Lineare Algebra und Analytische Geometrie I + II

Ergänzte Notizen zur Vorlesung an der
Humboldt-Universität zu Berlin
Version 0.92 / 0.82

Thorsten Rohwedder

Diese Vorlesungsnotizen sind eine überarbeitete Niederschrift meiner zweisemestrigen Vorlesung in den „Corona-Semestern" im Winter 2020/21 und Sommer 2021. Ich habe begonnen, das Skript an geeigneten Stellen um die Vorlesungsfolien und/oder nachträglich um erläuternde Textpassagen, fehlende Beweise oder illustrierende Beispiele, teils umfassend zu ergänzen. Für die Lineare Algebra I ist das recht weit gediehen, das erklärt hoffentlich den Untertitel „Version 0.92"; für die II (noch) nicht so weit, daher ist das etwa Version 0.82.

Inhaltsverzeichnis

Inhaltlich-didaktischer Kommentar

Schaut man in die Geschichte der Mathematik, so stellt man lustigerweise fest, dass sie während ihrer fachlichen Entwicklung der letzten Jahrhunderte öfter mal „auf die Schnauze gefallen ist“. Das lag erstaunlicherweise hauptsächlich daran, dass sie in ihren Begriffsbildungen nicht klar und objektiv genug war – und daran, dass die fachliche Struktur der betrachteten Sachverhalte dann manchmal doch nicht so klar war, wie man eigentlich angenommen hatte.[1] Die Konsequenz daraus war die formale Neuordnung und „Durchaxiomatisierung“ der höheren Mathematik am Anfang des 20. Jahrhunderts: Die Mathematik wurde systematisch-formal, ausgehend von den Grundlagen der Mengenlehre mit Hilfe der formalen Logik bis hin zu den höheren Fachbereichen wie Analysis und Linearer Algebra aufgebaut. Dabei wurde bei der Darstellung der fachlichen Zusammenhänge großflächig auf „suggestive Bestandteile“ wie Erklärungen zum Sinngehalt verzichtet; Definitionen, Sätze und formale Beweise standen für sich, und sollten für sich stehen und nicht durch irreführende, „sinnliche“ Erklärungen verwässert werden. Federführend für diese Entwicklung war damals die französische Mathemematiker:innengruppe *Nicolas Bourbaki*, und der heute nach ihnen benannte fachliche Stil des *Bourbakismus* durchströmt auch heute noch viele der Vorlesungen, die an deutschen Universitäten gehalten werden.

Dieser bourbakistische Stil führt auf vielen Ebenen leider zu dem, was der Mathematikdidaktiker Hans Freudenthal 1963 als „antididaktische Inversion“ bezeichnet: In so einer fachlich-systematischen Abhandlung finden sich typischerweise die Dinge, die eigentlich am Ende eines langen historischen Prozesses standen (wie z.B. die Mengenlehre oder Logik) jetzt am Anfang; oft sieht man Begriffsdefinitionen die Gründe dafür, warum die Begriffe so festgelegt sind, wie sie sind, nur noch implizit oder gar nicht mehr an, und Beweise mathematischer Sätze fallen, gespickt mit genialen Ideen, gewissermaßen vom Himmel, während der geneigte Rezipient sich oft fragt, wie zum Teufel man darauf kommt. Was man erhält, ist, wie das ehemalige Bourbaki-Mitglied Pierre Cartier es über die frühen Veröffentlichungen der Bourbaki-Gruppe formuliert, „eine Enzyklopädie der Mathematik, die alle notwendige Information enthält.“ Dass das skizzierte Vorgehen für Mathematikstudierende aber von der didaktischen Warte beurteilt nicht besonders günstig ist, ist wohl klar. Cartier selbst formuliert weiter: „Wenn man [das Bourbaki-Frühwerk] als Lehrbuch betrachtet, ist es ein Desaster.“ (Quelle).

[1] Falls Sie sich schon ein bisschen in der Mathematik auskennen: Berühmte Beispiele sind zum einen das Ringen um einen vernünftigen Grenzwert-, dann Funktions-, dann Mengenbegriff, zum anderen die Frage nach der Stellung des Parallelenaxioms in der euklidischen Geometrie.

Dennoch hat sich an dem vorherrschenden Bourbaki-Stil die letzten 100 Jahre relativ wenig geändert. Mathematikstudierende der letzten Jahrzehnte (mich eingeschlossen) haben sich daran gewöhnt, dass man die enthaltenen Ideen und Gründe für die gegebenen Formulierungen dann oft für sich selbst „reverse-engineeren“ muss, um beim enthaltenen Sinngehalt und bei den vermutlichen Gründen für das „Warum?“ anzukommen. Für den „Allgemeinsterblichen“ nährt jeder Kontakt mit diesem fachlich perfekt durchdesignten, aber recht unnahbaren Konstrukt der modernen Hochschulmathematik oft die ohnehin recht massiven Vorurteile gegenüber dem „Elfenbeinturm Mathematik“; für Lehramtsstudierende, die das Hauptzielpublikum der diesem Skript zugrundeliegenden Vorlesung waren, führt es zu zusätzlich dem, was Felix Klein schon 1930 als „doppelte Diskontinuität“ bezeichnet hat: Erstsemesterstudierenden ist unklar, was zum Teufel die neue, formal-axiomatische Uni-Mathematik mit der Schul-Mathematik zu tun hat;[2] die zweite „Diskontinuität nach Klein“ entsteht, wenn sie sich als junge Lehrer fragen, wie sie diese undidaktisch-axiomatische Mathematik, von der sie jetzt einige gelernt haben, denn nun in die Schule bringen sollen.

Trotz dieser offensichtlichen Schieflagen hat sich der fachlich-systematischen Aufbau von Erstsemestervorlesungen im allgemeinen und der Linearen Algebra im Besonderen, für Mathematikstudierende ebenso wie für solche des höheren Lehramts, in meiner Wahrnehmung die letzten Jahrzehnte über relativ wenig bewegt. Bei der Vorbereitung dieser Vorlesung war ich recht erstaunt, wie wenig Variation es etwa über die Literatur hinweg im fachlichen Aufbau der Linearen Algebra von 1930 bis heute gegeben hat. Dieser bewegt sich immer noch relativ strikt an den klassischen historischen Vorlagen, und das, obwohl wohl jeder Mathematikerin wohl klar ist, wie hilfreich verschiedene Zugänge und Beweise für das fachliche Verständnis mathematischer Zusammenhänge sein können. Der Grund für diesen relativen Stillstand ist neben der nicht zu bestreitenden Eleganz und Kompaktheit der axiomatisch-bourbakischen Darstellung im Großen wie im Kleinen wohl auch, dass es sich bei fachlich-axiomatischen Konstrukten wie dem im ersten Semester der Linearen Algebra unterrichteten um etwas handelt, was man neudeutsch als *Legacy System* bezeichnet: Das ist ein System, das einige nicht von der Hand zu weisende Mängel (in diesem Fall didaktischer Natur) aufweist; trotzdem traut man sich nicht, an irgendeiner Stelle etwas zu verändern, weil es sofort Auswirkungen auf viele andere Stellen im durchoptimierten fachlichen Aufbau hat, und man diese wegen der großen Komplexität des dargestellten Zusammenhangsgefüges für gewöhnlich nicht vollständig zu überblicken vermag.

[2] Vielleicht haben Sie zum Beginn Ihres Studiums auch mal den Spruch gehört, sie sollen „alles, was sie in der Schule über Mathematik gelernt haben, vergessen“.

Im vorliegenden Skript habe ich trotzdem versucht, einige behutsame Vorschläge festzuhalten, wie man den Grundkanon der Linearen Algebra an einigen Stellen modifizieren kann, hoffentlich ohne dabei allzu viel fachliche Strenge aufgeben zu müssen (die uns Mathematikerinnen und Mathematikern ja auch ganz lieb ist). Hierzu gehören die folgenden Leitlinien:

- Im Gesamtaufbau: Fachliche Restrukturierung, weg von der bourbakistischen zu einer eher *genetisch-historisierenden* Darstellungsweise, die, wo immer sinnvoll, versucht, sich an der historischen Entwicklung des Fachs zu orientieren und die dargestellten Fragestellungen sinnvoll von der fachlichen und/oder historischen Seite zu motivieren;

- Im Kleineren: Expliziteres Herausstellen der Grundideen von Definitionen und Zusammenhängen, die in bemerkenswert vielen Definitionen und Sätzen oft nur implizit enthalten sind. Definitionen sollten wann immer möglich so formuliert werden, dass man den Sinn der Begriffsbildung, möglicherweise sogar ihren Werdungsprozess (ihre *Genese*, daher wieder: genetischer Ansatz) ansieht. Das technisch-zweckmäßige Vorgehen ist damit nicht abgeschafft, wird aber nachrangig entwickelt.

- Beweise sollen so klar wie möglich formuliert werden. Damit ist nicht nur die Klarheit in Ausdrucksweise und logischen Schlüssen gemeint, wie man sie auch sonst ohne Zweifel allerorts findet; Beweise sollen jeweils aber so ausgewählt werden, dass die Beweisidee möglichst klar hervortritt und möglichst viel Einsicht in die untersuchten Zusammenhänge geben kann.

- Auch zwischen Definitionen und Sätzen werden stärker als sonst Sinn und Bedeutung der dargestellten fachlichen Zusammenhänge herausgestellt werden. Es gibt hier meines Erachtens recht viel implizites Wissen, das Kenner der Linearen Algebra teilen und das es lohnt, explizit festgehalten zu werden.

- Wann immer möglich sollen auch Anknüpfungspunkte an die Anschauung (insbesondere geometrischer Natur!), an die Schulmathematik und an heutige Anwendungen der Linearen Algebra genutzt werden. Auch meine persönlichen didaktische Erfahrungen als Lehrender der Linearen Algebra soll natürlich einfließen, insbesondere, was mir bekannte besondere Lernhindernisse angeht.

Mit Hilfe meines eigenen Hirns und einer Menge Literatur habe ich im vorliegenden Skript versucht, all diesen Zielstellungen in einem gewissen Maß nachzukommen. Herausgekommen ist als Zwischenergebnis die vorliegende Zusammenstellung, die der suchenden Leserin hoffentlich als ein gewinnbringender Fundort für anregende Darstellungen der Zusammenhänge der Linearen Algebra I und II im Gesamtbild wie im Detail dienen kann.

Zur Orientierung gibt die folgende Aufzählung einen Überblick über den fachlich-didaktischen Aufbau und einige Besonderheiten der einzelnen Kapitel.

- Kapitel 1 gibt eine auf Erstsemesterstudierende abzielende Einführung in die **Grundfragen der Mathematik**, der Algebra, der Linearen Algebra und der Analytischen Geometrie. Dazu gehört neben einem Ausflug in die Historie der Gebiete auch ein Umriss der mathematisch-axiomatischen Arbeitsweise und eine Skizze der Fragestellungen und Grundideen der einzelnen Gebiete.

- Kapitel 2: Nach der notwendigen Einführung einiger fundamentaler Grundbegriffe steht hier von der fachlichen Warte bereits die algebraische Grundfrage nach der **Lösbarkeit von Gleichungen** im Zentrum. Dabei spielt der Begriff der Bedingung (aka Aussageform) als zentralere Rolle als sonst: Es stellt sich für die fachliche Darstellung als nützlich heraus, Gleichungen das gesamte Skript hindurch als spezielle Bedingungen der Form $f(x) = b$ an Elemente $x \in X$ einer gegebenen Grundmenge X zu formulieren. Bei den dargestellten Aussagen über die Lösbarkeit von Gleichungen wird die Verbindung entsprechender Eigenschaften der beteiligten Funktion f (Injektivität, Surjektivität, Umkehrbarkeit) in das Zentrum gestellt.

- Kapitel 3 beschäftigt sich etwas ausführlicher als in den meisten Büchern zur Linearen Algebra mit **Polynomen und komplexen Zahlen**. Addition und Multiplikation komplexer Zahlen werden durch das Hankelsche Permanenzprinzip motiviert und die Definitionen daraus hergeleitet. Thematisiert werden außerdem die algebraische Abgeschlossenheit von $\mathbb{C}$, der Fundamentalsatz der Algebra und die Mandelbrotmenge.

- Kapitel 4: Hier werden einfache geometrische Abbildungen in Ebene, Raum und $\mathbb{R}^n$ (ohne, dass eine Theorie zu linearen Funktionen zur Verfügung steht) mit Hilfe von Matrizen beschrieben. Abbildungen der Form $x \mapsto A \cdot x$ erlauben eine natürliche Einführung der **Matrixschreibweise für geometrische Abbildungen**; auch rückt der wichtige Gedanke der linearen Fortsetzung in den Vordergrund. Die Definition von Matrix-Vektor- und Matrix-Matrix-Multiplikation lassen sich bei der Nutzung von Matrizen zur Darstellung von Abbildungen natürlicher einführen als andernorts, wo sie oft unmotiviert vom Himmel fallen.

- Kapitel 5 führt in die Grundbegriffe der **abstrakten Algebra** ein, wie sie sich vor etwas mehr als 100 Jahren entwickelt haben. Hier steht wieder der algebraisch-gleichungslösende Aspekt im Vordergrund: Welche Bedeutung haben die algebraische Strukturen (wie z.B. die Guppenstruktur) auf die (eindeutige?) Lösbarkeit/ äquivalente Umformbarkeit von Gleichungen in dieser Struktur?

- Kapitel 6 wird die Lösungstheorie für **lineare Gleichungssysteme**, so weit es geht, zunächst ohne Vektorraumtheorie entwickelt. Es geht schon ziemlich weit.

- Kapitel 7: Zu einer Einführung in die Lineare Algebra, die normalerweise im ersten Semester eines Mathematikstudiums stattfindet, gehört immer auch eine Einführung in die **Grundlagen logischer Schlussweisen und** der **Mengenlehre**. Diese wird, wie oben erwähnt, häufig zu Beginn der Abhandlung „vorgeschaltet“. Historisch ordnet sich die Beschäftigung mit der Grundlagenmathematik eher an dieser Stelle ein, daher ist es hier zu finden. Ausgehend von den Grunderkenntnissen der Aussagenlogik und Prädikatenlogik gibt das Kapitel auch auch eine ausführlichere Darstellung gängiger Aussagetypen, Schlussweisen und Beweistechniken, die im „täglichen Leben“ einer Mathematikerin häufig vorkommen.

- Kapitel 8, **Vektorräume**: Hier habe ich mich abweichend von vielen landläufigen Darstellungen an äquivalenten Umformulierungen einiger Definition (Lineare Unabhängigkeit, Basis, Dimension) versucht, die die Darstellung meiner Meinung nach genetischer und im Falle der Definition der Dimension eines Vektorraums V (als maximale Anzahl linear unabhängiger Vektoren in V, angelehnt an H. Weyl) auch an einigen Stellen einfacher machen. Neu ist wohl, dass man beim Beweis, dass alle Basen gleich viel Elemente haben, auf den klassischen, recht technischen Grassmann-Steinitz'schen Austauschsatz verzichten kann, indem man das Problem auf die Lösungstheorie für lineare Gleichungssysteme aus dem vorigen Kapitel zurückführt.

- Kapitel 9: **Lineare Abbildungen** sind dann die Verallgemeinerung der in Kapitel 4 einführten Matrixabbildungen für allgemeine Vektorräume.

- Kapitel 10: Der Themenbereich der **darstellenden Matrizen für lineare Abbildungen** sorgt in meiner Erfahrung bei Lernenden immer wieder für einen großen Haufen an Missverständnissen. Das liegt oft an recht technischen Einführung der beteiligten Abbildungen und der für viele Studierende verwirrenden Rollen von Basiswechselmatrix und Koordinatentransformationsmatrix, kombiniert mit dem Problem, dass bei Abbildungen im $\mathbb{K}^n$ Koordinatenabbildungen. Es wurde versucht, diese Problematiken deutlich anzusprechen und klar zu formulieren und sie mit geeigneten Notationen und kommutativen Diagrammen „problemsensibler“ als sonst zu erklären.

- Kapitel 11 formuliert die grundlegenden Erkenntnisse der Linearen Algebra zur Lösung linearer Systeme in Form einer **allgemeinen Lösungstheorie für Gleichungen** der Form $f(x) = b$ mit linearem f, worin lineare Gleichungssysteme als wichtiger Spezialfall mit inbegriffen sind.

- Kapitel 12, **Affine Räume**: Hier wird zunächst die „Vektorisierung von Punkträumen“ dargestellt, d.h., wie eine axiomatisch definierte Geometrie unter einigen üblichen Voraussetzungen mit Vektoren ausgestattet werden kann und wie diese dann genutzt werden können, um ein Koordinatensystem für diese Geometrie einzuführen. Affine Räume werden dann als Axiomatisierung der Vorgehensweise in $\mathbb{R}^3$ eingeführt; insbesondere wird versucht, eine klare Theorie zur Geometrie des Raumes und der Ebene in dem Sinne zu entwickeln, dass man sich dort „mit Hilfe von Vektoren auf den Punkten des Raumes bewegt“. Die aus der Schule bekannten Schnittprobleme in Raum und Ebene werden dann von einem höheren Standpunkt aus im abstrakten Setting der affinen Räume beleuchtet.

- Im Kapitel 13 zu den **affinen Abbildungen** werden nach Darstellung der grundlegend-axiomatischen Theorie als Beispiele einige illustrierende Anwendungen im $\mathbb{R}^2$ und $\mathbb{R}^3$, insbesondere grafische Anwendungen und so genannte Axonometrien zur perspektivischen Darstellung, skizziert.

- Kapitel 14 behandelt **Skalarprodukte** im $\mathbb{R}^n$ und $\mathbb{C}^n$. Hierbei spielen Orthogonalität und orthogonale Projektionen eine zentralere Rolle als sonst üblich, z.B. in der motivierten Herleitung der Definition des Skalarprodukts im $\mathbb{R}^n$, im Beweis der Cauchy-Schwarz-Ungleichung und in der Herleitung einiger grundlegender Resultate über Vektorgeometrie (wie das Prinzip der orthogonalen Zerlegung oder der verallgemeinerte Satz des Pythagoras, die zwar eher dem Kanon der Funktionalanalysis entstammen, aber nicht auf diesen Kontext beschränkt sind, sich in den behandelten Skalarprodukträumen geometrisch schön deuten lassen und m.E. auch dort sehr zum geometrischen Verständnis dieser Skalarprodukträume beitragen können).

- Kapitel 15 widmet sich **isometrischen linearen Abbildungen** (also solchen, deren darstellende Matrizen orthogonal bzw. unitär sind). Stellen sie den linearen Anteil der in Kapitel 12 betrachteten affinen Abbildungen, so lässt sich daraus schnell eine auf diesen Abbildungen basierende Darstellung der so genannten Abbildungsgeometrie der euklidischen Räume entwickeln, die eine problemlose Einführung diverser geometrischer Grundbegriffe wie Kongruenz, Ähnlichkeit und Symmetrie mit Hilfe von Abbildungen erlaubt. Dieses Vorgehen und eine daraus resultierende Klassifikation der Affinitäten des $\mathbb{R}^n$ wird skizziert. Zusätzlich beweisen wir den Satz von Mazur-Ulam: Jede Isometrie eines Vektorraums ist affin.

- Kapitel 16: **Determinanten** können, wie es hier geschieht, recht natürlich als Axiomatisierung einer Volumenfunktion im $\mathbb{K}^n$ eingeführt werden, indem man die Determinante als Volumenmessung in $\mathbb{R}^2$ und $\mathbb{R}^3$ erkennt und dann sinnvolle Eigenschaften einer Volumenmessung als definierende Eigenschaften wählt (im Gegensatz zur landläufigen Darstellung, in der dieser Absatz oft gewissermaßen nur „durchscheint“). Die heutige Bedeutung von Determinanten zur Lösung linearer Gleichungssysteme ist im Gegensatz zu der vor 400 Jahren doch sehr begrenzt, daher wird das hier nur so weit entwickelt, wie er zur Berechnung von Eigenwerten im folgenden Kapitel notwendig ist.

- Die im Kapitel 17 behandelten **Eigenwerte** und das in Kapitel 18 folgende Thema der **Diagonalisierbarkeit** verkommen leider oft gerade bei Lehramtsstudierenden zu einer recht kalkülorientierten Angelegenheit ohne viel inhaltliches Verständnis. Daher wird hier einerseits der Begriff der Eigenvektoren und Eigenwerte auf Grundlage der Entwicklungen der vergangenen Kapitel expliziter als sonst geometrisch motiviert, andererseits habe ich mich bemüht, die Bedeutung der Existenz einer Eigenbasis und den Zusammenhang zu den darstellenden Matrizen und Koordinatentransformationen ausführlicher als andernorts darzustellen.

- Der Rest von Kapitel 18 ist einem meiner besonderen Steckenpferde, der **Singulärwertzerlegung** beliebiger Matrizen $A \in \mathbb{R}^{n \times m}$ gewidmet, die als „krönender Abschluss“ der Linearen Algebra II die Antwort auf viele Grundfragen der Linearen Algebra bereit hält, ein exzellentes Werkzug zur der strukturellen Analyse von Matrizen bietet und auch in der Praxis mittlerweile eine immense Bedeutung hat. Leider wird stattdessen vielerorts immer noch der für praktische Zwecke irrelevante und von der theoretischen Seite recht komplexe Jordan’sche Normalform Vorzug gegeben, wenn es um eine sinnvolle Verallgemeinerung der Diagonalisierbarkeit von Matrizen geht. Das mag im Lichte des jahrhundertelangen Ringes um Aussagen zur Diagonalisierbarkeit nicht-symmetrischer Matrizen eine gewisse Berechtigung haben, aber: Die strukturell bei weitem schönere Lösung dieser Fragestellung ist die Singulärwertzerlegung.

- Ein lockerer Ausblick in weiterführende Fragestellungen der Algebra (**Gleichungen höheren Grades**, insbesondere der Kegelschnitte und Quadriken) bildet in Kapitel 19 das Ende der vorliegenden Darstellung.

Teil I

Elementare Algebra und Lineare Algebra I

1 Einführung

1.1 Was ist Mathematik?

Mathematik, so haben Sie es sicher schon öfter gehört, steckt „irgendwie“ in allen wichtigen Anwendungen von Wissenschaft und Technik, und daher geht es in der Mathematik also hauptsächlich darum, irgendwelche solche wichtigen Dinge auszurechnen bzw. Methoden dafür zu entwickeln. Das ist wahrscheinlich zumindest der Grund, der Ihnen in der Schule dafür verkauft wurde, dass Sie sich dort einen größeren Teil Ihrer Schulzeit mit diesem Fach auseinandergesetzt haben und in der Tat das, wofür die Mathematik von Anwendern benutzt wird. Wenn Sie aber eine Mathematikerin fragen, wie sie ihr Fach „von innen betrachtet“ sieht, wird die Antwort höchstwahrscheinlich sein, dass die Mathematik für sie eher eine Geisteswissenschaft als eine Anwendungswissenschaft ist. Das mag verwundern, ist aber seit etwa 4000 Jahren das Selbstverständnis der Mathematik: Mathematik ist eine „in sich geschlossene“, im wahrsten Sinne des Wortes „abstrahierte“ Ideen- und Gedankenwelt, deren Anwendbarkeit auf Probleme der „echten“ Welt ihr zwar zu ihrem Aufschwung verholfen hat, die meisten Mathematikerinnen und Mathematiker aber eher sekundär interessiert. Herzlich willkommen bei der Befassung mit dieser Ideenwelt! Um genauer zu verstehen, was damit gemeint ist, lohnt es sich, einen Blick in die Geschichte der Mathematik zu werfen.

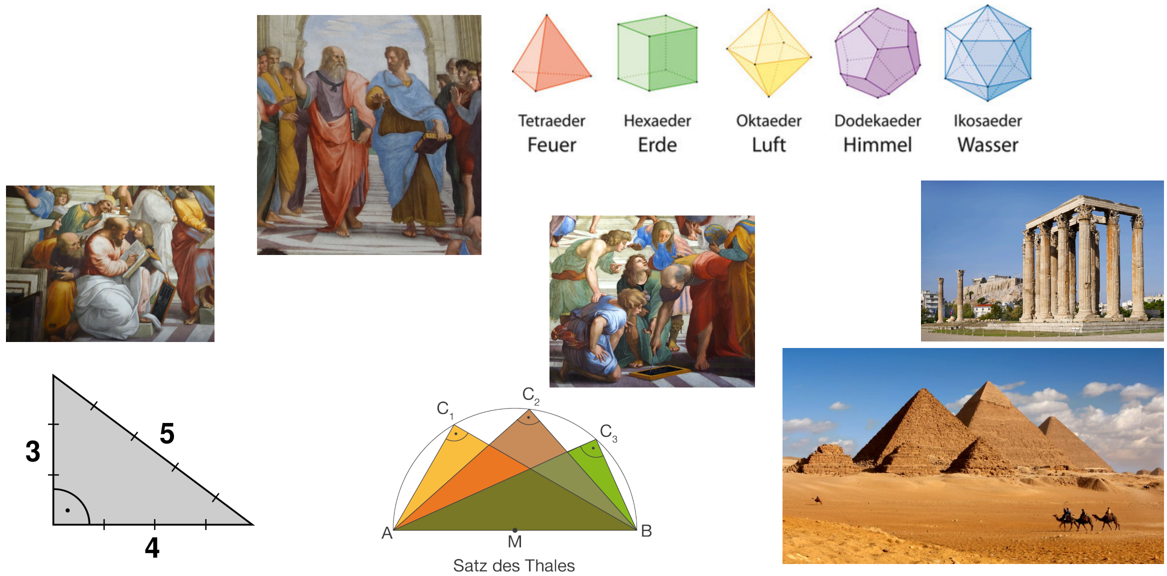
Im Vatikan, einen Raum vor der Sixtinischen Kapelle mit ihrem berühmten Deckenfresko mit Engelchen und Gottesfinger, findet sich ein Wandgemälde des Renaissancekünstlers Rafael (gemalt 1510-1511), auf dem das Who-is-Who der klassischen griechischen Gelehrten als „Wimmelbild“ darstellt ist – „Die Schule von Athen“.

Unter den abgebildeten „großen Denkern“ der Antike finden sich dort einige Mathematiker/Philosophen (so eindeutig waren die Berufsbezeichnungen damals nicht), deren Namen man wohl wenigstens schon mal gehört hat und die für unsere Darstellung hier interessant sind, zum Beispiel:

- Im Zentrum des Bildes (!): Platon („platonische Ideenwelt“) und Aristoteles („Begründer“ des logischen Schließens)
- Links mit Buch: Pythagoras (und seine Schüler, die Pythagoräer)
- Vermutlich (mit Tafel neben Pythagoras oder) ganz rechts, mit Zirkel: Euklid

Damals war die Mathematik zunächst als Anwendungswissenschaft (zum Beispiel als „Dienstleister“ für Architektur und Kunst) bedeutsam. Im alten Griechenland kristallisierte sich aber zum ersten Mal die geisteswissenschaftliche Auffassung von Mathematik heraus, die sich vielleicht so beschreiben lässt: Mathematik beschäftigt sich nicht mit Gegenständen des realen Lebens, sondern mit „Idealisierungen“ derselben. Sie haben in der Welt um Sie herum sicher noch nie einen Punkt, eine Gerade oder ein Dreieck in dem Sinne gesehen, wie sich die Mathematik diese vorstellt. Viel mehr sind solche mathematischen Gegenstände immer mehr oder minder abstrahiert, Platon würde sagen: Ideen (wie in „idealisiert“). Der große Gewinn für die Mathematik ist nun, dass diese Gegenstände so durch ein paar grundlegende Eigenschaften festgelegt sind, dass die Dinge also durch die der Mathematik Abstraktion (entgegen der landläufigen Meinung) oft *einfacher* werden, wenn man nur die gerade wesentlichen „Features“ dieser Objekte betrachtet. Erstaunlich ist, dass der menschliche Geist mit solchen abstrahierten Vorstellungen wie „Dreieck“ oder „Zahl“ (z.B. einer „Zwei“) relativ mühelos umgeht – und dass diese Reinheit der Gedanken die Menschen (manche jedenfalls) seit langem fasziniert. Zwei von vielen Beispielen für solche reinen, abstrakten Ideen, die aus der griechischen Antike stammen, sind die Platonischen Körper, die auf den Folien unten zu sehen sind – die einzigen fünf geometrischen Körper, die aus Kopien nur eines regelmäßigen Vielecks zusammengesetzt werden können – und die Untersuchung der „reinen Verhältnisse“ durch die Pythagoräer, die der Welt letztendlich die Bruchrechnung und die Saiteninstrumente geschenkt hat. (Was Ihnen besser gefällt, bleibt Ihnen überlassen.)

Ursprünge der modernen Mathematik



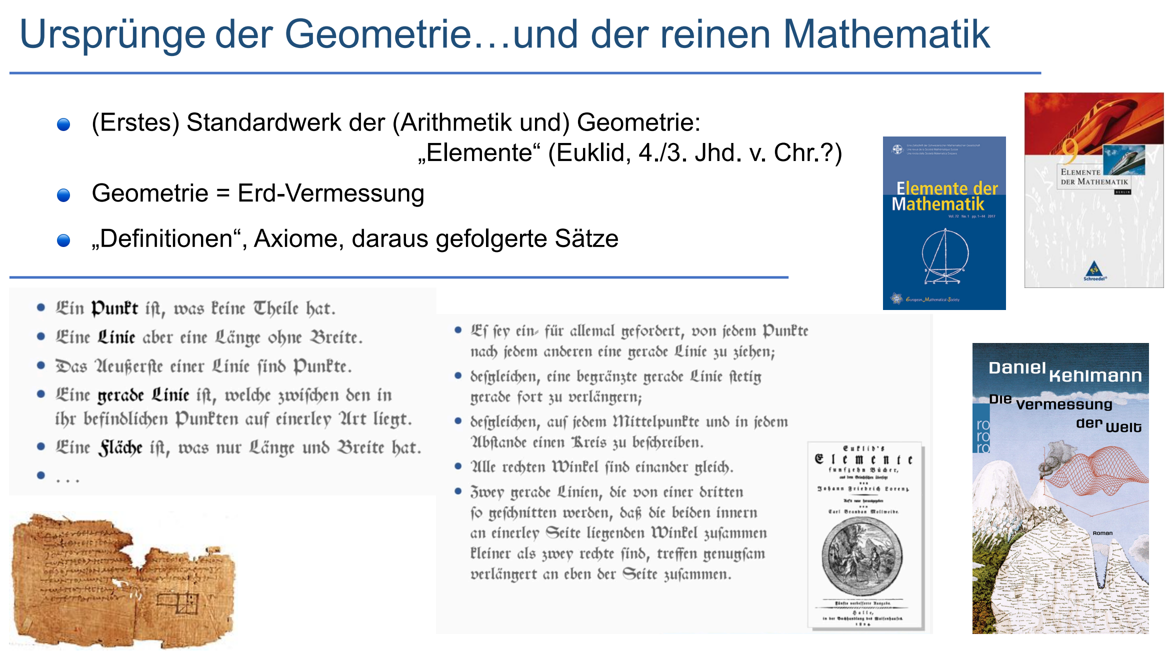
Das erste uns umfassend überlieferte Werk, das sich mit der Beschreibung solcher idealisierter Denkgegenstände und ihrer Beziehungen untereinander beschäftigte, sind die berühmten „Elemente“ von Euklid (3./2. Jhd. v. Chr.; für Leute, die Latein/Griechisch können oder einfach mal ein altes Buch sehen wollen: (Link); eine 1847 kolorierte, inzwischen sehr modern wirkende Web-Version: (Link)). Es enthält neben dem gesammelten, beeindruckenden geometrischen und zahlentheoretischen Wissen der damaligen Zeit (das noch immer einen recht großen Teil unserer Mittelstufenmathematik umfasst) auch das „Grundmuster“ für noch heute gültige Arbeitsweise der Mathematik, das für die Geometrie exemplarisch auch auf der Folie weiter unten dargestellt ist:

- Die Begriffe einer mathematischen Theorie (die „Denkgegenstände“) werden zunächst *definiert* (in der euklidischen Geometrie z.B. Linien, rechwinklige Dreiecke,...). Alle Eigenschaften eines Denkgegenstandes, die in der dargelegten Theorie benutzt werden, müssen sich aus dieser Definition ergeben.

- Zudem wird formuliert, welche Zusammenhänge man als für die Theorie offensichtlich wahr erachtet – das sind die *Axiome* der Theorie. (Z.B. bei Euklid „Durch zwei Punkte kann immer eine Gerade gezogen werden.“)

- Ein *mathematischer Satz* ist nun eine Aussage über die Gegenstände der Theorie (z.B. der Satz des Pythagoras einer über rechtwinklige Dreiecke). Ausführlich lautet ein solcher Satz immer „Unter Voraussetzung der geforderten Axiome ... und mit den gegebenen Definitionen ... gilt: $<$ Behauptung hier einfügen $>$“.

- In einer solchen mathematischen Theorie muss sich alles aus den postulierten Axiomen und Definitionen ableiten lassen können, das heißt: Ein jeder Satz muss, damit seine Gültigkeit akzeptiert wird, nur mit Hilfe der Definitionen und Axiomen *bewiesen werden*. Dazu „hangelt“ man sich mit Hilfe eines „Katalogs“ zulässiger Schlussregeln (der *formalen Logik*, als deren erster „Erfinder“ der zeitgenössische Aristoteles gilt) von den Axiomen und Definitionen der Theorie bis zur Behauptung. Schematisch könnte das so aussehen: „Aus den Axiomen und den Definitionen folgt... Daraus folgt... Dann stimmt auch... Daher muss dann schließlich auch die Behauptung gelten.“



Die Begriffsdefinitionen Euklids, zum Beispiel die für Punkt und Linie, erscheinen einem heute etwas eigentümlich und können auch den heutigen fachlichen Anforderungen nicht mehr standhalten (schauen Sie mal in die Links, insbesondere die Begriffsdefinitionen rein! (Link)). Auch an der Logik und den Grundlagen, die in den „Elementen“ zur Herleitung mathematischer Resultate benutzt wird, gab es noch einiges nachzubessern. Das alles wurde aber erst viel später klarer formuliert (19./Anfang 20. Jahrhundert, viel befördert durch David Hilbert, Göttingen). Aber was zählte war: Die *mathematische Methode*, basierend auf Idealisierung und logischer Strenge, war geboren, und „der Euklid“ wurde ein der Klassiker der griechischen Antike, der die Jahrhunderte überdauern sollte.

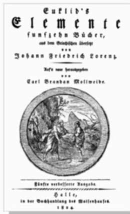
Mathematisches Schließen und Beweisen

Definitionen

- Ein **Punkt** ist, was keine Theile hat.
- Eine **Linie** aber eine Länge ohne Breite.
- Das Aeußerste einer Linie sind Punkte.
- Eine **gerade Linie** ist, welche zwischen den in ihr befindlichen Punkten auf einerley Art liegt.
- Eine **Fläche** ist, was nur Länge und Breite hat.
- . . .

Axiome

- Es sey ein- für allemal gefordert, von jedem Punkte nach jedem anderen eine gerade Linie zu ziehen;
- deßgleichen, eine begränzte gerade Linie stetig gerade fort zu verlängern;
- deßgleichen, auf jedem Mittelpunkte und in jedem Abstande einen Kreis zu beschreiben.
- Alle rechten Winkel sind einander gleich.
- Zwey gerade Linien, die von einer dritten so geschnitten werden, daß die beiden innern an einerley Seite liegenden Winkel zusammen kleiner als zwey rechte sind, treffen genugsam verlängert an eben der Seite zusammen.



„Math. Satz"

Unter Voraussetzung der geforderten Axiome… und mit den gegebenen Definitionen gilt:
< Behauptung hier einfügen >

Beweis:

- Mit Katalog zulässiger Schlussregeln (Logik, Aristoteles als „Erfinder") von Axiomen und Definitionen zur Behauptung
- „Aus den Axiomen und den Definitionen folgt … Daraus folgt … Dann stimmt auch… Daher muss dann auch die Behauptung gelten."

Mittelalter: Wiedergeburt alter Ideen (knapp 2000 Jahre später!)

Platon (Mitte li.) schrieb vier Lehrfächer („Mathema") als Vorbereitung zum Studium der Philosophie vor:

- Arithmetik
- Geometrie
- Harmonielehre
- Astronomie

Dies ist das *Quadrivium* („Vierweg") der mittelalterlichen Universität.

(*Trivium*: Grammatik, Dialektik, Rhetorik)

(„Die Schule von Athen", Rafael, im Vatikan)

Wir beamen uns zwei Jahrtausende weiter: Zwischendurch war, etwas polemisch gesagt, in der Mathematik in Mitteleuropa nicht viel los – das düstere Mittelalter hindurch wurde sich in Klosterschulen weiterhin hauptsächlich am Euklid festgeklammert, dieser handschriftlich kopiert und später auch als eines der ersten Bücher (nach der Bibel) per Buchdruck unters Volk gebracht. Gelobt wurde, wer unter den geistlichen Gelehrten die Inhalte der „Elemente“ Euklids vollends durchschaute; ansonsten war dort vielerorts das höchste der mathematischen Gefühle die Berechnung jährlichen Osterdatums, und in der weltlichen Welt wurde auf Märkten zwar handfest kaufmännisch gerechnet, und die Faszination für weitergehende Mathematik erschöpfte sich landläufig oft in Zahlenmystik, wie man sie etwa in „magischen Quadraten“ sieht (z.B. hier: (Link)).

Erst im Zeitalter der Renaissance (14-16. Jhd.), aus dessen Blütezeit in Italien unser Bild aus dem letzten Abschnitt stammt, wurden die Ideen der Griechen wieder auf einen Ehrenplatz gehoben (in der Mathematik natürlich besonders die „Elemente“). Nach einem Zwischenalter des geistigen „Verfalls“ sollte der nun durch Rückgriff auf antike Vorbilder das finstere Mittelalter überwunden werden. Schon früher waren die griechischen Lehranstalten zum Vorbild für die europäischen Universitäten geworden. Platon hatte dort einst vier Lehrfächer (auf griechisch so genannte „Mathema“, wie in Mathematik) als Vorbereitung zum Studium der Philosophie vorgeschrieben: Arithmetik (also die Grundrechenarten), Geometrie, Musiktheorie und Astronomie. Diese wurden zum klassischen Quadrivium („Vierweg“) der mittelalterlichen Universität zusammengefasst. Ergänzt wurden sie durch einen Trivium („Dreiweg“) aus Grammatik, Dialektik, Rhetorik, das waren insgesamt die klassischen „sieben freien Künste“ der mittelalterlichen Unis. Erhalten geblieben ist in der Mathematik das Wörtchen „trivial“, das von vielen für einen Sachverhalt benutzt wird, der nach kurzem Hinsehen leicht durchschaubar ist. Das zeugt wohl auch von der Wertschätzung, die manche Vierwegler gegenüber ihren Dreiweg-Kollegen hatten und haben.

1.2 Was ist analytische Geometrie? – Beschreibung von Ebene und Raum durch Koordinaten

Grob in die Zeit der Renaissance fällt auch die Formulierung einer der Hauptideen der *analytischen Geometrie* (René Descartes (1595-1650), Pierre de Fermat (1601-1665)): In der klassischen euklidischen Geometrie, die heute synthetische Geometrie genannt wird, wurden geometrische Objekte wie Kreise, Parabeln, Ellipsen und Hyperbeln über Eigenschaften ihrer Punkte beschrieben. Zum Beispiel wissen Sie wahrscheinlich, dass ein Kreis aus allen Punkten besteht, die von einem gegebenen Punkt (seinem Mittelpunkt) einen gegebenen Abstand (seinen Radius) haben; ebenso könnte man eine Parabel definieren als bestehend aus allen Punkten, die von einem gegebenen Punkt (ihrem Brennpunkt) denselben Abstand haben wie von einer gegebenen Linie (der Leitlinie). Eine Ellipse lässt sich über zwei Brennpunkte charakterisieren usw. Durch die Einführung von *Koordinatensystemen* in Ebene und Raum ändert sich die Herangehensweise gewaltig: Nun lassen sich geometrische Objekte/Probleme durch ihre *Koordinaten* beschreiben. Insbesondere kann man nun geometrische Fragestellungen lösen, indem man Gleichungen für ihre Koordinaten löst. Die Wissenschaft vom Gleichungslösen ist die *Algebra*, die sich als Lehre ab der Blütezeit des Islam (ab 800 n.Chr.) verbreitete und es durch Mithilfe von Mathematikern wie etwa Leonardo Fibonacci („Liber abaci", ca. 1200) auch nach Mitteleuropa geschafft hatte. Die Verknüpfung von Geometrie und Algebra, die nun stattfand, ist für viele Themen schon der Schulmathematik fundamental, z.B. gehören die quadratischen Funktionen und Gleichungen und ihre Verknüpfung mit dem geometrischen Objekt der Parabel zum Allgemeingut der Schulmathematik.

Was ist eigentlich Analytische Geometrie?

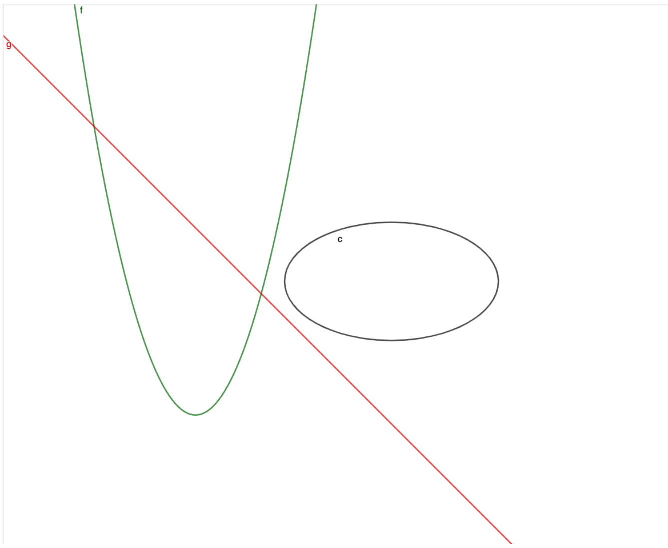
René Descartes
(1596-1650)

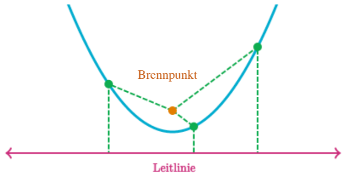
Pierre de Fermat
(1601-1665)

„Die Gleichungen kann man aber bequem versinnlichen, wenn man die beiden unbekannten Größen in einem Winkel (den wir meist gleich einem rechten nehmen) aneinandersetzt."

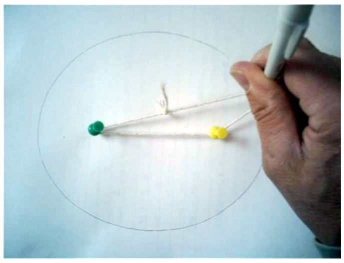
(Pierre de Fermat, Zitat aus *Neue Wege*)

Die Hauptidee der analytischen Geometrie





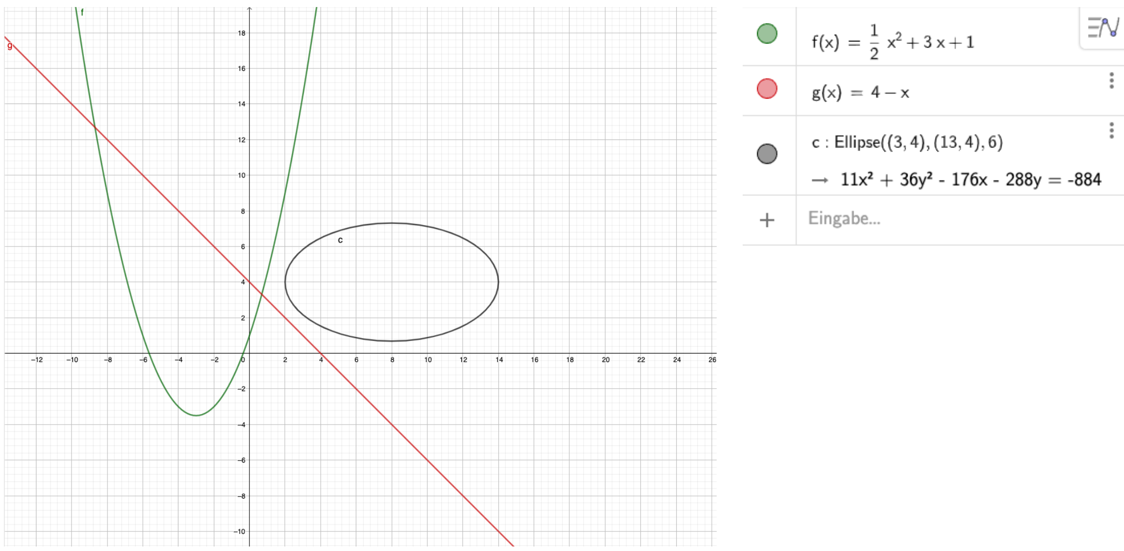
(khanacademy.org)



(Wikipedia: Ellipse)

$\Downarrow \quad \Downarrow \quad \Downarrow \quad \Downarrow \quad \Downarrow \quad \Downarrow \quad \Downarrow \quad \Downarrow \quad \Downarrow \quad \Downarrow \quad \Downarrow \quad \Downarrow \quad \Downarrow \quad \Downarrow \quad \Downarrow$

Die Hauptidee der analytischen Geometrie



(GeoGebra)

Geometrische Fragen lassen sich nun also mit Hilfe von Gleichungen formulieren; umgekehrt kann man Lösungsmengen von Gleichungen, also z.B. die Zusammenfassung derjenigen Zahlentupel (x, y, z), die eine gegebene Gleichung für drei Variablen erfüllen, durch geometrische Gebilde im Raum veranschaulichen (wie auf den Bildern aus der „Imaginary"-Ausstellung, s. Folie unten). Die Errungenschaft der Koordinatisierung von Ebene und Raum, so schlicht sie aus heutiger Perspektive erscheinen mag, kann tatsächlich in seiner Bedeutung für die weitere Entwicklung der Mathematik und ihrer Anwendungen in z.B. der Physik kaum überschätzt werden.

Umgekehrt: Lösungen von Gleichungen als geometrische Formen

(Ausstellung „Imaginary", Math. Forschungsinstitut Oberwolfach, imaginary.org)

1.3 Was ist Lineare Algebra?

Im ersten Teil des Skripts wird es hauptsächlich darum gehen, die Grundpfeiler des Theoriegebäudes der *Linearen Algebra* zu errichten. Was soll das nun wieder sein? Meistens bekommt man auf diese Frage von Leuten, die sich damit auskennen, eine der folgenden zwei Antworten:

- Die Lineare Algebra ist die Wissenschaft der linearen Gleichungssysteme.
- Die Lineare Algebra ist die Wissenschaft der Vektorräume und der linearen Abbildungen zwischen ihnen.

Am ehesten kann man wohl im Sinne der Versöhnung dieser beiden Ansichten sagen: Es geht um beides (bei uns in Kapitel VI bzw. in Kapitel VII und VIII), und später um den Zusammenhang beider Theorien (in Kapitel IX). Sie stellen fest: Das ist ja noch ziemlich lang hin! Das hat aber seinen Grund. Große mathematische Errungenschaften entstehen oft durch die *Verbindung* von Gebieten, die auf den ersten Blick nicht viel miteinander zu tun haben; zur Erinnerung: Das war ja schon bei der Analytischen Geometrie als Verbindung der Algebra mit der Geometrie so gewesen. Dazu müssen wir aber zunächst die Gebiete an sich kennen lernen. Daher wird es in den ersten Kapiteln darum gehen, zunächst einige grundlegende Ergebnisse der jahrhundertealten Lehre vom Gleichungslösen, der Algebra, und dann die Vektor- und Matrizenrechnung in Ebene, Raum und im $\mathbb{R}^n$ für sich kennen zu lernen. Wollen wir die Zusammenhänge zwischen diesen Gebieten (und später die grundlegenden Zusammenhänge der Linearen Algebra) erkennen und „auf der Höhe der Zeit“ formulieren, so benötigen wir eine allgemeine Einführung in einige grundlegende Strukturen der Algebra und auch der Mengenlehre und Logik, die sich gegen Anfang des 20. Jahrhunderts herausgebildet haben; all dies folgt in den nächsten Kapiteln.
Diese teilen sich in die zwei Hälften „Elementare Algebra und Lineare Algebra I“ und „Lineare Algebra II und Analytische Geometrie“ auf. Der erste Teil entspricht dem, was als normalerweise als Stoff für das erste Semester „Lineare Algebra“ gelehrt wird, der zweite dem zweiten. Viel Spaß!

2 Aussagen, Mengen, Funktionen und Gleichungen - Grundbegriffe der Algebra

In diesem ersten Kapitel wird unser Hauptziel sein, einige grundsätzlich wichtige Zusammenhänge zwischen Gleichungen und Funktionen herauszuarbeiten, der dann auch für die Fragestellungen der Linearen Algebra zentral sein wird.

2.1 Mengen und Aussagen

Im 19. Jahrhundert hat sich die Mengenlehre als grundlegendes Mittel zur Beschreibung mathematischer Zusammenhänge durchgesetzt. Unsere mathematischen Denkgegenstände (oder platonisch-mathematischen „Ideen“ des letzten Kapitels) werden daher immer Elemente klar definierter Mengen sein. Reden wir beispielsweise über Zahlen, so werden wir immer angeben, aus welcher Menge die betrachteten Zahlen kommen. Insbesondere werden ihre Eigenschaften und die mit ihnen durchführbare Operationen werden durch die Eigenschaften der Menge festgelegt.
Mathematische Erkenntnisse werden dann *immer* (!!!) in Form von Aussagen über die betrachteten Mengen bzw. ihre Elemente formuliert. Damit wir das auch so machen können, wollen wir die Begriffe der Menge, der Aussagen und der Bedingung genauer kennenlernen. Es ist Zeit für unsere ersten Begriffsdefinitionen.

Definition 2.1.1. *(Menge, Georg Cantor, 1895): Unter einer Menge M verstehen wir jede Zusammenfassung von bestimmten wohlunterschiedenen Objekten (m) unserer Anschauung oder unseres Denkens. Die Objekte in M heißen* Elemente von M.

Mengen werden auch heute noch teilweise relativ informell durch die Angabe dieser „Zusammenfassung“ oder der charakterisierenden Eigenschaften ihrer Elemente charakterisiert, z.B. „die Menge M aller EU-Mitgliedsstaaten“, „die Menge $\mathbb{N}$ der natürlichen Zahlen“ oder „die Menge N der natürlichen Zahlen, die kleiner als fünf sind“. Es ist üblich, in dem Stile wie bei den gerade genannten Beispielen auch gleich einen Bezeichner für die gerade definierte Menge einzuführen (hier: $M, \mathbb{N}, N$).

Eine weitere Möglichkeit ist die Angabe einer Menge durch Aufzählung ihrer Elemente. Diese ist präzise, falls die Aufzählung vollständig ist, und wieder etwas informell und möglicherweise missverständlich (aber auch gebräuchlich), wenn diese unvollständig ist:

- $M = \{\text{Deutschland, Frankreich, Dänemark}, \ldots\}$,
- $\mathbb{N} = \{1, 2, 3, \ldots\}$[1]
- $N = \{1, 2, 3, 4\}$
- Eine besonderes, an vielen Stellen praktisches formales Konstrukt ist die *leere Menge*, die keine Elemente enthält und die mit $\emptyset$ oder $\{\}$ bezeichnet wird.

Als Schreibweise für „m ist ein Element von M“ wird das „Kürzel“ $m \in M$ verwendet, für das Gegenteil die Schreibweise $m \notin M$. Zum Beispiel ist

- Deutschland $\in M$, Argentinien $\notin M$;
- $3 \in N, 3 \in \mathbb{N}, \pi \notin N$.

Mengen sind gleich genau dann, wenn sie dieselben Elemente enthalten. Es ist also z.B. auch $N = \{2, 3, 3, 4, 1, 3, 2\}$ eine weitere Darstellung der Menge N.

Wichtige Beispiele von Mengen, die wir oft benutzen werden, sind die Zahlbereiche

- $\mathbb{N} := \{1, 2, 3, \ldots\}$ – die natürlichen Zahlen,
- $\mathbb{Z} := \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$ – die ganzen Zahlen,
- $\mathbb{Q} := \{\frac{p}{q} \mid p \in \mathbb{Z}, q \in \mathbb{N}\}$ – die rationalen Zahlen oder Bruchzahlen,
- $\mathbb{R}$ – die reellen Zahlen, die eins zu eins den Zahlen auf der Zahlengeraden entsprechen (die also im Gegensatz zu $\mathbb{Q}$ auf dieser die Lücken schließt, an denen sich irrationale Zahlen wie $\sqrt{2}$, e oder π befinden.)

Sie haben vielleicht gemerkt, dass die Definitionen der Zahlbereiche hier recht wattig sind (Was soll z.B. diese Symbole „1“, „2“ usw. eigentlich sein?) und von Beispiel zu Beispiel immer schwammiger werden. Uns soll aber dieses „halbformale Niveau“ an dieser Stelle zunächst reichen. Wir werden es im Laufe der kommenden Kapitel begründet weiter formalisieren.

[1] Bei uns ist $\mathbb{N} = \{1, 2, 3, \ldots\}$ und $\mathbb{N}_0 = \{0, 1, 2, 3 \ldots\}$, das kann sich von Dozent zu Dozent unterscheiden.

Definition 2.1.2. *(Teilmenge): Eine Menge T heißt Teilmenge von M, wenn jedes Element aus T ist auch in M enthalten ist. Man schreibt dafür $T \subseteq M$.*

Im Beispiel oben gilt also $N \subseteq \mathbb{N}$, für die Zahlbereiche ist $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}$.

Bitte vermeiden Sie den Anfängerfehler, dass Sie eine *Menge* und die *Elemente dieser Menge* verwechseln. Es ist ein Unterschied, ob Sie von einer natürlichen Zahl $n \in \mathbb{N}$ reden (das ist eine Zahl, zu der Sie z.B. eine andere natürliche Zahl addieren könnten oder entscheiden könnten, ob $n \in N$ gilt) oder von der Menge der natürlichen Zahlen – das ist eine Menge, von der man z.B. Teilmengen betrachten kann. Es handelt sich bei n und $\mathbb{N}$ also um komplett verschiedene mathematische Gegenstände.
Die obige „Cantor-Definition" des Begriffs „Menge" ist wie die Definition der Zahlbereiche in heutigem Sinne zu unpräzise und führt, angewandt auf die richtigen Mengen, zu logischen Widersprüchen. Viele Leute sprechen daher heute von der *naiven Mengenlehre*, wenn sie die Cantor-Definition und darauf fußende Konventionen benutzen. Für unsere Zwecke wird sie jedoch ausreichen. Den heutzutage bevorzugten, „minimalistischeren" Aufbau, in dem nur Mengen erlaubt sind, die sich aus einigen wenigen klar definierten „Grundmengen" durch einen zulässigen Katalog von Operationen erzeugen lassen, lernen wir grob in Kapitel 7) kennen. Eine mit beiden Konzepten kompatible Vorgehensweise, die wir oft verwenden werden, ist die *Aussonderung* von Elementen einer Menge durch eine Bedingung; z.B. lässt sich $N = \{1, 2, 3, 4\}$ auch durch Aussonderung bestimmter natürlicher Zahlen charakterisieren: Es ist

$$M = \{n \in \mathbb{N} \mid n \leq 4\}.$$

sprich: M enthält diejenigen $n \in \mathbb{N}$, die die Bedingung $n \leq 4$ erfüllen. Wir definieren genau, was wir unter einer Bedingung verstehen wollen. Dazu benötigen wir zunächst den Begriff der mathematischen Aussage:

Definition 2.1.3. *(Aussage): Eine (mathematische) Aussage ist ein sprachlicher Ausdruck, dem eindeutig einer der beiden Wahrheitswerte „wahr" oder „falsch" zugeordnet werden kann.*

Die Ausdrücke „Dienstag ist ein Wochentag.", „$1 = 1$.", „Dienstag ist ein Monat.", „$1 = 2$." und „Wenn $x = 3$ ist, ist $x + 4 = 9$." sind Beispiele für (teils wahre, teils falsche) Aussagen. „1+3" und „x - 3 = 4" sind keine Aussagen. (Letzteres ist eine Bedingung, s.u.).

Gängige Sprechweisen in der Formulierung mathematischer Sätze sind die Sätzchen „Es gilt...“ oder „Dann ist...“, um zu signalisieren, dass im Folgenden eine wahre Aussage formuliert wird.

Definition 2.1.4. *(Bedingung):*

Bezeichne X eine Menge. Dann nennen wir einen sprachlichen Ausdruck $B(x)$ eine Bedingung an Elemente x von X, wenn $B(x)$ eine Variable x enthält (oder auch mehrere Variablen) und durch Einsetzen von Elementen $x \in X$ immer zu einer Aussage wird.

Eine Bedingung lässt sich also für Elemente von X immer als wahr oder falsch entscheiden. Die „Nicht-Aussage“ $x + 3 = 7$ von oben ist also eine Bedingung, die darüber per Aussonderung definierte Menge

$$S = \{x \in \mathbb{N} \mid x + 3 = 7\}$$

eine andere Schreibweise für die Menge $S = \{4\}$. Mit Hilfe von Bedingungen an Elemente von X lassen sich Teilmengen T einer Menge X folgendermaßen definieren:

$$T := \{x \in X \mid B(x)\}.$$

Andere Beispiele für solche Bedingungen an natürliche Zahlen, mit deren Hilfe man solche Teilmengen der natürlichen Zahlen definieren kann, sind z.B.: „n ist eine Primzahl.“, „n ist durch 3 teilbar.“, „$n \geq 5$“, „n ist eine Quadratzahl, d.h. $n = m^2$ für ein $m \in \mathbb{N}$.“.

Vielleicht kennen Sie die die Intervallschreibweise für Teilmengen von $\mathbb{R}$, die sich über Aussonderung folgendermaßen definieren lässt: Man setzt

$$[a, b] := \{x \in \mathbb{R} \mid a \leq x \leq b\} \quad \text{und} \quad [a, b[:= \{x \in \mathbb{R} \mid a \leq x < b\}.$$

Eine weitere, nicht sehr verbreitete, aber nützliche Schreibweise, die wir nutzen werden, ist die Schreibweise:

$$\underline{n} := \{i \in \mathbb{N} \mid 1 \leq i \leq n\}.$$

Ein großer Teil der Formulierung mathematischer Theorien besteht darin, zu erkennen und zu beweisen, dass auf den ersten Blick verschieden klingende Bedingungen gleichwertig, auf hochmathematisch: äquivalent sind oder, dass die Gültigkeit einer Bedingung immer eine die einer anderen nach sich zieht.

Definition 2.1.5. *(Implikation und Äquivalenz):*
Sei X eine Menge und $B_1(x)$ und $B_2(x)$ Bedingungen für $x \in X$.

- *Wir sagen „$B_1(x)$ impliziert $B_2(x)$", „Aus $B_1(x)$ folgt $B_2(x)$", „$B_1(x)$ ist hinreichend für $B_2(x)$" oder „unter der Voraussetzung $B_1(x)$ gilt $B_2(x)$" (für alle $x \in X$), falls gilt:*
$$\{x \in X \mid B_1(x)\} \subseteq \{x \in X \mid B_2(x)\}. \tag{2.1}$$
- *Wir sagen „$B_1(x)$ ist äquivalent zu $B_2(x)$" (für alle $x \in X$), falls aus $B_1(x)$ $B_2(x)$ folgt und umgekehrt auch $B_2(x)$ $B_1(x)$ impliziert, d.h. falls*
$$\{x \in X \mid B_1(x)\} = \{x \in X \mid B_2(x)\}$$
gilt.

Beispielsweise impliziert für $n \in \mathbb{N}$ die Bedingung „n ist durch 4 teilbar." die Bedingung „n ist gerade.", die Bedingung „n ist gerade." ist äquivalent zu „n^2 ist gerade." (Letzteres ist für Sie vielleicht nicht offensichtlich, und ein Beweis könnte Sie hier überzeugen.)

Wir hatten oben festgelegt, dass zwei Mengen M_1, M_2 genau dann gleich sind, wenn sie die gleichen Elemente enthalten. Das ist der Fall, wenn aus $x \in M_1$ auch immer $x \in M_2$ und umgekehrt aus $x \in M_2$ auch immer $x \in M_1$, und wir werden auf diese Weise oft die Gleichheit von Mengen beweisen. Insbesondere können wir zeigen, dass zwei Bedingungen $B_1(x)$ und $B_2(x)$ äquivalent sind, indem wir

$$\{x \in X \mid B_1(x)\} \subseteq \{x \in X \mid B_2(x)\} \quad \text{und} \quad \{x \in X \mid B_2(x)\} \subseteq \{x \in X \mid B_1(x)\}$$

zeigen.

Gleichungen, um die es in der Algebra und im Folgenden gehen soll, können wir als spezielle Bedingungen lesen. Zum Beispiel ist die quadratische Gleichung $x^2 + 3x + 7 = 0$ eine Bedingung an reelle Zahlen, und die Frage nach der/den Lösungen dieser Gleichung ist die Frage, wie die Menge $\{x \in \mathbb{R} \mid x^2 + 3x + 7 = 0\}$, die Lösungsmenge der Gleichung, explizit aussieht. Um zu sagen, was wir konkret unter Gleichungen verstehen wollen, brauchen wir zunächst den im Folgenden vorgestellten, fundamentalen Begriff.

2.2 Funktionen

Der Begriff der Funktion spielt in der Schule und auch in vielen Teildisziplinen der universitären Mathematik eine grundlegende Rolle. In diesem Abschnitt wird der Funktionsbegriff formal definiert[2] werden und wir untersuchen einige Eigenschaften von Funktionen – besonders die, die wir in den darauf folgenden Abschnitten in den Zusammenhang mit der Lösbarkeit von Gleichungen bringen wollen.

Wir werden dabei die Begriffe „Funktion“ und „Abbildung“ synonym benutzen. Die Begriffe haben sich zwar historisch getrennt entwickelt; ihre genauere Untersuchung und Abstrahierung hat aber gezeigt, dass diese beide von der folgenden Definition umfasst werden, der für Definitionsbereich und Wertevorrat beliebige Mengen zulässt.

Definition 2.2.1. *(Abbildung/Funktion)*

Seien X und Y Mengen. Eine <u>*Abbildung*</u> *oder* <u>*Funktion*</u> *von X nach Y (oder „von X in Y“) ist eine Vorschrift f, die* <u>*jedem*</u> *$x \in X$* <u>*genau ein*</u> *Element aus Y zuordnet; dieses Element heißt das Bild von x unter f und wird mit $f(x)$ bezeichnet. Wenn f eine Abbildung von X nach Y ist, so schreibt man dafür auch $f : X \to Y$. Man nennt X den* <u>*Definitionsbereich*</u> *und Y den* <u>*Wertevorrat*</u> *von f.*

Beispiele für Abbildungen, Bemerkungen:

- Abbildungen von endlichen Teilmengen von $\mathbb{N}$ in endliche Teilmengen von $\mathbb{N}$, also $f : \underline{n} \to \underline{m}$ mit $n, m \in \mathbb{N}$. Beispiele sind

$$f_1 : \underline{3} \to \underline{4}, \begin{cases} 1 & \mapsto & 2 \\ 2 & \mapsto & 2 \\ 3 & \mapsto & 1 \end{cases}, \quad f_2 : \underline{4} \to \underline{3}, \begin{cases} 1 & \mapsto & 1 \\ 2 & \mapsto & 3 \\ 3 & \mapsto & 2 \\ 4 & \mapsto & 1 \end{cases}, \quad \text{und} \quad f_3 : \underline{3} \to \underline{4}, \begin{cases} 1 & \mapsto & 3 \\ 2 & \mapsto & 2 \\ 3 & \mapsto & 1 \end{cases};$$

  keine Funktionen sind die Zuordnungen

$$f_4 : \underline{3} \to \underline{4}, \begin{cases} 1 & \mapsto & 2 \\ 2 & \mapsto & 3 \end{cases}, \quad \text{und} \quad f_5 : \underline{3} \to \underline{4}, \begin{cases} 1 & \mapsto & 2 \\ 1 & \mapsto & 3 \\ 2 & \mapsto & 2 \\ 3 & \mapsto & 1 \end{cases},$$

  da hier die Bedingungen „jedem $x \in X$“ bzw. „genau ein“ verletzt sind.

[2]Formalisierungstand ca. 1900, für eine weitere Formalisierungsstufe vgl. Kapitel 7

- Abbildungen werden oft mit Hilfe einer Berechnungs- oder Zuordnungsvorschrift angegeben. Für aus der Schule bekannte Funktionen, die von $\mathbb{R}$ nach $\mathbb{R}$ abbilden, sind
$$f(x) = x^2, \quad g(x) = \sin(x), \quad \text{und} \quad h(x) = a^x$$
(für $a > 0$) oder präziser
$$f : \mathbb{R} \to \mathbb{R}, \quad x \mapsto x^2, \qquad g : \mathbb{R} \to \mathbb{R}, \quad x \mapsto \sin(x), \quad \text{und} \quad h : \mathbb{R} \to \mathbb{R}, \quad x \mapsto a^x \tag{2.2}$$
Beispiele.

- Eine recht langweilige Abbildung, die aber später eine tiefergehende Bedeutung haben wird, ist die *Identität* oder *identische Abbildung* auf einer Menge X: $\mathrm{id}_X : X \to X, x \mapsto x$. Der Graph dieser Abbildung für $X = \mathbb{R}$ ist die Winkelhalbierende im ersten/dritten Quadranten des Koordinatensystems.

- In den kommenden Abschnitten werden wir uns Abbildungen von $\mathbb{R}^n$ nach $\mathbb{R}^m$ ansehen, z.B. $\mathbb{R}^2$ nach $\mathbb{R}^2$, die interpretierbar sind als Abbildungen der Ebene in sich selbst.

- Funktionen stecken überall, und es ist manchmal nicht ganz einfach, aber oft nützlich, sie zu erkennen. Zum Beispiel gilt: Ist X eine Menge und $B(x)$ eine Bedingung an $x \in X$, so ist
$$w_B : X \to \{w, f\}, \qquad x \mapsto \text{Wahrheitswert von } B(x)$$
eine Funktion.

Wir werden im Folgenden mehrere Begriffe für Eigenschaften und Attribute von Funktionen kennen lernen. So wird es oft sein – ist ein mathematischer Begriff erstmal definiert, beginnen wir, seine „Features“ kennen zu lernen, diese formal zu definieren und den Begriff weiter zu unterteilen, hier also bestimmte Funktionen mit Hilfe ihrer Eigenschaften zu charakterisieren und die Gesamtheit aller Funktionen weiter nach Eigenschaften zu untergliedern (z.B. unten in injektive/nicht injektive Funktionen, umkehrbare/nicht umkehrbare Funktionen, lineare/nicht lineare Funktionen). Bitte lernen Sie solche Begriffe wie Vokabeln – es ist das Vokabular, das Mathematikerinnen benutzen, um sich zu verständigen.

Einige solche Sprechweisen zu Funktionen sind die folgenden:

- Funktionen „werden angewandt auf Elemente“ $x \in X$,
- Elemente aus X werden durch f abgebildet – sie bilden nicht selbst ab, eine Sprechweise wie „x bildet ab auf x^2“ ist (außer unter manchen Erstsemestern) nicht gebräuchlich, besser ist „x *wird abgebildet* auf x^2“.

Ein weiterer wichtiger Begriff für Funktionen ist der des Bildes/Urbildes. Dieser ist so bedeutsam, dass wir dafür eine Definition formulieren, in der die verschiedene Gebrauchsformen zusammengefasst sind:

Definition 2.2.2. *(Bild einer Funktion; Bild und Urbild einer Menge)*

Seien X, Y Mengen, $f : X \to Y$ eine Funktion.

- *Das Bild von f ist die mengenmäßige Zusammenfassung aller möglichen Funktionswerte $f(x)$, die f annimmt, in Mengennotation:*

$$\begin{aligned} \mathit{Bild}(f) &= \{y \in Y \mid y \text{ ist Bild eines } x \in X \text{ unter } f\} &(2.3)\\ &= \{y \in Y \mid \text{Es existiert ein } x \in X \text{ mit } y = f(x)\}. &(2.4)\end{aligned}$$

  beziehungsweise etwas informeller[3], aber vielleicht verständlicher

$$\mathit{Bild}(f) := \{f(x) \mid x \in X\}.$$

- *Ist A eine „Teilmenge von X“, also eine Menge von Elementen aus X, so bezeichnet $f(A)$ die Menge aller zugehörigen Bilder, es ist also*

$$f(A)_{=}\{y \in Y \mid y = f(a) \text{ für ein } a \in A\} = \{f(a) \mid a \in A\}.$$

- *Ist $y \in Y$, so nennt man die Menge*

$$f^{-1}(y) := \{x \in X \mid f(x) = y\} \quad \subseteq X$$

  das Urbild von y (unter f).

[3] deshalb, weil es nicht genau unserer Definition der Aussonderungsschreibweise entspricht

- *Ist $B \subseteq Y$, so ist*
$$f^{-1}(B) := \{x \in X \mid f(x) \in B\} \quad \subseteq X$$
das Urbild der Menge B (unter f).

Das Urbild eines Element $y \in Y$ kann auch die leere Menge $\emptyset$ sein; das ist genau dann der Fall, wenn $y \notin \mathrm{Bild}(f)$ ist.

Bemerkung (Gleichheit im Allgemeinen, Gleichheit von Funktionen):

Wenn wir nicht über Zahlen reden (sondern über Mengen, Funktionen, ...), sollten wir uns Gedanken machen und *definieren*, was die Gleichheit dieser Objekte überhaupt bedeuten soll.[4] Das gibt uns dann auch eine Anleitung, wie man dann *zeigt*, dass zwei der jeweiligen Objekte gleich sind, vgl. z.B. den Nachweis der Gleichheit von zwei Mengen in Abschnitt 2.1.

Für zwei Funktionen $f, g : X \to Y$ etwa schreiben wir $f = g$, wenn sie in *allen ihren Funktionswerten* übereinstimmen, d.h. wenn für alle $x \in X$ gilt $f(x) = g(x)$. Insbesondere kann man $f = g$ durch nachrechnen von $f(x) = g(x)$ für alle x prüfen, z.B. durch Umformung des Funktionsterms von f in den von g oder durch Vergleich aller Wertepaare (falls das möglich ist).

[4] Wenn man es genau nimmt, sollte man sich das für Zahlen auch überlegen. Ist $\frac{3}{7} = \frac{27}{63}$, und wenn ja, nach welcher Definition eigentlich? Aber das ist eine andere Geschichte.

2.3 Gleichungen

In vielen Kulturen, schon seit Jahrtausenden wurden Problemstellungen mathematisch formuliert, die auf Gleichungen verschiedener Typen führen. Gleichungen, wie zum Beispiel

$$3x = 9, \qquad 3x + 4y = 2, \qquad x^2 + 3x + 7 = 0$$

sind in der Sprache des vergangenen Kapitels *Bedingungen* für Zahlen, aus denen sich aber die Zahlen für gewöhnlich nicht direkt ablesen lassen; kann man trotzdem alle Zahlen ermitteln, die die gegebene Bedingung erfüllen, so hat man die Gleichung gelöst. Die Algebra *ist* die „Wissenschaft vom Gleichungslösen"; ihr Name leitet sich ab aus dem Titel aus einer berühmten Sammlung praktischer Lösungsverfahren für Gleichungen, das aus der arabischen Blütezeit um 800 n.Chr. stammt und den Titel „Hisab al-dschabr wa-l-muqabala", zu deutsch „Buch vom Rechnen durch Ergänzung und Ausgleich" trägt. Verfasst wurde es durch den Mathematiker Al-Chwarizmi („Algorismi", wie in „Algorithmus"), der u.a. im heutigen Iran und Usbekistan lebte und dessen Schriften und dort zusammengefasste Methoden sich später in Richtung Europa verbreiteten. Dort wurden die algebraischen Methoden zur Lösung von Gleichungen weiterentwickelt; als weiteres bedeutendes Werk der prä-formalen Algebra (die ab dem 20. Jhd. bzw. ab Kapitel IV die Oberhand gewinnen soll) sei hier noch Leonhard Eulers „Vollständige Anleitung zur Algebra" von 1770 erwähnt und verlinkt (Link).

Um den Bogen zu unserem formalen Rahmen zu spannen, ist die folgende Feststellung wichtig: Mathematische Fragestellungen und die daraus abgeleiteten Gleichungen lassen sich gleichwertig als Frage nach dem Urbild eines gegebenen Funktionswertes einer Funktion ausdrücken, z.B.

- Ziehe die Wurzel aus b. $\rightsquigarrow$ Bestimme für $f(x) = x^2$ ein Urbild von b, d.h. ein $x \in \mathbb{R}$ mit $x^2 = b$. (Die Wurzel aus b ist dann diejenige Zahl, für die zusätzlich $x > 0$ gilt.)
- Löse die quadratische Gleichung $x^2 + 3x - 7 = 0$. $\rightsquigarrow$ Bestimme, wenn möglich, das Urbild der Null unter $f(x) = x^2 + 3x - 7$.
- Bestimme $\ln(b)$. $\rightsquigarrow$ Bestimme ein Urbild von b unter $f(x) = e^x$, d.h. ermittle ein $x \in \mathbb{R}$ mit $e^x = b$.
- Bestimme aus dem Verhältnis b von Gegenkathete und Hypotenuse, also $\sin(\alpha)$, den zugehörigen Winkel α $\rightsquigarrow$ Löse $\sin(\alpha) = b$.

Alle diese Aufgaben müssten ein bisschen präziser gestellt werden: Welche rechten Seiten b sind überhaupt zulässig (d.h., es gibt eine zugehörige Wurzel, einen zugehörigen Logarithmus, einen zugehörigen Winkel)? Wann ist eine Lösung eindeutig? Um uns diesem Ziel zu nähern, halten wir zunächst die allgemeine Form solcher Gleichungen, die wir im Folgenden behandeln wollen, in der folgenden Definition fest.

Definition 2.3.1. *(Gleichung, Lösungsmenge)*

Sei X eine Menge. Eine Bedingung (im Sinne von Def 2.1.4) heißt Gleichung für $x \in X$, wenn sie sich in der Form

$$f(x) = b \tag{2.5}$$

mit einer Funktion $f : X \to Y$ und mit fester, vorgegebener rechter Seite $b \in Y$ formulieren lässt. Ein $x \in X$, das diese Bedingung (2.5) erfüllt, heißt Lösung der Gleichung. Die zugehörige Lösungsmenge der Gleichung ist

$$\mathbb{L} = \{x \in X \mid f(x) = b\}. \tag{2.6}$$

Zwei Gleichungen $f(x) = b,\ g(x) = c$ heißen äquivalent über der Menge X, wenn sie als Bedingungen äquivalent sind, d.h. wenn ihre Lösungsmengen gleich sind,

$$\{x \in X \mid f(x) = b\} = \{x \in X \mid g(x) = c\}.$$

Ist man nur an der Lösungsmenge interessiert, kann man äquivalente Gleichungen gegeneinander austauschen, das nennt man eine Äquivalenzumformung. Äquivalenzumformungen haben meist das Ziel, eine Gleichung gegen eine *einfachere* zu ersetzen, am besten gegen eine, aus der man $\mathbb{L}$ direkt ablesen kann.

Auch wenn die Menge X, über der die Lösungen einer Gleichung $f(x) = b$ gesucht werden, oft nicht explizit genannt wird (meist, weil sie implizit klar ist, manchmal auch aus Schlampigkeit), ist sie doch entscheidend, wie die beiden folgenden Beispiele hoffentlich klarmachen:

- Sucht man die Lösungsmenge $\mathbb{L}$ der Gleichung $(x^2 + 4) \cdot (x^2 + 2) = 0$, so ist es entscheidend, ob diese für $x \in \mathbb{N}$ (dann ist $\mathbb{L} = \{2\}$), $x \in \mathbb{Z}$ (dann ist $\mathbb{L} = \{-2, 2\}$) oder $x \in \mathbb{R}$ gestellt wird (dann wäre $\mathbb{L} = \{-2, 2, -\sqrt{2}, \sqrt{2}\}$).

- Es gibt Gleichungen, die über einer Menge äquivalent sind, über einer anderen nicht. Für *positive* reelle Zahlen x ist das Ersetzen der Gleichung $x^2 = b$ (mit $b > 0$) durch die Gleichung $x = \sqrt{b}$ eine Äquivalenzumformung. Suchen wir alle Lösungen in $X = \mathbb{R}$, so ist $\mathbb{L} = \{-\sqrt{b}, \sqrt{b}\})$, die Lösungsmenge von $x = \sqrt{b}$.

Wegen solcher Subtilitäten ist im Zweifel eher mehr als weniger Genauigkeit angebracht, d.h. insbesondere schreiben wir immer dazu, über welcher Menge wir z.B. eine Gleichung lösen wollen, d.h. aus welcher Menge zulässige Lösungen kommen sollen!

Um Aussagen über die Lösbarkeit von Gleichungen machen zu können, werden sie werden oft zunächst durch die Eigenschaften der Funktion f klassifiziert, z.B.

- quadratische Funktion $\rightsquigarrow$ quadratische Gleichung,
- trigonometrische Funktion (aus sin, cos, tan gebaut) $\rightsquigarrow$ trigonometrische Gleichung,
- Exponentialfunktion (s.o.) $\rightsquigarrow$ Exponentialgleichung usw.

2.4 Lösbarkeit von Gleichungen und Eigenschaften der zugehörigen Funktionen

Uns interessiert nun die Frage nach Lösbarkeitsaussagen für Gleichungen wie $x^2 = b$, $e^x = b$, $\sin(x) = b$ (wobei b jeweils beliebige Werte aus $\mathbb{R}$ bezeichnet). Klar sollte sein:

- Damit es für Gleichungen $f(x) = b$ überhaupt eine Lösung geben kann, muss b im Bild von f liegen.
- Damit eine Lösung eindeutig ist, darf f nicht zwei verschiedene Werte x_1, x_2 gleichermaßen auf die rechte Seite $b \in Y$ abbilden.

Im Idealfall sind wir an globalen Aussagen interessiert, also Aussagen für *alle möglichen* rechten Seiten $b \in Y$, interessiert. Die folgenden beiden Fragestellungen sind typische mathematische Fragestellung, für Gleichungen und auch darüber hinaus:

- (globale) Existenz von Lösungen: Gilt eine Aussage der Form „Für alle $b \in Y$ existiert für die Gleichungen von Typ $f(x) = b$ stets eine Lösung“?
- (globale) Eindeutigkeitsaussage: Gilt eine Aussage der Form „Für alle $b \in Y$ gilt: Wenn $f(x) = b$ eine Lösung hat, ist sie dann eindeutig“?

Diese Fragen lassen sich für gewöhnlich entscheiden, indem man sich die involvierte Funktion f genauer ansieht. Dazu definieren wir die Begriffe der Injektivität und Surjektivität von Funktionen.

Für Gleichungen vom Typ 2.5 ist globale Lösbarkeit (also Existenz von Lösungen für beliebige rechte Seite $b \in Y$) garantiert, falls f eine so genannte surjektive Funktion ist.

Definition 2.4.1. *(Surjektivität):*
Eine Funktion $f : X \to Y$ heißt surjektiv, wenn eine der folgenden gleichwertigen Aussagen erfüllt ist:

- *Jeder Wert $b \in Y$ wird als Funktionswert angenommen.*
- *Zu jedem $b \in Y$ existiert mindestens ein $x \in X$ mit $f(x) = b$.*
- *Für jede rechte Seite $b \in Y$ hat die Gleichung $f(x) = b$ mindestens eine Lösung.*
- *Es ist $\text{Bild}(f) = Y$, d.h. das Bild von f ist der gesamte Wertevorrat der Funktion.*
- *Jedes $b \in Y$ besitzt mindestens ein Urbild unter f.*

Eindeutigkeit von Lösungen ist unabhängig von der rechten Seite $b \in Y$ gesichert, falls f injektiv ist.

Definition 2.4.2. *(Injektivität): Eine Funktion heißt injektiv, wenn eine der folgenden gleichwertigen Aussagen erfüllt ist:*

- *Verschiedene Werte aus X werden durch f stets auf verschiedene Bilder abgebildet, d.h. für alle $x_1, x_2 \in X$ gilt:*

$$\text{Wenn} \quad x_1 \neq x_2, \quad \text{dann auch} \quad f(x_1) \neq f(x_2).$$

- *Für jede rechte Seite $b \in Y$ hat die Gleichung $f(x) = b$ höchstens eine Lösung.*
- *Zu jedem $b \in Y$ existiert höchstens ein Urbild.*
- *Wenn zwei Funktionswerte $f(x_1)$ und $f(x_2)$ gleich sind, dann müssen schon x_1 und x_2 gleich gewesen sein, d.h. für alle $x_1, x_2 \in X$ gilt:*

$$\text{Wenn} \quad f(x_1) = f(x_2), \quad \text{dann auch} \quad x_1 = x_2.$$

Bemerkungen dazu:

- Von den Funktionen f_1, f_2, f_3 von Seite 30 ist die Funktionen f_1 also weder surjektiv noch injektiv, die Funktion f_2 nur surjektiv und die Funktion f_3 nur injektiv.
- Für die reellen Funktionen $f, g, h : \mathbb{R} \to \mathbb{R}$ aus 2.2 gilt:
  - f ist nicht injektiv, denn z.B. $f(1) = f(-1)$, und nicht surjektiv, denn z.B. $-1 \notin \text{Bild}(f)$.
  - g ist nicht injektiv, denn z.B. $\sin(0) = \sin(\pi)$, und nicht surjektiv, denn z.B. $2 \notin \text{Bild}(g)$.
  - h ist injektiv, aber nicht surjektiv, denn z.B. $-1 \notin \text{Bild}(h)$.

  Es lassen sich aber Definitions- und Wertebereich so wählen, dass f injektiv, surjektiv oder beides wird. Zum Beispiel ist die Funktion $x \mapsto x^2$ als Funktion $\mathbb{R}_0^+ \to \mathbb{R}$ injektiv, als Funktion $\mathbb{R} \to \mathbb{R}_0^+$ surjektiv und als Funktion $\mathbb{R}_0^+ \to \mathbb{R}_0^+$ beides.
- Ein gern genommenes Argument, mit dem nachweisen kann, dass eine Funktion $f : \mathbb{R} \to \mathbb{R}$ injektiv ist, ist sich zu überlegen, dass diese streng monoton steigend ist, d.h. dass für $x < y$ ist auch immer $f(x) < f(y)$ gilt. Hieraus folgt offensichtlich die Injektivität von f, ein Argument, das sich z.B. für die Exponentialfunktion h anwenden lässt.

Definition 2.4.3. *(Bijektivität)*

Eine Funktion f heißt bijektiv, wenn sie injektiv und surjektiv ist.

Setzt man die Eigenschaften injektiver und surjektiver Funktionen zusammen, so bedeutet das, dass für eine Funktion $f : X \to Y$ *jede* Gleichung $f(x) = b$ genau eine eindeutige Lösung hat genau dann, wenn f bijektiv ist. Wir wollen nun noch den Zusammenhang zur Umkehrbarkeit von Funktionen beleuchten.

2.5 Verkettung von Funktionen, Umkehrfunktionen und der Hauptsatz über bijektive Funktionen

Funktionen, deren Definitions- und Wertebereiche „kompatibel“ sind, lassen sich hintereinander ausführen. Führt man z.B. für

$$f : \mathbb{R} \to \mathbb{R}, x \mapsto x^2, \qquad \text{und} \qquad g : \mathbb{R} \to \mathbb{R}, x \mapsto \sin(x)$$

erst f, dann g, aus, so erhält man die Funktion $g \circ f : x \mapsto g(f(x)) = \sin(x^2)$.

Definition 2.5.1. *(Hintereinanderausführung $\circ$ von Funktionen):*
Seien $f : X \to Y$ und $g : Y \to Z$ zwei Abbildungen, dann ist die Hintereinanderausführung, Komposition oder Verknüpfung von f und g eine Funktion $g \circ f : X \to Z$, die punktweise durch

$$g \circ f : \quad X \to Z, \qquad x \mapsto (g \circ f)(x) := g(f(x))$$

eindeutig festgelegt ist.

Beachten Sie die Reihenfolge in der Notation $g \circ f$. Diese ist so gewählt, dass sie mit der Anwendung $g(f(x))$ auf einzelne Elemente $x \in X$ übereinstimmt, entspricht daher aber nicht der typischen Leserichtung von links nach rechts! Das ist wichtig, wenn selbst wenn $X = Y = Z$ gilt, ist die Reihenfolge der Anwendung von Funktionen nicht egal, die Hintereinanderausführungen $f \circ g$ ist die Funktion $x \mapsto (\sin(x))^2$. Es gilt also im Allgemeinen $f \circ g \neq g \circ f$, man sagt, die Hintereinanderausführung von Abbildungen ist nicht kommutativ.

Lemma 2.5.2. *(Hintereinanderausführungen injektiver und surjektiver Funktionen)*
Seien $f : X \to Y$ und $g : Y \to Z$ Abbildungen. Dann gelten folgende Aussagen:

(i) Wenn f und g injektiv sind, so ist auch die Hintereinanderausführung $g \circ f$ injektiv,

(ii) Wenn f und g surjektiv sind, so ist auch $g \circ f$ surjektiv,

(iii) Wenn $g \circ f$ injektiv ist, so ist f injektiv,

(iv) Wenn $g \circ f$ surjektiv ist, so ist g surjektiv.

Beweis: Wir beweisen nur (ii) und (iii), die Aussagen (i) und (iv) stimmen auch, die beweist man ähnlich, das bleibt Ihnen als Übungsaufgabe überlassen. Für (ii) seien f und g surjektiv, das heißt:

- Zu jedem $y \in Y$ gibt es ein $x \in X$ mit $f(x) = y$ (Surjektivität von f, (*)),
- Zu jedem $z \in Z$ gibt es ein $y \in Y$ mit $f(x) = y$ (Surjektivität von f, (**)).

Wir wollen zeigen: $g \circ f$ ist surjektiv, d.h. zu jedem $z \in Z$ gibt es ein $x' \in X$ mit $f(x') = z$. Sei dazu $z \in Z$ beliebig, dann besitzt z wegen (**) ein Urbild y unter g, und dieses y wegen (*) ein Urbild unter f. Insgesamt gilt

$$(g \circ f)(x) = g(f(x)) = g(y) = z,$$

x ist also ein gesuchtes Urbild von $g \circ f$.
Für (iii) ist unsere Voraussetzung, dass $g \circ f$ injektiv ist, d.h. aus $x_1 \neq x_2$ folgt $f(x_1) \neq f(x_2)$ für alle $x_1, x_2 \in X$ (†). Wir bedienen uns das erste Mal der Technik des Widerspruchsbeweises, die folgendermaßen funktioniert: Wir zeigen, dass die Annahme, dass f *nicht* injektiv ist, aber gleichzeitig unsere Voraussetzung (†) gilt, zu einem Widerspruch führt. Das bedeutet, dass die beiden Bedingungen nicht gleichzeitig erfüllt sein können, und da wir (†) voraussetzen, muss in solch einem Fall also f injektiv sein.[5] Wäre f nicht injektiv, so gäbe es $x_1, x_2 \in X$, für die $x_1 \neq x_2$ gilt, aber $f(x_1) = f(x_2)$ gilt. Wendet man darauf dann g an, so ergibt sich $(g \circ f)(x_1) g(f(x_1)) = g(f(x_2)) = (g \circ f)(x_2)$, das widerspricht aber der vorausgesetzten Injektivität von $g \circ f$.

□

Wir kommen zum Hauptresultat dieses Abschnitts (und zum ersten echten Beweis):

Satz 2.5.3. *Sei $f : X \to Y$ eine Abbildung, wobei X und Y nichtleere Mengen sind. Dann gilt*

(i) f ist genau dann injektiv, wenn es eine Abbildung $h : Y \to X$ gibt, so dass $h \circ f = id_X$ ist (d.h. f ist von links umkehrbar),

(ii) f ist genau dann surjektiv, wenn es eine Abbildung $h : Y \to X$ gibt, so dass $f \circ h = id_Y$ ist (d.h. f ist von rechts umkehrbar),

(iii) f ist genau dann bijektiv, wenn es eine Abbildung $h : Y \to X$ gibt, so dass $h \circ f = id_X$ und $f \circ h = id_Y$ (d.h. f ist von links und rechts umkehrbar).

[5] Bleiben Sie hier ruhig ein bisschen hängen und fragen Sie sich, ob Sie diese Argumentation so nachvollziehbar finden und glauben. Das sollten Sie sich ohnehin angewöhnen. Logische Schlussregeln werden wir uns dann genauer in Kapitel 7 ansehen.

Beweis: Hier wird es das erste Mal beweistechnisch etwas ausführlicher. Wir starten mit (i). Da dort insgesamt eine Äquivalenz behauptet wird, sind dort zwei Implikationen zu zeigen:

(i-a) Ist f injektiv, so gibt es eine Abbildung $h : Y \to X$ mit $h \circ f = id_X$, und

(i-b) Gibt es eine Abbildung $h : Y \to X$ mit $h \circ f = id_X$, so ist f injektiv.

Für (i-a) müssen wir eine Funktion $h : Y \to X$ geeignet definieren, indem wir angeben, was $h(y)$ für alle $y \in Y$ sein soll und dann zeigen, dass h eine Funktion ist. Das tun wir folgendermaßen : Wir definieren die Funktion h genau so, dass sie auf $\text{Bild}(f)$ gerade die Wirkung von f rückgängig macht: Für alle $y \in \text{Bild}(f)$ besitzt y ein eindeutiges Urbild $x_y := f^{-1}(y)$, und wir setzen $h(y) = x_y$. Für alle übrigen $y \notin \text{Bild}(f)$ können wir $h(y)$ beliebig festsetzen; sei dazu $x_0 \in X$ irgendein Element von X, wir setzen $h(y) = x_0$. Mit dieser Wahl gilt dann für beliebiges $x \in X$, dass $(h \circ f)(x) = h(f(x)) = x$ ist, d.h. es ist $h \circ f = id_X$. Zu (i-b) sei $h : Y \to X$ eine Abbildung mit $h \circ f = id_X$. Dann ist $h \circ g$ offensichtlich injektiv (denn mit $x_1 \neq x_2$ folgt $\text{id}_X(x_1) = x_1 \neq x_2 = \text{id}_X(x_2)$). Mit Lemma 2.5.2(iii) folgt, dass f injektiv ist.

Für den Teil (ii) haben wir wieder zwei Aussagen zu beweisen:

(ii-a) Ist f surjektiv, so gibt es eine Abbildung $h : Y \to X$ mit $f \circ h = id_Y$, und

(ii-b) Gibt es eine Abbildung $h : Y \to X$ mit $f \circ h = id_Y$, so ist f surjektiv.

Zu (ii-a) sei f als surjektiv vorausgesetzt. Dann existiert zu jedem $y \in Y$ mindestens ein Urbild unter f. Zu jedem $y \in Y$ wählen wir ein Urbild x_y aus und setzen $h(y) = x_y$. Dann gilt

$$(f \circ h)(y) = f(x_y) = y.$$

Umgekehrt können wir für (ii-b) wieder Lemma 2.5.2(iv) anwenden.
Auch für den Teil (iii) haben wir nochmal zwei Aussagen zu beweisen:

(iii-a) Ist f bijektiv, so gibt es eine Abbildung $h : Y \to X$ mit $f \circ h = id_Y$ und $h \circ f = id_X$, und

(iii-b) Gibt es eine Abbildung $h : Y \to X$ mit $f \circ h = id_Y$ und $h \circ f = id_X$, so ist f bijektiv.

Wir starten mit (iii-b): Gibt es eine Abbildung h mit den angegebenen Eigenschaften, so folgt durch Zusammensetzen der Resultate aus (i) und (ii) , dass f injektiv und surjektiv, also bijektiv ist. Zu (iii-a) sei f als bijektiv vorausgesetzt. Dann existiert nach (i) eine Funktion $h_1 : Y \to X$, die f von links umkehrt, und eine Funktion $h_2 : Y \to X$, die f

von rechts umkehrt. Diese könnten aber potenziell verschieden sein. Wir zeigen, dass sie in jedem Wert übereinstimmen, also gleich sind. Ist $y \in Y$, so ist

$$h_2(y) = id_X(h_2(y)) = h_1(f(h_2(y))) = h_1(\mathrm{id}_X(y)) = h_1(y).$$

$\square$

Sie fragen sich vielleicht gerade, was hier eigentlich gerade genau argumentiert wurde und wo der Unterschied in den beiden Argumentationen lag. Ja, das sollten Sie! Genau darin liegt die Arbeit, die Sie beim Lesen investieren müssen, um den Stoff zu verstehen. Dabei hilft in diesem Fall und auch allgemein zum Beispiel:

- Machen Sie sich zunächst die Aussage des Satzes klar. Versuchen Sie, den Satz auf einfache Beispiele, z.B. die Funktionen f, g und h bzw. f_1 bis f_3 aus Abschnitt anzuwenden. Welche von denen ist also von links- bzw. von rechts umkehrbar? Wie kann eine Funktion h aussehen?
- Versuchen Sie sich eine einfache Zeichnung zu machen, in der Sie die Situation bildlich darstellen. (Das habe ich in der Vorlesung auch gemacht.)
- Fragen Sie sich, an welcher Stelle eigentlich genau die Voraussetzung benutzt wurde (in diesem Fall, dass f injektiv bzw. surjektiv war).
- Beweisen Sie ähnliche Aussagen.

Um Sie zur Auseinandersetzung mit der Materie zu animieren, kommen hier gleich zwei geeignete Aufgaben.

Aufgabe 2.1.: Beweisen Sie spätestens jetzt die Ausagen (i) und (iv) aus Lemma 2.5.2.

Aufgabe 2.2.: Ist f nur injektiv oder nur surjektiv, so ist f nach 2.5.3 nur von links bzw. nur von rechts umkehrbar, und die Funktion h, die das leistet, für gewöhnlich nicht eindeutig. Finden Sie *zwei verschiedene* Funktionen, die die injektive Funktion f_3 von links umkehren, und zwei, die die surjektive Funktion f_2 von rechts umkehren.

Wir haben den entscheidenden Zusammenhang dieses Abschnitts bewiesen (uff), führen aber, um ihn griffig zu formulieren, noch eine Notation ein.

Definition 2.5.4. *(Umkehrfunktion)*

Sei $f : X \to Y$ eine Funktion. Eine Funktion $h : Y \to X$ heißt Umkehrfunktion oder inverse Funktion von f, falls $h \circ f = id_X$ und $f \circ h = id_Y$ ist. In diesem Fall sagen wir auch, f sei umkehrbar oder invertierbar.

Für spätere Zwecke sammeln wir noch einige Eigenschaften von Umkehrfunktionen:

Lemma/Definition 2.5.5. *(Eindeutigkeit von Umkehrfunktionen)*

Sei $f : X \to Y$ umkehrbar. Dann gilt:

(i) Es gibt genau eine Umkehrfunktion von f, man spricht daher von der Umkehrfunktion f^{-1} von f.

(ii) Ist $h : Y \to X$ eine Funktion mit $f \circ h = id_Y$, so ist h die Umkehrfunktion von f.

Beweis: Sei $f : X \to Y$ umkehrbar. Dann hat f mindestens eine Umkehrfunktion, die insbesondere $f \circ h = id_Y$ erfüllt. Wir zeigen: Es kann keine andere Funktion h_2 eine zweite Inverse von f sein. Es ist nämlich so, dass für jede andere Funktion $h_2 : Y \to X$ für mindestens ein $y^* \in Y$ $h_2(y^*) \neq h(y^*)$ gilt. Nach Satz 2.5.3 ist f bijektiv und insbesondere injektiv, d.h. es gilt $f(h_2(x^*)) \neq f(h(x^*)) = x^*$. Somit ist $(f \circ h_2)(x^*) \neq x^* = \mathrm{id}_Y(x^*)$. Alle anderen Funktionen $h_2 \neq h$ können also keine Umkehrfunktion von f sein. (ii) folgt, da f eine Umkehrfunktion besitzt, aber alle Funktionen $h_2 \neq h$ hierfür nicht in Frage kommen.

□

Lemma 2.5.6. *(Umkehrbarkeit der Umkehrfunktion)*

Sei $f : X \to Y$ eine bijektive Abbildung und $f^{-1} : Y \to X$ die Umkehrfunktion von f. Dann ist f auch die Umkehrfunktion von f^{-1} und f^{-1} ebenfalls bijektiv.

Beweis: Ist f^{-1} die Umkehrfunktion von f, so ist $f^{-1} \circ f = \mathrm{id}_X$ und $f \circ f^{-1} = \mathrm{id}_Y$, das heißt aber, dass f^{-1} eine Links- und Rechtsinverse hat (nämlich gerade f). Nach Satz 2.5.3(iii) ist f^{-1} damit auch bijektiv.

□

Durch „Zusammensetzen“ der vorher besprochenen Eigenschaften von injektiven und surjektiven Funktionen ergibt sich nun unser erster Hauptsatz:

Satz 2.5.7. *(Hauptsatz über bijektive Funktionen)*
Sei $f : X \to Y$ eine Funktion. Dann sind die folgenden Aussagen äquivalent:

(i) Die Gleichung $f(x) = b$ hat für jedes $b \in Y$ genau eine Lösung.

(ii) Die Funktion f ist bijektiv, d.h. injektiv und surjektiv.

(iii) Die Funktion f besitzt eine Umkehrfunktion $f^{-1} : Y \to X$.

Die in (i) angegebene Lösung ist dann gerade $f^{-1}(b)$. Die Elemente von X und Y entsprechen einander dann eins-zu-eins: Zu jedem $x \in X$ gehört genau ein $y \in Y$ (d.h. f ist eine Funktion), und auch umgekehrt gehört zu jedem $y \in Y$ genau ein $x \in X$ (d.h. f^{-1} ist eine Funktion). Man kann also, wenn f bijektiv ist, im Sinne der Denke oben für alle $b \in Y$ von der Lösung der Gleichung $f(x) = b$ sprechen. Wir werden nun sehen, dass das auch rechtfertigt, von dem Logarithmus, der Wurzel oder dem Arkussinus einer Zahl zu sprechen (wenn diese im Bild der entsprechenden Funktion liegt):

2.5.1 Gängige Anwendung: Bijektivmachen durch Einschränkung

Funktionen lassen sich oft durch sinnvolle Einschränkung des Definitionsbereichs injektiv machen, z.B. sind die reellen Funktion $x \mapsto x^2, x \mapsto \sin(x), x \mapsto a^x$ auf den Definitionsbereichen $\mathbb{R}_0^+ := [0, \infty[$, $[\frac{\pi}{2}, \frac{\pi}{2}]$ bzw. $\mathbb{R}$ injektiv. Sie lassen sich surjektiv machen, indem man den Wertevorrat der obigen Funktionen jeweils auf ihr Bild, also alle Werte, die tatsächlich angenommen werden, einschränkt. Für die Funktion $x \mapsto x^2$ ist das der Wertevorrat $\mathbb{R}_0^+$, für den Sinus das Intervall $[-1, 1]$, die Exponentialfunktion ist bereits surjektiv auf $\mathbb{R}$, daher lässt man $\mathbb{R}$ als Wertevorrat. Es sind also die Funktionen

$$f : \mathbb{R}_0^+ \to \mathbb{R}_0^+, \quad x \mapsto x^2, \qquad g :, [\frac{\pi}{2}, \frac{\pi}{2}] \to [-1, 1] \quad x \mapsto \sin(x), \quad \text{und} \quad h : \mathbb{R} \to \mathbb{R}^+, \quad x \mapsto a^x$$

bijektiv. Diese besitzen also jeweils eine eindeutige Umkehrfunktion; die gängigen Bezeichnungen dafür sind

$$\sqrt{x} : \mathbb{R}_0^+ \to \mathbb{R}_0^+, \qquad \arcsin(x) : [-1, 1] \to [\frac{\pi}{2}, \frac{\pi}{2}] \quad \text{und} \quad \log_a(x) : \mathbb{R}^+ \to \mathbb{R}.$$

Die Gesamtlösungsmengen entsprechender Gleichungen lassen sich dann sinnvoll daraus konstruieren (indem man beispielsweise bei x^2 die Symmetrie zur y-Achse und beim Sinus die Symmetrie bei Verschiebung um 2π ausnutzt).

Da insbesondere beim Wurzelziehen gelegentlich ein bisschen Verwirrung herrscht, was man darf und welche Notation wie gemeint ist, halten wir hier fest:

Definition 2.5.8. *(„Die Wurzel“ $\sqrt{a}$ und die n-te Wurzel $\sqrt[n]{a}$ in $\mathbb{R}$)*

Sei $a \in \mathbb{R}$ mit $a \geq 0$. Dann gibt es genau eine nichtnegative[6] reelle Zahl y mit $y^n = a$. Diese nennt man die n-te Wurzel von a, *man schreibt $y = \sqrt[n]{a}$ oder $y = a^{\frac{1}{n}}$.*

Für $x \in \mathbb{R}$ und $a \geq 0$ ist mit dieser Notation

$$x^2 = a \quad \Longleftrightarrow \quad x = \pm\sqrt{a} \tag{2.7}$$

eine zulässige Äquivalenzumformung; weiß man, das $x \geq 0$ ist (z.B., wenn x eine geometrische Länge bezeichnet), so gilt sogar

$$x^2 = a \quad \Longleftrightarrow \quad x = \sqrt{a}. \tag{2.8}$$

Vorsicht: Auch wenn für ungerade n auch eine Wurzel aus negativen Zahlen a Sinn zu machen scheint, und auch wenn im folgenden Kapitel u.a. darum gehen wird, auch „Quadratwurzeln aus negativen Zahlen zu ziehen“, werden die Notationen $\sqrt[n]{a}$ und $a^{\frac{1}{n}}$ nur für nichtnegative reelle Zahlen a verwendet.

Beweis: Die Funktion $f : \mathbb{R}_0+ \to \mathbb{R}_0^+, x \mapsto x^n$ ist für alle $n \in \mathbb{N}$ bijektiv (Die genaue, fachlich korrekte Begründung insbesondere der Surjektivität sei hierbei Analysis überlassen!), also $\sqrt{a}$ als die eindeutige Lösung der Gleichung $x^n = a$ mit $x \in \mathbb{R}_0^+$ wohldefiniert. Der oben skizzierte Weg, die weiteren Lösungen durch Symmetrieüberlegungen zu erhalten, lässt sich z.B. so bewerkstelligen: Neben $x = \sqrt{a}$ und $x = -\sqrt{a}$ Lösung der Gleichung $x^2 = a$ (denn es ist $(-\sqrt{a})^2 = (-1)^2 \cdot \sqrt{a}^2 = a$). Auch die Einschränkung der Funktion $x \mapsto x^2$ auf $\mathbb{R}^-$ ist injektiv, $x = -\sqrt{a}$ ist also die einzige negative Lösung; das zeigt (2.7). $\square$

2.6 Aufgaben zu Kapitel 2

Aufgabe 2.3. (Injektivität und Surjektivität):

Sei X eine Menge mit endlich vielen Elementen und $f : X \to X$ eine Abbildung von der Menge X in sich selbst. Begründen Sie:

(i) Wenn f surjektiv ist, so ist f auch injektiv.

(ii) Wenn f injektiv ist, so ist f auch surjektiv.

(iii) Ist f injektiv oder surjektiv, so ist f bijektiv.

Aufgabe 2.4. (Verkettung von Funktionen):

Seien X, Y, Z nicht leere Mengen und $f : X \to Y$ und $g : Y \to Z$ Abbildungen. Beweisen Sie, dass folgende Aussagen wahr sind:

(a) Wenn f und g injektiv sind, so ist auch $g \circ f$ injektiv.

(b) Wenn $g \circ f$ surjektiv ist, so ist g surjektiv.

Finden Sie dann Gegenbeispiele, mit denen Sie zeigen, dass folgende Folgerungen nicht für beliebige Funktionen f, g wahr sind:

(c) Wenn $g \circ f$ surjektiv ist, dann ist auch f surjektiv.

(d) Wenn $g \circ f$ injektiv ist, dann ist g injektiv.

Aufgabe 2.5. (Trigonometrische Funktionen):

In den Kapiteln werden wir die trigonometrischen[7] Funktionen Sinus und Cosinus und einige Zusammenhänge zwischen ihnen benötigen. Bitte stellen Sie sicher, dass Sie mit folgenden Begriffen bzw. Fragen etwas anfangen können (z.B. durch Recherche in Lehrbüchern oder Internet):

- Winkelmessung im Bogenmaß
- Definition von Sinus- und Cosinusfunktion am Einheitskreis
- Typische Werte und Symmetrien von Sinus- und der Cosinusfunktion
- Welche Koordinaten hat ein Vektor der Zeichenebene, der die Länge $a > 0$ hat (a bezeichne eine positive reelle Zahl) und mit der x-Achse den Winkel α (im Bogenmaß) einschließt?
- „trigonometrischer Pythagoras“
- Additionstheoreme für Sinus und Cosinus

Aufgabe 2.6. (Umkehrbarkeit der Sinusfunktion anschaulich):

(a) Begründen Sie: Ist ein Wert für $\sin(x)$ gegeben, so lässt sich der beteiligte Winkel $x \in [0, 2\pi[$ daraus nicht eindeutig ermitteln. Benutzen Sie hierbei, wenn möglich, geeignete Fachbegriffe aus dem letzten Kapitel.

(b) Schränken Sie die Funktion $f : \mathbb{R} \to \mathbb{R}, x \mapsto \sin(x)$ sinnvoll so ein, dass diese auf dem eingeschränkten Definitionsbereich und Wertevorrat umkehrbar ist. Geben Sie Definitionsbereich und Wertevorrat Ihrer Umkehrfunktion $\sin^{-1}$ an.

(c) Sei b ein gegebener „Sinuswert“, d.h. ein Wert aus dem Bild der Sinusfunktion. Geben Sie ein Verfahren an, wie man mit Hilfe des Wertes von $\sin^{-1}(b)$ (mit Ihrer Umkehrfunktion aus (b)) *alle* $x \in \mathbb{R}$ ermitteln kann, die die Gleichung $\sin(x) = b$ erfüllen.

[7] „Trigonometrie“ bezeichnet die „Dreiecksmessung“, also die Geometrie des Dreiecks.

3 Polynome und die komplexen Zahlen

Ein „Unterkapitel“ der Beschäftigung mit Gleichungen ist die Lösungstheorie für so genannte Polynomgleichungen, in denen die Funktion f, die in der Gleichung vorkommt, ein Polynom ist. Ein Beispiel sind die oft in der Mittelstufe behandelten quadratischen Gleichung (in denen das vorkommende Polynom zu einer quadratischen Funktion gehört). Die Beschäftigung mit ihnen führte im Laufe einiger Jahrhunderte u.a. zur stückweisen „Erfindung“ der so genannten komplexen Zahlen $\mathbb{C}$ (u.a. durch Girolamo Cardano (16. Jhd.), Leonhard Euler (18. Jhd.), Carl Friedrich Gauß (19. Jhd.)) Die komplexen Zahlen sind ein Zahlbereich, der die reellen Zahlen als Teilbereich enthält, in dem aber zusätzlich das „Wurzelziehen“ uneingeschränkt, zum Beispiel auch für negative reelle Zahlen, möglich ist, und in dem sich mathematische Fragestellungen (auch solche über reelle Zahlen) oft einfacher und eleganter lösen lassen als in den reellen Zahlen.

Komplexe Zahlen und Polynome

Girolama Cardano
(16. Jhd.)

Leonhard Euler
(18. Jhd.)

Carl Friedrich Gauß
(19. Jhd.)

3.1 Polynome

3.1.1 Grundbegriffe zu Polynomen

Definition 3.1.1. *(Polynome über $\mathbb{R}$, Vokabeln) Ein Ausdruck der Form*

$$p = \sum_{i=0}^{n} a_i \cdot x^i = a_n \cdot x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0,$$

wobei $n \in \mathbb{N}_0$ ist, x eine Variable ist und die Koeffizienten („Vorfaktoren") $a_n, \ldots, a_1, a_0$ des Polynoms reelle Zahlen sind, heißt ein Polynom über $\mathbb{R}$.

Das Polynom $p = 0$ (also mit $n = 0$ und $a_0 = 0$) heißt Nullpolynom. Alle anderen Polynome werden normalerweise so angegeben, dass der Leitkoeffizient $a_n \neq 0$ ist. Die Zahl n heißt dann Grad des Polynoms, kurz: Grad$(p) = n$.

Ein Polynom heißt normiert, wenn $a_n = 1$ ist.

Die Menge aller Polynome über $\mathbb{R}$ bezeichnet man mit $\mathbb{R}[x]$. Ist $p \in \mathbb{R}[x]$, so heißt die Funktion

$$p : \mathbb{R} \to \mathbb{R}, \quad x \mapsto p(x) := \sum_{i=0}^{n} a_i \cdot x^i,$$

die p durch Einsetzen von $x \in \mathbb{K}$ in x auswertet, die zu p gehörige Polynomfunktion.

Die formal definierte Addition und Multiplikation von Polynomen $p, q \in \mathbb{R}[x]$ wird so gewählt, dass sich bei Betrachtung der zugehörigen Polynomfunktionen die Summe $p+q \in \mathrm{Abb}(\mathbb{R})$ und das Produkt $p \cdot q \in \mathrm{Abb}(\mathbb{R})$ ergibt.

Definition 3.1.2. *(Addition und Multiplikation von Polynomen) Wir setzen für $p = \sum_{i=0}^{n} a_i \cdot x^i$, $q = \sum_{i=0}^{m} b_i \cdot x^i \in \mathbb{R}[x]$ fest:*

$$p + q = \sum_{i=0}^{\max\{m,n\}} (a_i + b_i) \cdot x^i.$$

Hierbei bezeichnet $\max\{m, n\}$ die größere der Zahlen n, m, im Fall $i > n$ sei in der obigen Summe $a_i = 0$, für $i > m$ analog $b_i = 0$. Polynome werden also koeffizientenweise addiert.

$$p \cdot q = \sum_{i=0}^{n+m} c_i \cdot x^i, \text{ wobei } \quad c_i = \sum_{j+k=i} a_j \cdot b_k.$$

In der hinteren Summe zur Berechnung von c_i kommen all die Produkte $a_j \cdot b_k$ von Koeffizienten vor, für die $j + k = i$ ist. Diese Festlegung ergibt sich durch die Forderung, dass $(p \cdot q)(x) = p(x) \cdot q(x)$ sein soll.

Lemma/Definition 3.1.3. *(Gradformel, Grad des Nullpolynoms)*
Sind p, q Polynome aus $\mathbb{R}[x]$, so gilt

$$\mathit{Grad}(p + q) \leq \max\{\mathit{Grad}(p), \mathit{Grad}(q)\}.$$

und

$$\mathit{Grad}(p \cdot q) = \mathit{Grad}(p) + \mathit{Grad}(q).$$

Damit die Formel oben auch für das Nullpolynom $p = 0$ richtig ist, setzen wir $\mathit{Grad}(0) = -\infty$ und definieren $-\infty + n = -\infty$ für alle natürlichen Zahlen n und $-\infty + (-\infty) = -\infty$.

Beweis: Übungsaufgabe.

□

3.1.2 Polynomgleichungen und Nullstellen

Sei $p \in \mathbb{R}[x]$ ein Polynom mit $p \neq 0$. Eine Gleichung der Form $p(x) = b$ ($x \in \mathbb{R}$ gesucht, $b \in \mathbb{R}$) heißt Polynomgleichung.

- Grad$(p) = 1$: Lineare Gleichung,
- Grad$(p) = 2$: Quadratische Gleichung,
- Grad$(p) = 3$: Kubische Gleichung,
- Grad$(p) = n \geq 4$: Algebraische Gleichung von Grad n.

Für jedes Polynom $p \neq 0$ ist die Division durch den Leitkoeffizient $a_n \neq 0$ eine Äquivalenzumformung, daher gilt

$$\sum_{i=0}^{n} a_i \cdot x^i = b \qquad \Longleftrightarrow \qquad x^n + \frac{a_{n-1}}{a_n} x^{n-1} + \ldots + \frac{a_1}{a_n} x + \frac{a_0 - b}{a_n} = 0.$$

Wir können uns daher auf Lösungsaussagen für $p(x) = 0$ mit normiertem p beschränken.

Für eine quadratische Gleichung $x^2 + px + q = 0$ (gemeint sind hier mit p, q Koeffizienten aus $\mathbb{R}$) findet man durch Äquivalenzumformung der Gleichung für $x \in \mathbb{R}$:

$$x^2 + px + q = 0 \qquad \Longleftrightarrow \qquad \left(x + \frac{p}{2}\right)^2 = \underbrace{\left(\frac{p}{2}\right)^2 - q}_{(*)}. \tag{3.1}$$

Ist der Radikant (*) größer als Null, so gibt es in den reellen Zahlen genau zwei Möglichkeiten für den Wert von $x + \frac{p}{2}$, nämlich $x + \frac{p}{2} = \pm\sqrt{\left(\frac{p}{2}\right)^2 - q}$, was wiederum gleichbedeutend dazu ist, dass die Gleichung hat genau zwei Lösungen hat, die sich folgendermaßen schreiben lassen:

$$x = -\frac{p}{2} \pm \sqrt{\left(\frac{p}{2}\right)^2 - q}.$$

Das ist die bekannte „p-q-Formel“ für quadratische Gleichungen. Ist der Radikant $(*)$ gleich Null, so ist $x + \frac{p}{2} = 0$, also $x = -\frac{p}{2}$ die einzige Lösung in $\mathbb{R}$. Ist der Radikant negativ, so hat die Gleichung keine reelle Lösung, da Quadrate von reellen Zahlen immer nichtnegativ (also null oder positiv) sind.

Bemerkung: Lösungsformeln für Polynomgleichungen höheren Grades?

Analog zur Lösung quadratischer Gleichungen durch quadratische Ergänzung gibt es auch Lösungsformeln für Polynomgleichungen dritten und vierten Grades (die „Formeln von Cardano“, (Wiki-Eintrag)). Eines der großen Erkenntnisse der Galois-Theorie (s.u.: Evariste Galois war Vorreiter der Gruppentheorie, die in Kapitel 5 behandelt wird) ist die, dass es für Polynomgleichungen höheren Grades keine allgemeine Lösungsformel mehr gibt, die wie die „p-q-Formel“ und die Cardano-Formeln mit den Operationen $+, \cdot$ und Wurzelziehen auskommt.

In der Lösungsformeln von Cardano kommen Wurzelausdrücke als Zwischenergebnisse vor. Bei der Untersuchung von Gleichungen dritten und vierten Grades stieß Cardano auf die Erkenntnis, dass er in diesen Zwischenergebnissen *Wurzeln aus negativen Zahlen* erhielt – vorgestellten, „imaginären“ Zahlen, füe die gar nicht definiert war, was das eigentlich sen soll. Diese hoben sich dann später durch Quadrieren wieder weg, und so erhielt Cardano ein *korrektes Endergebnis aus den reellen Zahlen*, also eine der gesuchten Nullstellen des Polynoms. Der Gedanke, Wurzeln aus negativen Zahlen zu definieren, könnte also nützlich sein. Diesen Gedanken wollen wir hier weiterverfolgen.

3.2 Die Zahl i und die komplexen Zahlen

3.2.1 „Herleitung“ und Definition der komplexen Zahlen

Wir wollen im Folgenden systematisch vorgehen und einen neuen Zahlbereich $(\mathbb{C}, +, \cdot)$ konstruieren, der die folgenden Forderungen erfüllt:

(A) Er enthält die reellen Zahlen $(\mathbb{R}, +, \cdot)$ mit den gewohnten Operationen $+$ und $\cdot$.

(B) In ihm gelten die für reelle Zahlen gewohnten Rechenregeln (z.B. Assoziativgesetze und Kommutativgesetze für $+$ und $\cdot$, Distributivgesetz) weiter („Permanenz der Rechengesetze“).

(C) $\mathbb{C}$ enthält Wurzeln aller positiven *und negativen* reellen Zahlen, d.h. zu *allen* $a \in \mathbb{R}$ gibt es ein $z \in \mathbb{C}$ mit $z^2 = a$.

Um die Forderung (C) auch für negative reelle Zahlen erfüllen zu können, „erfinden“ wir neue Zahlen (die natürlich keine reellen Zahlen sein können). Das ist nichts Besonderes, negative Zahlen und Brüche haben wir auch erfunden, weil sie nützlich sind, und dann Rechenoperationen so festgelegt, dass sie sinnvoll sind. Wir werden zusätzlich fordern, dass die gesuchten Zahlmenge $\mathbb{C}$ minimal ist, d.h. dass $\mathbb{C}$ keine zusätzlichen, für die Erfüllung der Forderungen nicht nötigen Zahlen erhält.
Mit dieser Heuristik[1] folgen wir weiter der geschichtlichen Entwicklung der komplexen Zahlen: Bei dem hier verwendeten „Permanenzprinzip“ handelt es sich um ein Verfahren, das der deutsche Mathematiker Herrmann Hankel 1867 systematisch vorstellte, um damit die Festlegung der Rechengesetze für komplexe Zahlen strukturell zu rechtfertigen. (Das Original kann man hier durchblättern: (Link)).

Zunächst stellen wir klar, dass es reicht, *eine* Zahl i zu definieren, für die $i^2 = -1$ gilt, um dann damit den Zahlbereich $\mathbb{C}$ mit den obigen Eigenschaften zu konstruieren.

Lemma 3.2.1. *Ist i eine Zahl mit $i^2 = -1$, $c > 0$ eine positive reelle Zahl und $\mathbb{C}$ eine Zahlmenge mit den oben geforderten Eigenschaften (A), (B) und (C), so gilt:*

$$x^2 = -c \qquad \Longleftrightarrow \qquad x = \pm i \cdot \sqrt{c}.$$

Insbesondere gilt: Ist in einer quadratischen Gleichung mit Koeffizienten p, q aus $\mathbb{R}$ der

[1] Das Wort kommt aus dem Griechischen und bedeutet in etwa „Finde-Strategie“. Wenn ein Grieche „Heureka!“ schreit und dabei nackt durch die Straßen Athens läuft, wie es vom Mathematiker und Physiker Archimedes behauptet wird, hat er etwas (heraus-)gefunden.

Radikant (*) in 3.1 negativ ist, so gilt:

$$x^2 + px + q = 0 \quad \Longleftrightarrow \quad x = -\frac{p}{2} \pm i \cdot \sqrt{q - \left(\frac{p}{2}\right)^2}.$$

Eine reelle quadratische Gleichung $x^2 + px + q = 0$ hat also, wenn der Radikant $(*)$ in (3.1) nicht Null ist, in $\mathbb{C}$ immer zwei Lösungen.

Beweis: Aus der Permanenz der Rechengesetze folgt, dass $x = i \cdot \sqrt{c}$ und $-i \cdot \sqrt{c}$ Lösungen der Gleichung $x^2 = -c$ sind: Es ist

$$(i \cdot \sqrt{c})^2 = i^2 \cdot \sqrt{c}^2 = -c \qquad \text{und} \qquad (-i \cdot \sqrt{c})^2 = (-1)^2 \cdot i^2 \cdot \sqrt{c}^2 = -c.$$

Wenn es eine andere Lösung y der Gleichung, also mit $y^2 = a$ gibt, dann ist für diese $y^2 = a = (i\sqrt{c})^2$ gelten. Da aus der Gültigkeit der gängigen Rechengesetze auch die Gültigkeit der binomischen Formeln folgt, haben wir:

$$y^2 = (i\sqrt{c})^2 \quad \Leftrightarrow \quad y^2 - (i\sqrt{c})^2 = 0 \quad \Leftrightarrow \quad (y + i\sqrt{c}) \cdot (y - i\sqrt{c}) = 0.$$

Damit das stimmt, muss eine der Klammern Null sein (auch diese Eigenschaft überträgt sich automatisch aus $\mathbb{R}$), also schon $y = i\sqrt{c}$ oder $y = i\sqrt{c}$ gegolten haben. Es kann also keine anderen Lösungen geben.

□

Wir konstruieren die komplexen Zahlen $\mathbb{C}$ nun als den kleinsten Körper, der die Zahl i und die Forderungen (A) und (B) von oben erfüllt. Um das zu tun, müssen wir angeben, welche Elemente $\mathbb{C}$ (neben i und den reellen Zahlen) überhaupt enthalten soll, und wie die beiden Operationen $+$ und $\cdot$ sinnvoll so auf ganz $\mathbb{C}$ ausgeweitet werden können, dass die Rechengesetze aus $\mathbb{R}$ erhalten bleiben.

Unser Vorgehen nach Permanenzprinzip ist nun folgendermaßen: Wir nehmen an, es gäbe einen Zahlbereich $\mathbb{C}$, der die obigen Bedingungen erfüllt. Daraus lassen sich dann durch „Herumrechnen" unter Benutzung der gewohnten „Rechengesetze" (Assoziativgesetz, Distributivgesetz,...) notwendige Bedingungen daran aufstellen, welche Elemente $\mathbb{C}$ zwingend enthalten muss und wie die Addition und Multiplikation auszusehen haben. Es ergeben sich z.B. die folgenden Forderungen (s. auch Vorlesung):

- Damit man in $\mathbb{C}$ vernünftig „wie in $\mathbb{R}$" rechnen kann, sollten Summen und Produkte von Zahlen aus $\mathbb{C}$ wieder Elemente aus $\mathbb{C}$ sein, also sollte für reelle a, b dann $a \cdot b, a \cdot i, a + b, a + b \cdot i, \ldots \in \mathbb{C}$ sein. (Abgeschlossenheit bzgl. $+$ und $\cdot$)

- Für $a, b \in \mathbb{R}$ weiterhin $a + b \in \mathbb{R}$ sein wie gehabt, wegen des Distributivgesetzes muss $c \cdot i + d \cdot i := (c + d)i$ gelten,
- Das Kommutativgesetz soll gelten: Das impliziert $a \cdot i = i \cdot a$.
- Für die Addition folgt $(a + b \cdot i) + (c + d \cdot i) = (a + b) + (c + d) \cdot i$, das liefert nur diese eine Möglichkeit, diese zu definieren!
- Für die Multiplikation muss gelten:
  $(a+bi) \cdot (c+di) = (\text{rechne, rechne}) = (a \cdot c - b \cdot d) + (a \cdot d - b \cdot c) \cdot i$ – auch hier ergibt sich das als einzige Möglichkeit, wenn die Forderungen oben erfüllt sein sollen!

So ergibt sich für unseren neu zu erschaffenden Zahlbereich $\mathbb{C}$ nur eine Möglichkeit:

Definition 3.2.2. *(Die komplexen Zahlen)*
Sei i eine Zahl mit $i^2 = -1 \in \mathbb{R}$. Das Tripel $(\mathbb{C}, +, \cdot)$ mit

$$\mathbb{C} := \{a + b \cdot i \mid a, b \in \mathbb{R}\},$$

der komponentenweisen Addition

$$(a + b \cdot i) + (c + d \cdot i) = (a + c) + (b + d) \cdot i$$

und der Multiplikation

$$(a + b \cdot i) \cdot (c + d \cdot i) = (a \cdot c - b \cdot d) + (a \cdot d + b \cdot c) \cdot i$$

als Verknüpfungen nennen wir die komplexen Zahlen, die Zahl i heißt die imaginäre Einheit von $\mathbb{C}$.

Man kann zeigen, dass mit der so gegebenen Definition in der Tat die gewohnten Rechengesetze aus den reellen Zahlen gelten, also zum Beispiel, dass $(\mathbb{C}, \cdot)$ assoziativ und kommutativ ist und eine Eins besitzt, dass jedes Element $z \neq 0$ einen „Kehrwert“ w mit $z \cdot w = 1$ besitzt, das Distributivgesetz gilt usw. (nur eins, da $(\mathbb{C}, \cdot)$ kommutativ ist). Einige dieser Tatsachen sind die Aussagen (i), (ii), (iii) und (ix) von Lemma 3.2.4, das wir weiter unten beweisen. Ein besserer, struktureller Blick gelingt, wenn wir in Kapitel 5 den Körperbegriff kennen gelernt haben: Die reellen Zahlen $(\mathbb{R}, +, \cdot)$ haben durch ihre Eigenschaften wie Assoziativität, Kommutativität usw. die algebraische Struktur eines so genannten Körpers; die obige Forderung lautet dann kurz und simpel, dass $\mathbb{C}$ auch einer sein soll. $\mathbb{C}$ enthält „eine Kopie“ der reellen Zahlen $\mathbb{R}$: Diese entsprechen der Teilmenge

$$\{z = a + b \cdot i \in \mathbb{C} \mid b = 0\}$$

von $\mathbb{C}$, und für $a, b \in \mathbb{R}$ stimmen $a + b$ und $a \cdot b$ in $\mathbb{R}$ mit dem Ergebnis der Addition und Multiplikation der entsprechenden Zahlen in $\mathbb{C}$ überein. Prüfen Sie das nach, indem Sie die entsprechenden Summen und Produkte für $a + 0 \cdot i$ und $b + 0 \cdot i$ in $\mathbb{C}$ berechnen!

Für reelle Zahlen, die wir als Elemente der komplexen Zahlen auffassen, schreiben wir z.B. auch einfach $z = 3$, und damit $3 + (-2 + 4i) = -1 + 4i$.

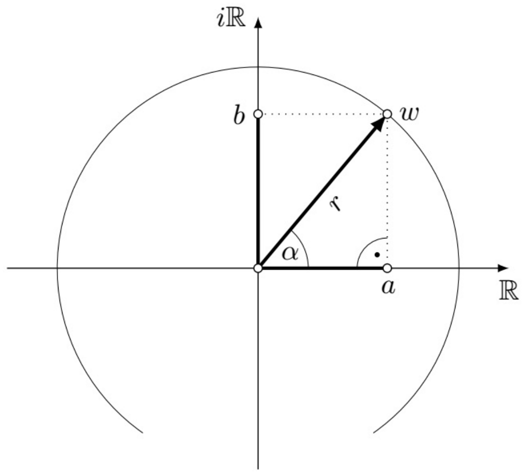
3.2.2 Die Gauß'sche Zahlenebene

Eine feine Erkenntnis über komplexe Zahlen, die wie viele Ergebnisse aus unserem mathematischen Erkenntnisschatz auf C. F. Gauß zurückgeht, ist, dass man komplexe Zahlen auch graphisch darstellen kann: Die Erweiterung der reellen zu komplexen Zahlen entspricht anschaulich der Erweiterung der reellen Zahlengerade zur komplexen Zahlenebene.

Jede komplexe Zahl z lässt sich mit eindeutigen $a, b \in \mathbb{R}$ als

$$z = a \cdot 1 + b \cdot i.$$

Daher können wir z in einem kartesischen Koordinatensystem mit Hilfe seiner „Koordinaten“ (a, b) anschaulich darstellen. Das ist die so genannte Gaußsche Zahlenebene, s. die Abbildung unten. Die ehemalige „x-Achse“ wird „reelle Achse“ genannt; das ist die aus der Schule bekannte reelle Zahlengerade, auf der gerade alle reellen Zahlen liegen.



$$\boxed{\begin{aligned} & w = a + ib = r(\cos\alpha + i\sin\alpha) \\ & |w| = \sqrt{a^2 + b^2} = r \\ & a = \mathrm{Re}(w), \quad b = \mathrm{Im}(w) \\ & \alpha = \arg(w) \end{aligned}}$$

3.2.3 Begriffe und Bezeichnungen zu komplexen Zahlen

- Komplexe Zahlen werden meist mit z oder w bezeichnet.
- Jede komplexe Zahl $z = a + bi \in \mathbb{C}$ lässt sich wie oben skizziert als Punkt in der *komplexen Ebene* deuten. Der zugehörige Ortsvektor von z, im Ursprung angeheftet, heißt dabei Zeiger von z.
- a heißt Realteil, b Imaginärteil von z, kurz $a = \mathrm{Re}(z), b = \mathrm{Im}(z)$.
- Die x-Achse bzw. y-Achse deutet man als *reelle* bzw. *imaginäre Achse*, auf ihnen wird $\mathrm{Re}(z)$ bzw. $\mathrm{Im}(z)$ abgetragen.
- Die Addition zweier komplexer Zahlen $z = a + bi, w = c + di$ lässt sich dann als Vektoraddition, also als „Hintereinanderhängen“ der Zeiger in der Ebene, deuten.
- Wir definieren den Betrag einer komplexen Zahl anschaulich als Länge ihres Zeigers. Diese berechnet sich nach Satz des Pythagoras als $|z| = (a^2 + b^2)^{\frac{1}{2}}$; dieser wird auch oft mit r abgekürzt.
- Das Argument einer komplexen Zahl $z \neq 0$ ist definiert als der Winkel $\alpha \in [0, 2\pi]$, den sie mit der reellen (x-)Achse einschließt.
- Ist $|z| = r$ und α das Argument von z, so gilt für die Koordinaten:

$$a = r \cdot \cos(\alpha) =: r \cdot c_\alpha, \quad b = r \cdot \sin(\alpha) =: r \cdot s_\alpha,$$

also insgesamt[2]

$$z = \overbrace{r}^{\text{Betrag von } z} \cdot \underbrace{(c_\alpha + i \cdot s_\alpha)}_{\text{„Einheitszeiger“ in Richtung } z.}.$$

Jede Zahl $z \neq 0$ ist so eindeutig durch die Angabe ihres Betrages und des Arguments $\alpha \in [0, 2\pi[$ festgelegt. Diese Darstellung über Winkel und Betrag heißt Polardarstelllung einer komplexen Zahl.

[2] In den folgenden Notationen wurde die Reihenfolge von i und dem „Vorfaktor“ $\mathrm{Im}(z)$ getauscht. Das dürfen wir, weil $(\mathbb{C}, \cdot)$ kommutativ ist, und machen wir, weil es bei der Notation $z = c_\alpha + i \cdot s_\alpha$ so üblich ist.

Lemma 3.2.3. *(Deutung der Multiplikation in Polarkoordinaten)*

Sind $z = r \cdot (c_\alpha + i \cdot s_\alpha)$ *und* $w = s \cdot (c_\beta + i \cdot s_\beta)$ *zwei komplexe Zahlen in Polardarstellung, so gilt*

$$z \cdot w = r \cdot s \cdot (c_{\alpha+\beta} + i s_{\alpha+\beta}),$$

d.h. das Resultat ist ein eine komplexe Zahl vom Betrag $r \cdot s$ *und mit dem Argument* $\alpha + \beta$*, komplexe Zahlen werden also multipliziert, indem man ihre Beträge multipliziert und ihre Argumente addiert.*

Beweis: Schreibe wie oben kurz $z = r \cdot (c_\alpha + i s_\alpha)$ und $w = s \cdot (c_\beta + i s_\beta)$, dann ergibt sich mit den Additonstheoremen für Sinus und Cosinus:

$$\begin{aligned} z \cdot w &= r \cdot s \cdot (c_\alpha + i s_\alpha) \cdot (c_\beta + i s_\beta) \\ &\overset{\text{Def. der Mult.}}{=} r \cdot s \cdot (c_\alpha c_\beta - s_\alpha s_\beta) + i \cdot (s_\alpha c_\beta + c_\alpha s_\beta)) \overset{\text{Add.th.}}{=} r \cdot s \cdot (c_{\alpha+\beta} + i s_{\alpha+\beta}). \end{aligned}$$

□

Lemma 3.2.4. *(Rechenregeln für komplexen Zahlen)*

Seien $z = a + bi = r \cdot (c_\alpha + i s_\alpha)$ *und* $w = c + di = s \cdot (c_\beta + i s_\beta)$ *komplexe Zahlen. Weiter sei die zu* $z = a + bi$ *komplex kojugierte Zahl* $\overline{z}$ *als* $\overline{z} = a - bi$ *definiert. (was einer Spiegelung von* z *an der reellen Achse entspricht). Mit diesen Bezeichnungen gilt:*

(i) $0 = 0 + 0 \cdot i$ *und* $1 = 1 + 0 \cdot i$ *sind aditiven bzw. multiplikativen die neutralen Elemente von* $(\mathbb{C}, +)$ *und* $(\mathbb{C}, \cdot)$ *(es ist also immer* $z + 0 = z$ *und* $z \cdot 1 = 1$*).*

(ii) Die Gegenzahl zu z *(seine „additive Inverse“) ist die Zahl* $-z = -a - bi$*, es ist also* $z + (-z) = 0$*.*

(iii) Für alle $z \neq 0$ *ist der Kehrwert zu* z *(seine „multiplikative Inverse“) gegeben durch* $\dfrac{1}{z} = \dfrac{a - bi}{a^2 + b^2} = \dfrac{\overline{z}}{|z|^2}$

(iv) $r = |z| = \sqrt{z \cdot \overline{z}}$ *(insb. ist* $r \in \mathbb{R}, r \geq 0$*),*

(v) $\frac{1}{2}(z + \overline{z}) = Re(z)$*,*

(vi) $\frac{1}{2i}(z - \overline{z}) = Im(z)$*,*

(vii) $\overline{w \cdot z} = \overline{w} \cdot \overline{z}$*,*

(viii) $\overline{w + z} = \overline{w} + \overline{z}$*.*

Für die Multiplikation gelten in Polarkoordinaten außerdem die Regeln

(ix) $\frac{1}{z} = \frac{1}{r} \cdot (c_{-\alpha} + is_{-\alpha})$ *für alle* $z \neq 0$,

(x) $z^n = r^n \cdot (c_{n\cdot\alpha} + is_{n\cdot\alpha})$ *für alle* $n \in \mathbb{N}$.

Beweis: Die Rechenregeln kann man leicht algebraisch, also durch Rechnung, nachvollziehen, aber auch gut in der komplexen Ebene deuten. S. Vorlesung/Übung. □

3.2.4 Exkurs: Die Mandelbrotmenge

Eine der faszinierendsten Entdeckungen zu den komplexen Zahlen, die wir mit unserem Wissen bis hier verstehen können, ist die so genannte Mandelbrotmenge, die eine Teilmenge der komplexen Ebene ist, die erstaunliche Eigenschaften aufweist. Um zu verstehen, was die Menge ist, betrachten wir zunächst eine andere, ähnliche Konstruktion:
Wir wählen zunächst irgendeinen Punkt der komplexen Ebene als Startwert z_1. Aus z_1 ergibt sich durch „immer wieder quadrieren" eine Folge komplexer Zahlen, die durch den Startwert z_1 und die *Rekursionsvorschrift* $z_{n+1} = z_n^2$. festgelegt ist, z.B.

$$\begin{array}{lcll}
z_1 & = & 2 & \rightsquigarrow \text{Folge} \quad 2, 4, 16, 256, \ldots \\
z_1 & = & i & \rightsquigarrow \text{Folge} \quad i, -1, 1, 1, 1, \\
z_1 & = & 3+i & \rightsquigarrow \text{Folge} \quad 3+i, 8+6i, 28+96i, \ldots \\
z_1 & = & \frac{1}{2}+\frac{1}{4}i & \rightsquigarrow \text{Folge} \quad \frac{1}{2}+\frac{1}{4}i, 0,1875+0,25i, -0,02734375+0,09375i, \ldots
\end{array}$$

Allgemein kann man die entstehende Folge meist besser mit der Polardarstellung ausrechnen. Es gilt nach den Potenzgesetzen und Lemma 3.2.4(x): Ist $z_1 = r \cdot (c_\alpha + is_\alpha)$, so ist

$$z_n = z^{2^n} = r^{2^n} \cdot (c_{2n\cdot\alpha} + is_{2n\cdot\alpha}).$$

Man sieht:

- Für Startwerte mit $|z_1| > 1$ ist diese Folge $(z_n)_{n\in\mathbb{N}}$ betragsmäßig unbeschränkt, die Punktefolge entfernt sich beliebig weit vom Ursprung.

- Für komplexe Zahlen z_1 mit $|z_1| = 1$ haben alle Glieder der Folge $(z_n)_{n\in\mathbb{N}}$ Betrag 1, die resultierenden Punkte befinden sich alle auf dem Rand des komplexen Einheitskreises (also des Kreises mit Radius 1 um den Ursprung der komplexen Ebene.
- Für komplexe Zahlen z_1 mit $|z_1| < 1$ werden die Beträge der Folge z_n immer kleiner, die Folge konvergiert gegen den Ursprung $(0 \mid 0)$ der komplexen Ebene.

Markiert man nun alle Punkte der komplexen Ebene, für die die resultierende Folge beschränkt bleibt, schwarz, und alle komplexen Zahlen, für die diese Folge vom Betrag her gegen unendlich strebt, in weiß, so ergibt sich ein zugegebenermaßen relativ langweiliges Bild: Der Einheitskreis (einschließlich Rand) ist schwarz (das sind gerade die letzten beiden Fälle), alle Zahlen jenseits dieses Kreises sind weiß. Klar, komplexe Zahlen vom Betrag größer 1 werden beim Quadrieren betragsmäßig immer größer, die anderen nicht.

Wir ändern die erzeugende Folge nun unscheinbar, indem wir nach jedem Quadrieren noch einmal den Startwert z_1 hinzuaddieren: So erhält man die Rekursion

$$z_{n+1} = z_n^2 + z_1.$$

Wieder färben wir alle Punkte, für die die Folge beschränkt bleibt, schwarz. Wie sieht die entstehende Menge aus? Das ist die Mandelbrotmenge (oder das „Apfelmännchen"), und Computer können dankenswerterweise ein Bild davon zeichnen: (Link). (Hierbei wird die weiße Fläche der divergierenden Punkte oft noch hübsch danach gefärbt, wie schnell die Punktefolge von dem jeweiligen Startwert aus divergiert.) Hätten Sie das erwartet?

3.3 Der Fundamentalsatz der Algebra

Polynome zweiten Grades lassen sich, wenn sie zwei oder eine Nullstellen in $\mathbb{R}$ haben, „über $\mathbb{R}$ in Linearfaktoren faktorisieren", z.B.

$$x^2 - 4 = (x+2)\cdot(x-2), \qquad -3x^2 + 3x = (-3)\cdot(x-0)\cdot(x-1),$$

$$3x^2 + 12x + 12 = 3\cdot(x+2)\cdot(x+2), \qquad 2x^2 = 2\cdot(x-0)\cdot(x-0), \quad \ldots$$

Von Polynomen ohne Nullstellen in $\mathbb{R}$ lassen sich keine reellen Nullstellen „abdividieren". In $\mathbb{C}$ ist eine analoge Zerlegung quadratischer Polynome möglich, wenn wir die komplexen Nullstellen berücksichtigen, z.B.

$$x^2 + 1 = (x+i)\cdot(x-i), \qquad -3x^2 - 12 = (-3)\cdot(x-2i)\cdot(x+2i).$$

Wir wollen diese Erkenntnis, die hier durchscheint, nun allgemeiner formulieren, indem wir die Aussage

- auf Polynome vom Grad n ausweiten und
- auch Polynome zulassen, die komplexe Zahlen als Koeffizienten haben. Die entsprechende Menge aller Polynome über $\mathbb{C}$ wird mit $\mathbb{C}[x]$ bezeichnet; die Definition der Additition und Multiplikation („koeffizientenweise") bleibt gleich.

Der Satz, den wir jetzt formulieren wollen, ist der so genannte Fundamentalsatz der Algebra 3.3.2. Dazu beweisen wir zunächst den folgenden Zusammenhang über das Abdividieren von Nullstellen:

Lemma/Definition 3.3.1. *(Abdividieren von Nullstellen, Linearfaktor)*
Sei p ein Polynom mit reellen oder komplexen Koeffizienten. Es sind äquivalent:

(i) *$c \in \mathbb{C}$ ist eine Nullstelle von p,*

(ii) *Es gibt ein Polynom $q \in \mathbb{C}[x]$ so, dass $p = (x-c) \cdot q$. Ist p ein reelles Polynom und c eine reelle Nullstelle, so ist auch q ein reelles Polynom.*

$(x-c)$ aus (ii) heißt Linearfaktor von p. Für das Polynom q ist Grad(q) = Grad$(p) - 1$, und alle Nullstellen $d \neq c$ von p sind auch Nullstellen von q.

Beweis:
(ii) $\Rightarrow$ (i) folgt wegen $p(c) = 0 \cdot q(c) = 0$. (i)$\Rightarrow$ (ii) ist etwas aufwändiger; hierfür prüft man zunächst nach, dass für alle $k \in \mathbb{N}$ und alle $c \in \mathbb{C}$ die Gleichheit

$$x^k - c^k = (x-c) \cdot \left(x^{k-1}c^0 + x^{k-2}c^1 + \ldots + x^1c^{k-2} + x^0c^{k-1}\right) =: q_{k-1}$$

gilt, wobei q_{k-1} also ein Polynom von Grad $k-1$ ist, das, falls c reell ist. Damit erhält man, falls $c \in \mathbb{C}$ eine Nullstelle von p ist,

$$\begin{aligned} p(x) &= p(x) - 0 = p(x) - p(c) \\ &= \sum_{k=0}^{n} a_k x^k - \sum_{k=0}^{n} a_k c^k = \sum_{k=1}^{n} a_k \left(x^k - c^k\right) \\ &= \sum_{k=1}^{n} a_k (x-c) \cdot q_{k-1}(x) = (x-c) \cdot \sum_{k=1}^{n} a_k q_{k-1}(x). \end{aligned}$$

wobei der hintere Term als Summe der Polynome q_{i-1}, $i \in \underline{n}$, ein Polynom vom Grad $n-1$ ist, das reell ist, falls c und die Koeffizenten a_k, $k \in \underline{n}$ reell sind.

Die letzte Bemerkung ergibt sich folgendermaßen: Ist c eine Nullstelle von p und $p(d) = 0$ für $d \neq c$, so ist mit der obigen Zerlegung in Linearfaktoren $p(d) = (d-c) \cdot q(d) = 0$. Damit das stimmen kann, obwohl ja $d - c \neq 0$ ist, muss $q(d) = 0$ sein, denn ein Produkt wird in $\mathbb{R}$ oder $\mathbb{C}$ Null, wenn mindestens einer der Faktoren Null ist. (Das ist die so genannten Nullteilerfreiheit in den reellen bzw komplexen Zahlen; für ihre allgemeine Formulierung in Körpern siehe 5.2). □

Satz 3.3.2. *(Fundamentalsatz der Algebra)*
Jedes nichtkonstante komplexe Polynom „zerfällt in $\mathbb{C}$ in Linearfaktoren“, d.h. ist $n \geq 1$ und $p \in \mathbb{C}[x]$ ein Polynom n-ten Grades, so es existieren $a, c_1, \dots, c_n \in \mathbb{C}$ so dass

$$p = a \cdot \prod_{k=1}^{n} (x - c_k). \tag{3.2}$$

Die Zahlen c_k, $k \in \underline{n}$, sind dabei gerade die Nullstellen von p (wobei eine Nullstelle auch mehrfach auftreten kann, d.h. es kann $c_i = c_j$ für $i \neq j$ sein).

Die Beweisidee zerlegt sich in zwei Schritte. Der erste, wesentliche ist die folgende Erkenntnis, die auf Gauß zurückgeht:

Satz 3.3.3. *Sei $p \in \mathbb{C}[x]$ ein nichtkonstantes Polynom (d.h. mit Grad$(p) \geq 1$). Dann hat p mindestens eine Nullstelle $x \in \mathbb{C}$.*

Beweis: Ist das Absolutglied a_0 von p Null, so ist $x = 0$ immer eine Nullstelle von p. Für den Fall, dass $a_0 \neq 0$ ist, findet sich ein schöner visueller Beweis, der die Grundargumente des ersten Beweises dieses Satzes (von Gauß, in seiner Doktorarbeit!) aufzeigt, auf der Seite von Edmund Weitz, die auch so sehr zu empfehlen ist. Wer es genauer wissen möchte, kann sich hier auch einen formalen Beweis sehr schön von Herrn Weitz erklären lassen. □

Zum Beweis von 3.3.2 lassen sich mit Hilfe dieses Satzes und Lemma 3.3.1 lassen sich nun sukzessive die Linearfaktoren abdividieren:

$$p(x) = (x - c_1) \cdot p_{n-1} = (x - c_1) \cdot (x - c_2) \cdot p_{n-2} = \ldots = (x - c_1) \cdot (x - c_2) \cdot a \cdot (x - c_n).$$

Dabei haben wir bei jedem Gleichheitszeichen ausgenutzt, dass die Polynome p_i, $i \in \underline{n}$, nach 3.3.2 wieder eine Nullstelle c_i von p besitzen, die nach 3.3.1 wieder abdividiert werden kann usw. Es bleibt schließlich ein Polynom p_1 vom Grad 1, d.h. von der Form vom $a \cdot (x - c_n)$, das die letzte Nullstelle für die Darstellung 3.2 liefert.

Wir müssen noch begründen, dass die Zahlen c_k ($k \in \underline{n}$) genau die Nullstellen von p sind. Ist c eine Nullstelle von p, also $p(c) = 0$, so folgt aus der Nullteilerfreiheit in $\mathbb{R}$ bzw. $\mathbb{C}$, dass für mindestens einer der Faktoren Null werden muss, für ein $k \in \underline{n}$ also $c - c_k = 0$ und demnach $c = c_k$ gilt. Umgekehrt folgt $p(c_k) = 0$ einfach durch Einsetzen von c_k in die Zerlegung in Linearfaktoren, da mindestens ein Faktor dann Null wird.

Bemerkungen dazu:

- Eine Nullstelle kann in der Darstellung mehrfach vorkommen $\rightsquigarrow$ „Vielfachheit von Nullstellen“, z.B. $p = x^n = (x - 0)^n$, also ist 0 eine Nullstelle der Vielfachheit n von p.
- Ein reelles Polynom hat also höchstens n verschiedene Nullstellen in $\mathbb{C}$.

Den letzten Punkt wollen wir noch genauer fassen, dafür als Vorbereitung:

Lemma 3.3.4. *Sei p ein Polynom mit <u>reellen Koeffizienten</u> und $c = a + ib \in \mathbb{C}$ mit $b \neq 0$ eine echt komplexe Nullstelle von p. Dann gilt:*

(i) $\overline{c}$ ist auch eine Nullstelle von p.

(ii) Für alle $c \in \mathbb{C}$ ist $(x - c) \cdot (x - \overline{c})$ ein reelles Polynom, also mit reellen Koeffizienten.

Beweis: Der Beweis beider Tatsachen folgt aus den Rechenregeln für komplexe Zahlen 3.2.4 (insbesondere denen für die komplex konjugierte) und ist Übungsaufgabe. □

Hieraus erhalten wir nun:

Satz 3.3.5. *(Vollständige Zerlegung eines Polynoms über $\mathbb{R}$)*

Jedes normierte Polynom n-ten Grades $p \in \mathbb{R}[x]$ vom Grad $n \geq 1$ lässt sich in $\mathbb{R}[x]$ in höchstens n lineare oder quadratische Faktoren zerlegen, so dass

- *die Faktoren $(x - c_i)$ vom Grad 1 gerade die reellen Nullstellen enthalten und*
- *die quadratischen Faktoren die echt komplexen Nullstellen enthalten, die man daraus mit Hilfe der „komplexen p-q-Formel“ aus 3.2.1 berechnen kann.*

Beweis: S. Vorlesung, Skizze: In der Zerlegung aus dem Fundamentalsatz lassen sich Linearfaktoren $(x - c_i)$ zu komplexen Nullstellen sukzessive immer mit den jeweiligen Faktoren $(x - \overline{c_i})$ zusammenfassen und ausmultiplizieren; Ergebnis ist nach Lemma 3.3.4 ein reelles Polynom vom Grad 2. Man kann zeigen, dass das Produkt der verbleibenden komplexen Faktoren wieder ein Polynom mit reellen Koeffizienten sein muss, das wieder mit jeder Nullstelle die komplex konjuguerte Zahl als Nullstelle enthalten muss. Hat man die höchstens $\frac{n}{2}$ komplex konjugierten Paare so zusammengefasst, bleiben noch die Linearfaktoren $(x - c_i)$, die zu den reellen Nullstellen gehören. □

Der Fundamentalsatz der Algebra ist deshalb so fundamental, weil er es (wie eigentlich alle weitreichenden Sätze der Mathematik) erlaubt, mit Hilfe einfacher, kurzer Argumente („folgt direkt aus dem Fundamentalsatz, weil...“) ansonsten nicht so offensichtliche Fragestellungen der Mathematik zu beweisen. Ein Beispiel dafür soll das folgende Korollar liefern.

Korollar 3.3.6. *Seien $p_1, p_2 \in \mathbb{R}[x]$ zwei Polynome vom gleichen Grad $n \geq 0$, deren Funktionswerte $p_1(x)$ und $p_2(x)$ an $n+1$ verschiedenen Stellen $x \in \mathbb{R}$ übereinstimmen. Dann sind p_1 und p_2 gleich.*

Beweis: Für $n = 0$ ist die Behauptung klar, da es dann um zwei konstante Polynome geht. Für $n \geq 1$ ist $p_1 - p_2$ ist ein Polynom mit mindestens $n+1$ Nullstellen, das sich als Polynom aus $\mathbb{C}[x]$ auffassen lässt, also nach Fundamentalsatz konstant sein muss. Wegen $(p_1 - p_2)(x) = 0$ an den übereinstimmenden Stellen x ist also $p_1 - p_2 = 0$, also $p_1 = p_2$. □

Abschlussbemerkung zum Fundamentalsatz. Der Fundamentalsatz der Algebra und auch der eben bewiesene Satz 3.3.5 sind theoretische Aussagen, die die Existenz solcher

Nullstellen und Zerlegungen sichern. Nach den Bemerkungen weiter oben gibt es aber (zumindest für Polynome von $\mathrm{Grad}(p) \geq 5$) kein allgemeines exaktes algebraisches Verfahren zur Berechnung dieser Nullstellen, sondern nur numerische Näherungsverfahren dafür. Das zeigt uns: Dass es etwas gibt und dass man es konkret und einfach mit einer allgemeingültige Formel ausrechnen kann, sind zwei verschiedene Sachen.

3.4 Algebraische Abgeschlossenheit von $\mathbb{C}$

Lemma 3.4.1. *Sei $z \neq 0$ eine komplexe Zahl und $n \in \mathbb{N}$. Dann gibt es genau n verschiedene Zahlen $w_k \in \mathbb{C}, k \in \underline{n}$ mit*

$$w_k^n = z,$$

d.h. jede komplexe Zahl $z \neq 0$ besitzt genau n n-te Wurzeln $w_k \in \mathbb{C}$. ($z = 0$ hat nur die n-te Wurzel $w = 0$.)

Bemerkungen dazu:

- Anders als in den reellen Zahlen, wo die Quadratwurzel einer Zahl a die *positive* Zahl x ist, die $x^2 = a$ erfüllt, spricht man im komplexen von *einer* Wurzel w von z, wenn $w^n = z$ gilt. Das ist vielleicht etwas verwirrend, aber so üblich.
- Eines schöne Visualisierung der n-ten Wurzeln einer (im Applet dynamisch änderbaren) komplexen Zahl z ist hier zu finden.

Beweis:

- Die Wurzeln von z sind genau die Nullstellen des Polynoms $p(x) = x^n - z \in \mathbb{C}[x]$, also gibt es nach Fundamentalsatz höchstens n Wurzeln von z.
- Sei $z = r \cdot (c_\alpha + is_\alpha)$. Die n komplexen Zahlen

$$w_k = r^{\frac{1}{n}} \cdot \left(\cos(\frac{\alpha + 2\pi \cdot k}{n}) + i \sin(\frac{\alpha + 2\pi \cdot k}{n}) \right)$$

  mit $k \in \{0, \ldots, n-1\}$ sind (für $r > 0$) alle verschieden, und man rechnet nach (z.B. mit der Rechenregel für Potenzen komplexer Zahlen), dass für sie $w_k^n = z$ gilt.

$\square$

Schlussbemerkung: Zahlbereichserweitungen und Umkehrbarkeit von Rechenoperationen

- Die Menge der ganzen Zahlen $(\mathbb{Z}, +)$ hat gegenüber $(\mathbb{N}, +)$ den Vorteil, dass sie bezüglich der Addition und ihrer Umkehroperation $-$ abgeschlossen: Die zur Addition einer Zahl a gehörige Funktion $x \mapsto x + a$ ist bijektiv und daher immer umkehrbar mit Umkehrfunktion $x \mapsto x - a$, wir haben

$$x + a = b \qquad \Longleftrightarrow \qquad x = b - a$$

  (vgl. nochmal den allgemeinen Satz 2.5.3).

- $(\mathbb{Q}\backslash\{0\}, \cdot)$ und $(\mathbb{R}\backslash\{0\}, \cdot)$ sind multiplikativ abgeschlossen: Die zur Multiplikation mit $a \neq 0$ gehörige die Funktion $x \mapsto a \cdot x$, ist im Gegensatz zur Multiplikation in $(\mathbb{Z}, +)$, bijektiv und daher immer (eindeutig) umkehrbar (mit Umkehrfunktion $x \mapsto \frac{x}{a}$): Wir haben

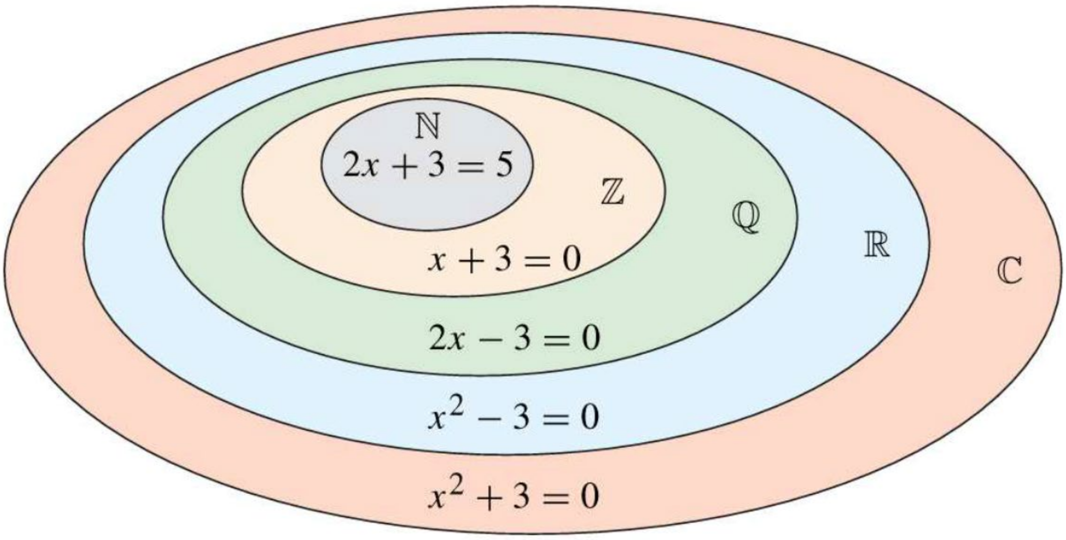
$$a \cdot x = b \qquad \Longleftrightarrow \qquad x = \frac{b}{a}.$$

  Das folgt ebenfalls aus 5.2.7.

- $(\mathbb{C}, +, \cdot)$ hat die zusätzliche Eigenschaft, dass es gegenüber Potenzieren abgeschlossen ist, d.h. die Abbildung $x \mapsto x^n$ $(n \in \mathbb{N})$ ist surjektiv (in $\mathbb{R}$ i.A. nicht, vgl, $x \mapsto x^2$), daher hat (vgl. nochmal Satz 2.5.3) eine Gleichung $w^n = z$ ($z \in \mathbb{C}$, x gesucht) in $\mathbb{C}$ immer mindestens eine Lösung. (Die Abbildung $x \mapsto x^n$ ist allerdings nicht injektiv, d.h. die Gleichung nicht eindeutig lösbar, es gilt also nicht „die“ n-te Wurzel (sondern n verschiedene, s. Lemma 3.4.1.)

In den komplexen Zahlen sind also die gängigen Rechenoperationen (Addition, Multiplikation, Potenzieren) umkehrbar, d.h. neben der (eindeutigen) Umkehrbarkeit von Addition und Multiplikation (bis auf den in Ringen immer ausgeschlossenen Fall $z = 0$) auch das Potenzieren immer umkehrbar (wenn auch nicht eindeutig). Man nennt den Körper $\mathbb{C}$ daher <u>algebraisch abgeschlossen</u>.

Ende „Komplexe Zahlen und Polynome“ (Kap. IV)



3.5 Aufgaben zu Kapitel 3

Aufgabe 3.7. (Umgang mit komplexen Zahlen):

(a) Gegeben sind die komplexen Zahlen $z = 1 + 2i, w = \frac{1}{2} + \frac{\sqrt{3}}{2}i$. Skizzieren Sie (jeweils in einem Koordinatensystem pro Unteraufgabe) in der Gauß'schen Zahlenebene die komplexen Zahlen

(i) $z,\ i \cdot z,\ i^2 \cdot z,\ i^3 \cdot z,\ i^4 \cdot z,$

(ii) $w,\ w^2,\ w^3,\ w^4,$

(iii) $z^{-1},\ z^{-2}, z^{-3},$

(iv) $2z - 3w,\ z \cdot w,\ \frac{w}{z}.$

(b) Sei

$$A = \{z \in \mathbb{C} \mid \frac{\pi}{4} \leq \arg(z) < \frac{\pi}{2}, 2 \leq |z| < 4\}.$$

Skizzieren Sie in der Gauß'schen Zahlenebene die Menge A sowie die Mengen

$$B = \{z^2 \mid z \in A\},$$

$$C = \{z^{-1} \mid z \in A\},$$

$$D = \{z \in \mathbb{C} \mid z^2 \in A\}.$$

Aufgabe 3.8. (Eigenschaften von Polynomen über $\mathbb{C}$):

(a) Beweisen Sie: Für alle $c \in \mathbb{C}$ ist $(x-c)\cdot(x-\overline{c})$ ein Polynom mit reellen Koeffizienten.

(b) Sei $p \in \mathbb{R}[x]$, also p ein Polynom mit reellen Koeffizienten, und $c = a + bi \in \mathbb{C}$ mit $b \neq 0$ eine echt komplexe Nullstelle von p. Zeigen Sie: Dann ist auch $\overline{c}$ eine Nullstelle von p.

Sie dürfen dazu neben in der Vorlesung bewiesenen Tatsachen auch alle auf diesem Übungsblatt zu zeigenden Resultate verwenden.

Aufgabe 3.9. (Rechenregeln und -gesetze für komplexe Zahlen):

(a) Seien $z = a + b \cdot i$, $w = c + d \cdot i$ komplexe Zahlen. Beweisen Sie die folgenden Rechenregeln
(aus Lemma 3.2.4):

(iv) $|z| = \sqrt{z \cdot \overline{z}}$,

(v) $\frac{1}{2}(z + \overline{z}) =Re(z)$,

(vi) $\frac{1}{2i}(z - \overline{z}) =Im(z)$,

(vii) $\overline{w \cdot z} = \overline{w} \cdot \overline{z}$,

(viii) $\overline{w + z} = \overline{w} + \overline{z}$.

(b) Um zu zeigen, dass die komplexen Zahlen mit der Addition

$$z + w = (a + c) + (b + d) \cdot i$$

und der Multiplikation

$$z \cdot w = (ac - bd) + (ad + bc) \cdot i$$

wirklich ein Körper sind, müssen für beide Verknüpfungen $+$ und $\cdot$ jeweils Assoziativgesetz und Kommutativgesetz geprüft und die Distributivgesetze verifiziert werden. Alle diese Gesetze gelten in $\mathbb{C}$. Prüfen Sie hier exemplarisch:

(i) Für alle $z, w \in (\mathbb{C}, \cdot)$ gilt das Kommutativgesetz, d.h. es ist $z \cdot w = w \cdot z$.

(ii) Für alle $z, w_1, w_2 \in (\mathbb{C}, +, \cdot)$ gilt das Distributivgesetz $z \cdot (w_1 + w_2) = z \cdot w_1 + z \cdot w_2$.

Aufgabe 3.10. (Nullstellen und Faktorisierungen eines Polynoms):

Gegeben ist das Polynom $p(x) = x^3 - 3x^2 + x + 5$.

a) Finden Sie eine reelle Nullstelle von p. Überprüfen Sie, dass es sich tatsächlich um eine Nullstelle handelt und reduzieren Sie so den Grad des zu untersuchenden Polynoms mit Hilfe von Polynomdivision.

b) Bestimmen Sie die Nullstellen des so entstandenen quadratischen Polynoms in $\mathbb{C}$.

c) Geben Sie für p eine Faktorisierung in komplexe Polynome vom Grad 1 (also eine Zerlegung in Linearfaktoren) und eine Faktorisierung in ausschließlich reelle Polynome vom Grad 1 und 2 an.

Aufgabe 3.11. (Polynome und ihre Nullstellen):

(a) Gibt es Polynome 3. Grades in $\mathbb{R}[x]$, die

  (i) genau drei verschiedene reelle Nullstellen haben?

  (ii) genau zwei verschiedene reelle Nullstellen haben?

  (iii) genau eine reelle Nullstelle haben?

  (iv) keine reelle Nullstelle haben?

  Wenn ja, finden Sie ein Beispiel und weisen Sie nach, dass es die gewünschte Anzahl reeller Nullstellen hat. Wenn nicht, begründen Sie (z.B. mit Hilfe geeigneter Sätze aus der Vorlesung), warum es kein solches Polynom gibt.

(b) (i) Zeigen Sie: Zwischen einem Polynom $p = x^2 + a_1 x + a_0 \in \mathbb{C}[x]$ vom Grad 2 und seinen Nullstellen[3] c_1, c_2 besteht der Zusammenhang

$$a_1 = -(c_1 + c_2), \qquad a_0 = c_1 \cdot c_2.$$

  (ii) Sei $p = \sum_{i=0}^{4} a_i x^i$ ein normiertes Polynom vierten Grades (also mit $a_4 = 1$), das die Nullstellen[1] $c_1, \ldots, c_4$ hat. Drücken Sie analog zu Teil (a) die Koeffizienten a_0, a_1, a_2 und a_3 durch die Nullstellen von p aus.

  (iii) Sei $n \in \mathbb{N}$ beliebig, $p = \sum_{i=0}^{n} a_i x^i$ ein normiertes Polynom vom Grad n und $c_1, \ldots, c_n$ die Nullstellen[1] von p. Drücken Sie die Koeffizienten a_{n-1} und a_0 durch die Nullstellen von p aus und begründen Sie diesen Zusammenhang.

[3] wobei die Nullstellen hier gemäß ihren Vielfachheiten aufgeführt werden, d.h. c_1 und c_2 bzw. c_i und c_j für $i \neq j$ können auch gleich sein.

Aufgabe 3.12. (n-te Einheitswurzeln):

Sei $n \in \mathbb{N}$. Die n-ten Einheitswurzeln in $\mathbb{C}$ sind die Zahlen $w \in \mathbb{C}$ mit $w^n = 1$, also gerade die Nullstellen des Polynoms $p \in \mathbb{C}[x]$ mit $p = x^n - 1$.

(a) Geben Sie für allgemeines n die $n-$ten Einheitswurzeln in Polardarstellung an. Skizzieren Sie die $n-$ten Einheitswurzeln für $n = 3, 5, 8$.

(b) Zerlegen Sie das Polynom $p = x^6 - 1$ in $\mathbb{R}[x]$ so in reelle Linearfaktoren und reelle quadratische Faktoren, dass die Linearfaktoren gerade die reellen Nullstellen enthalten und die quadratischen Faktoren die echt komplexen Nullstellen.

(c) Sei $n \in \mathbb{N}$ und $a + bi \in \mathbb{C}$ so, dass gilt: $(a + bi)^n = 2020$. Zeigen Sie, dass $(a - bi)^n = 2020$ ist.

Aufgabe 3.13. (Summen- und Indexschreibweise):

Wir werden im kommenden Kapitel die so genannte Summenschreibweise und Indizierung von Größen verwenden, z.B. zur Verkürzung für lange Summationen.

(a) Bitte machen Sie sich mit der Summenschreibweise vertraut, z.B. hier: (Link)

(b) Indizes werden verwendet, um gleichartige Objekte (Zahlen, Vektoren) systematisch durchzunummerieren. Wir betrachten drei Vektoren x_1, x_2, x_3 aus dem $\mathbb{R}^4$ und einen Vektor a aus dem $\mathbb{R}^3$ mit den Einträgen

$$x_1 = \begin{pmatrix} x_{11} \\ x_{21} \\ x_{31} \\ x_{41} \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}, \quad x_2 = \begin{pmatrix} x_{12} \\ x_{22} \\ x_{32} \\ x_{42} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} \quad x_3 = \begin{pmatrix} x_{13} \\ x_{23} \\ x_{33} \\ x_{43} \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ -2 \\ 2 \end{pmatrix}, \quad a = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 0 \\ -2 \\ 1 \end{pmatrix}.$$

Berechnen Sie aus diesen Größen konkret:

(i) $\sum_{i=1}^{3} x_i$,

(ii) $\sum_{i=1}^{4} x_{i,1} \cdot x_{1,2}$,

(iii) $\sum_{i=1}^{3} a_i \cdot x_i$,

(iv) $\sum_{i=1}^{4} \sum_{j=1}^{3} x_{i,j}$.

Überlegen Sie sich jeweils vorher, ob eine Zahl oder ein Vektor herauskommt!

4 Vektorrechnung und Matrixabbildungen

Nach dem ersten Rundumschlag zu einigen Themen der Algebra wollen wir uns nun den Kern der Linearen Algebra, den Vektorräumen und linearen Abbildungen nähern. Wir tun das, indem wir rekapitulieren, wie man sich zwischen den Punkten von Ebene und Raum mit Hilfe von Vektoren bewegen kann. Dazu wollen wir wieder auffrischen, was Sie dazu hoffentlich noch aus der Schule wissen: Wie man Ebene und Raum koordinatisiert, d.h. wir klären, was eigentlich ein Koordinatensystem ist und was das mit Punkten und Vektoren zu tun hat. Erfahrungsgemäß gilt es dabei gelegentlich, einige Vorstellungen zu diesen Konzepten wie z.B. dem Ortsvektor gerade zu rücken, die in der Schule vielleicht nicht ganz so klar herausgekommen sind. Wir werden dann so genannte Matrixabbildungen definieren, die z.B. die Ebene oder den Raum auf relativ einfach Art und Weise sich selbst abbilden – Sie könnten hier an geometrische Abbildungen wie Drehungen, Spiegelungen oder Streckungen als Beispiele denken. Wir wandeln hier in den Spuren des guten alten C. F. Gauß, der nach meinem Wissen der erste war, der geometrische Abbildungen mit Matrizen beschrieb.

4.1 Koordinatisierung von Raum und Ebene; die Koordinatenräume $\mathbb{R}^n$

Wir beginnen mir der Konstruktion eines Koordinatensystems für

- die zweidimensionale Ebene unserer Anschauung (die *Euklidische Punktebene*, die wir ab hier mit $\mathcal{A}_2$ bezeichnen wollen) bzw.
- den dreidimensionaler Raum unserer Anschauung (den *Euklidischen Punktraum*, den wir mit $\mathcal{A}_3$ bezeichnen).
- $\mathcal{A}_2$ und $\mathcal{A}_3$ enthalten also genau alle Punkte P der Ebene bzw. des Raumes. Sie sind formalisierbar durch das, was wir mathematisch als affine Punkträume bezeichnen werden. Die exakte Theorie dazu werden wir und tiefergehend in Teil II dieses Skripts ansehen.

Diese lässt sich folgendermaßen skizzieren:

- Durch Wahl eines Ursprung O (ein fest gewählter Punkt in $\mathcal{A}_2$ bzw. $\mathcal{A}_3$) plus zwei bzw. drei unabhängiger Richtungen $\vec{e}_1, \vec{e}_2(, \vec{e}_3)$ erhält man ein Koordinatensystem für $\mathcal{A}_2$ bzw. $\mathcal{A}_3$. Die Vektoren $\vec{e}_1, \vec{e}_2, \vec{e}_3$ geben dabei jeweils eine Raumrichtung *und* die Größe einer Einheit in diese Richtung vor.

- Jeder Punkt P ist nun vom Ursprung O aus eindeutig zu erreichen durch Anweisung der Form „Gehe von O aus x_1 Schritte in Richtung $\vec{e}_1$, x_2 Schritte in Richtung $\vec{e}_2$(, x_3 Schritte in Richtung $\vec{e}_3$)“. Durch dieses „Hintereinanderhängen“ von Vielfachen von $\vec{e}_1, \vec{e}_2(, \vec{e}_3)$ erhält man einen Vektor

$$\vec{v} = x_1 \cdot \vec{e}_1 + x_2 \cdot \vec{e}_2 + x_3 \cdot \vec{e}_3 \tag{4.1}$$

der den Ursprung O auf P verschiebt (*transportiert, vgl. lat.: vector*). Das ist der so genannte Ortsvektor $\overrightarrow{OP}$ von P. Hierbei sind alle reellen Zahlen als Schrittzahlen in jede Richtung zulässig (insb. auch nicht ganzzahlige oder negative).

Zu jedem Punkt gehört also (bei fest gewähltem Ursprung O) genau ein eindeutiger Ortsvektor und umgekehrt. Der Ortsvektor von P und somit auch P ist so eindeutig durch einen Koordinatenvektor

$$v = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \qquad \text{bzw.} \qquad v = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

beschreibbar, wobei die Einträge x_1, x_2, x_3 (die „Schrittweiten“ aus (4.1) oben) reelle Zahlen sind. Punkte des Raumes bzw. der Ebene entsprechen also mit der folgenden Definition eins-zu-eins den *Koordinatenvektoren* des $\mathbb{R}^2$ bzw. des $\mathbb{R}^3$.

Definition 4.1.1. *(Koordinatenvektoren; „der $\mathbb{R}^n$“)*

Eine sortierte Aneinanderreihung von n Zahlen nennt man ein n-Tupel von Zahlen. Ein n-Tupel von reellen Zahlen nennen wir einen Koordinatenvektor aus reellen Zahlen. Er wird für gewöhnlich als Spaltenvektor, also in der Form

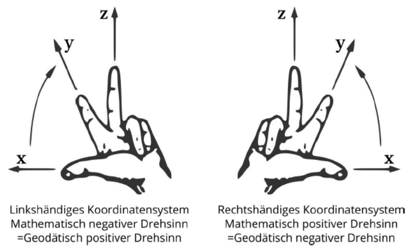
$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

notiert, wobei die Einträge $x_1, \ldots, x_n$ reelle Zahlen sind.[1] *Wir nutzen auch die Notation $x = (x_i)_{i \in \underline{n}}$ oder kürzer $x = (x_i)$ (wenn klar ist, was n ist). Die Zusammenfassung aller solcher Vektoren bezeichnet man mit dem Kürzel $\mathbb{R}^n$ (und spricht von „dem $\mathbb{R}^n$“).*

[1] Wir werden zunächst nicht zwischen x und seinem zugehörigen Zeilenvektor $(x_1, \ldots, x_n)$ unterscheiden und später darauf hinweisen, wenn diese Unterscheidung wichtig wird (in der LAAG II).

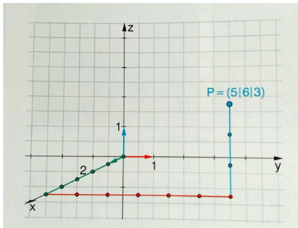
Koordinatensysteme

Anschauungsraum
$\hat{=}$
„affiner Punktraum“



- Wahl eines Ursprungs O
- Wahl zweier nicht kollinearer/drei nicht komplanarer Raum-/Bewegungsrichtungen und jeweils einer Einheitslänge für diese ($\rightsquigarrow$ Basisvektoren).

- Jeder Punkt kann durch „Aneinanderhängen“ positiver/negativer Vielfache dieser Einheitsrichtungen erreicht werden.



Wir haben gerade ein erstes Beispiel für einen formalen Vektorraum kennen gelernt: Es ist der oben definierte Vektorraum $\mathbb{R}^n$, der sich für $n = 3$ im Anschauungsraum deuten lässt, indem man dort ein Koordinatensystem eingeführt und jedem Vektor $x = (x_1, x_2, x_3)$ mit dem Vektor $\vec{v} = x_! \cdot \vec{e}_1 + x_2 \cdot \vec{e}_2 + x_3 \cdot \vec{e}_3$; identifiziert. Interpretiert man v als Ortsvektor, den man an den Ursprung O des Koordinatensystems ansetzt, so gehört dazu immer genau ein Punkt im Anschauungsraum $\mathcal{A}_3$.

Hier geht auch oftmals in der Literatur einiges durcheinander. Daher nochmal zur Klarheit: Der Punkt P, der Ortsvektor $\vec{v} = \overrightarrow{OP}$ von P (der vom gewählten Ursprung abhängt) und der ihnen gemeinsame Koordinatenvektor sind komplett verschiedene mathematische Denkgegenstände. Und zwar:

- P ist ein Punkt P der Anschauungsebene $\mathcal{A}_2$ bzw. des Raumes $\mathcal{A}_3$,
- $\vec{v}$ ist ein anderes geometrisches Objekt, ein „mobiler Pfeil“ mit festgelegter Länge, Richtung, Orientierung. Er lässt sich *frei im Raum bewegen. Wenn* man ihn im Ursprung O ansetzt, gelangt man zu P als Endpunkt.
- Der Koordinatenvektor von P und v ist ein 2- bzw. 3-Tupel aus reellen Zahlen, also ein Element des $\mathbb{R}^n$!

Eine Bezeichnung wie $P(3|5)$ steht hierbei z.B. in vielen Schulbüchern für einen Punkt (nämlich den mit den Koordinaten $(3,5)$). Dabei ist mehrfach Vorsicht angebracht: Das macht erstens nur dann Sinn, wenn das Koordinatensystem fest gewählt ist, bezüglich dem die Koordinaten angegeben werden; zweitens findet man leider oft die noch weiter verkürzende Sprechweise „Der Punkt $(3,5)$“, die P und seinen Koordinatenvektor quasi gleichsetzt. Das hat eine gewisse Rechtfertigung darin, dass räumliche Vektoren und Koordinatenvektoren sich gleich verhalten, wenn man sie addiert oder mit Skalaren multipliziert. Wenn wir aber exakte Mathematik betreiben wollen, sollten wir genau aufpassen, über welche Denkobjekte wir gerade reden. Genauso wird die Anschauungsebene (dieser sehr schwer zu definierende Haufen von Punkten, den wir und als euklidische Ebene vorstellen) vielerorts relativ unbedacht mit dem $\mathbb{R}^2$ gleichgesetzt. Ich hoffe, es ist nun klar, dass wir hier über verschiedene Objekte reden.

4.2 Was wir noch aus der Vektorrechnung der Schule brauchen

Im kommenden Abschnitt werden wir sehen, wie man in $\mathbb{R}^2$ und $\mathbb{R}^3$ Abbildungen dieser Mengen in sich definieren kann, die sich in $\mathcal{A}_2$ und $\mathcal{A}_3$ als geometrische Abbildungen deuten lassen. Diese sollen uns nun als erste Beispiele für die in der Lineare Algebra zentralen linearen Abbildungen dienen.

Dazu brauchen wir noch eine Zutat, nein, zwei: Zu der Definition eines Vektorraumes gehören neben der Definition seiner Elemente auch immer zwei charakteristische Verknüpfungen, die Sie beide aus der Schule kennen sollten und die wiederum anschauliche Interpretationen haben, wenn wir Koordinatenvektoren des $\mathbb{R}^2$ bzw. $\mathbb{R}^3$ mit Verschiebungspfeilen identifizieren: Die Addition $+$ und die skalare Multiplikation $\cdot$.

Definition 4.2.1. *(Addition und skalare Multiplikation von Vektoren des $\mathbb{R}^n$)*
Sind $x = (x_i)_{1\le i\le n}, y = (y_i)_{1\le i\le n}$ zwei Vektoren aus dem $\mathbb{R}^n$, so definieren wir ihre Summe $x+y$ durch den Vektor $x = (x_i+y_i)_{1\le i\le n}$, also durch

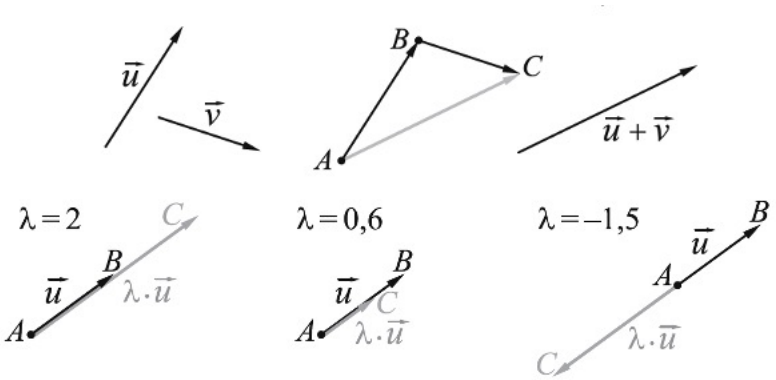
$$x + y := \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix},$$

und die Skalierung $a\cdot x$ von x mit einer reellen Zahl a durch

$$a \cdot x = (a\cdot x_i)_{1\le i\le n} = \begin{pmatrix} a\cdot x_1 \\ a\cdot x_2 \\ \vdots \\ a\cdot x_n \end{pmatrix}.$$

Graphisch lassen sich Addition und skalare Multiplikation von Vektoren in Raum und

Ebene damit konsistent folgendermaßen erklären:



(aus: A. Filler, Elementare Lineare Algebra [Filler].)

(Wie man die Summe $\vec{v}+\vec{w}$ zweier Vektoren $\vec{v},\vec{w}$ in der Anschauungsebene und Anschauungsraum bildet, ist hoffentlich klar. Genauso sollte Ihnen klar sein, wie man skalare Vielfache $a\cdot\vec{v}$ eines Vektors bildet, auch für negative Skalare a oder wenn $a=0$ ist.)

Wir werden im Folgenden noch einige Begriffe und Zusammenhänge für Vektoren des $\mathbb{R}^n$ beziehungsweise des Anschauungsraumes benutzen, die vielleicht aus der Schule bekannt sind; wenn nicht, sind sie hier der Vollständigkeit halber aufgeführt:

- Haben wir ein Koordinatensystem mit Achsen $\vec{e}_1,\vec{e}_2,\vec{e}_3$ gegeben, so lässt sich jeder Vektor $\vec{v}$ des Anschauungsraumes wie oben erwähnt als

  $$\vec{v}=x_1\cdot\vec{e}_1+x_2\cdot\vec{e}_2+x_3\cdot\vec{e}_3,$$

  also als Summe von Vielfachen der Vektoren $\vec{e}_1,\vec{e}_2,\vec{e}_3$ schreiben. Eine solche Summe bezeichnen wir als Linearkombination der Vektoren $\vec{e}_1,\vec{e}_2,\vec{e}_3$, der Vektor $\vec{v}$ ist also eine Linearkombination von $\vec{e}_1,\vec{e}_2,\vec{e}_3$. Da sich durch das „Anheften" des Vektors $\vec{v}$ am Ursprung Oso jeder Punkt durch Linearkombinationen der Raumrichtungen $\vec{e}_1,\vec{e}_2,\vec{e}_3$ erreichen lässt, sagt man, die drei Vektoren $\vec{e}_1,\vec{e}_2,\vec{e}_3$ „spannen den euklidischen Raum" $\mathcal{A}_3$ auf.

- Die drei Koordinatenvektoren

  $$\mathrm{e}_1=\begin{pmatrix}1\\0\\0\end{pmatrix},\quad \mathrm{e}_2=\begin{pmatrix}0\\1\\0\end{pmatrix},\quad \mathrm{e}_3=\begin{pmatrix}0\\1\\0\end{pmatrix}$$

bezeichnet man als die Standardbasisvektoren des $\mathbb{R}^3$. Beschreibt man den Anschauungsraum $\mathcal{A}_3$ mit Hilfe eines Koordinatensystems $(O, \vec{e_1}, \vec{e_2}, \vec{e_3})$, so sind diese Standardbasisvektoren des $\mathbb{R}^3$ gerade die Koordinatenvektoren der Vektoren $\vec{e}_1, \vec{e}_2, \vec{e}_3$ (denn es ist ja

$$\vec{e_1} = 1 \cdot \vec{e_1} + 0 \cdot \vec{e_2} + 0 \cdot \vec{e_3}$$

usw.).

- Analog zur Situation im Anschauungsraum gilt: Jeder Koordinatenvektor $v \in \mathbb{R}^3$ lässt sich als Linearkombination der Standardbasisvektoren des $\mathbb{R}^3$ schreiben:

$$v = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_1 \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + x_2 \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + x_3 \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = x_1 \cdot e_1 + x_2 \cdot e_2 + x_3 \cdot e_3$$

  Die drei Vektoren e_1, e_2, e_3 spannen also den Koordinatenraum $\mathbb{R}^3$ auf.

- Allgemein sollte anschaulich plausibel sein, dass n Vektoren im Anschauungsraum spannen eine n-dimensionales Objekt „aufspannen“ (für $n = 1$ ist das eine Ursprungsgerade, für $n = 2$ eine Ebene durch den Ursprung, für $n = 3$ der gesamte dreidimensionale Raum) -– es sei denn, einer der Vektoren ist Linearkombination der anderen, d.h. sie sind kollinear bzw. komplanar. (Zwei Vektoren in der Ebene bzw. im Anschauungsraum nennt man ja gerade kollinear, wenn sie keine Ebene aufspannen, also auf einer gemeinsamen Ursprungsgeraden liegen; drei Vektoren im Raum heißen komplanar, wenn sie nicht den gesamten Raum aufspannen (also nur eine Ebene bzw. Gerade bzw. nur den Nullpunkt erzeugen)).

Wenn Sie jetzt das Gefühl bekommen, dass es egal zu sein scheint, ob man mit räumlichen Vektoren oder mit ihren Koordinatenvektoren rechnet, da sich die Vektoren, was die beiden obigen Operationen angeht, „gleich verhalten“, obwohl es sich ja, wie oben erwähnt, um grundverschiedene mathematische Objekte handelt, haben Sie ein wichtiges Konzept der Linearen Algebra verstanden. (Wir werden dazu sagen, dass endlichdimensionale Vektorräume isomorph zu ihren, d.h. „von gleicher Gestalt wie ihre“ Koordinatenräume sind.)

4.3 Matrixabbildungen in $\mathbb{R}^n$

Wir betrachten die euklidische Ebene $\mathcal{A}_2$ und den zugehörigen Koordinatenraum $\mathbb{R}^2$. Auf dieser können wir Abbildungen dadurch definieren, dass wir eine Abbildungsvorschrift für die *Koordinaten* angeben, die jeden Vektor $(x_1, x_2) \in \mathbb{R}^2$ auf seinen Bildvektor $f(x_1, x_2) \in \mathbb{R}^2$ abbildet. Z.B. wäre ein Beispiel für eine solche Abbildung die Funktion

$$g : \mathbb{R}^2 \to \mathbb{R}^2, \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x^2 - \frac{1}{3}y^3 \\ x \cdot y \end{pmatrix}.$$

Solche Abbildungen (Transformationen der Ebene bzw. des Raumes) können sich sehr kompliziert verhalten, wenn es um Fragen geht wie:

- Welche Werte kann g überhaupt annehmen, d.h. was ist $\text{Bild}(g)$?
- Wird ein bestimmtes y als Bild von g angenommen (d.h. hat die Gleichung $g(x) = y$ eine Lösung)?
- Kann man die Wirkung von f sonst irgendwie vernünftig beschreiben? Worauf werden z.B. geometrische Standardformen (Geraden, Ebenen, Kreise,...) durch g abgebildet?

Sie können sich mit dem Applet hier (Link) überzeugen, dass die Funktion g oben nicht einfach ist: Bewegen Sie den Eingabepunkt im Fenster links. Rechts sehen Sie die Ausgabe nach Anwendung von g. Versuchen Sie vorherzusagen, wie sich die Ausgabe verhält!

4.3.1 Konstruktion von Matrixabbildungen

Eine sehr wichtige Klasse von Abbildungen von $\mathbb{R}^2$ nach $\mathbb{R}^2$ und allgemeiner von $\mathbb{R}^n$ nach $\mathbb{R}^m$ $(n, m \in \mathbb{N})$ sind die so genannten Matrixabbildungen – hauptsächlich, weil sie einfacher als das obige Beispiel sind und *trotzdem* (bzw. sogar eher) sinnvolle Sachen beschreiben. Sie können wir nach dem folgenden Bauplan definieren:

<u>Konstruktion einer Matrixabbildung $f : \mathbb{R}^2 \to \mathbb{R}^2$</u>

- Definiere die Bilder $f(e_1), f(e_2)$ der Standardbasisvektoren e_1, e_2.

$$\text{z.B. } f(\begin{pmatrix} 1 \\ 0 \end{pmatrix}) = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, f(\begin{pmatrix} 0 \\ 1 \end{pmatrix}) = \begin{pmatrix} 0 \\ 2 \end{pmatrix},$$

- Setze f linear fort, d.h.

- Vielfache von Basisvektoren werden auf das gleiche Vielfache des Bildes abgebildet, also ist
$$f(a \cdot e_i) := a \cdot f(e_i) \qquad (i = 1, 2),$$
$$z.B.\ f(\begin{pmatrix}3\\0\end{pmatrix}) = 3 \cdot f(\begin{pmatrix}1\\0\end{pmatrix}) = \begin{pmatrix}6\\3\end{pmatrix}, f(\begin{pmatrix}0\\-2\end{pmatrix}) = (-2) \cdot \begin{pmatrix}0\\2\end{pmatrix} = \begin{pmatrix}0\\-4\end{pmatrix}.$$
- Summen von solchen Vielfachen werden auf die Summe ihrer Bilder abgebildet, man setzt also fest:
$$f(x_1 \cdot e_1 + x_2 \cdot e_2) := f(x_1 \cdot e_1) + f(x_2 \cdot e_2) \overset{\text{s.oben}}{:=} x_1 \cdot f(e_1) + x_2 \cdot f(e_2).$$
$$z.B.\ f(\begin{pmatrix}x_1\\x_2\end{pmatrix}) = x_1 \cdot \begin{pmatrix}2\\1\end{pmatrix} + x_2 \cdot \begin{pmatrix}0\\-2\end{pmatrix} = \begin{pmatrix}2x_1\\x_1 - 2x_2\end{pmatrix}.$$

- Eine so definierte Abbildung lässt sich handlich und effizient notieren, indem man ihre Abbildungsmatrix angibt: Man schreibt die Bildvektoren $f(e_1), f(e_2)$ spaltenweise hintereinander.
$$z.B.\ A = \begin{pmatrix}2 & 0\\1 & -2\end{pmatrix}$$

Hier (Link) finden Sie ein GeoGebra-Applet dazu, in dem Sie mit Matrixabbildungen herumspielen können, die sich als Abbildungen der Ebene auf sich interpretieren lassen. Im linken Koordinatensystem sehen Sie den Definitionsbereich der zugehörigen Matrixabbildung mit einigen geometrischen Objekten, im rechten Koordinatensystem die zugehörigen Bilder. Ganz rechts können Sie die Abbildungsmatrix eintragen. Probieren Sie beispielsweise die folgenden

Beispiele für Matrixabbildungen im $\mathbb{R}^2$:

- Spiegelung an x-Achse: Abbildungsmatrix $A = \begin{pmatrix}1 & 0\\0 & -1\end{pmatrix}$
- Drehung um Winkel $\alpha \in [0, 2\pi[$. : Abbildungsmatrix $A = \begin{pmatrix}\cos(\alpha) & -\sin(\alpha)\\ \sin(\alpha) & \cos(\alpha)\end{pmatrix}$
- Punktspiegelung am Ursprung: $A = \begin{pmatrix}-1 & 0\\0 & -1\end{pmatrix}$
- Projektion auf die y-Achse: $A = \begin{pmatrix}0 & 0\\0 & 1\end{pmatrix}$

- Zentrische Streckung der Achsen mit Faktor 2: $A = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$
- Elliptische Streckung mit Koordinatenachsen als Ellipsenachsen: $A = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$
- Spiegelung an Gerade mit Winkel $\alpha \in [0, \pi[$ zur x-Achse : $A = \begin{pmatrix} \cos(2\alpha) & \sin(2\alpha) \\ \sin(2\alpha) & -\cos(2\alpha) \end{pmatrix}$

...

Auch im Raum lassen sich viele nützliche geometrische Abbildungen als Matrixabbildungen beschreiben, z.B. zyklische Vertauschung der Achsen, Drehung um feste Achse, Projektion,...

Um allgemeine Matrixabbildungen $f : \mathbb{R}^n \to \mathbb{R}^m$ sinnvoll notieren zu können, halten wir die oben eingeführten Ideen und Begriffe in allgemeinen Definitionen fest:

Definition 4.3.1. *(*Reelle $(m \times n)$-Matrix*):*
Seien m und n natürliche Zahlen. Eine reelle $(m \times n)$-Matrix ist ein rechteckiges Schema aus m Zeilen und n Spalten, deren Einträge reelle Zahlen sind. Also sieht eine $(m \times n)$-Matrix A mit den Einträgen $a_{ij} \in \mathbb{R}$ $(i \in \underline{m}, j \in \underline{n})$ folgendermaßen aus:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} | & | & \cdots & | \\ s_1 & s_2 & \cdots & s_n \\ | & | & \cdots & | \end{pmatrix} = \begin{pmatrix} — & z_1 & — \\ — & z_2 & — \\ \vdots & \vdots & \vdots \\ — & z_m & — \end{pmatrix}$$

Die Vektoren $s_1, \dots, s_n \in \mathbb{R}^m$ heißen die Spaltenvektoren oder einfach die Spalten von A, die Vektoren $z_1, \dots, z_m \in \mathbb{R}^n$ heißen die Zeilenvektoren oder die Zeilen von A. Wir schreiben für eine Matrix A oft auch

$$A = (a_{ij})_{i \in \underline{m}, j \in \underline{n}} \quad (a_{ij} \in \mathbb{R})$$

oder kürzer $A = (a_{ij})$ (wenn n, m klar sind). Die Menge aller reellen $(m \times n)$-Matrizen wird mit $\mathbb{R}^{m \times n}$ bezeichnet.

- Bitte beachten Sie, dass hier ähnliche Notationen für verschiedene Dinge benutzt werden: a_{ij} bezeichnet den Eintrag von A, der in der i-ten Zeile und j-ten Spalte von A steht, die Notation (a_{ij}) bezeichnet die ganze Matrix.
- Die Reihenfolge der Indizierung von Matrixeinträgen a_{ij} können Sie sich vielleicht folgendermaßen merken: „Zeilen zuerst, Spalten später“,
- Eine Matrix, viele Matrizen. „Eine Matritze“ oder „zwei Matrixen“ gibt es nicht.

Definition 4.3.2. *(*Matrixabbildung*):*

Sei $A \in \mathbb{R}^{m\times n}$. Eine Abbildung $f : \mathbb{R}^n \to \mathbb{R}^m$ heißt die zu A gehörige Matrixabbildung und A die zu f gehörige Abbildungsmatrix, wenn f folgendermaßen festgelegt ist:

(i) In der i-ten Spalte der Matrix steht das Bild des i-ten Vektors der Standardbasis, d.h. der Bildvektor $f(e_i) \in \mathbb{R}^m$.

(ii) Ist $x = (x_i) \in \mathbb{R}^n$, so ist $f(x) = \sum_{i=1}^n x_i \cdot f(e_i)$.

Die obige Festlegung von $f(x)$ durch A ist für alle $x \in \mathbb{R}^n$ eindeutig, denn für jeden Vektor $x \in \mathbb{R}^n$ sind die Einträge x_i sind eindeutig, und diese legen den Bildvektor $f(x)$ eindeutig fest (mit Hilfe der Spalten von A, per linearer Fortsetzung). $f(x)$ lässt sich nun leicht mit Hilfe der Abbildungsmatrix per *Matrix-Vektor-Multiplikation* ausrechnen.

Definition/Lemma 4.3.3. *(Matrix-Vektor-Multiplikation)*

Sei $A \in \mathbb{R}^{m\times n}(a_{ij}) \in \mathbb{R}^{m\times n}$ eine Matrix mit Spalten $s_1, \ldots, s_n \in \mathbb{R}^m$ und Zeilen $z_1, \ldots, z_m \in \mathbb{R}^n$. Dann definieren wir das Matrix-Vektor-Produkt von A mit x als den Vektor $A\cdot x \in \mathbb{R}^m$ mit

$$A \cdot x \;:=\; x_1 \cdot s_1 + \ldots + x_n \cdot s_n \;=\; \begin{pmatrix} \sum_{j=1}^n a_{1j}x_j \\ \sum_{j=1}^n a_{2j}x_j \\ \vdots \\ \sum_{j=1}^n a_{mj}x_j \end{pmatrix} = \begin{pmatrix} \langle z_1, x\rangle \\ \langle z_2, x\rangle \\ \vdots \\ \langle z_m, x\rangle \end{pmatrix}.$$

Hierbei haben wir für eine kompakte Notation das euklische Skalarprodukt

$$\langle x, y\rangle = \sum_{i=1}^n x_i \cdot y_i$$

zweier Vektoren x, y im $\mathbb{R}^n$ benutzt.

$A \cdot x$ definiert so das Bild von x unter der zugehörigen Matrixabbildung f. Es gilt also $f(x) = A \cdot x$, „$f(x)$ lässt sich durch Multiplikation von A mit dem Vektor x berechnen“. Dafür haben wir hier zwei Möglichkeiten angegeben, die das selbe Ergebnis liefern:

- Möglichkeit 1 (Aufsummieren der Spalten mit Vorfaktoren,
- Möglichkeit 2 (komponentenweise, „Zeile mal Spalte“, genauer

Möglichkeit 1 ist direkt die Definition von f als Matrixabbildung. Dass Möglichkeit 2 das selbe Ergebnis liefert, folgt durch komponentenweises Ausschreiben der Summe in Möglichkeit 1, vgl. Vorlesung.

Beispiele:

- Es ist

  $$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}, \quad B = \begin{pmatrix} 6 & 0 \\ 1 & 2 \\ 0 & 3 \end{pmatrix}.$$

  Dann gilt

  $$B \cdot A = \begin{pmatrix} 6 & 12 & 18 \\ 9 & 12 & 15 \\ 12 & 15 & 18 \end{pmatrix} \quad \text{und} \quad A \cdot B = \begin{pmatrix} 8 & 13 \\ 29 & 28 \end{pmatrix}.$$

- Die identische Abbildung $\mathrm{id}_{\mathbb{R}^n}$ ist eine Matrixabbildung. Die zugehörige Abbildungsmatrix $I_n \in \mathbb{R}^{n \times n}$ ist die so genannte <u>Einheitsmatrix</u>

  $$I_n = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}. \tag{4.2}$$

  – in der i-ten Spalte steht das Bild des Einheitsvektors e_i, also e_i selbst.

4.4 Eigenschaften von Matrixabbildungen

Lemma 4.4.1. *(Linearität von Matrixabbildungen)*
Sei $A \in \mathbb{R}^{m\times n}$ eine Matrix und f die zugehörige Matrixabbildung. Dann gilt für alle Vektoren $x, y \in \mathbb{R}^n$ und alle Skalare $a \in \mathbb{R}$

$$A \cdot (x + y) = A \cdot x + A \cdot y, \qquad A \cdot (a \cdot x) = a \cdot (A \cdot x)$$

und deshalb auch

$$f(x + y) = f(x) + f(y), \quad f(a \cdot x) = a \cdot f(x). \tag{4.3}$$

Bevor wir das beweisen – Folgendes:
Bemerkung und Ausblick: Lineare Funktionen, Vektorräume & Co.
Funktionen f, die die letzten beiden Bedingungen aus Lemma 4.4.1,

$$f(x + y) = f(x) + f(y), \quad f(a \cdot x) = a \cdot f(x), \tag{4.4}$$

für alle x, y aus einem Vektorraum und für alle zugelassenen Skalare a erfüllen, heißen linear und werden in dieser Vorlesung eine große Rolle spielen. Die in diesem Kapitel eingeführten Matrixabbildungen sind lineare Funktionen zwischen den Vektorräumen $\mathbb{R}^n$ nach $\mathbb{R}^m$, und wir werden viele Fragestellungen des vergangenen Kapitels II zur Algebra (Wann existieren Umkehrabbildungen?, Wie berechnet man Umkehrabbildungen? Was hat das mit der Lösbarkeit von Gleichungen zu tun?...) im allgemeineren Rahmen der Theorie für lineare Abbildungen zwischen Vektorräumen klären (s. Kapitel 9). Dabei werden auch die in diesem Kapitel entwickelten Zusammenhänge zwischen Matrixabbildungen und Abbildungsmatrizen werden auch im größeren Rahmen eine Rolle spielen.

Beweis: Der Eintrag in der i-ten Zeile des Vektors $A \cdot (x + y)$ ist gegeben durch

$$\sum_{j=1}^{n} a_{i,j} \cdot (x_j + y_j) = \sum_{j=1}^{n} a_{i,j} \cdot x_j + \sum_{j=1}^{n} a_{i,j} \cdot y_j.$$

Auf der rechten Seite erkennt man die Summe des i-ten Eintrags von $A \cdot x$ und des i-ten Eintrags von $A \cdot y$. $A(a \cdot x) = a \cdot (A \cdot x)$ zeigt man analog. Die Behauptung für f ergibt sich daraus durch

$$f(x+y) := A\cdot(x+y) = A\cdot x + A\cdot y =: f(x)+f(y)$$

und analog für $f(a\cdot x)$.

□

Satz 4.4.2. *(Hintereinanderausführung von Matrixabbildungen)*

Seien $f:\mathbb{R}^n\to\mathbb{R}^m, x\mapsto A\cdot x$ und $g:\mathbb{R}^m\to\mathbb{R}^k, x\mapsto B\cdot x$ Matrixabbildungen mit Abbildungsmatrizen

$$A=\begin{pmatrix} a_{11} & \dots & a_{1n}\\ \vdots & & \vdots\\ a_{m1} & \dots & a_{mn}\end{pmatrix} = \begin{pmatrix} | & | & \cdots & | \\ s_1 & s_2 & \cdots & s_n \\ | & | & \cdots & | \end{pmatrix}, \quad B=\begin{pmatrix} b_{11} & \dots & b_{1m}\\ \vdots & & \vdots\\ b_{k1} & \dots & b_{km}\end{pmatrix} = \begin{pmatrix} \text{---} & z_1 & \text{---}\\ \text{---} & z_2 & \text{---}\\ \vdots & \vdots & \vdots \\ \text{---} & z_k & \text{---}\end{pmatrix}.$$

Die Hintereinanderausführung

$$g\circ f:\mathbb{R}^n\to\mathbb{R}^k,\quad x\mapsto B\cdot(A\cdot x)$$

ist dann wieder eine Matrixabbildung. Definieren wir das Matrixprodukt $B\cdot A$ der zwei Matrizen B und A als die Matrix $C\in\mathbb{R}^{k\times n}$ mit den Spalten $B\cdot s_1, B\cdot s_2,\dots,B\cdot s_n$, so ist $B\cdot A$ die Abbildungsmatrix von $g\circ f$.

Die einzelnen Einträge c_{ij} der Matrix $(c_{ij})_{i\in\underline{k}, j\in\underline{n}} := B\cdot A = \in\mathbb{R}^{k\times n}$ lassen sich alternativ eintragsweise folgendermaßen berechnen: Für $i\in\underline{k}, j\in\underline{n}$ ist

$$c_{ij} = \sum_{\ell=1}^{m} b_{i\ell}a_{\ell j} = \langle z_i, s_j\rangle. \tag{4.5}$$

(„i-te Zeile von B mal j-te Spalte von A“, wobei die Sprechweise „mal“ bedeutet, dass das am Ende von 4.3.3 eingeführte euklidische Skalarprodukt $\langle\cdot,\cdot\rangle$ zweier Vektoren aus dem $\mathbb{R}^m$ benutzt wird.)

Beweis: (nur Skizze) Mit Hilfe von Lemma 4.4.1, angewandt auf f undf g, weist man nach, dass auch $(g\circ f)(x)$ eine Matrixabbildung im Sinne von 4.3.2(ii) ist. Dann stehen in der Abbildungsmatrix von $g\circ f$ die Vektoren $g(f(e_i))$. Wegen $f(e_i)=s_i$ ist $g(f(e_i)) = B\cdot s_i$ und die Matrix mit diesen Spalten also die zugehörige Abbildungsmatrix.

□

Bemerkungen dazu:

- Vektoren im $\mathbb{R}^n$ werden standardmäßig als Spaltenvektoren geschrieben. Das hat den Vorteil, dass das Matrix-Vektor-Produkt aus 4.3.3 für $A \in \mathbb{R}^{m\times n}$ und $x \in \mathbb{R}^n$ sich einfach als Matrix-Matrix-Produkt für $A \in \mathbb{R}^{m\times n}$ und $x \in \mathbb{R}^{n\times 1}$ deuten lässt. Beweisen wir also Aussagen über Matrix-Matrix-Produkte, ist also stets die Matrix-Vektor-Multiplikation als Spezialfall eingeschlossen.

- Die Multiplikation zweier Matrizen ist (auch für den Fall, dass $n = m = k$ ist) im Allgemeinen nicht kommutativ. Zum Beispiel ist

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 2 \\ 7 & 4 \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 4 & 7 \end{pmatrix}.$$

  Dass sich hier verschiedene Matrixprodukte ergeben, macht hoffentlich auch in dem Lichte Sinn, dass wir ja die Multiplikation von Matrizen so definiert hatten, dass wir die Abbildungsmatrix der Hintereinanderausführungen der zugehörigen Matrixabbildungen erhalten, und die Hintereinanderausführung von Abbildungen ja ebenfalls nicht kommutativ ist. Es ist lehrreich, sich z.B. für die Hintereinananderausführung einer Spiegelung einer Drehung im $\mathbb{R}^2$ klarzumachen, dass die Reihenfolge diesr Aktionen eine Rolle spielt.

4.4.1 Inverse Matrizen und Umkehrungen von Matrixabbildungen

Definition 4.4.3. *(Invertierbarkeit von Matrizen, inverse Matrix)*

Sei $A \in \mathbb{R}^{m\times n}$ eine Matrix. $I_n \in \mathbb{R}^{n\times n}$ bezeichne die in 4.2 eingeführte Einheitsmatrix.

(i) Eine Matrix $B \in \mathbb{R}^{n\times m}$ heißt linksinverse Matrix von A, falls $B \cdot A = I_n$.

(ii) Eine Matrix $B \in \mathbb{R}^{n\times m}$ heißt rechtsinverse Matrix von A, falls $A \cdot B = I_m$.

(iii) A heißt invertierbar, falls eine Matrix $B \in \mathbb{R}^{n\times m}$ existiert, die gleichzeitig Linksinverse und Rechtsinverse von A ist, die also

$$A \cdot B = I_m \quad \text{und} \quad B \cdot A = I_n$$

erfüllt. B nennt man eine inverse Matrix oder einfach Inverse von A.

Linksinverse und rechtsinverse Matrizen sind nicht zwingend eindeutig, z.B. sind mit

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 1 & 2 & 1 \end{pmatrix}, B_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ -1 & 1 \end{pmatrix} \quad \text{und} \quad B_2 = \begin{pmatrix} 2 & -2 \\ -\frac{1}{2} & 1 \\ -1 & 1 \end{pmatrix}$$

B_1 und B_2 zwei verschiedene rechtsinverse Matrizen von A (und von links ist A nicht invertierbar). Wir werden aber sehen: Ist A invertierbar, so gibt es genau eine Inverse („die Inverse von A"), die gleichzeitig Linksinverse und die Rechtsinverse von A ist.

Erinnert Sie das an etwas? Sie stellen vielleicht fest, dass sich die Zusammenhänge zur Umkehrbarkeit von Funktionen und zur Invertierung von Matrizen strukturell ähnliche Aussagen gelten (s. Lemma 2.5.3 und Lemma 2.5.5). Es ist also so, dass für verschiedene mathematische Objekte und zugehörige Operationen mit ihnen (die Hintereinanderausführung von Funktionen ja etwas anderes als die rechnerische Multiplikation zweier Matrizen) sich *strukturell gleich verhalten*: Betrachten wir die zugehörigen Matrixabbildungen

$$f : \mathbb{R}^3 \to \mathbb{R}^2, x \mapsto \begin{pmatrix} 1 & 2 & 0 \\ 1 & 2 & 1 \end{pmatrix} \cdot x,$$

$$g_1 : \mathbb{R}^2 \to \mathbb{R}^3, x \mapsto \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ -1 & 1 \end{pmatrix} \cdot x \quad \text{und} \quad g_2 : \mathbb{R}^2 \to \mathbb{R}^3, x \mapsto \begin{pmatrix} 2 & -2 \\ -\frac{1}{2} & 1 \\ -1 & 1 \end{pmatrix} \cdot x.$$

die Funktionen g_1 und g_2 beide Rechtsinverse von f, d.h. es gilt $f \circ g_1 = \mathrm{id}_{\mathbb{R}^2}$ und $f \circ g_2 = \mathrm{id}_{\mathbb{R}^2}$.

Solche Strukturgleichheiten zu Tage zu fördern und auszunutzen, ist eines der Hauptanliegen der modernen „abstrakten" Algebra, um deren Darstellung es schwerpunktmäßig im nächsten Kapitel gehen soll.

Den Zusammenhang zwischen Umkehrabbildungen von Matrixabbildungen und inversen Matrizen formulieren wir in dem folgenden Lemma.

Lemma 4.4.4. *(Umkehrabbildung von Matrixabbildungen und inverse Matrizen)*
Sei $A \in \mathbb{R}^{m\times n}$ eine Matrix und $f : \mathbb{R}^n \to \mathbb{R}^m, x \mapsto A \cdot x$ die zugehörige Matrixabbildung. Dann sind äquivalent:

(i) A ist invertierbar mit Inverser $B \in \mathbb{R}^{n\times m}$,

(ii) f ist umkehrbar mit Matrixabbildung

$$f^{-1} : \mathbb{R}^m \to \mathbb{R}^n, x \mapsto B \cdot x$$

als Umkehrfunktion von f,

(iii) f ist umkehrbar.

Beweis: Wir zeigen die Implikationen (i) $\Rightarrow$ (ii), (ii) $\Rightarrow$ (iii), (iii) $\Rightarrow$ (ii) und (ii) $\Rightarrow$ (i), damit ist alles gezeigt. Die Implikation (i) $\Rightarrow$ (iii) folgt nach Satz 4.4.2: Es gilt

$$\mathrm{id}_{\mathbb{R}^m} = f_{I_m} = f_{A\cdot B} = f_A \circ f_B$$

und

$$\mathrm{id}_{\mathbb{R}^n} = f_{I_n} = f_{B\cdot A} = f_B \circ f_A,$$

also ist f_B Links- und Rechtsinverse zu f. (ii) $\Rightarrow$ (iii) ist klar. Zeigen wir (iii) $\Rightarrow$ (ii). Wir nehmen dazu an, f besitzt eine Umkehrfunktion f^{-1} und müssen nun zeigen, dass f^{-1} eine Matrixabbildung ist. Ist das der Fall, so müssen nach Definition von Matrixabbildungen in den Spalten der Abbildungsmatrix von f^{-1} auf jeden Fall die Bilder von $e_1, \dots, e_n$ unter f^{-1} stehen. Wir definieren die Matrix

$$B = \begin{pmatrix} | & | & \cdots & | \\ f^{-1}(e_1) & f^{-1}(e_2) & \cdots & f^{-1}(e_n) \\ | & | & \cdots & | \end{pmatrix}.$$

Wenn f^{-1} eine Matrixabbildung ist, muss B die Abbildungsmatrix von f^{-1} sein. Wir zeigen, dass die Matrixabbildung $g : \mathbb{R}^m \to \mathbb{R}^n, x \mapsto B\cdot x$ eine Rechtsinverse der Abbildung f ist. Mit Lemma 2.5.5(ii) ist g dann die eindeutige Umkehrabbildung von f, also $g = f^{-1}$ eine Matrixabbildung. Sei dazu $x = (x_i)_{i\in \underline{m}} \in \mathbb{R}^m$, wir zeigen $f(g(x)) = x$: Es gilt

$$f(g(x)) = f(B \cdot x) = f(\sum_{i=1}^{n} x_i f^{-1}(e_i)) \overset{\text{Lemma 4.4.1}}{=} \sum_{i=1}^{n} x_i \cdot f(f^{-1}(e_i)) = \sum_{i=1}^{n} x_i \cdot e_i = x.$$

Zuletzt müssen wir noch (ii) $\Rightarrow$ (i) begründen: Ist f^{-1} eine Matrixabbildung mit zugehöriger Matrix B, so gilt für alle $x \in \mathbb{R}^n$

$$(B \cdot A) \cdot x = B \cdot (A \cdot x) = f^{-1}(f(x)) = x,$$

also insbesondere $(B \cdot A) \cdot e_i = e_i$. Damit folgt nach Definition der Matrixspalten von $B \cdot A$ als $(B \cdot A) \cdot e_i$ nun $B \cdot A = I_n$. Das selbe Argument zeigt $A \cdot B = I_m$, also ist B die inverse Matrix zu A.

$\square$

Wir wollen nun die dargestellten Zusammenhänge der Vektorräume $\mathbb{R}^n$ und der linearen (=Matrix)Abbildungen zwischen in einem allgemeineren Rahmen formulieren. Dazu werden wir einen Schritt vollführen, der sich Anfang des 20. Jahrhunderts als Grundmuster der Mathematik durchgesetzt hat: Wir werden bei der Untersuchung mathematischer Zusammenhänge eher auf die strukturellen Eigenschaften von Mengen mathematischer Gegenstände (wie Verknüpfbarkeit oder Umkehrbarkeit) als auf die Gegenstände an sich schauen. Das soll im kommenden Kapitel zur *abstrakten Algebra* geschehen. Aber keine Angst, abstrakt ist nicht schlecht, es macht die Dinge (im Gegensatz zur weit verbreiteten Meinung zur Mathematik) oftmals leichter und übersichtlicher als vorher, weil man sich auf das Wesentliche konzentrieren kann.

4.5 Aufgaben zu Kapitel 4

Aufgabe 4.14. (Linearkombination und Koordinatendarstellung, Komplanarität):

Gegeben sind im $\mathbb{R}^3$ die Vektoren

$$x = \begin{pmatrix} 3 \\ 4 \\ 7 \end{pmatrix}, \quad v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

(a) Schreiben Sie x als Linearkombination von v_1, v_2 und v_3, d.h. finden Sie reelle Zahlen a_1, a_2, a_3, so dass $x = \sum_{i=1}^{3} a_i \cdot v_i$ ist.

(b) Schreiben Sie einen allgemeinen Vektor $y = (y_1, y_2, y_3)$ des $\mathbb{R}^3$ als Linearkombination von v_1, v_2 und v_3, d.h. finden Sie reelle Zahlen a_1, a_2, a_3, so dass $y = \sum_{i=1}^{3} a_i \cdot v_i$ ist. (Hier müssen Sie a_1, a_2, a_3 in Abhängigkeit von y_1, y_2, y_3 angeben.)

(c) Sind die folgenden Vektoren w_1, w_2, w_3 komplanar?

$$w_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad w_2 = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}, \quad w_3 = \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix}.$$

Aufgabe 4.15. (Lesen von Abbildungsmatrizen):

Seien $f_1, f_2 \colon \mathbb{R}^2 \to \mathbb{R}^2$ und $g_1, g_2 \colon \mathbb{R}^3 \to \mathbb{R}^3$ Matrixabbildungen mit

$$f_1 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix}, \qquad f_2 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix},$$

$$g_1 \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \qquad g_2 \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Geben Sie eine geometrische Deutung der angegebenen Abbildungen in der Ebene bzw. im Raum. Mit anderen Worten: Was bewirken diese?

Aufgabe 4.16. (Abbildungen in der Ebene):

(a) Stellen Sie das Rechteck $PQRS$ mit den Eckpunkten $P(2|1), Q(5|1), R(5|3)$ und $S = (2|3)$ in einem kartesischen Koordinatensystem dar. Geben Sie dann jeweils graphische Veranschaulichungen der folgenden Abbildungen, indem Sie jeweils das Parallelogramm einzeichnen, dessen Eckpunktkoordinaten sich ergeben, wenn man die Abbildungen auf die Koordinaten der Eckpunkte von $PQRS$ anwendet.

(i) $f : \mathbb{R}^2 \to \mathbb{R}^2$ mit $f : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 2x \\ 2y \end{pmatrix} - \begin{pmatrix} 4 \\ 4 \end{pmatrix}$,

(ii) $g : \mathbb{R}^2 \to \mathbb{R}^2$ mit $f : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ y - \frac{3}{2}x \end{pmatrix}$,

(iii) $h : \mathbb{R}^2 \to \mathbb{R}^2$ mit $f : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x - 3y \\ y - x \end{pmatrix}$,

(iv) Drehung in der Ebene um den Koordinatenursprung mit dem Drehwinkel $\frac{\pi}{6} \mathrel{\widehat{=}} 30°$ (gegen den Uhrzeigersinn).

(b) Drehungen um den Koordinatenursprung in der Ebene lassen sich als Matrixabbildungen für ihre Koordinaten darstellen. Bestimmen Sie für eine Drehung d_α um einen Winkel α in der Ebene (gegen den Uhrzeigersinn) die Bildvektoren $d_\alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ und $d_\alpha \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ und geben Sie damit die Abbildungsmatrix von d_α an.

(c) Eine der Abbildungen aus (a) ist keine Matrixabbildung. Geben Sie an welche, und begründen Sie.

(d) Geben Sie zu den drei Abbildungen aus (a), die sich als Matrixabbildung schreiben lassen, jeweils die zugehörige Abbildungsmatrix an.

Aufgabe 4.17. (Matrixabbildungen mit gegebenen Bildvektoren):

In der Vorlesung haben wir gesehen, dass eine Matrixabbildung f vollständig durch die Bilder der Standardbasisvektoren $e_1, \ldots, e_n$ festgelegt wird. In dieser Aufgabe sollen Sie Abbildungen untersuchen, in denen die Bilder anderer Vektoren aus dem Definitionsbereich von f angegeben sind.

(a) Bestimmen Sie die Abbildungsmatrix A_f einer Matrixabbildung $f : \mathbb{R}^3 \to \mathbb{R}^3$ mit

$$f\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad f\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix}, \quad f\begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix}.$$

(Begründen Sie Ihre Antwort sinnvoll.)

(b) Eine Matrixabbildung $g : \mathbb{R}^3 \to \mathbb{R}^3$ ist durch Angabe der Bilder zweier Vektoren nicht eindeutig festgelegt. Geben Sie zwei verschiedene Matrixabbildungen $g_1, g_2 : \mathbb{R}^3 \to \mathbb{R}^3$ an, für die

$$g_1\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = g_2\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \\ -2 \end{pmatrix} \quad \text{und} \quad g_1\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = g_2\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$$

ist. (Stellen Sie Ihren Ansatz dabei sinnvoll dar.)

(c) Begründen Sie, dass es keine Matrixabbildung $h : \mathbb{R}^3 \to \mathbb{R}^3$ geben kann, für die

$$h\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ 0 \\ 1 \end{pmatrix}, \quad h\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ -1 \\ 0 \end{pmatrix} \quad \text{und} \quad h\begin{pmatrix} 0 \\ 3 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

ist.

Aufgabe 4.18. (Besondere Matrizen und Matrixmultiplikationen):

(a) Der *Nullvektor im* $\mathbb{R}^n$ ist der Vektor $0_{\mathbb{R}^n} = (\underbrace{0, 0, \ldots, 0}_{n \text{ Einträge}})$. Zeigen Sie, dass die *Nullabbildung*

$$\mathcal{O}_{\mathbb{R}^3} : \mathbb{R}^3 \to \mathbb{R}^3, \quad x \mapsto 0_{\mathbb{R}^3}$$

auf $\mathbb{R}^3$ eine Matrixabbildung ist, d.h. geben Sie eine Abbildungsmatrix A an, für die Sie begründen, dass $\mathcal{O}_{\mathbb{R}^3}(x) = A \cdot x$ ist.

(b) Zeigen Sie, dass die *identische Abbildung im* $\mathbb{R}^4$,

$$\mathrm{id}_{\mathbb{R}^4} : \mathbb{R}^4 \to \mathbb{R}^4, \quad x \mapsto x$$

eine Matrixabbildung ist, d.h. geben Sie eine Abbildungsmatrix I an, für die Sie begründen, dass $\mathrm{id}_{\mathbb{R}^4}(x) = I \cdot x$ ist.

(c) Sei

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

eine beliebige 3×3-Matrix. Berechnen Sie $A \cdot 0_{\mathbb{R}^3}$ und

$$A \cdot \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + A \cdot \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} + A \cdot \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

Aufgabe 4.19. (Hintereinanderausführung von Matrixabbildungen):

(a) Wenn man in der Ebene zweimal nacheinander um 180° um den Koordinatenursprung dreht, dann erwartet man, dass das Ergebnis so ist, als hätte man gar keine Bewegung durchgeführt. Ebenso wird man erwarten können, dass eine zweimalige Spiegelung an der y-Achse auch keinen Effekt hat. Überprüfen Sie beide Behauptungen anhand der entsprechenden Matrixabbildungen in $\mathbb{R}^2$, d.h. stellen Sie also zunächst Matrizen für die Drehung bzw. die Spiegelung auf und berechnen Sie dann die entsprechenden Produkte.

(b) Sei f_α eine Drehung um $\alpha = 30°$ und f_β eine Drehung um $\beta = 20°$, jeweils um den Koordinatenursprung. Es ist anschaulich klar, dass sowohl $f_\alpha \circ f_\beta$ als auch $f_\beta \circ f_\alpha$ einer Drehung um 50° entsprechen müssen. Zeigen Sie dies durch Multiplikation der entsprechenden Matrizen (und damit auch, dass die Komposition in diesem speziellen Fall kommutativ ist).

(c) Gegeben sind die Abbildungen

$$g_1 : \mathbb{R}^2 \to \mathbb{R}^2, \quad \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \qquad g_2 : \mathbb{R}^2 \to \mathbb{R}^2, \quad \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 \\ b & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

mit $a, b \neq 0$, g_1 ist also eine Scherung in Richtung x-Achse und g_2 eine Scherung in Richtung y-Achse. Wenn man die beiden hintereinander ausführt, kommt es dann auf die Reihenfolge an?

(d) Macht es einen Unterschied, ob man in der Ebene erst an der x-Achse und dann an der y-Achse spiegelt oder ob man es umgekehrt macht?

Aufgabe 4.20. (Injektive/surjektive/bijektive Matrixabbildungen):

(a) Geben Sie die Matrix A_f einer Matrixabbildung $f : \mathbb{R}^2 \to \mathbb{R}^2$ an, die *nicht* injektiv ist, und begründen Sie Ihre Behauptung!

(b) Geben Sie die Matrix A_f einer Matrixabbildung $f : \mathbb{R}^2 \to \mathbb{R}^2$ an, die *nicht* surjektiv ist, und begründen Sie Ihre Behauptung!

(c) Geben Sie die Matrix A_f einer Matrixabbildung $f : \mathbb{R}^2 \to \mathbb{R}^2$ an, die bijektiv ist, und begründen Sie Ihre Behauptung!

Aufgabe 4.21. (Matrixprodukte):

(a) Gegeben sind die folgenden Matrizen:

$$A = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}, \qquad C = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \\ \frac{1}{4} & 0 \end{pmatrix}.$$

Welche Produkte von zwei verschiedenen dieser drei Matrizen sind definiert? Berechnen Sie diese.

(b) Wir betrachten die folgenden beiden Abbildungen:

$$f : \mathbb{R}^3 \to \mathbb{R}^2, \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} 2x - 3y \\ z - x \end{pmatrix}, \qquad g : \mathbb{R}^3 \to \mathbb{R}^3, \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} x + 2z \\ 5y - z \\ x - y + z \end{pmatrix}$$

Begründen Sie (z.B. durch Zitieren eines geeigneten Satzes aus der Vorlesung), dass $f \circ g$ eine Matrixabbildung ist, und geben Sie die zu $f \circ g$ gehörende Matrix an.

Aufgabe 4.22. (Nur Rechtsinverse, aber davon mehrere):

Wir betrachten die Matrixabbildungen

$$f : \mathbb{R}^3 \to \mathbb{R}^2,\ x \mapsto \begin{pmatrix} 1 & 2 & 0 \\ 1 & 2 & 1 \end{pmatrix} \cdot x, \quad g_1 : \mathbb{R}^2 \to \mathbb{R}^3,\ x \mapsto \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ -1 & 1 \end{pmatrix} \cdot x$$

und

$$g_2 : \mathbb{R}^2 \to \mathbb{R}^3,\ x \mapsto \begin{pmatrix} 2 & -2 \\ -\frac{1}{2} & 1 \\ -1 & 1 \end{pmatrix} \cdot x.$$

(a) Zeigen Sie, dass f sowohl durch g_1 als auch durch g_2 von rechts umkehrbar ist (vgl. Satz 2.5.3).

(b) Begründen Sie, dass f surjektiv ist.

(c) Begründen Sie, dass f nicht von links umkehrbar ist, dass es also keine Funktion $h : \mathbb{R}^2 \to \mathbb{R}^3$ geben kann, für die $h \circ f = \mathrm{id}_{\mathbb{R}^3}$ gilt.

5 Abstrakte Algebra: Algebraische Strukturen

5.1 Einführendes und Beispiele

- Algebra ursprünglich: „Wissenschaft vom Gleichungslösen"
- Evariste Galois („Galoistheorie", 1811-1832(!)) als einer der ersten, der strukturelle Eigenschaften anderer Mengen erkennt und ausnutzt.
- Die drei „gängigen" Gleichungsumformungen zur Lösung von

  (i) $x + 4 = 3$,

  (ii) $5 \cdot x = 3$

  (iii) $x + 3 \cdot y - 4 = 2 \cdot y + 1$

  zeigen einige wichtige, „wünschenswerte" Eigenschaften der betrachteten algebraischen Strukturen $(\mathbb{R}, +)$ und $(\mathbb{R}, \cdot)$, die beim Lösen von Gleichungen eine Rolle spielen: Assoziativgesetz, Existenz von neutralem Element und inversen Elementen, Kommutativgesetz, bei Benutzung von zwei Verknüpfungen: Distributivgesetz.
- Analyse der algebraischen Struktur von (Zahlen-)mengen und anderen Mengen (Abbildungen, Matrizen,...), (Rechen-)operationen auf diesen Mengen (Addition und Multiplikation von Zahlen, Hintereinanderausführung von Abildungen, Matrixmultiplikation, zeigen: Diese Gesetzmäßigkeiten gelten auch in anderen Strukturen und ermöglichen es, dort analoge Aussagen zu treffen. $\rightsquigarrow$ Algebra als Strukturwissenschaft.
- Gauß benutzt in Abhandlungen über Matrixabbildungen der Ebene und des Raumes implizit, dass für (invertierbare) Matrizen analoge Rechengesetze wie in Zahlbereichen gelten.
- Dass in verschiedensten mathematischen Strukturen die gleichen Gesetze gelten, kristallisiert sich Mitte bis Ende des 19. Jahrhunderts heraus. Leopold Kronecker schreibt darüber, dass „Die überaus einfachen Prinzipien, auf denen die Gauß'sche Methode (die des letzten Spiegelstrichs, Bem. TR) beruht, finden nicht blos an der bezeichneten Stelle, sondern vielfach und zwar schon in den elementarsten Theilen der Zahlentheorie Anwendung. Dieser Umstand deutet darauf hin, und es ist leicht, sich davon zu überzeugen, dass die erwähnten Prinzipien einer allgemeineren, abstrakteren Ideensphäre angehören. Deshalb scheint es angemessen, die Entwicklung derselben von allen unwesentlichen Beschränkungen zu befreien, sodass man alsdann einer Wiederholung derselben Schlussweisen in den verschiedenen Fällen des Gebrauchs überhoben wird. Dieser Vortheil kommt sogar schon bei der Entwicklung

selbst zur Geltung und die Darstellung gewinnt dadurch, wenn sie in der zulässig allgemeinsten Weise gegeben wird, zugleich an Einfachheit und durch das deutliche Vortreten des allein Wesentlichen auch an Übersichtlichkeit.“ (Zitat aus [Wuß], S. 207/208).

- Ab hier beginnt die Ausformung abstrakter, axiomatisch festgelegter Begriffe für Strukturen (hier: „Gruppe“), die die wesentlichen Zusammenhänge einer Klasse von Strukturen herausstellen. Damit einher geht eine Ablösung von der Anschauung, z.B. exemplarisch in der Geometrie (Anfang 20. Jhd.: David Hilbert, Göttingen, „Bierseidelzitat“)
- Gewöhnungsbedürftig, aber weitreichend: Wir führen abstrakte Beweise, die nur auf den Eigenschaften der jeweiligen Struktur (z.B. einer Gruppe) basieren – so bewiesene Zusammenhänge gelten dann in jeder Gruppe.

Gleichungslösen

Al-Chwarizmi (um 800 n.Chr.)

Leonhard Euler (1707-1783)

Evariste Galois (1811-1832(!))

Emmy Noether (Anf. 20. Jhd.)

Einige Beispiele für abstraktere, ungewohntere mathematische Strukturen, in denen aus den Zalhbereichen bekannte Rechengesetze ihre Gültigkeit behalten, kennen wir schon. Wir schieben hier gleich zwei Beispiele aus unserem Stoff der letzten Kapitel hinterher, das Assoziativgesetz für die Hintereinanderausführung von Funktionen und das für die Multiplikation von Matrizen:

Lemma 5.1.1. *(Assoziativität von Funktionsverkettungen)*

Für alle Funktionen $f : X \to Y, g : Y \to Z, h : Z \to W$ (X, Y, Z, W Mengen) gilt bezüglich der Hintereinanderausführung das Assoziativgesetz

$$f \circ (g \circ h) = (f \circ g) \circ h.$$

Das bedeutet insbesondere, dass man, wie man das von Zahlen gewohnt ist, die Klammern in diesem Ausdruck weglassen kann.

Beweis: Das rechnet man werteweise nach: Es ist für alle $x \in X$

$$((f \circ g) \circ h)(x) \overset{\text{Def. von } \circ h}{:=} (f \circ g)(h(x)) \overset{\text{Def. von } f \circ g}{:=} f(g(h(x)))$$

und

$$((f \circ (g \circ h))(x) \overset{\text{Def. von } f \circ}{:=} (f(g \circ h)(x)) \overset{\text{Def. von } g \circ h}{:=} f(g(h(x))),$$

also gilt (mit der Bemerkung am Anfang dieses Abschnitts) $f \circ (g \circ h) = (f \circ g) \circ h$. □

Da wir die Matrixmultiplikation über die Hintereinanderausführung der entsprechenden Matrixabbildungen definiert haben, ist auch die Multiplikation von Matrizen assoziativ.

Lemma 5.1.2. *(Assoziativität der Matrixmultiplikation)*

Seien $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{k \times m}, C \in \mathbb{R}^{\ell \times k}$ Matrizen (mit „kompatiblen“ Zeilen- und Spaltenzahlen). Dann ist

$$C \cdot (B \cdot A) = (C \cdot B) \cdot A.$$

Auch hier können Klammern also weggelassen werden, und wir schreiben für die obige Operation oft einfach $C \cdot B \cdot A$.

Beweis: Wir argumentieren mit Hilfe der Eigenschaften der zugehörigen Matrix*abbildung*. Für diese gilt nach 4.4.2 und wegen des gerde bewiesenen Assoziativgesetzes 5.1.1 für Funktionen, angewandt auf die zugehörigen Matrixabbildungen:

$$f_{C \cdot (B \cdot A)} := f_C \circ f_{B \cdot A} := f_C \circ (f_B \circ f_A) \overset{\text{Ass.}}{=} (f_C \circ f_B) \circ f_A =: f_{C \cdot B} \circ f_A =: f_{(C \cdot B) \cdot A}.$$

Die Spalten der Matrix $C \cdot (B \cdot A)$ sind die Funktionswerte $f_{C\cdot(B\cdot A)}(e_i)$, die der Matrix $(C \cdot B) \cdot A$ sind die Funktionswerte $f_{(C\cdot B)\cdot A}(e_i)$. Da die diese Funktionen gleich sind, sind auch die Matrizen spaltenweise identisch, also insgesamt gleich.

Alternativ kann man die Einträge der Matrizen links und rechts mit Hilfe der Definition (4.5) auch eintragsweise ausrechnen und so zeigen, dass die Matrizen $C\cdot(B\cdot A) = (C\cdot B)\cdot A$ eintragsweise gleich sind.

□

5.2 Algebraische Strukturen – Grundbegriffe bis zum Gruppenbegriff

Es ist an der Zeit, die Begrifflichkeiten zu definieren, die die von Kronecker oben skizzierte abstrakte „Ideensphäre“ formalisieren.

Definition 5.2.1. *(Algebraische Struktur)*
Sei M eine nicht leere Menge. Dann heißt $*$ eine Verknüpfung auf M, wenn

$$* : (a, b) \mapsto a * b$$

eine Funktion mit Wertebereich M ist, d.h. für alle $a, b \in M$ ist $a*b$ definiert und ein Element aus M. Man spricht auch von „Abgeschlossenheit von M gegenüber der Verknüpfung $*$“.

Ein solches Paar $(M, *)$ aus Menge und zugehöriger Verknüpfung bezeichnet man als eine algebraische Struktur.

Einfache (aber sehr wichtige) Beispiele für algebraische Strukturen sind die Zahlbereiche $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$, jeweils mit der Addition als Verknüpfung (also z.B. $(\mathbb{N}, +), (\mathbb{Z}, +), \ldots$), und mit der Multiplikation als Verknüpfung (also $(\mathbb{N}, \cdot), (\mathbb{Z}, \cdot), \ldots$). Kompliziertere Beispiele sind

- der $\mathbb{R}^n$ mit der Vektoraddition $+$;
- die Menge $\mathrm{Abb}(X)$ aller Abbildungen $f : X \to X$ einer Menge X in sich selbst mit der Hintereinanderausführung $\circ$ aus 2.5.1 als Verknüpfung (Hier wird die Gleichheit von Definitions- und Wertbereich gefordert, damit für alle $f, g \in \mathrm{Abb}(X)$ ihre Hintereinanderausführung wieder eine Funkion aus derselben Menge ist.);

- die Menge $\mathbb{R}[x]$ aller reellen Polynome, die sowohl mit der 3.1.2 definierten Addition als auch mit der Multiplikation $\cdot$ jeweils eine algebraische Strukturen $(\mathbb{R}[x], +)$ bzw. $(\mathbb{R}[x], \cdot)$ ist;

- die Menge der quadratischen Matrizen $\mathbb{R}^{n\times n}$ mit der Matrizenmultiplikation $(A, B) \mapsto A \cdot B$ aus 4.4.2 (Hier wird die Gleichheit von Zeilen- und Spaltenanzahl gefordert, damit für alle $A, B \in \mathbb{R}^{n\times n}$ ihre Hintereinanderausführung wieder eine Matrix aus derselben Menge ist

- Bezeichnen wir mit $\mathbb{R}^-$ die negativen reellen Zahlen, so ist $(\mathbb{R}^-, \cdot)$ *keine* algebraische Struktur, da $\cdot$ keine Verknüpfung ist, d.h., dass das Ergebnis einer Multiplikation nicht zwingend wieder in $\mathbb{R}^-$ liegt (z.B. ist $(-3)\cdot(-5) = 15 \notin \mathbb{R}^-$). Man sagt auch, $(\mathbb{R}^-, \cdot)$ ist nicht gegenüber der Verknüpfung $\cdot$ abgeschlossen.

Definition 5.2.2. *(Eigenschaften algebraischer Strukturen)*
Zusätzliche Eigenschaften, die eine algebraische Struktur $(M, *)$ haben kann, sind u.a.:

(i) Gilt für alle $a, b, c \in M$ das Assoziativgesetz

$$(a * b) * c = a * (b * c),$$

so heißt $(M, *)$ eine <u>assoziative</u> algebraische Struktur. In diesem Fall kann die Klammerung bei einem Ausdruck wie $(a * b) * c$ weggelassen werden, man schreibt z.B. $3 + 4 + 5$.

(ii) Wenn für alle $a, b \in M$ das Kommutativgesetz gilt, also immer

$$a * b = b * a$$

ist, heißt $(M, *)$ eine <u>kommutative</u> algebraische Struktur.

Beispiele:

- Die gängigen Zahlbereiche mit $+$ und $\cdot$ sind jeweils assoziative <u>und</u> kommutative algebraische Strukturen. (Vgl. die einführenden Beispiel oben: Man nutzt das beim Rechnen dauernd aus, ohne darüber nachzudenken!).

- Die Hintereinanderausführung von Funktionen $f, g, h \in \mathrm{Abb}(X)$ ist assoziativ nach 5.1.1, also $(\mathrm{Abb}(X), \circ)$ eine assoziative algebraische Struktur

- $(\mathbb{R}^{n\times n}, \cdot)$ bildet (für festes $n \in \mathbb{N}$) eine nach 5.1.2 assoziative algebraische Struktur.
- Aber: $\mathrm{Abb}(X)$ ist nicht-kommutativ, falls X mehr als ein Element enthält; $\mathbb{R}^{n\times n}$ ist nicht kommutativ für $n \geq 2$, s. die Beispiele in Kapitel 2 und 4.

Die Assoziativität kann man auch beim Rechnen mit Funktionen (und der Hintereinanderausführung) und Matrizen (und die Matrixmultiplikation als Verknüpfung) nutzen.

Definition 5.2.3. *(Neutrale Elemente, Monoide)*
Sei $(M, *)$ eine algebraische Struktur. Dann heißt $e \in M$ ein neutrales Element von M, falls für alle $a \in M$

$$a * e = e * a = a$$

ist. Eine assoziative algebraische Struktur, in der ein solches neutrales Element existiert, heißt Monoid.

Beispiele:

- Bezüglich der Addition ist 0 ein neutrales Element in $\mathbb{N}_0, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$.
- Bezüglich der Multiplikation ist 1 ein neutrales Element in $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$.
- Bezüglich der Vektoraddition in $\mathbb{R}^n$ ist der Nullvektor $0_{\mathbb{R}^n} = (0, \ldots, 0)$ ein neutrales Element.
- Bezüglich der Matrixmultiplikation in $\mathbb{R}^{n\times n}$ ist die Einheitsmatrix I_n ein neutrales Element.
- Bezüglich der Hintereinanderausführung von Funktionen aus $\mathrm{Abb}(X)$ ist $\mathrm{id}_X : x \to x$ ein neutrales Element.

Definition 5.2.4. *(Inverse Elemente)*
Sei $(M, *)$ ein Monoid mit neutralem Element e. Ein Element $a \in M$ heißt invertierbar bezüglich $*$, wenn es ein $b \in M$ gibt mit

$$a * b = b * a = e.$$

Solch ein Element $b \in M$ nennt man ein Inverses oder eine Inverse von a in M (bezüglich $*$). Die Menge aller invertierbaren Elemente von $(M, *)$ bezeichnet man mit $M^\times$.

Beispiele:

- Im Gegensatz zu $(\mathbb{N}, +)$ ist in $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$ und $(\mathbb{R}, +)$ *jedes* Element bezüglich der Addition invertierbar durch seine *Gegenzahl* $-a \in \mathbb{Z}$. Neue Perspektive: **Das Addieren von Gegenzahlen $-a$ nennt man normalerweise Subtraktion von a.**

- In $(\mathbb{N}, \cdot)$ besitzt allein die 1 eine Inverse (nämlich sich selbst); in $(\mathbb{Z}, \cdot)$ sind 1 und -1 invertierbar (wiederum mit sich selbst als Inverse).

- In $(\mathbb{Q}, \cdot)$ und $(\mathbb{R}, \cdot)$ ist $\mathbb{Q}^\times = \{a \in \mathbb{Q} \mid a \neq 0\}$, $\mathbb{R}^\times = \{a \in \mathbb{R} \mid a \neq 0\}$, da für alle diese a der Kehrwert $\frac{1}{a}$ das Inverse ist. Neue Perspektive: **Das Multiplizieren mit solchen multiplikativen Inversen $\frac{1}{a}$ nennt man Division durch a.**

- Im Monoid $(\mathbb{R}^{n\times n}, \cdot)$ entspricht die Invertierbarkeit einer Matrix $A \in \mathbb{R}^{n\times n}$ genau unserer Definition 4.4.3. Das heißt insbesondere, dass viele Matrizen *nicht invertierbar* sind. Aus 4.4.4 wissen wir, dass eine Matrix invertierbar ist, genau dann, wenn die zugehörige Matrixabbildung umkehrbar/bijektiv ist; Z.B. liest man für die durch
$$A = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}$$
definierte Matrixabbildung durch Betrachtung der Spalten $A \cdot e_1 = A \cdot e_2$ ab. Die zugehörige Matrixabbildung ist also nicht injektiv und damit nicht bijektiv. $(\mathbb{R}^{n\times n})^\times$ besteht zusammengefasst also genau aus allen Matrizen, die zu bijektiven Matrixabbildungen gehören.

- In $\mathrm{Abb}(X)$ sind die invertierbaren Elemente gerade die umkehrbaren/bijektiven Funktionen $f : X \to X$, die zugehörigen Inversen genau die Umkehrfunktionen, vergleiche die allgemeine Definition und Definition 2.5.4.

Lemma 5.2.5. *Eine algebraische Struktur $(M, *)$ besitzt höchstens ein neutrales Element. Man sagt, das neutrale Element ist eindeutig.*

Beweis: Seien e_1 und e_2 neutrale Elemente einer algebraischen Struktur $(M, *)$. Dann lässt sich (egal, wie M aussieht, nur aufgrund der Eigenschaft, dass e_1 und e_2 neutrale Elemente sind) folgendermaßen rechnen:
$$e_1 \overset{e_2 \text{ neutral}}{=} e_1 * e_2 \overset{e_1 \text{ neutral}}{=} e_2.$$
□

Lemma/Definition 5.2.6. *In einem Monoid $(M, *)$ sind Inverse eindeutig, d.h. ein invertierbares Element $a \in M^\times$ hat* genau ein *Inverses. Wir sprechen daher von* dem *Inversen oder auch von* der *Inversen von a und schreiben dafür a^{-1}.*

Beweis: Sei a ein invertierbares Element eines beliebigen Monoids $(M, *)$ und b, c zwei Inverse von a. Wir zeigen, dass diese immer gleich sein müssen. Unter den gegebenen Voraussetzungen lässt sich auf Grundlage der Assoziativität und der Eigenschaften des Neutralen und der Inversen b, c immer folgendermaßen rechnen:

$$b = b * e = b * (a * c) = (b * a) * c = e * c = c.$$

□

Eines der Hauptresultate über invertierbare Elemente ist, dass mit ihnen formulierte Gleichungen immer eindeutig lösbar sind. Das ist der Inhalt des folgenden Satzes.

Satz 5.2.7. *Sei $(M, *)$ ein Monoid. Dann ist für beliebige $a \in M$ äquivalent:*

(i) a ist invertierbar.

(ii) Die Abbildung

$$\mathcal{L}_a : M \to M, x \mapsto a * x,$$

also die Multiplikation mit a von links, ist bijektiv.

(iii) Gleichungen vom Typ

$$a * x = b$$

(a von oben, $b \in M$ gegeben, x gesucht) haben immer, d.h. egal, für welches $b \in M$, die eindeutige Lösung $x = a^{-1} \circ b \in M$.

Die Aussagen von (ii) und (iii) gelten analog für die Multiplikation von rechts. Ist $(M, *)$ nicht kommutativ, muss diese zwar von der von links unterschieden werden, der Beweis läuft aber analog (Übungsaufgabe).

Der obige Satz lässt sich zum Beispiel anwenden auf

- eine lineare Gleichung $x + a = b$ in $\mathbb{R}$ (mit A beliebig, denn alle $a \in \mathbb{R}$ haben ein additives Inverses); dort ist die eindeutige Lösung $x = b - a$,
- eine multiplikative Gleichung $a \cdot x = b$ in $\mathbb{R}$ – diese Gleichung lässt genau dann eindeutig durch $x = \frac{b}{a}$ lösen, wenn $a \neq 0$ ist;

- alle Vektoren v des $\mathbb{R}^3$ – diese besitzen immer eine additive Inverse $-v$, also ist für alle v, w, x in $\mathbb{R}^3$ die Äquivalenzumformung $v+w = x \iff v = w+(-x)(= w - v)$ erlaubt,
- auf Gleichungen für Matrizen, die sich damit umformen lassen: Die Äquivalenzumformung
$$A \cdot B = C \qquad \iff \qquad B = A^{-1} \cdot C$$
ist genau dann erlaubt, wenn A invertierbar ist.

Beweis: (ii) ⇔ (iii) ist Satz 2.5.7. Bleibt (i) ⇔ (ii), zu zeigen, wir starten mit (i)⇒ (ii): Wenn a invertierbar ist, ist $\mathcal{L}_{a^{-1}}$ Links- und Rechtsinverse der Abbildung $\mathcal{L}_a$, also $\mathcal{L}_a$ bijektiv. Ist umgekehrt $\mathcal{L}_a$ bijektiv, so gibt es wegen der Surjektivität ein $b \in M$ mit $a * b = e$, also ein Rechtsneutrales. Wir zeigen, dass wegen der Injektivität dieses b auch Linksneutrales von a ist, d.h., dass $b \circ a = e$ gilt. Dazu stellen wir fest, dass

$$\mathcal{L}_a(b * a) = a * (b * a) \overset{\text{Ass.}}{=} (a * b) * a \overset{\text{s.o.}}{=} e * a = a = a * e = \mathcal{L}_a(e)$$

ist. Daraus folgt per Injektivität $b * a = e$.

□

Sind *alle* Elemente einer algebraischen Struktur invertierbar, so ist nach dem obigen Satz alles gut: Gleichungen können durch Äquivalenzumformungen (also allgemein durch „Verkringeln" mit Inversen äquivalent umgeformt werden, ohne sich in Gleichungen mit Variablen Gedanken machen zu müssen, was die beteiligten Elemente genau sind. Eine solche Struktur nennt man *Gruppe*.

Definition 5.2.8. *(Gruppe; abelsche Gruppe): Ein Monoid $(M, *)$ heißt Gruppe, falls jedes Element von M invertierbar ist, also falls $M = M^\times$ gilt.*

Ist die Verknüpfung $*$ zusätzlich kommutativ, so heißt $(M, *)$ abelsche Gruppe[1]*.*

Um zu zeigen, dass eine algebraische Struktur $(M, *)$ eine Gruppe ist, müssen also folgende vier Eigenschaften gezeigt werden:

(i) Abgeschlossenheit: Für alle $a, b \in M$ ist $a * b \in M$.

(ii) $*$ ist assoziativ.

[1] Der Däne Niels Henrik Abel (1802-1829 (!)) war einer der Pioniere der Gruppentheorie

(iii) Es gibt ein neutrales Element $e \in M$.

(iv) Zu jedem $a \in M$ gibt es ein zugehöriges Inverses $a^{-1} \in M$.

Bemerkungen dazu:

- Im Gegensatz zu $(\mathbb{N}, +)$ sind $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$ und $(\mathbb{R}, +)$ Gruppen.
- $(\mathbb{N}, \cdot)$ und $(\mathbb{Z}, \cdot)$ sind keine Gruppen, da bis auf die 1 (in $\mathbb{N}$) bzw. 1 und -1 (in $\mathbb{Z}$) keines der Elemente einen Kehrwert in der jeweiligen Menge besitzt.
- Auch $(\mathbb{Q}, \cdot)$ und $(\mathbb{R}, \cdot)$ sind keine Gruppen, da die Null kein multiplikatives Inverses besitzt. (Denn „durch Null kann man nicht teilen“ – warum das eigentlich genau so ist, begründen wir genauer unten in Lemma 5.4.3.)
- Ist $(M, *)$ ein Monoid, so ist die Menge $(M^\times, *)$ eine Gruppe (Übungsaufgabe). (Überlegen Sie: Was genau ist hierfür zu zeigen?)
- Insbesondere sind $(\mathbb{Q}^\times, \cdot)$ und $(\mathbb{R}^\times, \cdot)$, also die Mengen $\mathbb{Q}^\times = \{a \in \mathbb{Q} \mid a \neq 0\}$ mit der Multiplikation bzw. $\mathbb{R}^\times \{a \in \mathbb{R} \mid a \neq 0\}$ mit der Multiplikation, Gruppen.
- Die Menge der $(n \times n)$-Matrizen (für festes $n \in \mathbb{N}$) mit der Matrixmultiplikation bildet *keine Gruppe*, da nicht alle $(n \times n)$-Matrizen invertierbar sind (Alle anderen Forderungen für eine Gruppe sind erfüllt).
- Die Menge der *invertierbaren* $(n \times n)$-Matrizen mit der Matrixmultiplikation bildet eine Gruppe, die zum Beispiel für $\mathbb{R}^{2\times 2}$ Spiegelungen, Drehungen und zentrische Streckungen enthält; diese wird mit $\mathrm{GL}(\mathbb{R}^n)$ („general linear group“ von $\mathbb{R}^n$) bezeichnet.

Die wichtigste Eigenschaft von Gruppen ist die schon oben erwähnte Lösbarkeit von Gleichungen per Äquivalenzumformungen:

Korollar 5.2.9. *(Eindeutige Lösbarkeit von Gleichungen in Gruppen)*

In jeder Gruppe $(M, *)$ ist jede Gleichung der Form $a * x = b$ (mit gegebenen Elementen $a, b \in M$, gesuchtem $x \in M$) eindeutig lösbar durch $x = a^{-1} * b$. Ebenso gilt „von rechts“ die Äquivalenz

$$x * a = b \qquad \Longleftrightarrow \qquad x = b * a^{-1}.$$

Beweis: In (M,*) sind alle Elemente invertierbar, also folgt das aus der Aussage von 5.2.7. □

5.3 Spezielle Typen algebraischer Strukturen

Wir wollen nun einen Blick auf ein paar typische algebraische Strukturen werfen, die uns schon untergekommen sind und uns wieder unterkommen werden.

5.3.1 Additive Strukturen

Die Addition ist in den Zahlbereichen $\mathbb{N}$, $\mathbb{Z}$ $\mathbb{Q}$, $\mathbb{R}$ die grundlegende Verknüpfung. Viele Verknüpfungen von allgemeineren Objekte (Funktionen, Matrizen, Koordinatenvektoren, ...) sind mit Hilfe der Addition in den Zahlbereichen definiert, z.B.

(i) Addition im $\mathbb{R}^n$: Durch punktweise Addition der Einträge in $\mathbb{R}$,

$$x + y = (x_i)_{i\in\underline{n}} + (y_i)_{i\in\underline{n}} := (\underbrace{x_i + y_i}_{\text{Add. in } \mathbb{R}})_{i\in\underline{n}}.$$

($x + y$ ist komponentenweise definiert als Vektor, der in Komponente i den Eintrag $x_i + y_i$ hat.)

(ii) Addition von Funktionen, $f, g \in \mathrm{Abb}(\mathbb{R})$: Punktweise Addition der Funktionswerte, $f + g$ wird punktweise definiert durch

$$\underbrace{(f + g)(x)}_{\text{Wert von } f+g \text{ an der Stelle } x} \quad := \quad \underbrace{f(x) + g(x).}_{\text{punktweise Addition von } f(x) \text{ und } g(x) \text{ in } \mathbb{R}}$$

(iii) Addition von Matrizen $A, B \in \mathbb{R}^{m\times n}$: Analog definieren wir

$$A + B = (a_{ij})_{i\in\underline{m},j\in\underline{n}} + (b_{ij})_{i\in\underline{m},j\in\underline{n}} := (\underbrace{a_{ij} + b_{ij}}_{\text{Add. in } \mathbb{R}})_{i\in\underline{m},j\in\underline{n}}.$$

(Erinnerung, vgl. Kapitel 2.2: (i) und (iii) sind eigentlich Spezialfälle von (ii).) Solche additiven Strukturen sind für gewöhnlich kommutative Gruppen:

- Das Assoziativgesetz und Kommutativgesetz überträgt sich, da es komponentenweise gilt.
- Das neutrale Element ergibt sich, indem man alle Einträge Null setzt.
- Inverse ergeben sich, indem man in jeder Komponente die Inverse des Eintrags wählt.

In additiven Strukturen schreibt man für das Neutrale auch gerne etwas, was an die Null erinnert (Nullfunktion, Nullvektor, Nullmatrix), für Inverse etwas wie $-f, -v, -A$, und für $f + (-g), v + (-w), A + (-B)$ kurz $f - g, v - w, A - B$.

5.3.2 Multiplikative Strukturen

- Als zweite wichtige Verknüpfung in Zahlbereichen kennt man die Multiplikation. In einer multiplikativen Struktur $(X, \cdot)$ schreibt man für gewöhnlich kurz ab für $a \cdot b$ und, falls es sie gibt, $\frac{1}{a}$ für die Inverse von a, außerdem sind die Notationen

$$a^n := \underbrace{a \cdot a \cdot \ldots \cdot a}_{n-\text{mal}}, \quad a^{-n} := (a^{-1})^n \text{ (falls } a \text{ invertierbar ist) und } a^0 := 1 \qquad (5.1)$$

  gebräuchlich.

- Es gibt auch einige Beispiele, in denen die Multiplikation der Zahlbereiche zugrunde gelegt wird, um elementweise eine Multiplikation in einer abgeleiteten Menge zu definieren. Wichtigstes Beispiel: Multiplikation von Funktionen, z.B. $f, g \in \mathrm{Abb}(\mathbb{R})$.
  - $f \cdot g$ wird punktweise definiert durch punktweise Multiplikation der Funktionswerte

$$\underbrace{(f \cdot g)(x)}_{\text{Wert von } f \cdot g \text{ an der Stelle } x} := \underbrace{f(x) \cdot g(x)}_{\text{punktweise Multiplikation von } f(x) \text{ und } g(x) \text{ in } \mathbb{R}}.$$

  - Ist f invertierbar bezüglich der Multiplikation (nicht zu verwechseln mit der Hintereinanderausführung), so ist die Notationen $\frac{1}{f}$ für die multiplikative Inverse gängig,
  - „Einsfunktion“: $f(x) = 1$ für alle $x \in \mathbb{R}$.
  - Analog zur Addition überträgt sich die Gültigkeit des Kommutativgesetzes, da man es in jedem Funktionswert in $\mathbb{R}$ anwenden kann.

- Für Vektoren aus dem $\mathbb{R}^n$ gibt es für $n = 2$ und $n = 3$ sinnvolle Verknüpfungen, die an die Multiplikation in den Zahlbereichen erinnern (für $n = 2$ s. das Kapitel über komplexe Zahlen, für $n = 3$ vergleiche das möglicherweise aus der Schule bekannte Kreuzprodukt $\times$ für Vektoren aus dem $\mathbb{R}^3$). Für allgemeines $n \in \mathbb{N}$ gibt es für Vektoren aus dem $\mathbb{R}^n$ *keine sinnvolle multiplikative Verknüpfung.*

5.3.3 Algebraische Strukturen von Mengen von Funktionen und Matrizen

- Sei $\mathrm{Abb}(X)$ die Menge der Abbildungen von einer Menge X in sich. Dann ist $(\mathrm{Abb}(X), \circ)$ eine algebraische Struktur. Für Inverse bezüglich $\circ$ benutzt man dann für gewöhnlich Notationen wie f^{-1}, A^{-1}, für das Neutrale die Bezeichnung id_X. Die Komposition von Funktionen ist assoziativ (s. Satz 5.1.1), aber für gewöhnlich *nicht kommutativ* und *nicht distributiv*, vgl. Übung.

- Die oben eingeführte Potenzschreibweise (5.1) wird auch für Funktionen und Matrizen verwendet. Für Matrizen ist

$$A^n := \underbrace{A \cdot A \cdot \ldots \cdot A}_{n-\text{mal}}, \quad A^{-n} := (A^{-1})^n \text{ (falls } A \text{ invertierbar ist)}, \quad A^0 = I_n,$$

  z.B.

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \quad \Longrightarrow \quad A^2 = \begin{pmatrix} 4 & 0 \\ 0 & 9 \end{pmatrix}, \quad A^{-1} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{pmatrix}, \ldots$$

- Betrachtet man Strukturen von Funktionen mit mehreren Verknüpfungen, so geht es oft um die Addition und die Hintereinanderausführung $\circ$ (an Stelle der Multiplikation). So erhält man Struktur $(M, +, \circ)$ mit zwei Verknüpfungen, bei denen nur $(M, +)$ eine Gruppe ist.

5.4 Ringe und Körper

Definition 5.4.1. *(Ring)*

Ein Tripel $(R, +, \cdot)$ heißt Ring, falls $+$ und $\cdot$ Verknüpfungen auf R sind und folgende drei Eigenschaften erfüllt sind:

(i) *$(R, +)$ ist eine abelsche Gruppe.*

(ii) *$(R, \cdot)$ ist eine assoziative algebraische Struktur.*

(iii) *Es gelten die Distributivgesetze: Für alle $a, b, c \in R$ ist*

$$(a + b) \cdot c = a \cdot c + b \cdot c \qquad \textit{und} \qquad c \cdot (a + b) = c \cdot a + c \cdot b.$$

Beispiele:

Die wichtigsten Ringe (in denen nicht alle Elemente $a \neq 0$ multiplikativ invertierbar sind) sind:

- $(\mathbb{Z}, +, \cdot)$,
- $(\mathbb{K}[x], +, \cdot)$ ist ein kommutativer Ring mit dem konstanten Polynom $p = 1$ als Einselement; man spricht vom Polynomring $\mathbb{K}[x]$ über $\mathbb{K}$.
- $\mathbb{R}^{n \times n}$, also der Ring der quadratischen Matrizen mit $(n \times n)$ Einträgen,
- Ringe mit endlich vielen Elementen, d.h. die Menge $R_n = \{0, 1, \ldots, n-1\}$ mit Addition und Multiplikation *modulo* n, („Rechnen an der Uhr“, s. VL Algebra/Zahlentheorie)
- $(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$ sind auch Ringe, aber „können“ noch mehr (s.u., „Körper“)
- $\text{Abb}(\mathbb{R}, +, \cdot)$ ist ein Ring, aber $\text{Abb}(\mathbb{R}, +, \circ)$ ist *nicht* distributiv[2], also kein Ring.

Weitere Bemerkungen zum Begriff des Ringes:

- Die zweite Verknüpfung $\cdot$ bzw. $\circ$ muss in einem Ring nicht kommutativ sein, vgl. oben: Das wichtigste Beispiel von nichtkommutativen Strukturen sind algebraische Strukturen von Abbildungen fast jeglicher Art, z.B. die von Funktionen $\text{Abb}(X)$ einer Menge X in sich selbst, insbesondere die Struktur der Matrixabbildungen.
- In Ringen kann die Struktur $(R, \cdot)$ zusätzlich ein Monoid oder kommutativ sein. R heißt dann *Ring mit 1* (z.B. Abbildungen, Polynomringe) bzw. *kommutativer Ring* (z.B. P,olynomringe).
- In Ringen ist Teilbarkeitslehre von Bedeutung – für $\mathbb{Z}$ kennen Sie das vielleicht aus der Unter- bis Mittelstufe (Stichwort Primfaktorzerlegung usw.), für Polynome vielleicht aus der Oberstufe (Stichwort „Polynomdivision“).

[2] Es gilt zwar das erste Distributivgesetz ($(f+g)\circ h = f\circ h + g\circ h$), das zweite ($h\circ(f+g) = h\circ f + h\circ g$) ist aber verletzt. So kann's laufen in nicht-kommutativen algebraischen Strukturen.

Lemma 5.4.2. *(Einige Rechenregeln in Ringen)*

Sei $(R, +, \cdot)$ ein Ring. Dann gelten für alle $a, b \in R$ die folgenden Rechenregeln:

(i) $0 \cdot a = 0$ *und* $a \cdot 0 = 0$ *für alle* $a \in R$,

(ii) $(-1) \cdot a = -a$ *und* $a \cdot (-1) = -a$,

(iii) $a \cdot (-b) = (-a) \cdot b = -(a \cdot b)$,

(iv) $(-a) \cdot (-b) = a \cdot b$,

(v) Potenzgesetze: Mit der Festlegung (5.1) gilt für $n, m \in \mathbb{N}$:

$$a^{n+m} = a^n \cdot a^m \quad \textit{und} \quad (a^n)^m = a^{n \cdot m}.$$

(vi) Binomische Formeln in kommutativen Ringen: Ist $(R, \cdot)$ *kommutativ, so gilt*

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

(wobei hier $2ab$ *als* $a \cdot b + a \cdot b$ *zu lesen ist) und*

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

(Man beachte: $-1, -a, -(a \cdot b)$ *usw. muss man hier allgemein lesen als „die additive Inverse des Elements* $1, a$ *bzw.* $a \cdot b$*“ usw.)*

Lemma 5.4.3. *(Durch Null kann man in Ringen nie teilen (oder es wird langweilig)!)*

Sei $(R, +, \cdot)$ *ein Ring mit additivem neutralen Element* 0*. Wenn man in* R *durch Null teilen kann (d.h.* 0 *besitzt eine multiplikative Inverse), dann ist* $R = \{0\}$ *(„der Nullring“).*

Beweis: Wir zeigen, dass zwei Elemente $a, b \in R$ unter den gegebenen Voraussetzungen immer gleich sein müssen. Mit Rechenregel 5.4.2 (i) haben wir $a \cdot 0 = 0$ und $b \cdot 0 = 0$. Ist 0 invertierbar in $(R, \cdot)$, so hat die Gleichung $x \cdot 0 = 0$ nach Satz 5.2.7 genau eine Lösung, also gilt $a = b$ für alle $a, b \in R$, also $a = 0$ für alle Elemente von R. □

Interpretation der Rechenregeln in 5.4.2 und 5.4.3:

Wenn wir es für sinnvoll erachten, dass in einer algebraischen Struktur mit zwei Verknüpfungen die Rechengesetze Assoziativgesetz und Distibutivgesetz gelten sollen, dann

- muss nach 5.4.2 (iv) immer „minus mal minus gibt plus“ gelten, d.h. es gibt insbesondere keine „sinnvolle“ Erweiterung von $(\mathbb{N}, +, \cdot)$, in der das nicht so ist, und
- wird man nie durch Null teilen können, d.h es gibt insbesondere keine „sinnvolle“ Erweiterung von $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$, in der man das kann.

Das additive Element eines Ringes wird also in interessanten Fällen niemals eine multiplikative Inverse besitzen. Für alle anderen Elemente ist das durchaus möglich, z.B. in den Ringen $(\mathbb{Q}, +, \cdot)$ und $(\mathbb{R}, +, \cdot)$.

Definition 5.4.4. *(Körper)*

Ein Ring $(R, +, \cdot)$ *heißt* *Körper, wenn*

(i) $(R, \cdot)$ *kommutativ ist,*

(ii) $(R, \cdot)$ *ein neutrales Element („eine Eins“) enthält und*

(iii) das additive neutrale Element $0 \in R$ *das einzige Element ist, das in* $(R, \cdot)$ *keine multiplikative Inverse besitzt.*

Wenn wir ab jetzt über einen beliebigen Körper reden, werden wir diesen mit $\mathbb{K}$ *abkürzen (und damit einen Körper* $(\mathbb{K}, +, \cdot)$ *meinen).*

Denkaufgabe: Schreiben Sie das ausführlich hin, d.h. schreiben Sie die jeweiligen Definitionen von „Ring“ und im nächsten Schritt von „Gruppe“ aus. Welcher Katalog von Regeln muss insgesamt erfüllt sein, damit eine Menge $\mathbb{K}$ mit Verknüpfungen $+, \cdot$ ein Körper ist?

Die wichtigsten Beispiele für Körper:

- $(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$, $(\mathbb{C}, +, \cdot)$.

Neben diesen grundlegenden Zahlbereichen liefern die so genannten endlichen Körper, also Körper mit endlich vielen Elementen, ein weiteres wichtiges Beispiel. Diese werden wir aber in diesem Skript nicht groß weiter thematisieren, daher hier nur ein kurzer Exkurs: Auf einer endlichen Menge M lassen sich $+$ und $\cdot$ per Rechnen mit Rest, der so genannten „Modulo-Rechnung“ definieren. Die resultierende algebraische Struktur $(M, +, \cdot)$ ist ein Körper genau dann, wenn M p^n Elemente enthält (p Primzahl, $n \in \mathbb{N}$). Bis auf Umbenennungen gibt es dafür dann genau eine Möglichkeit. Diese Körper sind unter anderem deshalb von so großer praktischer Bedeutung, weil man sie im Gegensatz zu den Körpern oben in Computern exakt nachbauen und exakt mit ihnen rechnen kann. Sie sind deshalb viel besser als das ungenaue Abbild der rationalen/reellen Zahlen im Rechner und daher Grundlage für viele technische Anwendungen, allen voran die Codierung und Prüfung von Daten.

Wir halten noch zwei wichtige Rechenregeln fest, die wir in den reellen Zahlen schon benutzt haben und auch oft in allgemeinen Betrachtungen von Körpern benutzen werden.

Lemma 5.4.5. *(Nullteilerfreiheit in Körpern, Potenzgesetze)*

(i) Sei $\mathbb{K}$ ein beliebiger Körper. Dann gilt für alle $a, b \in \mathbb{K}$:

$$a \cdot b = 0 \quad \Longrightarrow \quad a = 0 \text{ oder } b = 0, \tag{5.2}$$

man sagt, Körper sind nullteilerfrei.

(ii) In Erweiterung der Potenzgesetze in Lemma 5.4.2 gilt für alle Körperelemente $a \neq 0$ für alle $p, q \in \mathbb{Z}$:

$$a^{p+q} = a^p \cdot a^q \quad \text{und} \quad (a^p)^q = a^{p \cdot q}.$$

Beweis: (i) beweisen wir, indem wir annehmen, beide Zahlen a, b wären ungleich Null. Da $(\mathbb{R}^\times, \cdot)$ eine Gruppe ist (vgl. Übungsaufgabe), müsste wegen der Abgeschlossenheit von $(\mathbb{R}^\times, \cdot)$ auch schon $a \cdot b \neq 0$ sein. (ii) ist Übungsaufgabe, die sich durch Prüfen der verschiedenen Fälle für Kombinationen von p, q erledigen lässt (jeweils positiv, negativ, Null).

□

Bemerkungen dazu:

- Die praktische Hauptanwendung der Nullteilerfreiheit ist, multiplikative Gleichungen zu „zerlegen“, z.B. löst $x \in \mathbb{R}$ die Gleichung $(x^2 - 2) \cdot (\sin(x) - \frac{1}{2}) = 0$ genau dann, wenn $x^2 = 2$ oder $\sin(x) - \frac{1}{2} = 0$ ist, und diese beiden Gleichungen lassen sich gut lösen. Diesen Zusammenhang wird in Rechnungen oft als Fallunterscheidung notiert: Ist im Beispiel $x^2 - 2 = 0$, so kann man diese Gleichung lösen, ist $x^2 - 2 \neq 0$, so ist die Division der Gleichung $(x^2 - 2) \cdot (\sin(x) - \frac{1}{2}) = 0$ durch $x^2 - 2$ eine Äquivalenzumformung, und man kann stattdessen $(\sin(x) - \frac{1}{2} = 0$ lösen. Insgesamt ergeben sich wegen

$$(x^2 - 2) \cdot (\sin(x) - \frac{1}{2}) = 0 \quad \Leftrightarrow \quad x^2 - 2 = 0 \text{ oder } \sin(x) - \frac{1}{2}) = 0$$

alle Lösungen der ursprünglichen Gleichung.

- Ringe sind nicht unbedingt nullteilerfrei, z.B. ist in $\mathbb{R}^{2\times 2}$

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

  und die Nullmatrix rechts ist die Null in $\mathbb{R}^{2\times 2}$.

5.4.1 Schlussbemerkungen: Rechnen in Gruppen, Ringen und Körpern

Der Deutlichkeit halber fassen wir nochmal die wesentlichen Erkenntnisse zusammen:

- In **Körpern** (insb. in $\mathbb{R}$ und $\mathbb{C}$) ist das Addieren (und Subtrahieren) und die **Multiplikation mit allen Körperelementen ungleich Null (und also auch Division durch Elemente ungleich Null) eine Äquivalenzumformung**.

  Beispiel: Für reelle Zahlen a, b, c, d gilt genau dann, wenn $c \neq 0$ ist:

$$a = b + c \cdot d \quad \Longleftrightarrow \quad d = \frac{a-b}{c}.$$

- In **Ringen** ist das Addieren und das Subtrahieren (d.h. Addieren von additiven Inversen) immer eine Äquivalenzumformung. **Das Multiplizieren mit Elementen (und ihren Inversen, den „Kehrwerten") ist genau dann eine Äquivalenzumformung, wenn das Element invertierbar ist.**

  Beispiel: Für Matrizen A, B, C, D gilt genau dann, wenn C invertierbar ist:

$$A = B + C \cdot D \quad \Longleftrightarrow \quad D = C^{-1} \cdot (A - B).$$

5.5 Aufgaben zu Kapitel 5

Aufgabe 5.23. (Verknüpfungen von Funktionen aus Abb($\mathbb{R}$)):

Wir betrachten die Menge Abb($\mathbb{R}$) der Abbildungen von $\mathbb{R}$ in sich. Diese lassen sich gleich mit drei Verknüpfungen betrachten, die alle werteweise definiert sind:

$$\begin{aligned}
+: &\quad (f,g) \mapsto f+g, &&\text{festgelegt durch } (f+g)(x) := f(x)+g(x),\\
\cdot: &\quad (f,g) \mapsto f\cdot g, &&\text{wobei } (f\cdot g)(x) := f(x)\cdot g(x),\\
\circ: &\quad (f,g) \mapsto f\circ g, &&\text{wie in der Vorlesung.}
\end{aligned}$$

Für $f(x) = \sin(x)$ und $g(x) = x^2$ beispielweise ist

$$(f+g)(x) = \sin(x) + x^2, \qquad (f\cdot g)(x) = \sin(x)\cdot x^2 \qquad \text{und } (f\circ g)(x) = \sin(x^2).$$

(a) Beweisen Sie: $(\mathrm{Abb}(\mathbb{R}), \cdot)$ ist eine assoziative algebraische Struktur[3].

(b) $(\mathrm{Abb}(\mathbb{R}), +)$, $(\mathrm{Abb}(\mathbb{R}), \cdot)$ und $(\mathrm{Abb}(\mathbb{R}), \circ)$ besitzen jeweils ein neutrales Element. Finden Sie es und beweisen Sie, dass es sich wirklich um das neutrale Element der jeweiligen algebraischen Struktur handelt.

(c) Eine der Verknüpfungen $+, \cdot, \circ$ ist nicht kommutativ. Zeigen Sie für die anderen beiden, dass sie kommutativ sind.

(d) Beweisen Sie, dass jedes Element $f \in \mathrm{Abb}(\mathbb{R})$ ein inverses Element bezüglich $+$ besitzt.

(e) Beweisen Sie, dass nicht jedes Element $f \in \mathrm{Abb}(\mathbb{R})$ ein inverses Element bezüglich $\cdot$ besitzt.

[3] Die anderen beiden auch, aber für $\circ$ wissen wir das schon aus der Vorlesung, und für $+$ sollen Sie es hier nicht beweisen.

Aufgabe 5.24. (Invertierbare Elemente in Monoiden):

Sei $(M, *)$ ein Monoid, also eine assoziative algebraische Struktur mit neutralem Element $e \in M$. Bezeichne

$$M^\times = \{x \in M \mid x \text{ besitzt ein inverses Element in } M\}$$

die Menge der invertierbaren Elemente aus M.

(i) Zeigen Sie, dass $e \in M^\times$ ist.

(ii) Zeigen Sie: Sind $x, y \in M^\times$ beliebig, so ist $x * y \in M^\times$.

(iii) Zeigen Sie: Ist $x \in M^\times$ beliebig, so ist auch $x^{-1} \in M^\times$, wobei x^{-1} das zu x inverse Element ist.

(iv) Begründen Sie, dass für alle $x, y, z \in M^\times$ das Assoziativgesetz gilt, d.h. es ist

$$x * (y * z) = (x * y) * z$$

(Damit haben Sie gezeigt, dass $(M^\times, *)$ immer eine Gruppe ist.)

(v) Zeigen Sie außerdem die folgende Kürzungsregel: Ist $a \in M^\times$, so gilt für alle $x, y \in M$:
Es ist $x = y$ genau dann, wenn $a * x = a * y$ ist.

Aufgabe 5.25. (Eine merkwürdige Gruppe):

Für zwei beliebige reelle Zahlen x, y sei

$$x * y := x + y + 2.$$

Zeigen Sie, dass $(\mathbb{R}, *)$ eine Gruppe ist.

Aufgabe 5.26. (Rechenregeln für Matrizen):

Wir betrachten das Tripel $(\mathbb{R}^{n\times n}, +, \cdot)$ mit der Menge $\mathbb{R}^{n\times n}$ der quadratischen Matrizen mit n Zeilen und n Spalten, der durch

$$A + B = (a_{ij})_{i\in\underline{n}, j\in\underline{n}} + (b_{ij})_{i\in\underline{n}, j\in\underline{n}} := (a_{ij} + b_{ij})_{i\in\underline{n}, j\in\underline{n}}$$

eintragsweise definierten Addition $+$ und der Matrixmultiplikation $\cdot$ als zweiter Verknüpfung.

(a) Beweisen Sie: $(\mathbb{R}^{n\times n}, +, \cdot)$ ist ein Ring, d.h. zeigen Sie:
   - $(\mathbb{R}^{n\times n}, +)$ ist eine kommutative Gruppe. Hierzu müssen Sie Argumente aus der Vorlesung vom 25.11. präzisieren.
   - $(\mathbb{R}^{n\times n}, \cdot)$ ist eine assoziative algebraische Struktur.
   - Es gelten die beiden Distributivgesetze.

   Für die letzten beiden Punkte können Sie, falls vorhanden, auch geeignete Aussagen aus der Vorlesung zitieren.

(b) Zeigen Sie: Sind $A, B \in (\mathbb{R}^{n\times n}, \cdot)^\times$, so ist auch $A \cdot B$ invertierbar.

Aufgabe 5.27. (Rechenregeln in Ringen, Beweis von Lemma 5.4.2):

Sei $(R, +, \cdot)$ ein Ring. Zeigen Sie, dass dann für alle $a, b \in R$ die folgenden Rechenregeln gelten[4]:

(iii) $a \cdot (-b) = -(a \cdot b)$ und $(-a) \cdot b = -(a \cdot b)$,

(iv) $(-a) \cdot (-b) = a \cdot b$,

(v) Potenzgesetze: Mit der Festlegung $a^n := \underbrace{a \cdot \ldots \cdot a}_{n-\text{mal}}$ gilt für $n, m \in \mathbb{N}$:

$$a^{n+m} = a^n \cdot a^m \quad \text{und} \quad (a^n)^m = a^{n\cdot m}.$$

(vi) Binomische Formeln in kommutativen Ringen: Ist $(R, \cdot)$ kommutativ, so gilt

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

[4]Die Nummerierung entspricht der von Lemma 5.4.2

(wobei hier $2ab$ als $a \cdot b + a \cdot b$ zu lesen ist) und

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

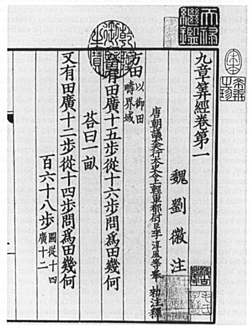
6 Lineare Gleichungssysteme (Teil 1)

Gleichungssysteme

Leonhard Euler (1707-1783)

C.F. Gauß (18. Jhd.)

Al-Chwarizmi (um 800 n.Chr.)



Jiu Zhang Suanshu, 1. Jhd. u.Z.

6.1 Lineare und nichtlineare Gleichungssysteme

Im folgenden sei $\mathbb{K}$ einer der Körper $\mathbb{K} = \mathbb{Q}, \mathbb{R}$ oder $\mathbb{C}$.

Definition 6.1.1. *(Gleichungssysteme) Seien $f_1, \ldots, f_m : \mathbb{K}^n \to \mathbb{K}$ Funktionen und $b_1, \ldots, b_m \in \mathbb{K}$. Die Aufgabe „Bestimme $x = (x_1, \ldots, x_n) \in \mathbb{K}^n$ so, dass x jede der Gleichungen*

$$(*) \left\{ \begin{array}{rcl} f_1(x) & = & b_1 \\ f_2(x) & = & b_2 \\ & \vdots & \\ f_m(x) & = & b_m \end{array} \right. \tag{6.1}$$

erfüllt“ (d.h. jede zu einer wahren Aussage macht) heißt Gleichungssystem für $x \in \mathbb{K}^n$. Wir sagen auch „x löst (*)“. Die Menge

$$\mathbb{L} = \{x \in \mathbb{K}^n \mid (f_1(x) = b_1) \wedge \ldots \wedge (f_m(x) = b_m)\}$$

ist die Lösungsmenge des Systems (*). Eine Äquivalenzumformung eines solchen Systems ist das Ersetzen des Systems (*) durch ein GS (**) für x mit gleicher Lösungsmenge.

Wir betrachten hier die standardisierte Form (*) eines GS, in der auf der rechten Seite nur konstante Terme auftauchen. Allgemeinere Gleichungen $f_i(x) = g_i(x)$ lassen sich durch Äquivalenzumformung in $\mathbb{K}$ in diese Form bringen, s.u.

Es ist eine alte Kunst (China 1. Jhd. n.Chr., Al Kwarizmi 8. Jhd., Gauß 19. Jhd.), die Zeilen eines Gleichungssystems per Äquivalenzumformungen so zu manipulieren, dass die Lösungsmenge erhalten bleibt, die Gleichungen aber so einfach werden, dass man die Lösung ablesen kann. Die wichtigsten erlaubten Umformungen sind die im folgenden Satz zusammengestellten.

Satz 6.1.2. *(Äquivalenzumformungen für Zeilen von Gleichungssystemen)*

Folgende Umformungen von Gleichungssystemen der Form (*) sind Äquivalenzumformungen:

(i) *Beliebige Vertauschung der Reihenfolge der Gleichungen,*

(ii) *Äquivalenzumformungen in einzelnen Zeilen, insbesondere Addition von $c \in \mathbb{K}$, Multiplikation einer Zeile mit $c \in \mathbb{K}\backslash\{0\}$,*

(iii) *Weglassen immer wahrer Gleichungen wie „$0 = 0$“,*

(iv) *Addition einer Zeile auf eine andere Zeile, d.h. die Ersetzung von $f_i(x) = b_i$ durch $f_i(x) + f_j(x) = b_i + b_j$ für Zeilenindizes $i, j \in \underline{m}$,*

(v) *Für $k, \ell \in \mathbb{K}^\times$: Ersetzen der i-ten Zeile $f_i(x) = b_i$ durch die Summe des k-fache von Zeile i und des ℓ-fachen von Zeile j, $k \cdot f_i(x) + \ell \cdot f_j(x) = k \cdot b_i + \ell \cdot b_j$,*

(vi) *Sind $f_i(x) = b$ und $f_j(x) = b$ zwei verschiedene Zeilen mit identischer rechter Seite: Ersetzen der i-ten Gleichung durch $f_i(x) = f_j(x)$,*

(vii) *Einsetzen einer beliebigen wahren Gleichung $x_k = g(x)$ für eine der Variablen x_k in Gleichung i,*

(viii) *Einsetzen der nach x_k umgestellten j-ten Gleichung für eine der Variablen x_k in Gleichung i.*

Bemerkungen dazu:

- Die Aussagen (iv), (vi) und (viii) legitimieren die (hoffentlich) aus der Schule bekannten Verfahrensweisen zur Lösung linearer Gleichungssysteme (in dieser Reihenfolge: Additionsverfahren, Gleichsetzungsverfahren, Einsetzungsverfahren) in größerer Allgemeinheit.
- Beachten Sie: Es müssen bei diesen Ersetzungen weiterhin alle anderen Gleichungen Teil des neuen Systems sein, damit der Tausch „altes System $\rightsquigarrow$ neues System“ eine Äquivalenzumformung ist.
- Die Multiplikation einer j-ten Zeile mit Gleichungen von der Form $\text{Term}_1 = \text{Term}_2$, z.B. mit $f_j(x) = b_j$, ist nur dann Äquivalenzumformung, falls die Term nie Wert Null annimmt. Für gewöhnlich muss man diesen Fall dann gesondert behandeln oder durch Argumente ausschließen, dass dieser überhaupt auftreten kann.

Beweis: (i) ist klar, da die Reihenfolge der Gleichungen keine Rolle spielt (Etwas formaler Betrachtet ist das die Kommutativität logischer Aussagen 7.5.3(ii).). (ii) ist klar, die umgeformte Zeile ist ja wahr genau dann, wenn die Ursprungszeile wahr ist; daher ändert sich bei einer solchen Umformung auch nichts daran, ob alle Gleichungen wahr sind oder nicht. (Auch das lässt sich strikter formal-logisch begründen, vgl. 7.5.4(v).), (iii) folgt daraus, dass die Hinzunahme oder das Weglassen einer immer wahren Aussage keinen Einfluss auf den gesamten Wahrheitswert hat (formal steht das in 7.5.4(iii)). Zu (iv): Ist x eine Lösung von (*), so gilt

$$f_i(x) = b_i \overset{\text{(ii) für Gl. } i}{\iff} f_i(x) + b_j = b_i + b_j \overset{\text{Gl. } j \text{ ist wahr}}{\iff} f_i(x) + f_j(x) = b_i + b_j,$$

d.h. x ist eine Lösung des neuen Gleichungssystems.
Ist x keine Lösung, so ist entweder eine der anderen Gleichungen falsch und bleibt falsch, also ist x keine Lösung des umgeformten Systems, oder nur die i-te Gleichung ist falsch, so dass wegen

$$f_i(x) \neq b_i \overset{\text{(ii) für Gl. } i}{\iff} f_i(x) + b_j \neq b_i + b_j \overset{\text{Gl. } j \text{ ist wahr}}{\iff} f_i(x) + f_j(x) \neq b_i + b_j$$

auch die neue i-te Gleichung falsch ist, x ist also auch keine Lösung des neuen Systems. Die Ersetzung in (v) ist eine Äquivalenzumformung, da sie die Hintereinanderausführung der Äquivalenzumformungen „Zeile i mal k“, „Zeile j mal ℓ“, „Zeile i ersetzt durch Zeile i

plus Zeile j“, „Zeile j mal $\frac{1}{\ell}$“ ist. (vi) und (viii) sind Übungsaufgaben, also bleibt (vii) zu zeigen: Ist $x_k = g(x)$, so ist $f_i(x_1, \ldots, g(x), \ldots, x_n) = f_i(x)$. Also gilt auch

$$f_i(x) = b_i \quad \Longleftrightarrow \quad f_i(x_1, \ldots, g(x), \ldots, x_n) = b_i,$$

und die Behauptung folgt mit (ii). □

Definition 6.1.3. *(Lineare Gleichungen)*
Eine Gleichung $f(x) = b$ für $x = (x_1, \ldots, x_n) \in \mathbb{K}^n$ mit $b \in \mathbb{K}$ heißt linear, falls die Funktion $f : \mathbb{K}^n \to \mathbb{K}$ die Form

$$f(x) = a_1 \cdot x_1 + a_2 \cdot x_2 + \ldots + a_n \cdot x_n \tag{6.2}$$

mit $a_1, \ldots, a_n \in \mathbb{K}$ hat. Nutzen wir als Kurzschreibweise das Skalarprodukt

$$\langle a, x \rangle = a_1 \cdot x_1 + a_2 \cdot x_2 + \ldots + a_n \cdot x_n,$$

für $a = (a_1, \ldots, a_n), x = (x_1, \ldots, x_n)$, so ist eine lineare Gleichung also von der Form $\langle a, x \rangle = b$, wobei $a \in \mathbb{K}^n, b \in \mathbb{K}$ gegeben und die Menge aller diese Bedingungen erfüllenden $x \in \mathbb{K}^n$ gesucht ist.
Eine Gleichung $f(x) = b$, in der $f : \mathbb{K}^n \to \mathbb{K}$ nicht die obige Form hat, heißt nichtlinear (oft nach der beteiligten Funktion f z.B. quadratische Gleichung, Polynomgleichung, trigonometrische Gleichung,...).

Definition 6.1.4. *(Lineares Gleichungssystem, (erweiterte) Koeffizientenmatrix)*
Ein lineares Gleichungssystem (LGS) für $x \in \mathbb{K}^n$ ist ein Gleichungssystem, in dem alle vorkommenden Gleichungen linear sind, d.h. ein System der Form

$$(*) \left\{ \begin{array}{rcl} a_{1,1} \cdot x_1 + a_{1,2} \cdot x_2 + \ldots + a_{1,n} \cdot x_n & = & b_1 \\ a_{2,1} \cdot x_1 + a_{2,2} \cdot x_2 + \ldots + a_{2,n} \cdot x_n & = & b_2 \\ & \vdots & \\ a_{m,1} \cdot x_1 + a_{m,2} \cdot x_2 + \ldots + a_{m,n} \cdot x_n & = & b_m \end{array} \right. . \tag{6.3}$$

Bei einem System mit m Gleichungen für n Unbekannte spricht man auch von einem $(m \times n)$-System. Sobald eine der Gleichungen eines Gleichungssystems nichtlinear ist, heißt das Gleichungssystem ebenfalls nichtlinear.

Fassen wir die Koeffizienten $a_{i,j}$ eines LGS der Form (6.3) in einer $m \times n$-Matrix A (der Koeffizientenmatrix des LGS) und die Einträge der rechten Seite in einem Vektor $b \in \mathbb{K}^m$ zusammen, so erhalten wir als kompakte Schreibweise für das LGS die so genannte erweiterte Koeffizientenmatrix $(A \mid b) \in \mathbb{K}^{m\times(n+1)}$ des Systems (6.3):

$$(A \mid b) = \left(\begin{array}{cccc|c} a_{1,1} & a_{1,2} & \dots & a_{1,n} & b_1 \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} & b_m \end{array}\right) \tag{6.4}$$

Auf dieser lassen sich Äquivalenzumformungen bequem und übersichtlich durchführen, auch bei der heute gängigen Lösung solcher Systeme mit Hilfe von Rechnern, und es lässt anhand der Matrix A die Lösbarkeit des Systems untersuchen.

6.2 Einige Begriffe für lineare Gleichungssysteme

Unser nächstes Ziel ist die Formulierung des so genannten Gauß-Jordan-Algorithmus, der das gängigste Verfahren ist, das die Lösungsmenge eines zugehörigen Linearen Gleichungssystems (6.3) berechnet. Es lässt sich mithilfe von elementaren Rechenoperationen an der erweiterten Koeffizientenmatrix durchführen und ist komplett am Rechner umsetzbar. Dazu formt der Algorithmus das gegebene LGS mit Hilfe von Äquivalenzumformungen in ein äquivalentes System um, aus dem sich die Lösbarkeit des LGS einfach entscheiden lässt und formuliert gegebenenfalls, d.h. wenn es eine nicht leere Lösungsmenge gibt, k unabhängige Bedingungen für k der n Variablen (die abhängigen Variablen der Lösungsmenge). Man nimmt dann die übrigen $n-k$ Variablen als Parameter; damit lässt sich aus den k Bedingungen der Wert der abhängigen Variablen berechnen.

6.2.1 Zeilenstufenform, normierte Zeilenstufenform, obere Dreiecksmatrizen

Beispiel zu Zeilenstufenform, normierter Zeilenstufenform: S. Vorlesung. $\rightsquigarrow$ Aus Matrizen in Zeilenstufenform lässt sich die Lösungsmenge ablesen, indem man die Variablen, $x_{k_1}, \dots, x_{k_r}$, deren Indizes als Zeilenköpfe vorkommen, durch die übrigen freien Variablen ausdrückt.

Definition 6.2.1. *Sei $A \in \mathbb{K}^{m\times n}$ eine beliebige Matrix (insbesondere kann A auch eine erweiterte Koeffizientenmatrix sein). Ein Spaltenindex k der Matrix A heißt <u>Pivotindex</u> der i-ten Zeile von A ($i \in \underline{m}$), falls $a_{i,j} = 0$ für alle $j < k$ und $a_{i,k} \neq 0$ gilt. Wir schreiben für den Pivotindex der i-ten Zeile $p(i)$. Gibt es keinen solchen Index, so setzen wir $p(i) = n+1$ und nennen die Zeile <u>Nullzeile</u>.*

Wir sagen von einer (erweiterten) Koeffizientenmatrix A, sie sei in <u>Zeilenstufenform</u>, falls für alle Zeilenindizes $i \in \underline{n-1}$ gilt: $p(i+1) > p(i)$ oder $p(i+1) = n+1$.

A ist in <u>normierter Zeilenstufenform</u>, falls A in Zeilenstufenform ist und zusätzlich gilt:

- *Für alle Nicht-Nullzeilen i ist der Eintrag am Pivotindex 1 ist, d.h. es gilt $a_{i,p(i)} = 1$ und*
- *Ist $p(i) < n+1$ ein Pivotindex, so gilt in der zugehörigen Spalte $a_{j,p(i)} = 0$ für <u>alle</u> Einträge $j \neq i$ gilt.*

6.3 Der Gauß-Jordan-Algorithmus

Aus einer Matrix in normierter Zeilenstufenform lässt sich also die Lösung eines LGS wie oben skizziert einfach ablesen und in der Mengenschreibweise wie im Beispiel (s. Vorlesung) angeben. Der Gauß-Jordan-Algorithmus, der eine beliebiges lineares Gleichungssystem auf normierte Zeilenstufenform bringt, läuft ab wie auf der folgenden Seite dargestellt.

1. Bilde die erweiterte Koeffizientenmatrix $(A|b) \in \mathbb{K}^{m\times(n+1)}$ des LGS.

Wiederhole Schritt 2 und 3 für alle Zeilen i, $i \in \underline{n}$:

2. Ist der Pivotindex $p(i)$ der Zeile i größer als der von einer der Zeilen darunter, tausche die Zeile i ganz nach unten; alle anderen Zeilen rutschen eine Zeile nach oben. (Damit sorgt man dafür, dass folgende Spaltenköpfe niemals links vom Spaltenkopf der Zeile i stehen.)
3. Ist der Pivotindex $p(i)$ der Zeile i kleiner als $n+1$, und gibt es unterhalb des Eintrags $a_{i,p(i)}$ Einträge $a_{\ell,p(i)} \neq 0$, ersetze in allen diesen Zeilen Z_ℓ die gesamte Zeile durch $Z_\ell - \frac{a_{\ell,p(i)}}{a_{i,p(i)}} \cdot Z_i$.

(Die Matrix ist, nachdem diese Schritten für alle Zeilen $i \in \underline{n}$ durchgeführt wurden, in Zeilenstufenform.)

4. Gehe die Zeilen von oben nach unten durch: Sollte der Eintrag $(a_{i,p(i)})$ am Pivotindex nicht 1 sein, ersetze die Zeile durch $\frac{1}{a_{i,p(i)}} Z_i$.
5. Statt von oben nach unten geht man die Matrix nun von unten nach oben durch und ändert alle Spalten analog zu Schritt 3 derart, dass auch über den Pivotelementen nur Nullen stehen. (Die Pivotvariablen x_k werden jetzt unabhängig von den übrigen Pivotvariablen ausdrückt.)
6. Gib die Normierte Zeilenstufenform $(\hat{A} \mid \hat{b})$ aus. Zur Auswertung gibt es drei mögliche Fälle:

   1. Fall: Ist einer der Pivotindizes der ausgeworfenen Matrix $n+1$, so ist der Eintrag in der Rechte-Seite-Spalte der erste von Null verschiedene Zeileneintrag einer Zeile i. Es ergibt sich hier rückübersetzt die Gleichung $0 = \hat{b}_i$ mit $\hat{b}_i \neq 0$; also ist das LGS nicht lösbar.

   2. Fall: Die Pivotindizes sind gerade *alle* zu $\hat{A}$ gehörigen Spaltenindizes $k \in \underline{n}$, (insbesondere ist $n+1$ keiner der Pivotindizes von $(\hat{A} \mid \hat{b})$). Dann lassen sich die einzelnen Zeilen als $x_1 = \hat{b}_1, \ldots, x_n = \hat{b}_n$ rückinterpretieren. Das LGS ist somit eindeutig lösbar, und die Lösung direkt in der letzten Spalte ablesbar.

   3. Fall: Keiner der Pivotindizes ist $n+1$ (das schließt Fall 1 aus), $\hat{A}$ hat zudem Spalten, die keine Pivotspalten sind. Die Variablen x_j, für die j kein Pivotindex ist, nennt man *freie Variablen* der normierten Zeilenstufenform. Die Werte der Variablen x_k, die zu Pivotindizes k gehören, lassen sich dann in Abhängigkeit der freien Variablen angegeben.

Beispiele s. Vorlesung.

Bemerkungen dazu:

- Die Schritte 1., 2., 3. nennt man den Gauß-Algorithmus, den gesamten Algorithmus den Gauß-Jordan-Algorithmus.

- Der Gauß-Jordan-Algorithmus liefert nach Konstruktion bei Anwendung auf eine erweiterte Koeffizientenmatrix $(A \mid b) \in \mathbb{R}^{m\times(n+1)}$ eine Matrix $(\hat{A} \mid \hat{b}) \in \mathbb{R}^{m\times(n+1)}$ in normierter Zeilenstufenform (vgl. Vorlesung/Übungsaufgabe).

- Anwendung des Gauß-Algorithmus (also von Schritten 1. bis 3.) auf eine erweiterte Koeffizientenmatrix $(A \mid b) \in \mathbb{R}^{m\times(n+1)}$, die bereits in Zeilenstufenform ist, gibt die Matrix $(A \mid b)$ aus; Anwendung des Gauß-Jordan-Algorithmus auf eine Matrix in normierter Zeilenstufenform gibt ebenfalls wieder diese Matrix aus.

- Überprüfen Sie: Alle im Gauß-Jordan-Algorithmus durchgeführten Schritte sind Äquivalenzumformungen nach Satz 6.1.2. Genauer gesagt kommt der Gauß-Algorithmus mit Kombinationen der Operationen
  - Vertauschen von zwei Zeilen,
  - Multiplikation einer Zeile mit $a \neq 0$,
  - Ersetzen von Zeile i, $i \in \underline{m}$, durch die Summe von Zeile i und Zeile $j \in \underline{m}$

  aus. Man spricht bei den obigen drei Umformungstypen von den drei <u>elementaren Zeilenumformungen</u> zur Lösung von LGS.

- Ist das LGS lösbar, so liefert die normierte Zeilenstufenform nach Festlegung der Nicht-Pivotvariablen als Parameter gerade r Bedingungen an diejenigen r Variablen x_k, deren Index als Pivotindex der Matrix $\hat{A}$ vorkommt. Offensichtlich lässt sich keine der r Bedingungen einfach weglassen, da dann einfach die Bedingung an die entsprechende Variable wegfallen würde, x_k wäre also frei wählbar; die Bedingungen sind also etwas wie ein minimaler Satz an Bedingungen.

- Nicht ganz offensichtlich (aber richtig) ist, dass es auch durch andere Äquivalenzumformungen nicht möglich ist, mehr der Bedingungen wegzulassen. Das heißt, dass der Gauß-Jordan-Algorithmus eine minimale Anzahl an linearen Bedingungen liefert, mit denen sich die Lösungsmenge beschreiben lässt. Das werden wir im Kapitel IX genauer sehen.

Um Lösbarkeitsaussagen für LGS zu treffen, ist der *Rang* einer Matrix der zentrale Begriff. Eine Möglichkeit, ihn festzulegen, ist die folgende. Andere Möglichkeiten, den Rang äquivalent zu definieren und zu bestimmen, kommen ebenfalls in Kapitel IX.

Definition 6.3.1. *(Rang einer Matrix, erste Charakterisierung)*

Sei $A \in \mathbb{K}^{m\times n}$, $b \in \mathbb{K}^m$ und $(\hat{A} \mid \hat{b})$ die Ausgabe nach Durchführung des Gauß-Jordan-Algorithmus mit Eingabe $(A \mid b)$. Als Rang von A, kurz rg(A), bezeichnen wir die Anzahl der Nicht-Null-Zeilen in der Matrix $\hat{A}$, also die Anzahl von Zeilen mit Pivotindizes $p(i) < n+1$. Der Rang einer erweiterten Koeffizientenmatrix $(A \mid b)$ ist entsprechend die Anzahl der Nicht-Null-Zeilen der Matrix $(\hat{A} \mid \hat{b})$, also die Anzahl von Zeilen mit Pivotindizes $p(i) < n+2$.

Bemerkung: Eigenschaften des Ranges

- Für $A \in \mathbb{K}^{m\times n}$ ist $rg(A)$ durch die Definition oben eindeutig festgelegt. Das liegt daran, dass der Gauß-Jordan-Algorithmus ist in der hier gegebenen Form deterministisch ist, d.h. bei Eingabe einer erweiterten Koeffizientenmatrix $(A \mid b)$ gibt es genau eine mögliche Ausgabe.
- Es gilt $\mathrm{rg}(A) \leq m$ und $\mathrm{rg}(A) \leq n$. ($\mathrm{rg}(A) \leq m$ gilt, da es höchstens m Zeilen gibt; $\mathrm{rg}(A) \leq n$ ergibt sich, da die Matrix $\hat{A}$ in normierter Zeilenstufenform ist, also spätestens ab Zeile $n+1$ nur noch Nullzeilen erlaubt sind.)

Mit Hilfe des Ranges lassen sich allgemeiner Aussagen über die Lösbarkeit von linearen Gleichungssystemen mit Systemmatrix A *unabhängig von der genauen Gestalt der rechten Seite* b machen.

Satz 6.3.2. *(Rang und Lösbarkeit)*

Ein gegebenes Gleichungssystem mit erweiterter Koeffizientenmatrix $(A \mid b)$, $b \in \mathbb{K}^m$, ist lösbar genau dann, wenn $rg(A) = rg(A \mid b)$ gilt.

Beweis: Das LSG $A \cdot x = b$ ist genau dann nicht lösbar, wenn in der Ausgabe des Gauß-Jordan-Algorithmus Fall 1 eintritt, d.h wenn in der Ausgabe $(\hat{A} \mid \hat{b})$ eine oder mehrere Zeilen vorkommen, bei der die ersten n Einträge Null sind, der $(n+1)$-te Eintrag aber nicht Null ist. Das passiert genau dann, wenn die erweiterte Matrix mindestens eine Nicht-Nullzeile mehr hat als die Matrix $\hat{A}$.

□

Wir präzisieren die Lösbarkeitsaussage des vorangegangenen Satzes weiter zu folgendem wichtigen Satz.

Satz 6.3.3. *(Rangkriterium für Lösbarkeit von LGS mit Systemmatrix A)*

Sei $A \cdot x = b$ ein gegebenes LGS mit $A \in \mathbb{K}^{m \times n}$, d.h. mit m Zeilen und n Spalten. Dann gilt:

(i) Ist $rg(A) = m$, so existiert zu jedem *LGS der Form (6.3), d.h. für jede rechte Seite $b \in \mathbb{K}^m$, mindestens eine Lösung.*

(ii) Ist $rg(A) < m$, so gibt es rechte Seiten $b \in \mathbb{K}^m$, für die das LGS keine Lösung hat.

(iii) Ist ein LGS (6.3) lösbar, so ist die Anzahl der freien Parameter in der Lösungsmenge unabhängig von b gegeben durch $n - rg(A)$.

(iv) Eine Lösung eines LGS (wenn es sie gibt) ist eindeutig genau dann, wenn $rg(A) = n$ ist.

(v) Jedes LGS der Form (6.3) (A gegeben, b beliebig) hat genau dann eine eindeutige Lösung, wenn A quadratisch ist und vollen Rang hat, d.h. es gilt $n = m = rg(A)$.

Beweis: Zu (i): Ist $\text{rg}(A) = m$, so kann $\text{rg}(A \mid b)$ nicht größer als m sein, da nicht mehr Zeilen zur Verfügung stehen. Nach 6.3.2 gibt es also eine Lösung.

Zu (ii) Wir betrachten das LGS mit Systemmatrix A, das nach der Umformung durch den Gauß-Jordan-Algorithmus auf die Form $(\hat{A} \mid \hat{b})$ mit $\hat{b} = (0, \ldots, 0, 1)$ gebracht wurde. Das System $(\hat{A} \mid \hat{b})$ ist wegen $\text{rg}(\hat{A}) < \text{rg}(A \mid b)$ nicht lösbar nach 6.3.2(ii). Da alle im Gauß-Jordan-Algorithmus durchgeführten Umformungen Äquivalenzumformungen waren, sind diese umkehrbar – machen wir diese in umgekehrter Reihenfolge wieder rückgängig, so erhalten wir als Systemmatrix A zurück, und als rechte Seite einen Vektor b, für den das LGS $(A \mid b)$ nicht lösbar ist (denn die Lösungsmenge ist ja die von $(\hat{A} \mid \hat{b})$).

Zu (iii): Die Indizes der freien Parameter der Lösungsmenge sind die Spaltenindizes, die nicht zu den Pivotspalten gehören, das sind $n - rg(A)$ Stück.

Ad (iv): Eine Lösung ist eindeutig genau dann, wenn keiner der Parameter frei wählbar ist; das passiert genau im Fall $\text{rg}(A) = n$.

(v) ergibt sich aus den Genau-dann-wenn-Bedingungen „Das LGS ist für jedes $b \in \mathbb{K}^m$ lösbar, genau dann, wenn rg(A) = m“ ((i) und (ii)) und „Jede Lösung (wenn es sie gibt) ist eindeutig genau dann, wenn $rg(A) = n$ ist.“ (iv). $\square$

6.4 „Zeilensicht“, „Abbildungssicht“ und „Spaltensicht“ auf LGS

Wir betrachten das LGS (6.3) noch einmal systematisch.

- Wir haben es bisher als m einzelne Bedingungen $\langle x, z_i\rangle = b_i$ interpretiert. Das ist die „Zeilensicht“ auf ein LGS, es entspricht der Frage, welche Vektoren $x \in \mathbb{R}^n$ die gegebenen m Bedingungen Zeile für Zeile erfüllen.)

- Die „Abbildungssicht“ auf ein LGS ergibt sich, wenn wir die linke und die rechte Seite des LGS als eine Gleichung für einen Vektor x aus dem $\mathbb{K}^n$ lesen. Bezeichnen wir die Zeilen von A wieder mit $z_1, \ldots, z_n \in \mathbb{K}^n$ und fassen wir $b_1, \ldots, b_m$ zu einem Vektor $b \in \mathbb{K}^m$ zusammen, so erhalten wir

$$\left\{\begin{array}{rcl} a_{1,1}\cdot x_1 + \ldots + a_{1,n}\cdot x_n &=& b_1\\ a_{2,1}\cdot x_1 + \ldots + a_{2,n}\cdot x_n &=& b_2\\ &\vdots& \\ a_{m,1}\cdot x_1 \ldots + a_{m,n}\cdot x_n &=& b_m \end{array}\right. \iff \begin{pmatrix} \langle z_1, x\rangle \\ \langle z_2, x\rangle \\ \vdots \\ \langle z_m, x\rangle \end{pmatrix} = \begin{pmatrix} b_1\\ b_2\\ \vdots \\ b_m\end{pmatrix} \iff A\cdot x = b$$

  Die Frage nach der Lösbarkeit eines linearen Gleichungssystems lässt sich in dieser Sicht interpretieren als die Frage „Gibt es ein $x \in \mathbb{K}^n$, das durch die durch $A \in \mathbb{K}^{m\times n}$ definierte Matrixabbildung auf $b \in \mathbb{K}^m$ abgebildet wird?“ Das ist wieder eine Frage nach der Lösbarkeit einer Gleichung $f(x) = b$ wie in Kapitel 2.3, die sich auf die Frage nach der Injektivität, Surjektivität bzw. Bijektivität der zugehörigen Matrixabbildung zurückführen lässt. Wir werden diese Fragen im übernächsten Kapitel allgemein für lineare Abbildungen untersuchen. Oft wird die Abbildungssicht auch genutzt, um ein LGS knapp in der Form „Sei $A\cdot x = b$ ein gegebenes LGS, dann...“ zu notieren.

- Eng damit verknüpft ist die „Spaltensicht“ auf ein LGS, die sich ergibt, wenn wir die einzelnen Spalten von A als Vektoren schreiben. Dann erhalten wir

$$\left\{\begin{array}{rcl} a_{1,1}\cdot x_1 + \ldots + a_{1,n}\cdot x_n &=& b_1\\ a_{2,1}\cdot x_1 + \ldots + a_{2,n}\cdot x_n &=& b_2\\ &\vdots& \\ a_{m,1}\cdot x_1 \ldots + a_{m,n}\cdot x_n &=& b_m \end{array}\right. \iff x_1\cdot \begin{pmatrix} a_{1,1}\\ a_{2,1}\\ \vdots \\ a_{m,1}\end{pmatrix} + \ldots + x_n\cdot \begin{pmatrix} a_{1,n}\\ a_{2,n}\\ \vdots \\ a_{m,n}\end{pmatrix} = \begin{pmatrix} b_1\\ b_2\\ \vdots \\ b_m\end{pmatrix}$$

Hier ergibt sich die in der Vektorraumtheorie (s. nächstes Kapitel) angesiedelte Fragestellung, ob bzw. unter welchen Umständen sich der Vektor b als Linearkombination der Spaltenvektoren $s_1, \ldots, s_n$ von A schreiben lässt, d.h. ob man Vorfaktoren finden kann, so dass die obige Vektorgleichung stimmt.

Ein flexibles Wechseln zwischen allen drei dargestellten Sichtweisen ist das, was die Lineare Algebra tut, und somit zentral für das Verständnis der Linearen Algebra!

Wir halten solch einen „sichtweisenüberschreitenden“ Zusammenhang fest und fassen damit einige wesentliche Erkenntnisse über Lineare Gleichungssysteme, Matrixabbildungen und Matrizen zusammen.

Satz 6.4.1. *(Invertierbarkeit von Matrizen und Lösbarkeit von LGS)*

Sei $f : \mathbb{K}^n \to \mathbb{K}^m, x \mapsto A \cdot x$ eine Matrixabbildung. Dann sind äquivalent:

(i) *Jede Vektorgleichung der Form $A \cdot x = b$ mit gegebener rechter Seite $b \in \mathbb{K}^m$ hat genau eine Lösung $x \in \mathbb{K}^n$,*

(ii) *Es ist $n = m$ (d.h. A ist quadratisch) und A hat vollen Rang, d.h. es ist $rg(A) = n$,*

(iii) *A hat eine Inverse A^{-1}, d.h. $A^{-1} \cdot A = I_n$ und $A \cdot A^{-1} = I_m$,*

(iv) *$f : \mathbb{K}^n \to \mathbb{K}^m, x \mapsto A \cdot x$ ist bijektiv.*

(v) *$f : \mathbb{K}^n \to \mathbb{K}^m, x \mapsto A \cdot x$ besitzt eine Umkehrabbildung.*

Der Satz 6.4.1 sagt insbesondere auch aus, dass es für nicht-quadratische Matrizen $A \in \mathbb{K}^{n\times m}$ mit $n \neq m$ *immer* lineare Gleichungssysteme $A \cdot x = b$ mit $b \in \mathbb{K}^m$ gibt, die entweder nicht oder nicht eindeutig lösbar sind.

Beweis: Die Äquivalenz (i)$\Leftrightarrow$(iv)$\Leftrightarrow$(v) ist die Aussage von Satz 2.5.7. Auch die Äquivalenz (i)$\Leftrightarrow$(ii) haben wir bereits in 6.3.3(v) bewiesen. Die Äquivalenz zwischen (iii) und (iv) ist gerade die Aussage von Lemma 4.4.4 gewesen (die Umkehrabbildung ist nach der Aussage dort dann gerade $f^{-1} : \mathbb{K}^n \to \mathbb{K}^m, x \mapsto A^{-1} \cdot x$). Damit sind alle Behauptungen bereits bewiesen; es ist aber trotzdem schön, den Überblick über die Zusammenhänge zu bekommen :)

$\square$

6.5 Berechnung inverser Matrizen

Der Gauß-Algorithmus liefert auch ein einfaches systematisches Verfahren zur Berechnung der Inversen einer invertierbaren Matrizen.

Lemma 6.5.1. *(Invertierbarkeit bei vollem Rang)*

Sei $A \in \mathbb{K}^{n\times n}$ eine Matrix mit vollem Rang. Bezeichnet $e_1, \ldots, e_n$ die Standardbasis[1] von $\mathbb{K}^n$, so sind die die Spalten $s_1, \ldots, s_n$ der (nach Satz 6.4.1 existierenden) Inversen A^{-1} von A gerade die Lösungen der n linearen Gleichungssysteme $A \cdot s_i = e_i$.

Beweis: Die nach Satz 6.4.1 existierende Inverse A^{-1} ist eindeutig. Sie ist insbesondere Rechtsinverse von A, d.h. es gilt $A \cdot A^{-1} = I_n$. Die i-te Spalte von $A \cdot A^{-1}$ ist also e_i. Bezeichnen wir mit $s_1, \ldots, s_n$ die Spalten der Matrix A^{-1}, so ist nach Definition der Matrix-Matrix-Multiplikation die i-te Spalte von $A \cdot A^{-1}$ aber auch $A \cdot s_i$, es gilt also $A \cdot s_i = e_i$, und da wegen des vollen Ranges von A alle LGS $A \cdot x = b$ eindeutig lösbar sind, sind damit die Spalten von A^{-1} (und damit A^{-1}) eindeutig bestimmt.

$\square$

[1] Wie in $\mathbb{R}^n$ bezeichnen wir auch in $\mathbb{K}^n$ die Vektoren

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

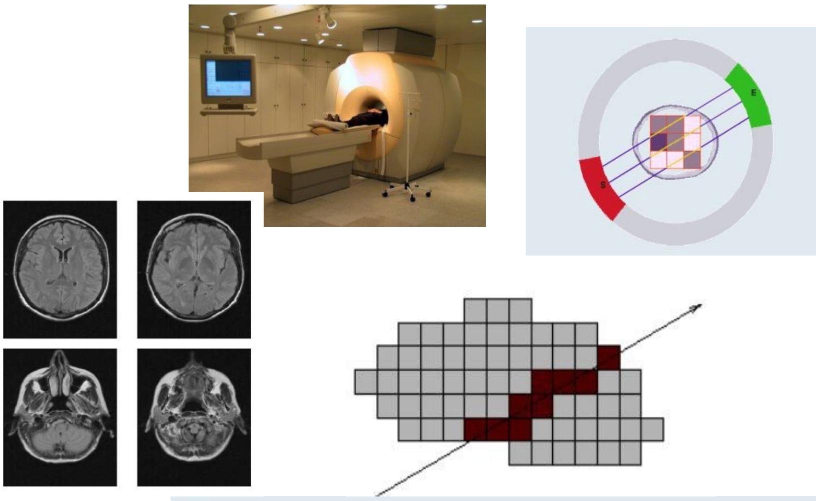
als Standardbasis von $\mathbb{K}^n$. Im kommenden Kapitel wird die Begrifflichkeit klarer werden.

Praktisches Verfahren zur Berechnung von inversen Matrizen:

Um aus einer quadratischen Matrix $A \in \mathbb{K}^{n \times n}$ mit vollem Rang ihre Inverse A^{-1} zu berechnen, müssen nach dem Lemma oben die n Gleichungssysteme $A \cdot s_i = e_i$, $i \in \underline{n}$, gelöst werden. Eine pfiffige Art und Weise, das zu tun, ist die Anwendung des Gauß-Jordan-Algorithmus auf die Koeffizientenmetrix $(A \mid I_n) \in \mathbb{K}^{n \times (2n)}$: Ist A invertierbar, so kann A auf normierte Zeilenstufenform mit vollem Rang gebracht werden, d.h. auf der linken Seite von $(A \mid I_n)$ ergibt sich die Einheitsmatrix, und insgesamt eine $(n \times 2n)$)-Matrix $(I_n \mid B)$. In den Spalten der Matrix B stehen nach Konstruktion des Gauß-Algorithmus nun die jeweiligen Lösungen des LGS $A \cdot s_i = e_i$, B ist also nach dem vorangegangenen Lemma die Inverse von A.

- **Beispiel** s. Vorlesung.
- Es scheint, als sei das Verfahren (der „Algorithmus“) ist auch für nicht-invertierbare quadratische Matrizen anwendbar – es ist allerdings dann nicht möglich, auf der linken Seite die Form der Einheitsmatrix zu erreichen. (Das liegt an Satz 11.1.3.)

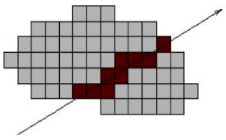
Beispiel: Ein großes LGS



"Tomographenformel" $\quad ln\, I(0) - ln\, I(d) = \mu_1 \cdot d_1 + \mu_2 \cdot d_2 + \ldots + \mu_n \cdot d_n$

Beispiel: Ein großes LGS

pro „Scheibe“:



"Tomographenformel" $\quad ln\, I(0) - ln\, I(d) = \mu_1 \cdot d_1 + \mu_2 \cdot d_2 + \ldots + \mu_n \cdot d_n$

Annahme: Auflösung 512×512 Pixel.
Lineares Gleichungssystem $Ax = b$ mit

- Einträge μ_i des gesuchten Vektors x:
  Absorptionskoeffizient eines Rasterquadrats
  ($\widehat{=}$ Grauwert des Pixels)
  $n = (512)^2 = 262.144$ Pixel $\widehat{\approx}$ 2 MB in *double*-Format.
- rechte Seite b ergibt sich aus Messungen (pro Strahl ankommende Intensität)
- Matrix A aus Geometrie des Querschnitts:
  $n^2 = (512)^4 \approx 68 \cdot 10^9$ Einträge, $\widehat{\approx}$ 550 GB

6.6 Aufgaben zu Kapitel 6

Für die linearen Gleichungssysteme in diesen Aufgaben (und denen in den folgenden Kapiteln) bietet es sich an, nach beendeter Rechnung eine Probe für Ihre Ergebnisse zu machen. Sie können dafür beispielsweise Online-Tools wie dieses oder dieses nutzen.

Aufgabe 6.28. (Lösung linearer Gleichungssysteme):

(a) Bestimmen Sie mit Hilfe des Gauß-Jordan-Algorithmus die Lösungsmenge des folgenden linearen Gleichungssystems für $x = (x_1, x_2, x_3, x_4) \in \mathbb{R}^4$.

$$\begin{array}{rcrcrcrcr} 4x_1 & - & x_2 & + & x_3 & - & x_4 & = & 1 \\ 4x_1 & + & 3x_2 & - & 3x_3 & + & 3x_4 & = & -1 \\ 7x_1 & + & x_2 & - & x_3 & + & x_4 & = & 0 \end{array}$$

(b) Bestimmen Sie alle Polynome dritten Grades

$$p = a_3x^3 + a_2x^2 + a_1x + a_0 \in \mathbb{R}[x],$$

welche die Eigenschaft

$$p(-2) = -16, \quad p(1) = -1, \quad p(4) = 50, \quad p(-3) = -41$$

besitzen.

Aufgabe 6.29. (Ein Kriterium für die Lösbarkeit von (2×2)-Systemen):

Gegeben sei ein lineares Gleichungssystem der Form

$$\begin{aligned} a_{1,1}\,x + a_{1,2}\,y &= b_1 \\ a_{2,1}\,x + a_{2,2}\,y &= b_2 \end{aligned}$$

für $(x, y) \in \mathbb{K}^2$ (wobei $\mathbb{K}$ ein Körper ist). Beweisen Sie, dass ein solches Gleichungssystem *genau dann* eindeutig lösbar ist (d.h. genau eine Lösung (x, y) besitzt), wenn für die Koeffizienten der Systemmatrix

$$a_{1,1}a_{2,2} - a_{1,2}a_{2,1} \neq 0$$

gilt. Begründen Sie sauber und beachten Sie alle möglichen auftretenden Fälle, z.B. mit Hilfe einer geeigneten Fallunterscheidung für die Matrixeinträge $a_{i,j}$.

Aufgabe 6.30. (Eine Matrix eines LGS über $\mathbb{C}$):

(a) Bestimmen Sie den Rang der Matrix

$$A = \begin{pmatrix} 1 & 2i & 3i+1 \\ 2i & -4 & 2i-6 \\ 3i+1 & 2i-6 & -8+6i \end{pmatrix} \in \mathbb{C}^{3\times 3}.$$

(b) Entscheiden Sie mit Hilfe Ihres Ergebnisses aus (a), ob lineare Gleichungssysteme der Form

$$\begin{array}{ccccccc} x_1 & + & (2i)\cdot x_2 & + & (3i+1)\cdot x_3 & = & b_1 \\ (2i)\cdot x_1 & - & 4\cdot x_2 & + & (2i-6)\cdot x_3 & = & b_2 \\ (3i+1)\cdot x_1 & + & (2i-6)\cdot x_2 & + & (-8+6i)\cdot x_3 & = & b_3 \end{array}$$

mit gesuchtem Vektor $x = (x_1, x_2, x_3) \in \mathbb{C}^3$ für beliebige rechte Seiten $b = (b_1, b_2, b_3) \in \mathbb{C}^3$ immer eindeutig lösbar sind. Begründen Sie Ihre Antwort.

Aufgabe 6.31. (Gleichungssysteme mit Parametern):

(a) Für welche $a, b \in \mathbb{R}$ hat das lineare Gleichungssystem

$$\begin{aligned} ax + y &= 0 \\ x + by &= 0 \end{aligned}$$

für $(x, y) \in \mathbb{R}^2$

(i) genau eine Lösung,

(ii) keine Lösung,

(iii) unendlich viele Lösungen?

Hinweis: Finden Sie durch eine geeignete Fallunterscheidung alle Möglichkeiten.

(b) Für welche der folgenden Werte von a, b, c ist das lineare Gleichungssystem

$$\begin{array}{rcrcrcl} x & + & y & + & 2z & = & a \\ 2x & - & y & - & 5z & = & b \\ x & & & - & z & = & c \end{array}$$

für $(x, y, z) \in \mathbb{R}^3$ lösbar? Geben Sie jeweils die Lösungsmenge an.

(i) $a = 1, b = -1, c = 0$.

(ii) $a = 0, b = 1, c = 2$.

(iii) $a = \lambda + 2, b = 2\lambda - 2, c = \lambda \quad$ (mit $\lambda \in \mathbb{R}$).

Tipp: Wenden Sie zuerst den Gauß-Jordan-Algorithmus auf das obige Gleichungssystem mit allgemeinen Parametern a, b und c an, um die für die Lösbarkeit notwendigen Beziehungen zwischen den Parametern festzustellen. Überprüfen Sie dann, ob die speziellen Werte in (i)–(iii) diese Beziehungen erfüllen.

Aufgabe 6.32. (Berechnung inverser Matrizen):

Berechnen Sie die Inverse der folgenden Matrix $A \in \mathbb{Q}^{4\times 4}$ mit dem Gauß-Jordan-Verfahren.

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

Aufgabe 6.33. (Eigenschaften linearer Gleichungssysteme):

Sei $A \in \mathbb{K}^{m \times n}$. Das lineare Gleichungssystem $A \cdot x = 0$ mit dem Nullvektor 0 des $\mathbb{K}^m$ als rechte Seite heißt das zu A gehörige homogene System.

(a) Sei $b \in \mathbb{K}^m$ und $x^* \in \mathbb{K}^n$ eine Lösung des LGS $A \cdot x = b$. Zeigen Sie: Ein Vektor $y \in \mathbb{K}^n$ ist eine Lösung des homogenen Systems $A \cdot x = 0$ genau dann, wenn $x^* + y$ das System $A \cdot x = b$ löst.

(b) Bestimmen Sie für das folgende Gleichungssystem für $(x_1, x_2, x_3, x_4) \in \mathbb{R}^4$ seine Lösungsmenge sowie die Lösungsmenge des homogenen Systems.

$$\begin{array}{rcrcrcrcl} 2\,x_1 & + & x_2 & - & x_3 & + & 5\,x_4 & = & 2 \\ & - & 2\,x_2 & & & + & 3\,x_4 & = & 2 \\ 2\,x_1 & + & 3\,x_2 & - & x_3 & + & 3\,x_4 & = & 1 \end{array}$$

(c) Sei $A \in \mathbb{K}^{m \times n}$. Beweisen Sie:

(i) Die Lösungsmenge

$$\mathbb{L}(A, 0) = \{x \in \mathbb{K}^n \mid A \cdot x = 0\}$$

des zu A gehörigen homogenen Systems ist ein Unterraum des $\mathbb{K}^n$.

(ii) Für $b \in \mathbb{K}^m$ mit $b \neq 0$ ist die Lösungsmenge des Systems $A \cdot x = b$ nie ein Unterraum des $\mathbb{K}^n$.

7 Grundlagen der Mathematik: Logik und Mengen

Wir sind nun fachlich und historisch (wenn auch nicht ganz chronologisch) so weit vorgedrungen, dass die für die Lineare Algebra zentrale Vektorraumtheorie zum Greifen nahe liegt. Zu einer Einführung in die Lineare Algebra, die normalerweise im ersten Semester eines Mathematikstudiums stattfindet, gehört neben einer vorgeschalteten Einführung in die Grundbegriffe der Algebra wie bis zu dieser Stelle dieses Skripts immer auch eine Einführung in die Grundlagen logischer Schlussweisen und der Mengenlehre. Warum das so ist, beleuchtet der historische Einstieg zum Kapitel; dieser versucht auch zu begründen, warum die Mathematik so eine formale, genaue Wissenschaft (geworden) ist. Der Rest des Kpitels gibt einen Überblick über die wichtigsten Grundlagen der Mengenlehre und Logik, insbesondere entwirft es einen Katalog von gängigen Aussagetypen, Schlussweisen und Beweistechniken und von mengentheoretischen Strukturen, die im „täglichen Leben“ einer Mathematikerin häufig vorkommen.

7.1 Die Grundlagenkrise der Mathematik

- Der Funktionsbegriff ist zu Beginnzeiten der Analysis (Newton, Leibniz,. . . , ca. spätes 17. Jhd.) ist anschauungsbasiert und dadurch „unscharf“.

- Ebenso ist der Stand der Dinge bei Grundbegriffen der Geometrie (z.B. „Punkt“ und „Ebene“ immer noch so „definiert“ wie bei Euklid, s. Einleitung) und großen Teilen der übrigen mathematischen Begriffsbildung.

- Im „formalen Zeitalter der Analysis“ ($\sim$ 19. Jhd.) beweisen große Mathematiker wie Auguste Cauchy daher auf Grundlage unscharfer Begriffe und Zusammenhänge Aussagen, für die wenig später Gegenbeispiele gefunden werden. . . siehr Folien.

- Durch Fundierung der Definition des Funktionsbegriffs auf Cantors Mengenbegriff (Felix Hausdorff, 1914, s. Def. 7.6.3 unten) werden diese Probleme weitestgehend gelöst.

- Der Mengenbegriff ist leider zu dieser Zeit genauso unscharf, z.B. führt die Betrachtung der „Menge aller Mengen“ (schon durch Cantor selbst) zu Widersprüchen. Eine prominentere Demonstration der logisch-formalen Ungereimheiten ist die mit dem damaligen Mengenverständnis mögliche Bildung der „Menge aller Mengen, die sich nicht selbst enthalten“. Die Frage, ob diese Menge sich selbst enthält, führt auf einen Widerspruch, egal ob das der Fall ist oder nicht. Das ist die so genannte Russell’sche Antinomie. Der namensgebende Philosoph und mathematische Grundlagenforscher

> Bertrand Russell fasst die Grundstimmung der Zeit (ca. Ende 19. Jhd.) folgendermaßen zusammen:
>
> „Seit man begonnen hat, die einfachsten Behauptungen zu beweisen, erwiesen sich viele von ihnen als falsch.“

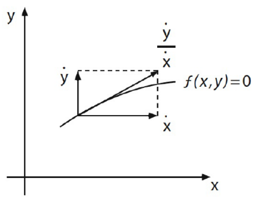
Einen Ausweg aus dieser *Grundlagenkrise der Mathematik* skizzierte der Göttinger Mathematikstar David Hilbert in den 1920ern im so genannten Hilbert-Programm: Der grundlegende Gedanke ist, dass es in der Mathematik genau möglich sein sollte, festzulegen

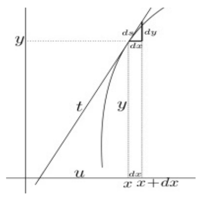
- welche Mengen erlaubt sind und
- welche logischen Schlüsse erlaubt sind,

so dass Begriffsdefinitionen prinzipiell nur auf Grundlage mengentheoretischer und aussagenlogischer Formulierungen formuliert werden dürfen (keine intuitiven Definitionen mehr!) und mathematische Sätze sich immer durch eine Hintereinanderreihung erlaubter logischer Schlüsse aus den Axiomen ableiten lassen. Bei Formulierung von Definitionen und Axiomsystemen sollte zudem Minimalismus ein führendes Prinzip sein – man sollte nur voraussetzen, was wirklich notwendig ist! (Das ist heute spätestens seit den Werken der Bourbaki-Gruppe, die Hilberts Forderungen um Anlass nahm, die Mathematik in eine neue Dimension des Formalismus zu katapultieren, immer noch ein „Grundmuster“ der Mathematik.)

Wenn die klare Fassung der mengentheoretischen und logischen Grundlagen erledigt ist, so die Hoffnung Hilberts und seiner Zeitgenossen, können auch diese Grundlagen der Mathematik mit mathematischen Methoden untersucht werden, um zu zeigen, das das so errichtete Gebäude widerspruchsfrei ist (d.h. es lassen sich darin keine widersprüchlichen Aussagen beweisen) und vollständig ist (d.h. alles was unter Voraussetzung der Axiome wahr ist, lässt sich beweisen). Das hat zwar nicht wirklich geklappt, s. die Bemerkungen zu den Gödelschen Unvollständigkeitssätzen am Ende von Abschnitt 7.8. Trotzdem sind die im Folgenden dargestellten Grundlagen aus der Logik und Mengenlehre die „Grundpfeiler“, die den Begriffsbildungen und Argumentationsmustern der heutigen Mathematik explizit oder implizit zugrunde liegen.

Wo der ganze Trubel anfängt: Die Geburt der Analysis (Ende 17. Jhd)





Isaac Newton

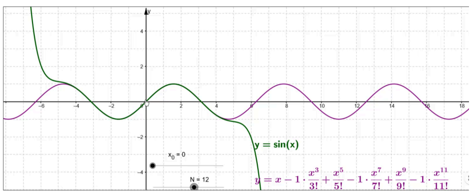
Gottfried Wilhelm Leibniz

- Leibniz und Newton erschaffen die Infinetisemalrechnung und die Grundlagen der Analysis auf Grundlage geometrischer/physikalischer Vorstellungen

---

Arbeit mit einem anschaulichen Konvergenzbegriff

- Weiterentwicklung der Erkenntniswelt der Analysis:
  - Ampere beweist 1806, dass stetige Funktionen fast überall differenzierbar sind.
  - Cauchy beweist 1821: Der Grenzwert einer Folge stetiger Funktionen ist stetig,

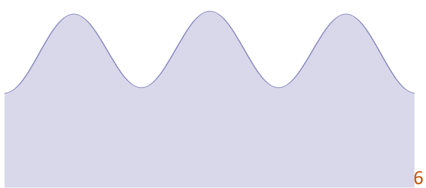


$$\begin{aligned} e^x &= 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \frac{1}{5!}x^5 + \frac{1}{6!}x^6 + \dots \\ \sin(x) &= x - \frac{1}{3!}x + \frac{1}{5!}x^5 + \frac{1}{7!}x^7 + \frac{1}{9!}x^9 + \dots \\ \cos(x) &= 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \frac{1}{6!}x^6 + \frac{1}{8!}x^8 + \dots \end{aligned}$$

4

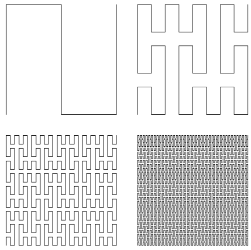
Arbeit mit einem anschaulichen Konvergenzbegriff

- Weiterentwicklung der Erkenntniswelt der Analysis:
  - ~~Ampere beweist 1806, dass stetige Funktionen fast überall differenzierbar sind.~~
  - ~~Cauchy beweist 1821: Der Grenzwert einer Folge stetiger Funktionen ist stetig~~,
- 1834: Bolzano konstruiert eine stetige Funktion, die nirgends differenzierbar ist,
- 1872 Weierstraß publiziert noch so ein Monster...
- Monstersperre? (Niels Henrik Abel)



6

Das „formale Zeitalter der Analysis“

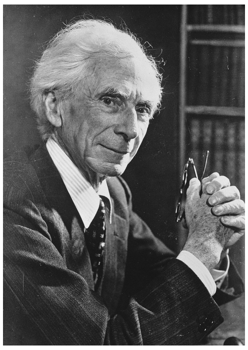


- Anpassung der Definition des Funktionsbegriffs
- Herausbildung vieler verschiedener Konvergenzbegriffe

- „Ist in jedem Intervalle jedem einzelnen Werte x durch irgendwelche Mittel ein bestimmter Wert y zugeordnet, dann soll y eine Funktion von x heißen.“ *(Peter Dirichlet, Anf. 19. Jhd.)*
- „Unter einer Menge verstehen wir jede Zusammenfassung M von bestimmten wohlunterschiedenen Objekten (m) unserer Anschauung oder unseres Denkens.“ *(Georg Cantor, Ende 19. Jhd.)*

8

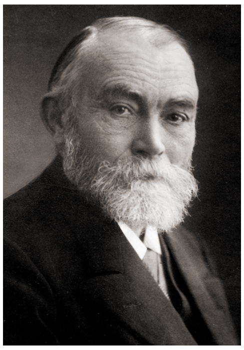
Grundlagenforschung, Logik und Formalismus (19. Jhd.)



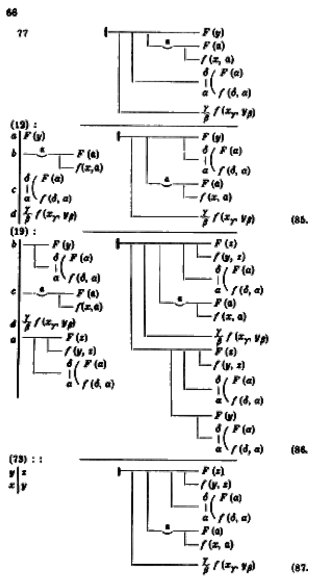
Bertrant Russell (19. Jhd.)
„Principia Mathematica“

Aristoteles (re.)

Gottlob Frege
(Mitte-Ende 19. Jhd.)

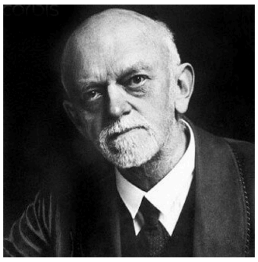


Auguste deMorgan
(Anfang – Mitte 19. Jhd.)



12

Hilberts Exempel: Neu-Axiomatisierung der Geometrie

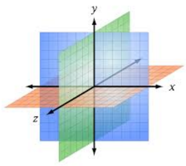


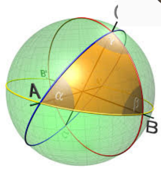
David Hilbert (Anf. 20. Jhd.)

Probleme mit anschauungsbasierten Begriffsdefinitionen:

- „Ein Punkt ist, was keine Theile hat.“ (Euklid)

- „drei verschiedene Systeme von Dingen“ Punkte, Geraden und Ebenen, und „drei grundlegende Beziehungen“ liegen, zwischen und kongruent.
- „Man muss jederzeit an Stelle von Punkten, Geraden, Ebenen, Tische, Stühle, Bierseidel sagen können.“





Formales Beweisen in Theorien

Axiome der Theorie (hier: Gruppentheorie)

$$\exists e\ \forall g : e * a = a * e = a$$

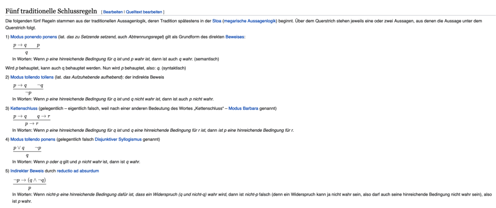
$$\forall g \exists h : g * h = h * g = e$$

$$\forall g \forall h \forall k : g * (h * k) = (g * h) * k$$

(im strikten Sinn Formulierung als abstrakte Zeichenkette OHNE JEDE Anbindung zur Wirklichkeit mehr)

+

Katalog formal-logischer Schlussregeln („Ableitungskalkül)



Fünf traditionelle Schlussregeln [Bearbeiten | Quelltext bearbeiten]

Die folgenden fünf Regeln stammen aus der traditionellen Aussagenlogik, deren Tradition spätestens in der Stoa (megarische Aussagenlogik) beginnt. Über dem Querstrich stehen jeweils eine oder zwei Aussagen, aus denen die Aussage unter dem Querstrich folgt.

1) Modus ponendo ponens (lat. *das zu Setzende setzend*, auch *Abtrennungsregel*) gilt als Grundform des direkten Beweises:

$$\frac{p \to q \qquad p}{q}$$

In Worten: Wenn *p eine hinreichende Bedingung für q ist* und *p wahr ist*, dann ist auch *q wahr*. (semantisch)

Wird p behauptet, kann auch q behauptet werden. Nun wird p behauptet, also: q. (syntaktisch)

2) Modus tollendo tollens (lat. *das Aufzuhebende aufhebend*): der indirekte Beweis

$$\frac{p \to q \qquad \neg q}{\neg p}$$

In Worten: Wenn *p eine hinreichende Bedingung für q ist* und *q nicht wahr ist*, dann ist auch *p nicht wahr*.

3) Kettenschluss (gelegentlich – eigentlich falsch, weil nach einer anderen Bedeutung des Wortes „Kettenschluss“ – Modus Barbara genannt)

$$\frac{p \to q \qquad q \to r}{p \to r}$$

In Worten: Wenn *p eine hinreichende Bedingung für q ist* und *q eine hinreichende Bedingung für r ist*, dann *ist p eine hinreichende Bedingung für r*.

4) Modus tollendo ponens (gelegentlich falsch Disjunktiver Syllogismus genannt)

$$\frac{p \vee q \qquad \neg p}{q}$$

In Worten: Wenn *p oder q* gilt und *p nicht wahr* ist, dann ist *q wahr*.

5) Indirekter Beweis durch reductio ad absurdum

$$\frac{\neg p \to (q \wedge \neg q)}{p}$$

In Worten: Wenn *nicht-p eine hinreichende Bedingung dafür ist, dass ein Widerspruch (q und nicht-q) wahr wird*, dann ist *nicht-p* falsch (denn ein Widerspruch kann ja nicht wahr sein, also darf auch seine hinreichende Bedingung nicht wahr sein), also ist p wahr.

Lassen sich so alle Aussagen, die man über G formulieren kann, entscheiden?

$$\forall g \exists! h : g * h = h * g = e$$

$$\forall a \forall g \forall h : g = h \Leftrightarrow a * g = a * h$$

Lassen sich alle in jeder Gruppe wahren Aussagen so als Zeichenketten ableiten?

7.2 Mengenlehre nach Zermelo und Fraenkel

- Heute weithin als Rahmen der Mengenlehre akzeptiert ist das Axiomensystem von Ernst Zermelo und Abraham Adolf Frankel („ZF", 1907/1921), zu dem heute zusätzlich noch das so genannte Auswahlaxiom hinzugenommen wird ($\rightsquigarrow$ „ZFC", „C" für „axiom of choice").
- Es hat sich gezeigt – dies ist aber nur eine empirische Feststellung –, dass sich so gut wie alle bekannten mathematischen Sätze so formulieren lassen, dass sie sich als beweisbare Aussagen aus den Axiomen des ZFC ableiten lassen.

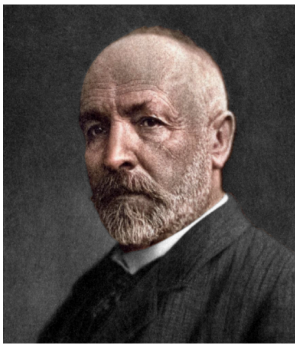
- „Unter einer Menge verstehen wir jede Zusammenfassung M von bestimmten wohlunterschiedenen Objekten (m) unserer Anschauung oder unseres Denkens."
(Georg Cantor, Ende 19. Jhd.)

Rückführung des Funktionsbegriffs auf den Mengenbegriff:

Bezeichnet

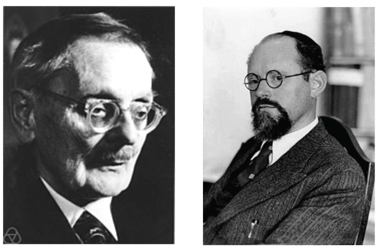
$$X \times Y := \{(x, y) \mid x \in X, y \in Y\}$$

die Menge aller möglichen Wertepaare in X und Y, so lässt sich eine Funktion immer als Teilmenge $f \subseteq X \times Y$ auffassen. So eine Teilmenge ist eine Funktion, wenn es zu es zu jedem $x \in X$ genau ein Pärchen $(x, y) \in f$ gibt.



(Georg Cantor, Ende 19. Jhd.)

$$\{(x, y) \in \mathbb{R} \times \mathbb{R} \mid y = x^2\}$$



Zermelo und Fraenkel

Felix Hausdorff, Anfang 20. Jhd., „Grundzüge der Mengenlehre"

Die genaue Formulierung von ZFC ist historisch gewachsen, recht abstrakt und technisch anspruchsvoll. Wir skizzieren diejenigen der sich daraus ergebenden Grundsätze, die für das Alltags-Mathematikerleben interessant sind.

- **Gleichheit von Mengen:** Mengen sind genau dann gleich, wenn sie dieselben Elemente enthalten. (Die Rückrichtung ist entscheidend, z.B. sind die Mengen

  $$\{1,2,3\}, \{3,2,1\}, \{1,1,2,1,3,2\}$$

  alle gleich.)

- **Erlaubte Konstruktionsschritte (Mengenoperationen):**[1]

  Sind X, Y Mengen, so sind

  - $X \cup Y := \{x \mid x \in X \text{ oder } x \in Y\}$ – die Vereinigung(smenge) von X und Y,
  - $X \cap Y := \{x \mid x \in X \text{ und } x \in Y\}$ – die Schnitt(smenge) von X und Y,
  - $X \setminus Y := \{x \mid x \in X \text{ und } x \notin Y\}$ – „X ohne Y", „Differenz(-menge)"

  Mengen.

  - Außerdem ist das kartesische Produkt $X \times Y$ von zwei Mengen X, Y eine Menge. Dies ist die Menge aller möglichen geordneten Paare von Elementen aus X und Y, also
    $$X \times Y := \{(x,y) \mid x \in X, y \in Y\}.$$
  - Aussonderung: Ist X eine Menge und $B(x)$ eine Bedingung[2] an $x \in X$, so ist
    $$\{x \in X \mid B(x)\}$$
    ebenfalls eine Menge.

[1] Im ZFC-„Katalog" sind mehr und z.T. andere Operationen erlaubt, die hier angegebenen Operationen lassen sich aber daraus ableiten.

[2] $B(x)$ haben wir Bedingung für $x \in X$ genannt, wenn $B(x)$ für jedes $x \in X$ einen eindeutigen Wahrheitswert hat, d.h. es gibt eine Funktion $w_B : X \to \{w, f\}$, die jedem $x \in X$ den Wahrheitswert von $B(x)$ zuordnet. In anderen Darstellungen werden Bedingungen auch als *Aussageformen* oder *einstellige Prädikate* bezeichnet, der Begriff der *Bedingung* ist meines Erachtens aber sehr treffend und griffig für den zu beschreibenden Sachverhalt.

- **Erlaubte Grundmengen:** Erlaubt sind in Zermelo-Fraenkel als Grundbausteine, auf die die obigen Operationen zum Konstruieren neuer Mengen angewendet werden, nur (!!)
  - die leere Menge $\emptyset$, die kein Element enthält (Schreibweise auch: $\{\}$; s.o., „Gleichheit“: Es gibt nur eine leere Menge.),
  - eine unendliche, rekursiv erzeugbare Menge.

Die Existenz dieser unendlichen, rekursiv erzeugbaren Menge erlaubt mit Hilfe der so genannten Peano-Axiome die Konstruktion der Menge der natürlichen Zahlen mit den zugehörigen Rechenoperationen $(+, \cdot)$ und Ordnungsrelation $(<)$ als algebraische Grundstruktur (Peano-Arithmetik). Die Existenz von $\mathbb{N}$ (und damit alles andere) basiert also auf einem *Axiom*, die natürlichen Zahlen werden quasi als gottgegeben vorausgesetzt.

Alle anderen Mengen, die in der Mathematik eine Rolle spielen (Zahlbereiche $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, mit jeweiligen Operationen, Funktionen auf solchen, $\mathbb{R}^n, \mathbb{K}[x], \mathrm{Abb}(X, Y), \ldots$), müssen hieraus *konstruiert werden können* („axiomatischer Aufbau der Zahlbereiche“). Das ist langwierig und umständlich, aber im letzten Jahrhundert weitestgehend erledigt worden. Eine schöne Darstellung findet man bei Oliver Deiser (Link).

> „Die ganzen Zahlen hat Gott gemacht, alles übrige ist Menschenwerk.“
>
> (Leopold Kronecker, 19. Jhd.),

7.3 Allaussagen und Existenzaussagen

Mathematische Erkenntnisse werden nach der im in den vergangenen beiden Abschnitten dargestellten Denke grundsätzlich in Form von (wahren) Aussagen über gewisse, klar definierte und nach dem ZMF-System konstruierbare Mengen bzw. ihre Elemente formuliert.[3] Simple Beispiele von Aussagen über Elemente von Mengen sind Ausdrücke „$2 < 5$“ oder „$\pi \notin \mathbb{Q}$“. Wenn man allerdings die Sätze in diesem Skript durch eine „logische Brille“ betrachtet, stellt man fest, dass so gut wie alle von ihnen so genannte Existenz- oder Allaussagen sind (bzw. sich im Sinne der formalen Logik äquivalent in diese Form bringen lassen). formulieren lassen. Daher nehmen wir diese diese als Startpunkt für die folgende Darstellung. Zunächst erinnern wir an ein paar Vokabeln, die schon in gesamten Skript eine Rolle gespielt haben:

- Eine Aussage ist ein Ausdruck, den ein eindeutiger Wahrheitswert („wahr“ oder „falsch“) zugeordnet werden kann.
- Sei X eine Menge. $B(x)$ haben wir Bedingung für $x \in X$ genannt, wenn $B(x)$ für jedes $x \in X$ einen eindeutigen Wahrheitswert hat, d.h. es gibt eine Funktion $w_B : X \to \{w, f\}$, die jedem $x \in X$ den Wahrheitswert von $B(x)$ zuordnet.[4]
- Genauso können wir Bedingungen $B(x, y)$ an *Paare* (x, y) *von Elementen* $x \in X, y \in Y$ definieren (wobei Y eine zweite Menge sei). Diesmal ist das eine Funktion

  $$w_B : X \times Y \to \{w, f\}, \qquad (x, y) \mapsto \text{Wahrheitswert von } B(x, y),$$

  wobei $X \times Y$ die Menge aller möglichen Pärchen (x, y) mit $x \in X, y \in Y$ bezeichnet.[5]
- Zwei Bedingungen $A(x)$, $B(x)$ heißen äquivalent in X, wenn sie dieselbe Wahrheitswertfunktion haben, d.h. für alle $x \in X$ gilt $w_A(x) = w_B(x)$.

[3] Das ist zwar nur die halbe Wahrheit, weil sich heutzutage der Klassenbegriff als sinnvolle Erweiterung des Mengenbegriffs etabliert hat; wir bleiben hier aber dabei, und wen das interessiert, schaut in [Dei] vorbei.

[4] In anderen Darstellungen werden Bedingungen auch als *Aussageformen* oder *einstellige Prädikate* bezeichnet, der Begriff der *Bedingung* ist meines Erachtens aber sehr treffend und griffig für den zu beschreibenden Sachverhalt.

[5] Das wäre dann ein zweistelliges Prädikat.

Definition 7.3.1. *(Tautologien, Allaussagen; erfüllbare Aussagen, Existenzaussagen)*

- *Eine Bedingung $A(x)$ an $x \in X$ heißt Tautologie (auf X), wenn $A(x)$ für alle $x \in X$ den Wahrheitswert w hat, d.h. $w_A(x) = w$ für alle $x \in X$.*

  Eine Allaussage über die Menge X ist eine Aussagen vom Typ „Für alle Elemente $x \in X$ ist $A(x)$ wahr", kurz:
  $$\forall x \in X : A(x).$$
  Einer Allaussage formal den Wahrheitswert wahr zu genau dann, wenn $A(x)$ eine Tautologie auf X ist; also machen wir das so.

- *Eine Bedingung $A(x)$ heißt erfüllbar in X, wenn es ein $x \in X$ gibt, so dass $w_A(x) = w$. Die zugehörige Aussage „Es gibt ein Element x aus X, so dass $A(x)$ wahr ist" nennt man Existenzaussage, kurz:*
  $$\exists x \in X : A(x).$$
  Eine Existenzaussage ist wahr genau dann, wenn $A(x)$ (durch ein Element $x \in X$) erfüllbar ist.

Das Symbol $\forall$ nennt man *Allquantor*, das Symbol $\exists$ nennt man *Existenzquantor*.

7.4 Logische Verknüpfungen von Aussagen und Bedingungen

Oft kommen auch „quantisierte Aussagen", d.h. All- oder Existenzaussagen vor, in denen verschiedene Bedingungen verknüpft werden, so z.B.

$$(1)\ \forall x \in X : \big(A(x) \text{ oder } B(x)\big), \qquad (2)\ \forall x \in X : \big(A(x) \text{ und } B(x)\big),$$

$$(3)\ \forall x \in X : \big(\text{Aus } A(x) \text{ folgt immer } B(x)\big), \qquad (4)\ \forall x \in X : \big(A(x) \text{ genau dann, wenn } B(x)\big).$$

Definition 7.4.1. *(Folgerung, Äquivalenz)*

Ein Ausdruck von Typ $\forall x \in X$: *Aus* $A(x)$ *folgt* $B(x)$ *heißt* *Folgerung* *oder* *Implikationen* *auf* X. *Man dass die Bedingung* $B(x)$ notwendig für $A(x)$ *ist, und dass die Bedingung* $A(x)$ hinreichend für $B(x)$ *ist.*[6]

Eine Aussage vom Typ $\forall x \in X : \big(A(x)$ *genau dann, wenn* $B(x)\big)$ *heißt* *Äquivalenz* *auf* X *und die Bedingungen* $A(x)$ *und* $B(x)$ *für* $x \in X$ *äquivalent*.

Wir wollen in diesem Abschnitt darstellen, wie sich in solchen Verknüpfungen logischer Bausteine $A(x), B(x), \ldots$, die sich jeweils mit Wahrheitswerten belegen lassen, der Gesamtwahrheitswert formal ermitteln lässt. Dazu definieren wir formal, was wir unter Verknüpfungen von Aussagen und Bedingungen verstehen wollen.

Definition 7.4.2. *(Formale Verknüpfung von Aussagen und Bedingungen)*

Seien $A, B, C, \ldots$ *im Folgenden Aussagevariablen, die die Wahrheitswerte* w *oder* f *annehmen können. Abhängig davon, ob Aussagevariablen* A *und* B *wahr (w) oder falsch (f) sind, definieren wir den Wahrheitswert der formalen logischen Verknüpfungen* $A \wedge B$, $A \vee B$, $A \Longrightarrow B$ *und* $A \Longleftrightarrow B$ *(als Formalisierungen von „A und B“, „A oder B“, „A folgt B“, und von „A genau dann, wenn B“) und die Verneinung* $\neg A$ *von* A *folgendermaßen:*

A	B	$\neg A$	$A \wedge B$	$A \vee B$	$A \Rightarrow B$	$A \Longleftrightarrow B$
w	w	f	w	w	w	w
w	f	f	f	w	f	f
f	w	w	f	w	w	f
f	f	w	f	f	w	w

[6] Diese Sprechweisen werden oft verwechselt. Eine gute, wenn auch etwas makabre Eselsbrücke zum Merken der Begriffe ist die bei Beutelspacher zu findende Formalisierung der Aussagen „Die Zerstörung unseres Planeten ist hinreichend für den Weltfrieden.“ und „Die Zerstörung unseres Planeten ist notwendig für den Weltfrieden.“, deren Formalisierung in der Form $A \Rightarrow B$ einem gut zeigt, welche Bedingung die hinreichende Bedingung und welche die notwendige ist. Ein weiteres Beispiel kennen Sie vermutlich aus dem schulischen Reich der Ableitungen. Dort nutzt man die für differenzierbare Funktionen $f : \mathbb{R} \to \mathbb{R}$ geltende Implikation, dass aus der Aussage $A(x)$: „f besitzt ein Extremum in x_0.“ folgt, dass $B(x)$: „Es ist $f'(x_0) = 0$.“ folgt. $B(x)$ ist notwendig für $A(x)$, d.h. an jeder Extremstelle muss $f'(x_0) = 0$ sein, umgekehrt stimmt das im Allgemeinen nicht (Stichwort Sattelpunkte). Ebenso ist $A(x)$ hinreichend für $B(x)$, das wird in der Schule aber oft nicht benannt, da man ja nicht wissen will, an welchen Stellen die Ableitung gleich Null ist, sondern wiederum eine hinreichende Bedingung für $A(x)$ haben möchte. An diesem Beispiel wird hoffentlich nochmal klar, dass eine notwendige Bedingung für $A(x)$ nur eine Menge von möglichen „Kandidaten“ $x \in X$ liefert, für die $A(x)$ überhaupt gelten kann, aber möglicherweise auch noch mehr Elemente – siehe oben, es ist $M_A \subseteq M_B$.

Bemerkungen dazu:

- Die Hauptanwendung im mathematischen Alltag ist die Folgende: Die Aussagenvariablen A, B können für die Wahrheitswerte von Bedingungen $A(x), B(x)$ für verschiedene Elemente auf einer Menge X stehen. Damit lassen sich dann elementweise für konkretes $x \in X$ die Wahrheitswerte der logisch verknüpften Bedingungen $(A \wedge B)(x)$, $(A \Rightarrow B)(x)$ usw. und damit die Wertebelegung $w_{A\wedge B}, w_{A\Rightarrow B}$ usw. der jeweiligen verknüpften Bedingungen bestimmen, und sich daraus wiederum ermitteln, welchen Wahrheitswert die zugehörigen Allaussagen oder Existenzaussagen haben.

- Das ist insbesondere die Motivation die möglicherweise erstaunliche Festlegung der Wahrheitswerte von „$\Rightarrow$“ in der obigen Tabelle: Damit eine Implikationen der Form

$$\forall x \in X : A(x) \Longrightarrow B(x)$$

  auf X wahr ist, muss die Bedingung $A(x) \Longrightarrow B(x)$ für alle $x \in X$ wahr sein. Das heißt, dass der Wahrheitswert von $A \Longrightarrow B$ so festgelegt werden muss, dass für Belegungen von $A(x)$ und $B(x)$ mit Wahrheitswerten folgendes gilt:

  - Wenn $A(x)$ wahr ist, muss auch $B(x)$ wahr sein.
  - Ist $A(x)$ falsch, so kann $B(x)$ einen beliebigen Wert haben.

  Daher müssen wir festlegen: Hat A den Wahrheitswert „falsch“, so muss $A \Rightarrow B$ den Wahrheitswert „wahr“ haben – nur so ergibt sich der Wahrheitswert „wahr“ für die zugehörige Allaussage af X. Das führt zu der in der Tabelle oben getätigten Festlegung, die als „Ex falsum quodlibet“ (aus Falschem folgt beliebiges) bekannt ist und oftmals in ihrer Festlegung recht dogmatisch und unmotiviert daherkommt und über merkwürdige Argumentationen motiviert wird.

- Der Wahrheitswert der Äquivalenz $A \Longleftrightarrow B$ wird analog so festgesetzt, dass sich für $A(x) \Longleftrightarrow B(x)$ genau dann eine Tautologie ergibt, wenn $A(x)$ und $B(x)$ für alle $x \in X$ immer denselben Wahrheitswert haben.

- $A \vee B$ wird so festgelegt, dass immer auch die Möglichkeit besteht, dass A *und* B wahr sind. Sonst nennt man das *Ausschließendes Oder*, kurz $A \dot{\vee} B$.

Die Verknüpfungen $\vee$ und $\wedge$ definieren eine algebraische Struktur auf der Menge aller Aussagen. Damit wird die Analyse der Mathematik mit Mitteln der Mathematik („Metamathematik“) möglich, wie Hilbert sie damals eingefordert hatte.

Durch wiederholte Anwendung der Verknüpfungsregeln lassen sich auch komplexere Formeln aus Aussagevariablen $A, B, C, \ldots$ bauen; wir nennen so etwas auch eine logische Verknüpfung von Aussagevariablen. Um sich dabei unnötige Klammern zu sparen, wird vereinbart, dass der „nicht“-Operator stärker bindet als die übrigen Junktoren, d.h. unter $\neg A \vee B$ verstehen wir die Formel $(\neg A) \vee B$ (und nicht die Verknüpfung $\neg(A \vee B)$), und die Klammern werden entsprechend weggelassen.

Als Beispiel betrachten wir die Verknüpfungen $\neg A \vee B$, $\neg B \Rightarrow \neg A$ und $\neg(A \wedge \neg B)$.

A	B	$\neg A$	$\neg B$	$A \wedge B$	$A \vee B$	$A \Rightarrow B$	$\neg A \vee B$	$\neg B \Rightarrow \neg A$	$A \wedge \neg B$	$\neg(A \wedge \neg B)$
w	w	f	f	w	w	w				
w	f	f	w	f	w	f				
f	w	w	f	f	w	w				
f	f	w	w	f	f	w				

Wir stellen fest: Die Wahrheitswerteinträge sind bei allen diesen Verknüpfungen ($\neg A \vee B$, $\neg B \Rightarrow \neg A$ und $\neg(A \wedge \neg B)$) gleich. Der zugehörige Begriff ist

Definition 7.4.3. *Zwei logische Verknüpfungen F_1 und F_2, die die Aussagevariablen $A, B, C, \ldots$ enthalten, heißen formal logisch äquivalent, wenn sich für jede Belegung von $A, B, C, \ldots$ mit Wahrheitswerten für F_1 und F_2 derselbe Wahrheitswert ergibt. Wir schreiben dafür $F_1 \equiv F_2$.*

7.5 Logische Schlussregeln

Die Bedeutung logisch äquivalenter Verknüpfungen F_1, F_2 für unseren „mathematischen Alltag“ ergibt sich dadurch, dass wir, beispielsweise, wenn in einem Allausdruck eine logische Verknüpfung $F_1(x)$ von Bedingungen $A(x)$, $B(x), \ldots$ vorkommt, $F_1(x)$ gegen $F_2(x)$ austauschen können und stattdessen F_2 beweisen können: Für alle $x \in X$ sind ja die Belegungen der Grundbausteinaussagen gleich, und deshalb auch die Wahrheitswerte von $F_1(x)$ und $F_2(x)$ für alle $x \in X$. Ist also $F_1(x)$ für alle $x \in X$ wahr, so auch $F_2(x)$ und umgekehrt. Analog lässt sich argumentieren, wenn eine Verknüpfung als Baustein einer Existenzaussage oder als eine von mehreren verknüpften Formeln vorkommt.

Wir haben also gerade gezeigt:

Lemma 7.5.1. *Die folgenden Aussagen sind formal logisch äquivalent, d.h. jede von ihnen ist wahr genau dann, wenn die drei anderen wahr sind:*

(i) $\forall x \in X : A(x) \Longrightarrow B(x)$ *(direkter Beweis)*

(ii) $\forall x \in X : \neg A(x) \vee B(x)$ *(Umformulierung)*

(iii) $\forall x \in X : \neg B(x) \Longrightarrow \neg A(x)$ *(Kontrapositionsbeweis)*

(iv) $\forall x \in X : \neg(A(x) \wedge \neg B(x))$ *(Widerspruchsbeweis)*

Die Beschäftigung mit logischen Aussagen hilft uns so, beim Beweisen das Spektrum möglicher logischer Schlüsse zu erweitern, diese besser analysieren zu können und unser Schließen auf eine solide Basis zu stellen: Die Basis all unseres mathematischen Schließens ist die axiomatische Festlegung in der Tabelle für die logischen Verknüpfungen $\vee, \wedge, \neg, \Rightarrow, \Leftrightarrow$ in 7.4.2 zusammen mit der Festlegung, wann Aussagen mit den Quantoren $\exists$ und $\forall$ über konkreten Mengen wahr werden, s. Abschnitt 7.3.[7] – das ist unser zulässiger *logisches Kalkül.* Alle von uns verwendeten Schlussregeln müssen sich daraus ableiten lassen!

Wir stellen einige solcher Schlussregeln im Folgenden zusammen: Die folgenden logischen Äquivalenzen der folgenden Verknüpfungen lassen sich zum Beweis von Aussagen der Form $\forall x \in X : A(x) \Longleftrightarrow B(x)$ nutzen:

Lemma 7.5.2. *(Beweisen von Äquivalenzen)*

Die folgenden logischen Verknüpfungen sind formal logisch äquivalent:

(i) $A \Longleftrightarrow B$

(ii) $\neg A \Longleftrightarrow \neg B$

(iii) $(A \Rightarrow B) \wedge (\neg A \Rightarrow \neg B)$

(iv) $(A \Rightarrow B) \wedge (B \Rightarrow A)$

Aus den Festlegungen aus 7.4.2 ergeben sich einige weitere logische Äquivalenzen verknüpfter Aussagen, von denen viele im mathematischen Alltag nützlich und einige hier zusammengestellt sind. Die Ergebnisse geben uns hoffentlich im Abgleich mit unserer Alltagslogik das Gefühl, das die Festlegungen oben sinnvoll waren.

Die folgenden beiden Lemmata über logisch äquivalente Formeln beweist man mittels Wahrheitswerttabelle oder durch Rückführung auf andere logische Äquivalenzen; exemplarische Beispiele dazu sehen wir in den Übungen.

[7] Die logischen Symbole $\forall, \wedge, \Rightarrow$ und $\Leftrightarrow$ lassen sich auch noch mit Hilfe von $\vee, \exists$ und $\neg$ definieren. Es reicht also sogar die formale Festlegung der Wahrheitswerte dieser letzten drei logischen Verknüpfungen.

Lemma 7.5.3. *(„Rechenregeln" für logische Junktoren)*
Für Aussagevariablen A, B, C gilt:

(i) $A \equiv \neg(\neg A)$,

(ii) $A \wedge B \equiv B \wedge A$, *(Kommutativgesetz I)*

(iii) $A \vee B \equiv B \vee A$, *(Kommutativgesetz II)*

(iv) $A \wedge (B \wedge C) \equiv (A \wedge B) \wedge C$ *(Assoziativgesetz I, Klammern weglassen)*

(v) $A \vee (B \vee C) \equiv (A \vee B) \vee C$ *(Assoziativgesetz II, Klammern weglassen)*

(vi) $A \wedge (B \vee C) \equiv (A \wedge B) \vee (A \wedge C)$ *(Distributivgesetz I)*

(vii) $A \vee (B \wedge C) \equiv (A \vee B) \wedge (A \vee C)$ *(Distributivgesetz II)*

(viii) $\neg(A \vee B) \equiv \neg A \wedge \neg B$ *(de Morgansche Regel I)*

(ix) $\neg(A \wedge B) \equiv \neg A \vee \neg B$ *(de Morgansche Regel II)*

So weit, so gut. All diese logisch-formalen Äquivalenzen werden über Gleichheit der Wahrweitswerttafeleinträge oder mit Hilfe schon bewiesener Regeln bewiesen. Besonders wichtig sind die de Morganschen Regeln (vii) und (viii), die beim Verneinen von und- oder oder-Verknüpfungen helfen, den Überblick zu behalten. Für das mathematische Beweisen interessant sind weiterhin die folgenden Schlussregeln, mit der wir diesen Teil zur Logik abschließen wollen:

Lemma 7.5.4. *(Zusammenstellung einiger gängiger Schlussweisen)*

(i) *Ist $\neg A$ falsch, so ist A wahr (ein Drittes gibt es nicht,* tertium non datur*),*

(ii) $A \Rightarrow (B \vee C) \equiv (A \wedge \neg B) \Rightarrow C$,

(iii) *Ist A wahr, so ist $B \wedge A \equiv B$,*

(iv) *Ist $(A \Rightarrow B) \wedge (B \Rightarrow C)$ wahr, so auch $A \Rightarrow C$ („Transitivität der Implikation"),*

(v) *Gilt $B_1 \equiv B_2$, so sind auch die Aussagen $A \wedge B_1$ und $A \wedge B_2$, die Aussagen $A \Rightarrow B_1$ und $A \Rightarrow B_2$ usw. logisch äquivalent,*

(vi) *Ist $(A \Leftrightarrow B) \wedge (B \Leftrightarrow C)$ wahr, so auch $A \Leftrightarrow C$,*

(vii) $(A \Rightarrow B) \wedge (B \Rightarrow C) \wedge (C \Rightarrow A)$ *ist logisch äquivalent zu*

$$(A \Leftrightarrow B) \wedge (B \Leftrightarrow C) \wedge (A \Leftrightarrow C).$$

7.6 Mehr Mengenlehre; Relationen, Funktionen und Äquivalenzklassen

Mit den in 7.4.2 entwickelten Grundlagen der Logik lassen sich jetzt auch die in 7.2 entwickelten Grundzüge der Mengenlehre etwas weiter ausführen.

Lemma 7.6.1. *(Rechengesetze für Durchschnitte und Vereinigungen von Mengen)*

Für beliebige Mengen A, B und C gelten die Kommutativgesetze

$$A \cap B = B \cap A \qquad \text{und} \qquad A \cup B = B \cup A,$$

die Assoziativgesetze

$$A \cap (B \cap C) = (A \cap B) \cap C \quad \text{und} \quad A \cup (B \cup C) = (A \cup B) \cup C$$

und die Distributivgesetze

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \qquad \text{und} \qquad A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

Beweis: In der Vorlesung werden exemplarisch einige Aussagen auf den zugehörigen logischen Ausdruck zurückgeführt, vgl. ansonsten auch das Skript von Jana Bielagk, Kapitel 2.

□

Definition/Bemerkung 7.6.1. *(Einige weitere Begriffe zu Mengen)*

- *Zwei Mengen heißen disjunkt, falls $X \cap Y = \emptyset$ ist. Sind X und Y disjunkte Mengen, so heißt die Vereinigung $Z = X \cup Y$ disjunkte Vereinigung; Schreibweise: $Z = X \dot{\cup} Y$.*
- *Wir haben oben die Mengendifferenz $X \backslash Y$ kennen gelernt. Ist $Y \subseteq X$, so nennt man diese Menge*
$$X \backslash Y := \{x \in X \mid x \notin Y\}$$
das Komplement von Y in X; es gilt $X = Y \dot{\cup} (X \backslash Y)$.
- *Auf Grund der Kommutativ- und Assoziativgesetze für Vereinigungen und Schnitte von Mengen können wir Vereinigungen und Schnittmengen von mehr als zwei Mengen definieren, ohne auf die Reihenfolge oder Klammern achten zu müssen. Dafür*

nutzen wir die folgende Notation:

$$\bigcup_{i\in I} A_i \qquad \textit{bzw.} \qquad \bigcap_{i\in I} A_i$$

wobei I eine beliebige Indexmenge sein kann. Ist $I = \underline{n}$, so schreibt man auch $\bigcup_{i=1}^{n} A_i$ für $\bigcup_{i\in I} A_i$ und $\bigcap_{i=1}^{n} A_i$ für $\bigcap_{i\in I} A_i$

- $X^n := X \times X \times \ldots \times X = \{(x_1, \ldots, x_n) \mid x_1, \ldots, x_n \in X\}$, *die Menge aller n-Tupel über X, z.B. sind $\mathbb{R}^n$ oder $\mathbb{K}^n$ Koordinatenvektoren über $\mathbb{R}$ bzw. über $\mathbb{K}$. Vorsicht: Bei n-Tupeln spielt die Reihenfolge eine Rolle, es ist $(1, 2) \neq (2, 1)$.*

Im Zermelo-Fraenkel-Aufbau der Mengenlehre ist das Bilden von kartesischen Produkten erlaubt. Das ermöglicht viele Konstruktionen, die beim Aufbau der Zalhbereiche (also ihrer Konstruktion auf Grundlage der Existenz der natürlichen Zahlen $\mathbb{N}$, deren Existenz ja axiomatisch gefordert wird) und beim Umgang mit Funktionen eine Rolle spielen. In diesem Abschnitt werden winige der dabei benutzten Grundkonstruktionen eingeführt.

Definition 7.6.2. *(Relation)*
Seien X, Y Mengen. Eine beliebige Teilmenge von $X \times Y$ (eine „Auswahl von x-y-Pärchen“) heißt Relation zwischen (den Elementen von) X und Y. Relationen lassen sich über Bedingungen $B(x, y)$ an Pärchen $(x, y) \in X \times Y$ mit Hilfe der Aussonderung

$$R = \{(x, y) \in X \times Y \mid B(x, y)\}$$

definieren.

Beispiele:

- Ist X die Menge aller Menschen, Y die Menge aller Automarken, so definiert die Bedingung $B(x, y)$: „x besitzt ein Auto der Marke y“ eine Relation zwischen X und Y.

- Die Menge $R \subseteq \mathbb{C}[x] \times \mathbb{C}$, definiert durch

$$R := \{(p, c) \mid c \text{ ist Nullstelle von } p\}$$

ist eine Relation auf zwischen Polynomen aus $\mathbb{C}[x]$ und Zahlen aus $\mathbb{C}$.

Definition 7.6.3. *(Funktion, V2.0)*
Sind X, Y Mengen und R eine Relation zwischen X und Y, so heißt R Funktion, wenn gilt: Zu jedem $x \in X$ gibt es genau ein $y \in Y$ so, dass $(x, y) \in R$ ist. y ist dann der jeweilige Funktionswert zu x. Profisprechweise: Eine Funktion ist eine linkstotale (zu jedem x-Wert ein y-Wert) und rechtseindeutige (y-Wert zu x ist eindeutig) Relation.

Die Definition des Funktionsbegriffs ist damit vollständig auf den Mengenbegriff zurückgeführt und verzichtet auf den unscharfen Begriff des „Zuordnens“ (und damit leider auch auf eine gewisse „dynamische“ Anschaulichkeit des Zuordnungsprozesses, die durch Betrachtung „statischer“ Relationspaare ersetzt wird.[8]). Insbesondere ist damit auch die Gleichheit von Funktionen $f, g : X \to Y$ auf Gleichheit von Mengen zurückgeführt: Es gilt $f = g$, wenn diese als Relationen, also als Teilmengen von $X \times Y$ übereinstimmen.

Beispiel:

- Die „Auto-Relation“ und auch die Relation R zwischen den Polynomen und komplexen Zahlen oben sind keine Funktionen (warum?).
- Die Relation
$$\{(x, y) \in \mathbb{R} \times \mathbb{R} \mid y = x^2\}$$
ein altes Beispiel einer Funktion in neuer Schreibweise.
- Ist M eine nicht leere Menge und $*$ eine Verknüpfung auf M, so ist $*$ eine Funktion $* : M \times M \to M$, also eine Relation auf $(M \times M) \times M$ (!).

Definition 7.6.4. *(Begriffe für Relationen, Ordnungsrelation, Äquivalenzrelation)*
Eine Relation $R \subseteq X \times X$ (also <u>für $X = Y$</u>) heißt Relation auf X oder Relation zwischen Elementen von X. Für $(x, y) \in R$ schreiben wir dann auch $x \sim_R y$ oder $x \sim y$, wenn klar ist, um welche Relation es sich handelt.
Hiermit lassen sich Mengen nach bestimmten Kriterien ordnen bzw. strukturieren. Wichtige Begriffe, um dies zu tun, sind die folgenden: Eine Relation R auf X heißt

- *<u>reflexiv</u>, falls für alle $x \in X$ gilt: $x \sim x$,*
- *<u>transitiv</u>, falls für alle $x, y, z \in X$ gilt: $(x \sim y \wedge y \sim z) \Rightarrow (x \sim z)$,*

[8] Diese Unterscheidung zwischen dynamischen und statischen Erklärungsmodellen ist in der Mathematikdidaktik gängig; in vielen Fällen ist wie hier die statische Herangehensweise mathematisch präziser, aber weniger zugänglich.

- *symmetrisch, falls für alle $x, y \in X$ gilt: $(x \sim y) \Rightarrow (y \sim x)$,*
- *antisymmetrisch, falls für alle $x, y \in X$ gilt: $\quad x \sim y \ \wedge \ y \sim x \implies x = y$,*

Falls R reflexiv, transitiv und symmetrisch ist, so nennen wir R eine Äquivalenzrelation auf X.
Ist R reflexiv, transitiv und antisymmetrisch, so nennen wir R eine Ordnungsrelation oder einfach Ordnung auf X.
Eine Ordnung R auf einer Menge X heißt total, falls alle Elemente von X durch R „vergleichbar" sind, d.h. für alle $x, y \in X$ gilt $(x, y) \in R$ oder $(y, x) \in R$.

Beispiele:

- Beispiele, die die Anforderungen für Ordnungsrelationen erfüllen, sind $\leq$ und $\geq$ auf den gängigen Zahlbereichen[9] und die Inklusion $\subseteq$ als Relation zwischen Mengen. Letzteres Beispiel zeigt uns, dass es ganz praktisch wäre, die Menge aller Mengen betrachten zu können, s. aber die Bemerkung unten.
- Äquivalenzrelationen sind z.B. $=$ für Elemente von Mengen, Funktionen, Mengen,..., die Äquivalenz von Bedingungen $A(x)$ an $x \in X$ und die Einsetzungsgleichheit für Terme über einer Menge X[10].

P.S.: S.o., die Betrachtung der Menge aller Mengen führt zu Widersprüchen, „=" für Mengen kann also nicht so einfach als Äquivalenzrelation auf dieser Menge angesehen werden, „$\subseteq$" nicht als Ordnungsrelation. Hierfür ist wieder eine Modifikation der Mengenlehre notwendig, die zwischen Mengen und so genannten Klassen unterscheidet.

Definition 7.6.5. *(Klasseneinteilung auf einer Menge)*
Ist X eine Menge und $\{X_i \mid i \in I\}$ eine disjunkte Zerlegung von X in Teilmengen $X_i \subset X$, d.h.

$$\bigcup_{i \in I} X_i = X \quad \text{und} \quad X_i \cap X_j = \emptyset \text{ für } i \neq j, \tag{7.1}$$

so nennt man $\{X_i \mid i \in I\}$ auch Klasseneinteilung auf X. (I kann hierbei eine beliebige Indexmenge sein.)

[9] Nicht auf $\mathbb{C}$, dort ist das Finden einer sinnvollen Ordnung schwierig.
[10] Zwei Terme nennt man einsetzungsgleich über einer Menge X, wenn sie beim Einsetzen eines jeden $x \in X$ den gleichen Wert haben, z.B. sind $x^2 - 1$ und $(x-1)(x+1)$ einsetzungsgleich über $\mathbb{R}$ (aber, wie man klar sieht, nicht derselbe Term).

Beispiele.

- Menge aller Schüler einer Schule, Klasseneinteilung.
- Für $n \in \mathbb{N}$ und $X = \mathbb{Z}$ ist die Einteilung in die n Mengen $X_i \subset X$ mit $i \in \{0, \dots, n-1\}$ und

  $$X_i = \{a \in \mathbb{Z} \mid a \text{ lässt beim Teilen durch } n \text{ den Rest } i\}$$

  eine Klasseneinteilung $\mathbb{Z}/n\mathbb{Z} := \{X_0, X_1, \dots, X_{n-1}\}$ in n so genannte Restklassen von $\mathbb{Z}$, nämlich

  $$\begin{aligned} X_0 &:= \{\dots, -2n, -n, 0, n, 2n, 3n, \dots\} \\ X_1 &:= \{\dots, -2n+1, -n+1, 1, n+1, 2n+1, 3n+1, \dots\} \\ X_2 &:= \{\dots, -2n+2, -n+2, 2, n+2, 2n+2, 3n+2, \dots\} \\ &\vdots \\ X_{n-1} &:= \{\dots, -2n+(n-1), -1, n-1, 2n, 3n, \dots\}. \end{aligned}$$

Definition 7.6.6. *Sei $\sim$ eine Äquivalenzrelation auf einer Menge X. Dann heißt für $x \in X$ die Menge*

$$[x]_\sim := \{y \in X \mid x \sim y\}$$

die Äquivalenzklasse von x (bezüglich $\sim$). Ist $y \in [x]$ ein beliebiges Element mit $x \sim y$, so heißt y ein Repräsentant der Äquivalenzklasse $[x]$. Die Menge aller Äquivalenzklassen

$$X/\!\sim\, := \{[x]_\sim \mid x \in X\}$$

bezüglich $\sim$ (das ist eine Menge von Mengen!) heißt die Quotientenmenge von X bezüglich $\sim$.

Kurzes Beispiel s. Vorlesung.

Satz 7.6.7. *(Äquivalenzrelationen induzieren Klasseneinteilungen)*

Ist $\sim$ eine Äquivalenzrelation auf einer Menge X, so ist $X/\!\sim$ eine Klasseneinteilung auf X („Einteilung von X in durch $\sim$ erzeugte Äquivalenzklassen“).

Beweis: Wir müssen die Bedingungen aus (7.1) zeigen. Für jedes $x \in X$ ist $x \in [x]_\sim$ (warum?) und deshalb $x \in \cup_{x\in M}[x]$ für alle $x \in X$, also $X = \cup_{x\in M}[x]$. Um für Äquivalenzklassen $X_i = [x]$, $X_j = [y]$ mit Repräsentanten $x, y \in X$ zu zeigen, dass im Fall $X_i \neq X_j$ die Bedingung $X_i \cap X_j = \emptyset$ erfüllt ist, zeigen wir: Enthalten diese ein gemeinsames Element, so sind sie gleich (also die Kontraposition der Aussage). Ist dieses gemeinsame Element $z \in X$, so gilt $x \sim z$ und $y \sim z$, also $z \sim y$, also $x \sim y$, also $y \in [x]$ (Jeweils: Warum?). Wegen der Transitivität impliziert das $[y] \subseteq [x]$. Das gleiche Argument lässt sich unter Vertauschung der Rollen von x und y formulieren und liefert $[x] \subseteq [y]$, also insgesamt $[x] = [y]$.

□

7.7 Mehr zu All- und Existenzaussagen

7.7.1 Kombinationen von All- und Existenzaussagen

In „mathematischen Alltag“ werden für komplexere Aussagen, insbesondere solche, die logische Beziehungen zwischen verschiedenen Objekten beschreiben, Verschachtelungen von All- und Existenzaussagen benutzt. Hierbei gibt es neben Ausdrücken der Form

$$\forall x \in X\ \forall y \in Y:\ A(x, y) \qquad \text{und} \qquad \exists x \in X\ \exists y \in Y:\ A(x, y),$$

zwei häufig besondere Formen, in denen die Quantoren gemischt vorkommen und über die man Folgendes wissen sollte:

(1.) „Zu jedem $x \in X$ findet man ein zugehöriges $y \in Y$“, formal:

$$\forall x \in X\ \ \exists y \in Y : A(x, y),$$

Beispiele aus dem vorliegenden Skript sind:

- „Zu jeder Funktion f aus der Menge der bijektiven Funktionen von X nach Y existiert eine zugehörige Umkehrfunktion $f^{-1} : Y \to X$“,
- „Jedes Element einer Gruppe besitzt ein Inverses“,
- „Jedes Polynom $p \in \mathbb{C}[x]$ besitzt eine komplexe Nullstelle“,
- „Zu jeder nichtnegativen reellen Zahl a existiert eine eindeutige nichtnegative reellen Zahl x mit $x^2 = a$“.

In der letzten Aussage ist die Existenzaussage sogar eine Existenz- und Eindeutigkeitsaussage, d.h. so ein x existiert und es gibt nur ein x mit dieser Eigenschaft. Das ist etwas anderes als die reine Existenzaussage, also eine Aussage von Typ „es gibt ein“ wie oben, die immer als „es gibt mindestens ein“ zu lesen ist. In Quantoren ausgedrückt findet man für solche Existenz- und Eindeutigkeitsaussagen („es existiert genau ein“) manchmal die Schreibweise

$$\forall x \in Y \; \exists! y \in Y : A(x, y).$$

Der zweite bedeutende

(2.) „Es gibt ein $x \in X$, das für alle $y \in Y$ die Aussage $A(x, y)$ wahr macht“, formal:

$$\exists x \in X \quad \forall y \in Y : A(x, y).$$

Ein Beispiel aus der Gruppentheorie ist die Forderung der Existenz eines neutralen Elements: „Es gibt ein $e \in M$, so dass für alle $a \in M$ die Aussagen $e \circ a = a$ und $a \circ e = a$ wahr sind.“

Vorsicht! Im Gegensatz zu einer Aussage vom Typ $\forall x \forall y$ darf man in (1) und (2) Existenz- und Allquantor nicht vertauschen! Es ergeben sich Aussagen mit anderem Sinngehalt und für gewöhnlich mit anderem Wahrheitswert (probiere das anhand der Beispiele oben!).

7.7.2 Verneinungen von Quantoren(-ketten)

Will man zeigen, dass eine Aussage falsch ist, so tut man das meistens, indem man ihre Verneinung formuliert und zeigt, dass diese wahr ist.

- **Verneinung von Allaussagen:** Eine Allaussage $\forall x \in X : A(x)$ ist genau dann wahr, wenn $A(x)$ für alle $x \in X$ wahr ist. Der komplementäre Fall tritt ein, wenn es mindestens ein x gibt, das die negierte Aussage $\neg A(x)$ erfüllt, also genau dann, wenn die Aussage $\exists x \in X : \neg A(x)$ wahr ist. Um eine Allaussage zu widerlegen, muss man also zeigen, dass $\neg A(x)$ erfüllbar ist – man sagt, man muss ein *Gegenbeispiel* zu $A(x)$ angeben.

- **Verneinung von Existenzaussagen:** Eine Existenzaussage $\exists x \in X : A(x)$ ist genau dann wahr, wenn es mindestens ein $x \in X$ gibt, für das $A(x)$ wahr ist. Die Aussage ist also genau dann falsch, wenn für alle $x \in X$ gilt, dass $A(x)$ falsch ist. Ihre Verneinung $\neg (\exists x \in X : A(x))$ ist also wahr genau dann, wenn $\forall x \in X : \neg A(x)$ gilt, und die hintere Bedingung lässt sich eventuell besser prüfen.

Man sieht in den vorigen beiden Punkten, das, was oft halbformal als „Hineinziehen einer Negation in Quantoren dreht den Quantor um." ausgedrückt wird. Für die Negation verschachtelter Quantorenausdrücke liefert das die folgende Vorgehensweise:

- **Verneinung von Quantorenketten:** Quantorenketten lassen sich negieren, indem man die Negation Schritt für Schritt hineinzieht; so ergibt sich zum Beispiel die Gleichwertigkeit der Aussagen

$$\neg\left(\forall x \in X \exists y \in Y : A(x,y)\right) \quad \text{und} \quad \exists x \in X \forall y \in Y : \neg A(x,y)$$

Das macht, wenn man die Aussagen mit seiner „Alltagslogik" abgleicht, hoffentlich auch Sinn.

7.8 Über- und Ausblick: Axiomatisierung und Formalisierung der Mathematik und ihre Grenzen

- Anfänge der Logik: Aristoteles (Aussagenlogik, 4 Jhd. v. Chr)
- Bertrand Russell, George Boole, Auguste de Morgan, Gottlob Frege (19. Jhd): Grundlagen der Logik und Mengenlehre (Russell: *Principia Mathematica*)
- 1920er (nach „logischen Katastrophen"): Hilbert'sches Programm zur Axiomatisierung der gesamten Mathematik:
  - Zermelo-Fraenkel-Mengenlehre, formale Logik (Schlussregeln, Paul Bernays, Albert Skolem, Kurt Gödel)
  - Axiomatik der natürlichen Zahlen (Guiseppe Peano, Ende 19. Jhd., „Welche Eigenschaften muss eine Menge haben, damit wir sie die natürlichen Zahlen nennen?")
  - Axiomatisierung der Geometrie, der Algebra, der Zahlbereiche, der Vektorraumtheorie,...
- Großes Ziel: Beweis der Widerspruchsfreiheit dieser Axiomsysteme (d.h. mathematisch zu beweisen, dass sich aus den jeweiligen Axiomensystemen mit Hilfe der verwendeten Schlussregeln keine Widersprüche formal ableiten lassen)
- Dazu: Abstraktion zu bedeutungslosen Zeichenketten, mit denen rein formal festgelegt wird, wann sich aus einer Zeichenkette eine andere ableiten („logisch-formal

beweisen“) lässt – dann Interpretation der Zeichenketten als Aussagen in konkreten Mengen (alles,was sich so formal beweisen lässt, stimmt dann auch in jeder „Real-Life“-Interpretation der Zeichenketten).

- „Metamathematik“ – Untersuchung der Grundlagen der Mathematik mit Mathematischen Methoden
- Das klappt nicht: Die Widerspruchsfreiheit des Theoriegebäudes „Mengenlehre plus Schlussregeln“ ist nicht mit Mengenlehre plus Schlussregeln beweisbar. (Zweiter Gödelscher Unvollständigkeitssatz, 1930er)
- Erster Gödelscher Unvollständigkeitssatz, 1930er: In jedem System von Axiomen und Schlussregeln, das die natürlichen Zahlen mit $+$ und $\cdot$ (Peano-Arithmetik) formal beschreiben kann, kann man Sätze formulieren, die nicht formal beweisbar oder widerlegbar, also nicht formal entscheidbar, sind.
- Trotzdem hat sich das axiomatisch-logisch-minimalistische Vorgehen durchgesetzt und bildet formalen Rahmen für unsere Mathematik: Jede Aussage, die wir formulieren, sollte prinzipiell als abstrakte Zeichenkette aus logischen Zeichen schreibbar sein, jeder Schluss sollte ein Schluss sein, der sich aus Schlüssen aus dem in diesem Kapitel umrissenen Katalog zulässiger Schlussregeln zusammensetzen lässt.
- Anwendung: „Proof Checker“ (z.B. Mizar-System, Beweis kann formalisiert eingegeben und auf Korrektheit geprüft werden)

Ausblick: Logische Katastrophen.

Aus unseren Regeln oben ergibt sich auch, dass eine Aussage der Form $A \wedge \neg A$ nie wahr sein kann (Warum? Woraus?). Daraus ergibt sich die folgende logische Problematik: Könnte man in einer mathematischen Struktur, die bestimmte Aussagen als Voraussetzungen annimmt (also z.B. in einer Gruppe unter Annahme der Gruppenaxiome G (z.B. als eine *und*-Aussage verkettet), für irgendeine Aussage A beweisen, dass unter der Voraussetzung von G die Aussagen A und $\neg A$ und somit die falsche Aussage $A \wedge \neg A$ folgt, so muss in dieser Struktur wegen der logischen Schlussregeln für $G \Rightarrow (A \wedge \neg A)$ die Aussage G den Wahrheitswert „falsch" gehabt haben, also zum Beispiel die Annahme der Gesamtheit der Gruppenaxiome in sich falsch gewesen sein, etwa, indem man dort in sich widersprüchliche Voraussetzungen angenommen hat.

Bleiben wir bei dem Beispiel mit den Gruppenaxiomen, so wäre unter der Annahme z.B. der Gruppenaxiome daher dann *alles* beweisbar (denn nach den Festlegungen der Logik gilt, wenn G nie wahr ist, dass $G \Rightarrow B$ für alle Aussagen B wahr ist). Also: Folgt aus einem mathematischen Axiomensystem ein Widerspruch, so folgt aus diesem Axiomensystem jede beliebige Aussage, die man überhaupt in der Symbolsprache, die man benutzt, formulieren kann! Das würde die Mathematik ad absurdum führen, und die Widerspruchsfreiheit von Axiomenystemen zu zeigen war ein großes Ziel des „Hilbert-Programms", das mit den Gödel'schen Unvollständigkeitssätzen spektakulär gescheitert ist, s. die etwas oberflächliche Formulierung der Unvollständigkeitssätze unten.

Kurt Gödel (Österreich, USA)

David Hilbert (früheres 20. Jhd.)

- Erster Gödelscher Unvollständigkeitssatz, 1930er:
  In jedem System von Axiomen und Schlussregeln, das die natürlichen Zahlen mit $+$ und $\cdot$ (Peano-Arithmetik) formal beschreiben kann, kann man Sätze formulieren, die nicht formal beweisbar oder widerlegbar, also nicht formal entscheidbar, sind.
- Zweiter Gödelscher Unvollständigkeitssatz:
  Die Widerspruchsfreiheit des Theoriegebäudes „Mengenlehre plus Schlussregeln" ist nicht mit Mengenlehre plus Schlussregeln beweisbar.

Kurt Gödel (früheres 20. Jhd.)

7.9 Aufgaben zu Kapitel 7

Aufgabe 7.34. (Wahrheitswerttafeln):

(a) (Beweis von Lemma 7.5.2) Seien A, B Aussagevariablen. Zeigen Sie, dass die folgenden logischen Verknüpfungen formal logisch äquivalent sind:

(i) $A \Longleftrightarrow B$,

(ii) $(A \Rightarrow B) \wedge (\neg A \Rightarrow \neg B)$,

(iii) $\neg A \Longleftrightarrow \neg B$,

(iv) $(A \Rightarrow B) \wedge (B \Rightarrow A)$.

(b) Beweisen Sie mit Hilfe einer Wahrheitswerttafel, dass für Aussagevariablen A, B, C gilt:

- $A \wedge (B \vee C) \equiv (A \wedge B) \vee (A \wedge C)$ (Distributivgesetz I)
- $\neg(A \vee B) \equiv \neg A \wedge \neg B$ (de Morgansche Regel I)
- Ist $(A \Rightarrow B) \wedge (B \Rightarrow C)$ wahr, so auch $A \Rightarrow C$ („Transitivität der Implikation“),
- Ist $(A \Leftrightarrow B) \wedge (B \Leftrightarrow C)$ wahr, so auch $A \Leftrightarrow C$.

(Das sind die Aussagen (vi) und (viii) von Lemma 7.5.3 und die Aussagen (iv) und (vi) von Lemma 7.5.4)

Sie können die Teilaufgaben mit Hilfe jeweils *einer einzigen* Wahrheitswerttafel bearbeiten; bitte halten Sie diese aber übersichtlich und geben Sie als weitere Tabellenspalten sinnvolle Zwischenschritte an, aus denen Ihre Ableitung der zu ermittelnden Wahrheitswerte aus denen von A, B, C ersichtlich wird.

Aufgabe 7.35. (All- und Existenzaussagen und logische Schlüsse):

(a) Formulieren Sie als Aussage mit All- und Existenzquantoren:

- Zu jeder positiven reellen Zahl gibt es eine natürliche Zahl, so dass der Abstand zwischen diesen beiden Zahlen kleiner als eins ist.
- Es gibt eine natürliche Zahl, so dass der Abstand dieser Zahl zu jeder positiven reellen Zahl kleiner als eins ist.

Sind die Wahrheitswerte der beiden Aussagen gleich oder verschieden?

(b) Beurteilen Sie, ob die folgenden Schlüsse im Sinne der formalen Logik richtig oder falsch sind.

   (i) „Ein Forscher im Mittelalter beobachtete einen bislang wenig bekannten Planeten in den Jahren 1537, 1540, 1543 und 1546. Nun vermutete er, dass er ihn wohl auch im Jahre 1549 sehen müsste. Und er hatte Recht, 3 Jahre später erschien der Planet wieder am Himmel. Damit hatte er bewiesen, dass der Planet alle 3 Jahre am Himmel erscheint.“

   (ii) „Frau Meiers Nachbar behauptete, in Brandenburg gäbe es Igel. „Das muss ich nachprüfen“, meinte Frau Meier und fuhr in die Uckermark. Kaum stellte sie ihr Auto an einem See ab, schon trottete frech ein Igel vorbei. Der Nachbar hatte also Recht, in Brandenburg gibt es tatsächlich Igel. Dieser hier ist der Beweis.“

   (iii) „Wenn ich 100 m unter 10 Sekunden laufe“, sagte Hänschen, „dann werde ich zur Olympiade delegiert. Leider laufe ich aber die 100 m nicht unter 10 Sekunden. Somit ist bewiesen, dass ich nicht zur Olympiade delegiert werde.“

Aufgabe 7.36. (Verneinung von Aussagen):

(a) Verneinen Sie formal Aussagen der folgenden Bauarten. Was genau müssen Sie zeigen, wenn Sie beweisen wollen, dass eine solche Aussage nicht gilt?

- $\forall x \in X : A(x) \Longrightarrow B(x)$
- $\forall x \in X : A(x) \Longleftrightarrow B(x)$
- $\forall x \in X : A(x) \Longrightarrow (B(x) \wedge C(x))$
- $\forall x \in X : A(x) \Longrightarrow (B(x) \vee C(x))$
- $\forall x \in X : (A(x) \wedge B(x)) \Longrightarrow C(x)$
- $\exists x \in X \ \forall y \in Y : A(x, y)$
- $\exists x \in X : (A(x) \wedge B(x) \wedge C(x))$
- $\exists e \in X \quad \forall x \in X \quad \exists y \in X : x * y = e$

(b) Bilden Sie die Negation der folgenden Ausdrücke so, dass das Negationszeichen $\neg$ höchstens noch direkt vor den Aussagenvariablen A, B, C vorkommt.

(i) $A \vee (B \wedge C)$,

(ii) $\neg A \Leftrightarrow (B \vee C)$

(iii) $(A \vee B) \wedge (\neg A \vee C)$

(iv) $(A \Rightarrow \neg B) \vee (\neg A \Rightarrow B)$

In (ii) bzw. (iv) bietet es sich zur Verneinung an, die Ausdrücke mit den logischen Junktoren $\Leftrightarrow$ bzw. $\Rightarrow$ zunächst durch logisch äquivalente Ausdrücke zu ersetzen, in denen die Junktoren $\neg, \wedge$ und $\vee$ vorkommen.

(c) Es sei G eine Menge und $* : G \times G \to G$ eine Abbildung. $(G, *)$ ist eine Gruppe genau dann, wenn die Aussage

$$A: \quad \big(\forall g \in G\ \forall h \in G : g * h \in G\big) \quad \wedge \quad \big(\forall g \in G\ \forall h \in G\ \forall \ell \in G : g * (h * \ell) = (g * h) * \ell\big) \quad \wedge$$
$$\wedge \quad \Bigg(\exists e \in G : \bigg(\big(\forall g \in G : \ (e * g = g \wedge g * e = g)\big) \ \wedge \ \big(\forall g \in G\ \exists h \in G : (g * h = e \wedge h * g = e)\big)\bigg)\Bigg)$$

wahr ist.

Behalten Sie einen kühlen Kopf und verneinen Sie die Aussage A. Formen Sie sie dann so um, dass die Negationszeichen in jeder der Teilaussagen direkt vor den einzelnen Aussagen ($g * h \in G$, $g * (h * \ell) = (g * h) * \ell$, $e * g = g$ usw.) stehen; Sie können dann noch die Verneinung dieser Aussagen durch die Symbole $\notin$ und $\neq$ ausdrücken.

Fassen Sie das Ergebnis kurz (!) in Worten zusammen: Was müssen Sie zeigen, um die Aussage $\neg A$ zu beweisen?

Aufgabe 7.37. (Äquivalenzrelation oder nicht?):

Entscheiden Sie, ob es sich bei folgenden Relationen $\sim$ auf der jeweiligen gegebenen Menge M um Äquivalenzrelationen handelt. Begründen Sie Ihre Entscheidung, d.h. falls es sich um eine Äquivalenzrelation handelt, beweisen Sie die dafür nötigen Eigenschaften, falls nicht, geben Sie eine Eigenschaften an, die verletzt ist.

(a) $M = \mathbb{Z}$, $a \sim b \quad :\Longleftrightarrow \quad (a - b \quad \text{ist gerade})$,

(b) $M = \mathbb{Z}$, $a \sim b \quad :\Longleftrightarrow \quad (a - b \quad \text{ist ungerade})$,

(c) $M = \mathbb{R}^2$, $\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \sim \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \quad :\Longleftrightarrow \quad \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$ liegt auf der Geraden $g = \Big\{ \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} + \lambda \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix} \mid \lambda \in \mathbb{R} \Big\}$

Aufgabe 7.38. (Mengenoperationen):

(a) Es seien A, B, C Mengen. Beweisen Sie durch Rückführung auf die formal logische Äquivalenz wie in der Vorlesung:

$$A\backslash(B \cup C) = (A\backslash B) \cap (A\backslash C).$$

(b) Beweisen oder widerlegen Sie durch Angabe eines Gegenbeispiels: Sind M_1, M_2, M_3 Mengen mit der Eigenschaft $\bigcap_{i=1}^{3} M_i = \emptyset$, so gibt es unter den Mengen M_1, M_2, M_3 zwei disjunkte Mengen, d.h. es gibt Indizes $i, j \in \underline{3}$ so, dass $M_i \cap M_j = \emptyset$.

Aufgabe 7.39. (Konstruktion der Bruchzahlen aus ganzen Zahlen):

Baut man wie umrissen die Zahlbereiche Stück für Stück aus den natürlichen Zahlen auf, so lassen sich, nachdem man die ganzen Zahlen $\mathbb{Z}$ konstruiert hat, daraus wiederum die Bruchzahlen $\mathbb{Q}$ konstruieren, indem man Paare ganzer Zahlen, genauer, Elemente aus

$$M := \mathbb{Z} \times (\mathbb{Z}\backslash\{0\})$$

betrachtet. Dabei ergibt sich das Problem, dass z.B. $(2, 5)$ und $(4, 10)$ als Paare verschieden sind, aber dieselbe Bruchzahl $\frac{2}{5} = \frac{4}{10}$ repräsentieren sollen. Dieses wird dadurch bewältigt, dass eine Äquivalenzrelation auf M definiert, deren Äquivalenzklassen genau die Paare enthalten, die dieselbe Bruchzahl repräsentieren. Sei $\sim$ die auf $(\mathbb{Z} \times \mathbb{Z} \setminus \{0\})$ wie folgt definierte Relation:

$$(a, b) \sim (c, d) :\Longleftrightarrow\; a \cdot d = c \cdot b.$$

(a) Zeigen Sie, dass $\sim$ eine Äquivalenzrelation ist.

(b) Eine Bruchzahl wird nun formal als Äquivalenzklasse der Relation $\sim$ definiert:

$$\frac{a}{b} := \{(c, d) \in \mathbb{Z} \times \mathbb{Z} \setminus \{0\} \mid (a, b) \sim (c, d)\}.$$

d.h. $\frac{a}{b}$ ist eine <u>Menge</u>, deren Elemente gerade die Pärchen (c, d) sind, die mit (a, b) in Relation stehen. Geben Sie einen Repräsentanten der Bruchzahl $\frac{1}{3}$ an.

(c) Zeigen Sie auf Basis der Definition aus (b), dass mit dieser Konstruktion die Mengengleichheit $\frac{2}{5} = \frac{4}{10}$ gilt.

8 Vektorräume

- Erste Nutzungen von Vektoren: u.a. Rowan Hamilton (England, physikalisch motiviert), Herrmann Grassmann (umtriebiger Lehrer in Stettin, „Rechnen mit Strecken“), beide Mitte 18. Jhd.

- Grassmanns Werk ist umfassend, inklusive axiomatischer Fassung eines Vektorraumes (eine ganze Weile vor Hilbert & Co.), aber sehr unverständlich (Universitätsgutachten zu seinen „Gebieten n-ter Stufe“ konstatiert, „... dass diese Schrift von den Mathematikern ferner ignoriert werden wird wie bisher, denn die Mühe, sich in dieselbe einzuarbeiten, erscheint zu groß in Beziehung auf den wirklichen Gewinn an Erkenntnis, welchen man aus derselben schöpfen zu können vermutet.“)

- Guiseppe Peano formuliert basierend auf Grassmanns Werk erste axiomatische Fassung

- heutige Version basiert auf Ausarbeitungen von Hermann Weyl („Raum, Zeit, Materie – Vorlesungen über Allgemeine Relativitätstheorie“, 1918) und Bartel Leenert van der Waarden („Moderne Algebra“, 1930)

- Vektorraumtheorie und Theorie der linearen Abbildungen (unsere Beispiele bisher: $\mathbb{R}^n$ und Matrixabbildungen mit darstellenden Matrizen $A \in \mathbb{R}^{m \times n}$) heute: Grundlegendes, vielseitiges Werkzeug der Mathematik, das in vielen Teilbereichen der Mathematik und angrenzender Anwenderdisziplinen zum Einsatz kommt.

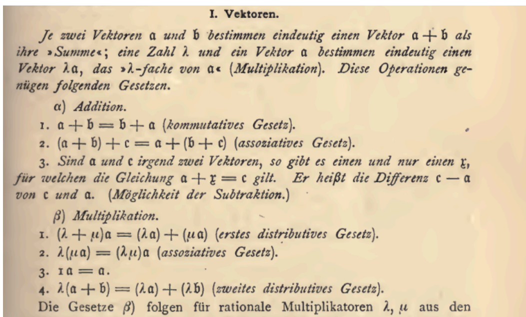
- Begrifflichkeiten und Zusammenhänge in diesem und dem folgenden Kapitel daher: Sehr grundlegend für den weiteren Verlauf der Vorlesung, des Studiums.

Rechnen mit Strecken und Axiomatisierung

Rowan Hamilton (19. Jhd.)

Herrmann Grassmann (19. Jhd.)

Hermann Weyl (20. Jhd)



I. Vektoren.

Je zwei Vektoren $\mathfrak{a}$ *und* $\mathfrak{b}$ *bestimmen eindeutig einen Vektor* $\mathfrak{a}+\mathfrak{b}$ *als ihre »Summe«; eine Zahl* λ *und ein Vektor* $\mathfrak{a}$ *bestimmen eindeutig einen Vektor* $\lambda\mathfrak{a}$, *das »*λ*-fache von* $\mathfrak{a}$*« (Multiplikation). Diese Operationen genügen folgenden Gesetzen.*

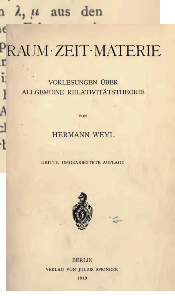
α) *Addition.*

1. $\mathfrak{a}+\mathfrak{b}=\mathfrak{b}+\mathfrak{a}$ (*kommutatives Gesetz*).
2. $(\mathfrak{a}+\mathfrak{b})+\mathfrak{c}=\mathfrak{a}+(\mathfrak{b}+\mathfrak{c})$ (*assoziatives Gesetz*).
3. *Sind* $\mathfrak{a}$ *und* $\mathfrak{c}$ *irgend zwei Vektoren, so gibt es einen und nur einen* $\mathfrak{x}$, *für welchen die Gleichung* $\mathfrak{a}+\mathfrak{x}=\mathfrak{c}$ *gilt. Er heißt die Differenz* $\mathfrak{c}-\mathfrak{a}$ *von* $\mathfrak{c}$ *und* $\mathfrak{a}$. (*Möglichkeit der Subtraktion.*)

β) *Multiplikation.*

1. $(\lambda+\mu)\mathfrak{a}=(\lambda\mathfrak{a})+(\mu\mathfrak{a})$ (*erstes distributives Gesetz*).
2. $\lambda(\mu\mathfrak{a})=(\lambda\mu)\mathfrak{a}$ (*assoziatives Gesetz*).
3. $1\mathfrak{a}=\mathfrak{a}$.
4. $\lambda(\mathfrak{a}+\mathfrak{b})=(\lambda\mathfrak{a})+(\lambda\mathfrak{b})$ (*zweites distributives Gesetz*).

Die Gesetze β) folgen für rationale Multiplikatoren λ, μ aus den Additionsaxiomen, falls wir die Multiplikation mit solche[…] oben aus der Addition *erklären*. Gemäß dem Prinzip[…] nehmen wir sie auch für beliebige reelle Zahlen in An[…] […] Axiome, da sie sich i[…] […]en Additionsaxiomen […] […]ultiplikation auf die A[…] […]Stand, aus dem logisc[…] Stetigkeit ganz zu verb[…]



RAUM · ZEIT · MATERIE

VORLESUNGEN ÜBER ALLGEMEINE RELATIVITÄTSTHEORIE

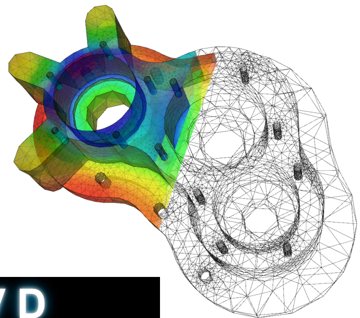
VON

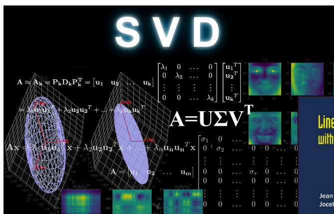
HERMANN WEYL

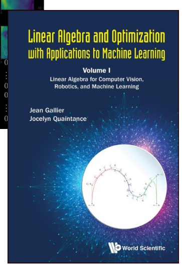
DRITTE, UMGEARBEITETE AUFLAGE

BERLIN
VERLAG VON JULIUS SPRINGER
1919

Hermann Weyl (20. Jhd)







Die Erfindung der linealen (!) Algebra

HERMANN GÜNTER GRASSMANN (1809–1877)

„Den ersten Anstoß gab mir die Betrachtung des Negativen in der Geometrie. Ich gewöhnte mich, die Strecken AB und BA als entgegengesetzte Größen aufzufassen. ... Strecken wurden nicht als bloße Längen aufgefasst, sondern an ihnen zugleich die Richtung festgehalten. So drängte sich der Unterschied auf zwischen der Summe der Längen und zwischen der Summe solcher Strecken, in denen zugleich die Richtung festgehalten war. .... Am Gesetz, dass AB + BC = AC sei, wurde auch dann noch festgehalten, wenn A, B, C nicht in einer geraden Linie lagen. Hiermit war der erste Schritt zu einer Analyse getan, welche in der Folge zu dem neuen Zweig der Mathematik führte, die hier vorliegt."

- Universitäres Gutachten zu Graßmanns erstem Werk: „...daß diese Schrift von den Mathematikern ferner ignoriert werden wird wie bisher; denn die Mühe, sich in dieselbe einzuarbeiten, erscheint zu groß in Beziehung auf den wirklichen Gewinn an Erkenntnis, welchen man aus derselben schöpfen zu können vermutet..."

(zit. nach Petsche, „Graßmann", Birkhäuser 2006)

Die Erfindung der linealen (!) Algebra

Es geht darum, die sinnlichen Anschauungen der Geometrie zu allgemeinen, logischen Begriffen zu erweitern und zu vergeistigen ... Ich sage, eine Größe a sei aus den Größen $b, c, \ldots$ durch die Zahlen $\beta, \gamma, \ldots$ abgeleitet, wenn $a = \beta b + \gamma c + \ldots$ Dabei seien $\beta, \gamma, \ldots$ reelle Zahlen.

Die Größen $a, b, c, \ldots$ stehen zueinander in einer Zahlbeziehung, wenn irgend eine sich aus den anderen numerisch berechnen lässt ... Einheit nenne ich jede Größe, welche dazu dienen soll, um aus ihr eine Reihe von Größen abzuleiten. Ein System von Einheiten nenne ich jeden Verein von Größen, welche in keiner Zahlbeziehung zueinander stehen und welche dazu dienen sollen, um aus ihnen durch beliebige Zahlen andere Größen abzuleiten. Die algebraischen Größen heißen auch extensive Größen.

Für extensive Größen gelten die Fundamentalformeln:

$$
\begin{aligned}
a + b &= b + a \\
a + (b + c) &= (a + b) + c \\
a + b - b &= a\,;
\end{aligned}
$$

a, b, c sind Größen, α, β reelle Zahlen:

$$
\begin{aligned}
\alpha a &= a\alpha \\
\alpha(\beta a) &= (\alpha\beta)a \\
\alpha(a + b) &= \alpha a + \alpha b \\
a(\alpha + \beta) &= a\alpha + a\beta \\
1a &= a\,.
\end{aligned}
$$

Die Gesamtheit der Größen, welche aus einer Reihe von Größen $a_1, a_2, \ldots, a_n$ numerisch ableitbar sind, nenne ich das aus jenen Größen ableitbare Gebiet n-ter Stufe, wenn jene Größen von erster Stufe sind und sich das Gebiet nicht aus weniger als n solchen Größen ableiten lässt. Jedes Gebiet n-ter Stufe kann aus n (ihm angehörenden) Größen erster Stufe, die in keiner Zahlbeziehung zueinander stehen, abgeleitet werden, und zwar aus beliebigen n solchen Größen des Gebiets.

(Graßmanns „Gebiete n-ter Stufe",
aus: Beutelspacher, Lineare Algebra, S. 93 (Link))

8.1 Grundbegriffe der Vektorraumtheorie

Definition 8.1.1. *(Vektorraum):*

Sei $\mathbb{K}$ ein Körper. Ein $\mathbb{K}$-Vektorraum (oder Vektorraum über $\mathbb{K}$, linearer Raum über $\mathbb{K}$, kurz oft einfach nur „Raum“) ist ein Tripel $(V, +, \cdot)$, bestehend aus einer Menge V, einer (Vektor-)Addition $+ : V \times V \longrightarrow V$ und einer Skalarmultiplikation oder skalaren Multiplikation $\cdot : \mathbb{K} \times V \longrightarrow V$, für die die folgenden Eigenschaften gelten:

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

(VS) Die skalare Multiplikation ist assoziativ und normiert: Für alle $a, b \in \mathbb{K}$ und alle $v \in V$ gilt

$$(a \cdot b) \cdot v = a \cdot (b \cdot v)$$

und

$$1_{\mathbb{K}} \cdot v = v.$$

(wobei $1_{\mathbb{K}}$ das neutrale Element bezüglich der Multiplikation in $\mathbb{K}$ bezeichnet.)

(VAS) Es gelten die Distributivgesetze für Vektorräume: Für alle $a, b \in \mathbb{K}$ und alle $v, w \in V$ gilt

$$(a + b) \cdot v = a \cdot v + b \cdot v \quad \text{und} \quad a \cdot (v + w) = a \cdot v + a \cdot w$$

Elemente von V heißen Vektoren.

Bemerkungen dazu:

- Die universitäre Antwort auf die Frage „Was ist ein Vektor?“ ist die obige strukturelle: Ein Vektor ist ein Element eines Vektorraumes.
- Oft unterdrückt man in der Notation auch die Verknüpfungen und den zugrundeliegenden Körper und sagt kurz (aber etwas schlampig), dass „V ein Vektorraum“ ist. Genauso wird in der Notation (wie oben schon) nicht zwischen der Multiplikation in $\mathbb{K}$ (z.B. $a \cdot b$) und der Skalarmultiplikation (z.B. $a \cdot v$) unterschieden. Es ist eine gute Übung, sich stets genau zu überlegen, welche Addition/Multiplikation in $\mathbb{K}$ bzw. V stattfindet.
- Die Elemente des zugrundeliegenden Körpers werden als Skalare bezeichnet.

- Es ist gängig, in der Vektorraumtheorie Körperelemente durch griechische Buchstaben $\alpha, \lambda, \mu, \ldots$ anzudeuten, Elemente des Vektorraums durch lateinische Buchstaben $v, w, u, x, b, \ldots$. Wir vertrauen in die Intelligenz der Leser und bleiben bei lateinischen Buchstaben, so dass auch Körperelemente meist mit lateinischen Buchstaben, meistens a oder c, bezeichnet werden. Aus dem Kontext sollte stets klar werden, was gemeint ist.
- Sei $\mathbb{K}$ ein Körper und sei $(V, +, \cdot)$ ein $\mathbb{K}$-Vektorraum. Man beachte, dass die Skalarmultiplikation Skalare mit Vektoren multipliziert (was Vektoren liefert). Im allgemeinen können in einem Vektorraum nicht zwei Vektoren „sinnvoll" miteinander multipliziert werden (um wieder einen Vektor zu erhalten).
- Die Forderungen (VA), (VS) und (VAS) nennt man die *Vektorraumaxiome.* Sie sind die strukturelle Grundlage der Vektorraumtheorie.

Die wichtigsten Beispiele für Vektorräume:

(i) $(\mathbb{R}^n, +, \cdot)$ ist ein $\mathbb{R}$-Vektorraum. Allgemeiner ist der $\mathbb{K}^n$ mit der punktweisen Addition und der analog zu $\mathbb{R}^n$ definierten skalaren Multiplikation ein $\mathbb{K}$-Vektorraum. (Komponentenweise Definition: Zurückführung auf Rechengesetze in $\mathbb{K}$)

(ii) Spezialfall von (i): Jeder Körper $\mathbb{K}$ ist selbst ein $\mathbb{K}$-Vektorraum (mit der ihm eigenen Addition und seiner Multiplikation als skalare Multiplikation).

(iii) Ist X eine Menge, so ist $\mathrm{Abb}(X, \mathbb{R})$ ein $\mathbb{R}$-Vektorraum, wenn wir die Summe zweier Funktionen $f + g$ wie vorher punktweise und die Skalierung mit $a \in \mathbb{R}$ ebenfalls punktweise definieren, also

$$f + g \quad \text{durch} \quad (f+g)(x) := \underbrace{f(x) + g(x)}_{\text{in } \mathbb{R}}, \qquad a \cdot f \quad \text{durch} \quad (a \cdot f)(x) := \underbrace{a \cdot f(x)}_{\text{in } \mathbb{R}}$$

setzen. Auch hier bestätigt man die Vektorraumaxiome durch Zurückführung auf die Addition und Multiplikation in $\mathbb{R}$.

(iv) Verallgemeinerung von (iii): Ist X eine Menge, V ein $\mathbb{K}$-Vektorraum, so ist $\mathrm{Abb}(X, V)$ ein $\mathbb{K}$-Vektorraum, wenn wir die Addition $f + g$ und skalare Multiplikation $a \cdot f$ von Funktionen jeweils punktweise auf die in V zurückführen (statt auf die in $\mathbb{R}$ im Beispiel oben).

(v) Die Menge aller $(m \times n)$-Matrizen $\mathbb{K}^{(m\times n)}$ über einem Körper $\mathbb{K}$ ist mit der komponentenweisen Addition $A+B$ und der komponentenweisen Skalierung $a \cdot A$ $(a \in \mathbb{K})$, also mit

$$\begin{aligned} A+B = (a_{i,j})_{i\in\underline{m},j\in\underline{n}} + (b_{i,j})_{i\in\underline{m},j\in\underline{n}} &:= (a_{i,j}+b_{i,j})_{i\in\underline{m},j\in\underline{n}}, \\ a \cdot A = a \cdot (a_{i,j})_{i\in\underline{m},j\in\underline{n}} &:= (a \cdot a_{i,j})_{i\in\underline{m},j\in\underline{n}} \end{aligned}$$

ein $\mathbb{K}$-Vektorraum. (Schränkt man sich auf Matrizen mit reellen Einträgen ein so ist dies wieder der Spezialfall des Definitionsbereichs $X = \{(i,j) \mid i \in \underline{m}, j \in \underline{n}\}$ von Beispiel (iii). Allgemeiner ergibt sich dieses Beispiel als Spezialfall von $\mathrm{Abb}(X,\mathbb{K})$, s. (iv).)

(vi) $\mathbb{C}$ ist ein $\mathbb{R}$-Vektorraum mit der normalen komplexen Addition $+ : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$,

$$(a+bi)+(c+di) = (a+c)+(b+d)i$$

und der skalaren Multiplikation mit $x \in \mathbb{R}$,

$$\cdot : \mathbb{R} \times \mathbb{C} \to \mathbb{C}, \qquad x \cdot z := x \cdot a + (x \cdot b)i$$

(Skalierung in der komplexen Ebene „wie im $\mathbb{R}^2$“)

(viii) Die Menge der Polynome über einem Körper $\mathbb{K}$, also die in 3.1.2 definierte Menge $\mathbb{K}[x]$ ist zusammen mit der dort angegebenen koeffizientenweisen Addition $+$ von Polynomen und mit der skalaren Multiplikation

$$a \cdot p = a \cdot (\sum_{i=0}^{n} c_i \cdot x^i) := a \cdot (\sum_{i=0}^{n} (a \cdot c_i) \cdot x^i)$$

ein Vektorraum, s. Übung.

(viii) Jede Verschiebung im zweidimensionalen bzw. dreidimensionalen Anschauungsraum lässt sich eindeutig durch einen nicht ortsfesten Verschiebungspfeil $\vec{v}$ (einen Vektor im Anschauungsraum wie in Kapitel IV) darstellen. Die Addition und skalare Multiplikation zweier Verschiebungen lassen sich mit Hilfe dieser Repräsentantenvektoren durch graphische Addition $+$ und durch graphische Skalierung $\cdot$ definieren, wie wir das dort schon gemacht haben und wie Sie es aus der Schule kennen. Die Menge der Verschiebungspfeile bildet in der Anschauungsebene mit $+$ und $\cdot$ einen Vektorraum. Dazu werden wir genauer am Anfang von Teil II dieses Skripts zurückkommen.

(XXX) **Kein** Vektorraum ist der zweidimensionale oder dreidimensionale Anschauungsraum $\mathcal{A}_2$ bzw. $\mathcal{A}_3$. Dieser lässt sich mathematisch als ein so genannter affiner Punktraum beschreiben, d.h. es handelt sich an sich erstmal „nur“ eine Menge (von Punkten), auf der man sich im graphischen Sinne mit Hilfe von Verschiebungen, also mit Hilfe des Vektorraumes aus (viii) bewegen kann. Auch zur Beschreibung von Punkten, geometrischer Objekte usw. werden die Verschiebungsvektoren zur Hilfe genommen und führen zur Analytischen Geometrie. Diese Formulierungen sind hier relativ schwammig und unpräzise; das gesamte Thema schauen wir uns genauer am Anfang der LAAG II an.

Im Folgenden werden wir die so genannte Vektorraumtheorie entwickeln. In dieser werden wir Aussagen formulieren, die allgemein in jedem Vektorraum richtig sind, also Aussagen der Form „Sei $(V, +, \cdot)$ ein $\mathbb{K}$-Vektorraum. Dann gilt: [...]“. Damit wir eine solche Allgemeingültigkeit garantieren können, dürfen wir bei unseren Beweisen nur die Vektorraumaxiome aus Definition 8.1.1 und schon bewiesene Folgerungen daraus benutzen.

Erste Rechenregeln für Umformungen von Gleichungen von Vektoren ermitteln wir im folgenden Lemma; einen guten Eindruck, welche Umformungen in Vektorräumen überhaupt erlaubt sind, gibt die Aufgabe 1 dieses Kapitels, in der gezeigt wird, dass schon die Umformung

$$u = a \cdot v + b \cdot w \qquad \Longleftrightarrow \qquad v = \frac{1}{a} \cdot u - \frac{b}{a} \cdot w.$$

(für $u, v, w \in V, a \in \mathbb{K}, a \neq 0$) eine Äquivalenzumformung ist – um sie zu rechtfertigen, werden aber alle Vektorraumaxiome sowie einige der Rechenregeln benötigt, die wir als nächstes aus den Axiomen ableiten:

Lemma 8.1.2. *(Rechenregeln in Vektorräumen)*

Ist V ein $\mathbb{K}$-Vektorraum und sind 0 bzw. 0_V die Nullelemente von $(\mathbb{K}, +)$ bzw. von $(V, +)$, dann gelten:

(i) $0 \cdot v = 0_V$ *für alle $v \in V$,*

(ii) $a \cdot 0_V = 0_V$ *für alle $a \in \mathbb{K}$, und*

(iii) $(-a) \cdot v = a \cdot (-v) = -(a \cdot v)$ *für alle $v \in V$ und $a \in \mathbb{K}$.*

(iv) Sind $a \in \mathbb{K}$, $v \in V$ und $a \cot v = 0$, so muss $a = 0$ oder $v = 0_V$ sein.

Beachte: In (iii) ist mit $-(a \cdot v)$ die additive Inverse des Vektors $a \cdot v$ in V gemeint.

Beweis:
(i) Für alle $v \in V$ gilt $0 \cdot v = (0+0) \cdot v = 0 \cdot v + 0 \cdot v$ (Nach welchen VR-Axiomen?). Nach Addition von $-(0 \cdot v)$ auf beiden Seiten dieser Identität erhalten wir, da $(V, +)$ eine Gruppe ist, $0_V = 0 \cdot v$.

(ii) Für alle $a \in \mathbb{K}$ gilt $a \cdot 0_V = a \cdot (0_V + 0_V) = a \cdot 0_V + a \cdot 0_V$. Nach Addition von $-(a \cdot 0_V)$ auf beiden Seiten dieser Identität erhalten wir $0_V = a \cdot 0_V$.

(iii) Für alle $a \in \mathbb{K}$ und $v \in V$ gilt $a \cdot v + (-a) \cdot v = (a-a) \cdot v = 0 \cdot v = 0_V$, sowie $a \cdot v + a \cdot (-v) = a \cdot (v - v) = a \cdot 0_V = 0_V$ (iv) Wir zeigen die logisch formal äquivalente Aussage: Ist $a \cdot v = 0$ und $a \neq 0$, so gilt $v = 0_V$. Unter dieser Voraussetzung folgt aus $a \cdot v = 0$ durch Anwendung der Vektorraumaxiome und (i), dass

$$v = 1 \cdot v = (\frac{1}{a} \cdot a) \cdot v = \frac{1}{a} \cdot (a \cdot v) = \frac{1}{a} \cdot 0_V = 0_V.$$

□

Im Folgenden ist mit der Abkürzung V immer ein $\mathbb{K}$-Vektorraum $(V, +, \cdot)$ gemeint. Außerdem werden wir anstatt $0_{\mathbb{K}}$ und 0_V oft einfach 0 schreiben – es sollte stets durch Nachdenken ersichtlich sein, welches Nullelement gemeint ist.

Vorwort zur Definition: Im $\mathbb{R}^3$ „erzeugen" Vektoren Ursprungsgeraden und Ursprungsebenen (sie „spannen sie auf"), formal heißt das: Die Menge aller möglichen Linearkombinationen aus gegebenen Vektoren lassen sich anschaulich als Ursprungsgeraden, -ebenen deuten, oder sie „spannen gleich den gesamten $\mathbb{R}^3$ auf". Die formale Verallgemeinerung dieser Anschauung führt zur Definition von zwei extrem wichtigen Begriffen der Vektorraumtheorie:

Definition 8.1.3. *(Linearkombinationen und Aufspann)*

Sei V ein $\mathbb{K}$-Vektorraum und $v_1, \ldots, v_n \in V$.

- *Jeder Vektor $v \in V$, der sich in der Form*

$$v \quad = \quad \sum_{i=1}^{n} a_i v_i \quad = \quad a_1 \cdot v_1 + \ldots + a_n \cdot v_n$$

mit $a_1, \ldots, a_n \in \mathbb{K}$ schreiben lässt, wird als <u>Linearkombination</u> der Vektoren $v_1, \ldots, v_n$ bezeichnet. Alternative Sprechweisen: „v lässt sich durch $v_1, \ldots, v_n$ erzeugen", „darstellen", „als Linearkombination von $v_1, \ldots, v_n$ schreiben", „ist aus $v_1, \ldots, v_n$ linear kombinierbar",...

- *Die Menge aller möglicher Linearkombinationen von $v_1, \ldots, v_n$, also die Menge*

$$\mathcal{L}(v_1, \ldots, v_n) := \{v \in V \mid v \textit{ ist Linearkombination von } v_1, \ldots, v_n\}$$

  heißt Aufspann, Spann, lineare Hülle oder (lineares) Erzeugnis von $\{v_1, \ldots, v_n\}$.

- *Allgemeiner definieren wir für eine Teilmenge $M \subseteq V$ den Aufspann von M als*

$$\mathcal{L}(M) := \{v \in V \mid \exists v_1, \ldots, v_m \in M, a_1, \ldots, a_m \in \mathbb{K} : v = \sum_{i=1}^{m} a_i v_i\}.$$

Bemerkung:

- Man beachte, dass eine Linearkombination im Sinne der Linearen Algebra *immer* aus *endlich vielen* Vektoren gebildet wird, wie in 8.1.3 gefordert. Reihendarstellungen wie die Taylorreihen $e^x = \sum_{i=0}^{\infty} \frac{1}{n!} x^n, \sin(x) = \sum_{i=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1}$, die in der Analysis gebräuchlich sind, könnte man zwar im erweiterten Sinne als Linearkombination von unendlich vielen Polynomen ansehen, tut dies aber nicht – hier ist ja von der Warte der Algebra zunächst unklar, was so eine Summe aus unendlich vielen Summanden überhaupt sein soll, und wann und wie diese vernünftig definiert ist. Solche Fragen nach der Sinnhaftigkeit solcher unendlicher Summen, d.h. nach Konvergenz von Folgen und Reihen und auch nach der Darstellbarkeit bzw. Approximierbarkeit von z.B. stetigen Funktionen durch andere Sätze von Funktionen (hier: Polynome) erfordern einige zusätzliche Begriffe und Hilfsmittel und sind das „Kernbusiness“ der Analysis.

Definition 8.1.4. *(Unterraum)*

Sei $(V, +, \cdot)$ ein Vektorraum und $U \subseteq V$. Ist das Tripel $(U, +, \cdot)$ mit der von V „vererbten“ Addition und skalaren Multiplikation selbst wieder ein Vektorraum, so heißt U (linearer) Teilraum, Unterraum, Untervektorraum oder kurz UVR von V.

Bemerkungen dazu:

- Eine Teilmenge ist nicht zwingend ein Teilraum. Das liegt daran, dass in Teilräumen U die Addition und die skalare Multiplikation nach Definition eines Vektorraumes Abbildungen in U sein müssen. Finden Sie eine Teilmenge von $\mathbb{R}^n$, die kein Teilraum ist!

- Ist U wiederum ein Unterraum und $u \in U$, so muss nach den obigen Überlegungen $a \cdot u \in U$ für alle $a \in \mathbb{K}$ gelten. $\rightsquigarrow$ Mit einem Vektor u enthält ein Unterraum gleich alle seine skalaren Vielfachen, also anschaulich gesehen gleich die gesamte von u erzeugte Gerade. Genauso gilt: Die Addition $+: U \times U \to U$ muss wieder nach U abbilden, ein Untervektorraum enthält also mit je zwei Vektoren $u_1, u_2 \in U$ auch immer ihre Summe $u_1 + u_2$, und in Kombination mit der vorherigen Beobachtung also den gesamten Aufspann $\mathcal{L}(u_1, u_2)$.

Es gilt auch die Umkehrung, die wir jetzt beweisen.

Satz 8.1.5. *(Unterraumkriterium)*

Eine Teilmenge $U \subseteq V$ ist ein Teil<u>raum</u> eines Vektorraumes V genau dann, wenn die folgenden drei Kriterien erfüllt sind:

(i) $0_V \in U$,

(ii) Für alle $u \in U$ und alle $a \in \mathbb{K}$ ist $a \cdot u \in U$.

(iii) Für alle $u_1, u_2 \in U$ ist $u_1 + u_2 \in U$.

Bemerkung: Will man zeigen, dass eine Teilmenge U eines Vektorraumes V ein Teil<u>raum</u> ist, lassen sich die Kriterien (ii) und (iii) sogar dadurch ersetzen, dass man beweist, dass $u_1 + a \cdot u_2 \in U$ für alle $a \in \mathbb{K}$ und alle $u_1, u_2 \in U$ gilt (Übungsaufgabe).

Beweis: Die „Hinrichtung“ der Äquivalenz wurde schon oben begründet, für die Rückrichtung, dass aus der Gültigkeit von (i), (ii) und (iii) folgt, dass U selbst ein Vektorraum ist, s. Vorlesung oder jedes Lineare-Algebra-Buch. Kurzerläuterung: (i) schließt $\emptyset$ als Unterraum aus ($\emptyset$ ist kein Vektorraum, würde aber die Kriterien (ii) und (iii) erfüllen). Das Kriterien (ii) und (iii) sichern die Abgeschlossenheit bzgl. + und $\cdot$. Der Rest der geforderten Axiome vererbt sich von V. □

Beispiele für Unterräume:

- Ist V ein beliebiger Vektorraum, so sind $\{0_V\}$ und V immer Unterräume von V.
- Der Aufspann eines Vektors $0_V \neq v \in \mathbb{R}^3$ bzw. zweier nicht kollinearer Vektoren $\vec{v}, \vec{w} \in \mathbb{R}^3$ (deutbar als „Ursprungsgeraden“ und „Ursprungsebenen“) sind Unterräume des $\mathbb{R}^3$. Vom Ursprung verschobene Geraden und Ebenen im Anschauungsraum sind dagegen **keine** Untervektorräume (sondern so genannte affine Unterräume, später!)

- Teilräume von Funktionenräumen, also von $\mathrm{Abb}(X, V)$ für eine Vielzahl von Mengen X und Vektorräumen V, z.B.
  - von $\mathrm{Abb}(\mathbb{R}, \mathbb{R})$: Raum der stetigen Funktionen, der differenzierbaren Funktionen, der stückweise linearen Funktionen, der Polynomfunktionen, der trigonometrischen Funktionen, der Treppenfunktionen, ...
  - von $\mathrm{Abb}(\mathbb{N}, \mathbb{R})$ (aller Funktionen $\mathbb{N} \to \mathbb{R}$,das sind die Folgen in $\mathbb{R}$): Raum der konvergenten Folgen, Raum der Nullfolgen, Raum der absolutsummierbaren Folgen, ...
- Sind U_1, U_2 Unterräume von V, so sind $U_1 \cap U_2$ und die Menge

$$U_1 + U_2 := \{u_1 + u_2 \mid u_1 \in U_1, u_2 \in U_2\}$$

  ein Untervektorraum, $U_1 \cup U_2$ im Allgemeinen nicht (Übungsaufgabe).
- Ist $A \in \mathbb{R}^{m \times n}$, so ist die Lösungsmenge

$$\mathbb{L}(A, 0) = \{x \in \mathbb{R}^n \mid A \cdot x = 0\}$$

  des linearen Gleichungssystems $A \cdot x = 0$ (also mit rechter Seite $b = 0 \in \mathbb{R}^m$) ein Unterraum des $\mathbb{R}^n$. Für $b \neq 0$ gilt das nicht (Übungsaufgabe).

Ein weiteres Beispiel wollen wir wegen seiner Bedeutung für den folgenden Vorlesungsstoff gesondert betrachten.

Lemma 8.1.6. *(Das Erzeugnis von Vektoren ist ein Vektorraum)*

Sei V ein Vektorraum und $M \subseteq V$ eine nichtleere Menge von Vektoren. Dann ist $\mathcal{L}(M)$ ein Unterraum von V, der so genannte von M erzeugte Unterraum in V.
$\mathcal{L}(M)$ ist der kleinste Unterraum, der M enthält, d.h. für jeden Unterraum $U \subseteq V$ mit $M \subseteq U$ gilt immer schon $\mathcal{L}(M)$.

Beweis: Das folgt aus dem Teilraumkriterium, ausführlicher Beweis s. Vorlesung. □

Wie eben schon betont ist der Begriff des Erzeugnisses einer der wichtigen Grundbegriffe der Linearen Algebra. Mit ihm eng verwandt ist der folgende Begriff, bei dem man nicht danach fragt, was das Erzeugnis einer gegebenen Menge ist, sondern umgekehrt nach einer Menge sucht, die einen gegebenen Vektorraum, z.B. den $\mathbb{R}^3$, erzeugt.

Definition 8.1.7. *(Erzeugendensysteme)*

Eine Teilmenge $M \subseteq V$ eines Vektorraumes V heißt <u>Erzeugendensystem</u> von V, wenn $V = \mathcal{L}(M)$ ist. Man sagt auch, M erzeugt V oder spannt V auf.

Beispiele:

- Jede Teilmenge M eines Vektorraumes V ist also Erzeugendensystem ihres Erzeugnisses $\mathcal{L}(M)$ (das nach 8.1.6 ein Teilraum von V, also selbst ein Vektorraum ist).
- $e_1, \ldots, e_n$ aus (8.1) ist ein Erzeugendensystem von $\mathbb{K}^n$.
- Jede Teilmenge von $\mathbb{K}^n$, die $e_1, \ldots, e_n$ enthält, ist ebenfalls Erzeugendensystem von $\mathbb{K}^n$.
- Die Vektoren $\begin{pmatrix}1\\2\\3\end{pmatrix}, \begin{pmatrix}4\\5\\6\end{pmatrix}, \begin{pmatrix}7\\8\\9\end{pmatrix} \in \mathbb{R}^3$ sind kein Erzeugendensystem des $\mathbb{R}^3$, denn beispielsweise gilt
$$\begin{pmatrix}1\\2\\4\end{pmatrix} \notin \mathcal{L}\left(\begin{pmatrix}1\\2\\3\end{pmatrix}, \begin{pmatrix}4\\5\\6\end{pmatrix}, \begin{pmatrix}7\\8\\9\end{pmatrix}\right),$$
wie man anhand des entsprechenden LGS erkennt.
- Die Monomfunktionen sind ein Erzeugendensystem des Vektorraumes aller Polynomfunktionen (denn jede Polynomfunktion ist eine (per Definition endliche) Linearkombination dieser Funktionen). Sie sind aber kein Erzeugendensystem von $\mathrm{Abb}(\mathbb{R}, \mathbb{R})$, denn nicht jede Funktion $f : \mathbb{R} \to \mathbb{R}$ ist eine (per Definition endliche) Linearkombination von Monomen. Das hieße ja, dass jede solche Funktion sich als Polynom schrebien lässt. Das könnte ja sein, ist aber nicht so: Jede Funktion mit unendlich vielen Nullstellen (wie z.B. der Sinus) kann nach dem Fundamentalsatz kein Polynom, also keine Linearkombination von Monomen sein. Wir werden zeigen, dass auch die Monomfunktionen ein minimales Erzeugendensystem für den Raum der Polynomfunktionen sind.

- Jeder Vektorraum erzeugt trivialerweise sich selbst, d.h. es gilt $\mathcal{L}(V) = V$. Das ist aber nicht Sinn der Sache. Wir werden meist nach *minimalen* Erzeugendensystemen suchen (also nach solchen mit möglichst wenigen Elementen, in denen sozusagen „keine Redundanzen bestehen“, die aber trotzdem den gesamten betrachteten Raum erzeugen). Auf solchen nichtredundanten, minimalen Erzeugendensystemen (keine Sorge, das wird unten alles noch genauer formuliert!) lässt sich dann im Sinne der Linearen Algebra „gut arbeiten“. Ein (sehr wichtiges!) Beispiel dafür haben wir schon gesehen: Die obige Standardbasis im $\mathbb{K}^n$ ist solch ein minimales Erzeugendensystem.

8.2 Lineare Unabhängigkeit und Dimension

In diesem Abschnitt beschäftigen wir uns mit einem weitere grundlegenden Begriff, der die am Schluss des letzten Abschnitts erwähnten „Redundanzen“ in Auswahlen von Vektoren formalisiert.

Eine Vorüberlegung zur folgenden Definition: Ein Vektor v spannt anschaulich etwas eindimensionales auf, es sei denn, es handelt sich bei v um den Nullvektor. Zwei bzw. drei Vektoren spannen etwas zwei- bzw. dreidimensionales auf, es sei denn, sie sind kollinear bzw. komplanar. $\rightsquigarrow$ Lineare Unabhängigkeit wollen wir definieren als: n linear unabhängige Vektoren „spannen etwas n-dimensionales auf“, d.h. es gibt keine Abhängigkeiten oder Redundanzen.

Definition 8.2.1. *(Lineare Unabhängigkeit)*

Sei V ein Vektorraum.

- *Einen Vektor $v \in V$ (also ein 1-Tupel von Vektoren) nennen wir linear unabhängig, wenn $v \neq 0_V$ gilt.*
- *Für $k \geq 2$ heißen endlich viele Vektoren $v_1, \ldots, v_k \in V$, linear unabhängig (oft kurz: l.u.), wenn sich keiner der Vektoren als Linearkombination der übrigen ausdrücken lässt.*

Sind Vektoren $v_1, \ldots, v_k \in V$ (mit $k \geq 1$) nicht linear unabhängig, so nennt man sie linear abhängig (oft kurz: l.a.).

Bemerkungen dazu:

- Die lineare Unabhängigkeit/Abhängigkeit endlich vieler Vektoren $v_1, \ldots, v_k$ aus V lässt sich etwas ausführlicher so formulieren: Ist v_i ein beliebiger dieser Vektoren, so lässt sich v_i nicht in der Form
$$v_i = \sum_{\substack{j=1 \\ j \neq i}}^{k} a_j v_j$$
mit $a_1, \ldots, a_{j-1}, a_{j+1}, \ldots, a_k \in \mathbb{K}$ schreiben. Sind die Vektoren $v_1, \ldots, v_k$ hingegen linear abhängig, so gibt es unter ihnen mindestens einen Vektor v_i, für den
$$v_i \in \mathcal{L}(v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_k)$$
gilt.

- In der Aufzählung „$v_1, \ldots, v_k \in V$" in der Definition der linearen Unabhängigkeit dieser Vektoren wollen wir auch erlauben, dass zwei dieser Vektoren in einer solchen Aufzählung gleich sein dürfen. Formal bedeutet das, dass der Begriff der linearen Unabhängigkeit nicht für eine Teilmenge $\{v_1, \ldots, v_k\} \subseteq V$, sondern für eine „Auswahl" von n (potenziell auch gleichen) Vektoren aus V definiert werden muss, was sich formal dadurch bewerkstelligen lässt, dass man k-Tupel $(v_1, \ldots, v_k)$ von Vektoren aus V betrachtet. Ungünstigerweise findet man in der Literatur oft die Definition einer „linear unabhängigen Menge" $\{v_1, \ldots, v_n\}$, die diesen Problem dann später weitestgehend ignoriert und z.B. Vektoren wie $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ und $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ als linear abhängig bezeichnet, obwohl die Menge ja nur den Vektor $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ enthält, also linear unabhängig sein müsste.
- Auch die teils zu hörende Sprechweise „v ist linear (un)abhängig von $v_1, \ldots, v_k$" eine *ungünstige, in der Fachliteratur nicht gängige* Umschreibung dafür, dass „v (k)eine Linearkombination von $v_1, \ldots, v_k$" ist.

Beispiele:

- Ist $\mathbb{K}$ ein Körper, so sind die Vektoren
$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \tag{8.1}$$
linear unabhängig in $\mathbb{K}^n$.
- Man stellt fest: Für Fragen wie
  - Sind die Vektoren $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}, \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix} \in \mathbb{R}^3$ linear unabhängig?
  - Sind in $\mathrm{Abb}(\mathbb{R}, \mathbb{R})$ sind die ersten $n+1$ Monomfunktionen
$$p_n : x \mapsto x^n, \quad p_{n-1} : x \mapsto x^{n-1}, \ldots, \quad p_1(x) : x \mapsto x^1, \quad p_0 : x \mapsto 1 = x^0$$
linear unabhängig?

ist die obige Definition recht unhandlich. Es hilft der folgende Satz.

Satz 8.2.2. *(Kriterium für lineare Unabhängigkeit)*

Sei V ein Vektorraum, $k \in \mathbb{N}$ und $v_1, \ldots, v_k \in V$. Es sind äquivalent:

- *Die Vektoren $v_1, \ldots, v_k$ sind linear unabhängig,*
- *Die Gleichung*

$$\sum_{i=1}^{k} a_i v_i = 0_V \tag{8.2}$$

hat nur die Lösung $a_1 = a_2 = \ldots = a_k = 0$.

Beweis: S. Vorlesung. Skizze: Für $n = 1$ klar. Für $n \geq 2$, Hinrichtung: Stelle Gleichung für als LK schreibbaren Vektor um auf Form aus (ii) mit Vorfaktor 1, Rückrichtung: Stelle nach Vektor mit Nicht-Null-Koeffizient um, teile durch den zugehörigen Vorfaktor. Alle Operationen sind durch die Vektorraum- und Körperaxiome legitimiert. □

Bemerkungen dazu:

- Das obige Kriterium wird (H. Weyl folgend) in vielen Lehrbüchern als Definition der linearen Unabhängigkeit gewählt. Diese Definition ist aber wenig *genetisch*, d.h. man sieht den Grund für die Begriffsbildung nicht mehr sehr deutlich. So etwas kommt in den Begriffsbildungen der Unimathematik häufiger vor, ich möchte so etwas in diesem Skript und der Vorlesung dazu aber, wenn es geht, vermeiden.
- Die Gleichung (8.2) ist für Vektoren $v_1, \ldots, v_k \in \mathbb{K}^m$ ein homogenes lineares Gleichungssystem $V \cdot x = 0$ mit Koeffizientenmatrix $V \in \mathbb{R}^{m \times k}$, die die Vektoren $v_1, \ldots, v_k$ in ihren Spalten trägt, und mit $x = (a_1, \ldots, a_k)$. Der Satz $v_1, \ldots, v_k$ ist also nach Satz 6.3.3(iv) genau dann linear unabhängig, wenn $\mathrm{rg}(V) = k$ gilt. Das wiederum lässt sich mit dem Gauß-Jordan-Algorithmus gut überprüfen.

Beispiele:

- Sind die Vektoren $\begin{pmatrix}1\\2\\3\end{pmatrix}, \begin{pmatrix}4\\5\\6\end{pmatrix}, \begin{pmatrix}7\\8\\9\end{pmatrix} \in \mathbb{R}^3$ linear unabhängig?
- In $\mathrm{Abb}(\mathbb{R}, \mathbb{R})$ sind die ersten $n+1$ Monomfunktionen

  $$p_n : x \mapsto x^n, \quad p_{n-1} : x \mapsto x^{n-1}, \ldots, \quad p_1(x) : x \mapsto x^1, \quad p_0 : x \mapsto 1 = x^0$$

  linear unabhängig. (Beweis: Sucht man Koeffizienten $a_i, i \in \underline{n}$, mit

  $$\sum_{i=0}^{n} a_i \cdot x^i = 0,$$

  wobei 0 die Nullfunktion im Vektorraum $\mathrm{Abb}(\mathbb{R}, \mathbb{R})$ ist, so kann das nur für den Fall $a_i = 0$ funktionieren – ansonsten steht auf der linken Seite entweder ein konstantes Polynom ohne Nullstellen, oder ein Polynom von Grad $n \geq 0$, das nach dem Fundamentalsatz der Algebra höchstens n Nullstellen hat, rechts aber eine Funktion mit unendlich vielen Nullstellen.)

Bevor wir uns dem Begriff der Basis eines Vektorraumes zuwenden, halten wir in den folgenden zwei Lemmata noch zwei wichtige Eigenschaften der linearen Unabhängigkeit fest, die wir später oft nutzen werden werden:

Lemma 8.2.3. *(Eindeutige Darstellbarkeit des Erzeugnisses und l.u. Vektoren)*
Seien $v_1, \ldots, v_k$ Vektoren aus einem Vektorraum V. Dann sind äquivalent:

(i) Die Vektoren $v_1, \ldots, v_k$ sind linear unabhängig.

(ii) Jeder Vektor $v \in \mathcal{L}(v_1, \ldots, v_k)$ lässt sich eindeutig als Linearkombination der Vektoren $v_1, \ldots, v_k$ schreiben, d.h. für alle $v \in \mathcal{L}(v_1, \ldots, v_k)$ gibt es genau einen Satz an Koeffizienten $a_1, \ldots, a_k \in \mathbb{K}$ so, dass $\sum_{i=1}^{k} a_i \cdot v_i$ gilt.

Beweis: Wir zeigen $\neg(ii) \Longrightarrow \neg(i)$ und $\neg(i) \Longrightarrow \neg(ii)$. Sind

$$v = \sum_{i=1}^{k} a_i \cdot v_i = \sum_{i=1}^{k} c_i \cdot v_i$$

zwei verschiedene Darstellungen eines Vektors $v \in \mathcal{L}(v_1, \ldots, v_k)$, so gilt

$$\sum_{i=1}^{k}(a_i - c_i) \cdot v_i = 0_V,$$

aber $a_i - c_i \neq 0$ für mindestens ein $i \in \underline{k}$, die Vektoren $v_1, \ldots, v_k$ sind also linear abhängig. Ist umgekehrt für $\neg(i)$ o.B.d.A. v_1 als Linearkombination von $v_2, \ldots, v_n$ schreibbar, so sind $v_1 = 1 \cdot v_1$ und $v_1 = \sum_{i=2}^{k} a_i \cdot v_i$ zwei offensichtlich verschiedene Darstellungen für $v_1 \in \mathcal{L}(v_1, \ldots, v_k)$ als Linearkombination von $v_1, \ldots, v_k$.

□

Lemma 8.2.4. *(Ergänzen linear unabhängiger Vektoren)*

Seien $v_1, \ldots, v_k$ linear unabhängige Vektoren aus einem Vektorraum V und $w \in V$. Dann sind äquivalent:

(i) Die Vektoren $v_1, \ldots, v_k, w$ sind linear unabhängig.

(ii) Es ist $w \notin \mathcal{L}(v_1, \ldots, v_k)$.

Beweis: Wir zeigen wieder die Kontrapositionen $\neg(i) \Longrightarrow \neg(ii)$ und $\neg(ii) \Longrightarrow \neg(i)$. Sind $v_1, \ldots, v_k, w$ linear abhängig, so gibt es eine Darstellung $\sum_{i=1}^{k} a_i \cdot v_i + a_{k+1}w = 0_V$ des Nullvektors, bei der mindestens einer der Vorfaktoren ungleich Null ist. Wäre $a_{k+1} = 0$, so würde das nach dem Kriterium für lineare Unabhängigkeit bedeuten, dass die Vektoren $v_1, \ldots, v_k$ l.a. sind, was aber vorausgesetzt wird. Also ist $a_{k+1} \neq 0$ und daher

$$w = \sum_{i=1}^{k}\left(-\frac{a_i}{a_{k+1}}\right) \cdot v_i$$

eine Darstellung von w als Linearkombination von $v_1, \ldots, v_k$. Ist umgekehrt $\neg(2)$ erfüllt, d.h. $w \in \mathcal{L}(v_1, \ldots, v_k)$, so ist w Linearkombination der Vektoren $v_1, \ldots, v_k$, also $v_1, \ldots, v_k, w$ linear abhängig.

□

Zur Vorbereitung der folgenden Definition: Anschaulich ist hoffentlich klar, dass man im Raum zu drei linear unabhängigen Vektoren keinen vierten hinzufügen kann, ohne dass das entstehende Tupel linear abhängig wird. Wir nehmen diese Anschauung als Grundlage für die folgende Definition des Dimensionsbegriffs für Vektorräume. Diese weicht von der gängigen Definition als Anzahl der Elemente einer Basis ab – sie folgt eher der Definition, wie sie sich bei Weyl finden lässt, benötigt netterweise zunächst keine Resultate über Mächtigkeiten von Basen von Vektorräumen (s. nächster Abschnitt) und führt dazu, dass einige der folgenden Sachverhalte zugehörigen Beweise meines Erachten natürlicher erscheinen.

Definition 8.2.5. *(Dimension eines Vektorraumes)*

Sei V ein Vektorraum. Die größte Zahl $n \in \mathbb{N}_0$, zu der es n linear unabhängige Vektoren $v_1, \ldots, v_n \in V$ gibt, nennt man die <u>Dimension von V</u>, kurz $\dim(V)$.

Gibt es eine solche Zahl nicht, so heißt V <u>unendlichdimensional</u>, kurz $\dim(V) = \infty$.

Bemerkungen dazu:

- Die Dimension des trivialen Unterraumes $\{0_V\}$ eines Vektorraumes V ist nach dieser Definition immer null.
- Die Dimension des $\mathbb{K}$-Vektorraums $\mathbb{K}^n$ ist n,[1] denn die Vektoren der Standardbasis sind linear unabhängig, also ist $\dim(\mathbb{K}^n) \geq n$, und nach der zweiten Bemerkungen nach Satz 8.2.2 sind k Vektoren im $\mathbb{K}^n$ linear unabhängig genau dann, wenn für die Matrix M mit diesen Vektoren als Spalten $rg(M) = k$ gilt. Da aber auch immer $rg(M)$ höchstens die Anzahl der n Zeilen von M sein kann, können mehr als n Vektoren im $\mathbb{K}^n$ nicht linear unabhängig sein.
- Der Vektorraum $\mathrm{Abb}(\mathbb{R}, \mathbb{R})$ ist unendlichdimensional, da für jedes $n \in \mathbb{N}$ die ersten $n+1$ Monomfunktionen

  $$p_n : x \mapsto x^n, \quad p_{n-1} : x \mapsto x^{n-1}, \ldots, \quad p_1(x) : x \mapsto x^1, \quad p_0 : x \mapsto 1 = x^0$$

  linear unabhängig sind, man also zu jeder Zahl n, die als Dimension in Frage käme, $n+1$ linear unabhängige Monomen finden kann. Die Untersuchung solcher unendlichdimensionaler Vektorräume (v.a. von Funktionenräumen, also von Vektorräumen, deren Elemente Funktionen sind) gehört eher in die Analysis, genauer gesagt in die sogenannte Funktionalanalysis.
- Wir werden im nächsten Abschnitt systematische Mittel und Wege kennenlernen, um die Dimension eines Vektorraumes zu bestimmen.
- Auch Teilräumen von Vektorräumen kann man eine Dimension zuweisen, denn sie sind ja selbst Vektorräume.

[1] Vorsicht! Hier ist wichtig, dass $\mathbb{K}^n$ als $\mathbb{K}$-Vektorraum betrachtet wird; betrachtet man etwa $\mathbb{C}^n$ als $\mathbb{R}$-Vektorraum, so stimmt das nicht.

8.3 Basen

Ab hier werden wir uns, wie in der Linearen Algebra üblich, weitestgehend auf endlichdimensionale $\mathbb{K}$-Vektorräume beschränken.

Definition 8.3.1. *(Basis, Koordinaten(-vektor))*

Sei V ein endlichdimensionaler $\mathbb{K}$-Vektorraum. Ein n-Tupel $B = (b_1, \ldots, b_n)$ von Vektoren aus V heißt (manchmal: geordnete) Basis von V, wenn sich jeder Vektor $v \in V$ eindeutig als Linearkombination der Vektoren $b_1, \ldots, b_n$ darstellen lässt, d.h. es gibt zu jedem Vektor $v \in V$ genau einen zugehörigen Satz an Koeffizienten $a_1, \ldots, a_n \in \mathbb{K}$ so, dass $\sum_{i=1}^n a_i \cdot b_i$ gilt.

Die Vorfaktoren $a_1, \ldots, a_n \in \mathbb{K}$ in dieser Darstellung heißen die Koordinaten von v bezüglich B, der Vektor $a = (a_i)_{i=1}^n \in \mathbb{K}^n$ heißt Koordinatenvektor von v bezüglich B.

Im Folgenden wollen wir untersuchen, welche Eigenschaften ein n-Tupel B von Vektoren haben muss, damit die geforderte eindeutigen Darstellbarkeit garantiert werden kann. Die Eindeutigkeit der Koordinaten a_i lässt sich auch folgendermaßen formulieren:

Lemma 8.3.2. *(Koordinatenabbildung)*
Die Koordinatenabbildung bezüglich B,

$$k_B : V \to \mathbb{K}^n, \qquad v \mapsto \textit{Koordinatenvektor von } v \textit{ bezüglich } B$$

ist wohldefiniert (d.h. sie ist, wie im Namen behauptet, eine Abbildung). Der Zusatz „bezüglich B“ ist dabei notwendig: Sind B, C Basen, so gilt i.A. $k_B(v) \neq k_C(v)$.

Da es bei der Angabe von Koordinaten auf die Reihenfolge der Zahlen $a_1, \ldots, a_n$ ankommt, werden wir geordnete Basen betrachten. Da wir eine Basis B aber auch immer als Menge von Vektoren auffassen können, werden wir im Folgenden etwas schlampig auch Notationen wie $b_i \in B$ benutzen.

Satz 8.3.3. *(Erste Charakterisierung von Basen)*

Sei V ein Vektorraum und $B = (b_1, \ldots, b_n)$ ein n-Tupel von Vektoren aus V. Dann sind äquivalent:

(i) B ist eine Basis von V, d.h. jedes Element von V lässt sich eindeutig als Linearkombination von Vektoren aus B darstellen.

(ii) B ist linear unabhängig und ein Erzeugendensystem von V.
(oft kurz: „ein linear unabhängiges Erzeugendensystem von V“)

Oft wird die Eigenschaft (ii) zur Definition des Basisbegriffs gewählt. Die Hauptmotivation für die Einführung des Begriffs und wesentliche Eigenschaft einer Basis ist aber (i).

Beweis: Nach Lemma 8.2.3 und der Definition von Erzeugendensystemen ist B ist linear unabhängig und Erzeugendensystem genau dann, wenn gleichzeitig jeder Vektor $v \in \mathcal{L}(b_1, \dots, b_n)$ eindeutig als Linearkombination der Vektoren in B schreibbar ist und $\mathcal{L}(b_1, \dots, b_n) = V$ ist, d.h. wenn B eine Basis von V ist.

□

Alternativ lassen sich Basen auch über die folgende wichtige Minimalitäts- bzw. Maximalitätseigenschaft beschreiben.

Satz 8.3.4. *(Zweite Charakterisierung von Basen)*

Sei V ein Vektorraum und $B = (b_1, \dots, b_n)$ ein n-Tupel von Vektoren aus V. Dann sind äquivalent:

(i) B ist eine Basis von V.

(ii) B ist eine maximal linear unabhängig, d.h. wird ein beliebiges weiteres Element $v \in V$ zu B hinzugefügt, so ist das neue Tupel $(b_1, \dots, b_n, v)$ nicht mehr linear unabhängig.

(iii) B ist ein minimales Erzeugendensystem von V, d.h. B ist Erzeugendensystem von V und es gibt keinen Vektor $b_i \in B$ finden, so dass $B \backslash \{b_i\}$ immer noch in Erzeugendensystem ist.

Beweis: Nach Satz 8.3.3 ist eine Basis ist ein linear unabhängiges Erzeugendensystem. Für (i) $\Leftrightarrow$ (ii) sei B also linear unabhängig, und wir zeigen, dass B genau dann maximal linear unabhängig ist, wenn B ein Erzeugendensystem ist: B ist erzeugend genau dann, wenn für alle $w \in V$ auch $w \in \mathcal{L}(B)$ gilt, was nach Lemma 8.2.4 genau dann wahr ist, wenn das Tupel $(b_1, \dots, b_n, w)$ für alle $w \in V$ linear abhängig ist, d.h. B ist maximal linear unabhängig.

Für (i) $\Leftrightarrow$ (iii) sei B ein Erzeugendensystem, und wir zeigen, dass das Tupel B linear unabhängig ist, genau dann, wenn es minimal ist, wiederum, indem wir die Kontrapositionen beider Äquivalenzen beweisen: Ist B kein minimales Erzeugendensystem, so ist o.B.d.A. $(b_2, \dots, b_n)$ Erzeugendensystem von V, also insbesondere b_1 Linearkombinnation dieser Vektoren und B somit linear abhängig. Ist B linear abhängig und o.B.d.A.

$b_1 = \sum_{i=2}^n c_i \cdot b_i$, so ist für alle $v = \sum_{i=1}^n a_i \cdot b_i \in V$ auch $v = \sum_{i=1}^n a_i \cdot b_i = \sum_{i=2}^n (a_1 c_i + a_i) \cdot b_i$ eine Darstellung in $B' = (b_2, \ldots, b_n)$, B war also nicht minimal.

□

Der obige Satz liefert Verfahren, um ausgehend von einer linear unabhängigen Menge oder einem endlichen Erzeugendensystem eine Basis eines endlichdimensionalen Vektorraumes V zu gewinnen. Das ist in den folgenden beiden Sätzen zusammengefasst:

Satz 8.3.5. *(Basisergänzungssatz)*

Sei V ein endlichdimensionaler Vektorraum und $M = (v_1, \ldots, v_k)$ linear unabhängig. Dann lässt sich M durch Hinzunehmen geeigneter Vektoren aus V zu einer Basis von V ergänzen.

Satz 8.3.6. *(Basisauswahlsatz)*

Sei V ein endlichdimensionaler Vektorraum und $M = (v_1, \ldots, v_k)$ ein Erzeugendensystem von V. Dann lässt sich M sich durch sukzessives Wegnehmen geeigneter Vektoren aus M in endlich vielen Schritten zu einer Basis „verkürzen“.

Beweis: Wir beweisen nur den Basisergänzungssatz; den Basisauswahlsatz beweist man analog. Wir setzen $B_k = (v_1, \ldots, v_k)$. Ist B_k eine Basis, sind wir fertig; wenn nicht, ist B_k kein Erzeugendensystem, also existiert ein $w \in V$ mit $w \notin \mathcal{L}(B_k)$. Nach Übungsaufgabe ist $B_{k+1} = (v_1, \ldots, v_k, w)$ linear unabhängig. Nun wiederholen wir den obigen Schritt. Das Verfahren endet entweder, wenn B_i eine Basis ist, oder spätestens nach $\dim(V) - k$ Schritten, denn dann haben wir einen l.u. Satz von $\dim(V)$ Vektoren, der nach Definition von $\dim(V)$ maximal ist. B_i ist also nach 8.3.4 die gesuchte Basis von V.

□

Jedes linear unabhängige Tupel in einem Vektorraum V lässt sich also durch Hinzunehmen geeigneter Vektoren zu einer Basis verlängern, umgekehrt geht das auch bei Erzeugendensystemen durch Wegnehmen geeigneter Vektoren. Das sich ergebende Verfahren ist auch

Schritt für Schritt in der Praxis durchführbar, aber meist recht mühsam. Ein besseres Verfahren, das sich aus der Theorie der kommenden Abschnitte ergibt, werden wir in 11.3 sehen.

Wir wollen uns zunächst einer damit zusammenhängenden theoretischen Frage widmen: Unklar ist bisher, ob dieses Hinzufüg- bzw. Wegnehmverfahren immer bei *gleich vielen* Vektoren endet, d.h. ob *alle* Basen eines Vektorraumes V *genau* $\dim(V)$ Elemente enthalten.
Das zu zeigen, ist das Hauptanliegen dieses restlichen Abschnittes.
Eine erste Einsicht auf dem Weg ist die folgende: Da eine Basis linear unabhängig ist, kann jede Basis nach Definition der Dimension höchstens $\dim(V)$ Elemente enthalten. Genauer gilt der folgende Satz:

Satz 8.3.7. *Jeder endlichdimensionale Vektorraum $V \neq \{0\}$ besitzt mindestens eine Basis mit genau $\dim(V)$ Elementen.*

Beweis: Wegen $\dim(V) = n$ existiert eine linear unabhängiges n-Tupel $(b_1, \ldots, b_n)$ aus Vektoren $b_i \in V$. Nach Definition der Dimension ist diese Anzahl maximal, d.h. $(b_1, \ldots, b_n)$ ist eine Basis nach Satz 8.3.3 (iv).

□

Bemerkung (Die Fälle dim$(V) = \infty$, dim$(V) = 0$):

- Auch jeder unendlichdimensionale Vektorraum besitzt eine Basis im obigen Sinne. Der Beweis basiert auf dem Zorn'schen Lemma, einem Axiom der ZFC-Mengenlehre (vgl. Abschnitt 7.2), das hier nicht weiter thematisiert werden soll. Für eine große Klasse von unendlichdimensionalen Räumen (so genannte Banachräume, Satz von Baire) lässt sich allerdings auch zeigen, dass solche Basen immer überabzählbar viele Basisvektoren enthalten müssen.

- In der Analysis betrachtet man im Gegensatz zur Linearen Algebra *analytische* Basen (Schauder-Basen) von unendlichdimensionalen Räumen. Mit Hilfe einer analytischen Basis werden Elemente von solchen Räumen, z.B. von Funktionenräumen, als Grenzwerte von unendlichen Reihen, also quasi als *unendliche* Summe von Basisfunktionen dargestellt werden, z.B. $f(x) = \sum_{k=0}^{\infty} a_k x^k$, $f(x) = \sum_{k=0}^{\infty} a_k \sin(k \cdot x)$ (Das erste ist eine so genannte Potenzreihe/Taylorreihe, das zweite eine spezielle Fourierreihe). Hier spielen wieder verschiedene Grenzwertbegriffe eine Rolle. Dieses Setting ist in der Linearen Algebra explizit <u>nicht</u> zugelassen, in der letzten Bemerkung ist eine Basis gemeint, in der sich jedes Vektorraumelement als <u>endliche</u> Summe von Basiselementen schreiben lässt, auch wenn diese Basis sehr viele Elemente enthält.

- Der triviale Vektorraum $V = \{0\}$ hat nach unserer Definition die Dimension 0. Trotzdem lässt sich eine von einem formalen Standpunkt sinnvolle Basis finden: Definiert man in einer Gruppe $(M, +)$ abstrakt die leere Summe in M, also die Summe über keinen Summanden, als $0 \in M$, so lässt sich die leere Menge $\emptyset$ formal als Basis von $V = \{0\}$ auffassen, denn der Nullvektor lässt sich über die leere Summe linear kombinieren. Das kann man als Taschenspielertrick für diesen pathologischen Fall abtun, es rundet aber das Bild ab.

Das zentrale Resultat unserer theoretischen Überlegungen ist nun das Folgende:

Satz 8.3.8. *Jede Basis eines n-dimensionalen Vektorraumes V hat genau n Elemente.*

Der Beweis folgt gleich, zunächst

Bemerkungen dazu:

- In vielen Vorlesungen wird das Resultat, dass Basen immer gleich viele Elemente haben, genutzt, um diese gemeinsame Zahl dann als Dimension eines Vektorraumes zu definieren. Ich finde meine gleichwertige Darstellung, die die Definition über die maximal mögliche Anzahl von l.u. Vektoren in V nutzt, schöner :).
- Die wichtigste Anwendung des obigen Satzes ist die folgende: Zur Bestimmung von $\dim(V)$ kann man eine beliebige Basis bestimmen (für gewöhnliche, indem man lineare Unabhängigkeit und Erzeugendeneigenschaft beweist) und dann ihre Elemente zählen. Z.B. gilt:
  - $\dim(\mathbb{K}^n) = n$, da die Vektoren der Standardbasis $e_1, \ldots, e_n$ erzeugend und linear unabhängig sind.
  - Ist $P_n[x] = \{f \in \mathit{Abb}(\mathbb{R}, \mathbb{R}) \mid f \text{ ist Polynomfunktion vom Grad } \leq n\}$, so ist $\dim(P_n[x]) = n + 1$ (!), da die Monome $(x^n, x^{n-1}, \ldots, x, 1)$ eine Basis von sind. Die Folge

    $$P_0[x] \subseteq P_1[x] \subseteq P_2[x] \subseteq P_3[x] \subseteq \ldots \subseteq \mathrm{Abb}(\mathbb{R}, \mathbb{R})$$

    von Unterräumen liefert also eine Folge von geschachtelten Unterräumen aufsteigender Dimension in $\mathrm{Abb}(\mathbb{R}, \mathbb{R})$, solche Folgen werden benutzt, um Funktionen $f \in \mathrm{Abb}(\mathbb{R}, \mathbb{R})$ anzunähern.
  - $\dim(\mathbb{R}^{m \times n}) = n \cdot m$, da die $(n \times m)$-Matrizen $M^{(k,\ell)}$, $k \in \underline{m}$, $\ell \in \underline{n}$, mit Einträgen

    $$m_{i,j}^{(k,\ell)} = \begin{cases} 1 & \text{für } (i,j) = (k,\ell), \\ 0 & \text{sonst} \end{cases}$$

    eine Basis sind.
  - $\mathbb{C}$ hat als $\mathbb{R}$-Vektorraum Dimension 2 (mit $B = (1, i)$ als Basis), als $\mathbb{C}$-Vektorraum Dimension 1 (mit $B = (1)$ als Basis).

Beweis von Satz 8.3.8: Sei V ein endlichdimensionaler Vektorraum mit $\dim V = n$ und $B = (b_1, \dots, b_n)$ eine Basis von V mit n Elementen (die nach 8.3.7 existiert). Sei nun $C = (v_1, \dots, v_k)$ eine weitere Basis von V; unser Ziel ist zu zeigen, dass dann $k = n$ sein muss. Wir bemerken dazu zunächst, dass, weil B eine Basis ist, jeder Vektor $v_j \in C$ eine Darstellung $v_j = \sum_{i=1}^n a_{i,j} b_i$ besitzt, ebenso besitzt ein beliebiges $v \in V$ eine eindeutige Darstellung $v = \sum_{i=1}^n c_i b_i$. Soll nun jeder solche Vektor $v \in V$ eindeutig als Linearkombination von $v_1, \dots, v_k$ schreibbar sein, so geht das genau dann, wenn für beliebige Koordinatentupel $(c_i)_{i \in \underline{n}}$ eindeutige Vorfaktoren $x_1, \dots, x_k \in \mathbb{K}$ für die Vektoren aus C existieren mit $\sum_{j=1}^k x_j \cdot v_j = \sum_{i=1}^n c_i \cdot b_i$, d.h. mit

$$\sum_{i=1}^{k} \Big(\sum_{j=1}^{n} a_{i,j} x_j\Big) \cdot b_i = \sum_{i=1}^{n} c_i \cdot b_i. \qquad (\star)$$

Da $B = (b_1, \dots, b_n)$ eine Basis ist, ist die Darstellung von Vektoren in B eindeutig. D.h., für $(x_1, \dots, x_k)$ gilt $(\star)$ genau dann, wenn das $(n \times k)$-System

$$\sum_{j=1}^{k} a_{i,j} x_j = c_i, \qquad i = 1, \dots, n$$

erfüllt ist. Es gibt also zu jedem Vektor $v \in V$ einen eindeutigen Koeffizientenvektor $x_1, \dots, x_k$ bezüglich C genau dann, wenn das obige LGS für $(x_1, \dots, x_k) \in \mathbb{K}^k$ eindeutig lösbar ist. Das geht nach Satz 6.3.3(ii),(iv),(v) nur, falls $k = n$ ist (und die Matrix $A = (a_{i,j})$ zusätzlich vollen Rang hat), d.h. nur dann kann C eine Basis sein.

Wir beschließen diese Kapitel mit einer oft recht nützlichen Erkenntnis:

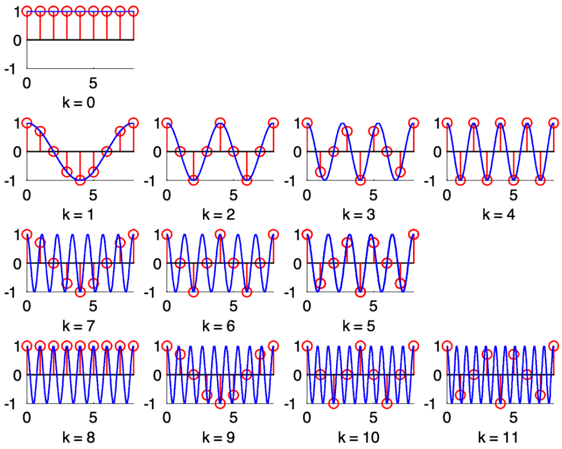
Korollar 8.3.9. *Seien $v_1, \dots, v_k$ Vektoren eines endlichdimensionalen Vektorraums V. Dann gilt:*

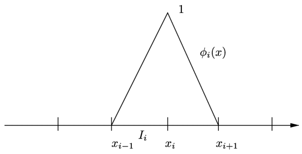
(i) Sind $v_1, \dots, v_k$ linear unabhängig, so sind sie eine Basis von V genau dann, wenn $k = dim(V)$ ist.

(ii) Sind $v_1, \dots, v_k$ ein Erzeugendensystem von V, so sind sie eine Basis von V genau dann, wenn $k = dim(V)$ ist.

Der Beweis stand zu Teilen schon oben, der Rest ergibt sich durch Zusammenfügen der Resultate dieses Kapitels und ist Übungsaufgabe.

Beispiele von Basisfunktionen in praktischen Anwendungen





oben: Linearkombinationen von solchen "Hutfunktionen" ergeben stückweise lineare Funktionen mit "Knicken" an den Stützstellen x_i. Hiermit lassen sich z.B. gegebene Datensätze auf einfache Weise verbinden ("linear interpolieren")

links: Der Aufspann von Kosinusfunktionen ist gut geeignet, um damit akustische Anwendungen zu meistern (z.B. MP3). Wenn man das im Rechner tun will, kann man die Wellen z.B. punktweise an den rot markierten Stellen "samplen" und erhält entsprechende Basisvektoren des $\mathbb{R}^n$.

Quelle: ti.tuwien.ac.at/cps/teaching/courses/dspv/files/DFT-FFT.pdf

8.4 Aufgaben zu Kapitel 8

Aufgabe 8.40. (Umformen von Gleichungen in allgemeinen Vektorräumen):

Es sei $(V, +, \cdot)$ ein $\mathbb{K}$-Vektorraum und $u, v, w \in V$, wobei

$$u = a \cdot v + b \cdot w \qquad \text{mit} \qquad a, b \in \mathbb{K}, \quad a \neq 0$$

gelte. Stellen Sie diese Gleichung so, wie sie es wahrscheinlich aus der Schule für Vektoren aus dem $\mathbb{R}^3$ gewohnt sind, nach v um. Genauer: Zeigen Sie, dass gilt:

$$u = a \cdot v + b \cdot w \qquad \Longleftrightarrow \qquad v = \frac{1}{a} \cdot u - \frac{b}{a} \cdot w.$$

Welche der Vektorraumaxiome haben Sie bei Ihren Umformungen benutzt? (Spoiler: Eigentlich fast alle, plus Eigenschaften aus Lemma 8.1.2:)

Aufgabe 8.41. (Erzeugnisse):

(a) Seien v_1, v_2, v_3 Vektoren eines Vektorraums V mit $\mathcal{L}(v_1, v_2, v_3) = V$. Beweisen Sie, dass V dann auch durch die Vektoren

$$w_1 := v_1 - v_2, \quad w_2 := v_2 - v_3, \quad w_3 := v_3$$

erzeugt wird. **Tipp:** Stellen Sie v_1, v_2, v_3 durch die Vektoren w_1, w_2, w_3 dar.

(b) Gegeben seien folgende Mengen von reellen Polynomen:

$$M_1 := \{1, x, x^2, x^3\} \subseteq \mathbb{R}[x]$$
$$M_2 := \{1 - x, 1 + x, x^2 + x^3, x^3\} \subseteq \mathbb{R}[x]$$

- Zeigen Sie, dass $M_1 \subseteq \mathcal{L}(M_2)$ und $M_2 \subseteq \mathcal{L}(M_1)$ ist.
- Zeigen Sie, dass $\mathcal{L}(M_1) = \mathcal{L}(M_2)$ ist.

(c) Seien v, w Vektoren eines $\mathbb{K}$-Vektorraumes V mit $v, w \neq 0_V$ und $z \in V$ ein weiterer Vektor, der sich auf zwei unterschiedliche Weisen $z = a_1 \cdot v + a_2 \cdot w = b_1 \cdot v + b_2 \cdot w$ als Linearkombination von v und w ausdrücken lässt (d.h. es ist $(a_1, a_2) \neq (b_1, b_2)$).

Beweisen Sie, dass v dann ein Vielfaches von w ist (d.h. es existiert ein $a \in \mathbb{K}$ mit $v = a \cdot w$).

Aufgabe 8.42. (Unterräume):

Seien U_1, U_2 Unterräume eines Vektorraumes V. Beweisen Sie:

(a) Die Menge $U_1 \cap U_2$ ist ein Unterraum von V.

(b) Die Menge $U_1 \cup U_2$ ist im Allgemeinen kein Unterraum von V.

Aufgabe 8.43. (Vektoren mit Parametern):

Bestimmen Sie eine Zahl $t \in \mathbb{R}$ so, dass die Vektoren

$$\begin{pmatrix} 3 \\ 1 \\ 4 \end{pmatrix}, \quad \begin{pmatrix} 2 \\ -3 \\ 5 \end{pmatrix}, \quad \begin{pmatrix} 5 \\ 9 \\ t \end{pmatrix}$$

linear abhängig in $\mathbb{R}^3$ sind.

Bilden die drei Vektoren mit dem von Ihnen bestimmten t eine Basis des reellen Vektorraums $\mathbb{R}^3$?

Aufgabe 8.44. (Rangkriterien zur Untersuchung von Vektoren):

Sei $\mathbb{K} = \mathbb{R}, \mathbb{C}$ oder $\mathbb{Q}$ und $v_1, \ldots, v_k \in \mathbb{K}^n$. $A \in \mathbb{K}^{n \times k}$ bezeichne die Matrix, deren Spalten gerade die Vektoren $v_1, \ldots, v_k$ sind. In der Vorlesung wurde gezeigt: $v_1, \ldots, v_k$ sind linear unabhängig genau dann, wenn $\mathrm{rg}(A) = k$ ist.

(a) Formulieren Sie ein analoges Rangkriterium an die Matrix A dafür, dass die Vektoren $v_1, \ldots, v_k$ ein Erzeugendensystem des $\mathbb{K}^n$ bilden. Begründen Sie Ihr Kriterium.

(b) Untersuchen Sie, ob die Vektoren

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}, \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix}, \begin{pmatrix} 10 \\ 11 \\ 12 \end{pmatrix} \in \mathbb{R}^3$$

ein Erzeugendensystem von $\mathbb{R}^3$ sind.

(c) Beweisen Sie durch ein geeignetes Argument für die Matrix A: Ist $v_1, \ldots, v_k \in \mathbb{K}^n$ ein Tupel von $k > n$ Vektoren, so können diese nicht linear unabhängig sein.

(d) Formulieren und begründen Sie ein Kriterium, mit dem man anhand der Matrix A erkennen kann, wann k Vektoren $v_1, \ldots, v_k \in \mathbb{K}^n$ eine Basis von $\mathbb{K}^n$ bilden.

(e) Sind drei Vektoren

$$\begin{pmatrix} 1 \\ i \\ 0 \end{pmatrix}, \begin{pmatrix} 1-i \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} -2i \\ -1+i \\ -2i \end{pmatrix}$$

eine Basis des $\mathbb{C}$-Vektorraums $\mathbb{C}^3$?

Aufgabe 8.45. (Lineare Abhängigkeit und Unabhängigkeit):

Beweisen oder widerlegen Sie die folgenden Aussagen.

(a) Sind $v_1, \ldots, v_k$ linear unabhängige Vektoren eines $\mathbb{K}$-Vektorraumes V und $a \in \mathbb{K}$ mit $a \neq 0$, so ist $a \cdot v_1, v_2, v_3, \ldots, v_k$ ebenfalls linear unabhängig.

(b) Sind $v_1, \ldots, v_k$ und $w_1, \ldots, w_k$ jeweils linear unabhängige Vektoren eines $\mathbb{K}$-Vektorraumes V, so sind $v_1 + w_1, \ldots, v_k + w_k$ ebenfalls linear unabhängig.

(c) Sind $v_1, \ldots, v_k$ linear unabhängige Vektoren eines $\mathbb{K}$-Vektorraumes V, so sind $v_1 + v_2, v_2, v_3, \ldots, v_k$ ebenfalls linear unabhängig.

(d) Ist V ein $\mathbb{K}$-Vektorraum und $M \subseteq V$ ein Erzeugendensystem von V, so ist auch jede Menge N mit $M \subset N$ ein Erzeugendensystem von V.

Aufgabe 8.46. (Aufspann und Summe von Vektorräumen):

Sei V ein $\mathbb{K}$-Vektorraum und seien $M_1, M_2 \subseteq V$ beliebige Mengen. Zeigen Sie, dass folgende Gleichheit gilt:

$$\mathcal{L}(M_1 \cup M_2) = \mathcal{L}(M_1) + \mathcal{L}(M_2).$$

Aufgabe 8.47. (Zusammenhänge zwischen Basen):

Sei V ein $\mathbb{K}$-Vektorraum (mit $\mathbb{K} = \mathbb{R}, \mathbb{C}$ oder $\mathbb{Q}$) und $v_1, v_2, v_3 \in V$ so, dass $B = (v_1, v_2, v_3)$ eine Basis von V ist. Beweisen Sie, dass dann auch $C := (w_1, w_2, w_3)$ mit

$$\begin{aligned} w_1 &:= v_1 + 2v_2 + 3v_3 \\ w_2 &:= v_1 + v_2 \\ w_3 &:= v_1 \end{aligned}$$

eine Basis von V ist.

Aufgabe 8.48. (Argumentieren mit Vektorraumbegriffen):

(a) Sei V ein endlichdimensionaler Vektorraum. Beweisen Sie mit Hilfe geeigneter Zusammenhänge aus diesem Kapitel:

(i) Ist $U \subseteq V$ ein Teilraum von V, so gilt $\dim U \leq \dim V$.

(ii) Ist $U \subseteq V$ ein Teilraum von V und $\dim U = \dim V$, so gilt $U = V$.

(iii) Sei $\dim V = n \in \mathbb{N}$ und E ein Erzeugendensystem von V.
Zeigen Sie: Wenn E aus n Vektoren besteht, so ist E eine Basis von V.

(b) Seien V, W $\mathbb{K}$-Vektorräume. Zeigen Sie mit Hilfe des Teilraumkriteriums (angewandt auf einen geeigneten Vektorraum), dass

$$L(V, W) := \{f : V \to W \mid f \text{ ist linear } \}$$

zusammen mit der gewöhnlichen, punktweise in W definierten Addition

$$+ : \; L(V,W) \times L(V,W) \to L(V,W), \qquad (f,g) \mapsto f+g \quad \text{mit} \quad (f+g)(v) := f(v)+g(v)$$

und der skalaren Multiplikation

$$\cdot : \; \mathbb{K} \times L(V,W) \to L(V,W) : \qquad (a,f) \mapsto a \cdot f \quad \text{mit} \quad (a \cdot f)(v) := a \cdot f(v)$$

ein Vektorraum über $\mathbb{K}$ ist.

Aufgabe 8.49. (Rechnen mit Basen):

(a) Zeigen Sie, dass drei Vektoren der Form

$$\begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix}, \begin{pmatrix} 1 \\ y \\ y^2 \end{pmatrix}, \begin{pmatrix} 1 \\ z \\ z^2 \end{pmatrix}$$

mit $x, y, z \in \mathbb{R}$ genau dann eine Basis von $\mathbb{R}^3$ bilden, wenn $x \neq y, x \neq z$ und $y \neq z$ gilt.

(b) Im $\mathbb{R}^4$ seien die Unterräume

$$U_1 = \Big\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \in \mathbb{R}^4 \mid -x_1 - x_2 + x_3 = 0 \Big\} \quad \text{und} \quad U_2 = \mathcal{L}\Big(\begin{pmatrix} -1 \\ 1 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ -1 \\ 2 \end{pmatrix} \Big)$$

gegeben. Bestimmen Sie jeweils Basen für die Unterräume $U_1, U_2, U_1 \cap U_2$ und $U_1 + U_2$. Welche Dimensionen haben diese Unterräume? Begründen Sie Ihre Ergebnisse.

Hinweis: Wie prüfen Sie, ob für einen Vektor $v \in U_2$ auch $v \in U_1$ gilt?

9 Lineare Abbildungen

Im folgenden bezeichnen wir mit V und W stets zwei $\mathbb{K}$-Vektorräume $(V, +, \cdot)$, $(W, \oplus, \odot)$ über dem gleichen Körper $\mathbb{K} = \mathbb{R}, \mathbb{C}$ oder $\mathbb{Q}$. Ihre Nullvektoren seien mit 0_V und 0_W bezeichnet.

9.1 Grundlegende Definitionen und Beispiele

Definition 9.1.1. *(Lineare Abbildung zwischen Vektorräumen)*

Seien V und W beliebige Vektorräume über dem gleichen Körper $\mathbb{K}$. Eine Abbildung $f : V \to W$ heißt $\underline{\mathbb{K}\text{-Homomorphismus}}$, $\underline{\mathbb{K}\text{-lineare Abbildung}}$, $\underline{\text{linear von } V \text{ nach } W}$ oder einfach linear (wenn $\mathbb{K}$ klar ist), falls gilt:

(i) Für alle $u, v \in V$ ist $f(u + v) = f(u) \oplus f(v)$ und

(ii) Für alle $a \in \mathbb{K}$, $u \in V$ ist $f(a \cdot u) = a \odot f(u)$.

Die Menge aller linearen Abbildungen von V nach W kürzen wir mit $L(V, W)$ ab.

Bemerkungen dazu:

- Es ist gängig, nicht ständig $\oplus$ und $\odot$ für die Addition und skalare Multiplikation in W zu schreiben, sondern in W ebenfalls die Symbole $+$ und $\cdot$ zu benutzen; auch wir werden ab hier nur dann $\oplus$ und $\odot$ schreiben, wenn es für das Verständnis notwendig ist.
- Statt der Bedingungen (i) und (ii) reicht es aus, zu zeigen, dass die Bedingung $f(u + a \cdot v) = f(u) \oplus a \odot f(v)$ für alle $u, v \in V, a \in \mathbb{K}$ erfüllt ist, damit f linear ist (Übung).
- Vorsicht! Funktionen der Bauart $f : \mathbb{R} \to \mathbb{R}, x \mapsto m \cdot x + b$, die in der Schule lineare Funktionen genannt werden, sind $\underline{\text{nur für } b = 0}$ linear im Sinne der obigen Definition (das sind die, die in der Schule als proportionale Zuordnungen genannt werden, also Funktionen $f : \mathbb{R} \to \mathbb{R}, x \mapsto m \cdot x$). Im Fall $b \neq 0$ sind sie keine linearen Funktionen nach Def. 9.1.1 (sondern in „Uni-Sprech“ so genannte affine Funktionen, kommt später).

Definition 9.1.2. *(Kern und Bild linearer Funktionen)*

Der Kern einer linearen Abbildung $f : V \to W$ ist das Urbild des Nullvektors von W, also die Menge

$$\mathrm{Kern}(f) = \{v \in V \mid f(v) = 0_W\}$$

aller Vektoren $v \in V$, die durch f auf den Nullvektor von W abgebildet werden.

Wie zuvor bezeichnen wir als das Bild von f die Menge

$$\mathrm{Bild}(f) = \{f(v) \mid v \in V\}$$

und nutzen auch die Schreibweise $f(M) = \{f(v) \mid v \in M\}$ für das Bild einer Teilmenge $M \subset V$ unter f.

9.1.1 Wichtige Beispiele linearer Abbildungen

Beispiel 0. Nullabbildung und identische Abbildung

Für alle $\mathbb{K}$-Vektorräume V, W sind

- die Nullabbildung $\mathcal{O} : V \to W, v \mapsto 0_W$ und
- die identische Abbildung id_V auf V, definiert durch $\mathrm{id}_V : V \to V, v \mapsto v$

$\mathbb{K}$-linear.
Es ist $\mathrm{Kern}(\mathcal{O}) = V$ und $\mathrm{Bild}(\mathcal{O}) = \{0_W\}$ sowie $\mathrm{Kern}(\mathrm{id}_V) = \{0_V\}$ und $\mathrm{Bild}(\mathrm{id}_V) = V$.

Beispiel 1. Matrixabbildungen

Sei $\mathbb{K} = \mathbb{Q}, \mathbb{R}$ oder $\mathbb{C}$. Die linearen Abbildungen $f : \mathbb{K}^n \to \mathbb{K}^m$ sind genau die Matrixabbildungen, d.h. die Abbildungen, die sich in der Form $f : x \mapsto A \cdot x$ mit einer Matrix $A \in \mathbb{K}^{m \times n}$ schreiben lassen. Um das einzusehen, ist es gut, die Resultate von Kapitel II zu rekapitulieren (Tun Sie das!):

- Dass jede Matrixabbildung linear ist, also Definition 9.1.1 erfüllt, haben wir für den Fall $\mathbb{K} = \mathbb{R}$ schon in 4.4.1 gesehen. In Fall eines allgemeinen Körpers $\mathbb{K}$ lässt sich ganz genauso argumentieren.

- Umgekehrt gilt für eine lineare Abbildung $f \in L(\mathbb{R}^n, \mathbb{R}^m)$ und alle $x = (x_i)_{i \in \underline{n}} \in \mathbb{K}^n$, dass

$$f(x) \quad = \quad f(\sum_{i=1}^{n} x_i \cdot e_i) = \sum_{i=1}^{n} f(x_i \cdot e_i) = \sum_{i=1}^{n} x_i \cdot f(e_i), \tag{9.1}$$

  also ist jede lineare Abbildung $f : \mathbb{K}^n \to \mathbb{K}^m$ wie in 4.3.2 durch eine Matrix beschreibbar, die Koordinaten bezüglich der Standardbasis $e_1, \ldots, e_n$ auf die Koordinaten des Bildes abbildet, d.h f ist eine Matrixabbildung.

- In Gleichung (9.1) sieht man auch, dass sich jedes Bild $f(x)$ eines Vektors $x \in \mathbb{K}^n$ als Linearkombination der Vektoren $f(e_1), \ldots, f(e_n)$ schreiben lässt. Das **Bild** einer Matrixabbildung $f \in L(\mathbb{R}^n, \mathbb{R}^m)$ ist also gerade der Aufspann der Spalten der Abbildungsmatrix

$$A = \begin{pmatrix} | & | & \cdots & | \\ f(e_1) & f(e_2) & \cdots & f(e_n) \\ | & | & \cdots & | \end{pmatrix},$$

  in denen eben $f(e_1), \ldots, f(e_n)$ zu finden sind.

- Vorsicht: Wir werden bald sehen, dass lineare Abbildungen auch bezüglich anderer Basen durch Matrizen dargestellt werden können. Die Spalten solch einer darstellenden Matrix $A_{B,C}$ spannen auch für Matrixabbildungen $f \in L(\mathbb{R}^n, \mathbb{R}^m)$ dann nicht mehr zwingend das Bild von f auf, s. Abschnitt 10.2.

- Der **Kern** einer Matrixabbildung ist die Menge aller $x \in \mathbb{K}^n$ mit $A \cdot x = 0 \in \mathbb{K}^m$, d.h. genau die Lösungsmenge des so genannten homogenen linearen Gleichungssystems zu A. Man sieht, dass $0 \in \mathbb{R}^n$ immer eine Lösung dieses Systems ist, also immer im Kern von f enthalten ist, der Kern aber auch mehr Elemente enthalten kann (wenn $\mathrm{rg}(A) < n$ gilt, vgl. 6.3.3(iv)).

Beispiel 2. Koordinatenabbildungen

Ist V ein endlichdimensionaler $\mathbb{K}$-Vektorraum und $(b_1, \ldots, b_n)$ eine Basis von V, so ist die Koordinatenabbildung

$$k_B : \quad V \to \mathbb{K}^n, \qquad \left(v = \sum_{i=1}^{n} a_i b_i\right) \mapsto \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

wohldefiniert (das ist gerade die Eigenschaft der Basis B, jedem Vektor wird genau ein Koordinatenvektor zugeordnet) und linear.

- Die Linearität von k_B rechnen wir in der Vorlesung nach.
- k_B ist zudem bijektiv, denn

$$k_B^{-1}: \quad \mathbb{K}^n \to V, \qquad \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \mapsto \left(v = \sum_{i=1}^{n} a_i b_i \right)$$

  ist Links- und Rechtsinverse zu k_B.
- Der Kern von k_B ist die Menge aller Vektoren aus V, die die Koordinaten $(0, \dots, 0)$ haben. Da k_B bijektiv ist, gehört zu jedem Vektor aus V genau ein Koordinatenvektor, also ist $\text{Kern}(k_B) = \{0_V\}$.
- Da f eine Inverse besitzt, ist f auch surjektiv, d.h. wir haben $\text{Bild}\, f = \mathbb{K}^n$.

Beispiel 3: Lineare Abbildungen auf Funktionenräumen

In der Analysis zeigt man, dass für das Ableiten differenzierbarer Funktionen f, g die Regeln

$$(f+g)' = f' + g', \quad (a \cdot f)' = a \cdot f'$$

gelten. Interpretiert man das Ableiten als eine Funktion (oder auf „hochmathematisch“: Operator) auf dem Vektorraum der differenzierbaren Funktionen, die einer Funktion f wiederum eine Funktion, nämlich ihre Ableitungsfunktion f', zuordnet, so ist dieser Operator $(\cdots)'$, der oft auch mit $\frac{d}{dx}$ abgekürzt wird, linear. Die möglicherweise erstaunliche Erkenntnis ist, dass Sie diese Linearität des Ableitens schon seit Jahren benutzt haben – denn Sie wissen höchstwahrscheinlich, was die Ableitung von $f(x) = 3 \cdot sinx + 2 \cdot x^2 - x^5$ ist, und der Grund dafür ist, dass sie die Linearität benutzen und die Ableitungen der Funktionen $\sin x$, x^2 und x^5 kennen, aus denben Sie dann $(3 \cdot sinx + 2 \cdot x^2 - x^5)'$ per Linearität berechnen können. Auch wenn es in der Analysis (noch) komplizierter wird, mehr Raumrichtungen betrachtet werden usw., bleiben solche Differentialoperatoren wie $\frac{d}{dx}$ für gewöhnlich lineare Abbildungen und passen damit in unsere Theorie. Da hier wie oft in unendlichdimensionalen Räumen aber auch viele Begriffe aus der Analysis eine Rolle spielen, beschränken wir uns als instruktives Beispiel auf einen bekannten endlichdimensionalen Teilraum von $\text{Abb}(\mathbb{R}, \mathbb{R})$: Wir betrachten die Menge aller reellen Polynomfunktionen zu Polynomen vom Grad höchstens n. Diese hatten wir mit $P_n[x]$ bezeichnet (für ein festes $n \in \mathbb{N}$).

- Die Abbildung

$$\frac{d}{dx} : P_n[x] \to P_n[x], \quad \sum_{i=0}^{n} c_i x^i \mapsto \sum_{i=1}^{n} i \cdot c_i \cdot x^{i-1},$$

  die sich als Ableitung der entsprechenden Polynomfunktion deuten lässt, ist linear. (Man beachte die verschiedenen Indizes!)

- Was ist hier $\text{Kern}(\frac{d}{dx})$? Was $\text{Bild}(\frac{d}{dx})$?

Auch das Integrieren lässt sich als linearer „Operator“ auffassen, den man z.B. auf den Raum der über einem Intervall $[a, b]$ definierten stetigen Funktionen $C([a, b], \mathbb{R})$ definieren kann (denn solche stetigen Funktionen sind immer integrierbar).

9.2 Eigenschaften linearer Abbildungen

Lineare Abbildungen besitzen eine Menge nützliche Eigenschaften, die immer wieder benutzt werden und die wir in diesem Abschnitt zusammenstellen. Erst geht es allgemein um lineare Abbildungen, später schauen wir uns an, wie sich Injektivität, Surjektivität und Bijektivität linearer Abbildungen charakterisieren lassen.

Lemma 9.2.1. *(Allgemeine Eigenschaften linearer Abbildungen)*

Sei $f : V \to W$ linear. Dann gilt u.a.:

(i) $f(0_V) = 0_W$.

(ii) Sind $v_i \in V, i \in \underline{n}$ und $v = \sum_{i=1}^{n} a_i v_i$ eine Linearkombination dieser Vektoren (mit $a_i \in \mathbb{K}, i \in \underline{n}$), so gilt

$$f(v) \quad = \quad f\left(\sum_{i=1}^{n} a_i v_i\right) = \sum_{i=1}^{n} a_i f(v_i).$$

(iii) Eine lineare Abbildung f ist durch ihre Wirkung auf eine Basis B von V, also durch Angabe von $f(b_i)$ für alle Basisvektoren $b_i \in B$, vollständig festgelegt.

(iv) Ist $B = (b_1, \ldots, b_n)$ eine Basis von V, so ist $\text{Bild}(f) = \mathcal{L}\big(f(b_1), \ldots, f(b_n)\big)$.

(v) *Ist U ein Untervektorraum von V, so ist auch $f(U)$ ein Untervektorraum von W.*

(vi) *Ist $M \subseteq V$ das Urbild eines Untervektorraum $U \subseteq W$, d.h.*

$$M = \{v \in V \mid f(v) \in U\},$$

so ist M ein Untervektorraum von V.

(vii) $\operatorname{Kern}(f)$ *ist ein Untervektorraum von V,* $\operatorname{Bild}(f)$ *ist ein Untervektorraum von W.*

(viii) *Sind $v_1, \ldots, v_k$ linear abhängige Vektoren von V, so sind $f(v_1), \ldots, f(v_k)$ ebenfalls linear abhängig in W.*

(ix) *Für alle Teilräume $U \subseteq V$ gilt* $\dim f(U) \leq \dim(U)$.

Beweis: Für den ausführlichen Beweis vgl. Vorlesung, hier werden nur die Ideen skizziert: Zu (i) stellen wir mit Hilfe der Linearität $f(0_V) = f(0_V + 0_V) = f(0_V) + f(0_V)$ fest. Subtrahiert man in der Gruppe $(W, +)$ auf beiden Seiten den Vektor $f(0_V) \in W$, so folgt $f(0_V) = 0_W$. (ii) folgt einfach mit Hilfe mehrfacher Anwendung der beiden Linearitätseigenschaften, genauer durch $(n-1)$-faches Auseinanderziehen der Summe und n-facher Anwendung der Skalariät. (iii) und (iv) folgen aus (ii), (v) ergibt sich aus den Unterraumkriterium plus Linearität.

Zu (vi): Zeigen wir das Unterraumkriterium. Da U ein Unterraum von W ist, ist $0_W \in U$ und daher nach (i) $0_V \in M$. Sind $u, v \in M$, und $a \in \mathbb{K}$, so sind $f(u)$ und $f(v)$ in U und, da U ein Unterraum ist, auch $f(u) + a \cdot f(v) \in U$. Wegen der Linearität haben wir

$$f(u) + a \cdot f(v) = f(u + a \cdot v),$$

also ist auch $f(u + a \cdot v) \in U$ und damit $u + a \cdot v \in M$. Damit erfüllt M das Unterraumkriterium und ist somit ein Teilraum von V.

(vii) folgt direkt aus (vi) angewandt auf den Teilraum $\{0_W\}$ von W bzw. aus (v) angewandt auf den Teilraum V von V.

Zu Aussage (viii) rechnet man nach: Lässt sich einer der Vektoren in einer Menge als Linearkombination der anderen schreiben, so bleibt das wegen der Linearität von f auch im Bild so.

Aussage (ix) folgt aus der Definition der Dimension und (viii) bzw. dessen Kontraposition: Sind $v_1, \ldots, v_k \in U$ und $f(v_1), \ldots, f(v_k)$ linear unabhängige Vektoren in $f(U)$, so sind nach (viii) $v_1, \ldots, v_k$ auch l.u. Elemente von U, also $\dim(U) \geq \dim(f(U))$.

□

Lemma 9.2.2. *(Verknüpfungen linearer Funktionen)*

Sei $f \in L(V,W)$, also $f : V \to W$ linear. Dann gilt:

(i) Ist f bijektiv, so ist $f^{-1} : W \to V$ linear.

(ii) Ist $g \in L(V,W)$ eine weitere lineare Abbildung, so ist auch $f+g$ linear.

(iii) Ist $a \in \mathbb{K}$, so ist auch $(a \cdot f) : v \mapsto a \cdot f(v)$ linear.

(iv) Ist Z ein weiterer $\mathbb{K}$-VR und $g : W \to Z$ linear, so ist auch $g \circ f : V \to Z$ linear.

Beweis: Wir zeigen nur (i), der Rest folgt ähnlich und leichter aus der Linearität der beteiligten Abbildungen und ist z.T. Übungsaufgabe. Sei f bijektiv. Dann existiert die Umkehrabbildung $f^{-1} : W \to V$. Seien $w_1, w_2 \in W$ und $a \in \mathbb{K}$; wir zeigen die verkürzte Linearitätsbedingung

$$f^{-1}(w_1 + a \cdot w_2) = f^{-1}(w_1) + a \cdot f^{-1}(w_2),$$

vgl. die Bemerkung nach Def. 9.1.1. Da f bijektiv und somit surjektiv ist, gibt $v_1, v_2 \in V$ mit $f(v_1) = w_1$ und $f(v_2) = w_2$ bzw. $v_1 = f^{-1}(w_1)$ und $v_2 = f^{-1}(w_2)$. Man nutzt die Linearität von f; dann gilt:

$$\begin{aligned} f^{-1}(w_1 + a \cdot w_2) &= f^{-1}(f(v_1) + a \cdot f(v_2)) = (f^{-1} \circ f)(v_1 + a \cdot v_2) \\ &= v_1 + a \cdot v_2 = f^{-1}(w_1) + a \cdot f^{-1}(w_2). \end{aligned}$$

□

Bemerkung (L(V,W) als Vektorraum und als Ring):

Aus dem obigen folgt insbesondere, dass die für $f, g \in L(V,W)$ und für $a \in \mathbb{K}$ definierten Abbildungen $(f,g) \mapsto f+g$ und $(a,f) \mapsto a \cdot f$ wieder lineare Abbildungen nach $L(V,W)$ sind. Da die Nullabbildung $V \to W, v \mapsto 0_V$ immer linear ist, folgt daraus mit dem Teilraumkriterium, dass auch die Struktur $(L(V,W), +, \cdot)$ ein Vektorraum (ein Teilraum des Raumes $\mathrm{Abb}(V,W)$) ist.

Analog zum Resultat aus Kapitel 5, dass die Menge der $(n \times n)$-Matrizen (für festes $n \in \mathbb{N}$) mit der Addition von Matrizen und der Matrixmultiplikation einen Ring bilden, kann man auch relativ leicht zeigen, dass im Fall $V = W$ das Tripel $(L(V), +, \circ)$ ein Ring ist.

Wir untersuchen nun injektive, surjektive und bijektive lineare Abbildungen. Wir beginnen mit einer einfachen Feststellung.

Lemma 9.2.3. *(Surjektivität linearer Abbildungen)*

Sei $f : V \to W$ linear. Dann kann f nur surjektiv sein, wenn $\dim W \leq \dim V$ gilt.

Beweis: Nach Definition ist f surjektiv genau dann, wenn $\mathrm{Bild}(f) = W$ ist, was genau dann stimmt (vgl. alte ÜA), wenn $\dim \mathrm{Bild}(f) = \dim W$ ist. Nach Lemma 9.2.1(ix) ist $\dim \mathrm{Bild}(f) = \dim f(V) \leq \dim V$; das ist nur miteinander kompatibel für $\dim W \leq \dim V$. □

Lemma 9.2.4. *(Kriterium für Injektivität linearer Abbildungen)*

Sei $f : V \to W$ linear. Es sind äquivalent:

(i) *f ist injektiv,*

(ii) $\mathrm{Kern}(f) = \{0_V\}$.

Weiter gilt, falls f injektiv ist[1]*:*

(iii) *Sind $v_1, \ldots, v_k \in V$ beliebige linear unabhängige Vektoren, so sind auch $f(v_1), \ldots, f(v_k) \in W$ linear unabhängig.*

(iv) *Für jede Basis B von V ist $f(b_1), \ldots, f(b_n)$ eine Basis von* $\mathrm{Bild}(f)$.

(v) *f ist dimensionserhaltend, d.h. für alle Teilräume $U \subseteq V$ gilt* $\dim f(U) = \dim U$.

Insbesondere kann f nur injektiv sein, wenn $\dim W \geq \dim V$ *gilt.*

Beweis: Wir zeigen nur die (sehr wichtige, weil oft benutzte) Äquivalenz (i) $\Leftrightarrow$ (ii) und die Schlussbemerkung. Der zweite Teil ist Übungsaufgabe.

Zu (i): Für eine lineare Abbildung gilt nach 9.2.1(i) immer $f(0_V) = 0_W$, also immer $0_V \in \mathrm{Kern}\, f$. Ist f injektiv, ist nach Definition der Injektivität $f(v) \neq f(0_V)$ für alle $v \neq 0$, also gilt (ii). Ist f nicht injektiv, so gibt es $v, w \in V$ mit $v \neq w$, aber $f(v) = f(w)$, und wegen der Linearität ist $f(v - w) = f(v) - f(w) = 0_W$, aber $v - w \neq 0_V$, also $\mathrm{Kern}(f) \neq \{0_V\}$.

Die Schlussbemerkung folgt, da für eine injektive Abbildung nach (iv) für den Unterraum $\mathrm{Bild}(f) \subseteq W$ insbesondere $\dim \mathrm{Bild}\, f = \dim f(V) = \dim V$ gelten muss, für alle Unterräume U von W aber $\dim U \leq \dim W$ gilt (vgl. alte Übungsaufgabe). □

---

[1] Es ist sogar so, dass die Bedingungen (ii) und (iv) zur Injektvität äquivalent sind; das wollen wir hier aber nicht zeigen.

9.3 Der Dimensionssatz für lineare Abbildungen und Anwendungen

Wie im vorangangenen Abschnitt kann man oft mit den Dimensionen von Vektorräumen bzw. bestimmten Teilräumen argumentieren. Das wichtigste Hilfsmittel dafür ist die folgende Dimensionsformel für lineare Abbildungen.

Satz 9.3.1. *(Dimensionssatz für lineare Abbildungen, „Kern-Bild-Satz“)*

Seien V, W endlichdimensionale Vektorräume und $f : V \to W$ linear.
Dann gilt für f die Dimensionsformel für lineare Abbildungen:

$$\dim V \quad = \quad \dim \operatorname{Kern}(f) \;+\; \dim \operatorname{Bild}(f). \tag{9.2}$$

Beweis: Sei $(b_1, \ldots, b_k)$ eine Basis des Kerns von f (d.h. insbesondere $\dim \operatorname{Kern}(f) = k$). Wir ergänzen diese Basis zu einer Basis $B = (b_1, \ldots, b_k, v_1, \ldots, v_\ell)$ von V, das dürfen wir nach Satz 8.3.5. Dann haben wir $\dim V = k + \ell = \dim \operatorname{Kern}(f) + \ell$, und wir haben also $\dim \operatorname{Bild}(f) = \ell$ zu zeigen. Das folgt, falls wir zeigen können, dass $B' = \{f(v_1), \ldots, f(v_\ell)\}$ eine Basis von $\operatorname{Bild}(f)$ ist. Also los:

(i): B' ist Erzeugendensystem des Bildes: Sei $w \in \operatorname{Bild}(f)$, dann gibt es ein $v \in V$ mit $f(v) = w$, und v lässt sich als Linearkombination der Vektoren aus B darstellen. Dann ist

$$w = f(v) = f(\underbrace{\sum_{i=1}^{k} c_i b_i + \sum_{i=1}^{\ell} c'_i v_i)}_{\text{Darstellung von } v} = \sum_{i=1}^{k} c_i \cdot \underbrace{f(b_i)}_{=0} + \sum_{i=1}^{\ell} c'_i \cdot f(v_i) = \sum_{i=1}^{\ell} c'_i \cdot f(v_i),$$

also ist $w \in \mathcal{L}(f(v_1), \ldots, f(v_\ell))$ wie behauptet.

(ii): B' ist linear unabhängig. Dazu zeigen wir, dass die Gleichung $\sum_{i=1}^{\ell} c_i f(v_i) = 0_W$ nur die Lösung $c_i = 0$ für alle $i = 1, \ldots, \ell$ haben kann. Wir haben mit dieser Gleichung

$$0_W = \sum_{i=1}^{\ell} c_i \cdot f(v_i) = f(\sum_{i=1}^{\ell} c_i \cdot v_i),$$

also ist $\sum_{i=1}^{\ell} c_i \cdot v_i$ demnach ein Element aus dem Kern von f. $(b_1, \ldots, b_k)$ ist eine Basis des Kerns, daher gibt Koeffizienten $a_1, \dots, a_k \in \mathbb{K}$ so, dass

$$\sum_{i=1}^{k} a_i \cdot b_i = \sum_{i=1}^{\ell} c_i \cdot v_i$$

gilt. Umstellen gibt

$$\sum_{i=1}^{k} a_i \cdot b_i + \sum_{i=1}^{\ell} (-c_i) \cdot v_i = 0,$$

woraus wegen der linearen Unabhängigkeit der „Gesamtbasis“ B dann $a_1 = \ldots = a_k = 0, c_1 = \ldots = c_\ell = 0$ folgt. Das zeigt die lineare Unabhängigkeit von B', somit $\ell = \dim \operatorname{Bild}(V)$ und insgesamt die Dimensionsformel. □

Eine wichtige, schnelle Folgerung aus dem Dimensionssatz ist die folgende:

Lemma 9.3.2. *Seien V, W endlichdimensionale Vektorräume mit $\dim V = \dim W$. Dann sind äquivalent:*

(i) f ist injektiv (d.h. eine der Bedingungen aus Lemma 9.2.4 ist erfüllt),

(ii) f ist surjektiv,

(iii) f ist bijektiv.

Beweis:
Wir zeigen zunächst $(i) \Leftrightarrow (ii)$ Nach Lemma 9.2.4 ist f ist injektiv genau dann, wenn $\operatorname{Kern}(f) = \{0_V\}$, also genau dann, wenn $\dim \operatorname{Kern}(f) = 0$ gilt. f ist surjektiv genau dann, wenn $\operatorname{Bild}(f) = W$ ist, also genau dann, wenn $\dim \operatorname{Bild}(f) = \dim W$ gilt, nach Voraussetzung ist das gleichwertig zu $\dim \operatorname{Bild}(f) = \dim V$. □

Mit dem Dimensionssatz

$$\dim V \quad = \quad \dim \operatorname{Kern}(f) \ + \ \dim \operatorname{Bild}(f).$$

haben wir daher

$$f \text{ ist injektiv } \Leftrightarrow \dim \operatorname{Kern}(f) = 0 \Leftrightarrow \dim \operatorname{Bild}(f) = \dim V \Leftrightarrow f \text{ ist surjektiv }.$$

Da jede lineare injektive Funktion automatisch surjektiv ist und umgekehrt, folgt für solche Funktionen automatisch auch ihre Bijektivität, d.h .(i) und (ii) implizieren auch (iii). Die umgekehrten Richtungen sind klar, bijektive Funktionen sind injektiv und surjektiv.

Definition/Bemerkung 9.3.1. *(Etwas Griechisch)*

Eine $\mathbb{K}$-lineare Abbildung $f : V \to W$ zwischen $\mathbb{K}$-Vektorräumen V, W bezeichnet man auch als $\mathbb{K}$-Homomorphismus zwischen V und W.

- *Injektive Homomorphismen heißen Monomorphismen,*
- *Surjektive Homomorphismen heißen Epimorphismen,*
- *Bijektive Homomorphismen (also „Epimonomorphismen“) heißen Isomorphismen.*
- *Ist $V = W$, so nennt man einn Homomorphismus auch Endomorphismus, Isomorphismen $f : V \to V$ heißen auch Automorphismus von V.*

Die algebraische Struktur $(\mathrm{Aut}(V), \circ)$ aller Automorphismen eines Vektorraumes V mit der Hintereinanderausführung $\circ$ bildet eine Gruppe (ÜA).

Wir haben also bereits gezeigt:

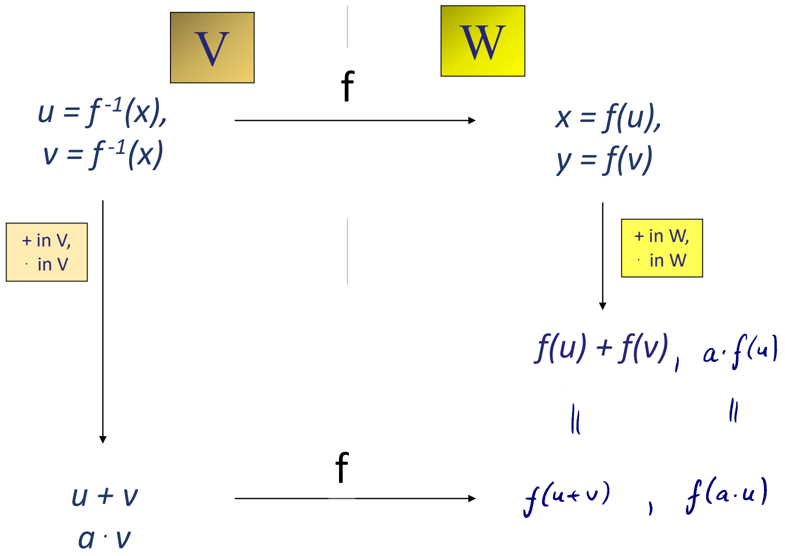
- Im Fall $\dim V = \dim W$ sind alle Monomorphismen und alle Epimporphismen automatisch Isomorphismen.
- Im Fall $\dim V < \dim W$ kann ein Homomorphimus kein Epimorphismus sein (Lemma 9.2.3), und damit kein Isomorphismus.
- Im Fall $\dim V > \dim W$ kann ein Homomorphimus kein Monomorphismus sein (Lemma 9.2.4), und damit kein Isomorphismus.

Von den obigen Begriffen werden wir im Folgenden nur die unterstrichenen benutzen, das reicht für gewöhnlich.

Kommutative Diagramme:
Wenn wir die Eigenschaften linearer Abbildungen und vor allem später ihrer Zusammenhänge verdeutlichen wollen, kann man dafür oft so genannte kommutative Diagramme verwenden. Hier werden die Vektorräume und die Abbildungen zwischen ihnen verdeutlicht, wie das in dem Diagramm unten für eine lineare Abbildung $f : V \to W$ zu sehen ist.
„Kommutativ“ ist leider eine etwas merkwürdige Sprechweise, die sich im Diagramm folgendermaßen beschrieben lässt: Kann man durch die Hintereinanderausführung von Abbildungen den Pfeilen folgend von einem Vektorraum zum anderen gelangen, so spielt die

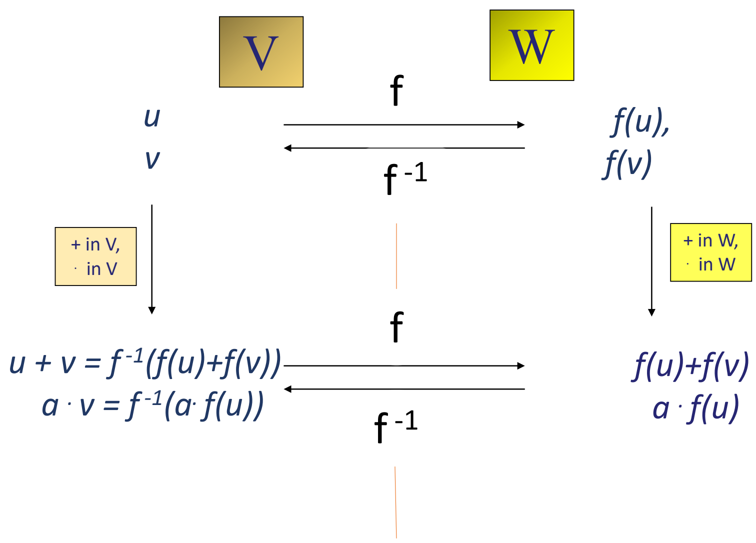
Art und Weise, wie man dorthin gelangt, keine Rolle, das Ergebnis ist immer dasselbe. Die Kommutativität des untenstehenden Diagramms bedeutet also genau, dass f linear ist – die Reihenfolge der Anwendung von f und der Vektorraumoperationen spielt für das Ergebnis keine Rolle.

f ist linear



Für einen Isomorphismus kann man das Diagramm weiter ausbauen. Die Doppelpfeile sollen andeuten, dass die die Abbildung bijektiv ist, also die Wirkung der jeweiligen Abbildung immer wieder rückgängig gemacht werden kann.

Isomorphie von V und W



9.4 Isomorphie

Definition 9.4.1. *(Isomorphe Vektorräume)*
Zwei Vektorräume V, W heißen strukturgleich oder isomorph, wenn es einen Isomorphismus $f : V \to W$ zwischen ihnen gibt, d.h eine lineare, bijektive Abbildung. Für isomorphe Vektorräume schreiben wir $V \rightleftharpoons W$ oder, wenn wir die Abbildung f betonen wollen, $V \overset{f}{\rightleftharpoons} W$.

Bemerkungen dazu:

- Sind $(V, +, \cdot)$ und $(W, \oplus, \odot)$ isomorph und $f : V \to W$ ein Isomorphismus zwischen ihnen, so hat f eine nach Kapitell II bijektive Umkehrabbildung f^{-1}. Wir können daher schreiben:

$$u + v = f^{-1}(f(u) \oplus f(v)), \qquad a \cdot u = f^{-1}(a \odot f(u))$$

  für alle $u, v \in V, a \in \mathbb{K}$ und

$$x + y = f(f^{-1}(x) \oplus f^{-1}(y)), \qquad a \cdot x = f(a \odot f^{-1}(x))$$

  für alle $x, y \in W, a \in \mathbb{K}$. Es ist also *egal*, ob man die Vektorraumoperationen $+, \cdot$ in V ausführt, oder ob man die Vektoren per f nach W abbildet, dort per $\oplus, \odot$ verknüpft und wieder zurück abbildet. Umgekehrt gilt das genauso. V und W entsprechen einander also eins zu eins, was (a) ihre Elemente angeht (die man per f miteinander identifiziert) und was (b) die Ergebnisse der Verknüpfungen der Addition und der skalare Multiplikation angeht. Das ist oben mit „strukturgleich“ gemeint.Das hat weit reichende Konsequenzen: Alle Begrifflichkeiten, die mit Hilfe der Vektorraumoperationen $+, \cdot$ definiert sind, lassen sich von dem einen Raum in den anderen übertragen.

Zum Beispiel gilt für isomorphe Vektorräume V, W mit einem Isomorphismus $f : V \to W$

- für Linearkombinationen

$$v = \sum_{i=1}^{n} a_i v_i \quad \Longleftrightarrow \quad f(v) = \sum_{i=1}^{n} a_i f(v_i). \qquad \text{(ÜA!)}$$

- $v_1, \ldots, v_n$ sind l.u. in $V \Longleftrightarrow f(v_1), \ldots, f(v_n)$ sind l.u. in W (nach Lemma 9.2.4).

- U ist Unterraum von $V \iff f(U)$ ist Unterraum von V; es gilt dann $\dim(U) = \dim(f(U))$ (nach Satz 9.2.1(vi), (vii), Lemma 9.2.4),

  [...]

Ist f nur linear, bleibt nur jeweils eine der beiden Implikationen in den obigen Äquivalemzen gültig. Es ist eine gute Übung, sich zu überlegen, welche das jeweils ist.

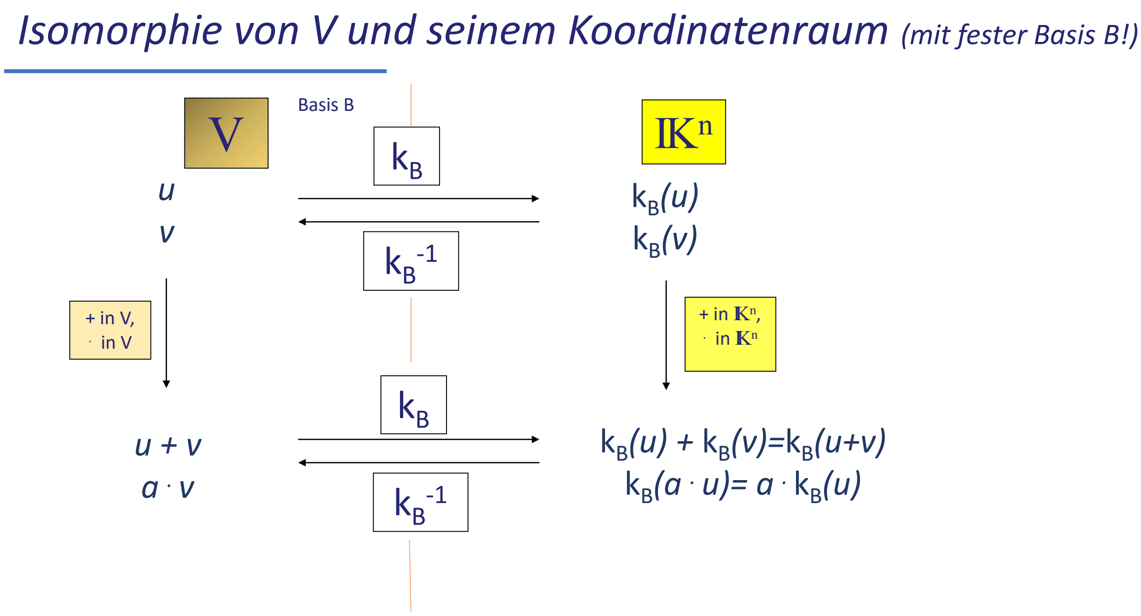
Eine weitere Bemerkung, die schon in Beispiel 2 in 9.1.1 steht, ist besonders hervorzuheben:

Lemma 9.4.2. *Jeder endlichdimensionale $\mathbb{K}$-Vektorraum V ist isomorph zu seinem Koordinatenraum $\mathbb{K}^{\dim V}$.*

Diese Isomorphie wird in der Linearen Algebra andauernd genutzt, z.B., wenn der zweidimensionale oder dreidimensionale Anschauungsraum mit Hilfe des Vektorraumes der Verschiebungen beschrieben wird: Durch Wahl einer Basis (Achsen des Koordinatensystems) kann man jeden Verschiebungsvektor eindeutig als Linearkombination der Einheitsrichtungen ausdrücken (diese sind hier die Basis des Raumes der Verschiebungen im Dreidimenionalen). Die Isomorphie zum Koordinatenraum $\mathbb{R}^3$ bedeutet nun: Es ist egal, ob wir mit anschaulichen Verschiebungen agieren (d.h. diese vektoriell addieren und per Streckung skalar multiplizieren), oder ob wir dies algebraisch (d.h. rechnerisch) mit den zugehörigen Koordinatenvektoren aus dem $\mathbb{R}^2$ bzw. $\mathbb{R}^3$ tun.

Weitergehend ist es ebenfalls egal, ob wir Beziehungen wie lineare Abhängigkeit, Dimensionsfragen usw. für Teilmengen von V oder für ihre Koordinatenvektoren untersuchen. Letzteres ist oft günstiger, da Problemstellungen der Linearen Algebra so auf der Koordinatenebene oft auf lineare Gleichungssysteme heraus laufen (vgl. z.B. das Kriterium für lineare Unabhängigkeit), deren Lösungsverhalten wir mit dem Apparat aus Kapitel 6 vollständig klassifiziert haben und mit den Gauß-Jordan-Algorithmus Lösungen der Probleme automatisiert berechnen können.

Im kommutativen Diagramm sieht das so aus:



Wir wollen den Isomorphiebegriff und isomorphe Vektorräume noch etwas genauer untersuchen.

Lemma 9.4.3. *Isomorphie ist eine reflexive, transitive und symmetrische Relation*[2] *zwischen $\mathbb{K}$-Vektorräumen, genauer gilt für Vektorräume V, W, Z über demselben Körper $\mathbb{K}$:*

- $V \overset{\mathrm{id}_V}{\rightleftharpoons} V$,
- *Gilt* $V \overset{f}{\rightleftharpoons} W$, *so gilt* $W \overset{f^{-1}}{\rightleftharpoons} V$,
- *Gilt* $V \overset{f}{\rightleftharpoons} W$ *und* $W \overset{g}{\rightleftharpoons} Z$, *so gilt* $V \overset{g\circ f}{\rightleftharpoons} Z$.

Beweis: id_V ist immer linear und bijektiv (mit sich selbst als Umkehrabbildung), also ist $V \rightleftharpoons V$. Sind $f : V \to W, g : V \to Z$ Isomorphismen, d.h. linear und bijektiv, so sind auch f^{-1} und $g \circ f$ linear nach Lemma 9.2.2, und f^{-1} und $g \circ f$ auch bijektiv (vgl. ÜA zu Kapitel II bzw. Lemma 2.5.2), also jeweils Isomorphismen. Damit ist gezeigt, dass die oben angegebenen Eigenschaften gelten. □

[2] zum Begriff vgl. das Grundlagenkapitel zu Logik und Mengen, oder sie lesen einfach die folgende Erläuterung dazu, was das hier konkret bedeutet

Satz 9.4.4. *(Isomorphe endlichdimensionale Vektorräume)*

Zwei endlichdimensionale $\mathbb{K}$-Vektorräume V, W sind genau dann isomorph, wenn $\dim V = \dim W$ *gilt.*

Schränkt man sich auf endlichdimensionale Vektorräume ein, so teilt die Relation „$V \rightleftharpoons W$“ also diese Vektorräume gerade in Klassen gleicher Dimension ein, vgl. Satz 7.6.7.

Beweis: Nach den Überlegungen am Ende des letzten Abschnitts 9.3 können nur Vektorräume gleicher Dimension zueinander isomorph sein. Umgekehrt seien V, W zwei $\mathbb{K}$-Vektorräume mit $\dim(V) = \dim(W) = n \in \mathbb{N}$. Dann gilt mit beliebigen Basen B von V und B' von W wegen der Transitivität von $\rightleftharpoons$, dass

$$V \overset{k_B}{\rightleftharpoons} \mathbb{K}^n \quad \text{und} \quad W \overset{k_{B'}}{\rightleftharpoons} \mathbb{K}^n, \quad \text{also} \quad V \overset{k_{B'}^{-1} \circ k_B}{\rightleftharpoons} W.$$

□

9.5 Koordinatenabbildungen und -transformationen im $\mathbb{K}^n$

Ein wichtiger, relativ einfacher, aber dennoch potenziell (gerade für „Einsteiger“, aber auch so) manchmal etwas verwirrender Spezialfall einer Koordinatenabbildung ist der von Koordinatenabbildungen im $\mathbb{K}^n$:

- Allgemein ist nach 8.3.9 jeder Satz aus n linear unabhängigen Vektoren des $\mathbb{K}^n$ eine Basis von $\mathbb{K}^n$. Ist $B = (b_1, \ldots, b_n)$ eine solche Basis, so besteht der Koordinatenvektor eines Vektors $x \in \mathbb{K}^n$ aus den Vorfaktoren in der Darstellung

  $$x = a_1 \cdot b_1 + a_2 \cdot b_2 + \ldots + a_n \cdot b_n, \qquad (\star)$$

  der Koordinatenvektor von $x \in \mathbb{K}^n$ ist also *wieder ein Vektor* $a = (a_1, \ldots, a_n) \in \mathbb{K}^n$ und die Koordinatenabbildung $k_B : \mathbb{K}^n \to \mathbb{K}^n$ ein Automorphismus (also ein Isomorphismus mit identischem Definitionsbereich und Wertevorrat $\mathbb{K}^n$) von $\mathbb{K}^n$.

Es empfiehlt sich also im Umgang mit linearen Abbildungen $f : \mathbb{K}^n \to \mathbb{K}^m$, genau im Auge zu behalten, ob man es gerade mit einem Vektor aus dem $\mathbb{K}^n$ oder mit seinem Koordinatenvektor bezüglich einer gegebenen Basis in $\mathbb{K}^n$ zu tun hat.

- Eine zweite bemerkenswerte Eigenart des $\mathbb{K}^n$ ergibt sich, wenn man die Koordinatenabbildung im $\mathbb{K}^n$ bezüglich der Standardbasis $I_n = (e_1, \ldots, e_n)$ betrachtet: Für alle $x = (x_1, \ldots, x_n) \in \mathbb{K}^n$ ist

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 \cdot e_1 + x_2 \cdot e_2 + \ldots + x_n \cdot e_n,$$

  jeder Vektor $x \in \mathbb{K}^n$ ist also *identisch mit seinem Koordinatenvektor* und die Koordinatenabbildung gegeben durch

$$k_{I_n}(x) \quad = \quad x \quad = \quad I_n \cdot x.$$

Nachdem das klargestellt ist, wollen wir jetzt die Koordinatenabbildungen k_B für eine beliebige Basis B des $\mathbb{K}^n$ bestimmen. Da $k_B : \mathbb{K}^n \to \mathbb{K}^n$ linear ist (vgl. Beispiel 2 in Abschnitt 9.1.1), ist k_B nach Beispiel 1 von dort eine Matrixabbildung, also mit Hilfe einer quadratischen Abbildungsmatrix $T_B \in \mathbb{K}^{n \times n}$ ausdrückbar. Diese wollen wir jetzt bestimmen.

Sei also B eine beliebige Basis des $\mathbb{K}^n$. Die Frage nach der Koordinatenabbildung k_B ist die nach einer Berechnungsvorschrift für die Koeffizienten $a_1, \ldots, a_n$ in ($\star$) oben. Schreiben wir die Vektoren $b_1, \ldots, b_n$ aus B in dieser Reihenfolge in die Spalten einer Matrix und bezeichnen diese wieder mit B, so ist[3]

$$x = a_1 \cdot b_1 + a_2 \cdot b_2 + \ldots + a_n \cdot b_n = \begin{pmatrix} | & | & \cdots & | \\ b_1 & b_2 & \cdots & b_m \\ | & | & \cdots & | \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = B \cdot a.$$

Das ist ein Gleichungssystem $B \cdot a = x$ für a (Vorsicht! x entspricht hier der rechten Seite, die sonst b heißt, a dem gesuchten Vektor), das für alle rechten Seiten x eine eindeutige Lösung hat: B ist eine Basis von $\mathbb{K}^n$, also gibt es genau einen Koordinatenvektor a von x bezüglich B. Die Matrix B ist also invertierbar nach Satz 6.4.1. Multipliziert man die Gleichung oben auf beiden Seiten mit B^{-1}, erhält man $a = B^{-1} \cdot x$, also insgesamt:

[3] Das nennt man oft „abuse of notation“: Es handelt sich bei der Basis B und der Matrix B um verschiedene mathematische Objekte (nämlich ein n-Tupel aus Vektoren bzw. eine $(n \times n)$-Matrix). Man kann die Objekte aber eindeutig (im Sinne einer Bijektion zwischen ihnen) miteinander identifizieren und darf daher die Notation doppelt verwenden, wenn man weiß, was man tut und wenn es wie hier bequem ist.

Lemma 9.5.1. *(Koordinatenabbildungen und -transformationen in $\mathbb{K}^n$)*

Sei $B = (b_1, \ldots, b_n)$ eine Basis des $\mathbb{K}^n$ und bezeichne B wieder ebenfalls[1] die Matrix $B \in \mathbb{K}^{n\times n}$ mit den Elementen von B als Spalten. Dann sind die Koordinatenabbildung k_B bezüglich B und ihre Inverse k_B^{-1} gegeben durch

$$k_B : \mathbb{K}^n \to \mathbb{K}^n, \quad x \mapsto k_B(x) = B^{-1} \cdot x, \qquad k_B^{-1} : \mathbb{K}^n \to \mathbb{K}^n, \quad a \mapsto B \cdot a,$$

Insbesondere rechnet die Abbildung k_B Koordinaten bezüglich der Standardbasis in Koordinaten bezüglich B um.

Ist $C = (c_1, \ldots, c_n)$ eine weitere Basis des $\mathbb{K}^n$ (die wir bei Bedarf wieder als Matrix lesen), so ist die Koordinatentransformation $\tau_{B,C} : \mathbb{K}^n \to \mathbb{K}^n, k_B(x) \mapsto k_C(x)$ mit der sich B-Koordinaten in C-Koordinaten umrechnen lassen, eine die lineare Abbildung mit der Abbildungsmatrix

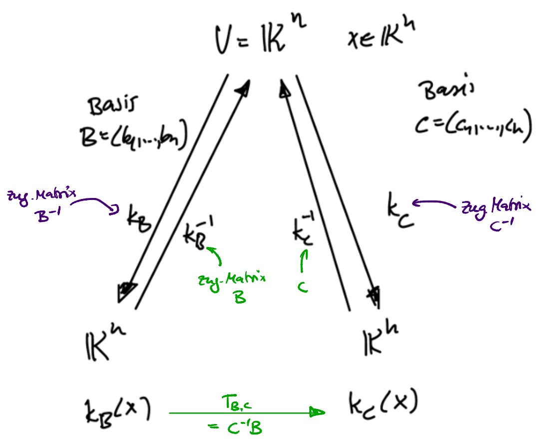
$$T_{B,C} = C^{-1} \cdot B,$$

es ist also für alle $x \in \mathbb{K}^n$

$$k_C(x) = C^{-1}B \cdot k_B(x) \quad \text{und umgekehrt} \quad k_B(x) = B^{-1}C \cdot k_C(x).$$

Beweis: Alle Behauptungen im ersten Absatz haben wir schon in der Herleitung oder vorher bewiesen; insbesondere steht in Lemma 4.4.4, dass die Inverse k_B^{-1} von k_B gerade die Matrixabbildung mit der Matrix $(B^{-1})^{-1} = B$ ist. Die Darstellung für die Koordinatentransformation $\tau_{B,C}$ ergibt sich dadurch, dass man $\tau_{B,C} = k_C \circ k_B^{-1}$ setzt und ausnutzt, dass sich die Verkettung zweier linearer Abbildungen in $\mathbb{K}^n$ nach 4.4.2 als Produkt der beiden Abbildungsmatrizen schreiben lässt. □

Im kommutativen Diagramm wird der Zusammenhang zwischen den einzelnen Transformationen hoffentlich noch klarer:



9.6 Aufgaben zu Kapitel 9

Aufgabe 9.50. (Lineare Abbildungen):

Welche der folgenden Abbildungen sind $\mathbb{R}$-linear? Begründen Sie, und geben Sie in den Fällen, in denen es sich um eine Matrixabbildung handelt, die zugehörige Abbildungsmatrix an.

(i) $f : \mathbb{R}^3 \longrightarrow \mathbb{R}^2, \quad f\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 - x_2 - x_3 \\ 2x_1 + x_2 + x_3 \end{pmatrix},$

(ii) $f : \mathbb{R}^3 \longrightarrow \mathbb{R}^3, \quad f\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_3 \\ x_1 + 1 \\ 0 \end{pmatrix},$

(iii) $f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \quad f\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \cdot x_2 \\ x_1 - x_2 \end{pmatrix},$

(iv) $f : \mathrm{Abb}(\mathbb{R}, \mathbb{R}) \longrightarrow \mathbb{R}, \quad f(g) = g(5),$

(v) $f : \mathbb{R}^2 \longrightarrow \mathbb{R}, \quad f\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 11x_1 + 121x_2.$

Aufgabe 9.51. (Kern und Bild I):

Betrachten Sie die lineare Abbildung

$$f : \mathbb{R}^3 \longrightarrow \mathbb{R}^3, \quad f\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} := \begin{pmatrix} 3x_1 + x_2 - 3x_3 \\ 2x_2 \\ 3x_1 + 2x_2 - 3x_3 \end{pmatrix}.$$

Bestimmen Sie die Zuordnungsvorschrift, welche die Abbildung $f \circ f$ beschreibt, und berechnen Sie $\mathrm{Kern}(f \circ f)$ und $\mathrm{Bild}(f \circ f)$.

Aufgabe 9.52. (Kern und Bild II):

(a) Sei $f : \mathbb{K}^n \to \mathbb{K}^m, x \mapsto A \cdot x$ eine Matrixabbildung. Begründen Sie, dass die Spalten von A ein Erzeugendensystem des Bildes von f sind.

(b) Gegeben sei die lineare Abbildung

$$g : \mathbb{R}^4 \to \mathbb{R}^3, \qquad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \mapsto \begin{pmatrix} x_1 + \frac{1}{2}x_2 - \frac{1}{2}x_3 \\ -x_1 + x_2 - 4x_3 - 3x_4 \\ 2x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_3 + x_4 \end{pmatrix}.$$

Bestimmen Sie jeweils Basen von Kern(g) und Bild(g). Prüfen Sie die Plausibilität Ihres Ergebnisses mit der Dimensionsformel.

(c) Sei $b = \begin{pmatrix} -1 \\ -8 \\ 1 \end{pmatrix} \in \mathbb{R}^3$.

Bestimmen Sie das Urbild von b unter der Funktion g aus Teil (b), d.h. die Menge $\{x \in \mathbb{R}^4 \mid g(x) = b\}$. Geben Sie es in der Form $\mathbb{L} = x^* + U$ an, wobei $x^* \in \mathbb{R}^4$, U ein geeigneter Unterraum von $\mathbb{R}^4$ und

$$x^* + U := \{x + u \mid u \in U\}$$

ist.

Aufgabe 9.53. (Einige Eigenschaften linearer Funktionen):

Seien V, W $\mathbb{K}$-Vektorräume, $v_1, \dots, v_k \in V$ und $f : V \to W$ linear.

(a) Zeigen Sie, dass gilt:

$$\mathcal{L}(f(v_1), \dots, f(v_k)) = f\big(\mathcal{L}(v_1, \dots, v_k)\big).$$

(b) Seien $v_1, \dots, v_k \in V$ linear unabhängig. Zeigen Sie durch ein geeignetes Beispiel, dass die Vektoren $f(v_1), \dots, f(v_k) \in W$ im Allgemeinen nicht linear unabhängig ist.

(c) Sei f injektiv. Zeigen Sie unter dieser Voraussetzung:

(i) Sind $v_1, \ldots, v_k$ linear unabhängig, so sind auch $f(v_1), \ldots, f(v_k) \in W$ linear unabhängig.

(ii) Für jede Basis $B = (b_1, \ldots, b_n)$ von V ist $(f(b_1), \ldots, f(b_n))$ eine Basis von $\mathrm{Bild}(f)$.

(iii) f ist dimensionserhaltend, d.h. für alle Teilräume $U \subseteq V$ gilt $\dim f(U) = \dim U$.

(d) Seien Z ein weiterer $\mathbb{K}$-Vektorraum und $g : W \to Z$ eine weitere lineare Abbildung. Zeigen Sie die Inklusionen $\mathrm{Kern}(f) \subseteq \mathrm{Kern}(g \circ f)$ und $\mathrm{Bild}(g \circ f) \subseteq \mathrm{Bild}(g)$.

Aufgabe 9.54. (Kern und Bild III):

Finden Sie für folgende Fragen entweder ein konkretes Beispiel oder geben Sie eine Begründung, falls es keine solche Abbildung gibt.

(a) Gibt es lineare Abbildungen $f : \mathbb{R}^2 \to \mathbb{R}^2$ mit $\mathrm{Bild}(f) = \mathrm{Kern}(f)$?

(b) Gibt es lineare Abbildungen $f : \mathbb{R}^3 \to \mathbb{R}^3$ mit $\mathrm{Bild}(f) = \mathrm{Kern}(f)$?

Aufgabe 9.55. (Koordinatentransformation in $\mathbb{C}^2$):

Wir betrachten den $\mathbb{C}$-Vektorraum $V = \mathbb{C}^2$.

(a) Das Tupel $B = (\begin{pmatrix} 1 \\ i \end{pmatrix}, \begin{pmatrix} i \\ 1 \end{pmatrix})$ ist eine Basis von V. Berechnen Sie Koordinatenvektor $k_B(v)$ des Vektors

$$v = \begin{pmatrix} 1+2i \\ -i \end{pmatrix} \in \mathbb{C}^2.$$

(b) Das Tupel

$$C := (\begin{pmatrix} i \\ 0 \end{pmatrix}, \begin{pmatrix} 2-i \\ -i \end{pmatrix})$$

ist eine weitere Basis von $\mathbb{C}^2$. Berechnen Sie die Koordinatentransformationsmatrix $T_{B,C} \in \mathbb{C}^{2\times 2}$, mit der sich Koordinaten bezüglich der Basis B in Koordinaten bezüglich der Basis C umrechnen lassen. Berechnen Sie anschließend mit ihrer Hilfe $k_C(v)$ für den Vektor v aus (a).

10 Darstellung linearer Abbildungen über Matrizen

10.1 Konstruktion darstellender Matrizen

In Kapitel 9.4 haben wir gezeigt, dass ein endlichdimensionaler $\mathbb{K}$-Vektorraum immer zu seinem Koordinatenraum $\mathbb{K}^{\dim V}$ isomorph ist, dass man also nichts von der Struktur des zu Grunde liegenden Vektorraumes V verliert, wenn man statt mit den Vektoren selbst mit ihren Koordinaten rechnet. Wir wollen nun zeigen, dass sich auch lineare Abbildungen zwischen Vektorräumen bequem auf der Koordinatenebene mit Hilfe von Matrizen darstellen lassen, ohne dabei Information zu verlieren[1].

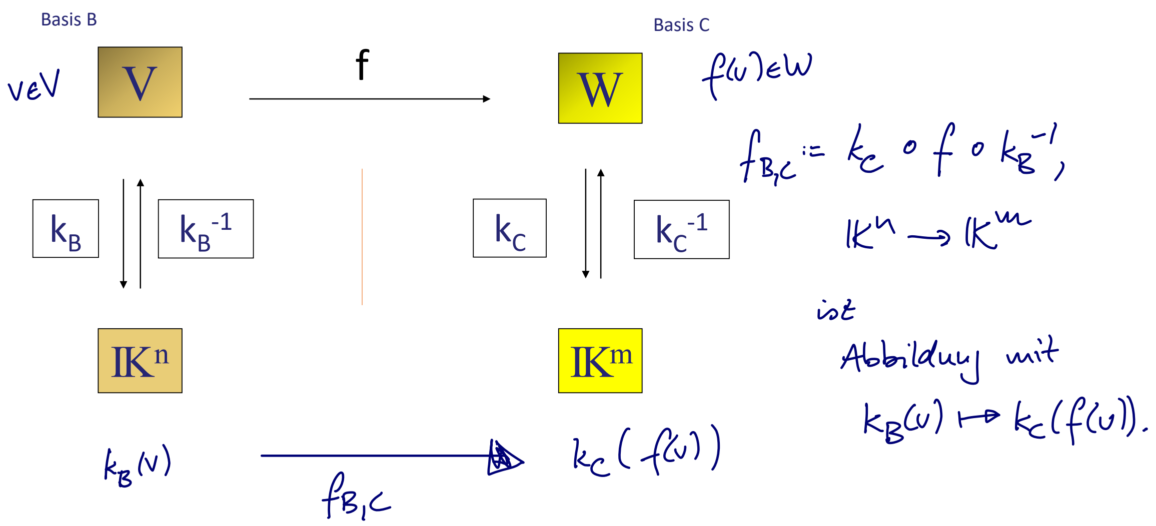
Genauer seien in diesem Abschnitt V und W beliebige endlichdimensionale K-Vektorräume mit $\dim V = n$ und $\dim W = m$. Wir wollen nun zeigen, dass sich jede lineare Abbildung $f : V \to W$ vollständig mit Hilfe einer Matrix $A^f \in \mathbb{K}^{m\times n}$ (also durch Angabe von $n \cdot m$ Zahlen!) beschreiben lässt. Wie genau die Matrix aussieht, hängt von der Wahl der Basen in V und W ab.

- Bezeichne $B = (b_1, \ldots, b_n)$ daher im Folgenden eine fest gewählte Basis von V und $C = (c_1, \ldots, c_m)$ eine Basis von W.

Wir erinnern uns zunächst an einige Fakten, die wir schon bewiesen haben. Wir werden diese dann geeignet zusammensetzen, um die Konstruktion darstellender Matrizen zu verstehen – und zu verstehen, wie die Information über f in ihnen „codiert“ wird.

Wenden Sie dazu bitte Ihren Blick dem Diagramm auf der nächsten Seite zu und versuchen Sie, sich die darauf folgenden Tatsachen daran klarzumachen.

[1] Disclaimer: Wir werden in diesem Kapitel vielen der Ergebnisse zur Darstellung von Matrixabbildungen $f : \mathbb{R}^n \to \mathbb{R}^m$ aus Kapitel II wieder begegnen. Weil wir dort noch nicht wussten, was ein Körper ist, hatten wir uns dort auf den „Schulkörper“ $\mathbb{R}$ beschränkt. Alle Resultate dort gelten aber analog auch allgemein für die Vektorräume $\mathbb{K}^n$ bzw. $\mathbb{K}^m$ über einem beliebigen Körper $\mathbb{K}$. Das lässt sich prüfen, indem man verifiziert, dass in allen im Kapitel II dargestellten Zusammenhängen nur die Körpereigenschaften von $\mathbb{R}$ und die Vektorraumeigenschaften von $\mathbb{R}^n, \mathbb{R}^m$ eine Rolle gespielt haben, wir aber nie die konkrete Gestalt von $\mathbb{R}$ ausgenutzt haben. Wir werden das hier nicht explizit nachprüfen; tun Sie das für sich, es ist eine gute Übung! Im Folgenden werden wir mit der vorangegangenen Rechtfertigung aber die Gültigkeit der dort bewiesenen Resultate für allgemeine Körper $\mathbb{K}$ und allgemeine Matrixabbildungen $f : \mathbb{K}^n \to \mathbb{K}^m$ benutzen und an entsprechender Stelle darauf hinweisen.



- Der endlichdimensionale Vektorraum V ist isomorph zu seinem Koordinatenraum $\mathbb{K}^n$, d.h. insbesondere gehört bezüglich einer festen Basis B zu jedem Vektor $v \in V$ genau ein Koordinatenvektor $k_B(v) \in \mathbb{K}^n$ und umgekehrt. Dasselbe gilt für Vektoren $w \in W$ und zugehörige Koordinatenvektoren $k_C(w) \in \mathbb{K}^m$. (Das zeigen die in beide Richtungen führenden Pfeile im Diagramm, die V und $\mathbb{K}^n$ bzw. W und $\mathbb{K}^m$ verbinden).

- Daher gilt (das werden wir gleich genauer zeigen): Statt für f eine Abbildungsvorschrift

$$f : V \to W, \qquad v \mapsto f(v)$$

  anzugeben, können wir alternativ auch die Abbildung

$$f_{B,C} : \mathbb{K}^n \to \mathbb{K}^m, \qquad k_B(v) \mapsto k_C(f(v))$$

  angeben, die festlegt, wie man jedem Koordinatenvektor eines Vektors aus V den entsprechenden Koordinatenvektor seines Bildes $f(v)$ in W zuordnet.

- Die Abbildung $f_{B,C} : \mathbb{K}^n \to K^m$ ist linear, lässt sich also als Matrixmultiplikation mit einer Matrix $A^f_{B,C} \in \mathbb{K}^{m \times n}$ schreiben. Die Wirkung der lineare Abbildung f lässt sich also dadurch berechnen, dass man Koordinatenvektoren der abzubildenden Vektoren aus v mit der Matrix $A^f_{B,C}$ multipliziert; man erhält die Koordinaten des Bildvektors $f(v)$.

Den Beweis der gerade ausgelisteten Tatsachen und die genaue Form der darstellenden Matrix $A^f_{B,C}$ liefert der folgende Satz:

Satz/Definition 10.1.1. *(Darstellende Matrizen)*

Seien V, W wie oben als endlichdimensional mit $\dim V = n$ und $\dim W = m$ vorausgesetzt, $f : V \to W$ $\mathbb{K}$-linear.

(i) Die Abbildung

$$f_{B,C} : \mathbb{K}^n \to \mathbb{K}^m, \qquad k_B(v) \mapsto k_C(f(v)),$$

die die B-Koordinaten eines Vektors v auf die C-Koordinaten seines Bildes $f(v)$ unter f abbildet. ist wohldefinierte, lineare Abbildung von $\mathbb{K}^n$ nach $\mathbb{K}^m$; es gilt

$$f_{B,C} \;:=\; k_C \circ f \circ k_B^{-1}. \tag{10.1}$$

(ii) Zu $f_{B,C} \in L(\mathbb{K}^n, \mathbb{K}^m)$ gibt es eine Matrix $A^f_{B,C} \in \mathbb{K}^{m \times n}$, so dass für alle $x \in \mathbb{K}^n$

$$f_{B,C}(x) = A^f_{B,C} \cdot x$$

ist. $A^f_{B,C}$ heißt die <u>darstellende Matrix der Abbildung f bezüglich der Basen B und C</u>. Sie besitzt die Form

$$A^f_{B,C} \;=\; \begin{pmatrix} | & | & \cdots & | \\ k_C(f(b_1)) & k_C(f(b_2)) & \cdots & k_C(f(b_2)) \\ | & | & \cdots & | \end{pmatrix}.$$

Für alle $v \in V$ gilt hiermit

$$k_C(f(v)) \;=\; A_{B,C} \cdot k_B(v).$$

(iii) Ist $A_{B,C} = (a_{i,j})_{i \in \underline{m}, j \in \underline{n}}$, so lässt sich aus der j-ten Spalte das Bild

$$f(b_j) = \sum_{i=1}^{m} a_{i,j} c_i$$

des j-ten Basisvektors $b_j \in B$ unter f ablesen.

Beweis: Zur Beweis von (i) sei $x \in \mathbb{K}^n$ beliebig. Wir bemerken, dass x wegen der Surjektivität der Koordinatenabbildung k_B der Koordinatenvektor genau eines Vektors $v \in V$ ist, also $x = k_B(v)$ gilt; damit ist $f_{B,C}(x) = k_C(f(v))$ für alle $x \in \mathbb{K}^n$ eindeutig festgelegt.

Es gilt

$$(k_C \circ f \circ k_B^{-1})(x) = (k_C \circ f \circ k_B^{-1} \circ k_B)(v) = (k_C \circ f)(v) = k_C(f(v)) = f_{B,C}(x)$$

wie behauptet; das lässt sich auch gut am Diagramm nachvollziehen. Da die Hintereinanderausführung linearer Abbildungen linear ist, zeigt das (i).

Zu (ii) konstruieren wir die Abbildungsmatrix $A_{B,C}^f$ der linearen Abbildung $f_{B,C} : \mathbb{K}^n \to \mathbb{K}^m$ wie in Def. 4.3.2: Die Spalten der Matrix sind genau die Bildvektoren Der Standardbasis des $\mathbb{K}^n$ unter f, d.h. $f_{B,C}(e_1), \ldots, f_{B,C}(e_n)$. Die Erkenntnis, dass die Elemente $b_1, \ldots, b_n$ von B per Koordinatenabbildung gerade den Standardbasisvektoren $e_1, \ldots, e_n$ entsprechen (d.h. $k_B(b_i) = e_i$ für $i \in \underline{n}$) liefert mit (i) für die i-te Spalte der Matrix

$$f_{B,C}(e_i) = (k_C \circ f \circ k_B^{-1})(k_B(b_i)) = k_C(f(b_1)).$$

Zu (iii): Wegen der Umkehrbarkeit der Koordinatenabbildungen gilt

$$f_{B,C} \; := \; k_C \circ f \circ k_B^{-1} \quad \Longrightarrow \quad f \; := \; k_C^{-1} \circ f_{B,C} \circ k_B.$$

Daher ist

$$f(b_j) = k_C^{-1} \circ f_{B,C} \circ k_B(b_j) = k_C^{-1}(A_{B,C} \cdot e_j) = k_C^{-1}(a_{i,j})_{i \in \underline{m}} = \sum_{i=1}^{m} a_{i,j} c_i.$$

□

Definition 10.1.2. *(Notationen für darstellende Matrizen)*
Ist $V = W$ und $B = C$, so spricht man kürzer von der darstellenden Matrix A_B von f bezüglich B. Wollen wir hervorheben, dass es sich um die darstellende Matrix von f handelt (etwa, wenn mehrere Abbildungen im Spiel sind), so schreiben wir $A_{B,C}^f$, wenn f klar ist, aber auch $A_{B,C}$ für ihre darstellende Matrix.

Bemerkungen dazu:

- Bei fest gewählten Basen B, C ist Abbildung

  $$L(V, W) \to \mathbb{K}^{m \times n}, \quad f \mapsto A_{B,C}^f,$$

  die jeder linearen Abbildung die zugehörige darstellende Matrix $A_{B,C}^f$ bzgl. B und C zuordnet, eine Bijektion zwischen diesen Räumen mit $A \mapsto k_C^{-1} \circ f_A \circ k_B$ als Umkehrabbildung (f_A bezeichnet die zu A gehörige Matrixabbildung).

- Man kann sogar zeigen, dass die obige Abbildung ein Vektorraumisomorphismus zwischen $L(V,W)$ und $\mathbb{K}^{m\times n}$ ist, d.h. es gilt $A_{B,C}^{f+g} = A_{B,C}^{f} + A_{B,C}^{g}$ und $A_{B,C}^{a\cdot f} = a \cdot A_{B,C}^{f}$. Wir werden den Beweis hier allerdings nicht führen, vgl. Literatur, z.B. Skript von C. Tischendorf.

- Berechnet man darstellende Matrizen $A_{B,C}$ und $A_{B',C'}$ für verschiedene Basen B, B' von V, C, C' von W, so ergeben sich für gewöhnlich verschiedene darstellende Matrizen. Daher reden wir allgemein von *einer* darstellenden Matrix von f, aber von *der* darstellenden Matrix von f (bezüglich B und C), wenn diese Basen fest gewählt sind.

Aus dem obigen Satz ergibt sich der folgende

Algorithmus zur Berechnung einer darstellenden Matrix für lineares $f : V \to W$

Gegeben: $f : V \to W$ linear, Basis B von V, Basis C von W.

1. Bestimme $f(b_1), f(b_2), \ldots, f(b_n)$,
2. Schreibe $f(b_1), f(b_2), \ldots, f(b_n)$ jeweils als Linearkombination der Vektoren aus C.
3. Lese die Koordinatenvektoren $k_C(f(b_1)), k_C(f(b_2)), \ldots, k_C(f(b_n))$ daraus ab; $k_C(f(b_j))$ ist die j-te Spalte von $A_{B,C}$.

Beispiel: (Vgl. Vorlesung)

Wir berechnen die darstellende Matrix der Ableitungsabbildung

$$\frac{d}{dx} : \mathbb{K}_4[x] \to \mathbb{K}_4[x], \qquad \left(p = \sum_{i=0}^{4} a_i x^i\right) \mapsto \left(p' := \sum_{i=1}^{4} i \cdot a_i x^{i-1}\right)$$

bezüglich der Monombasis $B = C = (1, x, x^2, x^3)$.

Man erhält

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

Lemma 10.1.3. *(Darstellung von Verknüpfungen von linearen Abbildungen)*

Seien $f : V \to W$ und $g : W \to Z$ linear, B eine Basis von V, C eine Basis von W und D eine Basis von Z. Weiterhin sei $A^f_{B,C}$ die darstellende Matrix von f bezüglich B und C, $A^g_{C,D}$ die darstellende Matrix von g bezüglich C und D. Dann ist

$$A^{g\circ f}_{B,D} \;=\; A^g_{C,D} \cdot A^f_{B,C}, \tag{10.2}$$

die darstellende Matrix der linearen Abbildung $g \circ f \in L(V, Z)$ bezüglich B und D, diese ergibt sich also einfach durch Multiplikation der darstellenden Matrizen von f und g.

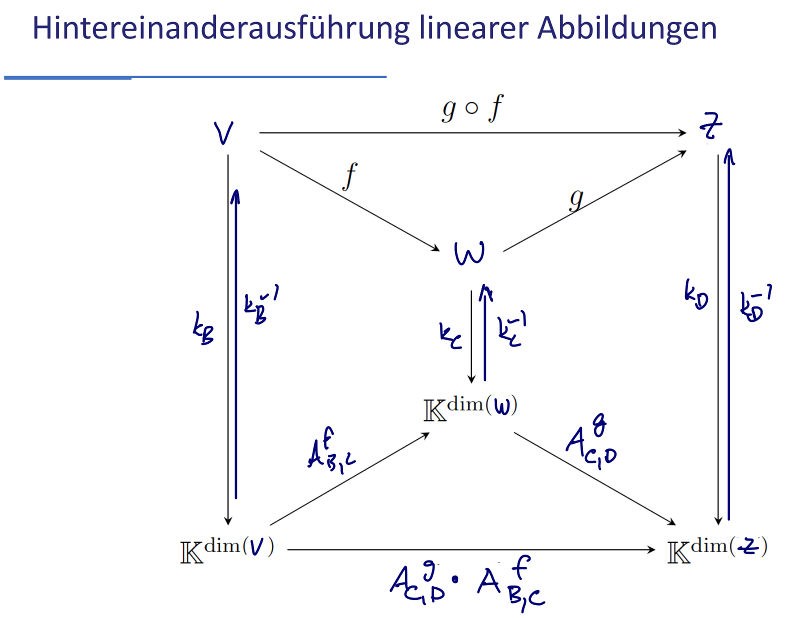


Bild: Jana Bielagk, HU Berlin

Beweis: Ist $\dim(Z) = \ell$, so ist mit den Bezeichnungen aus Definition 10.1.1 die Abbildung $(g \circ f)_{B,D}$ eine lineare Abbildung $\mathbb{K}^n \to K^\ell$, für die

$$g_{C,D} \circ f_{B,C} \;:=\; (k_D \circ g \circ k_C^{-1}) \circ (k_C \circ f \circ k_B^{-1}) \;=\; k_D \circ (g \circ f) \circ k_B^{-1} \;=:\; (g \circ f)_{B,D}$$

gilt, die sich also durch Hintereinanderausführung der Abbildungen $f \in L(\mathbb{R}^n, \mathbb{R}^m)$ und $g \in L(\mathbb{R}^m, \mathbb{R}^\ell)$ schreiben lässt. Die zu $(g \circ f)_{B,D}$ gehörige Abbildungsmatrix ergibt sich daher nach Satz 4.4.2 durch die Matrixmultiplikation 10.2.

□

Bemerkungen dazu:

- Bei der Verknüpfung von linearen Abbildungen über ihre darstellenden Matrizen ist wichtig, dass zur Darstellung von f und g im „mittleren“ Raum W dieselbe Basis C benutzt wird. Ist dies nicht der Fall, so kann man zwischen die Matrizen $A^g_{C',D}, A^f_{B,C}$ noch eine Koordinatentransformation von C auf C' schalten, s.u.

- Insbesondere gilt: Ist für lineares $f : V \to W$ eine darstellende Matrix durch $A_{B,C}$ gegeben, so ist f ein Isomorphismus zwischen V und W genau dann, wenn $A_{B,C}$ invertierbar ist; $(A_{B,C})^{-1}$ ist dann die darstellende Matrix der Abbildung $f^{-1} : W \to V$ bezüglich der Basen C und B nach Satz 6.4.1.

10.2 Darstellung von Matrixabbildungen $f : \mathbb{K}^n \to \mathbb{K}^m$ bezüglich „Nicht-Standardbasen“

Wir betrachten nun ncoh einmal genauer den auch für den Weitergang dieses Kurses wichtigen Spezialfall $V = \mathbb{K}^n$, $W = \mathbb{K}^m$, in dem lineare Abbildungen mit Hilfe ihrer Abbildungsmatrix $A \in \mathbb{K}^{m\times n}$ über

$$f : \mathbb{K}^n \to \mathbb{K}^m, \quad x \mapsto A \cdot x \qquad (\star)$$

beschrieben werden können: f bildet durch Multiplikation mit A Vektoren $x \in \mathbb{K}^n$ auf ihre Bilder $f(x) \in \mathbb{K}^n$ ab.

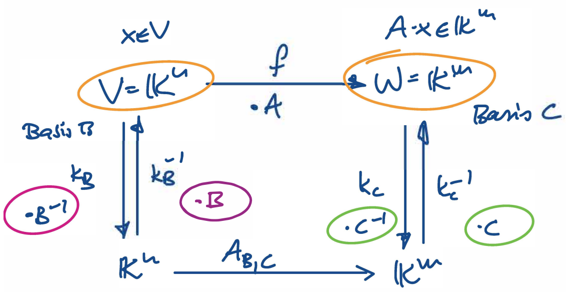
Unsere entwickelte Theorie befähigt uns nun dazu, f auch bezüglich anderer Basen B des $\mathbb{K}^n$ und C des $\mathbb{K}^m$ darzustellen; Ergebnis ist wiederum eine zu f und B, C gehörige darstellende Matrix $A_{B,C}$; diese „akzeptiert“ B-Koordinatenvektoren des Inputvektors x und liefert den C-Koordinatenvektor des Bildes $f(x)$, d.h. ergibt sich durch Multiplikation von $A_{B,C}$ mit dem Koordinatenvektor $k_B(x) \in \mathbb{R}^n$ der Koordinatenvektor $k_C(f(x)) \in \mathbb{R}^m$ von $f(x)$. Das folgende Diagramm verdeutlicht den Zusammenhang zwischen A als „kanonische“ (d.h. so etwas wie „Standard-“) Darstellung der Abbildung f und der Matrix $A_{B,C}$, die Koordinaten bezüglich B in Bildkoordinaten bezüglich C überführt.

Darstellung von Matrixabbildungen bezüglich anderer Basen

In diesem Abschnitt

$f: \mathbb{K}^n \to \mathbb{K}^m$ linear, d.h.

$f: x \mapsto A \cdot x$ (A: Abbildungsmatrix der Matrixabbildung f)



$$\boxed{A_{B,C} = C^{-1} \cdot A \cdot B}$$

$$A = C \cdot A_{B,C} \cdot B^{-1}$$

Wir werden die Matrix A aus $(\star)$ oben nach wie vor als die zu f gehörige Abbildungsmatrix bezeichnen. Alle Matrizen, die Koordinatenvektoren bezüglich irgendwelcher Basen auf Koordinaten der Bilder unter f abbilden, bezeichnen wir (wie vorher) als darstellende Matrizen (mit der Schreibweise $A_{B,C}$).

Da Vektoren aus dem $\mathbb{K}^n$ ihre eigenen Koordinatenvektoren bezüglich der Standardbasis I_n sind (vgl. Abschnitt 9.5), ist die Abbildungsmatrix A gleichzeitig die darstellende Matrix A_{I_n,I_n} von f bezüglich der Standardbasis I_n. Im Allgemeinen müssen die Matrix A und die Matrizen $A_{B,C}$ aber, wie gerade erörtert, unterschieden werden.

Für solche darstellende Matrizen $A_{B,C}$ erhalten wir durch Zusammensetzen von Lemma 9.5.1 und Lemma 10.3.1 die folgenden Matrixtransformationsformeln.

Lemma 10.2.1. *(Matrixtransformation für $A \in \mathbb{K}^{m\times n}$)*

Sei $f : \mathbb{K}^n \to \mathbb{K}^m, x \mapsto A \cdot x$ eine Matrixabbildung mit Abbildungsmatrix $A \in \mathbb{K}^{m\times n}$. Sei weiter B eine Basis von $\mathbb{K}^n$ und C eine Basis von $\mathbb{K}^m$. $A_{B,C}$ sei die darstellende Matrix von f bezüglich B und C.

Benutzen wir wieder die Notation B bzw. C auch für die Matrizen $B \in \mathbb{K}^{n\times n}$ bzw. $C \in \mathbb{K}^{m\times m}$ mit den Vektoren der Basen B bzw. C als Spalten, so gelten die Umrechnungsformeln

$$A_{B,C} = C^{-1} \cdot A \cdot B \quad \text{und} \quad A = C \cdot A_{B,C} \cdot B^{-1}.$$

Im wichtigen Spezialfall $\mathbb{K}^n = \mathbb{K}^m$ und $B = C$ vereinfacht sich das zu

$$A_{C,C} = C^{-1} \cdot A \cdot C \quad \text{und } A = C \cdot A_{C,C} \cdot C^{-1}.$$

Allgemeiner gilt, falls $A_{B,C}$ und $A_{B',C'}$ darstellende Matrizen von f bezüglich Basen B, B' von $\mathbb{K}^n$ bzw. C, C' von $\mathbb{K}^m$ sind, mit den in Lemma 9.5.1 definierten Koordinatentransformationsmatrizen $T_{B,C}$:

$$A_{B,C} = T_{C',C} \cdot A_{B',C'} \cdot T_{B,B'}.$$

An dieser Stelle passt die folgende Begriffseinführung gut:

Definition 10.2.2. *(Ähnliche Matrizen)*

Sei $n \in \mathbb{N}$ fest gewählt. Zwei quadratische Matrizen $M, M' \in \mathbb{K}^{n\times n}$ nennt man ähnlich, wenn es eine invertierbare Matrix C gibt, so dass gilt:

$$M = C \cdot M' \cdot C^{-1} \qquad (\star\star)$$

Bemerkungen dazu:

- Ähnlichkeit ist eine Äquivalenzrelation auf $\mathbb{K}^{n\times n}$ (Übungsaufgabe).
- Das Lemma 10.2.1 oben zeigt, dass zwei quadratische Matrizen genau dann ähnlich sind, wenn es eine Matrixabbildung $f : \mathbb{K}^n \to \mathbb{K}^n$ und Basen B, B' von $\mathbb{K}^n$ gibt, so dass $M = A_{B,B}$ und $M' = A_{B',B'}$ darstellende Matrizen von f sind.

Beweis der zweiten Bemerkung: In der Tat, gilt ($\star\star$), so interpretieren wir M als Abbildungsmatrix (also als die darstellende Matrix einer Matrixabbildung bezüglich der Standardbasis), und M' ist dann nach dem Lemma 10.2.1 oben die darstellende Matrix bezüglich der Basis C. Ist $M = A_{B,B}$ und $M' = A_{C,C}$, so ergibt sich mit der Transformationsformel am Ende von Lemma 10.2.1, dass $M = T_{B',B} \cdot M \cdot T_{B,B'}$, also folgt die Behauptung wegen $T_{B,B'} = T_{B',B}^{-1}$.

Beispiele für Matrixtransformationen

Die nun folgenden Beispiele sollen mögliche Anwendungen von Matrixtransformationen für $A \in \mathbb{R}^{m\times n}$ verdeutlichen. Wir betrachten dazu der Einfachheit halber $V = W = \mathbb{R}^2$ (so dass wir f als Abbildung in der Ebene deuten können) und neben der Standardbasis $I = I_2 = (e_1, e_2)$ die Basis

$$B = (b_1, b_2) = \left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right),$$

die wir uns als um 45° im Uhrzeigersinn gedrehtes, mit $\sqrt{2}$ skaliertes Koordinatensystem vorstellen können.

Ist B eine beliebige Basis eines Vektorraumes V, so ist eine lineare Abbildung $f : V \to W$ nach Lemma 9.2.1(iii) durch Angabe von $f(b_i)$, $i \in \underline{n}$, vollständig festgelegt. Wir können nun Abbildungsmatrizen von f wie in den folgenden Beispielen berechnen.

Beispiel 1: Typisches Rechenbeispiel

Sei $f : \mathbb{R}^2 \to \mathbb{R}^2$ eine lineare Abbildung, die durch

$$f\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}, \qquad f\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix} \qquad (\dagger)$$

festgelegt ist. Da die Bilder bezüglich der Standardbasis angegeben sind, ist mit dieser Festlegung

$$A_{B,I} = \begin{pmatrix} 3 & 0 \\ 2 & -1 \end{pmatrix},$$

also nach Lemma 10.2.1

$$A \;=\; I \cdot A_{B,I} \cdot B^{-1} \;=\; A_{B,I} \cdot B^{-1}.$$

Z.B. mit der Methode aus 6.5 berechnet man

$$B^{-1} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

und damit

$$A = \begin{pmatrix} \frac{3}{2} & -\frac{3}{2} \\ \frac{1}{2} & -\frac{3}{2} \end{pmatrix}$$

als die gesuchte Abbildungsmatrix der linearen Abbildung f, die durch (†) festgelegt wurde.

Beispiel 2: Eine geometrische Abbildung

Wir können das Verfahren aus Beispiel 1 nutzen. um z.B. geometrische Abbildungen wie Spiegelungen oder Projektionen auf „schief zu den Koordinatenachsen liegenden" Geraden oder Ebenen zu definieren. Wir verdeutlichen das, indem wir die Abbildungsmatrix einer Spiegelung an der durch $(1, 2)$ erzeugten Gerade berechnen. Dazu legen wir

$$f\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \qquad f\begin{pmatrix} -2 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

fest. Das ist dadurch motiviert, dass mit dieser Festlegung für diese beiden Basisvektoren, die wir c_1, c_2 nennen wollen, $f(c_1) = c_1$ und $f(c_2) = -c_2$ ist, d.h. die durch c_1 erzeugte Spiegelachse bleibt fest, die Richtung senkrecht zur Spiegelachse[2] wird per Multiplikation mit (-1) „auf die andere Seite geklappt". Analog zu Beispiel 1 haben wir

$$A_{C,I} = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix},$$

wegen

$$f(c_1) = c_1 = 1 \cdot c_1 + 0 \cdot c_2, \qquad f(c_2) = -c_2 = 0 \cdot c_1 + (-1) \cdot c_2,$$

erhält man die „schönere" Darstellung

$$A_{C,C} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

[2] Man kann sich mit Hilfe einer Zeichnung überzeugen, dass die beiden Vektoren senkrecht stehen, oder sich an die Schulzeit erinnern und erkennen, dass das Skalarprodukt der beiden Vektoren $1 \cdot (-2) + 2 \cdot 1 = 0$ ist.

Wie in Beispiel 1 berechnet man damit die Abbildungsmatrix

$$A = A_{I,I} = C \cdot A_{C,C} \cdot C^{-1} = \begin{pmatrix} -\frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{pmatrix}$$

d.h. die Darstellung dieser Spiegelung bezüglich der Standardbasis. Dieser sieht man die Eigenschaften der zugehörigen geometrischen Abbildung im Gegensatz zur Darstellung durch $A_{C,C}$ nicht mehr unmittelbar an. Weitere Möglichkeiten, auf ähnliche Weisen geometrische Abbildungen in der Ebene, im Raum und allgemeiner in $\mathbb{K}^n$ zu definieren, werden wir uns in der LAAG II ansehen, wenn wir uns genauer mit den so genannten affinen Punkträumen beschäftigen.

Beispiel 3: Diagonalisierung

Als drittes Beispiel betrachten wir die Matrixabbildung f mit Abbildungsmatrix

$$A = \begin{pmatrix} 5 & 1 \\ 1 & 5 \end{pmatrix}$$

Stellen wir f bezüglich der in Beispiel 1 benutzten Basis B dar, so ergibt sich

$$A_{B,B} = B^{-1} \cdot A \cdot B = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \cdot \begin{pmatrix} 5 & 1 \\ 1 & 5 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 0 \\ 0 & 6 \end{pmatrix}.$$

Es ist hier also

$$f(b_1) = 4 \cdot b_1 + 0 \cdot b_2 = 4b_1, \qquad f(b_2) = 0 \cdot b_1 + 6 \cdot b_2 = 6b_2.$$

Interpretiert man die Wirkung von f im durch B induzierten Koordinatensystem, so bewirkt f in x-Richtung eine Streckung („Skalierung“) um Faktor 4, in y-Richtung eine mit Faktor 6. Der Rest ergibt sich wie immer durch lineare Fortsetzung; der Einheitskreis wird auf eine Ellipse abgebildet, deren Hauptachsen auf den x- und y-Achsen des neuen Koordinatensystems liegen und die die Länge 4 bzw. 6 haben.

Gibt es eine Basis B, die durch die Wirkung von f nur skaliert wird, so ist die darstellende Matrix $A_{B,B}$ besonders einfach, sie hat wie die Matrix $A_{B,B}$ oben nur für die Diagonalelemente $a_{1,1}, a_{2,2}$ Einträge ungleich Null. Gibt es eine solche Basis, heißt A diagonalisierbar; Diagonalisierbarkeit ist ebenfalls ein größeres Thema der LAAG II (bei dem die Hauptfrage ist, ob alle Matrizen diagonalisierbar sind (nein) und woran man ggf. anhand einer Abbildungsmatrix ihre Diagonalisierbarkeit und ihre „Streckrichtungen“ (die so genannten Eigenräume von A) erkennt.)

10.3 Koordinatentransformationen in allgemeinen Vektorräumen

In Abschnitt 9.5 haben wir bereits gesehen, wie man im $\mathbb{K}^n$ Koordinaten bezüglich einer Basis B in Koordinaten bezüglich einer Basis C umrechnet (nämlich mit der Matrix $C^{-1}B$). Zum Schluss dieses Kapitels über Matrixdarstellungen wenden wir uns noch einmal dem Fall allgemeiner Vektorräume zu und entwickeln nun für die Koordinstentransformation ein analoges Verfahren:

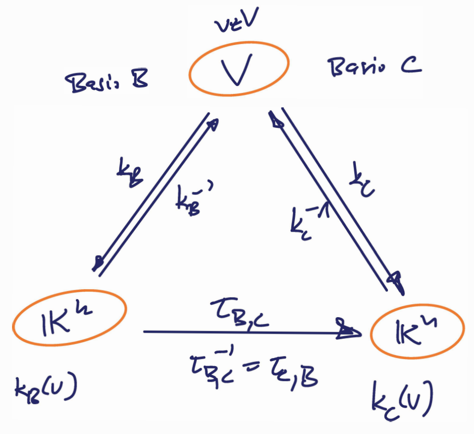
Für zwei Basen B, C eines beliebigen Vektorraumes V ist die Koordinatentransformation von Koordinaten bezüglich B auf solche bezüglich C gegeben durch

$$\tau_{B,C} : \quad \mathbb{K}^n \to \mathbb{K}^n, \quad a \mapsto k_C \circ k_B^{-1}(a)$$

.

Koordinatentransformationen in allgemeinen Vektorräumen

$\tau_{B,C} : \mathbb{K}^n \to \mathbb{K}^n, \; a \mapsto (k_C \circ k_B^{-1})(a)$



Beobachtungen

- $\tau_{B,C}$ ist eine Matrixabbildung (weil linear als Hintereinanderausführung von k_C, k_B^{-1}, Abb von $\mathbb{K}^n$ nach $\mathbb{K}^n$).
- Wegen

  $k_C \circ k_B^{-1} = k_C \circ id_V \circ k_B^{-1}$

  ist die Berechnung einer Basistrafo-matrix der Spezialfall $f = id_V$ im Berechnungsalgorithmus für darst. Matrizen von oben.

Durch $\tau_{B,C}$ wird der „B-Koordinatenvektor" eines Vektors $v \in V$ auf v abgebildet, v dann auf die Koordinaten bezüglich C. Da

$$k_C \circ k_B^{-1} = k_C \circ \mathrm{id}_V \circ k_B^{-1}$$

ist, ist das gerade der Spezialfall $f = \mathrm{id}_V$ in Definition 10.1.1: Die darstellende Matrix der identischen Abbildung id_V in V bezüglich B und C rechnet also Koordinaten bezüglich B in Koordinaten bezüglich C um.

Wendet man für die Berechnung der zu $\tau_{B,C}$ gehörigen Transformationsmatrix $T_{B,C}$ den Algorithmus auf Seite 222 auf $f = \mathrm{id}_V$ an, so bleibt davon:

Algorithmus zur Berechnung einer Basistransformationsmatrix $T_{B,C}$

Gegeben: Basen B, C von V.

2.' Schreibe $b_1, b_2, \ldots, b_n$ jeweils als Linearkombination der Vektoren aus C.

3. Lese die Koordinatenvektoren $k_C(b_1), k_C(b_2), \ldots, k_C(b_n)$ daraus ab;
   $k_C(b_i)$ ist die i-te Spalte von $T_{B,C}$.

Die resultierende Matrix $T_{B,C}$ leistet dann $k_C(v) = T_{B,C} \cdot k_B(v)$ für alle $v \in V$.

Als krönenden Abschluss der theoretischen Betrachtungen in diesem Kapitel leiten wir her, wie sich die darstellende Matrix $A_{B,C}$ einer linearer Abbildungen $f : V \to W$ zwischen endlichdimensionalen Vektorräumen bezüglich Basen B und C umrechnen lassen, wenn man zur Darstellung von f stattdessen eine Basis B' für V und eine Basis C' für W benutzen will, also die Matrix $A_{B',C'}$ sucht.

Lemma 10.3.1. *(Umrechnung darstellender Matrizen)*

Seien B, B' Basen von V, C, C' Basen von W, $f : V \to W$ linear und $A_{B,C}$ die darstellende Matrix von f bezüglich B, C. Dann ist die darstellende Matrix von f bezüglich B', C' gegeben durch

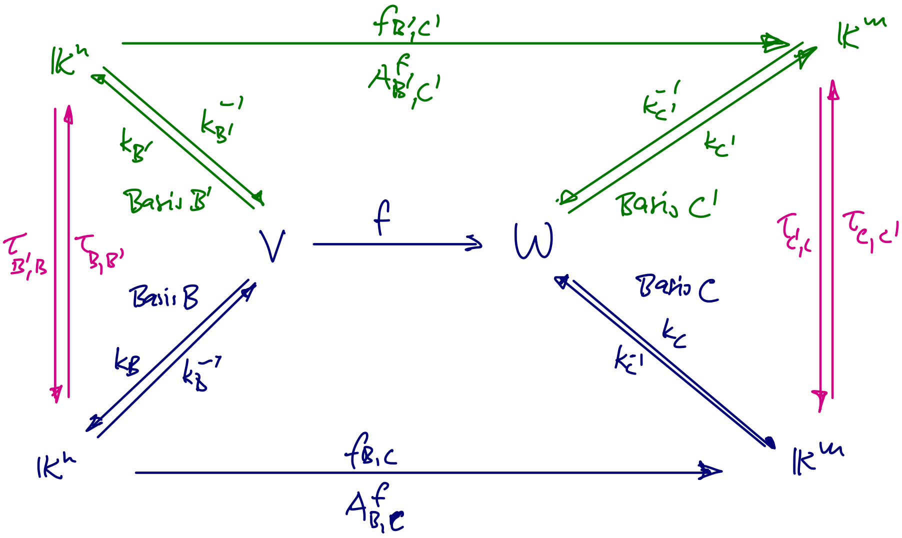
$$A_{B',C'} = T_{C,C'} \cdot A_{B,C} \cdot T_{B',B},$$

wobei $T_{B',B}$ die Transformationsmatrix der Koordinatentransformation $\tau_{B',B}$ und $T_{C,C'}$ die Transformationsmatrix der Koordinatentransformation $\tau_{C,C'}$ bezeichnet.

Beweis: Sei $v \in V$ beliebig. Wir argumentieren mit Hilfe des Diagramms: $T_{B',B}$ bildet die B'-Koordinaten $k_{B'}(v)$ von v auf B-Koordinaten von v ab, $A_{B,C}$ ordnet diesen die Koordinaten von $f(v)$ bezüglich C zu, $T_{C,C'}$ rechnet diese in Koordinaten bezüglich C' Insgesamt erhält man durch Multiplikation des dem B'-Koordinatenvektors von v mit $A_{B',C'}$ den Koordinatenvektor von $f(v)$ bezüglich C', das war gerade die Definition der Matrixabbildung $f_{B',C'} : \mathbb{K}^n \to \mathbb{K}^m$ in Definition 10.1.1.

□

Darstellende Matrizen von $f: V \to W$ linear



10.4 Aufgaben zu Kapitel 10

Aufgabe 10.56. (Matrixtransformation):

Die Matrixabbildung $f : \mathbb{R}^3 \to \mathbb{R}^4$ sei gegeben durch die Abbildungsvorschrift

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

Es seien B und C die Standardbasen des $\mathbb{R}^3$ und $\mathbb{R}^4$. Weiter seien Basen B' von $\mathbb{R}^3$ und C' von $\mathbb{R}^4$ gegeben durch

$$B' = \left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right) \quad \text{und} \quad C' = \left(\begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1/2 \\ 1/2 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -1/2 \\ 1/2 \end{pmatrix} \right).$$

(a) Bestimmen Sie die Koordinatentransformationsmatrizen $T_{B,B'}, T_{B',B}, T_{C,C'}$ und $T_{C',C}$.

(b) Bestimmen Sie die darstellenden Matrizen $A^f_{B,C}$, $A^f_{B,C'}$, $A^f_{B',C'}$ von f.

Aufgabe 10.57. (Darstellende Matrix einer Abbildung auf $P_2[x]$):

Gegeben sei der Vektorraum der reellen Polynomfunktionen zu Polynomen vom Grad höchstens zwei,

$$P_2[x] := \left\{ p : \mathbb{R} \to \mathbb{R} \mid p(x) = a_0 + a_1 x + a_2 x^2 \text{ mit } a_0, a_1, a_2 \in \mathbb{R} \right\}$$

mit der Basis $B := (1, x, x^2)$. Weiterhin sei f die lineare Abbildung, die durch

$$f : P_2[x] \to P_2[x], \qquad p \mapsto 3p + \left(-1 - 2x + x^2\right) \cdot p' + \left(2 + 3x - x^3\right) \cdot p''$$

gegeben ist (p' bzw. p'' bezeichnen hierbei wie gewöhnlich die erste bzw. zweite Ableitung der Polynomfunktion p).

(a) Prüfen Sie, dass für beliebige $p \in P_2[x]$ tatsächlich $f(p) \in P_2[x]$ gilt (d.h., dass f wohldefiniert ist) und bestimmen Sie ein Erzeugendensystem des Bildes von f.

(b) Bestimmen Sie die darstellende Matrix $A^f_{B,B}$ von f bezüglich der Basis B.

(c) Prüfen Sie anhand der Matrix $A^f_{B,B}$, ob die Abbildung f invertierbar ist.

Aufgabe 10.58. (Transformation auf Diagonalgestalt):

Gegeben sei die Basis

$$B = (b_1, b_2, b_3) = \left(\begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 2 \end{pmatrix} \right),$$

des $\mathbb{R}^3$ und die lineare Abbildung $f : \mathbb{R}^3 \to \mathbb{R}^3$, für die

$$f(b_1) = \begin{pmatrix} 2 \\ 2 \\ -2 \end{pmatrix}, \quad f(b_2) = \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}, \quad f(b_3) = \begin{pmatrix} 0 \\ -1 \\ 2 \end{pmatrix}$$

gilt.

(a) Bestimmen Sie die Abbildungsmatrix von f (also die darstellende Matrix von f bezüglich der Standardbasis; diese enthält fünf Bruchzahlen).

(b) Es sei C eine weitere Basis von $\mathbb{R}^3$, gegeben durch

$$C = (c_1, c_2, c_3) = \left(\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ -3 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 2 \end{pmatrix} \right).$$

Bestimmen Sie die darstellende Matrix $A^f_{C,C}$ von f bezüglich C. (Diese sollte mit der Überschrift der Aufgabe zusammenpassen.)

(c) Bestimmen Sie jeweils eine Basis von $\mathrm{Kern}(f)$ und von $\mathrm{Bild}(f)$.

(d) Entscheiden Sie begründet, ob f ein Automorphismus von $\mathbb{R}^3$ ist.

11 Lineare Gleichungssysteme (Teil 2) und lineare Gleichungen

Wir kehren mit unserem Wissen über lineare Abbildungen nun noch einmal zurück zum Begriff des Ranges einer Matrix und zur Theorie der Linearen Gleichungssysteme – beide Fragestellungen können wir von dieser Warte nun noch einmal klarer und umfassender beleuchten.

Grundlage für das Verständnis dieses Kapitels sind neben den Erkenntnissen der letzten beiden Kapitel die Definition des Ranges 6.3.1 einer Matrix (als Anzahl der Nichtnullzeilen in der (normierten) Zeilenstufenform der Matrix), die Folgerungen daraus für die Lösbarkeit zugehöriger linearer Gleichungssysteme 6.3.3 und die drei Sichtweisen auf Lineare Gleichungssysteme, s. Abschnitt 6.4. Werfen Sie bitte bei Bedarf noch einen auffrischenden Blick auf all diese Zusammenhänge, bevor Sie weiterlesen.

11.1 Der Rang von Matrizen (II) und linearen Abbildungen

Neben dem in Definition 6.3.1 eingeführten Rang $\operatorname{rg}(A)$ einer Matrix $A \in \mathbb{K}^{m\times n}$ gibt es noch drei weitere Rangbegriffe, die sich den drei Sichtweisen auf lineare Gleichungssysteme bzw. die zugehörigen Matrizen zuweisen lassen und deren Zusammenhänge untereinander und zum Rangbegriff aus Thema dieses Abschnitts sind.

Definition 11.1.1. *(Rang linearer Abbildungen; Spaltenrang und Zeilenrang)*

- *Sei f eine lineare Abbildung. Wir bezeichnen mit* $\operatorname{rg}(f) := \dim \operatorname{Bild}(f)$ *den Rang von f.*
- *Sei $A \in \mathbb{K}^{m\times n}$. Bezeichnen wir mit $s_1, \dots, s_n$ die Spalten von A, so nennen wir den Aufspann*

  $$\mathcal{L}(s_1, \dots, s_n) \subseteq \mathbb{K}^m$$

  der Spalten von A den Spaltenraum von A und bezeichnen die Dimension des Spaltenraumes von A, also $\dim \mathcal{L}(s_1, \dots, s_n)$*, als den Spaltenrang von A, kurz* $\operatorname{srg}(A)$*. Im Fall* $\operatorname{srg}(A) = n$ *sagt man, A habe vollen Spaltenrang.*
- *Analog definieren wir mit den Zeilenvektoren $z_1, \dots, z_m$ von A den Zeilenraum von A als*

  $$\mathcal{L}(z_1, \dots, z_m) \subseteq \mathbb{K}^n$$

  und nennen $\dim \mathcal{L}(z_1, \dots, z_m)$ *den Zeilenrang von A, kurz* $\operatorname{zrg}(A)$*. Im Fall* $\operatorname{zrg}(A) = m$ *sagt man, A habe vollen Zeilenrang.*

Wir halten zunächst in den folgenden beiden Lemmata einige Zusammenhänge zu den verschiedenen Rangbegriffen fest.

Lemma 11.1.2. *(Spaltenrang als maximale Anzahl linear unabhängiger Spalten)*
Sei $A \in \mathbb{K}^{m \times n}$. Der Spaltenrang von A gibt gerade die maximale Anzahl linear unabhängiger Spalten von A an, d.h. die größte Zahl k, zu der sich k der n Spalten $s_1, \ldots, s_n$ von A auswählen lassen, die linear unabhängig sind. Insbesondere gilt:

- *Jede Auswahl von $\operatorname{srg}(A)$ linear unabhängigen Spalten von A ist eine Basis von $\mathcal{L}(s_1, \ldots, s_n)$ und damit auch Basis des Bildes der zugehörigen Matrixabbildung. Es gilt somit $\operatorname{srg}(A) = \operatorname{rg}(f_A)$, wenn f_A die zu A gehörige Matrixabbildung bezeichnet.*
- *A hat vollen Spaltenrang genau dann, wenn die Spalten von A als Vektoren des $\mathbb{K}^m$ linear unabhängig sind.*

Analog gibt der Zeilenrang die maximal mögliche Anzahl linear unabhängiger Zeilen von A an. Voller Zeilenrang von A bedeutet also, dass die Zeilen von A, als Vektoren des $\mathbb{K}^n$ gelesen, linear unabhängig sind.

(Beweis siehe Vorlesung.)

Lemma 11.1.3. *(Erhaltung des Zeilenranges unter elementaren Zeilenumformungen)*
Sei $A \in \mathbb{K}^{m \times n}$ eine Matrix und $\hat{A} \in \mathbb{K}^{m \times n}$ eine Matrix, die aus A durch Anwendung von elementaren Zeilenumformungen, also durch Umformungen vom Typ

- *Vertauschung zweier Zeilen von A,*
- *Multiplikation einer der Zeilen von A mit $c \in \mathbb{K}$, $c \neq 0$,*
- *Ersetzung der j-ten Zeile von A, $j \in \underline{m}$, durch die Summe von Zeile i und Zeile j, $i \in \underline{n}$.*

oder Kombinationen dieser Umformungen entstanden ist. Dann ist

$$zrg(A) = zrg(\hat{A}).$$

Beweis: Bezeichne $z_1, \ldots, z_m$ die Zeilenvektoren von A. Die Vertauschung zweier Zeilen ändert nichts am Zeilenrang, da die Reihenfolge für die Anzahl linear unabhängiger Vektoren in diesem Tupel keine Rolle spielt. Ersetzt man in A eine Zeile z_i durch $z_i' = a \cdot z_i$

mit $a \neq 0$ bzw. durch die Summe von z_i und einer anderen Zeile, $z_i' = z_i + z_j$, so ist jede Auswahl von Vektoren aus diesem neuen Tupel linear unabhängig genau dann, wenn die entsprechende Auswahl von Vektoren aus $z_1, \ldots, z_m$ linear unabhängig ist (vgl. Übungsaufgabe zur linearen Unabhängigkeit). Die maximale Anzahl linear unabhängiger Zeilen von A bleibt also bei allen elementaren Zeilenumformungen erhalten, und damit nach Lemma 11.1.2 auch der Zeilenrang.

□

Der zentrale Zusammenhang zwischen all diesen Rangbegriffen ist der folgende (nämlich, dass sie alle gleich sind):

Satz 11.1.4. *(Zeilenrang = Spaltenrang = Rang = Rang von f)*
Sei $A \in \mathbb{K}^{m\times n}$ eine Matrix. Dann gilt

$$\operatorname{srg}(A) \;=\; \operatorname{zrg}(A) \;=\; \operatorname{rg}(A). \tag{11.1}$$

Zudem gilt: Sind V, W beliebige Vektorräume B und C beliebige Basen von V bzw. W, $f \in L(V,W)$ und $A_{B,C}$ die darstellende Matrix von f, so ist

$$\operatorname{rg}(f) \;=\; \operatorname{srg}(A_{B,C}) \qquad (\;=\; \operatorname{rg}(A_{B,C}) \;=\; \operatorname{zrg}(A_{B,C}))\,. \tag{11.2}$$

Beweis: Wir beginnen mit der zweiten Gleichheit, $\underline{\operatorname{rg}(f) = \operatorname{srg}(A_{B,C})}$. Die Gleichheiten in der Klammer dahinter folgen dann aus dem Beweis des ersten Teils, den wir unten beweisen.

Ist $B = (b_1, \ldots, b_n)$, so gilt nach Lemma 9.2.1(iv), dass $\operatorname{Bild}(f) = \mathcal{L}(f(b_1), \ldots, f(b_n)) =: L$ ist, also $\operatorname{rg}(f) = \dim L$. Nach Definition der darstellenden Matrix ist gerade

$$k_C(f(b_1)) = s_1, \quad \ldots, \quad k_C(f(b_n)) = s_n.$$

Die Koordinatenabbildung k_C ist ein Isomorphismus, also gilt wegen der Linearität von k_C (nach Übungsaufgabe zu linearen Abbildungen)

$$k_C(L) = \mathcal{L}(\; k_C(f(b_1)), \ldots \;, k_C(f(b_n)) \;) = \mathcal{L}(s_1, \ldots, s_n)$$

und somit

$$\operatorname{srg}(A_{B,C}) := \dim k_C(L).$$

Zum anderen ist k_C insbesondere injektiv, daher gilt nach Lemma 9.2.4(iv) $\dim k_C(L) = \dim L = \operatorname{rg}(f)$, also $\operatorname{srg}(A_{B,C}) = \operatorname{rg}(f)$ wie behauptet. Das zeigt die zweite Behauptung, zum Beweis des ersten Teils sei nun $A \in \mathbb{K}^{m\times n}$.

$\underline{\operatorname{zrg}(A) = \operatorname{rg}(A)}$: $\operatorname{zrg}(A)$ bleibt bei elementaren Zeilenumformungen unverändert und entspricht also insbesondere dem der Matrix $\hat{A}$ in normierter Zeilenstufenform, die man nach Beendigung des Gauß-Jordan-Algorithmus erhält (dieser besteht nur aus elementaren Zeilenumformungen, s. Abschnitt 6.3). Hier lässt sich der Zeilenrang, also die maximale Anzahl linear unabhängiger Zeilen aber direkt ablesen: Es ist die Anzahl Nicht-Nullzeilen von $\hat{A}$, also der Rang von A.

$\underline{\operatorname{srg}(A) = \operatorname{rg}(A)}$: Dazu betrachten wir die von A induzierte Matrixabbildung $f : x \mapsto A\cdot x$. Nach Beispiel 1 in Abschnitt 9.1.1 gilt $\operatorname{Bild}(f) = \mathcal{L}(s_1, \ldots, s_n)$ und deshalb

$$\operatorname{srg}(A) \;=\; \dim \operatorname{Bild}(f) \;=\; n - \dim \operatorname{Kern}(f). \qquad (\star)$$

Bestimmen wir den Kern von f, also die Lösungsmenge $\mathbb{L}_0$ von $A \cdot x = 0$ mit Hilfe des Gauß-Jordan-Algorithmus, so erhalten wir eine Menge $F = \{x_{i_1}, \ldots, x_{i_k}\}$ von k freien Variablen nach Satz 6.3.3; hierbei gilt $k = n - \operatorname{rg}(A)$. Setzt man jeweils für eine der freien Variablen $x_j = 1$ und für alle übrigen Null, so erhält man k linear unabhängige Vektoren $\ell_1, \ldots, \ell_k \in \mathbb{L}_0$ (denn für alle $j \in \underline{k}$ ist im Vektor ℓ_j der Eintrag in der j-ten zeile j-ten Zeile ungleich Null, in allen anderen Vektoren ℓ_m, $m \neq j$ gleich Null), und jede andere Lösung lässt sich nach Konstruktion des Algorithmus als Linearkombination dieser Vektoren schreiben. Es ist also $\mathbb{L}_0 = \mathcal{L}(\ell_1, \ldots, \ell_k)$ mit linear unabhängigen Vektoren $\ell_1, \ldots, \ell_k$ und deshalb $\dim \operatorname{Kern}(f) = k = n - \operatorname{rg}(A)$; aus $(\star)$ folgt hiermit die behauptete Gleichheit $\operatorname{srg}(A) = \operatorname{rg}(A)$.

□

Bemerkungen dazu:

- Der Rang einer Matrix lässt sich nach Lemma 11.1.3 in Kombination mit 11.1.4 also nicht nur durch die „starre“ Durchführung des Gauß-Jordan-Algorithmus ermitteln, sondern durch jede Abfolge von elementaren Zeilenumformungen, bei denen am Schluss eine Matrix steht, an der sich gemäß 11.1.2 die Anzahl der linear unabhängigen Spalten bzw. Zeilen ablesen lässt – und so bestimmt man den Rang einer Matrix für gewöhnlich in der Praxis.

- Wegen der eben bewiesenen Gleichheit der verschiedenen Rangbegriffe reden wir ab jetzt nur noch vom *Rang einer Matrix*, behalten aber im Hinterkopf, dass damit auch gleichzeitig die maximale Anzahl linear unabhängiger Zeilen und die maximale Anzahl linear unabhängiger Spalten der Matrix sowie die Dimension des Spaltenraumes von A und damit auch des Bildes der zugehörigen Matrixabbildung angegeben wird.

- Ähnliche Matrizen haben denselben Rang, denn sie stellen die gleiche Matrixabbildung dar.

Mit den oben entwickelten Rangbegrifflichkeiten lässt sich auch der Hauptsatz zur Lösbarkeit linearer Gleichungssysteme 6.3.3 griffig umformulieren:

Satz 11.1.5. *(Spalten- und Zeilenrang und Lösbarkeit von LGS)*
Sei $A \in \mathbb{K}^{m\times n}$.

- *Zu jeder rechten Seite $b \in \mathbb{K}^m$ existiert mindestens eine Lösung des LGS $A\cdot x = b$ genau dann, wenn A hat vollen Zeilenrang hat (d.h. es gilt $\mathrm{rg}(A) = m$).*

- *Zu jeder rechten Seite $b \in \mathbb{K}^m$ existiert höchstens eine Lösung des LGS $A\cdot x = b$ genau dann, wenn A hat vollen Spaltenrang hat (d.h. es gilt $\mathrm{rg}(A) = n$).*

- *Zu jeder rechten Seite $b \in \mathbb{K}^m$ existiert genau eine Lösung des LGS $A\cdot x = b$ genau dann, wenn A vollen Rang hat (d.h. es gilt $m = n = \mathrm{rg}(A)$).*

Beweis: Das ist genau die Aussage von Satz 6.3.3.

□

Die große praktische Bedeutung des Rangbegriffes rührt daher, dass sich nun viele *per se* relativ verschiedene Fragestellungen (die zu den verschiedenen Sichtweisen auf lineare Gleichungssysteme gehören) beantworten lassen, indem man für eine entsprechende Matrix A den Gauß-Jordan-Algorithmus durchführt und damit den Rang von A bestimmt. Bevor wir uns einem Beispiel dazu zuwenden, halten wir noch eine praktische Eigenschaft der Rangbestimmung per Gauß-Jordan-Algorithmus fest.

Lemma 11.1.6. *(Lokalisierung linear unabhängiger Spalten/Zeilen per Gauß-Jordan)*

Sei $A \in \mathbb{K}^{m \times n}$ und $\hat{A}$ die Matrix, die man nach Durchführung des Gauß-Jordan-Algorithmus für A (mit beliebiger rechter Seite, bzw. auch ohne) erhält.

Ist $\operatorname{rg}(A) = k$ und sind $j_1, \ldots, j_k$ die Spaltenindizes der Pivotspalten von $\hat{A}$, so findet man in den Spalten von A eine Basis des Spaltenraumes (also eine maximal linear unabhängige Auswahl der Spalten von A).

Der Gauß-Jordan-Algorithmus liefert also unter anderem Antworten auf die folgenden Fragestellungen:

- „Abbildungssicht“: Für Matrixabbildungen $f : \mathbb{R}^n \to \mathbb{R}^m, x \mapsto A \cdot x$, ist die Dimension ihres Bildes gerade $\operatorname{rg}(A)$ und nach dem Dimensionssatz daher $\dim \operatorname{Kern}(A) = n - \operatorname{rg}(A)$, das werden wir gleich unten noch benutzen. Die durch den Gauß-Algorithmus per Lemma 11.1.6 als Pivotspalten markierten Spalten von A (nicht von $\hat{A}$!) liefern eine Basis des Bildes von f.

- „Spaltensicht“: Sind $v_1, \ldots, v_k$ Vektoren des $\mathbb{K}^n$, so sind sie linear unabhängig, genau dann, wenn die zugehörige Matrix $A = (v_1 \ldots v_k) \in \mathbb{K}^{n \times k}$ vollen Spaltenrang k hat. Auch wenn nicht $\operatorname{rg}(A) = k$ gilt, entspricht der Rang der Matrix immer der maximalen Anzahl linear unabhängiger Spalten in M, und man kann dann mit Lemma 11.1.6 die Position entsprechend vieler linear unabhängiger Spalten von A berechnen, diese dann auswählen und erhält damit eine Basis des Spaltenraumes $\mathcal{L}(v_1, \ldots, v_k)$.

- „Zeilensicht“: Hier ist etwas mehr Vorsicht angebracht. Ist $\operatorname{rg}(A) = k$, so lassen sich aus den m Zeilen eines LGS k Zeilen so auswählen, dass das Weglassen der übrigen $m - k$ Zeilen eine Äquivalenzumformung ist (denn jede der anderen Zeilen ist Linearkombination der anderen, also lässt sich die Gültigkeit jeder der übrigen Gleichungen aus denen der anderen Zeilen folgern). Man erhält so einen „minimalen“ Satz an Gleichungen, die zum ursrünglichen LGS äquivalent sind. Diese k Zeilen sind aber (im Gegensatz zum Ergebnis für die Spalten, s. Lemma 11.1.6) im Allgemeinen *nicht* die (ersten k) Zeilen, in denen die Pivotelemente in $\hat{A}$ stehen, da beim Gauß-Algorithmus eventuell Zeilen vertauscht werden[1].

Beispiel dazu: S. Vorlesung.

[1] Führt man darüber Buch, so lassen sich die Pivotzeilen wieder eindeutig mit einem linear unabhängigen Satz Spalten identifizieren, aber das wollen wir hier nicht vertiefen.

11.2 LGS und verallgemeinerte lineare Gleichungen in Vektorräumen – Lösbarkeitsaussagen und Lösungsmengen

Die Untersuchung des Rangbegriffes im letzten Abschnitt zeigt, dass die Begrifflichkeiten und Strukturen der Linearen Algebra eng mit der Lösungstheorie für lineare Gleichungssysteme zusammenhängen. Zum Abschluss dieses ersten Teils des Vorlesungsskriptes werden diesem Gedanken folgend mit Hilfe der Vektorraumtheorie in diesem Abschnitt noch einige wichtige allgemeine Lösbarkeitsaussagen für lineare Gleichungssysteme und verallgemeinerte lineare Gleichungen formuliert. Wir verallgemeinern zunächst unsere Definition 6.2 für den Begriff einer linearen Gleichung.

Definition 11.2.1. *(Lineare Gleichung)*

Seien V, W $\mathbb{K}$-Vektorräume, $f: V \to W$ eine lineare Abbildung und $b \in W$. Dann heißt die Bedingung

$$f(x) = b$$

an $x \in V$ (verallgemeinerte) lineare Gleichung.

Bemerkungen dazu:

- Ist $\mathbb{K}$ ein Körper und $a_1, \ldots, a_n \in \mathbb{K}$, so ist die Funktion

$$f: \mathbb{K}^n \to \mathbb{K}, \quad f(x) = a_1 \cdot x_1 + a_2 \cdot x_2 + \ldots + a_n \cdot x_n \tag{11.3}$$

  linear. Eine Gleichung $f(x) = b$ mit dieser Funktion, die wir ja im Kontext der linearen Gleichungssysteme als lineare Gleichung bezeichnet haben, ist also auch mit der neuen Begriffsbildung weiterhin eine.

- Allgemeiner ist, da Matrixabbildungen linear sind, auch jedes lineare Gleichungssystem $A \cdot x = b$ eine lineare Gleichung für $x \in \mathbb{K}^n$ (mit der entsprechenden Matrixabbildung $f: x \mapsto A \cdot x$).

Für allgemeine lineare Gleichungen erhalten wir nun aus der Theorie für lineare Funktionen leicht die folgenden Lösbarkeitsaussagen, die wir anschließend noch einmal speziell für den Fall linearer Gleichungssysteme formulieren.

Satz 11.2.2. *(Lösbarkeit linearer Gleichungen)*

Seien V, W Vektorräume, $f : V \to W$ linear, $b \in W$ und

$$f(x) = b \tag{11.4}$$

eine gegebene (verallgemeinerte) lineare Gleichung für $x \in V$. Dann gilt:

(i) Die lineare Gleichung (11.4) ist genau dann lösbar, wenn $b \in \mathrm{Bild}(f)$ ist.

(ii) Sie ist genau dann eindeutig lösbar, wenn $b \in \mathrm{Bild}(f)$ ist und $\mathrm{Kern}(f) = \{0_V\}$ gilt.

(iii) Ist $f(x) = b$ lösbar (s. (i)), $x_0 \in V$ eine beliebige Lösung von 11.4, und $\mathbb{L}_0 = \mathrm{Kern}(f)$ die Lösungsmenge der zugehörigen homogenen linearen Gleichung $f(x) = 0$, so gilt

$$\begin{aligned} \mathbb{L} &= x_0 + \mathbb{L}_0 := \{x_0 + x \mid x \in \mathbb{L}_0\} \\ &= x_0 + \mathrm{Kern}(f) := \{x_0 + x \mid x \in \mathrm{Kern}(f)\}. \end{aligned}$$

Wenn also die Lösungsmenge der Gleichung $f(x) = b$ nicht leer ist, so ist sie immer der „um x_0 verschobene“ Kern von f, ein so genannter affiner Unterraum von V mit „Dimension“ $\dim \mathrm{Kern}(f)$.[a]

[a] Diese Beschreibung ist an dieser Stelle nur als ein Appell an Ihre mathematische Intuition zu verstehen. Genauer werden wir das in der LAAG II formulieren, insbesondere die Dimension solcher „verschobenen Unterräume“ als die Dimension des beteiligten Unterraumes selbst (hier $\dim \mathrm{Kern}(f)$) definieren.

Beweis: Der Satz fasst im Grunde genommen Resultate vorangegangener Sätze in anderen Worten zusammen. Zu (i): Es ist nach Definition

$$\mathbb{L} \;=\; \{x \in V \mid f(x) = b\}$$

$\mathbb{L}$ ist also genau dann nicht-leer ist, wenn es ein $x \in V$ gibt, für das $f(x) = b$ gilt, also (vgl. Kapitel II) wenn $b \in \mathrm{Bild}(f)$ gilt. Zu (ii): Eine gefundene Lösung einer Gleichung $f(x) = b$ ist nach 2.4.2 eindeutig, wenn f injektiv ist, das ist nach 9.2.4 genau dann der Fall, wenn $\mathrm{Kern}(f) = \{0_V\}$ gilt. Bezeichnet weiter $\mathbb{L}_0$ die Lösungsmenge der zugehörigen homogenen Gleichung $f(x) = 0_V$, so gilt nach Definition $\mathbb{L}_0 = \mathrm{Kern}\, f$, $\mathbb{L}_0$ ist also insbesondere ein

Unterraum. Sei nun $x_0 \in V$ eine spezielle Lösung von (11.4). Ist $x \in \mathbb{L}$, so ist $x - x_0 \in \mathbb{L}_0$, denn

$$f(x - x_0) = f(x) - f(x_0) = b - b = 0_V.$$

Ist umgekehrt $x \in \mathbb{L}_0$ so ist $x_0 + x \in \mathbb{L}$, denn

$$f(x_0 + x) = f(x_0) + f(x) = b + 0_V = b$$

Also ist $\mathbb{L} = x_0 + \mathbb{L}_0$.

□

Wir formulieren die Aussage des vorangegangenen Satzes wegen seiner Wichtigkeit einfach gleich nochmal für den Spezialfall linearer Gleichungssysteme und machen dabei von einigen der Zusammenhänge zwischen Matrizen und den zugehörigen Matrixabbildungen Gebrauch.

Satz 11.2.3. *(Struktur von Lösungsmengen linearer Gleichungssysteme)*

Sei $\mathbb{K}$ ein Körper, $m, n \in \mathbb{N}$ und $A \in \mathbb{K}^{m \times n}$ und $b \in \mathbb{K}^m$ gegeben. Wir betrachten das lineare Gleichungssystem „Finde alle $x \in \mathbb{K}^n$ so, dass $A \cdot x = b$ ist" und die zugehörige Matrixabbildung $f : \mathbb{K}^n \to \mathbb{K}^m, x \mapsto A \cdot x$. Mit $s_1, \dots, s_n \in \mathbb{K}^m$ seien die Spalten der Matrix A bezeichnet.

(i) Das lineare Gleichungssystem $A \cdot x = b$ ist genau dann lösbar, wenn $b \in \mathcal{L}(s_1, \dots, s_n)$ ist.

(ii) Es ist genau dann eindeutig lösbar, wenn $b \in \mathcal{L}(s_1, \dots, s_n)$ ist und $\mathrm{rg}(A) = n$ gilt.

(iii) Ist $A \cdot x = b$ lösbar, $\mathbb{L}$ die Lösungsmenge dieses Systems und $x_0 \in \mathbb{K}^n$ eine beliebige Lösung dieses Systems (d.h. $A \cdot x_0 = b$), und $\mathbb{L}_0$ die Lösungsmenge des zugehörigen homogenen linearen Gleichungssystems $A \cdot x = 0$, so ist

$$\begin{aligned} \mathbb{L} &= x_0 + \mathbb{L}_0 := \{x_0 + x \mid x \in \mathbb{L}_0\} \\ &= x_0 + \mathrm{Kern}(f) := \{x_0 + x \mid x \in \mathrm{Kern}(f)\}. \end{aligned}$$

$\mathbb{L}_0$ ist hierbei ein Untervektorraum von $\mathbb{K}^n$ mit $\dim \mathbb{L}_0 = n - \mathrm{rg}(A)$. Die Lösungsmenge $\mathbb{L}$ ist also (abhängig von b) entweder leer oder der „um x_0 verschobene" Lösungsraum $\mathbb{L}_0$, ein so genannter affiner Unterraum von $\mathbb{K}^n$ mit „Dimension" $\dim \mathrm{Kern}(f) = n - \mathrm{rg}(A)$.

Wie gibt man also Lösungsmengen zu linearen Gleichungssystemen an? Da die Lösungsmengen von homogenen linearen Gleichungssystemen Teilräume von $\mathbb{K}^n$ sind, gibt man in diesem Fall normalerweise eine Basis des Lösungsraumes an. Im allgemeinen Fall gibt man (falls vorhanden) eine spezielle Lösung an und eine Basis des Lösungsraums des zugehörigen homogenen Systems.

11.3 Drei Beispiele und praktische Verfahren

Siehe Vorlesung: Wir untersuchen mit Hilfe unserer Theorie

- ein $(4 \times 5)-$System $A \cdot x = b$ und die zugehörige Aufgabe „Finde ein Urbild von b unter A" aus der Abbildungssicht,
- einen Satz von Vektoren und ihren Aufspann im $\mathbb{K}^n$, für den eine Basis gefunden und diese zu einer Basis von $\mathbb{K}^n$ ergänzt werden soll.
- als Beispiel für ein allgemeineres lineares Problem eine bestimmte lineare Abbildung, die sogenannte Spurabbildung für reelle 3×3-Matrizen,

$$\mathrm{Spur} : \mathbb{R}^{3\times 3} \to \mathbb{R}, \qquad A \mapsto a_{1,1} + a_{2,2} + a_{3,3}.$$

Zur zweiten Aufgabe halten wir das in der Vorlesung entwickelte Verfahren hier noch abschließend fest:

Praktisches Verfahren:

Gegeben: Ein k-Tupel $M = (v_1, \ldots, v_k)$ von Vektoren aus $\mathbb{K}^n$.
Gesucht: Eine Basis B_0 von $\mathcal{L}(M)$ sowie eine Basis B von $\mathbb{K}^n$, die diese Basis B_0 enthält.

Diese Aufgabe lässt sich mit dem folgenden Algorithmus lösen:

- Stelle die $\mathbb{K}^{n\times(k+n)}$-Matrix $(M \mid I_n)$ auf, wobei $I_n \in \mathbb{K}^{n\times n}$ die Einheitsmatrix bezeichnet.
- Bringe $(M \mid I_n)$ auf Zeilenstufenform $\hat{M} \in \mathbb{K}^{n\times(k+n)}$ (z.B. mit dem Gauß-Algorithmus).
- Notiere die Pivotspaltenindizes von $\hat{M}$. Die zu diesen Indizes gehörigen Spalten der Matrix $(M \mid I_n)$ sind eine Basis B von $\mathbb{K}^n$, wobei genau die Spalten, die zu M gehören, eine Basis B_0 von $\mathcal{L}(M)$ bilden.

11.4 Aufgaben zu Kapitel 11

Aufgabe 11.59. (Ermittlung von Unterraumbasen per Gauß-Jordan):

Gegeben sind zwei Unterräume U_1 und U_2 von $\mathbb{R}^4$:

$$U_1 = \{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \in \mathbb{R}^4 \ \mid \ 2x_1+x_2+x_3 = 0, x_2-x_3+2x_4 = 0\} \ , \qquad U_2 = \mathcal{L}(\begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix}).$$

(a) Begründen Sie, dass es sich bei U_1 und U_2 um Teilräume des $\mathbb{R}^4$ handelt.

(b) Ermitteln Sie Basen der Unterräume U_1, U_2 und $U_1 \cap U_2$.

(c) Benutzen Sie die in diesem Kapitel vorgestellten Methoden, um mit Hilfe der Basen der Unterräume U_1, U_2 eine Basis von $U_1 + U_2$ zu ermitteln und letztere zu einer Basis des $\mathbb{R}^4$ zu ergänzen.

Aufgabe 11.60. (Kern, Bild und Lösungsmengen linearer Gleichungen):

Sei $\mathbb{K} = \mathbb{R}, \mathbb{C}$ oder $\mathbb{Q}$, $a_1, \ldots, a_n \in \mathbb{K}$ und $a_1 \neq 0$.

(a) Zeigen Sie, dass die Abbildung

$$f : \mathbb{K}^n \to \mathbb{K}, \qquad \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \mapsto \sum_{i=1}^{n} a_i x_i$$

linear ist.

(b) Bestimmen Sie Kern(f) und Bild(f) und die Dimensionen dieser Teilräume.

(c) Geben Sie für den Fall $n = 5$ die Menge aller $x \in \mathbb{K}^5$ an, für die $f(x) = 42$ gilt.

(d) Im Fall $\mathbb{K}^n = \mathbb{R}^2$ oder $\mathbb{R}^3$ lässt sich Kern(f) anschaulich deuten als die Menge der Vektoren aus $\mathbb{K}^n$, die auf den Vektor $(a_1, \ldots, a_n) \in \mathbb{K}^n$ senkrecht stehen; auch für größeren n bleibt man bei diesem Begriff (hat dafür aber keine Anschauung mehr). Wie lässt sich analog zu dieser Interpretation der Kern der Matrixabbildung

$$g : \mathbb{R}^3 \to \mathbb{R}^3, \qquad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ -2 & 0 & -2 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

durch „zeilenweises Lesen“ des Matrix-Vektor-Produktes deuten?

Aufgabe 11.61. (Überblick über einige wesentliche Zusammenhänge der LA):

Zum Abschluss dieses ersten Teils der LAAG folgt eine Aufgabe, die Ihnen helfen soll, die behandelten Zusammenhänge noch einmal für sich zu systematisieren.
Wir betrachten im Folgenden ein $(m \times n)$-LGS der Form

$$\begin{aligned} a_{1,1}x_1 + \ldots + a_{1,n}x_n &= b_1 \\ &\vdots \\ a_{m,1}x_1 + \ldots + a_{m,n}x_n &= b_m, \end{aligned} \tag{11.5}$$

die Matrix und die Vektoren

$$A = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix} \in \mathbb{K}^{m\times n}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{K}^n, \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in \mathbb{K}^m \tag{11.6}$$

und die lineare Abbildung

$$f\colon \mathbb{K}^n \to \mathbb{K}^m,\ x \mapsto A \cdot x. \tag{11.7}$$

Die Spalten von A bezeichnen wir mit $s_1, \ldots, s_n \in \mathbb{K}^m$.
Die unten gegebenen Begriffe und Aussagen beziehen sich jeweils auf LGS, Matrizen oder lineare Abbildungen. Ordnen Sie die Begriffe und Aussagen einer der drei Kategorien zu und sortieren Sie sie in einer Tabelle der folgenden Form derart, dass inhaltlich zusammengehörende Begriffe oder Aussagen in gleichen Zeilen stehen.

LGS (11.5)	Matrix (11.6)	lineare Abb. (11.7)
x ist Lösung von (11.5)	$A \cdot x = b$	$f(x) = b$
$\vdots$	$\vdots$	$\vdots$

Zu sortierende Begriffe und Aussagen:

- f ist surjektiv.
- $\{x \mid A \cdot x = 0\}$
- Eine Lösung von (11.5) existiert für alle $b \in \mathbb{K}^m$.
- $\mathrm{Bild}(f)$
- $\mathbb{L}$
- $\mathrm{Kern}(f)$
- A hat vollen Zeilenrang.
- A hat vollen Spaltenrang.
- $\mathcal{L}(s_1, \ldots, s_n) = \mathbb{K}^m$.
- A^{-1} existiert.
- $b \in \mathrm{Bild}(f)$.
- $\mathrm{rg}(f)$
- Für jedes $b \in \mathbb{K}^m$ existiert genau eine Lösung x von (11.5).
- f ist injektiv.
- $\mathcal{L}(s_1, \ldots, s_n)$
- die Lösungsmenge des zu (11.5) gehörigen homogenen Systems
- $\mathrm{rg}(f) = m$.
- $\mathrm{rg}(f) = n$.
- Die Spalten von A sind eine Basis des $\mathbb{K}^m$.
- f^{-1} existiert.
- $\dim(\mathrm{Bild}(f))$
- Mindestens eine Lösung von (11.5) existiert.
- f ist bijektiv.
- $\{x \mid A \cdot x = b\}$
- $\mathrm{rg}(A) = m = n$.
- Dimension des Spaltenraumes
- $\dim(\mathrm{Bild}(f)) = m$.
- A ist invertierbar.
- Die Spalten von A sind linear unabhängig.
- Menge aller $b \in \mathbb{K}^m$, für die (11.5) lösbar ist
- $\{x \in \mathbb{K}^n \mid f(x) = b\}$
- $b \in \mathcal{L}(s_1, \ldots, s_n)$.
- Lösung von (11.5) ist, wenn sie existiert, eindeutig für jedes $b \in \mathbb{K}^m$.
- $\mathrm{rg}(f) = m = n$.
- $\ker(f) = \{0\}$ mit $0 \in \mathbb{K}^n$.

Aufgabe 11.62. (Ablesen aus normierter Zeilenstufenform):

Die Durchführung des Gauß-Jordan-Algorithmus für die erweiterte Koeffizientenmatrix

$$(A \mid b) = \left(\begin{array}{rrrrrr|r} 0 & 1 & -1 & 2 & 3 & 1 & 4 \\ 3 & 2 & 1 & 0 & -1 & 2 & 0 \\ -3 & -5 & 2 & -7 & -2 & -6 & -20 \\ 0 & 1 & -1 & 4 & -1 & 4 & 14 \\ 0 & 1 & -1 & 1 & 1 & -1 & 2 \end{array}\right) \in \mathbb{R}^{5\times 7}$$

liefert die folgende Matrix in normierter Zeilenstufenform:

$$\widehat{(A \mid b)} = \left(\begin{array}{rrrrrr|r} 1 & 0 & 1 & 0 & 0 & \frac{21}{8} & \frac{1}{4} \\ 0 & 1 & -1 & 0 & 0 & -\frac{23}{8} & -\frac{3}{4} \\ 0 & 0 & 0 & 1 & 0 & \frac{7}{4} & \frac{7}{2} \\ 0 & 0 & 0 & 0 & 1 & \frac{1}{8} & -\frac{3}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Lösen Sie mit Hilfe der Matrizen $(A \mid b)$ und $\widehat{(A \mid b)}$ die folgenden Aufgaben. Geben Sie jeweils kurze, stichpunktartige Begründungen, beispielsweise geeignete Zusammenhänge aus der Vorlesung, an.

(a) Bestimmen Sie eine Basis des Bildes der zugehörigen Matrixabbildung

$$f : \mathbb{R}^6 \to \mathbb{R}^5, \quad x \mapsto A \cdot x.$$

(b) Ist f injektiv?

(c) Geben Sie die Lösungsmenge des linearen Gleichungssystems $A \cdot x = b$ an.

(d) Ist das LGS $A \cdot x = b'$ für beliebige rechte Seiten $b' \in \mathbb{R}^5$ lösbar?

(e) Lassen sich die Vektoren

$$\begin{pmatrix} 0 \\ 3 \\ -3 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 2 \\ -5 \\ 1 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} -1 \\ 1 \\ 2 \\ -1 \\ -1 \end{pmatrix}, \quad \begin{pmatrix} 2 \\ 0 \\ -7 \\ 4 \\ 1 \end{pmatrix}$$

zu einer Basis von $\mathbb{R}^5$ ergänzen?

Teil II

Lineare Algebra II und Analytische Geometrie

12 Affine Räume

Willkommen zurück! Wissen Sie noch? Wir waren in Kapitel 4 in die Vektorrechnung gestartet, indem Sie von mir genötigt wurden, sich zu erinnern, wie Sie damals in der Schule die („euklidische“) Ebene und den („euklidischen“) Raum mit Vektoren ausgestattet hatten. Vektoren können, so der Clou der ganzen Geschichte, benutzt werden, um in diese Punktemengen eine gewisse Ordnung zu bringen, indem man ein Koordinatensystem festlegt, das heißt, einen beliebigen, aber dann festen Ursprungspunkt O (das ist ein Punkt!) und drei frei gewählte linear unabhängigen Raumrichtungen $\vec{e}_1, \vec{e}_2, \vec{e}_3$ (das sind Vektoren!) definiert.

Ist das getan, können wir so nun Punkte P der Ebene/des Raumes gleichwertig über Angabe ihre Ortsvektoren $\overrightarrow{OP}$ beschreiben – und sie vor allem mit Hilfe der *Koordinaten* dieser Vektoren charakterisieren und so sinnvoll „mit Punkten rechnen“. So lassen sich dann auch im dreidimensionalen relativ leicht geometrische Objekte wie Ebenen, Kreise, Kugeln o.Ä. beschreiben. Mit Hilfe der Vektoren lassen sich dann auch weitergehend – auch das kennen Sie wahrscheinlich noch aus der Schule – Längen und Winkel geometrischer Objekte messen. Das grundlegende Vorgehen zur „Vektorisierung“ und Koordinatisierung von Punktmengen wollen wir im folgenden Kapitel axiomatisch formulieren; das führt uns zum Begriff des affinen Raumes. Ein affiner Raum besteht stets aus einer Punktmenge, auf der man mit Hilfe von Vektoren aus einem zugehörigen operiert, d.h sich anschaulich „fortbewegt“ („Vektor“ bedeutet in etwa so etwas wie „Beweger“).

12.1 „Vektorisierung“ am Beispiel der Ebene und des Anschauungsraums

Bevor wir in den axiomatischen Teil starten, skizzieren wir am Beispiel der Anschauungsebene $\mathcal{A}_2$ und dem Anschauungsraum $\mathcal{A}_3$ (euklidischer Raum/ euklidische Ebene), wie man mit den elementargeometrischen Begriffen und Axiomen, wie man sie bei Euklid und Hilbert findet, Vektoren als <u>Pfeilklassen</u> definieren kann. Diese kann dann benutzen (s. Abschnitt 12.5), um in Ebene und Raum Koordinatensysteme einzuführen. Die dargestellte anschauliche Vorgehensweise bei der Vektorisierung wird unser *paradigmatisches (etwa: „mustergebendes“) Beispiel* für die spätere abstrakte Formulierung solcher Zusammenhänge als affine Räume. Die euklidische Ebene bzw. euklidischen Raum bezeichnen wir im Folgenden kurz als $\mathcal{A}_2$ bzw. $\mathcal{A}_3$; wenn beide gleichzeitig gemeint sind, schreiben wir $\mathcal{A}$.

Für unsere Konstruktion brauchen wir einige Voraussetzungen, die in Anschauungsebene und -raum zwar augenscheinlich erfüllt sind, aber als Axiome betrachtet werden müssen, um darauf eine mathematische Theorie aufbauen zu können. Wir folgen damit zunächst der axiomatischen Beschreibung der euklidischen Geometrie in Ebene und Raum im Sinne von Euklid und Hilbert.

Voraussetzungen an die Geometrie auf $\mathcal{A}$ für die „Vektorisierbarkeit“:

- In $\mathcal{A}$ (Ebene oder Raum) haben wir Punkte $P, Q, \ldots \in \mathcal{A}$ und Geraden $g, h, \ldots \subseteq \mathcal{A}$ zur Verfügung.
- Durch zwei verschiedene Punkte $P, Q \in \mathcal{A}$ verläuft genau eine Gerade g_{PQ}.
- Wir können für je zwei Geraden entscheiden, ob sie parallel sind; jede Gerade ist parallel zu sich selbst.
- Wir können für je drei verschiedene Punkte auf einer Geraden entscheiden, welcher zwischen den zwei anderen liegt.
- Ist g_{PQ} die Gerade durch P und Q (wobei $P \neq Q$) und R liegt nicht auf g_{PQ}, so gibt es genau eine Gerade h mit $R \in h$ und so, dass h parallel zu g_{PQ} ist. („Euklidisches Parallelenaxiom“)
- Wir können Streckenlängen $|PQ|$ messen und an anderer Stelle abtragen.
- Wir können Streckenlängen $|PQ|$ ver-a-fachen (mit $a \in \mathbb{R}$, $a > 0$) und an anderer Stelle abtragen.
- In $\mathcal{A}$ ist es möglich, in der gewohnten Weise Winkel zwischen zwei Geraden zu messen, insbesondere kann man diese vergleichen.

Damit können wir in die Konstruktion der Vektorisierung und Koordinatisierung von $\mathcal{A}$ starten. Bevor wir starten, noch der folgende Disclaimer: Die folgende Darstellung soll hauptsächlich die Konstruktionsidee transportieren; daher werden einige Beweise, insbesondere zur Wohldefiniertheit der Operationen mit Vektoren und Nachweise der Vektorraumeigenschaften und Eigenschaften von Punkt-Vektor-Verknüpfungen hier eher schnell und unvollständig abgehandelt. Für eine ausführlichere Herleitung mit elementargeometrischen Beweisen der erwähnten Tatsachen siehe [Filler], S. 89ff.

12.1.1 Konstruktion von Pfeilen, Verschiebungen und Pfeilklassen

Wir definieren zunächst Pfeile in $\mathcal{A}$ per Anfang- und Endpunkt. Fassen wir alle gleichartigen Pfeile in einer Pfeilklasse zusammen, so entspricht das Resultat dem, was wir unter einem Vektor verstehen: Einer Pfeilklasse, die sich weniger mathematisch schwermütig als „nicht ortsfester Pfeil“ charaktierisieren lässt.

Lemma/Definition 12.1.1. *(Pfeile in $\mathcal{A}$)*
Seien $P, Q \in \mathcal{A}$.

(i) Das Paar (P, Q) heißt Pfeil von P nach Q.

(ii) Ist $P \neq Q$, so heißt (P, Q) echter Pfeil, ein Paar (P, P) heißt Nullpfeil.

(iii) Wir nennen zwei echte Pfeile (P, Q), (R, S), die auf parallelen Geraden liegen und für die $|PQ| = |RS|$ gilt, zusätzlich gleich orientiert

- *im Fall $(P, Q) = (R, S)$,*
- *im Fall $g_{P,Q} \neq g_{R,S}$, falls $g_{P,R} \parallel g_{Q,S}$ ist, und*
- *im Fall $g_{P,Q} = g_{R,S}$ mit $(P, Q) \neq (R, S)$, falls Q und R zwischen P und S liegen oder umgekehrt.*

(Beim Verständnis dieser Definition ist eine Zeichnung hilfreich!)

(iv) Jeder echte Pfeil (P, Q) legt folgendermaßen eine eindeutige Abbildung $p_{P,Q} : \mathcal{A} \to \mathcal{A}$ fest:

(1.) Ziehe eine parallele Gerade h zu g_{PQ} durch R.

(2.) Trage $|PQ|$ von R aus auf h ab ($\rightsquigarrow$ zwei Punkte).

(3.) Wähle denjenigen Punkt S der beiden Punkte, für den (P, Q) und (R, S) die gleiche Orientierung haben.

Hiermit ist für alle $R \in \mathcal{A}$ ein eindeutiger Bildpunkt $p_{P,Q}(R)$ festgelegt. Die Abbildung $p_{P,Q}$ lässt sich als Parallelverschiebung in $\mathcal{A}$ interpretieren.

Man sieht anhand der Konstruktion: Zwei Pfeile $(P, Q), (R, S)$ beschreiben dieselbe Verschiebung, wenn sie

- auf zueinander parallelen Geraden liegen (d.h. $g_{P,Q} \parallel g_{R,S}$),
- die Streckenlängen $|PQ|$ und $|RS|$ übereinstimmen und
- sie gleich orientiert sind.

Solche Pfeile nennen wir parallelgleich.

Wir fassen nun alle parallelgleichen Pfeile in je einer Äquivalenzklasse zusammen.

Lemma/Definition 12.1.2. *(Pfeilklassen definieren Vektoren auf $\mathcal{A}$)*

(i) *Parallelgleichheit ist eine Äquivalenzrelation auf der Menge aller echten Pfeile. Die Äquivalenzklasse eines echten Pfeiles (P, Q) ist die Pfeilklasse oder der Vektor*

$$\overrightarrow{PQ} := [(P, Q)] = \{(R, S) \mid (R, S) \textit{ ist parallelgleich zu } (P, Q)\}.$$

Zusätzlich definieren wir den Nullvektor als zusammenfassende Klasse aller Nullpfeile

$$\vec{0} := \overrightarrow{PP} := \{(P, P) \mid P \textit{ Punkt von } \mathcal{A}\}.$$

(ii) *Zu jedem Punkt R und jedem Pfeil (P, Q) (mit $P \neq Q$) existiert genau ein Punkt S und damit ein Pfeil (R, S) so, dass $\overrightarrow{PQ} = \overrightarrow{RS}$.*

(iii) *Die Vektoren von $\mathcal{A}$ definieren wir als die Menge*

$$\mathbb{V} := \{\overrightarrow{PQ} \mid P, Q \textit{ Punkte aus } \mathcal{A}\}.$$

Ihre Elemente schreiben wir in Anlehnung an die konstituierenden Pfeilklassen mit Vektorpfeil: $\vec{v} \in \mathbb{V}$.

Für den Beweis, dass es sich hierbei um eine Äquivalenzrelation handelt (d.h., dass die Definition von $\mathbb{V}$ legitim ist), verweise ich die interessierte Leserin auf [Filler].

Klar sollte aber werden: Die mathematisch etwas umständlich zu beschreibenden Pfeilklassen $\overrightarrow{PQ}$ sind (im Gegensatz zu dem, was man in einigen Schulbüchern vorfindet) eine mathematisch saubere Beschreibung dessen, was man sich unter den anschaulichen Vektoren der Ebene bzw. des Raumes vorstellt. Insbesondere sind diese Vektoren entgegen einigen landläufigen Darstellungen keine Pfeile (mit festem Anfangs- und Endpunkt), sondern Pfeilklassen: Eine Pfeilklasse ist somit durch viele Pfeile darstellbar, genauer gesagt, durch *jeden* zu (P, Q) parallelgleichen Pfeil repräsentierbar.Insbesondere bedeutet die Gleichheit $\overrightarrow{PQ} = \overrightarrow{RS}$, dass die Äquivalenzklasse der Pfeile (P, Q) und (R, S) dieselbe ist, also (P, Q) und (R, S) parallelgleiche Pfeile sind und dieselbe Verschiebung $p_{P,Q} = p_{R,S}$ in $\mathcal{A}$ beschreiben.

Ist P aus $\mathcal{A}$ und $\vec{v} \in \mathbb{V}$ eine Vektor auf $\mathcal{A}$, so gibt es genau einen Punkt $Q \in \mathcal{A}$ so, dass der Pfeil (P, Q) ein Repräsentant von $\vec{v}$ gehört (nämlich anschaulich genau das Resultat der Verschiebung von P um $\vec{v}$). Nutzen wir für diese eindeutige Zuordnung die Schreibweise $P + \vec{v} := Q$, so definiert das insgesamt eine Abbildung

$$+ : \quad \mathcal{A} \times \mathbb{V} \ \rightarrow \ \mathcal{A}, \qquad (P, \vec{v}) \mapsto P + \vec{v}.$$

Die Addition + (eine Summe von Punkt und Vektor) verknüpft Punkte und Vektoren, Resultat ist jeweils das Bild von P unter der zu $\vec{v}$ gehörigen Verschiebung. Das ist das Grundkonzept der affinen Räume: Wir operieren mit Hilfe von Vektoren (später eines allgemeinen Vektorraumes) auf Punktmengen (später allgemeinen Menge), das Ergebnis sind auch im allgemeinen Setting wieder Punkte.

12.1.2 Die Vektorraumstruktur der Vektoren von $\mathcal{A}$

Die Vektoren $\mathbb{V}$ verhalten sich natürlich, wie man es gewohnt ist. Insbesondere bilden sie als Struktur in sich einen Vektorraum $(\mathbb{V}, +, \cdot)$, der die axiomatischen Eigenschaften aus Kapitel 8 erfüllt. Das zu beweisen, ist nicht besonders schwer und erfordert nur die Anwendung einiger elementargeometrischer Kenntnisse, es ist aber eine recth längliche Angelegenheit, daher wollen wir hier darauf verzichten und nur demonstrieren, wie sich Addition und skalare Multiplikation mit Hilfe der geometrischen Eigenschaften der Pfeile und Pfeilklassne definieren lassen.

- *Addition:* Seien $\vec{v}, \vec{w} \in \mathbb{V}$. Um $\vec{v} + \vec{w}$ zu definieren, wählen wir einen Punkt P der Ebene. Dann gibt genau einen Punkt Q mit $\vec{v} = \overrightarrow{PQ}$. Zu $\vec{w}$ existiert nach Lemma 12.1.2 genauso ein Punkt R mit $\vec{v} = \overrightarrow{QR}$. Wir setzen

  $$\vec{v} + \vec{w} := \overrightarrow{QR}.$$

  Da der Punkt P und somit der Repräsentantenpfeil (P, Q) beliebig gewählt wurde, ist noch zu zeigen, dass sich für alle Repräsentantenpfeile dasselbe Ergebnis für $\vec{v}+\vec{w}$ ergibt, siehe [Filler], S. 92.

- *Skalare Multiplikation:* Sei $\vec{v} \in \mathbb{V}$ und $a \in \mathbb{R}$. Um einen Vektor $a \cdot \vec{v} \in \mathbb{V}$ zu definieren, wählen wir wieder einen Punkt P der Ebene und ein zugehöriges Q mit $\vec{v} = \overrightarrow{PQ}$. Wir tragen auf der Gerade durch P und Q von P aus das $|a|$-fache der Streckenlänge $|PQ|$ ab (das soll in der Geometrie auf $\mathcal{A}$ möglich sein, vgl. die Forderungen am Anfang des Abschnitts). Wir erhalten so zwei Punkte R, R' auf der Geraden, wobei einer der Pfeile (o.B.d.A. sei das (P, R)) mit (P, Q) gleich orientiert ist, der andere (o.B.d.A. sei das (P, R')) nicht. Wir setzen nun

$$a \cdot \vec{v} := \overrightarrow{PR} \quad \text{falls } a > 0, \quad a \cdot \vec{v} := \overrightarrow{PR'} \quad \text{falls } a < 0, \quad 0 \cdot \vec{v} := \vec{0}.$$

Man prüft wieder, dass die so definierte skalare Multiplikation wohldefiniert ist.

Das so definierte Tripel $(\mathbb{V}, +, \cdot)$ erfüllt die Vektorraumaxiome, ist also (Überraschung!) ein Vektorraum. Den Beweis hierzu führt man wiederum elementargeometrisch anhand von Repräsentanten, also anhand von Pfeilen in der Ebene oder im Raum, s. wieder [Filler]. Wir nutzen die Gelegenheit, die Vektorraumaxiome aus 8.1.1 hier zur Erinnerung noch einmal für $\mathbb{V}$ aufzulisten:

(VA) $(\mathbb{V}, +)$ ist eine abelsche Gruppe. Genauer gilt für alle $\vec{v}, \vec{w}, \vec{z} \in \mathbb{V}$

- (i) $(\vec{v} + \vec{w}) + \vec{z} = \vec{v} + (\vec{w} + \vec{z})$,
- (ii) $\vec{0}$ ist neutrales Element von $\mathbb{V}$, also $\vec{0} + \vec{v} = \vec{v} + \vec{0} = \vec{v}$,
- (iii) $(-1) \cdot \vec{v}$ ist inverses Element von $\vec{v}$,
- (iv) $\vec{v} + \vec{w} = \vec{w} + \vec{v}$.

(VS) Die skalare Multiplikation ist assoziativ und normiert: Für alle $a, b \in \mathbb{R}$ und alle $\vec{v} \in \mathbb{V}$ gilt

$$(a \cdot b) \cdot \vec{v} = a \cdot (b \cdot \vec{v}) \qquad 1 \cdot \vec{v} = \vec{v}.$$

(VAS) Es gelten die Distributivgesetze für Vektorräume: Für alle $a, b \in \mathbb{R}$ und alle $\vec{v}, \vec{w} \in \mathbb{V}$ gilt

$$(a + b) \cdot \vec{v} = a \cdot \vec{v} + b \cdot \vec{v} \quad \text{und} \quad a \cdot (\vec{v} + \vec{w}) = a \cdot \vec{v} + a \cdot \vec{w}$$

Operiert man auf den Punkten von $\mathcal{A}$, so lassen sich ebenfalls Gesetzmäßigkeiten feststellen. Es gilt für die Punkte-Vektor-Addition $+ : \mathcal{A} \times \mathbb{V} \to \mathcal{A}$ beispielsweise

(A1) $P + \vec{0} = P$ für alle $P \in \mathcal{A}$.

(A2) Für alle $P, Q \in \mathcal{A}$ existiert genau ein Vektor $\vec{v} \in \mathbb{V}$, für den $Q = P + \vec{v}$ gilt.

(A3) $P + (\vec{v} + \vec{w}) = (P + \vec{v}) + \vec{w}$ für alle $P \in \mathcal{A}$ und $\vec{v}, \vec{w} \in \mathbb{V}$.

Das sind genau die drei Eigenschaften, die wir fordern werden, wenn wir im kommenden Abschnitt axiomatisch definieren, wann wir einen Pärchen $(\mathcal{A}, \mathbb{V})$ aus Punkten und darauf operierendem Vektorraum als affinen Raum bezeichnen wollen.

12.2 Affine Räume

Definition 12.2.1. *(Affiner Raum, Dimension affiner Räume)*

Sei $\mathbb{K} = \mathbb{R}, \mathbb{C}$ *oder* $\mathbb{Q}$, V *ein* $\mathbb{K}$*-Vektorraum,* A *eine nichtleere Menge und* $+ \colon A \times V \to A$ *eine Verknüpfung. Das Tripel* $(A, V, +)$ *heißt affiner Raum über dem Vektorraum* V*, falls gilt:*

(A1) $P + 0_V = P$ *für alle* $P \in A$.

(A2) Für alle $P, Q \in A$ *existiert genau ein Vektor* $v \in V$*, für den* $Q = P + v$ *gilt.*

(A3) $P + (v + w) = (P + v) + w$ *für alle* $P \in A$ *und* $v, w \in V$.

Wir werden oft kürzer auch nur von A_V *oder gar von* A *als affinem Raum sprechen, wenn klar ist, was der beteiligte Vektorraum* V *ist. Die Elemente von* A *werden als Punkte bezeichnet. Als Dimension eines affinen Raumes* A *definieren wir*

$$\dim A := \dim V.$$

Bemerkungen dazu:

- Seien $P, Q \in A$ zwei Punkte. Für den nach (A2) eindeutigen Vektor v mit $Q = P + v$ schreiben wir auch $\overrightarrow{PQ}$. Nach (A1) ist dann insbesondere $\overrightarrow{PP} = 0_V$ für alle $P \in A$.
- Die Verknüpfung + liefert also durch Addition von Punkten und Vektoren wieder Punkte. In (A3) ist allerdings etwas Vorsicht angebracht: Das erste + entspricht der Addition + von Punkt und Vektor, das zweite der Addition zweier Vektoren in V. Wie hier wird meist zwischen den Verknüpfungssymbolen nicht unterschieden, da durch die beteiligten Punkte/Vektoren immer klar sein sollte, was gemeint ist.

Beispiele:

- Die im vorigen Kapitel behandelte Ebene bzw. der Anschauungsraum erfüllen mit den Verschiebungen $\mathbb{V}$ als $\mathbb{R}$-Vektorraum V und der dort betrachteten Punktverschiebung + die Axiome (A1)-(A3). Warum die Sache trotzdem schwierig ist, sehen wir weiter unten.
- Mit der Wahl $A = \mathbb{K}^n$ und $V = \mathbb{K}^n$ und der üblichen Vektoraddition + als Verknüpfung ist $(\mathbb{K}^n, \mathbb{K}^n, +)$ ein affiner Punktraum „über sich selbst“; die Axiome des affinen Raumes prüft man hierfür sehr leicht nach.
- Allgemein ist jeder Vektorraum V ein affiner Raum über sich selbst.

Satz/Definition 12.2.2. *(Darstellung affiner Räume)*

Sei $(A, V, +)$ ein affiner Raum und $P \in A$ ein beliebiger Punkt.

(i) Die Abbildung

$$ort_P : A \to V, \qquad Q \mapsto \overrightarrow{PQ},$$

ist eine Bijektion. Den zu $Q \in A$ gehörigen Vektor $ort_P(Q) = \overrightarrow{PQ}$ nennt man den Ortsvektor von Q bezüglich P.

(ii) Es gilt

$$A \quad = \quad P + V \quad := \quad \{P + v \mid v \in V\}. \tag{12.1}$$

Insbesondere kann man sich den Raum A als „ausgehend von einem beliebigen Punkt $P \in A$ durch „Anheften“ aller Vektoren aus V (also aller Ortsvektoren bzgl. P)“ vorstellen.

Man beachte, dass die Wahl des Punktes P für die Darstellung 12.1 vollkommen beliebig ist. Man findet daher gelegentlich den Ausspruch, ein affiner Raum sei „ein Vektorraum, der vergessen hat, wo sein Nullvektor ist“.

Beweis: Zu (i): Wir zeigen, dass die Abbildung ort_P injektiv und surjektiv ist. Für die Injektivität seien $R, S \in A$ mit $\overrightarrow{PR} = \overrightarrow{PS}$, dann ist $R = P + \overrightarrow{PR} = P + \overrightarrow{PS} = S$, also $R = S$. Das zeigt die Kontraposition der Injektivitätsbedingung $x \neq y \Rightarrow f(x) \neq f(y)$. Für die Surjektivität sei $v \in V$, dann existiert, da $+$ nach V abbildet, ein Punkt $R \in A$ mit $P + v = R$, also ist $v = \overrightarrow{PR} = ort_P(R)$.

Zu (ii): Für $R \in A$ ist $R = P + \overrightarrow{PR}$, also $A \subseteq P + V$. Umgekehrt ist $P + v \in A$ ist für jedes $v \in V$, da $+$ nach A abbildet. □

Bemerkungen dazu:

- Es ist nützlich, sich zu vergegenwärtigen, was die bisherigen Definitionen und der obige Satz 12.2.2(ii) im Fall $A = V$, also z.B. für $A = V = \mathbb{R}^3$ aussagen:

  Für den Fall $A = V$ bedeutet der obige Satz, dass wir den Vektorraum V für alle $v \in V$ als

  $$V = v + V = \{v + x \mid x \in V\}$$

  schreiben können. Ist $v \in V$, so sind die Ortsvektoren bezüglich v sind ebenfalls Vektoren aus V: Es ist $ort_v(y) = y - v$, denn das ist der eindeutig bestimmte Vektor mit $v + (y - v) = y$.

- Für $A = V = \mathbb{R}^3$ sind Punkte des affinen Raumes die Elemente des $\mathbb{R}^3$, also die „normalen" Vektoren $x = (x_1, x_2, x_3)$ aus dem $\mathbb{R}^3$ (bzw. V); diese könnte man auch mit Großbuchstaben bezeichnen, wie wir das bei Punkten gemacht haben. Die Ortsvektoren sind ebenfalls Vektoren aus dem $\mathbb{R}^3$. Wählt man als Referenzpunkt $x \in \mathbb{R}^3$, so ist wieder $ort_x(y) = y - x$ mit der gewöhnlichen Differenz von Vektoren des $\mathbb{R}^3$.

Auf die folgenden Rechenregeln und Zusammenhänge werden wir in diesem Kapitel öfter zurückgreifen.

Lemma 12.2.3. *(Rechenregeln und Zusammenhänge in affinen Räumen)*

Sei V ein $\mathbb{K}$-Vektorraum und sei A ein affiner Raum über V.
Seien $P, Q, R, S \in A$ und $v, w \in V$, $a \in \mathbb{K}$. Dann gelten folgende Aussagen:

(i) $P + v = P + w \Longrightarrow v = w$,

(ii) $P + v = Q + v \Longrightarrow P = Q$,

(iii) $P + v = Q \Longrightarrow P = Q + (-v)$,

(iv) *Sind $P, Q \in A$, so gilt* $\{P + a \cdot \overrightarrow{PQ} \mid a \in \mathbb{K}\} \subseteq A$,

(v) *Es ist* $\overrightarrow{PQ} + \overrightarrow{QR} = \overrightarrow{PR}$,

(vi) $\overrightarrow{PQ} + \overrightarrow{QP} = 0_V$, *insbesondere* $\overrightarrow{PQ} = -\overrightarrow{QP}$,

(vii) $\overrightarrow{PQ} = \overrightarrow{RS} \Longrightarrow \overrightarrow{PR} = \overrightarrow{QS}$.

Allgemeiner gilt für Punkte $P_1, \ldots, P_n \in A$ (mit $n \geq 2$) stets

(viii) $\overrightarrow{P_1P_2} + \overrightarrow{P_2P_3} + \ldots + \overrightarrow{P_{n-1}P_n} = \sum_{k=1}^{n-1} \overrightarrow{P_kP_{k+1}} = \overrightarrow{P_1P_n}$,

und für beliebiges $P \in A$ und jede beliebige Teilmenge $M \subseteq A$ auch

(ix) $P + \mathcal{L}(ort_P(M)) \subseteq A$.

In (ix) wurden die (hoffentlich!) suggestiven Schreibweisen

$$ort_P(M) = \{ort_P(Q) \mid Q \in M\}$$

und

$$P + \mathcal{L}(ort_P(M)) := \{P + v \mid v \in \mathcal{L}(ort_P(M))\}$$

verwendet.

Beweis: Die Beziehungen folgen aus den Axiomen (A1)-(A3) und Satz 12.2.2, insbesondere der Eindeutigkeit des Verbindungsvektors $\overrightarrow{PQ}$. Ein Teil des Beweises ist Übungsaufgabe, wir zeigen exemplarisch (i), (iii) und (ix). (i): S. Vorlesung. Zu (iii): Nach Voraussetzung ist $P + v = Q$, also

$$Q + (-v) \;=\; (P + v) + (-v) \overset{(A3)}{=} P + (v + (-v)) = P + 0_V \overset{(A3)}{=} P.$$

Zu (ix): Für beliebiges $Q \in M$ folgt wegen $P, Q \in A$ nach (A2) $\overrightarrow{PQ} \in V$. Da V ein Vektorraum ist, haben wir (nach Teilraumkriterium) auch $\mathcal{L}((\mathrm{ort}_P(M)) \mid Q \in M\}) \subseteq V$, also $P + \mathcal{L}((\mathrm{ort}_P(M)) \subseteq A$ nach Definition der Abbildung $+$.

□

Definition 12.2.4. *(Affine Teilräume)*
Sei A_V ein affiner Raum. Eine Teilmenge $N \subseteq A$ nennt man einen affinen Teilraum von A_V, wenn es einen Punkt $P \in A$ und einen Unterraum $U \subseteq V$ gibt, so dass $N = P + U$ ist.

Lemma 12.2.5. *(Charakterisierung von affinen Teilräumen, deren Gleichheit)*
Sei $(A, V, +)$ ein affiner Raum über einem Vektorraum V und $N \subseteq A$.

(i) Es sind äquivalent:

- *N ist ein affiner Teilraum von A_V.*
- *Es gibt einen Unterraum $U \subseteq V$ so, dass $(N, U, +)$ (mit der Addition $+$ aus $(A, V, +)$) selbst ein affiner Raum ist.*

(ii) Sind $N = P + U$, $M = Q + W$ zwei affine Teilräume von A mit Unterräumen U und W von V, so ist $N = M$ genau dann, wenn $P \in M$ und $U = W$ gilt.

Beweis: (i) Ist $N = P + U$, so gelten (A1) und (A3), da diese für alle Punkte $P \in A$ und alle Vektoren $v \in V$ gelten, also insbesondere für $P \in N$ und alle Vektoren $v \in U$. (Forderung (A1) macht insbesondere Sinn, da U ein Unterraum ist, also $0_V \in U$ ist.) Für (A2) seien $Q_1 = P + u_1, Q_2 = P + u_2 \in A$ mit $u_1, u_2 \in U$. Mit den Rechengesetzen in A_V ergibt sich $Q_1 = P + (u_2 - u_1)$ mit $u_2 - u_1 \in U$, und diese Darstellung ist eindeutig, da sie sonst in A_V nicht eindeutig wäre im Widerspruch zu (A2) für A_V). Wir müssen also nur prüfen, dass $+$ als Abbildung $+ : N \times U \to N$ wohldefiniert ist, d.h. Wertevorrat N hat. Seien dazu $Q \in N$ und $u \in U$; wir zeigen $Q + u \in N$: Es ist $Q = P + \overrightarrow{PQ}$ mit $\overrightarrow{PQ} \in U$, also

$$Q + u \;=\; (P + \overrightarrow{PQ}) + u \overset{(A3)\text{ in }V}{=} P + \underbrace{(\overrightarrow{PQ} + u)}_{\in U} \;\in\; P + U \;=\; N$$

und damit N ein affiner Raum.

Ist umgekehrt $(N, U, +)$ ein affiner Raum, so ist N nicht leer (vgl. die Def. affiner Räume). Sei also $P \in N$, dann hat N nach 12.1 die Darstellung $N = P + U$.

(ii) Ist $P \in M$ und $U = W$, so gilt nach Satz 12.1

$$M = P + W = P + U = N.$$

Umgekehrt ist klar, dass im Fall $M = N$ für alle $P \in N$ auch $P \in M$ gilt. Die Abbildung $\mathrm{ort}_P : A \to V$ ist bijektiv, daher ist

$$U \overset{\text{bij.}}{=} \mathrm{ort}_P(N) = \{\overrightarrow{PR} \mid R \in N\} \overset{N=M}{=} \{\overrightarrow{PR} \mid R \in M\} \overset{\text{bij.}}{=} \mathrm{ort}_P(M) = W.$$

□

Definition 12.2.6. *(Punkte, Geraden, Ebenen und Hyperebenen affiner Räume)*

Nach der Definition oben haben die affinen Teilräume N eines affinen Raums A_V gerade die Form $N = P + U$ mit einem Aufpunkt $P \in A$ und einem Unterraum $U \subseteq V$. Nach Satz 12.2.5 ist dann $(N, U, +)$ selbst ein affiner Raum und insbesondere die Dimension von N definiert als $\dim(N) := \dim(U)$.

(0) Die nulldimensionalen Teilräume von A_V haben gerade die Form

$$N = P + \{0_V\} := \{P\}$$

– sie entsprechen eins zu eins gerade den <u>Punkten von A</u> (wie in Def. 12.2.1).

(1) Die eindimensionalen Teilräume von A_V sind gerade die, bei denen U eindimensional ist. Ist $\{v\}$ eine Basis von U (d.h. insbesondere $v \neq 0$), so haben diese gerade die Form

$$N \;=\; P + \mathcal{L}(v) \;=\; \{Q \in A \mid Q = P + a \cdot v, a \in \mathbb{K}\}.$$

Wir nennen diese die <u>Geraden von A</u>.

(2) Ist $\dim N = 2$ *und damit* $\dim U = 2$, *$U = \mathcal{L}(v_1, v_2)$ mit l.u. $v_1, v_2 \in V$, so ist*

$$N \;=\; P + \mathcal{L}(v_1, v_2) \;=\; \{Q \in A \mid Q = P + a_1 \cdot v_1 + a_2 \cdot v_2, a_1, a_2 \in \mathbb{K}\}.$$

Wir nennen solche Teilräume die <u>Ebenen von A</u>.

(3) *Einen affinen Teilraum N von A_V mit $\dim N = \dim V - 1$ nennt man affine Hyperebene von A_V. Die Hyperebenen des $\mathbb{R}^2$ sind also die Geraden des $\mathbb{R}^2$, die Hyperebenen des $\mathbb{R}^3$ die Ebenen des $\mathbb{R}^3$.*

Bemerkungen dazu:

- Hiermit haben wir nun die aus der Schule bekannten Geraden und Ebenen des Anschauungsraumes und des $\mathbb{R}^3$ in die Theorie der affinen Teilräume eingebettet. Im $\mathbb{R}^n$ lassen sich für $n > 3$ nun analog affine Unterräume aller möglichen Dimensionen $d \in \{0, \dots, n\}$ betrachten.

- Da affine Teilräume nach 12.2.5 selbst affine Räume sind, zeigt Satz 12.2.2 eine hoffentlich aus der Schule bekannte Eigenheit in der Darstellung von affinen Unterräumen $N = P + U$ wie Geraden oder Ebenen: Ist $Q \in N$ ein beliebiger Punkt und B eine Basis von U, so ist nach (12.1) auch $N = Q + \mathcal{L}(B)$. *Jede* Kombination eines Punktes $Q \in N$ und einer Basis des „Richtungsraumes“ U liefert also eine mögliche Darstellung von N.

- Nach Lemma 12.2.5 sind zwei affine Teilräume $N = P + U$, $M = Q + W$ identisch, falls $P \in M$ und $U = W$ gilt; auch das ist ein in der Schulmathematik beliebtes Kriterium, um die Gleichheit von Geraden und Ebenen zu untersuchen.

Ein weiteres wichtiges Beispiel für affine Unterräume ist uns schon in der LAAG I begegnet.

Beispiel: Seien V, W Vektorräume, $f : V \to W$ ist linear und $\mathbb{L}$ die Lösungsmenge einer linearen Gleichung $f(x) = b$ mit $b \in W$. $\mathbb{L}_0$ sei wie immer die Lösungsmenge der zugehörigen homogenen Gleichung $f(x) = 0$, also der Kern von f. Wir haben am Schluss der LAAG I gesehen, dass folgende Beziehung gilt: Ist $x^* \in \mathbb{L}$ eine beliebige spezielle Lösung für $f(x) = b$, so ist

$$\mathbb{L} = x^* + \mathbb{L}_0$$

Zudem ist (s. LAAG I) $\mathbb{L}_0 = \mathrm{Kern}(f)$ ein UVR von V und somit selbst ein Vektorraum. Betrachten wir V wie oben als affinen Raum über sich selbst, so ist $\mathbb{L}$ nach Definition 12.2.4 ein affiner Unterraum von V und

$$\dim(\mathbb{L}) = \dim(\mathbb{L}_0) = n - \mathrm{rg}(f)$$

seine Dimension als affiner Raum.

12.2.1 Exkurs: Das Problem mit Anschauungsraum und -ebene

Der Anschauungsraum $\mathcal{A}_3$ und die Zeichenebene $\mathcal{A}_2$ erfüllen Axiome der Geometrie (u.a. die oben) und die Forderungen (A1)-(A3) für affine Räume, aber:

- (als affiner Raum:) Was genau *ist* $\mathcal{A}$? „Die Zeichenebene" bzw. „der Anschauungsraum" ist keine wirkliche Definition einer Menge $\mathcal{A}$.
- (als affiner Raum und als Geometrie:) Was sind hier eigentlich Punkte? Es sollten im anerkannten mathematischen Setting *Elemente* einer *Menge* $\mathcal{A}$ sein. Die Definition von $\mathcal{A}$ als Menge ist aber wie gesagt unklar, und zur anschauungsbasierten Definition von Punkten in der klassischen euklidischen Zeichenebene hat sich von Euklid bis heute relativ wenig getan, so dass wir nur aus heutiger Sicht relativ schwurbelige Festlegungen wie „Ein Punkt ist, was keine Teile hat", „Eine Gerade ist eine Länge ohne Breite" wie bei Euklid zur Verfügung haben.
- (als Geometrie:) Außerdem steckt in den Axiomen für unsere Geometrie aus Abschnitt 12.1 schon eine Abstandsmessung (in der Länge von Pfeilen), von der wir ebenfalls nicht genau wissen, wie wir diese eigentlich für unsortierte Punkte ohne Koordinaten definieren sollen.

Möglichkeit 1, dieser Falle zu entkommen: Abstraktion von der Vorstellung der Zeichenebene $\rightsquigarrow$ Elementargeometrie in heutigen Sinne, rein axiomatisch nach David Hilbert.

Folgt man diesem Weg, so sagt man nicht mehr, was Punkte, Geraden und Ebenen eigentlich sind, sondern fordert nur die Gültigkeit bestimmter Relationen zwischen ihnen, die zwar in der Anschauung fundiert sind, aber rein formal formuliert sind (ähnlich wie unsere Voraussetzungen oben). Mit Hilfe elementarer Relationen zwischen Gegenständen der Geometrie lässt sich dann eine große Zahl bekannter geometrische Sätze beweisen, ohne dass konkret gesagt wird, was die Menge der Punkte der Geometrie überhaupt sein soll (vgl. Gruppentheorie, Vektorraumtheorie: Ergebnisse über nicht konkret benannte Gruppen, Vektorräume, jetzt: Ergebnisse über nicht konkret benannte Geometrien). Falls Sie einen Eindruck vom Hilbertschen Axiomensystem und Werk gewinnen wollen, hier: (Link)

Möglichkeit 2, die wir hier weiter verfolgen: Wir haben oben gesehen, dass der $\mathbb{R}^n$ ist affiner Raum über sich selbst ist. Führen wir noch in Skalarprodukt ein, mit dem sich Winkel, Längen und Abstände messen lassen (vgl. Kap. XI) so lässt sich die euklidische Geometrie der Ebene und des Raumes so vollständig in $\mathbb{R}^2$ bzw. $\mathbb{R}^3$ abbilden, und man kann zeigen, dass der $\mathbb{R}^2$ und der $\mathbb{R}^3$ die Axiome der euklidischen Geometrie nach Hilbert (also der „normalen“ Geometrie in Ebene und Raum) mit $x \in \mathbb{R}^n$ als Punkten und mit seinen eindimensionalen affinen Unterräumen als Geraden erfüllt. Man sagt, der $\mathbb{R}^2$ bzw. $\mathbb{R}^3$ ist ein *Modell* für die euklidische Geometrie in der Ebene bzw. im Raum.

Der $\mathbb{R}^n$ liefert also für $\mathbb{R}^2$ bzw. $\mathbb{R}^3$ eine zwei- bzw. dreidimensionale euklidische Geometrie, ist ein affiner Raum *und* ist im Sinne der anerkannten Zermelo-Frankel-Axiome (vgl. Anhang) als Menge konstruierbar („axiomatischer Aufbau der Zahlbereiche“).

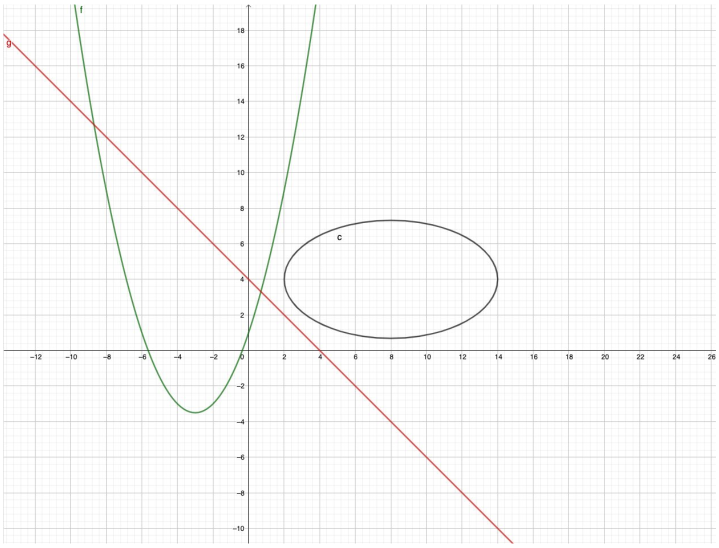
Die Analytische Geometrie treibt also mit Hilfe der Linearen Algebra Euklidische Geometrie mathematisch präzise im Zahlentupelraum $\mathbb{R}^n$ (insbesondere in $\mathbb{R}^2$ und $\mathbb{R}^3$).

Bemerkenswert ist hierbei: Die Geometrie im $\mathbb{R}^n$ braucht für ihre Gültigkeit keine Anschauung mehr, sie ist rein axiomatisch auf Grundlage der Mengenlehre definiert. Eine räumlich-geometrische Vorstellung hinter den Operationen im $\mathbb{R}^n$ ist trotzdem schön und hilfreich für das Verständnis und die Ideenfindung – diese überzeugt einen dann auch davon, dass unsere Rechnungen immer noch in guter Übereinstimmung mit unserer Wahrnehmung der Wirklichkeit sind, z.B. wenn man mit Hilfe der hier entwickelten Werkzeuge Computergrafik erzeugt. Zwei Zitate großer Denker fassen diesen Spagat zwischen Abstraktion und Realität dabei vielleicht ganz gut zusammen:

„Insofern sich die Sätze der Mathematik auf die Wirklichkeit beziehen, sind sie nicht sicher, und insofern sie sicher sind, beziehen sie sich nicht auf die Wirklichkeit.“ (Albert Einstein)

„So fängt denn alle menschliche Erkenntnis mit Anschauung an, geht von da zu Begriffen und endigt mit Ideen.“ (Immanuel Kant: Kritik der reinen Vernunft, Elementarlehre)

Die Hauptidee der analytischen Geometrie



As long as Algebra and Geometry were separated, their progress was slow and their use limited; but once these sciences were united, they lent each other mutual support and advanced rapidly together towards perfection. We owe to Descartes the application of Algebra to Geometry; this has become the key to the greatest discoveries in all fields of mathematics.

Joseph-Louis Lagrange

Euklidischer Raum

In der Mathematik ist der **euklidische Raum** zunächst der „Raum unserer Anschauung", wie er in Euklids *Elementen* durch Axiome und Postulate beschrieben wird (vgl. euklidische Geometrie). Bis ins 19. Jahrhundert wurde davon ausgegangen, dass dadurch der uns umgebende physikalische Raum beschrieben wird. Der Zusatz „euklidisch" wurde nötig, nachdem in der Mathematik allgemeinere Raumkonzepte (z. B. hyperbolischer Raum, riemannsche Mannigfaltigkeiten) entwickelt wurden und es sich im Rahmen der speziellen und allgemeinen Relativitätstheorie zeigte, dass zur Beschreibung des Raums in der Physik andere Raumbegriffe benötigt werden (Minkowski-Raum, Lorentz-Mannigfaltigkeit).

Im Laufe der Zeit wurde Euklids Geometrie auf verschiedene Arten präzisiert und verallgemeinert:

- axiomatisch durch Hilbert (*siehe Hilberts Axiomensystem der euklidischen Geometrie*),
- als **euklidischer Vektorraum** (einem über $\mathbb{R}$ definierten Vektorraum mit Skalarprodukt),
- als **euklidischer Punktraum** (einem affinen Raum, der über einem euklidischen Vektorraum modelliert ist),
- als Koordinatenraum $\mathbb{R}^3$ mit dem Standardskalarprodukt.

Wenn vom *euklidischen Raum* die Rede ist, dann kann jede dieser Definitionen gemeint sein oder auch eine höherdimensionale Verallgemeinerung. Den zweidimensionalen euklidischen Raum nennt man auch **euklidische Ebene**. In diesem zweidimensionalen Fall wird der Begriff in der synthetischen Geometrie etwas allgemeiner gefasst: *Euklidische Ebenen* können dort als affine Ebenen über einer

Wikipedia, „Euklidischer Raum"

12.3 Schnittprobleme und Lagebeziehungen in Vektorräumen und affinen Räumen

Wir untersuchen in diesem Abschnitt die Lagebeziehungen affiner Teilräume. Dazu betrachten wir im Folgenden Teilräume $N = P + U$, $M = Q + W$ eines affinen Raumes A_V. Wir werden sehen, dass die Fragestellung häufig etwas mit den beteiligten Unter*vektorräumen* U, W zu tun hat. Als Vorbereitung dient daher der folgende Abschnitt, der gleichzeitig als Erinnerung an Bekanntes aus der LAAG I und zur Einführung einiger neuer Begriffe und Zusammenhänge dienen soll.

12.3.1 Schnittbeziehungen zwischen Untervektorräumen

Lemma/Definition 12.3.1. *(Summe und Schnitt von Untervektorräumen)*

Es seien U, W Untervektorräume eines Vektorraumes V.

(i) Die Mengen

$$U + W := \{u + w \mid u \in U, w \in W\},$$

die sogenannte Summe von U und W, und der Schnitt $U \cap W$ sind dann ebenfalls Untervektorräume von V.

(ii) Die Vereinigungsmenge $U \cup W$ ist im Allgemeinen kein Unterraum. $U + W$ ist der kleinste Unterraum, der $U \cup W$ enthält in dem Sinne, dass jeder Unterraum, der $U \cup W$ enthält, auch ganz $U + W$ enthält.

(iii) Eine Summe zweier Unterräume heißt direkte Summe, in Zeichen: $U \oplus W$, wenn zusätzlich $U \cap W = \{0_V\}$ gilt.

(iv) Allgemein ist für eine Teilmenge $M \subseteq V$ der Aufspann (die „lineare Hülle“) $\mathcal{L}(M)$ von M definiert als die Menge aller möglicher Linearkombinationen der Elemente von M. $\mathcal{L}(M)$ ist der kleinste Teilraum von V, der die Menge M enthält; insbesondere ist für $M = U \cup W$

$$U + W = \mathcal{L}(U \cup W).$$

Beweis: Der Beweis von (i) ist eine Übungsaufgabe aus der LAAG I, die man schnell mit dem Teilraumkriterium erledigt hat.

Zur ersten Aussage zur Vereinigungsmenge betrachtet man z.B. die Vereinigung zweier Ursprungsgeraden in $\mathbb{R}^3$, die kein Untervektorraum ist, da die Menge z.B. gegenüber der Addition nicht abgeschlossen ist und so das Teilraumkriterium verletzt. Für die Hauptaussage aus (ii) sei ein Unterraum Z mit $U \cup W \subseteq Z$ gegeben. Für alle $u \in U$ und $w \in W$

muss nach dem Teilraumkriterium (LAAG I) auch immer $u + w \in Z$, also insgesamt $U + W \subseteq Z$ gelten.

(iv) folgt analog, da nach dem Teilraumkriterium alle Linearkombinationen von Elementen vom M in einem Teilraum T mit $M \subseteq T$ enthalten sein müssen (vgl. auch eine entsprechende Übungsaufgabe in der LAAG I).

$\square$

Wir halten eine zentrale Feststellung über die Dimension von Schnitten von Untervektorräumen fest.

Satz 12.3.2. *(Dimensionssatz für Untervektorräume)*

Es sei V ein Vektorraum, U, W seien endlichdimensionale Unterräume von V. Dann gilt für U und W der Dimensionssatz für Unterräume,

$$\dim(U + W) \;=\; \dim(U) + \dim(W) - \dim(U \cap W). \tag{12.2}$$

Nach Lemma 12.3.1 sind alle diese Dimensionen wohldefiniert.

Beweis: Sei $k = \dim(U \cap W)$ und $(b_1, \ldots, b_k)$ eine Basis von $U \cap W$. Ist $\dim U = n$ und $\dim W = m$, so gibt es nach dem Basisergänzungssatz $u_1, \ldots, u_{n-k} \in V$, so dass $B_U := (b_1, \ldots, b_k, u_1, \ldots, u_{n-k})$ eine Basis von U ist; genauso existieren $w_1, \ldots, w_{m-k} \in V$, so dass $B_W := (b_1, \ldots, b_k, w_1, \ldots, w_{m-k})$ eine Basis von W ist. Wir zeigen, dass

$$B := (b_1, \ldots, b_k, u_1, \ldots, u_{n-k}, w_1, \ldots, w_{m-k})$$

eine Basis von $U + W$ ist, dann gilt

$$\dim(U + W) = \dim(U \cap W) + (\dim(U) - k) + (\dim(W) - k),$$

was wegen $k = \dim(U \cap W)$ die Behauptung zeigt. Da sich alle $v \in U + W$ in der Form $v = u + w$ mit $u \in U$ und $w \in W$ schreiben lassen und sich u und w jeweils in den obigen Basen darstellen lassen, ist klar, dass B erzeugend ist. Wir müssen zeigen, dass B linear unabhängig ist. Sei also

$$\sum_{i=1}^{k} c_i \cdot b_i + \sum_{i=1}^{n-k} a_i \cdot u_i + \sum_{i=1}^{m-k} a'_i \cdot w_i = 0_V; \qquad (\star)$$

wir haben zu zeigen, dass alle beteiligten Koeffizienten gleich null sind. Zunächst ist die obige Gleichung äquivalent zu

$$\underbrace{\sum_{i=1}^{k} c_i \cdot b_i + \sum_{i=1}^{n-k} a_i \cdot u_i}_{:=u' \in U} = \underbrace{\sum_{i=1}^{m-k} (-a_i') \cdot w_i,}_{:=u' \in W}$$

woraus folgt, dass der eben definierte Vektor u' in $U \cap W$ enthalten ist. Es gibt also eindeutige Koeffizienten c_i', $i \in \underline{k}$, so dass

$$\sum_{i=1}^{m-k} a_i' \cdot w_i = \sum_{i=1}^{k} c_i' \cdot b_i, \qquad \text{d.h.} \quad \sum_{i=1}^{m-k} a_i' \cdot w_i + \sum_{i=1}^{k} (-c_i') \cdot b_i = 0_V,$$

woraus wegen der Basiseigenschaft von B_U dann $a_i' = 0$ für alle $i \in \underline{m-k}$ folgt. Dasselbe Argument umgekehrt hingeschrieben gibt auch $a_i = 0$ für alle $i \in \underline{n-k}$. Die Gleichung $(\star)$ reduziert sich so zu $\sum_{i=1}^{k} c_i \cdot b_i = 0_V$, woraus mit der linearen Unabhängigkeit von $(b_1, \ldots, b_k)$ auch $c_i = 0$ für alle $i \in \underline{k}$ folgt.

□

Bemerkungen dazu:

- Aus der Dimensionsformel folgt auch, dass $\dim U + W = \dim U + \dim W$ gilt genau dann, wenn $U \cap W = \{0_V\}$ ist, also im Fall, dass wir es mit einer direkten Summe $U \oplus W$ von Vektorräumen zu tun haben.
- Im Fall $V = \mathbb{K}^n$ lässt sich die Dimension von $U + W$ also dadurch bestimmen, dass man den Rang der Matrix $(B_U \mid B_W)$ bestimmt (denn das ist die maximale Anzahl linear unabhängiger Spalten von A). Bringt man $(B_U \mid B_W)$ auf Zeilenstufenform, um den Rang zu ermitteln, so steht bilden die Spalten $(B_U \mid B_W)$, die in der Zeilenstufenform zu Pivotspalten werden, auch eine Basis von $U + V$.
- Aus dem Kern dieser Matrix(-abbildung) lässt sich dann auch der Schnittraum $U \cap W$ bestimmen, das ist allerdings schwieriger: Es gilt $u = \sum_{i=1}^{n} a_i \cdot u_i \in U \cap W$ genau dann, wenn es $a_1', \ldots, a_m' \in \mathbb{K}$ gibt, so dass der Vektor $x = (a_1, \ldots, a_n, a_1', \ldots, a_m') \in \mathbb{K}^{n+m}$ im Kern der Matrix $(B_U \mid B_W)$ liegt. Man bestimmt also diese Koeffizienten als Teilraum des $\mathbb{K}^n$, setzt diese in die Darstellung für u von oben ein und bestimmt eine Basis der resultierenden Menge. Wir gehen hier nicht weiter ins Detail.

12.3.2 Lagebeziehungen affiner Unterräume

Nach diesen Erinnerungen und Vorbereitungen sind wir bereit, Schnitte von affinen Teilräumen systematisch zu untersuchen. Wir beginnen mit der folgenden allgemeinen Feststellung:

Lemma 12.3.3. *(Schnitt von affinen Teilräumen)*

Der Schnitt zweier affiner Teilräume $N = P + U$, $M = Q + W$ ist entweder leer oder wieder ein affiner Teilraum. Genauer gilt im Fall $N \cap M \neq \emptyset$ für jeden Punkt $O \in (N \cap M)$

$$N \cap M \;=\; O + (U \cap W), \tag{12.3}$$

insbesondere ist

$$\dim(N \cap M) = \dim(U \cap W).$$

Beweis: Ist $N \cap M \neq \emptyset$, so gibt es einen Punkt $O \in A$ mit $O \in N \cap M$, und wir können $N = O + U$ und $M = O + W$ schreiben (nach 12.1). Ist $R \in A_V$ ein anderer Punkt, so kann man diesen als $R = O + v$ mit $v \in V$ schreiben. Wegen der obigen Darstellung $N = O + U, M = O + W$ ist $R \in N \cap M$ genau dann, wenn $v \in U$ und $v \in W$ gilt (wobei bei der Hinrichtung hier noch die Eindeutigkeit des Verbindungsvektors eingeht), also wenn $v \in U \cap W$ gilt. Es ist also $N \cap M = O + (U \cap W)$ und $U \cap W$ wieder ein Unterraum (vgl. 12.3.1(i)), also $N \cap M$ ein affiner Teilraum.

□

Wie prüft man nun, wann der Schnitt zweier affiner Teilräume leer ist? Eigentlich wie in der Schule, man muss es nur wiedererkennen.

Lemma 12.3.4. *(Kriterium für nichtleeren Schnitt)*
Für zwei affine Teilräume $N = P + U$, $M = Q + W$ von A_V gilt

$$N \cap M \neq \emptyset \quad \Leftrightarrow \quad \overrightarrow{PQ} \in U + W.$$

Beweis: Es ist $N \cap M \neq \emptyset$ genau dann, wenn es ein $u \in U$ und ein $w \in W$ gibt, so dass $P + v = Q + w$ ist. Wegen Lemma 12.2.3(iii) gilt das genau dann, wenn $P + (u - w) = Q$ mit $u - w \in U + W$ ist, was gleichbedeutend mit $\overrightarrow{PQ} \in U + W$ ist.

□

Wir haben oben schon betont, dass der wichtigste Anwendungsfall der des affinen Punktraums $A = \mathbb{K}^n$ (insbesondere $A = \mathbb{R}^2, \mathbb{R}^3, \mathbb{R}^n$) über sich selbst als Vektorraum $V = \mathbb{K}^n$ ist. Wie jetzt werden wir allgemeine Resultate für diesen Spezialfall im Folgenden öfter noch einmal gesondert formulieren und untersuchen.

Lemma 12.3.5. *(Schnittberechnung für affine Teilräume von $\mathbb{K}^n$)*

Sei $A = V = \mathbb{K}^n$ mit $n < \infty$. Seien weiter $N = P + U, M = Q + W$ affine Teilräume von A, $B = (b_1, \dots, b_k)$ eine Basis von U, $C = (c_1, \dots, c_\ell)$ eine Basis von W (jeweils mit Vektoren $b_i, c_i \in \mathbb{K}^n$) und $(B \mid C) \in \mathbb{K}^{n \times (k+\ell)}$ die kanonische Matrix mit den Vektoren aus B und aus C als Spalten. Dann ist $N \cap M \neq \emptyset$ genau dann, wenn das lineare Gleichungssystem

$$(B \mid C) \cdot x = \overrightarrow{PQ}$$

lösbar ist, d.h. wenn

$$\operatorname{rg}(B \mid C \mid \overrightarrow{PQ}) = \operatorname{rg}(B \mid C) \tag{12.4}$$

gilt.

Beweis: Die Schnittbedingung aus dem vorangegangenen Beweis, nämlich, dass $u \in U$ und $w \in W$ existieren, so dass $P + u = Q + w$ ist, ist gleichwertig zur Frage, ob es Vektoren $a = (a_1, \dots, a_k) \in \mathbb{K}^k$, $a' = (a'_1, \dots, a'_\ell) \in \mathbb{K}^\ell$ gibt mit

$$R \;=\; P \;+\; \sum_{i=1}^{k} a_i b_i = Q + \sum_{i=1}^{\ell} a'_i c_i. \qquad (\star)$$

(Die Elemente b_i, c_i sind die jeweiligen Basisvektoren, s. Behauptung.) Das lässt sich gleichwertig mit Hilfe der Matrix-Vektor-Multiplikation darstellen, wodurch man die folgenden Äquivalenzen erhält: Die Vektoren $a \in \mathbb{K}^k, a' \in \mathbb{K}^\ell$ erfüllen

$$(\star) \;\Leftrightarrow\; P + B \cdot a = Q + C \cdot a' \;\Leftrightarrow\; B \cdot a + C \cdot (-a') = \overrightarrow{PQ} \;\Leftrightarrow\; (B \mid C) \cdot \begin{pmatrix} a \\ -a' \end{pmatrix} = \overrightarrow{PQ}.$$

Es ist also $N \cap M \neq \emptyset$ genau dann, wenn das letzte LGS eine Lösung $x = (a, -a') \in \mathbb{K}^{k+\ell}$ besitzt; das ist genau dann der Fall, wenn die behauptete Rangbedingung erfüllt ist.

□

Bemerkung dazu:

- Da Punkte von A nulldimensionale affine Teilräume von A sind, lässt sich mit Lemma 12.3.5 auch prüfen, ob ein Punkt R zu einem affinen Teilraum $N = P + U$ mit Basis B gehört: Hier nimmt die Bedingung (12.4) die folgende einfache Form an[1]:

$$\operatorname{rg}(B \mid \overrightarrow{PR}) = \operatorname{rg}(B).$$

Beispiel: S. Vorlesung.

Mit dem obigen Ergebnis aus Lemma 12.3.3 ist die Menge $N \cap M$ im Fall $N \cap M \neq \emptyset$ ein affiner Unterraum der Dimension $\dim(N \cap M) = \dim(U \cap W)$. Den Zusammenhang zwischen $\dim(U \cap W)$ und $\dim U$, $\dim W$ liefert die Dimensionsformel (12.2) für Untervektorräume.

Analog zum Vorgehen zu Lagebeziehungen im Anschauungsraum und in $\mathbb{R}^3$ in der schulischen Systematik lassen sich auch allgemeine affine Teilräume auf Parallelität untersuchen; das wollen wir hier auch tun.

Definition 12.3.6. *(Parallele affine Unterräume)*

Seien $N = P+U$ und $M = Q+W$ zwei affine Unterräume eines affinen Raumes $(A, V, +)$. Dann heißt N parallel zu M, wenn für die zugehörigen Unterräume (mit Notation wie oben) $U = W$ gilt; N heißt schwach parallel zu M, falls $U \subsetneq W$ gilt.

So können zwei Ebenen im $\mathbb{R}^3$ beispielsweise parallel sein, eine Gerade und eine Ebene sind nach der obigen Terminologie nie parallel, sondern die Gerade ggf. schwach parallel zur Ebene.

[1] Man erinnere sich, dass der nulldimensionale Vektorraum $\{0_V\}$ die leere Menge als Basis hat, d.h. unser gerade betrachteter Spezialfall ist genau genommen schon in der Formulierung von Lemma 12.3.5 enthalten.

Lemma 12.3.7. *(Schnittraum paralleler Teilräume)*

(i) *Seien N und M zwei parallele affine Unterräume eines affinen Raumes A_V. Dann gilt entweder $N \cap M = \emptyset$ oder $N = M$.*

(ii) *Ist N schwach parallel zu M, so haben wir $N \cap M = \emptyset$ oder $N \subsetneq M$, also $N \cap M = N$.*

Beweis: Sei $N \cap M \neq \emptyset$. Dann existiert ein Punkt $P \in N \cap M$, also kann man nach (12.1) $N = P + U$ mit $P \in M$ und $U = W$ schreiben; es ist also $N = M$ nach Lemma 12.2.5(ii). Mit derselben Argumentation folgt im schwach parallelen Fall $N = P + U \subsetneq P + W = M$.
□

Bemerkung dazu:

- Auch die (schwache) Parallelität zweier affiner Unterräume N, M von $\mathbb{K}^n$ lässt sich gut rechnerisch untersuchen, indem man Basen B und C für die zugehörigen Unterräume U und W ermittelt und dann den Rang der Matrix $A = (B \mid C)$ (also mit den Vektoren von B und C als Spalten) bestimmt. Es gilt: $N = P + U$ ist schwach parallel zu $M = Q + W$ genau dann, wenn $U \subsetneq W$ genau dann, wenn $W = U + W$ genau dann, wenn $\mathrm{rg}(C) = rg(B \mid C)$ gilt; genauso ist (da sich $U = W$ durch die zwei Bedingungen $U \subseteq W$ und $W \subseteq U$ schreiben lässt), N parallel zu M genau dann, wenn $\mathrm{rg}(C) = \mathrm{rg}(C) = rg(B \mid C)$ gilt.

Wieder betrachten wir noch einmal gesondert den Fall $A = V = \mathbb{K}^n$, in dem sich der Schnitt zweier affiner Teilräume $N = P + U, M = Q + W \subseteq A = \mathbb{K}^n$ effizient zum Beispiel mit Hilfe des auf der folgenden Seite dargestellten, auf dem Gauß-Jordan-Algorithmus aufbauenden Verfahren untersuchen lässt.

Algorithmus:
Untersuchung von Lagebeziehungen affiner Teilräume von $\mathbb{K}^n$

- Gegeben: $P, Q \in \mathbb{K}^n$, Basis $B = (b_1, \ldots, b_k)$ von U, Basis $C = (c_1, \ldots, c_m)$ von W.
- Damit gilt: $k = \dim(U), m = \dim(W)$.
- Bringe (z.B. per Gauß-Jordan-Algorithmus) die Matrix $(B \mid C \mid \overrightarrow{PQ})$ auf normierte Zeilenstufenform. Bestimme daraus $\operatorname{rg}(B \mid C \mid \overrightarrow{PQ})$ und $\operatorname{rg}(B \mid C)$.

Hieraus lassen sich nun alle gewünschten Ergebnisse ablesen:

- Es ist $r := \dim(U + V) = \operatorname{rg}(B \mid C)$, also sind die Räume N, M
  - parallel, falls $r = k = m$,
  - schwach parallel, falls $k \neq m$ und $r = \max\{k, m\}$.
- Es ist $N \cap M \neq \emptyset$ genau dann, wenn $\operatorname{rg}(B \mid C \mid \overrightarrow{PQ}) = \operatorname{rg}(B \mid C)$ (Lemma 12.3.5). Für die Dimension des Schnittraumes gilt in diesem Fall (vgl. Lemma 12.3.3, Dimensionssatz 12.3.2)

$$\dim(N \cap M) = \dim(U \cap V) = k + m - r.$$

- Der Schnittraum selbst ergibt sich aus der Lösungsmenge des Systems

$$(B \mid C) \cdot x = \overrightarrow{PQ},$$

  die man aus der normierten Zeilenstufenform ablesen kann: Bezeichnen wir mit

$$\mathbb{L} \ = \ \{x = (a, a') \mid a \in \mathbb{K}^k, a' \in \mathbb{K}^m\}$$

  die Lösungsmenge dieses Systems (zur Notation vgl. Lemma 12.3.5 und den zugehörigen Beweis), so ist

$$\mathbb{S} \ := \ \{P + B \cdot a \mid (a, a') \in \mathbb{L}\} \ = \ \{Q + C \cdot (-a') \mid (a, a') \in \mathbb{L}\}$$

  der affine Schnittraum der affinen Unterräume.

Beispiel dazu: Schnitt zweier Ebenen in $\mathbb{R}^4$. Aus [Filler], S. 213.

12.4 Affine Hüllen von Teilmengen affiner Räume

Am Anfang des letzten Abschnitts habe ich daran erinnert, dass für Untervektorräume U, W eines Vektorraumes V die Menge $U \cup W$ im Allgemeinen kein Vektorraum ist. Auch für affine Räume gilt, dass die Vereinigung affiner Räume im allgemeinen keinen affinen Raum ergibt. Es gibt aber ein analoges Konzept zur linearen Hülle von Vektormengen, die insbesondere einen affinen „Oberraum“ für die Vereinigung zweier affiner Räume liefern kann.

Definition/Satz 12.4.1. *(Affine Hülle)*

Sei A_V ein affiner Raum und $M \subseteq A$ eine beliebige nichtleere Teilmenge von A. Sei $P \in M$ beliebig gewählt und wie oben $ort_P(M) = \{\overrightarrow{PQ} \mid Q \in M\}$. Dann ist

$$A(M) := P + \mathcal{L}(ort_P(M)) \tag{12.5}$$

ein affiner Teilraum $A(M)$ von A_V mit den folgenden Eigenschaften:

(i) $M \subseteq A(M)$

(ii) *$A(M)$ ist der kleinste affine Raum, der M enthält in dem Sinne, dass jeder affine Raum $N \subseteq A_V$, der M enthält, auch $A(M)$ enthält.*

Die Menge $A(M)$ heißt affine Hülle von M.

Ist M elementweise angegeben, z.B. $M = \{P, Q\}$, so lassen wir die Mengenklammern weg und schreiben z.B. einfach $A(P, Q)$. Sind N, M affine Teilräume von A_V, so nennt man $A(N \cup M)$ auch den (affinen) Verbindungsraum von N und M.

Beweis: $A(M)$ ist ein affiner Raum nach Definition 12.2.4 und Lemma 12.2.5, und es gilt $Q \in A(M)$ für alle $Q \in M$ nach Definition. Ist N ein weiterer affiner Raum mit $M \subseteq N$, so gilt nach Lemma 12.2.3 wegen $P \in N$ und $M \subseteq N$ auch

$$A(M) = P + \mathcal{L}(\mathrm{ort}_P(M)) \subseteq N,$$

das zeigt die Behauptung.

□

Beispiele: Sei A wieder ein affiner Raum.

1. Sind $P, Q \in A$, so ist
$$A(P,Q) = P + \mathcal{L}(\overrightarrow{PQ})$$
einfach die durch P und Q festgelegte Gerade.

2. Für drei Punkte $P, Q, R \in A$ ist
$$A(P,Q,R) = P + \mathcal{L}(\overrightarrow{PQ}, \overrightarrow{PR}),$$
das ist, falls $\overrightarrow{PQ}$ und $\overrightarrow{PR}$ linear unabhängig sind, gerade die Ebene durch P, Q und R.

3. Sind $N = P + U$ und $M = Q + W$ affine Unterräume mit nicht-leerem Schnitt und O ein gemeinsamer Punkt, so hat man die Darstellung $N = O + U$, $M = O + W$, und es ist
$$A(N \cup M) \;=\; O + \mathcal{L}(U \cup V) \;=\; O + (U + W).$$
(Denn es ist $U = \mathrm{ort}_O(M), W = \mathrm{ort}_O(N)$, und nach Lemma 12.3.1(iv) gilt für Unterräume $U, W \subseteq V$ die Beziehung $\mathcal{L}(U \cup W) = U + W$.)

4. Haben zwei affine Teilräume $N = P + U$ und $M = Q + W$ keinen gemeinsamen Punkt, so ist
$$A(N \cup M) \;=\; P + \Big(U + W + \mathcal{L}(\overrightarrow{PQ})\Big)\,. \tag{12.6}$$
Hinten steht hier eine Summe von drei Vektorräumen, die analog zur Summe zweier Vektorräume definiert ist, in diesem Fall als
$$U + W + \mathcal{L}(\overrightarrow{PQ}) \;=\; \{u + v + a \cdot \overrightarrow{PQ} \mid u \in U, v \in V, a \in \mathbb{K}\}.$$
Man beachte, dass die Punkte $P \in N$ und $Q \in M$ hier wieder beliebig gewählt werden können, um die affine Hülle in (12.6) darzustellen.

Die letzte Behauptung 4. aus den obigen Beispielen, dass (12.6) die affine Hülle von $N \cup M$ darstellt, ist noch zu beweisen.

Nach Definition ist $A(N \cup M) = P + \mathcal{L}(\mathrm{ort}_P(N \cup M)$. Wir stellen zunächst $\mathrm{ort}_P(N \cup M)$ genauer dar: Es ist

$$\begin{aligned}
\mathrm{ort}_P(N \cup M) &= \{\overrightarrow{PS} \mid S \in N\} \cup \{\overrightarrow{PS'} \mid S' \in M\} \\
&\overset{\mathrm{ort}_P \text{ bij. auf } U}{=} U \cup \left(\overrightarrow{PQ} + \{\overrightarrow{QS'} \mid S' \in M\}\right) \\
&= U \cup \left(\overrightarrow{PQ} + \mathrm{ort}_Q(M)\right) \\
&\overset{\mathrm{ort}_Q \text{ bij. auf } W}{=} U \cup (\overrightarrow{PQ} + W)
\end{aligned}$$

Um den Aufspann dieser Menge zu ermitteln, bemerken wir zunächst, dass der kleinste Unterraum, der alle Vektoren der Form $\overrightarrow{PQ}+w$ mit $w \in W$ enthält, der Raum $\mathcal{L}(\overrightarrow{PQ})+W$ ist (mit der Summe von Vektorräumen aus Satz 12.3.2 – Warum? Bekommen Sie das hin?). Also ist

$$\begin{aligned}
A(N \cup M) &= P + \mathcal{L}(U \cup (\overrightarrow{PQ} + W)) \\
&= P + \mathcal{L}(U \cup (\mathcal{L}(\overrightarrow{PQ}) + W)) \\
&= P + \left(U + W + (\mathcal{L}(\overrightarrow{PQ})\right)
\end{aligned}$$

wie behauptet.

Satz 12.4.2. *(Dimensionssatz für affine Unterräume)*

Seien $N = P + U, M = Q + W$ wieder affine Unterräume eines affinen Raums A. Dann gelten die Schnittformeln für affine Unterräume:

(i) Ist $N \cap M \neq \emptyset$, so gilt

$$\dim(A(N \cup M)) = \dim(N) + \dim(M) - \dim(N \cap M).$$

(ii) Im Fall $N \cap M = \emptyset$ gilt

$$\dim(A(N \cup M)) = \dim(N) + \dim(M) - \dim(U \cap W) + 1.$$

Beweis: Zu (i): Nach den Überlegungen in Beispiel 3 oben ist mit den Bezeichnungen von dort $A(N \cup M) = O + (U + W)$, also mit dem Dimensionssatz für Unterräume

$$\begin{aligned}\dim(A(N \cup M)) = \dim(U + W) &= \dim(U) + \dim(W) - \dim(U \cap W) \\ &= \dim(N) + \dim(N) - \dim(N \cap M)\end{aligned}$$

wie behauptet. (Die letzte Gleichheit gilt wegen (12.3).)

Zu (ii) zeigen wir, dass aus $N \cap M = \emptyset$ folgen muss, dass $\overrightarrow{PQ} \notin U + W$ ist; damit haben wir dann $\dim(U + W + \overrightarrow{PQ}) = \dim(U + W) + 1$. Kontrapositionsbeweis: Ist $\overrightarrow{PQ} \in U + W$, so gäbe es $u \in U, w \in W$ mit $P + u + w = Q$. Nach 12.2.3(iii) ist dann $P + u = Q + (-w)$. Das linke ist ein Punkt in N, das rechte einer in M, also ist der Schnitt von N und M nicht leer. Damit haben wir $\dim(U + W + \overrightarrow{PQ}) = \dim(U + W) + 1$ und erhalten analog zur Rechnung aus (i):

$$\dim(A(N \cup M)) = \dim(U + W) + 1 \quad = \quad \dim(U) + \dim(W) - \dim(U \cap W) + 1.$$

Man kann hier nun noch $\dim(U) = \dim(N)$ und $\dim(W) = \dim(M)$ ersetzen, um die in der Behauptung angegebene Formel zu bekommen. (In diesem Fall besteht aber kein Zusammenhang zwischen $\dim(U \cap W)$ und $\dim(N \cap M)$ wie in (i)).

□

12.5 Koordinatensysteme und Koordinatentransformationen in affinen Räumen

12.5.1 Koordinatisierung von Punkträumen, anschaulich

Wir skizzieren zunächst für den Fall des Anschauungsraumes $\mathcal{A}_3$, wie sich das entwickelte „affine Setting" relativ nutzen lässt, um nun relativ einfach ein Koordinatensystem zu definieren und damit jedem Punkt $P \in \mathcal{A}$ eindeutige Koordinaten zuordnen lassen. Das Verfahren für $\mathcal{A}_2$ ist dasselbe mit einem Basisvektor/einer Koordinate weniger, und auch das Vorgehen für allgemeine affine Räume birgt dann wenig Überraschungen.

Wir wählen

- einen festen Punkt $O \in A$, den Ursprung unseres Koordinatensystems. Zu jedem Punkt P gehört dann genau ein Vektor $\overrightarrow{OP}$ (und entsprechend die Verschiebung, die O auf P abbildet). Hinzu kommt
- eine Basis von $\mathbb{V}$, die im Raum aus drei linear unabhängigen Pfeilklassen $\vec{e}_1, \vec{e}_2, \vec{e}_3$ besteht.

Heftet man die Vektoren $\vec{e}_1, \vec{e}_2, \vec{e}_3$ am Ursprung an (d.h. man betrachtet den Repräsentanten mit O als Anfangspunkt), so ergeben sich die Koordinatenachsen unseres Koordinatensystems $K = (O; \vec{e}_1, \vec{e}_2, \vec{e}_3)$. Hierdurch wird auch

- eine Orientierung des Raumes festgelegt (Anordnung der Achsen des gewählten Koordinatensystems $\widehat{=}$ „positive Orientierung", mit der andere Orientierungen vergleichen werden können)
- eine Einheitslänge auf der i-ten Achse festgelegt ($\widehat{=}$ Länge des Vektors $\vec{e}_i$).

Nach Definition einer Basis lässt sich jeder Vektor $\vec{v} \in \mathbb{V}$, also auch jeder Ortsvektor $\overrightarrow{OP}$, als Linearkombination

$$\overrightarrow{OP} = x_1 \cdot \vec{e}_1 + x_2 \cdot \vec{e}_2 + x_3 \cdot \vec{e}_3$$

schreiben. Die beteiligten Zuordnungen

$$P \quad \leftrightarrow \quad \overrightarrow{OP} \quad \leftrightarrow \quad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

sind beide bijektiv, zu jedem Punkt P des Raumes gehört also mit Hilfe dieser Konstruktion genau ein Koordinatenvektor von P aus dem $\mathbb{R}^3$.

Das war – etwas formaler gefasst – die gängige Konstruktion der Schule zur Koordinatisierung unserer anschaulichen Punktmengen, (der Zeichenebene und) des Anschauungsraums. Wir wollen dieses Konzept nun im Rest dieses Kapitel verallgemeinern und axiomatisch formalisieren und daraus einiges an Erkenntnissen über die „linear-algebraische Struktur“ solcher affiner Räume gewinnen.

Definition 12.5.1. *(Koordinatensysteme affiner Räume)*

Sei $(A, V, +)$ ein affiner Raum mit $\dim V = \dim A = n < \infty$. Sei O ein beliebiger Punkt aus A und $B = (b_1, \ldots, b_n)$ eine Basis von V. Dann heißt das Paar $K := (O, B)$ Koordinatensystem von A und der Punkt O Ursprung von K.

Nach Satz 12.2.2 ist damit

$$\begin{aligned} A \quad = \quad O + V \quad &:= \quad \{O + v \mid v \in V\} \\ &= \quad \{O + \textstyle\sum_{i=1}^{n} a_i \cdot b_i \mid a_1, \ldots, a_n \in \mathbb{K}\}. \end{aligned} \tag{12.7}$$

Die Abbildung

$$k_K : A \to \mathbb{K}^n \qquad P \mapsto k_B(ort_O(P)) = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

heißt die Koordinatenabbildung zum Koordinatensystem K, und $k_K(P)$ nennen wir die Koordinaten des Punktes P bezüglich K.

Bemerkungen dazu:

- Da die Abbildungen $ort_O : A \to V$ und $k_B : V \to \mathbb{K}^n$ bijektiv sind, ist auch jede Koordinatenabbildung k_K bijektiv.
- Man beachte, dass die Wahl eines Koordinatensystems in A, also des Punktes $O \in A$ und der Basis B von V absolut beliebig ist. Das wird im Anschauungsraum (z.B. in der Physik) oft genutzt, um ein für das gerade zu beschreibende Problem besonders günstiges Koordinatensystem zu wählen.
- Die erhaltenen Koordinaten eines Punktes $P \in A$ bezüglich $K = (O, B)$ hängen dann aber natürlich vom jeweils Koordinatensystem ab!
  (Das sieht man auch sehr gut in der Definition $k_K := k_B \circ \mathrm{ort}_O$.)

- Das uns bekannte kartesische Koordinatensystem des $\mathbb{R}^3$ ist $K = (O; e_1, e_2, e_3)$ mit

$$O = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad b_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \; b_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \text{ und } b_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

  In diesem konkreten Beispiel stimmt wie im Vektorraumfall jeder Punkt selbst mit seinen Koordinaten bezüglich der Standardbasis und auch mit seinen affinen Koordinaten überein.

12.5.2 Koordinatentransformationen im affinen Raum $\mathbb{K}^n$

Wir wollen als wichtige Vorbereitung auf das kommende Kapitel noch sehen, wie man „zwischen verschiedenen Koordinatensystemen im $\mathbb{K}^n$ wechselt“, da dies die Darstellung von geometrischen Abbildungen im $\mathbb{K}^n$ massiv vereinfachen kann.
Zur Motivation betrachten wir die Situation im Anschauungsraum $\mathcal{A}_3$. Die Koordinatisierung funktioniert hier folgendermaßen: In vielen Anwendungen ist im Raum ein festes „Hauptkoordinatensystem“ W gegeben, das auch oft als das Weltkoordinatensystem bezeichnet wird. Es ist

$$W = (O, \overrightarrow{OP_1}, \overrightarrow{OP_2}, \overrightarrow{OP_3}),$$

wobei der Ursprung O gewählt wird und und die drei Punkte P_1, P_2, P_3 im Raum so liegen, dass die Richtungsvektoren $\overrightarrow{OP_1}, \overrightarrow{OP_2}, \overrightarrow{OP_3}$ drei unabhängige Raumachsen sind.
Für die Koordinaten der Punkte O, P_1, P_2, P_3 bezüglich W gilt mit der Koordinatenabbildung k_W:

$$k_W(O) = 0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad k_W(P_1) = e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad k_W(P_2) = e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad k_W(P_3) = e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

und die Koordinaten aller anderen Punkte P der Ebene/des Raumes ergeben sich durch $k_W(\overrightarrow{OP})$.

Die Darstellung und Anwendung von geometrischen Abbildungen wird aber oft viel leichter, wenn z.B. die Achsen einer Spiegelung auf den Achsen liegen oder das Zentrum einer zentrischen Streckung der Koordinatenursprung ist. Dabei stellt sich die folgende Problematik: Haben wir ein weiteres Koordinatensystem $K = (O', \overrightarrow{O'Q_1}, \overrightarrow{O'Q_2}, \overrightarrow{O'Q_3})$ gewählt, wollen wir Koordinaten bezüglich W in Koordinaten bezüglich K umrechnen können und umgekehrt. Das geht mit Hilfe von Matrizen und Vektoren recht einfach folgendermaßen:

Lemma 12.5.2. *(Koordinatentransformationen in $\mathbb{K}^n$)*

Bezeichne $W = (0, e_1, \ldots, e_n)$ das Standard-(„Welt-“)Koordinatensystem des affinen Raumes $\mathbb{K}^n$ und $K = (v, b_1, \ldots, b_n)$ ein weiteres Koordinatensystem mit $v \in \mathbb{K}^n$ als Ursprung. B bezeichne die Matrix $B \in \mathbb{K}^{n\times n}$ mit den Basisvektoren $b_1, \ldots, b_n$ als Spalten.

Dann rechnet die Koordinatenabbbildung k_K die Weltkoordinatenvektoren von $x \in \mathbb{K}^n$ in Koordinaten bezüglich K um. Sie und ihre Inverse k_K^{-1} sind gegeben durch die Bijektionen

$$k_K : \mathbb{K}^n \to \mathbb{K}^n, \quad x \mapsto B^{-1} \cdot (x - v), \qquad k_K^{-1} : \mathbb{K}^n \to \mathbb{K}^n, \quad a \mapsto B \cdot a + v.$$

Bemerkungen dazu:

- Das obige Lemma ist eine Verallgemeinerung von Lemma9.5.1, das sich als Koordinatentransformation im affinen Raum $\mathbb{K}^n$ lesen lässt, bei der der Ursprung fest bleibt.

Beweis: Die angegebene Formel ergibt sich einfach aus dem Zusammenhang für die gesuchten Koordinaten bezüglich K: Ist $x \in \mathbb{K}^n$ (und somit sein eigener Koordinatenvektor bezüglich Standardbasis), so suchen wir einen Koordinatenvektor $a = (a_1, \ldots, a_n)$ mit

$$x = v + a_1 \cdot b_1 + \ldots + a_n \cdot b_n =: B \cdot a + v.$$

Das ist bereits die Umrechnungsformel von K auf W, die umgekehrte Umrechnung ergibt sich, indem man den Zusammenhang nach a auflöst.

□

12.6 Aufgaben zu Kapitel 12

Aufgabe 12.63. (Affine Räume):

(a) Zeigen Sie: Ist $(V, +, \cdot)$ ein $\mathbb{K}$-Vektorraum, so ist das Tripel $(V, V, +)$ mit der Addition $+$ aus V ein affiner Raum („über sich selbst“).

(b) Beweisen Sie die restlichen Aussagen aus Lemma X.2.3 (Die Nummerierung entspricht der aus der Vorlesung): Ist $(A, V, +)$ ein affiner Raum, so gilt für alle $P, Q, R, S \in A, v \in V$:

 (ii) $P + v = Q + v \Longrightarrow P = Q$,

 (v) Es ist $\overrightarrow{PQ} + \overrightarrow{QR} = \overrightarrow{PR}$,

 (vi) $\overrightarrow{PQ} + \overrightarrow{QP} = 0_V$,

 (vii) $\overrightarrow{PQ} = \overrightarrow{RS} \Longrightarrow \overrightarrow{PR} = \overrightarrow{QS}$.

 0_V bezeichnet hierbei den Nullvektor des Vektorraumes V.

(c) Sei $(A, V, +)$ ein affiner Raum. Zeigen Sie, dass dann für $P, Q \in A, v, w \in V$ gilt:

$$\overrightarrow{P + v\ Q + w} = \overrightarrow{PQ} + w - v.$$

Begründen Sie in jeder der Teilaufgaben jeden Ihrer Schritte sorgfältig mit Hilfe entsprechender Axiome für affine Räume bzw. Vektorräume.

Aufgabe 12.64. (Geometrisches):

Sei $\square PQRS$ dasjenige Viereck im Anschauungsraum $\mathcal{A}$, das durch vier verschiedene Punkte $P, Q, R, S \in \mathcal{A}$, von denen keine drei auf einer Geraden liegen, festgelegt ist.

$\square PQRS$ heißt in der der euklidischen Geometrie *Parallelogramm* genau dann, wenn die Geraden $g_{P,Q}$ und $g_{R,S}$ und die Geraden $g_{Q,R}$ und $g_{S,P}$ parallel sind. Stattet man $\mathcal{A}$ mit den in der Vorlesung vorgestellten Pfeilklassen $\mathbb{V}$ aus, so ist das äquivalent dazu, dass $\overrightarrow{PQ} = \overrightarrow{SR}$ gilt, vgl. Übung.

Benutzen Sie die über $\mathcal{A}$ definierten Pfeilklassen, die Punkt-Vektor-Addition $+ : \mathcal{A} \times \mathbb{V} \to \mathcal{A}$ und die für das Tripel $(\mathcal{A}, \mathbb{V}, +)$ geltenden Rechenregeln für affine Räume, um den so genannten „Satz von Varignon“ zu zeigen:

Die Mittelpunkte M_{AB}, M_{BC}, M_{CD} und M_{DA} der Seiten eines beliebigen (auch räumlichen) Vierecks $\Box ABCD$ bilden ein Parallelogramm.

Aufgabe 12.65. (Geometrisches II):

Gegeben sei ein Dreieck ΔABC in der Anschauungsebene $\mathcal{A}$. Als Seitenhalbierende des Dreiecks bezeichnet man die Gerade, die durch einen Eckpunkt des Dreiecks und den Mittelpunkt der gegenüberliegenden Seite verläuft. Als Seitenmittendreieck des Dreiecks ΔABC bezeichnet man das Dreieck, dessen Eckpunkte die Mittelpunkte M_{AB}, M_{BC} und M_{CA} der Seiten eines Dreiecks ΔABC sind.

Benutzen Sie wieder die über $\mathcal{A}$ definierten Pfeilklassen, die Punkt-Vektor-Addition $+ : \mathcal{A} \times \mathbb{V} \to \mathcal{A}$ und die für das Tripel $(\mathcal{A}, \mathbb{V}, +)$ geltenden Rechenregeln für affine Räume, um zu zeigen:

(a) Alle Seitenhalbierenden laufen durch den Punkt

$$S = A + \frac{1}{3}\overrightarrow{AB} + \frac{1}{3}\overrightarrow{AC}.$$

(b) S teilt alle Seitenhalbierenden im Verhältnis $2:1$.
(Den Punkt S nennt man den *Schwerpunkt* des Dreiecks.)

(c) Ist $O \in \mathcal{A}$ ein beliebiger Referenzpunkt, so gilt für die Ortsvektoren bezüglich O

$$ort_O(S) = \frac{1}{3}\big(ort_O(A) + ort_O(B) + ort_O(C)\big).$$

(d) Für ein beliebiges Dreieck ΔABC sind der Schwerpunkt des Dreiecks ΔABC und der Schwerpunkt seines Seitenmittendreiecks $\Delta M_{AB}M_{BC}M_{CA}$ identisch.

Aufgabe 12.66. (Direkte Summe von Untervektorräumen):

Es seien U, W zwei endlichdimensionale Untervektorräume eines Vektorraumes V mit $\dim U = n$, $\dim W = m$. Beweisen Sie, <u>ohne</u> den Dimensionssatz für Untervektorräume zu verwenden (Sie können sich aber an den Ideen des Satzes orientieren):

(a) Es gilt $\dim(U + W) = n + m$ genau dann, wenn $U \cap W = \{0_V\}$ ist.

(b) Ist $U \cap W = \{0_V\}$ und $v \in U + W$, so gibt es genau einen Vektor $u \in U$ und genau einen Vektor $w \in W$ so, dass $v = u + w$ gilt.

Einige Hinweise hierzu:

- Tipp zum Beweis: Wählen Sie Basen $(b_1, \dots, b_n)$ von U bzw. $(b'_1, \dots, b'_m)$ von W und untersuchen Sie die Vereinigung dieser Mengen auf Eigenschaften wie lineare Unabhängigkeit/Abhängigkeit.
- Zum Beweis von Genau-Dann-Wenn-Beziehungen lassen sich verschiedene Wege einschlagen, vgl. etwa Lemma 4.2. im Anhang; für Implikationen $A \Rightarrow B$ ist ein Kontrapositionsbeweis oft eine gute Idee.

Aufgabe 12.67. (Lagebeziehungen in $\mathbb{R}^7$):

Im $\mathbb{R}^7$ sind die affinen Teilräume N_1, N_2, N_3 als jeweilige Lösungsmengen der linearen Gleichungssysteme

$$
\begin{array}{ll}
(G1) & x_1 = 3, \quad x_5 = 7, \quad x_7 = 5, \quad x_2 + x_3 = 1 \\
(G2) & x_1 = 2, \quad x_7 = 3 \\
(G3) & x_2 + x_3 + x_4 = 0, \quad x_5 + x_6 + x_7 = 0
\end{array}
$$

definiert.

(a) Geben Sie jeden der drei affinen Teilräume in der Form $N = P + \mathcal{L}(B)$ an, wobei $P \in \mathbb{R}^7$ und B eine Basis des zu N gehörigen Unterraumes ist, und bestimmen Sie seine Dimension.

(b) Untersuchen Sie, welche der drei Räume N_1, N_2, N_3 zueinander parallel oder schwach parallel zueinander sind.

(c) Geben Sie für alle drei möglichen Konstellationen zweier dieser Unterräume ihren Schnitt $N_i \cap N_j$ an. Wenn dieser nichtleer ist, stellen Sie den Schnitt wie in (a) dar und geben Sie zusätzlich seine Dimension an.

Hinweis: Zur Berechnung von Schnittmengen können Sie den Algorithmus aus der Vorlesung verwenden. Möglicherweise ist es aber einfacher, das entsprechende Gleichungssystem zu ermitteln, dass Punkte im affinen Schnittraum erfüllen müssen.

Aufgabe 12.68. (Linear abhängige und unabhängige Punkte):

(a) Sei V ein Vektorraum und $v, w_1, \ldots, w_{n-1} \in V$ linear unabhängig. Zeigen Sie, dass dann auch das Tupel

$$-v, \quad -v + w_1, \ \ldots, -v + w_{n-1} \in V$$

linear unabhängig ist.

(b) Sei $(A, V, +)$ ein affiner Raum. Punkte $P_0, \ldots, P_n \in A$ heißen *affin unabhängig*, wenn die Vektoren

$$\overrightarrow{P_0P_1}, \ \overrightarrow{P_0P_2}, \ \ldots, \overrightarrow{P_0P_n} \in V$$

linear unabhängig sind. Zeigen Sie mit Hilfe von (a), dass diese Definition nicht von der Reihenfolge der Punkte $P_0, \ldots, P_n$ abhängt.

Aufgabe 12.69. (Schnitte zweier affiner Geraden):

Beweisen Sie: Sei $(A, V, +)$ ein affiner Raum mit Dimension $n \geq 3$ und seien $g = P + \mathcal{L}(v)$ und $h = Q + \mathcal{L}(w)$ zwei Geraden in A mit „Richtungsvektoren" $v, w \in V$, die nicht kollinear sind. Zeigen Sie, dass g und h genau dann einen Schnittpunkt haben, wenn die Vektoren v, w und $\overrightarrow{PQ}$ komplanar sind.

Aufgabe 12.70. (Affine Hüllen):

(a) Wir betrachten den affinen Raum $\mathbb{R}^3$. Gegeben seien die vier Punkte

$$P = \begin{pmatrix} -5 \\ -11 \\ -2 \end{pmatrix}, \qquad Q = \begin{pmatrix} 5 \\ 2 \\ 4 \end{pmatrix}, \qquad R = \begin{pmatrix} -3 \\ 10 \\ 0 \end{pmatrix} \qquad \text{und} \qquad S = \begin{pmatrix} -1 \\ -7 \\ 1 \end{pmatrix}.$$

Bestimmen Sie die affine Hülle $A(\{P, Q, R, S\})$ und ihre Dimension, und geben Sie sie in der Form $T + \mathcal{L}(B)$ an, wobei $T \in \mathbb{R}^3$ und B eine Basis des zugehörigen Unterraumes ist.

(b) Sei $(A, V, +)$ ein vierdimensionaler affiner Raum. Gegeben seien eine Gerade $g \subseteq A$ und eine Ebene $\varepsilon \subseteq A$ mit

$$g = P + \mathcal{L}(v), \qquad \varepsilon = Q + \mathcal{L}(u, w).$$

Untersuchen Sie alle möglichen Lagebeziehungen von g und ε. Bestimmen Sie jeweils die Schnittmenge $g \cap \varepsilon$ und, wenn möglich, ihre Dimension, sowie die Dimension von $A(g \cup \varepsilon)$. Treffen Sie außerdem jeweils eine begründete Aussage zur Parallelitätsbeziehung zwischen g und ε.

Aufgabe 12.71. (Koordinatenumrechnungen):

(a) Sei g eine Gerade und ε eine Ebene im affinen Raum $\mathbb{K}^n$, gegeben durch

$$g = P + \mathcal{L}(\overrightarrow{PQ}), \qquad \varepsilon = P + \mathcal{L}(\overrightarrow{PQ}, \overrightarrow{PR}).$$

Hierbei sind $P, Q, R \in \mathbb{K}^n$ drei jeweils paarweise verschiedene Punkte.

(i) Begründen Sie, dass g und ε selbst affine Räume sind.

(ii) Gegeben seien für g das Koordinatensystem $K_g = (P, \overrightarrow{PQ})$, und für ε das Koordinatensystem $K_\varepsilon = (P, \overrightarrow{PQ}, \overrightarrow{PR})$. Bestimmen Sie für den Punkt $S_g := P + a \cdot \overrightarrow{PQ} \in g$ die Koordinaten bezüglich K_g, und für den Punkt $S_\varepsilon := P + a \cdot \overrightarrow{PQ} + b \cdot \overrightarrow{PR} \in \varepsilon$ die Koordinaten bezüglich K_ε.

(b) Gegeben sind im affinen Raum $\mathbb{R}^3$ die Punkte bzw. Vektoren

$$P = \begin{pmatrix} 0 \\ 6 \\ 2 \end{pmatrix}, Q = \begin{pmatrix} -10 \\ 1 \\ 17 \end{pmatrix}, O = \begin{pmatrix} 5 \\ 6 \\ 7 \end{pmatrix}, b_1 = \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}, b_2 = \begin{pmatrix} -2 \\ -1 \\ 3 \end{pmatrix}, b_3 = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$$

Bestimmen Sie die Koordinaten der Punkte P und von Q bezüglich $K = (O; b_1, b_2, b_3)$.

(c) Betrachten Sie die folgende Menge:

$$M := \{\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \mid y = 2x^2 - 12x + 16\}.$$

Bestimmen Sie ein Koordinatensystem $K = (O_K; b_1, b_2)$ von $\mathbb{R}^2$ so, dass die Koordinaten (λ_1, λ_2) der Punkte aus M die Gleichung $\lambda_2 = \lambda_1^2$ erfüllen.

13 Affine Abbildungen

Interpretiert man, wie man es intuitiv oft tut und wie wir das in der LAAG I auch schon gemacht haben, ohne das weiter zu thematisieren, den $\mathbb{R}^3$ als affinen Punktraum, so lassen sich lineare Abbildungen $f : \mathbb{R}^3 \to \mathbb{R}^3$ oft geometrisch deuten, z.B. als zentrische Streckungen, Spiegelungen oder Drehungen um den Ursprung des jeweiligen Koordinatensystems. Eine weitere solche grundlegende geometrische Abbildung, die aber nicht linear ist, ist eine Parallelverschiebung, die sich im $\mathbb{R}^3$ als Abbildung $f : \mathbb{R}^3 \to \mathbb{R}^3, v \mapsto v + t$, mit festem Verschiebungsvektor $t \in \mathbb{R}^3$ schreiben lässt. Wir wollen nun eine Klasse von Abbildungen zwischen affinen Räumen betrachten, zu denen all diese wichtigen geometrischen Abbildungen gehören und mit denen sich im $\mathbb{R}^2$, $\mathbb{R}^3$ oder auch $\mathbb{R}^n$ grundlegende geometrische Eigenschaften wie Kongruenz oder Ähnlichkeit formal definieren, untersuchen und verallgemeinern lassen.

13.1 Allgemeine Form affiner Abbildungen

Definition 13.1.1. *(Affine Abbildung)*

Es seien $(A, V, +)$ und $(A', V', +)$ affine Räume. Eine Abbildung $\alpha : A \to A'$ heißt affine Abbildung, wenn es einen Punkt $R \in A$ gibt, so dass sich das Bild jedes Punktes $P \in A$ unter α in der Form

$$\alpha : A \to A', \qquad \alpha(P) \quad = \quad \alpha(R) + f(\overrightarrow{RP}) \tag{13.1}$$

schreiben lässt, wobei f eine lineare Abbildung $f : V \to V'$ ist.

Bemerkungen dazu:

- Eine affine Abbildung $\alpha : A \to A$ der Form

  $$\alpha(P) = R + f(\overrightarrow{RP})$$

  (also mit $\alpha(R) = R$) kann man sich als „im Punkt R angebrachte Wirkung“ von f vorstellen, das ist im Bild unten angedeutet.

- Der Referenzpunkt R in der Definition oben scheint eine besondere Bedeutung zu haben. Das stimmt so nicht: Ist $R' \in A$ ein weiterer Punkt, so hat man $\alpha(R') = \alpha(R) + f(\overrightarrow{RR'})$ und daher wegen der Linearität von f für alle $P \in A$, dass

  $$\alpha(P) = \alpha(R) + f(\overrightarrow{RR'} + \overrightarrow{R'P}) = \alpha(R) + f(\overrightarrow{RR'}) + f(\overrightarrow{R'P}) = \alpha(R') + f(\overrightarrow{R'P}).$$

Die Darstellung (13.1) gilt also, wenn α affin ist, für *jeden* Referenzpunkt $R \in A$.

- Eine affine Abbildung lässt sich auch in der Form

$$\alpha : A \to A', \qquad \big(P = R + \overrightarrow{RP}\big) \;\mapsto\; \big(\alpha(P) = \alpha(R) + f(\overrightarrow{RP}\big) \tag{13.2}$$

schreiben, das ist vielleicht ein wenig intuitiver und zeigt wegen der bijektiven Beziehung $ort_R(A) = V$ auch

$$\mathrm{Bild}(\alpha) = \alpha(R) + \mathrm{Bild}(f),$$

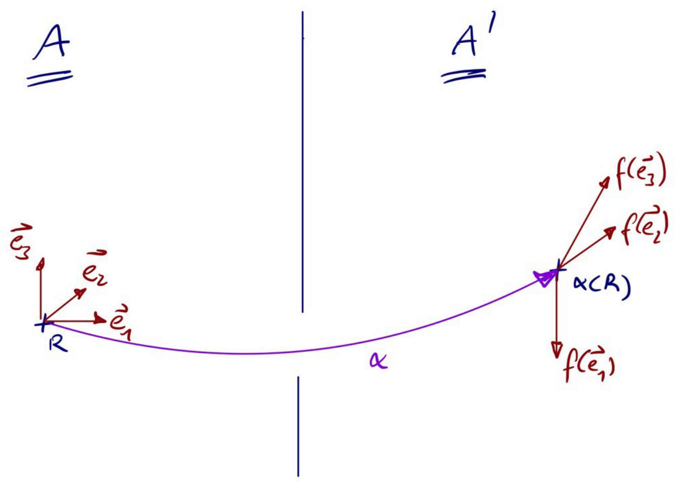
denn es ist ja

$$\mathrm{Bild}(\alpha) = \alpha(V) = \alpha(R) + f(ort_R(A)) = \alpha(R) + f(V) = \alpha(R) + \mathrm{Bild}(f).$$

- Die Darstellung (13.1) oben bedeutet, dass $f(\overrightarrow{RP})$ für alle $P, R \in A$ der Verbindungsvektor $\overrightarrow{\alpha(R)\alpha(P)}$ ist, d.h., dass α affin ist, ist äquivalent dazu, dass die Abbildung

$$f\colon V \to V', \quad \overrightarrow{RP} \mapsto \overrightarrow{\alpha(R)\alpha(P)}$$

linear ist. Diese Definition wird in einigen Lehrbüchern alternativ zur Definition 13.1.1 verwendet.



Lemma 13.1.2. *(Eigenschaften affiner Abbildungen)*

Seien A und A' affine Punkträume und $\alpha : A \to A'$ eine affine Abbildung.

(i) *Ist $M \subseteq A$ nichtleer und $P \in M$ beliebig, so ist gilt für das Bild der Punktmenge M unter α, dass*
$$\alpha(M) = \alpha(P) + f(ort_P(M)),$$
insbesondere gilt, wenn $M = P + U$ ein affiner Teilraum von A ist,
$$\alpha(M) = \alpha(P) + f(U).$$

(ii) $\mathrm{Bild}(\alpha)$ *ist ein affiner Teilraum von A'.*

(iii) *Allgemeiner bildet eine affine Abbildung affine Teilräume von A auf affine Teilräume von A' ab.*

(iv) *α ist parallelentreu, das heißt: Sind N, M parallele affine Teilräume, so sind $\alpha(N)$ und $\alpha(M)$ wieder parallel.*

(v) *Ist (A'', V'') ein weiterer affiner Raum und sind $\alpha\colon A \to A', \beta\colon A' \to A''$ affin, dann ist auch $\beta \circ \alpha\colon A \to A''$ affin und für die zugehörigen linearen Abbildungen gilt mit hoffentlich offensichtlichen Bezeichnungen $f_{\beta\circ\alpha} = f_\beta \circ f_\alpha$.*

(vi) *$\alpha : A \to A'$ ist genau dann bijektiv, wenn die zugehörige lineare Abbildung $f : V \to V'$ bijektiv ist. In diesem Fall ist die Umkehrabbildung $\alpha^{-1}\colon A' \to A$ ebenfalls affin und für die zu α und α^{-1} gehörigen linearen Abbildung gilt $f_{\alpha^{-1}} = f_\alpha^{-1}$.*

Ist $\alpha : A \to A'$ injektiv, so gilt außerdem:

(vii) $\dim(\alpha(N)) = \dim(N)$ *für alle affinen Teilräume $N \subseteq A$;*

(viii) *Sind N, M parallele affine Teilräume, so sind $\alpha(N)$ und $\alpha(M)$ wieder parallele Teilräume mit* $\dim \alpha(N) = \dim \alpha(M)$.

(ix) *Sind N, M schwach parallele affine Teilräume, so sind $\alpha(N)$ und $\alpha(M)$ wieder schwach parallel.*

Beweis: Übungsaufgabe.

□

Es wird hoffentlich klar, dass die Eigenschaften affiner Abbildungen eng mit denen der beteiligten linearen Abbildungen zusammenhängen. Wir merken hier nur an, dass sich weiter gehend zeigen lässt, dass eine affine Abbildung die beteiligte lineare Abbildung immer eindeutig festlegt; umgekehrt reicht die Angabe einer linearen Abbildung und eines Bildpunktes $\alpha(R)$ aus, um die affine Abbildung genauer festzulegen. Aus der LAAG I wissen wir, dass eine lineare Abbildung durch die Angabe von $n = \dim(V)$ Bildvektoren $f(v_1), \dots, f(v_n)$ für linear unabhängige Vektoren $v_1, \dots, v_n$ vollständig festgelegt ist. Analog kann man affine Unabhängigkeit von <u>Punkten in affinen Räumen</u> definieren: $P_1, \dots, P_n$ mit $N \geq 2$ heißen linear unabhängig, wenn die Vektoren $\overrightarrow{P_1P_2}, \overrightarrow{P_1P_2}, \overrightarrow{P_1P_2}$ l.u. sind. Daraus ergibt sich dann, dass affine Abbildungen durch die Angabe der Bildpunkte von $\dim(V) + 1$ affin unabhängigen Punkten in A festgelegt sind. Wir gehen hier aber nicht weiter ins Detail.

13.2 Affine Selbstabbildungen

Ab jetzt betrachten wir ausschließlich den wichtigsten Anwendungsfall $A = A'$, also affine Abbildungen *eines* affinen Raumes A „auf sich selbst“. Speziellere Fälle hiervon sind affine Selbstabbildungen $\alpha : \mathbb{K}^n \to \mathbb{K}^n$, noch spezieller diejenigen mit $\mathbb{K} = \mathbb{R}$, $n = 2, 3$. Die letzteren sind das anschauliche „Bread & Butter“ der affinen Geometrie, d.h. das, wo diese herkommt und wofür sie hauptsächlich benutzt wird: Die Welt der geometrischen Abbildungen in Ebene und Raum. Wir werden daher bald viel über diese geometrisch interpretierbaren affinen Abbildungen reden, wir bleiben aber erstmal beim allgemeinen Fall.

Definition 13.2.1. *(Affine Selbstabbildungen, Affinitäten, Fixpunkt)*

- *Eine Abbildung $\alpha : A \to A$ von einem affinen Raum A in sich selbst nennt man affine Selbstabbildung.*
- *Eine bijektive affine Abbildung $\alpha : A \to A$ von einem affinen Raum in sich selbst nennt man Affinität von A.*
- *Ein Punkt $P \in A$ mit $\alpha(P) = P$ heißt Fixpunkt von α.*
- *Sei $t \in V$. Dann heißt die affine Abbildung*

  $$\alpha : A \to A, \qquad \alpha(P) = P + t$$

  Verschiebung oder Translation um den Verschiebungsvektor t.

Einige Beobachtungen, um mit diesen Definitionen warm zu werden:

Bemerkungen dazu:

- Nicht jede affine Abbildung hat Fixpunkte; zum Beispiel hat jede Translation um $t \neq 0_V$ keine.
- Hat eine Abbildung Fixpunkte, so ist die Menge aller Fixpunkte von f ein affiner Teilraum von A, das ist eine nette Übungsaufgabe.
- Ist $\alpha : A \to A$ affin und Z ein Fixpunkt der affinen Abbildung, so ergibt sich für alle $P \in A$ die Abbildungsvorschrift

  $$\overrightarrow{Z(\alpha(P))} = f(\overrightarrow{ZP}), \quad \text{d.h.} \quad \alpha(P) = Z + f(\overrightarrow{ZP}) \tag{13.3}$$

- Betrachten wir einen Vektorraum V als affinen Raum $(V, V, +)$ über sich selbst, so lässt sich jede lineare Abbildung $f : V \to V$ als affine Abbildung

$$\alpha_f : V \to V, \qquad x \mapsto 0_V + f(\overrightarrow{0_V x}) = f(x)$$

auffassen. Diese hat dann immer zumindest den Fixpunkt 0_V.

- Eine Translation ist eine affine Abbildung mit linearem Anteil $f = \mathrm{id}_V$.

Affine Abbildungen in Vektorräumen und insbesondere im $\mathbb{K}^n$ lassen sich immer als „lineare Abbildung plus Verschiebung" deuten:

Lemma 13.2.2. *(Standarddarstellung für affine Abbildungen $\alpha : V \to V$)*

Wir betrachten einen Vektorraum V als affinen Raum $(V, V, +)$ über sich selbst. Dann sind die affinen Abbildungen $\alpha : V \to V$ genau diejenigen Abbildungen, die sich in der Form

$$\alpha : V \to V, \qquad x \mapsto f(x) + t \tag{13.4}$$

mit linearem $f : V \to V$ und $t \in V$ schreiben lassen.

Ist $V = \mathbb{K}^n$, so bedeutet das: Affine Abbildungen im $\mathbb{K}^n$ sind genau die Abbildungen der Form

$$\alpha : \mathbb{R}^n \to \mathbb{R}^n, \qquad x \;\mapsto\; A \cdot x + t \tag{13.5}$$

mit $A \in \mathbb{K}^{n\times n}, t \in \mathbb{K}^n$.

Bemerkung dazu:

- Insbesondere sind die in Lemma 12.5.2 hergeleiteten Koordinatentransformationen von der Form 13.5und sind somit affine Selbstabbildungen des $\mathbb{K}^n$.

Beweis: Hat α die angegebene Form und nimmt man $0_V \in V$ als Referenzpunkt, so ist $\alpha(0_V) = t$, so hat α die Darstellung

$$\alpha(x) = t + f(x) = \alpha(0_V) + f(\overrightarrow{0_V x}),$$

also die Form 13.1 einer affinen Abbildung. Umgekehrt gilt die Darstellung 13.1 ja für jeden Referenzpunkt $r \in V$; wählt man $r := 0_V$ und $t := \alpha(0_V)$, so ergibt sich gerade die obige Darstellung 13.4. Da die linearen Abbildungen im $\mathbb{K}^n$ gerade die Matrixabbildungen, also Abbildungen der Form $x \mapsto A \cdot x$ mit $A \in \mathbb{K}^{n \times n}$ sind, ergibt sich auch die Darstellung für $\alpha : \mathbb{R}^n \to \mathbb{R}^n$.

□

13.3 Affine Abbildungen in $\mathbb{R}^n$

13.3.1 Beispiele: Geometrische Abbildungen in $\mathbb{R}^n$

Neben der oben eingeführten Translationsabbildung $\alpha : \mathbb{R}^n \to \mathbb{R}^n, x \mapsto x + t$ um einen Vektor $t \in \mathbb{R}^n$ gibt es viele andere geometrische Abbildungen, die sich mit Hilfe affiner Abbildungen darstellen lassen. Einige Beispiele:

- Sei $z \in \mathbb{R}^n$ und $s \in \mathbb{R} \backslash \{0\}$. Die zentrische Streckung

  $$x \mapsto z + s \cdot \overrightarrow{zx}$$

  mit Streckzentrum Z und Streckfaktor s ist eine affine Abbildung. Will man für sie den Nullvektor $=_V \in \mathbb{K}^n$ als Referenzpunkt nehmen, so besitzt sie die nicht ganz so elegante Darstellung

  $$x \quad \mapsto \quad (1-s) \cdot \overrightarrow{O_V Z} \; + \; s \cdot x$$

  (Übungsaufgabe!)

- Eine Spiegelung an der x-Achse im $\mathbb{R}^2$ beziehungsweise an der $x - y$-Ebene im $\mathbb{R}^3$ lässt sich mit Hilfe der affinen Abbildungen

  $$P_x : \mathbb{R}^2 \to \mathbb{R}^2, \quad x \mapsto \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \cdot x, \qquad P_{(x,y)} : \mathbb{R}^3 \to \mathbb{R}^3, \qquad x \mapsto \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot x$$

  beschreiben.

- Eine Projektion auf die x-Achse des $\mathbb{R}^2$ beziehungsweise auf die $x-y$-Ebene des $\mathbb{R}^3$ lässt sich über

$$S_x : \mathbb{R}^2 \to \mathbb{R}^2, \qquad x \mapsto \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \cdot x, \qquad S_{(x,y)} : \mathbb{R}^3 \to \mathbb{R}^3, \qquad x \mapsto \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \cdot x$$

darstellen.

- Eine Scherung in der Ebene wird in $\mathbb{R}^2$ beispielsweise durch eine Abbildung

$$\sigma_2 : \quad \mathbb{R}^2 \to \mathbb{R}^2, \qquad x \mapsto \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \cdot x$$

beschrieben; dabei wird der Einheitsvektor in x-Richtung fest gelassen und der Einheitsvektor e_2 „geschert“. Allgemein kann man im $\mathbb{R}^n$ als Scherung in Richtung $e_1, \ldots, e_k$ jede Matrixabbildung

$$\sigma_n : \mathbb{R}^n \to \mathbb{R}^n, \quad x \mapsto \begin{pmatrix} I_k & A \\ 0 & I_\ell \end{pmatrix}$$

mit $k + \ell = n$ und $A \in \mathbb{R}^{k \times \ell}$ bezeichnen; hierbei bleibt der Spann der ersten k Standardbasisvektoren fest, und die anderen werden in Richtung $\mathcal{L}(e_1, \ldots, e_k)$ „geschert“. Die Begriffe hierzu werden in der Literatur für gewöhnlich nicht einheitlich gehandhabt; eine Scherung ist aber immer linear, d.h. eine affine Abbildung mit $t = 0$.

- Eine Skalierung der Achsen des Weltkoordinatensystems W ist eine affine Abbildung

$$s : \mathbb{R}^n \to \mathbb{R}^n, \qquad x \quad \mapsto \begin{pmatrix} \lambda_1 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 \\ 0 & 0 & \lambda_3 & \ldots & 0 \\ \vdots & \vdots & 0 & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \lambda_n \end{pmatrix} \cdot x$$

mit einer Diagonalmatrix mit Diagonaleinträgen $\lambda_1, \ldots, \lambda_n > 0$ (wobei die Achsen durch $\lambda_i \neq \lambda_j$ auch verschieden skaliert werden können).

- Zur Illustration betrachten wir in der Vorlesung einige weitere Beispiele geometrischer Abbildungen in Ebene und Raum (aus [Filler], S. 281-284 und Beispiele 7.28-7.31 und aus [FaHa]).

„Font Art“

Aus: Farin/Hansford, „Lineare Algebra“

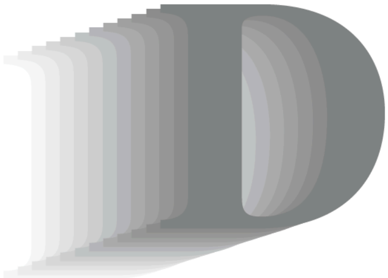


Abb. 6.3. Translationen: Der Buchstabe **D** wird mehrfach bewegt.

Abb. 6.4. Skalierung: der Buchstabe **D** wird mehrfach skaliert; der Ursprung liegt im Mittelpunkt.

Abb. 6.5. Skalierung: der Buchstabe **D** wird mehrfach skaliert; nach jedem Schritt wurde zusätzlich eine Translation angewendet.

„Font Art“

Aus: Farin/Hansford, „Lineare Algebra“

Abb. 6.6. Rotationen: Der Buchstabe **S** wird mehrfach rotiert; der Ursprung liegt unten links am Buchstaben.

Abb. 6.7. Rotationen: Skalierungen und Translationen werden ebenfalls angewendet.

13.3.2 Eine „Blaupause“ zur Konstruktion geometrischer Abbildungen in $\mathbb{R}^n$

Der vorangegangene Abschnitt zeigt, dass die die Konstruktion affiner Abbildungen einfach ist, wenn „die Musik entlang der Achsen spielt“, d.h. wenn Skalierungsrichtungen, Spiegel-, Scherungs- und Projektionsebenen so gewählt werden, dass sie entlang der Achsen liegen. Normalerweise, zum Beispiel in grafischen Anwendungen, ist das nicht so; wir haben ein Weltkoordinatensystem $W = (O, \vec{e_1}, \vec{e_2}, \vec{e_3})$ gegeben, aber die an der Abbildung beteiligten Achsen liegen irgendwie im Raum.

Hier helfen die in Lemma 12.5.2 entwickelten Koordinatentransformationen. Das generelle Procedere zur Berechnung einer affinen Abbildung (z.B. Skalierung, Scherung, Drehung, Projektion, Spiegelung; vgl. auch Abschnitt 15) ist das Folgende:

- Finde ein Koordinatensystem $K = (v, B)$, so dass sich die gewünschte affine Abbildung α möglichst einfach darstellen lässt, d.h.:
    - Wähle, wenn möglich, als neuen Ursprung v einen Fixpunkt von α und „ausgezeichnete Richtungen“ der Abbildung α (Streck- oder Spiegelachsen, Projektionsflächen, …) als Achsen des neuen Koordinatensystems.
- Transformiere die Koordinaten abzubildender Punkte P, die als $k_W(P)$ gegeben sind, auf das Koordinatensystem $K = (v, B)$, vgl. dafür Lemma 12.5.2.
- Wende die gewünschte affine Abbildung α auf die transformierten Koordinatenvektoren an. Diese hat dann die Form

  $$\alpha : \mathbb{R}^n \to \mathbb{R}^n, \qquad x \ \mapsto \ A_{B,B} \cdot x + t,$$

  wobei $A_{B,B}$, die darstellende Matrix der zu α gehörigen linearen Abbildung f ist. Für gewöhnlich hat $A_{B,B}$ dann, wie in Abschnitt 13.3.1 demonstriert, eine einfache Form. Für den Fall, dass der Ursprung ein Fixpunkt ist, ist zusätzlich $t = 0$.
- Transformiere „zurück in die Welt“ mit der Abbildung k_K^{-1} aus Lemma 12.5.2.

Insgesamt ergibt sich für so als Abbildungsvorschrift für die „im Koordinatensystem $K = (v, B)$ ausgeführte“ Matrixmultiplikation $A_{B,B}$ die affine Abbildung

$$\boxed{x \ \mapsto \ (BA_{B,B}B^{-1})(x - v) \ + \ (B \cdot t + v).}$$

Die Matrix $BA_{B,B}B^{-1}$ ist gerade die darstellende Matrix der linearen Abbildung f bezüglich Standardbasis, also die Abbildungsmatrix von f.

13.3.3 Axonometrien

Eine *Axonometrie* nennt man in Anwendungen wie im Bauingenieurswesen (vgl. z.B. [Hartm]) jede affine Abbildung α des Raumes in die Ebene. Viele gängige Möglichkeit zur Erstellung perspektivischer Zeichnungen (also zur Darstellung räumlicher Gegenstände und Phänomene in 2D) sind Axonometrien. Das Vorgehen zur Erstellung solcher Axonometrien ist das folgende:

- Man führt im Raum ein geeignetes Koordinatensystem $K = (O, \vec{e}_1, \vec{e}_2, \vec{e}_3)$ mit Ursprung O und Basis (also „Koordinatenachsen") $\vec{e}_1, \vec{e}_2, \vec{e}_3 \in \mathcal{A}_3$ ein, das für gewöhnlich kartesisch (rechtwinklig) ist und ein Rechtssystem bildet, d.h. ihre Orientierung entspricht der eines Koordinatensystems aus Daumen, Zeigefinger und Mittelfinger der rechten Hand.

- In der Ebene $\mathcal{A}_2$ gibt man nun das Bild $\alpha(O)$ des Ursprungs O und die Bilder der drei Achsen beliebig vor – wir bezeichnen die Bilder mit $\overline{O}, f(\vec{e}_1), f(\vec{e}_2), f(\vec{e}_3)$. Das axonometrische Bild eines Punktes $P = (x, y, z)$ der Ebene ist durch die affine Abbildung

  $$\alpha : \mathcal{A}_3 \to \mathcal{A}_2, \quad \alpha(P) := \overline{O} + x \cdot f(\vec{e}_1) + y \cdot f(\vec{e}_2) + z \cdot f(\vec{e}_3)$$

  gegeben, Bilder unter α lassen sich bequem in der Ebene darstellen.

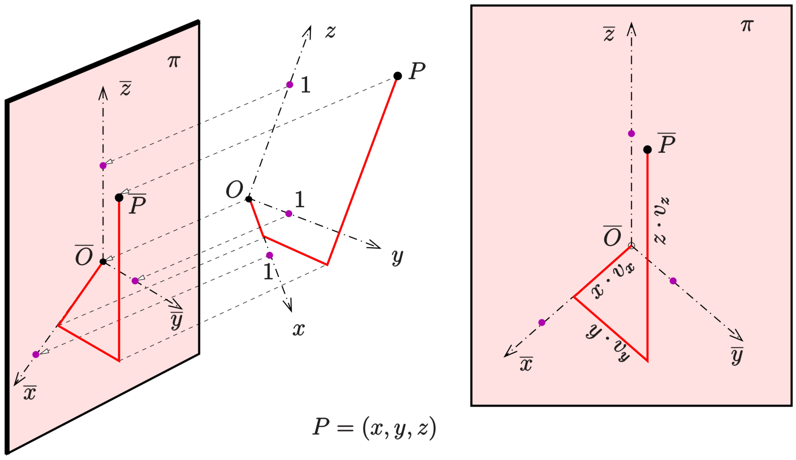
- Die Längen der Bildvektoren $f(\vec{e}_1), f(\vec{e}_2), f(\vec{e}_3)$ kann man als Maßstäbe der Achsen oder auch als Verzerrungsverhältnisse

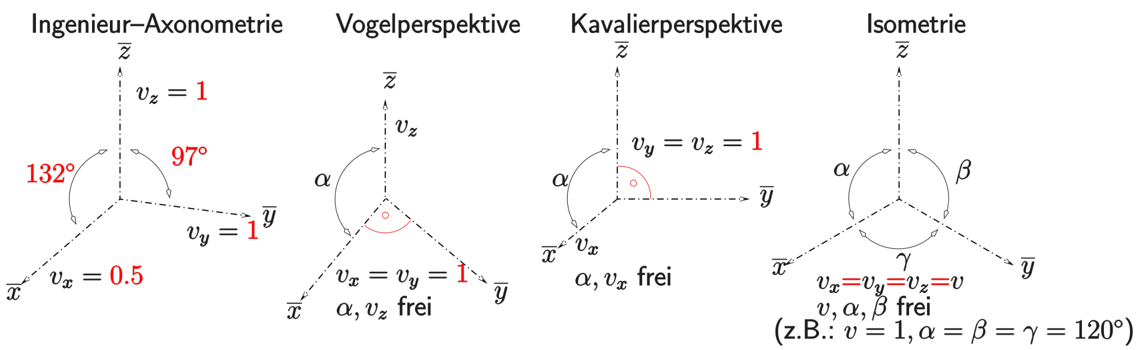
  $$v_x := \|f(\vec{e}_1)\|, \quad v_y := \|f(\vec{e}_2)\|, \quad v_z := \|f(\vec{e}_3)\|$$

  verstehen; $\|f(\vec{e_i})\|$ bezeichnet hier jeweils die Länge der projizierten Achsen.

- Die zugehörigen Abbildungen lassen sich bei gegebenen Koordinatensystemen auch mit Hilfe zugehöriger Koordinaten angeben: Beispielsweise ist der Grundriss eine Axonometrie mit $\alpha(O) = (0,0)$ und $f(e_1) = (1,0), f(e_2) = (0,1), f(e_3) = (0,0)$.

Bettet man die Projektionsebenen im $\mathbb{R}^3$ ein, d.h sie sind affine Unterräume des Raumes, von dem aus projeziert wird, so ergeben sich so genannte Parallelprojektionen wie auf den Bildern weiter unten.





(beide Grafiken: Wikipedia „Axonometrien“, Ag2gaeh - Eigenes Werk)

X.6.4 Beispiele affiner Abbildungen in $\mathbb{R}^n$, Teil 2 – Axonometrien

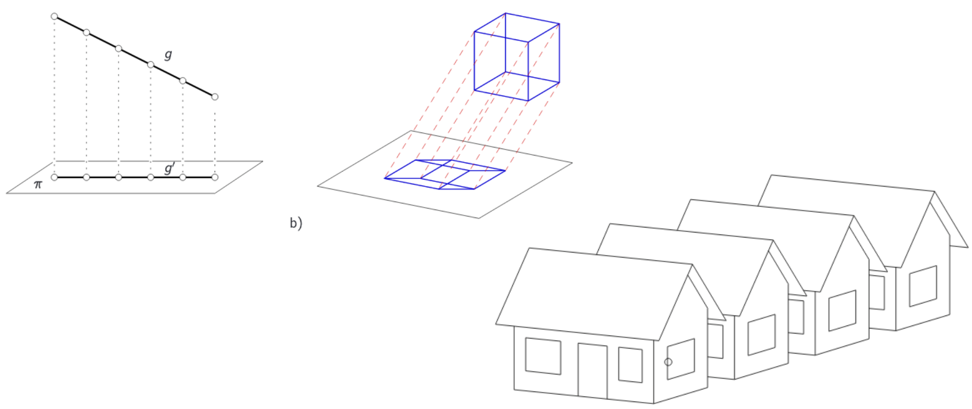


Abbildung 1.5: Häuser in Parallelprojektion: Parallelitäten und Teilverhältnistreue sind z.B. an den Fenstern ablesbar

Aus: Erich Hartmann, Darstellende Geometrie für Bauingenieure

13.3.4 Computergrafik, Level 1 bis 4

Mit den in diesem Abschnitt 13.3 vorgestellten Bauteilen zum Umgang mit geometrischen Abbildungen haben wir einige der Grundbausteine für die Erzeugung von 3D-Computergrafiken gesehen. Ein mögliches Vorgehen zur Erzeugung eines zweidimensionalen Ausgabe wollen wir uns zum Abschluss dieses geometrischen Kapitels skizzieren:

- Dreidimensionale geometrische Objekte wie eine Kugeloberfläche werden oft dadurch angenähert, dass eine „Punktwolke“ im $\mathbb{R}^3$ erzeugt wird, die aus einigen Punkten des Objekts besteht. Je mehr Punkte, dest besser die Darstellung, aber auch: Desto höher der Rechenaufwand im allem, was folgt.

- Solche Punktwolken werden meist aus Parameterdarstellungen der Objekte erzeugt: Eine Kugel vom Radius 1, die im Ursprung sitzt, ist etwa gegeben durch

  $$x = \sin\theta \cdot \cos\varphi, \quad y = \sin\theta \cdot \sin\varphi, \quad z = r \cdot \cos\theta \quad \text{mit} \quad 0 \leq \theta \leq \pi, 0 \leq \varphi < 2\pi.$$

  Durch Einsetzen einiger Werte aus dem gegebenen Intervall wird eine die Kugel beschreibende Punktwolke im $\mathbb{R}^3$ erzeugt.

- Die in den Abschnitten 13.3.2, 13.3.1 vorgestellten Mitteln zur einfachen Anwendung von geometrischen Abbildungen können nun auf jeden der Punkte dieser Punktwolke im $\mathbb{R}^3$ angewandt werden, um die Kugel zu strecken, zu stauchen, auf einer vorgegebenen Bahn fliegen zu lassen,...

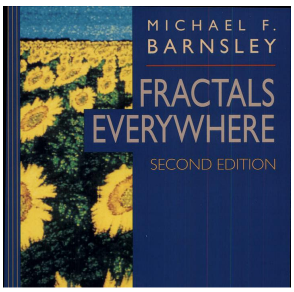
- Es gibt viele, viele Techniken, aus dem so generierten Zeitablauf dann versiertere Optiken zu erzeugen. Schattenwurf beispielsweise ließe sich prinzipiell mit Hilfe von Projektionstechniken erzeugen; auch die Reflexion von Lichtstrahlen („Ray-Tracing“) lässt sich mit Mitteln der Linearen Algebra erzeugen.

- Schließlich wird das dreidimensionale Objekt in eine Betrachtungsebene „herunterprojiziert“, dabei kann das Koordinatensystem durchaus zeitabhängig sein, sich also bewegen oder bewegbar sein. Zum Beispiel geht das mit den Mitteln der Axonometrie. Eine andere Form von Projektion, die wir hier nicht behandelt, die aber auch in Kunst und Technik häufig verwandt wird, ist die so genannte Zentralprojektion.

Seit den ersten Tagen ernstzunehmender Beschäftigung mit Computergrafik sind nun einige Jahrzehnte ins Land gegangen, und es gibt viele, viele Techniken, versiertere Optiken zu erzeugen. Zum Beispiel werden Pflanzen in moderneren Computerspielen nicht mehr wie hier erzeugt, sondern eher über so genannte Lindenmeyer-Systeme (oder kurz „L-Systeme"), die in das Gebiet der fraktalen Geometrie fallen. Eine ausführliche Darstellung all dieser Techniken ginge hier viel zu weit. Ich hoffe aber, das an der oben gegebenen Skizze klar wird: Dieses Vorgehen ist ein möglicher Startpunkt, von dem aus solche Grafikdarstellungen prinzipiell ermöglicht werden.

Zum Abschluss dieser Illustrationen betrachten wir last but not least noch ein spezielles dieser eben erwähnten Objekte aus der fraktalen Geometrie, weil es einen schönen Schlussbogen zu den affinen Transformationen spannt: Der so genannte „Barsley-Farn" ist eine selbstähnliche Figur, die durch die iterierte (also mehrfache) Anwendung affiner Transformationen entsteht. Er ist z.B. hier zu bewundern: (Link), seine Erzeugung ist als GeoGebra-Applet hier zu sehen: (Link).

Der „Barnsley-Farn"



Urheber: „Farry" auf Wikipedia

13.4 Aufgaben zu Kapitel 13

Aufgabe 13.72. (Eigenschaften affiner Abbildungen):

Seien A, A' affine Räume. Beweisen Sie die fehlenden Aussagen für affine Abbildungen $\alpha : A \to A'$ aus Lemma 13.1.2 aus der Vorlesung:

(ii) Bild(α) ist ein affiner Teilraum von A'.

(iii) Allgemeiner bildet eine affine Abbildung $\alpha : A \to A'$ affine Teilräume von A auf affine Teilräume von A' ab.

(iv) Sind N, M parallele affine Teilräume von A, so sind $\alpha(N)$ und $\alpha(M)$ wieder parallele Teilräume in A' („Parallelentreue affiner Abbildungen").

(v) Ist (A'', V'') ein weiterer affiner Raum und sind $\alpha\colon A \to A', \beta\colon A' \to A''$ affine Abbildungen, dann ist auch $\beta \circ \alpha\colon A \to A''$ affin und für die zugehörigen linearen Abbildungen gilt mit hoffentlich offensichtlichen Bezeichnungen $f_{\beta\circ\alpha} = f_\beta \circ f_\alpha$.

(vii) Ist $\alpha : A \to A'$ injektiv, so gilt $\dim(\alpha(N)) = \dim(N)$ für alle affinen Teilräume von A.

Aufgabe 13.73. (Affine Abbildungen der Ebene):

(a) Skizzieren Sie in der Anschauungsebene die Geradenschar

$$G = \{g_n \mid n \in \mathbb{N}\} \text{ mit } g_n = \begin{pmatrix} n \\ 0 \end{pmatrix} + \mathcal{L}(\begin{pmatrix} n \\ 2n \end{pmatrix})$$

und ihr Bild unter der affinen Transformation

$$\alpha : \mathbb{R}^2 \to \mathbb{R}^2, \qquad x \mapsto \begin{pmatrix} \sin\frac{\pi}{4} & -\cos\frac{\pi}{4} \\ \cos\frac{\pi}{4} & \sin\frac{\pi}{4} \end{pmatrix} \cdot x.$$

(b) Wir betrachten den affinen Raum $A = \mathbb{R}^2$ und in ihm die Figur

$$\mathbb{P} := \left\{(x, y) \in \mathbb{R}^2 \mid |x| + |y| \leq 1\right\}.$$

In der Vorlesung wurden als spezielle affine Abbildungen die Translationen

$$t_v : A \to A, \quad P \mapsto P + v$$

(mit $v \in V$ als Translationsvektor) und die zentrischen Streckungen

$$\sigma_{Z,s} \colon A \to A,\ P \mapsto Z + s \cdot \overrightarrow{ZP}$$

(mit Streckzentrum $Z \in A$ und Streckfaktor $s \in \mathbb{K}\backslash\{0\}$) betrachtet.

Skizzieren Sie die folgenden Mengen in der Anschauungsebene:

(i) $\mathbb{P}$

(ii) $t_v(\mathbb{P})$ für $v = \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$.

(iii) $\sigma_{Z,s}(\mathbb{P})$ für $s = 2$ und $Z = \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right)$,

(iv) $(\sigma_{Z,s} \circ t_v)(\mathbb{P})$ für $s = -1, Z = \left(\begin{smallmatrix} 0{,}5 \\ 0{,}5 \end{smallmatrix}\right)$ und $v = \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right)$.

(v) $(t_v \circ \sigma_{Z,s})(\mathbb{P})$ für $s = -1, Z = \left(\begin{smallmatrix} 0{,}5 \\ 0{,}5 \end{smallmatrix}\right)$ und $v = \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right)$.

Bitte nutzen Sie für jede Teilaufgabe eine eigene Skizze!

Aufgabe 13.74. (Hyperebenen):

Sei $(A, V, +)$ ein n-dimensionaler affiner Raum mit $n \geq 1$, und N eine Hyperebene von A.

(a) Beschreiben Sie in den Spezialfällen $A = \mathbb{R}^2$ und $A = \mathbb{R}^3$ alle affinen Unterräume, die N nicht schneiden (ohne Beweis).

(b) Beweisen Sie, dass jeder affine Teilraum $M \subset A$, der N nicht schneidet, zu N parallel oder schwach parallel ist.

Aufgabe 13.75. (Geometrische Abbildungen in $\mathbb{R}^2$):

Wir betrachten den $\mathbb{R}^2$ wie üblich als affinen Raum über dem Vektorraum $\mathbb{R}^2$. Gegeben sei neben dem „Weltkoordinatensystem“

$$W = (O; e_1, e_2) \qquad \text{mit} \qquad O = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

das Koordinatensystem

$$K = (O'; b_1, b_2) \qquad \text{mit} \qquad O' = \begin{pmatrix} 3 \\ 2 \end{pmatrix}, \qquad b_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \qquad b_2 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$

(a) Ermitteln Sie die expliziten Zuordnungsvorschriften für die Koordinatentransformation $k_K \colon \mathbb{R}^2 \to \mathbb{R}^2$, die Koordinaten bezüglich W auf Koordinaten bezüglich K umrechnet, sowie für ihre Inverse.

(b) Wir betrachten nun den $\mathbb{R}^2$ mit Weltkoordinatensystem W. Gegeben seien für $i = 1, 2$ die Abbildungen $\alpha_i \colon \mathbb{R}^2 \to \mathbb{R}^2$, die durch

$$x \ \mapsto \ k_K^{-1}(A_i \cdot (k_K(x)),$$

definiert sind. Hierbei sind die Matrizen A_1, A_2 durch

$$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad \text{und} \qquad A_2 = \begin{pmatrix} \frac{3}{2} & 0 \\ 0 & -1 \end{pmatrix}$$

gegeben. Geben Sie die Abbildungsvorschriften der affinen Abbildungen $\alpha_1, \alpha_2 \colon \mathbb{R}^2 \to \mathbb{R}^2$ in der Form

$$\alpha_i : x \mapsto C_i \cdot x + t$$

mit geeigneten Matrizen $C_i \in \mathbb{R}^{2\times 2}$ an.

(c) Skizzieren Sie die Geraden

$$g = O' + \mathcal{L}(b_1) \qquad \text{und} \qquad h = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \mathcal{L}(\begin{pmatrix} 1 \\ -1 \end{pmatrix})$$

und den Kreis

$$K = \big\{(x, y) \in \mathbb{R}^2 \mid (x-3)^2 + (y-2)^2 = 1\big\},$$

und veranschaulichen Sie in zwei weiteren Zeichnungen die Wirkung der affinen Abbildungen α_1, α_2 auf diese geometrischen Objekte.

Aufgabe 13.76. (Ingenieur-Axonometrie):

Im Bild nebenan sehen Sie den Grundriss und den Aufriss eines Turms im Anschauungsraum, d. h. die Bilder des Turms unter den beiden Axonometrien, die als Koordinatenabbildungen π_1, π_2 mit

$$\pi_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \pi_1 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad \pi_1 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

und

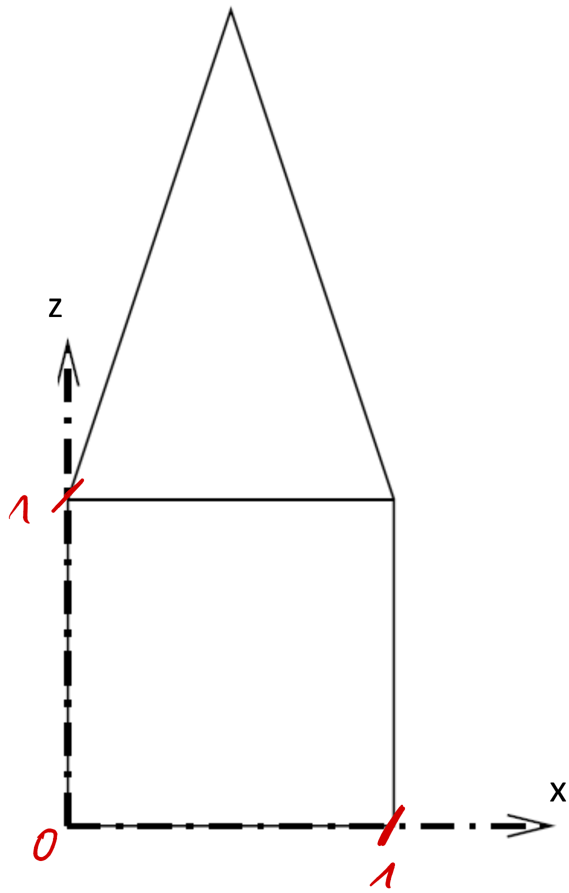
$$\pi_2 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \pi_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad \pi_2 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

beschrieben werden.

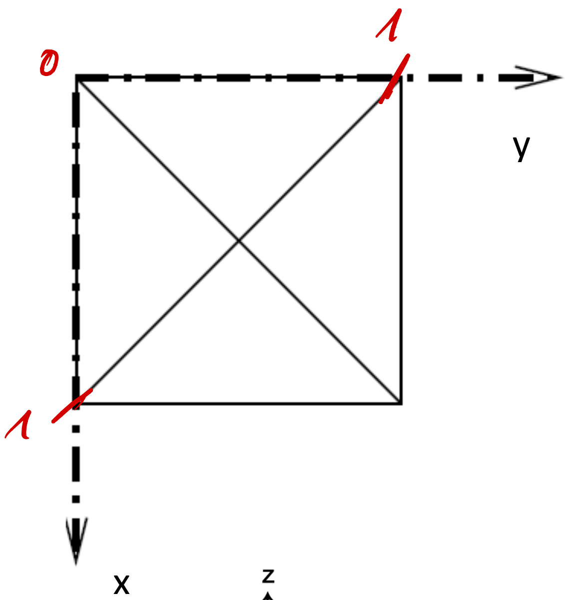
Skizzieren Sie das Bild dieses Turms unter der ebenfalls dort angegebenen Ingenieur-Axonometrie. Die Bilder der x-, y- und z-Achse unter dieser Axonometrie sowie die relativen Kantenlängen des Turms entnehmen Sie bitte der Vorlesung.

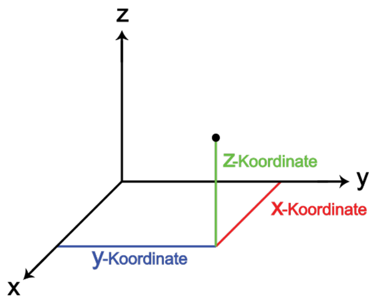
Zu Aufgabe 5.1

Aufriss:



Grundriss:

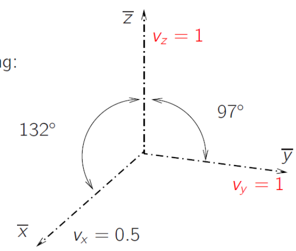




2.3 Ingenieur–Axonometrie

Dies ist eine einfach zu erstellende Axonometrie mit guter Wirkung:

1. Die Verzerrungen sind $v_x = 0.5$, $v_y = v_z = 1$.
2. In der Projektion ist der
   Winkel zwischen der z–Achse und der x–Achse 132°,
   Winkel zwischen der z–Achse und der y–Achse 97°.



Die Vorteile der Ingenieur–Axonometrie sind:

- Durch die einfachen Verzerrungsverhältnisse ist sie leicht zu erstellen.
- Das axonometrische Bild ist nahezu eine um den Faktor 1.06 skalierte **senkrechte** Parallelprojektion.
- Die hierzu notwendigen Winkel von 7° und 42° sind auf vielen Geodreiecken markiert.
- Der Umriss einer Kugel ist in guter Näherung ein Kreis.

Aus: Erich Hartmann, Darstellende Geometrie für Bauingenieure

14 Skalarprodukträume

Unser Leitgedanke in diesem zweiten Teil dieses Skripts war es bisher, die Geometrie der Ebene und des Raumes als Motivation zu nehmen, diese dann in affinen Räumen „abzubilden“ und mit Hilfe der Vektorräume und affinen Räume zu verallgemeinern. Dabei bleiben wir: Zu einer „vollständigen“ Geometrie nach Euklid fehlt noch ein Konzept der Längen- und Winkelmessung. Das entscheidende Hilfsmittel auf der formalen Ebene ist die Einführung eines Skalarproduktes in den betrachteten *Vektorräumen*. Das ist das Thema dieses Kapitels; auf affine Räume wird das Konzept dann im folgenden Kapitel übertragen.

14.1 Einführung und erste Grundbegriffe

Zur Motivation des Skalarproduktbegriffs schauen wir uns einige elementargeometrische Situationen an (Vorlesung).

- In Anschauungsraum leitet man her, dass die euklidische Länge eines Vektors $x \in \mathbb{R}^3$ als
$$\|x\| := \sqrt{x_1^2 + x_2^2 + x_3^2},$$
definiert werden sollte, wenn das mit der Länge eines entsprechenden Repräsentantenpfeils der Pfeilklasse $\vec{x}$ übereinstimmen soll.

- Aus dem Satz des Pythagoras und seiner Umkehrung ergibt sich, dass zwei Vektoren $\vec{x}, \vec{y}$ im Anschauungsraum senkrecht stehen, genau dann, wenn für ihre Koordinatenvektoren $x, y \in \mathbb{R}^3$ die Größe
$$\langle x, y \rangle := x_1y_1 + x_2y_2 + x_3y_3 \tag{14.1}$$
Null ist. Diese Größe nennt man das euklidische Skalarprodukt der Vektoren x, y im $\mathbb{R}^3$.

- Wählt man für zwei Vektoren $\vec{x}, \vec{y}$ Repräsentanten mit gleichen Startpunkt P, so schließen die Vektoren in der Ebene $\varepsilon : P + \mathcal{L}(\vec{x}, \vec{y})$ einen Winkel ein. Für diesen lässt sich (z.B mit Hilfe des Kosinussatzes, als Verallgemeinerung des letzten Punktes) beweisen, dass
$$\cos(\gamma) = \frac{\langle x, y \rangle}{\|x\| \|y\|}$$
gilt. (Hierbei sind x und y wieder Koordinatenvektoren von $\vec{x}$ und $\vec{y}$.) Mit dem Skalarprodukt lassen sich *aus Koordinaten von Ortsvektoren* also *Winkel zwischen*

diesen Vektoren berechnen! Über

$$\gamma = \arccos\left(\frac{\langle x, y\rangle}{\|x\| \cdot \|y\|}\right)$$

ergibt sich ein Winkel $\gamma \in [0, \pi]$, der dem kleineren der beiden Winkel zwischen $\vec{x}$ und $\vec{y}$ entspricht.

- Setzt man in (14.1) $x = y$, so ergibt sich

$$\langle x, x\rangle \;=\; x_1^2 + x_2^2 + x_3^2 = \|x\|^2, \quad \text{also} \quad \|x\|_2 = \langle x, x\rangle^{\frac{1}{2}}.$$

  Umgekehrt interpretiert definiert das Skalarprodukt dadurch also durch den Fall $x = y$ eine Längenmessung $\|.\| : \mathbb{R}^3 \to \mathbb{R}_0^+$, eine so genannte Norm für Vektoren.

- In der Anschauungsebene definiert die dadurch induzierte Funktion

$$d : \mathcal{A} \times \mathcal{A} \to \mathbb{R}, \quad (P, Q) \mapsto d(P, Q) := \|\overrightarrow{PQ}\|$$

  eine Abstandsmessung oder Metrik auf den Punkten des Anschauungsraumes.

Wir werden im nächsten Abschnitt den Begriff des Skalarproduktes für allgemeine Vektorräume definieren und später zeigen, das alle der oben benannten Eigenschaften im allgemeinen Fall erhalten bleiben, d.h. dass wir das Skalarprodukt einsetzen können um Winkel zwischen Vektoren, Längen von Vektoren und Abstände der Punkte eines affinen Punktraumes zu definieren.

Dazu wollen wir zunächst als „Leitbeispiel“ ansehen, welche Eigenschaften des obigen euklidischen Skalarprodukts wie an die induzierte Norm weitergegeben werden.

Lemma 14.1.1. *(Euklidische Norm auf $\mathbb{R}^3$, Eigenschaften)*

Die von (14.1) induzierte euklidische Länge $\|\cdot\|_2 : \mathbb{R}^3 \to \mathbb{R}$ ist eine Norm, d.h. es gilt:

(i) $\|x\| \geq 0; \qquad \|x\| = 0 \Leftrightarrow x = 0$, *für alle* $x \in \mathbb{R}^3$

(ii) $\|a \cdot x\| = |a| \cdot \|x\|$ *für alle* $x \in \mathbb{R}^3$ *und* $a \in \mathbb{R}$,

(iii) $\|x + y\| \leq \|x\| + \|y\|$ *für alle* $x, y \in \mathbb{R}^3$.

Beweis: Punkt (i) folgt, da nach Definition des euklidischen Skalarproduktes offensichtlich immer $\langle x, x\rangle \geq 0$ ist, und die Gleichheit nur im Fall $x = 0$ gilt. Punkt (ii) gilt, da

$$\|a \cdot x\|^2 = \langle a \cdot x, a \cdot x\rangle = (a \cdot x_1)^2 + (a \cdot x_2)2 + (a \cdot x_3)^2 = a^2 \cdot (x_1^2 + x_2^2 + x_3^2)$$

gilt, also wegen $|a| = \sqrt{a^2}$

$$\|a \cdot x\| = |a| \cdot \|x\|.$$

Allgemeiner halten wir fest, dass für das euklidische Skalarprodukt

$$\langle a \cdot x, y\rangle = a \cdot \langle x, y\rangle \quad \text{und} \quad \langle x, a \cdot y\rangle = a \cdot \langle x, y\rangle \qquad (\star)$$

gilt, woraus die Normeigenschaft (ii) folgt. Punkt (iii) ist am aufwändigsten. Hierzu stellen wir fest, dass das euklidische Skalarprodukt die additiven Eigenschaften

$$\begin{aligned} \langle x + y, z\rangle &= (x_1 + y_1)z_1 + (x_2 + y_2)z_2 + (x_3 + y_3)z_3 \\ &= x_1z_1 + x_2z_2 + x_3z_3 + y_1z_1 + y_2z_2 + y_3z_3 = \langle x, z\rangle + \langle y, z\rangle \end{aligned}$$

und analog

$$\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$$

hat, dass es symmetrisch ist, d.h.

$$\langle x, y\rangle = x_1y_1 + x_2y_2 + x_3y_3 = \langle y, x\rangle$$

und dass wegen $-1 \leq \cos(\gamma) \leq 1$ die Ungleichung

$$\pm\langle x, y\rangle \;=\; \pm\cos(\gamma) \cdot \|x\|\|y\| \;\leq\; \|x\| \cdot \|y\|$$

gilt. Wir erhalten zusammen mit der Eigenschaft $(\star)$ für Skalare

$$\begin{aligned} \|x + y\|^2 &= \langle x + y, x + y\rangle = \langle x, x + y\rangle + \langle y, x + y\rangle \\ &= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\ &\leq \langle x, x\rangle + 2\|x\|\|y\| + \langle y, y\rangle = (\|x\| + \|y\|)^2 . \end{aligned}$$

Da beide Seiten nach (i) positive Größen sind, bleibt die Ungleichung bei Wurzelziehen erhalten.

$\square$

Nach diesen einführenden Betrachtungen wollen wir den Begriff des Skalarprodukts nun so verallgemeinern, dass die Eigenschaften, die zu Längen-, Abstands- und Winkelmessung geführt haben, auch im allgemeinen Fall erhalten bleiben. Wir werden unten sehen, dass man aus bestimmten Gründen zwischen reellen Vektorräumen und komplexen Vektorräumen unterscheiden muss, wenn man Skalarprodukte zur Längen- und Abstandsmessung verwenden will. Wir beginnen mit dem reellen Fall, der das obige Vorgehen in der euklidischen Ebene verallgemeinert, und führen dabei einige Begrifflichkeiten für die im vorigen Abschnitt gesehenen Eigenschaften ein.

14.2 Skalarprodukte auf reellen und komplexen Vektorräumen

Wir betrachten im folgenden nur Skalarprodukte auf Vektorräumen über den Körpern $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$. Es lassen sich z.B. auch Skalarprodukte auf endlichen Körpern definieren. Da dabei relativ viele Besonderheiten zu beachten sind (siehe allein hier den Fall $\mathbb{R}$ vs. den Fall $\mathbb{C}$), würde die Darstellung hier den Rahmen sprengen. Mehr zum Thema findet man in [HuWi].

14.2.1 Skalarprodukte auf reellen Vektorräumen

Definition 14.2.1. *(Euklidisches Skalarprodukt, euklidischer Vektorraum)*

Sei V ein $\mathbb{R}$-Vektorraum. Eine Abbildung

$$b : V \times V \to \mathbb{R}, \qquad (v, w) \mapsto b(v, w),$$

heißt

(i) bilinear, falls bei festen v bzw. $w \in V$ die Abbildungen $b(v, \cdot) : V \to \mathbb{R}, w \mapsto b(v, w)$ und $b(\cdot, w) : V \to \mathbb{R}, v \mapsto b(v, w)$ $\mathbb{R}$-lineare Abbildungen sind. Ausgeschrieben bedeutet das, dass bei festem $v \in V$ für alle $w_1, w_2 \in V$ und $a \in \mathbb{R}$ gilt

- $b(v, w_1 + w_2) = b(v, w_1) + b(v, w_2)$ *und* $b(v, a \cdot w) = a \cdot b(v, w)$

(„Linearität in der zweiten Komponente"), und dass bei festem $w \in V$ für alle $v_1, v_2 \in V$ und $a \in \mathbb{R}$ gilt:

- $b(v_1 + v_2, w) = b(v_1, w) + b(v_2, w)$ *und* $b(a \cdot v, w) = a \cdot b(v, w)$.
  („Linearität in der ersten Komponente")

(ii) symmetrisch, falls gilt:

$$b(v, w) = b(w, v) \quad \textit{für alle} \quad v, w \in V.$$

(iii) positiv definit, wenn gilt:

$$b(v,v) \geq 0 \quad \textit{für alle} \quad v \in V, \quad \textit{und} \quad b(v,v) = 0 \Leftrightarrow \quad v = 0_V$$

Eine Abbildung $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$, die alle diese Eigenschaften vereint, also bilinear, positiv definit und symmetrisch ist, nennen wir Skalarprodukt oder (besonders im englischnahen Sprachgebrauch) inneres Produkt auf V.

Einen mit euklidischem Skalarprodukt ausgestatteten $\mathbb{R}$-Vektorraum $(V, \langle \cdot, \cdot \rangle)$ bezeichnet man auch als euklidischen Vektorraum.

Ist $(V, \langle \cdot, \cdot \rangle)$ ein euklidischer Vektorraum, so ist, da $\langle v, v \rangle \geq 0$ für alle $v \in V$ ist, die Abbildung

$$\|\cdot\| : V \to \mathbb{R}_0^+, \qquad v \mapsto \langle v, v \rangle^{\frac{1}{2}}$$

wohldefiniert. Sie heißt die von $\langle \cdot, \cdot \rangle$ induzierte euklidische Norm auf $(V, \langle \cdot, \cdot \rangle)$.

Wir werden später zeigen, dass die induzierte Norm $\|\cdot\|$ immer die drei Normeigenschaften erfüllt, die wir in Lemma 14.1.1 bereits für die vom Euklidischen Skalarprodukt im Raum induzierte Norm nachgewiesen haben, also als verallgemeinerte Längenmessung verstanden werden kann.

Beispiele:

- Das (verallgemeinerte) euklidische Skalarprodukt in $\mathbb{R}^n$ ist die Abbildung

  $$\langle x, y \rangle := \sum_{i=1}^{n} x_i y_i.$$

  Man bezeichnet es auch als das kanonische Skalarprodukt des $\mathbb{R}^n$. Für den Fall $n = 3$ haben im vorangegangenen Abschnitt gesehen, dass das euklidische Skalarprodukt wirklich die Bedingungen erfüllt, die wir gerade an ein Skalarprodukt gestellt haben. Analog verifiziert man den allgemeinen Fall.

- Das skalierte euklidische Skalarprodukt

  $$\langle x, y \rangle_n := \frac{1}{n} \left(\sum_{i=1}^{n} x_i y_i \right)$$

ist ebenfalls ein Skalarprodukt, die zugehörige Norm $\|x\|$ ist das so genannte *quadratische Mittel* über die Einträge des Vektors $x \in \mathbb{R}^n$.

- Bezeichnet

$$C(I) := \{f : [a, b] \to \mathbb{R} \mid f \ \text{ stetig}\}$$

die stetigen Funktionen von einem nichtleeren Intervall $I = [a, b]$ nach $\mathbb{R}$, so sind die Abbildungen

$$\langle f, g\rangle := \int_a^b f(x)g(x)dx \quad \text{und} \quad \langle f, g\rangle_I := \frac{1}{b-a}\int_a^b f(x)g(x)dx$$

wohldefiniert[1] und Skalarprodukte auf dem unendlichdimensionalen Raum $C([a, b])$.

14.2.2 Skalarprodukte auf komplexen Vektorräumen

Im Vektorraum $\mathbb{K} = \mathbb{C}^n$ liefert die einfache Übertragung des Skalarprodukts $\langle z, w\rangle = \sum_{i=1}^n z_i w_i$ für den Fall $z = w$ den Ausdruck $\langle z, z\rangle = \sum_{i=1}^n z_i^2 \in \mathbb{C}$, bei dem die Definition von $\|z\| \in \mathbb{R}$ zur Längenmessung schwierig ist: Der Ausdruck ist eine komplexe Zahl, für den man zwar zwei „Wurzeln“ findet (s. Lemma 3.2.1), diese allerdings ja im Allgemeinen nicht reell sind und sich so nicht als Längen bzw. Abstände deuten lassen können.

Definiert man stattdessen

$$\langle z, w\rangle = \sum_{i=1}^n \overline{z}_i w_i \qquad (\star),$$

so ergibt sich für den Fall $z = w$ den Ausdruck $\langle z, z\rangle = \sum_{i=1}^n |z_i|^2 \in \mathbb{R}$. Man prüft leicht, dass

$$\|z\|_2 \ := \ \langle z, z\rangle^{\frac{1}{2}}$$

eine Norm auf $\mathbb{C}$ definiert (Übungsaufgabe). Allerdings sind durch $(\star)$ die Eigenschaften $\langle a \cdot v, w\rangle = a \cdot \langle v, w\rangle$ und $\langle v, w\rangle = \langle w, v\rangle$ des reellen Skalarprodukts verletzt; stattdessen gilt $\langle a \cdot v, w\rangle = \overline{a} \cdot \langle v, w\rangle$ und $\langle v, w\rangle = \overline{\langle w, v\rangle}$ (der Querbalken symbolisiert wie vorher die komplexe Konjugation der Größe darunter). Wir passsen daher die Definition für Vektorräume über $\mathbb{K} = \mathbb{C}$ entsprechend an.

[1] Das folgt eigentlich erst aus der Cauchy-Schwarz'schen Ungleichung, die wir später beweisen werden, sowie einigen Sätzen zum Riemann-Integral, insbesondere dem, dass stetige Funktionen auf abgeschlossenen Intervallen immer integrierbar sind; es aber trotzdem ein wichtiges Beispiel und ist daher hier schon aufgeführt.

Definition 14.2.2. *(Skalarprodukte auf $\mathbb{C}$-Vektorräumen, unitäre Räume)*

Sei V ein $\mathbb{C}$-Vektorraum. Eine Abbildung

$$b : V \times V \to \mathbb{C}, \qquad (v, w) \mapsto b(v, w)$$

heißt unitäres Skalarprodukt, falls sie sesquilinear, hermitesch und positiv definit ist. Die Begriffe bedeuten dabei folgendes:

(i) b heißt sesquilinear[2], wenn mit festem $v \in V$

- *die Abbildung*

$$b(v, \cdot) : V \to \mathbb{C}, \qquad w \mapsto b(v, w)$$

  mit fester erster Komponente $\mathbb{C}$-linear ist, und

- *die Abbildung $v \mapsto b(v, w)$ mit festem w in der zweiten Komponente semilinear ist, d.h. es gilt für alle $v_1, v_2 \in V$ und $a \in \mathbb{C}$*

$$b(w_1 + w_2, w) = b(w_1, v) + b(w_2, v) \quad \text{und} \quad b(a \cdot w, v) = \overline{a} \cdot b(w, v).$$

  (Die rechte Gleichung ist die entscheidende Änderung!)

(ii) b heißt hermitesch, falls

$$b(v, w) = \overline{b(w, v)} \quad \text{für alle} \quad v, w \in V.$$

(iii) Aus (ii) folgt $b(v, v) \in \mathbb{R}$ für alle $v \in V$. b heißt zusätzlich positiv definit, genau dann, wenn

$$b(v, v) \geq 0 \quad \text{für alle} \quad v \in V, \quad \text{und} \qquad b(v, v) = 0 \quad \Leftrightarrow \quad v = 0_V.$$

Ein komplexer Vektorraum mit einem Skalarprodukt heißt unitärer Vektorraum. Ist $(V, \langle \cdot, \cdot \rangle)$ ein unitärer Vektorraum, so ist, da $\langle v, v \rangle \geq 0$ für alle $v \in V$ ist, die Abbildung

$$\|\cdot\| : V \to \mathbb{R}_0^+, \qquad v \mapsto \langle v, v \rangle^{\frac{1}{2}}$$

wohldefiniert. Sie heißt die von $\langle \cdot, \cdot \rangle$ induzierte Norm auf $(V, \langle \cdot, \cdot \rangle)$.

[2] Dass man „bilinear“ mit zwei-linear, also in zwei Komponenten linear übersetzen kann, und dass das unten auftretende „semilinear“ so etwas wie halb-linear bedeutet, ist wahrscheinlich einigermaßen klar und macht hier hoffentlich Sinn. Die Vorsilbe „Sesqui-“ ist wohl nicht so verbreitet: Übersetzt heißt es so etwas wie anderthalb, eine Sequilinearform ist also in $1\frac{1}{2}$ Komponenten linear. Man überzeuge sich, dass auch das einigermaßen genau wiedergibt, was hier gefordert wird.

Wir werden im folgenden euklidische Vektorräume und unitäre Vektorräume unter dem Begriff Skalarprodukträume zusammenfassen, wenn die Unterscheidung gerade keine Rolle spielt.

Bemerkungen dazu:

- Die oben erwähnte sesquilineare Abbildung

$$\langle\cdot,\cdot\rangle : \mathbb{C}^n \to \mathbb{C}, \quad (z,w) \mapsto \sum_{i=1}^{n} \overline{z}_i w_i$$

  für $Z = (z_1, \ldots, z_n), w = (w_1, \ldots, w_n) \in \mathbb{C}^n$ ist das sogenannte Standardskalarprodukt auf $\mathbb{C}^n$.

- Die stetigen Funktionen $C([a,b],\mathbb{C})$ mit Wertevorrat $[a,b]$ und Werten in $\mathbb{C}$ sind ein komplexer Vektorraum, der sich analog zu oben z.B. durch

$$\langle f,g\rangle = \int_a^b \overline{f(x)} \cdot g(x) dx$$

  zu einem unitären Raum machen lässt; insbesondere ist hier $\langle f,f\rangle \in \mathbb{R}$ für alle $f \in C([a,b],\mathbb{C})$.

14.3 Orthogonalität

Wir wollen nun den Begriff der Orhogonalität, also des „Senkrechtstehens“, der der Startpunkt unserer Überlegungen zum Skalarprodukt in $\mathbb{R}^3$ ausmachte, auf allgemeine Vektorräumen übertragen. Wir lassen uns für die folgende Definition der Orthogonalität per Skalarprodukt wieder von den Beziehungen im Anschauungsraum inspirieren und werden sehen, dass sich dadurch eine Fülle von Ergebnissen auf allgemeine Skalarprodukträume und affine Räume analog übertragen lassen.

Definition 14.3.1. *(Einige Begriffe zu Orthogonalität und Normen)*

Sei $(V, \langle\cdot,\cdot\rangle)$ ein Skalarproduktraum. $\|\cdot\|$ bezeichne die vom Skalarprodukt $\langle\cdot,\cdot\rangle$ induzierte Norm.

- *Ein Vektor $v \in V$ heißt normiert, wenn $\|v\| = 1$ ist.*
- *Zwei Vektoren u und v aus V heißen orthogonal, falls $\langle u, v\rangle = 0$ ist. Wir schreiben dafür kurz $u \perp v$.*
- *Zwei Untervektorräume U und W von V heißen orthogonal, falls*

$$u \perp w \quad \text{für alle} \quad u \in U, w \in W.$$

- *Sei $M \subseteq V$. Falls für $u \in V$, die Relation $u \perp v$ für alle $v \in M$ gilt, schreiben wir dafür auch $u \perp M$.*
- *Die Menge*

$$M^{\perp} := \{v \in V \mid v \perp M\}$$

  heißt das orthogonale Komplement von M in V.

Lemma 14.3.2. *(Einige Eigenschaften der Orthogonalität & Co.)*

Für beliebige Vektoren v, w eines Skalarproduktraums V und alle Teilmengen $M \subseteq V$ gilt:

(i) *Ist $v \neq 0_V$, so ist $\frac{1}{\|v\|} \cdot v$ normiert.*

(ii) $0_V \perp v$,

(iii) $v \perp w \Leftrightarrow w \perp v$,

(iv) $v \perp v \Longleftrightarrow v = 0_V$.

(v) *Ist $u \perp M$, so gilt auch $u \perp \mathcal{L}(M)$.*

(vi) *$M^{\perp}$ ist ein Teilraum von V.*

Beweis: (i) bis (v) ist mit Eigenschaften des Skalarproduktes schnell bewiesen. (vi) folgt aus dem Teilraumkriterium.

□

Definition 14.3.3. *(Orthogonal- und Orthonormalsysteme)*

Sei V ein Skalarproduktraum und $Q = (q_1, \ldots, q_m)$ ein m-Tupel von Vektoren, die alle ungleich dem Nullvektor sind. Dann heißt Q

- *Orthogonalsystem, falls $q_i \perp q_j$ für alle $i \neq j, i, j \in \underline{m}$,*
- *Orthonormalsystem, falls Q Orthogonalsystem ist und alle Vektoren q_i normiert sind, d.h. es gilt*

$$\langle q_i, q_j \rangle = \delta_{i,j} \quad \textit{mit dem „Kronecker-Delta"} \quad \delta_{i,j} := \begin{cases} 1 & \textit{für } \; i = j \\ 0 & \textit{sonst} \end{cases}$$

- *Orthogonalbasis, falls Q Orthogonalsystem und eine Basis von V ist,*
- *Orthonormalbasis (oft: ONB), falls Q Orthonormalsystem und eine Basis von V ist.*

Bemerkungen dazu:

- Die Elemente der Standardbasis $I_n = (e_1, \ldots, e_n)$ des $\mathbb{R}^n$ bzw. $\mathbb{C}^n$ können wir mit dem oben eingeführten Kronecker-Delta als $e_i = (\delta_{i,j})_{j \in \underline{n}}$ schreiben. Sie sind offensichtlich eine Orthonormalbasis bezüglich des Standardskalarprodukts (bezüglich Standardbasis) in $\mathbb{K}^n$.
- Es gibt noch andere Basen, die Orthonormalbasen in $\mathbb{K}^n$ bezüglich des Standardskalarprodukts sind, z.B. sind

$$B = \left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix} \right) \qquad \text{bzw.} \qquad C = \left(\begin{pmatrix} \cos\alpha \\ \sin\alpha \end{pmatrix}, \begin{pmatrix} -\sin\alpha \\ \cos\alpha \end{pmatrix} \right)$$

Basen von $\mathbb{R}^3$ bzw. $\mathbb{R}^2$ (und auch von $\mathbb{C}^3$ bzw. $\mathbb{C}^2$);

$$\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right), \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0 \right)$$

„nur" ein Orthonormalsystem in $\mathbb{R}^3$ und $\mathbb{C}^3$.

- Die Polynomfunktionen $1, x, x^2, \ldots$ sind i.A. nicht orthonormal. Es sind aber z.B. die Funktionen $1, x^2, x^4, \ldots$ in $C([-1,1],\mathbb{R})$ ein Orthogonalsystem bezüglich des Skalarprodukts
$$\int_{-1}^{1} f(x)g(x)dx.$$
- Ein Orthogonalsystem erhält man, wenn man in $C([-\pi,\pi],\mathbb{R})$, mit dem Skalarprodukt $\int_{-\pi}^{\pi} f(x)g(x)dx$ die Funktionen
$$1, \qquad x \mapsto \sin(kx) \quad (k \in \mathbb{N}), \qquad x \mapsto \cos(kx) \quad (k \in \mathbb{N})$$
betrachtet. (Das zeigt man mit Hilfe der Additionstheoreme und der Eigenschaften der jeweiligen Stammfunktionen.) Dieses Orthogonalsystem lässt sich zu einem ONS normieren, und man kann zeigen, dass sich jede stetige Funktion als Grenzwert einer unendlichen Reihe solcher Funktionen schreiben lässt (also als im Sinne der Analysis verallgemeinerte Linearkombination; die obigen Funktionen bezeichnet man auch als „Fourier-Basis“). Die Fourierbasis wird oft verwendet, um (meist akustische) Signale als Summe ihre verschiedenen Frequenzanteile zu zerlegen und dann nur einen wichtigen Teil davon zu speichern, etwa bei der Audiokompression mit MP3, eine vereinfachte Variante im $\mathbb{K}^n$ wurde grob in den LAAG-I-Videos vorgestellt. Die dafür nötigen Techniken lernen wir im Prinzip im kommenden Abschnitt kennen.

Mit Orthogonal- und -normalsystemen usw. können wir rechnerisch bequem umgehen, indem wir die Vektoren eines solchen Systems/einer solchen Basis Q wie zuvor in die Spalten einer Matrix schreiben (die wir dann wie immer auch Q nennen). Hier spiegeln sich einige Eigenschaften der Basis direkt in den Eigenschaften verschiedener Matrixoperationen wieder. In Abschnitt werden wir auch sehen, dass ein Skalarprodukt immer mit Hilfe einer Matrix mit bestimmten Eigenschaften beschrieben werden kann; daher definieren wir die folgenden Begriffe für Matrizen.

Definition 14.3.4. *(Matrixoperationen und besondere Matrizen)*

(i) Sei $A \in \mathbb{K}^{m\times n}$ eine Matrix. Dann heißt die Matrix $A^T \in \mathbb{K}^{n\times m}$ mit den Einträgen
$$a^T_{i,j} := a_{j,i}$$
für alle $i \in \underline{n}, j \in \underline{m}$, die Transponierte (oder transponierte Matrix) von A. A heißt symmetrisch, falls $A = A^T$ ist.

(ii) Sei $A \in \mathbb{C}^{m \times n}$, so heißt die Matrix $A^H \in \mathbb{C}^{n \times m}$ mit den Einträgen

$$a_{i,j}^H := \overline{a}_{j,i}$$

für alle $i \in \underline{n}, j \in \underline{m}$ die Adjungierte (oder adjungierte Matrix)[3] *von A. A heißt hermitesch, falls $A = A^H$ ist.*

Zusammenfassend spricht man auch davon, dass eine Matrix $A \in \mathbb{K}^{m \times n}$ selbstadjungiert ist, falls $A = A^T$ im reellen Fall bzw. $A = A^H$ im komplexen Fall gilt.

Bemerkungen dazu:

- Anschaulich entsteht die Transponierte einer Matrix durch Spiegelung der Einträge an der Hauptdiagonalen von A, Ergebnis ist eine Matrix mit vertauschten Zeilen- und Spalteneinträgen. Eine symmetrische Matrix erkennt man also auch daran, dass ihre Einträge symmetrisch zur Hauptdiagonalen sind.

- Offensichtlich kann A nur symmetrisch bzw hermitesch sein, wenn A quadratisch ist.

- Jede Matrix mit reellen Einträgen lässt sich auch als Matrix mit komplexen Einträgen auffassen. In diesem Fall gilt $A^T = A^H$.

- Mit Hilfe der Transponierten lässt sich das euklidische Skalarprodukt zweier Vektoren $x, y \in \mathbb{K}^n$ kurz als
$$x^T y \qquad \text{bzw.} \qquad x^H y$$
schreiben, wenn man davon ausgeht, dass Vektoren im $\mathbb{K}^n$ kanonisch Spaltenvektoren sind. Wir haben bisher nicht zwischen Spalten- und Zeilenvektoren unterschieden, wollen das aber aus diesem Grund ab hier tun:

Ab hier lesen wir Vektoren $x \in \mathbb{K}^n$ ausschließlich als Spaltenvektoren, also als Elemente von $\mathbb{K}^{n \times 1}$. Der zu x gehörige Zeilenvektor ist der Vektor $x^T \in \mathbb{K}^{1 \times n}$.

[3] aber nicht Adjunkte, das ist etwas anderes!

Lemma 14.3.5. *(„Matrixkriterium“ für Ortho-)*

Sei $V = \mathbb{R}^n$ mit dem Standardskalarprodukt ausgestattet und $Q = (q_1, \ldots, q_m)$ ein Tupel von Vektoren aus dem $\mathbb{R}^n$. $Q \in \mathbb{R}^{n \times m}$ bezeichne auch die Matrix mit den Basivektoren als Spaltenvektoren. Dann gilt:

- *Q ist Orthogonalsystem $\iff$ $Q^T Q$ ist eine Diagonalmatrix, d.h. für die Eintrage der Matrix gilt $a_{i,j} = 0$ für alle $i \neq j$, und alle Diagonaleinträge sind ungleich Null,*
- *Q ist Orthonormalsystem $\iff$ $Q^T Q = I_m$,*
- *Q ist Orthogonalbasis $\iff$ Es ist $m = n (= \dim \mathbb{R}^n)$, $Q^T Q$ ist Diagonalmatrix und alle Diagonaleinträge sind nicht Null.*
- *Q ist Orthonormalbasis $\iff$ Es ist $m = nV$ und $Q^T Q = I_n$.*

Die Aussagen gelten analog mit Q^H anstelle von Q^T, wenn $V = \mathbb{C}^n$ ist.

Beweis: Eine schöne Übungsaufgabe zum Warmwerden mit Orthogonalität und den Eigenschaften aus dem Lemma darüber. □

Lemma 14.3.6. *Sei V ein Skalarproduktraum.*

(i) Ist $Q = (q_1, \ldots, q_k)$ ein Orthonormalsystem (also $q_i \perp q_j$ für alle $i, j \in \underline{k}$), so ist Q linear unabhängig.

(ii) Ist V ein Skalarproduktraum mit $\dim V = n$, so ist ein Orthonormalsystem aus n Vektoren immer eine Orthonormalbasis von V.

Beweis: Für (i) berechnet man für alle $j \in \underline{k}$ das Skalarprodukt $\langle \sum_{i=1}^n a_i q_i, q_j \rangle$, um das Kriterium für lineare Unabhängigkeit zu zeigen. (ii) folgt, da orthogonale Vektoren nach (i) l.u. sind, also eine Basis nach Lemma 8.3.9.
□

Die große Bedeutung von Orthonormalbasen rührt daher, dass Sie theoretische wie praktische (also numerische) Rechnungen oft stark vereinfachen. Ein Beispiel für die theoretischen Vorteile von Orthonormalbasen zeigt das folgende Lemma.

Lemma 14.3.7. *(Koordinaten bezüglich ONBs)*
Sei $Q = (q_1, \dots, q_n)$ eine Orthonormalbasis eines endlichdimensionalen Skalarproduktraums $(V, \langle\cdot,\cdot\rangle)$ und $v \in V$ beliebig. Bezeichnen wir $c_i := \langle q_i, v\rangle$, so gilt

$$v = \sum_{i=1}^n c_i q_i \quad \text{und} \quad \|v\| = \left(\sum_{i=1}^n |c_i|^2\right)^{\frac{1}{2}} = \left\| \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \right\|,$$

die Norm von v ist also gerade die euklidische bzw. kanonische Norm seines Koeffizientenvektors $(c_1, \dots, c_n)^T$ in $\mathbb{K}^n$.

Beweis: Sei $v = \sum_{i=1}^n c_i q_i$ die Basisdarstellung von v. Dann sind die Koeffizienten c_i eindeutig bestimmt und es gilt für alle $j = 1, \dots, n$, dass

$$\langle q_j, v\rangle = \langle q_j, \sum_{i=1}^n c_i q_i\rangle = \sum_{i=1}^n c_i \langle q_j, q_i\rangle = c_j,$$

Die zweite Behauptung folgt durch Berechnung von $\|v\|^2 = \langle v, v\rangle$. $\square$

Bemerkungen dazu:

- Das Lemma bedeutet insbesondere dass man den Koordinatenvektor von v erhält, indem man die Skalarprodukte mit den Vektoren aus Q bildet. Ist $V = \mathbb{K}^n$, so ist das einfach die Matrix-Vektor-Multiplikation

$$k_Q(v) = Q^T \cdot v \qquad \text{bzw.} \qquad k_Q(v) = Q^H \cdot v.$$

- Wählen wir eine Teilmenge $Q' = (q_1, \dots, q_k)$ der obigen ONB aus und nennen den von Q' erzeugten Teilraum $U = \mathcal{L}(Q')$, so lässt sich v folgendermaßen zerlegen:

$$v = \underbrace{\sum_{i=1}^k \langle q_i, v\rangle q_i}_{\in U} + \underbrace{\sum_{i=k+1}^n \langle q_i, v\rangle q_i}_{\in U^\perp}; \tag{14.2}$$

  wir haben v also in einen Anteil „in Richtung U“ und einen Anteil senkrecht zu U zerlegt. Das ist der Grundgedanke der folgenden Überlegungen.

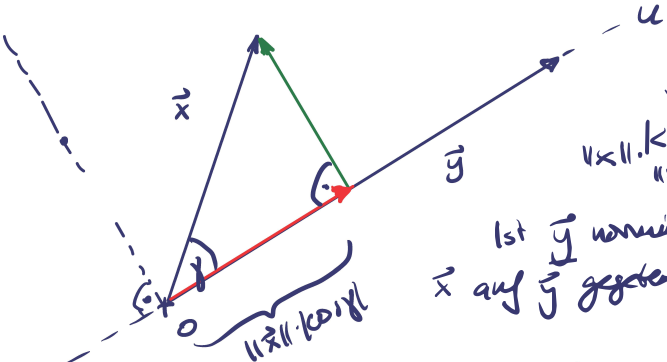
14.4 Orthogonale Zerlegungen

Wir lassen uns nun nochmal von der elementargeometrischen Herangehensweise am Anfang dieses Kapitels inspririeren.

- Dazu betrachten Vektoren $\vec{x}, \vec{y}$ in der Ebene (oder auch im Raum, ist nur komplizerter), die einen Winkel γ einschließen. Wir können den Kosinus dieses Winkels schreiben als
$$\cos(\gamma) = \langle \frac{x}{\|x\|}, \frac{y}{\|y\|} \rangle.$$
(Hierbei haben wir die Faktoren $\frac{1}{\|x\|}, \frac{1}{\|y\|}$ per Linearität in das Skalarprodukt hineingezogen, um noch einmal darauf hinzuweisen, dass es für $\cos(\gamma)$ nur auf die Richtung der beiden Vektoren ankommt.)

- Die *senkrechte Projektion* von $\vec{x}$ auf die von $\vec{y}$ erzeugte Gerade ist gerade $\|x\| \cos(\gamma) \cdot \vec{y}$. Ist $\vec{y}$ normiert, so ist die Länge der Projektion gerade $\langle x, y \rangle$ und die Projektion von $\vec{x}$ auf das Erzeugnis von $\vec{y}$ gerade der Vektor $P_{\vec{y}}\vec{x} = \langle x, y \rangle \vec{y}$, und es gilt mit der Einheitsmatrix I des $\mathbb{R}^3$
$$\vec{x} = P_{\vec{y}}\vec{x} + (I - P_{\vec{y}})\vec{x},$$
wobei für die zugehörigen Koordinatenvektoren
$$\langle (I - P_y)x, y \rangle = \langle x - \langle x, y \rangle y, y \rangle = \langle x, y \rangle - \langle x, y \rangle \langle y, y \rangle = 0$$
ist (Da $\vec{y}$ normiert ist, ist $\langle y, y \rangle = 1$.) Es ist also
$$(I - P_{\vec{y}})\vec{x} \perp \vec{y},$$
wir haben $\vec{x}$ orthogonal zerlegt in einen Vektor, der in der von $\vec{y}$ erzeugten Gerade liegt und einen, der senkrecht dazu liegt.

Diese Erkenntnis, dass zumindest im Anschauungsraum für einen gegebenen Unterraum U ein Vektor x in einen Anteil in U und einen senkrecht dazu zerlegt werden kann, wollen wir jetzt auf beliebige endlichdimensionale Unterräume U verallgemeinern.



Wegen

$$\cos\gamma = \frac{\langle x,y\rangle}{\|x\|\cdot\|y\|}$$

ist die Länge des roten Vektors gerade

$$\|x\|\cdot\frac{|\langle x,y\rangle|}{\|x\|\|y\|} = \left|\left\langle x, \frac{y}{\|y\|}\right\rangle\right|$$

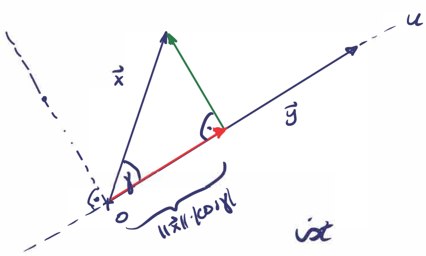
Ist $\vec y$ normiert, so ist die Projektion von $\vec x$ auf $\vec y$ gegeben durch

$$P_{\vec y}\vec x = \langle y,x\rangle\cdot\vec y$$

Der zu $\vec y$ orthogonale Anteil ist gerade

$$\vec x - P_{\vec y}\vec x = \underline{\vec x - \langle y,x\rangle\cdot\vec y}.$$

XI.4 Orthogonale Zerlegungen



Mit $\boxed{P_{\vec y}\vec x = \langle y,x\rangle\cdot\vec y}$

und $(I-P_{\vec y})\cdot\vec x = \vec x - P_{\vec y}\vec x = \underline{\vec x - \langle y,x\rangle\cdot\vec y}.$

ist

$$\langle P_{\vec y}\vec x, (I-P_{\vec y})\cdot\vec x\rangle = \langle\langle y,x\rangle\cdot\vec y, \vec x - \langle y,x\rangle\cdot\vec y\rangle$$

$$= \langle\langle y,x\rangle\cdot\vec y, \vec x\rangle - \langle\langle y,x\rangle\vec y, \langle y,x\rangle\vec y\rangle$$

$$= \langle y,x\rangle\cdot\langle\vec y,\vec x\rangle - \langle y,x\rangle\langle y,x\rangle\langle\vec y,\vec y\rangle$$

$$= \langle y,x\rangle^2 - \langle y,x\rangle^2$$

Satz 14.4.1. *(Orthogonale Projektionen auf Unterräume)*

Sei V ein Skalarproduktraum und U ein endlichdimensionaler Teilraum von V und $Q = (q_1, \dots, q_m)$ eine Orthonormalbasis von U. Dann gilt:

(i) Es gibt es eine eindeutige lineare Abbildung mit den drei Eigenschaften $\mathrm{Bild}(P) = U$, $\mathrm{Kern}(P) = U^\perp$ *und* $Pu = u$ *für alle* $u \in U$.
P heißt der orthogonale Projektor auf U.

(ii) P lässt sich folgendermaßen ausdrücken:

$$P : V \to V, \quad v \mapsto \sum_{i=1}^{m} \langle q_i, v \rangle q_i.$$

(iii) Die Abbildung $P^\perp = \mathrm{id}_V - P$ *ist der orthogonale Projektor auf das orthogonale Komplement von U, d.h es gilt* $P^\perp v = v$ *für alle* $v \in U^\perp$, $\mathrm{Bild}(P^\perp) = U^\perp$ *und* $\mathrm{Kern}(P^\perp) = U$.

Beweis, Teil 1. (i) & (ii): Wir zeigen zunächst, dass der in (ii) definierte Projektor die gewünschten Eigenschaften aus (i) hat. P ist linear: Wegen der Linearität des Skalarprodukts in der zweiten Komponente gilt

$$\langle q_i, v + a \cdot w \rangle q_i = \langle q_i, v \rangle q_i + a \langle q_i, w \rangle q_i,$$

d.h. die Abbildungen in der Summe alle linear sind und Summen linearer Abbildungen sind wiederum linear. Weiter gilt, da $(q_1, \dots, q_m)$ eine Orthonormalbasis von U ist, für alle $u \in U$ nach Lemma 14.3.7 die Darstellung

$$u = \sum_{i=1}^{m} \langle q_i, v \rangle q_i = Pu \in \mathrm{Bild}(P),$$

das zeigt $u = Pu$ für alle $u \in U$, und auch $U \subseteq \mathrm{Bild}(P)$.
Umgekehrt gilt für alle $v \in V$

$$Pv = \sum_{i=1}^{n} \langle q_i, v \rangle q_i,$$

das hintere ist eine Linearkombination der Vektoren aus Q, also gilt $Pv \in U$ und wir haben insgesamt $\operatorname{Bild}(P) = U$. Ist $v \in U^\perp$, so ist

$$Pv = \sum_{i=1}^{m} \underbrace{\langle q_i, v\rangle}_{\text{immer } =0_V, v\in U^\perp} q_i = 0.$$

Umgekehrt gilt, falls $v \in \operatorname{Kern} P$, dass

$$\sum_{i=1}^{m} \langle q_i, v\rangle\; q_i = 0_V.$$

Q ist als Basis von U l.u., nach dem Kriterium für lineare Unabhängigkeit also $\langle q_i, v\rangle = 0$ für alle $i \in \underline{m}$, das heißt $v \perp Q$. Mit Lemma 14.3.2(iii) folgt $v \perp \mathcal{L}(Q)$. P ist also ein Projektor mit den in (i) und (ii) angegebenen Eigenschaften.

Der Beweis von Teil (iii) ist Übungsaufgabe.

□

Einige Bemerkungen zu orthogonalen Projektoren:

- Blättern wir zurück zu (14.2) auf Seite 319, so sehen wir jetzt, dass wir dort die Basis Q in zwei Anteile zerlegt haben; die Anteile auf der rechten Seite sind gerade die Projektionen des Vektors v auf $U = \mathcal{L}(Q)$ und $U^\perp = \mathcal{L}(Q')$, und allgemeiner lässt sich der Ausdruck als eine Zerlegung von $v = \mathrm{id}_V(v)$ in die verschiedenen Richtungen q_i deuten; es ist also $\mathrm{id}_V = \sum_{i=1}^{n} \langle q_i, \cdot\rangle q_i$. Man nennt solch ein Vorgehen eine *Zerlegung der Eins* (mit „Eins" ist die Identität gemeint.)

- Nach (ii) sind für $v \in V$ die Koordinaten von Pv in der Basis $Q = (q_1, \ldots, q_m)$ des Unterraumes U gerade $c_i := \langle q_i, u\rangle$. Ist $V = \mathbb{K}^n$ und $\langle\cdot,\cdot\rangle$ das Standardskalarprodukt, so ergibt sich daraus

$$Pv \;=\; \sum_{i=1}^{n} q_i \cdot q_i^T v = QQ^T \cdot v, \;\text{ also } P = QQ^T, \tag{14.3}$$

und

$$P^\perp v \;=\; v - \sum_{i=1}^{n} q_i \cdot q_i^T v = (I - QQ^T)v, \;\text{ also } P^\perp = I - QQ^T.$$

- Man rechnet leicht nach, dass für die oben definierten orthogonalen Projektoren P die Beziehungen $P^2 = P$ und $\langle Pu, v\rangle = \langle u, Pv\rangle$ gelten (mit dem für den Projektor benutzten Skalarprodukt $\langle\cdot,\cdot\rangle$). Diese Eigenschaften sind zusammen äquivalent zur Forderung, dass es einen Unterraum gibt, so dass $\mathrm{Bild}(P) = U$, $Pu = u$ für alle $u \in U$ und $\mathrm{Kern}\, P = U^\perp$ ist, und werden manchmal alternativ als Definition für orthogonale Projektoren genutzt. Das hat den Vorteil, dass man U nicht zu kennen braucht, um zu prüfen, ob eine lineare Abbildung P ein Projektor ist (und den Nachteil, dass man schlechter sieht, was passiert).

Es bleiben im obigen Beweis zwei Lücken: Erstens ist noch die Eindeutigkeit des Projektors auf U (aus Teil (i)) zu zeigen. Das liefern wir weiter unten nach Lemma 14.4.4 nach.

Die zweite Lücke, die wir nun schließen wollen, ist die, dass wir oben vorausgesetzt haben, dass der endlichdimensionale Unterraum U eine Orthonormalbasis besitzt. Wir wollen nun zeigen, dass ein endlichdimensionaler Unterraum U eines Vektorraumes V *immer* eine Orthonormalbasis besitzt, dass wir also ohne Bedenken für *jeden* endlichdimensionalen Unterraum U den orthogonalen Projektor P_U benutzen können. Das zeigt der folgende Satz und das zugehörige *Orthonormalisierungsverfahren nach Gram und Schmidt.*

Satz 14.4.2. *(Existenz von Orthonormalbasen)*

Sei $(V, \langle\cdot,\cdot\rangle)$ ein Skalarproduktraum endlicher Dimension. Dann besitzt V eine Orthonormalbasis.

Beweis: Der Vektorraum V besitzt in jedem Fall eine eventuell nicht-orthogonale Basis $B = (b_1, \ldots, b_n)$, vgl. die entsprechende Aussage 8.3.7 aus der LAAG I. Diese können wir konstruktiv folgendermaßen schrittweise zu einer Orthonormalbasis machen:

Algorithmus: Gram-Schmidt'sches[3] Orthonormalisierungsverfahren

Gegeben: Basis $B = (b_1, \dots, b_n)$ eines Skalarproduktraumes V.

- Normiere b_1: Setze $q_1 := \frac{b_1}{\|b_1\|}$, $Q = (q_1)$
- Für $i = 2, \dots, n$ führe folgendes durch:
  - Orthogonalisiere b_i auf $\mathcal{L}(q_1, \dots, q_{i-1})$: Setze

$$q_i' := P^{\perp}_{\mathcal{L}(q_1,\dots,q_{i-1})} b_i \quad := \quad b_i - \sum_{j=1}^{i-1} \langle q_j, b_i \rangle q_j.$$

($P^{\perp}_{\mathcal{L}(q_1,\dots,q_{i-1})}$ ist der Projektor auf das orthogonale Komplement von $\mathcal{L}(q_1, \dots, q_{i-1})$, vgl. Satz 14.4.1(iii))

  - Normiere q_i': Setze $q_i := \frac{q_i'}{\|q_i'\|}$.
  - Hänge q_i an Q an $\rightsquigarrow Q = (q_1, \dots, q_i)$

Ausgabe: Orthonormalbasis $Q = (q_1, \dots, q_n)$ von U.

Wir zeigen die Korrektheit des Gram-Schmidt[4]-Verfahrens: Nach Konstruktion liefert der Algorithmus im ersten Schritt eine ONB $Q_1 = (q_1)$ des Raumes $\mathcal{L}(b_1)$. Wir zeigen im folgenden Absatz, dass diese Eigenschaft in jedem Schritt erhalten bleibt, d.h.: Ist $Q_{i-1} = (q_1, \dots, q_{i-1})$ eine ONB des Raumes $\mathcal{L}(b_1, \dots, b_{i-1})$, so entsteht im folgenden Schritt $i > 1$ eine ONB des Raumes $\mathcal{L}(b_1, \dots, b_i)$. Nach n Schritten ergibt sich dann die gewünschte ONB für U.

Sei also $Q_{i-1} = (q_1, \dots, q_{i-1})$ eine ONB des Raumes $\mathcal{L}(b_1, \dots, b_{i-1})$. Nach Konstruktion ist für $i \in \underline{n}$ $q_i' \perp q_j$ für $j < i$ nach Satz 14.4.1 (denn die dort angewandte Abbildung ist gerade der Projektor $P^{\perp}$ zum Unterraum Q_{i-1}. Da B linear unabhängig ist, gilt insbesondere $b_i \notin \mathcal{L}(q_1, \dots, q_{i-1}) = \mathcal{L}(b_1, \dots, b_{i-1})$, also $q_i' \neq 0_V$, so dass q_i wohldefiniert ist. Wegen $q_i \notin \mathcal{L}(q_1, \dots, q_{i-1})$ sind $q_1, \dots, q_i$ linear unabhängig und $q_m \in \mathcal{L}(b_1, \dots, b_i)$ für alle $m \in \underline{m}$, also eine i-dimensionale l.u. Teilmenge, d.h eine Basis des i-dimensionalen Raumes $\mathcal{L}(b_1, \dots, b_i)$.

□

[4] nach Erhard Schmidt (1876–1959), Deutschland, und Jørgen Pedersen Gram (1850-1916), Dänemark. Ersterer war ein Pionier der Funktionalanalysis ≈ Linearer Algebra unendlichdimensionaler Räume, letzterer beschäftigte sich mit u.a. orthogonalen Funktionensystemen, Zahlentheorie, Versicherungsmathematik, arbeitete aber nur im Versicherungswesen, nie an einer Universität.

Bemerkungen dazu:

- Das Gram-Schmidt'sche Orthonormalisierungsverfahren, funktioniert mit einer kleinen Modifikation auch, wenn man als Eingabe ein potenziell nicht linear unabhängiges Erzeugendensystem eingibt. In diesem Fall ergibt sich $q_i' = 0_V$, wenn b_i bereits im Aufspann der vorher berechneten Vektoren enthalten ist. Das lässt sich numerisch abfragen, und der Vektor wird dann einfach weggelassen. Man erhält als Ausgabe ein potenziell kürzere Vektormenge, die eine Basis von $\mathcal{L}(b_1, \ldots, b_n)$ ist.

- Das Gram-Schmidt-Verfahren funktioniert auch für Teilräume. Man kann es z.B. benutzen, um für affine Teilräume $N = P + U$ eine ONB Q_U des erzeugenden Unterraumes U zu berechnen.

- Ist B eine Basis von V, und wendet man dann das Gram-Schmidt-Verfahren auf $(Q_U \mid B)$ an, so lässt der Algorithmus das Orthonomalsystem Q_U wie es ist und orthogonalisiert den Rest. Es ergibt sich eine kürzere Vektormenge $Q = (Q_U \mid Q_{U^\perp})$, die eine Orthonormalbasis von $\mathcal{L}(Q_U \mid B) = V$ ist, die Basis Q_U von U also zu einer Basis des ganzen Raumes ergänzt.

Daraus folgt direkt ein Resultat, das jetzt ziemlich schlicht zu beweisen ist, aber oftmals angewandt wird:

Korollar 14.4.3. *(Ergänzung zur Orthogonalbasis)*

Sei V ein endlichdimensionaler Skalarproduktraum und seien $v_1, \ldots, v_k \in V$ orthonormale Vektoren. Dann kann man $v_1, \ldots, v_k$ zu einer Orthonormalbasis von V ergänzen.

Beweis: $v_1, \ldots, v_k \in V$ sind nach Voraussetzeung orthonormal, also insbesondere linear unabhängig. Man ergänzt $v_1, \ldots, v_k$ nach dem Basisergänzungssatz zu einer Basis $B = (v_1, \ldots, v_n)$ von V und wendet auf B das Gram-Schmidt-Verfahren an. (Wegen Punkt 3 in den Bemerkungen oben bleiben dabei die ersten k Vektoren erhalten.)

$\square$

Definition/Lemma 14.4.4. *(Orthogonale Zerlegung von V)*

Sei V ein endlichdimensionaler Skalarproduktraum und U ein Teilraum von V. Dann heißt das Paar $(U, U^\perp)$ orthogonale Zerlegung von V. Es hat die folgenden Eigenschaften:

(i) $U \cap U^\perp = \{0_V\}$

(ii) $\dim U + \dim U^\perp = \dim V$

(iii) Ist $B = (b_1, \dots, b_k)$ eine Basis von U und $B' = (c_1, \dots, c_\ell)$ eine Basis von $U^\perp$, so ist $B \cup B'$ eine Basis von V.

Beweis: (i) Ist $v \in U \cap U^\perp$, so folgt $v \perp v$, also $v = 0_V$. Umgekehrt gilt, da der Schnitt von Unterräumen immer ein Unterraum ist, auch $0_V \in U \cap U^\perp$.

(ii) Für den zu U gehörigen linearen Projektor P aus Satz 14.4.1 gilt nach (ii) dort $\operatorname{Kern} P = U$, $\operatorname{Bild} P = U^\perp$. Mit dem Dimensionssatz für lineare Abbildungen, angewandt auf $P : V \to V$, folgt die Behauptung.

(iii) Es ist $U = \mathcal{L}(B), U^\perp = \mathcal{L}(B')$. $U + U^\perp$ ist das Erzeugnis von $B \cup B'$, und diese Menge enthält nach (i) gerade $k + \ell$ Vektoren. Nach Dimensionssatz (jetzt für Unterräume) ist $U + U^\perp$ ein Raum der Dimension $k + \ell$ (da $U \cap U^\perp = \{0\}$), also muss $B \cup B'$ ein minimales Erzeugendensystem sein, also eine Basis von V.

□

Bevor wir nun den neu entwickelten „Apparat“ der Projektoren auf Unterräume nutzten wollen, liefern wir noch nach:

Beweis von Satz 14.4.1, Teil 2 (Eindeutigkeit des Projektors auf U):

Ist P' ein weiterer Projektor auf U mit den Eigenschaften aus (i) und $B' = (c_1, \dots, c_\ell)$ eine Basis von $U^\perp$, so gilt $P'(q_i) = q_i = P(q_i)$ und $P'(c_i) = 0 = P(c_i)$. P und P' stimmen also auf einer Basis von V überein, das bedeutet für lineare Abbildungen, dass sie gleich sind.

□

Projektoren sind oft praktische Werkzeuge. Wir werden Sie benutzen, um im Rest dieses Abschnitts leicht zwei grundlegende Resultate über Skalarprodukträume zu beweisen.

Satz 14.4.5. *(Orthogonale Zerlegung von Vektoren)*

Sei V ein Skalarproduktraum, U endlichdimensionaler Teilraum von V, P der orthogonale Projektor auf U und $v \in V$ beliebig. Dann nennt man die Zerlegung

$$v = Pv + P^{\perp}v$$

die orthogonale Zerlegung von v bezüglich U. Für sie ist

$$Pv \in U \quad \text{und} \quad P^{\perp}v \perp U.$$

Mit der vom Skalarprodukt induzierten Norm gilt der verallgemeinerte Satz des Pythagoras,

$$\|v\|^2 \;=\; \|Pv\|^2 + \|P^{\perp}v\|^2.$$

Beweis: Es ist $v = v - Pv + Pv = P^{\perp}v + Pv \in U$, und wir haben $Px \in U$ und $P^{\perp}x \perp U$ nach dem vorangegangenen Satz. Weiter gilt

$$\begin{aligned} \|v\|^2 = \langle v, v\rangle &= \langle Pv + P^{\perp}v, Pv + P^{\perp}v\rangle \\ &= \langle Pv, Pv\rangle + 2\langle \underbrace{Pv}_{\in U}, \underbrace{P^{\perp}v}_{\in U^{\perp}}\rangle + \langle Pv, Pv\rangle = \|Pv\|^2 + \|P^{\perp}v\|^2. \end{aligned}$$

□

Satz 14.4.6. *(Existenz und Eindeutigkeit von Bestapproximationen in Unterräumen)*

Sei V ein Skalarproduktraum, U ein endlichdimensionaler Teilraum und $v \in V$ beliebig. Dann existiert eine eindeutige Bestapproximation von v in U, d.h. es gibt einen eindeutigen Vektor $u^* \in U$, der

$$\|v - u^*\| \;=\; \min\{\, \|v - u\| \;\mid u \in U\}$$

erfüllt: Ist P der orthogonale Projektor auf U, so ist gerade $u^* = Pv$.

Beweis: Für alle $u \in U$ gilt nach Satz 14.4.5

$$\|v-u\|^2 = \|P(v-u)\|^2 + \|P^\perp(v-u)\|^2 = \|Pv-u\|^2 + \|P^\perp v\|^2,$$

da P und $P^\perp$ linear sind und mit $u \in U$ auch $Pu = u$ und $P^\perp u = 0$ ist. Der hintere Term ist unabhängig von u, der vordere ist nichtnegativ und genau dann Null, wenn $u = Pv$ gilt.

□

Eine Bemerkung zum Schluss: Wir haben im Rahmen der hier betrachteten Theorie nur senkrechte Projektionen betrachtet, bei denen man sich die Abbildung mit Hilfe von Projektionsstrahlen vorstellen kann, die *senkrecht* auf die Projektionsfläche (also den Unterraum U) treffen. Weitere wichtige, weil viel verwendete Typen von Projektionen sind die der Parallelprojektionen (hier sind die Projektionsstrahlen auch parallel, aber treffen nicht zwingend senkrecht auf die Projektionsfläche) und die der Zentralprojektion (hier gehen die Strahlen von einem festen Punkt aus).

- Allgemeine Parallelprojektionen $P : V \to V$ sind ebenfalls lineare Abbildungen, für die ebenfalls für einen geeigneten Unterraum U die Beziehungen $\mathrm{Bild}(P) = U$ und $Pu = u$ für alle $u \in U$ gelten (vgl. die Definition des orthogonalen Projektors auf U). Im Gegensatz zu orthogonalen Projektionen ist aber $\mathrm{Kern}(P) \neq U^\perp$; stattdessen gibt es einen zu U komplementären Unterraum $W \subseteq V$ (d.h. $U + W = V$) so, dass $Kern(P) = W$ ist. Den Unterraum W kann man sich als „Projektionsrichtung“ oder „Richtung der Projektionsstrahlen“ (z.B. beim Schattenwurf) vorstellen.

- Zentralprojektionen lassen sich ebenfalls mit den Techniken der Linearen Algebra berechnen, passen aber nicht direkt in unsere Theorie der linearen Abbildungen.

14.5 Skalarprodukt, Längen, Abstände und Winkel

Wir wollen nun allgemein beweisen, dass ein Skalarprodukt *immer* benutzt werden kann, um auf dem entsprechenden Vektorraum eine Längen- und Abstandsmessung mit „vernünftigen“ Eigenschaften zu definieren, und in euklidischen Vektorräumen auch dafür herhalten kann, ganz allgemein einen Winkel zwischen zwei Vektoren eines beliebigen Vektorraumes zu definieren.

Wir beginnen damit, uns noch einmal vertieft mit den durch das Skalarprodukt induzierten Längen und Abständen zu beschäftigen. Die Formalisierung und Abstraktion des Begriffs der Länge ist die *Norm*, die des Abstandsbegriffs die *Metrik*.

Definition 14.5.1. *(Norm)*

Sei V ein $\mathbb{R}$- oder $\mathbb{C}$-Vektorraum. Eine Abbildung $\|\cdot\| : V \to \mathbb{R}$ (!, auch über $\mathbb{C}$) heißt Norm auf V, wenn gilt:

(i) $\|v\| \geq 0; \qquad \|v\| = 0 \Leftrightarrow v = 0_V$, *für alle* $v \in V$ *($\|\cdot\|$ ist positiv),*

(ii) $\|a \cdot v\| = |a| \cdot \|v\|$ *für alle* $v \in V$ *und* $a \in \mathbb{R}$ *bzw.* $a \in \mathbb{C}$ *($\|\cdot\|$ ist skalierbar),*

(iii) $\|v + w\| \leq \|v\| + \|w\|$ *für alle* $v, w \in V$ *(Dreiecksungleichung).*

Das Paar $(V, \|\cdot\|)$ heißt normierter Vektorraum.

Bemerkungen dazu:

- Nicht jede Norm ist durch ein Skalarprodukt induziert, es gibt in der Tat „sehr viel mehr“ normierte Räume als Skalarprodukträume.
- Das Standardbeispiel für eine Norm auf den Vektorräumen $\mathbb{R}$ und $\mathbb{C}$ sind jeweils die Betragsfunktion $x \mapsto |x|$ bzw. $z = a + ib \mapsto |z| = (a^2 + b^2)^{\frac{1}{2}}$.
- Auf $\mathbb{K}^n$ definieren beispielsweise

  $$\|x\|_1 := \sum_{i=1}^{n} |x_i| \quad \text{und} \quad \|x\|_\infty := \max_{i \in \underline{n}} |x_i|$$

  verschiedene Normen für Vektoren aus $\mathbb{R}^n$, und der $\mathbb{K}^n$ ist mit der 1-Norm und auch mit der ∞-Norm ein normierter Vektorraum, bei dem die Norm nicht durch ein Skalarprodukt induziert wird.

- Allgemeiner ist $\mathbb{K}^n$ für $1 \leq p < \infty$ mit jeder Norm
$$\|\cdot\|_p := \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}$$
ein normierter Raum. Für $p = 1, p = 2$ und den Grenzfall $p = \infty$ erkennt man hoffentlich bekannte Normen, nur der Fall $p = 2$ kommt von einem Skalarprodukt (dem kanonischen). Die Wahl einer geeigneten Norm auf V hängt für gewöhnlich von der Art des Problems ab, das man mit den betrachteten Vektoren beschreibt.

- Es lässt sich zeigen, dass sich in jedem normierten Vektorraum $(V, \|\cdot\|)$, dessen Norm $\|\cdot\|$ die so genannte Parallelogrammgleichung für alle $v, w \in V$ erfüllt, ein Skalarprodukt definieren lässt, das die Norm $\|\cdot\|$ induziert (dieses definert man dann wiederum über die rechte Seite der Polarisationsformel aus Lemma 14.5.4).

Definition 14.5.2. *(Metrik)*

Sei M eine Menge. Eine Abbildung $d : M \times M \to \mathbb{R}$ mit den Eigenschaften

(i) $d(P, Q) \geq 0$ *für alle* $P, Q \in M$ *und* $d(P, Q) = 0 \Leftrightarrow P = Q$
(d ist positiv und trennt Punkte),

(ii) $d(P, Q) = d(Q, P)$ *für alle* $P, Q \in M$ *(Symmetrie),*

(iii) $d(P, Q) \leq d(P, R) + d(R, Q)$ *für alle* $P, Q, R \in M$ *(Dreiecksungleichung).*

heißt Metrik auf M.

Bemerkung dazu:

- Das Konzept der Metrik ist sehr allgemeine axiomatische Formulierung der Ideen der Abstandsmessung und nicht an einen Vektorraum als „zu metrisierende“ Menge gebunden. Es gibt viele verschiedene Metriken auf verschieden gearteten Mengen, und nicht durch Normen induziert werden. Zum Beispiel ist auf jeder Menge M die Funktion
$$d : \quad M \to \mathbb{R}, \qquad d(x, y) = \begin{cases} 0 & \text{falls} \quad x = y \\ 1 & \text{sonst} \end{cases}$$
eine „andere“ Metrik (die auch die Definition einer Metrik erfüllt); ein weiteres lustiges Beispiel für so eine Metrik ist die „Französische Eisenbahnmetrik“ (Link).

Der angekündigte Zusammenhang allgemeine Zusammenhang in Skalarprodukträumen ist nun der folgende:

Satz 14.5.3. *(Skalarprodukt induziert Norm induziert Metrik)*

(i) *Sei $(V, \langle\cdot,\cdot\rangle)$ ein Skalarproduktraum. Dann ist die durch $\langle\cdot,\cdot\rangle$ induzierte Norm*

$$V \to \mathbb{R}, \qquad v \mapsto \|v\| := \sqrt{\langle v, v\rangle}.$$

wirklich eine Norm im Sinne der obigen Definition.

(ii) *Sei $(V, \|\cdot\|)$ ein normierter Vektorraum und $(A, V, +)$ ein affiner Raum über V. Dann ist die Abbildung*

$$d_{\|\cdot\|} : A \times A \to \mathbb{R}, \qquad d(P,Q)_{\|\cdot\|} := \|\overrightarrow{PQ}\|$$

eine Metrik auf A, die so genannte durch $\|\cdot\|$ induzierte Metrik.

Beweis: S. Vorlesung. Man stellt fest, dass die Symmetrie/Hermitizität und Positivdefinitheit des Skalarprodukts die Wohldefinitertheit und Positivität der Norm implizieren, und dass die Skalarität aus der Linearität bzw. Sesquilinearität folgt. Die Dreiecksungleichung beweist man mit Hilfe der Cauchy-Schwarz-Ungleichung; das liefern wir sofort nach, nachdem wir im nächsten Abschnitt diese Ungleichung bewiesen haben. (ii) ist Übungsaufgabe. □

Beispiele für induzierte Normen:

- Wählt man das skalierte Skalarprodukt $\langle x, y\rangle_n$, so ist die induzierte Norm gegen durch

$$\|x\|_n = \frac{1}{\sqrt{n}} \left(\sum_{i=1}^{n} x_i^2 \right)^{\frac{1}{2}},$$

das ist das so genannte quadratische Mittel über die Einträge von x. Der induzierte Abstand zweier Vektoren x, y im affinen Raum $\mathbb{R}^n$ (über sich selbst) ist die mittlere quadratische Abweichung

$$\|x - y\|_n = \frac{1}{\sqrt{n}} \left(\sum_{i=1}^{n} (x_i - y_i)^2 \right)^{\frac{1}{2}}$$

zwischen x und y.

- Ganz analog kann man das Integral

$$\|f\|_I := (\frac{1}{b-a})^{\frac{1}{2}} \left(\int_a^b f^2(x)dx \right)^{\frac{1}{2}}$$

als quadratisches Mittel über die Werte von f und die Metrik

$$d(f,g) := \frac{1}{(b-a)^{\frac{1}{2}}} \left(\int_a^b (f-g)^2(x)dx \right)^{\frac{1}{2}}$$

als mittlere quadratische Abweichung zwischen f und g interpretieren.

- Nur so am Rande: Zwischen dem quadratischen Mittel für Vektoren und für Funktionen besteht eine sehr enge Verbindung. Ist im vorigen Beispiel $[a,b] = [0,1]$ und $x \in \mathbb{R}^n$ ein Vektor von Funktionswerten („Samples“) $x_i = f(\frac{i}{n})$ von f im Abstand $\Delta x = \frac{1}{n}$, so gilt für wachsende „Samplerate“

$$\|x\|^2 = \sum_{i=1}^n f(x_i) \cdot \Delta x \quad \overset{n\to\infty}{\longrightarrow} \int_0^1 f(x)dx = \|f\|_{[0,1]}.$$

Lemma 14.5.4. *(Rechenregeln: Binomische Formeln, Polarisationsformel)*

Sei $(V, \langle\cdot,\cdot\rangle)$ ein euklidischer Raum, und $\|\cdot\|$ die von $\langle\cdot,\cdot\rangle$ induzierte Norm. Dann gelten für alle $v, w \in V$

(i) die <u>binomischen Formeln</u> für $\mathbb{R}$-Vektorräume,

$$\|v \pm w\|^2 \;=\; \|v\|^2 \pm 2\langle v,w\rangle + \|w\|^2, \qquad \langle v+w, v-w\rangle = \|v\|^2 - \|w\|^2$$

und die

(ii) <u>Polarisationsformel</u> für $\mathbb{R}$-Vektorräume,

$$4\langle v,w\rangle = \|v+w\|^2 - \|v-w\|^2.$$

Ist $(V, \langle\cdot,\cdot\rangle)$ ein unitärer Raum und $\|\cdot\|$ die induzierte Norm, so gelten für $v, w \in V$ die analogen Formeln

(i) $$\|v\pm w\|^2 \;=\; \|v\|^2 \pm 2Re\langle v,w\rangle + \|w\|^2, \qquad \langle v+w, v-w\rangle \;=\; \|v\|^2 - 2iIm\langle v,w\rangle - \|w\|^2$$

und

(ii) $$4\langle v,w\rangle = \|v+w\|^2 - \|v-w\|^2 + i\|iv+w\|^2 - i\|iv-w\|^2.$$

Beweis: Beides zeigt man durch Rückführung auf das Skalarprodukt und Ausnutzen der Bilinearität bzw. Sesquilinearität. Übungsaufgabe.

□

Bemerkung dazu:

- Das Bemerkenswerte an diesen Binomischen Formeln für Skalarprodukte wie an der Polarisationsformel ist, dass man in ihnen einen Zusammenhang zwischen Skalarprodukten (und damit Winkel zwischen Vektoren) und Längen erhält.

Wir beweisen als nächstes ein Handwerkszeug, das bei der tieferen Beschäftigung mit Skalarprodukträumen fundamental ist. Im Anschauungsraum hatten wir heuristisch hergeleitet, dass dort

$$\langle x, y\rangle \quad = \quad \cos(\gamma)\ \|x\|_2\ \|y\|_2 \quad \leq \quad \|x\|_2\|y\|_2$$

gilt. Diese Herleitung war mit der Eigenschaft des Winkelmessens im Anschauungsraum verbunden und wurde dann benutzt, um zu zeigen, dass die induzierte euklidische Norm im $\mathbb{R}^3$ die in 14.1.1 angegebenen Eigenschaften hat.[5] Letztendlich rührte der Zusammenhang aber daher, dass im rechtwinkligen Dreieck, das in der Herleitung benutzt wurde, für den Vektor $\vec{x}^{\shortparallel}$, den man durch orthogonale Projektion des Vektors x auf den Vektor $\vec{y}$ gewonnen hatte, $\|\vec{x}^{\shortparallel}\| \leq \|\vec{x}\|$ gilt. Für allgemeine Skalarprodukte sieht der analoge Zusammenhang folgendermaßen aus:[6]

Satz 14.5.5. *(Cauchy-Schwarz-Bunjakowski'sche Ungleichung)*

Sei V ein Skalarproduktraum mit Skalarprodukt $\langle\cdot,\cdot\rangle$ und $\|\cdot\|$ die induzierte Normabbildung. Dann gilt für alle $v, w \in V$ die oft als Cauchy-Schwarz-Ungleichung bezeichnete Beziehung

$$|\langle v, w\rangle| \quad \leq \quad \|v\| \cdot \|w\|.$$

Gleichheit gilt genau dann, wenn v und w linear abhängig sind, also für $v \neq 0_V$ genau dann, wenn $w = a \cdot v$ ist.

[5] Daher war unser Beweis in im Lichte der Ausführungen in 12.2.1 auch nicht ganz lupenrein, aber die auch für $\mathbb{R}^3$ gültige Cauchy-Schwarz-Ungleichung zeigt, dass wir auch in diesem Fall axoimatisch-deduktiv vorgehen können, um zu beweisen, dass die induzierte Norm im $\mathbb{R}^3$ wirklich eine ist.

[6] Die folgende Ungleichung wird nach Augustin-Louis Cauchy (1789–1857), Hermann Amandus Schwarz (1843–1921) und manchmal auch nach Viktor Jakowlewitsch Bunjakowski (1804–1889) genannt. Sie wurde von Cauchy für Summen und von seinem Schüler Bunjakowski für allgemeinere Fälle bewiesen und benutzt, bei Schwarz taucht sie erst 50 Jahre später auf.

Beweis: Im Fall $v = 0_V$ erhalten wir das Ergebnis wegen $\langle 0_V, w\rangle = 0 = \|0_V\| \cdot \|w\|$ sofort. Im Fall $v \neq 0_V$ betrachten die Projektion von w in Richtung v. Dazu normieren wir v, der Projektor P_v auf den Aufspann von v ist dann gegeben durch

$$P_v(w) \quad = \quad \langle \frac{v}{\|v\|}, w\rangle \cdot \frac{v}{\|v\|} \quad = \quad \langle v, w\rangle \cdot \frac{v}{\|v\|^2}.$$

Nach dem Satz des Pythagoras (14.4.5) gilt nun

$$\|w\|^2 \;=\; \|P_v(w)\|^2 + \|P_v^{\perp}(w)\|^2 \;\geq\; \|P_v(w)\|^2 \;=\; \left(|\langle v, w\rangle| \cdot \frac{\|v\|}{\|v\|^2}\right)^2 = \frac{|\langle v, w\rangle|^2}{\|v\|^2},$$

und Gleichheit gilt genau dann, wenn $P_v^{\perp}(w) = 0$, also wenn $w \in \mathcal{L}(v)$.

□

Bemerkungen dazu:

- Für Vektoren $x, y \in \mathbb{K}^n$ bzw. Funktionen $f, g \in C([a, b], \mathbb{R})$ hat die Cauchy-Schwarz'sche Ungleichung die Form

$$\left(\sum_{i=1}^{n} x_i y_i\right)^2 \quad \leq \quad \left(\sum_{i=1}^{n} |x|_i^2\right)\left(\sum_{i=1}^{n} |y|_i^2\right)$$

beziehungsweise

$$\left(\int_a^b f(x)g(x)dx\right)^2 \leq \left(\int_a^b f^2(x)dx\right)\left(\int_a^b g^2(x)dx\right).$$

Schreiben Sie das Linke mal für Vektoren aus dem $\mathbb{R}^3$ aus – dann wird hoffentlich deutlich, dass diese Beziehung in dieser Form nicht unbedingt offensichtlich ist!

Bemerkungen dazu:

- Mit Hilfe dieser Ungleichung lässt sich nun beweisen, dass die induzierte Normabbildung auch die Dreiecksungleichung erfüllt und somit alle in 14.5.1 genannten Eigenschaften für induzierte Normen immer erfüllt sind. Das beweist Satz satz:skalnormd zuende, s. Vorlesung.

Definition 14.5.6. *(Winkel zwischen Vektoren)*

Die Cauchy-Schwarz'sche Ungleichung ist äquivalent dazu, dass für alle Vektoren $v, w \neq 0_V$ in einem euklidischen Vektorraum $(V, \langle \cdot, \cdot \rangle)$ gilt:

$$-1 \quad \leq \quad \frac{\langle v, w \rangle}{\|v\|\|w\|} \quad \leq \quad 1.$$

Somit existiert genau ein $\gamma \in [0, \pi]$ so, dass

$$\cos \gamma = \frac{\langle v, w \rangle}{\|v\|\|w\|}.$$

Wir definieren diese Zahl γ als den Winkel $\angle(v, w)$ zwischen den Vektoren v und w, also

$$\angle(v, w) := \arccos \frac{\langle v, w \rangle}{\|v\|\|w\|}.$$

Bemerkung dazu:

- Die vorangegangene Definition bedeutet, dass man mit dem Skalarprodukt z.B. auch Winkel zwischen Funktionen $f, g \in C(I)$ messen kann. Das entbehrt (wie auch ein Winkel im $\mathbb{R}^{42}$) jeder räumlich-anschaulichen Vorstellung, hat sich aber, insbesondere für das Konzept des Senkrechtstehens, als sinnvolle Erweiterung des Konzepts ds Winkels herausgestellt.

Mit Hilfe der vorgestellten allgemeinen Definitionen von Längen (also Normen) und Winkeln in euklidischen Vektorräumen lassen sich viele Sätze der klassischen Geometrie auch in Vektorräumen formulieren. In den Übungen tun wir das für den Satz des Thales; hier soll der folgende Satz als Beispiel dienen:

Lemma 14.5.7. *(Verallgemeinerter Cosinussatz in euklidischen Vektorräumen)*

Sei V ein euklidischer Vektorraum. Mit der obigen Definition (14.5.6) des Winkels zwischen zwei Vektoren aus V bzw. seines Cosinus gilt der verallgemeinerte Kosinussatz: Sind a, b und $c := b - a \in V$, so gilt

$$\|c\|^2 = \|a\|^2 + \|b\|^2 - 2 \cdot \|a\|\|b\| \cos \angle(a, b).$$

Beweis: Übungsaufgabe.

□

14.6 Bilinear- und Sesquilinearformen und ihre Darstellung durch Matrizen

Skalarprodukte und allgemeiner Bilinear- bzw. Sesquilinearformen auf endlichdimensionalen Vektorräuen lassen sich wie lineare Abbildungen effizient mit Hilfe von Matrizen darstellen. Die Eigenschaften des Skalarprodukts spiegeln sich dabei in Eigenschaften der darstellenden Matrix wieder.

Bevor wir damit beginnen, das genauer zu formulieren, definieren wir zunächst noch einmal allgemein:

Definition 14.6.1. *(Bilinearform, Sesquilinearform)*

Sei V ein $\mathbb{R}$-Vektorraum. Eine Abbildung $b : V \times V \to \mathbb{R}$ heißt <u>Bilinearform</u> auf V, falls für alle $v \in V$ die Abbildungen

$$b_1 : V \to \mathbb{R}, \quad w \mapsto b(w, v) \qquad \textit{und} \qquad b_2 : V \to \mathbb{R}, \quad w \mapsto b(v, w)$$

$\mathbb{R}$-linear sind, man sagt auch: „b ist linear in beiden Komponenten“.

Sei V ein $\mathbb{C}$-Vektorraum. Eine Abbildung $b : V \times V \to \mathbb{C}$ heißt <u>Sesquilinearform</u> auf V, falls für alle $v \in V$ die Abbildung

$$b_2 : V \to \mathbb{C}, \quad w \mapsto b(v, w)$$

$\mathbb{C}$-linear ist („b ist linear in der zweiten Komponente“) und

$$b_1 : V \to \mathbb{C}, w \mapsto b(w, v)$$

semilinear ist („b ist semilinear in der ersten Komponente“), d.h. es gilt

$$b(w_1 + w_2, v) = b(w_1, v) + b(w_2, v) \qquad \textit{und} \qquad b(a \cdot w, v) = \overline{a} \cdot b(w, v)$$

für alle $w_1, w_2 \in V$ und $a \in \mathbb{C}$.

Der grundlegende Zusammenhang zwischen diesen „Formen“ und Matrizen ist der folgende.

Lemma 14.6.2. *(Darstellungssatz für Bilinearformen/ Sesquilinearformen in $\mathbb{K}^n$)*
Sei $\mathbb{K} = \mathbb{R}$ bzw. $\mathbb{C}$ und $b : \mathbb{K}^n \times \mathbb{K}^n \to \mathbb{K}$ eine Bilinear- bzw. Sesquilinearform. Dann gibt es eine eindeutig bestimmte Matrix $B \in \mathbb{K}^{n\times n}$ so, dass für alle $x, y \in \mathbb{K}^n$ mit dem jewiligen Standardskalarprodukt

$$b(x, y) = \langle x, By\rangle$$

gilt, d.h. es ist

$$b(x, y) = x^T By \qquad \textit{beziehungsweise} \qquad b(x, y) = x^H By$$

im Fall $K = \mathbb{R}$ beziehungsweise $\mathbb{K} = \mathbb{C}$. Umgekehrt definiert jede Matrix $B \in \mathbb{K}^{n\times n}$ via $b(x, y) := \langle x, By\rangle$ eine zugehörige „kanonische“ Bilinearform bzw. Sesquilinearform.

Beweis: Wir beweisen die Behauptung für den komplexen Fall. Den reellen Fall beweist man analog, indem man einfach die komplexen Konjugationen weglässt und „H“ durch „T“ ersetzt. Ist $B = (b_{i,j})_{i,j\in\underline{n}} \in \mathbb{K}^{n\times n}$ eine Matrix und

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i=1}^{n} x_i e_i, \qquad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \sum_{i=1}^{n} y_i e_i \in \mathbb{C}^n,$$

so gilt

$$x^H By = (\overline{x_i})_{i\in\underline{n}} \cdot (\sum_{j=1}^{n} b_{i,j} y_j)_{i\in\underline{n}} = \sum_{i=1}^{n}\sum_{j=1}^{n} \overline{x_i} y_j b_{i,j}.$$

Gleichzeitig ist mit der Sesquilinearität von b

$$b(x, y) = b(\sum_{i=1}^{n} x_i e_i, \sum_{i=1}^{n} y_i e_i) = \sum_{i=1}^{n}\sum_{j=1}^{n} \overline{x_i} y_j b(e_i, e_j),$$

es ist also $b(x, y) = x^H By = \langle x, By\rangle$ für alle $x, y \in \mathbb{K}^n$ genau dann, wenn $b_{i,j} = b(e_i, e_j)$ ist. Das definiert die gesuchte Matrix B eindeutig. Ist umgekehrt $B \in \mathbb{K}^{n\times n}$, so rechnet man die Bilinearität der Abbildung b relativ einfach nach, wir sparen uns hier die Details. □

Bemerkungen dazu:

- Eine Bilinear- bzw. Sesquilinearform auf $\mathbb{K}^n$, insbesondere ein Skalarprodukt, ist also durch Angabe der n^2 Skalarprodukte zwischen den Standardbasisvektoren vollständig festgelegt. Das gilt auch allgemeiner für beliebige Basen und beliebige endlichdimensionale Vektorräume, s.u.

Satz 14.6.3. *(Darstellung von Skalarprodukten des $\mathbb{K}^n$ mit Gram'schen Matrizen)*

Sei $\mathbb{K} = \mathbb{R}$ bzw. $\mathbb{C}$, $b : \mathbb{K}^n \times \mathbb{K}^n \to \mathbb{K}$ eine Bilinear- bzw. Sesquilinearform und B die zugehörige Matrix aus Lemma 14.6.2.

Dann ist b ein Skalarprodukt auf $\mathbb{K}^n$ genau dann, wenn B symmetrisch bzw. hermitesch ist (vgl. Def. 14.3.4) und positiv definit ist. Letzteres bedeutet: Mit dem Standardskalarprodukt ist

$$\langle x, Bx \rangle > 0 \qquad \text{für alle} \qquad x \in \mathbb{K}^n \backslash \{0_V\}.$$

Ist B die zu einem Skalarprodukt symmetrisch positiv definite Matrix gehörige Matrix, so nennt man B die Gram'sche Matrix dieses Skalarprodukts.

Bemerkungen dazu:

- Die Gram'sche Matrix der kanonischen Skalarprodukte in $\mathbb{R}^n$ und $\mathbb{C}^n$ ist jeweils die Einheitsmatrix I_n.
- Da auch im Fall $\mathbb{K} = \mathbb{C}$ wegen der Hermitizität immer $\langle x, Bx \rangle \in \mathbb{R}$ gilt, macht die Ungleichung $\langle x, Bx \rangle > 0$ oben überhaupt Sinn.
- Die Symmetrie einer Matrix lässt sich ohne weiteres erkennen. Schwieriger ist es mit der Positivdefinitheit, die sich zwar bei Diagonalmatrizen direkt ablesen lässt (vgl. Übungen), aber im allgemeinen Fall anhand der Eigenwerte der Matrix B entschieden werden muss, s. Kapitel 17, insbesondere Abschnitt 18.4.

Bevor wir Satz 14.6.3 beweisen, schieben wir noch kurz ein:

Lemma 14.6.4. *(Rechenregeln für Transponierte und Adjungierte)*

(i) Ist $A \in \mathbb{K}^{m\times n}$, so gilt

$$(A^H)^H = A, \qquad (A+B)^H = A^H + B^H, \qquad (A \cdot B)^H = B^H \cdot A^H.$$

Im Fall $A \in \mathbb{R}^{m\times n}$ ist ja nach den Bemerkungen zu 14.3.4 $A^T = A^H$, also bedeuten diese Regeln dort $(A^T)^T = A$ usw.

(ii) Sei $A \in \mathbb{K}^{n\times n}$. Dann ist A symmetrisch bzw. hermitesch (für $\mathbb{K} = \mathbb{R}$ bzw. $\mathbb{K} = \mathbb{C}$) genau dann, wenn mit dem jeweiligen Standardskalarprodukt

$$\langle Ax, y\rangle = \langle x, Ay\rangle$$

für alle $x, y \in \mathbb{K}^n$ gilt.

Beweis: (a) war im reellen Fall schon eine Übungsaufgabe; den komplexen beweist man analog durch Betrachten der einzelnen Einträge. In (b) beweisen wir wieder den komplexen Fall: Ist $A = A^H$, so ist

$$\langle Ax, y\rangle := (Ax)^H y = x^H A^H x = x^H Ay = \langle x, Ay\rangle.$$

Für die Umkehrung halten wir zunächst fest, dass für die Einträge einer Matrix $A \in \mathbb{K}^{n\times n}$ für die Vektoren der jeweiligen Standardbasis $\langle e_i, Ae_j\rangle = a_{i,j}$ gilt. Ist nun $\langle Ax, y\rangle = \langle x, Ay\rangle$ für alle $x, y \in \mathbb{K}^n$, so ist insbesondere

$$a_{i,j} = \langle e_i, Ae_j\rangle = \langle Ae_i, e_j\rangle = \overline{\langle e_j, Ae_i\rangle} = \overline{a_{j,i}},$$

also ist A hermitesch. $\square$

Beweis: Wir beweisen wieder nur den komplexen Fall und zeigen, dass die Hermitizität von b zu der von B äquivalent ist, und dass und die Positivdefinitheit von b zu der von B äquivalent ist. Letzteres ist wegen $b(x, y) = \langle x, By\rangle$ offensichtlich. Zur Hermitizität: Ist B hermitesch, so ist auch b hermitesch, denn es gilt mit der Eigenschaft (ii) für hermitesche Matrizen aus dem Lemma oben und mit der Hermitezität in des kanonischen Skalarproduktes in $(\star)$, dass

$$b(x, y) = \langle x, By\rangle \overset{(ii)}{=} \langle Bx, y\rangle \overset{(\star)}{=} \overline{\langle y, Bx\rangle} = \overline{b(y, x)}.$$

Ist b hermitesch, so erhält man für alle $x, y \in \mathbb{C}^n$, dass

$$\langle x, By \rangle = b(x, y) = \overline{b(y, x)} = \overline{\langle y, Bx \rangle} \overset{(\star)}{=} \overline{\langle Bx, y \rangle}$$

gilt; B ist also hermitesch nach Lemma 14.6.4.

□

Wir verallgemeinern die oben entwickelte Theorie auf allgemeine endlichdimensionale Vektorräume.

Satz 14.6.5. *(Bilinear- und Sesquilinearformen und Skalarprodukte durch Matrizen)*

Sei V ein endlichdimensionaler Vektorraum über $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$, $b : V \times V \to \mathbb{K}$ eine Abbildung und $C = (c_1, \ldots, c_n)$ eine Basis von V. Dann ist analog zur oben entwickelten Theorie b bilinear genau dann, wenn b folgendermaßen durch die Matrix

$$B_C = (b(c_i, c_j))_{i,j \in \underline{n}}$$

darstellbar ist: Bezeichnen wir für beliebiges $v \in V$ seinen zugehörigen Koordinatenvektor bezüglich C mit $k_C(v)$, so gilt für alle $v, w \in V$

$$b(v, w) \quad = \quad k_C(v)^H \cdot B_C \cdot k_C(w).$$

Auch in diesem Fall gilt: Eine Bilinearform b ist ein Skalarprodukt auf V genau dann, wenn die Matrix B_C symmetrisch bzw. hermitesch und positiv definit ist; sie heißt in diesem Fall Gram'sche Matrix von b bezüglich der Basis C.

Beweis: Den Beweis werde ich in der Vorlesung nur skizzieren.

□

Bemerkung dazu:

- Bei der allgemeinen Definition von Gram'schen Matrizen B_C ist etwas Sorgfalt angebracht: Diese sind, wie der Wortlaut der Definition und der zugehörige Index zeigt, abhängig von der Wahl einer bestimmten Basis C. Bei Gram'schen Matrizen im $\mathbb{K}^n$ ist das als Spezialfall genauso zu; es gibt also je nach Wahl der Basis C des $\mathbb{K}^n$ auch dort viele verschiedene Gram'sche Matrizen. Die in Lemma definierte Gram'sche Matrix des Skalarproduktes $b : \mathbb{K}^n \times \mathbb{K}^n \to \mathbb{K}$ ist im Sinne des obigen Satzes genauer gesagt die Gram'sche Matrix B_{I_n} bezüglich der Standardbasis $I_n = (e_1, \ldots, e_n)$ des $\mathbb{K}^n$. In der Literatur ist diese Namensgebung leider eher etwas Wischi-Waschi; es lohnt sich also, genau darauf zu achten, was dort eigentlich steht und was gemeint ist.

Abschließend zeigen wir, dass sich zu jeder Basis C eines endichdimensionalen Vektorraumes V ein Skalarprodukt definieren lässt, so dass C in diesem Skalarprodukt eine Orthonormalbasis ist. Das zu prüfen, ist eine Übungsaufgabe; das folgende Lemma werde ich in der Vorlesung nur kurz erwähnen.

Lemma/Definition 14.6.6. *(Kanonisches Skalarprodukt zu gegebener Basis)*
Sei V ein endlichdimensionaler Skalarproduktraum, $C = (c_1, \dots, c_n)$ eine Basis von V und für $v \in V$ der Koordinatenvektor $k_C(v)$ wie oben. Die Matrix I_n ist offensichtlich symmetrisch und positiv definit. Die Abbildung

$$b_{\mathbb{R}} : \ V \times V \to \mathbb{R}, \qquad (v, w) \mapsto (k_C(v))^T I_n k_C(w) \ = \ (k_C(v))^T k_C(w)$$

im euklidischen bzw.

$$b_{\mathbb{C}} : \ V \times V \to \mathbb{C}, \qquad (v, w) \mapsto (k_C(v))^H I_n k_C(w) \ = \ (k_C(v))^H k_C(w)$$

im unitären Fall definiert nach Lemma 14.6.5 ein Skalarprodukt auf V, das sogenannte kanonische Skalarprodukt für die Basis C.

Bemerkungen dazu:

- Das kanonische Skalarprodukt für eine Basis C ist genau so gebaut, dass für die Basisvektoren aus C die Relationen

  $$s(c_i, c_j) = 0 \ \text{ für } \ i \neq j \qquad \text{und} \qquad \|b_i\|_s^2 := s(c_i, c_i) = 1$$

  gilt, dass sie also eine Orthonormalbasis bezüglich des zu C gehörigen kanonischen Skalarproduktes bilden.

- Im $\mathbb{K}^n$ ist das kanonische Skalarprodukt für die Standardbasis die Abbildung $(x, y) \mapsto x^T y$ bzw. $(x, y) \mapsto x^H y$ (also einfach das Standardskalarprodukt). Das kanonische Skalarprodukt für eine andere Basis C erhält man daraus mit der Abbbildung $(x, y) \mapsto (C^{-1}x)^T (C^{-1}) y$ (wenn wie immer C die Matrix mit den Vektoren aus C als Spalten bezeichnet; Übungsaufgabe). Allgemeiner lässt sich jedes Skalarprodukt mit Hilfe so einer Transformationsmatrix durch jedes andere ausdrücken, vgl. z.B. [BeuLA], S. 324ff.

14.7 Aufgaben zu Kapitel 14

Aufgabe 14.77. (Skalarprodukte im Anschauungsraum und in $\mathbb{R}^n$):

(a) Beweisen Sie elementargeometrisch, also ohne Verwendung der Vektorrechnung/Linearen Algebra, anhand einer Skizze den Cosinussatz:

Gegeben sei in der Anschauungsebene $\mathcal{A}_2$ ein Dreieck mit Seitenlängen $a, b, c \in \mathbb{R}^+$. Der Innenwinkel, der der Seite mit der Länge c gegenüberliegt, habe die Größe $\gamma \in (0, \pi)$. Dann gilt:

$$c^2 = a^2 + b^2 - 2ab\cos(\gamma).$$

Hinweis: Sie dürfen die Definitionen von Sinus und Cosinus sowie bekannte trigonometrische Zusammenhänge, wie zum Beispiel Additionstheoreme, verwenden.

(b) Zeigen Sie, dass die Verallgemeinerung des euklidischen Skalarprodukts auf den $\mathbb{R}^n$, also die Abbildung $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ mit

$$\langle x, y \rangle := \sum_{i=1}^{n} x_i y_i \quad \text{für} \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n,$$

die folgenden Eigenschaften hat (zu den Begriffen vgl. Bemerkungen am Ende des Abeitsblattes oder Vorlesung vom Mittwoch, 12.5.21):

(i) $\langle \cdot, \cdot \rangle$ ist bilinear,

(ii) $\langle \cdot, \cdot \rangle$ ist positiv definit,

(iii) $\langle \cdot, \cdot \rangle$ ist symmetrisch.

(c) Begründen Sie, dass $\|x\| := \sqrt{\langle x, x \rangle}$ für alle $x \in \mathbb{R}^n$ wohldefiniert ist. Zeigen sie, dass für alle $a \in \mathbb{R}$ und $x \in \mathbb{R}^n$

$$\|a \cdot x\| = |a| \cdot \|x\|$$

sowie

$$\|x\| \geq 0 \quad \text{und} \quad \|x\| = 0 \iff x = 0_{\mathbb{R}^n}.$$

gilt.

(d) Das kanonische Skalarprodukt im $\mathbb{R}^n$ aus (b) und das skalierte Skalarprodukt

$$\langle x, y\rangle_n := \frac{1}{n}\left(\sum_{i=1}^{n} x_i y_i\right)$$

lassen sich in der Form $(\star)$ aus Aufgabe 5.4 schreiben. Geben Sie (ohne Begründung) die entsprechenden Matrizen an.

Aufgabe 14.78. (Durch Matrizen definierte Bilinearformen):

(a) Sei $A \in \mathbb{R}^{n\times n}$ eine Matrix. Zeigen Sie, dass die Abbildung

$$b_A\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \quad (x, y) \mapsto x^{\mathrm{T}} \cdot A \cdot y \qquad (\star)$$

stets bilinear ist (s.u.).

(b) Bestimmen Sie jeweils eine Matrix A, für die b_A nicht symmetrisch bzw. nicht positiv definit ist (s.u.).

Aufgabe 14.79. (Spurprodukt):

Wir betrachten den reellen Vektorraum $\mathbb{R}^{n\times n}$. Die Spur einer Matrix $A \in \mathbb{R}^{n\times n}$ ist definiert als

$$\mathrm{Spur}(A) := \sum_{i=1}^{n} a_{ii},$$

also als die Summe der Diagonaleinträge von A.

(a) Zeigen Sie, dass $\mathrm{Spur}(A) = \mathrm{Spur}(A^{\mathrm{T}})$ gilt.

(b) Zeigen Sie, dass die Spur als Abbildung $\mathrm{Spur}\colon \mathbb{R}^{n\times n} \to \mathbb{R}$ eine lineare Abbildung ist.

(c) Sei $A = (a_{ij})_{i,j\in\underline{n}} \in \mathbb{R}^{n\times n}$. Geben Sie eine explizite Formel für die Einträge von $A^{\mathrm{T}}A$ an und berechnen Sie $\mathrm{Spur}(A^{\mathrm{T}}A)$.

(d) Zeigen Sie, dass der reelle Vektorraum $\mathbb{R}^{n\times n}$ zusammen mit der Abbildung $\langle\cdot,\cdot\rangle\colon \mathbb{R}^{n\times n}\times \mathbb{R}^{n\times n} \to \mathbb{R}$, definiert durch

$$\langle A, B\rangle := \mathrm{Spur}(A^{\mathrm{T}}B)$$

ein Skalarproduktraum ist.

Aufgabe 14.80. (Orthogonalität):

Beweisen Sie alle Aussagen von Lemma 14.3.2 für den reellen Fall, d.h. zeigen Sie, dass für beliebige Vektoren v, w eines reellen Skalarproduktraums V und alle Teilmengen $M \subseteq V$ gilt:

(i) Ist $v \neq 0_V$, so ist $\frac{1}{\|v\|} \cdot v$ normiert.

(ii) $0_V \perp v$,

(iii) $v \perp w \iff w \perp v$,

(iv) $v \perp v \iff v = 0_V$.

(v) Ist $u \perp M$, so gilt auch $u \perp \mathcal{L}(M)$.

(vi) $M^\perp$ ist ein Teilraum von V.

Aufgabe 14.81. (Orthonormalsysteme):

(a) Wir betrachten den $\mathbb{R}^n$ mit Standardskalarprodukt. Sei $Q = \big(q_1, \ldots, q_m\big)$ ein Tupel von Vektoren aus $\mathbb{R}^n$, wobei $m \in \mathbb{N}$. Mit $Q \in \mathbb{R}^{n\times m}$ bezeichnen wir auch die Matrix, in deren i-ter Spalte der Vektor q_i steht. Beweisen Sie die folgenden Aussagen:

   (i) Das Tupel Q ist genau dann ein Orthogonalsystem, wenn die Matrix Q^TQ eine Diagonalmatrix mit vollem Rang ist.

   (ii) Das Tupel Q ist genau dann ein Orthonormalsystem, wenn $Q^TQ = I_m$ ist.

   (iii) Das Tupel Q ist genau dann eine Orthogonalbasis von $\mathbb{R}^n$, wenn $m = n$ und die Matrix Q^TQ eine Diagonalmatrix ist und alle Diagonaleinträge ungleich Null sind.

   (iv) Das Tupel Q ist genau dann eine Orthonormalbasis von $\mathbb{R}^n$, wenn $m = n$ und $Q^TQ = I_n$ ist.

   Hinweis: Eine Diagonalmatrix $A = (a_{ij})_{i,j\in\underline{n}} \in \mathbb{R}^{n\times n}$ ist eine Matrix, die höchstens auf der Hauptdiagonalen Einträge ungleich Null besitzt, d. h. es ist $a_{ij} = 0$ für alle $i \neq j$.

(b) Wir betrachten den $\mathbb{R}^2$ bzw. den $\mathbb{C}^3$ mit dem jeweiligen Standardskalarprodukt. Gegeben seien die beiden Tupel von Vektoren

$$B = (\begin{pmatrix} \cos\alpha \\ \sin\alpha \end{pmatrix}, \begin{pmatrix} -\sin\alpha \\ \cos\alpha \end{pmatrix}) \qquad \text{und} \qquad C = (\begin{pmatrix} i \\ 0 \\ i \end{pmatrix}, \begin{pmatrix} 1 \\ 2i \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -i \\ -1 \end{pmatrix}),$$

mit $\alpha \in [0, 2\pi)$. Untersuchen Sie:

- Ist B eine Orthogonalbasis des reellen Vektorraums $\mathbb{R}^2$? Ist B eine Orthonormalbasis?
- Ist C eine Orthogonalbasis des komplexen Vektorraums $\mathbb{C}^3$? Ist C eine Orthonormalbasis?

Hinweis: Die Aussagen aus (a) gelten analog mit Q^H anstelle von Q^T, wenn $V = \mathbb{C}^n$ ist.

(c) Wir betrachten den reellen Vektorraum der stetigen Funktion $C([-1,1],\mathbb{R})$ mit dem Skalarprodukt

$$\langle f, g\rangle := \int_{-1}^{1} f(x)g(x)\mathrm{dx}.$$

Untersuchen Sie, ob die Menge der Monomfunktionen

$$\{f \mid f : [-1,1] \to \mathbb{R}, \text{ definiert durch } x \mapsto x^k,\ k \in \mathbb{N}\} \quad \subseteq \quad C([-1,1],\mathbb{R})$$

ein Orthogonalsystem bezüglich des obigen Skalarproduktes ist.

Aufgabe 14.82. (Hauptsatz über orthogonale Projektoren 14.4.1 (iii)):

Beweisen Sie Teil (iii) des Satzes 14.4.1: Sei V ein Skalarproduktraum, U ein endlichdimensionaler Teilraum von V und $Q = (q_1, \ldots, q_m)$ eine Orthonormalbasis von U. P bezeichne den orthogonalen Projektor auf U (vgl. Teil (i) und (ii) des Satzes). Zeigen Sie, dass dann die Abbildung

$$P^\perp\colon V \to V, \qquad P^\perp := \mathrm{id}_V - P$$

der orthogonale Projektor auf das orthogonale Komplement $U^\perp$ von U ist – das heißt, dass $P^\perp$ die Eigenschaften

$$\forall w \in U^\perp : P^\perp w = w, \qquad \mathrm{Bild}(P^\perp) = U^\perp \qquad \text{und} \qquad \mathrm{Kern}(P^\perp) = U$$

erfüllt.

Aufgabe 14.83. (Gram-Schmidt-Verfahren und orthogonale Projektion):

Wir betrachten den Vektorraum $\mathbb{R}^4$ mit dem Standardskalarprodukt $\langle \cdot, \cdot \rangle$ und den Untervektorraum $U := \mathcal{L}(B)$, wobei

$$B := \left(\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} \right).$$

(a) Orthonormalisieren Sie B mit Hilfe des Gram-Schmidt-Verfahrens zu einem Orthonormalsystem Q.

(b) Ergänzen Sie Q aus (a) zu einer Orthonormalbasis des $\mathbb{R}^4$.

(c) Berechnen Sie die Bestapproximation des Standardbasisvektors $e_1 \in \mathbb{R}^4$ in U.

Aufgabe 14.84. (Senkrechträume und orthogonale Zerlegungen):

(a) Sei $V = \mathbb{R}^n$, ausgestattet mit dem Standardskalarprodukt $\langle \cdot, \cdot \rangle$. Sei U ein Untervektorraum von V und $B = (b_1, \ldots, b_k)$ eine Basis von U.

Beweisen Sie: Es gilt

$$x \in U^{\perp} \iff B^{\mathrm{T}} x = 0,$$

das heißt der Senkrechtraum von U ist gerade die Lösungsmenge des linearen Gleichungssystems

$$\begin{aligned} \langle b_1, x \rangle &= 0, \\ \langle b_2, x \rangle &= 0, \\ &\vdots \\ \langle b_k, x \rangle &= 0. \end{aligned}$$

(b) Seien nun der Vektorraum $V = \mathbb{R}^3$ und der Untervektorraum $U := \mathcal{L}(v, w)$ gegeben, wobei

$$v := \begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix} \quad \text{und} \quad w := \begin{pmatrix} -2 \\ 0 \\ 0 \end{pmatrix}.$$

Berechnen Sie mit Hilfe von (a) und dem Gram-Schmidt-Verfahren eine Orthonormalbasis von U und eine von $U^{\perp}$.

(c) Seien V und U wie in (b) und P der orthogonale Projektor auf U. Geben Sie für $x = (1, 0, 1)^T$ die orthogonale Zerlegung $x = Px + P^{\perp}x$ bezüglich U an. Prüfen Sie für sie den verallgemeinerten Satz des Pythagoras.

Aufgabe 14.85. (Binomische Formel und Polarisationsformel):

Sei V ein Skalarproduktraum. Beweisen Sie

(a) die binomischen Formeln für $\mathbb{R}$-Vektorräume: Ist V ein $\mathbb{R}$-Vektorraum, so gilt für alle v, w in V

$$\|v \pm w\|^2 \;=\; \|v\|^2 \pm 2\langle v, w\rangle + \|w\|^2, \qquad \langle v+w, v-w\rangle = \|v\|^2 - \|w\|^2,$$

(b) die Polarisationsformel für $\mathbb{C}$-Vektorräume: Ist V ein $\mathbb{C}$-Vektorraum, so gilt für alle v, w in V

$$4\langle v, w\rangle = \|v+w\|^2 - \|v-w\|^2 + i\|iv+w\|^2 - i\|iv-w\|^2.$$

Aufgabe 14.86. (Norm induziert Metrik, verallgemeinerter Cosinussatz):

(a) Sei $(V, \|\cdot\|)$ ein normierter Vektorraum und $(A, V, +)$ ein affiner Raum über V. Zeigen Sie: Dann ist die Abbildung

$$d_{\|\cdot\|} \colon A \times A \to \mathbb{R}, \qquad d_{\|\cdot\|}(P, Q) := \|\overrightarrow{PQ}\|$$

eine Metrik auf A.

(b) Sei V ein euklidischer Vektorraum und $\|\cdot\|$ die durch das Skalarprodukt auf V induzierte Norm. Zeigen Sie: Dann gilt für alle $a, b \in V$ der verallgemeinerte Cosinussatz, d.h. mit der Definition für den Winkels $\angle(a, b)$ zwischen zwei Vektoren a, b und der Abkürzung $c := b - a \in V$ ist

$$\|c\|^2 \;=\; \|a\|^2 + \|b\|^2 - 2 \cdot \|a\| \cdot \|b\| \cdot \cos\angle(a, b).$$

Aufgabe 14.87. (Gram'sche Matrizen):

(a) Auf dem $\mathbb{R}^3$ sei für $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, $y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \in \mathbb{R}^3$ eine Abbildung $\langle \cdot, \cdot \rangle : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ durch

$$\langle x, y \rangle := x_1y_1 + 3x_2y_2 + 4x_3y_3 + x_1y_2 + x_2y_1 + x_1y_3 + x_3y_1 + x_2y_3 + x_3y_2.$$

definiert. Bestimmen Sie eine Matrix A, mit der $\langle x, y \rangle = x^T Ay$ gilt.

(b) Zeigen Sie, dass $\langle \cdot, \cdot \rangle$ aus (a) ein Skalarprodukt auf $\mathbb{R}^3$ ist.

(c) Wir betrachten den reellen Vektorraum

$$P_2[x] := \left\{ f\colon [-1, 1] \to \mathbb{R} \mid f(x) = ax^2 + bx + c, \quad a, b, c \in \mathbb{R} \right\}$$

der quadratischen Funktionen mit Standardbasis $B := (x \mapsto 1, x \mapsto x, x \mapsto x^2)$. Berechnen Sie die Gram'sche Matrix bezüglich B für das Skalarprodukt

$$\langle p, q \rangle = \int_{-1}^{1} p(x)q(x)dx.$$

15 Isometrien und Abbildungsgeometrie

15.1 Orthogonale und unitäre Abbildungen

Definition/Lemma 15.1.1. *(Isometrien: Orthogonale und unitäre Abbildungen)*

Seien $(V, \langle\cdot,\cdot\rangle)$ *und* $(W, [\cdot,\cdot])$ *zwei Skalarprodukträume über dem gleichen Grundkörper* $\mathbb{K} = \mathbb{R}$ *oder* $\mathbb{K} = \mathbb{C}$, *und* $\|\cdot\|_V$ *und* , $\|\cdot\|_W$ *die jeweils auf* V *und* W *induzierten Normen.*

Dann heißt eine lineare Abbildung $f : V \to W$ *lineare Isometrie, wenn eine der folgenden äquivalenten Eigenschaften erfüllt ist:*

(i) f *ist abstandserhaltend (isometrisch), d.h. für alle* $v, w \in V$ *gilt*

$$\|f(v) - f(w)\|_W \quad = \quad \|v - w\|_V,$$

(ii) f *ist längenerhaltend, d.h. für alle* $v \in V$ *gilt*

$$\|f(v)\|_W \quad = \quad \|v\|_V,$$

(iii) f *ist skalarprodukterhaltend, d.h. für alle* $v, w \in V$ *gilt*

$$[f(v), f(w)] \quad = \quad \langle v, w\rangle,$$

Ist $\mathbb{K} = \mathbb{R}$, so werden lineare Isometrien auch f oft als orthogonale Abbildung bezeichnet; für $\mathbb{K} = \mathbb{C}$ sagt man oft, f ist eine unitäre Abbildung.

Die obige Definition behauptet „durch die Blume“, dass eine der obigen Eigenschaften (für lineares f) reicht, damit alle anderen erfüllt sind. Das wollen wir natürlich beweisen.

Beweis: (ii) ist der Spezialfall $v = w$ von (iii), also gilt (iii) $\Rightarrow$ (ii). Umgekehrt haben wir mit der Polarisationsformel 14.5.4, der Linearität und der Isometrieeigenschaft

$$\begin{aligned} 4[f(v), f(v)] &= \|f(v) + f(w)\|_W^2 - \|f(v) - f(w)\|_W^2 \\ &= \|f(v + w)\|_W^2 - \|f(v - w)\|_W^2 = \|v + w\|_V^2 + \|v - w\|_V^2 = 4\langle v, w\rangle. \end{aligned}$$

und eine analoge Rechnung mit der komplexen Polarisationsformel im Fall $\mathbb{K} = \mathbb{C}$, also (ii) $\Rightarrow$ (iii). (i) $\Leftrightarrow$ (ii) folgt mit der Linearität von f, dem Nullvektor und der Differenz von Vektoren und ist eine Übungsaufgabe.

$\square$

Lemma 15.1.2. *(Weitere Kriterien für Isometrien)*

Ist mit den Notationen aus dem letzten Lemma V endlichdimensional, so ist eine lineare Abbildung $f : V \to W$ zudem eine lineare Isometrie genau dann, wenn eine der folgenden Eigenschaften erfüllt ist:

(iv) Es gibt eine Orthonormalbasis Q von V (bzgl. $\langle\cdot,\cdot\rangle$), so dass $f(Q)$ (bzgl. $[\cdot,\cdot]$ in W) ein Orthonormalsystem ist.

(v) Jedes Orthonormalsystem von V wird durch f auf ein Orthonormalsystem in W abgebildet.

Der Punkt für die Einschränkung auf endlichdimensionale Vektrräume ist, dass wir für unendlichdimensionale Vektorräume nicht die Existenz einer Orthonormalbasis voraussetzen können, und dass, falls es keine gibt, (v) in diesem Fall immer erfüllt ist.

Beweis: Wir zeigen mit den Nummerierungen aus der vorangegangenen Definition, dass (iii) $\Rightarrow$ (v) und (iv) $\Rightarrow$ (ii) gilt. Da (v) $\Rightarrow$ (iv) trivialerweise gilt, zeigt das die Behauptung.

Zu (iii) $\Rightarrow$ (v) sei $Q = (q_1, \ldots, q_k)$ ein gegebenes Orthonormalsystem. Dann gilt

$$[f(q_i), f(q_j)] = \langle q_i, q_j\rangle = \delta_{i,j},$$

$f(Q)$ ist also orthogonal wie behauptet. Sei schließlich $Q = (q_1, \ldots, q_n)$ eine gegebenes Orthonormalbasis so, dass $f(Q)$ ein Orthonormalsystem in W ist. Dann gibt es Koordinaten so, dass $v = \sum_{i=1}^n a_i q_i$ ist, und, da f linear ist, gilt $f(v) = \sum_{i=1}^n a_i f(q_i)$. Da Q und $f(Q)$ beides ONBs sind ($f(Q)$ ist eine von $\mathrm{Bild}(f)$, folgt nach Lemma 14.3.7

$$\|v\| = \left(\sum_{i=1}^n |a_i|^2\right)^{\frac{1}{2}} \quad \text{und} \quad \|f(v)\| = \left(\sum_{i=1}^n |a_i|^2\right)^{\frac{1}{2}},$$

also (ii).

$\square$

Lemma 15.1.3. *(Einige Eigenschaften linearer Isometrien)*

Für lineare Isometrien, also orthogonale Abbildungen (Fall $\mathbb{K} = \mathbb{R}$) oder unitäre Abbildungen (Fall $\mathbb{K} = \mathbb{C}$) haben wir die folgenden Eigenschaften (mit euklidischen bzw. unitären Vektorräumen V, W, Z):

(i) Sind $f : V \to W$ und $g : W \to Z$ lineare Isometrien, so ist auch $g \circ f$ eine lineare Isometrie.

(ii) Lineare Isometrien sind immer injektiv.

Ist $V = W$ und V endlichdimensional, so gilt zusätzlich:

(iii) Lineare Isometrien sind immer bijektiv und ihre Inverse ebenfalls eine lineare Isometrie,

(iv) Die orthononalen Endomorphismen eines euklidischen Vektorraumes

$$O(V) := \{f : V \to V \mid f \text{ ist orthogonal}\}$$

bzw. der unitären Endomorphismen eines unitären Vektorraumes

$$U(V) := \{f : V \to V \mid f \text{ ist unitär}\}$$

sind Gruppen mit der Hintereinanderausführung von Abbildungen als Verknüpfung.

(v) Ist Q eine Orthonomalbasis von V und $\mathbb{K}^n$ mit dem Standardskalarprodukt ausgestattet, so ist die Koordinatenabbildung $k_Q : V \to \mathbb{K}^n$ eine orthogonale bzw. unitäre Abbildung (und damit auch abstands- und längenerhaltend nach Lemma 15.1.1).

Beweis: Der Beweis von (i) bis (iii) ist Übungsaufgabe. Für die (iv) ist zu zeigen, dass (a) die Hintereinanderausführung zweier orthogonaler Abbildungen wieder orthogonal ist (das ist (i)), dass das Assoziativgesetzt für $\circ$ gilt (das folgt, da es für alle Abbildunge gilt), dass jedes $f \in O(V)$ bzw. $f \in U(V)$ eine Inverse in dieser Menge hat (das steht in (iii)), und dass es ein Neutrales gibt(das ist id_V, diese Abbildung ist trivialerweise orthogonal). Wir klären noch kurz (v): k_Q bildet die ONB Q auf die Standardbasis im $\mathbb{K}^n$ ab, welche eine Orthonormalbasis bezüglich des Standardskalarprodukt ist, daher ist k_B orthogonal nach Lemma 15.1.1.

□

Definition 15.1.4. *(Orthogonale und unitäre Matrizen)*

Eine quadratische Matrix $Q \in \mathbb{K}^{n\times n}$ *heißt* <u>*orthogonal*</u> *bzw.* <u>*unitär*</u>, *wenn ihre Spalten ein Ortho*<u>*normal*</u>*basis*[1] *des* $\mathbb{K}^n$ *bilden, also wenn (nach Lemma 14.3.5)*

$$Q^T Q = I_n \tag{15.1}$$

ist (im Fall $\mathbb{K} = \mathbb{R}$*) bzw. wenn*

$$Q^H Q = I_n \tag{15.2}$$

ist (im Fall $\mathbb{K} = \mathbb{C}$*).*

Bemerkungen dazu:

- Während das Invertieren von Matrizen ansonsten, wir erinnern uns, recht aufwendig ist, ist es bei orthogonalen Matrizen anders: Gleichung (15.1) bedeutet, dass $Q^T = Q^{-1}$ ist, dass man also durch bloßes Vertauschen der Zeilen- und Spaltenindizes die Inverse erhält. Das lässt sich – auch am Rechner – ziemlich schnell bewerkstelligen. Man vergleiche das z.B. für eine (20×20)-Matrix mit der Inversenberechnung per Gauß-Algorithmus wie in Abschnitt 6.5.[2]

- Das kommt einem auch bei Koordinatentransformationen zugute, falls die Basis Q, auf die man transformiert, eine Orthogonalbasis ist. In die Umrechnungverschrift geht ja (vgl. Lemma 12.5.2 für Koordinatentransformationen in Vektorräumen bzw. Lemma 12.5.2 für affine Transformationen) die Inverse der Q ein, und die muss man nicht extra berechnen, sondern erhält sie durch Transponieren.

Sie haben es sich vielleicht schon gedacht: Orthogonale bzw. unitäre Matrizen sind genau diejenigen Matrizen, die zu orthogonalen bzw. unitären linearen Abbildungen gehören.

[1] Ja, die Bezeichnung „orthogonale Matrix“ macht hier keinen Sinn. man sollte sie lieber „orthonormale Matrizen“ nennen. Es hat sich aber so eingebürgert.

[2] Für ein echtes praktisches Problem, z.B. aus den Ingenieurswissenschaften, ist das immer noch eine ziemlich niedliche Abmessung. „Echte“ Probleme haben eher die Abmessung $20.000 \times 20,000$.

Lemma 15.1.5. *(Orthogonale/unitäre Abbildungen und Matrizen)*

(i) *Sei $V = \mathbb{K}^n$ mit dem Standardskalarprodukt und $f : V \to V$ linear, also eine Matrixabbildung. Dann ist $f : x \mapsto A \cdot x$ eine lineare Isometrie genau dann, wenn die Matrix A orthogonal bzw. unitär ist.*

(ii) *Seien V, W endlichdimensional mit $\dim V = \dim W$ und Q, Q' ONBs von V, W. Dann ist eine lineare Abbildung $f : V \to W$ eine lineare Isometrie genau dann, wenn die darstellende Matrix $A_{Q,Q'}$ orthogonal bzw. unitär ist.*

Beweis: (i) In den Spalten der Matrix A stehen die Bilder der Standardbasis. Diese bilden ein Orthogonalsystem genau dann, wenn f isometrisch ist, das besagen die Implikationen (iv) $\Rightarrow$ (iii) und (iii) $\Rightarrow$ (v) aus Lemma 15.1.1. (ii): Die darstellende Matrix $A_{Q,Q'}$ von f ist die Abbildungsmatrix der Abbildung

$$f_{Q,Q'} : \quad \mathbb{K}^n \to \mathbb{K}^n, \qquad x \mapsto k_{Q'} \circ f \circ k_Q^{-1} \qquad (\star);$$

diese ist nach (i) genau dann eine lineare Isometrie, wenn die Matrix orthogonal/unitär ist. Also gilt, da die Koordinatenabbildung und ihre Inverse Isometrien sind, Hintereinanderausführungen isometrischer Abbildungen wieder isometrisch sind (beides Lemma 15.1.3) und wir neben $(\star)$ auch $f = k_{Q'}^{-1} \circ f_{Q,Q'} \circ k_Q$ haben,

$$A_{Q,Q'} \text{ orthogonal/unitär} \quad \Leftrightarrow \quad f_{Q,Q'} \text{ isometrisch} \quad \Leftrightarrow \quad f \text{ isometrisch.}$$

□

Bemerkungen dazu:

- Von diesem Standpunkt aus betrachtet beschreiben Isometrien und orthogonale Abbildungen also Wechsel zwischen Koordinatensystemen, bei denen die Maßstäbe sich nicht ändern und Achsen, die rechtwinklig aufeinander standen, zu Achsen werden, die rechtwinklig aufeinander stehen.

- Orthogonale Abbildungen und die damit verbundenen Matrizen sind (unter anderem aus diesen Gründen) unverzichtbare Werkzeuge in der Numerischen Linearen Algebra, in der vor allem große lineare Gleichungssysteme gelöst werden. Da sie Längen erhalten, bleibt auch der Fehler bei numerischen Rechnungen mit meßfehlerbehafteten Eingabedaten erhalten (und wird also nicht größer).

- Diese Bemerkungen waren alle für den reellen Fall formuliert, im Fall $\mathbb{K} = \mathbb{C}$ gilt es für unitäre Matrizen analog.

Zwei typische Beispiel für eine isometrische lineare Abbildung mit geometrischer Inspiration und Deutung und zugehörigen orthogonalen Matrizen sind die folgenden Spiegelungen und Drehungen. Wir beginnen mit Drehungen.

Verallgemeinerte Drehungen

- Sei $\alpha \in [0, 2\pi[$ und $c = \cos(\alpha), s := \sin(\alpha)$. Dann sind Drehmatrizen vom Typ

$$\begin{pmatrix} c & -s \\ s & c \end{pmatrix} \qquad \text{oder} \qquad \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & c & 0 & -s & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & s & 0 & c & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

  orthogonal. Die linke ist eine Drehung in $\mathbb{R}^2$ um den Winkel α, die rechte eine verallgemeinerte Drehung oder *Givens-Rotation* in der 2-4-Ebene im $\mathbb{R}^6$ (um den Winkel α). Allgemeiner lassen sich auch Givens-Rotationen in beliebigen Ursprungsebenen im $\mathbb{R}^n$ definieren; wir gehen hier aber nicht weiter ins Detail.

Definition/Lemma 15.1.6. *(Spiegelung an Unterräumen)*

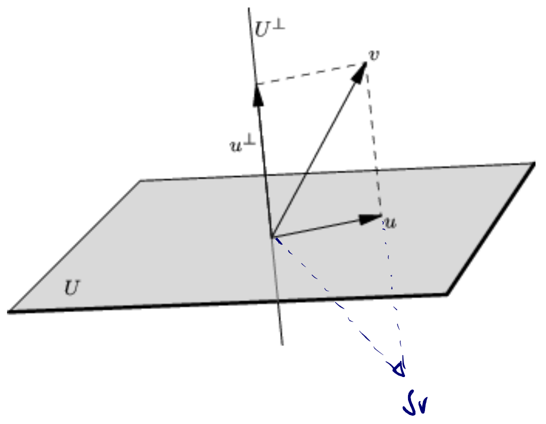
Sei V ein Skalarproduktraum, U ein endlichdimensionaler Teilraum von V und $P^\perp$ der orthogonale Projektor auf $U^\perp$. Dann heißt die lineare Abbildung

$$S := \mathrm{id}_V - 2P^\perp$$

Spiegelung am Unterraum U.

S ist eine isometrische lineare Abbildung mit der Eigenschaft, dass $Su = u$ für alle $u \in U$ und $Su = -u$ für alle $u \in U^\perp$ gilt.

Grenzbereiche zwischen Geometrie und Linearer Algebra…



$Sv = u - 2u^\perp$
ist Spiegelung an Ebene U.

Bild: C. Karpfinger –
Höhere Mathematik in Rezepten

Bemerkungen dazu:

- Ist $V = \mathbb{K}^n$, und Q eine ONB von $U^\perp$, dann lässt sich die Spiegelung an U als Matrixabbildung
$$S = I_n - 2QQ^T$$
schreiben. Das ergibt sich aus der dritten Bemerkung auf Seite 323.

- Ist insbesondere U eine Hyperebene von V, d.h. $\dim U = n-1$, und q ein normierter Normalenvektor von U, so lässt sich die Spiegelung an U als $H = I_n - 2qq^T$ schreiben. Das ist eine so genannte Householder-Spiegelung, die z.B. bei der numerischen Lösung linearer Gleichungssystemen eine große Rolle spielt.

Beweis: Die Isometrieeigenschaft folgt direkt aus dem verallgemeinerten Satz des Pythagoras für P. Der Rest ist Übungsaufgabe.

□

Aufgabe:

Gehen Sie für $V = \mathbb{R}^3$ alle Möglichkeiten $d = 0, 1, 2, 3$ für die Dimension des Unterraumes U durch. Was für eine „Spiegelung“ ergibt sich jeweils?

Lemma 15.1.7. *(Klassifizierung aller orthogonalen 2×2-Matrizen)*
Sei $U \in \mathbb{R}^{2\times 2}$ eine orthogonale Matrix. Dann gibt es genau einen Winkel $\alpha \in [0, 2\pi[$ so, dass

$$U = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix} \qquad \textit{oder} \qquad U = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ \sin(\alpha) & -\cos(\alpha) \end{pmatrix}$$

gilt.

Das rechte ist eine Spiegelung an der Geraden mit Winkel $\frac{\alpha}{2}$ zur x-Achse, also ist jede orthogonale Abbildung in $\mathbb{R}^2$ entweder eine Drehung oder eine Spiegelung.

Das Resultat lässt sich auch auf größere Dimensionen verallgemeinern, vgl. z.B. [BeuLA], S. 322. Dort lässt sich[3] ebenfalls zeigen, dass es immer zwei Typen von orthogonalen Abbildungen gibt: Die Drehungen und die sogenannten Drehspiegelungen, die sich als Kombination aus Drehung und Spiegelung darstellen lassen. In der Ebene sind die Drehspiegelungen einfach Spiegelungen, in höheren Dimensionen ist es etwas komplizierter, vgl. die angegebene Literatur.
Beweis: Wir skizzieren das Vorgehen, Details s. Vorlesung oder z.B. [Weitz], S. 368f. Sei

$$U = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in \mathbb{R}^{2\times 2}$$

orthogonal, also $U^T \cdot U = I_2$ und $U \cdot U^T = I_2$. Das bedeutet:

$$\begin{aligned} U^T \cdot U &= \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix} \cdot \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \\ &= \begin{pmatrix} a_{11}^2 + a_{21}^2 & a_{11}a_{12} + a_{21}a_{22} \\ a_{12}a_{11} + a_{22}a_{21} & a_{12}^2 + a_{22}^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ U \cdot U^T &= \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \cdot \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix} \\ &= \begin{pmatrix} a_{11}^2 + a_{12}^2 & a_{11}a_{21} + a_{12}a_{22} \\ a_{21}a_{11} + a_{22}a_{12} & a_{21}^2 + a_{22}^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

Aus der Gleichungen $a_{11}^2 + a_{21}^2 = 1$ liest man ab, dass der Vektor (a_{11}, a_{21}) auf dem Einheitskreis liegt, also gibt es genau einen Winkel $\alpha \in [0, 2\pi[$ mit $a_{11} = \cos(\alpha)$ und $a_{21} = \sin(\alpha)$. Aus den restlichen Gleichungen ergeben sich sich dann die restlichen behaupteten Einträge.
□

[3] mit verallgemeinerten Definitionen der Begriffe, vgl. die Definition der Spiegelung oben und mit verallgemeinerten Drehungen, die Hintereinanderausführungen von Drehungen in *beliebigen* Ebenen sind

15.2 Abbildungsgeometrie

Wir kehren mit den Erkenntnissen zur Winkel- und Längenmessung nochmal zurück in die affinen Räume. Ist $(A, V, +)$ ein affiner Raum, so kann man ein Skalarprodukt in V nutzen, um auf dem affinen Punktraum A eine Winkel-, Längen- und Abstandsmessung zu definieren und damit den affinen Raum A in den Spezialfälen $\mathbb{R}^2$ bzw. $\mathbb{R}^3$ zu dem zu machen, was im Hilbert'schen Sinne unter einer euklidischen Geometrie verstanden wird.

Dies ist ein Abschnitt, in dem noch einmal die Lineare Algebra/Analytische Geometrie und die klassische Geometrie nach Euklid und Hilbert aneinanderstoßen. Wir wollen hier keine vollwertige Geometrievorlesung versuchen; dieser Abschnitt soll allerdings vereutlichen, dass die euklidische Geometrie und damit auch die Geometrie der Schule mit Hilfe der Linearen Algebra des $\mathbb{R}^n$ effizient und vollwertig abgebildet werden kann. Wir schauen uns exemplarisch Bereiche aus der so genannten Abbildungsgeometrie an, in der es um die Begriffe der Kongruenz, der Symmetrien und der Ähnlichkeit und ihre Darstellung und Untersuchung mit den Mitteln der Linearen Algebra geht.

15.2.1 Affine euklidische und unitäre Räume

Definition 15.2.1. *(Affiner euklidischer/unitärer Raum)*

Sei $(V, \langle\cdot,\cdot\rangle)$ ein Skalarproduktraum und A ein affiner Raum über V. Dann heißt A

- *im Fall $\mathbb{K} = \mathbb{R}$ <u>euklidischer affiner Raum</u>,*
- *im Fall $\mathbb{K} = \mathbb{C}$ <u>unitärer affiner Raum</u>.*

Ist $A = V = \mathbb{R}^2$, so sprechen wir insbesondere von „der euklidischen Ebene $\mathbb{R}^2$“, für $A = V = \mathbb{R}^3$ von „dem euklidischen Raum $\mathbb{R}^3$“.

Die letzten beiden Formulierungen sind darin motiviert, dass die euklidische Ebene $\mathbb{R}^2$ und der euklidische Raum $\mathbb{R}^3$ sämtliche Axiome der euklidischen Geometrie nach Hilbert erfüllen, also „vollwertige“ euklidische Geometrien sind:

Bezeichnet $\|\cdot\|$ die durch das Skalarprodukt auf V induzierte Norm auf V, so lässt sich auf einem euklidischen affinen Raum A über V mit Hilfe von

$$d: \quad A \times A \to \mathbb{R}, \quad d(P, Q) := \|\overrightarrow{PQ}\|$$

eine Metrik definieren (Satz 14.5.3 und der bijektiven Beziehung $P \leftrightarrow ort_Q(P)$), um mit Hilfe der Norm von V Abstände von Punkten $P, Q \in A$ und Längen von durch ihre Endpunkte $P, Q \in A$ gegebenen Strecken (genau durch die reelle Zahl ($\|\overrightarrow{PQ}\|$) zu messen. Betrachten wir den für uns (und alle anderen auch) wichtigsten Fall, dass $V = A = \mathbb{R}^n$ ist, so sind die Punkte von A Vektoren $x, y \in \mathbb{R}^n$, und wir haben

$$d(x, y) = \|y - x\|$$

mit der „normalen“ euklidischen Norm des $\mathbb{R}^n$, die wir jetzt als Abstand zwischen den *Punkten* x, y lesen können.

Zudem haben wir über

$$\cos(\angle(v, w)) = \frac{\langle v, w\rangle}{\|v\|\|w\|}$$

ein Winkelmaß für Vektoren $v, w \neq 0_V$ aus V zur Verfügung, das uns in die Lage versetzt, auch Winkel zwischen Verbindungsvektoren $\overrightarrow{PQ}$, $\overrightarrow{RS}$ und damit zum Beispiel Winkel zwischen zwei sich schneidenden Geraden zu berechnen.[4]

Die für Vektorräume bewiesenen Sätze wie der Satz des Pythagoras in seiner allgemeineren Form und der Kosinussatz übertragen sich direkt auf den affinen Raum, alle übrigen Sätze der euklidischen Geometrie (z.B. „Klassiker“ wie der Stufenwinkelsatz, Strahlensatz, Kongruenzsätze, Satz des Thales,...) lassen sich nun im $\mathbb{R}^3$ mit den so definierten Längen und Winkelmaßen beweisen; da man nun die Macht der Linearen Algebra auf seiner Seite hat, ist es oft so, dass elementargeometrische Beweise sich oft stark vereinfachen.

Das sei hier nur erwähnt und wird in den Übungen etwas vertieft. Wir wollen uns statt einer tieferen Beschäftigung mit der Elementargeometrie noch einmal mit affinen Selbstabbildungen euklidischer Räume und deren geometrischer Deutung beschäftigen und so einen Einstieg in die so genannte Abbildungsgeometrie machen. Hierbei werden wir uns meistens auf einfachere, spezielle Fälle beschränken: Wie oben gilt das ganze restliche Kapitel $\mathbb{K} = \mathbb{R}$, wir werden oft einen Vektorraum als affinen Raum über sich selbst untersuchen, und oft wird dieser Vektorraum konkret $\mathbb{R}^n$ oder noch konkreter der euklidische Raum ($n = 3$) oder die euklidische Ebene ($n = 2$) sein.

[4] Den Winkel $\angle(PQR)$, der gegen den Uhrzeigersinn gesehen zwischen den Vektoren $\overrightarrow{QP}$ und $\overrightarrow{QR}$ liegt, können wir damit noch nicht definieren, da wir noch nicht wissen, wie wir den beteiligten Drehsinn mathematisch beschreiben. Dazu nutzt man i.A. die Determinante, s. Abschnitt 16.

15.2.2 Herangehensweise der Abbildungsgeometrie

Die so genannte Abbildungsgeometrie im $\mathbb{R}^n$ nutzt Abbildungen in Euklidischem Raum, Euklidischer Ebene und allgemeiner in $\mathbb{R}^n$ (z.B die jetzt zu definierenden Kongruenzabbildungen), um mit ihrer Hilfe im $\mathbb{R}^n$ (verallgemeinert, aber im Einklang mit den Axiomen der Geometrie) zu definieren, was man im $\mathbb{R}^n$ unter kongruenten Figuren verstehen will, also um allgemeine geometrische Grundbegriffe abstrakt zu definieren.

Definition 15.2.2. *(Euklidische Isometrien, Kongruenzabbildungen und Bewegungen)*

Sei $g : A \to A$ eine Abbildung eines affinen euklidischen Raumes in sich selbst.

g heißt euklidische Isometrie, Bewegung oder (besonders im Fall der euklidischen Ebene) Kongruenzabbildung von A, falls für alle P, Q in A gilt

$$d(g(P), g(Q)) \quad = \quad d(P, Q),$$

d.h. falls g Abstände zwischen Punkten erhält.

Zwei wichtige Begriffe, die sich jetzt mit Hilfe des Begriffs der Kongruenzabbildung in affinen Räumen definieren lassen, sind die folgenden:

Definition 15.2.3. *(Kongruenz, Symmetrie)*

Sei $(A, V, \langle\cdot,\cdot\rangle)$ ein euklidischer affiner Raum.

(a) Zwei beliebige Teilmengen $F_1, F_2 \subseteq A$ (oft „Figuren in A“ genannt) werden als kongruent bezeichnet, wenn es eine Kongruenzabbildung (= Bewegung = Isometrie) $g : A \to A$ gibt, die die eine auf die andere abbildet, für die also $g(F_1) = F_2$ gilt.

(b) Eine Figur $F \subseteq V$ heißt symmetrisch, wenn es eine Kongruenzabbildung $g : A \to A$ mit $g \neq \mathrm{id}_A$ und $g(F) = F$ gibt; man sagt, F ist oder bleibt fix unter g.

Bemerkungen dazu:

- Es ist wichtig zu verstehen, dass das obige in der Abbildungsgeometrie als Definition des Kongruenz- bzw. Symmetriebegriffs genommen wird. Alle anderen „bekannten“ Eigenschaften wie Längen- und Winkelerhaltung bei kongruenten Figuren folgen dann mit Hilfe der entsprechenden Eigenschaft der Kongruenzabbildung daraus.

- Natürlich kann man Symmetrien weiter unterteilen; auch das passiert mit Hilfe der oben eingeführten Abbildung g: Ist g beispielsweise eine Drehung, so heißt F drehsymmetrisch, ist g eine Achsenspiegelung, so heißt g achsensymmetrisch usw.

Die Frage, die sich unmittelbar stellt, wenn man wie oben den in der Geometrie grundlegenden Kongruenzbegriff mit Hilfe von Kongruenzabbildungen/Isometrien beschreibt, ist die Frage, wie man sich diese Abbildungen vorstellen kann bzw. ob sie irgendwie genauer klassifiziert oder beschrieben werden können. Das klärt der folgende Satz, der für reelle normierte Räume V als *Satz von Mazur-Ulam bekannt.* Wir setzen hier etwas weniger allgemein voraus, dass V ein euklidischer Vektorraum ist, den wir als affinen Raum über sich selbst auffassen; insbesondere schließt das unseren Hauptanwendungsfall $V = A = \mathbb{R}^n$ mit ein.

Satz 15.2.4. *(Satz von Mazur-Ulam: Euklidische Isometrien sind affine Abbildungen)*

Sei V ein euklidischer Vektorraum, den wir als affinen Raum $A = (V, V, \langle\cdot,\cdot\rangle)$ über sich selbst auffassen, und $g : A \to A$ eine Abbildung. Dann sind äquivalent:

(i) g ist eine Isometrie auf A, d.h. (mit der von V auf V induzierten Metrik) gilt

$$d(g(v), g(w)) = \|g(v) - g(w)\| = \|v - w\| \tag{15.3}$$

für alle $v, w \in A$.

(ii) Es gibt eine orthogonale lineare Abbildung $f : V \to V$ und einen Translationsvektor $t \in V$ so, dass für alle $v \in V$

$$g(v) = f(v) + t \quad (\star)$$

ist.

D.h. eine affine Isometrie $g : A \to A$ ist automatisch eine affine Abbildung mit orthogonalem Anteil (und auch umgekehrt, aber das ist nicht so spektakulär).

In euklidischen Vektorräumen, die man als affine Räume über sich selbst auffasst, insbesondere in Euklidischer Ebene und Euklidischem Raum, sind also

$$\boxed{\begin{gathered}\text{Kongruenzabbildungen} = \text{Bewegungen} = \text{Isometrien} \\ = \text{affine Abbildungen mit orthogonaler linearer Abbildung}\end{gathered}}$$

Beweis: Der Beweis wird in der Vorlesung ausgelassen, aber hier nachgeholt: Die Rückrichtung folgt aus

$$d(g(v), g(w)) \;=\; \|f(v) + t - (f(w) + t)\| \;=\; \|f(v - w)\| \;=\; \|v - w\| \;=\; d(v, w).$$

Für die andere, aufwändigere Richtung setzen wir $t := g(0_V)$ und betrachten die Abbildung $f = g - t$. Wie in der Rechnung oben erhält man, dass mit g auch f isometrisch ist. Wir wollen also nun zeigen, dass f linear ist; dazu leisten wir zunächst etwas Vorarbeit: Es gilt $f(0_V) = g(0_V) - g(0_V) = 0_V$ und deshalb

$$\|f(v)\| = \|f(v) - f(0_V)\| = \|v - 0_V\| = \|v\|$$

für alle $v \in V$, f ist also auch längenerhaltend in V. Wir nutzen wieder die Binomische Formel, um einen Zusammenhang zum Skalarprodukt herzustellen: Es ist mit der vorigen Überlegung und der Voraussetzung (15.3)

$$\begin{aligned} -2\langle f(v), f(w)\rangle &= \|f(v) - f(w)\|^2 - \|f(v)\|^2 - \|f(w)\|^2 \\ &= \|v - w\|^2 - \|v\|^2 - \|w\|^2 = -2\langle v, w\rangle \end{aligned}$$

also haben wir zudem $\langle f(v), f(w)\rangle = \langle v, w\rangle$ für alle $v, w \in V$.

Wir zeigen nun die Linearität von f, indem wir direkt die Linearitätsbedingungen beweisen, genauer zeigen wir

$$\|f(a \cdot v) - a \cdot f(v)\| = 0 \quad \text{und} \quad \|f(v + w) - f(v) - f(w)\| = 0$$

für alle $a \in \mathbb{R}$ und alle aus $v, w \in V$; aus der Positivität der Norm folgt dann jeweils die für die Linearität benötigte Gleichheit. Unter Benutzung der obigen Resultate und der

Eigenschaften von Skalarprodukt und Norm haben wir

$$\begin{aligned}
\|f(a\cdot v)-a\cdot f(v)\|^2 &= \|f(a\cdot v)\|^2-2\langle f(av),a\cdot f(v)\rangle+\|a\cdot f(v)\|^2\\
&= \|f(a\cdot v)\|^2-2a\cdot\langle f(av),f(v)\rangle+|a|^2\|f(v)\|^2\\
&= \|a\cdot v\|^2-2a\cdot\langle a\cdot v,v\rangle+|a|^2\|v\|^2\\
&= 2\|a\cdot v\|^2-2\langle a\cdot v,a\cdot v\rangle=0
\end{aligned}$$

und für die additive Bedingung analog und jetzt etwas abgekürzt

$$\begin{aligned}
\|f(v+w)-f(v)-f(w)\| &= \|v+w\|^2+\|v\|^2+\|w\|^2-2\langle v+w,v\rangle-2\langle v+w,w\rangle+2\langle v,w\rangle\\
&= \|v+w\|^2+\big(\|v\|^2+2\langle v,w\rangle+\|w\|^2\big)-2\langle v+w,v+w\rangle=0.
\end{aligned}$$

Damit ist f linear wie behauptet. $\square$

Beispiele für Kongruenzabbildungen in $\mathbb{R}^2,\mathbb{R}^3$ bzw. auch $\mathbb{R}^n$ sind

- Spiegelungen an affinen Ebenen durch den Ursprung ($A=I-2QQ^T$, $t=0$),
- Drehungspiegelungen (mir dem Ursprung des Koordinatensystems ($A=$ Drehmatrix, $t=0$),
- Translationen $A=I_n, t\neq 0$.

Die Bemerkung nach Lemma 15.1.7 am Schluss des letzten Abschnitts zeigt auch, dass das grundsätzlich alle sind, da es bis auf Verschiebungen damit keine anderen Isometrien als Drehungen und Drehspiegelungen gibt.
In der Elementargeometrie kann man mit den dortigen Methoden beweisen, dass sich jede Kongruenzabbildung in der Ebene als Hintereinanderausführung von drei Spiegelungen ist („Dreispiegelungssatz“, vgl. z.B. [BeGoMü]); jede Kongruenzabbildung im Raum lässt sich als Hintereinanderausführung von vier Spiegelungen an Ebenen) darstellen. Auch schön. So kann sich „im großen Garten der Geometrie (...) jeder nach seinem Geschmack einen Strauß pflücken.“ (D. Hilbert).

15.2.3 Ähnlichkeitsabbildungen

Analog zum Kongruenzbegriff wird in der Abbildungsgeometrie auch der Ähnlichkeitsbegriff mit Hilfe geeigneter Abbildungen eingeführt.

Definition 15.2.5. *(Ähnlichkeitsabbildungen)*

Sei V wieder ein euklidischer Vektorraum, den wir als affinen euklidischen Raum $A = (V, V, \langle \cdot, \cdot \rangle)$ über sich selbst auffassen.

(i) *Eine Abbildung $g : V \to V$ heißt Ähnlichkeitsabbildung oder Ähnlichkeitstransformation in V, falls es eine reelle Zahl $m > 0$ gibt, so dass für alle $v, w \in V$*

$$d(g(v), g(w)) \quad = \quad m \cdot d(v, w)$$

gilt. (Wieder haben wir wegen $V = A$ für die induzierte Metrik $d(v, w) = \|v - w\|$.) Die Zahl m heißt Maßstab der Ähnlichkeitsabbildung g.

(ii) *Zwei Figuren $F_1, F_2 \subseteq V$ heißen ähnlich, wenn sie durch eine Ähnlichkeitsabbildung aufeinander abgebildet werden, d.h. es gibt eine Ähnlichkeitsabbildung $g : V \to V$ mit $g(F_1) = F_2$.*

Bemerkungen dazu:

- Jede Kongruenzabbildung in einem euklidischen Vektorraum ist eine Ähnlichkeitsabbildung (mit $m = 1$).
- Für Ähnlichkeitsabbildungen g und v, w, x, y in V mit $v \neq w$, $x \neq y$ gilt

$$\frac{\|g(v) - g(w)\|}{\|v - w\|} \quad = \quad m \quad = \quad \frac{\|g(x) - g(y)\|}{\|x - y\|},$$

  d.h. g erhält Verhältnisse von Längen.
- Wir erinnern uns, dass eine zentrische Streckung in einem affnien Raum A eine affine Abbildung der Form

$$P \mapsto Z + s \cdot \overrightarrow{ZP}$$

  ist. Im Fall $A = \mathbb{R}^n$ wird eine zentrische Streckung also durch eine affine Transformation mit Transformationsmatrix

$$x \quad \mapsto \quad z + \begin{pmatrix} s & 0 & \dots & 0 \\ 0 & s & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & s \end{pmatrix} (x - z) \;=\; z + s \cdot I_n \cdot (x - z) \;=\; s \cdot x + (1 - s) \cdot z$$

beschrieben. Man prüft leicht nach, dass jede zentrische Streckung des $\mathbb{R}^n$ eine Ähnlichkeitstransformation des $\mathbb{R}^n$ ist.

Als Korollar des Satzes von Mazur-Ulam 15.2.4 beweisen wir für Ähnlichkeitsabbildungen sofort:

Lemma 15.2.6. *(Ähnlichkeitsabbildungen sind spezielle affine Abbildungen)*
Sei $A = (V, V, \langle\cdot,\cdot\rangle)$ ein affiner euklidischer Raum und $g : A \to A$ eine Ähnlichkeitsabbildung mit Maßstab m. Dann gilt:

(i) Es gibt eine orthogonale lineare Abbildung $f : V \to V$ und einen Translationsvektor $t \in V$ so, dass für alle $v \in V$

$$g(v) = m \cdot f(v) + t$$

ist.

(ii) g lässt sich als Hintereinanderausführung einer Kongruenzabbildung und einer zentrischen Streckung schreiben.

Bemerkungen dazu:

- Im $A = \mathbb{R}^n$ bedeutet das schließlich, dass eine Ähnlichkeitsabbildung die Form

$$x \quad \mapsto \quad m \cdot Q \cdot x \; + \; t$$

mit orthogonaler Matrix $Q \in \mathbb{R}^{n \times n}$ und $t \in \mathbb{R}^n$ hat.

Beweis: Zu (i): Man betrachtet die Funktion $\tilde{g} := \frac{1}{m} \cdot g$, die wegen

$$\|\tilde{g}(v) - \tilde{g}(w)\| \; = \; \frac{1}{m} \cdot m \cdot \|x - y\| = \|x - y\|$$

eine Isometrie ist und sich daher nach Satz 15.2.4 als $\tilde{g} = f(v) + \tilde{t}$ mit orthogonaler Abbildung $g : V \to V$, $t \in V$ schreiben lässt; das zeigt $g = m \cdot f(v) + m \cdot \tilde{t}$, also die Behauptung (mit $t := m \cdot \tilde{t}$). (ii) ist Übungsaufgabe.

□

15.2.4 Grobe Systematisierung geometrischer Abbildungen

Wir fassen abschließend einige Erkenntnisse über geometrische Abbildungen im euklidischen Raum $\mathbb{R}^n$ zusammen.

- Jede Kongruenzabbildung ist eine Ähnlichkeitsabbildung.
- Kongruenzabbildungen und Ähnlichkeitsabbildungen im $\mathbb{R}^n$ sind bijektiv (da die lineare Abbildung f bijektiv ist, vgl. Lemma 13.1.2).
- Man prüft leicht: Die Menge der Kongruenzabbildungen und die Menge der Ähnlichkeitsabbildungen sind daher jeweils mit der Hintereinanderausführung als Verknüpfung eine Gruppe (die Kongruenzabbildnugen eine so genannte Untergruppe der Ähnlichkeitsabbildungen).
- Erinnerung: Eine bijektive affine Selbstabbildung $f : V \to V$ heißt Affinität, auch diese bilden eine Gruppe.
- Jede Ähnlichkeitsabbildung (und damit jede Kongruenzabbildung) ist also eine Affinität, aber nicht jede Affinität eine Ähnlichkeitsabbildung.
- Die affinen Abbildungen von $\mathbb{R}^n$ in sich sind mit der Hintereinanderausführung als Verknüpfung keine Gruppe (da affine Abbildungen nicht zwingend invertierbar sind).

Insgesamt ergibt sich folgende grobe Klassifikation der geometrischen Abbildungen im euklidischen affinen Raum $\mathbb{R}^n$:

$$\begin{array}{c}\text{Gruppe der Isometrien/Kongruenzabbildungen/Bewegungen}\\ \cap\!\!| \\ \text{Gruppe der Ähnlichkeitsabbildungen}\\ \cap\!\!| \\ \text{Gruppe der Affinitäten von } \mathbb{R}^n \\ \cap\!\!| \\ \text{Algebraische Struktur der affinen Abbildungen von } \mathbb{R}^n \text{ in sich}\end{array}$$

Wir haben für jede dieser Teilmengenbeziehung ein Beispiel gesehen, das zu der Obermenge, aber nicht zu der anderen gehört; jede dieser Inklusionen ist also echt.

15.2.5 Konstruktion geometrischer Abbildungen in $\mathbb{R}^n$, orthogonale Version

Wir beenden dieses Kapitel mit einer praktischen Bemerkung, in der es noch einmal konkret um die Konstruktion gewollter geometrischer Abbildungen im $\mathbb{R}^n$ geht: Die in 13.3.2 vorgestellte „Blaupause“ zur Konstruktion geometrischer Abbildungen lässt sich oft mit Hilfe von Orthonormalsystemen rechnerisch vereinfachen.

Wir rekapitulieren diese zunächst (und passen sie so etwas an, dass wir mit Orthogonalbasen arbeiten). Sei dazu $W = (0, I_n)$ das „Weltkoordinatensystem“.

- Ist ein affiner Teilraum $N = v + U$ z.B. als Spiegel- oder Projektionsraum oder als Drehebene (mit $\dim U = 2$) gegeben und B_U ein Erzeugendensystem von U, so lässt sich eine Basis $B = (b_1, \ldots, b_n)$ des $\mathbb{R}^n$ bestimmen, in der die ersten $\dim U$ Vektoren eine Basis von U sind, die durch $n-k$ Vektoren zu einer Basis B_n des $\mathbb{R}^n$ ergänzt werden. (Das geschieht durch Anwendung des Gauß-Jordan-Verfahrens auf die Matrix $(B_U \mid I_n)$ und Auswahl der Pivotspalten der Ursprungsmatrix, vgl. 11.3).
- Wendet man nun das Gram-Schmidt-Verfahren auf B_n an, so erhält man eine ONB $Q = (Q_U \mid Q')$ des $\mathbb{R}^n$, für die $\mathcal{L}(Q_U) = U$, $\mathcal{L}(Q') = U^\perp$ gilt.
- Transformiere nun auf das Koordinatensystem $K = (v, Q)$, vgl. dafür Lemma 12.5.2.
- Wende die darstellende Matrix $A = A_{Q,Q}$ bezüglich der Basis Q an, die die gewünschte Abbildung (z.B. Projektion, Drehung, Spiegelmatrix usw. oder eine andere Matrixabbildung , z.B. Achsenskalierung, Scherung, meist in möglichst einfacher Darstellung) im Koordinatensystem K darstellt, auf die transformierten Koordinatenvektoren an.
- Transformiere zurück in die Welt, wieder mit Lemma 12.5.2.

Insgesamt ergibt sich für so eine Abbildungsvorschrift die affine Abbildung

$$\boxed{x \;\mapsto\; QA(Q^{-1}(x-v)) \;+\; v \;=\; QAQ^T x \;+\; \underbrace{(I - QAQ^T)v}_{=:t}.}$$

Diese Vorschrift ist analog zu der, die in Abschnitt 13.3.2 gewonnen wurde. Da die Koordinatensysteme orthogonal gewählt wurden, hat man hier aber keine Probleme mit der Invertierung der beteiligten Matrizen.

15.3 Aufgaben zu Kapitel 15

Aufgabe 15.88. (Spiegelungen):

In der Vorlesung wurde definiert: Ist V ein Skalarproduktraum, U ein endlichdimensionaler Teilraum von V und $P^\perp$ der orthogonale Projektor auf $U^\perp$, so heißt die Abbildung $S\colon V \to V$ mit

$$S := \mathrm{id}_V - 2P^\perp$$

Spiegelung am Unterraum U.

(a) Zeigen Sie, dass S eine lineare Isometrie ist.
Tipp: Nutzen Sie dazu orthogonale Zerlegungen für Vektoren aus V.

(b) Zeigen Sie für S die Eigenschaften $Su = u$ für alle $u \in U$ sowie $Su = -u$ für alle $u \in U^\perp$.

(c) Skizzieren Sie für $V = \mathbb{R}^3$ die Wirkung der Spiegelung S für die Fälle $U = \mathcal{L}(e_1)$ bzw. $U = \mathcal{L}(e_1, e_2)$.

(d) Beschreiben Sie für $V = \mathbb{R}^3$ kurz, welche „Spiegelung“ sich für Unterräume U der Dimension $d = 0$ und der Dimension $d = 3$ ergibt.

Aufgabe 15.89. (Lineare Isometrien):

Gegeben seien drei Skalarprodukträume V, W, Z. Beweisen Sie folgende Aussagen aus Lemma 15.1.3:

(i) Sind $f\colon V \to W$ und $g\colon W \to Z$ lineare Isometrien, so ist auch $g \circ f\colon V \to Z$ eine lineare Isometrie.

(ii) Ist $f\colon V \to W$ eine lineare Isometrie, so ist f injektiv.

(iii) Ist V endlichdimensional, so gilt: Jede lineare Isometrie $f : V \to V$ ist bijektiv und ihre Inverse f^{-1} ist ebenfalls eine lineare Isometrie.

Aufgabe 15.90. (Senkrechte Projektion auf eine affine Ebene im $\mathbb{R}^3$):

Wir betrachten den $\mathbb{R}^3$ wie üblich als affinen Raum über sich selbst. Es sei die Ebene $\varepsilon = v + \mathcal{L}(q_1, q_2)$ mit

$$v = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad q_1 = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}, \quad q_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}$$

gegeben. In dieser Aufgabe soll die affine Abbildung $P_\varepsilon \colon \mathbb{R}^3 \to \mathbb{R}^3$ bestimmt werden, die Punkte $x \in \mathbb{R}^3$ orthogonal auf die affine Ebene ε projiziert. Führen Sie dazu folgende Schritte aus:

(i) Zeigen Sie, dass die Vektoren q_1, q_2 ein Orthonormalsystem sind, und ergänzen Sie dieses zu einer Orthonormalbasis $Q = (q_1, q_2, q_3)$ des $\mathbb{R}^3$.

(ii) Geben Sie die Koordinatentransformation k_K vom Weltkoordinatensystem auf das Koordinatensystem $K = (v; Q)$ und ihre Inverse in der affinen Form $x \mapsto M \cdot x + b$ an. (Beachten Sie, dass Sie hierbei keine Matrix invertieren müssen!)

(iii) Für einen Vektor

$$x = v + a_1 q_1 + a_2 q_2 + a_3 q_3 \in \mathbb{R}^3$$

definieren wir seine senkrechte Projektion auf ε als den Vektor

$$P_\varepsilon x = v + a_1 q_1 + a_2 q_2 \in \mathbb{R}^3.$$

Ermitteln Sie die Matrix A_K, die die Koordinaten des Vektors x bezüglich K auf die Koordinaten von $P_\varepsilon x$ bezüglich K abbildet.

(iv) Berechnen Sie die Abbildungsvorschrift

$$P_\varepsilon \colon \quad x \mapsto k_K^{-1}(A_K \cdot k_K(x))$$

der gesuchten Projektion und projizieren Sie den Punkt $x = \begin{pmatrix} 7 \\ 1 \\ 5 \end{pmatrix}$ auf ε.

16 Determinanten

Die so genannten Determinanten, die Thema dieses Kapitels sind,

- sind eine „Kennzahl“ für quadratische Matrizen und n-Tupel von Vektoren des $\mathbb{K}^n$
- gehen historisch auf die Untersuchung der Lösbarkeit linearer Gleichungssysteme zurück und sind dort ein beliebtes Werkzeug, um eben diese für kleine quadratische Systemmatrizen schnell beurteilen zu können,
- sind eine Möglichkeit, Flächen/Volumen/verallgemeinerte Volumen für Körper zu definieren, die von n Vektoren in $\mathbb{R}^n$, allgemeiner $\mathbb{K}^n$ aufgespannt werden,
- können zur Beschreibung der Orientierung von Winkeln und Koordinatensystemen dienen
- werden auch für lineare Abbildungen definiert und geben dort an, in welchem Maße die lineare Abbildung Flächen/Volumina verzerrt,
- vereinfachen von der algebraischen Seite auch die Untersuchung von Matrizen, in denen Variablen vorkommen – insbesondere bei der Berechnung von Eigenwerten von Matrizen (s. Kapitel 17)

16.1 Orientierte Flächen und Volumina in $\mathbb{R}^2$ und $\mathbb{R}^3$

Wir wählen für die Determinante einen Einstieg über die Untersuchung von von Vektoren eingeschlossenen Flächen und Volumina in Ebene und Raum, untersuchen hier einige grundlegende Eigenschaften von Determinantenfunktionen und verallgemeinern den Ansatz dann in den folgenden Abschnitten.

16.1.1 Orientierte Flächen in $\mathbb{R}^2$

Seien

$$x = \begin{pmatrix} a \\ c \end{pmatrix}, \quad y = \begin{pmatrix} b \\ d \end{pmatrix} \in \mathbb{R}^2.$$

Deutet man die Vektoren als Koordinatenvektoren von zwei Vektoren $x, y \in \mathbb{V}$ in der Ebene, so ist (wenn man Repräsentanten mit identischem Anfangspunkt betrachtet) die Fläche des Parallelogramms, das von den beiden Vektoren aufgespannt wird, gerade der Betrag der reellen (2×2)-Determinante

$$\det_2 : \mathbb{R}^{2\times 2} \to \mathbb{R}, \qquad \det_2(x, y) := \det_2 \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad := \quad ad - bc.$$

Diese 2×2-Determinante hat drei grundlegende Eigenschaften, die im Folgenden auch allgemein die von uns betrachteten Flächen- bzw. Volumenmessungen in $\mathbb{K}^n$ charakterisieren werden:

(D1) Festlegung der „Einheitsfläche“ (Normiertheit): Es ist

$$\det_2(I_2) = \det_2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1.$$

(D2) Festlegung „Flächeninhalt Null“: Sind die Vektoren x, y linear abhängig, so ist $\det_2(x, y) = 0$.

(D3) Lineare Ausdehnung des Flächeninhalts in jede Richtung: Die Determinantenfunktion ist linear in jeder Spalte, d.h. es gilt für Vektoren $x, y, x_1, x_2, y_1, y_2 \in \mathbb{R}^2$ und alle $s \in \mathbb{R}$

$$\det_2(s \cdot x_1 + x_2, y) = s \cdot \det_2(x_1, y) + \det_2(x_2, y) \quad \text{und}$$
$$\det_2(x, s \cdot y_1 + y_2) = \det_2(x, y_1) + s \cdot \det_2(x, y_2).$$

Viele andere Eigenschaften werden uns bei der Untersuchung und Berechnung von Determinanten hilfreich sein. Für die obige (2×2)-Determinante $\det_2$ sind sie schnell offensichtlich; wir diskutieren sie im Folgenden für diesen Spezialfall. Der Beweis aller Eigenschaften (D1) -(D8) ergibt sich jeweils durch Einsetzen der Behauptung in die Definition der (2×2)-Determinante und leichtes Rechnen – siehe Vorlesung, oder wenn Sie das hier gerade so lesen: Schreiben Sie die jeweiligen (2×2)-Matrizen hin und überzeugen Sie sich!

(D4) Bei Addition des s-fachen eines Vektors auf den anderen (mit $s \in \mathbb{R}$) gilt

$$\det_2(x, y + s \cdot x) = \det(x + s \cdot y, y) = \det(x, y).$$

Dass diese Eigenschaft im Zusammenhang mit Flächenberechnungen sinnvoll ist, lässt sich an einer gescherten Fläche gut sehen: Hier ändert sich die Höhe der Fläche und damit die Fläche nicht, wenn der eine Vektor in Richtung des anderen geschert wird; berücksichtigt wird bei der Berechnung nur der Anteil senkrecht zum ersten Vektor.

(D5) Bei Multiplikation einer Spalte mit $s \in \mathbb{K}$ gilt

$$\det_2(s \cdot x, y) = \det_2(x, s \cdot y) = s \cdot \det_2(A).$$

Die Aussage ist anschaulich klar und folgt direkt aus (D3). Bei der Multiplikation mehrerer Spalten mit $s \in \mathbb{K}$ ist dann

$$\det_2(s \cdot x, s \cdot y) = s^2 \cdot \det_2(x, y);$$

das bedeutet, dass der Flächeninhalt von Parallelogrammen in $\mathbb{R}^2$ quadratisch mit den Längen der Seiten wächst: Ver-c-facht man die Seitenlängen („skaliert die Figur also mit Faktor c"), so wächst die Determinante um den Faktor c^2, also: Bei Verdoppelung der Seitenlängen vervierfacht sich die Fläche, verdreifacht man die Seitenlängen, verneunfacht sich die Fläche usw.

Interessant ist es auch, sich die ebenfalls aus (D3) folgenden Aussage für die Additivität in jeder Spalte anhand von Parallelogrammen und den jeweiligen Flächeninhalten zu veranschaulichen (also z.B. für die erste Spalte die Aussage $\det_2(x_1+x_2, y) = \det_2(x_1, y) = \det_2(x_2, y)$).

Aus der Linearität (D3) folgt, dass $\det_2(x, y)$ auch negativ sein kann. Genauer gilt:

(D6) Die (2×2)-Determinante ändert beim Tauschen von Spalten ihr Vorzeichen, d.h. es ist $\det_2(x, y) = -\det_2(y, x)$.

Die 2×2-Determinante $\det_2$ wird daher auch als *orientierte* Flächeninhaltsfunktion und der Wert $\det_2(x, y)$ als *orientierter* Flächeninhalt bezeichnet. Insbesondere ist bei Interpretation im Anschauungsraum

$$\det_2(x, y) = \sin(\angle(\vec{x}, \vec{y})) \cdot |\vec{x}| \cdot |\vec{y}|$$

und, wenn wir mit $A(\vec{x}, \vec{y})$ den Flächeninhalt des durch $\vec{x}, \vec{y}$ aufgespannten Parallelogramms bezeichnen,

$$A(\vec{x}, \vec{y}) = |\det_2(x, y)|.$$

- Zusammen mit der Definition des Kosinus für Vektoren im $\mathbb{R}^2$ via Skalarprodukt[1] liefert das eine Möglichkeit der Definition von Winkeln in $\mathbb{R}^2$: Aus dem Kosinus zweier Vektoren in $\mathbb{R}^2$ erhält man im Allgemeinen *zwei* mögliche Winkel $\alpha, 2\pi - \alpha$ im Intervall $[0, 2\pi]$ (also mit $\alpha \in [0, \pi]$). Ist $\sin(x, y) > 0$, so wählt man den ersteren, ist $\sin(x, y) < 0$, den letzteren; vgl. Vorlesung oder eine hilfreiche Zeichnung.

- Man erhält damit auch ein Maß für die *Orientierung* zweier Vektoren $\vec{x}, \vec{y}$: Sind diese rechtshändig orientiert (entsprechen also, wenn Sie auf die Innenseite Ihrer rechten Hand schauen, wie im „Standard-Koordinatensystem" des Anschauungsraumes in ihrer Reihenfolge dem Daumen und Zeigefinger der rechten Hand), so ist $\det_2(x, y) > 0$, andernfalls ist $\det_2(x, y) < 0$ (und es ist $\det_2(x, y) = 0$ genau dann, wenn die Vektoren x, y l.u. sind).

(D7) Bezeichnet $A^T = (a_{j,i}) \in \mathbb{R}^{2\times 2}$ die transponierte Matrix von $A = (a_{i,j})$, so gilt $\det_2(A) = \det_2(A^T)$. Insbesondere ist $\det_2$ auch linear in jeder Zeile, und alle obigen Aussagen über Spalten von A gelten genauso für die Zeilen von A.

Die Verbindung der Determinantenfunktion zur Lösbarkeitstheorie linearer Gleichungssysteme schlummert in der folgenden Eigenschaft: Wegen (D2) und der ebenfalls geltenden Umkehrung gilt: gilt

(D8) Die Matrix $A = (x, y) \in \mathbb{R}^{2\times 2}$ ist invertierbar genau dann, wenn $\det_2(A) \neq 0$ ist.

Das haben wir in der LAAG I schon mal gesehen; die Inverse einer Matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

ist in diesem Fall die Matrix

$$A^{-1} = \frac{1}{ad - cb} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \frac{1}{\det_2(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Wir haben nun schon einen guten Überblick über das, was die Determinante von n Vektoren im $\mathbb{K}^n$ auch im allgemeinen Fall leisten wird.

[1] Zur Erinnerung: Es ist $\cos(x, y) = \frac{\langle x,y \rangle}{|x|\cdot|y|}$.

16.1.2 Orientierte Volumina in $\mathbb{R}^3$

Analog zur Ebene $\mathcal{A}_2$ und zur zugehörigen Determinantenfunktion $\det_2$ für jeweils zwei Vektoren des $\mathbb{R}^2$ können wir auch im $\mathbb{R}^3$ das Volumen untersuchen, das von drei Vektoren $\vec{x}, \vec{y}, \vec{z}$ im Anschauungsraum $\mathcal{A}_3$ aufgespannt wird. Der Körper

$$P := \{v \in \mathcal{A} \mid v = s_1 \cdot \vec{x} + s_2 \cdot \vec{y} + s_3 \cdot \vec{z}, s_1, s_2, s_3 \in [0, 1]\}$$

heißt das von $\vec{x}, \vec{y}, \vec{z}$ erzeugte Parallelepiped oder Spat. Um das Volumen solcher Parallelepidede in $\mathcal{A}_3$ mithilfe der drei zugehörigen Koordinatenvektoren x, y, z aus $\mathbb{R}^3$ zu messen, kann man wie in der Schule die Formel „Grundfläche $\times$ Höhe“ anwerfen. Ein eleganter rechnerischer Weg, der Eigenschaften von Grundfläche und Höhe vereint, benutzt das vor allem in der Physik unabdingbare, aber möglicherweise auch als der Schule bekannte Vektorprodukt oder Kreuzprodukt für (Koordinaten von) Vektoren im dreidimensionalen Anschauungsraum.

Definition 16.1.1. *(Kreuzprodukt zweier Vektoren in $\mathbb{R}^3$)*

Sind $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \in \mathbb{R}^3$ *so definiert man das Kreuzprodukt $x \times y$ als den Vektor*

$$x \times y := \begin{pmatrix} x_2y_3 - x_3y_2 \\ x_3y_1 - x_1y_3 \\ x_1y_2 - x_2y_1 \end{pmatrix} \in \mathbb{R}^3. \tag{16.1}$$

Das Vektorprodukt unterscheidet sich insofern wesentlich vom Skalarprodukt, also dass es

- eben als Ergebnis kein Skalar, sondern einen Vektor liefert und
- ausschließlich im $\mathbb{R}^3$ (als Koordinatenraum des Anschauungsraunmes) definiert ist.

Deutet man die Vektoren $x, y \in \mathbb{R}^3$ als die Koordinatenvektoren von Pfeilklassen $\vec{x}, \vec{y}$ im Anschauungsraum $\mathcal{A}_3$ bezüglich eines gegebenen kartesische Koordinatensystems, so liefert das Kreuzprodukt $x \times y$ den Koordinatenvektor eines Vektors $\vec{z} = \vec{x} \times \vec{y} \in \mathcal{A}_3$ mit folgenden Eigenschaften:

(i) $\vec{z} \perp \mathcal{L}(\vec{x}, \vec{y})$,

(ii) Der Betrag von $\vec{z}$ ist genau der Flächeninhalt $A(\vec{x}, \vec{y})$ des von $\vec{x}, \vec{y}$ erzeugten Parallelogramms,

(iii) Die Orientierung des Vektortripels $(\vec{x}, \vec{y}, \vec{z})$ entspricht der Orientierung des gewählten Koordinatensystems in $\mathcal{A}$, d.h ist abhängig davon entweder ein linkshändiges oder rechtshändiges Koordinatensystem

Umgekehrt lässt sich die Formel (16.1) aus den Eigenschaften (i)- (iii) herleiten, vgl. [Filler], S. 133-135. Das Kreuzprodukt besitzt einige nützliche ergänzende Eigenschaften, die sich direkt aus den obigen ableiten lassen, zum Beispiel, dass das Kreuzprodukt zweier linear abhängiger Vektoren den Nullvektor ergibt, oder, dass das Kreuzprodukt zweier senkrechter Vektoren x, y einen Vektor vom Betrag $|x| \cdot |y|$ ergibt usw. Einige dieser Eigenschaften finden sich in den Übungsaufgaben zur Vorlesung.

Für uns reicht hier zunächst die folgende Erkenntnis: Aus den Eigenschaften (i)-(iii) ergibt sich mit dem Ansatz „Grundfläche $\times$ Höhe“ die folgende Formel für das Volumen eines von Vektoren $\vec{x}, \vec{y}, \vec{z} \in \mathcal{A}_3$ aufgespannten Spats P: Für den Vektor $\vec{z}$ gilt $\vec{z} = A(\vec{x}, \vec{y}) \cdot \vec{n}$, wobei $\vec{n} \perp \mathcal{L}(\vec{x}, \vec{y})$ und $|\vec{n}| = 1$ ist, $\vec{n}$ also ein normierter Normalenvektor ist. Daraus erhält man

$$V(P) \quad = \quad A(\vec{x}, \vec{y}) \cdot \langle \vec{n}, \vec{z} \rangle \quad = \quad |(x \times y) \cdot z|.$$

Die zugehörige Größe

$$\det_3 : \mathbb{R}^{3\times 3} \to \mathbb{R}, \qquad \det_3(x, y, z) := (x \times y) \cdot z$$

ist ein orientiertes Volumen in $\mathbb{R}^3$. Sie lässt sich berechnen durch

$$\det(x, y, z) \quad := \quad (x \times y) \cdot z = \begin{pmatrix} x_2 y_3 - x_3 y_2 \\ x_3 y_1 - x_1 y_3 \\ x_1 y_2 - x_2 y_1 \end{pmatrix} \cdot \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} \tag{16.2}$$

$$= \quad x_1 y_2 z_3 + y_1 z_2 x_3 + z_1 x_2 y_3 - x_3 y_2 z_1 - y_3 z_2 x_1 - z_3 x_2 y_1. \tag{16.3}$$

Das sieht auf den erstem Blick recht unübersichtlich aus, wird aber durchschaubarer, wenn man diese Sarrus'sche-(3×3)-Regel als „Lattenzaunregel“ interpretiert, s. Vorlesung. Die reelle (3×3)-Determinante besitzt wie die reelle (2×2)-Determinante des letzten Abschnitts die Eigenschaften (D1)-(D8) (wobei jetzt $\det_3(c \cdot x, c \cdot y, c \cdot z) = c^3 \det_3(x, y, z)$ gilt, Volumina also *kubisch* vmit der Skalierung der Kantenlängen wachsen; macht Sinn, oder?). Mit $\det_3$ und der Sarrus-Regel lassen sich daher gängige Aufgabenstellungen der Linearen Algebra I für drei Vektoren $x, y, z \in \mathbb{R}^3$ und Matrizen $A \in \mathbb{R}^{3\times 3}$ relativ zügig überblicken; in der Vorlesung sehen wir uns als Beispiel die Frage nach der Invertierbarkeit einer parameterabhängigen (3×3)-Matrix an.

Es fehlt noch der Beweis dafür, dass $\det_3$ wirklich die Eigenschaften (D1)-(D8) besitzt. Bevor wir hier anfangen, diese Eigenschaften für jede Determinantenfunktion einzeln nachzurechnen, machen wir uns jetzt auf zu allgemeinen, systematischen Betrachtungen zu Determinantenfunktionen.

16.2 Eigenschaften von Determinantenfunktionen in $\mathbb{K}^n$

Definition 16.2.1. *(Determinantenfunktion in $\mathbb{K}^n$; Determinante einer Matrix)*

Sei $\mathbb{K}$ ein Körper und $n \in \mathbb{N}$, $x_1, \ldots, x_n \in \mathbb{K}^n$ Vektoren. Dann heißt eine Abbildung

$$\det : \mathbb{K}^{n\times n} \to \mathbb{K}, \quad (x_1, \ldots, x_n) \mapsto \det(x_1, \ldots, x_n)$$

Determiantenfunktion des $\mathbb{K}^n$, wenn det *die folgenden drei Eigenschaften besitzt:*

(D1) Normiertheit: Es ist

$$\det(I_n) = \det(e_1, \ldots, e_n) = 1.$$

($I_n = (e_1, \ldots, e_n)$ bezeichnet wie immer die Standardbasis des $\mathbb{K}^n$.)

(D2) Definition von Nullvolumina: Ist $A = (x_1, \ldots, x_n) \in \mathbb{K}^{n\times n}$ ein n-Tupel von Vektoren und $\mathrm{rg}(A) < n$, so ist $\det(x_1, \ldots, x_n) = 0$.

(D3) Linearität in jeder „Richtung“: Für alle Spaltenindizes $i \in \underline{n}$, alle $x_1, \ldots, x_n, y_i \in \mathbb{K}^n$ und alle Skalare $s \in \mathbb{K}$ gilt

$$\det(x_1, \ldots, s \cdot x_i + y_i, \ldots, x_n) \quad = \quad s \cdot \det(x_1, \ldots, x_i, \ldots, x_n) \ + \ \det(x_1, \ldots, y_i, \ldots, x_n)$$

Den Funktionswert $\det(x_1, \ldots, x_n) \in \mathbb{K}$ nennt man auch Determinante oder orientiertes Volumen der n Vektoren $x_1, \ldots, x_n$.

Ist $A \in \mathbb{K}^{n\times n}$ eine quadratische Matrix mit Spalten $s_1, \ldots, s_n \in \mathbb{K}^n$, so spricht auch von der Determinante der Matrix A und meint damit

$$\det(A) := \det(s_1, \ldots, s_n).$$

Wir werden diese Notation auch oftmals als Kurzschreibweise benutzen.

Bevor wir fortfahren, stellen wir fest, dass dieser minimale Forderungskatalog bereits die Eigenschaften (D4)-(D6), die wir oben für $\det_2$ nachgerechnet haben, immer direkt impliziert:

Lemma 16.2.2. *(Eigenschaften von Determinantenfunktionen)*

Ist $\mathbb{K}$ ein Körper, $n \in \mathbb{N}$, $\det : \mathbb{K}^{n\times n} \to \mathbb{K}$ eine Determinantenfunktion und $A, B \in \mathbb{K}^{n\times n}$, so gilt für $\det$:

(D4) Geht B aus A durch Addition des c-fachen einer Spalte zu einer anderen hervor (mit $c \in \mathbb{K}$), so ist $\det(B) = \det(A)$.

(D5) Geht B aus A durch Multiplikation einer Spalte mit $c \in \mathbb{K}$ hervor, so ist $\det(B) = c \cdot \det(A)$.

(D6) Geht B aus A durch Vertauschung zweier Spalten hervor, so gilt $\det(B) = -\det(A)$.

Beweis: Zu (D4) sei $A = (s_1, \dots, s_n) \in \mathbb{K}^{n\times n}$ und $c \in \mathbb{K}$. Dann gilt

$$\begin{aligned}
& \det(s_1, \dots, s_i, \dots, s_j + c \cdot s_i, \dots, s_n) \\
\overset{\text{(D3)}}{=} \; & \det(s_1, \dots, s_i, \dots, s_j, \dots, s_n) + \det(s_1, \dots, s_i, \dots, c \cdot s_i, \dots, s_n) \\
\overset{\text{(D2)}}{=} \; & \det(A) + 0 = \det(A).
\end{aligned}$$

(D5) ist einfach die Linearität in der entsprechenden Spalte. Für (D6) wenden wir die vorigen Regeln kreativ an: bezeichnen wir mit s_i und s_j die beiden zu tauschenden Spalten, so ist

$$\begin{aligned}
\det(A) & \overset{\text{(D3)}}{=} \det(s_1, \dots, s_i + s_j, \dots, s_j, \dots, s_n) \\
& \overset{\text{(D3)}}{=} \det(s_1, \dots, s_i + s_j, \dots, s_j + (-1) \cdot (s_i + s_j), \dots, s_n) \\
& = \det(s_1, \dots, s_i + s_j, \dots, -s_i, \dots, s_n) \\
& \overset{\text{(D5)}}{=} -\det(s_1, \dots, s_i + s_j, \dots, s_i, \dots, s_n) \overset{\text{(D4)}}{=} -\det(s_1, \dots, s_j, \dots, s_i, \dots, s_n).
\end{aligned}$$

□

Für spätere Zwecke halten wir noch fest, dass man den Anforderungskatalog (D1)-(D3) für Determinantenfunktionen sogar noch etwas abschwächen kann:

Lemma 16.2.3. *(„Minimalere“ Forderungen an Determinantenfunktionen)*
Um zu zeigen, dass eine Abbildung

$$\det : \mathbb{K}^{n\times n} \to \mathbb{K}, \quad (x_1, \ldots, x_n) \mapsto \det(x_1, \ldots, x_n)$$

eine Determiantenfunktion des $\mathbb{K}^n$ ist, reicht es aus, wenn det *die Eigenschaften (D1), (D3) und die Eigenschaft*

> *(D2') Für alle Matrizen A mit zwei identischen Spalten ist* $\det(A) = 0$.

besitzt.

Beweis: Sei $\det : \mathbb{K}^{n\times n} \to \mathbb{K}$ eine Abbildung mit den Eigenschaften (D1), (D2') und (D3) und $A = (s_1, \ldots, s_n) \in \mathbb{K}^{n\times n}$ mit $\mathrm{rg}(A) < n$; wir haben zu zeigen, dass daraus dann $\det(A) = 0$ folgt. Ist $\mathrm{rg}(A) < n$, so lässt sich mindestens eine der Spalten von A als Linearkombination der übrigen schreiben. Sei also $i \in \underline{n}$ so gewählt, dass für die i-te Spalte $s_i = \sum_{j\neq i} a_j \cdot s_j$ für geeignete Koeffizienten a_j, $j \neq i$ gilt. Dann ist

$$\det(A) \;=\; \det(s_1, \ldots, \underbrace{\sum_{j\neq i} a_j \cdot s_j}_{i-\text{te Spalte}}, \ldots, s_n) \overset{\text{(D3)}}{=} \sum_{j\neq i} a_j \cdot \det(s_1, \ldots, \underbrace{s_j}_{i-\text{te Spalte}}, \ldots, s_n).$$

In der letzten Summe sind alle Determinanten nach (D2') null, da die i-te Spalte s_j gerade mit der Spalte s_j übereinstimmt; damit folgt (D2).

□

16.2.1 Existenz von Determinantenfunktionen

Wir kümmern uns in diesem Abschnitt um die Fragestellung, ob sich in jedem Koordinatenraum $\mathbb{K}^n$ eine Determinantenfunktion zur Messung (verallgemeinerter) orientierter Volumen finden lässt. Als Inspration betrachten wir nochmal die Determinantenfunktion $\det_3$ des $\mathbb{R}^3$: Es ist

$$\begin{aligned}
\det(x, y, z) &= x_1y_2z_3 + y_1z_2x_3 + z_1x_2y_3 - x_3y_2z_1 - y_3z_2x_1 - z_3x_2y_1 \\
&= x_1 \cdot \det\begin{pmatrix} y_2 & z_2 \\ y_3 & z_3 \end{pmatrix} - y_1 \cdot \det\begin{pmatrix} x_2 & z_2 \\ x_3 & z_3 \end{pmatrix} + z_1 \cdot \det\begin{pmatrix} x_2 & y_2 \\ x_3 & y_3 \end{pmatrix},
\end{aligned}$$

die reelle (3×3)-Determinante lässt sich also durch so genannte Entwicklung nach der ersten Zeile *rekursiv* aus den angegebenen (2×2)-Unterdeterminanten berechnen. Wir führen dafür eine Notation ein.

Definition 16.2.4. *(Untermatrizen $A_{i,j}$)*
Ist A eine $(n \times n)$-Matrix mit $n \geq 2$ Zeilen und $i, j, \in \underline{n}$, so bezeichnen wir mit $A_{i,j}$ die $((n-1) \times (n-1))$-Matrix, die durch Streichung der i-ten Zeile und j-ten Spalte von A entsteht.

Mit der Definition oben ist für $A = (x, y, z) \in \mathbb{R}^{3\times 3}$

$$\det(x, y, z) = (+1) \cdot x_1 \cdot \det(A_{1,1}) + (-1) \cdot y_1 \cdot \det(A_{1,2}) + (+1) \cdot z_1 \cdot \det(A_{1,3}).$$

Dieser Ansatz lässt sich benutzen, um rekursiv Determinantenfunktionen im $\mathbb{K}^n$ für beliebige große n zu definieren.

Satz 16.2.5. *(Existenz von Determinantenfunktionen in jedem Raum $\mathbb{K}^n$)*
Sei $\mathbb{K}$ ein Körper. Dann gilt:

(i) $\det_1 : \mathbb{K} \to \mathbb{K}, \quad \det(a) = a$ *definiert eine Determinantenfunktion auf* $\mathbb{K} = \mathbb{K}^1$,

(ii) Ist $n \in \mathbb{N}$, $x_1, \dots, x_n \in \mathbb{K}^n$ und A eine zugehörigen Matrix mit diesen Vektoren als Spalten, und wir schreiben dafür

$$x_1 = \begin{pmatrix} x_{1,1} \\ x_{2,1} \\ \vdots \\ x_{n,1} \end{pmatrix}, \dots, x_n = \begin{pmatrix} x_{1,n} \\ x_{2,n} \\ \vdots \\ x_{n,n} \end{pmatrix}, \qquad A = (x_1, \dots, x_n) \in \mathbb{K}^{n\times n},$$

so definiert die Rekursionsformel

$$\det_n : \mathbb{K}^{n\times n} \to \mathbb{K}, \qquad \det_n(x_1, \dots, x_n) := \sum_{j=1}^{n} (-1)^{j-1} \cdot x_{1,j} \cdot \det_{n-1}(A_{1,j}) \tag{16.4}$$

(die „Determinantenberechnung durch Entwicklung nach der ersten Zeile“) eine Determinantenfunktionen auf $\mathbb{K}^n$ für alle $n \in \mathbb{N}$.

So lässt sich die Determinante einer reellen (4×4)-Matrix durch Entwicklung nach der ersten Zeile und die Berechnung von vier (3×3)-Determinanten berechnen, ein Beispiel dazu sehen wir uns in der Vorlesung an. Da man dabei im Allgemeinen $n!$ Produkte mit je n Faktoren berechnen muss, ist das selbst für moderate n nicht mehr effizient berechenbar (z.B. ist $13! = 6227020800$); es geht hier aber um das theoretische Existenzresultat, zur Berechnung siehe Abschnitt 16.3.

Beweis: Für Punkt (i) prüft man direkt einfach anhand der Forderungen (D1), (D2), (D3) nach. Für (ii) nutzen wir 16.2.3 und zeigen, dass $\det_n$ die Determinantenaxiome (D1), (D2') und (D3) erfüllt. Zu (D1) erhält man für beliebiges $n \in \mathbb{N}$ und die Einheitsmatrix $I_n \in \mathbb{K}^{n\times n}$ durch Entwicklung nach der ersten Zeile, dass

$$\det_n(I_n) \;=\; 1 \cdot 1 \cdot \det_{n-1}(I_{n-1}) + \sum_{k=2}^{n} (-1)^{k-1} \cdot \underbrace{0}_{x_{1,k} \text{ für } k>1} \cdot \det_{n-1}(A_{1,k}) = \det_{n-1}(I_{n-1}),$$

also, da n beliebig war,

$$\det_n(I_n) = \det_{n-1}(I_{n-1}) = \ldots = \det_2(I_2) = \det_1(1) = 1.$$

Für (D2') und (D3) benutzen wir das Beweisverfahren der vollständigen Induktion über alle $n \in \mathbb{N}$: Aus (i) wissen wir, dass (D2') und (D3) für $n = 1$ gilt. Sei $n \in \mathbb{N}$ das kleinste n, für das wir (D2') und (D3) noch nicht bewiesen haben (also zunächst $n = 2$, dann $n = 3, \ldots$). Wir zeigen auf der Grundlage, dass (D2') und (D3) für kleinere n stimmt (was wir zunächst für $n = 1$ wissen, dann für $n = 2, \ldots$), dass (D2') und (D3) immer auch für die nächstgrößere Zahl n stimmen muss, und bestätigen so sukzessive für beliebig große n induktiv, dass $\det_n$ eine Determinantenfunktion des $\mathbb{K}^n$ ist.

Zu (D2') sei $A \in \mathbb{K}^{n\times n}$ und $s_i = s_j$ für zwei Spalten mit Indizes $i \neq j$. Dann ist zunächst

$$\begin{aligned} \det_n(s_1, \ldots, s_n) \;&:=\; \sum_{k=1}^{n} (-1)^{k-1} \cdot x_{1,k} \cdot \det_{n-1}(A_{1,k}) \\ &=\; (-1)^{i-1} \cdot x_{1,i} \cdot \det_{n-1}(A_{1,i}) + (-1)^{j-1} \cdot x_{1,j} \cdot \det_{n-1}(A_{1,j}) = (\star), \end{aligned}$$

da alle anderen $(n-1) \times (n-1)-$Untermatrizen $A_{1,k}$ mit den „Überresten“ der Spalten s_i und s_j zwei identische Spalten enthalten und wir wissen, dass (D2') für alle Matrizen mit weniger als n Spalten gilt, also $\det_{n-1}(A_{1,k}) = 0$ für $k \neq i, j$ gilt. Da $s_i = s_j$ ist, ist in der unteren Zeile oben $x_{1,i} = x_{1,j}$, und die Matrix $A_{1,j}$ eine Matrix, die man aus $A_{1,i}$ durch $|j - i - 1|$ Spaltenvertauschungen erhält (z.B., indem man die Spalte mit dem größeren Index Spalte für Spalte nach vorne tauscht). Da *(D6)* für $\det_{n-1}$ gilt (denn es folgt nach

Lemma 16.2.2 aus (D1)-(D3), die für $\det_{n-1}$ gelten), ist also

$$\begin{aligned}(\star) &= (-1)^{i-1}\cdot x_{1,i}\cdot \det_{n-1}(A_{1,i}) + (-1)^{j-1}\cdot(-1)^{|j-i-1|}\cdot x_{1,i}\cdot \det_{n-1}(A_{1,i})\\ &= \left((-1)^{i-1} + (-1)^{j-1}\cdot(-1)^{|j-i-1|}\right)\cdot x_{1,i}\cdot \det_{n-1}(A_{1,i}).\end{aligned}$$

Zum Beispiel durch die Fallunterscheidung, ob $|j-i-1|$ gerade oder ungerade ist, erhält man $(-1)^{j-1}\cdot(-1)^{|j-i-1|} = -(-1)^{i-1}$ und damit $\det_n(s_1,\ldots,s_n) = 0$ wie behauptet.

Für die Linearität (D3) sei ein Spaltenindex $i \in \underline{n}$ fixiert. Für $x_1,\ldots,x_n, v \in \mathbb{K}^n$ nutzen wir die Bezeichnung $A_v = (x_1,\ldots,\underbrace{v}_{i-\text{te Spalte}},\ldots,x_n)$. Zu zeigen haben wir dann, dass für alle $x_i, y_i \in \mathbb{K}^n$ und $s \in \mathbb{K}$

$$\det_n(A_{s\cdot x_i+y_i}) = s\cdot \det_n(A_{x_i}) + \det_n(A_{y_i})$$

gilt. Dazu spalten wir in (16.4) zunächst den Summanden für $k = i$ ab und berechnen unter Nutzung von (D3) für $\det_{n-1}$

$$\begin{aligned}\det_n(A_{s\cdot x_i+y_i}) &= (-1)^{i-1}\cdot(s\cdot x_{1,i}+y_{1,i})\cdot\det_{n-1}(A_{1,i}) + \sum_{k\neq i}(-1)^{k-1}\cdot x_{1,k}\cdot\det_{n-1}\left((A_{s\cdot x_i+y_i})_{1,k}\right)\\ &\overset{(D3)}{=} s\cdot(-1)^{i-1}\cdot x_{1,i}\cdot\det_{n-1}(A_{1,i}) + (-1)^{i-1}\cdot y_{1,i}\cdot\det_{n-1}(A_{1,i})\\ &\quad +s\cdot\sum_{k\neq i}(-1)^{k-1}\cdot x_{1,k}\cdot\det_{n-1}\left((A_{x_i})_{1,k}\right) + \sum_{k\neq i}(-1)^{k-1}\cdot x_{1,k}\cdot\det_{n-1}\left((A_{y_i})_{1,k}\right)\\ &= s\cdot\left((-1)^{i-1}\cdot x_{1,i}\cdot\det_{n-1}(A_{1,i}) + \sum_{k\neq i}(-1)^{k-1}\cdot x_{1,k}\cdot\det_{n-1}\left((A_{x_i+y})_{1,k}\right)\right)\\ &\quad +\left((-1)^{i-1}\cdot y_{1,i}\cdot\det_{n-1}(A_{1,i}) + \sum_{k\neq i}(-1)^{k-1}\cdot x_{1,k}\cdot\det_{n-1}\left((A_y)_{1,k}\right)\right)\\ &=: s\cdot\det_n(A_{x_i}) + \det_n(A_y).\end{aligned}$$

□

Man könnte vermuten, dass die Determinantenfunktion $\det_n$ bei weitem nicht die einzige Möglichkeit ist, im $\mathbb{K}^n$ unter Berücksichtigung der Forderungen (D1)-(D3) orientierte Volumina zu definieren. Falls man das denkt, ist der folgende Satz recht überraschend.

Satz 16.2.6. *(Eindeutigkeit der Determinantenfunktion)*

Sei $\mathbb{K}$ ein Körper und $n \in \mathbb{N}$. Dann gibt es auf $\mathbb{K}^n$ genau eine Determinantenfunktion/orientierte Volumenfunktion, also genau eine Funktion

$$\det : \mathbb{K}^{n\times n} \to \mathbb{K}$$

mit den Eigenschaften (D1)-(D3).

Im $\mathbb{R}^2$ und $\mathbb{R}^3$ ist diese Abbildung mit den Eigenschaften (D1) bis (D3) also durch die in Abschnitt 16.1 angegebenen Determinantenabbildungen $\det_2$ bzw. $\det_3$ gegeben, im $\mathbb{K}^n$ durch die Rekursionsformel (16.4) für $\det_n$ eindeutig festgelegt. Daher werden wir die Determinantenfunktionen $\det_2, \det_3$ und allgemeiner die Determinantenfunktion $\det_n$ des $\mathbb{K}^n$ für gewöhnlich kurz einfach durch det notieren.

Beweis: Wir zeigen: Sind $\det, \det' : \mathbb{K}^{n\times n} \longrightarrow \mathbb{K}$ zwei Determinantenfunktionen, also Abbildungen mit den Eigenschaften $(D1) - (D3)$, so ist $\det(A) = \det'(A)$ für jede Matrix $A \in \mathbb{K}^{n\times n}$. Ist rang $A < n$, dann ist, da det und $\det'$ beide (D2) erfüllen, $\det(A) = 0 = \det'(A)$. Sei also rang $A = n$. Durch Anwendung von (D4), (D5) und (D5) können wir A durch elementare Spaltenumformungen in die Einheitsmatrix I_n verwandeln, indem wir (a) zuerst Zeile für Zeile Nullen rechts der Diagonalen erzeugen, dann (b) die Diagonaleinträge auf 1 normieren, dann (c) die Einträge unter den Diagonaleinträgen zu Nullen machen. Hierbei entsteht die Einheitsmatrix, da $\mathrm{rg}(A) = n$ war und der Rang unter Spaltenumformungen erhalten bleibt. Bei Anwendung von (D5), (D6) erhält man eine Kette von Umformungen

$$A \to A_1 \to A_2 \to \ldots \to I_n$$

mit *denselben* Vorfaktoren $\alpha_1, \ldots, \alpha_m$ in den mit Hilfe von (D4), (D5) und (D6) resultierenden Beziehungen

$$\begin{aligned} \det(A) &= \alpha_1 \cdot \det(A_1) = \alpha_1 \cdot \alpha_2 \cdot \det(A_2) = \ldots = \alpha_1 \cdot \alpha_2 \cdot \ldots \cdot \alpha_m \cdot \det(I_n) \\ \det'(A) &= \alpha_1 \cdot \det(A_1) = \alpha_1 \cdot \alpha_2 \cdot \det(A_2) = \ldots = \alpha_1 \cdot \alpha_2 \cdot \ldots \cdot \alpha_m \cdot \det'(I_n). \end{aligned}$$

Nach (D1) ist $\det(I_n) = 1 = \det'(I_n)$, also auch $\det(A) = \alpha_1 \cdot \alpha_2 \cdot \ldots \cdot \alpha_m = \det'(A)$. □

16.3 Berechnung und weitere Eigenschaften von Determinanten

Zur praktischen Berechnung von Determinanten gibt es einige weitere technische Kniffe und Zusammenhänge, die weitere theoretische Einsichten bringen und die praktische Berechnung für spezielle Matrizen oft erleichtern und die in diesem Kapitel zusammengestellt sind. Wir beginnen mit der folgenden Verallgemeinerung der Eigenschaft (D8), die wir Anfang des Kapitels für (2×2)-Matrizen festgestellt haben, und die bei der Nutzung von Determinanten als Werkzeug in der Lösbarkeitstheorie von linearen Gleichungssystemen zentral ist.

Satz 16.3.1. *(Invertierbarkeit per Determinante prüfen)*

Sei $A \in \mathbb{K}^{n \times n}$. Es sind äquivalent:

(i) *A besitzt eine Inverse A^{-1},*

(ii) $\operatorname{rg}(A) = n$,

(iii) $\det(A) \neq 0$.

Beweis: Die Äquivalenz $(i) \Leftrightarrow (ii)$ ist hoffentlich ein alter Hut, siehe Satz 6.4.1. (Es sollte auch bekannt sein, dass zu diesen Aussagen noch einige andere Aussagen äquivalent sind, eindeutige Lösbarkeit aller Gleichungssysteme mit Systemmatrix A, dass die Spalten von A eine Basis des $\mathbb{K}^n$ sind usw.). Die Implikation $(iii) \Rightarrow (ii)$ (bzw. ihre Kontraposition) ist gerade die Aussage (D2) für det. Zu zeigen bleibt also $(ii) \Rightarrow (iii)$: Ist $\operatorname{rg}(A) = n$, so kann man die Matrix A wie im Beweis von Satz 16.2.6 in die Einheitsmatrix umwandeln, wobei sich aus dem Verfahren ergibt, dass alle Vorfaktoren α_i in der Gleichungskette

$$\det(A) = \alpha_1 \cdot \det(A_1) = \alpha_1 \cdot \alpha_2 \cdot \det(A_2) = \ldots = \alpha_1 \cdot \ldots \cdot \alpha_m \cdot \det(I_n) = \alpha_1 \cdot \alpha_2 \cdot \ldots \cdot \alpha_m$$

ungleich Null sind. Es folgt $\det(A) \neq 0$.

□

Der zweite Zusammenhang verallgemeinert schließlich die Eigenschaft (D7), die wir ebenfalls am Anfang dieses Kapitels bereits für (2×2)-Matrizen verifiziert haben.

Satz 16.3.2. *(Determinante transponierter Matrizen)*

Sei $A \in \mathbb{K}^{n \times n}$. Bezeichnet $A^T = (a_{j,i}) \in \mathbb{K}^{n \times n}$ die transponierte Matrix von $A = (a_{i,j})$, so gilt

$$\det(A) = \det(A^T). \tag{16.5}$$

Beweis: Man prüft analog zum Beweis von Satz 16.2.5 induktiv, dass die Determinantenabbildung $\det = \det_n$ auch linear in jeder Zeile ist (wobei der Begriff analog zur Linearität in jeder Spalte definiert ist). Hierzu hilft es, die Behauptung einzeln für die erste Zeile und beliebige der restlichen nachzurechnen; das ist eine Übungsaufgabe. Daraus folgt, dass die Abbildung

$$\det^T : \mathbb{K}^{n\times n} \to \mathbb{K}, \quad A \mapsto \det(A^T)$$

linear in jeder Spalte ist und somit (D3) erfüllt. Offensichtlich gilt für sie auch (D1) und wegen der Gleichheit von Spaltenrang und Zeilenrang auch (D2), also ist nach dem Eindeutigkeitssatz 16.2.6 $\det^T = \det$, also $\det(A) = \det^T(A) = \det(A^T)$ für alle Matrizen $A \in \mathbb{K}^{n\times n}$.

□

Der vorangegangene Satz ist deshalb wichtig, weil er impliziert, dass Eigenschaften der Determinante, die für Spalten von A gelten, analog auch immer für die Zeilen gelten (da man nach dem obigen Satz A durch A^T ersetzen kann und dort die entsprechende Eigenschaft für die Spalten anwenden kann (Genauer: Vgl. Vorlesung). Insbesondere gilt:

Korollar 16.3.3. *(Rechenregeln für Zeilenumformungen)*

Seien $A, B \in \mathbb{K}^{n\times n}$.

(D4') Geht B aus A durch Addition des c-fachen einer Zeile zu einer anderen hervor (mit $c \in \mathbb{K}$), so ist $\det(B) = \det(A)$.

(D5') Geht B aus A durch Multiplikation einer Zeile mit $c \in \mathbb{K}$ hervor, so ist $\det(B) = c \cdot \det(A)$.

(D6') Geht B aus A durch Vertauschung zweier Zeilen hervor, so gilt $\det(B) = -\det(A)$.

Aus dem obigen Korollar folgt insbesondere, dass elementare Zeilenumformungen und Durchführung des Gauß-Jordan-Algorithmus die Determinante eine Matrix nicht ändern. Das schafft einen effizienten Weg zur Berechnung von Determinanten, den wir jetzt entwickeln: Man bringt Matrizen damit auf die im folgenden definierte Diagonalform, um daraus die Determinante mit dem darauf folgenden Lemma abzulesen; das ist im darauf folgenden Algorithmus nochmal genauer zusammengefasst.

Definition 16.3.4. *(Obere und untere Dreiecksmatrizen, Diagonalmatrizen)*

Sei $A \in \mathbb{K}^{n \times n}$. Dann heißt A

- *obere Dreiecksmatrix, falls alle Einträge unterhalb der Diagonalen null sind, d.h. es gilt $a_{i,j} = 0$ für alle Indizes $i > j$,*
- *untere Dreiecksmatrix, falls alle Einträge oberhalb der Diagonalen null sind, d.h. es gilt $a_{i,j} = 0$ für alle Indizes $j > i$,*
- *Diagonalmatrix, falls A höchstens auf der Diagonalen Einträge ungleich null besitzt, d.h. es ist $a_{i,j} = 0$ für alle Indizes $i \neq j$.*

Man beachte, dass über die anderen Einträge von A nichts gefordert wird (insbesondere nicht, dass diese nicht null sind). Die Nullmatrix aus $\mathbb{K}^{n \times n}$ ist also beispielsweise eine obere und untere Dreiecksmatrix und eine Diagonalmatrix.

Lemma 16.3.5. *(Determinante von Dreiecksmatrizen, Diagonalmatrizen)*
Ist A eine obere oder untere Dreiecksmatrix, so ist ihre Determinante gerade das Produkt der Einträge auf der Diagonalen, d.h. es ist

$$\det(A) = \prod_{i=1}^{n} a_{i,i}.$$

Insbesondere sind Dreiecks- und Diagonalmatrizen nach 16.3.1 invertierbar genau dann, wenn alle Einträge auf der Diagonalen ungleich Null sind.

Beweis: Für obere Dreiecksmatrizen und Diagonalmatrizen folgt das einfach durch Anwendung der rekursiven Definition für die Determinante; für untere Dreiecksmatrizen betrachtet man stattdessen die obere Dreiecksmatrix A^T.

$\square$

Algorithmus: Determinante per Zeilenumformungen berechnen

Eingabe: Matrix $A \in \mathbb{K}^{n \times n}$.

- Bringe die Matrix $A \in \mathbb{K}^{n \times n}$ durch die Zeilenumformungen
  (i) Vertauschen von Zeilen,
  (ii) Multiplikation einer Zeile mit einem Faktor α ungleich Null,
  (iii) Addition des Vielfachen einer Zeile auf eine andere,

auf obere Dreiecksform.

- Bei jeder Durchführung von Schritt (i) notiere einen Faktor mit (-1), bei jeder von Schritt (ii) notiere den Faktor α.

- Ist $\det(D)$ die Determinante der resultierenden Diagonalamatrix, die man mit 16.3.5 einfach abliest, so ist

$$\det(D) = (-1)^n \cdot \alpha_1 \cdot \ldots \cdot \alpha_m \cdot \det(A),$$

also

$$\det(A) = \frac{\det(D)}{(-1)^n \cdot \alpha_1 \cdot \ldots \cdot \alpha_m},$$

wobei n die Anzahl der im vorigen Schritt ermittelten Faktoren (-1) bei Durchführungen von Schritt (i) und $\alpha_1, \ldots, \alpha_m$ die im vorigen Schritt notierten Faktoren α bei Durchführungen von Schritt (ii) sind. (die Faktoren α_i, die bei den Umformungen entstehen, sind hierbei alle ungleich Null.

Die Korrektheit des angegebenen Algorithmus ergibt sich aus den Rechenregeln (D4') - (D6) aus Korollar 16.3.3. Ein Beispiel zur Anwendung behandeln wir in der Vorlesung.

Satz 16.3.6. *(Laplace'scher Entwicklungssatz)*

Sei $n \geq 2$ und $A = (x_1, \ldots, x_n) \in \mathbb{K}^{n \times n}$ (mit der schon in 16.2.5 benutzten Notation $x_j = (x_{i,j})$ für die Spaltenvektoren x_j. Dann gilt:

(i) für jeden Zeilenindex $i \in \underline{n}$, dass

$$\det(A) = \sum_{k=1}^{n} (-1)^{i+k} x_{i,k} \det\left(A_{i,k}\right) \quad (\text{„}\textit{Entwicklung nach der i-ten Zeile}\text{“})$$

ist, und

(ii) für jeden Spaltenindex $j \in \underline{n}$, dass

$$\det(A) = \sum_{k=1}^{n} (-1)^{k+j} x_{k,j} \det\left(A_{k,j}\right) \quad (\text{„}\textit{Entwicklung nach der j-ten Spalte}\text{“})$$

ist.

Dabei sind die Matrizen A_{ij} die Untermatrizen aus Def. 16.2.4, die sich aus der Matrix A jeweils nach Streichung der i-ten Zeile und der j-ten Spalte ergeben.

Beispiel dazu: S. Vorlesung.

Beweis: Ist $i \in \underline{n}$ ein fester Zeilenindex, so lässt sich diese unter Beibehaltung der Reihenfolge der übrigen Zeilen ganz nach oben tauschen, indem man die Zeile in $i-1$ Schritten jeweils gegen die Zeile darüber tauscht. Bezeichne A' die resultierende Matrix, in der die i-te Zeile $(x_{i,1}, \ldots, x_{i,n})$ ganz oben steht. Dann gilt mit (D6) $\det(A) := (-1)^{i-1} \det(A')$ und daher mit Entwicklung von A' nach der ersten Zeile gemäß (16.4)

$$\begin{aligned}
\det(A) &= (-1)^{i-1} \cdot \sum_{k=1}^{n} (-1)^{k-1} \cdot x_{i,k} \cdot \det_{n-1}(A_{i,k}) \\
&= \sum_{k=1}^{n} \underbrace{(-1)^{i+k-2}}_{=(-1)^{i+k}} \cdot x_{i,k} \cdot \det_{n-1}(A_{i,k}),
\end{aligned}$$

das zeigt die erste Behauptung. Die zweite ergibt sich aus dem Zusammenhang $\det(A^T) = \det(A)$ zusammen mit den Zusammenhängen $A^T = (x_{j,i})$ und $(A^T)_{j,k} = (A_{k,j})^T$: Es ist für einen Spaltenindex $j \in \underline{n}$

$$\begin{aligned}
\det(A) &= \det(A^T) \overset{(i) \text{ für Zeile } j}{=} \sum_{k=1}^{n} (-1)^{j+k} x_{k,j} \det\left((A^T)_{j,k}\right) \\
&= \sum_{k=1}^{n} (-1)^{j+k} x_{k,j} \det\left((A_{j,k})^T\right) \\
&= \sum_{k=1}^{n} (-1)^{j+k} x_{k,j} \det\left(A_{k,j}\right),
\end{aligned}$$

das war die behauptete Formel aus (ii).

□

Satz 16.3.7. *(Determinantenmultiplikationssatz)*

Für $A, B \in \mathbb{K}^{n \times n}$ ist

$$\det(A \cdot B) = \det(A) \cdot \det(B). \tag{16.6}$$

Beweis: Ist $\operatorname{rg}(A) < n$, so ist $\det(A) = 0$ nach (D2), und $\det(A \cdot B) = 0$ zu zeigen. Das wiederum folgt mit (D2) aus $\operatorname{rg}(A \cdot B) = 0$, was wir beweisen. Wegen $\operatorname{rg}(A) < n$ ist

$\dim \operatorname{Bild} f_A < n$ für die zu A gehörige Matrixabbildung. Bezeichnet $b_1, \ldots, b_n$ die Spalten der Matrix B, so sind die Spalten der Matrix $A \cdot B$ gerade die Vektoren $A \cdot b_1, \ldots, A \cdot b_n$, die allesamt Elemente des Bildes von f_A sind. Daher gilt

$$\operatorname{rg}(A \cdot B) = \dim \mathcal{L}(A \cdot b_1, \ldots, A \cdot b_n) \leq \dim \operatorname{Bild} f_A < n,$$

womit die Behauptung folgt. Ein ganz ähnliches Argument zeigt auch, dass im Fall $\operatorname{rg}(B) < n$ schon $\det(A \cdot B) < n$ folgt.

Wir müssen den Multiplikationssatz also nur noch für den Fall $\operatorname{rg}(A) = n$ zeigen und können auch auf das obige Resultat für den zweiten Faktor B zurückgreifen; beides nutzen wir aus. Sei also $A \in \mathbb{K}^{n \times n}$ mit $\operatorname{rg}(A) = n$. Dann ist nach 16.3.1 $\det(A) \neq 0$. Wir zeigen nun, dass die Abbildung

$$\delta : \mathbb{K}^{n \times n} \longrightarrow \mathbb{K}, \quad B \longmapsto \frac{\det(A \cdot B)}{\det(A)},$$

die wegen $\det(A) \neq 0$ wohldefiniert ist, die Eigenschaften (D1), (D2) und (D3) erfüllt. Mit der Eindeutigkeit der Determinantenabbildung folgt dann

$$\det(B) = \delta(B) = \det(A \cdot B) \cdot \det(A)^{-1},$$

und damit die Behauptung.

(D1): Es ist $\delta(I_n) = \det(A \cdot I_n) \cdot \det(A)^{-1} = 1$.
(D2): Ist $\operatorname{rg}(B) < n$, so ist nach den Vorüberlegungen

$$\delta(B) = \det(A \cdot B) \cdot \det(A)^{-1} = 0 \cdot \det(A)^{-1} = 0.$$

Zu (D3) sei i ein fester Spaltenindex, für den die Linearität gezeigt werden soll. Mit der schon im Existenzbeweis benutzten Notation

$$B_v = (x_1, \ldots, \underbrace{v}_{i\text{-te Spalte}}, \ldots, x_n)$$

gilt:

$$A \cdot B_{s \cdot x_i + y_i} = (A \cdot x_1, \ldots, \underbrace{A \cdot (s \cdot x_i + y_i)}_{\text{Spalte } i}, \ldots, A \cdot x_n) = (A \cdot x_1, \ldots, s \cdot A \cdot x_i + A \cdot y_i, \ldots, A \cdot x_n).$$

(Die letzte Gleichheit gilt, da A als Matrixabbildung linear ist.) Daher ist

$$\begin{aligned}
\det(A \cdot B_{s \cdot x_i + y_i}) &= \det B \cdot x_1, \ldots, s \cdot A \cdot x_i + A \cdot y_i, \ldots, A \cdot x_n) \\
&\overset{(D3) \text{ für } \det}{=} s \cdot \det(A \cdot x_1, \ldots, A \cdot x_i, \ldots, A \cdot x_n) + \det(A \cdot x_1, \ldots, A \cdot y_i, \ldots, A \cdot x_n) \\
&= s \cdot \det(A \cdot B_{x_i}) + \det(A \cdot B_{y_i})
\end{aligned}$$

und damit

$$\delta(B_{s \cdot x_i + y_i}) = s \cdot \delta(B_{x_i}) + \delta(B_{y_i});$$

das zeigt, dass δ auch $(D3)$ erfüllt und somit $\delta = \det$ gilt, also folgt mit der Begründung oben die Behauptung.

□

Korollar 16.3.8. *(Folgerungen aus dem Determinatenmultiplikationssatz)*
Es gelten die folgenden Rechenregeln und Zusammenhänge für Determinanten:

(i) *Ist* $A \in \mathbb{K}^{n \times n}$ *invertierbar, so ist*

$$\det(A^{-1}) = \frac{1}{\det(A)},$$

(ii) *Sind* $A, A' \in \mathbb{K}^{n \times n}$ *ähnlich (d.h. es gibt eine invertierbare Matrix* $C \in \mathbb{K}^{n \times n}$ *mit* $A' = C \cdot A \cdot C^{-1}$*), so gilt*

$$\det(A) = \det(A'),$$

(iii) *Ist* U *eine orthogonale Matrix, so gilt* $\det(U) = \pm 1$.

(iv) *Sind* $A, B \in \mathbb{K}^{n \times n}$*, so ist* $\det(A \cdot B) = \det(B \cdot A)$.

Beweis: Alles folgt ziemlich direkt aus dem Determinantenmultiplikationssatz, angewandt auf das jeweils richtige Matrizenprodukt.

□

Der Vollständigkeit halber (und um mal wieder den Namen Leibniz fallenzulassen) beschließen wir diesen Abschnitt zur Berechnung von Determinanten mit der so genannten Leibnizformel für Determinanten, über die historisch gesehen die Determinante einer $(n \times n)-$Matrix zum ersten Mal definiert wurde. Die Definition ist zu der unseren (per Laplace-Entwicklung nach der ersten Zeile) äquivalent. Ist $A \in \mathbb{K}^{n \times n}$, so gilt

$$\det A = \sum_{\sigma \in S_n} \left(\operatorname{sgn}(\sigma) \cdot a_{1,\sigma(1)} \cdot a_{2,\sigma(2)} \cdot \ldots \cdot a_{n,\sigma(n)} \right).$$

Hierbei wird über alle möglichen Permutationen σ von n Elementen (also über alle Elemente der symmetrischen Gruppe S_n) summiert. $\operatorname{sgn}(\sigma)$ bezeichnet das so genannte *Signum* der Permutation: Eine Permutation hat Signum $\operatorname{sgn}(\sigma) = 1$, falls sich σ als Hintereinanderausführung einer *geraden* Anzahl von Vertauschungen schreiben lässt, ist σ Verkettung einer *ungeraden* Anzahl von Vertauschungen, so ist $\operatorname{sgn}(\sigma) = -1$. (Man zeigt dann auch, dass immer genau eines von beiden geht.)

Wir entschlüsseln die Formel in der Vorlesung, indem wir uns ansehen, was das für die Determinanten $\det_2$ und $\det_3$ bedeutet. Ansonsten wird die Leibnizsche Determinantenformel eher für theoretische Zwecke benutzt – einige der obigen Sätze lassen sich damit recht elegant beweisen. Ich hoffe, wir haben es auch ohne relativ elegant geschafft.

Wir wenden uns nun noch einmal verschiedenen Deutungen der „Kennzahl“ Determinante zu, die wir in den vorangegangenen Abschnitten als Kennzahl von n Vektoren des $\mathbb{K}^n$ bzw. von quadratischen Matrizen kennengelernt haben. Insbesondere werden wir auch definieren, was wir unter der Determinante einer linearen Abbildung $f : V \to V$ (also einem so genannten Endomorphismus von V) verstehen wollen; diese werden im Kapitel zu den Eigenwerten von Endomorphismen im kommenden Kapitel noch wichtig werden.

16.4 Die Determinante als Maßzahl für Vektoren, Matrizen und lineare Abbildungen $f : V \to V$

Wir haben am Anfang dieses Kapitels in den Fällen $V = \mathbb{R}^2$ und $V = \mathbb{R}^3$ gesehen, dass sich die Determinante als orientierte Flächen- bzw. Volumenfunktion deuten lässt. Wir nutzen nun unsere Ergebnisse zu Determinanten im $\mathbb{K}^n$, um diese Begrifflichkeiten auf andere Situationen zu verallgemeinern.

16.4.1 Verallgemeinerte Volumina und Orientierungen in $\mathbb{R}^n$

Die Determinante als verallgemeinertes Volumenmaß in $\mathbb{R}^n$:

Seien $x_1, \dots, x_n \in \mathbb{R}^n$ Vektoren. Dann heißt die Menge

$$P(x_1, \dots, x_n) := \{a_1 x_1 + \dots + a_n x_n \mid 0 \leq a_i \leq 1 \text{ für } i \in \underline{n}\} \subseteq \mathbb{R}^n$$

das von $x_1, \dots, x_n$ aufgespannte Parallelepiped in $\mathbb{R}^n$; in $\mathbb{R}^2$ ist es gerade das von zwei Vektoren aufgespannte Parallelogramm, im $\mathbb{R}^3$ gerade der von drei Vektoren erzeugte Spat. Wir definieren das Volumen von $P(x_1, \dots, x_n)$ als den Absolutbetrag

$$V(P(x_1, \dots, x_n)) \quad := \quad |\det(x_1, \dots, x_n)|.$$

Inbesondere ist damit das Volumen des n-dimensionalen „Einheitswürfels“ $P(e_1, \dots, e_n)$ gerade $\det(I_n) = 1$.

Die Determinante als Orientierungsmaß in $\mathbb{R}^n$:

Im $\mathbb{R}^2$ und $\mathbb{R}^3$ lässt sich die Standardbasis I_2 bzw. I_3 anschaulich als die Koordinatenvektoren eines fest gewählten „Standard“-Koordinatensystems „bezüglich sich selbst“ deuten. Ist B eine weitere Basis von $\mathbb{R}^2$ bzw. $\mathbb{R}^3$, so läst sich B als die Koordinaten eines zweiten

Koordinatensystems deuten (wobei die Vektoren aus B gerade die Koordinaten der Basisvektoren des zweiten Koordinatensystems bezüglich des „Standard“-Koordinatensystems sind). Man stellt fest: Es ist $\det(B) > 0$ genau dann, wenn die Orientierung (oder „Händigkeit“) des „Welt“-Koordinatensystems mit der des durch B gegebenen Koordinatensystems übereinstimmt. Da Standardkoodinatensysteme in 3D meistens rechtshändig sind (d.h. die $x-, y-$ und $z-$Achse entsprechen in dieser Reihenfolge Daumen, Zeigefinger und Mittelfinger der rechten Hand), nennt man Basen B mit $\det(B) > 0$ manchmal auch rechtshändig. Allgemein definieren wir für eine Basis B des $\mathbb{R}^n$:

Definition 16.4.1. *(Orientierung von Basen des $\mathbb{R}^n$)*

Ist $B = (x_1, \ldots, x_n)$ eine Basis des $\mathbb{R}^n$, so gilt nach 16.3.1 $\det(x_1, \ldots, x_n) \neq 0$, also $\det(B) > 0$ oder $\det(B) < 0$. Im Fall $\det(B) > 0$ heißt B positiv orientiert, im Fall $\det(B) > 0$ nennen wir B negativ orientiert.

16.4.2 Determinanten von Matrixabbildungen als orientiertes Verzerrungsmaß

Sei $f : \mathbb{R}^n \to \mathbb{R}^n, x \to A \cdot x$ eine lineare Abbildung des $\mathbb{R}^n$ in sich selbst. Ist $(x_1, \ldots, x_n)$ ein beliebiges n-Tupel von Vektoren des $\mathbb{R}^n$ und $X \in \mathbb{R}^{n \times n}$ wieder die zugehörige kanonische Matrix, so gilt nach dem Determinatenmultiplikationssatz

$$\det(A) = \frac{\det(A \cdot X)}{\det(X)} = \frac{\det(Ax_1, \ldots, Ax_n)}{\det(x_1, \ldots, x_n)}.$$

Insbesondere ist

$$|\det(A)| = \frac{V(P(Ax_1, \ldots, Ax_n))}{V(P(x_1, \ldots, x_n)))}$$

für *alle* Parallelepipede in $\mathbb{R}^n$, weshalb man den Betrag der Determinante als einheitliches Maß für die durch die lineare Abbildung A bewirkte *Volumenverzerrung* interpretieren kann: Volumina von Parallelepipeden vergrößern (bzw. verkleinern) sich durch Anwendung von A genau um den Faktor $|\det(A)|$. Insbesondere sind orthogonale Matrixabbildungen, also Abbildungen $f : \mathbb{R}^n \to \mathbb{R}^n, x \to U \cdot x$ mit orthogonaler Matrix U, nach 16.3.8(ii) volumenerhaltend.

Das *Vorzeichen* der Determinante einer Matrixabbildung gibt mit der Deutung aus Abschnitt 16.4.1 analog an, ob die Matrixabbildung A die Orientierung von Basen ändert. Wieder tut sie das für alle Basen gleich, d.h. alle positiv orientierten Basen werden auf negativ orientierte Basen abgebildet und umgekehrt (falls $\det(A) < 0$), oder sie erhält alle Orientierungen von Basen (falls $\det(A) > 0$). Für orthogonale Matrixabbildungen liefert die Klassifizierung in orthogonale Abbildungen mit Determinante $+1$ bzw. -1 gerade die in Lamma 15.1.7 und für den allgemeineren Fall danach erwähnte Aufteilung der orthogonalen Abbildungen des $\mathbb{K}^n$ in Drehungen (das sind die mit $\det(U) = 1$) und Drehspiegelungen (mit $\det(U) = -1$).

16.4.3 Die Determinante von linearen Abbildungen $f : V \to V$

Mit Hilfe darstellender Matrizen lässt sich der Begriff der Determinante auch für lineare Abbildungen $f : V \to V$ von Vektorräumen V in sich selbst (also so genannte *Endomorphismen*, dieser Begriff wird jetzt häufiger fallen) definieren, falls V endlichdimensional ist. Diese Verallgemeinerung soll unsere Untersuchungen zu Determinanten abschließen und wird im kommenden Kapitel bei der Untersuchung der Eigenwerte von Endomorphismen nützlich sein.

Definition 16.4.2. *(Determinante von Endomorphismen)*

Sei V ein endlichdimensionaler Vektorraum und $f : V \to V$ linear . Ist $B = (b_1, \dots, b_n)$ eine Basis von V und $A^f_{B,B}$ die darstellende Matrix von f bezüglich B, so definiert man die Determinante von f als

$$\det(f) \quad := \quad \det(A^f_{B,B}).$$

Bemerkungen dazu:

- Die obige Definition ist von der konkreten Auswahl der Basis B unabhängig, da
  - darstellende Matrizen derselben linearen Abbildung zu verschiedenen Basen B, B' ähnlich sind (mit der Basistransformationsmatrix $T_{B,B'}$ als Übergangsmatrix, es ist $A^f_{B',B'} = T_{B,B'} A^f_{B,B} T^{-1}_{B,B'}$, vgl. auch Definition 10.2.2 und die zugehörigen Bemerkungen), und
  - ähnliche Matrizen nach Lemma 16.3.8(ii) dieselbe Determinante besitzen.
- Für Matrixabbildungen $f : \mathbb{K}^n \to \mathbb{K}^n, \quad x \mapsto A \cdot x$ ist $\det(f) = \det(A)$.
- Viele Eigenschaften der Determinante von Matrizen übertragen sich so auf die entsprechenden linearen Abbildungen: Ist $\dim V < \infty$, so gilt u.a.:

- Sind $f, g : V \to V$ Endomorphismen, so überträgt sich durch Rückführung auf geeignete darstellende Matrizen von f und g der Determinantenmultiplikationssatz als
$$\det(f \circ g) = \det(f) \cdot \det(g).$$
- Für $\dim V < \infty$ und eine orthogonale Abbildung $f : V \to V$ ist $\det(f) = \pm 1$,
- f ist ein Isomorphimus genau dann, wenn $\det(f) \neq 0$ ist
- $[\ldots]$

16.5 Aufgaben zu Kapitel 16

Aufgabe 16.91. (Eigenschaften des Kreuzproduktes):

In der Vorlesung wurde das *Kreuzprodukt* oder *Vektorprodukt* $\times\colon \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ eingeführt, das den Vektoren $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, $y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \in \mathbb{R}^3$ den Vektor

$$x \times y := \begin{pmatrix} x_2y_3 - x_3y_2 \\ x_3y_1 - x_1y_3 \\ x_1y_2 - x_2y_1 \end{pmatrix} \in \mathbb{R}^3$$

zuordnet.

(a) Zeigen Sie, dass das Kreuzprodukt eine bilineare antikommutative Abbildung ist. Hierbei bedeutet *bilinear*, dass für alle Skalare $a, b \in \mathbb{R}$ und alle $x, x', y, y' \in \mathbb{R}^3$ die Gleichheiten

$$(ax + bx') \times y = a(x \times y) + b(x' \times y) \quad \text{und} \quad x \times (ay + by') = a(x \times y) + b(x \times y')$$

gelten, und *antikommutativ*, dass die Beziehung

$$y \times x = -(x \times y)$$

für alle $x, y \in \mathbb{R}^3$ besteht. (Aus der Antikommutativität folgt insbesondere $x \times x = 0_V$.)

(b) Zeigen Sie, dass für alle $x, y \in \mathbb{R}^3$ bezüglich des Standardskalarprodukts die Beziehungen $(x \times y) \perp x$ und $(x \times y) \perp y$ gelten.

(c) Beweisen Sie für alle $x, y \in \mathbb{R}^3$ die Äquivalenz

$$x \times y = 0_V \iff x, y \text{ sind linear abhängig.}$$

(d) Überprüfen Sie, ob das Assoziativgesetz

$$x \times (y \times z) = (x \times y) \times z$$

für alle $x, y, z \in \mathbb{R}^3$ erfüllt ist.

Aufgabe 16.92. (Eigenschaften von Determinanten):

Gegeben seien die folgenden vier reellen Matrizen:

$$A := \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{pmatrix}, \quad B := \begin{pmatrix} 2 & 3 & 5 \\ 3 & 4 & 6 \\ 1 & 1 & -1 \end{pmatrix}, \quad C := \begin{pmatrix} 0 & 0 & a & 0 \\ 0 & 0 & 0 & b \\ 0 & c & 0 & 0 \\ d & 0 & 0 & 0 \end{pmatrix}, \quad D := \begin{pmatrix} 2 & 0 & 0 & 0 & 2 \\ 0 & 2 & 0 & 2 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 0 & 2 \end{pmatrix}.$$

(a) Bestimmen Sie die Determinanten dieser Matrizen, indem Sie ausschließlich die Eigenschaften (D1) bis (D6) der Determinantenfunktion det des $\mathbb{K}^n$ verwenden. Begründen Sie jedes Gleichheitszeichen, indem Sie die entsprechende Eigenschaft angeben.

(b) Welche Bedingungen müssen die Einträge $a, b, c, d \in \mathbb{R}$ der Matrix C erfüllen, damit das Gleichungssystem $Cx = y$ für beliebiges $y \in \mathbb{R}^4$ stets eindeutig lösbar ist?

Aufgabe 16.93. (Berechnung von Determinanten):

(a) Berechnen Sie die Determinanten folgender reeller Matrizen, indem Sie sie per elementarer Zeilenumformungen auf obere Dreiecksform bringen.

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

(b) Berechnen Sie die Determinanten der folgenden reellen Matrizen mit der Methode Ihrer Wahl (ohne Technikeinsatz :)):

$$C := \begin{pmatrix} 1 & 2 \\ 2 & 8 \end{pmatrix}, \quad D := \begin{pmatrix} 1 & 1 & 0 \\ 2 & 3 & 1 \\ 1 & 1 & 2 \end{pmatrix}, \quad E = \begin{pmatrix} a_1 & 0 & b_1 & 0 \\ 0 & c_1 & 0 & d_1 \\ a_2 & 0 & b_2 & 0 \\ 0 & c_2 & 0 & d_2 \end{pmatrix}, \quad F := \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 5 & 6 & 7 \\ 3 & 4 & 7 & 8 & 9 \\ 4 & 5 & 9 & 0 & 1 \\ 5 & 4 & 9 & 0 & 1 \end{pmatrix}.$$

Aufgabe 16.94. (Invertierbarkeit einer Matrix mit Parameter):

Bestimmen Sie alle Werte $a \in \mathbb{C}$, für die die Matrix

$$\begin{pmatrix} a & -2 & 0 \\ -1 & 3 & a-1 \\ -1 & a & -1 \end{pmatrix} \in \mathbb{C}^{3\times 3}$$

nicht invertierbar ist. Nutzen Sie dafür die Determinante der gegebenen Matrix.

Aufgabe 16.95. (Laplace'scher Entwicklungssatz):

Berechnen Sie die Determinanten der folgenden Matrizen, indem Sie sie mit dem Laplace'schen Entwicklungssatz auf Determinanten von (3×3)- oder (2×2)-Matrizen zurückführen:

$$A := \begin{pmatrix} 1 & 0 & 0 & 7 \\ 2 & 0 & i & 0 \\ 0 & 3 & 6 & 0 \\ 0 & 4 & 0 & 8 \end{pmatrix} \in \mathbb{C}^{4\times 4}, \qquad B := \begin{pmatrix} 1 & 0 & -3 & 0 & 9 \\ 3 & 7 & 10 & 3 & 17 \\ 4 & 0 & 11 & 0 & 1 \\ 6 & 0 & 8 & 0 & -3 \\ 5 & 1 & 6 & -1 & 8 \end{pmatrix} \in \mathbb{R}^{5\times 5}.$$

Aufgabe 16.96. (Rechenregeln von Determinanten):

Beweisen Sie Korollar XIII.3.8, d. h., zeigen Sie, dass folgende Rechenregeln und Zusammenhänge für Determinanten gelten:

(i) Ist $A \in \mathbb{K}^{n\times n}$ invertierbar, so gilt $\det(A^{-1}) = \dfrac{1}{\det(A)}$.

(ii) Sind $A, A' \in \mathbb{K}^{n\times n}$ ähnlich, so gilt $\det(A) = \det(A')$.

Zur Erinnerung: Zwei Matrizen $A, A' \in \mathbb{K}^{n\times n}$ sind ähnlich, wenn eine Matrix $C \in GL_n(\mathbb{K})$ existiert, sodass $A' = C \cdot A \cdot C^{-1}$ gilt.

(iii) Ist $U \in \mathbb{R}^{n\times n}$ eine orthogonale Matrix, so gilt $\det(U) = \pm 1$.

(iv) Sind $A, B \in \mathbb{K}^{n\times n}$, so ist $\det(A \cdot B) = \det(B \cdot A)$.

17 Eigenwerte

17.1 Grundbegriffe und die Hauptaufgabe der Eigenwertberechnung

Mit der Eigenwerttheorie beschäftigen wir uns noch einmal mit einem unverzichtbaren Teilbereich der Linearen Algebra, der uns erlauben wird, lineare Abbildungen gewissermaßen „in ihre Einzelteile“ zu zerlegen und so für eine gegebene Abbildung (z.B. anhand der Matrix einer Matrixabbildung) zu erkennen, was diese Abbildung eigentlich „tut“. Entscheidend dafür sind für Selbstabbildungen $f : V \to V$ die so genannten Eigenwerte und Eigenräume von f und die daraus in vielen wichtigen Fällen gewinnbare Eigenwertzerlegung von f. Wenn das nicht geht, nud auch im allgemeineren Fall $f : V \to W$, hilft die so genannte Singulärwertzerlegung. Wir werden zunächst ausschließlich lineare *Selbstabbildungen* $f : V \to V$ (also lineare Abbildungen eines $\mathbb{K}$-Vektorraumes V in sich, so genannte Endomorphismen von V) betrachten. Die Menge solcher Abbildungen hatten wir schon früher als $L(V)$ abkürzt und werden das weiter tun.

Definition 17.1.1. *(Eigenwerte und Eigenvektoren; Spektrum)*

Sei V ein $\mathbb{K}$-Vektorraum und $f \in L(V)$. Dann heißt ein Vektor $v \in V$ mit $v \neq 0_V$ <u>Eigenvektor von f</u>, falls es ein $\lambda \in \mathbb{K}$ gibt, so dass

$$f(v) = \lambda \cdot v$$

ist. Den Wert $\lambda \in \mathbb{K}$ nennt man dann einen <u>Eigenwert von f</u>, das Pärchen (λ, v) auch oft <u>Eigenpaar</u> von f.

Speziell für Matrizen $A \in \mathbb{K}^{n\times n}$ definiert man, dass $\lambda \in \mathbb{K}$ Eigenwert von A heißt, falls es ein $v \neq 0_V$ in $\mathbb{K}^n$ gibt, so dass

$$A \cdot v = \lambda \cdot v$$

ist. Entsprechend heißt auch hier ein solcher Vektor v Eigenvektor zum Eigenwert λ.

Die Menge aller Eigenwerte von f bzw. A ist eine Teilmenge von $\mathbb{K}$, die man als das <u>Spektrum</u> von f bzw. A bezeichnet; wir schreiben dafür $\sigma(f)$ bzw. $\sigma(A)$.

Bemerkung dazu:

- Ist $f_A : \mathbb{K}^n \to \mathbb{K}^n, x \mapsto A \cdot x$, eine Matrixabbildung, so sind die Eigenwerte/-vektoren/-paare von f_A (als lineare Abbildung) genau die Eigenwerte /-räume/-vektoren der Abbildungsmatrix A.

Beispiele:

- Wir untersuchen in der Vorlesung die Eigenwerte und Eigenräume von Spiegelungen, Streckungen, insbesondere die Streck- und Spiegelachsen und die zu ihnen orthogonalen Richtungen in $\mathbb{R}^2$ und $\mathbb{R}^3$.
- Ist $V = C^\infty(\mathbb{R})$ der Vektorraum der unendlich oft differenzierbaren Funktionen $g : \mathbb{R} \to \mathbb{R}$ und
$$\frac{d}{dx} : \quad V \to V, \quad g \mapsto g'$$
der Ableitungsoperator auf V, so ist $\frac{d}{dx}$ linear und die obige Definition auch auf diese Situation anwendbar. In diesem Fall ist jedes beliebige $\lambda \in \mathbb{R}$ Eigenwert von $\frac{d}{dx}$, und die Funktionen $g(x) = e^{\lambda x}$ sind zugehörige Eigenvektoren von $\frac{d}{dx}$, denn es ist ja
$$\frac{d}{dx}(g) = \frac{d}{dx}(e^{\lambda x}) = \lambda \cdot e^{\lambda x} = \lambda \cdot g(x).$$
Ebenso sind die Sinus- und Cosinusfunktion auf $\mathbb{R}$ Eigenfunktionen des Zweifachen Ableitens $\frac{d^2}{d^2x} : V \to V, g \mapsto g''$ – finden Sie die zugehörigen Eigenwerte?

Wir wollen die Struktur von Mengen von Eigenvektoren untersuchen: Für $v \in V$ und $\lambda \in \mathbb{K}$ gilt

$$f(v) = \lambda \cdot v \quad \Leftrightarrow \quad f(v) - \lambda \cdot \mathrm{id}_V(v) = 0_V \quad \Leftrightarrow \quad v \in \mathrm{Kern}(f - \lambda \cdot \mathrm{id}_V).$$

d.h. λ ist ein Eigenwert von f genau dann, wenn $\mathrm{Kern}(f - \lambda \cdot \mathrm{id}_V) \neq \{0_V\}$ ist; für Matrizen $A \in \mathbb{K}^{n \times n}$ gilt analog, dass

$$A \cdot v = \lambda \cdot v \Leftrightarrow \quad A \cdot v - \lambda \cdot I_n \cdot v = 0_V \quad \Leftrightarrow \quad v \in \mathrm{Kern}(A - \lambda \cdot I_n),$$

d.h. λ ist ein Eigenwert von A genau dann, wenn $\mathrm{rg}(A - \lambda \cdot I_n) < n$ gilt.
Der Kern einer linearen Abbildung f bzw. die Lösungsmenge eines homogenen linearen Gleichungssystems ist immer ein Teilraum des Wertevorats V; das führt zu folgender Definition.

Definition 17.1.2. *(Eigenraum zu Eigenwerten)*

Sei $f \in L(V)$ und $\lambda \in \mathbb{K}$. Der zu λ und f gehörige Eigenraum, den wir als $E(\lambda, f)$ abkürzen, ist definiert als

$$E(\lambda, f) := \operatorname{Kern}(f - \lambda \operatorname{id}_V).$$

In anderen Worten: $E(\lambda, f)$ ist die Menge aller zum Eigenwert λ gehörigen Eigenvektoren von f, zusammen mit dem Nullvektor 0_V.

Analog definieren wir für $A \in \mathbb{K}^{n \times n}$ und $\lambda \in \mathbb{K}$ den zu λ und A gehörigen Eigenraum als

$$E(\lambda, A) := \operatorname{Kern}(A - \lambda I_n).$$

Wir halten dazu fest:

Lemma 17.1.3. *(Eigenräume sind Untervektorräume)*

Für $\lambda \in \mathbb{K}$, und $f \in L(V)$ ist der Eigenraum $E(\lambda, f)$ ein Teilraum von V. Genauso ist für eine Matrix $A \in \mathbb{K}^{n \times n}$ der Eigenraum $E(\lambda, A)$ ein Teilraum von $\mathbb{K}^n$.

Beweis: Falls Sie das überlesen haben sollten: Der steht für beide Fälle schon direkt über der Definition.

$\square$

Bemerkungen dazu:

- $E(\lambda, f)$ ist in diesem Fall gerade die Lösungsmenge des homogenen linearen Gleichungssystems $(A - \lambda I_n) \cdot x = 0$.
- Insbesondere besitzt $E(\lambda, f)$ immer eine Basis, die aus Eigenvektoren zum Eigenwert λ besteht, und die die Menge aller Eigenvektoren zu λ erzeugt. Diese kann, wenn ein Skalarprodukt auf V gegeben ist, nach Satz 14.4.2 sogar immer als *Orthonormalbasis* von $E(\lambda, f)$ gewählt werden.
- Insbesondere gilt in Skalarprodukträumen auch: Ist λ ein Eigenwert von f bzw. A, so gibt zu λ immer mindestens einen zugehörigen *normierten* Eigenvektor.

Beispiele:

- Der Eigenraum $E(0, f)$ zu $\lambda = 0$ ist gerade der Kern von f, eine lineare Abbildung f besitzt also den Eigenwert 0, wenn $\mathrm{Kern}(f) \neq \{0_V\}$ gilt;
- Der Eigenraum $E(1, f)$ zum Eigenwert $\lambda = 1$ gerade der so genannte Fixraum von f,
$$\mathrm{Fix}(f) := \{v \in V \mid f(v) = v\}.$$
Eine lineare Abbildung f hat den Eigenwert 1 genau dann, wenn es einen Fixvektor $v \neq 0_V$ gibt, also einen Vektor $v \neq 0_V$ mit $f(v) = v$, also wenn $\mathrm{Fix}(f) \neq \{0_V\}$ gilt.

Die Hauptaufgabe der Beschäftigung mit Eigenwerttheorie, um die es im nächsten Abschnitt gehen soll, lässt sich nun von ihrer praktischen Seite folgendermaßen fassen:

Hauptaufgabe der Eigenwertpraxis:

Für eine gegebene lineare Abbildung $f : V \to V$ oder eine gegebene Matrix $A \in \mathbb{K}^{n \times n}$ bestimme

(a) alle ihre Eigenwerte (d.h. $\sigma(f)$ bzw. $\sigma(A) \subseteq \mathbb{K}$) und

(b) die jeweiligen zugehörigen Eigenräume, was in der Praxis meist bedeutet: Bestimme zu jedem Eigenwert λ aus (a) eine Basis des zugehörigen Eigenraumes $E(\lambda, f) \subseteq V$ bzw. $E(\lambda, A) \subseteq \mathbb{K}^n$.

Die obige Hauptaufgabe ist recht „numerisch-praktisch“ formuliert; ihre Lösung dient z.B. Ingenieuren dazu, Eigenschwingungen von Materialien jeglicher Art (von Brücken bis Musikinstrumenten) vorauszuberechnen; in der Physik und Chemie liefern die Lösungen von Eigenwertproblemen zum Beispiel die Energien und Zustände von Atomen und Molekülen im Ruhezustand.

Wir werden im Folgenden ein Vorgehen zur Berechnung entwickeln, aber hauptsächlich, wie es die Natur einer „reinen“ Mathematikvorlesung ist, diese Aufgabe an sich klarer fassen und diverse theoretische Aussagen dazu machen, was wir als Ergebnis dieser Aufgabe zu erwarten haben.

Lemma 17.1.4. *(Lineare Unabhängigkeit von Eigenvektoren)*

Sei $f \in L(V)$, und $\lambda_1, \dots, \lambda_k$ seien k verschiedene Eigenwerte von f. Sind $v_1, \dots, v_k$ jeweils zugehörige Eigenvektoren, so sind $v_1, \dots, v_k$ linear unabhängig.

Beweis: Im Fall $k = 1$ ist nach Definition von Eigenvektoren $v_1 \neq 0_V$, also v_1 linear unabhängig. Für $k > 1$ zeigen wir die Kontraposition und starten mit einem linear abhängigen k-Tupel $(v_1, \ldots, v_k)$ von Eigenvektoren von f, für die wir zeigen, dass sie dann nicht Eigenvektoren zu verschiedenen Eigenwerten sein können. Sind die Vektoren linear abhängig, so können wir o.B.d.A. können wir die Vektoren so umnummerieren, dass ein Tupel $B = (v_1, \ldots, v_m)$ mit $m < k$ eine Basis von $\mathcal{L}(v_1, \ldots, v_k)$ ist. Insbesondere lässt sich der Vektor v_k als Linearkombination der Vektoren aus B schreiben, d.h. es gibt $a_1, \ldots, a_{k-1} \in \mathbb{K}$ so, dass

$$v_k = a_1 v_1 + \ldots + a_m v_m. \qquad (\star)$$

Nach Anwendung von f auf diese Gleichheit erhält man

$$\lambda_k v_k = a_1 \lambda_1 v_1 + \ldots + a_m \lambda_m v_m.$$

Multipliziert man beide Seiten von $(\star)$ mit λ_k und zieht von Ergebnis die letztere Gleichung ab, so ergibt sich

$$a_1 (\lambda_k - \lambda_1) v_1 + \cdots + a_m (\lambda_k - \lambda_m) v_m = 0_V.$$

Das Tupel $(v_1, \ldots, v_m)$ ist linear unabhängig. Daher müssen die Vorfaktoren $a_i(\lambda_k - \lambda_i)$ für alle $i \in \underline{m}$ Null sein. Da v_k ein Eigenvektor ist, ist $v_k \neq 0_V$, d.h. es kann in $(\star)$ nicht $a_i = 0$ für alle Koeffizienten a_i gelten. Das bedeutet aber (wegen der Nullteilerfreiheit in $\mathbb{K}$), dass $\lambda_k = \lambda_i$ für mindestens ein Indexpaar $k \neq i$ gilt, also v_i und v_k keine Eigenvektoren zu verschiedenen Eigenvektoren sind.

□

Korollar 17.1.5. *(Maximale Anzahl an verschiedenen Eigenwerten)*

Sei V endlichdimensional mit $\dim V = n$. Ist $f \in L(V)$, so hat f höchstens n verschiedene Eigenwerte. Insbesondere hat eine $(n \times n)$- Matrix höchstens n verschiedene Eigenwerte.

Beweis: Seien $\lambda_1, \ldots, \lambda_k$ verschiedene Eigenwerte von $f \in L(V)$ und $v_1, \ldots, v_k$ zugehörige Eigenvektoren. Dann ist $v_1, \ldots, v_k$ nach 17.1.4 linear unabhängig, d.h. $k \leq \dim V$.

□

Der obige Beweis benutzt explizit die Endlichdimensionalität von V. Im Unendlichdimensionalen sieht es tatsächlich anders aus: Hier kann man für geeignete Funktionen f sogar überabzählbar viele verschiedene Eigenwerte und zugehörige Eigenvektoren finden, vgl. das Beispiel $\frac{d}{dx}$ oben. (Diese sind dann nach 17.1.4 linear unabhängig, denn die Endlichdimensionalität wurde im Beweis von 17.1.4 nicht vorausgesetzt.)Etwas genauer als die Aussage von Korollar 17.1.5 ist das folgende Ergebnis, das noch sehr nützlich sein wird:

Lemma 17.1.6. *(Die Summe von Eigenräumen ist direkt)*

Sei V ein endlichdimensionaler Vektorraum, $f \in L(V)$ und $\lambda_1, \ldots, \lambda_k$ die Eigenwerte von f. Wir betrachten die Vektorraumsumme

$$\begin{aligned} \mathrm{Eig}(f) &:= E(\lambda_1, f) + \ldots + E(\lambda_k, f) \\ &= \{v \in V \mid v = v_1 + \ldots + v_k \text{ mit } v_i \in E(\lambda_i, f)\}, \end{aligned}$$

also den von allen Eigenvektoren von f aufgespannten Teilraum von V.

Dann gilt:

(i) $\mathrm{Eig}(f)$ *ist als Summe von Unterevektorräumen eine direkte Summe, d.h. es ist*

$$E(\lambda_i, f) \cap \left(\bigoplus_{j \neq i} E(\lambda_j, f) \right) = \{0_V\} \tag{17.1}$$

für alle $i \in \underline{k}$.[a]

(ii) Es ist

$$\dim E(\lambda_1, f) + \ldots + \dim E(\lambda_k, f) \quad \leq \quad \dim V.$$

(iii) Sind $B_1, \ldots, B_k$ Basen von $E(\lambda_1, f), \ldots, E(\lambda_k, f)$, so ist $B_1 \cup \ldots \cup B_k$ eine Basis von $\mathrm{Eig}(f)$.

(iv) Jeder Vektor $v \in \mathrm{Eig}(f)$ eine eindeutige Darstellung

$$v = u_1 + \ldots + u_k$$

mit $u_1 \in E(\lambda_1, f), \ldots, u_k \in E(\lambda_k, f)$.

[a] Man stellt fest, dass diese Bedingung die natürliche Verallgemeinerung der Bedingung $U \cap W = \{0_V\}$ für die direkte Summe zweier Vektorräume ist, die zu den hauptsächlich benutzten Folgerungen (iii) und (iv) führt. Insbesondere reicht dafür die Forderung $E(\lambda_i, f) \cap E(\lambda_j, f)$ nicht aus – sehen Sie, warum?

Beweis: Zu (i) beweisen wir die Kontraposition: Ist der angegebene Schnitt für ein $i \in \underline{k}$ nicht nur der Nullvektor, so existiert ein $u_i \in E(\lambda_i, f)$ mit $u_i \neq 0_V$, das in $\oplus_{j\neq i} E(\lambda_j, f)$ enthalten ist. Es gibt also für $j \neq i$ Vektoren $u_j \in E(\lambda_j, f)$ mit

$$u_i = \sum_{j\neq i} u_j,$$

u_i ist also Linearkombination der übrigen Vektoren und die Vektoren $u_1, \ldots, u_k$ sind also nicht linear unabhängig. Nach 17.1.4 können die Eigenwerte $\lambda_1, \ldots, \lambda_k$ dann nicht verschieden sein.

Zu (ii) wendet man $(n-1)$-mal den Dimensionssatz für Unterräume an, der wegen der Schnitteigenschaft 17.1

$$\dim \left(E(\lambda_1, f) + \cdots + \dim E(\lambda_k, f)\right) \quad = \quad \dim \operatorname{Eig}\big((\lambda_1, \ldots, \lambda_k), f\big)$$

gibt. Da $\operatorname{Eig}\big((\lambda_1, \ldots, \lambda_k), f\big)$ ein Teilraum von V ist, ist $\dim \operatorname{Eig}\big((\lambda_1, \ldots, \lambda_k), f\big) \leq \dim(V)$; das zeigt (ii).

Zu (iii) bemerken wir, dass wie im Fall der Summe von zwei Unterräumen auch für mehr Unterräume gilt: Sind $B_1, \ldots, B_k$ Basen von $E(\lambda_1, f), \ldots, E(\lambda_k, f)$, so ist $B_1 \cup \ldots \cup B_k$ ein Erzeugendensystem von $\operatorname{Eig}(f)$. Da aus (i) insbesondere $E(\lambda_i, f) \cap E(\lambda_j, f) = \{0_V\}$ für $i \neq j$ folgt, ist auch $B_i \cap B_j = \emptyset$ (denn keine Basis enthält den Nullvektor). Daher ist die Anzahl der Elemente in der Vereinigung der Basen B_i gerade

$$|B_1 \cup \ldots \cup B_k| = |B_1| + \ldots + |B_k| = \dim E(\lambda_1, f) + \cdots + \dim E(\lambda_k, f)\,.$$

Da das nach (ii) gerade $\dim \operatorname{Eig}(f), f\big)$ Elemente sind, ist $B_1 \cup \ldots \cup B_k$ nach Korollar 8.3.9 eine Basis von $\operatorname{Eig}(f)$.

Zu (iv): Um die Eindeutigkeit einer Darstellung $v \in \operatorname{Eig}(f)$ zu zeigen, betrachtet man die Differenz zweier potenziell verschiedener Darstellungen von v und erhält

$$0 = (u_1 - u_1') + \ldots + (u_k - u_k').$$

Für alle $i \in \underline{k}$ ist $u_i - u_i'$ jeweils ein Eigenvektor zum Eigenwert λ_i, oder es gilt $u_i - u_i' = 0_V$. Wäre $u_i - u_i' \neq 0_V$ für einen oder mehrere Indizes i, so wären diese Eigenvektoren nicht linear unabhängig im Widerspruch zu der Aussage von Lemma 17.1.4, also sind die Darstellungen gleich.

□

17.2 Eigenwerte und Eigenräume von Matrizen

Wir wenden uns nun der Frage zu, wie man Eigenwerte und Eigenvektoren von quadratischen Matrizen $A \in \mathbb{K}^{n \times n}$ bzw. Matrixabbildungen $f : \mathbb{K}^n \to \mathbb{K}^n, x \mapsto A \cdot x$ kann. Im folgenden Abschnitt werden wir diese Ergebnisse dann nutzen, um auch ein Verfahren für den allgemeineren Fall $f \in L(V)$ anzugeben.

Wir starten im folgenden Unterabschnitt mit einem neuen, zentralen Begriff.

17.2.1 Das charakteristische Polynom einer quadratischen Matrix

Wir formulieren dafür zunächst ein paar Zusammenhänge.

Lemma 17.2.1. *(Eigenwerte von Matrizen)*

Sei $A \in \mathbb{K}^{n \times n}$. Dann gilt: Bezeichnet $I := I_n$ die Einheitsmatrix des $\mathbb{K}^{n \times n}$, so sind sind äquivalent:

(i) $\lambda \in \mathbb{K}$ *ist ein Eigenwert von* A,

(ii) Die Gleichung $(A - \lambda \cdot I) \cdot x = 0$ *hat außer* $x = 0_V$ *noch weitere Lösungen* $x \in \mathbb{K}^n$ *(nämlich genau alle Eigenvektoren von* A *zu* λ*).*

(iii) Die Matrix $A - \lambda I$ *ist nicht invertierbar,*

(iv) Es ist $\operatorname{rg}(A - \lambda \cdot I) < n$,

(v) Es ist $\det(A - \lambda \cdot I) = 0$,

Beweis: Es ist $\lambda \in \mathbb{K}$ ein Eigenwert von A genau dann, wenn $\operatorname{Kern}(A - \lambda \cdot I_n) = \{0_V\}$ ist, vgl. die Überlegungen im Umfeld von 17.1.2. Die obigen Äquivalenzen folgen mit Hilfe der „großen Sätze“ der in der LAAG I entwickelten Theorie direkt aus diesem Zusammenhang. Kurze Erläuterung siehe Vorlesung.

$\square$

Bemerkungen dazu:

- Punkte (v) und (ii) werden wir konkret benutzen, um Eigenpaare auszurechnen. Punkt (ii) allein hilft allerdings nicht sehr viel, wenn λ nicht bekannt ist – denn das ist eine *nichtlineare* Gleichung für $n + 1$ Unbekannte $x_1, \ldots, x_n, \lambda \in \mathbb{K}$, für die ein Lösungsverfahren nicht wirklich offensichtlich ist.

- Wir nutzen stattdessen Punkte (v) und unsere guten alten Bekannten, die Determinanten aus dem letzten Kapitel, um zuerst die Eigenwerte einer Matrix zu bestimmen und dann mit *bekanntem* λ den Eigenraum $E(\lambda, A)$ zu berechnen, also die Lösungsmenge des homogenen LGS $(A - \lambda \cdot I_n) \cdot x = 0$, wobei die Matrix $A - \lambda \cdot I_n$ einfach durch Einsetzen des konkreten Wertes für λ aufgestellt werden kann.

Zur Berechnung von Eigenwerten ist der folgende Begriff bzw. der darauf folgende Satz die zentrale Beobachtung.

Lemma/Definition 17.2.2. *(Das charakteristische Polynom einer Matrix)*
Sei $A \in \mathbb{K}^{n \times n}$. Dann ist die Abbildung

$$p_A : \quad \mathbb{K} \to \mathbb{K}, \quad t \mapsto \det(A - t \cdot I)$$

eine Polynomfunktion. Das zugehörige Polynom p_A ist ein Polynom von Grad n mit Leitkoeffizient $a_n = (-1)^n$.

$p_A \in \mathbb{K}[x]$ nennt man das charakteristische Polynom der Matrix A.

Beweis: Berechnet man die Determinante der Matrix

$$A - t \cdot I = \begin{pmatrix} a_{11} - t & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - t & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} - t \end{pmatrix},$$

so ergibt sich aus der rekursiven Definition der Determinante (s. Vorlesung)

$$\det(A - t \cdot I) = (a_{11} - t) \cdot (a_{22} - t) \cdot \ldots \cdot (a_{nn} - t) + r_1, \qquad (\star)$$

wobei das vorne ein Polynom von Grad n und r_1 ein Polynom vom Grad kleiner n ist. Daraus ergibt sich wegen

$$(a_{11} - t) \cdot (a_{22} - t) \cdot \ldots \cdot (a_{nn} - t) \quad = \quad (-1)^n \cdot (t - a_{11}) \cdot (t - a_{22}) \ldots \cdot (t - a_{nn})$$

bei Umsortierung in die Standardform von Polynomen ein Polynom mit Leitkoeffizient $(-1)^n$. Das gilt damit auch für den obigen Ausdruck $(\star)$, da der Rest von Grad kleiner n ist, also die Behauptung zum Leitkoeffizient. $\square$

Der Beweis zur Aussage des folgenden Satzes sollte nun offensichtlich sein. Da er der zentrale Zusammenhang für die Berechnung von Eigenwerten ist, formulieren wir ihn dennoch nochmal deutlich.

Satz 17.2.3. *Sei $A \in \mathbb{K}^{n\times n}$. Die Nullstellen von p_A, die in $\mathbb{K}$ liegen, sind genau die Eigenwerte von $A \in \mathbb{K}^{n\times n}$.*

Beispiele:

1. Eigenwerte einer (3×3)-Matrix: S. Vorlesung oder [BeuLA], S. 250, Beispiele a)-c),
2. Eigenwerte der Drehmatrix

$$D = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix},$$

mit der sich Drehungen in der Ebene um den Winkel $\alpha \in \mathbb{R}$ beschreiben lassen. Das charakteristische Polynom p_D hat keine reellen Nullstellen, aber zwei in $\mathbb{C}$. Was macht man nun damit?

17.2.2 Eigenwertberechnungen über Teilkörpern von $\mathbb{C}$

Das zweite Beispiel oben zeigt, dass es für die Suche nach Eigenpaaren wichtig ist, über welchem Körper unsere Eigenwertaufgabe formuliert ist (wie es ja damals bei der Suche nach Nullstellen von Polynomen auch wichtig war, in welchem Körper man die Nullstellen sucht). Daher wollen wir die Hauptaufgabe im Folgenden noch genauer fassen und dabei im großen Stil die Ergebnisse zu Polynomen aus $\mathbb{K}[x]$ aus Kapitel 3 recyclen. Wir beschränken uns daher im Folgenden auf den Fall, dass $\mathbb{K}$ ein Teilkörper von $\mathbb{C}$ ist (d.h. eine Teilmenge von $\mathbb{C}$, die mit der Addition von $\mathbb{C}$ selbst ein Körper, also insbesondere abgeschlossen gegenüber diesen Operationen, ist). Das schließt den wichtigsten Fall $\mathbb{K} = \mathbb{R}$ mit ein.

Wir wollen wie folgt argumentieren: Ist $A \in \mathbb{K}^{n\times n}$, so ist wegen $\mathbb{K} \subseteq \mathbb{C}$ auch $A \in \mathbb{C}^{n\times n}$, A also als Matrix(-abbildung) „über $\mathbb{C}$“ deutbar und die Frage nach Eigenwerten von $A \in \mathbb{K}^{n\times n}$ auch „über $\mathbb{C}$“ stellbar: Gibt es $\lambda \in \mathbb{C}$ und $x \in \mathbb{C}^n$ so, dass $A \cdot x = \lambda \cdot x$ ist?

Das charakteristische Polynom p_A von A, mit dem wir Eigenwerte berechnen können, ist zwar ein Polynom mit Koeffizienten in $\mathbb{K}$, lässt sich aber auch als Polynom $p_A \in \mathbb{C}[x]$

mit $\mathrm{Grad}(p_A) \geq 1$ lesen. Jedes solche Polynom „zerfällt" nach dem Fundamentalsatz der Algebra 3.3.2 in $\mathbb{C}$ vollständig in Linearfaktoren, d.h. es existieren $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ so dass

$$p_A(t) = (-1)^n \prod_{k=1}^{n} (t - \lambda_n) \tag{17.2}$$

Die Zahlen λ_i, $i \in \underline{n}$, sind nach Satz 17.2.3 dabei gerade die Eigenwerte von A, wobei gemäß dem, was wir von der Zerlegung von Polynomen in Linearfaktoren kennen, ein Eigenwert auch mehrfach auftreten kann, d.h. es kann $\lambda_i = \lambda_j$ für $i \neq j$ sein. Die Vielfachheit $m_i \in \mathbb{N}$, mit der ein Eigenwert $\lambda_i \in \mathbb{C}$ als Nullstelle in p_A auftritt, bezeichnet man oft als algebraische Vielfachheit des Eigenwertes.

Da *jedes* charakteristische Polynom einer Matrix aus $\mathbb{K}^{n \times n}$ mindestens Grad 1 hat, ergibt sich sofort aus Ergebnissen aus Kapitel 3:

Lemma 17.2.4. *(Komplexe Eigenwerte)*

Sei $\mathbb{K}$ ein Teilkörper von $\mathbb{C}$ und $A \in \mathbb{K}^{n \times n}$. Dann gilt: Über $\mathbb{C}$ hat A mindestens einen und höchstens n verschiedene Eigenwerte.

Lemma 17.2.5. *(Reelle Eigenwerte)*

Ist $\mathbb{K} = \mathbb{R}$ und $A \in \mathbb{R}^{n \times n}$, so sind die reellen Nullstellen des charakteristischen Polynoms gerade die Eigenwerte $\lambda \in \mathbb{R}$ von A (als Matrixabbildung in $\mathbb{R}^n$ gelesen). Insbesondere gilt:

(i) Ist n ungerade, so hat jede Matrix $A \in \mathbb{R}^{n \times n}$ mindestens einen (reellen) Eigenwert;

(ii) Ist n gerade, so muss A keine Eigenwerte $\lambda \in \mathbb{R}$ besitzen.

(iii) Ist $\lambda \in \mathbb{R}$ eine Nullstelle des charakteristischen Polynoms, so besitzt $A \in \mathbb{R}^{n \times n}$ einen Eigenvektor $v \in \mathbb{R}^n$.

Beweis: Die ersten beiden Ergebnisse folgen mit Satz 17.2.3 aus den entsprechenden Aussagen zu Polynomen aus $\mathbb{R}[x]$ bzw. $\mathbb{C}[x]$ in Kapitel 3. Die dritte Aussage werden wir später noch einmal benötigen; sie folgt daraus, dass $\det(A - \lambda I) = 0$ bedeutet, dass die in $\mathbb{R}^n$ formulierte Gleichung $A \cdot x = \lambda \cdot x$ eine Lösung $x \in \mathbb{R}^n$ mit $x \neq 0_V$ haben muss. □

17.2.3 Berechnung von Eigenpaaren von Matrizen

Aus den vorangegangenen Ergebnissen ergibt sich nun ein praktisches Verfahren zur Berechnung der am Ende von Abschnitt 17.1 formulierten „Hauptaufgabe der Eigenwertpraxis“:

Algorithmus: Berechnung aller Eigenwerte/Eigenräume von $A \in \mathbb{K}^{n\times n}$

Gegeben: Matrix $A \in \mathbb{K}^{n\times n}$ (mit $\mathbb{K} \subseteq \mathbb{C}$).
Gesucht: Eigenwerte $\lambda \in \mathbb{K}$ von A, zugehörige Eigenräume $E(\lambda, A)$.

- Berechne $p_A(t) = \det(A - t \cdot I)$ und bestimme seine Nullstellen $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$.
- $\rightsquigarrow$ Eigenwerte von $A \in \mathbb{K}^{n\times n}$ sind diejenigen Nullstellen, die Elemente des Körpers $\mathbb{K}$ sind.
- Löse zu jedem Eigenwert das lineare Gleichungssystem $A - \lambda I = 0$
- $\rightsquigarrow$ Lösungsmenge ist der zugehörige Eigenraum, eine Basis lässt sich z.B. aus der normierten Zeilenstufenform von $A - \lambda I$ ablesen.

(3×3)**-Beispiel zur Berechung:** S. Vorlesung oder [BeuLA], S. 255.

Bemerkung zur „Eigenwertpraxis“ und dem vorgestellten Verfahren

Eigentlich handelt es sich bei dem oben angegebenen Verfahren in gewissem Sinne um einen „Pseudo-Algorithmus“, da das genaue Verfahren zur Berechnung von $\det(A - t \cdot I)$ und zur Bestimmung der Nullstellen p_A nicht konkret benannt wird. Während es für die Determinante im ersten Teil immerhin noch konkrete Rechenverfahren gibt, z.B. rekursive Entwicklung nach der ersten Zeile (die aber mit steigendem n schnell extrem aufwändig werden kann), gibt es, wie ich es im Kapitel über Polynome schon bemerkt hatte (Stichwort: Galois-Theoie), für die Bestimmung der Nullstellen von Polynomen von Grad größer 4, und damit für die Berechnung der Eigenwerte aller quadratischen Matrizen mit mehr als vier Zeilen/Spalten, keine allgemeine Lösungsformel. Man wird die Nullstellen von p_A für größeren Matrizen also numerisch annähern müssen, und auch das kann sehr aufwändig werden. Daher ist unser hier skizziertes Verfahren zwar der Standard in den LAAG-I-Vorlesungen der Welt und auch zur Berechnung der Eigenwerte kleiner Matrizen, für größere Matrizen sind die numerischen Standards das so genannte QR-Verfahren zur Eigenwertberechnung (für „mittelgroße“ Matrizen mit $n \approx 10^2 - 10^4$) bzw. das iterative

Lanczos-Verfahren für noch größere Probleme (die einem die z.B. Ingenieurswissenschaften bei genügend genauen Rechnungen ohne Probleme bescheren).

Wir beschließen diesen Abschnitt mit einem nützlichen Spezialfall und einem hoffentlich erhellenden Zusammenhang zwischen Determinante und Eigenwerten einer Matrix.

Lemma 17.2.6. *(Eigenwerte von Dreiecksmatrizen und Diagonalmatrizen)*

Sei $A \in \mathbb{K}^{n\times n}$ eine obere oder untere Dreiecksmatrix bzw. spezieller eine Diagonalmatrix (vgl. Def. 16.3.4). Dann sind die Eigenwerte von A genau die Einträge $a_{11}, \ldots, a_{nn}$ auf der Diagonalen von A.

Beweis: Wir berechnen p_A durch rekursive Entwicklung nach der jeweils ersten Spalte/Zeile. Es ergibt sich $p_A(t) = \prod_{i=1}^{n}(a_{ii} - t)$, und die Nullstellen dieses Polynoms liest man einfach ab.

□

Lemma 17.2.7. *(Determinante und Eigenwerte)*

Sei $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$ und $A \in \mathbb{K}^{n\times n}$ eine Matrix. $\lambda_1, \ldots, \lambda_k \in \mathbb{C}$ seien alle Eigenwerte von A (als komplexe Matrix gelesen), die jeweils mit den Vielfachheiten $m_1, \ldots, m_k$ als Nullstellen in der Linearfaktorzerlegung des charakteristischen Polynoms p_A von A auftreten. Dann gilt: Die Determinante von $A \in \mathbb{K}^{n\times n}$ ist gerade das Produkt der Eigenwerte von A (Vielfachheiten eingerechnet), d.h. es ist

$$\det(A) = \prod_{i=1}^{k} \lambda_i^{m_i} \in \mathbb{K}.$$

Beweis: Ist $p_A(t) = (-1)^n a_n t^n + a_{n-1}t^{n-1} + \ldots + a_1 x + a_0$, so ist nach Übungsaufgabe 6.1 der LAAG I ist der Koeffizient a_0 gerade das Produkt der Nullstellen von p (Vielfachhheiten jeweils entsprechend gezählt)– also ist $a_0 = \prod_{i=1}^{k} \lambda_i^{m_i}$ wie auf der rechten Seite der Formel oben. Andererseits ergibt sich das Absolutglied von p_A wie bei jedem anderen Polynom auch durch Einsetzen von $t = 0$, hier also zu $a_0 = p_A(0) = \det(A - 0 \cdot I) = \det(A)$.

□

17.3 Eigenwerte und Eigenräume von Endomorphismen

Ist V ein endlichdimensionaler $\mathbb{K}$-Vektorraum und $f : V \to V$ ein Endomorphismus, so lassen sich die Ergebnisse des letzten Abschnitts - wie eigentlich immer in der Linearen Algebra – direkt vom $\mathbb{K}^n$ auf V übertragen, indem man in V eine Basis einführt und ausnutzt, dass V und $\mathbb{K}^n$ per Koordinatenabbildung k_B zueinander isomorph sind. Es ist ja allgemein so, dass sich alle Ergebnisse, die nur die Vektorraumstruktur von V und $\mathbb{K}^n$ benutzen, aus diesem Grund eins zu ein miteinander identifizieren lassen. Wir wollen das für die Eigenwerttheorie in diesem Abschnitt genauer ausführen. Wenn wir im Fall der Eigenwertberechnung an Stelle von f als Abbildung in V dieselben Fragestellungen anhand der darstellenden Matrix $A^f_{B,B}$ von f, also anhand einer Matrix(-abbildung) auf $\mathbb{K}^n$ untersuchen, gilt der folgende Zusammenhang zwischen Eigenpaaren linearer Abbildungen und ihren darstellenden Matrizen:

Lemma 17.3.1. *(Eigenpaare von Endomorphismen und darstellenden Matrizen)*

Sei V endlichdimensional mit $\dim V = n$, B eine Basis von V und $k_B : V \to \mathbb{K}^n$ sei die Koordinatenabbildung bezüglich B. $f : V \to V$ sei linear und $A^f_{B,B}$ die darstellende Matrix von f bezüglich B. Dann gilt:

(i) f und $A^f_{B,B}$ haben dieselben Eigenwerte, d.h. $\sigma(f) = \sigma(A^f_{B,B})$.

(ii) Ein Vektor $v \in V$ ist ein Eigenvektor von f zum Eigenwert λ genau dann, wenn $x = k_B(v) \in \mathbb{K}^n$ Eigenvektor von $A^f_{B,B}$ zum selben Eigenwert λ ist.

(iii) Es gilt

$$\lambda \text{ ist Eigenwert von } f \quad \Longleftrightarrow \quad \det(f - \lambda \cdot \mathrm{id}_V) = 0.$$

Beweis: Es gilt nach Definition der darstellenden Matrix $A^f_{B,B}$ und der zugehörigen Matrixabbildung $f_B : \mathbb{K}^n \to \mathbb{K}^n$, dass

$$f = k_B^{-1} \circ f_A \circ k_B \quad \text{ und } \quad f_B = k_B \circ f \circ k_B^{-1}.$$

Ist also (v, λ) ein Eigenpaar von f, so gilt

$$A^f_{B,B} \cdot (k_B(v)) \;=\; f_B(k_B(v)) \;=\; (k_B \circ f \circ k_B^{-1})(k_B(v)) \;=\; k_B(\lambda \cdot v) = k_B(\lambda \cdot v);$$

ist umgekehrt $(x = k_B(v), \lambda)$ ein Eigenpaar von $A^f_{B,B}$, so gilt

$$f(v) = f(k_B^{-1}(x)) \;=\; (k_B^{-1} \circ f_B \circ k_B \circ k_B^{-1})(x) \;=\; k_B^{-1}(A^f_{B,B} \cdot x) = k_B^{-1}(\lambda \cdot x) = \lambda \cdot v.$$

Das zeigt bereits (i) und (ii), und (iii) folgt mit der Linearität der Koordinatenabbildung daraus, dass die darstellende Matrix der Abbildung $f - t \cdot \mathrm{id}_V$ wegen

$$k_B \circ (f - t \cdot \mathrm{id}_V) \circ k_B^{-1} \;=\; k_B \circ f \circ k_B^{-1} - t \cdot k_B \circ \mathrm{id}_V \circ k_B^{-1} = f_B - t \cdot \mathrm{id}_{\mathbb{K}^n}$$

gerade die Matrix $A^f_{B,B} - t \cdot I_n$ ist. Daraus ergibt sich wegen $\det(f - \lambda \cdot \mathrm{id}_V) := \det(A^f_{B,B} - \lambda I_n)$ die Äquivalenzkette

$$\begin{aligned} \det(f - \lambda \cdot \mathrm{id}_V) = 0 \quad &\iff \quad \det(A^f_{B,B} - \lambda I_n) = 0 \\ \iff \quad \lambda \text{ ist Eigenwert von } A^f_{B,B} \quad &\iff \quad \lambda \text{ ist Eigenwert von } f. \end{aligned}$$

□

Mit dem obigen Zusammenhang zwischen Endomorphismen endlichdimensionaler $\mathbb{K}$-Vektorräume und ihren darstellenden Matrizen können wir nun den folgenden Algorithmus zur „Hauptaufgabe der Eigenwerttheorie“ für allgemeine Endomorphismen $f : V \to V$ angeben.

Algorithmus: Berechnung aller Eigenwerte, Eigenräume von Endomorphismen

Gegeben: $\mathbb{K}$-Endomorphismus $f : V \to V$ (V endlichdimensional),
Basis $B = (b_1, \ldots, b_n)$ von V.
Gesucht: Eigenwerte $\lambda \in \mathbb{K}$ von f, zugehörige Eigenräume $E(\lambda, f)$.

- Bestimme die darstellende Matrix $A^f_{B,B}$ von f bezüglich B.
- Bestimme die Eigenwerte $\lambda \in \mathbb{K}$ und zugehörigen Eigenräume von $A^f_{B,B}$ (z.B. nach Algorithmus in Abschnitt 17.2.3)
- $\rightsquigarrow$ Die Eigenwerte von f sind genau die von $A^f_{B,B}$,
- $\rightsquigarrow$ Ist $x = (a_1, \ldots, a_n)^T \in \mathbb{K}^n$ Eigenvektor von $A^f_{B,B}$, so ist $v = \sum_{i=1}^n a_i \cdot b_i \in V$ Eigenvektor von f.

Viele Zusammenhänge zwischen Eigenwerten einer Matrix $A \in \mathbb{R}^{n\times n}$ und den Eigenschaften der Matrixabbildung f_A übertragen sich so wieder direkt auf Endomorphismen. Beispielsweise gilt:

- $\lambda \in \mathbb{K}$ ein Eigenwert von f ist genau dann, wenn $f - \lambda \operatorname{id}_V$ nicht injektiv/ nicht surjektiv/ nicht bijektiv ist, was wiederum genau dann stimmt, wenn der Rang von $f - \lambda \operatorname{id}_V$ kleiner als $\dim V$ ist.

Das folgt direkt aus dem Dimensionssatz für lineare Abbildungen zwischen endlichdimensionalen Vektorräumen (ÜA). Ist V unendlichdimensional, so gelten diese Beziehungen aber nicht mehr (z.B. ist die Ableitung $\frac{d}{dx}$ von oben surjektiv, aber nicht injektiv – letzteres können Sie z.B. für konstante Polynome schnell nachprüfen). Eigenwerte/-vektoren gehören hier nach wie vor zur Situation, in der $f-\lambda \operatorname{id}_V$ nicht injektiv ist, da die Herleitung auf Seite 17.1 auch in diesem Fall gültig ist.

In Hinblick auf das kommende Kapitel formulieren das obige Lemma 17.3.1 nochmal noch einmal für den wichtigen Fall von Matrixabbildungen und ihre Darstellung bezüglich anderer Basen des $\mathbb{K}^n$ als der Standardbasis. Hier hält das Lemma nämlich noch eine wesentliche Erkenntnis über ähnliche Matrizen bereit:

Lemma 17.3.2. *(Eigenwerte und -vektoren von Matrixabbildungen)*

Sei $f : \mathbb{K}^n \to \mathbb{K}^n, x \mapsto A \cdot x$ eine Matrixabbildung mit (quadratischer) Abbildungsmatrix A, d.h. A ist die darstellende Matrix von f bezüglich der Standardbasis.

Ist B eine weitere Basis des $\mathbb{K}^n$, so gilt für die darstellende Matrix $A^f_{B,B} = B \cdot A \cdot B^{-1}$ von f bezüglich B:

(i) A und $A^f_{B,B}$ haben dieselben Eigenwerte, d.h. $\sigma(A) = \sigma(A^f_{B,B})$.

(ii) Die Eigenvektoren von $A_{B,B}$ sind genau die Koordinatenvektoren der Eigenvektoren von f und A, d.h. es gilt: $x \in \mathbb{K}^n$ ist ein Eigenvektor von A zum Eigenwert λ genau dann, wenn $x' = k_B(x) \in \mathbb{K}^n$ Eigenvektor von $A^f_{B,B}$ zum Eigenwert λ ist.

Insbesondere bedeutet das:

Korollar 17.3.3. *Ähnliche Matrizen haben dieselben Eigenwerte und dasselbe charakteristische Polynom (aber i.A. nicht dieselben Eigenräume).*

– denn sie stellen dieselbe Matrixabbildung bezüglich verschiedener Basen dar.

Beweis: (i) und (ii) sind die entsprechenden Aussagen aus Lemma 17.3.1. Zur Klammerbemerkung halten wir genauer fest: Ist A die Abbildungsmatrix der Matrixabbildung $f : x \mapsto A \cdot x$, und $A' = CAC^{-1}$ mit einer invertierbaren Matrix C, so dass A' die darstellende Matrix von f bezüglich der Basis C, A die Abbildungsmatrix von f ist. Die beiden besitzen (ebenfalls nach Lemma 17.3.1, und da eine Matrixabbildung dieselben Eigenpaare besitzt wie ihre Abbildungsmatrix) also dieselben Eigenwerte und dasselbe charakteristische Polynom.

□

17.4 Aufgaben zu Kapitel 17

Aufgabe 17.97. (Determinanten/Eigenwerte von Endomorphismen):

Sei V ein $\mathbb{K}$ -Vektorraum und $f \in L(V)$ ein Endomorphismus von V. Wir schreiben

$$f^n := \underbrace{f \circ f \circ \cdots \circ f}_{n\text{-mal}}.$$

Wenn ein $n \in \mathbb{N}$ existiert, sodass f^n die Nullabbildung ist, dann heißt f *nilpotent.*
Beweisen Sie:

(a) Ist $f^2(v) = v$ für alle $v \in V$, dann gilt $|\det(f)| = 1$.

(b) Ist f nilpotent, dann gilt $\det(f) = 0$.

(c) Für alle $n \in \mathbb{N}$ gilt: Wenn λ Eigenwert von f ist, dann ist λ^n Eigenwert von f^n.

Überprüfen Sie außerdem, ob auch gilt: Existiert ein $k \in \mathbb{N}$, sodass λ^k Eigenwert von f^k ist, so ist auch λ Eigenwert von f.

(d) Sei V_3 der Vektorraum der Polynome vom Grad kleiner oder gleich 3 mit reellen Koeffizienten. Sei außerdem $f\colon V_3 \to V_3$ mit $p \mapsto p''$ der Endomorphismus, der einem Polynom seine zweite Ableitung zuordnet, das heißt,

$$f(\sum_{k=0}^{3} a_k x^k) = \sum_{k=2}^{3} k(k-1)a_k x^{k-2}.$$

Begründen Sie, dass $\lambda = 0$ der einzige Eigenwert von f ist, und bestimmen Sie den dazugehörigen Eigenraum $E(f, 0)$.

Aufgabe 17.98. (Basen aus Eigenvektoren):

Die Abbildungsmatrix A einer linearen Abbildung $f\colon \mathbb{R}^3 \to \mathbb{R}^3$ (also die darstellende Matrix von f bezüglich der Standardbasis) sei gegeben durch

$$A := \frac{1}{3} \cdot \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}$$

(a) Berechnen Sie jeweils eine Basis von Kern(f) und Bild(f). Ist f injektiv bzw. surjektiv?

(b) Bestimmen Sie die Determinante von A, ohne sie zu berechnen.

(c) Zeigen Sie dass die drei Vektoren

$$b_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \qquad b_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \qquad b_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Eigenvektoren von f sind, und geben Sie die dazugehörigen Eigenwerte an.

(d) Bestimmen Sie die darstellende Matrix $A_{B,B}$ von f bezüglich der Basis $B = (b_1, b_2, b_3)$.

Hinweis: Sie dürfen ohne Nachweis verwenden, dass B eine Basis ist.

Aufgabe 17.99. (Berechnung von Eigenräumen und Eigenwerten):

Bestimmen Sie für die Matrizen

$$A_1 := \begin{pmatrix} -2 & 1 & 2 \\ 6 & 2 & -3 \\ -2 & 2 & 3 \end{pmatrix} \in \mathbb{R}^{3\times 3} \quad \text{und} \quad A_2 := \begin{pmatrix} -1-i & -2-2i & -1-i \\ -i & 0 & -1 \\ 1+i & 2+2i & 1+i \end{pmatrix} \in \mathbb{C}^{3\times 3}$$

jeweils alle ihre Eigenwerte, deren algebraischen Vielfachheiten sowie Basen und Dimension der dazugehörigen Eigenräume.

Aufgabe 17.100. (Diagonalmatrizen):

Gegeben sei eine Diagonalmatrix $D = \operatorname{diag}(d_1, \ldots, d_n) \in \mathbb{K}^{n\times n}$, also eine Matrix

$$D = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & d_n \end{pmatrix}.$$

(a) Finden Sie ein Kriterium an die Diagonaleinträge von D, das äquivalent zur Invertierbarkeit von D ist.

(b) Bestimmen Sie die konkrete Gestalt der Matrix D^n, also der n-ten Potenz von D, sowie, für den Fall, dass D invertierbar ist, die konkrete Gestalt von D^{-1}.

(c) Bestimmen Sie alle Eigenwerte von D. Bestimmen Sie zudem eine Basis des $\mathbb{K}^n$, die nur aus Eigenvektoren von D besteht.

Aufgabe 17.101. (Eigenwerte und charakteristisches Polynom):

(a) Sei V ein $\mathbb{K}$-Vektorraum und $f\colon V \to V$ ein Endomorphismus mit $f^2 = \mathrm{id}_V$. Bestimmen Sie alle $\lambda \in \mathbb{K}$, die Eigenwert von f sein können.

(b) Sei $A \in \mathbb{K}^{n\times n}$. Zeigen Sie: A und A^T haben dieselben Eigenwerte.

(c) Zeigen Sie, dass ähnliche Matrizen das gleiche charakteristische Polynom haben.

Aufgabe 17.102. (Drehachsen von Hintereinanderausführungen):

Sie haben in einer früheren Aufgabe bereits gezeigt, dass für orthogonale Matrizen $Q \in \mathbb{R}^{n\times n}$ immer $\det(Q) = \pm 1$ gilt. Eine lineare Abbildung $f\colon \mathbb{R}^n \to \mathbb{R}^n$ mit orthogonaler Abbildungsmatrix Q, für die $\det(Q) = 1$ ist, nennt man Drehung.

(a) Begründen Sie, dass mit diesen Begrifflichkeiten die Hintereinanderausführung zweier Drehungen stets wieder eine Drehung ist.

(b) Es seien für $j = 1, 2$ die Abbildungen $f_j\colon \mathbb{R}^3 \to \mathbb{R}^3$ diejenigen Drehungen um die x_1-Achse bzw. die x_3-Achse des $\mathbb{R}^3$, die durch

$$f_1(\begin{pmatrix}0\\0\\1\end{pmatrix}) = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\1\\1\end{pmatrix} \qquad \text{bzw.} \qquad f_2(\begin{pmatrix}1\\0\\0\end{pmatrix}) = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\\0\end{pmatrix}$$

vollständig festgelegt sind. Bestimmen Sie die Drehwinkel (im Bogenmaß) und Abbildungsmatrizen von f_1 und f_2 bezüglich der Standardbasis des $\mathbb{R}^3$.

(c) Bestimmen Sie für die Drehung $f_1 \circ f_2$ die Drehachse sowie den Drehwinkel (im Bogenmaß).

Aufgabe 17.103. (Fehlende Beweise von 104 und 106):

Sei $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$. Beweisen Sie:

(ia) Ist $Q \in \mathbb{K}^{n\times n}$ eine orthogonale/unitäre Matrix, so haben alle Eigenwerte von Q Betrag 1.

(ib) Ist V ein beliebiger Skalarproduktraum und $f : V \to V$ ein isometrischer $\mathbb{K}$-Endomorphismus, so haben alle Eigenwerte $\lambda \in \mathbb{K}$ von f haben Betrag 1.

(ii) Mit den Voraussetzungen aus (ib) gilt: Sind u_1, u_2 Eigenvektoren von f, die zu verschiedenen Eigenwerten $\lambda_1, \lambda_2 \in \mathbb{K}$ gehören, so gilt $u_1 \perp u_2$.

Aufgabe 17.104. (Eigenwerte orth./unitärer Matrizen, „Achsen“ in $\mathbb{R}^3$):

Beweisen Sie die folgenden Aussagen über orthogonale und unitäre Matrizen:

(i) Für alle Eigenwerte λ einer orthogonalen/unitären Matrix $Q \in \mathbb{K}^{n\times n}$ ist $|\lambda| = 1$.

(ii) Eine orthogonale Matrix $Q \in \mathbb{R}^{3\times 3}$ mit $\det(Q) = 1$ hat immer $\lambda = 1$ als Eigenwert (also immer „mindestens eine Fixachse“),

(iii) Eine orthogonale Matrix $Q \in \mathbb{R}^{3\times 3}$ mit $\det(Q) = -1$ hat immer $\lambda = -1$ als Eigenwert (also immer „mindestens eine Spiegelrichtung“)„

(iv) „Satz vom Fußball“: Der Ball sei rund, und das Spiel dauere 90 Minuten. Unter diesen Annahmen gilt: Auf der Oberfläche eines Fußballs, der zu Beginn der ersten Halbzeit und zu Beginn der zweiten Halbzeit am selben Anstoßpunkt im Stadion liegt, gibt es mindestens zwei Punkte, die zu diesen zwei Zeitpunkten jeweils unverändert am selben Ort im umgebenden Raum liegen.

Bemerkung dazu:

- Unter anderem wegen der obigen Eigenschaften bezeichnet man orthogonale Matrizen $Q \in \mathbb{R}^{n\times n}$ mit $\det(Q) = 1$ als Drehungen, solche mit $\det(Q) = -1$ als Drehspiegelungen. Die entsprechenden Zusammenfassungen dieser Matrizen, bilden Untergruppen der orthogonalen Matrizen.
- Das charakteristische Polynom von Q ist ein Polynom dritten Grades. Eine Übungaufgabe aus der LAAG I und das vorangegangene Lemma helfen weiter.

Aufgabe 17.105. (Bestimmung der Rotationsachse des Mondes):

Bei der Beobachtung des Mondes werden drei Punkte A, B, C auf seiner Oberfläche gewählt messen und ihre Koordinaten zu einem und einem späteren festen Zeitpunkt gemessen. Der Einfachheit halber sei der Radius des Mondes gleich Eins und

$$A = (1, 0, 0), \quad B = (0, 1, 0), \quad C = (0, 0, 1)$$

und zum späteren Zeitpunkt

$$\bar{A} = \left(-\frac{3}{5}, 0, \frac{4}{5}\right), \quad \bar{B} = \left(\frac{4}{5}, 0, \frac{3}{5}\right), \quad \bar{C} = (0, 1, 0)$$

Geben Sie eine orthogonale Matrix Q an, die die Punkte ineinander überführt. Bestimmen Sie einen geeigneten Eigenvektor von Q, der die Rotationsachse des Mondes beschreibt.

Aufgabe 17.106. (Eigenpaare orthogonaler/unitärer Abbildungen):

Ist V ein Skalarproduktraum und $f : V \to V$ ein orthogonaler/unitärer $\mathbb{K}$-Endomorphismus. Beweisen Sie, dass dann gilt:

(i) Alle Eigenwerte $\lambda \in \mathbb{K}$ von f haben Betrag 1.

(ii) Sind u_1, u_2 Eigenvektoren zu verschiedenen Eigenwerten $\lambda_1, \lambda_2 \in \mathbb{K}$, so gilt $u_1 \perp u_2$.

18 Diagonalisierbarkeit und die Singulärwertzerlegung

18.1 Diagonalmatrizen und Diagonalisisierbarkeit

18.1.1 Der Begriff der Diagonalisierbarkeit – Einführung

Wir haben schon gesehen, dass es einige wichtige lineare Abbildungen $f : V \to V$ gibt, bei denen sich der abgebildete Vektorraum V in „Richtungen“ zerlegen lässt, die zu bestimmten Eigenwerten gehören Zum Beispiel bildet bei einer Spiegelung im $\mathbb{R}^2$ die Spiegelachse einen Eigenraum (zum Eigenwert 1, Vektoren auf der Spiegelachse bleiben fest); die Richtung senkrecht dazu gehört zum Eigenwert -1 (Ein Vektor senkrecht zur Spiegelachse wird „auf ihre andere Seite geklappt“). Wenn wir bei der geometrisch-anschaulichen Interpretation bleiben, so gilt: Liegen diese „Streckrichtungen“ in Richtung der Koordinatenachsen, so heißt das auf der Koordinatenebene, dass die Standardbasisvektoren $e_1, \ldots, e_n$ Eigenvektoren sind; beispielsweise beschreiben für

$$f: \quad \mathbb{R}^2 \to \mathbb{R}^2, \quad x \mapsto \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \cdot x \qquad \text{und} \qquad g: \quad \mathbb{R}^2 \to \mathbb{R}^2, \quad x \mapsto \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \cdot x$$

die Abbildung f eine Spiegelung an der y-Achse, die Abbildung g eine Projektion auf die x-Achse. Allgemein gilt: Sind die Standardbasisvektoren $e_1, \ldots, e_n$ Eigenvektoren zu Eigenwerten $\lambda_1, \ldots, \lambda_n$ einer Matrixabbildung $f : \mathbb{K}^n \to \mathbb{K}^n, x \mapsto A \cdot x$, so ist

$$A = \begin{pmatrix} \lambda_1 & 0 & \ldots & \ldots & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 \\ \vdots & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \lambda_{n-1} & 0 \\ 0 & 0 & \ldots & 0 & \lambda_n \end{pmatrix} \tag{18.1}$$

die zugehörige Abbildungsmatrix. Eine so einfache Darstellung einer linearen Abbildung in Diagonalgestalt hat viele Vorteile; einige davon werde ich in diesem ersten Abschnitt weiter unten darstellen.

Liegen die Eigenvektoren einer Matrixabbildung f nun nicht gerade „in Richtung der Achsen“, bilden aber dennoch eine Basis $C = (c_1, \ldots, c_n)$ des $\mathbb{K}^n$, so lässt sich die Matrixabbildung trotzdem in Diagonalgestalt schreiben, indem man die *darstellende Matrix* $A_{C,C}$ der Matrixabbidung f bezüglich dieser *Eigenbasis* betrachtet. Das fassen wir noch ein Stück allgemeiner für allgemeine Endomorphismen zusammen.

Definition 18.1.1. *(Eigenbasis)*
Sei $f : V \to V$ linear und $C = (c_1, \ldots, c_n)$ eine Basis von V. Wir nennen B eine Eigenbasis von V bezüglich f, falls alle Vektoren aus C Eigenvektoren von f sind.

Lemma 18.1.2. *(Darstellende Matrizen bezüglich Eigenbasen sind Diagonalmatrizen)*
Sei $f : \mathbb{K}^n \to \mathbb{K}^n, x \mapsto A \cdot x$ linear und $C = (c_1, \ldots, c_n)$ eine Basis von V. Dann ist C eine Eigenbasis von V bezüglich f genau dann, wenn die darstellende Matrix $A^f_{C,C}$ diagonal ist, d.h. wenn sie die folgende Gestalt hat:

$$A^f_{C,C} = \begin{pmatrix} \lambda_1 & 0 & \ldots & \ldots & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 \\ \vdots & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \lambda_{n-1} & 0 \\ 0 & 0 & \ldots & 0 & \lambda_n \end{pmatrix}. \tag{18.2}$$

Beweis: Wir zeigen: Der i-te Vektor c_i des Tupels C ist ein Eigenvektor von f mit zugehörigem Eigenwert $\lambda \in \mathbb{K}$ genau dann, wenn die i-te Spalte der darstellenden Matrix $A_{C,C} = (a_{i,j})_{i,j \in \underline{n}}$ von f bezüglich C gerade

$$(a_{i,j})_{j \in \underline{n}} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \lambda \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \leftarrow i\text{-te Zeile}$$

ist: In der i-ten Spalte der der darstellenden Matrix $A^f_{C,C}$ steht der Vektor $k_C(f(c_i))$. Es ist

$$f(c_i) = \lambda \cdot c_i = 0 \cdot c_1 + \ldots + 0 \cdot c_{i-1} + \lambda \cdot c_i + 0 \cdot c_{i+1} + \ldots + 0 \cdot c_n,$$

also $k_C(f(c_i)) = \lambda \cdot e_i$, das ist genau die oben angegebene Spalte.

□

Um diese Diagonaldarstellung für Matrixabbildungen bezüglich ihrer Eigenbasen hoffentlich erhellend weiter mit Leben zu füllen, schauen wir uns noch einmal ein Beispiel an, das wir auf Seite 228 als Anwendung darstellender Matrizen gesehen hatten: Dort hatten wir eine Spiegelung im $\mathbb{R}^2$ an der durch $c_1 = (1,2)^T$ erzeugten Ursprungsgerade dadurch definiert, dass wir die Wirkung der Spiegelung auf den Vektor c_1 und dem dazu senkrechten Vektor c_2 festlegen; genauer war

$$C = (c_1, c_2) = (\begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}-2\\1\end{pmatrix}) \qquad \text{und} \qquad f\begin{pmatrix}1\\2\end{pmatrix} = \begin{pmatrix}1\\2\end{pmatrix}, \qquad f\begin{pmatrix}-2\\1\end{pmatrix} = \begin{pmatrix}2\\-1\end{pmatrix}$$

In der Sprache unserer Zeit sind c_1 und c_2 also Eigenvektoren der Spiegelung f zu den Eigenwerten $\lambda_1 = 1$ und $\lambda_2 = -1$, und die darstellende Matrix von f bezüglich C hat wegen

$$f(c_1) = c_1 = 1 \cdot c_1 + 0 \cdot c_2, \qquad f(c_2) = -c_2 = 0 \cdot c_1 + (-1) \cdot c_2$$

die Diagonalform

$$A_{C,C} = \begin{pmatrix}1 & 0\\0 & -1\end{pmatrix}.$$

Die Abbildungsmatrix von f (also die darstellende Matrix bezüglich Standardbasis) erhält man daraus mit Hilfe von Satz 10.2.1; dort steht, dass

$$A = A_{I,I} = C \cdot A_{C,C} \cdot C^{-1} = \begin{pmatrix}-\frac{3}{5} & \frac{4}{5}\\ \frac{4}{5} & \frac{3}{5}\end{pmatrix}$$

ist, wobei in der Matrix C die Basisvektoren c_1, c_2 als Spalten stehen.

Kennt man nur die Abbildungsmatrix A, so sieht man ihr ihre Eigenwerte und Spiegelachsen nicht wirklich an. Wir werden uns im aktuellen Kapitel Gedanken darüber machen, wie man anhand A erkennen kann, ob eine Matrix eine Basis aus Eigenvektoren besitzt. Solche Matrizen heißen diagonalisierbar:

Definition 18.1.3. *(Diagonalisierbare Matrix, Diagonalisierung einer Matrix)*

Eine Matrix $A \in \mathbb{K}^{n\times n}$ heißt <u>über $\mathbb{K}$ diagonalisierbar</u>, wenn es eine Basis C von $\mathbb{K}^n$ gibt, die nur aus Eigenvektoren von A besteht. Die daraus resultierende Darstellung

$$A = C \cdot D \cdot C^{-1}$$

(mit der Diagonalmatrix $D = A_{C,C}$) nennt man eine <u>Diagonalisierung von A</u>.

Bemerkung dazu:

- Wie im Beispiel oben steht den Spalten der Matrix C dann gerade eine zugehörige Eigenbasis von A, und die Einträge auf der Diagonalen von D sind (in entsprechender Reihenfolge) genau die Eigenwerte von A.
- In anderen Worten ist A also diagonalisierbar genau dann, wenn A ähnlich zu einer Diagonalmatrix D ist.

Für allgemeine Endomorphismen ist die Definition analog.

Definition 18.1.4. *(Diagonalisierbarer Endomorphismus)*

Ein Endomorphismus $f : V \to V$ eines endlichdimensionalen K-Vektorraums V heißt diagonalisierbar, wenn V eine Basis C besitzt, die nur aus Eigenvektoren von f besteht.

Bemerkung dazu:

- $f \in L(V)$ ist also genau dann diagonalisierbar, wenn mit es eine Basis C gibt, so dass die darstellende Matrix $A^f_{C,C}$ von f bezüglich C eine Diagonalmatrix ist.
- Ein Endomorphismus $f \in L(V)$ ist genau dann diagonalisierbar, wenn eine zugehörige darstellende Matrix von f diagonalisierbar ist. Zwischen den entsprechenden Vektoren der Eigenbasen b_i^A für A und b_i^f für f besteht bei geeigneter Nummerierung die Beziehung $b_i^A = k_C(b_i^f)$, vgl. Lemma 17.3.1.

18.1.2 Was ist toll an Diagonalmatrizen?

Sei f ein diagonalisierbarer Endomorphismus, C eine Basis aus Eigenvektoren von f, und $A_{C,C}$ die zugehörige darstellende Matrix. Insbesondere wollen wir den Fall von Matrixabbildungen $f_A : \mathbb{K}^n \to \mathbb{K}^n, x \mapsto A \cdot x$ betrachten, dann ist $A = CA_{C,C}C^{-1}$ (vgl. 10.2.1). Wir führen einige Folgerungen daraus auf, dass f bzw. A diagonlaisierbar ist.

- „Zerlegung“ der linearen Abbildung: Ist λ ein Eigenwert, so wird der Eigenraum $E(\lambda, f)$ durch f in sich selbst abgebildet (das passt ja auch namentlich, oder?): Vektoren aus $E(\lambda, f)$ werden ja jeweils nur mit einem Faktor (nämlich λ) skaliert.

Das bedeutet, dass man f, eingeschränkt auf $E(\lambda)$, jeweils auch als lineare Abbildung des Raumes $E(\lambda, f)$ *in sich* verstehen kann. Existiert nun eine Basis an Eigenvektoren, so ist der gesamte Raum aus diesen Eigenräumen als direkte Summe „zusammengesetzt", vgl. Lemma 17.1.6, und f selbst ist also gewissermaßen „zusammengesetzt" aus lauter solchen Abbildungen, die die einzelnen Eigenräume „durch Skalieren" in sich selbst abbilden. Haben wir also eine solche Darstellung Diagonalisierung eines Endomorphismus f berechnet (mit Eigenwerten und Eigenbasen), so ist damit ziemlich gut beschrieben, was diese Abbildung f eigentlich tut. (Man vergleiche das mit z.B. der Situation, in der man für eine beliebige (20×20)-Matrix entscheiden soll, wie die zugehörige Matrixabbildung eigentlich wirkt.

- Einfaches Rechnen: Ist $A^f_{C,C}$ diagonal wie in 18.2, so gilt $f(c_i) = \lambda_i \cdot c_i$ für alle Basisvektoren $c_i \in C$. Stellt man beliebige Vektoren $v \in V$ als $v = \sum_{i=1}^n a_i \cdot c_i$ in der Basis C dar, so wird wegen der Linearität in dieser Darstellung jeder Basisvektor einfach mit dem entsprechenden Eigenwert multipliziert, $f(v)$ ist also einfach gegeben durch

$$f(v) = \sum_{i=1}^n \lambda_i a_i \cdot c_i.$$

Rechnet man nur noch mit Koordinatenvektoren, so bedeutet das: Ist $k_C(v) = (a_1, \ldots, a_n) \in \mathbb{K}^n$ ein Koordinatenvektor von $v \in V$ bezüglich C, so entsteht der von $f(v)$ daraus gerade durch Skalierung mit den Eigenwerten $k_C(f(v)) = (\lambda_1 a_1, \ldots, \lambda_n a_n) \in \mathbb{K}^n$. Das heißt insbesondere:

  - Für eine Matrixabbildung mit diagonalisierbarer Matrix A lässt sich die Multiplikation

$$A \cdot x = C \cdot D \cdot C^{-1} \cdot x.$$

  gerade (von rechts nach links) interpretieren als

    (i) die Umrechnung der Koordinaten bezüglich Standardbasis auf Koordinaten bezüglich C (durch C^{-1}),

    (ii) das anschließende „Durchskalieren" der Einträge mit den entsprechenden Eigenwerten (durch Anwenden von D), und

    (iii) der anschließenden Rücktransformation auf Standardbasiskoordinaten (durch Multiplikation mit C).

  - Beim Umgang mit den „normalen" Operationen wie Matrix-Vektor-Multiplikation sind in der Diagonalmatrix fast alle Einträge Null, das lässt sich beim Berechnen von Matrix-Vektor-Produkten ausnutzen, wenn sich die Matrix C einfach invertieren lässt (S. Abschnitt 18.3 unten: Das ist z.B. für symmetrische Matrizen der Fall!).

- Mehrfache Anwendung von f auf $v \in V$, also den Vektor $f^n(v) = (f \circ \ldots \circ f)(v)$, erhält man folgendermaßen: Man prüft, dass die Potenz A^n einer Matrix A über

$$A^n = C \cdot D^n \cdot C^{-1}$$

gegeben ist. Die Potenzen der Diagonalmatrix D ist leicht berechenbar, denn Diagonalmatrizen multipliziert man einfach miteinander, indem man ihre Diagonaleinträge miteinander multipliziert. Genauer gilt: Ist $k_C(v) = (a_1, \ldots, a_n) \in \mathbb{K}^n$ ein Koordinatenvektor von $v \in V$ bezüglich C, so ist der von $f(v)$ gerade der, der durch Skalierung dieses Vektors mit den Eigenwerten entsteht, also $k_C(f(v)) = (\lambda_1 a_1, \ldots, \lambda_n a_n) \in \mathbb{K}^n$, und daher $k_C(f^n(v)) = (\lambda_1^n a_1, \ldots, \lambda_n^n a_n) \in \mathbb{K}^n$.

- f ist invertierbar, besitzt also eine Umkehrfunktion, genau dann, wenn Null kein Eigenwert von f ist. In diesem Fall ist $k_C(f^{-1}(v)) = (\lambda_1^{-1} x_1, \ldots, \lambda_n^{-1} x_n) \in \mathbb{K}^n$. Im Fall von Matrixabbildungen ist die Inverse von A ist dann

$$A^{-1} = C \cdot D^{-1} \cdot C^{-1},$$

wobei die Inverse der Diagonalmatrix D einfach die Diagonalmatrix mit den Einträgen $\lambda_1^{-1}, \ldots, \lambda_n^{-1}$ ist.

- „Strukturelle Einsicht“: Für eine Matrixabbildung mit diagonalisierbarer Matrix A lässt sich die Multiplikation

$$A \cdot x = C \cdot D \cdot C^{-1} \cdot x.$$

gerade (von rechts nach links) interpretieren als

(i) die Umrechnung des Koordinatenvektors x bezüglich Standardbasis auf Koordinaten bezüglich C (durch C^{-1}),

(ii) das anschließende „Durchskalieren“ der Einträge mit den entsprechenden Eigenwerten (durch Anwenden von D), und

(iii) der anschließenden Rücktransformation auf Standardbasiskoordinaten (durch Multiplikation mit C).

18.1.3 Diagonalisierbar oder nicht?

An unseren bisherigen Beispielen haben wir einige Eigenschaften von Matrixabbildungen und ihren Eigenräumen sehen können. Daraus ergeben sich die folgenden beiden Fragen, die wir im Folgenden genauer untersuchen wollen.

(i) Wir haben schon gesehen, dass nicht jede Matrix reell diagonalisierbar ist - z.B. die Drehmatrix oben, die aber über $\mathbb{C}$ diagonalisierbar war. Ist eine Matrix $A \in \mathbb{R}^{n\times n}$ oder $A \in \mathbb{R}^{n\times n}$ *immer* über $\mathbb{C}$ diagonalisierbar?

(ii) In unseren Beispielen fällt auf, dass Eigenräume diagonalisierbarer Matrizen (z.B. bei Spiegelungen, Streckungen,...) oft senkrecht zueinander sind. Ist das immer so?

Unglücklicherweise lauten die Antworten auf diese beiden Fragen nein und nein. Wir geben zunächst entsprechende Gegenbeispiele und formulieren dann genauere Fragen, um deren Klärung es in den Abschnitten 18.2 und 18.3 gehen wird.

Als Gegenbeispiel zu (i) betrachten wir zunächst gleich zwei Gegenbeispiele, nämlich die Matrizen

$$A_0 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{und} \quad A_a = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \text{ mit } a \neq 0$$

die nach 17.2.6 jeweils nur den Eigenwert 0 bzw. 1 besitzen, der Eigenraum dazu aber nur eindimensional ist, es also keine Basis des $\mathbb{R}^2$ aus Eigenvektoren von A_0 bzw. A_1 geben kann. (Die Matrix rechts ist gerade die einer Scherung in $\mathbb{R}^2$.)

Verallgemeinern lässt sich dieses Beispiel durch Matrizen der Form

$$A_\lambda = \begin{pmatrix} \lambda & 1 & \dots & \dots & 0 \\ 0 & \lambda & 1 & \dots & 0 \\ \vdots & 0 & \ddots & 1 & 0 \\ \vdots & \vdots & \ddots & \lambda & 1 \\ 0 & 0 & \dots & 0 & \lambda \end{pmatrix} \in \mathbb{K}^{n\times n},$$

für beliebig große $n \in \mathbb{N}$, die nur λ als Eigenwert besitzen, wobei $A_\lambda - \lambda$ aber offensichtlich Rang $n-1$ hat, also $\dim E(\lambda, A) = \dim \operatorname{Kern}(A - \lambda) = 1$ gilt.

Als Gegenbeispiel zu (ii) konstruieren wir umgekehrt einfach eine Abbildung mit nicht senkrechten Eigenvektoren. Konkret ist z.B. im $\mathbb{R}^2$ die durch $c_1 = (1,0)$, $c_2 = (1,1)^T$ gegebene Basis nicht orthogonal bezüglich des Standardskalarprodukts. Definieren wir eine lineare Abbildung durch $f(c_1) = c_1$ und $f(c_2) = 2 \cdot c_2$, so ist $E(A,1) = \mathcal{L}(c_1)$ und $E(A,2) = \mathcal{L}(c_2)$ Es kann also keine Orthogonalbasis (c_1', c_2') von C geben, da für solche Eigenvektoren immer $c_1 \in E(A,1) = \mathcal{L}(c_1)$ und $c_1 \in E(A,2)$ gelten muss. Die Abbildungsmatrix berechnet man dann als

$$A = C \cdot \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \cdot C^{-1}.$$

Wir sparen uns das hier und formulieren lieber genauere Versionen unserer Leitfragen oben, die wir wie angekündigt jetzt in 18.2 und 18.3 klären wollen:

(i') *Wann* ist eine Matrix $A \in \mathbb{K}^{n \times n}$ über $\mathbb{K}$ diagonalisierbar?

(ii') *Wann* sind Eigenräume einer diagonalisierbaren Matrix senkrecht zueinander?

18.2 Kriterien für Diagonalisierbarkeit

Die Frage, welche Eigenschaften eine Matrix besitzen muss, damit sie diagonalisierbar ist und weiter, ob man Matrizen, die nicht diagonalisierbar sind, auf ähnliche Weise sinnvoll „zerlegen“ kann, hat die Mathematiker des späteren 18. Jahrhunderts einige Jahrzehnte beschäftigt, siehe den historischen Überblick in Abschnitt 18.5.1. Die wichtigsten Resultate, insbesondere mehrere allgemeines Kriterien und der Hauptsatz über die Diagonalisierbarkeit symmetrischer und hermitescher Matrizen, sind das Thema dieses Abschnitts.

Wir starten mit dem folgenden einfachen, gelegentlich auftretender Fall:

Lemma 18.2.1. *Sei* $\dim(V) = n$ *und* $f \in L(V)$ *ein Endomorphismus, der* n *verschiedene Eigenwerte besitzt. Dann ist* f *diagonalisierbar.*

Beweis: Da es zu jedem der n Eigenwerte $\lambda_1, \ldots, \lambda_n$ mindestens einen zugehörigen Eigenvektor gibt, und diese nach 17.1.4 linear unabhängig sind, sind sie wegen $\dim(V) = n$ nach 8.3.9 eine Eigenbasis von V.

□

Um die allgemeine Frage nach der Diagonalisierbarkeit zu eines Endomorphismus bzw. einer Matrix zu beantworten, müssen wir etwas tiefer einsteigen. Das zentrale Resultat ist das folgende. Wir geben danach zunächst ein Beispiel für die Anwendung; dann folgt der etwas längere Beweis.

Satz 18.2.2. *(Diagonalisierbarkeit von Endomorphismen/quadratischen Matrizen)*

Für einen Endomorphismus $f : V \to V$ eines endlichdimensionalen $\mathbb{K}$-Vektorraums V sind die folgenden Aussagen äquivalent:

(i) *f ist diagonalisierbar, d.h. V besitzt eine Basis aus Eigenvektoren von f.*

(ii) *(a) Das charakteristische Polynom p_f von f zerfällt in $\mathbb{K}[x]$ in Linearfaktoren, und*

(b) ist $p_f(t) = (-1)^n \cdot \prod_{i=1}^{k}(t - \lambda_i)^{m_i}$ diese Zerlegung von p_f mit den Eigenwerten $\lambda_1, \ldots, \lambda_k$ von f, und die Zahl $m_i \in \mathbb{N}$ deren jeweilige algebraische Vielfachheit, *so ist* $\dim E(\lambda_i, f) = m_i$ *für jeden der Eigenwerte λ_i.*

(iii) *Sind $\lambda_1, \ldots, \lambda_k \in \mathbb{K}$ die verschiedenen Eigenwerte von f, so ist V die direkte Summe aller Eigenräume, d.h. es ist mit dem in 17.1.6 eingeführten Unterraum* $\mathrm{Eig}(f)$

$$V = \mathrm{Eig}(f) = E(\lambda_1, f) \oplus \ldots \oplus E(\lambda_k, f)\,.$$

(iv) *Sind $\lambda_1, \ldots, \lambda_k \in \mathbb{K}$ die verschiedenen Eigenwerte von f, so ist*

$$\dim V = \sum_{i=1}^{k} \dim E(\lambda_i, f)\,.$$

All diese Äquivalenzen gelten genauso für die Diagonalisierbarkeit einer Matrix $A \in \mathbb{K}^{n \times n}$, wenn man in den obigen Aussagen f durch A ersetzt.

Bemerkung dazu:

- Ist λ ein Eigenwert eines Endomorphismus f, so bezeichnet man die Dimension des Eigenraumes $E(\lambda, f)$ auch als die geometrische Vielfachheit des Eigenwertes λ. Nach der obigen Äquivalenz (i) $\Leftrightarrow$ (ii) ist ein Endomorphismus also in dieser Sprechweise diagonalisierbar, wenn sein charakteristisches Polynom in Linearfaktoren zerfällt und für jeden Eigenwert seine algebraische und geometrische Vielfachheit gleich sind.

Anwendung des Kriteriums: Um die Diagonalisierbarkeit einer quadratischen Matrix $A \in \mathbb{K}^{n\times n}$ zu prüfen, kann man also

- die algebraischen Vielfachheiten m_i alle Eigenwerte λ_i von f bestimmen (mit Hilfe des charakteristischen Polynoms), und dann jeweils $\dim E(\lambda_i, A) = \operatorname{rg}(A - \lambda_i \cdot I)$ berechnen und prüfen, ob $\dim E(\lambda_i, A) = m_i$ gilt.
- inbesondere hierbei jeden Eigenwert der algebraischen Vielfachheit 1 überspringen, da hier nach Lemma 18.2.3 ohnehin $\dim E(\lambda_i, f) = 1 = m_i$ gilt;
- insbesondere aufhören, falls man einen Eigenwert λ gefunden hat, für den die geometrische Vielfachheit $\dim E(\lambda, A)$ nicht gleich der algebraischen Vielfachheit von λ als Nullstelle des charakteristischen Polynoms ist, denn dann kann (iv) aus Satz 18.2.2 wegen Lemma 18.2.3 nicht erfüllt sein.

Beispiel: Um die Diagonalisierbarkeit von

$$A = \begin{pmatrix} -1 & 0 & 0 & 6 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

zu untersuchen, berechnen wir die Eigenwerte und Eigenräume von A (nach dem Verfahren in Abschnitt 17.2.3). Das charakteristische Polynom von A ist $p_A(t) = (t+1)(t-1)^3$, die Eigenwerte also -1 und 1, mit algebraischen Vielfachheiten 1 und 3. Als zugehörige Eigenräume ergeben sich

$$E(-1, A) = \ker \begin{pmatrix} 0 & 0 & 0 & 6 \\ 0 & 2 & 0 & -3 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix} \quad \text{und} \quad E(1, A) = \ker \begin{pmatrix} -2 & 0 & 0 & 6 \\ 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Der zweite hat nur Dimension 2, das reicht nicht für Diagonalisierbarkeit.

Wir bereiten den Beweis von Satz 18.2.2 mit dem folgenden Zusammenhang zwischen geometrischer und algebraischer Vielfachheit vor, der im dann folgenden Beweis benötigt wird.

Lemma 18.2.3. *(Maximale Dimension eines Eigenraumes)*

Sei V endlichdimensional und λ ein Eigenwert einer Abbildung $f \in L(V)$. $m \geq 1$ bezeichne die Vielfachheit, mit der der Eigenwert λ als Linearfaktor in der Linearfaktorzerlegung des charakteristischen Polynoms auftritt. Dann gilt

$$1 \;\leq\; \dim E(\lambda, f) \;\leq\; m.$$

Beweis: Zunächst ist $\dim E(\lambda, f) \geq 1$ klar, da λ ein Eigenwert ist, zu dem ein Eigenvektor $v_\lambda \neq 0$ gehört, also gilt $E(\lambda, f) \neq \{0_V\}$ und damit $\dim E(\lambda, f) \neq 0$. Für die Aussage $\dim E(\lambda, f) \leq m$ sei $k := \dim E(\lambda, f)$ und $u_1, \ldots, u_k$ eine Basis von $E(\lambda, f)$. Ergänzen wir diese zu einer Basis B von ganz V, so ist die darstellende Matrix $A_{B,B}$ von f bezüglich B von der Form

$$A_{B,B} = \left(\begin{array}{cccc|c} \lambda & 0 & \cdots & 0 & \\ 0 & \lambda & 0 & \vdots & * \\ 0 & \ddots & \ddots & 0 & \\ 0 & \ldots & 0 & \lambda & \\ \hline 0 & 0 & \cdots & 0 & \\ \vdots & \vdots & & \vdots & C \\ 0 & 0 & \cdots & 0 & \end{array}\right).$$

Sukzessives k-faches Entwickeln der Determinante nach der ersten Spalte ergibt

$$p_A(t) = \det(A_{B,B} - t \cdot I) = (\lambda - t)^k \cdot \det(C - t \cdot I),$$

λ ist also eine Nullstelle der algebraischen Vielfachheit mindestens k des charakteristischen Polynoms von f, d.h. $m \geq k$.

□

Beweis von Satz 18.2.2: Wir bezeichnen mit $\lambda_1, \ldots, \lambda_k$ die Eigenwerte von f und mit m_i, $i \in \underline{k}$, ihre jeweiligen algebraischen Vielfachheiten. Zum Beweis nutzen wir hauptsächlich 17.1.6 und den dort eingeführten Unterraum von V,

$$\begin{aligned} \operatorname{Eig}(f) &:= E(\lambda_1, f) + \ldots + E(\lambda_k, f) \\ &= \{v \in V \mid v = v_1 + \ldots + v_k \text{ mit } v_i \in E(\lambda_i, f)\} \quad \subseteq V, \end{aligned}$$

der von allen Eigenvektoren von f aufgespannt wird. Für die Dimension dieses Raumes ist nach 17.1.6(iii) auch

$$\dim \operatorname{Eig}(f) = \sum_{i=1}^{k} \dim E(\lambda_i, f). \qquad (\star)$$

Als zweite Zutat halten wir fest, dass nach Lemma 18.2.3 für die Eigenräume $E(\lambda_i, f)$ die Ungleichung

$$\sum_{i=1}^{k} \dim E(\lambda_i, f) \leq \sum_{i=1}^{k} m_i \leq \dim V \qquad (\star\star)$$

gilt. Damit sind wir gut gerüstet und lesen die behaupteten Implikationen einfach ab:

(i) $\Rightarrow$ (ii): Besitzt V eine Basis aus Eigenvektoren $b_1, \ldots, b_n$, so ist $V = \mathcal{L}(b_1, \ldots, b_n) \subseteq \operatorname{Eig}(f)$. Nach $(\star)$ und $(\star\star)$ ist daher

$$\dim V \leq \sum_{i=1}^{k} \dim E(\lambda_i, f) \leq \sum_{i=1}^{k} m_i \leq \dim V;$$

damit das geht, müssen all diese Ungleichungen mit Gleichheit erfüllt sein, das heißt, es muss $\dim E(\lambda_i, f) = m_i$ (f.a. $i \in \underline{k}$) und $\sum_{i=1}^{k} m_i = \dim V$ gelten, was (ii) impliziert.

(ii) $\Rightarrow$ (iv): Gilt (ii), so gilt in $(\star\star)$ an beiden Ungleichungszeichen Gleichheit, das gibt (iv).

(iv) $\Rightarrow$ (iii): Gilt (iv), so ist nach $(\star)$ ist $\dim \operatorname{Eig}(f) = \sum_{i=1}^{k} \dim E(\lambda_i, f) = \dim V$, also wegen $\operatorname{Eig}(f) \subseteq V$ dann $\operatorname{Eig}(f) = V$.

(iii) $\Rightarrow$ (i): Wählt man in jedem der Eigenräume $E(\lambda_i, f)$ eine zugehörige Basis B_i, so ist $B = \cup_{i=1}^{k} B_i$ nach 17.1.6(iii) eine Basis von $\operatorname{Eig}(f)$, also, falls (iii) erfüllt ist, eine Basis von V, die nur aus Eigenvektoren besteht.

18.3 Diagonalisierbarkeit symmetrischer/hermitescher Matrizen

Der letzte Abschnitt hat gezeigt, dass im Allgemeinen relativ viel zu tun ist, wenn man über die Diagonalisierbarkeit einer Matrix entscheiden möchte. Umso erstaunlicher ist daher, dass man einer Matrix die Antwort auf die Frage nach der Diagonalisierbarkeit *mit orthogonalen Eigenräumen* einfach ansehen kann. Wir formulieren das Ergebnis zunächst für den wichtigen Fall reeller Matrizen $A \in \mathbb{R}^{n \times n}$, direkt darunter dann für den Fall komplexer Matrizen.

Satz 18.3.1. *(Spektralsatz für reelle, symmetrische Matrizen)*

Sei $A \in \mathbb{R}^{n \times n}$. Es sind äquivalent:

(i) A ist symmetrisch, d.h. es ist $A = A^T$,

(ii) A ist reell und orthogonal diagonalisierbar, d.h. A ist diagonalisierbar, wobei

- *alle Eigenwerte von A sind reell sind und*
- *eine zugehörige Eigenbasis so gewählt werden kann, dass sie eine Orthonormalbasis des $\mathbb{R}^n$ ist.*

(iii) Es gibt eine Diagonalmatrix $D \in \mathbb{R}^{n \times n}$ und eine orthogonale Matrix $Q \in \mathbb{R}^{n \times n}$ so, dass

$$A = QDQ^T$$

eine orthogonale Diagonalisierung von A ist.

Satz 18.3.2. *(Spektralsatz für komplexe, hermitesche Matrizen)*

Sei $A \in \mathbb{C}^{n \times n}$. Es sind äquivalent:

(i) A ist hermitesch, d.h. es ist $A = A^H$,

(ii) A ist reell und unitär diagonalisierbar, d.h. A ist diagonalisierbar, wobei

- *alle Eigenwerte von A reell sind und*
- *eine zugehörige Eigenbasis so gewählt werden kann, dass sie eine Orthonormalbasis des $\mathbb{C}^n$ ist.*

(iii) Es gibt eine Diagonalmatrix $D \in \mathbb{R}^{n \times n}$ und eine unitäre Matrix $Q \in \mathbb{C}^{n \times n}$ so, dass

$$A = QDQ^H$$

eine unitäre Diagonalisierung von A ist.

Im Folgenden fassen wir den reellen und den komplexen Fall, wo immer es sinnvoll ist, zusammen; insbesondere reden wir von selbstadjungierten Matrizen $A \in \mathbb{K}^{n\times n}$ und meinen damit symmetrische (im Fall $\mathbb{K} = \mathbb{R}$) bzw. hermitesche (im Fall $\mathbb{K} = \mathbb{R}$).

Der obige Satz ist dann der „Spektalsatz über die reellen Diagonalisierbarkeit selbstadjungierter Matrizen“. Ich habe hier beide Versionen formuliert, um die Aussage in beiden Fällen klarer herausstehen zu lassen. Bevor wir den Satz beweisen, einige Bemerkungen dazu:

Bemerkungen dazu:

- Zusammengefasst werden symmetrische Matrizen $A \in \mathbb{R}^{n\times n}$ und hermitesche Matrizen $A \in \mathbb{C}^{n\times n}$ auch als selbstadjungierte Matrizen bezeichnet, und
- Die Formel $A = QDQ^T$ bzw. $A = QDQ^H$ zeigt einen wesentlichen Vorteil der orthogonalen/unitären Diagonalisierbarkeit – die Transformationsmatrix Q^{-1}, die Koordinaten bezüglich der Standardbasis auf Koordinaten bezüglich der Basis Q transformiert, muss man nicht mehr mühsam ausrechnen, es ist dank Orthogonalität einfach Q^T/Q^H.
- Die Aussage (ii) und (iii) sind offensichtlich äquivalent, und aus (iii) folgt wegen $(QDQ^H)^H = (Q^H)^H D^H Q^H = QDQ^H$ im komplexen Fall und $Q^T = Q^H$ im reellen Fall auch direkt (i). Der Hauptklops ist aber, die Implikation (i) $\Longrightarrow$ (ii) zu zeigen, s.u.
- Das Beispiel
$$\begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix}$$
zeigt, dass symmetrische Matrizen mit komplexen Einträgen im allgemeinen nicht diagonalisierbar sind.
- Der Satz lässt sich auch auf Endomorphismen verallgemeinern, wenn man einen $\mathbb{R}$-Endomorphismus $f : V \to V$ eines Skalarproduktraumes $(V, \langle , \rangle)$ symmetrisch bzw. hermitesch nennt, falls für alle $v, w \in V$ die Beziehung
$$\langle f(v), w\rangle = \langle v, f(w)\rangle \quad \text{bzw} \quad \langle f(v), w\rangle = \overline{\langle v, f(w)\rangle}$$
(im Fall eines $\mathbb{R}$-Vektorraumes bzw. eines $\mathbb{C}$-Vektorraumes) gilt. Diese fasst man dann wieder unter dem Begriff „selbstadjungierte Endomorphismen“ zusammen. Für solche sind auch alle darstellenden Matrizen $A_{B,B}$ von f selbstadjungiert (mit der

entsprechenden obigen Definition der (doppelt verwendeten) Begriffe für Matrizen), und f daher dann reell und orthogonal bzw. reell und unitär diagonalisierbar.

- Verallgemeinert auf selbstadjungierte lineare Funktionen nun noch auf unendlichdimensionalen Vektorräumen (so genannte Operatoren) heißt das obige Ergebnis der „Spektralsatz für selbstadjungierte Operatoren“ und ist ein zentrales, wenn nicht das zentrale, Ergebnis der Funktionalanalysis (das ist die „Lineare Algebra von unendlichdimensionalen Vektorräumen“). Das liegt unter anderem daran, dass viele Differentialoperatoren selbstadjungiert auf geeigneten Funktionenräumen sind, und z.B. selbstadjungierte Operatoren nach den Postulaten der Quantenmechanik dort auch genau den Messoperatoren und somit den messbaren Größen der Physik entsprechen.

- Ist $Q = (q_1, \ldots, q_n)$ eine Orthonormalbasis, so ist nach Lemma 14.3.7 für $x \in \mathbb{K}^n$ die Darstellung in der Basis Q gegeben durch

$$x = \sum_{i=1}^{n} \langle q_i, x \rangle q_i = q_i q_i^H x$$

(wobei im reellen Fall $q_i^H = q_i^T$ ist). Hiermit rechnet man nach, dass dann

$$A \cdot x = \sum_{i=1}^{n} \lambda_i q_i q_i^H x = \big(\sum_{i=1}^{n} \lambda_i q_i q_i^T \big) x$$

ist, sich also A auch als

$$A \;=\; \sum_{i=1}^{n} \lambda_i q_i q_i^H \;=\; \lambda_1 q_1 q_1^H + \ldots + \lambda_n q_n q_n^H$$

schreiben lässt, also als Summe von skalierten Projektoren auf die von den jeweiligen Eigenvektoren erzeugten eindimensionalen Unterräume $\mathcal{L}(q_i)$; Genaueres s. Vl/Übung.

Zum Beweis des Spektralsatzes 18.3.1 bzw. 18.3.2 für selbstadjungierte Matrizen fangen wir mit einem Lemma an, das zwei Resultate sammelt, die wir im Beweis benötigen.

Lemma 18.3.3. *(Eigenwerte und -vektoren selbstadjungierter Matrizen)*
Sei $A \in \mathbb{K}^{n \times n}$ selbstadjungiert. Dann gilt:

(i) A besitzt nur reelle Eigenwerte (und im Fall $\mathbb{K} = \mathbb{R}$ auch einen zugehörigen Eigenvektor $x \in \mathbb{R}^n$).

(ii) Ist λ ein Eigenwert von A und $x \in \mathbb{K}^n$, so gilt:

$$x \perp E(\lambda, A) \quad \Longrightarrow \quad A \cdot x \perp E(\lambda, A). \tag{18.3}$$

(iii) Sind λ_1, λ_2 verschiedene Eigenwerte von A, so gilt $E(\lambda_1, A) \perp E(\lambda_2, A)$, d.h. $u_1 \perp u_2$ für alle $E(\lambda_1, A) \in U_1, u_2 \in E(\lambda_2, A)$.

Beweis: Zu (i) bemerken wir zunächst zum Fall $\mathbb{K} = \mathbb{R}$: Ist $A = A^T$, so ist, da alle Einträge von A reell sind, auch $A = A^H$. Daher gilt allgemein für selbstadjungierte Matrizen im symmetrischen wie im hermiteschen Fall nach 14.6.4(ii) mit dem *komplexen* Skalarprodukt für alle $x, y \in \mathbb{C}^n$ die Gleichheit $\langle Ax, y \rangle = \langle x, Ay \rangle$. Ist nun $\lambda \in \mathbb{C}$ ein (möglicherweise komplexer) Eigenwert von A und $x \in \mathbb{C}^n$ ein normierter Eigenvektor dazu, so gilt also (wieder mit dem komplexen Skalarprodukt), dass

$$\lambda = \lambda \langle x, x \rangle = \langle x, \lambda x \rangle = \langle x, Ax \rangle = \langle Ax, x \rangle = \langle \lambda x, x \rangle = \bar{\lambda} \langle x, x \rangle = \bar{\lambda},$$

also $\lambda \in \mathbb{R}$ sein muss. Dass es dazu im Fall $A \in \mathbb{R}^{n \times n}$ auch einen Eigenvektor mit reellen Einträgen gibt, steht schon in 17.2.5(iii).
Zu (ii) sei $x \perp E(\lambda, A)$, d.h. $\langle x, u \rangle = 0$ für alle $u \in E(\lambda, A)$. Dann ist für alle diese u auch

$$\langle A \cdot x, u \rangle = \langle x, A \cdot u \rangle = \lambda \langle x, u \rangle = 0.$$

Für die Aussage (iii) seien λ_1, λ_2 verschiedene Eigenwerte von A und $u_1 \in E(\lambda_1, A), u_2 \in E(\lambda_2, A)$ jeweils zugehörige Eigenvektoren. Dann gilt mit dem komplexen Skalarprodukt $\langle \cdot, \cdot \rangle$ im symmetrischen wie im hermiteschen Fall, dass

$$\lambda_2 \langle u_1, u_2 \rangle = \langle u_1, \lambda_2 u_2 \rangle = \langle u_1, Au_2 \rangle = \langle Au_1, u_2 \rangle = \langle \lambda_1 u_1, u_2 \rangle = \overline{\lambda_1} \langle u_1, u_2 \rangle = \lambda_1 \langle u_1, u_2 \rangle,$$

wobei der letzte Schritt folgt, da λ_1 nach (i) reell ist. Damit gilt $(\lambda_1 - \lambda_2) \cdot \langle u_1, u_2 \rangle = 0$, also wegen $\lambda_1 - \lambda_2 \neq 0$ und der Nullteilerfreiheit in $\mathbb{K}$ die Aussage $\langle u_1, u_2 \rangle = 0$. $\square$

Beweis von Satz 18.3: Sei $A \in \mathbb{K}^{n \times n}$ eine gegebene selbstadjungierte Matrix mit Einträgen in $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$. Das charakteristische Polynom von A zerfällt dann über $\mathbb{C}$ in Linearfaktoren, die den Eigenwerten $\lambda_1, \ldots, \lambda_k \in \mathbb{C}$ von A entsprechen (mit entsprechenden algebraischen Vielfachheiten in der Zerlegung von p_A). Nach Lemma 18.3.3(i) gilt $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$. Wir betrachten wieder die (direkte) Summe $Eig(f_A) = \oplus_{i=1}^{k} E(\lambda_i, A)$ der Eigenräume von A (bzw. der zugehörigen Matrixabbildung f_A; Wir zeigen nun (a) $V = \text{Eig}(f_A)$ – dann ist A nach 18.2.2(iii)$\Rightarrow$(i) diagonalisierbar –, und konstruieren (b) eine Orthonormalbasis Q von $\text{Eig}(f_A) = V$, die nur aus Eigenvektoren von A besteht.

Für (a) stellen wir fest, dass $\text{Eig}(f_A) = V$ äquivalent zur Aussage $(\text{Eig}(f_A))^\perp = \{0_V\}$ ist und führen die Annahme $(\text{Eig}(f_A))^\perp \neq \{0_V\}$ zum Widerspruch: In diesem Fall wäre die Abbildung

$$f^\perp : \quad (\text{Eig}(f_A))^\perp \to (\text{Eig}(f_A))^\perp, \qquad x \mapsto A \cdot x$$

linear und nach Lemma 17.1.6(ii) wohldefiniert ist (d.h. der Wertebereich ist wie angegeben $(\text{Eig}(f_A))^\perp$. Da $f^\perp$ ein Endomorphismus von $(\text{Eig}(f_A))^\perp$ ist, hat $f^\perp$ also einen Eigenwert $\mu \in \mathbb{C}$ mit zugehörigem Eigenvektor $v \in (\text{Eig}(f_A))^\perp \subseteq \mathbb{K}^n$, für den $v \neq 0_V$ gelten muss. Nach Definition von $f^\perp$ gilt dann aber auch $A \cdot v = \mu \cdot v$, also ist v auch Eigenvektor von A und daher $\mu \in \mathbb{R}$, und damit $v \in \text{Eig}(f_A)$ *und* $v \in (\text{Eig}(f_A))^\perp$. Daraus folgt $\langle v, v \rangle = 0$ und somit $v = 0_V$ im Widerspruch dazu, dass v ein Eigenvektor von A ist, also insbesondere $v \neq 0_V$ gelten muss.

Zu (b) wählen wir (nach Satz 14.4.2) für jedem der Eigenräume $E(\lambda_i)$ eine Orthonormalbasis Q_i. Die Vereinigung $Q = \cup_{i=1}^{k} Q_i$ ist nach Lemma 17.1.6(iii) eine Basis von V, die nur aus Eigenvektoren besteht. Da Eigenvektoren aus verschiedenen Eigenräumen nach Lemma 18.3.3(iii) aufeinander senkrecht stehen, ist Q orthonormal. $\square$

Bemerkung dazu:

- Die Diagonalisierbarkeit selbstadjungierter Matrizen lässt sich auch so formulieren: Symmetrische Matrizen beschreiben gerade diejenigen linearen Abbildungen, die gerade einfach einer Umskalierung $u_i \mapsto \lambda_i \cdot u_i$ der Achsen eines geeignet gewählten kartesischen Koordinatensystems (nämlich gerade der Eigenbasis Q von A) entsprechen. Wichtig ist hierbei, dass auch Werte $\lambda_i < 0$ oder $\lambda_i = 0$ möglich ist.

Diese Ergebnisse zur Diagonalisierbarkeit selbstadjungierter Matrizen werden wir in den diese Vorlesung abschließenden Abschnitten ausnutzen, um noch einiges über Matrizen und so genannte algebraische Gleichungen zu erfahren. Zunächst schauen wir uns aber noch einige besondere „Unterklassen“ selbstadjungierter Matrizen an, die oft, insbesondere auch in der Analysis, von Bedeutung sind.

18.4 Definite Matrizen

18.4.1 Definitheit und Eigenwerte

In Abschnitt 14.6 hatten wir gesehen, dass Gram'sche Matrizen, also die Matrizen, die über die Zuordnung

$$\mathbb{K}^n \times \mathbb{K}^n \to \mathbb{K}, \qquad (x, y) \mapsto x^T A y$$

ein Skalarprodukt definieren, besondere symmetrische/hermitesche Matrizen sind – nämlich solche, die zusätzlich *positiv definit* sind. Allgemeiner definieren wir die folgenden Klassen von symmetrischen Matrizen:

Definition 18.4.1. *(Definitheit von Matrizen)*

Sei $\langle , \rangle$ das Standardskalarprodukt auf $\mathbb{K}^n$ und $A \in \mathbb{K}^{n\times n}$ symmetrisch bzw. hermitesch. A heißt

(i) positiv (semi-)definit (kurz $A > 0$ bzw. $A \geq 0$, falls $\langle Ax, x\rangle > 0$ (bzw. $\langle Ax, x\rangle \geq 0$) für alle $x \in \mathbb{K}^n$ mit $x \neq 0_V$ gilt,

(ii) negativ (semi-)definit (kurz $A < 0$ bzw. $A \leq 0$), falls $\langle Ax, x\rangle < 0$ (bzw. $\langle Ax, x\rangle \leq 0$) für alle $x \in \mathbb{K}^n$ mit $x \neq 0_V$ gilt,

(iii) indefinit, falls A weder positiv noch negativ semidefinit ist.

Die Frage, die sich schon im Kontext unserer Beschäftigung mit Skalarprodukten stellte, war die, woran man die positive Definitheit einer gegebenen Matrix erkennen kann. Die meist einfachste Antwort lautet:

Lemma 18.4.2. *(Definitheit und Eigenwerte)*

Eine symmetrische/hermitesche Matrix $A \in \mathbb{K}^{n \times n}$ ist genau dann positiv definit, wenn sie nur positive Eigenwerte hat.

Analog ist A negativ definit, wenn sie nur negative Eigenwerte hat, semi-(...) wenn zusätzlich $\lambda = 0$ als Eigenwert erlaubt ist usw.
Insbesondere ist A indefinit genau dann, wenn A einen positiven und einen negativen Eigenwert besitzt.

Beweis: Sei A selbstadjungiert. Es bezeichne $\lambda_1, \dots, \lambda_n$ die Eigenwerte von A und $(q_1, \dots, q_n)$ eine zugehörige Orthonormalbasis. Hat man eine Definitheit von A, also eine Bedingung $\langle Ax, x \rangle$, gegeben, so folgt daraus wegen $\langle Aq_i, q_i \rangle = \lambda_i$ die entsprechende Bedingung für die Eigenwerte. Umgekehrt ist für alle $x = \sum_{i=1}^n c_i q_i \in \mathbb{R}^n$ dann $\langle Ax, x \rangle = \sum_{i=1}^n \lambda_i c_i^2$, hat man also eine Bedingung an die Eigenwerte gegeben, so folgt unmittelbar die entsprechende Definitsheitsbedingung.

□

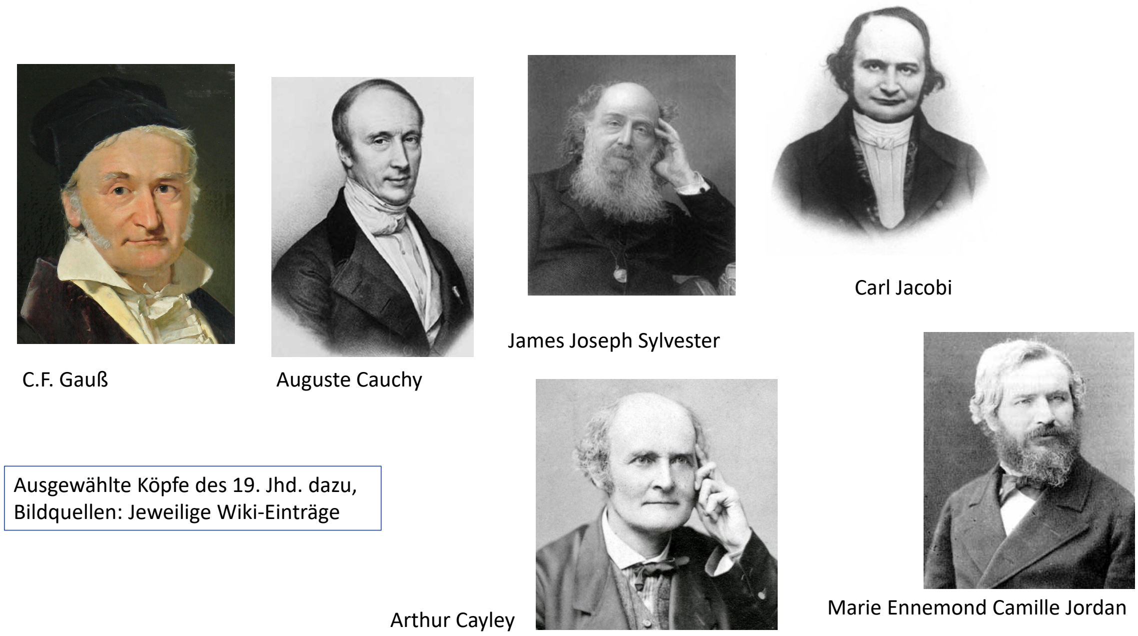
18.4.2 Skalarprodukte, revisited

In Lichte des vorangegangenen Lemmas lässt sich ein Skalarprodukt also immer durch den folgenden Prozess definieren/anschaulich interpretieren:

- Wähle eine Orthonormalbasis $Q = (q_1, \dots, q_n)$ des $\mathbb{R}^n$, also „die Achsen eines kartesischen Koordinatensystems“.
- „Eiche“ jeden der Vektoren q_i durch Festlegung seiner „Länge“, d.h. einer Zahl $\|q_i\| := \ell_i > 0$.
- Sei D die Diagonalmatrix mit den Einträgen $\ell_1, \dots, \ell_n$, kurz: $D = \operatorname{diag}(\ell_1, \dots, \ell_n)$. Die Gramsche Matrix des konstruierten Skalarprodukts (bezüglich der Standardbasis) ist dann $G = QDQ^T$.
- Winkel- und Längenmessungen funktionieren dann „wie in der klassischen Geometrie“, wenn man als kartesisches Koordinatensystem das durch Q festgelegte nimmt.

18.5 Lineare Algebra und Matrixzerlegungen

18.5.1 Überblick: Lineare Gleichungssysteme, Eigenwerte und Rechnen mit Matrizen



- Lineare Gleichungssysteme in Europa: Erste bedeutende Werke von Gottfried Wilhelm Leibniz (17. Jhd.), Gabriel Cramer (18. Jhd.)
- Carl Friedrich Gauß (1777-1857) entwirft unglaublich viele Methoden zur Lösung von Problemen der linearen Algebra, benutzt dafür aber selbst noch keine Matrizen! (damit beschreibt er nur geometrische Abbildungen)
- Allgemein werden Ergebnisse zu Linearen Gleichungssystemen und umliegenden Problemstellungen eher mit Hilfe von diesen Systemen, Determinanten, Bilinearformen und Quadratischen Formen (s.u.) beschrieben und gelöst
- Erste Beschäftigung mit Eigenwertproblemen für „kleine“ Systeme ($2 \times 2, 3 \times 3$) bei Leonhard Euler (1748), Lagrange (1788)
- Ab Anfang 19. Jahrhundert durch Anwendungen motivierte Beschäftigung mit größeren Eigenwertproblemen, z.B. für symmetrische Systeme
- Augustin-Louis Cauchy (1829): Eigenschaften von Eigenvektoren und Eigenwerten symmetrischer Systeme (noch nicht: Matrizen);

- Carl Gustav Jacob Jacobi (1848): Diagonalisierungsalgorithmus für asymmetrische Systeme
- Diagonalisierbarkeit von Systemen mit n verschiedenen Eigenwerten (Lemma 18.2.1): Riemann 1857, „von Jacobi gelernt“
- James Sylvester prägt den Begriff „Matrix“, bearbeitet damit aber nicht wirklich Probleme der Linearen Algebra
- Arthur Cayley benutzt ca. 1850 das erste Mal Matrizen, um lineare Gleichungssysteme damit auszudrücken, und untersucht auch die algebraische Strukturen der Verknüpfung von Matrizen (z.B. Nichtkommutativität, und zeigt eigentlich, dass diese einen nichtkommutativen Ring mit Eins bilden (ohne dass dieser Begriff dafür schon existiert), Klassifizierung in verschiedene Typen wie orthogonale, symmetrische,..., hauptsächlich motiviert durch die Untersuchung von Eigenwerten.
- Ab da ist der Siegeszug der Matrixdarstellung nicht mehr aufzuhalten. Die obigen Ergebnisse werden „rückwirkend“ in die handliche Form der Matrixdarstellung gebracht.
- Spätestens ab hier spielen auch die Matrixzerlegungen eine große Rolle, wobei die ersten Ideen, Aussagen und Beweise dazu wieder bei Gauß zu finden sind (nur eben ohne Matrizen).
- Jordan (1870): Allgemeine Zerlegung von Matrizen $A \in \mathbb{C}^{n\times n}$ in verallgemeinerte Eigenräume („Jordan-Zerlegung“, s.u., von ihm ursprünglich über endlichen Körpern entwickelt!)

18.5.2 Matrixzerlegungen

Die Beschäftigung mit linearen Gleichungssystemen, das wurde hoffentlich gerade klar, ist viel älter als ihre Darstellung mit Hilfe von Matrizen. Genauso ist die Frage nach der Diagonalisierbarkeit von Matrizen älter als die Frage nach so genannten Matrixzerlegungen. Nennt man aber eine Darstellung einer Matrix A als Summe, Produkt anderer Matrizen Matrixzerlegung (ersteres genauer Matrixfaktorisierung), so ist z.B. die Diagonalisierung

$$A = CDC^{-1}$$

eine solche Matrixzerlegung. Man stellt fest, dass sich viele wichtige Verfahren der Linearen Algebra als Zerlegung der betrachteten Matrix A interpretieren lassen. Zunächst ein paar Beispiele aus Kontext der Gleichungslösungen und anderen Bereichen, ohne groß ins Detail zu gehen:

- Die Zerlegung $A = LR$, $A = LL^T$ in obere (R) und untere (L) Dreiecksmatrix entspricht dem Gauß-Algorithmus,
- $A = QR$ in obere Dreiecks-(R) und orthogonale Matrix Q gibt das Gram-Schmidt-Verfahren wieder
- $A = D+L+R$ ist eine additive Zerlegung in obere (R) und untere (L) Dreiecksmatrix und Diagonalmatrix D die bei klasischen Iterationsverfahren zur Glecihungslösung, z.B. dem Gauß-Seidel- und dem Jacobi-Verfahren vorkommt,
- Eine Polarzerlegung $A = UP$ mit U unitär und P positiv semidefinit exisitiert für alle quadratischen Matrizen; dies liefert eine Analogie zur Darstellung $z = e^{i\phi}|z|$ für komplexe Zahlen!
-

Beispiele aus dem Kontext der „Strukturtheorie für Matrizen“, d.h. in denen es um Darstellung der strukturellen Eigenschaften der zugehörigen linearen Abbildung geht, sind die Zerlegungen

- $A = CDC^{-1}$ (ex. nur, falls diagonalisierbar)
- $A = QDQ^T$ für symmetrische Matrizen
- Es lässt sich zeigen, dass jede Matrix über $\mathbb{C}$ trigonalisierbar ist, d.h. $A = QRQ^T$ mit oberer Dreiecksmatrix R. Genauer ist über allgemeinen Körpern A trigonalisierbar gdw. p_A in Linearfaktoren zerfällt.

- Jordan-Normalform (1870, parallel von Karl Weierstrass entwickelt):

$$A = CJC^{-1} = C(J_1 \dots J_K)C^{-1} = A$$

  (komplette Zerlegung des $\mathbb{R}^n$ in Haupträume ($\widehat{=}$ Jordanblöcke in der Jordan-Matrix J), besonders „einfache" Trigonalform.

Die als letztes erwähnte „Jordan-Zerlegung" lieferte Ende des 19. Jahrhunderts die „Komplettlösung" der Frage, was eine sinnvolle Verallgemeinerung der Diagonalisierung $A = CDC^{-1}$ sein kann. Eine solche Jordan-Zerlegung existiert für alle quadratischen Matrizen. In dieser und allen anderen genannten Zerlegungen lässt sich aber eine allgemeine Matrix(-abbildung)nicht so schön deuten wie bei der simplen „Durchskalierung" einer orthogonalen Eigenbasis (eines „kartesischen Koordinstensystems") mit den verschiedenen Eigenwerten, die man im symmetrischen Fall in der Form

$$A = QDQ^T \qquad (\star)$$

erhalten hatte.

Um auch für nicht-symmetrische Matrizen immer eine Zerlegung zu erhalten, verfolgt der durch Jordan beschrittene Weg die Idee, die Diagonalmatrix D entsprechend anzupassen (in der Jordan-Zerlegung durch eine „Block-Diagonalmatrix" aus so genannten Jordankästchen). Die Jordanzerlegung gibt dabei auch die Orthogonalität der zugehörigen Eigenbasis C in der Zerlegung $A = CJC^{-1}$ auf. Im folgenden sehen wir uns die andere Möglichkeit an: Für nicht-symmetrische Matrizen auch die Symmetrie der Darstellung $(\star)$ zu relaxieren, indem man links und rechts *verschiedene* Orthonormalbasen zulässt. Das führt uns zur Singulärwertzerlegung, die immer, auch für nichtquadratische Matrizen existiert und für die Theorie der Linearen Algebra und der linearen Abbildnugen den krönenden Abschluss dieses Skripts bilden soll.

18.6 Die Singulärwertzerlegung

- Vielfach erfunden, neu entdeckt und erweitert: Eugenio Beltrami (1873), Camille Jordan (erster „echter“ Beweis, 1874), Sylvester (1889), Erhard Schmidt (1907, für Operatoren), Herrmann Weyl (1912, für Näherungsverfahren), Gene Golub, Charles van Loan (numerisch effiziente und stabile Berechnung)

Hauptfinder der SVD

Eugenio Beltrami

Marie Ennemond Camille Jordan

Erhard Schmidt

Gene Golub,
Charles Van Loan

Bildquellen: Jeweilige Wiki-Einträge

18.6.1 Existenz einer Singulärwertzerlegung

Definition 18.6.1. *Sei $A \in \mathbb{R}^{m\times n}$. Dann heißt eine Zerlegung*

$$A = U \cdot \Sigma \cdot V^T$$

von A in drei Matrizen $U \in \mathbb{R}^{m\times m}$, $\Sigma \in \mathbb{R}^{m\times n}$ und $V \in \mathbb{R}^{m\times m}$ eine Singulärwertzerlegung von A, falls

- *U und V orthogonal sind und*
- *$\Sigma = (s_{i,j})$ eine (für $m \neq n$ rechteckige) Diagonalmatrix ist (d.h. es ist $s_{i,j} = 0$ für alle $i \neq j$),*
- *für deren Diagonaleinträge $\sigma_i := s_{i,i}$ zusätzlich $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_k \geq 0$ für $k = \min\{m, n\}$ gilt.*

Die (nicht notwendigerweise voneinander verschiedenen) Diagonaleinträge $\sigma_1, \ldots, \sigma_r$, für die $\sigma_i > 0$ gilt, heißen die Singulärwerte von A.

Bemerkungen dazu:

- Im Fall $n < m$ hat die Matrix Σ mehr Zeilen als Spalten also Null*zeilen unter* dem quadratischen Diagonalblock $\mathrm{diag}(\sigma_1, \dots, \sigma_n)$, im Fall $n > m$ Null*spalten rechts* der Diagonalen. Wir decken im Folgenden beide Fälle ab.

Entscheidend ist nun:

Satz 18.6.2. *Sei $A \in \mathbb{R}^{m \times n}$. Dann existiert eine Singulärwertzerlegung von A.*

Bevor wir zum Beweis kommen, wollen wir den obigen Satz in verschiedenen äquivalenten Formulierungen wiedergeben: Der obige Satz bedeutet für jede lineare Abbildung $f : \mathbb{R}^n \to \mathbb{R}^m, x \mapsto A \cdot x$:

- Es gibt ein Orthonormalsystem V des $\mathbb{R}^n$, das durch f auf ein skaliertes Orthogonalsystem U des $\mathbb{R}^m$ abgebildet wird. (Die Skalierung ist gerade Multiplikation mit den Singulärwerten aus Σ).
- Es gibt „zwei orthogonale Eigenbasen“ $U = (u_1, \dots, u_n), V = (v_1, \dots, v_m)$ so, dass
$$f(u_i) = \sigma_i \cdot v_i$$
für alle $i \leq \min n, m$ gilt (für die anderen Indizes i ist $f(v_i) = 0$). Diese werden tatsächlich oft als Links- und Rechtseigenvektoren von A bezeichnet.
- Es gibt zwei „verallgemeinerte orthogonale Eigenbasen“ $U = (u_1, \dots, u_n), V = (v_1, \dots, v_m)$ gibt, so dass die darstellende Matrix von f bezüglich V, U diagonal ist (wenn auch rechteckig); es ist nach Lemma 10.2.1 ja $A = U \cdot A_{V,U} \cdot V^T$, also Σ gerade die darstellende Matrix von f bezüglich V und U.
- f lässt sich immer lesen als Transformation eines Vektors des $\mathbb{R}^n$ auf seine Koordinaten bezüglich V, dann Durchskalieren mit Singulärwerten, um daraus Bildkoordinaten bzgl. U zu erhalten; die Multiplikation mit U transformiert das Ergebnis wieder in den $\mathbb{R}^n$ in Standarddarstellung.

Beweis: Sei $A \in \mathbb{R}^{m \times n}$. Zum Beweis konstruieren wir konkret die Matrizen U, Σ und V und zeigen, dass sie das Gewünschte liefern. Dazu betrachten wir die Matrix $A^T A \in \mathbb{R}^{n \times n}$: Wegen $(A^T A)^T = A^T (A^T)^T = A^T A$ ist diese Matrix symmetrisch und ist daher nach Satz 18.3.1 reell diagonalisierbar, d.h. $A^T A$ besitzt n reelle Eigenwerte (wobei wir mehrfache Eigenwerte gemäß ihrer algebraischen Vielfachheit zählen). Diese lassen sich der Größe nach sortieren lassen zu $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. Zur Diagonalmatrix $D := \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ gibt es (wieder mit Satz 18.3.1) eine zugehörige orthogonale Matrix $V \in \mathbb{R}^{n \times n}$ so, dass wir $A^T A = V D V^T$ schreiben können. Sind v_i, $i \in \underline{n}$, die Spalten von V (also jeweils die normierten Eigenvektoren zum Eigenwert λ_i von $A^T A$), so ist

$$\lambda_i = \lambda_i \langle v_i, v_i \rangle = \langle v_i, A^T A v_i \rangle = v_i^T A^T A v_i = (A v_i)^T A v_i = \langle A v_i, A v_i \rangle = \|A v_i\|^2 \geq 0;$$

insbesondere existiert ein $r \leq n$ so, dass

$$\lambda_1 \geq \lambda_2 \geq \ldots \lambda_r > \lambda_{r+1} = \ldots = \lambda_n = 0. \qquad (\star)$$

Wir konstruieren mit Hilfe dieser Erkenntnisse nun die Matrizen U und Σ. Dazu berechnen wir die Wirkung von A auf die Orthonormalbasis V, um ein Erzeugendensystem des Bildes von A zu erhalten (vgl. LAAG I): Es ist $A \cdot V = (A v_1, \ldots, A v_n)$. Die Vektoren $A v_i$, $i \in \underline{n}$ sind wegen

$$\langle A v_i, A v_j \rangle = \langle v_i, A^T A v_j \rangle = \langle v_i, \lambda_j v_j \rangle = \lambda_j \delta_{i,j}$$

orthogonal. Weiter ist mit der Sortierung $(\star)$ dann

$$A \cdot V = (A v_1, \ldots, A v_r, 0_V, \ldots, 0_V).$$

Wir halten kurz inne und sehen, dass

- die Vektoren $v_{r+1}, \ldots, v_n \in \mathbb{R}^n$ eine Basis des Kerns von A sind,
- die Vektoren $A v_1, \ldots, A v_r$ eine Basis des Bildes von A sind und
- die Zahl r gerade der Rang von A ist.

In der obigen Rechnung steht auch, dass für $i \in \underline{n}$ dann $\|A v_i\| = \sqrt{\lambda_i}$ ist; wir bezeichnen diese Werte mit $\sigma_i := \sqrt{\lambda_i}$, das werden die gesuchten Singulärwerte von A sein. Die orthogonale Matrix U definieren wir nun, indem wir zunächst für alle $i \in \underline{r}$ die Bildvektoren $A v_i$ normieren zu $u_i := \sigma_i^{-1} A v_i$ – das gibt zunächst $r \leq n$ Vektoren $u_1, \ldots, u_r$, die eine Orthonormalbasis des Bildes von A sind –, und diese dann gegebenenfalls noch zu einer

Orthonormalbasis $U = (u_1, \dots, u_m)$ des $\mathbb{R}^m$ ergänzen. Mit dieser Definition gilt dann

$$\begin{aligned} A \cdot V &= (Av_1, \dots, Av_n) = (Av_1, \dots, Av_r, 0_V, \dots, 0_V) \\ &= (\sigma_1 u_1, \dots, \sigma_r u_r, 0_V, \dots, 0_V) \qquad (\star). \end{aligned}$$

Definieren wir nun die Matrix $\Sigma := (s_{ij}) \in \mathbb{R}^{m \times n}$ mit Einträgen

$$s_{ij} = \begin{cases} \sigma_i & \text{falls } i = j \\ 0 & \text{sonst} \end{cases}$$

so sieht man z.B. durch Ausmultiplizieren der folgenden Matrizen, dass auch

$$U \cdot \Sigma = (\sigma_1 u_1, \dots \sigma_1 u_r, 0_V, \dots, 0_V) = U \cdot \Sigma$$

ist. Zusammen mit der Zeile darüber ergibt sich daraus $A \cdot V = U \cdot \Sigma$, also durch Multiplikation der Gleichung mit der orthogonalen Matrix V^T von rechts die Darstellung $A = U \Sigma V^T$.

$\square$

Die Beobachtungen aus dem Beweis sind in der folgenden schematischen Darstellung festgehalten. Im darauf folgenden Lemma halten wir sie noch einmal separat fest und ergänzen sie ohne Beweis um einige der vielen theoretischen Eigenschaften der SVD.

Satz XV.6.2. *Sei $A \in \mathbb{R}^{m \times n}$. Dann existiert eine Singulärwertzerlegung von A.*

Fazit:

Ist $A = U \Sigma V^T$, so ist

$$A = \begin{pmatrix} \underbrace{u_1 \cdots u_r}_{\text{ONB von Bild}(f_A)} & \underbrace{u_{r+1} \cdots u_m}_{\text{erg.}} \end{pmatrix} \cdot \underbrace{\begin{pmatrix} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_r & \\ & 0 & & 0 \end{pmatrix}}_{\text{Skalierungsfaktoren}} \cdot \begin{pmatrix} \underbrace{v_1 \cdots v_r}_{Av_i = \sigma_i u_i} & \underbrace{v_{r+1} \cdots v_n}_{\text{Basis von Kern}(f_A).} \end{pmatrix}^T$$

und $Au_i = \sigma_i v_i$ für $i = 1, \dots, r$

Lemma 18.6.3. *(Einige theoretische Eigenschaften der SVD)*

Sei $A = U\Sigma V^T$ eine Singulärwertzerlegung von $A \in \mathbb{R}^{n\times m}$.

(i) *Die Anzahl r der von Null verschiedener Singulärwerte von A ist gerade der Rang von A.*

(ii) *Für die durch A definierte Matrixabbildung f_A ist $(u_1, \ldots, u_r)$ eine Basis des Bildes von f_A und $(v_{r+1}, \ldots, v_n)$ eine Basis des Kerns von f_A.*

(iii) *Bei A^T ist es genau umgekehrt: Wegen $A^T = V\Sigma^T U^T$ kriegt man die SVD von A^T geschenkt, und es ist $(v_1, \ldots, v_r)$ eine Basis des Bildes von f_{A^T} und $(u_{r+1}, \ldots, u_m)$ eine Basis des Kerns von f_{A^T}.*

(iv) *Für $i \le r$ ist für die obigen Vektoren $Au_i = \sigma_i v_i$ (und für $i > r$ ist $Au_i = 0_V$).*

(v) *Ist $r < \min\{n, m\}$, so lässt sich die Darstellung auch als sparsamere „Thin SVD“ schreiben, d.h. als Zerlegung $A = U'\Sigma'(V')^T$ mit $U' \in \mathbb{R}^{m\times r}, \Sigma' \in \mathbb{R}^{r\times r}, V' \in \mathbb{R}^{m\times r}$.*

(vi) *Strukturelle „Zerlegung“ der Matrixabbildung, zu jeder Matrixabbildung A gibt es ein Orthonormalsystem, das durch A elementweise entweder injektiv auf Elemente des Bild-Orthogonalsystem oder auf den Nullvektor abgebildet wird.*

(vii) *Sind dann wie oben u_i bzw. v_i die Spalten von U bzw. V sind, so ist*

$$A = \sum_{k=1}^{r} \sigma_k u_k v_k^T.$$

(viii) *„Kreise werden unter linearen Abbildungen zu Ellipsen.“*

(ix) *Die Pseudoinverse oder Penrose[1]-Inverse A^+ von A ist folgendermaßen definiert: Ist $A = U'\Sigma'(V')^T$ eine Thin SVD von A (s. (v)), so ist $A^+ := (V')\Sigma^+(U')^T$, wobei die Diagonaleinträge von Σ^+ gerade die Kehrwerte derer von Σ' sind. Durch A^+ wird A „auf dem Bild invertiert“ (d.h. als Abbildung $\mathcal{L}(u_1, \ldots, u_r) \to \mathcal{L}(v_1, \ldots, v_r)$), fortgesetzt durch die Nullabbildung auf Bild$(A)^\perp$.*

(x) *Ist A invertierbar, so ist $A^+ = A^{-1}$; ist A reell und orthogonal diagonalisierbar, so sind die Singulärwerte von A gerade die Beträge der Eigenwerte von A.*

[1] Roger Penrose: Englischer Mathematiker; Professor am Birkbeck College in London (1964) und seit 1973 Professor an der Universität Oxford; fundamentale Beiträge in der Mathematik, Kosmologie, Relativitäts- und Quantentheorie, frischer Nobelpreisträger (2020)

18.6.2 Lineare Algebra als Anwendungswissenschaft am Beispiel der SVD

Die Lineare Algebra ist über die Disziplinen hinweg das *working horse* in vielen praktischen Anwendungen. Wie schon im letzten und vorletzten Jahrhundert lassen sich Probleme oft mit Hilfe von linearen Gleichungssystemen, Matrizen und Eigenwertproblemen usw. Die Analyse der speziellen mathematischen Struktur dieser Probleme führt dann oft auf eine entsprechende genauere Klassifikation (z.B. von Typen von Matrizen), von denen die Lösungstheorie und praktische Anwendbarkeit oftmals entscheidend abhängt.

Die Breite des Anwendungspektrums will ich hier anhand meines Favourites (hat man vielleicht schon gemerkt), der im letzten Kapitel entwickelten Singulärwertzerlegung, illustrieren. Wir beginnen mit der Aufzählung einiger der praktischen Anwendungsgebiete der SVD. Dazu gehört – wie immer im interdisziplinären Feld – dass die SVD in so gut wie jeder anwendenden Wissenschaft einen eigenen Namen trägt. Das ist leider ein weit verbreitetes Charakteristikum von erfolgreichen Anwendungen: Oft ähneln sich Verfahren in verschiedenen Fachgebieten sehr, aber das zu erkennen, wenn diese unterschiedlich heißen, ist tatsächlich oftmals das Hauptproblem!)

Ein paar Anwendungsfelder und ihre Bedeutung für die Disziplinen:

- Singulärwertzerlegung, SVD (Mathematik)
- direkter Zugang zu wichtigen Kenngrößen von Matrizen (Kern, Bild, Rang; Kondition, Matrixnorm,...)
- = Schmidt-Entwicklung, Schmidt decomposition (Physik)
- A^+ löst Bestapproximationsproblem und Ausgleichsprobleme (die *immer* im Umgang mit Messdaten auftreten, vgl. z.B. CT-Beispiel in der LAAG I)
- = Karhunen-Loeve-Entwicklung oder -Transformation (Stochastik und Anwendungen stochastischer Prozesse)
- = Hotelling-Transformation (Bildverarbeitung und Kompression)
- = Principal Component Analysis (PCA, „Hauptkomponentenanalyse“; Statistik und Datenanalyse)
- = Berechnung „verallgemeinerter Eigenwerte“, der „Links- und Rechts-Eigenvektoren von A“[2] in vielen Anwendungen

[2] Vorsicht! Verallgemeinerte Eigenräume gibt es auch, die sind aber etwas anderes.

18.6.3 Anwendungen der SVD - Zwei genauere Beispiele

Die obige Liste macht das breite Anwendungsspektrum der SVD wohl deutlich. Bedeutsam sind hierbei neben den Eigenschaften aus Korollar auch die folgenden vorteilhaften Eigenschaften der SVD:

- Abgeschnittene („truncated“) SVD

  $$A \approx \sum_{k=1}^{s} \sigma_k u_k v_k^T$$

  (für $s << r$) liefert Beste Rang-s-Approximation von A in Spur(-Frobenius-) und euklidischer 2-Norm (Kompression von A).

- Rang-s-Kompression von Matrizen hat vielseitige Anwendungen, u.a. Datenanalyse und Mustererkennung, Stabilisierung und Beschleunigung (Vorkonditionierung) von Gleichungslösern, z.B. „robuste Regler“ in der Regelungstechnik, automatisierte Modellreduktion, ...

- Berechnung der SVD, der „thin SVD“ und der Rang-k-Approximation: Numerisch stabil mit den Golub-Van-Loan-Algorithmus (1960er).

Anhand von zwei Beispielen bekommen wir einen kleinen Einblick, wie die einzelnen Punkte gemeint sind bzw. wie die angegebenen Methoden grob funktionieren:

- Die Singulärwertzerlegung liefert eine einfache Methode zur Bildkompression, vgl. z.B. [Demmel].

- Wir schauen uns ein Beispiel zur Mustererkennung an, die Erkennung von handgeschriebenen Ziffern (vgl. [Elden])

---

19 „Nichtlineare Algebra“ – Algebraische Kurven und Flächen höherer Ordnung

Nach diesem eher praktisch angehauchten Ende des letzten Abschnitts kehren wir nun noch einmal zurück zu rein algebraischen Fragestellung, d.h. Fragen, in denen es um strukturelle Erkenntnisse über das Lösen bestimmter Typen von Gleichungen geht. Die *Lineare* Algebra wird in diesem Kontext oft als die Wissenschaft der Linearen Gleichungssysteme bezeichnet. So gesehen beschäftigt sie sich mit (Systemen von) den einfachsten Gleichungen in Körpern $\mathbb{K}$, die die Mathematik so zu bieten hat,

$$f(x_1, \ldots, x_n) = a_1x_1 + \ldots + a_nx_n - b = 0 \qquad (\star)$$

und ihrer geometrischen Deutung. Diese Fragestellung lässt sich in verschiedene Richtungen weiterdenken: Z.B. eine Lösungstheorie für Koeffizienten und Unbekannte, die aus allgemeinen/speziellen Körper $\mathbb{K}$ kommen (die z.T. von unserer Vorlesung mit abgedeckt wurde), oder Erweiterung der Fragestellung nach einer Lösungstheorie auf Gleichungen höheren Grades, wenn man f in $(\star)$ als Polynom ersten Grades über dem Körper $\mathbb{K}$, aber in n Variablen interpretiert.

Letzteren Gedanken wollen wir hier zum Abschluss dieses Skripts noch ein wenig verfolgen; wir werden sehen, dass auch hier die Objekte und Methoden der Linearen Algebra hilfreich und in der Theorie allgegenwärtig sind.

Dabei gehören wie oft in der Linearen Algebra und Analytischen Geometrie zu einer algebraischen Gleichung für gewöhnlich auch ein oder mehrere geometrische Probleme, die durch die Gleichungen beschrieben werden:

- Bisher haben wir uns mit Kurven/Flächen *erster* Ordnung beschäftigt; das sind solche, die durch eine lineare Gleichung in den Unbekannten beschrieben werden:
  - Die Unbekannten treten nur in in erster Potenz auf
  - Lösungen: Gerade in $\mathbb{R}^2$, Ebene in $\mathbb{R}^3$
  - Interpretierbar als vorgegebenes Skalarprodukt mit dem Vektor $(a_1, \ldots, a_n)$ (oder mehrere Skalarprodukte mit mehreren Vektoren des $\mathbb{K}^n \rightsquigarrow$ LGS)
  - Geometrisch: interpretierbar als Lösungsmenge von Schnittproblemen zwischen affinen Teilräumen
- Kurve/Fläche *zweiter Ordnung*, so genannte *Quadriken*, sind solche, die durch eine quadratische Gleichung in den Unbekannten beschrieben werden:

- In jedem der einzelnen Summanden treten Variablen insgesamt höchstens zweimal als Faktor auf, also z.B. $x_1^2 + 2x_1x_2 + 5 = 0$ (in ersten Summanden tritt die Variable x_1 zweimal auf, Summanden der Form $x_1^2x_2$ oder x_1^3 wären nicht erlaubt.
- Geometrische Schnittprobleme, z.B. Kreis-Kreis, Kreis-Ebene, Zylinder-Ebene und auch die so genannten Kegelschnitte (Schnitt Doppelkegel - Ebene), die die Menschen in Alltag und Kunst verfolgen, führen auf Quadriken in zwei Unbekannten, also quadratische Gleichungen im $\mathbb{R}^2$.
- Für die Quadriken im $\mathbb{R}^2$ findet man hierbei für diese Schnittprobleme durch grafisches Probieren (z.B. mit GeoGebra oder einer ähnlichen Software) als Lösungen für manche Konstellationen zum Beispiel die klassischen Kegelschnitte Hyperbel, Ellipse und Parabel und „entartete“ oder „ausgeartete“ Sonderfälle[1].

Zur Theorie von Quadriken und ihren Lösungsmengen, den Kurven und Flächen zweiter Ordnung, insbesondere den Kegelschnitten, wollen wir nun einen Überblick verschaffen.

19.1 Quadratische Formen, quadratische Funktionen und quadratische Gleichungen im $\mathbb{R}^n$

Die Verallgemeinerung einer quadratischen Potenzfunktion in $\mathbb{R}$,

$$f : \mathbb{R} \to \mathbb{R}, \qquad x \mapsto a \cdot x^2$$

mit $a \in \mathbb{R}$ auf den $\mathbb{K}^n$ für $n > 1$ ist die zu einer quadratischen Form $f : \mathbb{K}^n \to \mathbb{K}$:

Definition 19.1.1. *(Quadratische Form)*
Sei $A \in \mathbb{K}^{n\times n}$ eine symmetrische Matrix. Die Abbildung

$$Q_A : \mathbb{K}^n \to \mathbb{K}, \qquad x \mapsto x^T A x$$

nennt man die zu A gehörige quadratische Form.

[1] Ich persönlich bevorzuge die Bezeichnung „Grenzfälle“, der Begriff der „Entartung“ hält sich aber in vielen Gebieten wacker als Terminus für Sonderfälle, in denen erwartete Eigenschaften verletzt sind.

Bemerkungen dazu:

- Eine zu $A \in \mathbb{K}^{n\times n}$ gehörige quadratische Form lässt sich auch für nicht-symmetrische quadratische Matrizen A definieren. Es lässt sich aber (durch geeignetes „Umverteilen“ der Einträge) zeigen, dass sich jede solche quadratische Form auch mit Hilfe einer *symmetrischen* Matrix $B \in \mathbb{K}^{n\times n}$ über $Q(x) = x^T Bx$ darstellen lässt.

- Gut darstellen lassen sich quadratische Formen für Matrizen $A \in \mathbb{R}^{2\times 2}$, für die man den Funktionswert $Q_A(x)$ als Höhe über der $x_1 - x_2-$Ebene plotten lässt. Hierbei wird deutlich, dass die Funktionen $x \mapsto x^T Ax$ eine Verallgemeinerung der quadratischen Funktion $f : \mathbb{R} \to \mathbb{R}, f(x) = ax^2$ darstellen. Abhängig von der Definitheit, also den Eigenwerten von A erhält man hier verschiedene Formen, s. Vorlesung.

Eine allgemeine quadratische Funktion im $\mathbb{R}^n$ lässt sich nun folgendermaßen definieren: Fügt man wie in $f(x) = ax^2 + bx + c$ noch einen linearen und einen konstanten Term hinzu, so erhält man als Analogon quadratischer Funktionen $f : \mathbb{R} \to \mathbb{R}$ Funktionen der Form

$$f : \mathbb{K}^n \to \mathbb{K}, \qquad x \mapsto x^T Ax + b^T x + c \tag{19.1}$$

(Lineare Funktionen $\mathbb{K}^n \to \mathbb{K}$ sind immer darstellbar durch Multiplikation mit einer Matrix $b^T \in \mathbb{K}^{1\times n}$, also durch den oben hinzugefügten Term $b^T x$, der dem Skalarprodukt mit dem Vektor $b \in \mathbb{K}^n$ entspricht. Die Konstante c muss eine einfache reelle Zahl sein.)

Dies führt für invertierbare Matrizen A anschaulich wie im eindimensionalen Fall zu einer Verschiebung des Graphen von $x \mapsto x^T Ax$, nur jetzt im Raum $\mathbb{R}^{n+1}$. Ist $\lambda = 0$ ein Eigenwert von A, so ist etwas Vorsicht angesagt, s.u.

Quadratische Gleichungen im $\mathbb{R}^n$ sind nun Gleichungen der allgemeinen Form

$$x^T Ax + b^T x + c = 0 \qquad (\star)$$

d.h. von unserer allgemeinen Form $f(x) = 0$, wobei f nun eine quadratische Funktion der eben diskutierten Form (19.1) ist. Für $\mathbb{K}^n = \mathbb{R}^3$ (mit symmetrischem $A \in \mathbb{R}^{3\times 3}$ und $b \in \mathbb{R}^3$) ergibt sich beispielsweise:

$$\begin{aligned} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \mapsto \qquad & a_{1,1}x_1^2 + a_{2,2}x_2^2 + a_{3,3}x_3^2 \\ & +2a_{1,2}x_1x_2 + 2a_{2,3}x_2x_3 + 2a_{1,3}x_1x_3 \\ & +b_1x_1 + b_2x_2 + b_3x_3 + c = 0. \end{aligned}$$

Die Lösungsmengen dieser verallgemeinerten quadratischen Gleichungen $(\star)$ sind die so genannten *Quadriken*, um deren Klassifizierung es in der Lösungstheorie der Gleichungen zweiter Ordnung geht:

> Welche möglichen Lösungsmengen gibt es für $(\star)$ überhaupt, und für welche Arten von A, b, c ergeben sich welche Möglichkeiten?

19.2 Alle Quadriken des $\mathbb{R}^2$

Die vollständige Klassifikation der Quadriken im $\mathbb{R}^2$ gelingt mit Hilfe von Isometrien des $\mathbb{R}^2$: Jede Quadrik der Form $(\star)$ oben lässt sich mit Hilfe einer affinen Isometrie (also einer längenerhaltenden Koordinatentransformation, d.h. einer Abbildung $\mathbb{R}^2 \to \mathbb{R}^2, x \mapsto Qx+t$ mit einer orthogonalen Matrix Q) auf eine der folgenden Formen bringen:

- Ist A regulär (d.h. invertierbar, null ist kein Eigenwert von A) , so lässt sich die Gleichung zu
$$\begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \lambda_1 x_1^2 + \lambda_2 x_2^2 = d,$$
umschreiben;

- Ist A singulär (d.h. nicht invertierbar, null ist Eigenwert von A), so lässt sich die Gleichung zu
$$\begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + b_2 x_2 = \lambda_1 x_1^2 + b_2 x_2 = d$$
umschreiben (hierbei wird $A \neq 0$ vorausgesetzt, ansonsten hätten wir keine „echte“ quadratische Gleichung.

Der Haupttrick dabei ist, dass die Basis Q für die Koordinatentransformation nach 18.3.1 als eine Basis aus Eigenwerten der symmetrischen Matrix A gewählt werden kann, so dass die transformierte darstellende Matrix wie oben diagonal ist.

Die Eigenwerte λ_1, λ_2 sind also genau die Eigenwerte der Matrix A in der zu lösenden quadratischen Gleichung, und diese bestimmen die Form der Lösungsmenge (also der Quadrik): Im $\mathbb{R}^2$ ergibt sich in den Koordinaten der Eigenbasis mit einem geeignet gewählten Ursprung durch „Durch-x-en“ der verschiedenen möglichen Fälle von Kombinationen von

λ_1 und λ_2, d und ggf. b_2 (jeweils gleich Null, kleiner Null oder größer als Null) immer eine der folgenden Normalformen von 2D - Quadriken als Lösungsmenge der obigen Gleichung, die sich nach der Definitheit der obigen Diagonalmatrix klassifizieren lassen (mit Zahlen $a, b > 0$):

definite Fälle, d.h. $\lambda_1, \lambda_2 > 0$ oder $\lambda_1, \lambda_2 < 0$:

- Ellipse: $ax_1^2 + bx_2^2 = 1$
- Punkt: $ax_1^2 + bx_2^2 = 0$
- leere Menge: $ax_1^2 + bx_2^2 = -1$

indefinite Fälle, d.h. $\lambda_1 > 0, \lambda_2 < 0$ oder umgekehrt:

- Hyperbel: $ax_1^2 - bx_2^2 = 1$
- Zwei sich schneidende Geraden: $ax_1^2 - bx_2^2 = 0$

semidefinite Fälle λ_1 oder $\lambda_2 = 0$:

- Zwei parallele Geraden: $ax_1^2 = 1$
- Eine Gerade: $x_1^2 = 0$
- Parabel: $ax_1^2 - x_2 = 0$
- leere Menge $ax_1^2 = -1$

Hierbei kommt nur im semidefiniten Fall ein linearer Term vor, vgl. den Term mit $b_2 \neq 0$ am Anfang dieses Abschnitts. Der Fall $A = 0$ wurde hier ausgespart, das ist bei der Definition quadratischer Funktionen gängig (analog dazu, dass $0 \cdot x^2 + bx + c$ nicht als quadratische Funktion definiert wird), ansonsten käme hier durch die Gleichung $0 \cdot x = 0$ noch die gesamte Ebene als mögliche Quadrik hinzu.

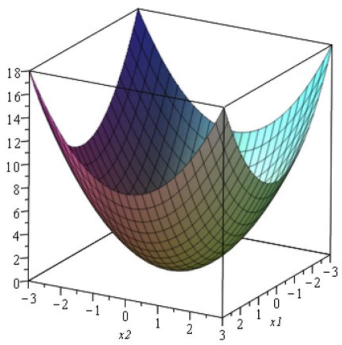
Beispiele

$$A = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$Q_A(\mathbf{x}) = \mathbf{x}^T \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mathbf{x} = (x_1, x_2) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (x_1, x_2) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1^2 + x_2^2.$$

analog zu

$f: \mathbb{R} \to \mathbb{R}, \ y(x) = x^2$



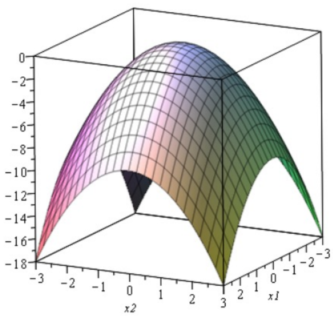
Prototyp
„positiv definite Matrix“

Quelle: T. Zehrt, U Basel, „Quadr. Formen und Definitheit“

Beispiele

$$B = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$Q_B(\mathbf{x}) = (x_1, x_2) \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (x_1, x_2) \begin{pmatrix} -x_1 \\ -x_2 \end{pmatrix} = -x_1^2 - x_2^2.$$



Prototyp
„negativ definite Matrix“

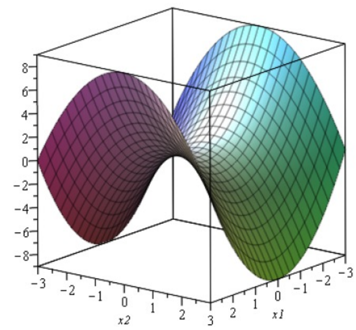
(analog zu
$y(x) = -x^2$)

Quelle: T. Zehrt, U Basel, „Quadr. Formen und Definitheit“

Beispiele

$$C = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$Q_C(\mathbf{x}) = (x_1, x_2)\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (x_1, x_2)\begin{pmatrix} x_1 \\ -x_2 \end{pmatrix} = x_1^2 - x_2^2.$$



Prototyp „indefinite Matrix“

Quelle: T. Zehrt, U Basel, „Quadr. Formen und Definitheit“

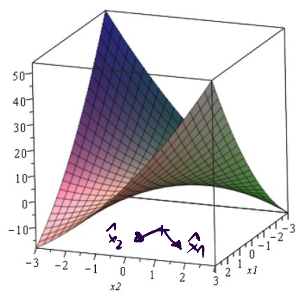
Beispiele

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

$$Q_D(\mathbf{x}) = \mathbf{x}^T\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}\mathbf{x} = x_1^2 + 4x_1x_2 + x_2^2. = 3\hat{x}_1^2 - \hat{x}_2^2$$

$\hat{=}$ Parabel $3\hat{x}_1^2$ für Richtung $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$\hat{=}$ Parabel $-\hat{x}_2^2$ in Richtung $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$



Quelle: T. Zehrt, U Basel, „Quadr. Formen und Definitheit“

19.3 Ein genauerer Blick auf Kegelschnitte: Herleitung der Kegelschnittgleichung und Diskussion möglicher Lösungen

Dass das Schnittproblem zwischen (Doppel-)kegel und Ebene auf eine Quadrik führt, sieht man am besten ein, wenn man in der Modellierung die Ebene fest wählt und den Kegel als beweglich ansieht (im Gegensatz zu den meisten Darstellungen, in denen der fest stehende Kegel mit einer beweglichen Ebene geschnitten wird). Dann lässt sich der euklidische Anschauungsraum $\mathcal{A}_3$ mit den in Kapitel 12 beschriebenen Methoden so koordinatisieren, dass die Ebene die $x_1 - x_2$-Ebene ist und die Kegelspitze S senkrecht über dem Ursprung des Koordinatensystems liegt. Damit erhält man die Koordinatenausdrücke

$$E: \quad x_3 = 0, \quad (\text{Normalenvektor von E: } n = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}), \quad \text{Koordinaten der Kegelspitze:} s = \begin{pmatrix} 0 \\ 0 \\ h \end{pmatrix}.$$

Um den Kegel vollständig zu beschreiben, benötigt man noch seine Achse $\vec{a} \in \mathcal{A}_3$, die sich über einen normierten Vektor $a \in \mathbb{R}^3$ beschrieben lässt, und den Öffnungswinkel $\alpha \in]0, \pi[$ des Kegels. Ein Punkt X liegt dann auf dem Mantel des Doppelkegels genau dann, wenn der Verbindungsvektor zwischen Kegelspitze und X mit der Achse des Kegels den Öffnungswinkel α bildet, d.h. wenn

$$\angle(\overrightarrow{SX}, \vec{a}) = \alpha \quad \text{oder} \quad \angle(\overrightarrow{SX}, \vec{a}) = \pi - \alpha$$

gilt, also genau dann, wenn der Cosinuswert der beiden Winkel bis aufs Vorzeichen übereinstimmt, d.h.

$$\left| \frac{\langle \overrightarrow{SX}, \vec{a} \rangle}{\|\overrightarrow{SX}\| \|\vec{a}\|} \right| = \cos \alpha, \qquad \text{d.h. wenn} \qquad \alpha^2 = \frac{\langle \overrightarrow{SX}, \vec{a} \rangle^2}{\|\overrightarrow{SX}^2\| \|\vec{a}\|}.$$

Für einen Punkt X, der gleichzeitig auf E, also in der x_1-x_2-Ebene liegt, hat $\overrightarrow{SX}$ die Koordinaten $(x_1, x_2, -h)$Wählen wir den Achsenvektor mit $\|\vec{a}\| = 1$ und Koordinatenvektor $a = (a_1, a_2, a_3)$, so ergibt sich für Punkte im Schnitt von Kegelmantel und Ebene, dass diese genau die Punkte sind, die (mit der Abkürzung $c_\alpha := \cos \alpha$ die Beziehung

$$c_\alpha^2 = \frac{(x_1 a_1 + x_2 a_2 - h a_3)^2}{x_1^2 + x_2^2 + h^2}$$

erfüllen. Ausmultiplizieren gibt die Gleichung

$$(a_1^2 - c_\alpha^2)x_1^2 + (a_2^2 - c_\alpha^2)x_2^2 + 2a_1a_2x_1x_2 - 2ha_1a_3x_1 - 2ha_2a_3x_2 - (a_3^2 - c_\alpha^2)h^2 = 0,$$

eine Quadrik im $\mathbb{R}^2$. Diese lässt sich (wie jede Quadrik in $\mathbb{R}^2$) auch elegant mit Hilfe einer (3×3)-Matrix ausdrücken: Es ist $(x_1, x_2)^T$ ein Punkt auf der $x_1 - x_2$-Ebene, der auf dem Kegelmantel des Kegels mit Spitze $s = (0, 0, h)$, Achse $a = (a_1, a_2, a_3)$ und Öffnungswinkel α liegt, genau dann, wenn für den Vektor $x = (x_1 x_2 h)$ und die Matrix

$$A = \begin{pmatrix} a_1^2 - c_\alpha^2 & a_1a_2 & -a_1a_3 \\ a_1a_2 & a_2^2 - c_\alpha^2 & -a_2a_3 \\ -a_1a_3 & -a_2a_3 & a_1^2 - c_\alpha^2 \end{pmatrix}$$

die Gleichung $x^T Ax = 0 \in \mathbb{R}^3$ gilt – rechnen Sie's nach!

Bei dem oben erwähnten Schnittproblemen Kreis-Kreis, Kreis-Ebene, Zylinder-Ebene ist klar, dass dort nicht jede der möglichen Quadriken auch als Lösung dieses Schnittproblems auftreten kann; das liegt an den sich aus den Problemen ergebenden Parametern A, b, c der entstehenden Gleichungen. Für Kegelschnitte gilt das genauso; es ergibt sich die folgende Einordnung in die Lösungstheorie aus Abschnitt 19.2:

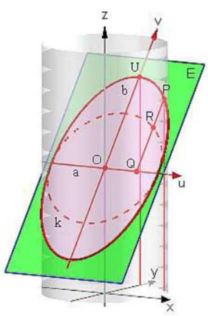
- Es stellt sich heraus, dass sich als Kegelschnitte im „nicht entarteten“ Fall (das bedeutet hier, dass die Kegelspitze nicht in der schneidenden Ebene liegt) nur Parabeln, Ellipsen (mit Spezialfall Kreis) oder eine Hyperbeln ergeben können, je nachdem, ob der Winkel zwischen Kegelachse und Schnittebene gleich, größer oder kleiner ist als der Öffnungswinkel des Kegels. Die Quadrik gibt hierbei Gleichungen für die Koordinaten der „Schnittspuren“ auf der $x_1 - x_2$-Ebene an.

- Umgekehrt sind entgegen mancher Darstellungen nicht alle Quadriken Kegelschnitte, so kommen die zwei parallelen Geraden und die leere Menge (und die gesamte Ebene, falls erlaubt) nicht als mögliche Kegelschnitte vor. Eine saubere Darstellung findet man z.B. bei Günther Ziegler, [Ziegler].

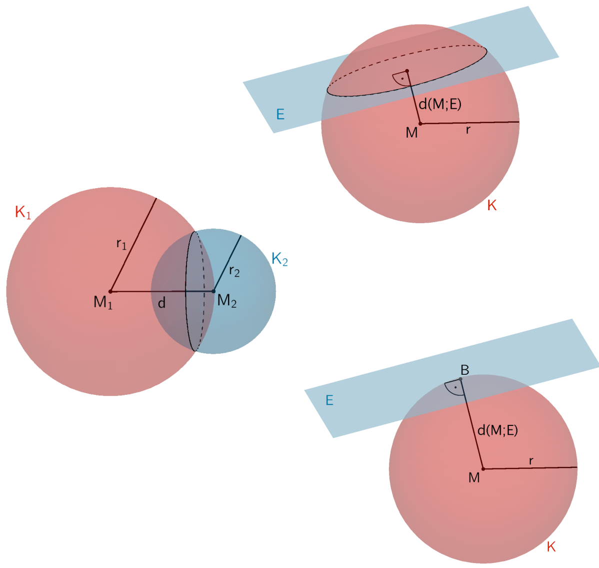
Weiter geht es in der Theorie mit Flächen zweiter Ordnung im $\mathbb{R}^3$: So nennt man die „nicht entarteten“ Lösungsmengen quadratischer Gleichungen $x^T Ax + bx + c = 0$ mit $A \in \mathbb{R}^{3\times3}$, also Quadriken im $\mathbb{R}^3$. Die sind auch nett und mit derselben Theorie klassifizierbar, kurze Darstellung s. Vorlesung.

Unter den Kurven und Flächen höherer Ordnung ($n \geq 3$), d.h. Lösungsmengen von Gleichungen, in denen in den einzelnen Summanden höchstens Produkte von n Unbekannten vorkommen, finden sich ebenfalls einige possierliche Tierchen, unter anderem die in der Kryptographie eingesetzten elliptischen Kurven, Kurven mit Spitzen und Überschneidungen usw. Einige bunte Bilder sehen wir zum Ende der letzten Vorlesung an.

Kurven der Ordnung 2 in IR^2

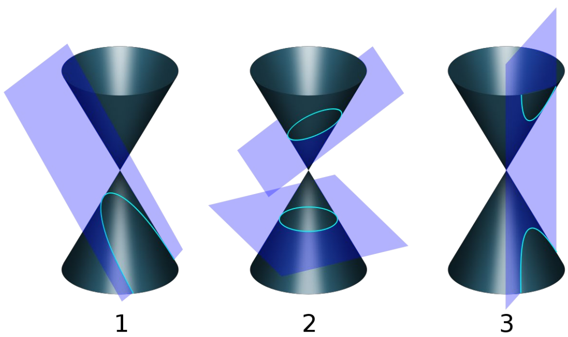


Links: Jugend trainiert Mathematik
Bennewitz, I. Hachtel, D. Herden

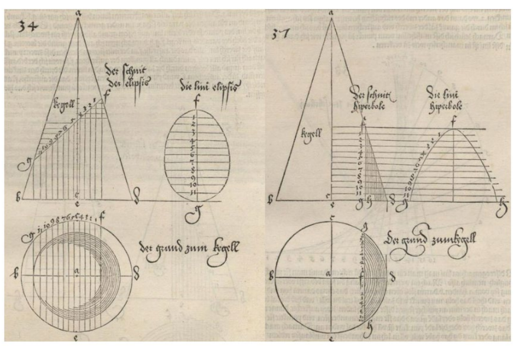


Rechts: mathelike.de

Kurven der Ordnung 2 in IR^2



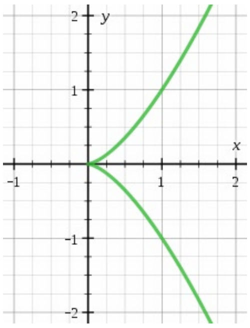
Wikipedia, „Kegelschnitt“, Pbroks13



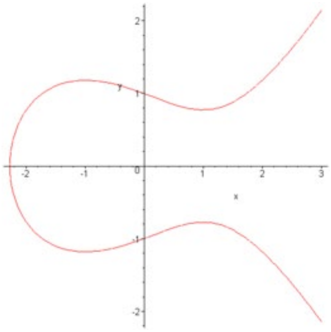
Aus: Das Wunderland der Kegelschnitte (Teil I), Georg Glaeser

Kurven der Ordnung 3 in IR²

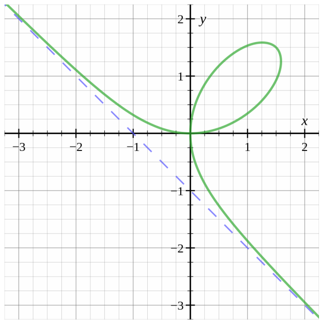
$y^3 = x^2$ (Neilsche Parabel)



$y^2 = x^3 + ax + b$ (elliptische Kuven)

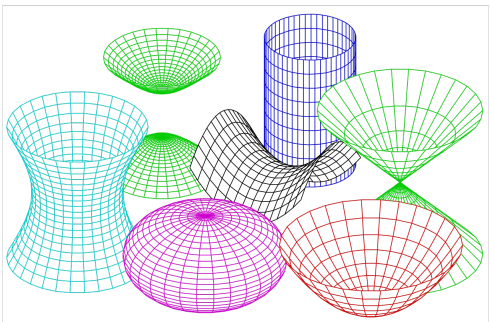


$x^3 + y^3 - 3xy = 0$ (kartesisches Blatt)

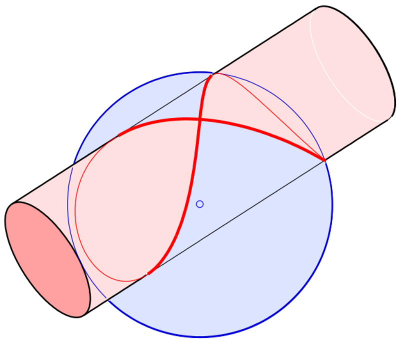


Quelle: Wikipedia: Algebraische Kurve/Elliptische Kurven

Alle Flächen der Ordnung 2 und eine Kurve der Ordnung 4 (beide in IR³)



Lösungen von $x^TAx + \langle b,x\rangle + c = 0$ in IR³



„Vivianisches Fenster“ (1692)

Schluss

So, das war's für jetzt. Es gäbe noch so viel zu erzählen. :)

Ich hoffe, mit diesen letzten Ausführungen über theoretische und praktische „Andockmöglichkeiten“ habe ich einigermaßen klar machen können, dass das Spektrum der Dinge, die man von hier aus mit dem in den Vorlesungen LAAG I & II vorgestellten Apparat und seinen Erweiterungen bewerkstelligen kann, ziemlich umfassend und vielseitig ist.

Ich hoffe auch, dass der Ansatz, möglichst viel *genetisches Prinzip* (Freudenthal) und möglichst wenig unmotivierten Bourbakismus in der Vorlesung vorkommen zu lassen, also insbesondere Begriffsbildungen ausführlich zu motivieren und Beweisideen nicht vom Himmel fallen zu lassen, einigermaßen sichtbar und nachvollziehbar war und dazu beitragen konnte, einige erhellende Momente mehr zu bescheren als dies in manchen anderen Abhandlungen über Lineare Algebra der Fall ist.

Rückmeldungen zum Skript nehme ich immer gern entgegen.

Thorsten Rohwedder, Juli 2021-Februar 2023.
thorstenrohwedder@gmx.de

19.4 Aufgaben zu Kapitel 19

Aufgabe 19.107. (Diagonalisierbarkeit von Matrizen):

Die komplexe Matrix $A_1 \in \mathbb{C}^{3\times 3}$ und die reellen Matrizen $A_2, A_3, A_4 \in \mathbb{R}^{3\times 3}$ seien gegeben durch

$$A_1 = \begin{pmatrix} 0 & 0 & 1+i \\ 1 & 0 & 1 \\ 1+i & 0 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} -5 & -3 & -3 \\ 2 & 2 & 4 \\ 0 & 0 & 1 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 2 & 0 \\ -1 & -1 & 1 \end{pmatrix}, \quad A_4 = \begin{pmatrix} 3 & 0 & 0 \\ -4 & -1 & -4 \\ 0 & 0 & 3 \end{pmatrix}.$$

Überprüfen Sie, ob die Matrizen A_j $(j = 1, 2, 3, 4)$ diagonalisierbar sind und bestimmen Sie gegebenenfalls eine invertierbare Matrix C_j und eine Diagonalmatrix D_j so, dass $A_j = C_j D_j C_j^{-1}$ ist.

Aufgabe 19.108. (Diagonalisierbarkeit von Endomorphismen):

Es bezeichne V_3 den Vektorraum der Polynome vom Grad kleiner oder gleich 3 mit reellen Koeffizienten und p'' die zweite Ableitung des Polynoms p. Die lineare Abbildung $f : V_3 \to V_3$ sei gegeben durch die Vorschrift

$$f\big(p\big) = (X^2 + X + 1) \cdot p''.$$

Zeigen Sie, dass f diagonalisierbar ist und bestimmen Sie eine Zerlegung von V_3 in Eigenräume von f.
Tipp: Wählen Sie eine geeignete Basis von V_3.

Aufgabe 19.109. (Theoretisches zu Diagonalisierbarkeit):

(a) Die Matrix $A \in \mathbb{C}^{n\times n}$ besitze $\lambda \in \mathbb{C}$ als einzigen Eigenwert. Zeigen Sie, dass A genau dann diagonalisierbar ist, wenn A von der Form $A = \lambda \cdot I_n$ ist.

(b) Sei V ein endlichdimensionaler $\mathbb{K}$-Vektorraum, $f \in \mathcal{L}(V)$ diagonalisierbar und $B = (b_1, \ldots, b_n)$ eine Basis von V, die aus Eigenvektoren von f besteht.

Beweisen Sie: Ist $g \in \mathcal{L}(V)$ und jeder Vektor aus B auch ein Eigenvektor von g (nicht notwendigerweise zu denselben Eigenwerten), so ist $f \circ g = g \circ f$.

Tipp: Es hilft z.B., für beliebiges $v \in V$ eine Darstellung von v als Linearkombination der Vektoren aus B zu betrachten.

Aufgabe 19.110. (Diagonalisierbarkeit symmetrischer Matrizen):

(a) Sei $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \in \mathbb{R}^{2\times 2}$.

Zeigen Sie durch konkretes Nachrechnen einer der Bedingungen aus Satz 18.2.2, also ohne Anwendung der Sätze 18.3.1, 18.3.2, dass A diagonalisierbar ist.

(b) Beweisen Sie, dass die Matrix

$$B = \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix}$$

nicht diagonalisierbar ist. Begründen Sie, warum das kein Widerspruch zum Satz über die reelle, orthogonale Diagonalisierbarkeit symmetrischer/hermitescher Matrizen (Satz 18.3.1, 18.3.2) ist.

Aufgabe 19.111. (Diagonalisierungen symmetrischer Matrizen):

(a) Bestimmen Sie zu jeder der folgenden reellen Matrizen $A_j (j = 1, 2, 3)$ jeweils eine orthogonale Matrix Q_j derart, dass $Q_j^T A_j Q_j$ Diagonalform besitzt:

$$A_1 = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

(b) Seien $q_1, \ldots, q_n$ in $\mathbb{R}^n$ und $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$. Begründen Sie, dass die Abbildung

$$\mathbb{R}^n \to \mathbb{R}^n, \qquad x \mapsto \sum_{i=1}^{n} \lambda_i \cdot \langle q_i, x \rangle \cdot q_i = \sum_{i=1}^{n} \lambda_i q_i q_i^T x$$

linear ist.

(c) Sei $A \in \mathbb{R}^{n \times n}$ eine symmetrische Matrix und eine $A = QDQ^T$ eine orthogonale reelle Diagonalisierung von A. Zeigen Sie: Sind $q_1, \ldots, q_n$ die Spalten von Q und $\lambda_1, \ldots, \lambda_n$ die Diagonaleinträge von D, so ist für alle $x \in \mathbb{R}^n$

$$Ax = \sum_{i=1}^{n} \lambda_i q_i q_i^T x,$$

es gilt also $A = \sum_{i=1}^{n} \lambda_i q_i q_i^T$. Diese Darstellung heißt Spektraldarstellung oder Spektralzerlegung von A. *Tipp:* Nutzen Sie (b) und dass lineare Abbildungen durch ihre Wirkung auf eine Basis eindeutig festgelegt sind.

(d) Geben Sie die Spektralzerlegung der Matrix A_2 aus Aufgabe (a) an. Summanden mit Eigenwert $\lambda_i = 0$ können hierbei weggelassen werden.

Aufgabe 19.112. (Diagonalisierbarkeit, Definitheit und Eigenwerte):

Gegeben sind die reellen Matrizen

$$A = \begin{pmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 3 & 3 \\ 3 & 1 & 3 \\ 3 & 3 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} 3 & 3 & 3 \\ 0 & 2 & 2 \\ 0 & 0 & 1 \end{pmatrix}.$$

(a) Entscheiden Sie für jede der Matrizen, ob sie diagonalisierbar ist.

(b) Bestimmen Sie die Eigenwerte der gegebenen Matrizen und ihre Vielfachheiten.

(c) Welche der Matrizen sind positiv oder negativ (semi-)definit oder indefinit?

(d) Welche der Matrizen ist Gram'sche Matrix eines Skalarprodukts auf $\mathbb{R}^3$?

Aufgabe 19.113. (Berechnung einer Singulärwertzerlegung):

Berechnen Sie eine Singulärwertzerlegung $A = U\Sigma V^T$ der Matrix

$$A = \begin{pmatrix} 2 & -3 \\ 1 & 6 \end{pmatrix} \in \mathbb{R}^{2\times 2}.$$

Gehen Sie dabei wie im Existenzbeweis der Singulärwertzerlegung vor:

- Eine Diagonalisierung der symmetrischen Matrix $A^T A$ liefert die Matrix $V = (v_1, \dots, v_n)$,
- Aus den Bedingungen $Av_i = \sigma_i u_i$, $\|u_i\| = 1$ und $\sigma_i \geq 0$ (für $i = 1, 2$) erhält man U und Σ.

Benutzte Literatur und Weiterführendes

[Ax] Sheldon Axler, *Linear Algebra Done Right*, 3. Auflage, 2015, Springer. (Link).

[BeGoMü] Ralf Benölken, Hans-Joachim Gorski, Susanne Müller-Philipp, *Leitfaden Geometrie, Für Studierende der Lehrämter*, 7. Auflage, 2018, Springer. (Link)

[BeuLA] Albrecht Beutelspacher, *Lineare Algebra*, 8. Auflage, 2014, Springer. (Link).

[Biel] Jana Bielagk, Skript zur Vorlesung *Lineare Algebra I* im Wintersemester 2018/19, HU Berlin.

[Biel] Jana Bielagk, Skript zur Vorlesung *Lineare Algebra II* im Sommersemester 2019, HU Berlin.

[Demmel] James Demmel, Applied numerical linear algebra, SIAM, 1997.

[Dei] Oliver Deiser, Einführung in die Mengenlehre, Online-Skript auf „Aleph 1“, (Link).

[Elden] Lars Elden, Matrix Methods in Data Mining an Pattern Recognition, SIAM, 2007.

[FaHa] Gerald Farin, Dianne Hansford, *Lineare Algebra: Ein geometrischer Zugang*, 2003, Springer (Link).

[Filler] Andreas Filler, *Elementare Lineare Algebra*, 2011, Springer (Link).

[Hanke-B] Martin Hanke-Borgeouis, Grundlagen der Numerischen Mathematik und des Wissenschaftlichen Rechnens, Vieweg und Teubner, 3. Auflage, 2009.

[Hartm] Erich Hartmann, Darstellende Geometrie für Bauingenieure, Skript zur Vorlesung von Karsten Große-Brauckmann, Frühjahr 2015, (Link)

[HuWi] Bertram Huppert, Wolfgang Willems, *Lineare Algebra*, 2010, Vieweg & Teubner, (Link).

[LPGö] Ina Kersten, Analytische Geometrie und Lineare Algebra, LP Göttingen, Online-Kurs, (Link)

[MehrLie] Jens Liesen, Volker Mehrmann, *Lineare Algebra*, 2. Auflage, 2015, Springer (Link).

[Lind] Detlef Lind, *Koordinaten, Vektoren, Matrizen*, 1997, Spektrum. (Link).

[Loeh] Clara Löh, Skript zur Vorlesung *Lineare Algebra I* im Wintersemester 2016/17, Uni Regensburg. Per Google-Suche gut zu finden.

[Loeh2] Clara Löh, Skript zur Vorlesung *Lineare Algebra II* im Sommersemester 2017, Uni Regensburg. Per Google-Suche gut zu finden.

[Muth] Herbert J. Muthsam, Lineare Algebra und ihre Anwendungen, Elsevier/Spektrum, 2006.

[Jähn] Klaus Jähnich, *Lineare Algebra*, 11. Auflage, 2011, Springer. (Link).

[Tisch] Caren Tischendorf, Skript zur Vorlesung *Lineare Algebra I* im Wintersemester 2018/19, HU Berlin.

[Tisch2] Caren Tischendorf, Skript zur Vorlesung *Lineare Algebra II* im Sommersemester 2019, HU Berlin.

[SchSch] Harald Scheid, Wolfgang Schwarz, Elemente der Linearen Algebra und der Analysis, Spektrum-Verlag, 2009.

[Weitz] Edmund Weitz, *Konkrete Mathematik (nicht nur) für Informatiker*, 2018, Springer. (Link)

[Wuß] Hans Wußing, *6000 Jahre Mathematik – Eine kulturgeschichtliche Zeitreise, Band 1*, 2008, Springer.

[Wuß] Hans Wußing, *6000 Jahre Mathematik – Eine kulturgeschichtliche Zeitreise, Band 2*, 2008, Springer.

[Ziegler] Günter M. Ziegler, Geometrie. Skript zur Vorlesung an der Freien Universität Berlin, Teil 3 (Kap. 4). (Link)