrixmultiplikation.

Sei jetzt $f : V \rightarrow W$ ein Isomorphismus mit Darstellungsmatrix A . Die Darstellungsmatrix des Isomorphismus f^{-1} sei B . Dann gilt $f \circ f^{-1} = f^{-1} \circ f = id$, und damit nach Beispiel 4.61 auch $AB = BA = E$. Da die Einheitsmatrix E sich wie ein neutrales Element bezüglich der Matrixmultiplikation verhält, spielt B die Rolle des inversen Elements zu A . Dies legt folgende Definition nahe.

Definition 4.63: Sei A eine quadratische Matrix. Gibt es eine Matrix A^{-1} mit

$$AA^{-1} = A^{-1}A = E,$$

so heißt A *invertierbar* oder auch *regulär*. A^{-1} wird als *Inverse* von A bezeichnet.

Folgerung 4.64: Eine lineare Abbildung $f : K^n \rightarrow K^m$ ist genau dann invertierbar, wenn ihre Darstellungsmatrix invertierbar ist.

Folgerung 4.65: Jede invertierbare Matrix ist quadratisch.

BEWEIS Ist $A \in K^{m \times n}$ invertierbar, dann ist $f : K^n \rightarrow K^m, f(x) = Ax$ invertierbar, also ein Isomorphismus. Dann gilt nach Folgerung 4.26 aber $\dim(K^n) = \dim(K^m)$, also $n = m$. ■

Beispiel 4.66: Nicht jede quadratische Matrix ist invertierbar: Die Nullmatrix ist die Darstellungsmatrix der Abbildung $f : K^n \rightarrow K^n, f(x) = 0$, die offensichtlich nicht injektiv und damit nicht invertierbar ist. Also ist die Nullmatrix auch nicht invertierbar.

Wir stellen wesentliche Eigenschaften im Hinblick auf Inverse in folgendem Satz zusammen.

Satz 4.67: Seien $A, B \in K^{n \times n}$ invertierbar. Dann gilt:

1.

$$AB = E \Leftrightarrow BA = E \Leftrightarrow B = A^{-1}$$

2. AB ist invertierbar, und es gilt $(AB)^{-1} = B^{-1}A^{-1}$.

3. A^{-1} ist invertierbar, und es gilt $(A^{-1})^{-1} = A$.

4. A^T ist invertierbar, und es gilt $(A^T)^{-1} = (A^{-1})^T$.

5. Für $\lambda \in K \setminus \{0\}$ ist λA invertierbar, und es gilt $(\lambda A)^{-1} = \frac{1}{\lambda}A^{-1}$.

BEWEIS

1. Wurde schon oben bewiesen.

2. Die Invertierbarkeit von AB erhält man aus Folgerung 4.54. Weiter gilt

$$\begin{aligned} B^{-1}A^{-1} &= B^{-1}A^{-1}(AB)(AB)^{-1} \\ &= B^{-1}A^{-1}A(B(AB)^{-1}) \\ &= B^{-1}E(B(AB)^{-1}) \\ &= (B^{-1}B)(AB)^{-1} \\ &= E(AB)^{-1} = (AB)^{-1}. \end{aligned}$$

3. Die Aussage folgt sofort aus $(f^{-1})^{-1} = f$ angewendet auf $f(x) = Ax$ und Teil 1.
4. Es ist $AA^{-1} = E$, also nach Satz 4.59 $(AA^{-1})^T = (A^{-1})^T A^T = E^T = E$. Andererseits gilt $A^{-1}A = E \Rightarrow (A^{-1}A)^T = A^T(A^{-1})^T = E^T = E$. Also erfüllt $(A^{-1})^T$ die definierenden Bedingungen einer inversen Matrix zu A^T ; es ist daher A^T invertierbar mit einer inversen Matrix $(A^T)^{-1} = (A^{-1})^T$.
5. Man hat $E = AA^{-1} = \lambda A \frac{1}{\lambda} A^{-1} = (\lambda A)(\frac{1}{\lambda} A^{-1})$, aber auch $E = A^{-1}A = \frac{1}{\lambda} A^{-1} \lambda A$. Also erfüllt $\frac{1}{\lambda} A^{-1}$ die Bedingungen an eine inverse Matrix zu λA ; daher ist λA invertierbar, und es gilt $(\lambda A)^{-1} = \frac{1}{\lambda} A^{-1}$.

■

Die Eigenschaften der inversen Matrix aus Satz 4.67 haben folgende interessante Konsequenz.

Satz 4.68: Die Menge aller invertierbaren Matrizen in $K^{n \times n}$ bilden zusammen mit der Matrixmultiplikation eine Gruppe.

BEWEIS Wir bezeichnen die Menge aller invertierbaren $n \times n$ -Matrizen mit $GL(n; K)$. Da das Produkt von invertierbaren Matrizen und die Inverse einer invertierbaren Matrix nach Satz 4.67 wiederum invertierbar sind, ist die Multiplikation zweier Matrizen eine Verknüpfung auf $GL(n; K)$. Die Assoziativität wurde in Satz 4.59 gezeigt, ebenso wie die Existenz des neutralen Elements, der Einheitsmatrix E . Die Existenz einer Inversen ist nach Konstruktion klar. ■

Bemerkung 4.69:

1. In der englischen Literatur heißt $GL(n; K)$ „general linear group“, was die Bezeichnung erklärt. Sie ist eines der wichtigsten Beispiele von nicht-abelschen Gruppen.
2. $GL(n; K)$ zusammen mit der Addition und dem Produkt von Matrizen bildet *keinen* Körper. Das scheitert u. a. daran, dass die Summe zweier

invertierbarer Matrizen nicht unbedingt invertierbar sein muss. So sind z. B.

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

offenbar invertierbar, aber

$$A + B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

nicht: Nach Folgerung 4.50 ist das Bild der von $A + B$ definierten linearen Abbildung f die lineare Hülle der Spaltenvektoren von $A + B$. Daher gilt $\text{rg}(f) = 1$. Damit ist f nicht surjektiv und daher nicht invertierbar. Nach Folgerung 4.64 ist dann $A + B$ nicht invertierbar.

Wir wollen jetzt die Inverse einer gegebenen Matrix $A \in GL(n; K)$ berechnen. Sei $A^{-1} = B = (b_1, \dots, b_n)$ mit Spaltenvektoren $b_i, 1 \leq i \leq n$. Nach Satz 4.67 entspricht die Berechnung der inversen Matrix genau der Lösung der Matrixgleichung $AB = E$. Sei $E = (e_1, \dots, e_n)$ mit den kanonischen Einheitsvektoren $e_i, 1 \leq i \leq n$. Dann folgt $AB = E \Leftrightarrow A(b_1, \dots, b_n) = (e_1, \dots, e_n)$, also $Ab_i = e_i$ für $1 \leq i \leq n$. Man erhält damit n lineare Gleichungssysteme mit jeweils n Unbekannten und derselben Matrix A . Löst man diese Gleichungssysteme mit dem Gauß-Verfahren, wird die Abfolge der Umformungen nur von A bestimmt, sie ist also bei allen hier vorkommenden Gleichungssystemen dieselbe. Daher kann man alle Gleichungssysteme simultan lösen, indem man alle rechten Seiten e_i zu einer Einheitsmatrix zusammenfasst und auf alle Spalten dieser Matrix dieselben Umformungen anwendet wie auf A .

Beispiel 4.70: Gegeben sei

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

Gesucht ist A^{-1} . Wir ergänzen die Matrix um eine gleichgroße Einheitsmatrix auf der rechten Seite.

$$\begin{array}{cccc|cccc} 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 2 & 1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 2 & 0 & 0 & 0 & 1 \end{array}$$

Wir erzeugen nun die Stufenform.

$$\begin{array}{ccc|cccc}
 1 & 0 & 11 & 1 & 0 & 0 & 0 & \\
 1 & 1 & 21 & 0 & 1 & 0 & 0 & \quad | - (I) \\
 0 & -1 & 01 & 0 & 0 & 1 & 0 & \\
 1 & 0 & 02 & 0 & 0 & 0 & 1 & \quad | - (I) \\
 \hline
 1 & 0 & 11 & 1 & 0 & 0 & 0 & \\
 0 & 1 & 10 & -1 & 1 & 0 & 0 & \\
 0 & -1 & 01 & 0 & 0 & 1 & 0 & \quad | + (II) \\
 0 & 0 & -11 & -1 & 0 & 0 & 1 & \\
 \hline
 1 & 0 & 11 & 1 & 0 & 0 & 0 & \\
 0 & 1 & 10 & -1 & 1 & 0 & 0 & \\
 0 & 0 & 11 & -1 & 1 & 1 & 0 & \\
 0 & 0 & -11 & -1 & 0 & 0 & 1 & \quad | + (III) \\
 \hline
 1 & 0 & 11 & 1 & 0 & 0 & 0 & \\
 0 & 1 & 10 & -1 & 1 & 0 & 0 & \\
 0 & 0 & 11 & -1 & 1 & 1 & 0 & \\
 0 & 0 & 02 & -2 & 1 & 1 & 1 &
 \end{array}$$

Die Diagonalelemente werden durch Multiplikation auf den Wert 1 gebracht.

$$\begin{array}{ccc|cccc}
 1 & 0 & 11 & 1 & 0 & 0 & 0 & \\
 0 & 1 & 10 & -1 & 1 & 0 & 0 & \\
 0 & 0 & 11 & -1 & 1 & 1 & 0 & \\
 0 & 0 & 02 & -2 & 1 & 1 & 1 & \quad | \cdot 1/2 \\
 \hline
 1 & 0 & 11 & 1 & 0 & 0 & 0 & \\
 0 & 1 & 10 & -1 & 1 & 0 & 0 & \\
 0 & 0 & 11 & -1 & 1 & 1 & 0 & \\
 0 & 0 & 01 & -1 & 1/2 & 1/2 & 1/2 &
 \end{array}$$

Durch die entsprechenden Zeilenoperationen werden oberhalb der Diagonalen Nullen erzeugt:

$$\begin{array}{ccc|cccc}
 1 & 0 & 11 & 1 & 0 & 0 & 0 & \quad | - (IV) \\
 0 & 1 & 10 & -1 & 1 & 0 & 0 & \\
 0 & 0 & 11 & -1 & 1 & 1 & 0 & \quad | - (IV) \\
 0 & 0 & 01 & -1 & 1/2 & 1/2 & 1/2 & \\
 \hline
 1 & 0 & 10 & 2 & -1/2 & -1/2 & -1/2 & \quad | - (III) \\
 0 & 1 & 10 & -1 & 1 & 0 & 0 & \quad | - (III) \\
 0 & 0 & 10 & 0 & 1/2 & 1/2 & -1/2 & \\
 0 & 0 & 01 & -1 & 1/2 & 1/2 & 1/2 & \\
 \hline
 1 & 0 & 00 & 2 & -1 & -1 & 0 & \\
 0 & 1 & 00 & -1 & 1/2 & -1/2 & 1/2 & \\
 0 & 0 & 10 & 0 & 1/2 & 1/2 & -1/2 & \\
 0 & 0 & 01 & -1 & 1/2 & 1/2 & 1/2 &
 \end{array}$$

Die Matrix auf der rechten Seite ist die Inverse A^{-1} .

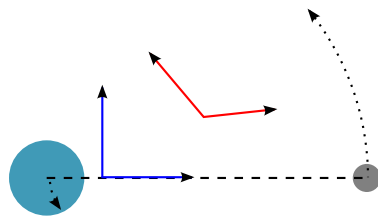


Abbildung 4.4: Ungünstiges (rot) und günstiges (blau) Koordinatensystem zur Beschreibung der Mondbewegung

4.5 Koordinatentransformationen

Ein wesentlicher Aspekt bei der mathematischen Modellbildung besteht darin, das gegebene Problem in Gleichungen zu fassen, die eine möglichst einfache und handhabbare Gestalt haben sollen. Dazu gehört in vielen Fällen die Wahl eines geeigneten Koordinatensystems. Wollte man z. B. die Bewegung des Mondes um die Erde beschreiben, d. h. eine einfache Bewegungsgleichung angeben, wird man wahrscheinlich nicht das rote Koordinatensystem in Abb. 4.4, sondern viel eher das blau dargestellte wählen. (Die Erde und der Mond kreisen um ihren gemeinsamen Schwerpunkt, der sich als Koordinatenursprung daher besonders empfiehlt.) Beide Koordinatensysteme eignen sich prinzipiell zur Beschreibung der Bahnkurven, nur werden die Berechnungen im blau dargestellten Koordinatensystem viel einfacher werden als im rot dargestellten. Sollte sich im Laufe der Modellbildung herausstellen, dass das gewählte Koordinatensystem doch nicht optimal gewählt war, wird man eine *Koordinatentransformation* in ein neues Koordinatensystem durchführen.

Allgemein versteht man in der Analysis unter einer Koordinatentransformation eine invertierbare stetig differenzierbare Abbildung, deren Inverse ebenfalls stetig differenzierbar ist. Man spricht von einem *Diffeomorphismus*. Im Rahmen der Linearen Algebra beschränken wir uns auf *lineare* Koordinatentransformationen; dies ist eine wesentliche Vereinfachung, die unter anderem impliziert, dass die Lage des Nullpunkts sich bei der Koordinatentransformation nicht ändert, weil ja lineare Abbildungen immer den Nullvektor auf sich selbst abbilden.

Beispiel 4.71: Bisher haben wir die Koordinaten eines Vektors in K^n immer bezüglich der kanonischen Einheitsvektoren e_1, \dots, e_n angegeben. Zum Beispiel bedeutete $x = (3, 1, 4)^T$, dass $x = 3e_1 + e_2 + 4e_3$. Auf diese Weise konnte der Vektor x mit seinen Koordinaten identifiziert werden. Wir

wählen jetzt mit

$$b_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, b_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, b_3 = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}$$

eine weitere Basis des \mathbb{R}^3 . Dann hat derselbe Vektor x bezüglich der Basis $\mathcal{B} = (b_1, b_2, b_3)$ die Koordinaten $(2, 1, 3)^T$. Um den Vektor von seinen Koordinaten unterscheiden zu können, schreibt man die *Koordinaten von x bezüglich \mathcal{B}* als

$$K_{\mathcal{B}}(x) = (2, 1, 3)^T.$$

Fehlt im Fall $V = K^n$ die Angabe der Basis, ist immer die Basis \mathcal{E} gemeint, die aus den kanonischen Einheitsvektoren e_1, \dots, e_n besteht. Die Schreibweise $V; \mathcal{A}$ soll im Folgenden bedeuten, dass im Vektorraum V alle Koordinaten bezogen auf die Basis \mathcal{A} zu verstehen sind. Somit handelt es sich in diesem Zusammenhang bei K^n um eine abkürzende Schreibweise für $K^n; \mathcal{E}$.

Wir betrachten nun beliebige endlich erzeugte Vektorräume und zeigen die Existenz von Koordinaten.

Satz 4.72: Sei V ein K -Vektorraum mit einer Basis $\mathcal{B} = (v_1, \dots, v_n)$. Dann existiert genau ein Isomorphismus $\varphi_{\mathcal{B}} : K^n \rightarrow V$ mit $\varphi_{\mathcal{B}}(e_i) = v_i, 1 \leq i \leq n$.

BEWEIS Man wähle in Satz 4.42 $V = K^n$ und $W = V$. ■

Es existiert somit stets eine eindeutige Zuordnung eines abstrakten Vektors v zu einem Tupel von Elementen aus K und umgekehrt. Dieses Tupel von Elementen aus K lässt sich als die Koordinaten des Vektors v auffassen.

Definition 4.73: Der Isomorphismus $\varphi_{\mathcal{B}}$ aus Satz 4.72 heißt *Koordinatenabbildung*, und für $v \in V$ heißen $K_{\mathcal{B}}(v) := \varphi_{\mathcal{B}}^{-1}(v) \in K^n$ die *Koordinaten von v bezüglich \mathcal{B}* .

Bereits in Kapitel 3.4.3 wurden die Koordinaten eines Vektors definiert (Definition 3.55) als die Vorfaktoren der Linearkombination der Basisvektoren und in Bemerkung 3.56 ihre Eindeutigkeit gezeigt. Wir stellen jetzt den Zusammenhang zwischen dieser Definition von Koordinaten und der hier getroffenen her. Für irgendein $v \in V$ und eine beliebige Basis $\mathcal{B} = (v_1, \dots, v_n)$ hat man $v = \sum_{i=1}^n \alpha_i v_i$ mit eindeutigen $\alpha_i \in K$ nach Bemerkung 3.56. Die Abbildung $\varphi_{\mathcal{B}}$ andererseits ist linear, und deswegen gilt

$$\varphi_{\mathcal{B}}((\lambda_1, \dots, \lambda_n)^T) = \varphi_{\mathcal{B}}\left(\sum_{i=1}^n \lambda_i e_i\right) = \sum_{i=1}^n \lambda_i \varphi_{\mathcal{B}}(e_i).$$

Nach Konstruktion gilt $\varphi_{\mathcal{B}}(e_i) = v_i$, und daher auch $\varphi_{\mathcal{B}}((\lambda_1, \dots, \lambda_n)^T) = \sum_{i=1}^n \lambda_i v_i$. Soll jetzt $\varphi_{\mathcal{B}}((\lambda_1, \dots, \lambda_n)^T) = v$ gelten, ist das aufgrund der Eindeutigkeit der Linearkombination nur möglich für $\alpha_i = \lambda_i, 1 \leq i \leq n$. Damit liefern beide Definitionen von Koordinaten dasselbe, nur der Blickwinkel auf Koordinaten ist ein anderer.

Beispiel 4.74: Im Fall von $K^n; \mathcal{E}$ gilt $\varphi_{\mathcal{E}} = id$, und der Vektor ist gleich seinen Koordinaten bezüglich der Standardbasis \mathcal{E} .

In der Situation von Beispiel 4.71 erhält man die Koordinatenabbildung $\varphi_{\mathcal{B}}(x) = Bx$ mit der Matrix $B = (b_1, b_2, b_3)$, weil offenbar $Be_j = b_j$ für $1 \leq j \leq 3$ gilt.

Bemerkung 4.75:

1. Ohne Angabe einer Basis sind Koordinaten wertlos: Wir betrachten den Vektorraum P_2 der Polynome vom Höchstgrad 2. Legt man die Monombasis $(1, x, x^2)$ zugrunde (vgl. Satz 3.99), verbirgt sich hinter den Koordinaten $(1, 0, -1)$ die Funktion $p(x) = 1 - x^2$, wählt man als Basis aber $(1, 1 - x, (1 - x)^2)$, dann entspricht denselben Koordinaten die Funktion $q(x) = 2x - x^2$. Im Fall von $V = K^n$ blieb dieser Sachverhalt bisher deswegen weitgehend verborgen, weil wir immer implizit Koordinaten auf die Standardbasis \mathcal{E} bezogen haben (vgl. auch Beispiel 4.74). In abstrakten Vektorräumen jedoch existiert kein Analogon zur Standardbasis in dem Sinne, dass sich immer eine der vielen Basen als „natürliche Wahl“ empfehlen würde.
2. Alternativ zur Basis selbst lässt sich die Koordinatenabbildung $\varphi_{\mathcal{B}}$ angeben, denn man kann ja einen beliebigen Basisvektor v_j nach Satz 4.72 durch $\varphi_{\mathcal{B}}(e_j)$ berechnen. Umgekehrt legt nach demselben Satz eine Basis \mathcal{B} die Koordinatenabbildung eindeutig fest. Kennt man also die Koordinatenabbildung, dann kennt man auch die Basis und umgekehrt.

Wir wollen nun zwischen verschiedenen Koordinatendarstellungen eines Vektors umrechnen und betrachten dazu einen Vektorraum der Dimension n mit den Basen $\mathcal{A} = (a_1, \dots, a_n)$ und $\mathcal{B} = (b_1, \dots, b_n)$. Für einen Vektor v existieren die Darstellungen $K_{\mathcal{A}}(v)$ und $K_{\mathcal{B}}(v)$. Wir nutzen aus, dass die Basen \mathcal{A} und \mathcal{B} zwei Koordinatenabbildungen $\varphi_{\mathcal{A}}$ und $\varphi_{\mathcal{B}}$ induzieren. Wir erhalten das folgende Diagramm:

$$\begin{array}{ccc}
 & V & \\
 \varphi_{\mathcal{A}} \nearrow & & \nwarrow \varphi_{\mathcal{B}} \\
 K^n & \xrightarrow{\quad} & K^n
 \end{array}$$

$$T_{\mathcal{B}}^{\mathcal{A}} = \varphi_{\mathcal{B}}^{-1} \circ \varphi_{\mathcal{A}}$$

Da alle vorkommenden Abbildungen Isomorphismen sind, kann man das Diagramm in beliebiger Richtung durchlaufen, ohne dass sich das Ergebnis ändert. Ein Diagramm, bei dem die Wahl des Weges keinen Einfluss auf die resultierende Abbildung hat, nennt man *kommutatives Diagramm*. Die Abbildung $\varphi_{\mathcal{B}}^{-1} \circ \varphi_{\mathcal{A}} : K^n \rightarrow K^n$ lässt sich als Automorphismus von K^n mithilfe einer Matrix formulieren; diese Matrix $T_{\mathcal{B}}^{\mathcal{A}}$ heißt *Transformationsmatrix des Basiswechsels von \mathcal{A} nach \mathcal{B}* .

Bemerkung 4.76: Sei $v \in V$ beliebig, $K_{\mathcal{A}}(v) = (x_1, \dots, x_n)^T$ und $K_{\mathcal{B}}(v) = (y_1, \dots, y_n)^T$. Dann gilt

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = T_{\mathcal{B}}^{\mathcal{A}} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

Sind die Koordinaten von v bezüglich \mathcal{A} bekannt, kann man mithilfe der Matrix $T_{\mathcal{B}}^{\mathcal{A}}$ die Koordinaten von v bezüglich \mathcal{B} berechnen.

Bemerkung 4.77: Es gilt $T_{\mathcal{A}}^{\mathcal{B}} = (T_{\mathcal{B}}^{\mathcal{A}})^{-1}$. Dies zeigt das kommutative Diagramm. Man führt also die Berechnung von Koordinaten „in umgekehrter Richtung“ auf die Invertierung einer Matrix zurück bzw. auf das Lösen eines linearen Gleichungssystems.

Wir betrachten den Spezialfall $V = K^n$. Dann lässt sich für jede Basis aus ihren Basisvektoren eine quadratische Matrix bilden, die wiederum die Koordinatenabbildung festlegt. Es mögen also die Vektoren der Basis \mathcal{A} die Matrix A bilden. Dann gilt $\varphi_{\mathcal{A}}(x) = Ax$ aufgrund von $Ae_i = a_i, 1 \leq i \leq n$. Andererseits gilt für die Wahl $\mathcal{B} = \mathcal{E}$, dass $T_{\mathcal{E}}^{\mathcal{A}} = \varphi_{\mathcal{A}}$ wg. $\varphi_{\mathcal{E}} = id$ nach Bemerkung 4.74, so dass $T_{\mathcal{E}}^{\mathcal{A}} = A$ folgt. Eine Umrechnung von Koordinaten eines Vektors x bzgl. \mathcal{A} in Koordinaten bzgl. \mathcal{E} erfolgt demnach durch Multiplikation mit A :

$$x = K_{\mathcal{E}}(x) = A K_{\mathcal{A}}(x)$$

Nach Bemerkung 4.77 folgt sofort

$$K_{\mathcal{A}}(x) = A^{-1} K_{\mathcal{E}}(x) = A^{-1}x.$$

Gerüstet mit diesen Vorüberlegungen betrachten wir eine weitere Basis \mathcal{B} von K^n und die zugehörige Matrix B der Basisvektoren. Wir erhalten das folgende kommutative Diagramm.

$$\begin{array}{ccc} x & \xrightarrow{E} & x \\ \uparrow A & & \uparrow B \\ K_{\mathcal{A}}(x) & \xrightarrow{T_{\mathcal{B}}^{\mathcal{A}}} & K_{\mathcal{B}}(x) \end{array}$$

Man erkennt $T_{\mathcal{B}}^{\mathcal{A}} = B^{-1}A$.

Beispiel 4.78: Bezüglich der Basen

$$\mathcal{A} = \left(\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} \right) \quad \text{und} \quad \mathcal{B} = \left(\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right)$$

hat der Vektor $x = (3, 1, 4)^T$ die Koordinaten $K_{\mathcal{A}}(x) = (2, 1, 3)^T$ und $K_{\mathcal{B}}(x) = (3, -4, 0)^T$. Die Koordinaten eines beliebigen Vektors x lassen sich mit Hilfe der Transformationsmatrix

$$\begin{aligned} T_{\mathcal{B}}^{\mathcal{A}} = B^{-1}A &= \begin{pmatrix} 1 & 0 & 1 \\ -1 & -1 & 1 \\ 0 & -1 & 1 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 1 & 1 & 0 \\ 2 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & -1 & 1 \\ 1 & 1 & -2 \\ 1 & 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & 0 \\ 2 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} -2 & 1 & 2 \\ 3 & -1 & -3 \\ 3 & 0 & -2 \end{pmatrix} \end{aligned}$$

umrechnen. Für die Umrechnung von \mathcal{B} nach \mathcal{A} dient

$$T_{\mathcal{A}}^{\mathcal{B}} = (T_{\mathcal{B}}^{\mathcal{A}})^{-1} = \begin{pmatrix} -2 & 1 & 2 \\ 3 & -1 & -3 \\ 3 & 0 & -2 \end{pmatrix}^{-1} = \begin{pmatrix} -2 & -2 & 1 \\ 3 & 2 & 0 \\ -3 & -3 & 1 \end{pmatrix}.$$

In der Tat gilt dann für obigen Vektor x :

$$\begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} = K_{\mathcal{A}}(x) = T_{\mathcal{A}}^{\mathcal{B}} \cdot K_{\mathcal{B}}(x) = \begin{pmatrix} -2 & -2 & 1 \\ 3 & 2 & 0 \\ -3 & -3 & 1 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ -4 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}.$$

Wir kommen jetzt zur Frage, inwiefern sich Darstellungsmatrizen von linearen Abbildungen (vgl. Bemerkung 4.44) ändern, wenn man von einer Basis zu einer anderen wechselt. Wir gehen von einer linearen Abbildung $f : V \rightarrow W$ aus, wobei \mathcal{A} eine Basis von V sei und \mathcal{B} eine Basis von W . Mittels der Koordinatenabbildungen $\varphi_{\mathcal{A}} : K^n \rightarrow V$ und $\varphi_{\mathcal{B}} : K^m \rightarrow W$ lässt sich folgendes kommutatives Diagramm aufstellen.

$$\begin{array}{ccc} K^n & \xrightarrow{M_{\mathcal{B}}^{\mathcal{A}}(f)} & K^m \\ \varphi_{\mathcal{A}} \downarrow & & \downarrow \varphi_{\mathcal{B}} \\ V & \xrightarrow{f} & W \end{array}$$

Dass das kommutative Diagramm wirklich kommutativ ist, entnimmt man folgendem Satz.

Satz 4.79: Seien V, W endlich erzeugte K -Vektorräume mit Basen \mathcal{A} und \mathcal{B} sowie $f \in \text{Hom}(V; W)$. Dann gilt

$$M_{\mathcal{B}}^{\mathcal{A}}(f) = \varphi_{\mathcal{B}}^{-1} \circ f \circ \varphi_{\mathcal{A}}.$$

BEWEIS Wir rechnen nach, dass die Abbildungen $f \circ \varphi_{\mathcal{A}} : K^n \rightarrow W$ und $\varphi_{\mathcal{B}} \circ M_{\mathcal{B}}^{\mathcal{A}}(f) : K^n \rightarrow W$ tatsächlich identisch sind. Da die Bilder der kanonischen Basisvektoren von K^n beide Abbildungen eindeutig festlegen, betrachten wir nur diese. Zur vereinfachten Notation sei $M_{\mathcal{B}}^{\mathcal{A}}(f) = A = (a_{ij})$. Dann ist

$$\left(\varphi_{\mathcal{B}} \circ M_{\mathcal{B}}^{\mathcal{A}}(f)\right)(e_j) = \varphi_{\mathcal{B}}\left((a_{1j}, \dots, a_{mj})^T\right) = \sum_{i=1}^m a_{ij} w_i$$

nach Konstruktion der Koordinatenabbildung. Andererseits gilt auch

$$(f \circ \varphi_{\mathcal{A}})(e_j) = f(v_j) = \sum_{i=1}^m a_{ij} w_i$$

nach Satz 4.43. ■

Es ist also gleich, ob man direkt die Abbildung f auf v anwendet oder zunächst die Koordinaten von v errechnet, dann mithilfe der Darstellungsmatrix die Koordinaten des Bildes $f(v)$ bestimmt und damit letztlich das Bild $f(v)$ selbst ermittelt.

Eine lineare Abbildung bildet einen Vektor v auf sein Bild $f(v) = w$ ab, ihre Darstellungsmatrix bildet die Koordinaten von v auf die Koordinaten von w ab.

Folgerung 4.80: Im Spezialfall $V = W$ mit Basen \mathcal{A} und \mathcal{B} sowie $f = \text{id}$ folgt aus dem kommutativen Diagramm, dass

$$M_{\mathcal{B}}^{\mathcal{A}}(\text{id}) = \varphi_{\mathcal{B}}^{-1} \circ \text{id} \circ \varphi_{\mathcal{A}} = T_{\mathcal{B}}^{\mathcal{A}}.$$

Es lassen sich also Darstellungsmatrizen als eine Verallgemeinerung von Transformationsmatrizen des Basiswechsels auffassen.

Nach diesen Vorüberlegungen kommen wir jetzt zum Hauptergebnis dieses Kapitels.

Satz 4.81: Seien V und W endlich erzeugt mit Basen \mathcal{A} und \mathcal{A}' bzw. \mathcal{B} und \mathcal{B}' . Sei weiter $f : V \rightarrow W$ linear. Dann gilt

$$M_{\mathcal{B}'}^{\mathcal{A}'}(f) = T_{\mathcal{B}'}^{\mathcal{B}} \cdot M_{\mathcal{B}}^{\mathcal{A}}(f) \cdot \left(T_{\mathcal{A}'}^{\mathcal{A}}\right)^{-1}.$$

BEWEIS Wir betrachten folgendes Diagramm:

$$\begin{array}{ccc}
 K^n & \xrightarrow{M_B^A(f)} & K^m \\
 \downarrow T_{A'}^A & \searrow \varphi_A & \swarrow \varphi_B \\
 & V & \xrightarrow{f} & W \\
 & \swarrow \varphi_{A'} & \searrow \varphi_{B'} & \\
 K^n & \xrightarrow{M_{B'}^{A'}(f)} & K^m \\
 & \downarrow T_{B'}^B &
 \end{array}$$

Nach den Überlegungen zu Transformationsmatrizen weiter oben und Satz 4.79 sind alle vier Teildiagramme kommutativ, und deswegen auch das gesamte Diagramm. ■

Folgerung 4.82: Seien $V = K^n = W$. Seien \mathcal{A} und \mathcal{B} Basen von K^n und S die Matrix des Basiswechsels von \mathcal{A} nach \mathcal{B} . Sei weiter f eine lineare Abbildung mit Darstellungsmatrix A bezogen auf Basis \mathcal{A} . Für die Darstellungsmatrix B von f bezogen auf \mathcal{B} gilt dann

$$B = SAS^{-1}.$$

Beispiel 4.83: Die Abbildung $f : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ mit $f((x, y, z)^T) = (x, y)^T$ projiziert \mathbb{R}^3 auf die (x, y) -Ebene. Die Abbildungsmatrix bezüglich der kanonischen Basen A und B ist

$$M_B^A(f) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Um die Abbildungsmatrix bezüglich der Basen

$$A' = \left(\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} \right) \quad \text{und} \quad B' = \left(\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \end{pmatrix} \right)$$

zu bestimmen, nutzen wir also

$$\begin{aligned}
 M_{B'}^{A'}(f) &= T_{B'}^B \cdot M_B^A(f) \cdot T_A^{A'} = (T_B^{B'})^{-1} \cdot M_B^A(f) \cdot T_A^{A'} \\
 &= \begin{pmatrix} 2 & -3 \\ -1 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & 0 \\ 2 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix} \\
 &= \begin{pmatrix} -4 & 2 & 3 \\ 3 & -1 & -2 \end{pmatrix}
 \end{aligned}$$

Beispiel 4.84: Wir kehren zur linearen Abbildung f aus Beispiel 4.46 zurück, deren Darstellungsmatrix A bezogen auf die Standardbasis $\mathcal{E} = (e_1, e_2)$ gegeben ist durch

$$A = \begin{pmatrix} 1,5 & 0,5 \\ 0,5 & 1,5 \end{pmatrix}.$$

Verwendet man stattdessen eine um 45° gegen den Uhrzeigersinn gedrehte Basis $\tilde{\mathcal{E}} = (\tilde{e}_1, \tilde{e}_2)$, dann führten uns geometrische Überlegungen (vgl. Abb. 4.3) zur Darstellungsmatrix

$$\tilde{A} = M_{\tilde{\mathcal{E}}}^{\tilde{\mathcal{E}}}(f) = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}.$$

Wir wollen nun dasselbe Ergebnis durch Rechnung erreichen. Die Drehung um 45° gegen den Uhrzeigersinn führt zu den Basisvektoren $\tilde{e}_1 = \frac{1}{\sqrt{2}}(1, 1)^T$ und $\tilde{e}_2 = \frac{1}{\sqrt{2}}(-1, 1)^T$ und daher zur Matrix des Basiswechsels

$$T_{\tilde{\mathcal{E}}}^{\tilde{\mathcal{E}}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Damit erhält man

$$T_{\tilde{\mathcal{E}}}^{\mathcal{E}} = \left(T_{\tilde{\mathcal{E}}}^{\tilde{\mathcal{E}}}\right)^{-1} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

Es folgt

$$\begin{aligned} M_{\tilde{\mathcal{E}}}^{\tilde{\mathcal{E}}}(f) &= T_{\tilde{\mathcal{E}}}^{\tilde{\mathcal{E}}} A T_{\tilde{\mathcal{E}}}^{\mathcal{E}} \\ &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1,5 & 0,5 \\ 0,5 & 1,5 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} = \tilde{A}. \end{aligned}$$

Kapitel 5

Determinanten

5.1 Motivation und Einführung

In Kapitel 2.3 haben wir die *Determinante* für 2×2 - und 3×3 -Matrizen in Form einer Rechenvorschrift eingeführt und untersucht. Als wesentliche Anwendung konnten wir in Satz 2.95 zeigen, dass ein lineares Gleichungssystem $Ax = b$ mit $A \in \mathbb{R}^{n \times n}$ und $n \in \{2, 3\}$ genau dann eine eindeutige Lösung besitzt, wenn $\det(A) \neq 0$ gilt. Die Determinante liefert also ein einfaches Kriterium zur Lösbarkeit eines linearen Gleichungssystems. Zudem führte die Determinante in Satz 2.91 zu einer einfachen Formel für das orientierte Volumen für ein durch drei Vektoren aufgespanntes Spat. Wir wollen nun den Begriff der Determinante für ein beliebiges $n \in \mathbb{N}$ erklären und hoffen dabei die obigen nützlichen Anwendungen in ähnlicher Form auch im allgemeinen Fall zu erhalten. Dazu schlagen wir den bewährten Weg bei Verallgemeinerungen ein: Wir gehen von den wesentlichen Eigenschaften des zu verallgemeinernden Objekts aus, erheben diese zur Definition und bezeichnen alles, was diese Eigenschaften hat, mit demselben Namen wie den (vorherigen) Spezialfall.

Wir betrachten nun Eigenschaften der Determinante in \mathbb{R}^3 , die sich durch ihre Charakterisierung als Volumenformel ergeben:

1. Vertauscht man die Reihenfolge zweier Matrixspalten, dann wird aus einem Rechts- ein Linkssystem oder umgekehrt, und das Vorzeichen der Determinante ändert sich.
2. Schneidet man ein Spat parallel zu einer seiner Flächen, dann ist das Volumen des Spats die Summe der Volumina der beiden Teilspate (vgl. Abb. 5.1).
3. Verlängert man eine Seite eines Spats um einen Faktor λ , dann vervielfacht sich sein Volumen um λ .
4. Das Volumen des Einheitswürfels beträgt eins.

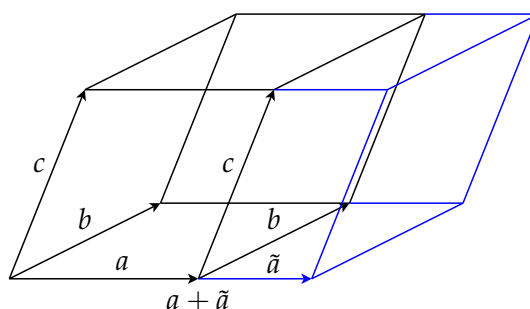


Abbildung 5.1: Das Volumen des Spats ist gleich der Summe der Volumina der beiden Teilspace

Sei im ganzen Kapitel stets K ein Körper und $n \in \mathbb{N}$.

Definition 5.1: Eine Abbildung $\det : K^{n \times n} \rightarrow K$ heißt *Determinante*, wenn gilt

1. \det ist *alternierend*: Vertauscht man in $A = (a_1, \dots, a_n) \in K^{n \times n}$ zwei benachbarte Spalten, dann ändert $\det(A)$ ihr Vorzeichen:

$$\det(\dots, a_i, a_{i+1}, \dots) = -\det(\dots, a_{i+1}, a_i, \dots)$$

2. \det ist *homogen*: Multipliziert man eine Spalte von A mit $\lambda \in K$, so gilt

$$\det(a_1, \dots, \lambda a_i, \dots, a_n) = \lambda \det(a_1, \dots, a_i, \dots, a_n)$$

3. \det ist *additiv*: Stimmen zwei Matrizen überall bis auf die i -te Spalte überein, so addiert man ihre Determinanten, indem man die beiden i -ten Spalten addiert und dann die Determinante der Summe bildet:

$$\begin{aligned} & \det(a_1, \dots, a_{i-1}, a_i, a_{i+1}, \dots, a_n) \\ & + \det(a_1, \dots, a_{i-1}, a_i^*, a_{i+1}, \dots, a_n) \\ & = \det(a_1, \dots, a_{i-1}, a_i + a_i^*, a_{i+1}, \dots, a_n) \end{aligned}$$

4. \det ist *normiert*: Für die Einheitsmatrix E gilt $\det(E) = 1$.

Bemerkung 5.2: Definiert man wie bei Definition 5.1 ein Objekt abstrakt über einige seiner Eigenschaften, spricht man von einer *axiomatischen Definition*. Hierbei sind zumindest die in der Definition postulierten Eigenschaften eines Objekts klar, unklar bleibt jedoch, ob es überhaupt ein solches Objekt gibt und ob es durch die festgelegten Eigenschaften eindeutig bestimmt ist. Bei axiomatischen Definitionen muss man also die Wohldefiniertheit streng beweisen. Alternativ dazu kann man ein Objekt z. B. über

eine konkrete Rechenvorschrift definieren. Ist diese immer ausführbar, sind Existenz und Eindeutigkeit klar, dafür muss man dann die Eigenschaften des Objekts nachweisen. Historisch wurden Determinanten zuerst über Rechenvorschriften definiert, Weierstraß¹ behandelte um 1870 erstmals die Determinante axiomatisch; der „summierte Beweisaufwand“ ist bei beiden Zugängen in etwa gleich. Wir werden nach etwas Theorie durch Angabe einer Rechenvorschrift die Frage der Wohldefiniertheit beantworten und so den axiomatischen Zugang rechtfertigen.

Bemerkung 5.3: Gelegentlich verwendet man für die Determinante einer Matrix A statt $\det(A)$ die kürzere Schreibweise $|A|$, z. B.

$$\begin{vmatrix} 1 & -1 \\ 2 & 5 \end{vmatrix} = \det \begin{pmatrix} 1 & -1 \\ 2 & 5 \end{pmatrix}.$$

5.2 Vorbereitung: Elementarmatrizen

Die Untersuchung weiterer Eigenschaften der Determinante gestaltet sich viel einfacher, wenn man *Elementarmatrizen* verwendet. Zudem werden sie uns die Analyse des Gauß-Verfahrens in Kapitel 6 sehr erleichtern. Daher führen wir jetzt Elementarmatrizen ein.

Definition 5.4: Seien $1 \leq i, j \leq n$ mit $i \neq j$ und $\lambda \in K \setminus \{0\}$ gegeben. Dann sei

$$C1 := \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & \lambda & \\ & & & \ddots \\ & & & & 1 \end{pmatrix} \in K^{n \times n},$$

wobei der (i, j) -te Eintrag den Wert λ annehmen soll und alle anderen Einträge außerhalb der Hauptdiagonalen 0 sein sollen. Sei $C2$ die Matrix, die man aus der Einheitsmatrix gewinnt, indem man die i -te und j -te Spalte

¹Karl Weierstraß, (1815–1897), dt. Mathematiker; fundamentale Beiträge zur Analysis

vertauscht, also

$$C2 := \left(\begin{array}{ccc|ccc|ccc} 1 & & & & & & & & & & \\ & \ddots & & & & & & & & & \\ & & 1 & & & & & & & & \\ - & - & - & 0 & - & - & - & 1 & - & - & - \\ & & & & 1 & & & & & & \\ & & & & & \ddots & & & & & \\ - & - & - & & & & 1 & & & & \\ - & - & - & 1 & - & - & - & 0 & - & - & - \\ & & & & & & & & 1 & & \\ & & & & & & & & & \ddots & \\ & & & & & & & & & & 1 \end{array} \right).$$

Zuletzt definieren wir

$$C3 := \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & \lambda & \\ & & & 1 \end{pmatrix} \in K^{n \times n}.$$

Matrizen der Gestalt $C1$, $C2$ oder $C3$ nennt man *Elementarmatrizen*.

Bemerkung 5.5: Die Multiplikation einer Matrix A von links mit einer Elementarmatrix C_i entspricht der Anwendung einer elementaren Zeilenoperation Z_i des Gauß-Verfahrens auf A (vgl. Kapitel 1.6.1).

Beispiel 5.6: Sei

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

Um A auf Stufenform zu bringen, eliminiert man zunächst den Eintrag $a_{21} = 2$, indem man das (-2) -fache der ersten zur zweiten Spalte addiert und so die Matrix

$$A^{(1)} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & -2 \\ 2 & 1 & 1 \end{pmatrix}$$

erhält. Dieser Zeilenoperation vom Typ $Z1$ entspricht die Anwendung der Matrix

$$C1^{(1)} = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

mit $\lambda = -2$ von links auf A ; durch Ausrechnen überzeugt man sich von $A^{(1)} = C1^{(1)}A$. Man eliminiert jetzt $a_{31}^{(1)} = 2$ durch Addition des (-2) -fachen der ersten Zeile von $A^{(1)}$ zur dritten und erhält

$$A^{(2)} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & -2 \\ 0 & -1 & -1 \end{pmatrix},$$

die zugehörige Elementarmatrix lautet

$$C1^{(2)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix},$$

und es gilt wiederum $A^{(2)} = C1^{(2)}A^{(1)} = C1^{(2)}C1^{(1)}A$. Vertauscht man die zweite und dritte Zeile von $A^{(2)}$ (Zeilenoperation Z2 im Gauß-Verfahren), ist die Stufenform erreicht. Die Zeilenvertauschung entspricht der Anwendung der Elementarmatrix

$$C2^{(3)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

mit $i = 2$ und $j = 3$; man erhält

$$A^{(3)} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & -1 \\ 0 & 0 & -2 \end{pmatrix} = C2^{(3)}C1^{(2)}C1^{(1)}A.$$

Möchte man eine reduzierte Stufenform errechnen, setzt man das Gauß-Verfahren entsprechend fort; dies entspricht weiteren Multiplikationen mit Elementarmatrizen von links.

Das Beispiel zeigt, dass sich eine Folge von k elementaren Zeilenoperationen im Gauß-Verfahren als Multiplikation von k Elementarmatrizen von links darstellen lässt. Damit erhalten wir sofort:

Folgerung 5.7: Eine Matrix A lässt sich genau dann als Produkt von Elementarmatrizen darstellen, wenn sie sich mit dem Gauß-Verfahren in die reduzierte Stufenform überführen lässt.

Bemerkung 5.8: Elementarmatrizen nach Definition 5.4 sind invertierbar, ihre Inversen sind wiederum Elementarmatrizen.

BEWEIS Man hat

$$C1^{-1} = \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & -\lambda & \ddots \\ & & & & 1 \end{pmatrix}, \quad C3^{-1} = \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1/\lambda & \\ & & & & 1 \end{pmatrix},$$

man beachte $\lambda \neq 0$, und $C2^{-1} = C2$ (doppelt Vertauschen heißt gar nicht Vertauschen). ■

Wir betrachten kurz die Multiplikation mit einer Elementarmatrix von *rechts*. Es stellt sich heraus, dass dies zur Anwendung von *Spalten-* statt *Zeilenoperationen* auf A führt. Analog zur jeweiligen Zeilenoperationen werden sie mit $S1 - S3$ bezeichnet. Elementares Ausrechnen beweist den folgenden Satz.

Satz 5.9: Die Spaltenoperation S_i auf A entspricht genau der Multiplikation mit der Elementarmatrix C_i von rechts, $i \in \{1, 2, 3\}$.

5.3 Eigenschaften der Determinante

Wir stellen die Frage nach der Wohldefiniertheit der Determinanten zurück und werden nun aus den vier definierenden Eigenschaften der Determinante weitere Eigenschaften herleiten. Der Einfachheit wegen beschränken wir uns in diesem Unterkapitel auf die Körper $K \in \{\mathbb{R}, \mathbb{C}\}$. Der Beweis der folgenden Aussage wird nachgeholt.

Satz 5.10: Für $A \in K^{n \times n}$ gilt

$$\det(A) = \det(A^T),$$

wobei A^T die zu A transponierte Matrix ist.

Es folgt, dass die definierenden Eigenschaften 1–4 der Determinante auch für Zeilen an Stelle von Spalten gelten. Daher können wir uns für die Beweise der folgenden Aussagen auf Spalten beschränken und erhalten die analogen Aussagen für Zeilen sofort durch Transposition von A .

Bemerkung 5.11:

1. Vertauscht man zwei beliebige Spalten/Zeilen, so ändert die Determinante ihr Vorzeichen.
2. Besitzt eine Matrix A zwei gleiche Spalten/Zeilen, so gilt $\det(A) = 0$.

BEWEIS

1. Ist der Abstand zweier Spalten k , so lässt sich die Vertauschung der beiden Spalten schreiben als $2k - 1$ Vertauschungen von benachbarten Spalten. Bezeichnen wir die Matrix mit den vertauschten Spalten als \tilde{A} , dann gilt $\det(\tilde{A}) = (-1)^{2k-1} \det(A) = (-1)^{2k-2} (-1) \det(A) = [(-1)^2]^{k-1} (-1) \det(A) = -\det(A)$.
2. Durch Vertauschung der beiden gleichen Spalten erhält man $\det(A) = -\det(A)$. ■

Wir untersuchen die Wirkung der Äquivalenzumformungen Z1 – Z3 bzw. S1 – S3 des Gauß-Verfahrens auf die Determinante einer Matrix A .

Bemerkung 5.12:

1. S1 und Z1 ändern die Determinante einer Matrix nicht.
2. S2 und Z2 kehren das Vorzeichen der Determinante um.
3. S3 und Z3: Multiplikation einer Zeile oder Spalte mit einem Faktor λ vervielfacht den Wert der Determinante um den Faktor λ .

BEWEIS Die Aussagen für S2 und S3 sind klar. Wir ersetzen Spalte a_i durch $a_i + \alpha a_j$ und errechnen

$$\begin{aligned} & \det(a_1, \dots, a_{i-1}, a_i + \alpha a_j, a_{i+1}, \dots, a_n) \\ &= \det(a_1, \dots, a_n) + \det(a_1, \dots, a_{i-1}, \alpha a_j, a_{i+1}, \dots, a_n) \\ &= \det(a_1, \dots, a_n) + \underbrace{\alpha \det(a_1, \dots, a_{i-1}, a_j, a_{i+1}, \dots, a_n)}_{=0} \\ &= \det(a_1, \dots, a_n) \end{aligned}$$

Mit Satz 5.10 folgen sofort die Aussagen für Z1, Z2 und Z3. ■

Satz 5.13:

1. Für die Elementarmatrizen aus Definition 5.4 gilt $\det(C1) = 1$, $\det(C2) = -1$ und $\det(C3) = \lambda$.
2. Für eine Elementarmatrix C und eine beliebige quadratische Matrix A gilt $\det(CA) = \det(C) \det(A) = \det(AC)$.

BEWEIS Weil die Determinante alternierend ist, gilt $\det(C2) = -1$, aus der Homogenität folgt $\det(C3) = \lambda$, und man errechnet

$$\begin{aligned} \det(C1) &= \det(e_1, \dots, e_i + \lambda e_j, e_{i+1}, \dots) \\ &= \det(E) + \lambda \det(\dots e_j, \dots, e_j, \dots) = 1. \end{aligned}$$

Aussage 2. folgt dann mit Bemerkung 5.12. ■

Folgerung 5.14: Für eine obere Dreiecksmatrix

$$A = \begin{pmatrix} \lambda_1 & \cdots & & \\ & \ddots & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

gilt $\det(A) = \lambda_1 \cdot \dots \cdot \lambda_n$.

BEWEIS Sind alle $\lambda_i \neq 0$, dann lässt sich A nur mit Zeilenumformungen vom Typ Z1 auf Diagonalgestalt bringen. Die Determinante ändert sich dadurch nach Bemerkung 5.12 nicht. Man erhält aufgrund der Homogenität der Determinante

$$\det(A) = \det \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} = \lambda_1 \cdot \dots \cdot \lambda_n \det(E)$$

und mit der Normiertheit der Determinante die Behauptung.

Gibt es $\lambda_i = 0$, wählen wir von allen das mit dem größten Index. Dann ist $\lambda_{i+1}, \dots, \lambda_n \neq 0$. Mit Z1 und den Zeilen $i+1, \dots, n$ lassen determinanterhaltend sich alle Einträge der i -ten Zeile z_i rechts der Hauptdiagonalen zu 0 machen. Damit gilt $z_i = 0$ und deswegen $\det(A) = 0$. ■

Wir kommen nun zur zentralen Anwendung der Determinante als Kriterium zur eindeutigen Lösbarkeit eines linearen Gleichungssystems.

Satz 5.15: Sei $A \in K^{n \times n}$. Dann gilt

$$A \text{ invertierbar} \Leftrightarrow \det(A) \neq 0.$$

BEWEIS Es ist A genau dann invertierbar, wenn die von A induzierte Abbildung $L_A(x) := Ax$ invertierbar ist, wenn also $\text{rg}(L_A) = n$ gilt. Mittels Zeilenumformungen vom Typ Z1 und Z2 kann man A in Stufenform transformieren, es existiert somit eine obere Dreiecksmatrix B mit $B = C_k \cdot \dots \cdot C_1 A$, und es gilt nach Satz 4.55 $\text{rg}(L_B) = \text{rg}(L_A)$, weil Elementarmatrizen nach Bemerkung 5.8 invertierbar sind und deswegen Isomorphismen beschreiben. Es kann L_B aber nur maximalen Rang besitzen, wenn alle Hauptdiagonalelemente von B nicht Null sind, also genau dann, wenn nach Folgerung 5.14 $\det(B) \neq 0$ gilt. Nach Folgerung 5.13 gilt aber auch $\det(B) = 0 \Leftrightarrow \det(A) = 0$. Daher erhält man

$$\begin{aligned} A \text{ invertierbar} &\Leftrightarrow \text{rg}(A) = n \\ &\Leftrightarrow \text{rg}(B) = n \\ &\Leftrightarrow \det(B) \neq 0 \\ &\Leftrightarrow \det(A) \neq 0. \end{aligned}$$

Satz 5.16: Für $A, B \in K^{n \times n}$ gilt $\det(AB) = \det(A) \det(B)$.

BEWEIS Ist $\det(A) = 0$ oder $\det(B) = 0$, dann ist $\text{rg}(L_A) < n$ oder $\text{rg}(L_B) < n$. Damit ist auch $\text{rg}(L_{AB}) < n$, und die Aussage ist trivial. Seien also A und B invertierbar, d. h. $\text{rg}(A) = \text{rg}(B) = n$. Wir unterstellen jetzt, dass in einem solchen Fall sich eine Matrix mit den Zeilenoperationen des

Gauß-Verfahrens immer in die reduzierte Stufenform überführen lässt (wir beweisen dies in Kapitel 6). Dann gilt $A = C_k \cdot \dots \cdot C_1$ für gewisse Elementarmatrizen C_i , und deswegen nach Folgerung 5.13

$$\begin{aligned} \det(AB) &= \det(C_k \cdot \dots \cdot C_1 B) \\ &= \det(C_k) \det(C_{k-1} \cdot \dots \cdot C_1 B) \\ &= \dots = \det(C_k) \dots \det(C_1) \det(B) \\ &= \dots = \det(C_k \cdot \dots \cdot C_1) \det(B) \\ &= \det(A) \det(B). \end{aligned}$$

Folgerung 5.17: Ist A invertierbar, dann gilt

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

BEWEIS Es gilt $\det(A) \det(A^{-1}) = \det(AA^{-1}) = \det(E) = 1$. ■

5.4 Verfahren zur Berechnung der Determinante

5.4.1 Leibniz-Formel

Wir wollen nun eine explizite Formel zur Berechnung der Determinante entwickeln und beginnen mit einer 2×2 -Matrix $A = (a, b)$. Sei

$$|A| = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}.$$

Mit den kanonischen Einheitsvektoren e_1 und e_2 gilt

$$a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = a_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + a_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = a_1 e_1 + a_2 e_2$$

und daher

$$\det(a, b) = \det(a_1 e_1 + a_2 e_2, b) = a_1 \det(e_1, b) + a_2 \det(e_2, b)$$

wegen der Additivität und Homogenität der Determinante. Analog erhält man für b :

$$\begin{aligned} \det(a, b) &= a_1 (b_1 \det(e_1, e_1) + b_2 \det(e_1, e_2)) + \\ &\quad a_2 (b_1 \det(e_2, e_1) + b_2 \det(e_2, e_2)) \\ &= a_1 b_1 \underbrace{\det(e_1, e_1)}_{=0} + a_1 b_2 \underbrace{\det(e_1, e_2)}_{=1} + \\ &\quad a_2 b_1 \underbrace{\det(e_2, e_1)}_{=-1} + a_2 b_2 \underbrace{\det(e_2, e_2)}_{=0}. \end{aligned}$$

Zwei der Determinanten besitzen den Wert 0, weil die Matrizen gleiche Spalten aufweisen. Wegen $\det(E) = 1$ und des Vorzeichenwechsels bei Vertauschung von Spalten gilt $\det(e_1, e_2) = 1$, $\det(e_2, e_1) = -1$ und damit

$$|A| = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = a_1 b_2 - a_2 b_1.$$

Für $n = 3$ erhält man durch analoge Rechnung die *Sarrussche Regel*. (vgl. Kap. 2.3). Entsprechend kann man auch für $n \times n$ -Matrizen vorgehen.

1. Man schreibt a_1 bis a_n als $a_i = \sum_{k=1}^n a_{ki} e_k$.
2. Man nutzt Additivität der Determinante für alle Spalten und erhält eine Summe aus n^n Summanden. Jeder Summand besitzt n Vorfaktoren (je einen Vorfaktor aus jeder Spalte) und eine Determinante aus Einheitsvektoren.
3. Alle Summanden, die doppelte Einheitsvektoren in der Determinante haben, fallen weg. Es bleiben nur noch Summanden, die Permutationen von (e_1, \dots, e_n) in der Determinante haben.
4. Die Determinanten sind 1 oder -1, je nachdem, ob sie sich durch eine gerade oder eine ungerade Anzahl von Vertauschungen in $\det(E)$ überführen lassen.

Wir gehen kurz auf Permutationen näher ein.

Definition 5.18: Eine bijektive Abbildung $\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ heißt *Permutation*.

Beispiel 5.19:

1. Die identische Abbildung $\sigma(i) = i \forall 1 \leq i \leq n$ ist eine Permutation.
2. Die Inverse einer Permutation σ ist wiederum eine Permutation wegen $(\sigma^{-1})^{-1} = \sigma$.
3. Sind σ und ν Permutationen, dann auch $\sigma \circ \nu$, weil Verkettungen bijektiver Funktionen wieder bijektiv sind.

Bemerkung 5.20: Die Menge aller Permutationen auf $\{1, \dots, n\}$ bilden mit der Abbildungsverkettung eine *Gruppe*; diese Gruppe wird die *symmetrische Gruppe* S_n genannt, ist für $n > 2$ nicht abelsch und besitzt $n!$ Elemente.

BEWEIS Die Gruppeneigenschaft ist nach Beispiel 5.19 klar. Weil eine Permutation injektiv sein muss, muss jedes $\sigma(i) \neq \sigma(j)$ für $i \neq j$ gelten. Man hat n mögliche Bilder für 1, nämlich $1, \dots, n$, aber nur $n - 1$ mögliche Bilder für 2, weil ja das Bild von 1 nicht gewählt werden darf. So fährt man fort, bis für n noch genau ein mögliches Bild bleibt; insgesamt sind dies $n(n - 1) \cdots 1 = n!$ Möglichkeiten. ■

Man erhält als Ergebnis die Leibniz-Formel²

Satz 5.21 (Leibniz-Formel): Für $A \in K^{n \times n}$ gilt

$$\det(A) = \sum_{\sigma \in S_n} \delta(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)}$$

mit

$$\delta(\sigma) = \begin{cases} 1, & \text{falls } \sigma(1), \dots, \sigma(n) \text{ durch eine gerade Zahl von Vertauschungen zu } 1, 2, \dots, n \text{ umgeordnet werden kann} \\ -1, & \text{sonst} \end{cases}$$

Durch die Angabe einer konkreten Berechnungsformel ist die Frage nach der Existenz der Determinante jetzt klar. Für den Nachweis, dass die Leibnizsche Formel in der Tat die definierenden Eigenschaften der Determinante aufweist und zur Eindeutigkeit verweisen wir auf einschlägige Literatur wie z. B. Beutelspacher [2] oder Fischer [3].

Es folgt der angekündigte Beweis von $|A| = |A^T|$ (Satz 5.10).

BEWEIS Sei $\hat{a}_{ij} = a_{ji}$. Wir nutzen in der folgenden Rechnung ohne Beweis, dass $\delta(\sigma^{-1}) = \delta(\sigma)$ gilt und erhalten mit der Leibnizschen Formel

$$\begin{aligned} \det(A^T) &= \sum_{\sigma \in S_n} \delta(\sigma) \hat{a}_{1\sigma(1)} \cdots \hat{a}_{n\sigma(n)} \\ &= \sum_{\sigma \in S_n} \delta(\sigma) a_{\sigma(1)1} \cdots a_{\sigma(n)n} \\ &= \sum_{\sigma \in S_n} \delta(\sigma^{-1}) a_{1\sigma^{-1}(1)} \cdots a_{n\sigma^{-1}(n)} \\ &= \det(A), \end{aligned}$$

weil man sowohl bei Indizierung mit σ als auch mit σ^{-1} alle Elemente von S_n durchläuft. ■

Man beachte, dass zum Beweis keine Eigenschaften der Determinante benutzt wurden, die man mit Hilfe von Satz 5.10 gezeigt hat.

5.4.2 Laplacescher Entwicklungssatz

Wir geben nun eine weitere Möglichkeit an, die Determinante zu berechnen. Für Herleitung und Hintergründe verweisen wir auf die Literatur [2, 3].

Definition 5.22: Für $A \in K^{n \times n}$ bezeichne A_{ij} die Matrix in $K^{(n-1) \times (n-1)}$, die durch Streichen der i -ten Zeile und der j -ten Spalte aus A hervorgeht.

²Gottfried Wilhelm Leibniz (ohne „t“ im Namen; 1646–1716), dt. Universalgelehrter

Satz 5.23 (Laplacescher Entwicklungssatz): Es sei $A = (a_{ij}) \in K^{n \times n}$ und j mit $1 \leq j \leq n$. Dann gilt

$$\det(A) = \sum_{i=1}^n (-1)^{i+j} a_{ij} \det(A_{ij}).$$

Man spricht von der *Entwicklung der Determinante nach der j -ten Spalte*. Ebenso ist eine *Entwicklung der Determinante nach einer Zeile* möglich:

$$\det(A) = \sum_{j=1}^n (-1)^{i+j} a_{ij} \det(A_{ij}).$$

Diese Berechnungsformel für eine Determinante kann auch als rekursive Definition der Determinante aufgefasst werden.

Beispiel 5.24: Wir berechnen die Determinante durch Entwicklung nach der 1. Zeile.

$$\begin{aligned} \begin{vmatrix} 1 & 0 & 2 \\ 2 & 4 & -1 \\ 2 & 3 & -2 \end{vmatrix} &= 1 \cdot \begin{vmatrix} 4 & -1 \\ 3 & -2 \end{vmatrix} - 0 \cdot \begin{vmatrix} 2 & -1 \\ 2 & -2 \end{vmatrix} + 2 \cdot \begin{vmatrix} 2 & 4 \\ 2 & 3 \end{vmatrix} \\ &= 1(-8 + 3) - 0(-4 + 2) + 2(6 - 8) = -9 \end{aligned}$$

5.4.3 Gauß-Algorithmus

Zur Berechnung mit dem Gauß-Algorithmus bringt man die gegebene Matrix A mittels äquivalenter Zeilen- oder Spaltenumformungen Z1 – Z3 bzw. S1 – S3 auf Stufenform B und errechnet dann nach Folgerung 5.14 $\det(A)$ leicht als Produkt der Hauptdiagonalelemente von B . Operationen vom Typ Z2 bzw. S2 verändern aber das Vorzeichen der Determinante; das muss man bei der Berechnung ausgleichen, ebenso multipliziert sich der Wert der Determinante bei Anwendung von Z3 oder S3 mit dem Faktor der Skalierung.

Beispiel 5.25: Es gilt

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \frac{1}{2} \begin{vmatrix} 2a & b \\ 2c & d \end{vmatrix}$$

und

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = - \begin{vmatrix} c & d \\ a & b \end{vmatrix} = - \begin{vmatrix} b & a \\ d & c \end{vmatrix}.$$

Beispiel 5.26:

$$\begin{aligned} \begin{vmatrix} 1 & 0 & 2 \\ 2 & 4 & -1 \\ 2 & 3 & -2 \end{vmatrix} &\stackrel{\text{(II)} \rightarrow \text{(II)} - 2\text{(I)}}{\stackrel{\text{(III)} \rightarrow \text{(III)} - 2\text{(I)}}{=}} \begin{vmatrix} 1 & 0 & 2 \\ 0 & 4 & -5 \\ 0 & 3 & -6 \end{vmatrix} = \frac{3}{4} \begin{vmatrix} 1 & 0 & 2 \\ 0 & 4 & -5 \\ 0 & 4 & -8 \end{vmatrix} \\ &\stackrel{\text{(III)} \rightarrow \text{(III)} - \text{(II)}}{=} \frac{3}{4} \begin{vmatrix} 1 & 0 & 2 \\ 0 & 4 & -5 \\ 0 & 0 & -3 \end{vmatrix} = \frac{3}{4} 1 \cdot 4 \cdot (-3) = -9 \end{aligned}$$

Beispiel 5.27: Die Kombination von Zeilen- und Spaltenoperationen ist möglich und kann bei geschickter Anwendung die Berechnung sehr vereinfachen. Wir bezeichnen mit Z Zeilen- und mit S Spaltenoperationen; es sei $t \in \mathbb{R}$ ein Parameter.

$$\begin{aligned} \begin{vmatrix} 1 & 1+t & 1 \\ 1 & 1 & 1+t \\ 1+t & 1 & 1 \end{vmatrix} &\stackrel{S1 \rightarrow S1 \pm S2 + S3}{=} \begin{vmatrix} 3+t & 1+t & 1 \\ 3+t & 1 & 1+t \\ 3+t & 1 & 1 \end{vmatrix} \\ &\stackrel{Z3 \rightarrow Z3 - Z2}{\stackrel{Z2 \rightarrow Z2 - Z1}{=}} \begin{vmatrix} 3+t & 1 & 1 \\ 0 & -t & t \\ 0 & 0 & -t \end{vmatrix} = t^2(3+t) \end{aligned}$$

5.4.4 Determinantenberechnung in der Praxis

Welches Rechenverfahren soll man denn nun in der Praxis wählen, wenn die Determinante einer konkreten Matrix A berechnet werden muss? Die Antwort hängt u. a. davon ab, ob man mit der Hand rechnet oder einen Computer verwendet. Üblicherweise lassen sich Determinanten für $n > 5$ kaum noch sinnvoll per Hand berechnen, wenn nicht gerade die spezielle Struktur der Matrix die Berechnung stark vereinfacht. Für die Handrechnung lassen sich nur Erfahrungswerte angeben, weil ja die Wahl des Rechenverfahrens auch von Gewohnheit und persönlichen Vorlieben abhängt. Folgendes lehrt die Praxis:

1. Die Leibnizsche Formel ist als Rechenverfahren nicht konkurrenzfähig, wir raten von ihrer Verwendung für konkrete Berechnungen ab.
2. Es gibt zur ersten Faustregel eine Ausnahme: Für 2×2 - und für 3×3 -Matrizen empfehlen sich besonders das aus Kapitel 2.3 bekannte explizite Rechenschema für $n = 2$ und die Sarrus-Regel. Dies sind natürlich Spezialfälle der Leibnizschen Formel.
3. Grundsätzlich sind sowohl die Laplace-Entwicklung als auch die Methode nach Gauß für praktische Rechnungen gut geeignet. Die

Laplace-Entwicklung ist dann vorzuziehen, wenn in einer Spalte oder Zeile nur wenige Nicht-Null-Einträge vorhanden sind, weil bei einer Entwicklung nach dieser Zeile/Spalte die meisten Summanden erst gar nicht berechnet werden müssen; sie haben ohnehin den Wert 0.

4. Es können zur Berechnung der Determinanten mehrere Verfahren kombiniert werden. Für 4×4 -Matrizen bietet sich z. B. ein Schritt der Laplace-Entwicklung an, die dabei entstehenden Determinanten von 3×3 -Matrizen werden direkt mit der Sarrus-Regel berechnet. Sinnvoll kann es auch sein, sich zunächst mit (unvollständiger) Gauß-Elimination eine Zeile oder Spalte mit wenigen Nicht-Null-Einträgen zu schaffen und dann mit einer Laplace-Entwicklung fortzusetzen.
5. Übung und Routine sind für eine sichere und schnelle Handrechnung nicht zu ersetzen.

Beispiel 5.28: Mit der Operation „ $S_3 \rightarrow S_3 - 2S_1$ “ wird für die Matrix aus Beispiel 5.24 erreicht, dass nur eine einzige Zahl $\neq 0$ in der ersten Zeile steht. Es folgt eine Laplace-Entwicklung nach der ersten Zeile.

$$\begin{vmatrix} 1 & 0 & 2 \\ 2 & 4 & -1 \\ 2 & 3 & -2 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 2 & 4 & -5 \\ 2 & 3 & -6 \end{vmatrix} = -24 + 15 = -9$$

Wir kommen zur Berechnung mit dem Computer. Als (grobes) Maß für den Rechenaufwand und die Rechenzeit wählen wir die Anzahl der arithmetischen Operationen.

Beim Gauß-Verfahren werden durch Zeilenoperationen, die den Wert der Determinanten erhalten, unterhalb der Diagonalen Nullen erzeugt. Eine Analyse ergibt, dass dazu etwa $\frac{n^3-n}{3}$ Operationen erforderlich werden. Der Aufwand der Laplace-Entwicklung beträgt etwa $n!$, weil man zunächst eine Summe von n Determinanten von $(n-1) \times (n-1)$ -Matrizen erhält, die jeweils auf die Berechnung einer Summe von $n-1$ Determinanten von $(n-2) \times (n-2)$ -Matrizen zurückgeführt werden usw. Die Leibnizsche Formel besteht aus $n!$ Summanden von Produkten von n Faktoren, zur Berechnung sind daher $(n-1)n!$ arithmetische Operationen nötig.

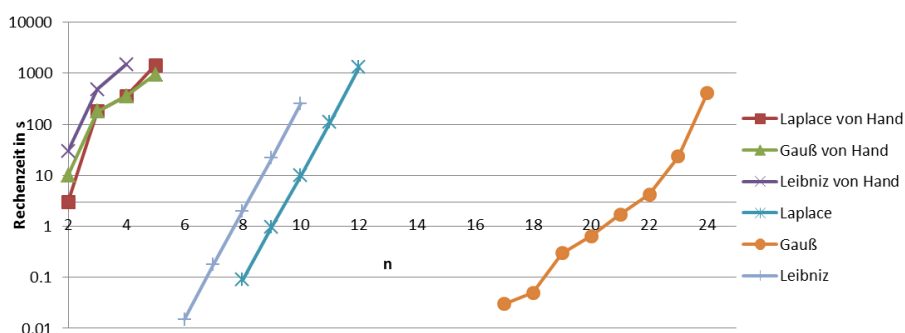
Wir veranschaulichen das Wachstumsverhalten der drei Methoden anhand der zu erwartenden Rechenzeit t und der Anzahl von arithmetischen Operationen Ops in Tabelle 5.1. Wir gehen davon aus, dass der verwendete Computer 1 GFlop/s Rechenleistung aufweist, also 10^9 Multiplikationen oder Additionen pro Sekunde durchführen kann. Man erkennt den verheerenden Anstieg von t in Millisekunden (ms), Minuten (m), Stunden (h) oder Jahren (a) bei wachsender Problemgröße bei der Leibnizschen Formel und der Laplace-Entwicklung. Daher eignet sich von den drei vorgestellten Verfahren allein das Gauß-Verfahren zur Berechnung von Determinanten größerer Matrizen.

5.5. EXKURS: INVERTIERUNG VON MATRIZEN MITTELS UNTERDETERMINANTEN 175

n	5	10	15	20	50	100
Ops	40	330	1,1E3	2,7E3	4,2E4	3,3E5
t	< 1 ms	< 1 ms	< 1 ms	< 1 ms	< 1 ms	< 1 ms
Ops	120	3,6E6	1,3E12	2,4E18	3,0E64	9,3E157
t	< 1 ms	3,6 ms	22 m	77 a	> 10^{12} a	> 10^{12} a
Ops	480	3,3E7	1,8E13	4,6E19	1,5E66	9,2E159
t	< 1 ms	33 ms	5,1 h	1,4E3 a	> 10^{12} a	> 10^{12} a

Tabelle 5.1: Schematischer Vergleich von Rechenzeiten für verschiedene Verfahren zur Determinantenberechnung (oben: Gauß-Verfahren, Mitte: Laplace-Entwicklung, unten: Leibnizsche Formel)

Beispiel 5.29: Für einen realistischen Vergleich wurden die drei obigen Verfahren zur Berechnung der Determinante in python implementiert. Obwohl sich python nicht unbedingt durch herausragende Performanz auszeichnet, zeigen die (nur bedingt) repräsentativen Laufzeitvergleiche den Rechenzeitgewinn gegenüber Handrechnungen, die für den vorliegenden Vergleich von einer Scientific-Programming-Studentin durchgeführt wurden. Man beachte im Diagramm die logarithmische Skalierung.



5.5 Exkurs: Invertierung von Matrizen mittels Unterdeterminanten

Wir stellen nun eine Methode zur Berechnung der inversen Matrix vor, die auf der Berechnung von Unterdeterminanten beruht und sich damit grundsätzlich vom Gauß-Verfahren unterscheidet.

Definition 5.30: Unter der zu A komplementären Matrix \tilde{A} versteht man die Matrix mit den Elementen

$$\tilde{a}_{ij} = (-1)^{i+j} \cdot \det(A_{ji}).$$

Satz 5.31: Sei $A \in GL(n; K)$. Dann gilt

$$A^{-1} = \frac{1}{\det(A)} \tilde{A}.$$

Für den Beweis verweisen wir auf z. B. Fischer [3, Kap.3.3].

Beispiel 5.32: Wir invertieren die Matrix

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

aus Beispiel 5.24 mittels Satz 5.31 und gehen dazu in mehreren Schritten vor.

1. Schritt: Wir stellen eine Hilfsmatrix $B = (b_{ij})$ auf mit $b_{ij} = |A_{ij}|$.

$$B = \begin{pmatrix} \left| \begin{array}{cc|cc|cc} 4 & -1 & 2 & -1 & 2 & 4 \\ 3 & -2 & 2 & -2 & 2 & 3 \end{array} \right| \\ \left| \begin{array}{cc|cc|cc} 0 & 2 & 1 & 2 & 1 & 0 \\ 3 & -2 & 2 & -2 & 2 & 3 \end{array} \right| \\ \left| \begin{array}{cc|cc|cc} 0 & 2 & 1 & 2 & 1 & 0 \\ 4 & -1 & 2 & -1 & 2 & 4 \end{array} \right| \end{pmatrix} = \begin{pmatrix} -5 & -2 & -2 \\ -6 & -6 & 3 \\ -8 & -5 & 4 \end{pmatrix}$$

2. Schritt: Wir ändern die Vorzeichen der Einträge von B nach dem Schachbrettmuster und erhalten eine weitere Hilfsmatrix C .

$$C = \begin{pmatrix} -5 & 2 & -2 \\ 6 & -6 & -3 \\ -8 & 5 & 4 \end{pmatrix}$$

3. Schritt: Man berechnet $\tilde{A} = C^T$.

$$\tilde{A} = \begin{pmatrix} -5 & 6 & -8 \\ 2 & -6 & 5 \\ -2 & -3 & 4 \end{pmatrix}$$

4. Schritt: Wir berechnen $A^{-1} = |A|^{-1} \tilde{A}$. Mit $|A| = -9$ (vgl. Bsp. 5.24) folgt

$$A^{-1} = \frac{1}{9} \begin{pmatrix} 5 & -6 & 8 \\ -2 & 6 & -5 \\ 2 & 3 & -4 \end{pmatrix}.$$

Wir nutzen jetzt Satz 5.31, um eine explizite Formel für die Inverse einer 2×2 -Matrix herzuleiten. Die Anwendung der obigen Rechenschritte auf

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad (5.1)$$

ergibt

$$B = \begin{pmatrix} d & c \\ b & a \end{pmatrix} \Rightarrow C = \begin{pmatrix} d & -c \\ -b & a \end{pmatrix} = \tilde{A}^T \Rightarrow \tilde{A} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

und damit folgenden Satz.

Satz 5.33: Sei A wie in Formel (5.1) und invertierbar. Dann gilt

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Beispiel 5.34:

$$A = \begin{pmatrix} 5 & 2 \\ 4 & 1 \end{pmatrix} \Rightarrow A^{-1} = -\frac{1}{3} \begin{pmatrix} 1 & -2 \\ -4 & 5 \end{pmatrix}$$

In der Rechenpraxis ist der Gauß-Algorithmus dem hier gezeigten Verfahren zur Berechnung der Inversen mindestens ebenbürtig. Dennoch lassen sich einige interessante Resultate aus dem hier gezeigten Verfahren ableiten.

1. Ist man in der Lage, eine Determinante zu berechnen, ist man auch fähig, ein lineares Gleichungssystem zu lösen.
2. Das Gauß-Verfahren ist mit allen seinen Varianten nicht die einzige Verfahrensklasse zum Lösen von linearen Gleichungssystemen, es existieren auch grundsätzlich andersartige Verfahren.

Wir diskutieren eine dritte Konsequenz. In der Praxis kommt es vor, dass die Matrix A eines linearen Gleichungssystems $Ax = b$ nicht exakt bekannt ist, weil z. B. Koeffizienten von A aus Messungen bestimmt werden, die ja immer fehlerbehaftet sind. Stattdessen liegt eine gestörte Version \hat{A} vor. Ist es überhaupt unter diesen Umständen sinnvoll, ein lineares Gleichungssystem zu lösen, weil man ja nicht hoffen kann, die exakte Lösung x zu erhalten, sondern nur eine gestörte Version \hat{x} ? Im Prinzip ja, wenn man gewährleisten kann, dass $\|x - \hat{x}\| < \varepsilon$ erreicht werden kann. Dabei gibt der Anwender je nach erforderlicher Genauigkeit eine Fehlerschranke ε vor. Der Wert von ε wird implizit festlegen, wie genau A bekannt sein muss. Fatal wäre es dagegen, wenn schon winzigste Störungen in A zu erheblichen Abweichungen in x führten, weil das eine zuverlässige Berechnung von x unmöglich machte. Ist die rechte Seite b bekannt, gilt $\hat{x} = \hat{A}^{-1}b$ und $x = A^{-1}b$, also $x - \hat{x} = (A^{-1} - \hat{A}^{-1})b$. Prinzipielle Berechenbarkeit in der obigen Lage setzt also voraus, dass $\|A^{-1} - \hat{A}^{-1}\|$ beliebig klein wird, wenn man nur $\|A - \hat{A}\|$ klein genug wählt.

Satz 5.35:

1. Die Determinante $\det : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}$ ist stetig.
2. Sei $A \in GL(n, \mathbb{C})$ und seien alle Matrizen in einer Umgebung von A ebenfalls regulär. Dann ist $\Psi : GL(n; \mathbb{C}) \rightarrow GL(n; \mathbb{C}), \Psi(B) = B^{-1}$ stetig im Punkt A , und es gilt

$$\forall \varepsilon > 0 \exists \delta > 0 : \|A - \hat{A}\| < \delta \Rightarrow \|A^{-1} - \hat{A}^{-1}\| < \varepsilon. \quad (5.2)$$

BEWEIS 1. Die Abbildung $f_{ij}(A) := a_{ij}$ ist offenbar linear und deswegen stetig. Daher ist für jede Permutation σ die Funktion $f_\sigma(A) = a_{1\sigma(1)} \cdot \dots \cdot a_{n\sigma(n)}$ stetig als Produkt stetiger Funktionen vom Typ f_{ij} . Nach der Leibnizschen Formel ist dann \det stetig als Summe stetiger Funktionen vom Typ f_σ .

2. Weil die Inverse eindeutig ist, kann man das Berechnungsverfahren für die Inverse frei wählen; alle Verfahren liefern ja dasselbe Ergebnis. Wir wählen die Berechnung nach Satz 5.31. Es ist $f(A) := (\det(A))^{-1}$ stetig in A nach 1. und wegen $\det \neq 0$ in einer Umgebung von A . Aus 1. und der Stetigkeit der Norm folgt die Stetigkeit von $g : A \rightarrow \tilde{A}$ und damit die Stetigkeit von $\Psi = f \cdot g$ als Produkt stetiger Funktionen. Gleichung (5.2) entspricht genau der Definition der Stetigkeit von Ψ . ■

Kapitel 6

Lineare Gleichungssysteme

6.1 Lösbarkeit eines linearen Gleichungssystems

Wir verallgemeinern die Definitionen aus Kapitel 1.5. Sei dazu stets K ein Körper.

Definition 6.1: Seien $A = (a_{ij}) \in K^{m \times n}$ und $b = (b_1, \dots, b_m)^T \in K^m$. Dann heißt

$$\begin{array}{r} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ \vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = b_m \end{array}$$

lineares Gleichungssystem bzgl. (x_1, \dots, x_n) mit Koeffizienten a_{ij} in K . Hierbei sind x_1, \dots, x_n die *Unbekannten* des Systems. Für $b = 0$ nennt man das lineare Gleichungssystem *homogen*, sonst *inhomogen*.

Bemerkung 6.2: Jedes lineare Gleichungssystem kann in der Form $Ax = b$ geschrieben werden.

Definition 6.3: In der Situation von Definition 6.1 ist die *Lösungsmenge* $L(A, b)$ des zu (A, b) gehörigen Gleichungssystems festgelegt durch

$$L(A, b) := \{x \in K^n \mid Ax = b\}.$$

Wir werden nun einfache Kriterien zur Lösbarkeit eines beliebigen linearen Gleichungssystems angeben. Dazu benötigen wir den *Rang* einer Matrix, den wir auf den bereits bekannten Rang einer linearen Abbildung (vgl. Definition 4.21) zurückführen. Sei im folgenden immer $A = (a_1, \dots, a_n) \in K^{m \times n}$.

Definition 6.4: Die lineare Abbildung $L_A : K^n \rightarrow K^m$ sei gegeben durch $L_A(x) := Ax$. Dann sei $\text{rg}(A) := \text{rg}(L_A)$. Der *Spaltenrang* $\text{rg}_S(A)$ sei die maximale Anzahl linear unabhängiger Spaltenvektoren von A .

Satz 6.5: Es gilt $\text{rg}(A) = \text{rg}_S(A)$.

BEWEIS Für $x = (x_1, \dots, x_n)^T \in K^n$ gilt $Ax = \sum_{i=1}^n x_i a_i$. Es folgt

$$\begin{aligned} \text{rg}(A) &= \text{rg}(L_A) = \dim(\text{Bild}(L_A)) \\ &= \dim(\{L_A(x) \mid x \in K^n\}) \\ &= \dim(\{Ax \mid x \in K^n\}) \\ &= \dim\left(\left\{\sum_{i=1}^n x_i a_i \mid x_i \in K\right\}\right) \\ &= \dim(L(a_1, \dots, a_n)) = \text{rg}_S(A) \end{aligned}$$

Satz 6.6: Das lineare Gleichungssystem $Ax = b$ ist genau dann lösbar, wenn $\text{rg}(a_1, \dots, a_n) = \text{rg}(a_1, \dots, a_n, b)$ gilt. Kürzer schreibt man

$$\text{rg}(A) = \text{rg}(A, b).$$

BEWEIS Man hat $Ax = (a_1, \dots, a_n)x = x_1 a_1 + \dots + x_n a_n = b$. Also existiert ein Lösungsvektor x genau dann, wenn

$$b \in L(a_1, \dots, a_n) \Leftrightarrow \text{rg}(A) = \text{rg}(A, b)$$

gilt, wobei $L(a_1, \dots, a_n)$ für die lineare Hülle von a_1, \dots, a_n steht. ■

Satz 6.7: Sei $x_s \in K^n$ Lösung von $Ax = b$. Dann gilt

$$L(A, b) = x_s + \ker(A) = \{x_s + x \mid x \in \ker(A)\}.$$

BEWEIS

$$x \in \ker(A) \Leftrightarrow Ax = 0 \Leftrightarrow A(x_s + x) = Ax_s + Ax = Ax_s = b \quad \blacksquare$$

Bemerkung 6.8: Sei x_s eine Lösung von $Ax = b$ und (v_1, \dots, v_r) eine Basis von $\ker(A)$. Dann gilt

$$L(A, b) = \{x_s + \lambda_1 v_1 + \dots + \lambda_r v_r \mid \lambda_i \in K\}.$$

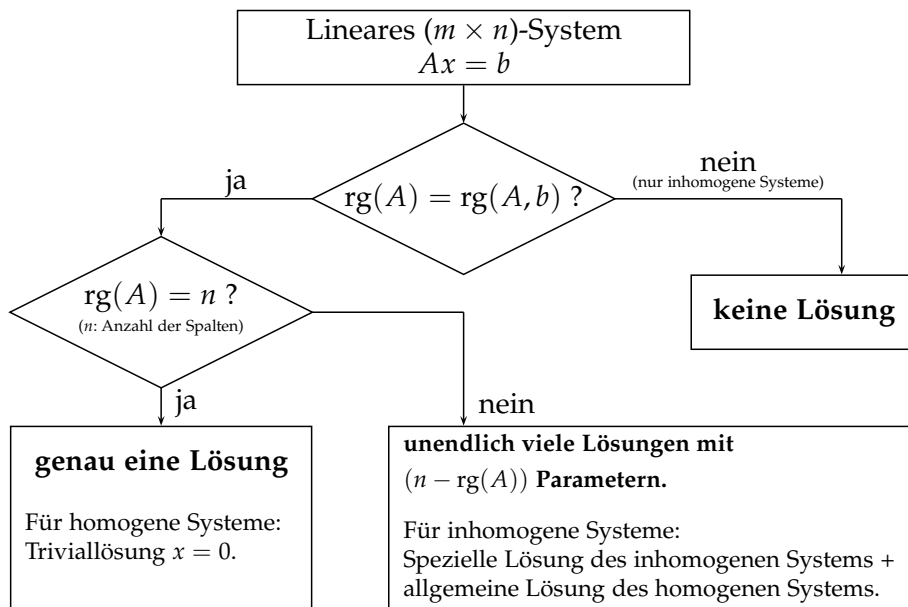
Wegen der Dimensionsformel Satz 4.23 gilt $r = \dim(\ker(A)) = n - \text{rg}(A)$.

Gelegentlich spricht man von einer *speziellen Lösung*, wenn ein konkreter Lösungsvektor angegeben wird und von einer *allgemeinen Lösung*, wenn der Lösungsvektor maximal viele unabhängig voneinander frei wählbare Parameter enthält (und damit die Lösungsmenge parametrisiert). Mit dieser Begriffsbildung gewinnt man aus Bemerkung 6.8 folgenden Merksatz:

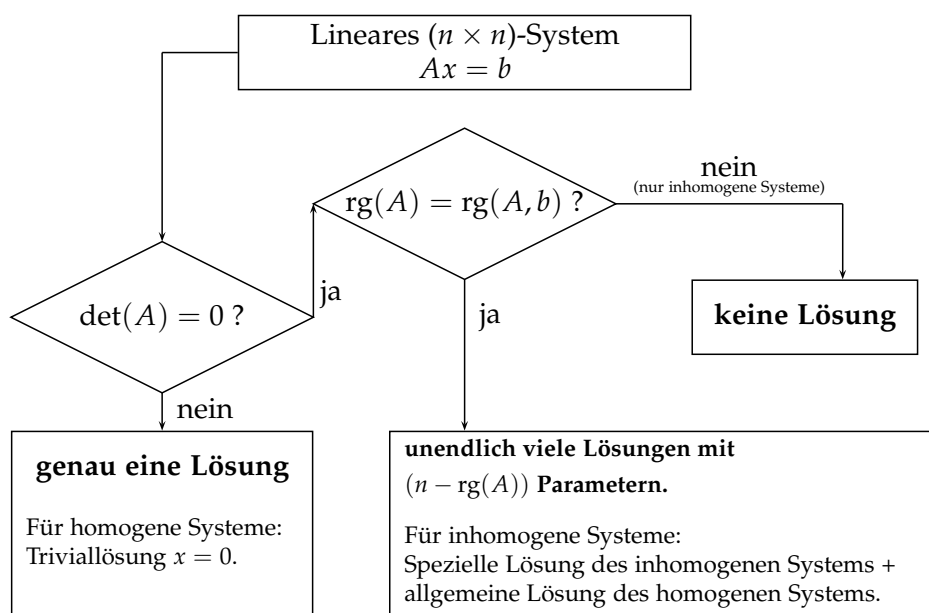
Ist x_s eine *spezielle Lösung des inhomogenen Gleichungssystems* $Ax = b$ und ist x_h die *allgemeine Lösung des homogenen Systems* $Ax = 0$, dann ist $x_a = x_h + x_s$ die *allgemeine Lösung des inhomogenen Gleichungssystems* $Ax = b$.

Folgerung 6.9: Das lineare Gleichungssystem $Ax = b$ sei lösbar. Es ist genau dann eindeutig lösbar, falls $\ker(A) = \{0\}$, d. h. $\operatorname{rg}(A) = n$ gilt, also genau dann, wenn $\operatorname{rg}(A) = \operatorname{rg}(A, b) = n$ ist.

Folgerung 6.10: Sei die Abbildung L_A gegeben durch $L_A(x) = Ax$. Dann ist $Ax = b$ genau dann lösbar, wenn $b \in \operatorname{Bild}(L_A)$ gilt.



Bemerkung 6.11: Für quadratisches A ist das Gleichungssystem $Ax = b$ genau dann eindeutig lösbar, wenn $|A| \neq 0$ gilt.



Beispiel 6.12: Gegeben sei

$$A = \begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix}; \quad b = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Es ist $n = 2$. Aus $|A| = 5$ folgt $\text{rg}(A) = \text{rg}(A, b) = 2$. Daher existiert eine eindeutige Lösung.

Beispiel 6.13:

$$A = \begin{pmatrix} 1 & 2 & 1 \\ -1 & 3 & 1 \end{pmatrix}; \quad b = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Es ist $n = 3$. Da die beiden Zeilen keine Vielfachen voneinander sind, ist $\text{rg}(A) = \text{rg}(A, b) = 2$, und es gibt unendlich viele Lösungen.

Beispiel 6.14:

$$A = \begin{pmatrix} 1 & -1 & 2 \\ 2 & -2 & 4 \end{pmatrix}; \quad b = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Es ist $n = 3$. Die zweite Zeile von A entspricht dem Doppelten der ersten Zeile von A , also gilt $\text{rg}(A) = 1$. Die beiden Zeilen von (A, b) sind keine Vielfachen voneinander, es folgt $\text{rg}(A, b) = 2$. Es existiert keine Lösung.

Beispiel 6.15:

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

Es ist $n = 2$. Die beiden Spalten sind keine Vielfachen voneinander, also gilt $\text{rg}(A) = 2$. Da $\text{rg}(A, b) \geq \text{rg}(A)$ und $\text{rg}(A, b) \leq (n + 1)$, gilt $\text{rg}(A, b) \in \{2, 3\}$. Es ist $\text{rg}(A, b) = 3$ genau dann, wenn $\det(A, b) \neq 0$. Man errechnet

$$\det(A, b) = \begin{vmatrix} 1 & 3 & -2 \\ -1 & 1 & -2 \\ 0 & 1 & -1 \end{vmatrix} = -1 + 2 - (-2 + 3) = 0,$$

und es folgt $\text{rg}(A, b) = 2$. Es existiert eine eindeutige Lösung.

Bemerkung 6.16: Man muss zwei verschiedene Situationen klar trennen:

1. Ein Gleichungssystem $Ax = b$ mit $A \in K^{n \times n}$ ist genau dann eindeutig lösbar, wenn $\det(A) \neq 0$.
2. Sei $Ax = b$ gegeben mit
 - (a) $A \in K^{(n+1) \times n}$ (eine Zeile mehr als Spalten) und

(b) die Spalten von A sind linear unabhängig.

Dann ist $Ax = b$ genau dann eindeutig lösbar, wenn $\det(A, b) = 0$ ist.

Beispiel 6.17: Wir betrachten die Matrix A aus Beispiel 6.15 zusammen mit $b = (0, 0, 1)^T$. Wiederum gilt $n = 2$ und $\text{rg}(A) = 2$ und weiter $2 \leq \text{rg}(A, b) \leq 3$. Mit einer Laplace-Entwicklung berechnen wir

$$\det(A, b) = \begin{vmatrix} 1 & 3 & 0 \\ -1 & 1 & 0 \\ 0 & 1 & 1 \end{vmatrix} = 1 \begin{vmatrix} 1 & 3 \\ -1 & 1 \end{vmatrix} = 4 \neq 0$$

und erhalten $\text{rg}(A, b) = 3$. Daher existiert keine Lösung.

Es sollen jetzt sämtliche äquivalente Bedingungen für die eindeutige Lösbarkeit eines linearen Gleichungssystems angegeben werden.

Satz 6.18: Sei $K \in \{\mathbb{R}, \mathbb{C}\}$. Für $A \in K^{n \times n}$ und die dadurch gegebene lineare Abbildung L_A sind folgende Bedingungen äquivalent:

1. A ist invertierbar.
2. $Ax = 0$ hat nur die triviale Lösung $x = 0$.
3. Durch Zeilen- und Spaltenumformungen kann A auf die Einheitsmatrix transformiert werden.
4. A ist darstellbar als Produkt von Elementarmatrizen.
5. $Ax = b$ besitzt für jedes $b \in K^n$ mindestens eine Lösung.
6. $Ax = b$ hat genau eine Lösung für jedes $b \in K^n$.
7. $\det(A) \neq 0$.
8. $\text{Bild}(L_A) = K^n$.
9. L_A ist bijektiv.
10. Die Spaltenvektoren von A sind linear unabhängig.
11. Die Zeilenvektoren von A sind linear unabhängig.
12. Die Spaltenvektoren von A bilden eine Basis von K^n .
13. Die Zeilenvektoren von A bilden eine Basis von K^n .
14. $\text{rg}(A) = n$.
15. $\ker(L_A) = \{0\}$.

16. $(\ker(L_A))^\perp = K^n$.
17. Das orthogonale Komplement des von den Zeilen von A aufgespannten Raums ist $\{0\}$.
18. $A^T A$ ist invertierbar.

6.2 Das Gaußsche Eliminationsverfahren reloaded

Wir haben bisher das Gauß-Verfahren in vielfältiger Weise angewendet, ein strenger Beweis, dass das Verfahren tatsächlich eine Lösung liefert, wenn eine solche existiert, steht aber aus. Das holen wir jetzt nach. Für eine invertierbare Matrix $A \in K^{n \times n}$ besteht der Gauß-Algorithmus zur Lösung des linearen Gleichungssystems $Ax = b$ aus zwei Schritten:

1. Man formt die erweiterte Matrix (A, b) mit Zeilenoperationen Z1 und Z2 zu einem neuen linearen Gleichungssystem (\tilde{A}, \tilde{b}) um, so dass \tilde{A} eine obere Dreiecksmatrix ist.
2. Man löst das System $\tilde{A}x = \tilde{b}$ durch Rückwärtseinsetzen. Die so gewonnene Lösung x löst $Ax = b$.

Wir zeigen zunächst, dass die Äquivalenzumformungen des Gauß-Verfahrens tatsächlich solche sind, dass also das Gauß-Verfahren keine Lösungen „erfindet oder unterschlägt“. Das gilt sogar für beliebige lineare Gleichungssysteme.

Satz 6.19: Die Zeilenumformungen Z1 – Z3 des Gauß-Verfahrens erhalten die Lösungsmenge.

BEWEIS Wir betrachten das lineare Gleichungssystem $Ax = b$, $A \in K^{m \times n}$ mit irgendeiner Lösung x . Da die rechte Seite im Gauß-Verfahren wie eine Matrixspalte behandelt wird, überführt nach Bemerkung 5.5 eine elementare Zeilenoperation $Ax = b$ in $CAx = Cb$. Dabei bleibt Gleichheit natürlich erhalten; es folgt $L(A, b) \subseteq L(CA, Cb)$. Sei jetzt $x \in L(CA, Cb)$. Dann gilt

$$CAx = Cb \Rightarrow C^{-1}CAx = C^{-1}Cb \Rightarrow Ax = b$$

aufgrund der Invertierbarkeit von Elementarmatrizen (vgl. Bemerkung 5.8), also $L(CA, Cb) \subseteq L(A, b)$.

Wir haben oben die Existenz einer Lösung implizit angenommen. Sei jetzt $L(A, b) = \emptyset$. Wäre dann $L(CA, Cb) \neq \emptyset$, gäbe es ein $x \in L(CA, Cb)$ mit $CAx = Cb$. Multiplikation von links mit C^{-1} liefert dann $Ax = b$ und damit $L(A, b) \neq \emptyset$. Also gilt $L(CA, Cb) = \emptyset$. In der umgekehrten Richtung argumentiert man analog, und die Gleichheit der Lösungsmengen gilt auch im Fall leerer Mengen. ■

Es bleibt noch zu zeigen, dass eine reguläre Matrix mit dem Gauß-Verfahren immer in Stufenform, sogar in die reduzierte Stufenform transformiert werden kann. Zur Vorbereitung diskutieren wir zunächst die Rangerhaltung im Gauß-Verfahren.

Satz 6.20: Die Zeilen- bzw. Spaltenoperationen des Gauß-Verfahrens erhalten den Rang einer Matrix.

BEWEIS Sei C eine Elementarmatrix und $A \in K^{m \times n}$ beliebig. Die Aussage $\text{rg}(CA) = \text{rg}(A)$ folgt aus der Invertierbarkeit von Elementarmatrizen (Bemerkung 5.8) und der Ranggleichung Satz 4.55.

Sei $A = (a_1, \dots, a_n)$. Spaltenvertauschung (S2) erhält $L := L(a_1, \dots, a_n)$ und deswegen auch $\text{rg}_S(A)$. Wegen $L := L(a_1, \dots, \lambda a_i, \dots, a_n)$ erhält auch S3 $\dim(L) = \text{rg}_S(A)$. Da auch $L = L(a_1, \dots, a_j + \lambda a_i, \dots, a_i, \dots, a_n)$ gilt, erhält S1 L und damit $\text{rg}_S(A) = \text{rg}(A)$ nach Satz 6.5. Damit ist $\text{rg}(AC) = \text{rg}(A)$ gezeigt. ■

Wir kommen zur Durchführbarkeit des Gauß-Verfahrens zurück, beschränken uns aber der Einfachheit halber auf den Fall einer regulären Matrix A .

Satz 6.21: Sei $A \in K^{n \times n}$ invertierbar. Dann lässt sich A durch eine Folge von Operationen vom Typ Z1 oder Z2 in eine obere Dreiecksmatrix umformen, wobei kein Hauptdiagonalelement den Wert 0 annimmt.

BEWEIS Wir wollen durch Z1 und Z2 zunächst die Einträge der ersten Spalte unterhalb der Hauptdiagonalen eliminieren, danach jene der zweiten Spalte usw. Wir nehmen an, dies sei bis zur $(k - 1)$ -ten Spalte gelungen. Wir haben also eine Matrix $A^{(\ell)} = C_\ell \cdot \dots \cdot C_1 A$ erzeugt mit

$$A^{(\ell)} = \begin{pmatrix} a_{11}^{(\ell)} & \cdots & a_{1k-1}^{(\ell)} & a_{1k}^{(\ell)} & a_{1k+1}^{(\ell)} & \cdots & a_{1n}^{(\ell)} \\ 0 & \ddots & \vdots & \vdots & \vdots & & \vdots \\ \vdots & \ddots & a_{k-1k-1}^{(\ell)} & a_{k-1k}^{(\ell)} & \vdots & & \vdots \\ 0 & \cdots & 0 & a_{kk}^{(\ell)} & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & a_{nk}^{(\ell)} & a_{nk+1}^{(\ell)} & \cdots & a_{nn}^{(\ell)} \end{pmatrix}.$$

Aufgrund von

$$\text{rg}(A) = \text{rg}(A^{(\ell)}) = n \tag{6.1}$$

muss $a_{ii}^{(\ell)} \neq 0$ für $1 \leq i < k$ gelten; diese Elemente werden im weiteren Verlauf des Gauß-Verfahrens nicht mehr verändert. Daher muss jetzt durch Zeilenumformungen eine neue Matrix $A^{(\mu)}$ erzeugt werden mit $a_{kk}^{(\mu)} \neq 0$

und $a_{ik}^{(\mu)} = 0$ für $i > k$. Ist bereits $a_{kk}^{(\ell)} \neq 0$, dann wird der (i, k) -te Eintrag zu 0, indem man das $a_{ik}^{(\ell)} / a_{kk}^{(\ell)}$ -fache der k -ten von der i -ten Zeile abzieht; so lässt sich die k -te Spalte mit (einer Folge von) Z1 auf die gewünschte Gestalt bringen. Gilt aber $a_{kk}^{(\ell)} = 0$, dann muss man mit einer Zeile unterhalb tauschen. Das ist der einzige kritische Punkt im Algorithmus: Was, wenn es dazu keine geeignete Zeile gibt, also $a_{ik}^{(\ell)} = 0$ für $i \geq k$? Dann hätte $A^{(\ell)}$ die Gestalt

$$A^{(\ell)} = \begin{pmatrix} a_{11}^{(\ell)} & \cdots & a_{1k-1}^{(\ell)} & a_{1k}^{(\ell)} & a_{1k+1}^{(\ell)} & \cdots & a_{1n}^{(\ell)} \\ 0 & \ddots & \vdots & \vdots & \vdots & & \vdots \\ \vdots & \ddots & a_{k-1k-1}^{(\ell)} & a_{k-1k}^{(\ell)} & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & a_{nk+1}^{(\ell)} & \cdots & a_{nm}^{(\ell)} \end{pmatrix}.$$

Dann aber läge die k -te Spalte in der Linearen Hülle der ersten $k - 1$ Spalten, und es gälte $\text{rg}(A^{(\ell)}) \leq n - 1$ im Widerspruch zu $\text{rg}(A^{(\ell)}) = n$ nach (6.1).

■

Mit ähnlichen Argumenten, aber technisch deutlich komplizierter zeigt man folgendes Ergebnis.

Satz 6.22: Sei $A \in K^{m \times n}$ mit $\text{rg}(A) = r$. Durch eine Folge von Zeilen- und Spaltenoperationen der Form Z1 – Z3 bzw. S1 – S3 kann man die Matrix A in eine Matrix der Form

$$E_r = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

umformen. Es gilt $a_{ii} = 1$ für $i \leq r$.

Neben dem *Spaltenrang* einer Matrix nach Definition 6.4 wird häufig auch der *Zeilenrang* einer Matrix betrachtet.

Definition 6.23: Für $A \in K^{m \times n}$ sei die maximale Anzahl linear unabhängiger Zeilenvektoren von A der *Zeilenrang* $\text{rg}_Z(A)$ von A .

Als Anwendung der Theorie zum Gauß-Verfahren zeigen wir die Gleichheit aller drei Rangbegriffe für Matrizen.

Satz 6.24 (Zeilenrang = Spaltenrang): Für $A \in K^{m \times n}$ gilt $\text{rg}(A) = \text{rg}_S(A) = \text{rg}_Z(A)$.

BEWEIS Nach Satz 6.22 gilt $E_r = SAT$, wobei $S = C_k \cdots C_1$ und $T = \tilde{C}_1 \cdots \tilde{C}_\ell$ Produkte von Elementarmatrizen sind. Damit erhält man nach Satz 6.20

$$\text{rg}_S(A) = \text{rg}_S(E_r) = \text{rg}_S(E_r^T) = \text{rg}_S(T^T A^T S^T). \quad (6.2)$$

Transponiert man eine Elementarmatrix, erhält man erneut eine Elementarmatrix vom selben Typ. Deswegen sind $S^T = C_1^T \cdots C_k^T$ und $T^T = \tilde{C}_\ell^T \cdots \tilde{C}_1^T$ Produkte von Elementarmatrizen, und wieder nach Satz 6.20 gilt $\text{rg}_S(A^T) = \text{rg}_S(E_r)$. Offenbar gilt $\text{rg}_Z(A) = \text{rg}_S(A^T)$. Damit folgt $\text{rg}_S(A) = \text{rg}_S(A^T) = \text{rg}_Z(A)$. ■

Satz 6.22 liefert ein Rechenverfahren zur Rangbestimmung einer Matrix. Man bringt die Matrix mittels Zeilen- und Spaltenumformungen auf die Gestalt E_r und liest dann den Rang anhand der Anzahl der vorkommenden Nicht-Nulleinträge ab.

Beispiel 6.25: Wir notieren Zeilenoperationen in schwarz und Spaltenoperationen in blau.

$$\begin{aligned} & \text{rg} \begin{pmatrix} 1 & 2 & -3 & 3 \\ 3 & -1 & 5 & 2 \\ 2 & 1 & 0 & 3 \\ 1 & 0 & 1 & 1 \end{pmatrix} \begin{array}{l} \text{(II)} \rightarrow \text{(II)} - 3\text{(I)} \\ \text{(III)} \rightarrow \text{(III)} - 2\text{(I)} \\ \text{(IV)} \rightarrow \text{(IV)} - \text{(I)} \\ = \end{array} \text{rg} \begin{pmatrix} 1 & 2 & -3 & 3 \\ 0 & -7 & 14 & -7 \\ 0 & -3 & 6 & -3 \\ 0 & -2 & 4 & -2 \end{pmatrix} \\ & \begin{array}{l} \text{(II)} \rightarrow \text{(II)} - \text{(IV)} \\ \text{(III)} \rightarrow \text{(III)} + 2\text{(IV)} \\ = \end{array} \text{rg} \begin{pmatrix} 1 & -1 & 3 & 3 \\ 0 & 0 & 0 & -7 \\ 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & -2 \end{pmatrix} \\ & \begin{array}{l} \text{(II)} \rightarrow -1/7\text{(II)} \\ \text{(III)} \rightarrow \text{(III)} - 3\text{(II)} \\ \text{(IV)} \rightarrow \text{(IV)} - 2\text{(II)} \\ = \end{array} \text{rg} \begin{pmatrix} 1 & -1 & 3 & 3 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\ & \begin{array}{l} \text{(II)} \rightarrow \text{(II)} + \text{(I)} \\ \text{(III)} \rightarrow \text{(III)} - 3\text{(I)} \\ \text{(IV)} \rightarrow \text{(IV)} - 3\text{(I)} \\ = \end{array} \text{rg} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{aligned}$$

Offenbar ist der Rang der Matrix 2, was man bereits nach der 2. Umformung erkennt; Vertauschung der Spalten 2 und 4 liefert die in Satz 6.22 postulierte Gestalt E_2 .

Zur Abrundung des Bildes diskutieren wir kurz ein alternatives Rechenverfahren zur Rangbestimmung von Matrizen (vgl. Fischer [3, Kap. 3.3]).

Bemerkung 6.26: Es sei $A \in K^{m \times n}$ und r die Zeilen/Spaltenanzahl der größten $(r \times r)$ -Untermatrix von A mit Determinante ungleich Null. Dann ist $\text{rg}(A) = r$.

Beispiel 6.27: Gesucht ist der Rang von

$$A = \begin{pmatrix} 5 & -2 & 4 & 1 \\ 3 & 1 & 4 & 2 \\ -2 & 0 & 6 & 3 \end{pmatrix}.$$

Durch Streichen der Spalte i erhält man vier 3×3 -Untermatrizen A_i :

$$A_1 = \begin{pmatrix} -2 & 4 & 1 \\ 1 & 4 & 2 \\ 0 & 6 & 3 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 5 & 4 & 1 \\ 3 & 4 & 2 \\ -2 & 6 & 3 \end{pmatrix}$$

$$A_3 = \begin{pmatrix} 5 & -2 & 1 \\ 3 & 1 & 2 \\ -2 & 0 & 3 \end{pmatrix}, \quad A_4 = \begin{pmatrix} 5 & -2 & 4 \\ 3 & 1 & 4 \\ -2 & 0 & 6 \end{pmatrix}.$$

Nach Bemerkung 6.26 gilt $\text{rg}(A) = 3$ genau dann, wenn $|A_i| \neq 0$ für ein $1 \leq i \leq 4$ ist. Man errechnet $|A_1| = -24 + 6 - (-24 + 12) = -6 \neq 0$, und es folgt $\text{rg}(A) = 3$. Da A_1 aus den ersten drei Spalten von A besteht, sind die drei Spalten linear unabhängig.

6.3 Die Cramersche Regel

Die *Cramersche Regel*¹ führt die Lösung von linearen Gleichungssystemen auf Determinantenberechnungen zurück.

Satz 6.28 (Cramersche Regel): Es seien $A = (a_1, \dots, a_n) \in K^{n \times n}$ und $x, b \in K^n$ sowie $Ax = b$ ein lineares Gleichungssystem, und es gelte $\det(A) \neq 0$. Seien

$$A_i := (a_1, \dots, a_{i-1}, b, a_{i+1}, \dots, a_n), i = 1, \dots, n.$$

Dann gilt

$$x_i = \frac{\det(A_i)}{\det(A)}, i = 1, \dots, n.$$

¹G. Cramer, 1704–1752; Schweizer Mathematiker, veröffentlichte 1750 die nach ihm benannte Formel, die aber schon 1678 Leibniz kannte, ohne dass Cramer dies wusste.

BEWEIS Sei $X_i \in K^{n \times n}$ gegeben durch

$$X_i = \begin{pmatrix} 1 & x_1 & & & \\ & \ddots & \vdots & & 0 \\ & & 1 & x_{i-1} & \\ & & & x_i & \\ 0 & & x_{i+1} & 1 & \\ & & \vdots & & \ddots \\ & & x_n & & 1 \end{pmatrix}, \quad (6.3)$$

also $X_i = (e_1, \dots, e_{i-1}, x, e_{i+1}, \dots, e_n)$. Eine n -malige Laplace-Entwicklung nach der jeweils ersten Zeile (vgl. Formel (6.3)) zeigt $|X_i| = x_i$. Weiter ist

$$\begin{aligned} A X_i &= A (e_1, e_2, \dots, e_{i-1}, x, e_{i+1}, \dots, e_n) \\ &= (a_1, a_2, \dots, a_{i-1}, Ax, a_{i+1}, \dots, a_n) \\ &= (a_1, a_2, \dots, a_{i-1}, b, a_{i+1}, \dots, a_n) = A_i. \end{aligned}$$

Daraus folgt $|A_i| = |AX_i| = |A| |X_i|$, also $x_i = |A_i|/|A|$. ■

Beispiel 6.29: Wie berechnen das lineare Gleichungssystem

$$\begin{array}{r|l} 1 & 2 & 5 \\ 3 & 2 & 11 \end{array}$$

mit der Cramerschen Regel und erhalten

$$|A| = \begin{vmatrix} 1 & 2 \\ 3 & 2 \end{vmatrix} = 2 - 6 = -4;$$

$$|A_1| = \begin{vmatrix} 5 & 2 \\ 11 & 2 \end{vmatrix} = 10 - 22 = -12;$$

$$|A_2| = \begin{vmatrix} 1 & 5 \\ 3 & 11 \end{vmatrix} = 11 - 15 = -4,$$

also $x_1 = |A_1|/|A| = -12/-4 = 3$ und $x_2 = |A_2|/|A| = -4/-4 = 1$.

Beispiel 6.30: Die Cramersche Regel kann bei Handrechnung vorteilhaft sein, wenn bei kleinen linearen Gleichungssystemen „krumme Zahlen“ in den Daten oder der Lösung auftauchen. Gegeben sei

$$\begin{array}{r|l} 1 & 2 & 7 & -2 \\ -3 & 1 & -4 & 1 \\ 5 & 11 & 5 & -1 \end{array}$$

Mit der Regel von Sarrus berechnet man

$$|A| = \begin{vmatrix} 1 & 2 & 7 \\ -3 & 1 & -4 \\ 5 & 11 & 5 \end{vmatrix} = 5 - 40 - 7 \cdot 33 - (35 - 44 - 30) = -227.$$

Wieder mit der Regel von Sarrus erhält man

$$|A_1| = \begin{vmatrix} -2 & 2 & 7 \\ 1 & 1 & -4 \\ -1 & 11 & 5 \end{vmatrix} = -10 + 8 + 77 - (-7 + 88 + 10) = -16,$$

$$|A_2| = \begin{vmatrix} 1 & -2 & 7 \\ -3 & 1 & -4 \\ 5 & -1 & 5 \end{vmatrix} = 5 + 40 + 21 - (35 + 4 + 30) = -3,$$

$$|A_3| = \begin{vmatrix} 1 & 2 & -2 \\ -3 & 1 & 1 \\ 5 & 11 & -1 \end{vmatrix} = -1 + 10 + 66 - (-10 + 11 + 6) = 68.$$

Die Lösung lautet $x_1 = 16/227$, $x_2 = 3/227$, $x_3 = -68/227$. Man überzeugt sich leicht davon, dass eine Lösung mit dem Gauß-Verfahren bei Handrechnung deutlich aufwändiger gerät.

6.4 Überbestimmte lineare Gleichungssysteme

Allen recht getan ist eine Kunst, die niemand kann.

(alte Volksweisheit)

In der Praxis sind häufig mehrere Datenpunkte gegeben, die physikalische Messwerte, wirtschaftliche Größen oder Ähnliches repräsentieren und von denen man z. B. einen linearen Zusammenhang annimmt. Aufgrund von Ungenauigkeiten wie z. B. Messfehlern liegen die Punkte aber nicht genau auf einer Geraden. Es existiert also keine Gerade, die alle Punkte genau trifft. Man versucht deswegen, diejenige Gerade zu finden, die „im Ganzen“ zu den gegebenen Punkten am besten passt, auch wenn sie womöglich nicht einen der Punkte exakt trifft (Fitting). Unter „passen“ versteht man, dass die Summe der Quadrate der Abweichungen der Kurve von den gegebenen Punkten minimiert wird. Man kann solche Überlegungen natürlich für beliebige Abhängigkeiten (polynomiell, exponentiell, trigonometrisch etc.) anstellen, bei Geraden spricht man von einer *linearen Regression*.

Beispiel 6.31: Durch die Punkte $(0;0)$, $(1;1)$, $(3;1)$, $(4;4)$ (vgl. Abb. 6.1) soll eine Ausgleichsgerade gelegt werden. Wir setzen also $g(x) = \alpha x + \beta$ und erhalten das lineare Gleichungssystem

$$\alpha \cdot 0 + \beta = 0$$

$$\alpha \cdot 1 + \beta = 1$$

$$\alpha \cdot 3 + \beta = 1$$

$$\alpha \cdot 4 + \beta = 4$$

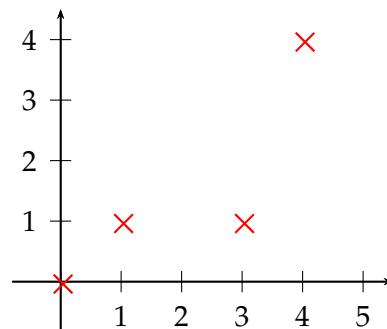


Abbildung 6.1: Datenpunkte aus Beispiel 6.31

in (α, β) bzw. in Matrixform:

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 4 \end{pmatrix}.$$

Da die Punkte nicht auf einer Geraden liegen, hat dieses lineare Gleichungssystem keine Lösung.

In der Praxis muss das zugrundeliegende Ausgleichsproblem aber gelöst werden, so dass wir im folgenden einen geeigneten Lösungsbegriff für diese Situation diskutieren und Rechenverfahren zur Bestimmung einer solchen Lösung angeben wollen.

Ein überstimmtes Gleichungssystem $Ax = b$, wie es bei obigem Ausgleichsproblem auftrat, besitzt nach Folgerung 6.10 nur eine Lösung, wenn b im Bild von L_A liegt, der von A induzierten linearen Abbildung $x \rightarrow Ax$, kurz $b \in \text{Bild}(L_A)$. Es liegt nahe, für beliebiges b eine (Ersatz)lösung zu generieren, indem man einfach b durch eine rechte Seite aus $\text{Bild}(L_A)$ ersetzt. Dabei sollte der Ersatz dem wahren b möglichst nahe kommen, um die entstehende Verfälschung zu minimieren. Es handelt sich nach Satz 4.18 bei $\text{Bild}(L_A)$ um einen Untervektorraum, so dass nach Satz 3.151 die *orthogonale Projektion* $p_A(b)$ von b auf $\text{Bild}(L_A)$ die bestmögliche Wahl bezogen auf die vom Skalarprodukt induzierte Norm ist.

Nach Satz 3.138 lässt sich die orthogonale Projektion eines Vektors auf einen Unterraum einfach berechnen, wenn eine Orthonormalbasis des Unterraums bekannt ist. Wir betrachten jetzt den allgemeineren Fall, dass lediglich irgendeine Basis bekannt ist.

Bemerkung 6.32: Sei $p_A(b)$ die Projektion eines Vektors $b \in \mathbb{R}^m$ auf den von den Vektoren $A = (a_1, \dots, a_n) \in \mathbb{R}^{m \times n}$ aufgespannten Unterraum U ,

also das Bild von A . Damit existiert ein $x \in \mathbb{R}^n$ mit

$$p_A(b) = \sum_{k=1}^n x_k a_k = Ax.$$

Dann gilt

$$\begin{aligned} b - p_A(b) \perp U &\Leftrightarrow b - Ax \in U^\perp & (6.4) \\ &\Leftrightarrow b - Ax \perp a_k \quad \forall k \\ &\Leftrightarrow \langle a_k, b - Ax \rangle = 0 \quad \forall k \\ &\Leftrightarrow A^T(b - Ax) = 0 \\ &\Leftrightarrow A^T Ax = A^T b. \end{aligned}$$

Die Gleichungen $A^T Ax = A^T b$ heißen *Normalgleichungen*.

Satz 6.33: Die Normalgleichungen sind für jede reelle Matrix $A \in \mathbb{R}^{m \times n}$ lösbar. Im Falle $\text{rg}(A) = n$ existiert mit

$$x = (A^T A)^{-1} A^T b \quad (6.5)$$

sogar eine eindeutige Lösung. In diesem Fall heißt $(A^T A)^{-1} A^T$ *verallgemeinerte Inverse* von A .

BEWEIS Die Lösbarkeit der Normalgleichungen folgt aus der Existenz der orthogonalen Projektion (vgl. Folgerung 3.144) und Rechnung (6.4). Sei ab jetzt $\text{rg}(A) = n$. Dann bilden die Spalten von A eine Basis des Bildes von A , so dass $p_A(b)$ eindeutige Koordinaten in dieser Basis besitzt. Dabei handelt es sich aber gerade um x , und die Normalgleichungen besitzen eine eindeutige Lösung. Folgerung 6.9 besagt für $A^T A \in \mathbb{R}^{n \times n}$, dass $\text{rg}(A^T A) = n$, also ist $A^T A$ invertierbar. Damit hat die Lösung der Normalgleichungen x die Gestalt (6.5). ■

Der folgende Satz rechtfertigt den Ansatz, bei einem überbestimmten linearen Gleichungssystem $Ax = b$ das b durch $p_A(b)$ zu ersetzen.

Satz 6.34 (Methode der kleinsten Quadrate): Gegeben ist das Gleichungssystem

$$Ax = b, \quad A \in \mathbb{R}^{m \times n}, \quad b \in \mathbb{R}^m, \quad m \geq n.$$

Im Fall $\text{rg}(A) = n$ gilt für

$$x_s = (A^T A)^{-1} A^T b,$$

dass

$$\|b - Ax_s\| = \min_{z \in \mathbb{R}^n} \|b - Az\|.$$

Der Vektor x_s heißt *Näherungslösung nach der Methode der kleinsten Quadrate*.

BEWEIS Nach Satz 6.33 ist $x = (A^T A)^{-1} A^T b$ die eindeutige Lösung der Normalgleichungen und $p_A(b) = Ax$ ist die orthogonale Projektion von b auf $\text{Bild}(A)$. Mit dem Satz des Pythagoras ergibt sich für beliebige $z \in \mathbb{R}^n$:

$$\begin{aligned} \|b - Az\|^2 &= \|(b - p_A(b)) + \underbrace{(p_A(b) - Az)}_{\perp (b - p_A(b))}\|^2 \\ &= \|b - p_A(b)\|^2 + \|p_A(b) - Az\|^2 \\ &\geq \|b - p_A(b)\|^2 \\ &= \|b - Ax\|^2 \end{aligned}$$

Somit liefert $p_A(b)$ den kleinsten Abstand aller Elemente des Bildes von A zu b . ■

Gelegentlich spricht man von einer klassischen Lösung eines linearen Gleichungssystems, wenn $Ax = b$ erfüllt ist in Abgrenzung von einer Lösung nach der Methode der kleinsten Quadrate.

Beispiel 6.35: Das Gleichungssystem

$$\begin{aligned} x &= 1 \\ x &= -1 \end{aligned}$$

ist überbestimmt und besitzt keine klassische Lösung. Für eine Lösung nach der Methode der kleinsten Quadrate $x \in \mathbb{R}^1$ muss $\|Ax - b\|$ minimal sein. Hier kann man x leicht direkt bestimmen:

$$\|Ax - b\|^2 = (x - 1)^2 + (x + 1)^2 = x^2 - 2x + 1 + x^2 + 2x + 1 = 2x^2 + 2,$$

also ist $\|Ax - b\|^2$ minimal für $x = 0$.

Wir wenden auf dasselbe Problem Satz 6.34 an und schreiben das lineare Gleichungssystem als $Ax = b$ mit

$$A = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad x = (x), \quad b = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Wir lösen die Normalgleichungen:

$$\begin{aligned} A^T A x &= A^T b \\ \Leftrightarrow (1 \ 1) \begin{pmatrix} 1 \\ 1 \end{pmatrix} x &= (1 \ 1) \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\ \Leftrightarrow 2x &= 0. \end{aligned}$$

Beispiel 6.36: Das Gleichungssystem

$$\begin{aligned} x_1 &= 1 \\ x_2 &= 2 \\ x_1 + x_2 &= -1 \end{aligned}$$

ist überbestimmt und im klassischen Sinn unlösbar. Für eine Lösung nach der Methode der kleinsten Quadrate schreiben wir das lineare Gleichungssystem in der Form $Ax = b$ mit

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}.$$

Die Minimalstelle von $\|Ax - b\|^2$ löst die Normalgleichungen $(A^T A)x = A^T b$. Wir errechnen

$$A^T A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

und weiter

$$A^T b = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Mit dem Gauß-Verfahren löst man das Normalgleichungssystem

$$\begin{array}{c|c} 2 & 1 & 0 \\ 1 & 2 & 1 \end{array} \longleftrightarrow \begin{array}{c|c} 2 & 1 & 0 \\ 0 & 3/2 & 1 \end{array} \longleftrightarrow \begin{array}{c|c} 2 & 1 & 0 \\ 0 & 1 & 2/3 \end{array} \longleftrightarrow \begin{array}{c|c} 2 & 0 & -2/3 \\ 0 & 1 & 2/3 \end{array}$$

und erhält $x = (-1/3, 2/3)^T$. Weiter gilt

$$\begin{aligned} Ax - b &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \frac{1}{3} \begin{pmatrix} -1 \\ 2 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} \\ &= \frac{1}{3} \left(\begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix} - \begin{pmatrix} 3 \\ 6 \\ -3 \end{pmatrix} \right) = \frac{4}{3} \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} \end{aligned}$$

und deswegen $\|Ax - b\|^2 = 5 \frac{1}{3}$.

Wir kehren zum Einstiegsbeispiel 6.31 zurück, bei dem wir auf ein überbestimmtes lineares Gleichungssystem der Form $Ax = b$ geführt wurden. Wählt man als Norm die euklidische Norm, dann entspricht $\|Ax - b\|^2$ genau der Summe der Quadrate der Abweichungen der Ausgleichsgerade zu den Datenpunkten, die minimal ist, wenn die Parameter α und β der Ausgleichsgeraden durch $A^T A(\alpha, \beta)^T = A^T b$ bestimmt werden. Dies rechtfertigt die Bezeichnung „Methode der kleinsten Fehlerquadrate“. Man errechnet

$$A^T A = \begin{pmatrix} 26 & 8 \\ 8 & 4 \end{pmatrix} \quad \text{und} \quad A^T b = \begin{pmatrix} 20 \\ 6 \end{pmatrix}$$

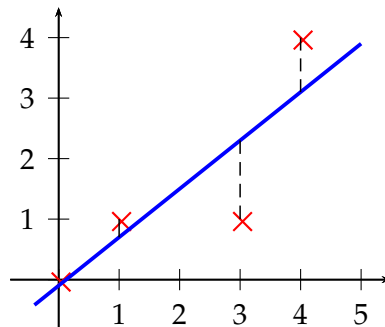


Abbildung 6.2: Ausgleichsgerade durch gegebene Datenpunkte

und löst das Normalgleichungssystem z. B. mit dem Gauß-Verfahren:

$$\begin{array}{c|c} 26 & 8 \\ \hline 8 & 4 \end{array} \left| \begin{array}{c} 20 \\ 6 \end{array} \right. \longleftrightarrow \begin{array}{c|c|c} 13 & 4 & 10 \\ \hline 4 & 2 & 3 \end{array} \left| \begin{array}{c} \cdot 4 \\ \cdot 13 \end{array} \right. \longleftrightarrow \begin{array}{c|c|c} 52 & 16 & 40 \\ \hline 52 & 26 & 39 \end{array} \longleftrightarrow \begin{array}{c|c|c} 13 & 4 & 10 \\ \hline 0 & -10 & 1 \end{array}$$

Man liest $\beta = -0,1$ ab. Durch Einsetzen in die erste Gleichung erhält man $\alpha = 0,8$. Somit lautet die Ausgleichsgerade $y = 0,8x - 0,1$.

Setzt man als Ausgleichsfunktion eine Parabel („quadratischer Fit“) an, führt der Ansatz $p(x) = \alpha x^2 + \beta x + \gamma$ auf das lineare Gleichungssystem

$$\begin{aligned} \alpha \cdot 0^2 + \beta \cdot 0 + \gamma &= 0 \\ \alpha \cdot 1^2 + \beta \cdot 1 + \gamma &= 1 \\ \alpha \cdot 3^2 + \beta \cdot 3 + \gamma &= 1, \quad \text{bzw.} \\ \alpha \cdot 4^2 + \beta \cdot 4 + \gamma &= 4 \end{aligned}$$

$$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 9 & 3 & 1 \\ 16 & 4 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 4 \end{pmatrix},$$

woraus man $p(x) = 1/3x^2 - 8/15x + 2/5$ erhält (Abb. 6.3).

Bemerkung 6.37: Legt man durch die Datenpunkte (x_k, y_k) , $k = 1, \dots, m$ eine Ausgleichsgerade $g(x) = \alpha x + \beta$ mit der Methode der kleinsten Quadrate, wird man auf

$$\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

geführt. Durch Ausrechnen erhält man die Normalgleichungen

$$\begin{pmatrix} \sum_{k=1}^m x_k^2 & \sum_{k=1}^m x_k \\ \sum_{k=1}^m x_k & m \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^m x_k y_k \\ \sum_{k=1}^m y_k \end{pmatrix}.$$

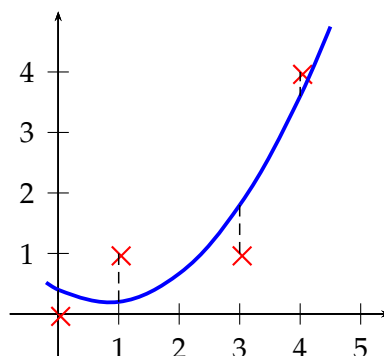


Abbildung 6.3: Ausgleichsparabel durch gegebene Datenpunkte

Dasselbe Gleichungssystem erhält man, wenn man das Minimum der quadratischen Funktion $e(\alpha, \beta) = \sum_{k=1}^m (g(x_k) - y_k)^2$ bestimmt, indem man die Nullstelle von ∇e errechnet.

6.5 Unterbestimmte lineare Gleichungssysteme

Zur Motivation betrachten wir wiederum die Aufgabe, durch n Datenpunkte ein Polynom p als Ausgleichskurve zu legen. Im Unterschied zu Kapitel 6.4 sei jetzt aber $\deg(p) \geq n$, so dass diese Aufgabe keine eindeutige Lösung hat und viele Polynome als Interpolanden in Frage kommen. Gegeben seien zur Illustration die Punkte $(-1, 1)$, $(0, 1)$ und $(1, 1)$. Wir wählen $p(x) = a + bx + cx^2 + dx^3$ und erhalten als Interpolanden $p_1(x) = 1$, aber auch $p_2(x) = x^3 - x + 1$. Beide Funktionen erfüllen die Interpolationsaufgabe vollauf, aber man wird in der Praxis sicherlich die Funktion p_1 vorziehen. Zur genaueren Analyse stellen wir das lineare Gleichungssystem der Interpolationsaufgabe auf und erhalten

$$\begin{pmatrix} 1 & -1 & 1 & -1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Sowohl der Koeffizientenvektor $v_1 = (1, 0, 0, 0)^T$ der Funktion p_1 als auch der Koeffizientenvektor $v_2 = (1, -1, 0, 1)^T$ von p_2 erfüllen natürlich obiges System, aber v_1 zeichnet sich von allen Lösungsvektoren dadurch aus, dass er die kleinstmögliche Norm besitzt: Wegen des zweiten Punktes muss $a = 1$ gewählt werden, die anderen Komponenten von v_1 sind bereits null. Dieser Befund motiviert den folgenden Satz:

Satz 6.38: Gegeben ist ein lineares Gleichungssystem $Ax = b$ mit $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ und $m \leq n$. Für $\text{rg}(A) = m$ löst

$$x_s = A^T(AA^T)^{-1}b$$

das Gleichungssystem $Ax = b$, und es gilt:

1. x_s ist die Lösung mit der kleinsten Norm aller Lösungen x , d. h.

$$\|x_s\| = \min_{Ax=b} \{\|x\|\}.$$

2. x_s steht orthogonal zu sämtlichen Lösungen z von $Az = 0$, d. h.

$$x_s \perp \ker(A)$$

Für $\text{rg}(A) = m$ heißt $A^T(AA^T)^{-1}$ die *verallgemeinerte Inverse* von A .

BEWEIS Wir zeigen zunächst die Invertierbarkeit von $AA^T \in \mathbb{R}^{m \times m}$, denn dann ist x_s definiert. Sei $x \in \ker(AA^T)$. Es folgt $AA^T x = 0 \Rightarrow x^T AA^T x = \|A^T x\|^2 = 0$, also $x \in \ker(A^T)$. Wegen $\text{rg}(A^T) = \text{rg}(A) = m$ folgt aus der Dimensionsformel 4.23 $\dim(\ker(A^T)) = 0$ und daher $x = 0$. Somit gilt $\ker(AA^T) = \{0\}$. Wiederum mit der Dimensionsformel erhält man $\text{rg}(AA^T) = m$, und deswegen ist AA^T invertierbar.

Weiter ist x_s eine Lösung von $Ax = b$, denn

$$Ax_s = A(A^T(AA^T)^{-1}b) = (AA^T)(A^T)^{-1}b = b.$$

Man hat $x_s \perp \ker(A)$, denn für $z \in \ker(A)$ gilt

$$\langle z, x_s \rangle = z^T A^T (A A^T)^{-1} b = \langle \overbrace{Az}^{=0}, (A A^T)^{-1} b \rangle = 0.$$

Nach Satz 6.7 gilt für jede Lösung x , dass $x = x_s + z$ mit $z \in \ker(A)$. Aus dem Satz des Pythagoras und $x_s \perp \ker(A)$ folgt

$$\|x\|^2 = \|x_s\|^2 + \|z\|^2 \geq \|x_s\|^2,$$

Gleichheit liegt nur vor für $\|z\|^2 = 0$, also $z = 0$. Damit ist x_s die eindeutig bestimmte Lösung von $Ax = b$ mit der kleinsten Norm. ■

Beispiel 6.39: Das lineare Gleichungssystem

$$2x_1 + 2x_2 - x_3 = 6$$

ist unterbestimmt, alle Lösungen bilden eine Ebene in \mathbb{R}^3 . Die normminimale Lösung x ist der Vektor, der vom Nullpunkt aus senkrecht auf die Lösungsebene trifft. Aus der Hesseschen Normalform der Ebene

$$\frac{1}{3}(2x_1 + 2x_2 - x_3) = 2$$

erhält man sofort $x = \frac{2}{3}(2, 2, -1)^T$. Dieselbe Lösung x erhält man auch durch $x = A^T(AA^T)^{-1}b$ mit

$$A = (2, 2, -1), \quad x = (x_1, x_2, x_3)^T, \quad b = (6).$$

Man errechnet

$$\begin{aligned} x = A^T(AA^T)^{-1}b &= \begin{pmatrix} 2 \\ 2 \\ -1 \end{pmatrix} \left((2, 2, -1) \begin{pmatrix} 2 \\ 2 \\ -1 \end{pmatrix} \right)^{-1} \cdot 6 \\ &= \begin{pmatrix} 2 \\ 2 \\ -1 \end{pmatrix} (9)^{-1} \cdot 6 \\ &= \frac{2}{3} \begin{pmatrix} 2 \\ 2 \\ -1 \end{pmatrix}. \end{aligned}$$

Kapitel 7

Geometrie linearer Abbildungen

Die Begriffe „Matrix“ und „Lineare Abbildung“ durchziehen die gesamte Lineare Algebra. Wir wissen, dass (bei endlich erzeugten Vektorräumen) jede Matrix A eine lineare Abbildung f_A induziert und umgekehrt jede lineare Abbildung Matrixdarstellungen besitzt. Für ein konkretes A ist es jedoch oftmals schwierig zu verstehen, was genau f_A „mit den Vektoren macht“. Es geht darum, aus algebraischen Eigenschaften von A geometrische Eigenschaften von f_A abzuleiten und umgekehrt. Derartige Zusammenhänge werden in diesem Kapitel behandelt. Wir gehen dabei im gesamten Kapitel von der euklidischen Norm und dem euklidischen Skalarprodukt aus.

7.1 Orthogonale Abbildungen und Matrizen

Wir betrachten eine Drehung D in der Ebene um 0. Jede Drehung erhält die euklidische Länge eines Vektors, es gilt also $\|D(x)\| = \|x\| \forall x \in \mathbb{R}^2$. Man spricht allgemeiner bei einer abstandserhaltenden Abbildung (also $\|f(x) - f(y)\| = \|x - y\| \forall x, y$) von einer *Isometrie*. Dreht man zwei Vektoren auf die gleiche Weise, ändert sich ihre Lage zueinander nicht, insbesondere bleiben rechte Winkel erhalten. Dies impliziert, dass die Standardbasis als Orthonormalbasis auf eine Orthonormalbasis abgebildet wird. Nach Bem. 4.29 sind Drehungen lineare Abbildungen und besitzen eine Darstellungsmatrix A . Da in den Spalten von A die Bilder der Basisvektoren stehen (vgl. Kapitel 4.3), bilden die Spalten von A bzgl. der Standardbasis eine Orthonormalbasis. Obige Überlegungen gelten allgemeiner für lineare winkelerhaltende Isometrien, die auch Spiegelungen umfassen und motivieren die folgende Definition.

Definition 7.1:

1. Eine Matrix $A \in \mathbb{R}^{n \times n}$ heißt *orthogonal*, wenn ihre Spaltenvektoren eine Orthonormalbasis bilden.
2. Die Menge aller orthogonalen Matrizen in $\mathbb{R}^{n \times n}$ heie $O(n)$.

Man beachte die Tcke dieser Definition: Eine *Orthogonal*matrix liegt vor bei einer *Orthonormal*basis in den Spalten, eine *Orthogonal*basis gengt nicht.

Beispiel 7.2:

1. Die Einheitsmatrix E liegt in $O(n)$.
2. In \mathbb{R}^2 haben nach Beispiel 4.40 Drehungen um den Winkel α bzgl. der Standardbasis die Matrixdarstellung

$$A_\alpha = (a_1 \ a_2) = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}.$$

Wegen $\cos^2 \alpha + \sin^2 \alpha = 1$ haben beide Matrixspalten die Lnge 1, zudem gilt $\langle a_1, a_2 \rangle = 0$, also $A_\alpha \in O(2)$ wie nach obiger Diskussion zu erwarten.

3. Eine Spiegelung eines Vektors in \mathbb{R}^2 an der x -Achse ist eine lineare Abbildung mit (bzgl. der Standardbasis) der Darstellungsmatrix

$$S_x = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

offensichtlich gilt $S_x \in O(2)$.

Satz 7.3: Fr $A \in \mathbb{R}^{n \times n}$ sind quivalent:

1. $A \in O(n)$
2. A ist invertierbar, und es gilt $A^{-1} = A^T$.
3. $A^T \in O(n)$

BEWEIS 1. \Leftrightarrow 2.: Sei $A = (a_1, \dots, a_n)$. Es folgt

$$\begin{aligned} A^T = A^{-1} &\Leftrightarrow A^T A = E \\ &\Leftrightarrow (A^T A)_{ij} = \langle a_i, a_j \rangle = \delta_{ij}, 1 \leq i, j \leq n \\ &\Leftrightarrow a_i \perp a_j, i \neq j \quad \text{und} \quad \|a_i\| = 1, 1 \leq i \leq n \\ &\Leftrightarrow A \in O(n). \end{aligned}$$

1. \Leftrightarrow 3.: Mit 1) \Leftrightarrow 2) erhält man

$$\begin{aligned} A^T \in O(n) &\Leftrightarrow (A^T)^{-1} = (A^T)^T \\ &\Leftrightarrow (A^{-1})^T = A \\ &\Leftrightarrow A^{-1} = A^T \Leftrightarrow A \in O(n). \end{aligned}$$

Satz 7.4: $O(n)$ bildet mit der Matrizenmultiplikation die sog. *orthogonale Gruppe*.

BEWEIS Wir rechnen die Gruppeneigenschaften nach Definition 3.1 direkt nach und zeigen zunächst, dass die Matrixmultiplikation eine Verknüpfung auf $O(n)$ darstellt, also für $A, B \in O(n)$ auch $AB \in O(n)$ gilt. Nach Satz 7.3 sind A und B invertierbar, somit auch AB . Es folgt $(AB)^{-1} = B^{-1}A^{-1} = B^T A^T = (AB)^T$ und daher wiederum nach Satz 7.3 $AB \in O(n)$. Wir zeigen die Gruppenaxiome. G1: Die Matrixmultiplikation ist für alle quadratischen Matrizen assoziativ, also insbesondere für orthogonale. G2 zeigt Beispiel 7.2.1. Zu G3: Sei $A \in O(n)$. Dann gilt $(A^{-1})^{-1} = (A^T)^{-1} = (A^{-1})^T$, und daher nach Satz 7.3 $A^{-1} \in O(n)$. ■

Wir kehren zu unseren anfänglichen Betrachtungen zurück. Deren Ergebnis lautet zusammengefasst: Jede winkelerhaltende lineare Isometrie wird bezogen auf die Standardbasis durch eine orthogonale Matrix dargestellt. Wir untersuchen die Umkehrung.

Bemerkung 7.5: Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ beliebig. Dann sind äquivalent:

1.

$$\langle f(x), f(y) \rangle = \langle x, y \rangle \quad \forall x, y \in \mathbb{R}^n \quad (7.1)$$

2. f ist eine winkelerhaltende Isometrie.

BEWEIS „ \Rightarrow “: Die Isometrie-eigenschaft zeigt

$$\begin{aligned} \|f(x) - f(y)\|^2 &= \langle f(x) - f(y), f(x) - f(y) \rangle \\ &= \langle f(x), f(x) \rangle - 2\langle f(x), f(y) \rangle + \langle f(y), f(y) \rangle \\ &= \langle x, x \rangle - 2\langle x, y \rangle + \langle y, y \rangle \\ &= \langle x - y, x - y \rangle = \|x - y\|^2. \end{aligned}$$

Damit folgt

$$\angle(f(x), f(y)) = \arccos \frac{\langle f(x), f(y) \rangle}{\|f(x)\| \|f(y)\|} = \arccos \frac{\langle x, y \rangle}{\|x\| \|y\|} = \angle(x, y).$$

„ \Leftarrow “: Mit $\alpha = \angle(x, y) = \angle(f(x), f(y))$ gilt

$$\langle f(x), f(y) \rangle = \cos(\alpha) \|f(x)\| \|f(y)\| = \cos(\alpha) \|x\| \|y\| = \langle x, y \rangle. \quad \blacksquare$$

Offenbar ist Eigenschaft (7.1) die hier maßgebliche. Daher definieren wir:

Definition 7.6: Eine Abbildung $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ heißt *orthogonal*, wenn (7.1) gilt.

Wir sind zu Beginn davon ausgegangen, dass alle Drehungen und Spiegelungen als orthogonale Abbildungen linear seien. Dies trifft zu.

Satz 7.7: Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ orthogonal. Dann ist f linear.

BEWEIS Wir skizzieren den Beweisgang. Sei $\alpha \in \mathbb{R}$. Man rechnet mit den Rechenregeln des Skalarproduktes $\|f(x+y) - f(x) - f(y)\|^2 = 0$ direkt aus und gewinnt so die Additivität von f . Ebenso errechnet man $\|f(\alpha x) - \alpha f(x)\|^2 = 0$ und erhält die Homogenität von f . Damit ist f linear. ■

Allgemeine (d. h. nicht unbedingt winkelerhaltende) Isometrien sind natürlich nicht zwangsläufig linear, man betrachte z. B. $r : \mathbb{R} \rightarrow \mathbb{R}, r(x) = x + 1$.

Satz 7.8: Die Abbildung $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ist genau dann orthogonal, wenn die zugehörige Darstellungsmatrix Q (bzgl. der Standardbasis) eine Orthogonalmatrix ist, also $Q \in O(n)$.

BEWEIS „ \Rightarrow “: Nach Satz 7.7 ist f linear, besitzt also eine Darstellungsmatrix Q . Es folgt $\langle Qx, Qy \rangle = \langle Q^T Qx, y \rangle = \langle x, y \rangle \forall x, y \in \mathbb{R}^n$, also $\langle Q^T Qx - Ex, y \rangle = 0 \forall x, y$. Mit der speziellen Wahl $y = Q^T Qx - Ex$ erhält man $\langle Q^T Qx - Ex, Q^T Qx - Ex \rangle = 0 \forall x$ und deswegen $(Q^T Q - E)x = 0 \forall x$. Es folgt $Q^T Q = E$ und nach Satz 7.3 $Q \in O(n)$.

„ \Leftarrow “: Folgt mit derselben Rechnung wie oben. ■

Die Gestalt der Darstellungsmatrix A einer linearen Abbildung f hängt von der Wahl der Basis ab, geometrische Eigenschaften von f selbst dagegen nicht. Deswegen dürfen algebraische Eigenschaften von A , wenn sie geometrische Bedeutung haben sollen, nicht von der Wahl der Basis abhängen. Für die Orthogonalität einer Matrix gilt dies zumindest beim Wechsel zwischen Orthonormalbasen.

Bemerkung 7.9: Sei f orthogonal. Dann ist die Darstellungsmatrix von f bezogen auf *jede* Orthonormalbasis orthogonal.

BEWEIS Sei \mathcal{E} die Standardbasis und \mathcal{B} eine Orthonormalbasis, sei $Q = M_{\mathcal{E}}^{\mathcal{E}}(f)$ und S die Matrix des Basiswechsels zwischen \mathcal{E} und \mathcal{B} . Es gilt $S \in O(n)$, weil in den Spalten von S die Basisvektoren von \mathcal{B} stehen. Für $\tilde{Q} = M_{\mathcal{B}}^{\mathcal{B}}(f)$ gilt nach Folgerung 4.82 $\tilde{Q} = SQS^{-1} \in O(n)$ aufgrund der Gruppeneigenschaft von $O(n)$. ■

Bemerkung 7.10: Bemerkung 7.9 ist für allgemeine Basen falsch. Mit

$$A = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$$

sei $A = M_{\mathcal{E}}^{\mathcal{E}}(f)$, B die Matrix des Basiswechsels in die Basis \mathcal{B} , die aus den Spalten von B besteht. Dann gilt

$$M_{\mathcal{B}}^{\mathcal{B}} = BAB^{-1} = \frac{1}{\sqrt{2}} \begin{pmatrix} 3 & -7 \\ 1 & -3 \end{pmatrix},$$

und $M_{\mathcal{B}}^{\mathcal{B}}$ ist offenbar nicht orthogonal.

Wir untersuchen jetzt, wie sich Volumina unter einer orthogonalen Abbildung verhalten.

Bemerkung 7.11: Für $A \in O(n)$ gilt $|\det(A)| = 1$.

BEWEIS $1 = \det(E) = \det(AA^{-1}) = \det(AA^T) = (\det(A))^2$. Wurzelziehen liefert die Behauptung. ■

Bemerkung 7.12: Nicht jede Matrix A mit $\det(A) = 1$ ist eine Orthogonalmatrix, so z. B.

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

Aus der Analysis ist die *Transformationsformel* für Integrale bekannt, eine Verallgemeinerung der Substitutionsregel für eindimensionale Integrale. Sei $\Omega \in \mathbb{R}^n$ ein glatt berandetes beschränktes Gebiet, $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ stetig differenzierbar mit stetig differenzierbarer Umkehrfunktion und $f : \Omega \rightarrow \varphi(\Omega)$ integrierbar. Dann gilt

$$\int_{\Omega} (f \circ \varphi)(x) |\det(J\varphi(x))| dx = \int_{\varphi(\Omega)} f(z) dz$$

mit der Jacobi-Matrix $J\varphi$ von φ . Das Volumen $m(\Omega)$ lässt sich mittels $m(\Omega) = \int_{\Omega} 1 dx$ berechnen, analog $m(Q(\Omega))$. Wir setzen also $\varphi(x) = Qx$ mit einer orthogonalen Matrix und $f = 1$. Durch Ausrechnen erhält man $J\varphi(x) = Q$ und deswegen

$$m(Q(\Omega)) = \int_{Q(\Omega)} 1 dz = \int_{\Omega} 1 |\det(Q)| dx = m(\Omega).$$

Winkelerhaltende Isometrien erhalten also zwingend auch Volumina. Winkelerhaltende Isometrien erhalten nicht zwangsläufig die Orientierung, weil eine Spiegelung S ein Rechts- in ein Linkssystem transformiert und umgekehrt (man denke an die Spiegelschrift). Eine Drehung D dagegen erhält die Orientierung, es gilt daher nach Bemerkung 7.11 $\det(S) = -1$ und $\det(D) = 1$ (vgl. auch Kapitel 2.3). Wegen $\det(E) = 1$ und $\det(AB) = \det(A) \det(B)$ gilt:

Bemerkung 7.13: Die *spezielle orthogonale Gruppe*

$$SO(n) := \{Q \in O(n) \mid \det(Q) = 1\}$$

bildet eine Untergruppe von $O(n)$. Die Darstellungsmatrix einer beliebigen Drehung um 0 (bzgl. einer Orthonormalbasis) liegt in $SO(n)$, die einer Spiegelung in $O(n) \setminus SO(n)$.

Wir fassen zusammen: Eine orthogonale Abbildung ist linear und erhält Abstände, Winkel und Volumina. Ihre Darstellungsmatrix bezogen auf eine beliebige Orthonormalbasis ist orthogonal. Umgekehrt induziert jede Orthogonalmatrix bezogen auf eine Orthonormalbasis eine orthogonale Abbildung.

7.2 Exkurs: QR-Zerlegung und Anwendungen

Die Bedeutung orthogonaler Matrizen reicht weit über die algebraische Beschreibung einfacher geometrischer Transformationen hinaus. Wir betrachten ein lineares Gleichungssystem $Ax = b$ mit $A \in \mathbb{R}^{m \times n}$, $m \geq n$ und $\text{rg}(A) = n$. Wir nehmen an, es läge eine *QR-Zerlegung von A*

$$A = QR$$

vor mit einer oberen Dreiecksmatrix $R \in \mathbb{R}^{n \times n}$ und einer in den Spalten orthogonalen Matrix $Q = (q_1, \dots, q_n) \in \mathbb{R}^{m \times n}$, d. h. es sei $\langle q_i, q_j \rangle = \delta_{ij}$. Für die eindeutige Lösung x_S im Sinne der kleinsten Quadrate des obigen Systems gilt nach Satz 6.34 $x_S = (A^T A)^{-1} A^T b$. Aufgrund der Orthogonalität der Spalten von Q gilt $\text{rg}(Q) = n = \text{rg}(A)$ und nach Satz 4.55 deswegen $\text{rg}(R) = n$. Damit ist R invertierbar. Wir setzen $A = QR$ ein und erhalten

$$\begin{aligned} x_S &= (A^T A)^{-1} A^T b \\ &= \left((QR)^T (QR) \right)^{-1} (QR)^T b \\ &= (R^T Q^T QR)^{-1} R^T Q^T b \\ &= R^{-1} (R^T)^{-1} R^T Q^T b = R^{-1} Q^T b. \end{aligned}$$

Somit lässt sich x_S als Lösung des linearen Gleichungssystems

$$R x_S = Q^T b \tag{7.2}$$

errechnen. Man führt die Berechnung der Ausgleichslösung auf ein „klassisches“ lineares Gleichungssystem (7.2) zurück, das sich zudem einfach durch Rückwärtseinsetzen lösen lässt, weil R obere Dreiecksgestalt besitzt. Der Fall quadratischer invertierbarer Matrizen ist als Spezialfall $m = n$ enthalten, so dass hier eine Alternative zur Gauß-Elimination vorliegt. Die bis jetzt angenommene QR-Zerlegung lässt sich in der Tat z. B. mittels des Verfahrens von Gram-Schmidt erreichen.

Satz 7.14: Sei $A = (a_1, \dots, a_n) \in \mathbb{R}^{m \times n}$ mit $m \geq n$ und $\text{rg}(A) = n$. Dann gibt es eine in den Spalten orthogonale Matrix $Q \in \mathbb{R}^{m \times n}$ und eine obere Dreiecksmatrix $R \in \mathbb{R}^{n \times n}$ mit $A = QR$. Hierbei können die Spalten von Q mithilfe des Verfahrens von Gram-Schmidt aus den Spalten von A erzeugt werden, und es gilt $\text{rg}(R) = n$.

BEWEIS Aufgrund von $\text{rg}(A) = n$ sind alle Spalten von A linear unabhängig. Nach Satz 3.140 liefert das Verfahren von Gram-Schmidt n orthonormale Vektoren $q_i \in \mathbb{R}^m$ und damit $Q = (q_1, \dots, q_n)$. Es bleibt die Existenz von $R = (r_{ij})_{i,j=1}^n$. Man hat

$$(a_1, \dots, a_n) = (q_1, \dots, q_n) \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ & r_{22} & \dots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{pmatrix} \quad (7.3)$$

Aus (7.3) erhält man die n linearen Gleichungssysteme

$$a_i = (q_1, \dots, q_i) \begin{pmatrix} r_{1i} \\ \vdots \\ r_{ii} \end{pmatrix}, \quad 1 \leq i \leq n \quad (7.4)$$

für die Koeffizienten von R . Sind diese linearen Gleichungssysteme alle lösbar, ist die Existenz von R gezeigt. Ein Gleichungssystem (7.4) ist dann lösbar, wenn $a_i \in L(q_1, \dots, q_i)$ gilt. Dies ist aber der Fall, denn für das Verfahren von Gram-Schmidt gilt nach Satz 3.140 $L(a_1, \dots, a_i) = L(q_1, \dots, q_i)$ für $1 \leq i \leq n$. Die Aussage $\text{rg}(A) = \text{rg}(Q) = n$ wurde schon bewiesen. ■

Beispiel 7.15: Gesucht ist eine QR-Zerlegung von

$$A = (a_1, a_2) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 0 & -1 \end{pmatrix}.$$

Wir berechnen Q mithilfe des Verfahrens von Gram-Schmidt gemäß

$$q_1 = \frac{a_1}{\|a_1\|}; \quad r_2 = a_2 - \langle a_2, q_1 \rangle q_1; \quad q_2 = \frac{r_2}{\|r_2\|} \quad (7.5)$$

und erhalten

$$Q = \begin{pmatrix} 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 0 \\ 0 & -1 \end{pmatrix}.$$

Zur Berechnung von R lösen wir $a_1 = r_{11}q_1 = r_{11}a_1/\|a_1\|$ und erhalten $r_{11} = \|a_1\| = \sqrt{2}$. Man hat $a_2 = r_{12}q_1 + r_{22}q_2$, aber nach (7.5) auch $a_2 =$

$r_2 + \langle a_2, q_1 \rangle q_1 = \|r_2\| q_2 + \langle a_2, q_1 \rangle q_1$. Ein Koeffizientenvergleich liefert $q_{12} = \langle a_2, q_1 \rangle = \sqrt{2}$ und $q_{22} = \|r_2\| = 1$ und damit

$$R = \begin{pmatrix} \sqrt{2} & \sqrt{2} \\ 0 & 1 \end{pmatrix}.$$

Bemerkung 7.16:

1. Die Matrix Q aus Satz 7.14 ist für $m \neq n$ keine Orthogonalmatrix im Sinne von Definition 7.1, weil sie nicht quadratisch ist.
2. Die QR-Zerlegung einer Matrix $A \in \mathbb{R}^{m \times n}$, $m \geq n$ und $\text{rg}(A) = n$ ist nur bis auf Vorzeichen eindeutig. So gilt für die Matrix A aus Beispiel 7.15 auch $A = \tilde{Q}\tilde{R}$ mit $\tilde{Q} = (-q_1, q_2)$ und

$$\tilde{R} = \begin{pmatrix} -q_{11} & -q_{12} \\ 0 & q_{22} \end{pmatrix}.$$

3. Die Einschränkungen an A in Satz 7.14 sind dem Verfahren von Gram-Schmidt geschuldet. In Wahrheit besitzt jede reelle rechteckige Matrix eine QR-Zerlegung, die dann aber auf anderem Wege als hier gezeigt gefunden werden muss, und es gilt $\text{rg}(R) = \text{rg}(A)$.
4. Die QR-Zerlegung mithilfe des Verfahrens von Gram-Schmidt eignet sich nicht für Berechnungen mit dem Computer, weil sich beim Gram-Schmidt-Verfahren die in der Computerarithmetik unvermeidlichen Rundungsfehler im Laufe der Berechnungen verheerend auswirken können. Man spricht von *numerischer Instabilität*. Numerisch stabile und effiziente Verfahren für die QR-Zerlegung stellt die numerische Mathematik bereit.
5. In der Praxis wird man bei Handrechnung ein lineares Gleichungssystem mit dem Gauß-Verfahren lösen. Dies gilt auch für Berechnungen mit dem Computer bei eindeutig lösbaren linearen Gleichungssystemen, moderne Verfahren zur Lösung von überbestimmten Gleichungssystemen basieren aber häufig auf dem hier gezeigten Zugang über eine QR-Zerlegung.

7.3 Eigenwerte und Eigenvektoren

Wir beginnen wieder bei dem Problem, zu verstehen, was eine lineare Abbildung konkret „mit den Vektoren macht“, so wie zuvor. Wir beziehen uns im folgenden auf \mathbb{C} -Vektorräume, weil wegen $\mathbb{R} \hookrightarrow \mathbb{C}$ der reelle Fall als Spezialfall erscheint. In diesem Kapitel sei V stets ein endlich erzeugter \mathbb{C} -Vektorraum. Als Sprechweise dient die folgende Definition.

Definition 7.17: Eine lineare Abbildung $f : V \rightarrow V$ heißt *Endomorphismus*.

Ab jetzt bezeichne f immer einen Endomorphismus. Im Sinne einer Wunschvorstellung unterstellen wir, es gäbe zu einem f Vektoren v_i mit

$$f(v_i) = \lambda_i v_i, \quad \lambda_i \in \mathbb{C} \quad (7.6)$$

so, dass $\mathcal{B} = (v_1, \dots, v_n)$ eine Basis von V bildet. In den Spalten der Darstellungsmatrix $M_{\mathcal{B}}^{\mathcal{B}}(f)$ von f bzgl. \mathcal{B} stehen die Koordinaten der Bilder der Basisvektoren v_i . Wegen $K_{\mathcal{B}}(v_i) = e_i$ gilt dann aber wegen (7.6), dass $K_{\mathcal{B}}(f(v_i)) = \lambda_i K_{\mathcal{B}}(v_i) = \lambda_i e_i$. Damit erhalten wir

$$M_{\mathcal{B}}^{\mathcal{B}}(f) = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}.$$

Damit ist die „Wirkung“ von f leicht zu verstehen: f skaliert jede Koordinate bezogen auf \mathcal{B} eines Vektors mit λ_i ; eine einfachere Darstellungsmatrix als oben ist kaum möglich. Man fragt sich natürlich, ob es solche Vektoren v_i immer gibt und wie man sie ggf. berechnen soll.

Definition 7.18: Existiert für f ein $\lambda \in \mathbb{C}$ und $v \in V \setminus \{0\}$ mit

$$f(v) = \lambda v, \quad (7.7)$$

dann heißt v *Eigenvektor* von f zum *Eigenwert* λ .

Bemerkung 7.19:

1. Eigenvektoren lassen sich als Verallgemeinerung von Fixpunkten auffassen, während letztere unter f invariant bleiben, behalten erstere wenigstens ihre Richtung unter f , auch wenn sich die Länge ändert.
2. Der Nullvektor wird als Eigenvektor ausgeschlossen, weil sonst wegen $0 = f(0) = \lambda 0$ jedes $\lambda \in \mathbb{C}$ ein Eigenwert wäre und der Begriff damit bedeutungslos würde.

Wir stellen einige elementare Eigenschaften von Eigenvektoren zusammen.

Bemerkung 7.20:

1. Sei λ ein Eigenwert von f und v_1, \dots, v_k Eigenvektoren von f zu λ . Dann ist auch $v \in L(v_1, \dots, v_k) \setminus \{0\}$ ein Eigenvektor von f zu λ .
2. Für $\lambda \in \mathbb{C}$ ist $\text{Eig}(f; \lambda) := \{v \in V \mid f(v) = \lambda v\}$, der *Eigenraum* von f zu λ , ein Untervektorraum von V .

3. Für $\lambda \neq \gamma$ gilt $Eig(f; \lambda) \cap Eig(f; \gamma) = \{0\}$.
4. Eigenvektoren zu unterschiedlichen Eigenwerten sind linear unabhängig.

BEWEIS 1.) Sei $v \in L(v_1, \dots, v_k) \Rightarrow v = \sum_{i=1}^k \alpha_i v_i$ mit $\alpha_i \in \mathbb{C}$. Aus der Linearität von f folgt $f(v) = \sum_{i=1}^k \alpha_i f(v_i) = \sum_{i=1}^k \alpha_i \lambda v_i = \lambda v$.

2.): Für $v, w \in Eig(f; \lambda)$ und $\mu \in \mathbb{C}$ gilt $f(\mu v) = \mu f(v) = \mu \lambda v = \lambda(\mu v)$, also $\mu v \in Eig(f; \lambda)$. Weiterhin gilt $f(v + w) = f(v) + f(w) = \lambda v + \lambda w = \lambda(v + w)$, also $v + w \in Eig(f; \lambda)$.

3.): Sei $v \in Eig(f; \lambda) \cap Eig(f; \gamma)$. Dann gilt $f(v) = \lambda v = \gamma v \Rightarrow 0 = f(v - v) = f(v) - f(v) = (\lambda - \gamma)v \Rightarrow v = 0$, da $\lambda - \gamma \neq 0$.

4.) Sei m die Anzahl unterschiedlicher Eigenwerte. Wir führen den Beweis per Induktion über m . Der Fall $m = 1$ ist klar. Die Behauptung gelte für $m - 1$. Seien v_1, \dots, v_m Eigenvektoren zu den verschiedenen Eigenwerten $\lambda_1, \dots, \lambda_m$ und $\sum_{i=1}^m \alpha_i v_i = 0 \Rightarrow \sum_{i=1}^m \lambda_1 \alpha_i v_i = 0$. Wegen $A0 = 0$ gilt andererseits $\sum_{i=1}^m \lambda_i \alpha_i v_i = 0$, weil die v_i Eigenvektoren sind. Subtraktion liefert $0 = \sum_{i=2}^m \alpha_i (\lambda_i - \lambda_1) v_i = 0$. Die $m - 1$ Vektoren v_2, \dots, v_m sind aber nach Voraussetzung linear unabhängig, so dass $\alpha_i = 0, i \geq 2$ wegen $\lambda_i - \lambda_1 \neq 0$. Mit $v_1 \neq 0$ folgt noch $\alpha_1 = 0$ und so die lineare Unabhängigkeit der Eigenvektoren. ■

Zur Berechnung von Eigenwerten und Eigenvektoren beschränken wir uns zunächst auf $V = \mathbb{C}^n$ mit der Standardbasis. Dann lässt sich ein Endomorphismus f mit seiner Darstellungsmatrix A identifizieren. Aus (7.7) folgt

$$Ax = \lambda x \Leftrightarrow (A - \lambda E)x = 0. \quad (7.8)$$

Gleichung (7.8) hat genau dann nichttriviale Lösungen, wenn $\det(A - \lambda E) = 0$ gilt. Damit ist die entscheidende Bedingung an die Eigenwerte gefunden.

Satz 7.21: Sei $A \in \mathbb{C}^{n \times n}$.

1. Die Funktion

$$\chi_A(\lambda) := \det(A - \lambda E) \quad (7.9)$$

ist ein Polynom mit $\deg(\chi_A) = n$ und heißt *charakteristisches Polynom* von A .

2. $\lambda \in \mathbb{C}$ ist Eigenwert von $A \Leftrightarrow \chi_A(\lambda) = 0$.
3. A hat (mit Vielfachheit) genau n Eigenwerte $\lambda_i \in \mathbb{C}$.
4. $Eig(f; \lambda) = \ker(A - \lambda E)$.

BEWEIS 1.) folgt aus der Darstellung der Determinante durch die Leibniz-Formel. 2.) ist klar. 3.) entspricht Folgerung 3.94. 4.) gilt aufgrund Konstruktion von $Eig(f; \lambda)$ und Satz 4.18. ■

Möchte man zu einer konkreten Matrix $A \in \mathbb{C}^{n \times n}$ Eigenwerte und -vektoren berechnen, geht man (bei Handrechnung) so vor:

1. Man finde alle Nullstellen $\lambda_1, \lambda_2, \dots, \lambda_n$ von $\chi_A(\lambda) \in \mathbb{C}$. Diese sind die Eigenwerte von A ; mehrfache Eigenwerte sind möglich.
2. Zu einem λ_i bestimmt man die zugehörigen Eigenvektoren als Lösungen von $(A - \lambda_i E)x = 0$. Für jeden Eigenwert findet man *unendlich viele Lösungen*. Man muss somit so viele lineare Gleichungssysteme lösen, wie es verschiedene Eigenwerte gibt.
3. Zur effizienten Beschreibung von $\text{Eig}(A; \lambda)$ gebe man eine Basis an. Löst man das unterbestimmte Gleichungssystem $(A - \lambda_i E)x = 0$, ergibt sich diese implizit durch die Parametrisierung der Lösungsmenge.

Beispiel 7.22: Gesucht sind alle Eigenwerte und Eigenvektoren von

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

Wir beginnen mit den Eigenwerten und berechnen

$$\begin{aligned} \chi_A(\lambda) = |A - \lambda E| &= \begin{vmatrix} -\lambda & -1 & 1 \\ 0 & 2 - \lambda & 0 \\ -2 & -1 & 3 - \lambda \end{vmatrix} = (2 - \lambda) \begin{vmatrix} -\lambda & 1 \\ -2 & 3 - \lambda \end{vmatrix} \\ &= (2 - \lambda)((-\lambda)(3 - \lambda) + 2) \end{aligned}$$

durch eine Laplace-Entwicklung nach der zweiten Zeile. Man erhält $\lambda_1 = 2$ und aus dem zweiten Faktor $\lambda^2 - 3\lambda + 2$ die weiteren Eigenwerte $\lambda_2 = 2$ und $\lambda_3 = 1$.

Zur Bestimmung von $\text{Eig}(A; 2)$ lösen wir das homogene Gleichungssystem $(A - 2E)x = 0$, also

$$\begin{array}{ccc|c} -2 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ -2 & -1 & 1 & 0 \end{array} \longleftrightarrow \begin{array}{ccc|c} -2 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ -0 & -0 & 0 & 0 \end{array}.$$

Aus $-2x_1 - x_2 + x_3 = 0 \Leftrightarrow 2x_1 + x_2 = x_3$ erhält man

$$\text{Eig}(A; 2) = \left\{ \begin{pmatrix} \alpha \\ \beta \\ 2\alpha + \beta \end{pmatrix}, \alpha, \beta \in \mathbb{C} \right\} = \left\{ \alpha \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \alpha, \beta \in \mathbb{C} \right\};$$

für den reellen Teil des Eigenraums wählt man die Koeffizienten $\alpha, \beta \in \mathbb{R}$. Die gesuchte Basis des Eigenraums ist z. B. $((1, 0, 2)^T, (0, 1, 1)^T)$. Alle nicht-trivialen Linearkombinationen dieser Eigenvektoren sind allerdings auch

Eigenvektoren zum Eigenwert 2.

Analog verfährt man für den Eigenwert 1. Man erhält hier Eigenvektoren der Form

$$\alpha \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad \alpha \in \mathbb{C} \setminus \{0\}.$$

Beispiel 7.23: Wir bestimmen die Eigenwerte und Eigenvektoren der Drehmatrix

$$A_\alpha = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}$$

aus Beispiel 7.2. Wir setzen $c := \cos \alpha$, $s := \sin \alpha$ und errechnen

$$\chi_{A_\alpha}(\lambda) = |A_\alpha - \lambda E| = (c - \lambda)^2 + s^2 = \lambda^2 - 2\lambda c + c^2 + s^2.$$

Mit $c^2 + s^2 = 1$ folgt

$$\begin{aligned} \chi_{A_\alpha}(\lambda) = 0 &\Leftrightarrow \lambda^2 - 2\lambda c + 1 = 0 \\ &\Leftrightarrow \lambda_{1,2} = c \pm \sqrt{c^2 - 1} \\ &\Leftrightarrow \lambda_{1,2} = \cos \alpha \pm i\sqrt{\sin^2 \alpha} = \cos \alpha \pm i|\sin \alpha| \end{aligned}$$

Lässt man den Betrag weg, ändert man höchstens die ohnehin willkürliche Reihenfolge der beiden Eigenwerte und erhält dann mit dem Eulerschen Satz $\lambda_{1,2} = e^{\pm i\alpha}$. Es gilt $\lambda_1 = \overline{\lambda_2}$ und $|\lambda_i| = 1$, weil beide Eigenwerte auf dem komplexen Einheitskreis liegen. Es existieren nur reelle Eigenwerte, wenn $\sin \alpha = 0$, also für $\alpha = 0$ oder $\alpha = 180^\circ$. Der erste Fall entspricht der identischen Abbildung, der zweite der Abbildung $f(x) = -x$, der Punktspiegelung um 0. Da es das Wesen einer Drehung ist, die Richtung von Vektoren zu ändern, können natürlich keine reellen Eigenwerte außer in den beiden Spezialfällen existieren, weil die dazu gehörigen reellen Eigenvektoren gerade ihre Richtung behalten müssten.

Wir berechnen die Eigenvektoren für $\sin \alpha \neq 0$:

$$\begin{aligned} A - \lambda_1 E | 0 &\longleftrightarrow \left(\begin{array}{cc|c} c - c - is & -s & 0 \\ s & c - c - is & 0 \end{array} \right) \\ &\longleftrightarrow \left(\begin{array}{cc|c} -is & -s & 0 \\ s & -is & 0 \end{array} \right) \longleftrightarrow \left(\begin{array}{cc|c} -i & -1 & 0 \\ 1 & -i & 0 \end{array} \right) \end{aligned}$$

Multiplikation der zweiten Zeile mit i und Addition der Zeilen ergibt

$$A - \lambda_1 E | 0 \longleftrightarrow \left(\begin{array}{cc|c} -i & -1 & 0 \\ i & 1 & 0 \end{array} \right) \longleftrightarrow \left(\begin{array}{cc|c} -i & -1 & 0 \\ 0 & 0 & 0 \end{array} \right)$$

und deswegen Eigenvektoren v_1 der Gestalt

$$v_1 = \mu \begin{pmatrix} 1 \\ -i \end{pmatrix}, \quad \mu \in \mathbb{C} \setminus \{0\}.$$

Analog bestimmt man die Eigenvektoren zu λ_2 zu

$$v_2 = v \begin{pmatrix} 1 \\ i \end{pmatrix}, \quad v \in \mathbb{C} \setminus \{0\}.$$

Man beachte, dass nur die Eigenwerte, nicht aber die Eigenvektoren vom Drehwinkel α abhängen, zudem gilt $v_1 \perp v_2$ in \mathbb{C}^2 , da $v_1 \cdot \overline{v_2} = 0$.

Bemerkung 7.24:

1. Dass die Eigenwerte der Drehung aus Beispiel 7.23 den Betrag eins besitzen, ist kein Zufall: Für jede lineare Isometrie f mit Eigenwert λ und Eigenvektor x gilt $\|x\| = \|f(x)\| = \|\lambda x\| = |\lambda| \|x\|$, also $|\lambda| = 1$.
2. Der „kritische“ Teil der Berechnung von Eigenwerten und -vektoren per Hand ist das Finden der Nullstellen des charakteristischen Polynoms. Für den Polynomgrad 3 und 4 kann man die überaus komplizierten *Cardanischen Formeln*¹ zur Nullstellenberechnung verwenden oder versuchen, Nullstellen zu erraten und dann mit Polynomdivision fortzufahren. Für Polynome vom Grad mindestens 5 existiert nach einem berühmten Resultat von Abel kein Rechenverfahren, mit dem sich mit endlich vielen Rechenschritten alle Nullstellen eines beliebigen Polynoms bestimmen ließen. Bei praktischen Eigenwertproblemen, hier gilt häufig $n \geq 5$, bescheidet man sich daher mit Näherungslösungen.

Wir kehren zum allgemeinen Fall eines Endomorphismus f zurück. Sei A die Darstellungsmatrix von f bezogen auf irgendeine Basis und B die Darstellungsmatrix bzgl. irgendeiner anderen Basis, weiter S die Matrix des Basiswechsels. Dann gilt nach Folgerung 4.82 $B = SAS^{-1}$. Es folgt

$$\begin{aligned} \chi_B(\lambda) = 0 &\Leftrightarrow |B - \lambda E| = 0 \\ &\Leftrightarrow |SAS^{-1} - \lambda SES^{-1}| = 0 = |S(A - \lambda E)S^{-1}| = 0 \\ &\Leftrightarrow |S| |A - \lambda E| |S^{-1}| = 0 \Leftrightarrow \chi_A(\lambda) = 0 \end{aligned}$$

wegen $|S| |S^{-1}| = |SS^{-1}| = 1$. Darstellungsmatrizen von f zu verschiedenen Basen haben somit alle die gleichen Eigenwerte. Um die Eigenwerte von f zu bestimmen, kann man also *irgendeine* Darstellungsmatrix A von f wählen und von dieser die Eigenwerte wie oben gezeigt bestimmen. Dies sind zugleich die Eigenwerte von f . Die Eigenvektoren der Darstellungsmatrix sind dann die Koordinaten der Eigenvektoren von f in der Basis, auf die sich die Darstellungsmatrix bezieht. Für einen formalen Beweis und vertiefte Theorie verweisen wir auf die Literatur [3]. Wir betrachten den Zusammenhang zwischen der Determinante einer Matrix und ihren Eigenwerten.

¹Gerolamo Cardano, 1501–1576; italienischer Arzt und Mathematiker

Satz 7.25: Für $A = (a_1, \dots, a_n) \in \mathbb{C}^{n \times n}$ mit Eigenwerten $\lambda_i, 1 \leq i \leq n$ gilt

$$\det(A) = \prod_{i=1}^n \lambda_i.$$

BEWEIS Man errechnet

$$\begin{aligned} \chi_A(\lambda) &= \det(A - \lambda E) \\ &= \det(a_1 - \lambda e_1, \dots, a_n - \lambda e_n) \\ &= \det(a_1, a_2 - \lambda e_2, \dots, a_n - \lambda e_n) + \lambda \det(-e_1, a_2 - \lambda e_2, \dots, a_n - \lambda e_n) \\ &= \dots = \det(a_1, \dots, a_n) + \dots + \lambda^n \det(-e_1, \dots, -e_n) \\ &= (-1)^n \lambda^n + \dots + \det(A). \end{aligned}$$

Andererseits gilt

$$\chi_A(\lambda) = c \prod_{i=1}^n (\lambda - \lambda_i) = c \lambda^n + \dots + c(-1)^n \prod_{i=1}^n \lambda_i.$$

Ein Vergleich der führenden Koeffizienten liefert $c = (-1)^n$, damit liefert ein Vergleich der Koeffizienten des Absolutglieds die Behauptung. ■

Für allgemeine Matrizen kann es durchaus schwierig sein, die Eigenwerte zu finden. Ganz anders ist das bei Dreiecksmatrizen.

Bemerkung 7.26: Die Eigenwerte einer Dreiecksmatrix sind die Werte auf der Hauptdiagonalen.

BEWEIS Mit A ist auch $A - \lambda E$ eine Dreiecksmatrix mit Einträgen $a_{ii} - \lambda$ auf der Hauptdiagonalen. Mit Folgerung 5.14 erhält man die Behauptung für eine obere Dreiecksmatrix und durch Übergang zu $(A - \lambda E)^T$ auch für eine untere Dreiecksmatrix. ■

Bemerkung 7.27: Sei x ein Eigenvektor zum Eigenwert λ von A . Dann gilt

$$A^k x = \lambda^k x.$$

BEWEIS

$$A^k x = A^{k-1} A x = A^{k-1} (\lambda x) = \lambda A^{k-1} x = \dots = \lambda^k x \quad \blacksquare$$

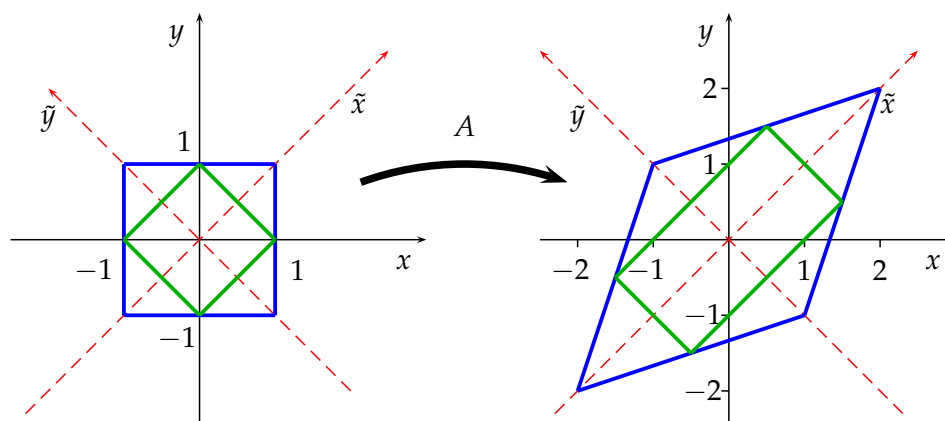


Abbildung 7.1: Quadrate unter der linearen Abbildung f aus den Beispielen 4.46, 4.84 und 7.29

7.4 Diagonalisierung linearer Abbildungen

Am Beginn des Kapitels 7.3 stand die Überlegung, dass eine Basis \mathcal{B} aus Eigenvektoren einer linearen Abbildung f zu einer diagonalen Darstellungsmatrix $D = M_{\mathcal{B}}^{\mathcal{B}}(f)$ führt, auf deren Hauptdiagonale die Eigenwerte λ_i stehen und die daher sehr einfach zu interpretieren ist. Sei A die Darstellungsmatrix von f bzgl. irgendeiner Basis \mathcal{B} und B die Matrix des Basiswechsels von \mathcal{B} nach $\tilde{\mathcal{B}}$. Nach Folgerung 4.82 gilt dann $D = B^{-1}AB$ bzw. $A = BDB^{-1}$. Offen ist die Frage, ob stets eine Basis aus Eigenvektoren existiert.

Definition 7.28: Ein $A \in \mathbb{C}^{n \times n}$ heißt *diagonalisierbar* $:\Leftrightarrow \exists B \in GL(n; \mathbb{C}), D$ Diagonalmatrix mit

$$A = BDB^{-1}. \quad (7.10)$$

Offenbar ist A genau dann diagonalisierbar, wenn eine Basis aus Eigenvektoren existiert. Bildet man mit ihnen als Spalten eine Matrix B , dann ist B genau die oben diskutierte Matrix des Basiswechsels.

Beispiel 7.29: Die lineare Abbildung $f(x) = Ax$ mit

$$A = \begin{pmatrix} 1,5 & 0,5 \\ 0,5 & 1,5 \end{pmatrix}$$

wurde schon in den Beispielen 4.46 und 4.84 diskutiert. Geometrische Überlegungen (vgl. Abb. 7.1) ließen uns eine um 45° gedrehte Basis $\tilde{\mathcal{E}} = (\tilde{e}_1, \tilde{e}_2)$ wählen, bzgl. derer die Darstellungsmatrix

$$\tilde{A} = T_{\tilde{\mathcal{E}}}^{\mathcal{E}} A T_{\tilde{\mathcal{E}}}^{\tilde{\mathcal{E}}} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \quad (7.11)$$

Diagonalgestalt besitzt, wobei $T_{\tilde{\mathcal{E}}}^{\mathcal{E}}$ die Matrix des Basiswechsels von der Standardbasis \mathcal{E} nach $\tilde{\mathcal{E}}$ bezeichnet. Damit ist A diagonalisierbar, denn mit $B = T_{\tilde{\mathcal{E}}}^{\mathcal{E}}$ und $\tilde{A} = D$ hat man $D = B^{-1}AB$, also $A = BDB^{-1}$.

Wir diskutieren dieses Beispiel im Licht der neu entwickelten Theorie. Es gilt

$$\chi_A(\lambda) = \left| \begin{pmatrix} 1,5 - \lambda & 0,5 \\ 0,5 & 1,5 - \lambda \end{pmatrix} \right| = (1,5 - \lambda)^2 - 0,25 = \lambda^2 - 3\lambda + 2$$

mit den Nullstellen $\lambda_1 = 2$ und $\lambda_2 = 1$. Wir berechnen $Eig(A; \lambda_1)$:

$$\begin{array}{ccc|c} 1,5 - 2 & 0,5 & 0 & \\ 0,5 & 1,5 - 2 & 0 & \end{array} \longleftrightarrow \begin{array}{ccc|c} -0,5 & 0,5 & 0 & \\ 0,5 & -0,5 & 0 & \end{array} \longleftrightarrow \begin{array}{ccc|c} -0,5 & 0,5 & 0 & \\ 0 & 0 & 0 & \end{array}$$

$$\Rightarrow Eig(A; 2) = \left\{ \alpha \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \alpha \in \mathbb{C} \right\}$$

Analog dazu erhält man

$$Eig(A; 1) = \left\{ \beta \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \beta \in \mathbb{C} \right\}.$$

Man erkennt $Eig(A; 1) \perp Eig(A; 2)$. Wählen wir die normierten Eigenvektoren $\tilde{e}_1 = 1/\sqrt{2}(1, 1)^T$ und $\tilde{e}_2 = 1/\sqrt{2}(-1, 1)^T$, dann ist

$$T_{\tilde{\mathcal{E}}}^{\mathcal{E}} = B = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

orthogonal, und es gilt

$$A = BDB^{-1} = BDB^T$$

mit

$$D = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} = \tilde{A},$$

was äquivalent zu (7.11) ist. Entscheidend ist aber, dass die Eigenwerttheorie einen systematischen Weg legt, um eine Basis wie $\tilde{\mathcal{E}}$ zu finden. Die algebraisch gefundene Zerlegung $A = BDB^T$ erlaubt zudem eine geometrische Interpretation von $f(x) = Ax$. Die Abbildung f wirkt in drei Schritten:

1. Durch $B^T = B^{-1}$ Drehung um 45° gegen den Uhrzeigersinn, denn wegen $|B| = 1$ gilt $B \in SO(2)$, also auch $B^T \in SO(2)$. Es handelt sich also um eine Drehmatrix; für den Drehwinkel α gilt $\cos \alpha = 1/\sqrt{2}$, also $\alpha = 45^\circ$.
2. Durch D Skalierung längs der gedrehten x_1 -Achse um 2, die gedrehten x_2 -Koordinaten bleiben unverändert.

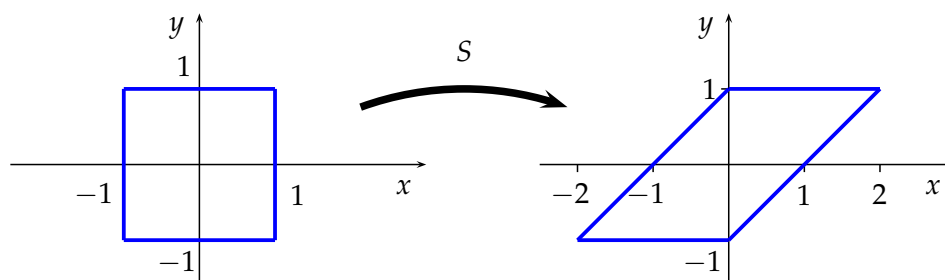


Abbildung 7.2: Einheitsquadrat unter der linearen Abbildung s aus Beispiel 7.30

3. Durch B Rückdrehung um 45° im Uhrzeigersinn.

Nicht jede Matrix lässt sich diagonalisieren.

Beispiel 7.30: Sei $s(x) = Sx$ mit

$$S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

An Abbildung 7.2 erkennt man, dass es sich bei s um eine *Scherung* handelt, die die x_1 -Achse invariant lässt. Alle anderen Vektoren ändern unter s ihre Richtung, können also keine Eigenvektoren sein. Aufgrund von $\chi_S(\lambda) = (1 - \lambda)^2$ ist 1 doppelte Nullstelle von χ_S und einziger Eigenwert von S . Die Eigenvektoren errechnet man zu

$$\begin{array}{l|l} 0 & 1 & 0 \\ 0 & 0 & 0 \end{array} \Rightarrow \text{Eig}(S; 1) = \left\{ \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, \alpha \in \mathbb{C} \right\}.$$

Dieses Ergebnis entspricht genau der geometrischen Anschauung oben. Damit ist S nicht diagonalisierbar, weil alle Eigenvektoren auf der x -Achse liegen und man so keine Basis von \mathbb{C}^2 aus Eigenvektoren bilden kann.

Wir diskutieren nun Kriterien zur Diagonalisierbarkeit. Weil Eigenvektoren zu verschiedenen Eigenwerten nach Satz 7.20 linear unabhängig sind, gilt:

Satz 7.31: Sei $A \in \mathbb{C}^{n \times n}$, und es mögen n verschiedene Eigenwerte existieren. Dann ist A diagonalisierbar.

Bemerkung 7.32: Satz 7.31 garantiert nicht die Diagonalisierbarkeit in \mathbb{R} , d. h. eine Zerlegung einer reellwertigen Matrix A in ausschließlich reelle Matrizen. So besitzt die Drehmatrix A_α aus Beispiel 7.23 die beiden verschiedenen komplexen Eigenwerte $e^{\pm i\alpha}$, und die zugehörigen Eigenvektoren sind komplexwertig.

Nicht-Diagonalisierbarkeit ist nach Satz 7.31 nur dann zu befürchten, wenn mehrfache Eigenwerte auftauchen. Wir untersuchen diesen Fall näher.

Definition 7.33: Sei $A \in \mathbb{C}^{n \times n}$ und λ ein Eigenwert. Die Vielfachheit der Nullstelle λ von χ_A heie *algebraische Vielfachheit* $a(\lambda)$, weiter sei $g(\lambda) := \dim(\text{Eig}(A; \lambda))$ die *geometrische Vielfachheit* von λ .

Satz 7.34: Sei A wie oben. Existiert ein Eigenwert $\tilde{\lambda}$ mit $a(\tilde{\lambda}) > g(\tilde{\lambda})$, dann ist A nicht diagonalisierbar.

BEWEIS Wir zeigen zunchst $a(\lambda) \geq g(\lambda)$. Wie in Kapitel 7.3 gesehen, sind Eigenwerte unabhngig von der Wahl der Basis und deswegen ebenso $a(\lambda)$. Weiterhin lsst ein Basiswechsel die Dimension von Unterrumen unverndert, daher hngt auch $g(\lambda)$ nicht von der Wahl der Basis ab. Sei $g(\lambda) = k$ und (v_1, \dots, v_k) eine Basis von $\text{Eig}(A; \lambda)$. Wir ergnzen sie mithilfe des Basisergnzungssatzes 3.57 zu einer Basis \mathcal{B} von \mathbb{C}^n . Es seien die Spalten von $B \in \mathbb{C}^{n \times n}$ die Vektoren in \mathcal{B} . Nach Konstruktion folgt dann

$$BAB^{-1} = \left(\begin{array}{ccc|c} \lambda & & & 0 \\ & \ddots & & \vdots \\ & & \lambda & 0 \\ \hline 0 & \cdots & 0 & * \end{array} \right).$$

Es ist $\chi_{BAB^{-1}}(x) = |BAB^{-1} - xE| = (\lambda - x)^k p(x)$ mit einem geeigneten Polynom p , und deswegen ist λ mindestens eine k -fache Nullstelle von $\chi_{BAB^{-1}}$, somit $a(\lambda) \geq k = g(\lambda)$. Nach dem Fundamentalsatz der Algebra gilt $\sum_{\lambda \text{ EW}} a(\lambda) = n$, also mit $a(\lambda) \geq g(\lambda)$ die Aussage $\sum_{\lambda \text{ EW}} g(\lambda) = \sum_{\lambda \text{ EW}} \dim(\text{Eig}(A; \lambda)) < n$. Damit kann man keine n linear unabhngigen Eigenvektoren whlen, und es existiert keine Basis aus Eigenvektoren. ■

Beispiel 7.35: Im Beispiel 7.30 gilt fr den einzigen Eigenwert $\lambda = 1$, dass $a(1) = 2 > g(1) = 1$. In der Tat war S nicht diagonalisierbar.

Viel bersichtlicher als bei allgemeinen quadratischen Matrizen wird die Situation bei symmetrischen Matrizen, die wir nachfolgend betrachten.

Satz 7.36: Eigenwerte reeller symmetrischer Matrizen sind immer reell. Zu jedem Eigenwert einer reellen symmetrischen Matrix existieren reelle Eigenvektoren.

BEWEIS Sei $\lambda \in \mathbb{C}$ ein Eigenwert von $A \in \mathbb{R}^{n \times n}$, $v \in \mathbb{C}^n \setminus \{0\}$ ein Eigenvektor zu λ und $\langle \cdot, \cdot \rangle$ das Standardskalarprodukt in \mathbb{C}^n . Wir benutzen $\overline{\overline{Av}} = \overline{A\overline{v}}$, was man fr reellwertige Matrizen A direkt ausrechnet. Man erhlt

$$\begin{aligned} \lambda \langle v, v \rangle &= (\lambda v)^T \overline{v} = (Av)^T \overline{v} = v^T A^T \overline{v} \\ &= v^T A \overline{v} = v^T \overline{A\overline{v}} = \langle v, \lambda v \rangle = \overline{\lambda} \langle v, v \rangle \end{aligned}$$

und daher $\lambda \in \mathbb{R}$.

Sei $v = x + iy$ ein Eigenvektor zu λ und $x \neq 0$. Dann folgt $Av = Ax + iAy = \lambda x + i\lambda y$. Wegen $Ax, Ay, \lambda x, \lambda y \in \mathbb{R}^n$ sind hier Real- und Imaginärteil vollständig separiert, so dass Gleichheit sowohl für Real- als auch Imaginärteil gilt. Daher gilt $Ax = \lambda x$, und x ist der gesuchte reellwertige Eigenvektor. Für $x = 0$ wähle man $y \neq 0$. ■

Satz 7.37: Sei $A \in \mathbb{R}^{n \times n}$ symmetrisch und $\lambda \neq \mu$ zwei Eigenwerte von A mit Eigenvektoren v bzw. w . Dann gilt $v \perp w$.

BEWEIS Sei $v \in \mathbb{R}^n$ ein Eigenvektor von A zum Eigenwert λ , $w \in \mathbb{R}^n$ ein Eigenvektor von A zum Eigenwert μ und $\lambda \neq \mu$. Man errechnet

$$\langle w, Av \rangle = w^T Av = (A^T w)^T v = (Aw)^T v = \langle Aw, v \rangle = \langle v, Aw \rangle.$$

Es folgt $\lambda \langle v, w \rangle = \langle Av, w \rangle = \langle v, Aw \rangle = \mu \langle v, w \rangle$ und deswegen $(\lambda - \mu) \langle v, w \rangle = 0$. Wegen $\lambda \neq \mu$ erhält man $\langle v, w \rangle = 0$. ■

Wir kommen nun zum Hauptergebnis dieses Kapitels.

Satz 7.38: Zu jeder reellsymmetrischen Matrix A gibt es eine Orthogonalmatrix Q und eine Diagonalmatrix D mit

$$A = QDQ^T.$$

Auf der Hauptdiagonalen von D stehen die Eigenwerte von A .

BEWEIS Wir führen den Beweis per Induktion über n . Der Fall $n = 1$ ist klar, weil jede 1×1 -Matrix bereits eine Diagonalmatrix ist. Die Behauptung gelte nun für $(n - 1) \times (n - 1)$ -Matrizen, gegeben sei $A \in \mathbb{R}^{n \times n}$. Nach dem Fundamentalsatz der Algebra und Satz 7.36 existiert ein reeller Eigenvektor v_1 mit $\|v_1\| = 1$ zum Eigenwert $\lambda_1 \in \mathbb{R}$. Nach Satz 3.140 existiert eine Orthonormalbasis v_2, \dots, v_n des orthogonalen Komplements v_1^\perp in \mathbb{R}^n von v_1 (vgl. Def. 3.123). Es folgt $V = (v_1, \dots, v_n) \in O(n)$. Da in den Spalten der Matrix die Koordinaten der Bilder der Basisvektoren stehen, hat man

$$\tilde{A} := V^T AV = \left(\begin{array}{c|c} \lambda_1 & \\ \hline 0 & * \\ \vdots & \end{array} \right)$$

Wegen $\tilde{A}^T = (V^T AV)^T = V^T A^T V = A$ ist \tilde{A} symmetrisch. Daher erkennt man

$$\tilde{A} = \left(\begin{array}{c|c} \lambda_1 & 0 \\ \hline 0 & B \end{array} \right)$$

mit irgendeiner symmetrischen Matrix $B \in \mathbb{R}^{(n-1) \times (n-1)}$. Nach Induktionsvoraussetzung existieren dann aber $W \in O(n-1)$ und eine Diagonalmatrix $D_{n-1} \in \mathbb{R}^{(n-1) \times (n-1)}$ mit $B = WD_{n-1}W^T$. Wir definieren

$$Q = V \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & W \end{array} \right).$$

Dann ist $Q \in O(n)$ und

$$AQ = AV \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & W \end{array} \right),$$

also

$$Q^T AQ = \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & W^T \end{array} \right) V^T AV \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & W \end{array} \right)$$

und weiter

$$\begin{aligned} Q^T AQ &= \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & W^T \end{array} \right) \cdot \left(\begin{array}{c|c} \lambda_1 & 0 \\ \hline 0 & B \end{array} \right) \cdot \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & W \end{array} \right) \\ &= \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & W^T \end{array} \right) \cdot \left(\begin{array}{c|c} \lambda_1 & 0 \\ \hline 0 & BW \end{array} \right) \\ &= \left(\begin{array}{c|c} \lambda_1 & 0 \\ \hline 0 & W^T BW \end{array} \right) = \left(\begin{array}{c|c} \lambda_1 & 0 \\ \hline 0 & D_{n-1} \end{array} \right). \end{aligned}$$

Somit ist $Q^T AQ \in \mathbb{R}^{n \times n}$ eine Diagonalmatrix, die auf der Diagonalen die Eigenwerte λ_i von A enthält. ■

Beispiel 7.39: Für die symmetrische Matrix

$$A = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & 1 \end{pmatrix}$$

existieren nach Satz 7.38 drei orthonormale Vektoren v_1, v_2, v_3 zu den drei reellen Eigenwerten $\lambda_1, \lambda_2, \lambda_3$, so dass mit $Q = (v_1, v_2, v_3)$ gilt:

$$Q^T AQ = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}.$$

Wir berechnen zunächst die Eigenwerte:

$$\begin{aligned} \left| \begin{array}{ccc|c} 1-\lambda & -1 & -1 & \\ -1 & 1-\lambda & 1 & \\ -1 & 1 & 1-\lambda & \end{array} \right| & \begin{array}{l} + \text{(II)} \\ - \text{(III)} \end{array} = \left| \begin{array}{ccc|c} -\lambda & -\lambda & 0 & \\ 0 & -\lambda & \lambda & \\ -1 & 1 & 1-\lambda & \end{array} \right| \\ & = \lambda^2(1-\lambda) + \lambda^2 - (-\lambda^2) \\ & = \lambda^2(1-\lambda+2) = 0, \end{aligned}$$

also $\lambda_1 = 3$ und $\lambda_{2,3} = 0$. Wir bestimmen die Eigenvektoren zu $\lambda_1 = 3$:

$$\begin{array}{ccc|c} -2 & -1 & -1 & 0 \\ -1 & -2 & 1 & 0 \\ -1 & 1 & -2 & 0 \end{array} \begin{array}{l} | \cdot 2; -\text{(I)} \\ | \cdot 2; -\text{(I)} \end{array} \leftrightarrow \begin{array}{ccc|c} -2 & -1 & -1 & 0 \\ 0 & -3 & 3 & 0 \\ 0 & 3 & -3 & 0 \end{array} \leftrightarrow \begin{array}{ccc|c} -2 & -1 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array} \leftrightarrow \begin{array}{ccc|c} -2 & 0 & -2 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}$$

Man liest $-x_1 = x_2 = x_3$ für die Lösungsmenge ab. Als normierten Eigenvektor kann man z. B. $v_1 = 1/\sqrt{3}(-1, 1, 1)^T$ wählen.

Wir betrachten $\lambda_{2/3} = 0$: Das lineare Gleichungssystem für die Eigenvektoren lautet

$$\begin{array}{ccc|c} 1 & -1 & -1 & 0 \\ -1 & 1 & 1 & 0 \\ -1 & 1 & 1 & 0 \end{array}$$

Alle Lösungen genügen daher der Gleichung $x_1 - x_2 - x_3 = 0$, die Dimension des Lösungsraumes ist 2. Wir wählen zwei linear unabhängige Lösungen, z. B.

$$\tilde{v}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \quad \text{und} \quad \tilde{v}_3 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}.$$

Eine Orthonormalisierung mit dem Verfahren von Gram-Schmidt liefert

$$v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad v_3 = \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ 1 \\ -2 \end{pmatrix}.$$

Damit ist

$$Q = \begin{pmatrix} -1/\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -2/\sqrt{6} \end{pmatrix}$$

und

$$D = Q^T A Q = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Es gilt die Umkehrung von Satz 7.38.

Bemerkung 7.40: Gibt es zu $A \in \mathbb{R}^{n \times n}$ ein $Q \in O(n)$ und eine Diagonalmatrix D mit $A = QDQ^T$, dann ist A symmetrisch.

BEWEIS

$$A^T = (QDQ^T)^T = (Q^T)^T D^T Q^T = QDQ^T = A$$

Wir diskutieren noch ein weiteres hinreichendes Kriterium zur Diagonalisierbarkeit.

Definition 7.41: $A \in \mathbb{C}^{n \times n}$ heißt *normal*, wenn gilt:

$$A\bar{A}^T = \bar{A}^T A. \quad (7.12)$$

Satz 7.42: Sei $A \in \mathbb{C}^{n \times n}$ normal. Dann existiert eine bzgl. des Standardskalarprodukts in \mathbb{C}^n orthonormale Basis aus Eigenvektoren, d. h. A ist diagonalisierbar.

Für den Beweis verweisen wir auf die Literatur (z. B. [3]).

Folgerung 7.43:

1. Jede reelle symmetrische Matrix ist diagonalisierbar. Die Eigenwerte sind reell.
2. Jede reelle antisymmetrische Matrix (d. h. $A^T = -A$) ist diagonalisierbar. Die Eigenwerte sind rein imaginär.
3. Jede reelle orthogonale Matrix ist diagonalisierbar.

BEWEIS 1.) Sei $A \in \mathbb{R}^{n \times n}$. Weil A reellwertig ist, hat man $A = \bar{A}$. Mit $A = A^T$ folgt die Normalität und nach Satz 7.42 die Diagonalisierbarkeit. Nach Satz 7.36 sind die Eigenwerte reell.

2.) Aufgrund der Antisymmetrie gilt $A^T = -A$. Mit $\bar{A} = A$ folgt $A\bar{A}^T = -A^2 = \bar{A}^T A$, also ist A normal. Analog zum Beweis des Satzes 7.36 errechnet man $\lambda = -\bar{\lambda}$ für einen Eigenwert λ , der deswegen keinen Realteil besitzen kann.

3.) Für $Q \in O(n)$ gilt $\bar{Q} = Q$ und weiter $Q\bar{Q}^T = QQ^{-1} = E = Q^{-1}Q = \bar{Q}^T Q$, also ist Q normal und daher diagonalisierbar. ■

7.5 Definitheit und Skalarprodukte

Die zentrale Rolle des Skalarprodukts für die analytische Geometrie hat sich im Laufe der Vorlesung hinreichend gezeigt. Es ist bekannt, dass auf \mathbb{R}^n neben dem bekannten Standardskalarprodukt viele Skalarprodukte existieren, nicht bekannt ist bisher aber eine systematische Methode, ein Skalarprodukt zu konstruieren. Dies holen wir hier nach. Für ein Skalarprodukt $(\cdot, \cdot) : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ müssen drei Bedingungen erfüllt sein:

1. (\cdot, \cdot) ist linear in den Spalten.
2. (\cdot, \cdot) ist symmetrisch.
3. $(x, x) > 0$ für $x \neq 0$.

Wir wählen eine beliebige Matrix $A \in \mathbb{R}^{n \times n}$ und betrachten die Abbildung $(\cdot, \cdot)_A : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$, $(x, y)_A := \langle x, Ay \rangle$, wobei $\langle \cdot, \cdot \rangle$ für das Standardskalarprodukt steht. Die Bedingung 1. ist für jede Matrix $A = (a_1, \dots, a_n)$ nach den Rechenregeln für Matrix-Vektorprodukt und Skalarprodukt erfüllt. Mit den Einheitsvektoren e_i und e_j erhält man

$$(e_i, e_j)_A = e_i^T A e_j = e_i^T a_j = a_{ij},$$

also $(e_i, e_j)_A = (e_j, e_i)_A \Leftrightarrow a_{ij} = a_{ji}$. Die Abbildung $(\cdot, \cdot)_A$ ist somit genau dann symmetrisch, wenn A symmetrisch ist.

Bemerkung 7.44: Die Einschränkung eines Skalarprodukts auf die Struktur $(\cdot, \cdot)_A$ ist in Wahrheit keine Einschränkung: Sei $[\cdot, \cdot]$ irgendein Skalarprodukt und A definiert durch $a_{ij} = [e_i, e_j]$. Dann gilt $[\cdot, \cdot] = \langle x, Ay \rangle$, jedes Skalarprodukt wird somit durch eine Matrix vollständig beschrieben.

BEWEIS Mit $x = (x_i)_{i=1}^n, y = (y_i)_{i=1}^n \in \mathbb{R}^n$ gilt

$$\begin{aligned} [x, y] &= \left[\sum_{i=1}^n x_i e_i, \sum_{j=1}^n y_j e_j \right] = \sum_{i=1}^n \sum_{j=1}^n x_i y_j [e_i, e_j] \\ &= \sum_{i=1}^n x_i \underbrace{\left(\sum_{j=1}^n a_{ij} y_j \right)}_{=(Ay)_i} = \langle x, Ay \rangle. \end{aligned}$$

Natürlich definiert nicht jede symmetrische Matrix ein Skalarprodukt, so etwa die Nullmatrix nicht. Wir kommen zu der Frage, welche Matrizen das obige dritte Kriterium erfüllen.

Definition 7.45: Sei A eine symmetrische Matrix $A \in \mathbb{R}^{n \times n}$.

1. Die Abbildung $x \rightarrow \langle x, Ax \rangle$ wird *quadratische Form* genannt.
2. A heißt *positiv definit*, wenn $\langle x, Ax \rangle > 0 \forall x \in \mathbb{R}^n \setminus \{0\}$.
3. A heißt *negativ definit*, wenn $\langle x, Ax \rangle < 0 \forall x \in \mathbb{R}^n \setminus \{0\}$.
4. A heißt *positiv semidefinit*, wenn $\langle x, Ax \rangle \geq 0 \forall x \in \mathbb{R}^n \setminus \{0\}$.
5. A heißt *negativ semidefinit*, wenn $\langle x, Ax \rangle \leq 0 \forall x \in \mathbb{R}^n \setminus \{0\}$.

6. A heißt *indefinit*, falls sie weder positiv noch negativ (semi-)definit ist, d. h.

$$\exists x, y \in \mathbb{R}^n \setminus \{0\} : \langle x, Ax \rangle > 0 \quad \wedge \quad \langle y, Ay \rangle < 0.$$

7. Häufig kürzt man „symmetrisch positiv definit“ mit „spd“ ab.

Beispiel 7.46: Die quadratische Form trägt ihren Namen zu Recht. Es sei

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

$$\begin{aligned} \langle x, Ax \rangle &= \left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right\rangle \\ &= \left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} x_1 + 2x_2 \\ 2x_1 + 4x_2 \end{pmatrix} \right\rangle \\ &= x_1^2 + 2x_1x_2 + 2x_2x_1 + 4x_2^2 \\ &= x_1^2 + 4x_1x_2 + 4x_2^2 = (x_1 + 2x_2)^2 \end{aligned}$$

Folgerung 7.47: Die Abbildung $(\cdot, \cdot)_A$ ist genau dann ein Skalarprodukt, wenn A spd ist.

Es fehlen noch handhabbare Kriterien für die Definitheit einer symmetrischen Matrix. Diese werden nachfolgend entwickelt.

Definition 7.48: Für $A \in \mathbb{R}^{n \times n}$ seien A_k die links oben beginnenden $k \times k$ -Untermatrizen $A_k = (a_{ij})_{i,j=1}^k$ von A . Dann heißen $D_k = \det(A_k)$ die *Hauptunterdeterminanten* oder *Hauptminoren* von A .

Beispiel 7.49: Die vier Hauptminoren $D_1 - D_4$ von

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

sind die Determinanten der vier quadratischen Teilmatrizen, die das linke obere Element von A enthalten:

$$\begin{aligned} D_1 &= |1| = 1 \\ D_2 &= \begin{vmatrix} 1 & -1 \\ -1 & 3 \end{vmatrix} = 3 - 1 = 2 \\ D_3 &= \begin{vmatrix} 1 & -1 & 0 \\ -1 & 3 & 4 \\ 0 & 4 & -2 \end{vmatrix} = -6 - (16 - 2) = -20 \\ D_4 &= \begin{vmatrix} 1 & -1 & 0 & 2 \\ -1 & 3 & 4 & 0 \\ 0 & 4 & -2 & 1 \\ 2 & 0 & 1 & 1 \end{vmatrix} = 82 \end{aligned}$$

Satz 7.50: Für eine reelle symmetrische Matrix A sind äquivalent:

1. A ist positiv definit.
2. A besitzt nur positive Eigenwerte.
3. Alle Hauptminoren von A sind positiv.

Weiter ist A genau dann positiv semidefinit, wenn alle Eigenwerte von A nicht negativ sind. In diesem Fall sind alle Hauptminoren von A nicht negativ.

Für den Beweis und die dabei verwendete Theorie der *Hauptachsentransformationen* verweisen wir z. B. auf Fischer [3, Kap. 5.7].

Beispiel 7.51: Für

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

erhält man $D_1 = 2$ und $D_2 = 4 - 1 = 3$. Nach Satz 7.50 ist A positiv definit. In der Tat bestimmt man die Eigenwerte von A zu

$$\det(A - \lambda E) = \det \begin{pmatrix} 2 - \lambda & 1 \\ 1 & 2 - \lambda \end{pmatrix} = \lambda^2 - 4\lambda + 3 = 0,$$

also $\lambda_1 = 1$ und $\lambda_2 = 3$. Beide Eigenwerte sind positiv. Wir betrachten die quadratische Form:

$$\begin{aligned} \langle x, Ax \rangle &= \left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right\rangle \\ &= \left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} 2x_1 + x_2 \\ x_1 + 2x_2 \end{pmatrix} \right\rangle \\ &= 2x_1^2 + 2x_1x_2 + 2x_2^2 \\ &= (x_1 + x_2)^2 + x_1^2 + x_2^2 > 0, \quad x \neq 0 \end{aligned}$$

Bemerkung 7.52: Eine spd-Matrix $A \in \mathbb{R}^{n \times n}$ ist invertierbar.

BEWEIS Nach Satz 7.50.3 ist $\det(A) = D_n \neq 0$ und daher A invertierbar. ■

Bemerkung 7.53: Ist A negativ definit, dann gilt $-\langle x, Ax \rangle = \langle x, -Ax \rangle > 0$ d. h. $-A$ ist positiv definit. Analog erkennt man, dass A genau dann negativ semidefinit ist, wenn $-A$ positiv semidefinit ist.

Bemerkung 7.54: Für die Hauptminoren D_k einer negativ definiten Matrix gilt:

$$(-1)^k D_k = (-1)^k \det(A_k) = \det(-A_k) > 0, \quad (7.13)$$

weil $-A$ positiv definit ist. Es sind also die Hauptminoren abwechselnd positiv und negativ, beginnend mit negativem Vorzeichen.

Eine hinreichend oft differenzierbare Funktion $f: \mathbb{R}^n \rightarrow \mathbb{R}$ weist in einem Punkt x_0 sicher ein lokales Maximum auf, wenn $\nabla f(x_0) = 0$ gilt und die nach dem Satz von Schwarz stets symmetrische Hessematrix $\nabla^2 f$ in x_0 negativ definit ist. Dies wird üblicherweise mit dem Kriterium (7.13) geprüft. Man spricht in der Analysis vom *Hurwitz-Kriterium*.

Wir fassen die oben entwickelten Kriterien zur Definitheit in einer Tabelle zusammen.

	D_1	D_2	D_3	D_4	...	Kriterium
positiv definit	> 0	> 0	> 0	> 0	...	hinreichend
negativ definit	< 0	> 0	< 0	> 0	...	hinreichend
positiv semidefinit	≥ 0	≥ 0	≥ 0	≥ 0	...	notwendig
negativ semidefinit	≤ 0	≥ 0	≤ 0	≥ 0	...	notwendig
indefinit	sonst					hinreichend

Die letzte Spalte der Tabelle bezieht sich darauf, ob die Vorzeichenfolge nur notwendig oder aber hinreichend für die Art der Definitheit ist.

Bemerkung 7.55:

1. Die Matrix A aus Beispiel 7.49 ist aufgrund der Vorzeichenfolge der Hauptminoren $D_1 - D_4$ indefinit. In der Tat weist A 2 positive und zwei negative Eigenwerte auf ($\lambda_1 \approx -4,37, \lambda_2 \approx -1,12$ sowie $\lambda_3 \approx 3,10, \lambda_4 \approx 5,39$).
2. Die Kriterien zur Semidefinitheit sind nur notwendig: So ist

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

offenbar indefinit und nicht etwa positiv semidefinit, dennoch gilt aber $D_1 = D_2 = D_3 = 0 \geq 0$.

Wir betrachten abschließend ein spezielles Beispiel einer symmetrischen positiv definiten Matrix.

Bemerkung 7.56: Sei $m \geq n$ und $A \in \mathbb{R}^{m \times n}$ mit $\text{rg}(A) = n$. Dann ist $A^T A$ positiv definit. Für $\text{rg}(A) < n$ ist $A^T A$ positiv semidefinit.

BEWEIS Wegen $(A^T A)^T = A^T (A^T)^T = A^T A$ ist $A^T A$ symmetrisch. Weiter gilt

$$\langle x, A^T A x \rangle = x^T A^T A x = (Ax)^T Ax = \langle Ax, Ax \rangle = \|Ax\|^2 \geq 0.$$

Also ist $A^T A$ positiv semidefinit. Nach Satz 6.33 ist $A^T A$ invertierbar für $\text{rg}(A) = n$. Gäbe es ein $x \neq 0$ mit $\|Ax\| = 0$, dann folgte $Ax = 0$, also $A^T(Ax) = 0$ und damit $x \in \ker(A^T A)$. Dann aber wäre $A^T A$ nicht invertierbar. Also $\langle x, A^T Ax \rangle > 0 \forall x \neq 0$, und $A^T A$ ist spd. ■

Bemerkung 7.57: Für $A \in \mathbb{R}^{m \times n}$ mit $m \leq n$ ist $A^T A$ positiv semidefinit und sogar positiv definit, falls $\text{rg}(A) = m$. Dies zeigt man analog zu Bemerkung 7.56.

Index

- $GL(n; K)$, 152
- $Hom(V, W)$, 128
- $O(n)$, 202, 203
- QR-Zerlegung, 206
- $SO(n)$, 206
- Abbildung, 125
 - bijektiv, 126
 - Bild, 125
 - Definitionsmenge, 125
 - injektiv, 126
 - inverse, 126
 - linear, 128
 - orthogonale, 204
 - Scherung, 217
 - surjektiv, 126
 - Urbild, 125
 - Wertebereich, 125
 - wohldefiniert, 68
- Abstand, 57
- Additivität, 128
- Additivität (Determinante), 164
- Äquivalenzumformungen, 20
- algebraische Vielfachheit, 218
- allgemeine Ebenengleichung, 49
- allgemeine Geradengleichung, 47
- allgemeine Lösung, 182
- alternierend (Determinante), 164
- Assoziativität, 67
- Aufpunkt, 43
- Austauschlemma, 86
- Automorphismus, 134
- Basis, 84, 91
- Basisergänzungssatz, 85
- Basiswechsel, 158
- Betragssummennorm, 31
- Bijektivität, 126
- Bild, 125
- Cardanische Formeln, 213
- Cauchy-Schwarzsche Ungl., 107
- charakteristisches Polynom, 210
- Cramersche Regel, 190
- Definition
 - axiomatisch, 164
- Definitionsmenge, 125
- Determinante, 61, 163
 - Additivität, 164
 - Definition, 164
 - Gauß-Verfahren, 174
 - Homogenität, 164
 - Inversenberechnung, 178
 - Laplace-Entwicklung, 174
 - normiert, 164
 - stetig, 180
- Diagonalisierbarkeit, 215
- Diagonalisierung, 219
- Diagonalmatrix, 220
- Diffeomorphismus, 155
- Differenzpolynom, 100
- Differenzvektor, 16
- Dimension, 87
- Dimensionsformel, 133
- direkte Summe, 77
- Distributivgesetz, 73
- Drehmatrix, 139
- Dreiecksmatrix
 - obere, 18
 - untere, 18
- Dreiecksungleichung, 30, 37, 106

- Ebene, 13, 49
 - Normalform, 49
 - Punkt-Richtungsgleichung, 49
 - Richtungsvektor, 49
- Eigenraum, 209
- Eigenvektor, 209
- Eigenwert, 209
 - Dreiecksmatrizen, 214
 - Potenzen einer Matrix, 214
- eindeutige Lösbarkeit, 185
- Eindeutigkeit der Basislänge, 87
- Einernorm, 31
- Einheitsmatrix, 19
- Einheitssphäre, 32
- Einheitsvektor, 32
 - kanonisch, 32
- Eisen, 7
- Eisenerzeugung, 7
- Elementarmatrizen, 165
 - Definition, 166
- Endomorphismus, 209
- Entwicklungsprozess, 9
- Entwicklungssatz nach Laplace, 174
- Erzeugendensystem, 79, 90
 - minimal, 84, 91
- Euklidischer Vektorraum, 105

- Faktorzerlegung, 97
- Familie, 90
- Folge, 27
- Fries, 10

- Gauß-Algorithmus, 19
- Gauß-Jordan-Algorithmus, 19
- geometrische Vielfachheit, 218
- Gerade, 43
 - Aufpunkt, 43
 - Normalform, 46
 - Parameterform, 45
 - Punkt-Richtungsgleichung, 43
 - Richtungsvektor, 43
 - Zweipunktform, 45
- Geradengleichung
 - allgemeine, 47
- Gleichungssystem
 - äquivalent, 19
 - Äquivalenzumformungen, 20
 - allgemeine Lösung, 182
 - homogen, 181
 - inhomogen, 181
 - Koeffizientenmatrix, 137
 - linear, 19, 181
 - Lösbarkeit, 185
 - Lösung mit MATLAB, 21
 - reduzierte Stufenform, 20
 - spezielle Lösung, 182
 - Stufenform, 20
 - überbestimmtes, 23, 192
 - unterbestimmtes, 23, 198
- Graßmannsche Identität, 42
- Grad, 95
- Gram-Schmidt, 116
- Gruppe, 9, 67
 - abelsch, 68
 - Automorphismengruppe, 10
 - kommutativ, 68
 - orthogonale, 202, 203
 - Raumgruppe, 10
 - spezielle orthogonale, 206
 - triviale, 69
- Gruppenisomorphismus, 72

- Hauptachsentransformation, 225
- Hauptminor, 224
- Hauptunterdeterminante, 224
- Hessesche Normalform, 47
 - Ebene, 49
- homogenes Gleichungssystem, 181
- Homogenität, 128
- Homogenität (Determinante), 164
- Homomorphismus, 128
- Hurwitz-Kriterium, 226
- Hyperebene, 54, 91, 92
 - Aufpunkt, 92
 - Parameterform, 92

- indefinit, 224
- Indexmenge, 27

- inhomogenes Gleichungssystem, 181
- Injektivität, 126
- Interpolation, 100
- Interpolationspolynom, 100
- Inverse, 151
 - Berechnung, 153
 - verallgemeinerte, 194, 199
- Inverse Matrix, 151
- inverses Element, 67
- invertierbare Matrix, 185
- Invertierbarkeit, 126
- Isometrie, 201
- isomorph (Gruppe), 72
- Isomorphismus, 134
- Isomorphismus v. Vektorräumen, 143
 - $K^{m \times n}$, 75
 - Körper, 73
 - kanonische Einheitsvektoren, 32
 - Kartesisches Produkt, 27
 - Kern, 131
 - kleinste Quadrate, 194
 - Koeffizienten, 95
 - Koeffizientenmatrix, 17
 - erweiterte, 18
 - Körper, 73
 - kommutatives Diagramm, 158
 - komplementäre Matrix, 177
 - komplexe Ebene, 72
 - komplexer Vektorraum, 75
 - Komponente eines Vektors, 35
 - Koordinaten, 85, 156
 - Existenz, 156
 - Koordinatenabbildung, 156
 - Koordinatensystem, 13
 - Kreuzprodukt, 39
 - Kristallographie, 10
 - Kronecker-Symbol, 112
 - Kurbel, 8
- Laplace-Entwicklung, 174
- Legierung, 7
- Leibnizsche Formel, 173
- Leitkoeffizient, 95
- Lineare Abbildung, 128
- Lineare Hülle, 78
- lineare Regression, 192
- Lineare Unabhängigkeit, 79
- Lineares Gleichungssystem, 8, 16, 19,
 - 64, 181, 190, 199
 - überbestimmtes, 23
 - Koeffizienten, 17
 - Lösbarkeit, 182
 - unterbestimmtes, 23
- Linearkombination, 78
- Linkssystem, 42
- Lösungsmenge, 181
- Matrix, 17, 137
 - äquivalent, 19
 - Äquivalenzumformungen, 20
 - Diagonal-, 18
 - diagonalisierbar, 215
 - Einheits-, 19
 - Elemente, 17
 - Gleichheit, 17
 - Hauptdiagonale, 18
 - Hauptminor, 224
 - Hauptunterdeterminante, 224
 - indefinit, 224
 - Inverse, 151
 - invertierbar, 185
 - Koeffizienten-, 17
 - komplementär, 177
 - MATLAB, 21
 - Multiplikation, 148
 - normal, 222
 - Null-, 17
 - obere Dreiecks-, 18
 - orthogonal, 202
 - positiv definit, 223
 - quadratisch, 18
 - Rang, 181
 - semidefinit, 223
 - Spaltenindex, 17
 - spd, 224
 - transponiert, 17
 - untere Dreiecks-, 18

- Zeilenindex, 17
- Matrix-Vektor-Multiplikation, 137
- Matrixprodukt, 148
- Maximumnorm, 31
- Menge, 125
- Methode der kl. Quadrate, 194
- minimales Erzeugendensystem, 84, 91
- Monobasis, 99
- negativ definit, 223, 225
- negativ semidefinit, 223
- Neutralelement, 67
- Norm, 31, 106
 - Betragssummen-, 31
 - Einer-, 31
 - euklidische, 29
 - Maximum-, 31
- Normalenvektor, 46
 - Ebene, 49
- Normalform, 46
 - Ebene, 49
- Normalgleichungen, 194
- normiert (Determinante), 164
- Nullpolynom, 95
- numerische Instabilität, 208
- numerische Lineare Algebra, 9
- Numerische Simulation, 8
- Octave, 22
- Orthogonalbasis, 112
- orthogonale Abbildung, 204
- orthogonale Gruppe, 202, 203
- orthogonale Matrix, 202
- orthogonale Projektion, 36, 193
- orthogonales Komplement, 110
- Orthogonalität
 - Ebene, 53
 - von Geraden, 51
- Orthogonalität von Vektoren, 34
- Orthogonalsystem, 112
- Orthonormalbasis, 112
- Orthonormalisierungsverfahren, 116
- Orthonormalsystem, 112
- Ortsvektor, 14
- Parallelepiped, 63
- Parallelität
 - Ebene, 53
 - von Geraden, 51
 - von Vektoren, 16
- Parameterform, 45
- Permutation, 172
- physikalisches System, 8
- Pivot-Element, 23
- Pivot-Spalte, 23
- Polynom, 95
 - charakteristisches, 210
 - Faktorzerlegung, 97
 - Grad, 95
 - Interpolations-, 100
 - Koeffizienten, 95
- positiv definit, 223
- positiv semidefinit, 223
- Problemlösestrategien, 10
- Projektion, 35
 - orthogonale, 35
 - senkrechte, 35
- Prototyp, 8
- Punkt-Richtungsgleichung
 - Ebene, 49
 - Gerade, 43
- Punktprodukt, 28
- python, 21
- quadratische Form, 223
- Rückwärtsarbeiten, 11
- Rang, 133, 181
- Rechtssystem, 42
- reeller Vektorraum, 75
- Regression
 - linear, 192
- Restklasse, 70
- rg, 133
- Richtungsvektor, 43
- Sarrussche Regel, 62, 172
- Satz des Pythagoras, 34

- Scherung, 217
- Schnittmengen, 53
- Schwarzsche Ungleichung, 36
- semidefinit, 223
- Skalar, 75
- Skalarprodukt, 27, 28, 103
 - euklidisches, 28
- Spaltenrang, 181
- span, 79
- Spat, 63
- Spatprodukt, 62
- spezielle Lösung, 182
- Standardnorm, 29, 106
- Standardskalarprodukt, 28
- Starrkörperbewegungen, 10
- Stufenform, 20
- Summe (von Vektorräumen), 77
- Surjektivität, 126
- symmetrische Gruppe, 172
- Transformationsformel, 205
- Transformationsmatrix, 158
- transponierte Matrix, 168
- Transpositionsabbildung, 131
- überbestimmtes Gl.-system, 23, 192
- Umkehrabbildung, 126
- Unitärer Vektorraum, 105
- unterbestimmtes Gl.-system, 23, 198
- Untergruppe, 71
- Unterraum, 76
- Untervektorraum, 76
- Urbild, 125
- Vektor, 12, 13, 75
 - Addition, 14
 - Differenz-, 16
 - Einheits-, 32
 - Gleichheit, 13
 - Multiplikation mit Skalar, 15
 - normiert, 33
 - orthogonal, 34
 - parallel, 16
 - senkrecht, 34
 - transponiert, 13
- Winkel, 37
- Vektorprodukt, 39
- Vektorraum, 74
 - endlich erzeugt, 79
 - Isomorphismus, 143
 - komplexer, 75
 - reeller, 75
- verallgemeinerte Inverse, 194, 199
- Verknüpfung, 67
- Verknüpfungstafel, 70
- Vielfachheit
 - algebraische, 218
 - geometrische, 218
- Volumen, 63
- Vorwärtsarbeiten, 11
- Wertebereich, 125
- windschief, 56
- Winkel
 - Ebene, 53
 - zwischen Geraden, 52
- Winkel zwischen Vektoren, 37, 38
- Wirtschaftswissenschaften, 13
- Zahlraum, 13
- Zeilenrang, 188
- Zielmenge, 125
- Zweipunktform, 45
- Zwischenziele, 12

Literaturverzeichnis

- [1] A. Beutelspacher. *Das ist o. B. d. A. trivial!* Vieweg, 8 edition, 2006.
- [2] A. Beutelspacher. *Lineare Algebra*. Springer Spektrum, 8 edition, 2014.
- [3] G. Fischer. *Lineare Algebra*. Springer, 18 edition, 2014.
- [4] D. Grieser. *Mathematisches Problemlösen und Beweisen*. Springer Spektrum, 2013.
- [5] W. Kabbalo. *Einführung in die Analysis 1*. Spektrum akademischer Verlag, 2 edition, 2000.
- [6] G. Pólya. *Schule des Denkens*. Narr Francke Attempto, 2010. Sonderausgabe der 4. Auflage.