FSU Jena
Fakultät für Mathematik und Informatik

Lineare Algebra für IB, AIB, BIB

FMI-MA0022
Wintersemester 2022/23

Simon King
10. Februar 2023

Inhaltsverzeichnis

0 Vorrede **1**
- 0.1 Organisatorisches . . . 1
- 0.2 Vor- und Frühgeschichte der Mathematik . . . 2
- 0.3 Studium als Gegensatz zur Schule . . . 3
- 0.4 Prioritäten . . . 4
- 0.5 Folgerungen für das Lehrkonzept . . . 6
- 0.6 Rechnerische Anforderungen . . . 7

1 Grundlagen **8**
- 1.1 Naive Mengenlehre . . . 8
- 1.2 Sprachliche Ausdrucksmittel . . . 11
- 1.3 Beweisstrategieen . . . 12
  - 1.3.1 Direkte Beweise . . . 14
  - 1.3.2 Indirekte Beweise . . . 15
- 1.4 Mengenkonstruktionen . . . 18

2 Lineare Gleichungssysteme **22**
- 2.1 Matrixnotation . . . 22
- 2.2 Matrixarithmetik . . . 25
- 2.3 Lösungsräume . . . 29
- 2.4 Gauß-Elimination . . . 34

3 Begriffe der Algebra **38**
- 3.1 Gruppen . . . 38
- 3.2 Ringe und Körper . . . 40
- 3.3 Restklassenringe . . . 43
- 3.4 Komplexe Zahlen . . . 46
  - 3.4.1 Die Gaußsche Zahlenebene . . . 48
- 3.5 Warum Restklassenkörper? . . . 51

4 Vektorräume **53**
- 4.1 Vektorraumaxiome . . . 53
- 4.2 Lineare Abbildungen . . . 56
- 4.3 Untervektorräume . . . 59
- 4.4 Basen . . . 61
- 4.5 Isomorphismen . . . 64
- 4.6 Basisauswahl und -ergänzung . . . 67

5 Matrizen – Teil 2 **70**
- 5.1 Rang und Rangformel . . . 70
- 5.2 Determinanten . . . 71
  - 5.2.1 Invertierbarkeit . . . 77

6 Eigenwertprobleme **80**
6.1 Eigenwerte, -vektoren und -räume 80
6.2 Eigenräume sind komplementär 84
6.3 Basiswechsel 85

7 Euklidische Räume **88**
7.1 Skalarprodukte 88
7.2 Definitheitstest 90
7.3 Orthonormalbasen 91
7.3.1 Das orthogonale Komplement 94
7.4 Besondere Abbildungen 96
7.4.1 Orthogonale Abbildungen 96
7.4.2 Symmetrische Endomorphismen 100

Index **105**

Dies sind Vorlesungsnotizen, kein voll ausgearbeitetes Skript. Weitere Erklärungen gibt es nur in der Vorlesung.

Literaturhinweise: Zur linearen Algebra gibt es viele Lehrbücher in verschiedenen Sprachen. Hier sind drei in deutscher Sprache erschienene Standardwerke:

- Gerd Fischer, „Lineare Algebra“, vieweg+teubner
- Max Koecher, „Lineare Algebra und analytische Geometrie“, Springer
- Hans-Joachim Kowalski, „Lineare Algebra“, Walter de Gruyter

Ich empfehle Ihnen, in der Bibliothek weitere Lehrbücher zu suchen, um zu vergleichen und zu sehen, mit welchen Sie am besten zurecht kommen.

0 Vorrede

0.1 Organisatorisches

Prüfungszulassungsvoraussetzung ist das Erreichen von insgesamt mindestens 50% der nominellen Gesamtpunktzahl in den Hausaufgaben. Wenn am Ende wenige Punkte fehlen, können diese ggf. durch Bearbeitung weiterer Aufgaben ausgeglichen werden. Um die Prüfungsanmeldung (das ist etwas anderes als die *Modul*anmeldung) müssen Sie sich innerhalb einer Frist selbst kümmern.

Hausaufgaben erfordern einen höheren Aufwand, als Sie wahrscheinlich aus der Schule gewohnt sind. Die Bearbeitungszeit beträgt jeweils rund eine Woche. Hausaufgaben sollen dazu anregen, sich jeweils einige Stunden mit den Vorlesungsinhalten zu beschäftigen sowie mit Ihren Kommiliton*innen zu diskutieren. Die Hausaufgaben werden in den **Übungsgruppen** besprochen. Zudem sind jeweils Skriptabschnitte zur Vorbereitung der nächsten **Vorlesung** zu lesen. Übrigens: Laut Modulbeschreibung ist für selbständige Arbeit mindestens die doppelte Zeit wie für die Präsenzlehre vorgesehen.

Im Tutorium sollen Sie in Kleingruppen „Präsenzaufgaben“ bearbeiten, zudem dürfte auch noch Zeit für die Beantwortung von Fragen sein. Man sollte nicht erwarten, dass Präsenzaufgaben stets eine klare Lösung haben. Stattdessen geht es noch mehr als bei Hausaufgaben darum, diskutieren und argumentieren zu lernen, denn dies ist Grundlage der mathematischen Methode und daher meines Erachtens für den Studienerfolg entscheidend.

„Operatoren“ wie „erörtern“ oder „bestimmen“ sind Schlüsselreize im Sinne behavioristischer Lerntheorien. Obwohl schuldidaktische Operatoren erst 2004/5 in den einheitlichen Prüfungsanforderungen erwähnt wurden und bereits 2007 in den diese ersetzenden Bildungsstandards nicht mehr vorkamen, wurden Sie vermutlich jahrelang auf ihren Gebrauch konditioniert. Daher muss ich betonen: Schuldidaktische Operatoren spielen im MINT-Studium keine Rolle (außer für das Lehramt).[1] *Die Bewertungsgrundlage sind schlüssige Begründungen und Rechenwege.* Daher müssen gründsätzlich alle Antworten begründet und Rechnungen nachvollziehbar dargestellt werden. Bei einer Entscheidungsfrage wird eine korrekte Antwort mit null Punkten bewertet, wenn eine Begründung fehlt, und bei einer Rechenaufgabe wird das korrekte Ergebnis mit null Punkten bewertet, wenn der Rechenweg nicht erkennbar ist. Das gilt auch in Prüfungen, man sollte sich also in den Hausaufgaben schon daran gewöhnen.

[1] In der Mathematik steht das Wort „Operator“ für verschiedene Begriffe, die alle mit dem schuldidaktischen Begriff nichts zu tun haben.

0.2 Vor- und Frühgeschichte der Mathematik

> Der materielle Fortschritt befriedigt keines der Bedürfnisse, die der Mensch wirklich hat.
>
> Winston Churchill [1874-1965]

Arithmetik[2], ***Geometrie***[3] und ***Logik***[4] entsprechen meines Erachtens gesellschaftlichen Bedürfnissen. Zahlverständnis ist im Handel nützlich, der bereits im ***Paläolithikum*** nachweisbar ist. Gleichwohl ist Spekulation, bis zu welchem Grad sich Zahlverständnis durch Artefakte wie den 20 000 Jahre alten Ishango–Knochen oder sogar schon bei den Neandertalern (die aber wohl noch keinen Handel trieben) nachweisen lässt. Das ***Neolithikum*** (zuerst im Fruchtbaren Halbmond[5] ab ca. 9500 v. Chr.) ist durch Landwirtschaft, Vorratshaltung und Sesshaftigkeit charakterisiert. Für Landwirtschaft ist nicht nur Geometrie (wörtlich „Landmessung“), sondern auch Arithmetik in Form von Kalendern nützlich, und auch bei der Vorratshaltung hilft Arithmetik.

Ich halte es nicht für Zufall, dass sich mathematische Kenntnisse bald auch in anderen Lebensbereichen zeigten, etwa in geometrischen Verzierungen auf keramischen Vorratsgefäßen (z.B. Linearbandkeramik ab ca. 5700 v. Chr.), großen Häusern mit rechteckigem Grundriss (z.B. bandkeramisches Langhaus ab 5500 v. Chr.) oder Kreisgrabenanlagen mit astronomischem Bezug (z.B. Kreisgrabenanlage von Goseck 4900 v. Chr., Stonehenge ab 3100 v. Chr.). Doch sogar in den frühen Hochkulturen fehlte die eigentliche mathematische Methode: Anscheinend entstand das Bedürfnis nach Rationalität erst in späteren Gesellschaftsformen.

> Mathematik ist die Königin der Wissenschaften.
>
> Carl Friedrich Gauß [1777–1855]

Die *Methode* der Mathematik beruht auf schlüssigem Argumentieren auf Grundlage von Begriffsdefinitionen. Ich halte es nicht für Zufall, dass die Grundlagen der mathematischen Methode und der Demokratie als einer auf Argumentationsfähigkeit basierenden Gesellschaftsform etwa zur gleichen Zeit etwa am gleichen Ort gelegt wurden (Thales von Milet [ca. 624–547 v. Chr.] bzw. Solon [ca. 640–560 v. Chr. in Athen]; beide wurden zu den Sieben Weisen gezählt).

In den meisten Wissenschaften wird bisweilen schlüssig begrifflich argumentiert. In der Mathematik ist man im begrifflichen Argumentieren sehr konsequent, es ist hier anders als in Literatur- oder Naturwissenschaften die *einzige* Erkenntnisquelle. Dadurch wird die mathematische Arbeitsweise in den meisten Wissenschaften relevant, weshalb ich das obige Gauß-Zitat für berechtigt halte.

[2] Zahlen und Zahlverhältnissen. Nach griech. ἀριθμός: Zahl
[3] Maße. Nach griech. γεωμετρία: Landmessung
[4] Schlussfolgerungslehre. Nach griech. λογικὴ τέχνη: Kunst des Denkens/Argumentierens
[5] von Südirak über Nordsyrien, Libanon, Israel und Palästina bis Jordanien

0.3 Studium als Gegensatz zur Schule

Die Schule erzieht zur Dummheit.

Noam Chomsky [geb. 1928]

In der Einführungsveranstaltung meines Studiums in Frankfurt/Main brachte die Fachschaft den Studienanfänger*innen unter anderem bei:

a) Die Studienmotivation sollte Interesse an den Fächern sein. „Ich studiere das Fach, weil ich in der Schule gut darin war“ oder „Mit diesem Fach bekomme ich später einen gut bezahlten Job“ bringt auf Dauer nichts.

b) In den ersten drei bis vier Wochen rauscht in den Vorlesungen der gesamte mathematische Schulstoff vorüber.

c) Die Inhalte des n-ten Semesters versteht man im $(n+2)$-ten Semester.

Früher war also nicht alles besser. Erläuterungen:

a) Von „discipulus“ (lateinisch: Schüler) stammt der Begriff „Disziplin“, von „studere“ (lateinisch: sich um etwas bemühen) der Begriff „***Studium***“. Sich um seine Fächer (auch: *Neben*fächer) zu bemühen, heißt insbesondere, sie nicht als Mittel zum Zweck zu missbrauchen.

   Anders gesagt: In der Schule ist das Lernen angeleitet, es genügt, diszipliniert zu tun, was einem gesagt wird. Im Studium genügt Disziplin nicht, sondern eigene Anstrengung ist erforderlich. Die Erfahrung zeigt, dass eine rein extrinsische Motivation dafür nicht ausreicht. Ziehen Sie Ihre Motivation stattdessen aus den studierten Fächern.

b) Das Unterrichtstempo ist an Hochschulen generell drastisch höher als an Schulen.[6] Was in der Schule wochenlang wiedergekäut würde, wird in einer Vorlesung teilweise in 15 Minuten abgehandelt; es ist Ihr Job, es danach gemäß Ihrer eigenen Lernbedürfnisse einzuüben.

c) Bis ca. 2005 wäre Ihnen allen völlig klar gewesen, dass das mit dem n-ten und $(n+2)$-ten Semester eine humoristische Anspielung auf das Beweisprinzip der vollständigen Induktion ist. Die humorfreie Übersetzung: Man benötigt eine hohe Frustrationstoleranz und einen sehr langen Atem. Man kann nicht erwarten, gleich alles zu verstehen. Aber: Wenn Sie gedanklich stets dabei bleiben, können Sie es schließlich in der Rückschau verstehen!

Ich empfehle Ihnen, Ihre Schulkenntnisse möglichst schnell hinter sich zu lassen und sie quasi im Rückspiegel zu betrachten, ohne sie ganz aus den Augen zu

[6] Ich spreche aus Erfahrung: In meinem Studium besuchte ich auch Lehrveranstaltungen in Physik, Linguistik, Musikwissenschaft, Meteorologie, Archäologie, Philosophie und Psychologie.

verlieren; auf diese Weise ist es meines Erachtens am ehesten möglich, von Schule auch noch im Studium zu profitieren. Vielleicht haben Sie den Eindruck, dass Hochschulmathematik und Schulmathematik nichts miteinander zu tun haben — das wäre nicht neu.[7] Doch ich versichere Ihnen, dass auch heute noch Bezüge zur Schulmathematik bestehen.

0.4 Prioritäten

> Das naive menschliche Denken geht von der Sache aus, das wissenschaftliche von der Methode.
>
> Carl Friedrich von Weizsäcker [1912–2007]

Fachmethoden sind die Verfahren zur Erlangung von Erkenntnissen, nach griech. μέθοδος, „Weg zu etwas hin“. Zur praktischen Anwendung von Methoden werden Techniken eingesetzt, nach griech. τέχνη, „Kunstfertigkeit“. Methoden und Techniken spielen nicht nur in der Forschung die wichtigste Rolle, sondern auch beim Lernen: Es ist sinnvoll, den historischen Erkenntnisprozess im eigenen Erkenntnisprozess nachzuvollziehen. Sie sollten Ihre Aufmerksamkeit also besonders hierauf lenken. In der Mathematik beruht Erkenntnis seit dem 7. Jahrhundert v. Chr. auf schlüssigen begrifflichen Argumentationen: Beweisen.

Praktische Methodenkenntnis fällt im Mathematikunterricht an heutigen deutschen Schulen komplett unter den Tisch, lediglich einige mathematische Techniken werden in der Schule unterrichtet, allerdings im Vergleich zu anderen Fächern um mehrere Jahrhunderte veraltet.[8] Das ist ein *bildungspolitischer* Skandal, so, als würde man im Deutschunterricht nie einen Aufsatz schreiben oder im Kunstunterricht lernen, dass Zeichnen nur etwas für Nerds sei. Darüber hinaus ist es ein *gesamtgesellschaftlicher* Skandal, denn der in einer Demokratie erwünschte „eigentümlich zwanglose Zwang des besseren Argumentes“[9] kann sich nur entfalten, wenn man die Korrektheit von Argumenten zu beurteilen gelernt hat.

> Wenn das einzige Werkzeug, das Du besitzt, ein Hammer ist, bist Du geneigt, jedes Problem als Nagel anzusehen.
>
> Abraham Maslow [1908–1970]

Fachbegriffe sind die Werkzeuge der Wissenschaften, ohne sie ist man im Studium verloren. Auch beim Einüben von Begriffen ist der Einsatz der Fachmethoden nützlich, denn das Zusammenwirken der Begriffe und die Gründe für die Details in ihrer Definition sieht man, wenn man mit den Begriffen arbeitet, also

[7] Felix Klein [1849–1925] in „Elementarmathematik vom höheren Standpunkt“ (1908).
[8] In der Vorlesung werde ich die Inhalte meist auch zeitlich einordnen.
[9] Jürgen Habermas [geb. 1929]

die Methoden auf sie anwendet. Durch Schwierigkeiten beim Einsatz der Methoden können u. U. neue bzw. verfeinerte Begriffsbildungen nötig werden.

Man kann darüber streiten, ob die im 19. Jahrhundert geprägten Begriffsdefinitionen der Analysis und der Algebra Schulstoff sein sollten, doch *überhaupt* mit Begriffen zu arbeiten, muss auch an Schulen selbstverständlich sein. Mich erfüllt mit großer Sorge, dass seit wenigen Jahren Studienanfänger*innen nicht nur keinen Grenzwert ausrechnen können, sondern ihnen sogar das Konzept fehlt, dass „Grenzwert“ ein Begriff mit Definition und Berechnungstechniken sein könnte.

In Geisteswissenschaften wird teilweise die Auffassung vertreten, die Bedeutung eines Begriffes ergebe sich aus dessen (alltagssprachlicher) Verwendung. Mathematische Begriffe beziehen sich hingegen auf Definitionen, aus denen sich ihr korrekter Gebrauch ergibt. **Halten Sie beim Lernen immer das Skript und Bücher bereit, um die Definitionen nachzulesen und zu vergleichen.**

Die Theorie ist Mutter der Praxis

Louis Pasteur [1822–1895]

Der Begriff *Theorie* kommt vom griechischen θεωρεῖν (anschauen, betrachten). Durch eine Theorie wird es überhaupt erst möglich, einen Gegenstand anschaulich zu erfassen. Insofern ist die modische Arroganz gegenüber Theorie („Das ist ja nur eine Theorie“, „Theorie bringt mir sowieso nichts“) nicht nur peinlich, sondern ein Lernhindernis.

Beispiele sind für sich genommen nutzlos — erst durch die Verbindung mit der Theorie wird ein Beispiel anschaulich und damit lehrreich. Beispiele sollte man sich selbst erarbeiten und möglichst auch selbst ausdenken. Das galt wohl schon in der Antike, denn in den „Elemente“ des Euklid [um 300 v. Chr.], einem der einflussreichsten Werke der Weltliteratur, fehlten Beispiele und Erläuterungen völlig, man musste sich also selbst um sie kümmern.

Im Skript wird die Extension[10] eines Begriffs meist durch Beispiele und zum Teil auch Nicht-Beispiele erläutert, also Konstruktionen, die unter einen Begriff fallen bzw. die trotz einer gewissen Verwandschaft mit dem Begriff doch nicht unter den Begriff fallen. Derartige Beispiele werde ich anfangs auch in der Vorlesung besprechen, langfristig sollen Sie sie sich aber aus dem Skript selbst erarbeiten. Beispiele für die Intension[10] eines Begriffs ergeben sich aus seinem Gebrauch: *Jeder Beweis ist ein Beispiel für den produktiven Gebrauch von Begriffen.*

Mathematische Techniken stelle ich wie Euklid meist in Form eines ***Problems***[11] vor: Das ist ein Aufgabentyp mit zugehörigem Lösungsschema. Beispiele dafür gibt es nicht nur im Skript, sondern auch in vielen Übungsaufgaben.

[10] Was das ist, kommt später.

[11] Nach griechisch πρόβλημα: das zur Lösung vorgelegte

0.5 Folgerungen für das Lehrkonzept

Aus den vorigen Abschnitten sollte ersichtlich sein, dass man im Studium besonders viel Wert auf eigene Aktivität legen sollte. Die mathematische Methode können Sie gut einüben, indem Sie sich in Lerngruppen gegenseitig den Lehrstoff erklären, denn dazu muss man argumentieren. Kommunikation ist wichtig. **Fragen zu stellen** ist ein Zeichen von Mitdenken und **ist ausdrücklich erwünscht**. Dafür sollten Rahmenbedingungen geschaffen werden.

Traditionell würde man in der großen Gruppe in der Vorlesung ein neues Thema rezipieren, würde in Kleingruppen zu diesem Stoff Übungsaufgaben bearbeiten und schließlich alleine versuchen, die Zusammenhänge zu schaffen. Ein bekanntes Konzept zur Förderung des aktiven Lernens ist der flipped classroom, welches das traditionelle Schema weitgehend umkehrt: Zuerst erarbeiten sich die Lernenden zu Hause das Thema selbst und im eigenen Tempo, beispielsweise anhand kurzer Lehrvideos. Dann werden in Kleingruppen Übungsaufgaben bearbeitet. In der Vorlesung wird dann versucht, die Zusammenhänge aufzuzeigen, insbesondere auch durch Beantwortung der Fragen, die beim vorherigen selbständigen Lernen aufgekommen sind. Dadurch wird einerseits zu eigener Arbeit angeregt, andererseits werden die Lernenden beim schwierigsten Teil des Lernprozesses, nämlich dem Schaffen von Zusammenhängen, nicht allein gelassen.

Ich bin skeptisch gegenüber *kurzen* Lernvideos, denn zum Studium gehört gedankliche Vielfalt, und diese zu vermitteln dauert länger als 10 Minuten. Außerdem ist fraglich, ob man in der Praxis wirklich im eigenen Tempo statt in dem durch das Video vorgegebenen Tempo arbeitet. Daher werde ich Sie Skriptabschnitte (typischerweise Definitionen und Beispiele) zur Vorbereitung der nächsten Vorlesung lesen statt Videos ansehen lassen. Sie sollten möglichst mehrere Bücher verwenden (Vergleichen gehört zur Medienkompetenz). Sie sollten skeptisch sein, wenn jemand im Brustton der Überzeugung nur eine einzelne Technik zur Lösung eines Aufgabentyps propagiert, obwohl andere Quellen auch andere Techniken diskutieren; Sie sollten dann versuchen, die Vor- und Nachteile der verschiedenen Techniken zu vergleichen und auch mich um einen solchen Vergleich bitten, wenn ich das nicht sowieso schon in der Vorlesung mache. Durch die Wahl ungeeigneter Techniken kann man nämlich in Prüfungen viel Zeit vergeuden.

Es wurde empirisch gezeigt,[12] dass *aktives* Lernen effizienter als das rein *rezipierende* Lernen ist, aber interessanterweise bisweilen subjektiv den falschen Eindruck hervorruft, dabei weniger zu lernen. Lassen Sie also die Schule hinter sich und lassen Sie sich auf das Studium ein! Wenn Sie meinen, dass Ihnen das nichts bringt, befinden Sie sich wahrscheinlich im Irrtum.

[12] Louis Deslauriers *et al.*: Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom, PNAS September 24, 2019 116 (39) 19251–19257

0.6 Rechnerische Anforderungen

Ein CAS liefert lediglich Ergebnisse und trägt daher zum Verständnis von Methoden und Techniken nichts bei. In einer Phase des Studiums, in der Sie die Techniken nicht schon sicher beherrschen, wären CASe also ein Lernhindernis. In der Prüfung zur linearen Algebra sind keine elektronischen Hilfsmittel zugelassen.

Wenn nicht ausdrücklich etwas anderes verlangt wird, ist exakt zu rechnen, mit Brüchen und Wurzeln, aber nicht mit gerundeten Kommazahlen. Das hat mehrere Gründe:

- Schulkenntnisse über das Runden sind defizitär, da sich die Schulmathematik dabei nach Verwaltungsvorschriften richtet, die an Schulen anscheinend für wichtiger als die davon abweichenden Fachstandards erachtet werden.
- Es mag Sie überraschen: Gerundetes Rechnen ist viel zu kompliziert! Viele nützliche Rechengesetze, etwa das Assoziativgesetz $a \cdot (b \cdot c) = (a \cdot b) \cdot c$, gelten bei gerundetem Rechnen nicht mehr.
- Gerundetes Rechnen werden Sie in der ***Numerik*** lernen. Doch dafür brauchen Sie Vorkenntnisse aus der linearen Algebra und vor allem aus der Analysis, weshalb die Numerik erst für das vierte Semester vorgesehen ist.
- Im ersten Semester können aufgrund fehlender Vorkenntnisse numerische Techniken noch nicht berücksichtigt werden. Dadurch könnten kleine Rundungsfehler in Zwischenergebnissen zu großen Fehlern in der Lösung oder sogar zur Unlösbarkeit führen.

Wer also im ersten Semester ohne ausdrücklichen Arbeitsauftrag rundet, beispielsweise $\sqrt{2}$ durch 1.4142 ersetzt, erhält einen Punktabzug, der ggf. mehr als die Hälfte der Gesamtpunktzahl einer Aufgabe betragen kann.

1 Grundlagen

Mengenlehre ist in der Mathematik allgegenwärtig. Leider kann man sich seit wenigen Jahren nicht mehr darauf verlassen, dass Schulen eine anschauliche Vorstellung wenigstens der einfachsten Begriffe (Teilmenge, Schnittmenge, Vereinigungsmenge) vermitteln. Der erste Abschnitt soll hier Abhilfe zu schaffen (ohne saubere Begriffsbildungen). Um Sprachliche Ausdrucksmittel für die Formulierung von Begriffen geht es im zweiten Abschnitt. Im dritte Abschnitt geht es um Beweistechniken. Und im letzten Abschnitt werden die neu gewonnenen Ausdrucksmittel für Mengenkonstruktionen verwendet.

1.1 Naive Mengenlehre

Nach einer Schulreform in den 1960erjahren unterrichtete man in Westdeutschland Mengenlehre in der ersten Klasse der Grundschule, noch vor dem Rechenunterricht.[13] Ein spielerischer[14] Zugang zur Mengenlehre wäre mit Unterichtsmaterialien möglich, die heute leider nur zum Sortieren und Ordnen eingesetzt, aber nicht mit dem Begriff der Menge verknüpft werden. Mengen wie $\mathbb{Z}$, $\mathbb{Q}$ und $\mathbb{R}$ müssten Ihnen dennoch bekannt sein.

In den drei Lehrbüchern, die nach dem Inhaltsverzeichnis genannt sind, ist naive Mengenlehre kein Thema. Zum Ausgleich fehlender Vorkenntnisse sollten Sie in Schulbüchern stöbern — denn Mengenlehre steht auch heute noch an mehreren Stellen in den Lehrplänen für Regelschulen und Gymnasien.

Zur Einführung elementarer Mengenlehre an Grundschulen entwickelte Zoltan P. Dienes [1916–2014] die ***Logischen Blöcke***. Das sind Plättchen aus Holz oder Kunststoff, in verschiedenen Formen (Dreiecke, Quadrate, längliche Rechtecke, Kreise), Farben (rot, blau, gelb) und Größen (einerseits klein oder groß, andererseits dick oder dünn). Hier vereinfache ich dies auf nur zwei Formen (Dreieck, Quadrat), zwei Farben (rot, blau) und zwei Größen.

In der naiven Mengenlehre werden Mengen als Zusammenfassungen von Elementen aufgefasst (zum Beispiel die aus der Schule bekannte Menge $\mathbb{Z}$ der ganzen Zahlen) und Konstruktionen erörtert, mit denen man aus gewissen Grundmengen weitere Mengen bilden kann (Schnittmenge, Vereinigung). Dies entspricht folgendem Zitat von Georg Cantor [1845–1918], dem Begründer der Mengenlehre:

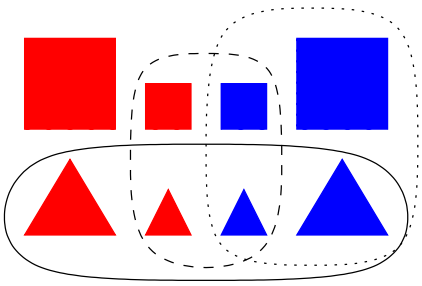
> *Unter einer ‚Menge' verstehen wir jede Zusammenfassung von bestimmten wohlunterschiedenen Objekten unserer Anschauung oder unseres Denkens zu einem Ganzen.* (G. Cantor, 1895)

Ist das Objekt x Element der Menge M, schreibt man $x \in M$, andernfalls $x \notin M$.

[13] Von Didaktiktiker*innen wurde dies leider mit einer Verballhornung mathematischer Ausdrucksweise verknüpft, wie im hier verlinkten Spiegel-Artikel von 1974 dargestellt.

[14] Viele Eltern und Lehrkräfte glaubten damals, dass Kinder nur durch Drill lernen können.

In der Abbildung sei $\mathbb{P}$ die Menge *aller* Plättchen, die Menge D der dreieckigen Plättchen ist mit einer durchgezogenen Linie, die Menge K der kleinen Plättchen mit einer unterbrochenen Linie und die Menge B der blauen Plättchen mit einer gepunkteten Linie markiert:



„p ist ein blaues Plättchen“ kann man jetzt als $p \in B$ ausdrücken. Die Anzahl der Elemente einer Menge nennt man ihre ***Kardinalität*** oder ***Mächtigkeit***. Wir haben hier also $|\mathbb{P}| = 8$ und $|B| = |K| = |D| = 4$.

Jede der Mengen entspricht einem Begriff (hier: „dreieckiges Plättchen“, „kleines Plättchen“ bzw. „blaues Plättchen“). Der Begriff ist sprachlich durch eine definierende Eigenschaft festgelegt (hier: dreieckig, klein bzw. blau); dies nennt man die ***Intension*** des Begriffs. Ebenso ist der Begriff aber auch gegenständlich festgelegt, nämlich durch die Gesamtheit aller Gegenstände des betrachteten Universums (hier: logische Blöcke), die unter den Begriff fallen; dies nennt man die ***Extension*** des Begriffs. Die Extension hängt natürlich sehr davon ab, in welchem Universum man sich befindet: Enthält dieses die betreffenden Formen jeweils auch gelb, wäre $|\mathbb{P}| = 12$, $|D| = |K| = 6$ und $|B| = 4$.

In Begriffsdefinitionen wird die Intension festgelegt, auch Beweise beziehen sich in der Regel nur auf Intensionen. Auf diese Weise wird die Mathematik allgemeingültig, also vom Universum unabhängig. Es ist sehr nützlich, wenn man nachweisen kann, dass verschiedene Intensionen auf die gleiche Extension führen. Eine Menge hingegen ist allein durch ihre Elemente bestimmt, egal wie die Elemente sprachlich beschrieben werden: Das Intervall $[-1, 1]$ ist das gleiche wie die Menge aller reellen Zahlen, deren Quadrat höchstens 1 ist, und ist das gleiche wie die Menge aller reellen Zahlen z, für die $z^2 \leq |z|$ gilt.

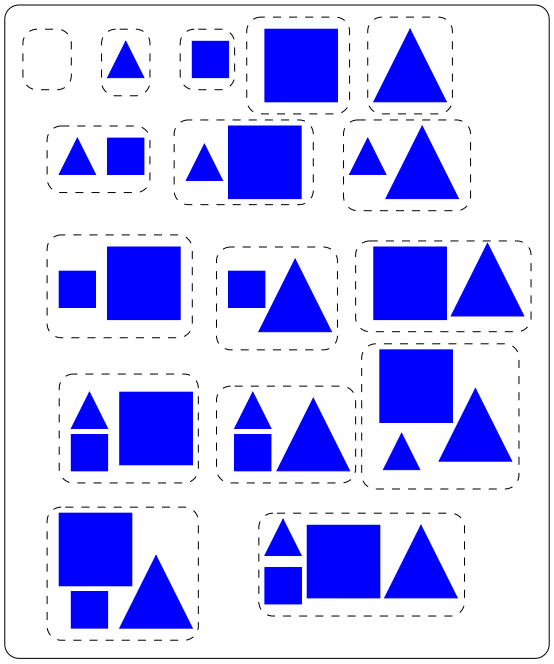
Definierende Eigenschaften kann man kombinieren. Die Menge der Plättchen, die sowohl zu B als auch zu K gehören, ergibt die Menge der „blauen kleinen“ Plättchen (alle Plättchen, die *sowohl* blau *als auch* klein sind). Sie ist durch die ***Schnittmenge*** $B \cap K$ gegeben. Es gilt $|B \cap K| = 2$ und $|B \cap K \cap D| = 1$. Beachte: Das kleine blaue Dreieck ist das einzige Element von $B \cap K \cap D$, aber es wäre falsch, eine ein-elementige Menge mit ihrem Element zu verwechseln (Analogie: Ein Portemonnaie, das genau eine Münze enthält, ist selbst keine Münze).

Die Menge der Plättchen, die zu B oder zu K gehören, enthält alle Plättchen außer den beiden großen roten Plättchen, also alle Plättchen, die blau oder klein (oder beides) sind: Man erhält die ***Vereinigungsmenge*** $B \cup K$, mit $|B \cup K| = 6$.

Beobachtung 1.1 (Prinzip von Inklusion und Exklusion)
Sind M, N Mengen, so gilt $|M| + |N| = |M \cup N| + |M \cap N|$.

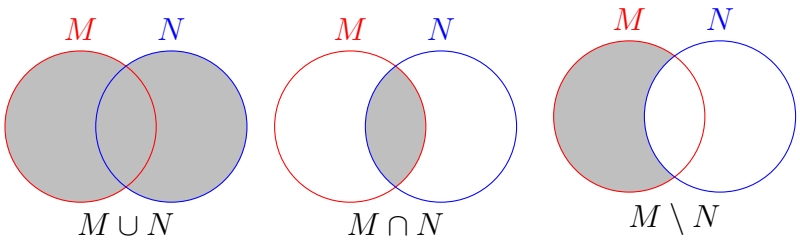
Die Menge der blauen Plättchen, die nicht dreieckig sind, wird durch die ***Differenzmenge*** $B \setminus D$ beschrieben und enthält die beiden blauen Quadrate. Es gilt also $|B \setminus D| = 2$.

Wir veranschaulichen einen weiteren Begriff: Eine Menge M heißt ***Teilmenge*** einer Menge N, wenn N alle Elemente von M (und möglicherweise noch weitere) enthält. Notation[15]: $M \subseteq N$. **Beachte:** Die ***leere Menge***, die *kein* Element enthält, ist Teilmenge jeder Menge. Die Abbildung zeigt alle Teilmengen von B.



Die Gesamtheit aller Teilmengen ist selbst eine Menge und ist im Bild mit durchgezogener Linie umrandet. Ihre 16 Elemente sind jeweils Teilmengen von B und sind mit unterbrochenen Linien umrandet. Jede Teilmenge besteht ausschließlich aus einigen (null bis allen) Elementen von B.

Die genannten „naiven“ Mengenbegriffe kann man sich leicht mit so genannten ***Venn-Diagrammen***[16] veranschaulichen:



[15] Wenn der Begriff formalisiert wird, stelle ich auch Notationsvarianten vor.
[16] John Venn [1834–1923]

Venn-Diagramme helfen lediglich bei der Veranschaulichung. In Beweisen würde man mit formalen Definitionen der Mengenkonstruktionen arbeiten, zu denen wir noch kommen werden.

1.2 Sprachliche Ausdrucksmittel

Für häufig vorkommende Formulierungen bietet sich die Verwendung von Formelzeichen an. Diese möchte ich hier einführen und dabei weitgehend darauf vertrauen, dass sich ihr korrekter Gebrauch aus dem natürlichen Sprachverständnis ergibt. Daher beschränke ich mich hier weitgehend auf Beispiele.

In der Mathematik arbeitet man meist mit ***Aussageformen***. Dies sind sprachliche Ausdrücke, in denen ggf. noch unbestimmte Platzhalter (***Variablen***) vorkommen, so dass durch Einsetzen konkreter Werte für die Variablen jeweils ein Aussagesatz entsteht. Das heißt: Eine Frage oder ein Befehl ist keine Aussageform.

Beispiel: $a^2 + b^2 = c^2$ ist eine Aussageform mit den drei Variablen a, b, c.

Das Gleichheitszeichen $=$ bedeutet, dass die Ausdrücke auf der linken und der rechten Seite des Gleichheitszeichen den gleichen Wert haben; diese Aussage kann wahr oder falsch sein. Hingegen bedeutet $\overset{!}{=}$ bzw. $\overset{?}{=}$ den Auftrag, die Gleichung zu lösen bzw. die Gültigkeit der Gleichung zu beweisen. Durch $:=$ wird dem Ausdruck auf der Seite, auf der der Doppelpunkt steht, ein Wert zugewiesen.

Beispiel: „Sei $a := 3$, $b := 4$, $c := 5$. Dann $a^2 + b^2 = c^2$“ hat keine Variable, sondern ist aufgrund der Zuweisungen eine wahre Aussage.

Logische Operatoren lassen sich auf Aussageformen anwenden, wodurch neue Aussageformen entstehen.

Negation, $\neg$, „nicht“: **Beachte:** Die Negation ist oft etwas anderes als das „gefühlte Gegenteil“. *Beispiel:* $\neg(M \subseteq \{1, 2, 3\})$ ist die Aussageform, dass M keine Teilmenge von $\{1, 2, 3\}$ ist; sie ist z. B. für $M := \{1, 2, 4\}$ wahr.

Konjunktion, $\wedge$, „und“: **Beachte:** „und“ heißt „sowohl als auch“.
Beispiel: $(a > b) \wedge (a < c)$ ist eine Aussageform mit drei Variablen, und ist z. B. wahr für $a := 4$, $b := 3$, $c := 5$, aber falsch für $a := 2$, $b := 3$, $c := 5$.

Disjunktion, $\vee$, „oder“: **Beachte:** „oder“ heißt in der Mathematik etwas anderes als „entweder - oder “, nämlich: Die Disjunktion ist genau dann wahr, wenn mindestens eine der Teilaussagen wahr ist. *Beispiel:* $(a > b) \vee (a < c)$ ist wahr z. B. für $a := 2$, $b := 3$, $c := 4$, aber auch für $a := 2$, $b := 1$, $c := 4$

Implikation, $\Rightarrow$, „wenn, dann“: **Beachte:** Mehr dazu im Kapitel über Beweise. *Beispiel:* Für alle $x \in \mathbb{R}$ gilt $1 < x \Rightarrow 1 < x^2$. Jedoch gilt nicht für alle $x \in \mathbb{R}$ $1 < x^2 \Rightarrow 1 < x$, wie man am Gegenbeispiel $x := -2$ sieht.

Äquivalenz, $\iff$, „genau dann, wenn“: **Beachte:** $P \iff Q$ besteht aus zwei Implikationen, sowohl $P \Rightarrow Q$ als auch $Q \Rightarrow P$. Ein typischer Fehler ist, das Symbol $\iff$ zu verwenden, obwohl nur eine der beiden Implikationen gilt. *Beispiel:* Für alle $x, y \in \mathbb{R}$ gilt $x < y \iff x + 1 < y + 1$.

Quantoren drücken die Erfüllbarkeit von Aussageformen aus. Im Folgenden seien P, Q Aussageformen mit Variablen, die sich aus dem Kontext ergeben. Steht ein Allquantor vor einer Aussageform mit mehreren Variablen, entsteht eine Aussageform mit einer Variable weniger. Meist gibt man zusätzlich eine Menge als Wertebereich des Quantors an.

„für alle“, $\forall$, ***Allquantor***: Die Aussage $\forall x\colon\ P(x)$ bedeutet, dass $P(x)$ für jeden konkreten Wert von x wahr (also allgemeingültig) ist. *Beispiel:* $\forall x \in \mathbb{R}\colon\ x^2 \geq 0$ ist wahr, allerdings wäre $\forall x\colon\ x^2 \geq 0$ falsch, da man für x dann auch etwas einsetzen könnte, für das x^2 oder $\geq$ nicht definiert ist.

„es gibt“, $\exists$, ***Existenzquantor***: Die Aussage $\exists x\colon\ P(x)$ bedeutet, dass $P(x)$ für mindestens einen (vielleicht sogar mehrere) Werte von x wahr ist. Auch hier gibt man meist einen Wertebereich vor. *Beispiel:* $\exists y \in \mathbb{R}\colon\ y^2 = 2$ ist die wahre Aussage, dass $y^2 \stackrel{!}{=} 2$ lösbar ist (es macht nichts, dass die Lösung nicht eindeutig ist). Hingegen ist $\exists y \in \mathbb{Q}\colon\ y^2 = 2$ falsch, denn $\sqrt{2}$ ist keine rationale Zahl, wie wir noch sehen werden.

„es gibt genau ein“, $\exists!$, ***Einzigkeitsquantor***: Die Aussage $\exists! x\colon\ P(x)$ bedeutet, dass es ein x gibt, so dass $P(x)$ wahr ist, und zudem x dadurch eindeutig bestimmt ist. *Beispiel:* $\exists! x \in \mathbb{R}\colon\ x^2 = 2$ ist falsch, da die Gleichung $x^2 \stackrel{!}{=} 2$ zwei Lösungen in den reellen Zahlen hat. Hingegen ist $\exists! x \in \mathbb{R}\colon\ (x > 0 \wedge x^2 = 2)$ wahr, ebenso $\exists! x \in \mathbb{R}\colon\ (x < 0 \wedge x^2 = 2)$.

Weitere übliche Notation für Quantoren: $\bigvee x$ statt $\exists x$; $\bigwedge x$ statt $\forall x$ — siehe Vorlesung „Diskrete Strukturen“.

Geschichtliche Einordnung: Die klassische zweiwertige ***Aussagenlogik*** geht bis Aristoteles [384–322 v.Chr] und Philon von Megara [ca. 4./3. Jhdt. v.Chr.] zurück. George Boole [1815–1865], Gottlob Frege [1848–1925] und Bertrand Russell [1872–1970] formalisierten die Aussagenlogik. Voneinander unabhängig erweiterten Frege sowie Charles Santiago Sanders Peirce [1839-1914] sie zur ***Prädikatenlogik***, zu der auch Aussageformen und Quantoren gehören. Intuitiv wurden Quantoren bereits im 4. Jhdt. v.Chr. verwendet, die Bezeichnung „Quantor“ und die heutigen Notationen stammen jedoch aus dem 19. und 20. Jahrhundert.

1.3 Beweisstrategieen

Eine spezielle Beweistechnik für Aussagen über natürliche Zahlen, nämlich „vollständige Induktion“, wird ausführlich in der Vorlesung über Diskrete Strukturen

behandelt. Hier geht es um allgemeine Strategien für die Konstruktion von Beweisen. Wenn ich im folgenden $P(\underline{x})$ schreibe, ist gemeint, dass P eine Aussageform mit einer oder mehreren Variablen ist, und $\underline{x}$ steht für die Liste dieser Variablen.

Mathematische Aussagen haben meist die Form $\forall \underline{x}\colon\ V(\underline{x}) \Rightarrow B(\underline{x})$ — und zwar in Definitionen, Sätzen und deduktiven Argumenten:

Definitionen: Es gibt verschiedene Definitionstheorien. In der mathematischen Praxis ist $V(\underline{x})$ ein Kontext, in dem ein Begriff steht, und $B(\underline{x})$ hat oft die Form einer Äquivalenz, die die Intension des zu definierenden Begriffs festlegt. Da hier dem Begriff ein (Bedeutungs-)Wert zugewiesen wird, schreiben wir einen Doppelpunkt vor dem Äquivalenzzeichen.
Beispiel: Sei $x \in \mathbb{Z}$. x heißt gerade $:\Leftrightarrow {}^{x}/_{2} \in \mathbb{Z}$.
Beachte: Recht oft liest man eine Abkürzung: „$x \in \mathbb{Z}$ heißt ***gerade***, falls ${}^{x}/_{2} \in \mathbb{Z}$.“ Hier ist das Wort „falls“ als $\Leftrightarrow$ und nicht als $\Leftarrow$ zu verstehen. Der Kontext muss berücksichtigt werden: Wäre x keine ganze Zahl, sondern eine Kurve, hätte der Begriff „gerade“ eine ganz andere Intension.

In Sätzen[17] der Form $\forall \underline{x}\colon\ V(\underline{x}) \Rightarrow B(\underline{x})$ ist $V(\underline{x})$ die Voraussetzung, zu der (wichtig!!) auch die Definitionen der in der Formulierung des Satzes verwendeten Begriffe gehören, und $B(\underline{x})$ die zu beweisende Behauptung.

Deduktive Argumente bestehen in der Anwendung einer Implikation auf eine gültige Voraussetzung, wodurch sich eine Schlussfolgerung ergibt (z.B. modus ponendo ponens: Wenn sowohl $\forall \underline{x}\colon\ V(\underline{x}) \Rightarrow B(\underline{x})$ als auch $V(\underline{y})$ für ein $\underline{y}$ gelten, dann gilt auch $B(\underline{y})$). Es ist wichtig, dass die Implikation bereits vor ihrer argumentativen Verwendung etabliert wurde (etwa als Definition oder als vorher bewiesener Satz).

Zu Beginn sollten Sie Beweise möglichst explizit und ausführlich aufschreiben; erst mit etwas Übung kann man davon ausgehen, dass die Leser*innen „offensichtliche“ Details selbst ergänzen.

Es ist sowohl in der wissenschaftlichen Arbeit als auch beim Lernen essentiell, beim Lesen eines Beweises jeden einzelnen Schritt in Frage zu stellen: Ist erkennbar, auf welcher bereits bewiesenen Implikation das Argument basiert? Ist die Voraussetzung, auf die die Implikation angewandt wird, wirklich gültig, d.h. wurde auch die Voraussetzung in vorherigen Beweisschritten bewiesen?

Das In-Frage-Stellen von Beweisen ist aber (anders als landläufig behauptet wird) generell wohlwollend! Erweist sich ein Beweisschritt nämlich als angreifbar, heißt das noch lange nicht, dass man den Beweis einfach insgesamt für grundfalsch erklären und schmollen darf. Stattdessen überlegt man sich, welche Teile des Beweises korrekt sind, ob man die inkorrekten Teile durch andere Argumente überbrücken kann, oder ob im Beweis versteckte Zusatzannahmen verwendet

[17] Je nach Situationen kann ein mathematischer Satz nicht nur Satz, sondern auch Lemma, Proposition, Theorem oder Korollar heißen.

wurden, so dass sich ein korrekter Beweis ergibt, wenn man die Zusatzannahmen noch zu den ursprünglich formulierten Annahmen hinzufügt.

Ein extremes Beispiel: Euklids Anspruch war, alle Voraussetzungen (Definitionen, Axiome, Postulate) schon zu Beginn seiner „Elemente“ zu formulieren und dann in den Beweisen nur diese zu verwenden. Doch schon im Beweis der ersten Proposition sind mehrere nicht formulierte Zusatzannahmen versteckt. Erst Ende des 19. Jahrhunderts gelang es, die Lücken zu stopfen — doch das schmälert den Wert von Euklids Werk nicht im geringsten! Es hat die Mathematik nämlich entscheidend voran gebracht, dass Euklid überhaupt so hohe Ansprüche stellte, auch wenn er selbst ihnen nicht vollständig gerecht wurde.

Aber wie kommt man nun auf einen Beweis? Nun: Hierfür sind Kreativität und Phantasie nötig und diese kann man bis zu einem gewissen Grad trainieren. Natürlich fängt man einfach an: Mit kurzen Beweisen, die nur ein oder zwei Einzelargumente umfassen, oder man nimmt sich Beweise aus Büchern oder der Vorlesung zum Vorbild und modifiziert sie so, dass etwas neues entsteht.

1.3.1 Direkte Beweise

Ein direkter Beweis von $\forall \underline{x}\colon\ (V(\underline{x}) \Rightarrow B(\underline{x}))$ geht von der Gültigkeit der Voraussetzung aus und erreicht durch Umformungen die Gültigkeit der Behauptung. Insbesondere arbeitet man im Beweis mit einem beliebigen $\underline{x}$, für das $V(\underline{x})$ wahr ist — falls $V(\underline{x})$ durch mehr als nur endlich viele $\underline{x}$ erfüllt wird, wäre es ein grober Fehler, im Beweis *konkrete* Werte für $\underline{x}$ zu betrachten![18]

Sind die Umformungsschritte konstruktiv, ergibt sich aus einem direkten Beweis ein rechnerisches Verfahren. Wenn die Voraussetzung im Beweis nicht vollständig verwendet wird, ist der Beweis falsch oder man hat etwas neues entdeckt.

Für das erste Beispiel eines direkten Beweises muss ich mich auf schulische Vorkenntnisse zur Dezimaldarstellung reeller Zahlen berufen: Jedes $x \in \mathbb{R}$ lässt sich als Folge von Ziffern von 0 bis 9 darstellen, nämlich endlich viele Vorkommastellen und unendlich viele Nachkommastellen (die allerdings alle null sein können). Die Nachkommastellen können sich periodisch wiederholen, dies wird durch einen Querstrich über der sich wiederholenden Ziffernfolge gekennzeichnet.

Beispiel: $\frac{1}{7} = 0{,}\overline{142857}$

Aus der Schule ist auch der Begriff der rationalen Zahl bekannt:

Definition 1.2
Es gilt $x \in \mathbb{Q}$ genau dann, wenn es $a, b \in \mathbb{Z}$ mit $b \neq 0$ gibt, so dass $x = \frac{a}{b}$.

Wir formulieren nun einen Satz und führen einen direkten Beweis. Dabei kommt es uns auf die grundsätzliche Struktur des Beweises an, weshalb ich auf

[18] Wenn aber nur endlich viele $\underline{x}$ die Voraussetzung erfüllen, kann man eine Fallunterscheidung machen.

eine formale Notation für Dezimaldarstellungen verzichte und stattdessen die Konstruktion nur durch ein Beispiel andeute, was man normalerweise natürlich nicht tut. Ich nenne es „Lemma“, da ich später daraus eine Folgerung ziehen werde. Anmerkungen zur Vorgehensweise sind in roter Farbe.

Lemma 1.3
Es sei $x \in \mathbb{R}$ und $\ell \in \mathbb{N}$ so, dass ab der ℓ-ten Nachkommastelle alle Ziffern in der Dezimaldarstellung von x gleich sind. Dann ist $x \in \mathbb{Q}$.

Beweis:
Zunächst wird die Voraussetzung etwas konkretisiert: Es sei $z \in \{0, ..., 9\}$ die Ziffer, die sich ab der ℓ-ten Nachkommastelle von x wiederholt. Der Fall $z = 0$ entspricht einer endlichen Dezimaldarstellung.

Kreative Konstruktion: $y \in \mathbb{R}$ sei aus dem Vorzeichen, den Vorkommaziffern und den ersten $\ell-1$ Nachkommaziffern von x gebildet. Beobachtung: $y \cdot 10^{\ell-1} \in \mathbb{Z}$. Es gilt also $y \in \mathbb{Q}$. Hier wurde die Definition rationaler Zahlen eingesetzt.

In der Dezimaldarstellung von $x - y$ sind die Vorkommastellen und die ersten $\ell - 1$ Nachkommastellen null. Bis auf das Vorzeichen gilt also $(x - y) \cdot 10^{\ell-1} = \pm 0{,}\bar{z} = \pm\frac{z}{9}$. Folglich ist $x = y \pm \frac{z}{9 \cdot 10^{\ell-1}}$ rational, denn es ist eine Summe zweier rationaler Zahlen. Wir haben x explizit als rationale Zahl dargestellt. □

Beispiel: Für $x := -17{,}242\bar{3}$ haben wir $\ell := 4$, $z = 3$, $y := -17{,}242 = \frac{-17242}{1000}$, $x - y = -0{,}000\bar{3}$ und damit $x = \frac{-17242}{1000} - \frac{3}{9000}$.

Das Zeichen □ zeigt an, dass dort der Beweis endet. Die eigentliche mathematische Arbeit besteht nun darin, den Beweis zu verbessern: Für Aussageformen $V, \tilde{V}, B, \tilde{B}$ sei $(*) := \forall \underline{x}\colon\ (V(\underline{x}) \Rightarrow B(\underline{x}))$. Wenn $\forall \underline{x}\colon\ \tilde{B}(\underline{x}) \Rightarrow B(\underline{x})$, so nennt man die Aussage $\forall \underline{x}\colon\ V(\underline{x}) \Rightarrow \tilde{B}(\underline{x})$ eine ***Verschärfung*** von $(*)$. Wenn $\forall \underline{x}\colon\ V(\underline{x}) \Rightarrow \tilde{V}(\underline{x})$, so nennt man die Aussage $\forall \underline{x}\colon\ \tilde{V}(\underline{x}) \Rightarrow B(\underline{x})$ eine ***Verallgemeinerung*** von $(*)$. Durch den Beweis einer Verallgemeinerung oder einer Verschärfung ist dann auch (*) bewiesen.

In vorliegenden Fall besteht die Verallgemeinerung darin, dass man nicht unbedingt eine Dezimaldarstellung braucht, sondern der Beweis in einem beliebigen Stellenwertsystem mit fester Basis funktioniert. An den Beweis von Lemma 1.3 werde ich eine Hausaufgabe anschließen.

1.3.2 Indirekte Beweise

Man kann eine Aussage beweisen, indem man beweist, dass keine Widerlegung der Aussage möglich ist — zumindest ist dies in der klassischen zweiwertigen Logik möglich. Andere Sichtweise: Man ersetzt die zu beweisende Aussage durch eine logisch äquivalente, aber ggf. leichter beweisbare Aussage.

Definition 1.4. *Seien P, Q Aussageformen mit Variablenliste $\underline{x}$.*
P, Q heißen **gleichbedeutend** *oder* **logisch äquivalent** *(Sprechweise: „$P(\underline{x})$ gdw. (genau dann wenn) $Q(\underline{x})$“)* $:\Leftrightarrow \forall \underline{x}\colon\ (P(\underline{x}) \iff Q(\underline{x}))$.

In einer Logik-Vorlesung würde man sorgfältig zwischen Objektsprache (in der man Aussagen über Objekte formuliert) und Metasprache (in der man Aussagen über die Objektsprache formuliert) unterscheiden. Das soll hier nicht Thema sein, aber dennoch ist dieser Wikipedia-Artikel lesenswert. Hier ist eine Liste[19] oft in Beweisen verwendeter Umformungsregeln:

Axiom 1.5. *Seien V, B Aussageformen mit Variablenliste $\underline{x}$.*

a) $\neg\,(\forall \underline{x}\colon\ V(\underline{x}))$ *gdw.* $\exists \underline{x}\colon\ \neg V(\underline{x})$.

b) $\neg\,(\exists \underline{x}\colon\ V(\underline{x}))$ *gdw.* $\forall \underline{x}\colon\ \neg V(\underline{x})$.

c) $\neg\,(V(\underline{x}) \wedge B(\underline{x}))$ *gdw.* $(\neg V(\underline{x})) \vee (\neg B(\underline{x}))$

d) $\neg\,(V(\underline{x}) \vee B(\underline{x}))$ *gdw.* $(\neg V(\underline{x})) \wedge (\neg B(\underline{x}))$

e) $\neg\,(V(\underline{x}) \Rightarrow B(\underline{x}))$ *gdw.* $V(\underline{x}) \wedge (\neg B(\underline{x}))$.

f) $V(\underline{x}) \Rightarrow B(\underline{x})$ *gdw.* $(\neg B(\underline{x})) \Rightarrow (\neg V(\underline{x}))$.

g) $V(\underline{x}) \iff B(\underline{x})$ *gdw.* $(V(\underline{x}) \Rightarrow B(\underline{x})) \wedge (B(\underline{x}) \Rightarrow V(\underline{x}))$.

h) $V(\underline{x}) \wedge W$ *gdw.* $V(\underline{x})$.

c) und d) heißen **de Morgansche Regeln**.

Beweise durch Kontraposition basieren auf Axiom 1.5.f). Im Beispiel wird als bekannt vorausgesetzt: Jede ganze Zahl hat eine bis auf Faktorreihenfolge eindeutige Primfaktorzerlegung (mit Vorzeichen).

Lemma 1.6. *Sei $x \in \mathbb{R}$. Wenn $x^2 = 2$, dann $x \notin \mathbb{Q}$. Insbesondere ist $\sqrt{2} \notin \mathbb{Q}$.*

Beweis:
Zunächst setzen wir die Definition von $\mathbb{Q}$ ein: Zu zeigen ist $\forall x \in \mathbb{R}\colon\ x^2 = 2 \Rightarrow \nexists a, b \in \mathbb{Z}\colon\ x = \frac{a}{b}$. Kontraposition: Dazu zeigen wir $\forall a, b \in \mathbb{Z}\colon\ \left(\frac{a}{b}\right)^2 \neq 2$, oder gleichbedeutend $a^2 \neq 2b^2$.

Wir berufen uns nun auf Vorkenntnisse. Der Primfaktor 2 komme m-mal bzw. n-mal in der Primfaktorzerlegung von a bzw. von b vor. Dann kommt der Primfaktor 2 genau $2m$-mal in a^2 und $(2n+1)$-mal in $2b^2$ vor. Da $2m$ gerade und $2n+1$ ungerade ist, gilt $2m \neq 2n+1$. Als letztes ist noch $2m \neq 2n+1 \Rightarrow a^2 \neq 2b^2$ zu zeigen. Erneute Kontraposition: Wenn $a^2 = 2b^2$, dann folgt wegen der Eindeutigkeit der Primfaktorzerlegung $2m = 2n+1$. □

Aufgabe 1.7
Formulieren und beweisen Sie Verallgemeinerungen von Lemma 1.6.

[19] Sie ist allerdings weder vollständig noch frei von Redundanzen!

Widerspruchsbeweise basieren auf Axiom 1.5.e). Die zu beweisende Aussage $\forall \underline{x}\colon\ V(\underline{x}) \Rightarrow B(\underline{x})$ ist genau dann falsch, wenn es ein Gegenbeispiel gibt, also $\exists \underline{x}\colon\ V(\underline{x}) \wedge (\neg B(\underline{x}))$. Also kann man die Aussage dadurch beweisen, dass man die Existenz eines Gegenbeispiels widerlegt, formell: $\forall \underline{x}\colon\ V(\underline{x}) \wedge (\neg B(\underline{x}))) \Rightarrow F$.

Proposition 1.8
Sei $x \in \mathbb{R}$, $x \notin \mathbb{Q}$. Es gibt mindestens zwei Ziffern, die jeweils unendlich oft in der Dezimaldarstellung von x auftreten.

Beweis:
Wir nehmen an, dass es ein Gegenbeispiel gibt. **Annahme:** Es gibt $x \in \mathbb{R}$, $x \notin \mathbb{Q}$, in dessen Dezimaldarstellung nur höchstens eine Ziffer unendlich oft auftritt. Nach Lemma 1.3 würde das aber $x \in \mathbb{Q}$ bedeuten. Widerspruch. Also war die Annahme falsch, d.h. es gibt kein Gegenbeispiel. □

Bemerkung Für ein konkretes $x \in \mathbb{R}$, $x \notin \mathbb{Q}$, so wie $x = \sqrt{2}$ oder $x = \pi$, liefert der Beweis nicht den geringsten Anhaltspunkt dafür, welche zwei Ziffern es nun sind, die unendlich oft in der Dezimaldarstellung auftreten. Dieses Problem tritt bei indirekten Beweisen häufig auf. Direkte konstruktive Beweise sind zu bevorzugen — aber nicht immer lässt sich ein indirekter Beweis vermeiden.

Fallunterscheidungen basieren auf Axiom 1.5.h) in Verbindung mit den de Morganschen Regeln. Oft ist ein Beweis einfacher zu führen, wenn zusätzliche Voraussetzungen gelten (allerdings ist dann der Beweis weniger allgemein). Bei einer Fallunterscheidung betrachtet man nacheinander mehrere Zusatzvoraussetzungen, von denen stets mindestens eine erfüllt ist (man weiß aber nicht, welche). Man hat also Zusatzvoraussetzungen $Z_1(\underline{x}), ..., Z_n(\underline{x})$ mit $\forall \underline{x}\colon\ Z_1(\underline{x}) \vee ... \vee Z_n(\underline{x})$, und dann ist $V(\underline{x}) \Rightarrow B(\underline{x})$ logisch äquivalent zu $((V(\underline{x}) \wedge Z_1(\underline{x})) \Rightarrow B(\underline{x})) \wedge ... \wedge ((V(\underline{x}) \wedge Z_n(\underline{x})) \Rightarrow B(\underline{x}))$. Hier ist ein Beispiel:

Theorem 1.9. $\exists a, b \in \mathbb{R}\colon\ a, b \notin \mathbb{Q} \wedge a^b \in \mathbb{Q}$.

Beweis:
Entweder ist $\sqrt{2}^{\sqrt{2}}$ in $\mathbb{Q}$ oder nicht. Fallunterscheidung:

Fall 1: $\sqrt{2}^{\sqrt{2}} \in \mathbb{Q}$. Dann ist die zu beweisende Aussage wahr mit $a = b = \sqrt{2}$, denn $\sqrt{2} \notin \mathbb{Q}$ nach Lemma 1.6.

Fall 2: $\sqrt{2}^{\sqrt{2}} \notin \mathbb{Q}$. Dann ist die zu beweisende Aussage wahr mit $a = \sqrt{2}^{\sqrt{2}}$ und $b = \sqrt{2}$, denn $a^b = \sqrt{2}^{\sqrt{2}\cdot\sqrt{2}} = \sqrt{2}^2 = 2 \in \mathbb{Q}$, $a \notin \mathbb{Q}$ in Fall 2, und $b \notin \mathbb{Q}$ nach Lemma 1.6. □

Bemerkung Der Beweis gibt keinen Aufschluss darüber, ob $\sqrt{2}^{\sqrt{2}}$ rational ist. Der Satz von Gelfond-Schneider [1934] impliziert aber, dass es tatsächlich nicht rational ist.

1.4 Mengenkonstruktionen

Inzwischen haben wir die sprachlichen Mittel, um die Konstruktionen der naiven Mengenlehre aus Abschnitt 1.1 zu formalisieren.

Notation 1.10
Sei $P(x)$ eine Aussageform mit einer Variable x. Mit $\{x \mid P(x)\}$ bezeichnet man die Gesamtheit aller x, für die $P(x)$ wahr ist. $P(x)$ entspricht der Intension, $\{x \mid P(x)\}$ der Extension eines Begriffs.

Bemerkung 1.11
Es ist Absicht, dass ich $\{x \mid P(x)\}$ nicht als Menge, sondern als Gesamtheit bezeichnet habe. Frege dachte tatsächlich, dass dadurch immer eine Menge beschrieben werde. Das nennt man das **Abstraktionsprinzip**. *Frege schrieb auf dieser Grundlage ein Buch, das sich 1902 bereits im Druck befand, als sein Werk durch einen Brief von Bertrand Russell [1872–1970] in seinen Grundfesten erschüttert wurde: Durch die* **Russellsche Antinomie** *$R := \{x \mid x \notin x\}$. Die Antinomie besteht darin, dass $R \in R$ genau dann gilt, wenn $R \notin R$.*

Man umgeht die Antinomie wie folgt: Gesamtheiten bezeichnet man generell als **Klasse**. *Eine Klasse, die zudem als Element einer Klasse auftreten kann, nennt man* **Menge**. *Das* **Paarmengenaxioms** *besagt: Sind M, N Mengen, dann ist auch $\{M, N\} := \{x \mid x = M \vee x = N\}$ eine Menge. Folglich ist auch $\{M, M\} = \{M\}$ eine Menge (das ist der Fall $M = N$).*

In der Notation $\{x \mid P(x)\}$ bezieht sich x nicht auf beliebige Klassen, sondern ausschließlich auf Mengen. Die Auflösung der Russellschen Antinomie ist dann: $R := \{x \mid x \notin x\}$ ist zwar eine Klasse, aber keine Menge; es entsteht kein Widerspruch, denn da R keine Menge ist, ist einfach $R \notin R$.

Es gab im 20. Jahrhundert mehrere Ansätze zu einer Formalisierung der Mengenlehre. Das große Problem: Kurt Gödel [1906–1978] bewies, dass kein *hinreichend komplexes formales System die eigene Widerspruchsfreiheit beweisen kann. Dies gilt auch für jede Art von axiomatischer Mengenlehre, die die natürlichen Zahlen enthält.*

Die Gleichheit von Mengen (oder generell Klassen) soll nur von den enthaltenen Elementen abhängen. Dies drückt man wie folgt aus:

Extensionalitätsaxiom
Zwei Mengen X, Y sind gleich ($X = Y$) $\iff \forall x\colon\ (x \in X \Leftrightarrow x \in Y)$.

Es ist also $\{x \mid P(x)\} = \{x \mid Q(x)\}$ gleichbedeutend dazu, dass die Aussageformen P, Q logisch äquivalent sind. Endliche Mengen kann man durch Auflistung der Elemente angeben, etwa $M = \{1, 3, 4, 6\}$. Bei unendlichen Mengen behilft man sich manchmal mit Auslassungszeichen, etwa $\{1, 2, 3, 4, 5, \ldots\}$. Das Problem ist, dass die Auslassungszeichen nur dann einen nachvollziehbaren Sinn ergeben,

wenn man die Menge bereits kennt. Mit Hilfe der On-Line Encyclopaedia of Integer Sequences können Sie beispielsweise feststellen, dass in der mathematischen Fachliteratur mindestens 6986 *verschiedene* Zahlenfolgen untersucht wurden, die hintereinander die Zahlen $1, 2, 3, 4, 5$ enthalten; jede von ihnen ergibt prinzipiell eine mathematisch sinnvolle Möglichkeit, „$1, 2, 3, 4, 5, ...$“ zu interpretieren.

Definition 1.12. *Seien M, N Mengen.*
N heißt eine **Teilmenge** *von M (Notation: $N \subseteq M$ oder synonym $N \subset M$)* $:\Leftrightarrow \forall x \in N\colon\ x \in M$.
N heißt **echte Teilmenge** *von M (Notation $N \subsetneq M$)* $:\Leftrightarrow N \subset M \wedge N \neq M$.

Ich verwende die Symbole $\subseteq$ und $\subset$ gleichbedeutend. In manchen Büchern wird $\subset$ aber nur für echte Teilmengen verwendet. Beim Lesen von Lehrbüchern müssen Sie immer darauf achten, welche Konventionen darin verwendet werden — in Aufgaben gelten die Konventionen aus der Vorlesung.

Beobachtung 1.13
Seien M, N Mengen. $M = N \iff (M \subseteq N \wedge N \subseteq M)$.

Wir haben durch die Russell-Antinomie gesehen, dass nicht alle Gesamtheiten Mengen sind, doch die Axiome der Mengenlehre implizieren, dass die folgenden Konstruktionen angewandt auf Mengen stets wieder Mengen ergeben. Manche der Konstruktionen haben wir bereits mit Venn-Diagrammen visualisiert, bei den anderen geht das leider nicht. Es ist zu hoffen, dass sie noch weitgehend Schulstoff sind (bis vor einigen Jahren waren sie es).

Definition 1.14
Seien M, N Mengen und P eine einstellige Aussageform. Axiome der Mengenlehre implizieren, dass die folgenden Konstruktionen wieder Mengen ergeben:

a) $\{x \in M \mid P(x)\} := \{x \mid x \in M \wedge P(x)\}$ *ist eine Menge; dies ist der Inhalt des* **Aussonderungsaxioms**.

b) $M \cap N := \{x \mid x \in M \wedge x \in N\} = \{x \in M \mid x \in N\} = \{x \in N \mid x \in M\}$ *heißt* **Schnitt** *oder* **Durchschnitt** *oder* **Schnittmenge**; *nach dem Aussonderungsaxiom ist es eine Menge. $M \cap N$ besteht also aus allen Elementen von M, die zugleich Elemente von N sind.*

c) $M \setminus N := \{x \in M \mid x \notin N\}$ *heißt* **Differenzmenge** *von M und N; nach dem Aussonderungsaxiom ist es eine Menge. $M \setminus N$ besteht also aus allen Elementen von M, die zugleich* nicht *Elemente von N sind.*

d) $M \cup N := \{x \mid x \in M \vee x \in N\}$ *ist eine Menge, genannt die* **Vereinigung** *(oder Vereinigungsmenge) von M und N; dies folgt aus dem Paarmengenaxiom zusammen mit dem* **Vereinigungsaxiom**.

e) $\mathscr{P}(M) := \{T \mid T \subset M\}$ *ist eine Menge, genannt* **Potenzmenge** *von* M*. Dies ist der Inhalt des* **Potenzmengenaxioms**

f) Das **kartesische Produkt** *(auch: direkte Produkt) von* M *und* N *besteht aus allen* **geordneten Paaren** (m, n) *mit* $m \in M$ *und* $n \in N$*:*

$$M \times N := \{(m, n) \mid m \in M,\ n \in N\}$$

g) $f \subset M \times N$ *heißt* **Abbildung** *von* M *(***Definitionsmenge***) nach* N *(***Zielmenge***)* $:\Leftrightarrow \forall m \in M \exists! n \in N\colon (m, n) \in f$*. Notation:* $f\colon M \to N$*, und statt* $(m, n) \in f$ *schreibt man üblicherweise* $n = f(m)$ *oder* $m \overset{f}{\mapsto} n$*. Abbildungen bezeichnet man synonym auch als* **Funktionen**.

h) Sei $f\colon M \to N$*. Das* **Ersetzungsaxiom** *besagt, dass auch* $\text{Bild}(f) := \{f(m) \mid m \in M\} := \{n \in N \mid \exists m \in M\colon (m, n) \in f\}$ *eine Menge ist. Man nennt sie die* **Bildmenge** *von* f*.*

Bemerkung

Zu f): „Geordnet“ heißt beispielsweise $(1, 2) \neq (2, 1)$; man beachte den Unterschied zu Mengen: $\{1, 2\} = \{2, 1\}$. Analog bezeichnet Ausdrücke der Form (a, b, c) als ***Tripel*** und $(x_1, ..., x_n)$ mit $n \in \mathbb{N}$ als n***-Tupel***. Sind $M_1, ..., M_n$ Mengen, so ist deren kartesisches Produkt $M_1 \times ... \times M_n := \{(m_1, ..., m_n) \mid m_1 \in M_1, ..., m_n \in M_n\}$. Statt $\underbrace{M \times ... \times M}_{n\text{-mal}}$ schreibt man auch M^n.

Zu g): In der Schule haben Sie die Menge $\{(m, f(m)) \mid m \in M\}$ als ***Funktionsgraph*** bezeichnet. Man kann den Begriff der Abbildung intuitiv auch so verstehen: Durch f wird jedem $m \in M$ ein eindeutiger Funktionswert $f(m) \in M$ zugeordnet.

Um die obigen Konstruktionen anwenden zu können, braucht man natürlich einen „Vorrat“ folgenden Mengen denken Sie vermutlich aus der Schule zu kennen.

$\emptyset$**,** $\{\}$ Die *leere* Menge, die *keine* Elemente enthält. Dass es die leere Menge gibt, wird im ***Leermengenaxiom*** gefordert.

$\mathbb{N}$ Die Menge $\mathbb{N} = \{0, 1, 2, 3, 4, 5, \ldots\}$ aller natürlichen Zahlen — zumindest nach DIN–Norm 5473 gehört die Null zu den natürlichen Zahlen. Die Existenz von $\mathbb{N}$ folgt aus dem ***Unendlichkeitsaxiom***.

$\mathbb{N}^*$ Die Menge $\mathbb{N}^* = \{1, 2, 3, 4, 5, \ldots\}$ aller positiven natürlichen Zahlen — wieder nach DIN-Norm 5473.

$\mathbb{Z}$ Die Menge $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ aller ganzen Zahlen

$\mathbb{Q}$ Die Menge der rationalen Zahlen

$\mathbb{R}$ Die Menge der reellen Zahlen

$\mathbb{Q}_{>0}$ Menge der positiven rationalen Zahlen (analog $\mathbb{Q}_{\geq 0}$, $\mathbb{R}_{<0}$ etc.)

Intervalle: Für $a, b \in \mathbb{R}$, $a \leq b$, bezeichnet man $[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$, $]a, b] = \{x \in \mathbb{R} \mid a < x \leq b\}$, analog $[a, b[$ und $]a, b[$, ferner $]-\infty, a] = \{x \in \mathbb{R} \mid x \leq a\}$ etc.

$\mathbb{C}$ Die Menge der komplexen Zahlen (kommt noch)

DIN–Norm 5473 wird von vielen Mathematikern ignoriert. Sie sollten also immer auf den Kontext achten, ob $0 \in \mathbb{N}$. Wer $\mathbb{N} = \{1, 2, 3, 4, ...\}$ verwendet, der schreibt meistens $\mathbb{N}_0 = \{0, 1, 2, 3, ...\}$.

Beispiele:

a) Die Menge der Quadratzahlen kann man einerseits mit dem Ersetzungs-, andererseits mit dem Aussonderungsaxiom darstellen:
$\{n^2 \mid n \in \mathbb{Z}\} = \{m \in \mathbb{Z} \mid \exists k \in \mathbb{N}\colon\ k^2 = m\}$

b) Auch die leere Menge lässt sich auf verschiedene Arten darstellen:
$\emptyset = \{x \in \mathbb{R} \mid x^2 < 0\} = \{x \mid x \neq x\}$.

c) $\mathscr{P}(\{1, 2, 3\}) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$. Am Ende von Abschnitt 1.1 wurde die Potenzmenge der Menge B der blauen Plättchen bildlich dargestellt.

d) In der Schule modellierten Sie den „Anschauungsraum“ als $\mathbb{R}^3$: Jeder Raumpunkt entspricht eindeutig einem Koordinatentripel (x, y, z) reeller Zahlen.

Wenn M, N Mengen sind, ist es eine gute Idee, sich $M \times N$ als ein Rechteck vorzustellen, dessen eine Seite M und die dazu senkrechte Seite N entspricht. Analog lässt sich ein dreifaches kartesisches Produkt als Quader veranschaulichen.

2 Lineare Gleichungssysteme

In diesem Kapitel geht es darum, die gedanklichen Grundlagen für die lineare Algebra zu legen. Wir beginnen mit Gleichungssystemen, wie Sie sie in der Schule betrachteten: Systeme von bis zu drei linearen Gleichungen mit bis zu drei Unbekannten, mit Koeffizienten in $\mathbb{R}$. Die Schulnotation dafür wäre zum Beispiel:

$$\begin{aligned} -2\,x + y - 2\,z &= 1 && \text{(I)}\\ 2\,x + 0\,y - 2\,z &= 1 && \text{(II)}\\ 0\,x - 4\,y - 4\,z &= -3 && \text{(III)} \end{aligned}$$

Hierfür werden wir zunächst zwei geometrische Sichtweisen entwickeln. Außerdem führen wir die Matrixnotation ein, die Sie in Zukunft statt der ungeschickten Schulnotation verwenden sollen.[20] Die Matrixnotation macht es möglich, beliebig große lineare Gleichungssysteme zu betrachten, und sie gibt auch Anlass zu weiterführenden algebraischen Themen. Die hier betrachteten Lösungstechniken sind in jedem Kontext durchführbar, in denen die Grundrechenarten mit den üblichen Rechengesetzen zur Verfügung stehen — und das führt auf die Begriffe „Ring“ und „Körper“, die wir im nächsten Kapitel betrachten werden.

Zur geschichtlichen Einordnung: Bereits 100 n.Chr. wurden in Kapitel 8 des chinesischen Buches Jiŭ Zhāng Suànshù Systeme von bis zu drei Gleichungen mit bis zu drei Unbekannten gelöst. Dies ging insofern über heutige Abiturkenntnisse hinaus, als dafür schon die Matrix-Notation und das zuerst von Gauß in voller Allgemeinheit formulierte Eliminationsverfahren verwendet wurde.

Im Rest des Kapitels sei $\mathbb{K}$ ein Ring.[21]

2.1 Matrixnotation

Notation 2.1. *Seien $m, n \in \mathbb{N}^*$.*

a) $\mathbb{K}^n$ ist ja das n-fache kartesische Produkt von $\mathbb{K}$ mit sich selbst. $\vec{v} \in \mathbb{K}^n$ heißt, dass $\vec{v}$ ein Tupel aus n Elementen $v_1, ..., v_n \in \mathbb{K}$ ist, die man als die **Komponenten** *von $\vec{v}$ bezeichnet. Normalerweise schreiben wir Elemente von $\mathbb{K}^n$ mit einem Pfeilakzent als* **Spaltenvektor**, *also $\vec{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$. Manchmal stellen wir sie jedoch auch als* **Zeilenvektor** *dar; dafür werde ich meist einen Unterstrich statt einen Pfeilakzent verwenden, also etwa $\underline{v} = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$. Beim* **Nullvektor** *$\vec{0} \in \mathbb{K}^n$ sind alle Komponenten Null.*

[20] Wie gesagt: Betrachten Sie Ihre Schulkenntnisse möglichst nur noch im Rückspiegel.

[21] Die Passagen in roter Farbe wurden angepasst, nachdem in Kapitel 3 die Begriffe Ring und Körper eingefügt wurden. Ursprünglich sollte $\mathbb{K}$ für die Menge $\mathbb{Q}$ rationalen Zahlen oder die Menge $\mathbb{R}$ der reellen Zahlen stehen.

b) *Eine* $(m \times n)$-**Matrix** A *über* $\mathbb{K}$ *besteht aus* $m \cdot n$ *Elementen von* $\mathbb{K}$*, die in einem rechteckigen Schema mit* m *Zeilen und* n *Spalten aufgestellt sind. Die Menge aller* $(m \times n)$*-Matrizen über* $\mathbb{K}$ *bezeichnen wir als* $\mathbb{K}^{m \times n}$*. Im Spezialfall* $m = n$ *nennt man eine Matrix* **quadratisch***; Notation* $M_n(\mathbb{K}) := \mathbb{K}^{n \times n}$*.*

c) *Für* $i \in \{1, ..., m\}$ *und* $j \in \{1, ..., n\}$ *und* $A \in \mathbb{K}^{m \times n}$ *sei* $A_{i,*} \in \mathbb{K}^n$ *der aus der* i*-ten Zeile von* A *gebildeten Zeilenvektor und* $A_{*,j} \in \mathbb{K}^m$ *der aus der* j*-ten Spalte von* A *gebildete Spaltenvektor.* $A_{i,j} \in \mathbb{K}$ *bezeichnet den Eintrag in der* j*-ten Spalte der* i*-ten Zeile von* A*, die* (i, j)-**Komponente** *von* A*.*

d) *Schreibt man Spaltenvektoren* $\vec{v}_1, ..., \vec{v}_n \in \mathbb{K}^m$ *nebeneinander, so entsteht die Matrix* $(\vec{v}_1, ..., \vec{v}_n) \in \mathbb{K}^{m \times n}$*.*

e) *Für* $A \in \mathbb{K}^{m \times n}$ *ist die* **transponierte Matrix** ${}^tA \in \mathbb{K}^{n \times m}$ *definiert durch* $\forall i \in \{1, ..., m\},\ j \in \{1, ..., n\}$: ${}^tA_{j,i} := A_{i,j}$ *(Tausche Zeilen und Spalten).*

In manchen Büchern, besonders aus dem englischen Sprachraum, werden Vektoren und Matrizen mit eckigen statt runden Klammern notiert, also etwa $\left[\begin{smallmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \end{smallmatrix}\right]$ statt $\left(\begin{smallmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \end{smallmatrix}\right)$. In manchen Büchern und in meinen früheren Vorlesungen wird Transposition mit einem $\top$-Exponent notiert, also $A^\top$ statt ${}^tA$; in der englischen Wikipedia sind noch weitere Notationen genannt.

Beispiele:

- Für $A := \left(\begin{smallmatrix} 1 & 1 & 1 \\ 2 & 3 & 1 \\ 1 & -1 & 3 \end{smallmatrix}\right) \in M_3(\mathbb{R})$ haben wir Einträge $A_{3,2} = -1$ und $A_{2,3} = 1$, $A_{2,*} = \left(\begin{smallmatrix} 2 & 3 & 1 \end{smallmatrix}\right)$ und $A_{*,2} = \left(\begin{smallmatrix} 1 \\ 3 \\ -1 \end{smallmatrix}\right)$.
- ${}^t\left(\begin{smallmatrix} 1 & 3 & 7 & 4 \\ 3 & 1 & 2 & 9 \\ 8 & 0 & 7 & 3 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 & 3 & 8 \\ 3 & 1 & 0 \\ 7 & 2 & 7 \\ 4 & 9 & 3 \end{smallmatrix}\right) \in \mathbb{R}^{4 \times 3}$. Man beachte: ${}^tA = A$.

Wir sind nun in der Lage, lineare Gleichungssysteme besser als in der Schule zu notieren. Zunächst einmal sind die Namen der Variablen völlig egal, es ist besser, sie durchzunummerieren und als Komponenten eines Vektors anzusehen. Die Unbekannten im obigen Beispiel stellen wir also durch einen Spaltenvektor $\vec{x} \in \mathbb{K}^3$ mit den Komponenten x, y, z dar. Die Vorfaktoren der Variablen lassen sich in eine Matrix schreiben, die ***Koeffizientenmatrix*** des Gleichungssystems. Die Zahlen auf der rechten Seite des Gleichungssystems bilden ebenfalls einen Vektor $\vec{b} \in \mathbb{K}^3$, die ***Inhomogenität*** des Gleichungssystems.

Beispiel: Obiges Gleichungssystem hat die Koeffizientenmatrix $A := \left(\begin{smallmatrix} -2 & 1 & -2 \\ 2 & 0 & -2 \\ 0 & -4 & -4 \end{smallmatrix}\right) \in M_3(\mathbb{R})$ und die Inhomogenität $\vec{b} := \left(\begin{smallmatrix} 1 \\ 1 \\ -3 \end{smallmatrix}\right) \in \mathbb{R}^3$.

Die Zuordnung eines Vorfaktors zur richtigen Variable ergibt sich aus seiner Position: Steht ein Vorfaktor in der zweiten Spalte der Koeffizientenmatrix, so ist er in der betreffenden Gleichung mit der zweiten Variable zu multiplizieren; es wäre eine völlig überflüssige Schreibarbeit, die Variablen in die Gleichungen zu schreiben: Die volle Information über das Gleichungssystem steckt in der Koeffizientenmatrix A und der Inhomogenität $\vec{b}$ und wir schreiben das Gleichungssystem daher einfach als $A \cdot \vec{x} \stackrel{!}{=} \vec{b}$. Die Notation sieht so aus, als werde hier die Matrix A mit dem Vektor $\vec{x}$ multipliziert. Dieser Multiplikation soll nun geometrisch motiviert und im nächsten Abschnitt definiert werden. Und wenn man schon eine Multiplikation definiert, könnte man erwarten, dass es vielleicht für A ein Inverses A^{-1} gibt, so dass man das Gleichungssystem durch $\vec{x} = A^{-1} \cdot \vec{b}$ auflösen kann. Die Frage, wann A^{-1} existiert und wie man es ggf. berechnet, wird uns noch beschäftigen; vorsichtshalber sei gesagt, dass man in der Praxis so gut wie nie A^{-1} explizit berechnet, denn es gibt effizientere Methoden zur Lösung eines linearen Gleichungssystems.

Zeilenbild: Die Lösungsmenge des obigen Gleichungssystems ist als die Schnittmenge der Lösungsmengen der drei einzelnen Gleichungen. Sind nicht alle Koeffizienten einer linearen Gleichung Null, so hat mindestens eine Variable einen nicht-verschwindenden Koeffizient, und nach jeder solchen Variable kann man auflösen, sie also als Funktion der anderen Variablen darstellen.

Beispiel: $2\,x + 0\,y - 2\,z = 1$ kann man wahlweise nach $x = \frac{1}{2} + z$ oder nach $z = -\frac{1}{2} + x$ auflösen, jedoch nicht nach y. Man erhält also eine zur y-Achse parallele Ebene als Lösungsmenge der zweiten Gleichung. Sie enthält nicht $\vec{0}$, denn aufgrund der Inhomogenität schneidet die Ebene die x-Achse in $\begin{pmatrix} \frac{1}{2} \\ 0 \\ 0 \end{pmatrix}$.

Allgemein werden die Lösungen einer einzelnen linearen Gleichung mit n Variablen und nicht-verschwindendem Koeffizientenvektor durch $n-1$ frei wählbare Variablen beschrieben. In der Schule lernten Sie, wie man prüft, ob zwei Vektoren aufeinander senkrecht stehen; hier kann man sehen, dass die Lösungsmenge einer linearen Gleichung senkrecht auf dem Koeffizientenvektor steht.

Schneidet man zwei Ebenen in $\mathbb{R}^3$, so erhält man eine Gerade, eine Ebene oder die leere Menge. Schneidet man dementsprechend drei Ebenen in $\mathbb{R}^3$, so ist die Schnittmenge leer, ein einzelner Punkt, eine Gerade oder eine Ebene. Einen einzelnen Punkt, also eine eindeutige Lösung des linearen Gleichungssystems, erhält man genau dann, wenn die Zeilen von A nicht in einer gemeinsamen Ebene liegen.

Spaltenbild: Ist $\vec{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$ eine Lösung unseres Gleichungssystems, dann ist das x-Fache von $A_{*,1}$ plus das y-Fache von $A_{*,2}$ plus das z-Fache von $A_{*,3}$ gleich der Inhomogenität $\vec{b}$. Man sagt dazu, dass $\vec{b}$ als ***Linearkombination*** der

Spalten von A dargestellt wird (dazu wird es noch eine ordentliche Definition geben). Man erahnt: Wenn nicht alle drei Spalten der Matrix in einer gemeinsamen Ebene in $\mathbb{R}^3$ liegen, so bilden sie ein dreidimensionales Koordinatensystem und das Gleichungssystem hat für jede Inhomogenität eine eindeutige Lösung.

Als Ziel für die Definition der Matrixmultiplikation halten wir fest: Die Multiplikation einer Matrix mit n Spalten und eines Spaltenvektors mit n Komponenten soll eine Linearkombination der n Spalten ergeben.

Vereinfachung des Gleichungssystems: Aus der Schule kennen Sie das Additionsverfahren, bei dem man ein Gleichungssystem schrittweise vereinfacht, indem man geschickt Vielfache einer Gleichung zu den anderen Gleichungen addiert.

Beispiel: $\begin{pmatrix} -2 & 1 & -2 \\ 2 & 0 & -2 \\ 0 & -4 & -4 \end{pmatrix} \cdot \vec{x} \stackrel{!}{=} \begin{pmatrix} 1 \\ 1 \\ -3 \end{pmatrix} \rightsquigarrow \begin{pmatrix} -2 & 1 & -2 \\ 0 & 1 & -4 \\ 0 & -4 & -4 \end{pmatrix} \cdot \vec{x} \stackrel{!}{=} \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix} \rightsquigarrow \begin{pmatrix} -2 & 1 & -2 \\ 0 & 1 & -4 \\ 0 & 0 & -20 \end{pmatrix} \cdot \vec{x} \stackrel{!}{=} \begin{pmatrix} 1 \\ 2 \\ 5 \end{pmatrix}$. Wir können nun $\vec{x}$ von unten nach oben berechnen (***Rückwärtssubstitution***), indem wir die letzte Komponente mittels der letzten Gleichung berechnen, dies in die vorletzte Gleichung einsetzen und nach der vorletzten Kompontente auflösen, und so weiter: $\vec{x} = \begin{pmatrix} \frac{1}{4} \\ 1 \\ -\frac{1}{4} \end{pmatrix}$ ist die einzige Lösung.

Die Gleichungen sind in den Matrixzeilen codiert. Die obigen Operationen beruhen also auf Summen von Vielfachen von Matrixzeilen, also wieder Linearkombinationen.

Als Ziel für die Definition der Matrixmultiplikation halten wir fest: Die Multiplikation eines Zeilenvektors mit m Komponenten und einer Matrix mit m Zeilen soll eine Linearkombination der m Zeilen ergeben.

2.2 Matrixarithmetik

In diesem Abschnitt sei jeweils $k, \ell, m, n, p \in \mathbb{N}^*$ und $\mathbb{K}$ sei ein kommutativer Ring. Addition von Matrizen sowie Multiplikation einer Matrix mit einem Körperelement ist genau so, wie Sie es aus der Schule für Vektoren kennen:

Definition 2.2
*Für $c \in \mathbb{K}$, $A, B \in \mathbb{K}^{m \times n}$ definiert man $A + B := \begin{pmatrix} A_{1,1}+B_{1,1} & \dots & A_{1,n}+B_{1,n} \\ \vdots & & \vdots \\ A_{m,1}+B_{m,1} & \dots & A_{m,n}+B_{m,n} \end{pmatrix}$ und $c \cdot A := \begin{pmatrix} cA_{1,1} & \dots & cA_{1,n} \\ \vdots & & \vdots \\ cA_{m,1} & \dots & cA_{m,n} \end{pmatrix}$ (**Skalarmultiplikation**[22]).*

Da Spalten- und Zeilenvektoren ebenfalls Matrizen sind, ist dadurch die Addition und Skalarmultiplikation auf $\mathbb{K}^n$ ebenfalls definiert. Für die Multiplikation von Matrizen — und auch sonst! — sollte man das Summenzeichen kennen:

[22] Beachte: Es gibt auch *Skalarprodukte*, aber das ist etwas anderes!

Definition 2.3
Sei $n \in \mathbb{N}$ und $a_1, ..., a_n \in \mathbb{K}$. Wir definieren $\sum_{i=1}^{0} a_i := 0$ bzw. $\prod_{i=1}^{0} a_i := 1$. Für $n > 0$ sei $\sum_{i=1}^{n} a_i := \left(\sum_{i=1}^{n-1} a_i\right) + a_n$ bzw. $\prod_{i=1}^{n} a_i := \left(\prod_{i=1}^{n-1} a_i\right) \cdot a_n$. Analog für andere Indexbereiche.

Beispiele: a) $\sum_{k=-2}^{3} k^3 = -8 - 1 + 0 + 1 + 8 + 27 = 27$;

b) $\sum_{i=-100}^{100} 2 = 201 \cdot 2 = 402$.

c) ***Fakultät***: Für $n \in \mathbb{N}$ sei $n! := \prod_{i=1}^{n} i = 1 \cdot 2 \cdot ... \cdot (n-1) \cdot n$.

d) Summen bzw. Produkte mit *unendlich* vielen Summanden bzw. Faktoren sind nicht definiert! Allenfalls lassen sich Ausdrücke wie $\sum_{n \in \mathbb{N}^*} \frac{(-1)^n}{n}$ mit Hilfe von Grenzwerten, also mit Mitteln der Analysis, erfassen. Man nennt das dann nicht mehr Summe, sondern Reihe. Aber dann kann man im Allgemeinen nicht die Summanden beliebig vertauschen. Übrigens: Obige Reihe hat den Wert $-\ln(2)$, sofern man die Indizes aufsteigend $(1, 2, 3, 4, ...)$ sortiert.

Definition 2.4
Seien $M_1, ..., M_k \in \mathbb{K}^{m \times n}$. Die **Linearkombination** *von $M_1, ..., M_k$ mit Koeffizienten $c_1, ..., c_k \in \mathbb{K}$ ist $\sum_{i=1}^{k} c_i M_i \in \mathbb{K}^{m \times n}$.*

Die Definition umfasst insbesondere Linearkombinationen von Spalten- bzw. von Zeilenvektoren.

Definition 2.5 (Matrixmultiplikation)
Für $A \in \mathbb{K}^{m \times k}$ und $B \in \mathbb{K}^{k \times n}$ definiert man $AB \in \mathbb{K}^{m \times n}$ durch

$$(AB)_{i,j} := A_{i,1}B_{1,j} + A_{i,2}B_{2,j} + A_{i,3}B_{3,j} + \cdots + A_{i,k}B_{k,j} = \sum_{\ell=1}^{k} A_{i,\ell}B_{\ell,j}\,.$$

Damit das Produkt AB existiert, muss A genau so viel Spalten haben, wie B Zeilen hat — andernfalls ist das Matrixprodukt nicht definiert! Folgende Beobachtung ist äußert nützlich:

Beobachtung 2.6
Seien $A \in \mathbb{K}^{m \times k}$ und $B \in \mathbb{K}^{k \times n}$. Für alle $i \in \{1, ..., m\}$ und alle $j \in \{1, ..., n\}$ gilt $(A \cdot B)_{i,*} = A_{i,*} \cdot B$ (Linearkombination der Zeilen von B mit Koeffizienten $A_{i,1}, ..., A_{i,k}$) und $(A \cdot B)_{*,j} = A \cdot B_{*,j}$ (Linearkombination der Spalten von A mit Koeffizienten $B_{1,j}, ..., B_{k,j}$).

Beispiel: $\left(\begin{smallmatrix} 1 & 0 \\ 3 & 1 \end{smallmatrix}\right)\left(\begin{smallmatrix} 0 & 2 & 4 \\ 1 & 3 & 5 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1\cdot0+0\cdot1 & 1\cdot2+0\cdot3 & 1\cdot4+0\cdot5 \\ 3\cdot0+1\cdot1 & 3\cdot2+1\cdot3 & 3\cdot4+1\cdot5 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0 & 2 & 4 \\ 1 & 9 & 17 \end{smallmatrix}\right)$.

Lemma 2.7
Seien $A \in \mathbb{K}^{m\times k}$, $B \in \mathbb{K}^{k\times n}$. *Dann* ${}^{\mathrm{t}}(AB) = {}^{\mathrm{t}}B\,{}^{\mathrm{t}}A$.

Beweis:
$\forall i \in \{1,...,m\},\ j \in \{1,...,n\}$:

$${}^{\mathrm{t}}(AB)_{j,i} = (AB)_{i,j} = \sum_{\ell=1}^{k} A_{i,\ell}B_{\ell,j} = \sum_{\ell=1}^{k} {}^{\mathrm{t}}B_{j,\ell}\,{}^{\mathrm{t}}A_{\ell,i} = ({}^{\mathrm{t}}B\,{}^{\mathrm{t}}A)_{j,i}. \qquad \square$$

Definition 2.8

a) $\mathbb{0} \in \mathbb{K}^{m\times n}$ *bezeichnet die* **Nullmatrix** *mit* m *Zeilen und* n *Spalten, deren Einträge alle Null sind.*

b) $D \in M_n(\mathbb{K})$ *heißt* **Diagonalmatrix** $:\Leftrightarrow \forall i \neq j \in \{1,...,n\}\colon D_{i,j} = 0$. *Für* $\gamma_1,...,\gamma_n \in \mathbb{K}$ *sei* $\operatorname{diag}(\gamma_1,...,\gamma_n) := \begin{pmatrix} \gamma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \gamma_n \end{pmatrix} \in M_n(\mathbb{K})$.

c) Die **Einsmatrix**[23] *ist* $\mathbb{1}_n = \operatorname{diag}(1,...,1) \in M_n(\mathbb{K})$.

Aufgabe 2.9
In den folgenden Aussagen sei $\gamma, \delta \in \mathbb{K}$ *und* A, B, C *seien Matrizen, für die die angegebenen Rechenoperationen definiert sind (d.h. die Zeilen- und Spaltenzahlen „passen“).*

a) Assoziativität: $(\gamma\delta)A = \gamma(\delta A)$, $(\gamma A)B = \gamma(AB)$ *und* $(AB)C = A(BC)$; $(A+B)+C = A+(B+C)$.

b) Kommutativität der Addition: $A+B = B+A$. *Multiplikation ist i.A. nicht kommutativ! Jedoch gilt* $A(\gamma B) = \gamma(AB)$.

c) Distributivität: $\gamma(A+B) = \gamma A + \gamma B$, $A(B+C) = AB + AC$, $(A+B)C = AC + BC$ *und* $(\gamma+\delta)A = \gamma A + \delta A$.

d) Für $A \in R^{m\times n}$ *gelten* $A \cdot \mathbb{1}_n = A$, $\mathbb{1}_m \cdot A = A$ *und* $\mathbb{0} + A = A$. *Falls* $\mathbb{0}A$ *bzw.* $A\mathbb{0}$ *definiert ist, ist das Ergebnis* $\mathbb{0}$.

e) $\forall A \in R^{m\times n}\colon\ \exists -A \in R^{m\times n}\colon\ A + (-A) = \mathbb{0}$.

Zeilenoperationen und Elementarmatrizen

Da wir in diesem Kapitel auch dividieren müssen, sei nun $\mathbb{K}$ ein Körper. In vielen Rechenverfahren der linearen Algebra treten drei Typen von ***Zeilenoperationen*** und zum Teil auch die entsprechenden ***Spaltenoperationen*** auf:

I) Zeilen i und ℓ miteinander vertauschen $(i \neq \ell)$.

[23] Auch ***Einheitsmatrix*** genannt.

II) Zeile i mit $\lambda \in \mathbb{K}^*$ multiplizieren.

III) Das γ-Fache von Zeile i zu Zeile ℓ addieren ($\gamma \in \mathbb{K}$, $i \neq \ell$).

Lemma 2.10 (und Definition und Notation)

a) Seien $M \in \mathbb{K}^{m\times n}$, $N \in \mathbb{K}^{n\times \ell}$ Wenn M' aus M durch eine Zeilenoperation enteht, dann entsteht $M' \cdot N$ aus $M \cdot N$ durch dieselbe Zeilenoperation.

b) Eine Matrix, die aus $\mathbb{1}_m$ durch eine einzelne Zeilenoperation entsteht, bezeichnet man als die **Elementarmatrix** *für diese Zeilenoperation. Es sei $T_{i,\ell} \in M_m(\mathbb{K})$ die Elementarmatrix für die Vertauschung der Zeilen $i \neq \ell$, $D_i(\lambda) \in M_m(\mathbb{K})$ die Elementarmatrix für die Multiplikation der Zeile i mit $\lambda \in \mathbb{K}\setminus\{0\}$, und $L_{i,\ell}(\gamma) \in M_m(\mathbb{K})$ mit $i \neq \ell$ und $\gamma \in \mathbb{K}$ die Elementarmatrix für die Addition des γ-Fachen der i-ten Zeile zur ℓ-ten Zeile.*

Ist $A \in \mathbb{K}^{m\times n}$ und ist $E \in M_m(\mathbb{K})$ die Elementarmatrix zu einer Zeilenoperation, so entsteht EA aus A durch Anwendung der Zeilenoperation.

Beispiel: In $M_4(\mathbb{R})$: $\begin{pmatrix} 1&0&0&0\\ 0&0&0&1\\ 0&0&1&0\\ 0&1&0&0 \end{pmatrix} = T_{2,4}$, $\begin{pmatrix} 1&0&0&0\\ 0&1&0&0\\ 0&0&5&0\\ 0&0&0&1 \end{pmatrix} = D_3(5)$, $\begin{pmatrix} 1&0&0&0\\ 0&1&0&0\\ -2&0&1&0\\ 0&0&0&1 \end{pmatrix} = L_{1,3}(-2)$.

Beweis:

a) Mit Beobachtung 2.6 folgt für die drei Typen der Zeilenoperationen:

I) Sei $M'_{i,*} = M_{\ell,*}$ und $M'_{\ell,*} = M_{i,*}$. Dann $(M'\cdot N)_{i,*} = M'_{i,*}\cdot N = M_{\ell,*}\cdot N = (M\cdot N)_{\ell,*}$. Analog $(M'\cdot N)_{\ell,*} = (M\cdot N)_{i,*}$.

II) Sei $M'_{i,*} = \lambda \cdot M_{i,*}$. Dann $(M'\cdot N)_{i,*} = (\lambda\cdot M_{i,*})\cdot N = \lambda\cdot(M\cdot N)_{i,*}$.

III) Sei $M'_{\ell,*} = M_{\ell,*}+\gamma\cdot M_{i,*}$. Dann $(M'\cdot N)_{\ell,*} = M'_{\ell,*}\cdot N = (M_{\ell,*}+\gamma\cdot M_{i,*})\cdot N = M_{\ell,*}\cdot N + \gamma\cdot M_{i,*}\cdot N = (M\cdot N)_{\ell,*} + \gamma\cdot(M\cdot N)_{i,*}$.

b) Direkte Konsequenz von a), denn $A = \mathbb{1}_m A$. □

Jede Zeilenoperation lässt sich durch eine Zeilenoperation rückgängig machen. Entsprechend lässt sich auch die Multiplikation mit einer Elementarmatrix rückgängig machen, was auf den folgenden Begriff führt.

Definition 2.11

$A \in M_n(\mathbb{K})$ heißt **invertierbar** $:\Leftrightarrow \exists A^{-1} \in M_n(\mathbb{K})\colon\ AA^{-1} = A^{-1}A = \mathbb{1}_n$. *Ggf. heißt A^{-1} die zu A* **inverse Matrix**. *Die Menge $GL_n(\mathbb{K}) := \{A \in M_n(\mathbb{K}) \mid A$ invertierbar$\}$ heißt* **allgemeine lineare Gruppe**.

Wir werden noch lernen, wie man ggf. die inverse Matrix berechnen kann. Bei den zugehörigen Beweisen helfen Elementarmatrizen, denn sie sind invertierbar:

Lemma 2.12

a) Wenn $A \in GL_n(\mathbb{K})$, dann ist A^{-1} eindeutig bestimmt und $A = (A^{-1})^{-1}$.

b) $\forall E_1, ..., E_m \in M_n(\mathbb{K})$ *Elementarmatrizen:* $G := E_1 \cdots E_m \in GL_n(\mathbb{K})$.

Bemerkung Wir werden noch sehen, dass sich jede invertierbare Matrix als Produkt von Elementarmatrizen darstellen lässt.

Beweis:

a) Seien $B, C \in M_n(\mathbb{K})$, so dass $AB = BA = AC = CA = \mathbb{1}_n$. Dann ist $B = B\mathbb{1}_n = BAC = \mathbb{1}_n C = C$. Die zweite Aussage folgt direkt aus der Definition.

b) Wir bemerken zunächst, dass jede Elementarmatrix invertierbar ist. Mit den Notationen aus Lemma 2.10: $T_{i,\ell} = T_{i,\ell}^{-1}$, $D_i(\lambda^{-1}) = D_i(\lambda)^{-1}$ (beachte: $\lambda \neq 0$, und da in $\mathbb{K}$ auch Divisionen durchführbar sind, ist $\lambda^{-1} \in \mathbb{K}$; für $\mathbb{K} = \mathbb{Z}$ würde das also nicht funktionieren.), und $L_{i,\ell}(-\gamma) = L_{i,\ell}(\gamma)^{-1}$ (beachte: In $\mathbb{K}$ existiert zu jedem Element ein Negatives; für $\mathbb{K} = \mathbb{N}$ würde das also nicht funktionieren.).

Es ist $G^{-1} = E_m^{-1} E_{m-1}^{-1} \cdots E_1^{-1}$, denn $E_1 \cdots E_m E_m^{-1} E_{m-1}^{-1} \cdots E_1^{-1} = \mathbb{1}_n = E_m^{-1} E_{m-1}^{-1} \cdots E_1^{-1} E_1 \cdots E_m E_m^{-1}$ (von der Mitte ausgehend heben sich jeweils eine Elementarmatrix und ihr Inverses auf). □

2.3 Lösungsräume

Definition 2.13. *Sei $\mathbb{K}$ ein Körper, $A \in \mathbb{K}^{m\times n}$ und $\vec{b} \in \mathbb{K}^m$.*

a) $A\vec{x} \stackrel{!}{=} \vec{b}$ *heißt* **lineares Gleichungssystem** *mit* **Koeffizientenmatrix** A, **Inhomogenität** $\vec{b}$ *und* **Unbekannter** $\vec{x} \in \mathbb{K}^n$.

b) Ist die Inhomogenität Null, also $A\vec{x} = \vec{0}$, so heißt das lineare Gleichungssystem **homogen**.

c) $\mathrm{LR}(A; \vec{b}) := \{\vec{x} \in \mathbb{K}^n \mid A\vec{x} = \vec{b}\}$ *heißt* **Lösungsraum** *des linearen Gleichungssystems* $A\vec{x} \stackrel{!}{=} \vec{b}$.

d) Die aus den Spalten von A und zusätzlich $\vec{b}$ gebildete Matrix $(A \mid \vec{b}) \in \mathbb{K}^{m\times(n+1)}$ *heißt* **erweiterte Matrix**. *Durch sie ist das lineare Gleichungssystem eindeutig bestimmt.*

Lineare Gleichungssysteme haben nicht immer eine eindeutige Lösung:

Beispiel: Sei $A := \left(\begin{smallmatrix} -2 & 1 & -2 \\ 2 & 0 & -2 \end{smallmatrix}\right) \in \mathbb{R}^{2\times 3}$ und $\vec{b} := \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right) \in \mathbb{R}^2$. Das lineare Gleichungssystem wird in einer erweiterten Matrix codiert. Wir wissen aus einem früheren Beispiel, wie man das Gleichungssystem durch eine Zeilenoperation vereinfacht: $(A|\vec{b}) = \left(\begin{array}{ccc|c} -2 & 1 & -2 & 1 \\ 2 & 0 & -2 & 1 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|c} -2 & 1 & -2 & 1 \\ 0 & 1 & -4 & 2 \end{array}\right)$. Wir stellen fest, dass man

für jeden Wert der dritten Variable eine Lösung erhält, das heißt wir können die Lösungen (die wieder mit ***Rückwärtssubstitution*** berechnet werden) parametrisieren: $\mathrm{LR}(A;\vec{b}) = \{\begin{pmatrix} \frac{1}{2}+c \\ 2+4\,c \\ c \end{pmatrix} \mid c \in \mathbb{R}\}$. Ein freier Parameter bedeutet wohl, dass der Lösungsraum eine Gerade in $\mathbb{R}^3$ darstellt. Gibt es dafür vielleicht eine schönere Notation? Es besteht ja eigentlich kein Grund, dem Parameter einen Namen zu geben.

Ich habe Sie bereits darauf hingewiesen, dass Sie die in der Schule üblichen Notationen für lineare Gleichungssysteme nicht mehr verwenden sollen. Um auch in der Angabe von Lösungsräumen auf günstige Notationsweisen zu kommen, befassen wir uns nun mit der allgemeinen Struktur von Lösungsräumen linearer Gleichungssysteme:

Satz 2.14 (Algebr. Eigenschaften von Lösungsräumen)
Sei $\mathbb{K}$ ein Körper, $A \in \mathbb{K}^{m\times n}$ und $\vec{b} \in \mathbb{K}^m$.

a) $\forall \vec{x}_1, \vec{x}_2 \in \mathrm{LR}(A;\vec{0})\colon\ \vec{x}_1 + \vec{x}_2 \in \mathrm{LR}(A;\vec{0})$.

b) $\forall c \in \mathbb{K},\ \vec{x} \in \mathrm{LR}(A;\vec{0})\colon\ c\vec{x} \in \mathrm{LR}(A;\vec{0})$.

c) Sei $\vec{x}_{\mathrm{inh}} \in \mathrm{LR}(A;\vec{b})$. *Dann* $\mathrm{LR}(A;\vec{b}) = \left\{\vec{x}_{\mathrm{inh}} + \vec{x}_{\mathrm{h}} \mid \vec{x}_{\mathrm{h}} \in \mathrm{LR}(A;\vec{0})\right\}$.

Beweis:

a) $A\cdot(\vec{x}_1+\vec{x}_2) = A\cdot\vec{x}_1 + A\cdot\vec{x}_2 = \vec{0}+\vec{0} = \vec{0}$.

b) $A\cdot(c\vec{x}) = c(A\vec{x}) = c\vec{0} = \vec{0}$.

c) „$\subset$“: Sei $\vec{x} \in \mathrm{LR}(A;\vec{b})$. Zu zeigen: $\vec{x}_{\mathrm{h}} := \vec{x} - \vec{x}_{\mathrm{inh}} \in \mathrm{LR}(A;\vec{0})$:

$$A\cdot(\vec{x}-\vec{x}_{\mathrm{inh}}) = A\vec{x} - A\vec{x}_{\mathrm{inh}} = \vec{b} - \vec{b} \overset{\checkmark}{=} \vec{0}$$

„$\supset$“: Sei $\vec{x}_{\mathrm{h}} \in \mathrm{LR}(A;\vec{0})$. Zu zeigen: $\vec{x} := \vec{x}_{\mathrm{inh}} + \vec{x}_{\mathrm{h}} \in \mathrm{LR}(A;\vec{b})$:

$$A\cdot(\vec{x}_{\mathrm{inh}}+\vec{x}_{\mathrm{h}}) = A\vec{x}_{\mathrm{inh}} + A\vec{x}_{\mathrm{h}} = \vec{b}+\vec{0} \overset{\checkmark}{=} \vec{b}$$ □

Definition 2.15. *Es sei $\mathbb{K}$ ein komm. Ring, $k \in \mathbb{N}^*$ und $\vec{v}_0, \vec{v}_1, ..., \vec{v}_k \in \mathbb{K}^n$.*

a) $\mathrm{Span}_{\mathbb{K}}(\vec{v}_1, ..., \vec{v}_k) := \{\sum_{j=1}^{k} c_j\vec{v}_j \mid c_1, ..., c_k \in \mathbb{K}\}$, *also die Menge der Linearkombinationen, die aus $\vec{v}_1, ..., \vec{v}_k$ gebildet werden können, heißt* **Erzeugnis** *oder* **$\mathbb{K}$-Span** *von $\vec{v}_1, ..., \vec{v}_k$.*

b) $\vec{v}_0 + \mathrm{Span}_{\mathbb{K}}(\vec{v}_1, ..., \vec{v}_k) := \{\vec{v}_0 + \vec{w} \mid \vec{w} \in \mathrm{Span}_{\mathbb{K}}(\vec{v}_1, ..., \vec{v}_k)\}$.

c) $\vec{v}_1, ..., \vec{v}_k$ *heißen* **linear unabhängig** *gdw.* $\forall c_1, ..., c_k \in \mathbb{K}$: $\sum_{j=1}^{k} c_j \vec{v}_j = \vec{0} \Rightarrow c_1 = ... = c_k = 0$. *Andernfalls heißen sie* **linear abhängig**.

Beispiele:

a) $\text{Span}_{\mathbb{R}}(\binom{1}{0}, \binom{0}{1}) = \{\binom{x_1}{x_2} \mid x_1, x_2 \in \mathbb{R}\} = \mathbb{R}^2$, aber $\binom{1}{0} + \text{Span}_{\mathbb{R}}(\binom{0}{1}) = \{\binom{1}{x_2} \mid x_2 \in \mathbb{R}\}$: Im Allgemeinen ist $\vec{v}_0 + \text{Span}_{\mathbb{K}}(\vec{v}_1, ..., \vec{v}_k) \neq \text{Span}_{\mathbb{K}}(\vec{v}_0, \vec{v}_1, ..., \vec{v}_k)$.

b) Oben lösten wir ein Gleichungssystem und erhielten $\text{LR}(A; \vec{b}) = \begin{pmatrix} \frac{1}{2} \\ 2 \\ 0 \end{pmatrix} + \text{Span}_{\mathbb{R}}(\begin{pmatrix} 1 \\ 4 \\ 1 \end{pmatrix})$. Dem entspricht die vermutlich aus der Schule bekannte Darstellung einer Gerade durch einen Stütz- und einen Richtungsvektor.

c) Seien $\vec{v}_1 = \binom{1}{0}$, $\vec{v}_2 = \binom{0}{1}$, $\vec{v}_3 = \binom{1}{1}$. Wegen $\vec{v}_1 + \vec{v}_2 - \vec{v}_3 = \vec{0}$ sind $\vec{v}_1, \vec{v}_2, \vec{v}_3$ linear abhängig.

Beobachtung 2.16. *Seien* $\vec{v}_1, ..., \vec{v}_k \in \mathbb{K}^n$.

a) $\text{Span}_{\mathbb{K}}(\vec{v}_1, ..., \vec{v}_k) = \{(\vec{v}_1, ..., \vec{v}_k) \cdot \vec{c} \mid \vec{c} \in \mathbb{K}^k\}$.

b) Somit kann man wie folgt testen, ob $\vec{v}_1, ..., \vec{v}_k$ *linear unabhängig sind: Es sei* $A := (\vec{v}_1, ..., \vec{v}_k)$. $\vec{v}_1, ..., \vec{v}_k$ *sind linear unabhängig* $\iff \text{LR}(A; \vec{0}) = \{\vec{0}\}$.

Beispiel: Sei wieder $\vec{v}_1 = \binom{1}{0}$, $\vec{v}_2 = \binom{0}{1}$, $\vec{v}_3 = \binom{1}{1}$. Weil $(\vec{v}_1, \vec{v}_2, \vec{v}_3)$ eine besonders einfache Gestalt hat (Zeilenstufenform, siehe unten), können wir den Lösungsraum bereits berechnen: $\text{LR}(\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}; \vec{0}) = \text{Span}_{\mathbb{R}}(\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}) \neq \{\vec{0}\}$.

Wir wollen Lösungsräume möglichst sparsam beschreiben. Wir werden einige so genannte Basislösungen berechnen, so dass jede Lösung von $A\vec{x} = \vec{0}$ eine *eindeutige* Beschreibung als Linearkombination von Basislösungen besitzt. Wenn wir dann noch eine einzelne Lösung von $A\vec{x} = \vec{b}$ berechnen (oder nachweisen, dass es keine Lösung gibt), ist dadurch wegen des vorigen Satzes $\text{LR}(A; \vec{b})$ berechnet.

Zeilenstufenform

Wir behandeln zuerst einen Spezialfall, der eine relativ einfache Ablesung des Lösungsraums erlaubt.

Definition 2.17. *Es sei* $\mathbb{K}$ *ein Körper und* $A \in \mathbb{K}^{m \times n}$.
Für $i = 1, ..., m$ *sei* j_i *die Nummer der ersten Spalte, in der ein von Null verschiedener Eintrag in Zeile* i *steht.* A *ist in* **Zeilenstufenform** *(kurz:* **ZSF***), wenn es ein* $0 \leq r \leq m$ *gibt, so dass die Zeilen* $1, ..., r$ *nicht Null sind*[24]*, die Zeilen* $r+1, ..., m$ *Null sind, und* $j_1 < j_2 < ... < j_r$. j_i *heißt* **Pivotspalte** *der Zeile* i.

[24] Für $r = 0$ wäre diese Bedingung leer.

Beispiele: $\left(\begin{array}{c|cccc} 0 & 2 & 1 & 0 & 3 & 4 \\ \hline 0 & 0 & 0 & 3 & 1 & 1 \\ 0 & 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$ ist ZSF, $\left(\begin{array}{c|cc} 3 & 4 & 5 \\ \hline 0 & 2 & 3 \\ 0 & -1 & 2 \end{array}\right)$ ist keine ZSF.

Ist $A \in \mathbb{K}^{m\times n}$ in ZSF, so berechnet man $\mathrm{LR}(A; \vec{0})$, indem man diejenigen Unbekannten frei wählt, die nicht den Pivotspalten der Matrix entsprechen, und dann nach den übrigen Unbekannten schrittweise „von unten nach oben“ auflöst.

Definition 2.18. *Sei $\mathbb{K}$ ein Körper.*
Sei $A \in \mathbb{K}^{m\times n}$ in ZSF mit Pivotspalten $j_1 < j_2 < \cdots < j_r$. Sei $\vec{b} \in \mathbb{K}^m$. Wir nennen $x_{j_1}, x_{j_2}, ..., x_{j_r}$ **gebundene Variablen** *und die anderen Komponenten von $\vec{x}$* **freie Variablen** *des linearen Gleichungssystems $A\vec{x} = \vec{b}$.*

Problem 2.19 (Lösungsraum für Matrizen in Zeilenstufenform)
Sei $\mathbb{K}$ ein Körper, $A \in \mathbb{K}^{m\times n}$ ZSF mit Pivotspalten $j_1 < j_2 < \cdots < j_r$ und $\vec{b} \in \mathbb{K}^m$.

Wenn es $i \in \{r+1, ..., m\}$ mit $b_i \neq 0$ gibt, so ist $\mathrm{LR}(A; \vec{b}) = \emptyset$. Andernfalls ist durch jede Wahl von Werten aus $\mathbb{K}$ für die freien Variablen eine Lösung von $A\vec{x} = \vec{b}$ eindeutig bestimmt.

Lösung:
Wenn $\vec{b}$ in einer der Zeilen $r+1, r+2, ..., m$ nicht Null ist, ist die betreffende Gleichung nicht erfüllbar (auf der linken Seite steht Null, auf der rechten nicht Null). Andernfalls sind die letzten $m - r$ Gleichungen $0 = 0$, sind also sicher erfüllt. Wir streichen konzentrieren uns nun auf die obersten r Zeilen.

Wir betrachten Zeile i für $i = r, r-1, ..., 1$ (also von unten nach oben). In der i-ten Gleichung tritt genau eine noch nicht bestimmte Unbekannte auf, nämlich die gebundene Variable x_{j_i}. Alle $x_{j_i+1}, ..., x_n$ wurden als freie Variable gewählt oder bereits vorher als gebundene Variable berechnet.

Wir lösen die i-te Gleichung nach x_{j_i} auf (alle anderen Variablen auf die rechte Seite bringen, durch A_{i,j_i} dividieren) und können somit x_{j_i} einen eindeutigen Wert zuweisen, durch den die i-te Gleichung erfüllt ist. Durch diese so genannte ***Rückwärtssubstitution*** berechnet man die Werte aller gebundenen Variablen in Abhängigkeit von den freien Variablen. □

Korollar 2.20 (und Definition). *Sei $\mathbb{K}$ ein Körper.*
Sei $A \in \mathbb{K}^{m\times n}$ in ZSF und $\vec{b} \in \mathbb{K}^m$ so, dass $\mathrm{LR}(A; \vec{b}) \neq \emptyset$. Sei $J \subset \{1, ..., n\}$ die Menge der Indizes der Nichtpivotspalten (also der freien Variablen).

a) Das $\vec{x}_{\text{spez}} \in \mathrm{LR}(A; \vec{b})$, bei dem alle freien Variablen 0 sind, heißt **spezielle Lösung** *des inhomogenen Gleichungssystems $A\vec{x} \stackrel{!}{=} \vec{b}$.*

b) Für $j \in J$ heißt dasjenige $\vec{\beta}_j \in \mathrm{LR}(A;\vec{0})$, bei dem die freie Variable $x_j = 1$ und alle anderen freien Variablen 0 sind, heißt die zu x_j gehörende **Basislösung** *des homogenen Gleichungssystems $A\vec{x} \overset{!}{=} \vec{0}$.*

Es gilt $\mathrm{LR}(A;\vec{b}) = \vec{x}_{\text{spez}} + \mathrm{Span}_{\mathbb{K}}(\{\vec{\beta}_j \mid j \in J\})$. □

Beispiel: $A := \left(\begin{array}{cccccc} \boxed{0} & 2 & 1 & 0 & 3 & 4 \\ 0 & \boxed{0} & 0 & 3 & 1 & 1 \\ 0 & 0 & \boxed{0} & \boxed{0} & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \in \mathbb{R}^{5\times 6}$. Für $\vec{a} = \begin{pmatrix} 2 \\ -1 \\ 3 \\ 0 \\ 1 \end{pmatrix}$ ist die letzte Zeile von $A\vec{x} = \vec{a}$ ist die letzte Gleichung nicht erfüllbar, also $\mathrm{LR}(A;\vec{a}) = \emptyset$.

Sei nun $\vec{b} = \begin{pmatrix} 2 \\ -1 \\ 3 \\ 0 \\ 0 \end{pmatrix}$. Diesmal sind alle Gleichungen erfüllbar. Wir haben $r = 3$ und streichen die letzten beiden Zeilen. Somit sind die Gleichungen

$$\left(\begin{array}{cccccc|c} 0 & 2 & 1 & 0 & 3 & 4 & 2 \\ 0 & 0 & 0 & 3 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 3 & 1 & 3 \end{array}\right).$$

Die gebundenen Variablen sind x_2, x_4, x_5, und die freien sind x_1, x_3, x_6.

$\vec{x}_{\text{spez}} = \begin{pmatrix} 0 \\ -1/2 \\ 0 \\ -2/3 \\ 1 \\ 0 \end{pmatrix}$. Zunächst setzt man die freien Variablen Null und Rückwärtssubstitution liefert $\frac{3-1\cdot 0}{3} = 1$, $\frac{-1-1\cdot 1-1\cdot 0}{3} = -2/3$, $\frac{2-1\cdot 0-0\cdot\frac{-2}{3}-3\cdot 1-4\cdot 0}{2} = -1/2$.

$\vec{\beta}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$. Rückwärtssubstitution des homogenen(!) Systems liefert in den gebundenen Variablen nur 0.

$\vec{\beta}_3 = \begin{pmatrix} 0 \\ -1/2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$. Rückwärtssubstitution liefert nämlich $\frac{0-1\cdot 1}{2} = -\frac{1}{2}$.

$\vec{\beta}_6 = \begin{pmatrix} 0 \\ -3/2 \\ 0 \\ -2/9 \\ -1/3 \\ 1 \end{pmatrix}$, denn Rückwärtssubstitution liefert $\frac{0-1\cdot 1}{3} = -\frac{1}{3}$, $\frac{0-1\cdot\frac{-1}{3}-1\cdot 1}{3} = -\frac{2}{9}$ und $\frac{0-1\cdot 0-0\cdot\frac{-2}{9}-3\cdot\frac{-1}{3}-4\cdot 1}{2} = -\frac{3}{2}$.

Ergebnis: $\mathrm{LR}(A;\vec{b}) = \vec{x}_{\text{spez}} + \mathrm{Span}_{\mathbb{R}}(\vec{\beta}_1, \vec{\beta}_3, \vec{\beta}_6)$.

In der Lösung von Problem 2.19 wurden nur Grundrechenarten verwendet — einschließlich Division durch $A_{i,j_i} \neq 0$. Das geht, denn wir setzten voraus, dass $\mathbb{K}$ ein Körper ist.

2.4 Das Gaußsche Eliminationsverfahren

Wieder sei $\mathbb{K}$ ein Körper. Was macht das Eliminationsverfahren mit einem linearen Gleichungssystem mit Koeeffizientenmatrix $A \in \mathbb{K}^{m\times n}$? Dazu gibt es mindestens vier Sichtweisen:

- Es nutzt Äquivalenzumformungen linearer Gleichungssysteme.
- Es nutzt die schon bekannten Zeilenoperationen.
- Es liefert $G \in GL_m(\mathbb{K})$, so dass GA Zeilenstufenform hat.
- Es beruht auf einer Koordinatentransformation in $\mathbb{K}^m$.

Die vorletzte Sichtweise ist unter anderem deshalb praktisch, weil man G nur einmal berechnen braucht — und dann kann man $\mathrm{LR}(A;\vec{b})$ für *beliebige* Inhomogenitäten $\vec{b}$ lösen. Das wird für Anwendungen in der Numerik wichtig werden, denn löst man $A\vec{x} \stackrel{!}{=} \vec{b}$ nur näherungsweise, so kann man die Näherungslösung durch einen Korrekturterm verbessern, der sich aus der Lösung des Gleichungssystems mit einer anderen Inhomogenität ergibt.

Algorithmus 2.21 (Das Gaußsche Eliminationsverfahren)
Forme $A \in \mathbb{K}^{m\times n}$ durch **Zeilenoperationen** *in ZSF um. Beginne mit $i := 1$:*

1) *Abbruch, falls die Zeilen $i, ..., m$ alle Null sind. Sonst: Sei $j \in \{1, ..., n\}$ minimal, so dass ein $A_{\ell,j} \neq 0$ mit $\ell \in \{i, ..., m\}$ existiert. Erzwinge $A_{i,j} \neq 0$, indem ggf. Zeilen i, ℓ für ein $\ell > i$ vertauscht werden.* ***Anmerkung:*** *Hier besteht manchmal eine Wahlmöglichkeit.*
2) *Optional: Ersetze Zeile i durch ihr $1/A_{i,j}$-Faches ($\rightsquigarrow A_{i,j} = 1$).*
3) *Für alle $\ell \in \{i+1, ..., m\}$: Ziehe das $\frac{A_{\ell,j}}{A_{i,j}}$-Fache der Zeile i von Zeile ℓ ab ($\rightsquigarrow A_{\ell,j} = 0$). Optional: Führe dies auch für alle $\ell \in \{1, ..., i-1\}$ durch.*
4) *Erhöhe i um 1 und gehe zurück zu Schritt 1).*

Beispiel: Gauß-Algorithmus für $A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 6 & 10 & 11 \\ -1 & -3 & -1 & 1 \\ 1 & 1 & 7 & 7 \end{pmatrix}$: Es ist $A_{1,1} \neq 0$. Wir ziehen die erste Zeile dreimal von der zweiten und einmal von der vierten Zeile ab und addieren die erste zur dritten: $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & -1 \\ 0 & -1 & 2 & 5 \\ 0 & -1 & 4 & 3 \end{pmatrix}$. Jetzt $i := 2$: Wegen $A_{1,2} = 0$ vertauschen wir Zeilen 2 und 3 von A: $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -1 & 2 & 5 \\ 0 & 0 & 1 & -1 \\ 0 & -1 & 4 & 3 \end{pmatrix}$. Jetzt ziehen wir die zweite von der vierten Zeile ab: $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -1 & 2 & 5 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 2 & -2 \end{pmatrix}$ und wiederholen mit $i := 3$. Diesmal ist $j = 3$. Indem wir das Doppelte der dritten Zeile von der vierten abziehen, erhalten wir die Zeilenstufenform $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -1 & 2 & 5 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$.

Korollar 2.22. *Sei $\mathbb{K}$ ein Körper.*
Sei $A \in \mathbb{K}^{m\times n}$, $\vec{b} \in \mathbb{K}^m$. Das Gauß-Verfahren bringt die erweiterte Matrix $B := (A|\vec{b})$ nach endlich vielen Schritten auf $B' = (A'|\vec{b'})$ mit einer Zeilenstufenform $A' \in \mathbb{K}^{m\times n}$ und $\vec{b'} \in \mathbb{K}^m$, und es gilt $\mathrm{LR}(A;\vec{b}) = \mathrm{LR}(A';\vec{b'})$.

Beweis:
Schritte 1)–4) erfordern jeweils nur endlich viele Rechenoperationen und werden nacheinander auf Zeilen 1 bis $m-1$ angewandt. Spätestens dann erfolgt Abbruch.

Bei Anwendung der Schritte auf Zeile i sei $A_{i,j}$ das erste von Null verschiedene Element. Mit Induktion folgt: Spalten $1,...,j-1$ sind auch unterhalb von Zeile i Null. Nach Anwendung von Zeilenoperationen bleiben sie Null und zudem gilt $\forall \ell \in \{i+1,...,m\}\colon\ A_{\ell,j} = 0$. Also entsteht eine Zeilenstufenform. Jede Zeilenoperation entspricht der Multiplikation mit einer Elementarmatrix. Es gibt also ein $G \in GL_m(\mathbb{K})$, nämlich das Produkt dieser Elementarmatrizen, mit $A' = GA$.

Wendet man die gleichen Zeilenoperationen auf $B = (A|\vec{b})$ an, entsteht $B' = (A'|\vec{b'}) = GB = (GA \mid G\vec{b})$. Für alle $\vec{x} \in \mathbb{K}^n$ gilt $A\vec{x} = \vec{b} \iff GA\vec{x} = G\vec{b} \iff A'\vec{x} = \vec{b'}$, denn G ist invertierbar. Also ist $\mathrm{LR}(A;\vec{b}) = \mathrm{LR}(A';\vec{b'})$. □

Beispiel 2.23 (und Definition)
Führt man im Gauß-Algorithmus auch alle optionalen Schritte durch, nennt man dies den **Gauß-Jordan-Algorithmus**[25]*. Es entsteht eine Matrix in Zeilenstufenform, bei der zusätzlich jede Pivotspalte* genau ein *von Null verschiedenes Element enthält, und dieses hat den Wert* 1 *(***reduzierte Zeilenstufenform***).*

Sei $A = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 6 & 10 \\ -1 & -3 & -1 \\ 1 & 1 & 7 \end{pmatrix}$ *und* $\vec{b} = \begin{pmatrix} 4 \\ 11 \\ 1 \\ 7 \end{pmatrix}$. *Dann* $B := (A|\vec{b}) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 6 & 10 & 11 \\ -1 & -3 & -1 & 1 \\ 1 & 1 & 7 & 7 \end{pmatrix}$. *Wir sahen oben, dass der Gauß-Algorithmus* $B' = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -1 & 2 & 5 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ *ergibt. Beim Gauß-Jordan-Algorithmus geht es noch etwas weiter:* $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & -2 & -5 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 7 & 14 \\ 0 & 1 & -2 & -5 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 0 & 21 \\ 0 & 1 & 0 & -7 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} = B''$. *Man liest ab:* $\mathrm{LR}(A;\vec{b}) = \{\begin{pmatrix} 21 \\ -7 \\ -1 \end{pmatrix}\}$.

Ein Gleichungssystem in reduzierter Zeilenstufenform ist besonders einfach lösbar, allerdings wiegt dies nicht immer den zusätzlichen Aufwand für die optionalen Schritte auf. Wahlmöglichkeiten bei den optionalen Schritten können beim gerundeten Rechnen (also erst im vierten Semester in Numerik) zur Reduzierung von Rundungsfehlern genutzt werden.

[25] Wilhelm Jordan [1842–1899] war ein deutscher Geodät.

Wir können nun auch das Matrixinvertierungsproblem rechnerisch lösen.

Satz 2.24. *Sei $\mathbb{K}$ ein Körper.*
Die folgenden Aussagen über $A \in M_n(\mathbb{K})$ sind zueinander logisch äquivalent[26]

a) A ist invertierbar.

b) $\forall \vec{b} \in \mathbb{K}^n : \exists! \vec{x} \in \mathbb{K}^n : A\vec{x} = \vec{b}$.

c) $\mathrm{LR}(A; \vec{0}) = \{\vec{0}\}$.

d) Der Gauß-Jordan-Algorithmus formt A in $\mathbb{1}_n$ um.

e) A ist gleich einem Produkt von Elementarmatrizen.

Der Beweis erfolgt über einen ***Zirkelschluss***, denn wir können und wollen ja nicht beweisen, dass eine der Aussagen tatsächlich gilt (für manche A gelten sie, für manche nicht), sondern dass man jede Aussage aus jeder anderen folgern kann.

Beweis:

a) $\Rightarrow$ b): $A \cdot (A^{-1}\vec{b}) = (AA^{-1}) \cdot \vec{b} = \mathbb{1}_n\vec{b} \overset{\checkmark}{=} \vec{b}$, also gibt es eine Lösung. Sie ist eineutig, denn aus $\vec{x} \in \mathbb{K}^n$ mit $A\vec{x} = \vec{b}$ folgt $\vec{x} = A^{-1}A\vec{x} = A^{-1}\vec{b}$.

b) $\Rightarrow$ c): Spezialfall $\vec{b} := \vec{0}$.

c) $\Rightarrow$ d): Wir bringen A durch den Gauß-Jordan-Algorithmus auf eine reduzierte ZSF $A' \in M_n(\mathbb{K})$. A' hat keine Nichtpivotspalten, denn sonst gäbe es freie Variablen und $\mathrm{LR}(A; \vec{0})$ würde außer $\vec{0}$ noch weitere Lösungen enthalten (nämlich das Erzeugnis der Basislösungen).

Da A' also n Pivotspalten hat und eine reduzierte ZSF ist, gilt $\forall i \in \{1, ..., n\}$: $A'_{i,i} = 1$ und $\forall i \neq j : A'_{i,j} = 0$. Also ist $A' = \mathbb{1}_n$ und entstand aus A durch Zeilenoperationen.

d) $\Rightarrow$ e): Zeilenoperationen entsprechen Multiplikation mit den zugehörigen Elementarmatrizen. Es gibt also Elementarmatrizen $E_1, ..., E_m \in GL_n(\mathbb{K})$, so dass $E_m E_{m-1} \cdots E_1 A = \mathbb{1}_n$. Also $A = E_1^{-1} \cdots E_m^{-1}$, wobei die Inversen von Elementarmatrizen ebenfalls Elementarmatrizen sind.

d) $\Rightarrow$ a): Siehe Lemma 2.12. □

Der obige Beweis liefert im Prinzip ein Berechnungsverfahren: Man wendet den Gauß-Jordan-Algorithmus auf A an und multipliziert die dabei auftretenden Elementarmatrizen. Doch lässt sich das etwas effizienter machen, denn man kann das Produkt der Elementarmatrizen „nebenher“ berechnen:

[26] Es gelten also alle oder keine.

Problem 2.25 (Inverse Matrix)
Sei $\mathbb{K}$ ein Körper und $A \in M_n(\mathbb{K})$. Erkenne, ob $A \in GL_n(\mathbb{K})$, und berechne ggf. A^{-1}.

Lösung:
Wir starten mit der erweiterten Matrix $(A, \mathbb{1}_n)$, wenden darauf den Gauß-Jordan-Algorithmus an und erhalten eine reduzierte Zeilenstufenform (A', B) mit $A', B \in M_n(\mathbb{K})$. Jede Zeilenoperation entspricht der Multiplikation mit einer Elementarmatrix. Es gibt also ein $G \in GL_m(\mathbb{K})$, nämlich das Produkt dieser Elementarmatrizen, mit $(A', B) = G \cdot (A, \mathbb{1}_n)$. Man erkennt: $G = B$, das heißt, das Produkt der Elementarmatrizen ergibt sich automatisch aus der erweiterten Matrix.
Wenn $A' \neq \mathbb{1}_n$, dann ist A' nicht invertierbar. Andernfalls ist $G = B = A^{-1}$. □

Beispiel: Für $A := \left(\begin{smallmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{smallmatrix}\right) \in M_3(\mathbb{Q})$ ist $(A, \mathbb{1}_3) = \left(\begin{smallmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{smallmatrix}\right)$ mit Zeilenstufenform $\left(\begin{smallmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 & 1 & 1 \end{smallmatrix}\right)$. Man erkennt, dass A nicht invertierbar ist.

Für $A := \left(\begin{smallmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{smallmatrix}\right)$ ist $(A, \mathbb{1}_3) = \left(\begin{smallmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{smallmatrix}\right)$ ergibt Gauß-Jordan hingegen $\left(\begin{smallmatrix} 1 & 0 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{smallmatrix}\right)$. Also invertierbar, mit $A^{-1} = \left(\begin{smallmatrix} 1 & 0 & -1 \\ 0 & 0 & 1 \\ -1 & 1 & 1 \end{smallmatrix}\right)$.

3 Begriffe der Algebra

Im vorigen Kapitel haben wir an keiner Stelle auf eine konkrete Konstruktion von $\mathbb{Q}$ bzw. $\mathbb{R}$ Bezug genommen — kein Wunder, denn $\mathbb{R}$ wird erst in der Analysis konstruiert. All unsere Beweise (und damit auch die Rechentechniken) basierten auf abstrakten Rechenregeln wie Assoziativ- und Distributivgesetze. Aus fachwissenschaftlicher Sicht liegt es daher nahe, basierend auf diesen Rechengesetzen abstrakte Begriff zu definieren:

- Gruppen: Abstrahiert von $GL_n(\mathbb{K})$, aber auch von Permutationen.
- Ringe: Abstrahiert von $\mathbb{Z}$, $\mathbb{K}[X]$ und $M_n(\mathbb{K})$.
- Körper: Abstrahiert von $\mathbb{Q}$, $\mathbb{R}$. Wir werden weitere Körper konstruieren, die in der Informatik besonders relevant sind.
- Vektorräume: Abstrahiert von den Lösungsräumen homogener linearer Gleichungssysteme, aber auch von der Menge der stetigen Funktionen, und hat auch mit dem Übergang von $\mathbb{Q}$ nach $\mathbb{R}$ nach $\mathbb{C}$ zu tun.
- Lineare Abbildungen: Abstrahiert von der Matrixmultiplikation und von algebraischen Eigenschaften der Differential- und Integralrechnung.
- Basen: Abstrahiert von Koordinatensystemen sowie den Basislösungen homogener linearer Gleichungssysteme.

Aus Anwendungssicht kann man die abstrakten Begriffe als Werkzeugkästen betrachten: In ihnen sind jeweils Rechengesetze formuliert, und man sollte sich bei jedem Beweis und jedem Lösungsverfahren darüber im klaren sein, welche Rechengesetze man dafür benötigt, also welche Werkzeuge der Werkzeugkasten bereitstellen muss. Es gibt noch viel mehr algebraische Strukturen als ich hier vorstelle (z.B. Magmen, Halbgruppen, Monoide).

3.1 Gruppen

Die aus der Schule bekannten Rechenbereiche sind zunächst einmal *Mengen*, nämlich Mengen von Zahlen. Sowohl durch Addition als auch durch Multiplikation wird jedem Paar von Zahlen eine neue Zahl zugeordnet.

Definition 3.1
Eine **innere Verknüpfung** *auf einer Menge M ist eine Abbildung $M \times M \to M$.*

Anders als sonst für Abbildungen üblich wird nicht die Funktionenschreibweise $f(a, b)$ verwendet, sondern eines von vielen möglichen ***Verknüpfungssymbolen***, also zum Beispiel $a * b$, $a \otimes b$ etc.

Beispiel: Auf $\mathbb{N}$ ist durch $\forall a, b \in \mathbb{N}\colon\ a * b := 0$ eine (wenig interessante) innere Verknüpfung gegeben, auch $+$ und $\cdot$ sind innere Verküpfungen auf $\mathbb{N}$.

Beispiel: Matrixmultiplikation ist eine innere Verknüpfung auf $M_n(\mathbb{Q})$ sowie auf $GL_n(\mathbb{Q})$.

Nicht-Beispiel: $-$ ist auf $\mathbb{N}$ keine innere Verknüpfung, denn $0 - 1 \notin \mathbb{N}$. Hingegen ist $-$ eine innere Verknüpfung auf $\mathbb{Z}$. Ebenso ist Division $\div$ auf $\mathbb{Q}$ keine innere Verknüpfung, denn $1 \div 0$ ist nicht definiert.

Nicht-Beispiel: Matrixaddition ist keine innere Verknüpfung auf $GL_n(\mathbb{Q})$.

Definition 3.2
Eine **Gruppe** $(G, *, e)$ *(oder kurz: G, wenn $*$ und e aus dem Kontext klar sind) ist eine Menge G mit einer inneren Verknüpfung $*$ und einem $e \in G$, so dass die folgenden* **Gruppenaxiome** *gelten:*

a) $\forall a, b, c \in G\colon\ (a * b) * c = a * (b * c)$ *(***Assoziativgesetz***)*

b) $\forall a \in G\colon\ a * e = a$

c) $\forall a \in G\colon\ \exists b \in G\colon\ a * b = e$

Die Gruppe heißt **abelsch**[27]*, falls zudem auch das* **Kommutativgesetz** $\forall a, b \in G\colon\ a * b = b * a$ *gilt.*

Für allgemeine Gruppen verwendet man meist ein Multiplikationssymbol wie $\cdot, *, \odot, \otimes$; Additionssymbole wie $+, \oplus, \#$ sollte man nur für abelsche Gruppen verwenden.

Beispiele: $(\mathbb{Z}, +, 0)$ und $(\mathbb{R}^*, \cdot, 1)$ sind abelsche Gruppen, $(GL_n(\mathbb{Q}), \cdot, \mathbb{1}_n)$ ist für $n \geq 2$ eine nicht-abelsche Gruppe.

Beispiel: $(\{e\}, *)$ mit $e * e := e$ ist eine abelsche Gruppe, die ***triviale Gruppe***.

Nicht-Beispiele: $(\mathbb{N}, +, 0)$, $(\mathbb{Z}, \cdot, 1)$, $(M_n(\mathbb{Q}), \cdot, \mathbb{1}_n)$.

Bei invertierbaren Matrizen forderten wir ursprünglih, dass die inverse Matrix sowohl von rechts als auch von links invers ist. In der Gruppendefinition fordern wir das nicht — der Grund ist, dass dies aus den Gruppenaxiomen bereits folgt:

Lemma 3.3
Sei $(G, *, e)$ eine Gruppe.

a) Rechtsinverse sind linksinvers: $\forall a, b \in G$: Wenn $a * b = e$, dann $b * a = e$.

b) e ist auch linksneutral: $\forall a \in G\colon\ e * a = a$.

[27] Niels Henrik Abel [1802–1829]

c) $\forall a, b \in G\colon\ a * b = b \Rightarrow a = e$. *Also ist* e *eindeutig bestimmt und man nennt es das* **neutrale Element** *der Gruppe.*

d) Inverse sind eindeutig: $\forall a, b, c \in G\colon\ a * b = e = a * c \ \Rightarrow\ b = c$. *Notation: Für* $a \in G$ *bezeichnet* $a^{-1} \in G$ *das durch* $a * a^{-1} = e$ *eindeutig bestimmte* **Inverse** *von* a.

e) Inversion von Produkten: $\forall a, b \in G\colon\ (a * b)^{-1} = b^{-1} * a^{-1}$.

f) Doppelte Inversion: $\forall a \in G\colon\ (a^{-1})^{-1} = a$.

Beweis:

a) $\exists c \in G\colon\ b * c = e$. Dann $b * a = (b * a) * e = b * (a * e) = b * (a * (b * c)) = b * ((a * b) * c) = b * (e * c) = (b * e) * c = b * c = e$.

b) Nach a) $\exists b \in G\colon\ a * b = b * a = e$. $e * a = (a * b) * a = a * (b * a) = a * e = a$.

c) $\exists c \in G\colon\ b * c = e$. Es folgt $a * b = b \ \Rightarrow\ a * b * c = b * c \ \Rightarrow\ a * e = e$.

d) Nach a) $\exists d \in G\colon\ a * d = d * a = e$. $a * b = a * c \ \Rightarrow\ d * (a * b) = d * (a * c) \ \Rightarrow\ e * b = e * c \ \Rightarrow\ b = c$.

e) $(a * b) * (b^{-1} * a^{-1}) = a * (b * (b^{-1} * a^{-1})) = a * ((b * b^{-1}) * a^{-1}) = a * (e * a^{-1}) = a * a^{-1} = e$. Wegen d) folgt die Behauptung.

f) Nach a) gilt $a^{-1} * a = e$, also ist wegen d) auch $a = (a^{-1})^{-1}$. □

Beispiel: Die Symmetrien geometrischer Körper (zum Beispiel regulärer n-Ecke) bilden Gruppen. Eine Symmetrie ist eine längenerhaltende Abbildung, die den Körper als Menge auf sich selbst abbildet (also etwa eine Spiegelung an einer Symmetrieachse oder eine Drehung um ein Symmetriezentrum), und man sieht leicht, dass die Hintereinanderausführung zweier Symmetrieabbildungen wieder eine Symmetrieabbildung ist.

3.2 Ringe und Körper

In einer Gruppe hat man nur eine Verknüpfung, also nur ein Werkzeug. Es erleichtert die Arbeit sehr, wenn man zwei Werkzeuge hat: Addition und Multiplikation.

Definition 3.4

a) $(R, +, \cdot, 0, 1)$ *(oder kurz* R*, falls* $+$ *und* $\cdot$ *aus dem Kontext klar sind) heißt* **Ring**, *falls* $(R, +, 0)$ *eine abelsche Gruppe und* $\cdot$ *eine innere Verknüpfung auf* R *ist und zudem gilt:*

i) $1 \in R$ *und* $\forall a \in R\colon\ a \cdot 1 = 1 \cdot a = a$ *(***Neutralität** *der* 1*)*

ii) $\forall a, b, c \in R\colon\ (a \cdot b) \cdot c = a \cdot (b \cdot c)$ *(***Assoziativgesetz** *der Multiplikation)*

iii) $\forall a, b, c \in R\colon\ (a+b)\cdot c = a\cdot c + b\cdot c$ *und* $a\cdot(b+c) = a\cdot b + a\cdot c$ *(die beiden* **Distributivgesetze***).*

Man nennt 0 *das* **Nullelement** *und* 1 *das* **Einselement** *des Rings.*

b) Ein Ring $(R, +, \cdot, 0, 1)$ *heißt* **kommutativer Ring***, falls das* **Kommutativgesetz** *der Multiplikation erfüllt ist, also* $\forall a, b \in R\colon\ a\cdot b = b\cdot a$*.*

c) Ein Ring $(R, +, \cdot, 0, 1)$ *heißt* **Divisionsring** *oder* **Schiefkörper***, falls* $(R\setminus\{0\},\ \cdot, 1)$ *eine Gruppe ist, und heißt* **Körper***, falls* $(R\setminus\{0\},\ \cdot, 1)$ *sogar eine abelsche Gruppe ist. Insbesondere gilt dann* $0 \neq 1$*.*

Die Gruppenaxiome für $(R, +, 0)$ *zusammen mit der Neutralität der* 1*, dem Assoziativgesetz der Multiplikation und den Distributivgesetzen heißen* **Ringaxiome***. Die Ringaxiome zusammen mit dem Kommutativgesetz der Multiplikation und der Invertibilität der von* 0 *verschiedenen Elemente heißen* **Körperaxiome***. Notation: Ist* R *ein Ring und* $a, b \in R$*, so sei* $-b \in R$ *das additive Inverse von* b *(also* $b + (-b) = 0$*) und* $a - b := a + (-b)$*.*

Beispiele: a) $\{0\}$ mit $0+0 = 0\cdot 0 = 0$ ist ein Ring, der ***triviale Ring*** (auch: ***Nullring***). Hier ist 0 gleichzeitig Null- und Einselement, $\{0\}$ ist also kein Körper.

b) $\mathbb{Z}$ ebenso wie die Menge $\mathbb{R}[X]$ aller Polynome sind bezüglich der gewohnten Addition und Multiplikation kommutative Ringe, jedoch keine Körper.

c) $\mathbb{R}$ und $\mathbb{Q}$ sind Körper.

d) $\mathbb{F}_2 := \{0, 1\}$ mit folgenden inneren Verknüpfungen ist ein Körper:

$+$	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

e) Da Computer zur Darstellung reeller Zahlen nur eine begrenzte Zahl von Stellen zur Verfügung stellen, ist es in vielen Anwendungen unvermeidbar, gerundet zu rechnen. Dabei gelten Distributiv- und Assoziativgesetze leider nicht, und nicht jede von Null verschiedene Zahl hat ein Inverses!

f) Die komplexen Zahlen bilden einen Körper — diesen werden wir bald näher betrachten.

g) Es gibt Varianten des Ringbegriffs, die wir hier aber nicht näher betrachten (höchstens in einigen Übungsaufgaben), beispielsweise Ringe ohne 1. So etwas bezeichnen wir als ***Rng*** (kein Schreibfehler). In manchen Büchern wird für einen Ring tatsächlich nicht die Existenz von $1 \in R$ gefordert; dort würde man von einem ***unitären Ring*** reden, wenn es eben doch ein Einselement gibt. Beispiel: $(\{2n \mid n \in \mathbb{Z}\}, +, \cdot, 0)$ ist ein Rng.

h) Wir definieren auf beliebigen Mengen eine innere Verknüpfung $*_r$ durch $\forall a, b\colon\ a *_r b := b$. Dann ist in $(\mathbb{Z}, +, *_r, 0)$ jedes Elemente eine „Links-1“, aber keines ist eine „Rechts-1“. Auch gilt nur eines der beiden Distributivgesetze, nämlich $\forall a, b, c \in \mathbb{Z}\colon\ a *_r (b + c) = b + c = a *_r b + a *_r c$. Das andere Distributivgesetz ist verletzt, etwa $(1 + 1) *_r 1 = 1 \neq 1 *_r 1 + 1 *_r 1 = 2$.

In Kapitel 2 war $\mathbb{K} \in \{\mathbb{Q}, \mathbb{R}\}$, aber tatsächliche hätte immer ein Körper und in den meisten Fällen sogar ein kommutativer Ring genügt. Ich habe entsprechende Anpassungen in roter Schrift vorgenommen. Auch beim Umgang mit Ringen hat man es bisweilen mit Gruppen zu tun.

Definition 3.5 (und Notation). *Sei R ein Ring.*

a) $R^* := R \setminus \{0\}$.

b) $a \in R$ *heißt* **invertierbar** *oder eine* **Einheit** $:\Leftrightarrow \exists b \in R\colon\ a \cdot b = b \cdot a = 1$.

c) $R^\times := \{a \in R \mid a \text{ invertierbar}\}$ *heißt die* **Einheitengruppe** *von R.*

Beobachtung 3.6
Für jeden Ring R ist $R^\times$ eine Gruppe. Es ist nämlich $1 \in R^\times$, denn $1 \cdot 1 = 1$, und zudem ist $\forall a, b \in R^\times\colon\ (ab) \cdot (b^{-1}a^{-1}) = (b^{-1}a^{-1}) \cdot (ab) = 1$, d.h. $ab \in R^\times$: Die Multiplikation in $R^\times$ ist eine innere Verknüpfung. Zudem ist $\cdot$ assoziativ, nicht nur in $R^\times$, sondern in ganz R.

Beispiele: a) Ist R eine Divisionsalgebra, dann $R^\times = R^*$.

b) Sei $\mathbb{K}$ ein Körper. Dann ist $M_n(\mathbb{K})^\times = GL_n(\mathbb{K})$.

Satz 3.7. *Sei R ein Ring.*

a) $\forall x \in R\colon\ 0 \cdot x = 0$. *b)* $\forall x \in R\colon\ (-1) \cdot x = -x$.

Beweis:
a) $0 + 0 \cdot x \overset{\text{Neutr.}}{=} 0 \cdot x \overset{\text{Neutr.}}{=} (0 + 0) \cdot x \overset{\text{Distr.}}{=} 0 \cdot x + 0 \cdot x \overset{\text{Kürzung}}{\Rightarrow} 0 = 0 \cdot x$.

b) $x + (-1) \cdot x \overset{\text{Neutr.}}{=} 1 \cdot x + (-1) \cdot x \overset{\text{Distr.}}{=} (1 + (-1)) \cdot x \overset{\text{Negat.}}{=} 0 \cdot x \overset{\text{a)}}{=} 0 = x + (-x) \overset{\text{Kürzung}}{\Rightarrow} (-1) \cdot x = -x$. □

Beispiel: Es folgt die bekannte Regel $(-1) \cdot (-1) = -(-1) = 1$.

3.3 Restklassenringe

Wir kommen nun zu einer reichhaltigen Klasse von Ringen, die besonders auch in der Informatik relevant sind. Dazu beginnen wir mit dem Begriff der Relation.

Definition 3.8
Sei X eine Menge.[28] *Eine (binäre bzw. zweistellige)* **Relation** *auf X ist eine Teilmenge $R \subset X \times X$. Für $x, y \in X$ schreibt man meist nicht $(x, y) \in R$, sondern verwendet dafür ein Relationssymbol. Wir verwenden hier $\sim$. Es ist also $R = \{(x, y) \in X \times X \mid x \sim y\}$. Eine Relation $\sim$ auf X heißt*

- reflexiv, *falls* $\forall x \in X\colon\ x \sim x$.
- symmetrisch, *falls* $\forall x, y \in X\colon\ x \sim y \iff y \sim x$.
- transitiv, *falls* $\forall x, y, z \in X$ *gilt: Wenn* $x \sim y$ *und* $y \sim z$, *dann* $x \sim z$.

Eine reflexive symmetrische transitive Relation heißt **Äquivalenzrelation**.

Beispiele:

a) $X =$ alle Städte Deutschlands: Relation „liegt im gleichen Bundesland wie“ ist eine Äquivalenzrelation.

b) Auf $\mathbb{R}$ ist die Relation $<$ weder reflexiv noch symmetrisch, aber sie ist transitiv; es handelt sich um eine ***Ordnungsrelation***, die ich hier zunächst nicht weiter thematisieren möchte.

Im Rest des Abschnitts sei R ein kommutativer Ring.

Definition 3.9
$\forall a, b \in R\colon\ a|b :\Leftrightarrow \exists c \in R\colon\ b = a \cdot c$ *(d.h. „a* **teilt** *b“).*
Sei $n \in R^*$. $\forall a, b \in R\colon\ a \equiv_n b :\Leftrightarrow n|(b - a)$ (d.h. a ist **kongruent** *zu b* **modulo** *n). Andere Schreibweise: $a \equiv b \pmod{n}$.*

Bemerkung In einer Präsenzübung lernten Sie eine auf Division mit Rest basierende Definition von Kongruenz kennen. In $\mathbb{Z}$ sind beide Definitionen gleichbedeutend, doch in allgemeinen kommutativen Ringen gibt es keine Division mit Rest. Daher ist Definition 3.9 die Definition, mit der wir ab jetzt arbeiten werden.

Lemma 3.10
$\forall n \in R^*\colon$ *„Kongruenz modulo n“ ist eine Äquivalenzrelation auf R.*

Beweis:
Reflexivität: $\forall a \in R\colon\ (a - a) = 0 = n \cdot 0$, also $n|(a - a)$, also $a \equiv_n a$.

[28] Mit echten Klassen geht das im Prinzip auch.

Symmetrie: Sei $a, b \in R$. Für $q \in R$ ist $b - a = q \cdot n$ genau dann wenn $a - b = (-q) \cdot n$. Folglich $n|(b-a) \iff n|(a-b)$, und das heißt $a \equiv_n b \iff b \equiv_n a$.

Transitivität: Seien $a, b, c \in R$ mit $a \equiv_n b$ und $b \equiv_n c$. Dann $\exists q_1, q_2 \in R$: $(b - a) = q_1 \cdot n$ und $(c-b) = q_2 \cdot n$. Daher $(c-a) = (c-b)+(b-a) = q_2 \cdot n + q_1 \cdot n = (q_2 + q_1) \cdot n$, also $a \equiv_n c$. □

Definition 3.11
Ist $\sim$ eine Äquivalenzrelation auf der Menge X und ist $y \in X$, so heißt $[y] := [y]_\sim := \{x \in X \mid x \sim y\}$ die **Äquivalenzklasse** *von y. Ist K eine Äquivalenzklasse und $x \in K$, so heißt x ein* **Repräsentant** *von K.*

Lemma 3.12
Sei $\sim$ eine Äquivalenzrelation auf der Menge X. Wegen Reflexivität gilt $\forall x \in X\colon\ x \in [x]$. Außerdem sind für $x, y \in X$ folgende drei Aussagen gleichbedeutend:

a) $x \sim y$ *b)* $[x] \cap [y] \neq \emptyset$. *c)* $[x] = [y]$

Beweis:
Wir zeigen a) ⇒ b) ⇒ c) ⇒ a).

a) ⇒ b) Aus $x \sim y$ folgt $x \in [y]$. Wegen Reflexivität ist aber auch $x \in [x]$. Daher ist $x \in [x] \cap [y]$, also $[x] \cap [y] \neq \emptyset$.

b) ⇒ c) Wenn $[x] \cap [y] \neq \emptyset$, dann gibt es ein $z \in [x] \cap [y]$. Wir wollen nun zeigen, dass $[x] \subseteq [y]$; wenn also $w \in [x]$, so wollen wir zeigen, dass auch $w \in [y]$. Aus $w \in [x]$ folgt $w \sim x$. Wegen $z \in [x]$ ist $z \sim x$, also $x \sim z$ wegen Symmetrie. Aus $w \sim x$ und $x \sim z$ folgt $w \sim z$ wegen Transitivität. Aufgrund von $z \in [y]$ gilt $z \sim y$, also aus $w \sim z$ folgt $w \sim y$ wegen Transitivität, also $w \in [y]$, was zu zeigen war.

Analog folgt $[y] \subseteq [x]$, also $[y] = [x]$, also c).

c) ⇒ a) Wenn $[x] = [y]$, dann $x \in [x] = [y]$, daher $x \sim y$. □

Auch wenn Ihnen das nicht bewusst ist: Bereits seit früher Kindheit sind Sie es gewohnt, in Äquivalenzklassen zu denken. Zum Beispiel haben Sie verstanden, dass eine Menge von 5 Spielzeugautos und eine Menge von 5 Gummibärchen etwas gemeinsam haben: Sie haben die Zahl 5 als Äquivalenzklasse der Relation „gleichmächtig“ entdeckt, und diese Äquivalenzklasse wird unter anderem durch eine Menge von 5 Gummibärchen repräsentiert.

Später lernten Sie, mit diesen Äquivalenzklassen zu rechnen. Und das entscheidende ist: Wenn man „$5 + 3$“ ausrechnen möchte, gelangt man zum gleichen Ergebnis, egal welche Repräsentanten man für 5 bzw. 3 wählt: Man kann mit Gummibärchen, Fingern oder Taschenrechnern rechnen („5 Taschenrechner plus 3 Taschenrechner sind 8 Taschenrechner“).

Wir wollen nun auch mit Äquivalenzklassen der Kongruenz modulo n rechnen. Auch hierbei darf das Ergebnis einer mit zwei Äquivalenzklassen durchgeführten Rechenoperation nicht davon abhängen, welche Repräsentanten man wählt. Man sagt dazu, dass die Rechenoperationen ***wohldefiniert*** sind.

Satz 3.13 (und Definition)
Sei R ein kommutativer Ring und $n \in R^*$. Wir betrachten die durch $\forall x, y \in R$ durch $x \equiv_n y$ gegebene Äquivalenzrelation und definieren ${}^R/_{nR} := \{[x] \mid x \in R\}$, wobei man $[x]$ in diesem Zusammenhang als **Restklasse** *von x modulo n bezeichnet.*

a) Für $x, y \in R$ sind durch $[x]+[y] := [x+y]$ und $[x] \cdot [y] := [x \cdot y]$ zwei innere Verknüpfungen auf ${}^R/_{nR}$ definiert.

b) Durch diese beiden inneren Verknüpfungen wird ${}^R/_{nR}$ zu einem kommutativen Ring, dem so genannten **Restklassenring** *von R modulo n.*

Beispiel: Um Ihnen zu verdeutlichen, wie wenig selbstverständlich Wohldefiniertheit ist, betrachten wir eine andere Rechenoperation, nämlich das Potenzieren. Seien $A, B \in {}^{\mathbb{Z}}/_{5\mathbb{Z}}$ und seien $a \in A$ und $b \in B$ Repräsentanten. Dann ist $A^B := [a^b]$ nicht wohldefiniert, denn 2 und 7 sind zwei Repräsentanten der gleichen Restklasse $[2]$, aber $[2^2] = [4] \neq [2^7] = [128] = [3]$.

Beweis:

a) Wir weisen die Wohldefiniertheit nach. Seien $x', y' \in R$ mit $[x] = [x']$ und $[y] = [y']$. Nach dem vorigen Lemma[29] ist das gleichbedeutend zu $x \equiv_n x'$ und $y \equiv_n y'$. Nach Definition von Kongruenz modulo n und Teilbarkeit ist das gleichbedeutend zu $\exists q_x, q_y \in R\colon\ x' = x + q_x \cdot n$ und $y' = y + q_y \cdot n$.

   Behauptung: $[x+y] = [x'+y']$ und $[x \cdot y] = [x' \cdot y']$.

   - $x'+y' = x+q_x \cdot n+y+q_y \cdot n = (x+y)+(q_x+q_y) \cdot n$, also $x+y \equiv_n x'+y'$, was zu zeigen war.
   - $x' \cdot y' = (x + q_x \cdot n) \cdot (y + q_y \cdot n) = x \cdot y + (x \cdot q_y + q_x \cdot y + q_x \cdot q_y \cdot n) \cdot n$, also $x \cdot y \equiv_n x' \cdot y'$, was zu zeigen war.

b) Die kommutativen Ringaxiome lassen sich für Äquivalenzklassen leicht durch Wahl von Repräsentanten nachweisen, da die entsprechenden Axiome in R gelten. Wir führen dies nur am Beispiel des Distributivgesetzes vor, den Rest kann man sich als Übung überlegen: Seien $X, Y, Z \in {}^R/_{nR}$ mit Repräsentanten x, y, z. Wir haben also $x, y, z \in R$ mit $[x] = X$, $[y] = Y$ und $[z] = Z$. Es gilt $X \cdot (Y+Z) = [x] \cdot ([y]+[z]) \overset{\text{Def}}{=} [x] \cdot [y+z] \overset{\text{Def}}{=} [x \cdot (y+z)] \overset{\text{Distr}}{=} [x \cdot y + x \cdot z] \overset{\text{Def}}{=} [x \cdot y] + [x \cdot z] \overset{\text{Def}}{=} [x] \cdot [y] + [x] \cdot [z]$. □

[29] Beachte: Die hier betrachtete Äquivalenzrelation wird nicht als $x \sim y$, sondern als $x \equiv_n y$ notiert.

Beispiel 3.14

a) Wenn Sie ausrechnen wollen, welcher Wochentag in 23 *Tagen ist, dann rechnen Sie in* $\mathbb{Z}/7\mathbb{Z}$*:* $[23] = [2]$*, wenn also heute Montag ist, dann ist in* 23 *Tagen zwei mehr als Montag, also Mittwoch.*

b) Sei $x \in \mathbb{N}$ *mit Dezimalziffern* $z_k, z_{k-1}, ..., z_0 \in \{0, ..., 9\}$*. Möglicherweise*[30] *kennen Sie aus der Schule folgende Teilbarkeitsregel:* x *ist durch* 11 *teilbar genau dann, wenn seine* **alternierende Quersumme** $z_0 - z_1 + z_2 - ... \pm z_k$ *durch* 11 *teilbar ist. Dies beweist man durch eine kleine Rechnung in* $\mathbb{Z}/11\mathbb{Z}$*: Wir haben* $x = \sum_{i=0}^{k} z_i \cdot 10^i$*. Also ist*

$$[x] = \left[\sum_{i=0}^{k} z_i \cdot 10^i\right] = \sum_{i=0}^{k} [z_i] \cdot [10]^i = \sum_{i=0}^{k} [z_i] \cdot [-1]^i \qquad \text{denn } 10 \equiv_{11} -1$$
$$= \left[\sum_{i=0}^{k} (-1)^i z_i\right] .$$

Das bedeutet: Jede Zahl hat bei Division durch 11 *den gleichen Rest wie ihre alternierende Quersumme.*

Beispiel: 7129 hat die alternierende Quersumme $9 - 2 + 1 - 7 = 1$, also lässst 7129 bei Division durch 11 den Rest 1.

Bemerkung 3.15
Sei $n \in \mathbb{N}^*$*.* $\mathbb{Z}/n\mathbb{Z}$ *ist genau dann ein Körper, wenn* n *eine Primzahl ist.*

Ein Teil dieser Aussage wird in den Übungen bewiesen. Übrigens: $n \in \mathbb{N}^*$ heißt ***Primzahl*** $:\Leftrightarrow n \nmid 1$ und $\forall a, b \in \mathbb{Z}\colon\ n|a \cdot b \ \Rightarrow\ (n|a) \vee (n|b)$. Der Primzahlbegriff, den Sie vermutlich in der Schule lernten („n heißt prim genau dann wenn n genau zwei Teiler hat“ oder „$n > 1$ heißt prim genau dann wenn n sich nicht als Produkt zweier von 1 verschiedener natürlicher Zahlen schreiben lässt“) heißt in der Algebra nicht „prim“ sondern ***irreduzibel***. Das ist in $\mathbb{Z}$ dasselbe, aber in manchen Ringen ist das ein Unterschied.

3.4 Komplexe Zahlen

In diesem Abschnitt geht es um ein Zahlbereichserweiterung, die die reellen Zahlen und zudem die Lösungen beliebiger algebraischer Gleichungen umfasst. Schon im 16. Jhdt. war es ein etablierter Rechentrick, mit Wurzeln aus negativen Zahlen zu rechnen. Auch in der Physik und in der Elektrotechnik ist dies sehr nützlich. Bis vor 30 Jahren war das normaler Schulstoff, schließt sich unmittelbar an das Thema „quadratische Gleichungen“ an und besitzt sogar eine ähnlich schöne Veranschaulichung wie das Rechnen im Zahlenstrahl.

[30] Im Lehrplan steht sie glaube ich nicht mehr.

Die komplexen Zahlen entstehen aus $\mathbb{R}$, indem man ein Symbol i einführt (***imaginäre Einheit***, nach Norm DIN 1302). In der Elektrotechnik darf auch j als Symbol verwendet werden; fieserweise sind i und i verschiedene Symbole). Mit i rechnet man wie mit der Unbekannten von Polynomen, mit einem wichtigen Unterschied: Zusätzlich hat man $\mathrm{i}^2 = -1$.

Beispiel: Es folgt $\mathrm{i}^3 = \mathrm{i} \cdot \mathrm{i}^2 = -\mathrm{i}$ und $\mathrm{i}^4 = (\mathrm{i}^2)^2 = (-1)^2 = 1$.

Auf diese Weise kann man alle höheren Potenzen von i beseitigen. Man erhält:

Definition 3.16
$\mathbb{C} := \{a + b\,\mathrm{i} \mid a, b \in \mathbb{R}\}$ *ist die Menge der* **komplexen Zahlen**. *Der* **Realteil** *von* $z = a + b\,\mathrm{i} \in \mathbb{C}$ *ist* $\mathrm{Re}(z) := a \in \mathbb{R}$, *der* **Imaginärteil** *ist* $\mathrm{Im}(z) := b \in \mathbb{R}$.

Beispiel 3.17 (und Definition)
In komplexen Zahlen kann man auch Wurzeln aus negativen Zahlen ziehen. Die Gleichung $x^2 = -1$ *hat nämlich in* $\mathbb{C}$ *die beiden Lösungen* $x_{1,2} = \pm\,\mathrm{i}$. *Man erweitert daher die Definition der Wurzelfunktion: Für* $a \in \mathbb{R}_{>0}$ *setzt man* $\sqrt{-a} := \mathrm{i}\sqrt{a}$. *Das ist sinnvoll, denn* $(\mathrm{i}\sqrt{a})^2 = \mathrm{i}^2(\sqrt{a})^2 = (-1) \cdot a = -a$.

Rechenregeln für komplexe Zahlen
$\mathbb{C}$ *wird zu einem Körper, indem man für* $z_1 = a_1 + b_1\,\mathrm{i} \in \mathbb{C}$ *und* $z_2 = a_2 + b_2\,\mathrm{i} \in \mathbb{C}$ *mit* $a_1, a_2, b_1, b_2 \in \mathbb{R}$ *wie folgt rechnet:*

- $z_1 + z_2 = (a_1 + a_2) + (b_1 + b_2)\,\mathrm{i}$ *sowie* $z_1 - z_2 = (a_1 - a_2) + (b_1 - b_2)\,\mathrm{i}$
- $z_1 \cdot z_2 = (a_1 a_2 - b_1 b_2) + (a_1 b_2 + a_2 b_1)\,\mathrm{i}$
- *Falls* $a_2 + b_2\,\mathrm{i} \neq 0$: $\dfrac{z_1}{z_2} = \dfrac{a_1 a_2 + b_1 b_2}{a_2^2 + b_2^2} + \dfrac{a_2 b_1 - a_1 b_2}{a_2^2 + b_2^2}\,\mathrm{i}$.

Beweis: Übung. □

Definition 3.18
Für $z = a + b\,\mathrm{i} \in \mathbb{C}$ *mit* $a, b \in \mathbb{R}$ *ist* $\bar{z} := a - b\,\mathrm{i} \in \mathbb{C}$ *die* **konjugiert komplexe Zahl**; *in der Physik schreibt man auch* z^* *statt* $\bar{z}$. *Ferner ist* $|z| := \sqrt{z \cdot \bar{z}} = \sqrt{a^2 + b^2} \in \mathbb{R}_{\geq 0}$ *der* **Betrag** *von* z.

Beobachtung 3.19
Weil stets $a^2 + b^2 \in \mathbb{R}_{\geq 0}$ *gilt, ist* $|z| \in \mathbb{R}_{\geq 0}$. *Ferner ist* $|z| = 0 \iff z = 0$. *Die Divisionsregel kann man sich wie folgt merken: Für* $z_1, z_2 \in \mathbb{C}$ *und* $z_2 \neq 0$ *gilt*

$$\frac{z_1}{z_2} = \frac{z_1 \overline{z_2}}{z_2 \overline{z_2}} = \frac{z_1 \overline{z_2}}{|z_2|^2}\,.$$

Bemerkung 3.20. *Es gibt mindestens zwei Wege,* $\mathbb{C}$ *als Ring zu konstruieren:*

a) Sei $R := \mathbb{R}[X]$ und $d := X^2 + 1$. Dann kann man $\mathbb{C} := {}^{R}/_{dR}$ definieren. Die Restklassen bei Division durch d sind nämlich durch Elemente der Form $a + bX$ mit $a, b \in \mathbb{R}$ repräsentiert. Zudem ist $X^2 \equiv_d -1$.

b) Man kann $\mathbb{C}$ auch durch Multiplikation gewisser (2×2)-Matrizen beschreiben. So haben Sie es in den Übungen bewiesen.

Der folgende wichtige Satz lässt sich auf unterschiedlichen Wegen beweisen, doch dabei müssten Anleihen bei der Analysis in einem Umfang gemacht werden, der den Rahmen einer Anfängervorlesung sprengen würde.

Fundamentalsatz der Algebra
*Jedes Polynom vom Grad ≥ 1 mit Koeffizienten aus $\mathbb{C}$ hat mindestens eine Nullstelle in $\mathbb{C}$. Es folgt: Für jedes $p \in \mathbb{C}[X]$ mit $\deg(p) = n$ gibt es $c, \lambda_1, ..., \lambda_n \in \mathbb{C}$, so dass $p(X) = c \cdot \prod_{k=1}^{n} (X - \lambda_k)$ (**Linearfaktorzerlegung**).* □

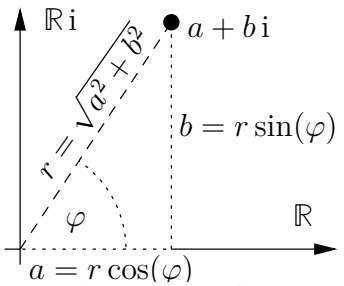
Ein zentrales Ergebnis der Galois-Theorie ist, dass es für $n = \deg(p) \geq 5$ beweisbar unmöglich ist, die Nullstellen $\lambda_1, ..., \lambda_n$ mit einer allgemeinen Lösungsformel (wie für quadratische Gleichungen) zu finden.

3.4.1 Die Gaußsche Zahlenebene

Bereits in der Mathematik des 16. oder 17. Jhdts. benutzte man komplexe Zahlen. Eine Veranschaulichung von $\mathbb{C}$ als Zahlenebene etablierte sich erst im 19. Jhdt.

Definition 3.21
$\mathbb{C}$ mit der Ebene $\mathbb{R}^2$ identifiziert, indem man $a + b\,\mathrm{i} \in \mathbb{C}$ als $(a, b) \in \mathbb{R}^2$ darstellt. Dies heißt **Gaußsche Zahlenebene**[31] *oder* **Arganddiagramm**[32]*, wurde aber zuerst 1797 von Caspar Wessel [1745–1818] beschrieben.*

Darstellungsarten komplexer Zahlen
Sei $z = a + b\,\mathrm{i} \in \mathbb{C}$. Seien (r, φ) die Polarkoordinaten (Bogenmaß) des Punkts $(a, b) \in \mathbb{R}^2$, also $r \geq 0$ und $a = r\cos(\varphi)$, $b = r\sin(\varphi)$. Dann $z = a + b\,\mathrm{i} = r(\cos(\varphi) + \mathrm{i}\sin(\varphi))$. Nach dem Satz des Pythagoras ist $r^2 = a^2 + b^2$, also $r = |z|$.

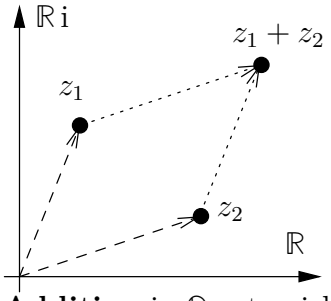


Man nennt φ das ***Argument*** $\arg(z)$ von z; es gibt verschiedene Konventionen, ob $\varphi \in]-\pi, \pi]$ oder $\varphi \in [0, 2\pi[$ gelten soll. Die Darstellung $z = r \cdot (\cos\varphi + \mathrm{i}\sin\varphi)$ heißt ***Polardarstellung*** oder ***trigonometrische Darstellung***, die Darstellung $z = a + b\,\mathrm{i}$ heißt ***Standarddarstellung*** oder ***kartesische Darstellung***.

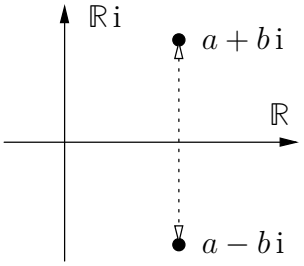
[31] Carl Friedrich Gauß [1777–1855] beschrieb sie in einem Brief von 1811.
[32] Jean–Robert Argand [1768–1822] beschrieb sie schon 1806

Übrigens: Oft lasse ich bei trigonometrischen Funktionen oder Logarithmus die Klammern weg, also $\sin\varphi$ statt $\sin(\varphi)$ oder $\ln\pi$ statt $\ln(\pi)$.

Veranschaulichung von Addition und Konjugation



Addition in $\mathbb{C}$ entspricht Vektoraddition in $\mathbb{R}^2$ (Kräfteparallelogramm).



Komplexe Konjugation ist Spiegelung an der reellen Achse.

Die ***Dreiecksungleichung*** besagt: Für $z_1, z_2 \in \mathbb{C}$ gilt $|z_1 + z_2| \leq |z_1| + |z_2|$. In der Mathematik müsste man solche Dinge eigentlich beweisen, doch hier berufen wir uns auf die Anschauung (siehe linkes Bild).

Rechnen in Polarkoordinaten

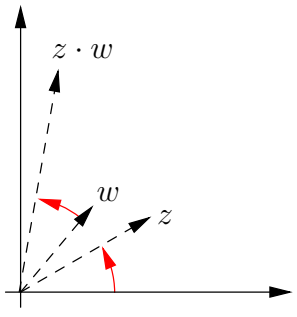
Addition und Konjugation sind in Standarddarstellung leicht, Multiplikation oder gar Wurzelziehen hingegen schwer. Zwar ist Addition in Polarkoordinaten schwer, aber Konjugation ist leicht: Weil Konjugation die Spiegelung an der reellen Achse ist, gilt $|\bar{z}| = |z|$ und $\arg(\bar{z}) = -\arg(z)$ für alle $z \in \mathbb{C}$.

Auch Multiplikation, Division und sogar Wurzelziehen sind in Polardarstellung ziemlich leicht. Denn für $z = r(\cos\varphi + \mathrm{i}\sin\varphi)$ und $w = s(\cos\psi + \mathrm{i}\sin\psi)$, mit $r, s, \varphi, \psi \in \mathbb{R}$ ist nach den ***Additionstheoremen***[33] für Sinus und Kosinus, die wir hier nicht beweisen:

$$\begin{aligned} z \cdot w &= rs(\cos(\varphi) + \mathrm{i}\sin(\varphi))(\cos(\psi) + \mathrm{i}\sin(\psi)) \\ &= rs((\cos(\varphi)\cos(\psi) - \sin(\varphi)\sin(\psi)) + \mathrm{i}(\cos(\varphi)\sin(\psi) + \sin(\varphi)\cos(\psi))) \\ &= rs(\cos(\varphi + \psi) + \mathrm{i}\sin(\varphi + \psi)) \end{aligned}$$

Also $|zw| = |z|\,|w|$ und $\arg(zw) = \arg(z) + \arg(w)$. Falls $w \neq 0$, erhält man ebenso $\left|\frac{z}{w}\right| = \frac{|z|}{|w|}$ und $\arg(\frac{z}{w}) = \arg(z) - \arg(w)$.

Im nebenstehenden Bild ist $|z| = 2$ und $|w| = 1.5$, daher $|zw| = 3$, und der Winkel zwischen positiver reeller Achse und z ist wie zwischen w und zw.

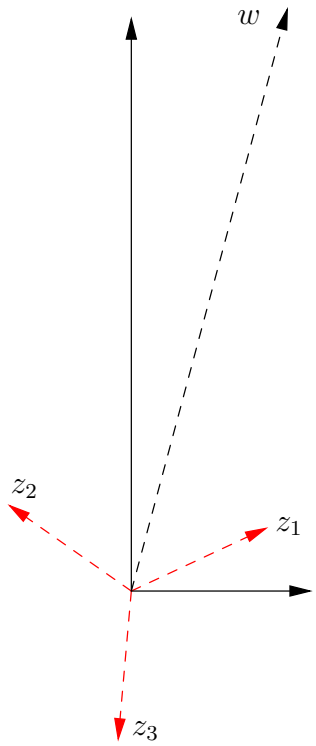


[33] In der Schule werden sie meist ausgelassen, obwohl sie elementargeometrisch behandelt werden können und für die Berechnung der Ableitung von Sinus und Kosinus (also $(\sin x)' = \cos x$ und $(\cos x)' = -\sin x$ — das ist noch Schulstoff!) gebraucht werden.

Beispiel 3.22 (Wurzelziehen)
Sei $\Phi := \{0, \frac{2\pi}{5}, \frac{4\pi}{5}, \frac{6\pi}{5}, \frac{8\pi}{5}\}$. *Dies sind alle Winkel, die mit* 5 *multipliziert ein Vielfaches von* 2π *ergeben. Daher gilt* $z^5 = 1$ *gdw.* $|z| = 1$ *und* $\arg(z) \in \Phi$, *oder anders formuliert:* $z = \cos(\varphi) + \mathrm{i}\sin(\varphi)$ *für* $\varphi \in \Phi$.
Ist nun $w \in \mathbb{C}$ *beliebig und* $n \in \mathbb{N}^*$, *so können wir alle Lösungen von* $z^n = w$ *(also die* n*-ten* **Wurzeln** *von* w*) finden: Es muss* $|z| = \sqrt[n]{|w|}$ *und* $n \cdot \arg(z) = \arg(w)$ *(bis auf Vielfache von* 2π*) gelten. Das bedeutet*

$$z = \sqrt[n]{|w|} \cdot \left(\cos\left(\frac{\arg(w)}{n} + \varphi\right) + \mathrm{i}\sin\left(\frac{\arg(w)}{n} + \varphi\right)\right)$$

mit $\varphi = \frac{2k\pi}{n}$ *und* $k \in \{0, ..., n-1\}$.
Im nebenstehenden Bild ist w *mit* $|w| = 8$ *und* $\arg(w) = 75°$. *Die Gleichung* $z^3 = w$ *hat die drei Lösungen* z_1, z_2, z_3 *mit* $|z_i| = 2$, $\arg(z_1) = 25°$, $\arg(z_2) = 25° + 120°$ *und* $\arg(z_3) = 25° + 240°$.



Man kann also aus jeder komplexen Zahl beliebige Wurzeln ziehen.

Umrechnungsformeln
Zwischen Standard- und Polardarstellung gelten folgende Umrechnungsformeln: Es ist $a + b\,\mathrm{i} = r(\cos\varphi + \mathrm{i}\sin\varphi)$ *genau dann wenn*

$$a = r \cdot \cos\varphi \qquad b = r \cdot \sin\varphi$$

$$r = \sqrt{a^2 + b^2} \qquad \varphi = \begin{cases} \arccos\left(\frac{a}{\sqrt{a^2+b^2}}\right) & \text{falls } b \geq 0 \\ -\arccos\left(\frac{a}{\sqrt{a^2+b^2}}\right) & \text{falls } b < 0 \end{cases}$$

Für das Argument gibt es auch Umrechnungsformeln, die auf dem Arkustangens basieren, aber eine Fallunterscheidung hat man immer.

Im Hinblick auf die Umrechnungsformeln ist folgende Tabelle praktisch, wobei man noch $\forall\varphi \in \mathbb{R}$: $\sin(-\varphi) = -\sin\varphi$, $\cos(-\varphi) = \cos\varphi$ sowie $\sin(\varphi + \frac{\pi}{2}) = \cos\varphi$ beachten muss.

Spezielle Werte von Sinus und Kosinus

φ	0	$\frac{\pi}{12}$	$\frac{\pi}{8}$	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{5\pi}{12}$	$\frac{\pi}{2}$
$\sin\varphi$	0	$\frac{\sqrt{6}-\sqrt{2}}{4}$	$\frac{\sqrt{2-\sqrt{2}}}{2}$	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{6}+\sqrt{2}}{4}$	1
$\cos\varphi$	1	$\frac{\sqrt{6}+\sqrt{2}}{4}$	$\frac{\sqrt{2+\sqrt{2}}}{2}$	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	$\frac{\sqrt{6}-\sqrt{2}}{4}$	0

3.5 Warum Restklassenkörper?

Lineare Algebra über Restklassenkörpern tritt beispielsweise in der ***Codierung*** auf, also der Darstellung von Daten als Symbolfolgen.[34] Die Symbole sollen Elemente eines endlichen Körpers $\mathbb{K}$ sein und die Symbolfolgen „Blöcke“ einer festen Länge n: Ein Block entspricht dann einem Element von $\mathbb{K}^n$.

Beispiel: Die 16 Elemente von $\mathbb{F}_2^4$ reichen zur Codierung einer Hexadezimalziffer.

Wir schreiben Blöcke hier als Zeilenvektoren, rechnen aber weiterhin mit Spaltenvektoren. Bei der Datenübertragung können Fehler passieren. Wir nehmen an, dass die Blocklänge erhalten bleibt (es werden keine Bits eingefügt oder entfernt), also durch den Fehler lediglich ein Block aus $\mathbb{K}^n$ durch einen anderen ersetzt wird, wobei pro Block höchstens r Stellen verändert werden.

Beispiel: Gesendet $(0,1,1,0)$, empfangen $(0,1,0,0) \rightsquigarrow r = 1$ Fehler.

Übertragungsfehler können nur erkannt oder gar automatisch korrigiert werden, wenn der Code redundant ist: Man wählt eine Teilmenge $C \subset \mathbb{K}^n$ von ***Codewörtern***. Gesendet wird $c \in C$, und wenn $c' \notin C$ empfangen wird, weiß man, dass es einen Übertragungsfehler gab. Und wenn es genau ein Codewort c'' gibt, das sich von c' in höchstens r Stellen unterscheidet, so weiß man, dass $c'' = c$ gilt: Der Fehler wurde korrigiert. Die ***Korrekturrate*** ist $\frac{r}{n}$.

Beispiel: Wird $d \in \mathbb{F}_2^4$ durch $(d,d,d) \in \mathbb{F}_2^{12}$ codiert, kann $r = 1$ Fehler korrigiert werden: Wird $(0,1,1,0,0,1,0,0,0,1,1,0)$ empfangen, so ist klar, dass an der rot markierten Stelle 1 durch 0 ersetzt wurde. Die Korrekturrate ist $\frac{1}{12}$.

Bei einem ***linearen Code*** wählt man eine ***Kontrollmatrix*** $H \in \mathbb{K}^{m\times n}$, und dann $C := \mathrm{LR}(H;\vec{0})$. Ist k die Anzahl der Basislösungen, so ist die ***Informationsrate*** $\frac{k}{n}$ (im obigen Beispiel ist also die Informationsrate $\frac{4}{12}$). Wird der Block $c' \in \mathbb{K}^n$ empfangen, so heißt $Hc' \in \mathbb{K}^m$ das ***Syndrom*** von c'.

Beispiel 3.23
$\mathbb{F}_2^m$ enthält 2^m-1 von $\vec{0}$ verschiedene Elemente, die wir in eine $m\times(2^m-1)$-Matrix H in ZSF schreiben. Für $m = 3$ beispielsweise $H := \left(\begin{smallmatrix}1&0&0&1&1&0&1\\0&1&0&1&0&1&1\\0&0&1&0&1&1&1\end{smallmatrix}\right) \in \mathbb{F}_2^{3\times 7}$.

Die Spalten $m+1$ bis 2^m-1 sind Nicht-Pivotspalten; sei $n := 2^m-1$ und $k := 2^m - m - 1$. Die **Generatormatrix** *sei $G := (\vec{\beta}_{m+1}, ..., \vec{\beta}_{2^m-1}) \in \mathbb{K}^{n\times k}$, und dann $C := \mathrm{LR}(H;\vec{0}) = \{G\cdot\vec{d} \mid \vec{d} \in \mathbb{K}^k\}$. Dies ist der (n,k)-***Hamming-Code***[35]. Für $m = 3$ beispielsweise $G := \left(\begin{smallmatrix}1&1&0&1\\1&0&1&1\\0&1&1&1\\1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{smallmatrix}\right)$. Interpretation: Die ersten drei Bit sind Prüfbits und die letzten vier Bit Datenbits. So wird $(0,1,1,0)$ durch $c = (1,1,0,0,1,1,0) \in \mathbb{F}_2^7$ codiert.*

[34] ***Chiffrierung***, also die Verschlüsselung codierter Daten, ist etwas anderes.
[35] Richard Wesley Hamming [1915–1998]

Lemma 3.24
Ein (n,k)-Hamming-Code hat Informationsrate $\frac{k}{n}$ und die Korrekturrate $\frac{1}{n}$.

Beweis:
Die Informationsrate ist klar. Sei $c \in \mathbb{F}_2^n$ das gesendete Codewort und $c' \in \mathbb{F}_2^n$ der empfangene Block, so dass der Fehler $f := c' - c$ nur an der Stelle $i \in \{1, ..., n\}$ von Null verschieden ist. Das Syndrom ist $Hc' = H(c' - c) = Hf$, und das ist die i-te Spalte von H. Man korrigiert also den Fehler, indem man an der i-ten Stelle von c' 0 durch 1 bzw. 1 durch 0 ersetzt. □

Beispiel: $c' = (1,1,0,1,1,1,0)$ hat das Syndrom $Hc' = \begin{pmatrix}1\\1\\0\end{pmatrix} = H_{*,4}$, also wird $c'' = c = (1,1,0,0,1,1,0)$ decodiert.

Praktische Anwendungen fehlerkorrigierender linearer Codes:

- Mit einem zusätzlichen Paritätsbit entsteht ein „erweiterter“ Hamming-Code, der einen Fehler pro Block korrigieren und zwei Fehler pro Block erkennen kann. Das genügt für schwach verrauschte Kanäle, etwa beim fehlerkorrigierenden RAM. Nutzt man von einem $(127, 120)$-Hammingcode $(m = 7)$ mit zusätzlichem Paritätsbit 64 der 120 Datenbits, erhält man einen 1-Fehler-korrigierenden 2-Fehler-erkennenden 64-Bit-Speicher und benötigt dafür die Busbreite $64 + m + 1 = 72$.

- Der erweiterte binäre Golay-Code[36] erreicht mit der Blocklänge 24 die Informationsrate $\frac{1}{2}$ und die Korrekturrate $\frac{1}{8}$, kann also pro Block bis zu drei Fehler korrigieren. Dieser Code wurde für die Bildübertragung der Voyager-1- und -2-Sonden verwendet.

- Der ternäre Golay-Code nutzt $\mathbb{F}_3 = \mathbb{Z}/3\mathbb{Z}$ statt $\mathbb{F}_2$ und erreicht mit Blocklänge 11 die Informationsrate $\frac{6}{11}$ und kann zwei Fehler pro Block korrigieren. Unabhängig von Golay wurde dieser Code von einem Fußballfan zum Knacken von Sportwetten entwickelt. Er kann zur Fehlerkorrektur im quantum computing genutzt werden.

- Hadamard-Codes[37] erreichen nur eine sehr geringe Informationsrate $\frac{k}{2^k}$, aber mit wachsendem k nähert man sich der Korrekturrate $\frac{1}{2}$. Hadamard-Codes werden daher für sehr stark verrauschte Kanäle verwendet, etwa 1971 für die Bildübertragung der Mariner-9-Marssonde.

- Das hier beschriebene Verfahren zum Decodieren (nämlich: Nachsehen, in welcher Spalte der Kontrollmatrix man das Syndrom findet) ist bei großen Kontrollmatrizen nicht effizient genug. Es gibt lineare Codes, die auf möglichst effizientes Codieren und Decodieren optimiert wurden, etwa Turbo-Codes in 3G- und 4G- oder Polare Codes in 5G-Netzen.

[36] Marcel Jules Edouard Golay [1902–1989]
[37] Jacques Salomon Hadamard [1865–1963]

4 Vektorräume

Auch in diesem Kapitel sei $\mathbb{K}$ ein Körper. In den vorigen Kapiteln behandelten wir typische Rechentechniken der linearen Algebra.[38] Deren Grundlagen sollen jetzt begrifflich erfasst werden.

- (Unter-)Vektorräume: Abstrahiert von $\mathbb{K}^{m\times n}$ und Lösungsräumen homogener linearer Gleichungssysteme und ist dann auch auf die Menge der stetigen Funktionen und auf den Übergang von $\mathbb{Q}$ nach $\mathbb{R}$ nach $\mathbb{C}$ anwendbar.
- Lineare Abbildungen: Abstrahiert von der Matrixmultiplikation und ist dann auch auf Differential- und Integralrechnung sowie geometrische Abbildungen anwendbar.
- Basen: Abstrahiert von der „sparsamen“ Erzeugung von Lösungsräumen mittels Linearkombinationen sowie von Koordinatensystemen und ist eine wichtige Grundlage für rechnerische Methoden.

Ich verwende Pfeilakzente für Spaltenvektoren. Für Elemente allgemeiner Vektorräume verwende ich Kleinbuchstaben ohne Pfeilakzent, also b statt $\vec{b}$. Elemente von $\mathbb{K}$ nennt man auch ***Skalare***, und um sie von Vektoren zu unterscheiden, verwenden wir für sie kleine griechische Buchstaben.

4.1 Vektorraumaxiome

Definition 4.1. *Es sei $\mathbb{K}$ ein Körper mit Einselement $1 \in \mathbb{K}$. Ein* **$\mathbb{K}$–Vektorraum** *V (oder deutlicher $(V, +, \cdot)$) besteht aus einer Menge V mit*

- *einer inneren Verknüpfung $+$ auf V (***Vektor-Addition***) zusammen mit einem Element $o \in V$, durch die $(V, +, o)$ zu einer abelschen Gruppe wird:*

  (V1) Kommutativ: $u + v = v + u$ für alle $u, v \in V$

  (V2) Assoziativ: $u + (v + w) = (u + v) + w$ für alle $u, v, w \in V$

  (V3) Nullvektor: Für jedes $v \in V$ gilt $v + o = v$.

  (V4) Negation: Zu jedem $v \in V$ gibt es $-v \in V$ mit $v + (-v) = o$.

- *einer* **Skalarmultiplikation**[39] $\cdot\colon \mathbb{K} \times V \to V$, $(\lambda, v) \mapsto \lambda v$ *(auch $\lambda \cdot v$ geschrieben), so dass gilt:*

  (V5) Assoziativität: $(\lambda\mu)v = \lambda(\mu v)$ für alle $\lambda, \mu \in \mathbb{K}$ und für alle $v \in V$;

  (V6) Eins: $1v = v$ für alle $v \in V$.

[38] Das meiste davon war bis vor wenigen Jahren Schulstoff, d.h. eigentliche Hochschulthemen beginnen erst jetzt.

[39] Das ist i.A. keine innere Verknüpfung!

- *Addition in $\mathbb{K}$ ebenso wie Addition in V erfüllen mit der Skalarmultiplikation ein Distributivgesetz. Also für alle $\lambda, \mu \in \mathbb{K}$ und $u, v \in V$:*

$$(V7) \quad \lambda(u+v) = \lambda u + \lambda v \qquad (V8) \quad (\lambda+\mu)v = \lambda v + \mu v\,.$$

Bemerkung An mehreren Stellen besteht eine Verwechslungsgefahr: Das neutrale Element der Vektoraddition ist der Nullvektor o (bisher hieß er $\vec{0}$), das natürlich vom Nullelement $0 \in \mathbb{K}$ zu unterscheiden ist. Und sowohl in V als auch in $\mathbb{K}$ verwendet man meist die gleichen Verknüpfungssymbole $+$ bzw. $\cdot$. Man sollte sich des Unterschieds bewusst sein!

Beispiel 4.2

a) Herkömmliche Vektoren im Anschauungsraum bzw. Kräfte in der Physik: Das ist etwas ganz anderes als Spaltenvektoren, denn Spaltenvektoren sind bereits eine rechnerische Abstraktion von Vektoren!

b) $\mathbb{K}^{m\times n}$ sowie die Lösungsräume homogener linearer Gleichungssysteme.

c) $\mathrm{Abb}(\mathbb{R},\mathbb{R})$ mit **punktweiser** *Addition und Skalarmultiplikation, d.h. für alle $f, g\colon \mathbb{R} \to \mathbb{R}$ und $\lambda \in \mathbb{R}$ sind $f+g\colon \mathbb{R} \to \mathbb{R}$ und $\lambda f\colon \mathbb{R} \to \mathbb{R}$ definiert durch*

$$(f+g)(x) := f(x) + g(x) \qquad (\lambda f)(x) := \lambda \cdot f(x)\,.$$

d) Der Vektorraum $\mathbb{K}[X]$ der Polynome über $\mathbb{K}$. Der Vektorraum $\mathbb{K}[X]_n$ aller Polynome von Grad $\leq n$ über $\mathbb{K}$.

e) Der Vektorraum $\mathscr{C}^d(\mathbb{R},\mathbb{R})$ der mindestens d-mal stetig differenzierbaren Abbildungen von $\mathbb{R}$ nach $\mathbb{R}$. Beispielsweise sind die Sinusfunktion und alle Polynomfunktionen beliebig oft stetig differenzierbar, aber die Abbildung $x \mapsto \sqrt{x^2} = |x|$ ist zwar stetig, aber nicht differenzierbar. Es handelt sich um einen Vektorraum, denn wenn $f, g\colon \mathbb{R} \to \mathbb{R}$ differenzierbar sind und $\lambda, \mu \in \mathbb{R}$, dann ist auch die Abbildung $x \mapsto \lambda f(x) + \mu g(x)$ differenzierbar, wegen $(\lambda f + \mu g)' = \lambda f' + \mu g'$.

f) Man sollte sich bei der Arbeit mit Begriffen stets auch fragen, ob die Axiome des Begriffs vielleicht redundant sind, also ob ein Axiom aus den anderen Axiomen folgt: Durch die Skalarmultiplikation mit $-1 \in \mathbb{K}$ erhält man $(-1)v$, das müsste sicherlich $-v$ sein, also muss man die Existenz des additiven Inversen nicht extra fordern, oder etwa doch?

Zum Nachweis der Redundanzfreiheit sucht man nach „pathologischen Beispielen“, in denen ein Axiom verletzt ist und alle anderen gelten. Hier zeige ich, dass (V4) nicht redundant ist:

Wir setzen $V = \mathbb{R}^2 \cup \{@\}$*, wobei* @ *ein zusätzliches Element bezeichnet, das* nicht *schon in* $\mathbb{R}^2$ *enthalten ist. Für die Addition und Skalarmultiplikation ergänzen wir die üblichen Operationen auf* $\mathbb{R}^2$ *durch*

$$\binom{x}{y} + @ := @ + \binom{x}{y} = \binom{x}{y} \qquad @ + @ := @ \qquad \lambda @ = @$$

für alle $\binom{x}{y} \in \mathbb{R}^2$, $\lambda \in \mathbb{R}$*. Es ist dann etwas mühsam aber nicht schwer, nachzuweisen, dass (V1), (V2) und (V5)–(V8) weiterhin erfüllt werden. Ferner wird (V3) erfüllt, beachten Sie aber, dass diesmal nicht* $\binom{0}{0}$ *der Nullvektor ist, sondern* @*. Aber (V4) wird nicht erfüllt, denn die Gleichung* $v + w = @$ *ist nicht für alle* $v \in V$ *lösbar, sondern die einzige Lösung ist* $v = w = @$*. Also folgt (V4) nicht aus den anderen Axiomen.*

Lemma 4.3
Sei V *ein* $\mathbb{K}$*–Vektorraum.*

a) Für alle $v \in V$ *und für alle* $\lambda \in \mathbb{K}$ *gilt* $0 \cdot v = o = \lambda \cdot o$*.*

b) Für alle $v \in V$ *und für alle* $\lambda \in \mathbb{K}$ *gilt* $(-\lambda)v = -(\lambda v) = \lambda(-v)$*.*

c) $\forall \lambda \in \mathbb{K}$ *und* $\forall v \in V$*: Aus* $\lambda v = o$ *folgt* $\lambda = 0$ *oder* $v = o$*.*

Beweis:

a) $\lambda v + o = \lambda v = (\lambda + 0)v = \lambda v + 0v$. Also $0v = o$ durch Addition von $-(\lambda v)$ auf beiden Seiten. Außerdem $\lambda o + o = \lambda o = \lambda(o + o) = \lambda o + \lambda o \Rightarrow \lambda o = o$ durch Addition von $-\lambda o$ auf beiden Seiten.

b) $\lambda v + (-\lambda)v = (\lambda - \lambda)v = 0v = o$, und $\lambda v + \lambda(-v) = \lambda(v - v) = \lambda o = o$.

c) Wir zeigen: Ist $\lambda v = o$ aber $\lambda \neq 0$, dann gilt $v = o$. Wegen $\lambda \neq 0$ existiert $\frac{1}{\lambda} \in \mathbb{K}$. Dann $o = \frac{1}{\lambda} o = \frac{1}{\lambda}(\lambda v) = \left(\frac{1}{\lambda}\lambda\right) v = 1v = v$. □

Bemerkung Man beachte, dass die Gültigkeit von (V4) aus b) und den übrigen Vektorraumaxiomen gefolgert werden könnte. Allerdings wurde b) mittels a) bewiesen, und dies wiederum nutzt (V4), um λv zu kürzen. Und wie wir am „pathologischen Beispiel“ sahen, kann man (V4) wirklich nicht einfach weglassen. Aber man könnte es durch das Axiom $\forall v \in V\colon\ 0 \cdot v = o$ ersetzen, oder auch $\forall v \in V\colon\ \exists \lambda \in \mathbb{K}\colon\ \exists w \in V\colon\ w + \lambda v = o$.

Der zweite Teil von b) wäre nicht geeignet, (V4) zu beweisen, denn er ist von der Form: Wenn $-v$ existiert, dann existiert auch $-(\lambda v)$ und hat den Wert $\lambda(-v)$.

Notation 4.4
Sei V *ein* $\mathbb{K}$*-Vektorraum; dabei ist stets implizit vorausgesetzt, dass* $\mathbb{K}$ *ein Körper ist. Für* $u, v \in V$ *schreiben wir* $u - v := u + (-v)$*.*

4.2 Lineare Abbildungen

Wir wollen nun von der Matrixarithmetik abstrahieren. Ist $A \in \mathbb{K}^{m\times n}$, dann ist eine Abbildung $L_A\colon \mathbb{K}^n \to \mathbb{K}^m$ mit $L_A(\vec{v}) = A \cdot \vec{v}$ definiert. Dabei zeigt sich, dass das Assoziativgesetz der Matrixmultiplikation eng mit allgemeinen Regeln für die Hintereinanderausführung von Abbildungen korrespondiert, die Sie teilweise bereits in einer Hausaufgabe kennenlernten.

Definition 4.5

a) Abbildungen $f, g\colon X \to Y$ sind **gleich** *($f = g$) gdw. $\forall x \in X\colon f(x) = g(x)$.*

b) Eine Abbildung $f\colon X \to Y$ heißt...

i) **injektiv** $\iff \forall x_1, x_2 \in X\colon x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2)$.

ii) **surjektiv** $\iff \forall y \in Y\, \exists x \in X\colon f(x) = y$.

iii) **bijektiv** *gdw. f ist injektiv und surjektiv.*

c) Die **Identitätsabbildung** $\mathrm{Id}_X\colon X \to X$ *ist dadurch definiert, dass $\forall x \in X\colon \mathrm{Id}_X(x) := x$.*

d) Für $f\colon X \to Y$ und $g\colon Y \to Z$ wird die **Verknüpfung** *oder* **Komposition** *$g \circ f\colon X \to Z$ dadurch definiert, dass $\forall x \in X\colon (g \circ f)(x) := g(f(x))$.*

Definition 4.6 (und Axiom und Lemma)
Sei $f\colon X \to Y$, $A \subset X$ und $B \subset Y$.

a) Die **Einschränkung** *oder* **Restriktion** *von f auf A, notiert als $f\big|_A\colon A \to Y$, wird definiert durch $\forall x \in A\colon f\big|_A(x) := f(x)$.*

b) $f(A)$ bezeichnet die **Bildmenge** *$f(A) := \{f(x) \mid x \in A\}$ von A. Insbesondere nennt man $\mathrm{Bild}(f) := f(X) = \{f(x) \mid x \in X\}$ das* **Bild**[40] *von f.*

c) Wenn $\mathrm{Bild}(f) \subset B$, so ist die **Corestriktion** *von f auf B die Abbildung $f\Big|^B\colon X \to B$ definiert durch $\forall x \in X\colon f\Big|^B(x) := f(x)$.*

d) Das **Urbild**[41] *von B unter f ist $f^{-1}(B) := \{x \in X \mid f(x) \in B\} \subseteq X$.*

e) Auswahlaxiom: Ist f surjektiv, so gibt es eine Abbildung $g\colon Y \to X$ mit $f \circ g = \mathrm{Id}_Y$.

f) Ist f bijektiv, so gibt[42] *es eine durch f eindeutig bestimmte Abbildung $f^{-1}\colon Y \to X$ mit $f \circ f^{-1} = \mathrm{Id}_Y$ und $f^{-1} \circ f = \mathrm{Id}_X$. Man nennt sie die* **inverse Abbildung** *oder* **Umkehrabbildung** *von f.*

[40] Auf Englisch "*image*"
[41] Auf Englisch "*preimage*". Andere Bezeichnung: „*Faser*“, englisch "*fibre*“
[42] Auch ohne Verwendung des Auswahlaxioms!

Beispiele:

a) $f\colon \mathbb{R} \to \mathbb{R}_{\geq 0}$ mit $f(x) := x^2$ ist nicht bijektiv, aber ist surjektiv. Für $g\colon \mathbb{R}_{\geq 0} \to \mathbb{R}$ mit $g(x) := \sqrt{x}$ gilt $f \circ g = \mathrm{Id}_{\mathbb{R}_{\geq 0}}$, aber nicht $g \circ f = \mathrm{Id}_{\mathbb{R}}$.

b) $\cos\big|_{[0,\pi]}^{[-1,1]}$ ist bijektiv, die Umkehrabbildung ist arccos.

c) $\exp\colon \mathbb{R}_{>0} \to \mathbb{R}$ mit $\exp(x) := \mathrm{e}^x$ ist bijektiv, $\exp^{-1} = \ln\colon \mathbb{R} \to \mathbb{R}_{>0}$.

Trotz der ähnlichen Notation sind natürlich das Urbild einer Menge und die inverse Abbildung zwei verschiedene Dinge!

Beweis:
Zu e) sei angemerkt, dass wegen Surjektivität $\forall y \in Y\colon\ f^{-1}(\{y\}) \neq \emptyset$. Die Abbildung g hat die Eigenschaft $\forall y \in Y\colon\ g(y) \in f^{-1}(\{y\})$.

Wir beweisen f): Für alle $y \in Y$ besteht $f^{-1}(\{y\})$ wegen Bijektivität von f aus genau einem Element. Durch die Bedingung $f^{-1}(y) \in f^{-1}(\{y\})$ ist die Abbildung $f^{-1}\colon Y \to X$ also eindeutig und ohne Auswahlproblem bestimmt, woraus zunächst $f \circ f^{-1} = \mathrm{Id}_Y$ folgt. Ferner $\forall x \in X\colon\ f^{-1}(\{f(x)\}) = \{x\}$ und daher $f^{-1}(f(x)) = x$. Das bedeutet $f^{-1} \circ f = \mathrm{Id}_X$. □

Lemma 4.7
Verknüpfung von Abbildungen ist assoziativ, d.h. sind $f\colon X \to Y$, $g\colon Y \to Z$ und $h\colon Z \to W$, so gilt $h \circ (g \circ f) = (h \circ g) \circ f$.

Beweis:
Für jedes $x \in X$ haben beide Abbildungen den Wert $h(g(f(x)))$. □

Beispiel: Sowohl die Definition der Komposition von Abbildungen als auch die Assoziativität von Abbildungen widerspiegelt sich im Assoziativgesetz der Matrixmultiplikation: Sind $A \in \mathbb{K}^{m\times n}$, $B \in \mathbb{K}^{k\times m}$, $C \in \mathbb{K}^{\ell\times k}$ und $\vec{v} \in \mathbb{K}^n$, dann gilt $L_{BA}(\vec{v}) = (BA)\vec{v} = B(A\vec{v}) = L_B \circ L_A(\vec{v})$ und $(L_C \circ L_B) \circ L_A(v) = (CB)(A\vec{v}) = ((CB)A)\vec{v} = (C(BA))\vec{v} = C((BA)\vec{v}) = L_C \circ (L_B \circ L_A)(\vec{v})$.

Um weitere Eigenschaften der Matrixarithmetik (etwa Distributivgesetze) zu erfassen, müssen wir eine besondere Art von Abbildungen definieren:

Definition 4.8. *Seien V, W $\mathbb{K}$-Vektorräume.*

a) Eine Abbildung $f\colon V \to W$ heißt **linear** *(oder auch $\mathbb{K}$-linear oder* **Homomorphismus** *von $\mathbb{K}$-Vektorräumen) $:\Leftrightarrow \forall u, v \in V$ und $\forall \lambda \in \mathbb{K}$ gilt $f(u + v) = f(u) + f(v)$ und $f(\lambda v) = \lambda f(v)$.*

b) $\mathrm{Hom}_{\mathbb{K}}(V, W) := \{f\colon V \to W \mid f$ ist $\mathbb{K}$-linear$\}$.

c) $V^* := \mathrm{Hom}_{\mathbb{K}}(V, \mathbb{K})$ heißt **Dualraum** *von V.*

Beispiele:

a) Für $A \in \mathbb{K}^{m\times n}$ ist $L_A \in \mathrm{Hom}_{\mathbb{K}}(\mathbb{K}^n, \mathbb{K}^m)$. Insbesondere entspricht ein Zeilenvektor $\underline{u} \in \mathbb{K}^n$ einem Element von $(\mathbb{K}^n)^*$, nämlich der linearen Abbildung $\mathbb{K}^n \ni \vec{v} \to \underline{u}\vec{v} \in \mathbb{K}^1 \mathrel{\hat{=}} \mathbb{K}$.

b) $\mathbb{R} \to \mathbb{R}$ mit $x \mapsto x^2$ ist *nicht* linear: $(1+1) \mapsto 2^2 = 4 \neq 1^2 + 1^2 = 2$

c) In der Schule haben Sie eine Funktion der Form wie $f(x) = 5\,x + 3$ vermutlich als „lineare Funktion“ bezeichnet. Das ist jedoch *keine* lineare Abbildung im Sinne der Definition, denn $f(1+1) = 13 \neq f(1)+f(1) = 8+8 = 16$.

d) Für differenzierbare Funktionen $f \in \mathscr{C}^1(\mathbb{R}, \mathbb{R})$ ist durch $D(f) := f'$ eine Abbildung $D\colon \mathscr{C}^1(\mathbb{R}, \mathbb{R}) \to \mathscr{C}^0(\mathbb{R}, \mathbb{R})$ definiert. Es gilt bekanntlich $(c_1 f_1 + c_2 f_2)' = c_1 f_1' + c_2 f_2'$ für $f_1, f_2 \in \mathscr{C}^1(\mathbb{R}, \mathbb{R})$ und $c_1, c_2 \in \mathbb{R}$. Also ist D linear.

e) Sei $\mathbb{K} := \mathbb{Z}/2\mathbb{Z}$ und sei $V := \mathbb{K}[X]$. Siehe Übungen: Die Abbildung $f\colon V \to V$ definiert durch $\forall p \in V\colon\ f(p) := p^2$ ist eine lineare Abbildung. Das ist erstaunlich, denn wenn man Polynome mit *reellen* Koeffizienten betrachtet, so ist die Abbildung $p \mapsto p^2$ *nicht* linear!

f) $\mathbb{C}$ ist ein $\mathbb{R}$-Vektorraum (wir haben ja $\mathbb{C}$ mit $\mathbb{R}^2$ dargestellt). Die Abbildungen $\kappa\colon \mathbb{C} \to \mathbb{C}$ mit $\kappa(z) := \bar{z}$ und $\mathrm{Re}\colon \mathbb{C} \to \mathbb{R}$ sind $\mathbb{R}$-linear. Sind nämlich $a, b, c, d, \lambda, \mu \in \mathbb{R}$ und $w := a + b\,\mathrm{i}$ sowie $z := c + d\,\mathrm{i}$, so gilt $\lambda w + \mu z = (\lambda a + \mu c) + (\lambda b + \mu d)\,\mathrm{i}$ und daher

- $\kappa(\lambda w + \mu z) = (\lambda a + \mu c) - (\lambda b + \mu d)\,\mathrm{i} = \lambda(a - b\,\mathrm{i}) + \mu(c - d\,\mathrm{i}) = \lambda\kappa(w) + \mu\kappa(z)$
- $\mathrm{Re}(\lambda w + \mu z) = \lambda a + \mu c = \lambda\,\mathrm{Re}(w) + \mu\,\mathrm{Re}(z)$.

Lemma 4.9. *Sei $f \in \mathrm{Hom}_{\mathbb{K}}(V, W)$.*
Sind $o_V \in V$ und $o_W \in W$ die Nullvektoren, so gilt $f(o_V) = o_W$.

Beweis:
Nach Lemma 4.3 und Linearität von f gilt $f(o_V) = f(0o_V) = 0f(o_V) = o_W$. □

Bemerkung Für $A \in \mathbb{K}^{m\times n}$ und $\vec{b} \in \mathbb{K}^m$ ist offenbar $L_A^{-1}(\{\vec{b}\}) = \mathrm{LR}(A; \vec{b})$. Die Aufgabe, das Urbild unter einer linearen Abbildung zu berechnen, verallgemeinert also das Berechnen des Lösungsraums eines linearen Gleichungssystems.

Aufgabe 4.10 (und Definition). *Sei $f\colon V \to W$ $\mathbb{K}$-linear. Zeigen Sie:*

a) $\mathrm{Bild}(f)$ *ist ein $\mathbb{K}$-Vektorraum bezüglich der in W definierten Vektoraddition und Skalarmultiplikation.*

b) $\ker(f) := \{v \in V\colon f(v) = o_W\}$ *ist ein* $\mathbb{K}$*-Vektorraum bezüglich der in* V *definierten Vektoraddition und Skalarmultiplikation. Man nennt* $\ker(f)$ *den* **Kern** *von* f*.*

Aufgabe 4.11 (und Definition). *Seien* V, W $\mathbb{K}$*-Vektorräume.*
Für $f, g \in \mathrm{Hom}_{\mathbb{K}}(V, W)$ *und* $\lambda \in \mathbb{K}$ *definieren wir* $(f+g)\colon V \to W$ *und* $(\lambda f)\colon V \to W$ *durch* $(f+g)(v) := f(v)+g(v)$ *und* $(\lambda f)(v) := \lambda f(v)$*. Zeigen Sie, dass dadurch* $\mathrm{Hom}_{\mathbb{K}}(V, W)$ *zu einem* $\mathbb{K}$*-Vektorraum wird.*

4.3 Untervektorräume

Bei der Lösung der vorigen Aufgaben hilft folgender Begriff:

Definition 4.12. *Sei* V *ein* $\mathbb{K}$*-Vektorraum.*
$U \subset V$ *heißt* **Untervektorraum** *von* V*, Notation* $U \leq V$*, gdw.* U *ist bezüglich der Vektoraddition und Skalarmultiplikation von* V *selbst ein* $\mathbb{K}$*-Vektorraum.*

Beispiele:

a) Sei $c \leq d \in \mathbb{N}$. Dann $\mathbb{K}[X]_c \leq \mathbb{K}[X]_d \leq \mathbb{K}[X]$ und $\mathscr{C}^d(\mathbb{R}, \mathbb{R}) \leq \mathscr{C}^c(\mathbb{R}, \mathbb{R})$.

b) Ist $f \in \mathrm{Hom}_{\mathbb{K}}(V, W)$, dann $\ker(f) \leq V$ und $\mathrm{Bild}(f) \leq W$.

Charakterisierung von Untervektorräumen
Sei V *ein* $\mathbb{K}$*-Vektorraum und* $U \subset V$*. Es gilt* $U \leq V \iff U \neq \emptyset$ *und* $\forall \lambda, \mu \in \mathbb{K}$ *und* $u, v \in U\colon \lambda u + \mu v \in U$*.*

Beweis:
Wenn $U \leq V$, dann gelten die beiden Aussagen offenbar. Sei nun $U \subset V$ und die beiden Aussagen seien wahr. Dann folgt $\forall u, v \in U\colon u + v = 1u + 1v \in U$ und $\forall u \in U, \lambda \in \mathbb{K}\colon \lambda u = \lambda u + 0u \in U$, also ist Vektoraddition eine innere und Skalarmultiplikation eine äußere Verknüpfung auf U.

Wegen $U \neq \emptyset$ gibt es ein $u \in U$. Dann auch $o = 0u \in U$ und $-u = (-1)u \in U$. Also gelten (V3) und (V4). Die übrigen Vektorraumaxiome gelten sogar in V, sie gelten also erst recht eingeschränkt auf die Elemente von U. □

Wir reformulieren nun den Begriff der Linearkombination so, dass wir prinzipiell auch mit unendlich vielen Vektoren umgehen könnten. Technische Probleme:

- Vektoren und Vorfaktoren gehören zusammen, wir würden gerne $\lambda_1 v_1 + \lambda_2 v_2 + ...$ schreiben. Aber was macht man, wenn man die Vektoren nicht durchnummerieren kann?
- Summen dürfen nur endlich viele Summanden haben (anders als konvergente Reihen in der Analysis). Wir wollen aber auch einen unendlichen Pool potentieller Summanden zulassen; wie drückt man aus, dass nur endlich viele davon tatsächlich in der Summe auftreten?

Notation 4.13. *Seien M und I Mengen.*

a) *Eine* **Familie** $(a_i)_{i\in I} \subset M$ *mit* **Indexmenge** I *ist eine Abbildung* $I \ni i \mapsto a_i \in M$. *Wir schreiben* $[a_1, ..., a_n] \subset M$ *statt* $(a_i)_{i\in\{1,...,n\}} \subset M$.

b) *Ist* $(\gamma_i)_{i\in I} \subset \mathbb{K}$ *so, dass* $\gamma_i \neq 0$ *nur für endlich viele* $i \in I$ *gilt, dann sagt man* $\gamma_i = 0$ *für* **fast alle** $i \in I$.

Da in der Abbildung $I \to M$ Elemente von M mehrfach getroffen werden können, entspricht eine Familie in M einer „Teilmenge mit Wiederholungen".

Definition 4.14 (und Lemma)
Sei V ein $\mathbb{K}$-Vektorraum, $S = (v_i)_{i\in I} \subset V$ und $(\gamma_i)_{i\in I} \subset \mathbb{K}$.

a) *Wenn $\gamma_i = 0$ für fast alle $i \in I$, dann heißt $\sum\limits_{i\in I} \gamma_i v_i$ eine* **Linearkombination** *von S mit den* **Koeffizienten** $(\gamma_i)_{i\in I}$. *Statt „$\sum\limits_{i\in I} \gamma_i v_i$ mit $\gamma_i = 0$ für fast alle $i \in I$" schreiben wir einfach $\sum\limits_{i\in I}{}' \gamma_i v_i$ (der Strich am Summenzeichen bedeutet, dass nur endlich viele Summanden auftreten). Sind alle Koeffizienten null, so heißt die Linearkombination* **trivial**.

Wenn wir die Indexmenge nicht explizit benennen möchten, schreiben wir einfach $\sum\limits_{v\in S}{}' \gamma_v v$ wobei man im Hinterkopf behalten sollte, dass zu jedem Vektor der Familie ein Index gehört und der gleiche Vektor in der Summe genau so oft wie in der Familie auftreten kann.

b) *S heißt* **linear unabhängig** $:\Leftrightarrow$ *die einzig mögliche Darstellung von o als Linearkombination von S ist die triviale Linearkombination.*

Andernfalls heißt S **linear abhängig***; dann gäbe es also eine nicht-triviale Linearkombination von S mit Wert o.*

c) *Die Menge aller Linearkombinationen von S heißt* **Erzeugnis** *oder* $\mathbb{K}$-**Span** *von S oder der von S erzeugte Untervektorraum. Notation:*

$$\mathrm{Span}_{\mathbb{K}}(S) := \{\sum_{i\in I}{}' \gamma_i v_i \mid (\gamma_i)_{i\in I} \subset \mathbb{K},\ \gamma_i = 0 \text{ für fast alle } i \in I\} \leq V$$

Ist $V = \mathrm{Span}_{\mathbb{K}}(S)$, so heißt S **Erzeugendensystem** *von V über $\mathbb{K}$.*

Beweis:
Es ist zu zeigen, dass $\mathrm{Span}_{\mathbb{K}}(S) \leq V$. Eine leere Summe hat den Wert Null, d.h. sogar im Fall $S = \emptyset$ ist $o \in \mathrm{Span}_{\mathbb{K}}(S)$.

Seien $u, v \in \mathrm{Span}_{\mathbb{K}}(S)$, also $u = \sum\limits_{w\in S}{}' c_w w$ und $v = \sum\limits_{w\in S}{}' d_w w$. Weil $c_w = d_w = 0$ für fast alle $w \in S$ gilt, gilt für $\lambda, \mu \in \mathbb{K}$ auch $\lambda c_w + \mu d_w = 0$ für fast alle $w \in S$. Also $\lambda u + \mu v = \sum\limits_{w\in S}{}' (\lambda c_w + \mu d_w) w \in \mathrm{Span}_{\mathbb{K}}(S)$. □

4.4 Basen

Sei V ein $\mathbb{K}$-Vektorraum. Es gilt immer $\text{Span}_{\mathbb{K}}(V) = V$, d.h. V ist sein eigenes Erzeugendensystem. Aber diese Sichtweise ist nicht besonders effizient: Wir möchten nun Erzeugendensysteme mit möglichst wenig Vektoren untersuchen.

Definition 4.15
Ein linear unabhängiges Erzeugendensystem eines $\mathbb{K}$-Vektorraums heißt **Basis**. *Wenn sich $\mathbb{K}$ nicht eindeutig aus dem Kontext ergibt, sagt man $\mathbb{K}$-Basis.*

Bemerkung 4.16
Wir wissen bereits, wie man prüfen kann, ob $[\vec{v}_1, ..., \vec{v}_k] \subset \mathbb{K}^n$ linear abhängig ist: Man bringt $A := (\vec{v}_1, ..., \vec{v}_k) \in \mathbb{K}^{n \times k}$ mit Gauß auf Zeilenstufenform A'. Die Familie ist genau dann linear unabhängig, wenn in A' jede Spalte Pivotspalte ist. Es folgt: Jede linear unabhängige Familie in $\mathbb{K}^n$ hat höchstens n Elemente.

Wir können auch prüfen, ob $[\vec{v}_1, ..., \vec{v}_k]$ ein Erzeugendensystem von $\mathbb{K}^n$ ist: $\vec{b} \in \mathbb{K}^n$ ist genau dann in $\text{Span}_{\mathbb{K}}(\vec{v}_1, ..., \vec{v}_k)$, wenn $\text{LR}(A; \vec{b}) \neq \emptyset$. Dies ist genau dann für alle $\vec{b} \in \mathbb{K}^n$ der Fall, wenn A' keine Nullzeile hat. Es folgt: Jedes Erzeugendensystem von $\mathbb{K}^n$ hat mindestens n Elemente.

Also hat jede Basis von $\mathbb{K}^n$ genau n Elemente und $[\vec{v}_1, ..., \vec{v}_n] \subset \mathbb{K}^n$ ist genau dann eine Basis von $\mathbb{K}^n$, wenn $(\vec{v}_1, ..., \vec{v}_n) \in GL_n(\mathbb{K})$.

Lemma 4.17 (und Definition)
Sei V ein $\mathbb{K}$-Vektorraum und $S \subset V$ eine Familie.
S ist Basis von $V \iff$ jedes $v \in V$ hat genau eine Darstellung als Linearkombination von S. Ggf. bezeichnet man die Koeffizienten dieser Linearkombination als die **Koordinaten** *von v bezüglich S.*

Beweis:
S ist genau dann ein Erzeugendensystem, wenn jedes $v \in V$ mindestens eine Darstellung als Linearkombination von S hat.

Wenn $V \ni v = \sum_{s \in S}{}' \lambda_s s = \sum_{s \in S}{}' \mu_s s$ mit $(\lambda_s)_{s \in S} \neq (\mu_s)_{s \in S}$ dann ist $o = \sum_{s \in S}{}' (\lambda_s - \mu_s) s$ eine nicht-triviale Linearkombination, also S linear abhängig. Und wenn S linear abhängig ist, dann hat o zwei verschiedene Darstellungen als Linearkombination von S, nämlich trivial und nicht-trivial. □

Beispiele:

a) Die Basislösungen des Lösungsraumes eines homogenen linearen Gleichungssystems bilden eine Basis dieses Lösungsraums, die freien Variablen einer Lösung sind die Koordinaten bezüglich dieser Basis.

b) Für $i \in \{1, ..., n\}$ sei $\vec{e}_i \in \mathbb{K}^n$ der Spaltenvektor, der in der i-ten Zeile den Eintrag 1 und überall sonst die Einträge 0 hat. Dann ist $[\vec{e}_1, ..., \vec{e}_n]$ eine Basis von $\mathbb{K}^n$. Man bezeichnet sie als die ***Standardbasis*** von $\mathbb{K}^n$.

c) $[\binom{1}{0}, \binom{1}{1}]$ und $[\binom{1}{0}, \binom{0}{-1}]$ sind jeweils Basen von $\mathbb{R}^2$. Die Koordinaten von $\binom{x}{y}$ bezüglich der Standardbasis sind $\binom{x}{y}$, bezüglich $[\binom{1}{0}, \binom{1}{1}]$ hingegen $\binom{x-y}{y}$ und bezüglich $[\binom{1}{0}, \binom{0}{-1}]$ sind die Koordinaten $\binom{x}{-y}$.

d) $[1]$ ist eine $\mathbb{C}$-Basis von $\mathbb{C}$. Hingegen ist $[1, \mathrm{i}]$ eine $\mathbb{R}$-Basis von $\mathbb{C}$; Real- und Imaginärteil sind die Koordinaten bezüglich dieser $\mathbb{R}$-Basis.

e) Für eine Menge I ist $\mathbb{K}^I_{\text{fin}} := \{(c_i)_{i\in I} \subset \mathbb{K} \mid c_i = 0 \text{ für fast alle } i \in I\}$ ein Untervektorraum von $\mathbb{K}^I$ (mit punktweiser Addition und Skalarmultiplikation). Wir nennen ihn den $\mathbb{K}$-Vektorraum der ***Abbildungen mit endlichem Träger***. Im Fall $I = \{1, ..., n\}$ ist $\mathbb{K}^I = \mathbb{K}^I_{\text{fin}} = \mathbb{K}^n$.

$(i^*)_{i\in I}$ mit $i^* := \begin{pmatrix} \begin{cases} 1 & (i=j) \\ 0 & (i\neq j) \end{cases} \end{pmatrix}_{j\in I} \subset \mathbb{K}$ ist eine Basis von $\mathbb{K}^I_{\text{fin}}$ (Zur Erläuterung: Die Basis ist eine mit $i \in I$ indizierte Familie, und jedes Element der Familie ist wiederum eine mit $j \in I$ indizierte Familie mit endlichem Träger), aber für unendliches I nicht von $\mathbb{K}^I$.[43]

f) Für die Theorie der $\mathbb{K}$-Vektorräume ist wichtig, dass $\mathbb{K}$ ein Körper ist — obwohl Division in den Vektorraumaxiomen (V1)–(V8) nicht auftritt. Fordert man nur, dass $\mathbb{K}$ ein Ring ist, so erhält man den Begriff des $\mathbb{K}$***-Moduls***. Linearkombinationen, Untermoduln und Erzeugendensysteme sind analog definiert. Aber ein $\mathbb{Z}$-Modul besitzt im Allgemeinen kein linear unabhängiges Erzeugendensystem. Beispielsweise ist $\mathbb{Z}/3\mathbb{Z}$ ein $\mathbb{Z}$-Modul mit Erzeugendensystem $\{[1]\}$, aber dieses ist linear abhängig: $3 \cdot [1] = 0 \cdot [1]$.

Lemma 4.18
Sei V ein $\mathbb{K}$-Vektorraum, $S \subset V$ eine Familie, $w \in S$ und $S' := S \setminus [w]$ (damit ist gemeint: Wenn w k-fach in S auftritt, dann tritt es noch $(k-1)$-fach in S' auf). Folgende drei Aussagen sind gleichbedeutend:

a) Es gibt eine Linearkombination $\sum_{v\in S}{}' \lambda_v v = o$ mit $\lambda_w \neq 0$.

b) $w \in \mathrm{Span}_{\mathbb{K}}(S')$.

c) $\mathrm{Span}_{\mathbb{K}}(S) = \mathrm{Span}_{\mathbb{K}}(S')$.

***Zusatz:** Ist S' linear unabhängig, so sind a), b), c) gleichbedeutend dazu, dass S linear abhängig ist.*

Beweis:
c) $\Rightarrow$ b): Klar.

b) $\Rightarrow$ a): Ist $\mathrm{Span}_{\mathbb{K}}(S') \ni w = \sum_{v\in S'}{}' \lambda_v v$, dann $o = \sum_{v\in S}{}' \lambda_v v$ mit $\lambda_w = -1$.

[43] Tatsächlich ist es in diesem Fall sehr schwer, eine Basis von $\mathbb{K}^I$ explizit anzugeben.

a) $\Rightarrow$ c): Aus jeder Linearkombination von Elementen von S kann man w löschen, indem man $w = \sum_{v \in S'}{}' \frac{-\lambda_v}{\lambda_w} v$ einsetzt. Also $\subseteq$, und $\supseteq$ ist klar.

Zusatz: Wenn a), dann ist S linear abhängig. Sei nun S' linear unabhängig. Wir zeigen: Wenn es $o = \sum_{v \in S}{}' \lambda_v v$ nicht-trivial gibt, dann folgt a). Widerspruchsbeweis: Wäre $\lambda_w = 0$, dann auch $o = \sum_{v \in S'}{}' \lambda_v v$ nicht-trivial — Widerspruch, da S' linear unabhängig. □

Aus dem vorigen Lemma folgt:

Charakterisierung von Basen
Sei V ein $\mathbb{K}$-Vektorraum. Für eine Familie $\mathcal{S} \subseteq V$ ist gleichbedeutend:

a) $\mathcal{S}$ ist eine Basis von V.

b) $\mathcal{S}$ ist ein minimales Erzeugendensystem von V: Entfernt man ein Element von $\mathcal{S}$, liegt danach kein Erzeugendensystem mehr vor.

c) $\mathcal{S}$ ist eine maximal linear unabhängige Familie in V: Fügt man ein Element zu $\mathcal{S}$ hinzu, liegt danach keine linear unabhängige Menge vor. □

Beweis:
a) $\iff$ b) Sei $\mathcal{S}$ ein Erzeugendensystem. Zu zeigen ist: $\mathcal{S}$ ist linear unabhängig gdw. $\mathcal{S}$ ist ein *minimales* Erzeugendensystem.

$\mathcal{S}$ ist nicht minimal gdw. $\exists w \in \mathcal{S}$, so dass $V = \mathrm{Span}_{\mathbb{K}}(\mathcal{S}) = \mathrm{Span}_{\mathbb{K}}(\mathcal{S} \setminus [w])$, gdw. $\mathcal{S}$ ist linear abhängig nach Lemma 4.18.

a) $\iff$ c) Sei $\mathcal{S}$ linear unabhängig. $\mathcal{S}$ ist *kein* Erzeugendensystem $\iff \exists w \in V \setminus \mathrm{Span}_{\mathbb{K}}(\mathcal{S}) \iff \exists w \in V \setminus \mathcal{S}\colon\ \mathrm{Span}_{\mathbb{K}}(\mathcal{S} \cup [w]) \supsetneq \mathrm{Span}_{\mathbb{K}}(\mathcal{S}) \iff \exists w \in V \setminus \mathcal{S}\colon\ \mathcal{S} \cup [w]$ ist linear unabhängig (Zusatz von Lemma 4.18). □

Bemerkung 4.19
Aus den mengentheoretischen Axiomen einschließlich des Auswahlaxioms folgt das so genannte Lemma von Zorn und aus diesem wiederum folgt, dass es in jedem Vektorraum eine maximale linear unabhängige Familie gibt, dass also jeder Vektorraum eine Basis besitzt. Dies aber wäre ein Beispiel für einen inkonstruktiven Beweis, der keinen Anhaltspunkt für die Bestimmung einer Basis gibt.

4.5 Isomorphismen

In diesem Abschnitt seien V, W $\mathbb{K}$-Vektorräume.

Definition 4.20

a) $f \in \mathrm{Hom}_{\mathbb{K}}(V,W)$ *heißt* **Isomorphismus** $:\Leftrightarrow f$ *ist bijektiv. Dafür schreibt man auch* $f\colon V \xrightarrow{\cong} W$.

b) V, W *sind zueinander* **isomorph**, *Notation* $V \cong W :\Leftrightarrow \exists f\colon V \xrightarrow{\cong} W$.

c) Eine lineare Abbildung $f\colon V \to V$ *nennt man auch* **Endomorphismus** *von* V. $\mathrm{End}_{\mathbb{K}}(V) := \mathrm{Hom}_{\mathbb{K}}(V,V)$ *heißt* **Endomorphismenring** *von* V.

d) Einen bijektiven Endomorphismen nennt man **Automorphismus**. $\mathrm{Aut}_{\mathbb{K}}(V) := \{f \in \mathrm{Hom}_{\mathbb{K}}(V,V) \mid f \text{ bijektiv}\}$ *heißt* **Automorphismengruppe** *von* V.

Bemerkung 4.21. *Aus den Ergebnissen einiger Hausaufgaben folgt:*

a) Wenn $f, g \in \mathrm{Aut}_{\mathbb{K}}(V)$, *dann* $f^{-1} \in \mathrm{Aut}_{\mathbb{K}}(V)$ *und* $f \circ g \in \mathrm{Aut}_{\mathbb{K}}(V)$. *Insbesondere ist* $(\mathrm{Aut}_{\mathbb{K}}(V), \circ, \mathrm{Id}_V)$ *eine Gruppe.*

b) Für $f, g \in \mathrm{End}_{\mathbb{K}}(V)$ *ist* $f \circ g \in \mathrm{End}_{\mathbb{K}}(V)$.

Übung: $\mathrm{End}_{\mathbb{K}}(V)$ *ist bezüglich Addition und Komposition ein Ring.*

Lemma 4.22. *Sei* $f \in \mathrm{Hom}_{\mathbb{K}}(V,W)$ *und* $S \subset V$ *eine Familie.*

a) Wenn f *injektiv und* S *linear unabhängig ist, dann ist* $f(S) \subset W$ *linear unabhängig.*

b) Wenn f *surjektiv und* $\mathrm{Span}_{\mathbb{K}}(S) = V$, *dann* $\mathrm{Span}_{\mathbb{K}}(f(S)) = W$.

c) Sei f *bijektiv.* S *ist genau dann eine Basis von* V, *wenn* $f(S)$ *eine Basis von* W *ist.*

Beweis:

a) Sei $o_W = \sum_{v\in S}{}' \gamma_v f(v)$; zu zeigen: $\forall v \in S\colon \gamma_v = 0$. Wegen f linear folgt $o_W = f(\sum_{v\in S}{}' \gamma_v v)$. Wegen f injektiv folgt $o_V = \sum_{v\in S}{}' \gamma_v v$. Wegen S linear unabhängig folgt $\forall v \in S\colon \gamma_v = 0$.

b) Sei $w \in W$. f surjektiv $\Rightarrow \exists x \in V\colon f(x) = w$. $\mathrm{Span}_{\mathbb{K}}(S) = V \Rightarrow \exists \gamma \in \mathbb{K}^S_{\mathrm{fin}}\colon x = \sum_{v\in S}{}' \gamma_v v \Rightarrow w = \sum_{v\in S}{}' \gamma_v f(v)$.

c) Folgt aus a) und b). □

Lemma 4.23 (und Definition)
Sei V ein $\mathbb{K}$-Vektorraum mit Basis $B = (b_i)_{i\in I} \subset V$. $\kappa_B\colon V \to \mathbb{K}^I_{fin}$ sei die Abbildung, die jedes $v \in V$ auf die Koordinaten von v bezüglich B abbildet. Man nennt κ_B die **Koordinatenabbildung** *bezüglich B. Sie ist ein Isomorphismus. Nach Definition der Koordinaten bezüglich B gilt $\forall v \in V\colon\ v = \sum_{i\in I}{}' \kappa_B(v)_i b_i$.*

Beispiel: Sei $i \in I$. Dann $\kappa_B(b_i) = i^* \in \mathbb{K}^I_{\text{fin}}$, denn $b_i = 1b_i = \sum_{j\in I}{}' \gamma_j b_j$, wobei $\gamma_j = \begin{cases} 1 & (j = i) \\ 0 & (j \neq i) \end{cases}$. Spezialfall: Wenn $B = [b_1, ..., b_n]$, dann ist $\kappa_B(b_i) = \vec{e}_i \in \mathbb{K}^n$.

Beweis:
$\forall \gamma \in \mathbb{K}^I_{\text{fin}}\colon\ \exists v \in V$ mit $\kappa_B(v) = \gamma$, nämlich $v = \sum_{i\in I}{}' \gamma_i b_i \in V$, also ist κ_B surjektiv. Injektivität ist klar, denn zwei verschiedene Vektoren können keinesfalls die gleichen Koordinaten haben, sonst wäre ja $\sum_{i\in I}{}' \gamma_i b_i \neq \sum_{i\in I}{}' \gamma_i b_i$. Also κ_B bijektiv.
$\forall \gamma, \delta \in \mathbb{K}^I_{\text{fin}},\ \lambda, \mu \in \mathbb{K}\colon\ \kappa_B^{-1}(\lambda\gamma + \mu\delta) = \sum_{i\in I}{}' (\lambda\gamma_i + \mu\delta_i) b_i = \lambda \sum_{i\in I}{}' \gamma_i b_i + \mu \sum_{i\in I}{}' \delta_i b_i$.
Also ist κ_B^{-1} und damit laut einer Hausaufgabe auch $\left(\kappa_B^{-1}\right)^{-1} = \kappa_B$ linear. □

Korollar 4.24 (und Definition). *Sei V ein $\mathbb{K}$-Vektorraum. Wenn V eine endliche Basis B hat, heißt V* **endlichdimensional**. *Jede Basis von V hat genau $|B|$ Elemente. Die* **Dimension** *von V ist $\dim(V) := |B|$ oder (wenn sich $\mathbb{K}$ nicht aus dem Kontext ergibt) $\dim_{\mathbb{K}}(V) := |B|$.*

Beispiele: $\dim(\mathbb{K}^n) = n$. $\dim_{\mathbb{C}}(\mathbb{C}) = 1$, aber $\dim_{\mathbb{R}}(\mathbb{C}) = 2$.

Beweis:
Sei $n := |B|$, also $\kappa_B\colon V \to \mathbb{K}^n$. Sei $C \subset V$ eine Basis. Dann ist auch $\kappa_B(C) \subset \mathbb{K}^n$ eine Basis von $\mathbb{K}^n$, also nach Bemerkung 4.16 $|C| \overset{\text{bij.}}{=} |\kappa_B(C)| = n$. □

Satz von der linearen Fortsetzung
Sei $B = (b_i)_{i\in I} \subset V$ eine Basis und $(w_i)_{i\in I} \subset W$. Es gibt genau ein $\varphi \in \text{Hom}_{\mathbb{K}}(V, W)$ mit $\forall i \in I\colon\ \varphi(b_i) = w_i$, nämlich $\forall v \in V\colon\ \varphi(v) = \sum_{i\in I}{}' \kappa_B(v)_i w_i$.

Beweis:
Eindeutigkeit: $\psi \in \text{Hom}_{\mathbb{K}}(V, W)$ mit $\forall i\colon\ \psi(b_i) = w_i \Rightarrow \forall v \in V\colon\ \psi(v) = \psi(\sum_{i\in I}{}' \kappa_B(v)_i b_i) \overset{\text{lin.}}{=} \sum_{i\in I}{}' \kappa_B(v)_i \psi(b_i) = \sum_{i\in I}{}' \kappa_B(v)_i w_i$.
Existenz: Wir definieren $\varphi\colon V \to W$ durch $\varphi(v) := \sum_{i\in I}{}' \kappa_B(v)_i w_i$. Es ist φ linear, denn $\forall \lambda, \mu \in \mathbb{K},\ u, v \in V\colon\ \sum_{i\in I}{}' \kappa_B(\lambda u + \mu v)_i w_i = \sum_{i\in I}{}' (\lambda\kappa_B(u)_i + \mu\kappa_B(v)_i) w_i$ wegen der Linearität von κ_B. □

Beobachtung 4.25 (und Notation)
Für jedes $\varphi \in \mathrm{Hom}_{\mathbb{K}}(\mathbb{K}^n, \mathbb{K}^m)$ gibt es genau ein $M_\varphi \in \mathbb{K}^{m\times n}$ mit $\varphi = L_{M_\varphi}$, nämlich $M_\varphi = (\varphi(\vec{e}_1), ..., \varphi(\vec{e}_n))$. Zudem ist φ genau dann ein Isomorphismus, wenn A_φ invertierbar ist.

Korollar 4.26 (und Definition)
Seien V, W endlichdimensionale $\mathbb{K}$-Vektorräume mit einer Basis $B = [b_1, ..., b_n]$ von V und $C = [c_1, ..., c_m]$ von W. Sei $f \in \mathrm{Hom}_{\mathbb{K}}(V, W)$. Wir setzen $\varphi := \kappa_C \circ f \circ \kappa_B^{-1} \colon \mathbb{K}^m \to \mathbb{K}^n$, und dann nennt man ${}_CM_B(f) := M_\varphi \in \mathbb{K}^{m\times n}$ die **Darstellungsmatrix** *(auch:* **Abbildungsmatrix***) von f bzgl. B und C.*

Die i-te Spalte von ${}_CM_B(f)$ entspricht den Koordinaten von $f(b_i)$ bezüglich C. Für alle $v \in V$ gilt $\kappa_C(f(v)) = {}_CM_B(f) \cdot \kappa_B(v)$

Beweis:
$(M_\varphi)_{*,i} = \varphi(\vec{e}_i) = \kappa_C(f(b_i))$, denn $\kappa_B(b_i) = \vec{e}_i$. Nach Beobachtung 4.25 gilt also $\forall v \in V \colon {}_CM_B(f) \cdot \kappa_B(v) = \varphi(\kappa_B(v)) = \kappa_C(f(\kappa_B^{-1}(\kappa_B(v)))) = \kappa_C(f(v))$. □

Beispiel 4.27
$f \colon \mathbb{R}^2 \to \mathbb{R}^2$ sei gegeben durch $\forall \binom{x}{y} \in \mathbb{R}^2 \colon f\binom{x}{y} := \binom{x+2y}{2x+4y}$. Man sieht leicht, dass f linear ist. Sei $S = [\vec{e}_1, \vec{e}_2]$ die Standardbasis. Dann gilt also $f(\vec{e}_1) = \vec{e}_1 + 2\vec{e}_2$ und $f(\vec{e}_2) = 2\vec{e}_1 + 4\vec{e}_2$. Daher ${}_SM_S(f) = \left(\begin{smallmatrix}1 & 2\\ 2 & 4\end{smallmatrix}\right)$.

Sei nun $\vec{b}_1 = \binom{1}{2}$ und $\vec{b}_2 = \binom{-2}{1}$. Dann ist $B = [\vec{b}_1, \vec{b}_2]$ eine Basis von $\mathbb{R}^2$, und wir können ${}_BM_B(f)$ berechnen: $f(\vec{b}_1) = \binom{1+2\cdot 2}{2\cdot 1+4\cdot 2} = \binom{5}{10} = 5\vec{b}_1 + 0\vec{b}_2$ und $f(\vec{b}_2) = \binom{-2+2\cdot 1}{2\cdot(-2)+4\cdot 1} = \vec{0} = 0\vec{b}_1 + 0\vec{b}_2$. Daher ${}_BM_B(f) = \left(\begin{smallmatrix}5 & 0\\ 0 & 0\end{smallmatrix}\right)$. Offenbar ist die Basis B besser zur Beschreibung von f geeignet als S.

Beobachtung 4.28 (und Definition)
Seien B und C Basen eines endlichdimensionalen $\mathbb{K}$-Vektorraums V, $n := \dim(V)$. Die **Basiswechselmatrix** *für die Transformation von B nach C ist ${}_C\mathbb{T}_B := {}_CM_B(\mathrm{Id}_V) \in M_n(\mathbb{K})$.*

a) Für alle $\vec{v} \in V$ gilt $\kappa_C(\vec{v}) = {}_C\mathbb{T}_B \cdot \kappa_B(\vec{v})$.

b) ${}_B\mathbb{T}_C = {}_C\mathbb{T}_B{}^{-1}$.

c) Ist $f \colon V \to W$ linear, B_1, B_2 endliche Basen von V, C_1, C_2 endliche Basen von W, dann ${}_{C_2}M_{B_2}(f) = {}_{C_2}\mathbb{T}_{C_1} \cdot {}_{C_1}M_{B_1}(f) \cdot {}_{B_1}\mathbb{T}_{B_2}$.

Beispiel: In Beispiel 4.27 kann man ${}_S\mathbb{T}_B = (\vec{b}_1, \vec{b}_2) = \left(\begin{smallmatrix}1 & -2\\ 2 & 1\end{smallmatrix}\right)$ direkt ablesen. Dann ${}_B\mathbb{T}_S = \left(\begin{smallmatrix}1 & -2\\ 2 & 1\end{smallmatrix}\right)^{-1} = \frac{1}{5}\left(\begin{smallmatrix}1 & 2\\ -2 & 1\end{smallmatrix}\right)$. Die Abbildungsmatrix von f bezüglich der „günstigen“ Basis B lässt sich also aus der gegebenen „ungünstigen“ Basis berechnen durch $\frac{1}{5}\left(\begin{smallmatrix}1 & 2\\ -2 & 1\end{smallmatrix}\right) \cdot \left(\begin{smallmatrix}1 & 2\\ 2 & 4\end{smallmatrix}\right) \cdot \left(\begin{smallmatrix}1 & -2\\ 2 & 1\end{smallmatrix}\right) = \left(\begin{smallmatrix}5 & 0\\ 0 & 0\end{smallmatrix}\right)$.

Aufgabe 4.29
Seien V, W endlichdimensionale $\mathbb{K}$-Vektorräume mit Basen B, C, wobei $\dim(V) = n$ und $\dim(W) = m$. Zeigen Sie ${}_CM_B \colon \mathrm{Hom}_{\mathbb{K}}(V, W) \xrightarrow{\cong} \mathbb{K}^{m\times n}$.

4.6 Basisauswahl und -ergänzung

Wiederum mit dem Lemma von Zorn könnte man zeigen, dass man aus jedem Erzeugendensystem eines Vektorraums eine Basis auswählen kann und dass man jede linear unabhängige Familie zu einer Basis ergänzen kann. Für Untervektorräume des $\mathbb{K}^m$ kann man die genannten Probleme auch rechnerisch lösen. Diese rechnerische Lösung überträgt sich auch auf allgemeine endlichdimensionale Vektorräume, sofern man jeweils einen Isomorphismus auf einen $\mathbb{K}^m$ kennt.

Problem 4.30 (Basisauswahl für Untervektorräume von $\mathbb{K}^m$)
Sei $[\vec{v}_1, ..., \vec{v}_n] \subset \mathbb{K}^m$ und $V := \mathrm{Span}_{\mathbb{K}}(\vec{v}_1, ..., \vec{v}_n)$. Sei $[\vec{v}_1, ..., \vec{v}_k]$ linear unabhängig (ggf. $k = 0$). Berechne eine Auswahl von $[\vec{v}_1, ..., \vec{v}_n]$, die $[\vec{v}_1, ..., \vec{v}_k]$ umfasst und eine Basis von V bildet.

Lösung:
Bringe $A := (\vec{v}_1, ..., \vec{v}_n) \in \mathbb{K}^{m\times n}$ mittels Gauß-Elimination auf Zeilenstufenform A' mit Pivotspalten $j_1, ..., j_r$. Dann ist $[\vec{v}_{j_1}, ..., \vec{v}_{j_r}]$ die gesuchte Basis.

Beweis: Jedes $\vec{b} \in V$ hat genau eine Darstellung als Linearkombination von $[\vec{v}_{j_1}, ..., \vec{v}_{j_r}]$, nämlich die, bei der alle freien Variablen (die ja als Koeffizienten vor den Nicht-Pivotspalten stehen) null sind. Die entstehende Basis umfasst $[\vec{v}_1, ..., \vec{v}_k]$, denn weil dies linear unabhängig ist, gibt es in den ersten k Spalten von A' keine Nichtpivotspalten. □

Beispiel: Die Vektoren $\vec{v}_1 = \begin{pmatrix}1\\2\\1\end{pmatrix}$, $\vec{v}_2 = \begin{pmatrix}-1\\1\\0\end{pmatrix}$, $\vec{v}_3 = \begin{pmatrix}2\\1\\1\end{pmatrix}$ und $\vec{v}_4 = \begin{pmatrix}2\\2\\2\end{pmatrix}$ bilden ein Erzeugendensystem von $\mathbb{R}^3$, die ersten beiden Vektoren sind linear unabhängig. $\begin{pmatrix}1&-1&2&2\\2&1&1&2\\1&0&1&2\end{pmatrix} \rightsquigarrow \begin{pmatrix}1&-1&2&2\\0&3&-3&-2\\0&0&0&2/3\end{pmatrix}$. Also bildet $[\vec{v}_1, \vec{v}_2, \vec{v}_4]$ eine Basis von $\mathbb{R}^3$.

Problem 4.31 (Basisergänzung für Untervektorräume von $\mathbb{K}^m$)
Sei $[\vec{w}_1, ...\vec{w}_\ell] \subset \mathbb{K}^n$ ein Erzeugendensystem von $V \leq \mathbb{K}^n$ und sei $[\vec{v}_1, ..., \vec{v}_k]$ linear unabhängig. Ergänze $[\vec{v}_1, ..., \vec{v}_k]$ zu einer Basis von V.

Lösung:
Die obige Lösung des Basisauswahlproblems angewandt auf das größere Erzeugendensystem $[\vec{v}_1, ..., \vec{v}_k, \vec{w}_1, ...\vec{w}_\ell]$ liefert auch die Lösung des Basisergänzungsproblems, denn wegen der linearen Unabhängigkeit von $[\vec{v}_1, ..., \vec{v}_k]$ ist jede der ersten k Spalten der entstehenden Zeilenstufenform eine Pivotspalte. □

Beispiel: Sei $\vec{v}_1 = \begin{pmatrix}1\\2\\3\end{pmatrix}$, $\vec{v}_2 = \begin{pmatrix}1\\4\\6\end{pmatrix}$. Ergänze $[\vec{v}_1, \vec{v}_2]$ zu einer Basis von $\mathbb{R}^3$:
$\begin{pmatrix}1&1&1&0&0\\2&4&0&1&0\\3&6&0&0&1\end{pmatrix} \rightsquigarrow \begin{pmatrix}1&1&1&0&0\\0&2&-2&1&0\\0&0&0&-3/2&1\end{pmatrix}$, also ist $[\vec{v}_1, \vec{v}_2, \vec{e}_2]$ eine Basis von $\mathbb{R}^3$.

Bemerkung 4.32
Mit Hilfe der Koordinatenabbildung bzw. ihrer Inversen erkennt man, dass der Basisauswahl- und der Basisergänzungssatz nicht nur für $\mathbb{K}^n$, sondern für jeden endlichdimensionalen Vektorraum gelten.

Korollar 4.33. *Sei V ein n-dimensionaler Vektorraum.*

a) Jedes Erzeugendensystem von V hat Länge $\geq n$. Jedes Erzeugendensystem der Länge n ist eine Basis.

b) Jedes linear unabhängige System in V hat Länge $\leq n$. Jedes linear unabhängige System der Länge n ist eine Basis.

Beweis:
a): Auswahlsatz: Länge n nach Auswahl einer Basis.

b): Basisergänzungssatz: Länge n nach Fortsetzung zu einer Basis. □

Satz 4.34
Sei V ein endlich dimensionaler $\mathbb{K}$-Vektorraum, $U \leq V$. Auch U endlich dimensional, $\dim(U) \leq \dim(V)$. *Wenn* $\dim(U) = \dim(V)$, *dann $U = V$.*

Bemerkung Auch für Gruppen betrachtet man Unterstrukturen, so genannte *Untergruppen*, sowie Erzeugendensysteme. Untergruppen einer Gruppe mit endlichem Erzeugendensystem haben aber nicht notwendig ebenfalls ein endliches Erzeugendensystem. Dass obiger Satz gilt, ist also alles andere als selbstverständlich.

Beweis:
Setze $n = \dim(V)$, und sei $[u_1, \ldots, u_r] \subset U$ linear unabhängig. Nach Korollar 4.33.b) für V ist $r \leq n$. Wir wählen nun r so groß wie möglich. In diesem Fall ist $[u_1, \ldots, u_r]$ ein maximales linear unabhängiges System in U, daher eine Basis von U (Charakterisierung von Basen). Insbesondere ist es ein endliches Erzeugendensystem für U, und $\dim(U) = r \leq n$. Ist $r = n$, dann ist $u_1, \ldots, u_r$ eine Basis von V, nach Korollar 4.33, also $U = V$. □

Sei V ein nicht notwendig endlichdimensionaler $\mathbb{K}$-Vektorraum.

Definition 4.35
Seien $U, W \leq V$. Die **Summe** *von U, W ist $U + W := \{u + w \mid u \in U,\ w \in W\}$.*

Aufgabe 4.36
Seien $U, W \leq V$. Dann ist $U \cap W \leq V$, $U + W \leq V$, $U \leq U + W$ und $W \leq U + W$.

Dimensionsformel
Seien U, W zwei endlich dimensionale Untervektorräume von V. Dann ist $U + W$ endlichdimensional, und $\dim(U + W) + \dim(U \cap W) = \dim(U) + \dim(W)$.

Beispiel: Sind U, W zwei 3-dimensionale Untervektorräume von $V = \mathbb{R}^5$, dann ist $\dim(U \cap W) \geq 1$: denn wegen $U + W \subseteq \mathbb{R}^5$ ist $\dim(U + W) \leq 5$, also

$$\dim(U \cap W) = \dim(U) + \dim(W) - \dim(U + W) \geq 3 + 3 - 5 = 1\,.$$

Vor dem Beweis ist noch ein Lemma hilfreich.

Lemma 4.37
Sind B, C Basen von $U, W \leq V$, so ist $U \cap W = \{o\} \iff B \cup C$ ist linear unabhängig. Anmerkung: Bei der Vereinigung von Familien addieren sich die Vielfachheiten jedes Elements.

Beweis:
$\sum_{v \in B \cup C}{}' \lambda_v v = o \iff z := \sum_{v \in B}{}' \lambda_v v = \sum_{v \in C}{}' -\lambda_v\, v$. Weil z sowohl in $\mathrm{Span}_{\mathbb{K}}(B)$ als auch $\mathrm{Span}_{\mathbb{K}}(C)$ liegt, ist $z \in U \cap W$. Weil B, C jeweils linear unabhängig sind, ist $z \neq o \iff \exists v \in B \cup C\colon\ \lambda_v \neq 0$, und wegen $\sum_{v \in B \cup C}{}' \lambda_v v = o$ ist dies genau dann der Fall, wenn $B \cup C$ linear abhängig ist. □

Beweis der Dimensionsformel:
Sei A eine Basis von $U \cap W$. Basisergänzung: Es gibt $B, C \subset V$, so dass $A \cup B$ bzw. $A \cup C$ Basen von U bzw. W sind. Offenbar ist $\mathrm{Span}(A \cup B \cup C) = U + W$.

Wegen $\mathrm{Span}_{\mathbb{K}}(C) \subset W$ gilt $\mathrm{Span}_{\mathbb{K}}(A \cup B) \cap \mathrm{Span}_{\mathbb{K}}(C) \subset U \cap W \cap \mathrm{Span}_{\mathbb{K}}(C) = \mathrm{Span}_{\mathbb{K}}(A) \cap \mathrm{Span}_{\mathbb{K}}(C) = \{o\}$ nach dem vorigen Lemma, denn $A \cup C$ ist linear unabhängig. Wiederum nach dem vorigen Lemma folgt, dass $A \cup B \cup C$ linear unabhängig ist, insbesondere sind A, B, C ***paarweise disjunkt***[44]. Also ist es eine Basis von $U + W$. $\dim(U + W) = |A \cup B \cup C| \overset{\text{disj.}}{=} |A| + |B| + |C| = (|A| + |B|) + (|A| + |C|) - |A| = \dim(U) + \dim(W) - \dim(U \cap W)$. □

[44] Das heißt $A \cap B = B \cap C = A \cap C = \emptyset$.

5 Matrizen – Teil 2

5.1 Rang und Rangformel

In diesem Abschnitt sei wieder $\mathbb{K}$ ein Körper.

Definition 5.1 (und Lemma). *Sei $A \in \mathbb{K}^{m\times n}$.*

a) $\mathrm{SR}(A) := \{A\vec{c} \mid \vec{c} \in \mathbb{K}^{n\times 1}\} = \mathrm{Bild}(L_A)$, *heißt* **Spaltenraum** *von A.*

b) $\mathrm{ZR}(A) := \{\underline{c}A \mid \underline{c} \in \mathbb{K}^{1\times m}\}$ *heißt* **Zeilenraum** *von A ($\mathbb{K}$-Span der Zeilen).*

c) Bringt man A durch Gauß-Elimination auf ZSF A' mit r Pivotspalten, dann $\dim(\mathrm{ZR}(A)) = \dim(\mathrm{ZR}(A')) = r = \dim(\mathrm{SR}(A')) = \dim(\mathrm{SR}(A))$.
Wir definieren den **Rang** *von A als* $\mathrm{Rang}(A) := \dim(\mathrm{SR}(A))$.

d) $\mathrm{Rang}(A) = \mathrm{Rang}({}^{\mathrm{t}}A)$.

e) $\forall B \in M_n(\mathbb{K})\colon\ B \in GL_n(\mathbb{K}) \iff \mathrm{Rang}(B) = n$.

Beweis:

c) In der Lösung des Basisauswahlproblems sahen wir, dass diejenigen Spalten von A, die den Pivotspalten von A' entsprechen, eine Basis von $\mathrm{SR}(A)$ bilden. Und offenbar bilden die Pivotspalten von A' auch eine Basis von $\mathrm{SR}(A')$. Davon gibt es jeweils r Stück.

In Übungen zeigen Sie gerade $\mathrm{ZR}(A) = \mathrm{ZR}(A')$, und offenbar bilden die r von $\underline{0}$ verschiedenen Zeilen von A' eine Basis von $\mathrm{ZR}(A')$.

d) Offenbar entspricht $\mathrm{ZR}({}^{\mathrm{t}}A)$ dem $\mathrm{SR}(A)$ und umgekehrt.

e) Darauf läuft der bereits bekannte Invertierbarkeitstest hinaus. □

Es ist ein viel zu häufiger Fehler, die Definition des Rangs mit der nun folgenden Rangformel für Matrizen und diese wiederum mit der Rangformel für lineare Abbildungen zu verwechseln. Vermeiden Sie diese Fehler bitte in der Klausur.

Rangformel für Matrizen
Sei $A \in \mathbb{K}^{m\times n}$. Es gilt $\mathrm{Rang}(A) + \dim(\mathrm{LR}(A; \vec{0})) = n$.

Beweis:
Bringe A mit Gauß-Elimination auf ZSF A' mit r Pivotspalten. Den Nicht-Pivotspalten entsprechen die Basislösungen von $\mathrm{LR}(A;\vec{0})$. Also ist $\dim(\mathrm{LR}(A;\vec{0})) = n - r$, aber $r = \mathrm{Rang}(A') = \mathrm{Rang}(A)$. □

Aufgabe 5.2. *Seien $A \in \mathbb{K}^{m\times n}$, $\vec{b} \in \mathbb{K}^m$.*
Das lineare Gleichungssystem $A\vec{x} \stackrel{!}{=} \vec{b}$ ist lösbar $\iff \mathrm{Rang}(A) = \mathrm{Rang}((A|\vec{b}))$.

Lemma 5.3. *Für alle* $A \in \mathbb{K}^{\ell \times m}$ *und* $B \in \mathbb{K}^{m \times n}$ *gilt:*
$\mathrm{Rang}(AB) \leq \mathrm{Rang}(A)$ *und* $\mathrm{Rang}(AB) \leq \mathrm{Rang}(B)$.

Beweis:
Für alle $\vec{c} \in \mathbb{K}^n$ ist $B\vec{c} \in \mathbb{K}^m$, und wegen $(AB)\vec{c} = A(B\vec{c})$ ist $\mathrm{SR}(AB) \subset \mathrm{SR}(A)$, daher $\mathrm{Rang}(AB) \leq \mathrm{Rang}(A)$.

Eine Basis von $\mathrm{SR}(B)$ wird durch Multiplikation mit A (also Anwendung von L_A) auf ein Erzeugendensystem von $\mathrm{SR}(AB)$ abgebildet, aus dem dann eine Basis ausgewählt werden kann. Daher $\mathrm{Rang}(AB) \leq \mathrm{Rang}(B)$. □

Definition 5.4
Sei V *ein endlichdimensionaler und* W *ein beliebiger* $\mathbb{K}$*-Vektorraum. Der* **Rang** *von* $f \in \mathrm{Hom}_{\mathbb{K}}(V, W)$ *ist* $\mathrm{Rang}(f) := \dim(\mathrm{Bild}(f))$.

Rangformel für lineare Abbildungen
Sei $f \in \mathrm{Hom}_{\mathbb{K}}(V, W)$, V *ein endlichdimensionaler und* W *ein beliebiger* $\mathbb{K}$*-Vektorraum.* $\mathrm{Bild}(f)$ *ist endlichdimensional und* $\mathrm{Rang}(f) + \dim(\ker(f)) = \dim(V)$.

Beispiel: Als Übung können Sie zeigen: f ist injektiv $\iff \ker(f) = \{o\}$, und wenn auch W endlichdimensional ist, dann ist f surjektiv $\iff \mathrm{Rang}(f) = \dim(W)$. Eine typische Anwendung: Wenn $f\colon \mathbb{R}^7 \to \mathbb{R}^3$ linear ist, dann ist $\mathrm{Rang}(f) \leq 3$ (denn $\mathrm{Bild}(f) \leq \mathbb{R}^3$), also ist $\dim(\ker(f)) = 7 - \mathrm{Rang}(f) \geq 4$ und daher $\ker(f) \neq \{o\}$; folglich ist f nicht injektiv. Es gibt also keine bijektive lineare Abbildung von $\mathbb{R}^7$ nach $\mathbb{R}^3$, aber interessanterweise gibt es nicht-lineare bijektive Abbildungen $\mathbb{R}^m \to \mathbb{R}^n$ für alle $m, n \in \mathbb{N}^*$.

Beweis der Rangformel für lineare Abbildungen:
Sei $\tilde{f} := f\big|^{\mathrm{Bild}(f)}$. Offenbar ist $\tilde{f} \in \mathrm{Hom}_{\mathbb{K}}(V, \mathrm{Bild}(f))$ surjektiv. Ist also $B \subset V$ eine endliche Basis, dann ist $\tilde{f}(B) = f(B)$ ein endliches Erzeugendensystem von $\mathrm{Bild}(\tilde{f})$. Da Basen minimale Erzeugendensysteme sind, gibt es eine endliche Teilfamilie von $\tilde{f}(B)$, die Basis von $\mathrm{Bild}(\tilde{f})$ ist. Wegen $\mathrm{Bild}(f) = \mathrm{Bild}(\tilde{f})$ und $\ker(f) = \ker(\tilde{f})$ folgt die Rangformel für f aus der Rangformel für eine beliebige Darstellungsmatrix von $\tilde{f}$. □

5.2 Determinanten

Sei R ein kommutativer Ring. Ziel: Ordne jedem $A \in M_n(R)$ die *Determinante* $\det(A) \in R$ zu, so dass A invertierbar $\Leftrightarrow$ $\det(A)$ invertierbar in R. Für $\vec{v}_1, ..., \vec{v}_n \in \mathbb{R}^n$ kann man die Determinante geometrisch deuten: $|\det(\vec{v}_1, ..., \vec{v}_n)|$ ist das Volumen von $\{\sum_{i=1}^{n} \lambda_i \vec{v}_i \mid \lambda_1, ..., \lambda_n \in [0, 1]\}$, und das Vorzeichen von $\det(\vec{v}_1, ..., \vec{v}_n)$ bestimmt die „Händigkeit“ von $[\vec{v}_1, ..., \vec{v}_n]$.

Nach der Definition formulieren wir einige Eigenschaften von Determinanten, insbesondere verschiedene Berechnungsmethoden. Insgesamt muss ein großer Aufwand betrieben werden, um zu zeigen, dass die Determinante überhaupt existiert.

Definition 5.5. *Sei R ein kommutativer Ring und $n \in \mathbb{N}$.*
Eine Abbildung $\det\colon M_n(R) \to R$ *heißt* **Determinantenfunktion** *(kurz: „Determinante“), wenn sie folgende Eigenschaften hat:*

a) Sie ist **multilinear***, das heißt: Wenn $i \in \{1, ..., n\}$ und wenn sich $A, B, C \in M_n(R)$ nur in Zeile i unterscheiden, wobei $C_{i,*} = \lambda A_{i,*} + \mu B_{i,*}$, dann* $\det(C) = \lambda \det(A) + \mu \det(B)$.

b) Sie ist **alternierend***: Wenn $A \in M_n(R)$ zwei gleiche Zeilen hat, dann* $\det(A) = 0$.

c) Sie ist **normiert***:* $\det(\mathbb{1}_n) = 1$.

Für $n \geq 2$ ist auch die Notation $|A| := \det(A)$ üblich.

Vorsicht: $\det(\lambda A) = \lambda^n \det(A)$ und im Allg. $\det(A + B) \neq \det(A) + \det(B)$. Für $A = (a) \in M_1(R)$ ist $|A| = a$ — bitte **NICHT** $|A| = |a|$!!

Satz 5.6. *Sei* $\det\colon M_n(R) \to R$ *eine Determinantenfunktion und $A \in M_n(R)$.*

a) $\det$ **schiefsymmetrisch** *(auch: antisymmetrisch), das heißt: Wenn A' aus A durch Tausch zweier Zeilen $i \neq j$ entsteht, dann* $\det(A') = -\det(A)$.

b) Sei $\lambda \in R$. Entsteht A' aus A, indem Zeile i durch ihr λ-Faches ersetzt wird, dann $\det(A') = \lambda \det(A)$.

c) Sei $\gamma \in R$. Entsteht A' aus A, indem das γ-Fache der Zeile i zur Zeile $j \neq i$ addiert wird, dann ist $\det(A') = \det(A)$.

Beweis:

a) Sei $a := A_{i,*}$ und $b := A_{j,*}$. Dann (die Notation ist hoffentlich klar)

$$0 \overset{\text{alt.}}{=} \det\begin{pmatrix} \cdots \\ a+b \\ \cdots \\ a+b \\ \cdots \end{pmatrix} \overset{\text{lin.}}{=} \det\begin{pmatrix} \cdots \\ a \\ \cdots \\ a+b \\ \cdots \end{pmatrix} + \det\begin{pmatrix} \cdots \\ b \\ \cdots \\ a+b \\ \cdots \end{pmatrix}$$
$$\overset{\text{lin.}}{=} \det\begin{pmatrix} \cdots \\ a \\ \cdots \\ a \\ \cdots \end{pmatrix} + \det\begin{pmatrix} \cdots \\ a \\ \cdots \\ b \\ \cdots \end{pmatrix} + \det\begin{pmatrix} \cdots \\ b \\ \cdots \\ a \\ \cdots \end{pmatrix} + \det\begin{pmatrix} \cdots \\ b \\ \cdots \\ b \\ \cdots \end{pmatrix} \overset{\text{alt.}}{=} \det\begin{pmatrix} \cdots \\ a \\ \cdots \\ b \\ \cdots \end{pmatrix} + \det\begin{pmatrix} \cdots \\ b \\ \cdots \\ a \\ \cdots \end{pmatrix}$$

b) Spezialfall von „Linear in Zeile i“.

c) $$\det\begin{pmatrix} \cdots \\ a \\ \cdots \\ b+\gamma a \\ \cdots \end{pmatrix} \overset{\text{lin.}}{=} \det\begin{pmatrix} \cdots \\ a \\ \cdots \\ b \\ \cdots \end{pmatrix} + \gamma \det\begin{pmatrix} \cdots \\ a \\ \cdots \\ a \\ \cdots \end{pmatrix} \overset{\text{alt.}}{=} \det\begin{pmatrix} \cdots \\ a \\ \cdots \\ b \\ \cdots \end{pmatrix}.$$ □

Wenn also R ein Körper ist, können wir die Berechnung der Determinante mittels des Gauß-Algorithmus auf Determinanten von Zeilenstufenformen zurückführen.

Definition 5.7

a) $A \in M_n(R)$ *heißt* **obere Dreiecksmatrix** $:\Leftrightarrow \forall i,j \in \{1,...,n\}\colon (i > j \Rightarrow A_{i,j} = 0)$. $A \in M_n(R)$ *heißt* **untere Dreiecksmatrix** $:\Leftrightarrow \forall i,j \in \{1,...,n\}\colon (i < j \Rightarrow A_{i,j} = 0)$. *Eine Dreiecksmatrix* $A \in M_n(R)$ *heißt* **strikt**, *wenn zudem* $\forall i \in \{1,...,n\}\colon A_{i,i} = 0$.

b) *Ist* $A \in M_n(R)$ *mit* $A = \left(\begin{smallmatrix} B & C \\ 0 & D \end{smallmatrix}\right) \in M_n(\mathbb{K})$ *oder* $A = \left(\begin{smallmatrix} B & 0 \\ C & D \end{smallmatrix}\right)$ *mit quadratischen Matrizen* B, D, *so hat* A **Blockgestalt**.

c) *Für* $A \in M_n(R)$ *und* $i,j \in \{1,...,n\}$ *sei* $A(i,j) \in M_{n-1}(R)$ *die Matrix, die aus* A *durch Entfernung der* i*-ten Zeile und der* j*-ten Spalte entsteht.*

Beispiel: $\begin{pmatrix} 1&2&3&4&5\\2&3&4&5&6\\0&0&1&2&3\\0&0&2&3&4\\0&0&3&4&5 \end{pmatrix}$ hat Blockgestalt, $\begin{pmatrix} 1&2&3&4\\0&5&6&7\\0&0&0&8\\0&0&0&9 \end{pmatrix}$ ist obere Dreiecksmatrix.

Theorem 5.8. *Sei* R *ein kommutativer Ring,* $n \in \mathbb{N}^*$ *und* $A \in M_n(R)$.

a) $f\colon M_n(R) \to R$ *multilinear* $\Rightarrow f(A) = \sum_{j_1=1}^{n} ... \sum_{j_n=1}^{n} (\prod_{i=1}^{n} A_{i,j_i}) f \begin{pmatrix} \underline{e}_{j_1} \\ ... \\ \underline{e}_{j_n} \end{pmatrix}$ *mit* $\underline{e}_j := {}^{t}\vec{e}_j \in R^n$.

b) $\forall n \in \mathbb{N}^*$*: Es gibt höchstens eine Determinantenfunktion* $\det\colon M_n(R) \to R$.

c) *Im Spezialfall* $n = 1$ *gilt* $\forall (a) \in M_1(R)\colon \det((a)) = a$.

d) *Im Spezialfall* $n \in \{2,3\}$ *gilt die* **Sarrus-Regel**[45]*:* $\left|\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right| = ad - bc$ *und* $\left|\begin{smallmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{smallmatrix}\right| = a_1b_2c_3 + a_2b_3c_1 + a_3b_1c_2 - a_3b_2c_1 - a_1b_3c_2 - a_2b_1c_3$.

e) *Durch c) und* $\det(A) := \sum_{i=1}^{n} (-1)^{i+j} A_{i,j} \det(A(i,j))$ *für* $n \geq 2$ *(***Laplace-Entwicklung**[46] *nach Spalte* j*) ist rekursiv eine Determinantenfunktion definiert. Wegen b) gibt es also genau eine Determinantenfunktion. Auch gilt die Laplace-Entwicklung nach Zeile* i, $\det(A) = \sum_{j=1}^{n} (-1)^{i+j} A_{i,j} \det(A(i,j))$. *Die Determinante einer Untermatrix von* A *bezeichnet man als* **Minore**. *Merkregel für die Vorzeichen: Schachbrett-Muster* $\begin{pmatrix} + & - & + & - & \cdots \\ - & + & - & & \\ + & - & & & \\ - & & & & \\ \vdots & & & & \end{pmatrix}$.

f) *Hat* $A = \left(\begin{smallmatrix} B & C \\ 0 & D \end{smallmatrix}\right)$ *Blockgestalt, dann* $\det(A) = \det(B) \cdot \det(D)$.

g) *Ist* A *eine obere oder untere Dreiecksmatrix, dann* $\det(A) = \prod_{i=1}^{n} A_{i,i}$.

[45] Pierre Frédéric Sarrus [1798–1861]
[46] Pierre-Simon Marquis de Laplace [1749–1827]

h) **Multiplikationssatz***:* $\forall B \in M_n(R)\colon\ \det(AB) = \det(A)\det(B)$.

i) $\det({}^tA) = \det(A)$. *Satz 5.6 gilt analog für Spaltenoperationen.*

Bemerkung a) **Wichtig:** Wer in einer Prüfung ein Diagonalschema zur Determinantenberechnung für $n \geq 4$ einsetzt, erhält erhebliche Punktabzüge, denn ein solches Schema wäre im Allgemeinen falsch! Die Sarrus-Regel für $n = 3$ ist oft weniger effizient als andere Rechenwege. Aber die Sarrus-Regel für $n = 2$ ist wirklich nützlich.

b) Ist R ein Körper, ist es sinnvoll, $\det(A)$ zu berechnen, indem man A mit Zeilen- und Spaltenoperationen auf Blockgestalt bringt und dann die Block- und Dreiecksgestalt nutzt. Dieser Rechenweg hat Komplexität $O(n^3)$, während das sture Beharren auf der Laplace-Entwicklung Komplexität $O(n!)$ hat. Laplace-Entwicklung ist dann nützlich, wenn in einer Zeile bzw. Spalte die meisten Einträge null sind. Die Wahl effizienter Rechenwege erfordert Erfahrung — rechnen Sie also viele Beispiele!

Beispiel: Effizientes Rechnen erfordert die Kombination von Rechentechniken.

$$\begin{aligned}
\begin{vmatrix} 1 & 2 & 3 & 0 & 4 \\ 5 & 6 & 7 & 0 & 8 \\ \pi & e & 13 & 3 & -1 \\ 2 & 6 & 4 & 0 & 8 \\ 3 & 1 & 1 & 0 & 2 \end{vmatrix} &= -3 \cdot \begin{vmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 2 & 6 & 4 & 8 \\ 3 & 1 & 1 & 2 \end{vmatrix} && \text{Laplace 4. Spalte} \\
&= -3 \cdot 2 \cdot \begin{vmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 1 & 3 & 2 & 4 \\ 3 & 1 & 1 & 2 \end{vmatrix} && \text{Skalierung 3. Zeile} \\
&= -6 \cdot \begin{vmatrix} 1 & 2 & 3 & 4 \\ 0 & -4 & -8 & -12 \\ 0 & 1 & -1 & 0 \\ 0 & -5 & -8 & -10 \end{vmatrix} && \text{Gauß-Elimination} \\
&= 6 \cdot \begin{vmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & -1 & 0 \\ 0 & -4 & -8 & -12 \\ 0 & -5 & -8 & -10 \end{vmatrix} && \text{Zeilentausch} \\
&= 6 \cdot \begin{vmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & -12 & -12 \\ 0 & 0 & -13 & -10 \end{vmatrix} && \text{Gauß-Elimination} \\
&= 6 \cdot 1 \cdot 1 \cdot \begin{vmatrix} -12 & -12 \\ -13 & -10 \end{vmatrix} && \text{Block-/Dreiecksgestalt} \\
&= 6 \cdot (12 \cdot 10 - 12 \cdot 13) = 6 \cdot 12 \cdot (-3) = -216 && \text{Sarrus-Regel } (2 \times 2)
\end{aligned}$$

Zum Beweis des Theorems benötigen wir noch etwas mehr Gruppentheorie:

Definition 5.9
Für eine Menge Ω *sei* $\mathrm{Sym}(\Omega) := \{f\colon \Omega \to \Omega \mid f \text{ ist bijektiv}\}$. $(\mathrm{Sym}(\Omega), \circ, \mathrm{Id}_\Omega)$ *ist eine Gruppe (wurde bereits in Übungen gezeigt), die* **symmetrische Gruppe** *von* Ω. *Für* $n \in \mathbb{N}$ *sei* $S_n := \mathrm{Sym}(\{1, ..., n\})$.
Die Elemente von $\mathrm{Sym}(\Omega)$ *nennt man auch* **Permutationen** *von* Ω. *Eine Permutation* $\tau \in \mathrm{Sym}(\Omega)$ *heißt* **Transposition**, *falls es* $i \neq j \in \Omega$ *gibt, so dass* $\tau(i) = j$, $\tau(j) = i$ *und* $\tau(k) = k$ *für alle* $k \in \Omega \setminus \{i, j\}$.

Aufgabe 5.10
Jedes $\sigma \in S_n$ ist gleich einer Verknüpfung von endlich vielen Transpositionen.

Notation 5.11. *Sei $\sigma \in S_n$ und $A \in M_n(R)$.*
$A^\sigma \in M_n(R)$ ist durch $\forall i \in \{1, ..., n\}\colon\ A^\sigma_{i,*} = A_{\sigma(i),*}$ definiert.

Beobachtung 5.12
A^σ entsteht aus A, indem jeweils die i-te Zeile durch die $\sigma(i)$-te ersetzt wird. Für $\rho, \sigma \in S_n$ und $A, B \in M_n(R)$ gilt $(AB)^\sigma = (A^\sigma)B$, denn $((AB)^\sigma)_{i,*} = (AB)_{\sigma(i),*} = A_{\sigma(i),*}B = A^\sigma_{i,*}B = (A^\sigma B)_{i,*}$, und $A^{\rho\circ\sigma} = (A^\rho)^\sigma$, denn $A^{\rho\circ\sigma}_{i,*} = A_{\rho(\sigma(i)),*} = A^\rho_{\sigma(i),*} = (A^\rho)^\sigma_{i,*}$.

Beweis Thm. 5.8:

a) $f(A) \overset{\text{linear}}{=} \sum_{j_1=1}^{n} A_{1,j_1} f\begin{pmatrix} \underline{e}_{j_1} \\ A_{2,*} \\ \dots \end{pmatrix} \overset{\text{Ind.}}{=} \sum_{j_1=1}^{n} A_{1,j_1} \sum_{j_2=1}^{n} \dots \sum_{j_n=1}^{n} (\prod_{i=2}^{n} A_{i,j_i}) f\begin{pmatrix} \underline{e}_{j_1} \\ \dots \\ \underline{e}_{j_n} \end{pmatrix}$.
Zusatz: Ist f zudem alternierend, so tragen zur Summe nur solche Summanden bei, bei denen die $j_1, ..., j_n$ paarweise verschieden sind, für die also $\sigma\colon \{1, ..., n\} \to \{1, ..., n\}$ mit $\sigma(i) := j_i$ eine Permutation ist. Dann: $f(A) = \sum_{\sigma \in S_n} (\prod_{i=1}^{n} A_{i,\sigma(i)}) f(\mathbb{1}^\sigma_n)$.

b) Seien det und D Determinantenfunktionen. Nach dem Zusatz von a) genügt zu zeigen: $\forall \sigma \in S_n\colon\ \det(\mathbb{1}^\sigma_n) = D(\mathbb{1}^\sigma_n)$. Sei σ eine Verknüpfung von k Transpositionen; das geht, nach Übung 5.10. Weil det und D schiefsymmetrisch und normiert sind, folgt $\det(\mathbb{1}^\sigma_n) = (-1)^k = D(\mathbb{1}^\sigma_n)$.

c) Multilinear und normiert: $\det((a)) = a \det(\mathbb{1}_1) = a \cdot 1$.

d) $n = 2$: $\left|\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right| \overset{\text{a)}}{=} ad \left|\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right| + bc \left|\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right| \overset{\text{schiefs.}}{=} ac \left|\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right| - bc \left|\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right| \overset{\text{norm.}}{=} ac - bc$.
$n = 3$ überlasse ich Ihnen zur Übung.

e) $M_1(R) \ni A \mapsto A_{1,1}$ normiert, multilinear und (wegen $n = 1$) auch alternierend. Induktionsannahme für $n > 1$: $\det\colon M_{n-1}(R) \to R$ ist eine Determinantenfunktion. Behauptung: $\det\colon M_n(R) \to R$ mit $\det(A) := \sum_{i=1}^{n} (-1)^{i+j} A_{i,j} \det(A(i,j))$ ist eine Determinantenfunktion.

Normiert: Für $A = \mathbb{1}_n$ und $i \neq j$ ist $A_{i,j} = 0$. Außerdem $A(i,i) = \mathbb{1}_{n-1}$. Daher $\det(\mathbb{1}_n) = 1 \cdot \det(\mathbb{1}_{n-1}) = 1$.

Multilinear in Zeile k: Sei $A_{k,*} = \lambda B_{k,*} + \mu C_{k,*}$ und in den anderen Zeilen seien $A, B, C \in M_n(R)$ gleich. Sei $i \in \{1, ..., n\}$. Wenn $k = i$, dann $A(i,j) = B(i,j) = C(i,j)$, denn die einzige Zeile, in der sich die drei Matrizen unterscheiden, wird entfernt. Wenn $k \neq i$, dann

$\det(A(i,j)) = \lambda\det(B(i,j)) + \mu\det(C(i,j))$ nach Induktionsannahme, und $A_{i,j} = B_{i,j} = C_{i,j}$. Daher

$$\begin{aligned}\det(A) &= (-1)^{k+j}A_{k,j}\det(A(k,j)) + \sum_{i\neq k}(-1)^{i+j}A_{i,j}\det(A(i,j))\\ &= \lambda(-1)^{k+j}B_{k,j}\det(B(k,j)) + \mu(-1)^{k+j}C_{k,j}\det(C(k,j))\\ &\quad + \sum_{i\neq k}(-1)^{k+j}A_{i,j}\left(\lambda\det(B(i,j)) + \mu\det(C(i,j))\right)\\ &= \lambda\sum_{i=1}^{n}(-1)^{i+j}B_{i,j}\det(B(i,j)) + \mu\sum_{i=1}^{n}(-1)^{i+j}C_{i,j}\det(C(i,j))\\ &= \lambda\det(B) + \mu\det(C)\end{aligned}$$

Alternierend: Für $k < \ell$ sei $A_{k,*} = A_{\ell,*}$. Ist $k \neq i \neq \ell$, dann hat $A(i,j)$ zwei gleiche Zeilen, also $\det(A(i,j)) = 0$ nach Induktionsannahme. Zwischen der k-ten und der ℓ-ten Zeile leigen $\ell - k - 1$ Zeilen. Also entsteht $A(k,j)$, indem man die k-te Zeile von $A(\ell,j)$ durch $\ell - k - 1$ Vertauschungen in die ℓ-te Zeile bringt. Also $A_{k,j}\det(A(k,j)) = (-1)^{\ell-k-1}A_{\ell,j}\det(A(\ell,j))$ und

$$\begin{aligned}\det(A) &= \sum_{i=1}^{n}(-1)^{i+j}A_{i,j}\det(A(i,j))\\ &= (-1)^{k+j}A_{k,j}\det(A(k,j)) + (-1)^{\ell+j}A_{\ell,j}\det(A(\ell,j))\\ &= ((-1)^{k+j+\ell-k-1} + (-1)^{\ell+j})A_{\ell,j}\det(A(\ell,j)) = 0\end{aligned}$$

Prinzipiell könnte man zeigen, dass auch die Laplace-Entwicklung nach Zeile i eine rekursive Definition einer Determinantenfunktion erlaubt, doch das ist technisch aufwändig. Es folgt leichter aus i) und Laplace-Entwicklung nach Spalte i.

f) Übung: Induktion nach der Zeilen-/Spaltenzahl von B. Im Induktionsschritt wird Laplace-Entwicklung nach Spalte 1 verwendet.

g) Offensichtlich, durch mehrfache Anwendung von f).

h) Sei $D\colon M_n(R) \to R$ definiert durch $D(A) := \det(AB)$. Sind zwei Zeilen von A gleich, dann sind auch die entsprechenden Zeilen von AB gleich, also ist D alternierend. Übung: D ist multilinear.

$\det(AB) = D(A) \overset{\text{a)}}{=} \sum_{\sigma\in S_n}(\prod_{i=1}^{n}A_{i,\sigma(i)})D(\mathbb{1}_n^\sigma) = \sum_{\sigma\in S_n}(\prod_{i=1}^{n}A_{i,\sigma(i)})\det(\mathbb{1}_n^\sigma B)$ $\overset{5.12}{=} \sum_{\sigma\in S_n}(\prod_{i=1}^{n}A_{i,\sigma(i)})\det(B^\sigma) = \sum_{\sigma\in S_n}(\prod_{i=1}^{n}A_{i,\sigma(i)})\det(\mathbb{1}_n^\sigma)\det(B) \overset{\text{a)}}{=}$ $\det(A)\det(B)$. Die vorletzte Gleichung folgt, da σ als ein Produkt von k Transpositionen darstellbar ist und wegen Schiefsymmetrie $\det(B^\sigma) = (-1)^k\det(B) = \det(\mathbb{1}_n^\sigma)\det(B)$ gilt.

i) Ist $\sigma \in S_n$ eine Verknüpfung von k Transpositionen, dann auch σ^{-1}. Daher $\det(\mathbb{1}_n^\sigma) = (-1)^k = \det(\mathbb{1}_n^{\sigma^{-1}})$. Jedes $\sigma \in S_n$ hat genau ein Inverses in S_n, also $\det({}^tA) = \sum_{\sigma \in S_n} (\prod_{i=1}^n {}^tA_{i,\sigma(i)}) \det(\mathbb{1}_n^\sigma) = \sum_{\sigma \in S_n} (\prod_{i=1}^n A_{\sigma(i),i}) \det(\mathbb{1}_n^\sigma)$ $= \sum_{\sigma^{-1} \in S_n} (\prod_{i=1}^n A_{i,\sigma^{-1}(i)}) \det(\mathbb{1}_n^{\sigma^{-1}}) = \det(A)$. □

Bemerkung 5.13 (und Definition)
Für $\sigma \in S_n$ heißt $\operatorname{sgn}(\sigma) := \det(\sigma(A))$ *das* **Vorzeichen** *von σ. Wenn man σ durch k Transpositionen darstellen kann, dann ist* $\operatorname{sgn}(\sigma) = (-1)^k$*, doch ohne den Umweg über die Determinante wäre nicht klar, dass das wohldefiniert ist. Es ist allerdings auch möglich,* $\operatorname{sgn}(\sigma)$ *rein gruppentheoretisch zu definieren. Es handelt sich um einen Homomorphismus* $\operatorname{sgn}\colon S_n \to \{\pm 1\}$ *von Gruppen, nämlich* $\forall \sigma, \varphi \in S_n\colon \operatorname{sgn}(\sigma \circ \varphi) = \operatorname{sgn}(\sigma) \cdot \operatorname{sgn}(\varphi)$*, denn wenn man σ durch k und φ durch ℓ Transpositionen darstellen kann, kann man $\sigma \circ \varphi$ durch $k + \ell$ Transpositionen darstellen, und* $\operatorname{sgn}(\sigma \circ \varphi) = (-1)^{k+\ell} = (-1)^k \cdot (-1)^\ell$*. Man erhält die* **Leibnizformel**

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^n A_{i,\sigma(i)}\,.$$

5.2.1 Invertierbarkeit

Definition 5.14. *Die* **Adjunkte**[47] *von $A \in M_n(R)$ ist* $\operatorname{adj}(A) \in M_n(R)$ *mit* $\operatorname{adj}(A)_{j,i} = (-1)^{i+j} \det(A(i,j))$.

Lemma 5.15
$\forall A \in M_n(R)\colon\ A \cdot \operatorname{adj}(A) = \operatorname{adj}(A) \cdot A = \det(A) \cdot \mathbb{1}_n$.

Beweis:

- Sei $B := A \cdot \operatorname{adj}(A)$. Dann $B_{i,k} = \sum_{j=1}^n A_{i,j} \cdot (-1)^{k+j} \det(A(k,j))$. Daraus folgt zunächst $B_{i,i} = \det(A)$ wegen der Laplace-Entwicklung nach Zeile i. Ist $i \neq k$, so entstehe $\tilde{A}$ aus A, indem man Zeile k durch eine Kopie von Zeile i ersetzt. Dann ist $B_{i,k} = \sum_{j=1}^n \tilde{A}_{k,j} \cdot (-1)^{k+j} \det(\tilde{A}(k,j)) = \det(\tilde{A}) = 0$, wieder wegen Laplace-Entwicklung.
- Dass auch $\operatorname{adj}(A) \cdot A = \det(A) \cdot \mathbb{1}_n$ gilt, kann man mit Spaltenentwicklung nachrechnen. □

Korollar 5.16
$A \in M_n(R)$ ist invertierbar in $M_n(R) \iff \det(A) \in R^\times$. Dann gilt $\det(A^{-1}) = (\det(A))^{-1}$ *und* $A^{-1} = (\det(A))^{-1} \operatorname{adj}(A)$.
$GL_n(R) := \{A \in M_n(R) \mid \det(A) \in R^\times\}$ ist eine Gruppe bzgl. Matrixmultiplikation.

[47] Bitte nicht „Adjungierte“, das wäre etwas anderes.

Beweis:
Ist A invertierbar, dann wegen Produktregel $\det(A) \cdot \det(A^{-1}) = \det(A \cdot A^{-1}) = \det(\mathbb{1}_n) = 1 = \det(A^{-1} \cdot A) = \det(A^{-1}) \cdot \det(A)$, also $\det(A)$ invertierbar in R und zudem $(\det(A))^{-1} = \det(A^{-1})$.
Ist $\det(A)$ in R invertierbar, dann $A^{-1} = (\det(A))^{-1} \cdot \mathrm{adj}(A)$ nach Lemma 5.15. Es bleibt zu zeigen: $A, B \in GL_n(R) \Rightarrow AB \in GL_n(R)$. Wenn aber $\det(A)$ und $\det(B)$ in R invertierbar sind, dann ist auch $\det(AB) = \det(A)\det(B)$ invertierbar, mit Inversem $\det(B)^{-1}\det(A)^{-1}$. □

Beispiel: Für $A := \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in M_2(R)$ mit $\det(A) \neq 0$ ist $A^{-1} = \frac{1}{\det(A)} \left(\begin{smallmatrix} d & -b \\ -c & a \end{smallmatrix}\right)$, denn $\mathrm{adj}(A)_{1,1} = (-1)^2 \det(A(1,1)) = d$, $\mathrm{adj}(A)_{1,2} = (-1)^3 \det(A(2,1)) = -b$, $\mathrm{adj}(A)_{2,1} = (-1)^3 \det(A(1,2)) = -c$, $\mathrm{adj}(A)_{2,2} = (-1)^4 \det(A(2,2)) = a$.

Zur Illustration leite ich eine explizite Formel für die eindeutige Lösung des Gleichungssystems $A \cdot \vec{x} = \vec{b}$ mit einer invertierbaren Matrix A her. Diese ist nach Gabriel Cramer [1704–1752] benannt, der sie 1750 beschrieb. Gottfried Wilhelm Leibniz [1646–1716] war sie sogar schon 1678 bekannt, aber einen Beweis der Formel lieferte erst 1815 Augustin Louis Cauchy [1789–1857].

Cramersche Regel
Sei $A \in GL_n(R)$ und $\vec{b} \in R^n$. Für $i \in \{1, ..., n\}$ sei $A_i \in M_n(R)$ die Matrix, die durch Ersetzung der Spalte i von A durch $\vec{b}$ entsteht. Für die eindeutige Lösung $\vec{x} \in R^n$ von $A \cdot \vec{x} = \vec{b}$ gilt $\forall i \in \{1, ..., n\}$: $x_i := \frac{\det(A_i)}{\det(A)}$.

Beispiel: Für $A := \left(\begin{smallmatrix} 1 & 2 \\ 4 & 5 \end{smallmatrix}\right)$ und $\vec{b} = \left(\begin{smallmatrix} 3 \\ 6 \end{smallmatrix}\right)$ hat $A \cdot \left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right) = \vec{b}$ die eindeutige Lösung

$$x_1 = \frac{\left|\begin{smallmatrix} 3 & 2 \\ 6 & 5 \end{smallmatrix}\right|}{\left|\begin{smallmatrix} 1 & 2 \\ 4 & 5 \end{smallmatrix}\right|} = \frac{3}{-3} = -1 \qquad x_2 = \frac{\left|\begin{smallmatrix} 1 & 3 \\ 4 & 6 \end{smallmatrix}\right|}{\left|\begin{smallmatrix} 1 & 2 \\ 4 & 5 \end{smallmatrix}\right|} = \frac{-6}{-3} = 2$$

Beweis der Cramerschen Regel:
Seien $\vec{a}_1, ..., \vec{a}_n \in R^n$ die Spalten von A. Aus der eindeutigen Lösung des Gleichungssystems ergibt sich $x_1 \cdot \vec{a}_1 + ... + 1 \cdot (x_i\vec{a}_i - \vec{b}) + ... + x_n \cdot \vec{a}_n = \vec{0}$, d.h. wir können durch Addition einer Linearkombination der anderen Spalten zur i-ten Spalte von C (wodurch sich die Determinante nicht ändert!) eine Nullspalte erreichen. Daher und wegen Linearität in der i-ten Spalte:

$$0 = \det(C) = x_i \det(A) - \det(\vec{a}_1, ..., \vec{a}_{i-1}, \vec{b}, \vec{a}_{i+1}, ..., \vec{a}_n). \qquad \square$$

Ich empfehle die Verwendung der Cramerschen Regel ausdrücklich *nicht*. Erstens nämlich nutzt sie nichts, wenn A nicht invertierbar ist. Und vor allem ist der Rechenaufwand unsinnig groß.

Bemerkung 5.17 (Vergleich von Gauß und Cramer)
a) Zur Lösung des obigen Problems mit Gauß-Elimination sind ungefähr n^3 Rechenschritte erforderlich.

b) Für die Anwendung der Cramerschen Regel muss man $n+1$ *Determinanten von ausrechnen, was mit dem Gauß-Algorithmus jeweils wieder* $\frac{2}{3}n^3$ *Rechenschritte erfordert. Das sind insgesamt mehr als* $\frac{2}{3}n^4 \gg n^3$.

c) Wenn man den Gauß-Algorithmus wirklich nicht mag und daher die Determinanten mit der Leibnizformel berechnet, braucht man jeweils etwa $n!$ *Operationen, also insgesamt* $(n+1)!$.

20 *Gleichungen mit* 20 *Unbekannten lassen sich auch ohne einen Computer noch in vertretbarer Zeit mit dem Gauß-Algorithmus in ca.* 8000 *Schritten lösen. Mit Cramer und Leibniz wären es ca.* $5.1 \cdot 10^{19}$. *Selbst mit einem Computer, der* 10^8 *Schritte pro Sekunde ausführen kann, würde dies* $5.1 \cdot 10^{11}$ *Sekunden dauern — rund* 16200 *Jahre!*

6 Eigenwertprobleme

In diesem Abschnitt sei $\mathbb{K}$ wieder ein Körper. Sei V ein $\mathbb{K}$-Vektorraum, $\dim(V) = n \in \mathbb{N}$. Wir wissen, dass man nach Wahl einer Basis B von V jedem $\varphi \in \mathrm{End}_{\mathbb{K}}(V)$ die Darstellungsmatrix ${}_B M_B(\varphi) \in M_n(\mathbb{K})$ zuordnen kann, so dass die Anwendung des Endomorphismus der Multiplikation mit der Darstellungsmatrix entspricht. Insbesondere lassen sich Eigenschaften von φ (etwa Injektivität) durch Betrachtung von ${}_B M_B(\varphi)$ untersuchen.

Welche Eigenschaften von ${}_B M_B(\varphi)$ hängen nur von φ, aber nicht von B ab? Wenn $A, A' \in M_n(\mathbb{K})$ gegeben sind, wie kann man entscheiden, ob es Basen B, C von V und ein $\varphi \in \mathrm{End}_{\mathbb{K}}(V)$ gibt, so dass $A = {}_B M_B(\varphi)$ und $A' = {}_C M_C(\varphi)$? Wie sich zeigte, hilft hierfür die Untersuchung von Untervektorräumen $U \leq V$ mit $\varphi(U) \subset U$, so genannten ***φ-invarianten Untervektorräumen***.

6.1 Eigenwerte, -vektoren und -räume

In voller Allgemeinheit wäre die Lösung der eingangs genannten Probleme Thema einer weiterführenden Vorlesung über lineare Algebra (siehe Frobenius-Normalform bzw. Jordan-Normalform). Hier wollen wir uns auf besonders einfache invariante Untervektorräume beschränken. Ist nämlich $U \leq V$ eindimensional und $[b]$ eine Basis von U, so gilt $\varphi(U) \subset U \iff \exists \lambda \in \mathbb{K}\colon \varphi(b) = \lambda b$, und dies führt unmittelbar auf den folgenden Begriff:

Definition 6.1
Sei V ein $\mathbb{K}$-Vektorraum und $\varphi \in \mathrm{End}_{\mathbb{K}}(V)$.

a) Sei $\lambda \in \mathbb{K}$. $E_\lambda(\varphi) = \{v \in V \mid \varphi(v) = \lambda v\}$ ist ein **Eigenraum** *von φ.*

b) $v \in V$ heißt **Eigenvektor** *zum* **Eigenwert** *$\lambda \in \mathbb{K}$ von $\varphi :\Leftrightarrow v \in E_\lambda(\varphi) \backslash \{o\}$.*

c) Für $A \in M_n(\mathbb{K})$ und $\lambda \in \mathbb{K}$ heißt $E_\lambda(A) := E_\lambda(L_A) = \{\vec{v} \in \mathbb{K}^n \mid A\vec{v} = \lambda\vec{v}\}$ der Eigenraum von A. Eigenvektoren bzw. Eigenwerte von A sind gleich denen von L_A.

Der Nullvektor ist niemals ein Eigenvektor!!
Zwar ist stets $o \in E_\lambda(A)$, aber λ ist nur Eigenwert, falls $E_\lambda(A) \neq \{o\}$.

Bemerkung 6.2

a) $v \in V$ ist genau dann Eigenvektor von φ, wenn $\mathrm{Span}_{\mathbb{K}}(v)$ ein eindimensionaler φ-invarianter UVR ist. $E_\lambda(\varphi)$ ist ebenfalls φ-invariant, allerdings ggf. höherdimensional.

b) Ein **zyklischer** *φ-invarianter Unterraum ist von der Form $\mathrm{Span}_{\mathbb{K}}(\{\varphi^k(v) \mid k \in \mathbb{N}\})$ für ein $v \in V$. Er kann aber bisweilen noch als Summe kleinerer invarianter Untervektorräume zerlegt werden.*

c) *Sei* $\varphi\colon \mathscr{C}^\infty(\mathbb{R}) \to \mathscr{C}^\infty(\mathbb{R})$ *definiert durch* $\varphi(f) := f''$ *für* ∞*-oft differenzierbare Funktionen* $f\colon \mathbb{R} \to \mathbb{R}$*. Jedes* $\lambda \in \mathbb{R}_{\geq 0}$ *ist Eigenwert von* φ*, nämlich mit Eigenvektor* $f\colon x \mapsto \mathrm{e}^{\sqrt{\lambda}x}$*. Auch jedes* $\lambda \in \mathbb{R}_{<0}$ *ist Eigewert von* φ*, nämlich mit Eigenvektor* $f\colon x \mapsto \cos(\sqrt{|\lambda|}x)$*. Tatsächlich standen derartige Beispiele am Anfang der Betrachtung von Eigenvektoren.*

d) **Hauptkomponentenanalyse** *zur Analyse experimenteller statistischer Daten: Aus den Daten wird zunächst die* Kovarianzmatrix *berechnet. An den Eigenvektoren und den zugehörigen Eigenwerten erkennt man, durch welche Einflussgrößen die Daten bestimmt werden (Hauptkomponenten). Kleine Eigenwerte der Kovarianzmatrix entsprechen statistischem Rauschen. Das lässt sich auch in der Datenkompression nutzen.*

e) *Resonanzfrequenzen berechnet man durch Eigenwertprobleme.*

f) *Die Eigenvektoren des Trägheitstensors sind die Rotationsachsen, bezüglich denen ein Körper keine Unwucht hat.*

g) *Quantenmechanischen Systeme werden mit dem* **Hamilton-Operator** *beschrieben. Dessen Eigenwerte sind die möglichen Energiezustände und lassen sich im Spektrum beobachten. Manchmal erlauben Zusatzannahmen, die Quantenzustände angenähert als* endlich*dimensionale Eigenwertprobleme zu modellieren.*

h) *Anfänglich verwendete Google bei der Bewertung von Suchergebnissen den PageRank-Algorithmus. Webseiten und ihre gegenseitige Verlinkung werden in einer gigantischen Matrix kodiert. Man berechnet den Eigenvektor zum größten Eigenwert dieser Matrix. Die Webseite, die zum größten Koeffizienten dieses Eigenvektors gehört, ist das „beste“ Suchergebnis.*

Mit den hier behandelten Methoden kann man kleinere Beispiele per Hand berechnen. Für große oder gar unendlichdimensionale Eigenwertprobleme benötigt man Methoden der Numerik oder der Analysis.

Definition 6.3 (und Lemma). *Sei* $A \in M_n(\mathbb{K})$ *und* $\lambda \in \mathbb{K}$*.*

a) $X\mathbb{1}_n - A \in M_n(\mathbb{K}[X])$ *heißt* **charakteristische Matrix** *von* A*.* $E_\lambda(A) = \mathrm{LR}(\lambda\mathbb{1}_n - A; \vec{0})$*.*

b) *Das* **charakteristische Polynom** *von* A *ist* $\chi_A(X) := \det(X\mathbb{1}_n - A) \in \mathbb{K}[X]$*.* ***Beweis:*** *Leibnizformel.*

c) $\lambda \in \mathbb{K}$ *ist Eigenwert von* $A \iff E_\lambda(A) \neq \{\vec{0}\} \iff \chi_A(\lambda) = 0$ *(***Säkulargleichung***).* ***Beweis:*** $A\vec{x} = \lambda\vec{x} \iff (\lambda\mathbb{1}_n - A)\vec{x} = \vec{0}$*, und dies hat genau dann eine Lösung* $\vec{x} \neq 0$*, wenn* $\lambda\mathbb{1}_n - A$ *nicht invertierbar ist, was man an* $\chi_A(\lambda) = 0$ *erkennt.*

Beispiele: a) Für $A := \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ ist $\chi_A(X) = \left|\begin{smallmatrix} X & 1 \\ -1 & X \end{smallmatrix}\right| = X^2+1$. Ist $\mathbb{K} = \mathbb{R}$, so hat A keine Eigenwerte. Ist $\mathbb{K} = \mathbb{C}$, so hat A die Eigenwerte i mit Eigenvektor $\left(\begin{smallmatrix} -1 \\ \mathrm{i} \end{smallmatrix}\right)$ und $-\mathrm{i}$ mit Eigenvektor $\left(\begin{smallmatrix} 1 \\ \mathrm{i} \end{smallmatrix}\right)$. Die Eigenwerte hängen also vom betrachteten Körper ab.

b) Für $A := \left(\begin{smallmatrix} 2 & 1 \\ 0 & 2 \end{smallmatrix}\right)$ ist $\chi_A(X) = (X-2)^2$ (wegen Dreiecksgestalt). $E_2(A) = \mathrm{LR}(\left(\begin{smallmatrix} 0 & -1 \\ 0 & 0 \end{smallmatrix}\right); \vec{0}) = \mathrm{Span}_{\mathbb{K}}(\vec{e}_1)$.

Bemerkung So wie bei uns wird das charakteristische Polynom in den Büchern von unter anderem Gramlich, Koecher, Lang und Lipschutz definiert. Mit dieser Definition ist $\chi_A(X)$ ***normiert***, d.h. der Koeffizient des führenden Terms ist 1.

Sehr oft findet man aber auch $\chi_A(X) = \det(A - X\mathbb{1}_n)$ in der Literatur. Das unterscheidet sich von der hier verwendeten Definition um das Vorzeichen $(-1)^n$ und hat aus meiner Sicht nur die Rechtfertigung, dass Vorzeichenfehler beim Hinschreiben von $X\mathbb{1}_n - A$ etwas wahrscheinlicher als bei $A - X\mathbb{1}_n$ sind.

Definition 6.4. *Die* **Spur** *von $A \in M_n(\mathbb{K})$ ist* $\mathrm{Spur}(A) := \sum_{i=1}^n A_{i,i}$.

Beispiel: $\mathrm{Spur}\left(\begin{smallmatrix} 2 & -1 \\ 3 & -2 \end{smallmatrix}\right) = 2 + (-2) = 0$.

Lemma 6.5. *Für $A \in M_n(\mathbb{K})$ hat $\chi_A(X)$ die Gestalt*

$$\chi_A(X) = X^n - \mathrm{Spur}(A)X^{n-1} + (\textit{Terme vom Grad } n-2 \geq r \geq 1) + (-1)^n \det(A)\,.$$

Insbesondere hat A höchstens n verschiedene Eigenwerte.

Beweis:
Sei $B := X\mathbb{1}_n - A$. $\chi_A(X) = \det(B) = \sum\limits_{\sigma \in S_n} \mathrm{sgn}(\sigma) \prod\limits_{i=1}^n B_{i,\sigma(i)} = \prod\limits_{i=1}^n (X - A_{i,i}) + \sum\limits_{\substack{\sigma \in S_n \\ \sigma \neq \mathrm{Id}}} \mathrm{sgn}(\sigma) \prod\limits_{i=1}^n B_{i,\sigma(i)}$. Ist $\sigma \neq \mathrm{Id}$, dann $\exists k \neq \ell \in \{1, ..., n\}\colon\ \sigma(k) \neq k$ und $\sigma(\ell) \neq \ell$. Weil X nur auf der Hauptdiagonale von B vorkommt, ist $\prod\limits_{i=1}^n B_{i,\sigma(i)}$ vom Grad $\leq n-2$. Zudem ist $\prod\limits_{i=1}^n (X - A_{i,i}) = X^n - \mathrm{Spur}(A)X^{n-1} + \text{Terme kleinerer Grade}$. Zudem $\det(A) = \chi_A(0) = \det(\mathbb{0} - A) = (-1)^n \det(A)$. □

Beispiel 6.6

a) Für $A \in M_2(\mathbb{K})$ gilt $\chi_A(X) = X^2 - \mathrm{Spur}(A) \cdot X + \det(A)$ — ***eine nützliche Formel, die man sich merken sollte!***

Für $A = \left(\begin{smallmatrix} 2 & -1 \\ 3 & -2 \end{smallmatrix}\right)$ ist $\chi_A(X) = X^2 - 4 + 3 = X^2 - 1 = (X-1)(X+1)$. Also sind 1 und -1 die einzigen Eigenwerte. Berechnung der Eigenvektoren:

$\boxed{\lambda = 1}$ $B := \mathbb{1}_2 - A = \left(\begin{smallmatrix} -1 & 1 \\ -3 & 3 \end{smallmatrix}\right)$, $E_1(A) = \mathrm{LR}(B;\ \vec{0}) = \mathrm{Span}_{\mathbb{R}}(\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right))$.

$\boxed{\lambda = -1}$ $B := -\mathbb{1}_2 - A = \left(\begin{smallmatrix} -3 & 1 \\ -3 & 1 \end{smallmatrix}\right)$, $E_{-1}(A) = \mathrm{LR}(B;\ \vec{0}) = \mathrm{Span}_{\mathbb{R}}(\left(\begin{smallmatrix} 1/3 \\ 1 \end{smallmatrix}\right))$.

b) Für $A = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \in M_3(\mathbb{R})$ *ist*

$$\chi_A(X) = \begin{vmatrix} X & -1 & 0 \\ 1 & X & -1 \\ 0 & 1 & X \end{vmatrix} = X \begin{vmatrix} X & -1 \\ 1 & X \end{vmatrix} + |\begin{smallmatrix} 1 & -1 \\ 0 & X \end{smallmatrix}| = X(X^2+1) + X = X(X^2+2)$$

Der einzige reelle Eigenwert ist 0*, mit* $E_0(A) = \mathrm{LR}(A;\ \vec{0}) = \mathrm{Span}_{\mathbb{R}}(\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix})$. *In* $\mathbb{C}$ *gibt es noch die Eigenwerte* $\pm\sqrt{2}\,\mathrm{i}$.

c) Bei der Berechnung von $\det(X\mathbb{1} - A)$ *sollte man kreativ sein und Brüche von Polynomen möglichst vermeiden. Sei* $A = \begin{pmatrix} -3 & -6 & 22 & -4 \\ -6 & -7 & 28 & -6 \\ -2 & -2 & 7 & -2 \\ 6 & 10 & -38 & 7 \end{pmatrix}$. *Dann ist*

$$\begin{aligned}
\chi_A(X) &= \begin{vmatrix} X+3 & 6 & -22 & 4 \\ 6 & X+7 & -28 & 6 \\ 2 & 2 & X-7 & 2 \\ -6 & -10 & 38 & X-7 \end{vmatrix} = \begin{vmatrix} X+3 & 3-X & -22 & 1-X \\ 6 & X+1 & -28 & 0 \\ 2 & 0 & X-7 & 0 \\ -6 & -4 & 38 & X-1 \end{vmatrix} \\
&= \begin{vmatrix} X+3 & -X+3 & -22 & -X+1 \\ 6 & X+1 & -28 & 0 \\ 2 & 0 & X-7 & 0 \\ X-3 & -X-1 & 16 & 0 \end{vmatrix} = (X-1) \cdot \begin{vmatrix} 6 & X+1 & -28 \\ 2 & 0 & X-7 \\ X+3 & 0 & -12 \end{vmatrix} \\
&= -(X-1) \cdot (X+1) \cdot \begin{vmatrix} 2 & X-7 \\ X+3 & -12 \end{vmatrix} \\
&= -(X-1) \cdot (X+1) \cdot (-X^2 + 4X - 3) \\
&= (X-1) \cdot (X+1) \cdot (X-1) \cdot (X-3)
\end{aligned}$$

Die Eigenwerte von A sind also $1, -1, 3$. *Für jeden Eigenwert muss man nun noch eine Basis für den Eigenraum berechnen. Ergebnisse:*

$$\begin{aligned}
E_1(A) &= \mathrm{LR}(\mathbb{1}_4 - A;\ \vec{0}) = \mathrm{Span}_{\mathbb{R}}(\begin{pmatrix} -2 \\ 5 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}) \\
E_{-1}(A) &= \mathrm{LR}(-\mathbb{1}_4 - A;\ \vec{0}) = \mathrm{Span}_{\mathbb{R}}(\begin{pmatrix} -1 \\ -1 \\ 0 \\ 2 \end{pmatrix}) \\
E_3 &= \mathrm{LR}(3 \cdot \mathbb{1}_4 - A;\ \vec{0}) = \mathrm{Span}_{\mathbb{R}}(\begin{pmatrix} -3 \\ -4 \\ -1 \\ 5 \end{pmatrix})
\end{aligned}$$

Rechnung für $E_1(A)$*:* $\mathbb{1}_4 - A = \begin{pmatrix} 4 & 6 & -22 & 4 \\ 6 & 8 & -28 & 6 \\ 2 & 2 & -6 & 2 \\ -6 & -10 & 38 & -6 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 4 & 6 & -22 & 4 \\ 0 & -1 & 5 & 0 \\ 0 & -1 & 5 & 0 \\ 0 & -1 & 5 & 0 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 4 & 6 & -22 & 4 \\ 0 & -1 & 5 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$, *und* $E_1(A) = \mathrm{LR}(\mathbb{1}_4 - A;\ \vec{0})$ *wird aufgespannt von den beiden Basislösungen* $\vec{\beta}_3 = \begin{pmatrix} -2 \\ 5 \\ 1 \\ 0 \end{pmatrix}$ *und* $\vec{\beta}_4 = \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$.

Guter Rat: Behalten Sie die während der Berechnung von $\chi_A(X)$ gefundenen Faktoren bei. Also *NICHT* ausmultiplizieren, wenn es sich vermeiden lässt! Grund: Die Eigenwerte kann man an den Faktoren leicht ablesen.

Der Hintergrund dieses Ratschlags ist, dass zwar über den komplexen Zahlen jedes Polynom in Linearfaktoren zerfällt, dass es aber im Allgemeinen beweisbar unmöglich ist, diese Linearfaktoren durch eine Lösungsformel zu berechnen. Seien Sie also froh, wenn Sie eine Nullstelle gefunden haben!

Hauptsatz der Algebra
Jedes Polynom $p \in \mathbb{C}[X]$ mit $\deg(p) = n \in \mathbb{N}^*$ und führendem Term X^n kann man vollständig in n Linearfaktoren zerlegen.

Die Nullstellenmenge $N_p := \{\lambda \in \mathbb{C} \mid p(\lambda) = 0\}$ ist endlich und für jede Nullstelle $\lambda \in N_p$ gibt es ein $d_\lambda \in \mathbb{N}^*$ (die **Vielfachheit** *der Nullstelle λ von p), so dass $p(X) = \prod_{\lambda \in N_p} (X - \lambda)^{d_\lambda}$, insbesondere $\sum_{\lambda \in N_p} d_\lambda = n$ und $|N_p| \leq n$.*

Definition 6.7
Sei $A \in M_n(\mathbb{K})$ und $\lambda \in \mathbb{K}$ ein Eigenwert von A.

a) Die Vielfachheit von λ als Nullstelle von $\chi_A(X)$ heißt **algebraische Vielfachheit** *von λ.*

b) $\dim_{\mathbb{K}}(E_\lambda(A))$ heißt **geometrische Vielfachheit** *von λ.*

In der Schule lernten Sie Lösungsformeln für quadratische Gleichungen, jedoch nicht die Formeln von Gerolamo Cardano [1501–1576], Scipione del Ferro [1465–1526], Niccolò Fontana Tartaglia [1500–1557] und Lodovico Ferrari [1522–1565] zur Lösung algebraischer Gleichungen vom Grad 3 und 4. Jedoch gibt es darüber hinaus keine allgemeingültigen Lösungsformeln:

Satz von Abel-Ruffini
Die Nullstellen eines Polynoms vom Grad ≥ 5 lassen sich im Allgemeinen nicht *durch Grundrechenarten und Wurzelziehen berechnen. Dies gilt zum Beispiel für die Nullstelle $x = -1.1673...$ von $x^5 - x + 1$.*

Paolo Ruffini [1765–1822] hatte dafür 1799 einen lückenhaften Beweis. Niels Henrik Abel [1802–1829] gelang 1824 der erste vollständige Beweis. Heute wird der Satz (z.B. in Algebra–Vorlesungen) im Rahmen der *Galois-Theorie*[48] bewiesen.

Berechnen Sie Eigenwerte immer exakt! Wäre λ nur *näherungsweise* Eigenwert von $A \in M_n(\mathbb{K})$, so wäre $E_\lambda(A) = \{\vec{0}\}$: Man fände keinen Eigenvektor, auch nicht näherungsweise! In numerischen Lösungen des Eigenwertproblems werden daher die Eigenwerte und die Eigenvektoren *gleichzeitig* approximiert.

In einem Dokument in Moodle finden Sie praktische Tipps zur Nullstellenberechnung von Polynomen, die manche von Ihnen vermutlich bereits aus der Schule (ggf. mit anderen Begriffsbildungen) kennen.

6.2 Eigenräume sind komplementär

Lemma 6.8
Seien $\lambda_1, \ldots, \lambda_r$ paarweise verschiedene Eigenwerte von $A \in M_n(\mathbb{K})$.

[48] Das gesamte Werk von Évariste Galois [1811–1832] umfasst nur rund 60 Seiten, doch er begründet darin unabhängig von Abel die Gruppentheorie, gibt mit der „Galois-Theorie“ notwendige und hinreichende Bedingungen dafür, welche algebraischen Gleichungen eine Lösung durch Grundrechenarten und Wurzelziehen besitzen, und konstruiert alle endlichen Körper.

a) Sind $\vec{u}_i \in E_{\lambda_i}(A)\setminus\{\vec{0}\}$ für $i = 1, ..., r$, dann ist $[\vec{u}_1, ..., \vec{u}_r]$ linear unabhängig.

b) Die Vereinigung von Basen von $E_{\lambda_1}(A), \ldots, E_{\lambda_r}(A)$ ist linear unabhängig. Insbesondere ist $\sum_{i=1}^{r} \dim_{\mathbb{K}}(E_{\lambda_i}(A)) \leq n$.

Beweis:
b) folgt aus a).
a): Sei $\sum_{i=1}^{r} \alpha_i \vec{u}_i = \vec{0}$ mit $\alpha_1, ..., \alpha_r \in \mathbb{K}$. Zu zeigen: $\alpha_1 = ... = \alpha_r = 0$.
Induktion über r. Klar für $r = 1$. Ist $r \geq 2$, dann

$$\vec{0} = A\left(\sum_{i=1}^{r} \alpha_i \vec{u}_i\right) - \lambda_r \sum_{i=1}^{r} \alpha_i \vec{u}_i = \sum_{i=1}^{r} (\lambda_i - \lambda_r)\alpha_i \vec{u}_i = \sum_{i=1}^{r-1} (\lambda_i - \lambda_r)\alpha_i \vec{u}_i \,.$$

Also (Induktionsannahme) $(\lambda_i - \lambda_r)\alpha_i = 0$ für alle $i < r$. Wegen $\lambda_i \neq \lambda_r$ für $i < r$ folgt daraus $\alpha_i = 0$ für alle $i < r$. Dann folgt auch $\vec{0} = \sum_{i=1}^{r} \alpha_i \vec{u}_i = \alpha_r \vec{u}_r$ und damit $\alpha_r = 0$ wegen $\vec{u}_r \neq \vec{0}$. □

6.3 Basiswechsel

Sind B, C Basen eines endlichdimensionalen $\mathbb{K}$-Vektorraums V, so haben wir gemäß Beobachtung 4.28 Basiswechselmatrizen ${}_C\mathbb{T}_B$ und ${}_B\mathbb{T}_C = ({}_C\mathbb{T}_B)^{-1}$.

Lemma 6.9
Sei $A \in M_n(\mathbb{K})$. Sei $F\colon \mathbb{K}^n \to \mathbb{K}^n$ der Endomorphismus $F = L_A$. Für eine Matrix $A' \in M_n(\mathbb{K})$ sind dann die folgenden beiden Aussagen äquivalent:

a) Es gibt $S \in GL_n(\mathbb{K})$ mit $A' = S^{-1}AS$.

b) Es gibt eine Basis B des $\mathbb{K}^n$ derart, dass $A' = {}_BM_B(F)$ ist.

Konkret ist der Zusammenhang zwischen S und B wie folgt: $S = {}_EM_B(\mathrm{Id}) = {}_E\mathbb{T}_B$, wobei E die Standardbasis ist; B besteht aus den Spalten von S.

Beweis:
$F = L_A$ heißt $A = {}_EM_E(F)$, für E die Standardbasis. Sei $A' = S^{-1}AS$. $\mathrm{Rang}(S) = n$, also bilden die Spalten von S eine Basis B von $\mathbb{K}^n$. Dann ist $S = {}_EM_B(\mathrm{Id}) = {}_E\mathbb{T}_B$ und $S^{-1} = {}_B\mathbb{T}_E$, also $A' = {}_B\mathbb{T}_E\,{}_EM_E(F)\,{}_E\mathbb{T}_B = {}_BM_B(F)$. Ist dagegen $A' = {}_BM_B(F)$, dann ist mit $S = {}_E\mathbb{T}_B$ auch $A' = S^{-1}\,{}_EM_E(F)S = S^{-1}AS$. □

Beispiel: $A = \left(\begin{smallmatrix} 2 & -1 \\ 3 & -2 \end{smallmatrix}\right)$ hat Eigenvektoren $\vec{v}_1 = \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$ und $\vec{v}_2 = \left(\begin{smallmatrix} 1 \\ 3 \end{smallmatrix}\right)$ mit Eigenwert 1 bzw. -1. Nun, $B := [\vec{v}_1, \vec{v}_2]$ ist eine Basis von $\mathbb{R}^2$. Wegen $F(\vec{v}_1) = A \cdot \vec{v}_1 = 1 \cdot \vec{v}_1$ und $F(\vec{v}_2) = (-1)\vec{v}_2$ ist ${}_BM_B(F) = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. Ferner ist ${}_E\mathbb{T}_B = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 3 \end{smallmatrix}\right)$. Nach dem Lemma gilt $\left(\begin{smallmatrix} 1 & 1 \\ 1 & 3 \end{smallmatrix}\right)^{-1} \left(\begin{smallmatrix} 2 & -1 \\ 3 & -2 \end{smallmatrix}\right) \left(\begin{smallmatrix} 1 & 1 \\ 1 & 3 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. Wir haben A *diagonalisiert*. Zur Kontrolle *kann* man die Gleichung direkt prüfen.

Lemma 6.10 (und Definition)
$A \in M_n(\mathbb{K})$ heißt **diagonalisierbar** $:\Leftrightarrow \exists S \in GL_n(\mathbb{K})$*: $S^{-1}AS$ ist diagonal. Ggf. nennt man S eine* **diagonalisierende Matrix** *für A. Dies ist gleichbedeutend zu jeder der folgenden Charakterisierungen:*

a) $\mathbb{K}^n$ hat eine Basis $B = [\vec{b}_1, ..., \vec{b}_n]$, so dass ${}_BM_B(L_A)$ Diagonalmatrix ist.

b) $\mathbb{K}^n$ hat eine Basis, die aus Eigenvektoren $\vec{b}_1, ..., \vec{b}_n$ von A besteht.

c) Mit $N := \{\lambda \in \mathbb{K} \mid \chi_A(\lambda) = 0\}$ gilt $\sum_{\lambda \in N} \dim_{\mathbb{K}}(E_\lambda(A)) = n$.

Beweis:
Def $\Leftrightarrow$ a): Lemma 6.9, $S = (\vec{b}_1, ..., \vec{b}_n)$.

a) $\Leftrightarrow$ b): ${}_BM_B(L_A)$ ist genau dann diagonal, wenn B aus Eigenvektoren besteht.

b) $\Leftrightarrow$ c): Lemma 6.8 und Basisergänzung. □

Problem 6.11 (Diagonalisierbarkeit)
Gegeben $A \in M_n(\mathbb{K})$, gesucht eine diagonalisierende Matrix, falls sie existiert.

Lösung:
Berechne $N := \{\lambda \in \mathbb{K} \mid \chi_A(\lambda) = 0\}$ und berechne für jedes $\lambda \in N$ eine Basis von $E_\lambda(A)$. Wenn man nicht insgesamt n Basisvektoren findet, ist A nicht diagonalisierbar. Andernfalls bilden die Basen der Eigenräume zusammen eine Basis $[\vec{b}_1, ..., \vec{b}_n]$ von $\mathbb{K}^n$. Mit $S := (\vec{b}_1, ..., \vec{b}_n)$ ist $S^{-1}AS$ diagonal, das i-te Diagonalelement ist ist der Eigenwert von $\vec{v}_i$. □

Beispiel: Die Matrix aus Bsp. 6.6.c) ist diagonalisierbar. Mit $S := \begin{pmatrix} -2 & -1 & -1 & -3 \\ 5 & 0 & -1 & -4 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 2 & 5 \end{pmatrix}$ ist $S^{-1} \cdot A \cdot S = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$. Die Diagonalmatrix lässt sich ohne Rechnung anhand der Eigenwerte angeben.

Beispiel: $A = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 2 \end{smallmatrix}\right) \in M_2(\mathbb{R})$ hat Eigenvektoren $\left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$ mit Eigenwert 1 und $\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$ mit Eigenwert 2, also diagonalisierbar. Mit $S = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ ist $S^{-1}AS = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 2 \end{smallmatrix}\right)$.

Beispiel: $A = \left(\begin{smallmatrix} 3 & -1 \\ 4 & -1 \end{smallmatrix}\right) \in M_2(\mathbb{R})$ ist nicht diagonalisierbar, denn $\chi_A(X) = X^2 - 2X + 1 = (X-1)^2$, d.h. 1 ist der einzige Eigenwert und $E_1(A) = \mathrm{Span}_{\mathbb{R}}(\left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right))$.

Ist nicht explizit nach einer diagonalisierenden Matrix gefragt, kann man sich bei der Prüfung der Diagonalisierbarkeit manchmal etwas Arbeit sparen. Dafür brauchen wir eine kleine Vorbereitung:

Aufgabe 6.12
Für $A' = S^{-1}AS$ gilt $\chi_{A'}(X) = \chi_A(X)$.

Korollar 6.13
Ist $\dim_{\mathbb{K}}(E_\lambda(A)) = r$*, dann ist* $\chi_A(X)$ *durch* $(X-\lambda)^r$ *teilbar. Für jeden Eigenwert ist also die geometrische kleiner oder gleich der algebraischen Vielfachheit.*

Beweis:
Sei $[\vec{v}_1, \ldots, \vec{v}_r]$ eine Basis von $E_\lambda(A)$. Basisergänzungssatz: Setze zu einer Basis $B = [\vec{v}_1, \ldots, \vec{v}_n]$ von $\mathbb{K}^n$ fort. Wie in Lemma 6.9 sei $F = L_A$, $S = {}_E\mathbb{T}_B$ und $A' = {}_BM_B(F) = S^{-1}AS$. Nach Aufgabe 6.12 ist $\chi_{A'} = \chi_A$. Wegen $F(\vec{v}_i) = \lambda\vec{v}_i$ für $i \leq r$ hat A' Blockgestalt $\left(\begin{smallmatrix} \lambda\mathbb{1}_r & C \\ 0 & D \end{smallmatrix}\right)$, also $X\mathbb{1}_n - A' = \left(\begin{smallmatrix} (X-\lambda)\mathbb{1}_r & -C \\ 0 & X\mathbb{1}_s - D \end{smallmatrix}\right)$ für $s = n - r$, und nach der Blockmatrix-Regel ist $\chi_{A'}(X) = (X-\lambda)^r\chi_D(X)$. □

Spezialfälle von Problem 6.11, wenn nur nach der Existenz einer diagonalisierenden Matrix gefragt ist:

- Hat A n verschiedene Eigenwerte in $\mathbb{K}$, ist A diagonalisierbar, denn jeder Eigenraum ist mindestens eindimensional.
- Zerfällt $\chi_A(X)$ in $\mathbb{K}[X]$ nicht in Linearfaktoren, ist A nicht diagonalisierbar.
- Ist die algebraische Vielfachheit eines Eigenwertes größer als seine geometrische Vielfachheit, ist A nicht diagonalisierbar.

Beispiel: Für $A = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ ist $\chi_A(X) = X^2 + 1$. Also nicht diagonalisierbar über $\mathbb{R}$ (aber schon über $\mathbb{C}$).

Beispiel: Für $A = \left(\begin{smallmatrix} 1 & 5 & -2 \\ 0 & 4 & 7 \\ 0 & 0 & -1 \end{smallmatrix}\right)$ ist $\chi_A(X) = (X-1)(X-4)(X+1)$ (Dreiecksgestalt). Also drei verschiedene Eigenwerte $1, 4, -1$, daher diagonalisierbar.

Der folgende Satz wird hier nicht bewiesen. Mehr Hintergrund dazu wird es im Abschlusskapitel über euklidische Geometrie geben. Man beachte, dass der Satz nicht für $\mathbb{K} = \mathbb{Q}$ oder für endliche Körper gilt.

Satz 6.14 (und Definition)
$A \in M_n(\mathbb{K})$ *heißt* **symmetrisch**, *gdw.* $A^\top = A$. *Jede symmetrische Matrix* $A \in M_n(\mathbb{R})$ *ist diagonalisierbar.*

Es gilt sogar eine Verallgemeinerung für Matrizen über den komplexen Zahlen. Ist $A \in M_n(\mathbb{C})$ und $A = {}^{\mathrm{t}}\overline{A}$ (Erinnerung: $\overline{a + b\,\mathrm{i}} := a - b\,\mathrm{i}$ bezeichnet die komplexe Konjugation, und das soll auf jeden Eintrag von A angewandt und das Ergebnis transponiert werden), so heißt A ***hermitesch***[49]. Jede hermitesche Matrix ist diagonalisierbar und alle ihre Eigenwerte sind *reell* (obwohl ja die Matrixeinträge komplex sind!). Das hat wichtige Anwendungen in der Physik, denn beobachtbare physikalische Größen in der Quantenmechanik entsprechen Eigenwerten hermitescher Operatoren — sind also reell, und das ist gut, denn ein nicht-reeller Messwert wäre seltsam.

[49] Nach Charles Hermite [1822–1901]

7 Euklidische Räume

In diesem Kapitel müssen wir in der Lage sein, aus positiven Körperelementen Wurzeln zu ziehen. Weil es in $\mathbb{C}$ keinen Begriff von „Positivität“ gibt, scheidet $\mathbb{C}$ aus. Weil man in $\mathbb{Q}$ nicht immer Wurzeln ziehen kann, scheidet $\mathbb{Q}$ aus. In diesem Kapitel sei daher $\mathbb{K} = \mathbb{R}$. In der Schule sollten Sie gelernt haben, wie man in der Anschauungsebene und im Anschauungsraum (mehr macht man ja heute nicht mehr) basierend auf „dem“ Skalarprodukt Längen und Winkel berechnet. Nun, es gibt tatsächlich mehr Skalarprodukte als Ihre Schulweisheit sich träumen lässt...

7.1 Skalarprodukte

In der Schule haben Sie vermutlich gelernt, dass man die Länge von $\vec{v} \in \mathbb{R}^n$ ($n \leq 3$) durch $\sqrt{\vec{v} \bullet \vec{v}}$ berechnet, mit dem „Skalarprodukt“ $\vec{v} \bullet \vec{w} = \sum_{i=1}^{n} v_i w_i$. Vielleicht haben Sie damit auch Winkel berechnet. Das wollen wir verallgemeinern. Wegen der Verwechslungsgefahr mit dem Matrixprodukt verwende ich für Skalarprodukte die in der Quantenmechanik übliche Notation $\langle v|w\rangle$.

Definition 7.1. *Sei V ein $\mathbb{R}$-Vektorraum.*

a) Eine **Bilinearform** *auf V ist eine multilineare Abbildung $b\colon V \times V \to \mathbb{R}$.*

b) Eine Bilinearform b auf V heißt...

i) **positiv definit** $:\Leftrightarrow \forall v \in V\colon \quad b(v, v) > 0 \iff v \neq 0$.

ii) **symmetrisch** $:\Leftrightarrow \forall v, w \in V\colon\ b(v, w) = b(w, v)$.

c) Ein **Skalarprodukt** *auf V ist eine positiv definite symmetrische Bilinearform. Ein Skalarprodukt notieren wir hier meist als $\langle v \mid w\rangle$ statt $b(v, w)$.*

d) Ein reeller Vektorraum V mit einem Skalarprodukt bezeichnet man als **Skalarproduktraum**[50]*. Ein* **euklidischer Raum** *ist ein endlichdimensionaler Skalarproduktraum.*

Beispiel 7.2 (und Definition)

a) Das **Standardskalarprodukt** *auf $\mathbb{R}^n$ ist durch $\langle \vec{v}|\vec{w}\rangle := \sum_{i=1}^{n} v_i w_i = {}^{\mathrm{t}}\vec{v}\vec{w}$ definiert. Es ist ein Skalarprodukt.*

b) Ist b eine Bilinearform auf einem endlich-dimensionalen $\mathbb{R}$-VR V und $B = [v_1, ..., v_n]$ eine Basis von V, so ist die **Darstellungsmatrix** ${}_B b \in M_n(\mathbb{R})$ *von b bzgl. B gegeben durch ${}_B b_{i,j} = b(v_i, v_j)$. Das Standardskalarprodukt ist der Fall ${}_E b = \mathbb{1}_n$ mit der Standardbasis E.*

[50] Auch: ***Prä-Hilbertraum***; David Hilbert [1862–1943]

Für $u, w \in V$ gilt $b(u,w) = \sum_{i,j=1}^{n} \kappa_B(u)_i b(v_i, v_j) \kappa_B(w)_j = {}^{\mathrm{t}}\kappa_B(u) \cdot {}_B b \cdot \kappa_B(w)$. Ist $A \in M_n(\mathbb{R})$ vorgegeben, so gibt es eine eindeutig bestimmte Bilinearform b auf V mit ${}_B b = A$, nämlich b definiert durch die obige Formel.

Offenbar ist b genau dann symmetrisch, wenn ${}_B b$ symmetrisch ist.

c) *Die durch $A = \begin{pmatrix} 1 & -2 \\ -2 & 1 \end{pmatrix}$ definierte symmetrische Bilinearform ist kein Skalarprodukt: Zwar ist $b(\vec{e}_1, \vec{e}_1) = b(\vec{e}_2, \vec{e}_2) = 1 > 0$, aber für $\vec{v} = \binom{1}{1}$ ist $b(\vec{v}, \vec{v}) = 1 - 2 - 2 + 1 = -2 < 0$.*

d) *Wir werden (hoffentlich) noch zeigen: Eine durch $A \in M_n(\mathbb{R})$ gegebene symmetrische Bilinearform auf $\mathbb{R}^n$ ist genau dann ein Skalarprodukt, wenn alle Eigenwerte von A positive reelle Zahlen sind.*

e) *Quantenmechanische Zustände werden durch eine so genannte* Wellenfunktion *$\varphi\colon \mathbb{R}^3 \to \mathbb{C}$ beschrieben. Für Wellenfunktionen φ, ψ definiert man $\langle \varphi \mid \psi \rangle := \int_{\mathbb{R}^3} \bar{\varphi}(\vec{x}) \psi(\vec{x}) \, \mathrm{d}\vec{x}$. Das ist nicht symmetrisch, aber* **hermitesch***: $\langle \varphi \mid \psi \rangle = \overline{\langle \psi \mid \varphi \rangle}$ und linear in der zweiten Komponente. Falls wir beweisen, dass symmetrische reelle Matrizen diagonalisierbar sind, werden wir auf „hermitesch“ näher eingehen. Diese Konstruktion liefert ein „echtes“ Skalarprodukt auf $\mathscr{C}^0([0,1], \mathbb{R})$.*

Lemma 7.3
Sind B, C zwei Basen eines endlichdimensionalen $\mathbb{R}$-Vektorraums V und ist $b\colon V \times V \to \mathbb{R}$ bilinear, so gilt ${}^{\mathrm{t}}{}_B\mathbb{T}_C \, {}_C b \, {}_B\mathbb{T}_C = {}_C b$.

Beweis:
Sei $C = [c_1, ..., c_n]$. Dann ist ${}_C b_{i,j} = b(c_i, c_j) = {}^{\mathrm{t}}\kappa_B(c_i) \, {}_B b \, \kappa_B(c_j)$
$= {}^{\mathrm{t}}\kappa_C(c_i) \, {}^{\mathrm{t}}{}_B\mathbb{T}_C \, {}_B b \, {}_B\mathbb{T}_C \, \kappa_C(c_i) = {}^{\mathrm{t}}\vec{e}_i \, {}^{\mathrm{t}}{}_B\mathbb{T}_C \, {}_B b \, {}_B\mathbb{T}_C \, \vec{e}_j$. □

In allgemeinen Skalarproduktträumen berechnet man Längen und Winkel genau so wie in der Schule bezüglich des Standardskalarprodukts:

Definition 7.4. *Sei $(V, \langle | \rangle)$ ein Skalarproduktraum und $v, w \in V$.*

a) *$\|v\| := \sqrt{\langle v | v \rangle}$, die* **Länge** *oder* **Norm** *von v. Einen Vektor der Länge 1 nennt man auch* **normiert**.

b) *Falls $v, w \neq 0$: $\sphericalangle(v,w) := \arccos \frac{\langle v|w \rangle}{\|v\| \cdot \|w\|}$, der* **Winkel** *zwischen v und w.*

c) *$v \perp w$, d.h. v und w sind zueinander* **orthogonal**[51] *$:\Leftrightarrow \langle v | w \rangle = 0$.*

Offenbar $\vec{u} \perp \vec{v} \iff \vec{v} \perp \vec{u}$ wegen Symmetrie des Skalarprodukts.

[51] Nach griech. ὀρθός „richtig, recht-“ und γωνία „Ecke, Winkel“.

7.2 Definitheitstest

Symmetrie einer Bilinearform kann man leicht erkennen. Aber wie kann man die positive Definitheit prüfen? Prinzipiell kann auch eine nicht-symmetrische Bilinearform positiv definit sein.[52] Doch man führt den Definitheitstest für allgemeine Bilinearformen und Matrizen auf den symmetrischen Fall zurück.

Lemma 7.5 (und Definition). *Sei $M \in M_n(\mathbb{R})$.*

a) M heißt **positiv definit** $:\Leftrightarrow \forall \vec{v} \in \mathbb{R}^n \setminus \{\vec{0}\}: {}^t\vec{v}M\vec{v} > 0$. *Offenbar ist eine Bilinearform genau dann positiv definit, wenn ihre Darstellungsmatrix bezüglich einer beliebigen Basis positiv definit ist.*

b) M heißt **schiefsymmetrisch** $:\Leftrightarrow {}^tM = -M$.
In diesem Fall gilt $\forall \vec{v} \in \mathbb{R}^n: {}^t\vec{v}M\vec{v} = 0$.

c) M ist genau dann positiv definit, wenn $M + {}^tM$ positiv definit ist.

Beweis:
b) ${}^t\vec{v}M\vec{v} = ({}^t\vec{v}M\vec{v}) = {}^t\vec{v}{}^tM\vec{v} = -{}^t\vec{v}M\vec{v}$. Also $2 \cdot {}^t\vec{v}M\vec{v} = 0 \overset{2\neq 0}{\Longrightarrow} {}^t\vec{v}M\vec{v} = 0$.

c) Sei $M_+ := \frac{1}{2}(M+{}^tM)$ und $M_- := \frac{1}{2}(M-{}^tM)$. Dann $M = M_+ + M_-$. Wegen $M_- = -{}^tM_-$ ist $\forall \vec{v} \in \mathbb{R}^n: {}^t\vec{v}M\vec{v} = {}^t\vec{v}M_+\vec{v} + {}^t\vec{v}M_-\vec{v} = \frac{1}{2}{}^t\vec{v}(M + {}^tM)\vec{v}$. □

Satz 7.6
Sei $A \in M_n(\mathbb{R})$, ${}^tA = A$. Es ist A genau dann positiv definit, wenn man man A mit dem Gauß-Algorithmus ohne Zeilenvertauschungen und optionale Schritte auf eine ZSF bringen kann, deren Diagonaleinträge alle positiv sind.

Bemerkung Vorsicht! Ist ${}^tA \neq A$, so testet man die Definitheit von A durch Anwendung von Satz 7.6 auf $A + {}^tA$ (nicht direkt auf A!).

Beweis:
Wenn $A' \in M_n(\mathbb{R})$ durch Zeilenoperationen aus A entsteht, dann gibt es $G \in GL_n(\mathbb{R})$ mit $A' = GA$. Sei B die Basis, die aus den Spalten von ${}^tG$ besteht, und sei E die Standardbasis, d.h. ${}_E\mathbb{T}_B = {}^tG$. Die Bilinearform, die bezüglich E die Darstellungsmatrix A hat, hat nach Lemma 7.3 bezüglich B die Darstellungsmatrix $A'' := GA{}^tG$. Also ist A positiv definit $\iff$ A'' ist positiv definit.

A'' entsteht aus A, indem man zu jeder Zeilenoperation auch die entsprechende Spaltenoperationen durchführt. Wenn also A' eine ZSF ist, dann ist A'' eine Diagonalmatrix. Eine Diagonalmatrix ist genau dann positiv definit, wenn

[52] In Anwendungen der Definitheit (Bestimmung von Extremwerten in der Analysis, Skalarprodukte in der Geometrie, verschiedene numerische Methoden) hat man es aber meist mit symmetrischen Matrizen zu tun; daher gibt es Lehrbücher, in denen der Begriff „positiv definit“ nur für symmetrische Matrizen definiert wird. also anders als bei uns.

alle Diagonaleinträge positiv sind (wurde in der Vorlesung nachgewiesen). Um den Satz zu beweisen, fehlt nur noch der Fall, dass G einige Zeilenoperationen beschreibt und danach im Gauß-Algorithmus ein Zeilentausch nötig wäre. Dann gäbe es ein i mit $A''_{i,i} = 0$, und wegen $A''_{i,i} = {}^{\mathrm{t}}\vec{e}_i A''\vec{e}_i$ wäre dann A'' (also auch A) nicht positiv definit. □

Beispiele: a) Sei $A := \begin{pmatrix} 1 & -1 & -1 \\ -1 & 3 & 5 \\ -1 & 5 & 17 \end{pmatrix} = {}^{\mathrm{t}}A$. $A \rightsquigarrow \begin{pmatrix} 1 & 1 & -1 \\ 0 & 2 & 4 \\ 0 & 4 & 16 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & -1 & 1 \\ 0 & 2 & 4 \\ 0 & 0 & 8 \end{pmatrix}$, also ist A positiv definit.

b) Sei $A := \begin{pmatrix} 4 & -2 & -2 \\ -2 & 2 & 0 \\ -2 & 0 & 1 \end{pmatrix} = {}^{\mathrm{t}}A$. $A \rightsquigarrow \begin{pmatrix} 4 & -2 & -2 \\ 0 & 1 & -1 \\ 0 & -1 & 0 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 4 & -2 & -2 \\ 0 & 1 & -1 \\ 0 & 0 & -1 \end{pmatrix}$, also nicht positiv definit.

c) Sei $A := \begin{pmatrix} 1 & -2 & 3 \\ -2 & 4 & 3 \\ 3 & 3 & 2 \end{pmatrix} = {}^{\mathrm{t}}A$. $A \rightsquigarrow \begin{pmatrix} 1 & -2 & 3 \\ 0 & 0 & 9 \\ 0 & 9 & -7 \end{pmatrix}$ und danach wäre ein Zeilentausch nötig; also nicht positiv definit.

7.3 Orthonormalbasen

In der Schule standen Koordinatenachsen paarweise aufeinander senkrecht und waren mit Einheitslängen markiert. Dies führt auf folgende Begriffsbildung:

Definition 7.7. *Sei V ein Skalarproduktraum.*

a) $(v_i)_{i\in I} \subset V$ *heißt* **Orthogonalsystem**, *gdw.* $v_i \perp v_j$ *für alle* $i \neq j \in I$.

b) $(v_i)_{i\in I} \subset V$ *heißt* **Orthonormalsystem** $:\Leftrightarrow \forall i, j \in I\colon \langle v_i \mid v_j\rangle = \delta_{i,j} := \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases}$, *mit dem so genannten* **Kronecker-Delta**[53] $\delta_{i,j}$.

c) Eine **Orthonormalbasis** *(***ONB***) bzw.* **Orthogonalbasis** *(***OGB***) eines euklidischen Raums V ist eine Basis von V, die zugleich ein Orthonormalsystem bzw. Orthogonalsystem ist.*

Eine Basis B eines eulidischen Raums ist genau dann eine OGB (bzw. ONB), wenn die Darstellungsmatrix des Skalarprodukts bzgl. B eine Diagonalmatrix (bzw. Einheitsmatrix) ist.

Beispiele: a) $[\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n]$ ist eine ONB von $\mathbb{R}^n$ bzgl. Standardskalarprodukt.

b) In $\mathbb{R}^3$ mit dem Standardskalarprodukt ist $\begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix} \perp \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}$ und $\begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix} \not\perp \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}$. $[\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}]$ ist ein Orthogonalsystem und $[\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}]$ ein Orthonormalsystem.

[53] Manchmal wird es auch Kronecker-Symbol genannt, aber leider wird „Kronecker-Symbol“ noch für einen ganz anderen Begriff verwendet!

Wir wollen aus einer gegebenen Basis eine Orthonormalbasis berechnen. Gemäß der folgenden Beobachtung ist es sinnvoll, als zwischenschritt zunächst eine ggf. nicht-normierte OGB zu bestimmen:

Beobachtung 7.8. *Sei V ein euklidischer Raum.*
Wenn $[b_1, ..., b_n]$ eine OGB von V ist, dann ist $[\frac{1}{\|b_1\|}b_1, ..., \frac{1}{\|b_n\|}b_n]$ eine ONB.

Beweis:
$\forall i, j \in \{1, ..., n\}\colon \langle \frac{1}{\|b_i\|}b_i \mid \frac{1}{\|b_j\|}b_j\rangle = \frac{1}{\|b_i\|\|b_j\|}\langle b_i|b_j\rangle = \begin{cases} 0 & (i \neq j) \\ \frac{\langle b_i|b_i\rangle}{\|b_i\|^2} = 1 & (\text{sonst}) \end{cases}$. □

Problem 7.9
Gegeben sei eine Basis $[u_1, \ldots, u_d]$ eines euklidischen Vektorraums V, so dass $[u_1, ..., u_k]$ für ein $0 \leq k \leq d$ ein Orthogonalsystem ist.
*Berechne eine OGB $[v_1, ..., v_d]$ von V, so dass zudem $v_1 = u_1$, ..., $v_k = u_k$ und $\forall r \in \{k+1, ..., d\}\colon \operatorname{Span}_{\mathbb{R}}(v_1, ..., v_r) = \operatorname{Span}_{\mathbb{R}}(u_1, ..., u_r)$ gelten. Berücksichtige den **Spezialfall** $V \leq \mathbb{R}^n$.*

Bemerkung Unsere Herleitung eines Orthogonalisierungsverfahrens basiert auf dem Definitheitstest aus Satz 7.6. Für manche Anwendungen ist wichtig, dass man $\operatorname{Span}_{\mathbb{R}}(u_1, ..., u_r)$ beibehält. Daher halte ich es für einen Fehler, dass in manchen Quellen erlaubt wird, die Reihenfolge der gegebenen Vektoren vor der Orthogonalisierung zu ändern.

Das Gram-Schmidt-Verfahren ist anfällig gegen Rundungsfehlerverstärkung, daher werden in der Numerik oft andere Orthogonalisierungsverfahren eingesetzt, die allerdings die Bedingung $\operatorname{Span}_{\mathbb{R}}(v_1, ..., v_r) = \operatorname{Span}_{\mathbb{R}}(u_1, ..., u_r)$ verletzen.

Lösung (Gram-Schmidt-Verfahren):
Sei $A \in M_d(\mathbb{R})$ die Darstellungsmatrix des Skalarprodukts von V bezüglich der Basis $B = [u_1, ..., u_d]$. Analog zu Satz 7.6 wenden wir auf $(A, \mathbb{1}_d) \in \mathbb{K}^{d\times 2d}$ den Gauß-Algorithmus ohne Zeilenvertauschung und optionale Schritte an und erhalten dadurch (A', G), wobei A' eine ZSF und $G \in GL_d(\mathbb{R})$ ist, so dass $GA = A'$ und so dass $GA^{\mathrm{t}}G$ eine Diagonalmatrix ist. Zudem ist G eine untere Dreiecksmatrix, da nur „nach unten“ eliminiert wurde, und die Diagonalelemente von G sind 1, sofern keine Zeilenskalierungen verwendet wurden.

Es gibt eine Basis $C = [v_1, ..., v_d]$, so dass ${}^{\mathrm{t}}G = {}_B\mathbb{T}_C$: Dabei ist v_i ist die Linearkombination von $[u_1, ..., u_d]$ gegeben durch die i-te Zeile von G. Die Darstellungsmatrix bzgl. C ist $GA^{\mathrm{t}}G$.
Variante: Man führt nach jeder Zeilenoperation auch die entsprechende Spaltenoperation durch. Dadurch entsteht automatisch $GA^{\mathrm{t}}G$ aus A. Die Zeilen $k+1, ..., d$ darf man skalieren, um Brüche zu vermeiden. Wird nämlich die i-te Zeile mit $\lambda \in \mathbb{R}^*$ skaliert, so wird auch die i-te Spalte mit λ skaliert; also ändert sich das i-te Diagonalelement um den Faktor $\lambda^2 > 0$: Das Vorzeichen bleibt gleich. Zeilenvertauschung und Elimination „nach oben“ bleibt aber verboten.

Behauptung: $[v_1, ..., v_d]$ ist die Lösung des Orthogonalisierungsproblems. Weil $[u_1, ..., u_k]$ ein Orthogonalsystem ist, sind die ersten k Zeilen und Spalten von A bereits diagonal und werden durch den Gauß-Algorithmus nicht verändert. Also $G = \begin{pmatrix} \mathbb{1}_k & \mathbb{0} \\ * & D \end{pmatrix}$ mit einer unteren Dreiecksmatrix $D \in GL_n(\mathbb{R})$. Insbesondere ist $\forall i \in \{1, ..., k\}\colon\ v_i = u_i$ und $\forall r \in \{1, ..., n\}\colon\ \mathrm{Span}_{\mathbb{R}}(v_1, ..., v_r) = \mathrm{Span}_{\mathbb{R}}(u_1, ..., u_r)$. Zudem ist C ein Orthogonalsystem, denn die Darstellungsmatrix des Skalarproduktes bzgl. C ist $GA^{\mathrm{t}}G$, also eine Diagonalmatrix.

Spezialfall: Sei $A \in M_d(\mathbb{K})$ mit $A_{i,j} = \langle u_i | u_j \rangle$ und sei $X \in \mathbb{K}^{d \times n}$ die Matrix, deren Zeilen durch $[u_1, ..., u_d]$ gegeben sind. Man transformiert (A, X) ohne optionale Schritte und Zeilentausch auf ZSF (A', X'), wendet aber für jede Zeilenoperation auch die zugehörige Spaltenoperation an. Es entsteht (A'', X'), wobei A'' eine Diagonalmatrix mit positiven Diagonaleinträgen ist, und es gibt $G \in GL_d(\mathbb{R})$ mit $A'' = GA^{\mathrm{t}}G$ und $GX = X'$. Die Zeilen von X' entsprechen der gesuchten OGB C. □

Beispiele: a) Sei $V \leq \mathbb{R}^4$ der Lösungsraum der Gleichung $x_1 + x_2 + x_3 + x_4 = 0$. Konstruiere eine Orthonormalbasis von V bzgl. Standardskalarprodukt.

$B := [\vec{u}_1, \vec{u}_2, \vec{u}_3]$ mit $\vec{u}_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}$, $\vec{u}_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix}$, $\vec{u}_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}$ ist eine Basis von V. Die Darstellungsmatrix des Standardskalarprodukts bzgl. B ist $A := \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$. Wir führen die Gram-Schmidt-Orthogonalisierung durch:

$$\left(\begin{array}{ccccccc} 2 & 1 & 1 & 1 & -1 & 0 & 0 \\ 1 & 2 & 1 & 1 & 0 & -1 & 0 \\ 1 & 1 & 2 & 1 & 0 & 0 & -1 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccccccc} 2 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & \frac{3}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & -1 & 0 \\ 0 & \frac{1}{2} & \frac{3}{2} & \frac{1}{2} & \frac{1}{2} & 0 & -1 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccccccc} 2 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 6 & 2 & 1 & 1 & -2 & 0 \\ 0 & 2 & 6 & 1 & 1 & 0 & -2 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccccccc} 2 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 6 & 0 & 1 & 1 & -2 & 0 \\ 0 & 0 & \frac{16}{3} & \frac{2}{3} & \frac{2}{3} & \frac{2}{3} & -2 \end{array}\right).$$

Dies ergibt (jeweils nach Division durch die Wurzel aus den Diagonalelementen) die ONB $[\begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \\ 0 \end{pmatrix}, \begin{pmatrix} 1/(2\sqrt{3}) \\ 1/(2\sqrt{3}) \\ 1/(2\sqrt{3}) \\ -3/(2\sqrt{3}) \end{pmatrix}]$.

b) Sei $V := \mathbb{R}[x]_{\leq 3}$ mit der Basis $B := [1, x, x^2, x^3]$ und dem Skalarprodukt $\langle f | g \rangle := \int_{-1}^{1} f(x)g(x)\,\mathrm{d}x$. Für $k \in \mathbb{N}$ gilt $\int_{-1}^{1} x^k\,\mathrm{d}x = \begin{cases} 0 & k \text{ ungerade} \\ \frac{2}{k+1} & \text{sonst} \end{cases} \rightsquigarrow$

Das Skalarprodukt hat bzgl. B die Darstellungsmatrix $\begin{pmatrix} 2 & 0 & \frac{2}{3} & 0 \\ 0 & \frac{2}{3} & 0 & \frac{2}{5} \\ \frac{2}{3} & 0 & \frac{2}{5} & 0 \\ 0 & \frac{2}{5} & 0 & \frac{2}{7} \end{pmatrix}$. Wir erweitern die Matrix, indem wir jeweils in der i-ten Zeile den i-ten Basisvektor anhängen, und führen dann Zeilen- und Spaltenoperationen durch:

$$\left(\begin{array}{ccccc} 2 & 0 & \frac{2}{3} & 0 & 1 \\ 0 & \frac{2}{3} & 0 & \frac{2}{5} & x \\ \frac{2}{3} & 0 & \frac{2}{5} & 0 & x^2 \\ 0 & \frac{2}{5} & 0 & \frac{2}{7} & x^3 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccccc} 2 & 0 & 0 & 0 & 1 \\ 0 & \frac{2}{3} & 0 & \frac{2}{5} & x \\ 0 & 0 & \frac{8}{45} & 0 & x^2 - \frac{1}{3} \\ 0 & \frac{2}{5} & 0 & \frac{2}{7} & x^3 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccccc} 2 & 0 & 0 & 0 & 1 \\ 0 & \frac{2}{3} & 0 & 0 & x \\ 0 & 0 & \frac{8}{45} & 0 & x^2 - \frac{1}{3} \\ 0 & 0 & 0 & \frac{8}{175} & x^3 - \frac{3}{5}x \end{array}\right)$$

Also ist $[1, x, x^2 - \frac{1}{3}, x^3 - \frac{3}{5}x]$ eine OGB von V.

Bemerkung 7.10
Zeilenoperationen entsprechen einem Basiswechsel: Addiert man in der Darstellungsmatrix bzgl. $[v_1, ..., v_{j-1}, u_j, ..., u_d]$ *das* λ*-Fache der* i*-ten Zeile und Spalte zur* j*-ten Zeile und Spalte* $(i < j)$*, so entspricht dies der Ersetzung von* u_j *durch* $u_j + \lambda v_i$*. Bei der Elimination ist* $\lambda = -\frac{\langle u_j | v_i \rangle}{\langle v_i | v_i \rangle}$.

Bis auf eine geänderte Summationsreihenfolge läuft das Gram-Schmidt-Verfahren darauf hinaus, für $j = 1, ..., d$ *iterativ* $v_j := u_j - \sum_{i=1}^{j-1} \frac{\langle u_j | v_i \rangle}{\langle v_i | v_i \rangle} v_i$ *zu berechnen. Dies ist in der Literatur die übliche Sichtweise auf das Gram-Schmidt-Verfahren. Dann muss man allerdings wiederholt Skalarprodukte berechnen; das ist fehleranfällig. Daher halte ich meine Herangehensweise für besser, zumal sich daraus auch ein Zusammenhang mit dem Definitheitstest ergibt.*

Lemma 7.11
Jedes Orthonormalsystem S *in einem Skalarproduktraum* V *ist linear unabhängig.*

Beweis:
Sei $v = \sum_{b \in S}{}' \lambda_b b$. Dann $\forall w \in S\colon\ \langle v | w \rangle = \sum_{b \in S}{}' \lambda_b \langle b | w \rangle = \lambda_w$ (diese Formel sollten Sie sich merken!). Daher: $\sum_{b \in S}{}' \lambda_b b = o \ \Rightarrow\ \lambda_b = \langle 0 | b \rangle = 0$ für alle $b \in S$. □

Daraus und aus dem Gram-Schmidt-Verfahren folgt unmittelbar:

Orthonormalisierungssatz
Jeder euklidische Vektorraum hat eine ONB und jedes Orthonormalsystem in einem euklidischen Vektorraum kann man zu einer ONB fortsetzen. □

7.3.1 Das orthogonale Komplement

Definition 7.12. *Sei* $\mathbb{K}$ *ein Körper,* V *ein* $\mathbb{K}$*-Vektorraum,* $U \leq V$.
$W \leq V$ *heißt* **Komplement** *von* U *in* $V :\Leftrightarrow V = U + W$ *und* $U \cap W = \{o\}$. *Notation:* $V = U \oplus W$ *(***direkte Summe** *von* U *und* W*). Ist* V *endlichdimensional, so folgt* $\dim(V) = \dim(U) + \dim(W)$ *aus der Dimensionsformel.*

Im Rest dieses Abschnitts sei $(V, \langle | \rangle)$ ein Skalarproduktraum.

Definition 7.13 (und Übung). *Sei* $M \subset V$.
$M^\perp := \{v \in V \mid \forall u \in M\colon\ v \perp u\} \leq V$ *heißt* **Orthogonalraum** *von* M.

Beispiel: Ist $[\vec{u}, \vec{v}] \subseteq \mathbb{R}^3$ linear unabhängig, dann ist $[\vec{u}, \vec{v}]^\perp$ eine Ursprungsgerade, die senkrecht zur Ebene $\mathrm{Span}_{\mathbb{R}}(\vec{u}, \vec{v})$ steht.

Lemma 7.14. *Sei* $M \subset V$.

a) $M_1 \subset M_2 \subset V \ \Rightarrow\ M_2^\perp \subset M_1^\perp$

b) $M^\perp = (\mathrm{Span}_{\mathbb{R}}(M))^\perp$

c) $M \subset (M^\perp)^\perp$

Beweis:

a) Für Elemente von $M_2^\perp$ sind auch die Bedingungen für das Enthaltensein in $M_1^\perp$ erfüllt.

b) $\supset$ folgt aus a). Sei $v \in M^\perp$ und $u = \sum_{b\in M}{}' \lambda_b b \in \mathrm{Span}_{\mathbb{R}}(M)$. Dann $\langle v \mid u\rangle = \sum_{b\in M}{}' \lambda_b \langle v|b\rangle = 0$.

c) Übung □

Satz 7.15 (und Definition)
Sei V euklidisch. Dann gilt für jedes $U \leq V$:

$$U \oplus U^\perp = V \qquad \textit{und} \qquad (U^\perp)^\perp = U$$

Man bezeichnet daher $U^\perp$ auch als **orthogonales Komplement** *von U.*

Bemerkung Für unendlichdimensionale Skalarprodukträume können die beiden Aussagen verletzt sein!

Beweis:
Sei B eine ONB von U. Orthonormalisierungssatz $\Rightarrow \exists C \subset V$ mit $B \cap C = \emptyset$, so dass $B \cup C$ eine ONB von V ist. Nach Lemma 4.37 gilt $V = U \oplus \mathrm{Span}_{\mathbb{R}}(C)$.
Behauptung: $U^\perp = \mathrm{Span}_{\mathbb{R}}(C)$.
Sei dazu $V \ni v = \sum_{b\in B}{}' \lambda_b b + \sum_{c\in C}{}' \lambda_c c$.

- $v \in U^\perp \Rightarrow \forall b \in B\colon \langle v|b\rangle = \lambda_b = 0 \Rightarrow v \in \mathrm{Span}_{\mathbb{R}}(C)$.
- $v \in \mathrm{Span}_{\mathbb{R}}(C) \Rightarrow \forall b \in B\colon \langle v|b\rangle = \sum_{c\in C}{}' \lambda_c \langle c|b\rangle = 0 \Rightarrow v \in U^\perp$.

Damit ist die Behauptung bewiesen. Es folgt $V = U \oplus U^\perp$. Außerdem folgt $U = (U^\perp)^\perp$, indem man in der Behauptung die Rollen von B und C vertauscht.□

Lemma 7.16. *Sei V euklidisch. Für alle $v, w \in V$ gelten:*

- $|\langle v|w\rangle| \leq \|v\| \cdot \|w\|$ ***(Cauchy-Schwarz-Ungleichung)***
- $\|v + w\| \leq \|v\| + \|w\|$ ***(Dreiecksungleichung)***

In beiden Teilen gilt: Im Gleichheitsfall ist $[v, w]$ linear abhängig.

Beweis:
Wir nutzen $V = \mathrm{Span}_{\mathbb{R}}(v) \oplus (\mathrm{Span}_{\mathbb{R}}(v))^{\perp}$.
Cauchy-Schwarz: $w = \lambda v + w'$ für ein $w' \in (\mathrm{Span}_{\mathbb{R}}(v))^{\perp}$ und $\lambda \in \mathbb{R}$, also $\|w\|^2 = |\lambda|^2 \|v\|^2 + \|w'\|^2$. Daher $|\langle v|w\rangle|^2 = |\lambda|^2 \|v\|^4 \leq |\lambda|^2 \|v\|^4 + \|v\|^2 \|w'\|^2 = \|v\|^2 \|w\|^2$. Für Gleichheit brauchen wir $v = o$ oder $w' = o$, in jedem Fall $[v, w]$ linear abhängig.
Dreieck: $\|v + w\|^2 = \|v\|^2 + \|w\|^2 + 2\langle v|w\rangle \leq \|v\|^2 + \|w\|^2 + 2\|v\| \cdot \|w\| = (\|v\| + \|w\|)^2$ wegen Cauchy-Schwarz. □

Ich möchte diesen Abschnitt mit der Anmerkung schließen, dass man in $\mathbb{R}^3$ mit dem Standardskalarprodukt zu linear unabhängigem $[\vec{v}, \vec{w}] \in \mathbb{R}^3$ eine Basis von $\mathrm{Span}_{\mathbb{R}}(\vec{v}, \vec{w})^{\perp}$ findet:

Bemerkung 7.17 (und Definition)
Für $\vec{v}, \vec{w} \in \mathbb{R}^3$ definiert man das **Kreuzprodukt** *(auch* **Vektorprodukt** *genannt) von $\vec{v}$ und $\vec{w}$ durch $\vec{v} \times \vec{w} := \begin{pmatrix} v_2w_3 - v_3w_2 \\ v_3w_1 - v_1w_3 \\ v_1w_2 - v_2w_1 \end{pmatrix}$. Sie zeigen gerade in einer Übung, dass $\vec{v} \times \vec{w} = \vec{0}$ gdw. $[\vec{v}, \vec{w}]$ linear abhängig.*

Bezüglich des Standardskalarprodukts ist $\vec{v} \perp (\vec{v} \times \vec{w})$ und $\vec{w} \perp (\vec{v} \times \vec{w})$. Ferner $\vec{w} \times \vec{v} = -\vec{v} \times \vec{w}$ (anti-kommutativ) Für das Kreuzprodukt gilt das Distributivgesetz, jedoch im Allgemeinen nicht das Assoziativgesetz.

Die folgenden Eigenschaften des Kreuzprodukts sind nicht so einfach nachzurechnen. Da es sich nur um eine Illustration handelt, beweise ich sie nicht:

a) $\|\vec{v} \times \vec{w}\| = \|\vec{v}\|\|\vec{w}\| \, |\sin \alpha|$, wobei $\alpha := \sphericalangle(\vec{v}, \vec{w})$.

b) **Rechte-Hand-Regel***: Zeigen Daumen bzw. Zeigefinger der rechten Hand in Richtung $\vec{v}$ bzw. $\vec{w}$, dann zeigt $\vec{v} \times \vec{w}$ senkrecht von der Handfläche weg.*

7.4 Besondere Abbildungen

7.4.1 Orthogonale Abbildungen

In diesem Abschnitt betrachten wir lineare Abbildungen, die Längen und Winkel erhalten. Welche derartige Endomorphismen von $\mathbb{R}^2$ und $\mathbb{R}^3$ mit dem Standardskalarprodukt gibt es (Drehungen, Spiegelungen etc.), wie erkennt man sie?

Definition 7.18. *Seien V, W Skalarprodukträume.*
$\varphi \in \mathrm{Hom}_{\mathbb{R}}(V, W)$ heißt **orthogonal** *$:\Leftrightarrow \forall v, w \in V\colon \langle v|w\rangle = \langle \varphi(v) \mid \varphi(w)\rangle$. Eine orthogonale Abbildung ist also längen- und winkelerhaltend.*

Ab jetzt geht es um den endlichdimensionalen Fall.

Lemma 7.19. *Sei $B = [v_1, ..., v_n]$ eine ONB des euklidischen Vektorraums V.*
$\varphi \in \mathrm{End}_{\mathbb{R}}(V)$ ist orthogonal $\iff$ $\varphi(B)$ ist eine ONB.

Beispiel: Drehungen oder Spiegelungen im Anschauungsraum.

Beweis:
$\Rightarrow$: φ orthogonal $\Rightarrow \langle\varphi(v_i) \mid \varphi(v_j)\rangle = \langle v_i|v_j\rangle = \delta_{i,j} \Rightarrow \varphi(B)$ ist ein Orthonormalsystem. Insbesondere ist es linear unabhängig. Es enthält n Elemente, ist also eine Basis (d.h. insgesamt eine ONB).

$\Leftarrow$: Seien B und $\varphi(B)$ Orthonormalbasen. $\forall v \in V\colon\ \kappa_B(v) = \kappa_{\varphi(B)}(\varphi(v))$, denn $\varphi(\sum_{i=1}^{n} \gamma_i v_i) = \sum_{i=1}^{n} \gamma_i \varphi(v_i)$. Dann $\forall v, w \in V\colon\ \langle v|w\rangle \overset{B \text{ ONB}}{=} {}^{\mathrm{t}}\kappa_B(v)\kappa_B(w) = {}^{\mathrm{t}}\kappa_{\varphi(B)}(\varphi(v))\kappa_{\varphi(B)}(\varphi(w)) \overset{\varphi(B) \text{ ONB}}{=} \langle\varphi(v) \mid \varphi(w)\rangle$. Also φ orthogonal. $\square$

Korollar 7.20 (und Definition)
Bezüglich des Standardskalarprodukts auf $\mathbb{R}^n$ gilt für $A \in M_n(\mathbb{R})$: L_A orthogonal $\iff$ die Spalten von A bilden eine ONB $\iff {}^{\mathrm{t}}AA = \mathbb{1}_n$.

a) Ist ${}^{\mathrm{t}}AA = \mathbb{1}_n$, so nennt man A eine **orthogonale Matrix**. *Man definiert die* **orthogonale Gruppe** $O_n := \{A \in M_n(\mathbb{R}) \mid {}^{\mathrm{t}}AA = \mathbb{1}_n\}$.

b) Für alle $A \in O_n$ gilt $\det(A) \in \{-1, 1\}$. Man definiert die **spezielle orthogonale Gruppe** $SO_n := \{A \in O_n \mid \det(A) = 1\}$.

Sowohl O_n als auch SO_n sind Gruppen bezüglich Matrixmultiplikation.

Beweis:
L_A orthogonal $\iff$ die Spalten bilden eine ONB gilt nach dem vorigen Lemma. Übung: Dies ist genau dann der Fall, wenn ${}^{\mathrm{t}}AA = \mathbb{1}_n$.
$A \in O_n \Rightarrow 1 = \det(\mathbb{1}_n) = \det({}^{\mathrm{t}}AA) = \det({}^{\mathrm{t}}A)\det(A) = (\det(A))^2$. Die Gruppeneigenschaft soll als Übung bewiesen werden. $\square$

Unsere geometrische Anschauung besagt, dass man jeden orthogonalen Endomorphismus des Anschauungsraums als Kombination von Drehungen und Spiegelungen darstellen kann. Das wollen wir nun systematisieren. Zunächst werde ich die Ergebnisse formulieren, damit Sie damit auch schon Aufgaben lösen können. Die Beweise führe ich danach.

Lemma 7.21
a) $SO_2 = \left\{\begin{pmatrix} \cos\vartheta & -\sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{pmatrix} \mid \vartheta \in \mathbb{R}\right\}$. Solche Matrizen nennt man **Drehmatrizen**; *beschrieben wird dabei die Drehung um $\vec{0}$ im Winkel ϑ gegen den Uhrzeigersinn. Ist $\sin\vartheta \neq 0$, so gibt es keinen reellen Eigenwert.*

b) $O_2 \setminus SO_2 = \left\{\begin{pmatrix} \cos\vartheta & \sin\vartheta \\ \sin\vartheta & -\cos\vartheta \end{pmatrix} \,\middle|\, \vartheta \in \mathbb{R}\right\}$. Multiplikation mit einer solchen Matrix enspricht einer Spiegelung an $\mathrm{Span}_{\mathbb{R}}(\begin{pmatrix} \cos\frac{\vartheta}{2} \\ \sin\frac{\vartheta}{2} \end{pmatrix}$, also Eigenwerte 1 und -1.

Lemma 7.22. *Für $A \in O_3$ mit $A \neq \pm\mathbb{1}_3$ (also L_A weder die identische Abbildung noch die* **Punktspiegelung***) sei $U_\pm := E_{\pm 1}(A) \leq \mathbb{R}^3$. Es gilt:*

a) $\det(A) = 1 \Rightarrow \dim(U_+) = 1$ *und L_A ist eine* **Drehung** *mit Achse U_+.*

b) $\det(A) = -1 \Rightarrow \dim(U_-) = 1$ *und L_A ist eine* **Drehspiegelung** *mit Achse U_-, also eine Drehung um U_- im Winkel ϑ gefolgt von der Spiegelung an der Ebene ${U_-}^\perp$.*
Im Spezialfall $\vartheta = 0$ stellt A die **Spiegelung** *an ${U_-}^\perp$ dar.*

Wir fassen zusammen:

Problem 7.23 (Klassifikation orthogonaler Endomorphismen von $\mathbb{R}^3$)
Sei $f\colon \mathbb{R}^3 \to \mathbb{R}^3$ ein Endomorphismus mit Abbildungsmatrix $F \in M_3(\mathbb{R})$ bezüglich der Standardbasis. Prüfe, ob f orthogonal bzgl. des Standardskalarprodukts ist, bestimme den Typ von f sowie ggf. Drehachse und den Betrag des Drehwinkels.

Lösung:
f orthogonal $\iff {}^tFF = \mathbb{1}_3$. Wir setzen ab jetzt ${}^tFF = \mathbb{1}_3$ voraus. $F = \mathbb{1}_3 \iff f = \mathrm{Id}_{\mathbb{R}^3}$ und $F = -\mathbb{1}_3 \iff f$ ist Punktspiegelung. Sei ab jetzt $F \neq \pm\mathbb{1}_3$.

- $\det(F) = 1 \Rightarrow f$ ist eine Drehung mit eindimensionaler Achse $E_1(F)$. Der Drehwinkel φ hat den Betrag $\arccos\left(\frac{1}{2}(\mathrm{Spur}(F) - 1)\right)$: Ergänzt man nämlich einen normierten Eigenvektor mit Eigenwert 1 zu einer ONB D von $\mathbb{R}^3$, dann ${}_DM_D(f) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix}$. Beim Basiswechsel ändert sich die Spur nicht, d.h. es gilt $\mathrm{Spur}({}_DM_D(f)) = \mathrm{Spur}(F)$, und die Spurformel trifft für ${}_DM_D(f)$ offenbar zu.

- $\det(F) = -1 \Rightarrow f$ ist eine Drehspiegelung mit eindimensionaler Achse $E_{-1}(F)$. Der Drehwinkel φ hat den Betrag $\arccos\left(\frac{1}{2}(\mathrm{Spur}(F) + 1)\right)$: Ergänzt man nämlich einen normierten Eigenvektor mit Eigenwert -1 zu einer ONB D von $\mathbb{R}^3$, dann ${}_DM_D(f) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix}$. Die Spurformel gilt offenbar für ${}_DM_D(f)$ und zudem $\mathrm{Spur}({}_DM_D(f)) = \mathrm{Spur}(F)$. Spiegelebene: $(E_{-1}(F))^\perp$. □

Beispiel 7.24
Es sei $f\colon \mathbb{R}^3 \to \mathbb{R}^3$ eine lineare Abbildung mit der Abbildungsmatrix $F = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \end{pmatrix}$ bezüglich der Standardbasis. Man prüft ${}^tFF = \mathbb{1}_3$.

- $\det(F) = 1$. *Wegen $F \neq \mathbb{1}_3$ ist* $\dim(E_1(F)) = 1$*: $E_1(F) = \mathrm{LR}(\mathbb{1}_3 - F; \vec{0}) = \mathrm{Span}_{\mathbb{R}}(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix})$. Es handelt sich um eine Drehung um die Achse* $\mathrm{Span}_{\mathbb{R}}(\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix})$.

- $\mathrm{Spur}(F) = 1/3$. *Betrag des Drehwinkels:* $\arccos\left(\frac{1}{2}(\frac{1}{3} - 1)\right) = \arccos(-1/3) \approx 1.910633$ *oder circa* $109.47°$*; dies ist der so genannte* **Tetraederwinkel**.

Beispiel 7.25
Es sei $f\colon \mathbb{R}^3 \to \mathbb{R}^3$ *eine lineare Abbildung mit der Abbildungsmatrix* $F = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ -\frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{pmatrix}$ *bezüglich der Standardbasis. Man prüft* ${}^{\mathrm{t}}FF = \mathbb{1}_3$.

- $\det(F) = -1$. *Wegen* $F \neq -\mathbb{1}_3$ *ist* $\dim(E_{-1}(F)) = 1$*:* $E_{-1}(F) = \mathrm{LR}(-\mathbb{1}_3 - F; \vec{0}) = \mathrm{Span}_{\mathbb{R}}(\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix})$.
- *Wegen* $\mathrm{Spur}(F) = 1$ *ist der Drehwinkel Null, es handelt sich also um die Spiegelung an der Ebene* $E_1(F) = \mathrm{Span}_{\mathbb{R}}(\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix})$.

Additionstheoreme für Sinus und Kosinus

$$\forall \varphi, \vartheta \in \mathbb{R}\colon \ \sin(\vartheta + \varphi) = \sin(\vartheta)\cos(\varphi) + \cos(\vartheta)\sin(\varphi) \quad \textit{und}$$
$$\cos(\vartheta + \varphi) = \cos(\vartheta)\cos(\varphi) - \sin(\vartheta)\sin(\varphi)\,.$$

Bemerkung Das ist kein Schulstoff mehr, obwohl man es eigentlich für Themen benötigt, die noch Schulstoff sind. Auch für die Polardarstellung komplexer Zahlen haben wir sie bereits verwendet. Am Ende des Skriptes werde ich noch einen elementargeometrischen Beweis der Additionstheoreme bereitstellen.

Beweis von Lemma 7.21:

a) Die Matrix $\begin{pmatrix} \cos\vartheta & -\sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{pmatrix}$ ist orthogonal mit Determinante 1. Für $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SO_2$, ist $\begin{pmatrix} a & c \\ b & d \end{pmatrix} \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ und $ad - bc = 1$, also

$$a^2 + c^2 = 1 \qquad b^2 + d^2 = 1 \qquad ab + cd = 0 \qquad ad - bc = 1$$

Aufgrund der ersten beiden Gleichungen gibt es Zahlen ϑ, φ mit $a = \cos\vartheta$, $c = \sin\vartheta$, $b = \cos\varphi$, $d = \sin\varphi$. Wegen $ad - bc = 1$ ist $\sin(\varphi - \vartheta) = 1$, also oBdA $\varphi = \vartheta + \frac{\pi}{2}$. Deshalb ist $b = -\sin\vartheta$, $d = \cos\vartheta$.

Das charakteristische Polynom ist $X^2 - 2\cos\vartheta X + 1 = (X - \cos\vartheta)^2 + \sin^2\vartheta$. Für $\sin\vartheta \neq 0$ hat das Polynom keine Nullstellen in $\mathbb{R}$.

b) Aus $A \in O_2 \setminus SO_2$ folgt $\det(A) = -1$. Sei $B = A \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Dann $A = B \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, und $B \in SO_2$. Nach a) ist $A = \begin{pmatrix} \cos\vartheta & \sin\vartheta \\ \sin\vartheta & -\cos\vartheta \end{pmatrix}$. Dann ist $\begin{pmatrix} \cos\frac{\vartheta}{2} \\ \sin\frac{\vartheta}{2} \end{pmatrix}$ ein Eigenvektor mit Eigenwert 1:

$$\begin{pmatrix} \cos\vartheta & \sin\vartheta \\ \sin\vartheta & -\cos\vartheta \end{pmatrix} \cdot \begin{pmatrix} \cos\frac{\vartheta}{2} \\ \sin\frac{\vartheta}{2} \end{pmatrix} = \begin{pmatrix} \cos\left(\frac{\vartheta}{2}\right)\cos(\vartheta) + \sin\left(\frac{\vartheta}{2}\right)\sin(\vartheta) \\ -\cos(\vartheta)\sin\left(\frac{\vartheta}{2}\right) + \cos\left(\frac{\vartheta}{2}\right)\sin(\vartheta) \end{pmatrix}$$
$$= \begin{pmatrix} \cos\left(\frac{\vartheta}{2}\right)\cos(-\vartheta) - \sin\left(\frac{\vartheta}{2}\right)\sin(-\vartheta) \\ -\cos(-\vartheta)\sin\left(\frac{\vartheta}{2}\right) - \cos\left(\frac{\vartheta}{2}\right)\sin(-\vartheta) \end{pmatrix} = \begin{pmatrix} \cos\left(-\frac{\vartheta}{2}\right) \\ -\sin\left(-\frac{\vartheta}{2}\right) \end{pmatrix} = \begin{pmatrix} \cos\frac{\vartheta}{2} \\ \sin\frac{\vartheta}{2} \end{pmatrix}$$

Ebenso ist $\begin{pmatrix} -\sin\frac{\vartheta}{2} \\ \cos\frac{\vartheta}{2} \end{pmatrix}$ ein Eigenvektor mit Eigenwert -1. Man erkennt: A stellt eine Spiegelung an der Gerade $E_1(A)$ dar. □

Beweis von Lemma 7.22:
$\deg(\chi_A(X)) = 3$ ist ungerade, also gibt es $\lambda \in \mathbb{R}$ mit $\chi_A(\lambda) = 0$. Wegen $A \in O_3$ gilt $\forall \vec{v} \in \mathbb{R}^3\colon\ \|A\vec{v}\| = \|\vec{v}\|$, dies gilt insbesondere, wenn $\vec{v}$ ein Eigenvektor zum Eigenwert λ ist. Daher $|\lambda| = 1$. Wenn möglich, so wähle $\lambda = 1$. Sei $B = [\vec{b}_1, \vec{b}_2, \vec{b}_3]$ eine ONB von $\mathbb{R}^3$ mit $\vec{b}_1$ ein Eigenvektor zum Eigenwert λ.

a) Setze $U = \vec{b}_1^{\perp} = \mathrm{Span}_{\mathbb{R}}(\vec{b}_2, \vec{b}_3)$. Für $\vec{u} \in U$ ist wegen A orthogonal

$$0 = \langle \vec{b}_1 \mid \vec{u} \rangle = \langle A\vec{b}_1 \mid A\vec{u} \rangle = \lambda \langle \vec{b}_1 \mid A\vec{u} \rangle \,.$$

Wegen $\lambda = \pm 1$ ist $\langle \vec{b}_1 | A\vec{u} \rangle = 0$, also $A\vec{u} \in U = \vec{b}_1^{\perp} = \mathrm{Span}_{\mathbb{R}}(\vec{b}_2, \vec{b}_3)$, und $F := L_A\big|_U^U$ ist ein orthogonaler Endormorphismus von U. Bezüglich B hat L_A die Matrix $\left(\begin{smallmatrix} \lambda & 0 \\ 0 & C \end{smallmatrix}\right)$, mit der Abbildungsmatrix C von F. Also $1 = \det A = \lambda \det(C)$, wegen $\lambda^2 = 1$ also $\det(C) = \lambda$.

$\lambda = -1$ ist unmöglich, denn sonst wäre $\det(C) = -1$ und nach Lemma 7.21.b) hätte C und damit auch A den Eigenwert $+1$ — wir hätten also $\lambda = +1$ gewählt.

b) Hier ist $-A \in SO_3$. Wenn $A \neq -\mathbb{1}_3$, dann ist $U = E_1(-A) = E_{-1}(A)$ eindimensional und $-A$ ist eine Drehung um die Achse U. Somit ist A eine Drehspiegelung mit Achse U, bzw. eine Spiegelung, falls der Drehwinkel Null ist.

Also $1 = \lambda = \det(C) = \det(F)$. Wegen $A \neq \mathbb{1}_3$ ist $F \neq \mathrm{Id}$ und nach Lemma 7.21.a) ist L_A eine echte Drehung der Ebene U um Achse $\mathrm{Span}_{\mathbb{R}}(\vec{b}_1)$. Insbesondere $U_+ = \mathrm{Span}_{\mathbb{R}}(\vec{b}_1)$, alle anderen Eigenräume haben Dimension 0. □

7.4.2 Symmetrische Endomorphismen

In diesem Abschnitt geht es um Endomorphismen eines euklidischen Raumes, die durch einen orthogonalen Basiswechsel diagonalisiert werden können. Das hat überraschend viele Anwendungen, beispielsweise in der Geometrie und in der numerischen Mathematik.

Definition 7.26
Sei V ein Skalarproduktraum. $\varphi \in \mathrm{End}(V)$. heißt **symmetrisch** *oder* **formal selbstadjungiert** $:\Leftrightarrow \forall v, w \in V\colon\ \langle v \mid \varphi(w) \rangle = \langle \varphi(v) \mid w \rangle$.

Für Anwendungen in der Physik müsste man die Begriffsbildung auf den Fall unendlichdimensionaler Skalarprodukträume übertragen, und dafür braucht man Analysis. Man würde dann von „selbstadjungierten Operatoren auf Hilberträumen“ reden.

Lemma 7.27
Sei V ein Skalarproduktraum und $\varphi \in \mathrm{End}_{\mathbb{R}}(V)$ symmetrisch. Seien $u, v \in V$ Eigenvektoren von φ zu Eigenwerten $\lambda \neq \mu$. Dann $u \perp v$.

Beweis:
$\mu\langle u|v\rangle \overset{\text{bilin.}}{=} \langle u|\mu v\rangle \overset{\text{EV}}{=} \langle u|\varphi(v)\rangle \overset{\text{symm.}}{=} \langle \varphi(u)|v\rangle \overset{\text{EV}}{=} \langle \lambda u|v\rangle \overset{\text{bilin.}}{=} \lambda\langle u|v\rangle$.
Also $(\lambda - \mu) \cdot \langle u|v\rangle = 0$ und wegen $\lambda \neq \mu$ folgt $\langle u|v\rangle = 0$. □

Bemerkung In der Vorlesung schien es, als sei ***partielle Integration*** (was man im nächsten Beispiel braucht) kein Schulstoff mehr. Nach der Vorlesung zeigte sich aber, dass Sie partielle Integration wahrscheinlich doch aus der Schule kennen, aber nicht unter diesem Namen. Partielle Integration wird aus der Ableitungsregel für Produkte hergeleitet: Seien $u, v\colon [a,b] \to \mathbb{R}$ stetig differenzierbar. Dann $(uv)' = u'v + uv'$. Folglich $uv\Big|_a^b = \int_a^b (uv)' \,\mathrm{d}x = \int_a^b u'v \,\mathrm{d}x + \int_a^b uv' \,\mathrm{d}x$. Man kann umstellen zu

$$\int_a^b u'v \,\mathrm{d}x = uv\Big|_a^b - \int_a^b uv' \,\mathrm{d}x$$

Insbesondere hat partielle Integration nichts mit partiellen Ableitungen zu tun.

Beispiel 7.28
$V := \{f \in \mathscr{C}^\infty([0,1],\mathbb{R}) \mid f(0) = f(1) = 0\}$ ist mit $\langle f|g\rangle := \int_0^1 f(x)g(x) \,\mathrm{d}x$ ein Skalarproduktraum. Die Abbildung $\varphi\colon V \to V$ mit $\varphi(f) := -f''$ ist symmetrisch, denn für alle $f, g \in V$ folgt mit zweimaliger partieller Integration unter Ausnutzung von $f(0) = f(1) = g(0) = g(1) = 0$:

$$\langle -f'' \mid g\rangle = \int_0^1 -f''(x)g(x) \,\mathrm{d}x = -f'(x)g(x)\Big|_{x=0}^{x=1} - \int_0^1 (-f'(x))g'(x) \,\mathrm{d}x$$

$$= f(x)g'(x)\Big|_{x=0}^{x=1} - \int_0^1 f(x)g''(x) \,\mathrm{d}x = \langle f \mid -g''\rangle$$

Für $n \in \mathbb{N}$ sei $f_n := ((x \mapsto \sin(n\pi x))$. Dann ist $f_n \in V$ ein Eigenvektor von φ zum Eigenwert $n^2\pi^2$. Für $m \neq n \in \mathbb{N}$ gilt also nach dem vorigen Lemma: $0 = \langle f_m \mid f_n\rangle = \int_0^1 \sin(m\pi x)\sin(n\pi x) \,\mathrm{d}x$. Das könnte man natürlich auch direkt nachrechnen, aber das wäre weniger elegant als der abstrakte Zugang über Eigenvektoren einer symmetrischen linearen Abbildung.

Lemma 7.29. *Sei V ein euklidischer Vektorraum.*
Sei $\varphi \in \mathrm{End}_{\mathbb{R}}(V)$, C eine Basis von V und P die Darstellungsmatrix des Skalarprodukts bezüglich C. φ ist symmetrisch $\iff {}^{\mathrm{t}}{}_C M_C(\varphi) \cdot P = P \cdot {}_C M_C(\varphi)$.

Bemerkung Ist C eine ONB, so hat das Skalarprodukt die Darstellungsmatrix $P = \mathbb{1}_n$. In diesem Fall gilt: φ ist symmetrisch $\iff$ ${}_CM_C(\varphi)$ ist symmetrisch.

Beweis:
Für alle $v, w \in V$ gilt: $\langle v|\varphi(w)\rangle = {}^t\kappa_C(v) \cdot P \cdot ({}_CM_C(\varphi)\kappa_C(w))$ und
$\langle \varphi(v)|w\rangle = {}^t({}_CM_C(\varphi)\kappa_C(v)) \cdot P \cdot \kappa_C(w) = {}^t\kappa_C(v)\, {}^t({}_CM_C(\varphi))\, P\kappa_C(w)$. □

Als nächstes zeigen wir, dass symmetrische Endomorphismen eines euklidischen Vektorraums reelle Eigenwerte haben.

Lemma 7.30 (und Definition). *Sei V ein $\mathbb{R}$-Vektorraum.*
Für $\varphi \in \mathrm{End}_{\mathbb{R}}(V)$ sei $\sigma(\varphi) := \{\lambda \in \mathbb{C} \mid \chi_\varphi(\lambda) = 0\}$ das **Spektrum** *von φ; das ist also die Menge aller komplexen Eigenwerte von φ.*
Wenn φ symmetrisch ist, dann $\sigma(\varphi) \subset \mathbb{R}$.

Beweis:
Erinnerung: $\chi_\varphi := \chi_A$ mit $A := {}_CM_C(\varphi) \in M_n(\mathbb{R})$ für eine beliebige Basis C von V. Sei nun C eine ONB. Wegen Lemma 7.29 folgt, dass A symmetrisch ist.

Wir wenden komplexe Konjugation komponentenweise auf Matrizen und Vektoren an. Wegen $A \in M_n(\mathbb{R})$ haben wird $\bar{A} = A$. Sei nun $\lambda \in \sigma(\varphi)$, also $\lambda \in \mathbb{C}$ mit $\chi_A(\lambda) = 0$. Dann $\exists \vec{0} \neq \vec{z} \in \mathbb{C}^n\colon\ A\vec{z} = \lambda\vec{z}$.

Aus $A = {}^tA$ und $\bar{A} = A$ folgt ${}^t(\overline{A\vec{z}})\vec{z} = {}^t\bar{\vec{z}}A\vec{z}$. Also $\bar{\lambda}\,{}^t\bar{\vec{z}}\vec{z} = \lambda\,{}^t\bar{\vec{z}}\vec{z}$. Wegen ${}^t\bar{\vec{z}}\vec{z} = \sum_{i=1}^{n} |z_i|^2 \neq 0$ folgt $\bar{\lambda} = \lambda$, also $\lambda \in \mathbb{R}$. □

Lemma 7.31
Sei V ein euklidischer Vektorraum und $\varphi \in \mathrm{End}_{\mathbb{R}}(V)$ symmetrisch. Wenn $U \leq V$ **φ-invariant**[54] *ist, dann ist auch $U^\perp$ φ-invariant.*

Beweis:
$v \in U^\perp \Rightarrow \forall u \in U\colon\ \langle u|v\rangle = 0 \Rightarrow \forall u \in U\colon\ \langle u \mid \varphi(v)\rangle \overset{\text{symm.}}{=} \langle \underbrace{\varphi(u)}_{\in U} \mid v\rangle = 0$. □

Spektralsatz
Ist V ein euklidischer Raum und $\varphi \in \mathrm{End}_{\mathbb{R}}(V)$ ein symmetrischer Endomorphismus, dann hat V eine ONB aus Eigenvektoren von φ.

Beweis:
Induktion nach $\dim(V)$. Verankerung für $\dim(V) = 1$: Sei $0 \neq u \in V$ beliebig, dann ist $\varphi(u) \in V = \mathrm{Span}_{\mathbb{R}}(u)$, d.h. $\exists\lambda \in \mathbb{R}\colon\ \varphi(u) = \lambda u$, d.h. u ist ein Eigenvektor, und $[\frac{1}{\|u\|}u]$ ist eine ONB aus Eigenvektoren.

Sei nun $n := \dim V > 1$. Nach dem Hauptsatz der Algebra hat $\chi_\varphi(X)$ mindestens eine Nullstelle in $\mathbb{C}$, d.h. φ hat einen Eigenwert $\lambda \in \mathbb{C}$. Nach Lemma 7.30 gilt sogar $\lambda \in \mathbb{R}$.

[54] Das heißt $\varphi(U) \subset U$.

Sei $U := E_\lambda(\varphi) \leq V$ mit einer ONB $[v_1, ..., v_r]$. Ware $U = V$, so wäre nichts mehr zu zeigen; sei also $U \neq V$, also $r = \dim(U) < n$. Weil λ ein Eigenwert ist, gilt $\dim(U) \geq 1$. Wegen $V = U \oplus U^\perp$ folgt $n = \dim(V) = \dim(U) + \dim(U^\perp)$, daher $1 \leq \dim(U^\perp) = n - r < n$. Durch das auf V definierte Skalarprodukt wird auch $U^\perp$ zu einem euklidischen Raum.

Nach Lemma 7.31 ist $\varphi(U^\perp) \subseteq U^\perp$, d.h. $\varphi' := \varphi\big|_{U^\perp}^{U^\perp} \in \mathrm{End}_\mathbb{R}(U^\perp)$. Induktionsannahme: $U^\perp$ hat eine aus Eigenvektoren von φ bestehende ONB $[v_{r+1}, ..., v_n]$. Wegen $V = U \oplus U^\perp$ ist $[v_1, ..., v_n]$ die gesuchte ONB von V. □

Bemerkung 7.32
Wir zeigen: Jede symmetrische Matrix $A \in M_n(\mathbb{R})$ hat eine speziell-orthogonale diagonalisierende Matrix.
Laut des Spektralsatzes gibt es nämlich Eigenvektoren $\vec{b}_1, ..., \vec{b}_n$ von A mit Eigenwerten $\lambda_1, ..., \lambda_n$, so dass $B := [\vec{b}_1, ..., \vec{b}_n]$ eine ONB bezüglich des Standardskalarprodukts ist. Sei $S := (\vec{b}_1, ..., \vec{b}_n) \in M_n(\mathbb{R})$.

- *Wenn E die Standardbasis bezeichnet, ist ${}_E\mathbb{T}_B = S$.*
- *Weil B eine ONB ist, ist S orthogonal; folglich $S^{-1} = {}^tS$ und $\det(S) = \pm 1$.*
- *Auch $[\vec{b}_1, ..., \vec{b}_{n-1}, -\vec{b}_n]$ ist eine ONB aus Eigenvektoren; wir können also oBdA annehmen, dass $\det(S) = +1$ und damit $S \in SO_n$.*
- *Weil $\vec{b}_1, ..., \vec{b}_n$ Eigenvektoren von A sind, ist S eine diagonalisierende Matrix, und daher $S^{-1}AS = \mathrm{diag}(\lambda_1, ..., \lambda_n)$.*
- *Wegen $S^{-1} = {}^tS$ ist ${}^tSAS = \mathrm{diag}(\lambda_1, ..., \lambda_n)$ diagonal.*

Im Spezialfall, dass A die Darstellungsmatrix eines Skalarprodukts b bezüglich der Standardbasis ist, dann ist die Darstellungsmatrix von b bezüglich der Basis B gleich ${}^tSAS = {}^t{}_E\mathbb{T}_B \, A \, {}_E\mathbb{T}_B$, ist also diagonal. Und dann ist B nicht nur eine Orthonormalbasis bezüglich des Standardskalarprodukts, sondern zugleich eine Orthogonalbasis (aber ggf. nicht normiert) bezüglich des Skalarprodukts b. Und wegen $\det(S) = 1$ gilt sogar, dass die B die gleiche Orientierung wie die Standardbasis hat.
Dies lässt sich wie folgt rechnerisch nachvollziehen (beachte: Wir setzen $A = {}^tA$ voraus!):

- *Berechne $\chi_A(X)$ und dessen Nullstellen, also die Eigenwerte von A. Der Spektralsatz garantiert, dass $\chi_A(X)$ in $\mathbb{R}[X]$ in Linearfaktoren zerfällt (findet man also nicht-reelle Eigenwerte, so hat man sich verrechnet).*
- *Für jeden Eigenwert λ: Berechne eine ONB von $E_\lambda(A)$. Berechne also zunächst eine gewöhnliche Basis, wende (falls $\dim(E_\lambda(A)) > 1$) das Gram-Schmidt-Verfahren an, und normiere die Basisvektoren. Der Spektralsatz garantiert, dass die Dimension jeweils gleich der algebraischen Vielfachheit ist (sonst hat man sich verrechnet).*

- *Da die verschiedenen Eigenräume eine symmetrischen Matrix gemäß Lemma 7.27 zueinander orthogonal sind, ergeben die ONBn der einzelnen Eigenräume zusammen genommen ein Orthonormalsystem in $\mathbb{R}^n$. Der Spektralsatz garantiert, dass dieses Orthonormalsystem eine Basis $[\vec{b}_1, ..., \vec{b}_n]$ bestehend aus Eigenvektoren zu Eigenwerten $\lambda_1, ..., \lambda_n \in \mathbb{R}$ ist.*

- *$\tilde{S} := (\vec{b}_1, ..., \vec{b}_n) \in O_n$, denn die Spalten bilden eine ONB, und $\tilde{S}$ ist eine diagonalisierende Matrix für A, denn die Spalten sind Eigenvektoren für A.*

- *Man berechnet $\det(\tilde{S})$. Ist dies nicht ± 1, so hat man sich irgendwo verrechnet. Ist $\det(\tilde{S}) = 1$, dann ist $S := \tilde{S} \in SO_n$ die Lösung. Ist $\det(\tilde{S}) = -1$, dann erhält man zum Beispiel durch Negation einer Spalte oder alternativ durch Vertauschen zweier Spalten aus $\tilde{S}$ eine speziell orthogonale diagonalisierende Matrix S und damit die Lösung denn durch Negation einer bzw. Tausch zweier Spalten ändert die Determinante ihr Vorzeichen, aber die Spalten bilden nach wie vor eine ONB aus Eigenvektoren.*

Beispiel: Sei $A := \left(\begin{smallmatrix} 1 & 3 \\ 3 & 1 \end{smallmatrix}\right)$, $A = {}^tA$. $\chi_A(X) = X^2 - 2X - 8 = (X-4) \cdot (X+2)$, $E_4(A) = \mathrm{LR}(\left(\begin{smallmatrix} 3 & -3 \\ -3 & 3 \end{smallmatrix}\right); \vec{0}) = \mathrm{Span}(\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right))$, $E_{-2}(A) = \mathrm{LR}(\left(\begin{smallmatrix} -3 & -3 \\ -3 & -3 \end{smallmatrix}\right); \vec{0}) = \mathrm{Span}(\left(\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right))$.

Die Eigenräume sind eindimensional. Um also jeweils eine ONB in jedem Eigenraum zu berechnen, ist die Gram-Schmidt-Orthogonalisierung nicht nötig, man braucht bloß jeweils den gefundenen Basisvektor normieren. Also ist $\tilde{S} := \left(\begin{smallmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{smallmatrix}\right)$. Wegen $\det(\tilde{S}) = -\frac{1}{2} - \frac{1}{2} \neq 1$ ist das noch nicht die Lösung. Aber $S := \left(\begin{smallmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{smallmatrix}\right) \in SO_2$ ist eine diagonalisierende Matrix für A.
Es gilt ${}^tSAS = \left(\begin{smallmatrix} 4 & 0 \\ 0 & -2 \end{smallmatrix}\right)$.

Index

Äquivalenzklasse, 44
Äquivalenzrelation, 43

Abbildung, 20
Abbildungen mit endlichem Träger, 62
Abbildungsmatrix, 66
abelsch, 39
Abstraktionsprinzip, 18
Additionstheoreme, 49
Additionstheoreme für Sinus und Kosinus, 99
Adjunkte, 77
allgemeine lineare Gruppe, 28
Allquantor, 12
alternierend, 72
alternierende Quersumme, 46
Arganddiagramm, 48
Argument, 48
Arithmetik, 2
Assoziativgesetz, 39, 40
Aussageform, 11
Aussagenlogik, 12
Aussonderungsaxiom, 19
Auswahlaxiom, 56
Automorphismengruppe, 64
Automorphismus, 64

Basis, 61
    Charakterisierung, 63
Basislösung, 33
Basiswechselmatrix, 66
Betrag, 47
bijektiv, 56
Bild, 56
Bildmenge, 20, 56
Bilinearform, 88
Blockgestalt, 73

charakteristische Matrix, 81
charakteristisches Polynom, 81
Chiffrierung, 51
Code
    linearer, 51
Codewort, 51
Codierung, 51
Corestriktion, 56
Cramersche Regel, 78

Darstellungsmatrix, 66, 88
de Morgansche Regeln, 16
Definitionsmenge, 20
Determinantenfunktion, 72
diagonalisierbar, 86
diagonalisierende Matrix, 86
Diagonalmatrix, 27
Differenzmenge, 10, 19
Dimension, 65
Dimensionsformel, 68
direkte Summe, 94
disjunkt
    paarweise, 69
Distributivgesetze, 41
Divisionsring, 41
Drehmatrizen, 97
Drehspiegelung, 98
Drehung, 98
Dreiecksmatrix
    obere, 73
    strikte, 73
    untere, 73
Dreiecksungleichung, 49
Dualraum, 57
Durchschnitt, 19

Eigenraum, 80
Eigenvektor, 80
Eigenwert, 80
Einheit, 42
Einheitengruppe, 42
Einheitsmatrix, 27
Einschränkung, 56

Einselement, 41
Einsmatrix, 27
Einzigkeitsquantor, 12
Elementarmatrix, 28
endlichdimensional, 65
Endomorphismenring, 64
Endomorphismus, 64
Ersetzungsaxiom, 20
erweiterte Matrix, 29
Erzeugendensystem, 60
Erzeugnis, 30, 60
euklidischer Raum, 88
Existenzquantor, 12
Extension, 9
Extensionalitätsaxiom, 18

Fakultät, 26
Fallunterscheidung, 17
Familie, 60
fast alle, 60
formal selbstadjungiert, 100
freie Variablen, 32
Fundamentalsatz der Algebra, 48
Funktion, 20
Funktionsgraph, 20

Gauß-Jordan-Algorithmus, 35
Gaussche Zahlenebene, 48
gebundene Variablen, 32
Generatormatrix, 51
Geometrie, 2
gerade, 13
gleich, 56
gleichbedeutend, 15
Gruppe, 39
    abelsche, 39
    symmetrische, 74
    triviale, 39
Gruppenaxiome, 39

Hamilton-Operator, 81
Hamming-Code, 51
Hauptkomponentenanalyse, 81
Hauptsatz der Algebra, 84
hermitesch, 87, 89
homogen, 29
Homomorphismus, 57

Identitätsabbildung, 56
imaginäre Einheit, 47
Imaginärteil, 47
Indexmenge, 60
Informationsrate, 51
Inhomogenität, 23, 29
injektiv, 56
innere Verknüpfung, 38
Intension, 9
Inverse, 40
inverse Abbildung, 56
invertierbar, 42
irreduzibel, 46
isomorph, 64
Isomorphismus, 64

Körper, 41
Körperaxiome, 41
Kardinalität, 9
kartesische Darstellung, 48
kartesische Produkt, 20
Kern, 59
Klasse, 18
Koeffizienten, 60
Koeffizientenmatrix, 23, 29
Kommutativgesetz, 39, 41
Komplement, 94
    orthogonales, 95
komplexen Zahlen, 47
Komponente, 23
Komponenten, 22
Komposition, 56
kongruent, 43
konjugiert komplexe Zahl, 47
Kontraposition, 16
Kontrollmatrix, 51
Koordinaten, 61
Koordinatenabbildung, 65
Korrekturrate, 51

Kreuzprodukt, 96
Kronecker-Delta, 91

Länge, 89
Lösung
    Basis-, 33
    spezielle, 32
Lösungsraum, 29
Laplace-Entwicklung, 73
leere Menge, 10
Leermengenaxiom, 20
Leibnizformel, 77
linear, 57
linear abhängig, 31, 60
linear unabhängig, 31, 60
lineares Gleichungssystem, 29
    homogenes, 29
Linearfaktorzerlegung, 48
Linearkombination, 24, 26, 60
    triviale, 60
Logik, 2
logisch äquivalent, 15
Logische Blöcke, 8

Mächtigkeit, 9
Matrix, 23
    inverse, 28
    invertierbare, 28
    orthogonale, 97
    quadratische, 23
    transponierte, 23
Menge, 8, 18
    leere, 20
Methode, 2
Minore, 73
Modul, 62
modulo, 43
multilinear, 72
Multiplikationssatz, 74

Neolithikum, 2
neutrale Element, 40
Neutralität, 40
Norm, 89
normiert, 72, 82, 89
Nullelement, 41
Nullmatrix, 27
Nullring, 41
Nullvektor, 22
Numerik, 7

OGB, 91
ONB, 91
Ordnungsrelation, 43
orthogonal, 89, 96
Orthogonalbasis, 91
orthogonale Gruppe, 97
Orthogonalraum, 94
Orthogonalsystem, 91
Orthonormalbasis, 91
Orthonormalisierungssatz, 94
Orthonormalsystem, 91

Paar
    geordnetes, 20
Paarmengenaxiom, 18
Paläolithikum, 2
partielle Integration, 101
Permutation, 74
Pivotspalte, 31
Polardarstellung, 48
positiv definit, 88, 90
Potenzmenge, 20
Potenzmengenaxiom, 20
Prä-Hilbertraum, 88
Prädikatenlogik, 12
Primzahl, 46
Problem, 5
Punktspiegelung, 98
punktweise, 54

Rückwärtssubstitution, 25, 30, 32
Rang, 70, 71
Rangformel
    für Matrizen, 70
Rangformel für lineare Abbildungen, 71
Realteil, 47
Rechenregeln für komplexe Zahlen, 47

Rechte-Hand-Regel, 96
reduzierte Zeilenstufenform, 35
Relation, 43
    reflexiv, 43
    symmetrisch, 43
    transitiv, 43
Repräsentant, 44
Restklasse, 45
Restklassenring, 45
Restriktion, 56
Ring, 40
    kommutativer, 41
    unitärer, 41
Ringaxiome, 41
Rng, 41
Russellsche Antinomie, 18

Säkulargleichung, 81
Sarrus-Regel, 73
Satz
    lineare Fortsetzung, 65
Satz von Abel-Ruffini, 84
Schiefkörper, 41
schiefsymmetrisch, 72, 90
Schnitt, 19
Schnittmenge, 9, 19
Skalar, 53
Skalarmultiplikation, 25, 53
Skalarprodukt, 88
Skalarproduktraum, 88
Spaltenoperationen, 27
Spaltenraum, 70
Spaltenvektor, 22
Span, 30, 60
Spektralsatz, 102
Spektrum, 102
spezielle orthogonale Gruppe, 97
Spezielle Werte von Sinus und Kosinus, 50
Spiegelung, 98
Spur, 82
Standardbasis, 61
Standarddarstellung, 48
Standardskalarprodukt, 88
Studium, 3
Summe, 68
    direkte, 94
surjektiv, 56
symmetrisch, 87, 88, 100
symmetrische Gruppe, 74
Syndrom, 51

Teilmenge, 10, 19
    echte, 19
teilt, 43
Tetraederwinkel, 98
Theorie, 5
Transposition, 74
trigonometrische Darstellung, 48
Tripel, 20
triviale Ring, 41
Tupel, 20

Umkehrabbildung, 56
Umrechnungsformeln, 50
Unbekannter, 29
Unendlichkeitsaxiom, 20
Untervektorraum, 59
    Charakterisierung, 59
    invarianter, 80, 102
    zyklischer, 80
Urbild, 56

Variablen, 11
Vektor-Addition, 53
Vektorprodukt, 96
Vektorraum, 53
Venn-Diagrammen, 10
Verallgemeinerung, 15
Vereinigung, 19
Vereinigungsaxiom, 19
Vereinigungsmenge, 9
Verknüpfung, 56
Verknüpfungssymbol, 38
Verschärfung, 15
Vielfachheit, 84
    algebraische, 84

geometrische, 84
Vorzeichen, 77

Widerspruchsbeweis, 17
Winkel, 89
wohldefiniert, 45
Wurzeln, 50

Zeilenoperationen, 27, 34
Zeilenraum, 70
Zeilenstufenform, 31
Zeilenvektor, 22
Zielmenge, 20
Zirkelschluss, 36
ZSF, 31