

# Mathematik für Physiker I

Michael Dreher  
Fachbereich für Mathematik und Statistik  
Universität Konstanz

Studienjahr 2011/12

**Etwas Juristisches:**

Dieses Werk ist unter einem *Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported Lizenzvertrag* lizenziert. Um die Lizenz anzusehen, gehen Sie bitte zu <http://creativecommons.org/licenses/by-nc-nd/3.0/de/> oder schicken Sie einen Brief an Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA.

Wer aus den Büchern nicht mehr lernt, als was in den Büchern steht,  
der hat die Bücher nicht halb genutzt.

Gotthold Ephraim Lessing, 1729–1781



# Inhaltsverzeichnis

<b>1 Grundlagen</b>	<b>11</b>
1.1 Wiederholung aus der Schulanalysis	11
1.1.1 Reelle Zahlen	11
1.1.2 Funktionen	12
1.2 Die komplexen Zahlen	13
1.2.1 Zur Einstimmung: Die Cardanische Formel	13
1.2.2 Die komplexen Zahlen	14
1.2.3 Funktionen komplexer Zahlen	18
1.2.4 Ausblick: Elektrotechnik	21
1.2.5 Ausblick: Mechanik	22
1.3 Die Ebene und der $\mathbb{R}^2$	22
1.3.1 Allgemeine Eigenschaften	22
1.3.2 Drehungen	27
1.4 Gruppentheorie	29
1.4.1 Einführung	29
1.4.2 Ausblick: Gruppen in der Physik	33
1.4.3 Ausblick: Mobilfunk	35
1.5 Der Raum und der $\mathbb{R}^3$	37
1.5.1 Allgemeines	37
1.5.2 Vektorprodukt und Spatprodukt	37
1.5.3 Drehungen im $\mathbb{R}^3$	42
1.5.4 Der affine Raum	46
1.6 Schlüsselbegriffe	46
<b>2 Vektorräume</b>	<b>47</b>
2.1 Allgemeine Eigenschaften	47
2.2 Linearkombinationen, Erzeugendensysteme usw.	49
2.2.1 Linearkombinationen und Unterräume	49
2.2.2 Lineare Unabhängigkeit und Basen	50
2.2.3 Dimension eines Vektorraumes	53
2.3 Vektorräume mit Skalarprodukt	57
2.3.1 Skalarprodukte, Normen, Orthogonalsysteme	57
2.3.2 Approximationsprobleme	60
2.4 Ausblick: Vektorräume in der Physik 1	62
2.5 Ausblick: die HELMHOLTZ-Projektion	63
2.6 Schlüsselbegriffe	64

<b>3 Matrizen</b>	<b>65</b>
3.1 Operationen mit Matrizen	65
3.2 Gleichungssysteme	67
3.3 Schlüsselbegriffe	72
<b>4 Homomorphismen</b>	<b>73</b>
4.1 Allgemeine Eigenschaften	73
4.2 Geometrische Aspekte	76
4.3 Lineare Gleichungen	80
4.4 Basistransformationen	81
4.5 Differentialgleichungen	84
4.6 Ausblick: Lineare Abbildungen in der Physik	87
4.7 Ausblick: Vektorräume in der Physik 2	90
4.8 Schlüsselbegriffe	92
<b>5 Normierte Räume, Reelle Zahlen, Folgen, Reihen</b>	<b>93</b>
5.1 Folgen im $\mathbb{R}^d$ und $\mathbb{C}^d$	93
5.2 Folgen und Reihen in normierten Räumen	96
5.2.1 Vollständigkeit	97
5.2.2 Reihen in normierten Räumen	99
5.2.3 Konvergenzkriterien	100
5.3 Folgen und Reihen reeller Zahlen	102
5.3.1 Schranken und Grenzen	103
5.3.2 Beispiele für konvergente Folgen	105
5.3.3 Nichtabsolute Konvergenz und Umordnungen	107
5.4 Potenzreihen	108
5.5 Beispiel: Die Exponentialfunktion	111
5.6 Schlüsselbegriffe	114
<b>6 Funktionen</b>	<b>115</b>
6.1 Grenzwerte von Funktionen	115
6.2 Stetigkeit	118
6.3 Differenzierbarkeit	123
6.4 Mittelwertsatz und TAYLORScher Satz	129
6.5 Elementare Funktionen	137
6.5.1 Der geometrische Zugang zu den Winkelfunktionen	137
6.5.2 Der analytische Zugang zu den Winkelfunktionen	141
6.5.3 Die Hyperbelfunktionen	146
6.5.4 Wurzeln aus komplexen Zahlen	147
6.6 Verfahren zur numerischen Lösung nichtlinearer Gleichungen	149
6.6.1 Das Halbierungsverfahren	150
6.6.2 Funktionaliteration und der BANACHSche Fixpunktsatz	150
6.6.3 Das NEWTONverfahren	152
6.7 Schlüsselbegriffe	154
<b>A Algebraische Strukturen</b>	<b>155</b>

# Einleitung

## Herzlich willkommen im Physikstudium !

Zu jedem Physikstudium gehört ein Mathematikurs, und dieser Versuch eines Vorworts soll beschreiben, welche Ziele wir mit diesem Mathematikurs anpeilen. Im Schnelldurchlauf: es sind etwas andere Ziele als im gymnasialen Mathematikunterricht (wobei hier nicht erörtert werden soll, ob der schulische Mathematikunterricht sinnvoll konzipiert ist und ob er seine Ziele überhaupt erreicht).

Bekanntlich ist die Physik eine Wissenschaft. Das Gegenteil davon ist die Unwissenschaft, wie sie vertreten wird von Wünschelrutengängern, Schamanen, Esoterikern und sonstigen verwirrten Persönlichkeiten. Zwischen Wissenschaftlern und Nichtwissenschaftlern besteht der ausschlaggebende Unterschied, daß die Wissenschaftler ihre Arbeitsmethoden präzise begründen und ihre erzielten Ergebnisse rechtfertigen können, und zwar (z.B.) auf folgenden Wegen:

**Experiment:** die Voraussagen einer Theorie werden geprüft (Arbeitsweise in der Experimentalphysik),

**Theorie:** eine Aussagen wird logisch präzise aus elementaren Prinzipien hergeleitet, die man als wahr voraussetzt (Arbeitsweise in der theoretischen Physik und der Mathematik),

**numerische Simulationen:** das sind Rechenexperimente auf dem Computer, z.B. um eine Theorie zu testen.

Da wir nun als Wissenschaftler wahrgenommen werden wollen, kommen wir nicht drumherum, diese Ansprüche an wissenschaftliches Arbeiten auch gegen uns gelten zu lassen, woraus sich ein erster Unterschied zum Schulunterricht ergibt:

Die schultypische Frage nach dem *Wie* (*Wie* verläuft der Rechenweg ?) wird ersetzt durch die Frage nach dem *Warum*: *Warum* glaube ich eigentlich, daß mein Rechenweg mich überhaupt zum Rechenziel führt ? Gelegentlich artet die Abiturvorbereitung aus in ein Antrainieren halbverstandener Rechenrituale; diese Phase soll im Physikstudium nicht wiederkehren.

## Was ist eigentlich Mathematik ?

Wie eben schon angedeutet, geht es in der Mathematik nicht um das Abspulen von immergleichen Rechen-schemata, ganz im Gegenteil:

- (a) Mathematik ist die Kunst, stumpfsinniges Rechnen zu vermeiden. Es gibt keinen Platz für häßliche Mathematik, genauso wie es keinen Platz gibt für häßliche Physik.
- (b) Mathematik ist eine Wissenschaft, die Strukturen erforscht und durchschaubar macht.

Dabei strebt die Mathematik zuallererst nach Erkenntnis und versucht zu erklären, warum die (von ihr betrachtete Facette der höchst vielfältigen) Wirklichkeit so ist, wie sie ist; und genau dasselbe gilt natürlich für jede andere Wissenschaft. Ein besonderes Merkmal einer jeden Naturwissenschaft ist es, daß ein Forscher<sup>1</sup> sich zwar ein bestimmtes Forschungsziel stellen kann, aber es ist zunächst nicht klar, ob das Problem überhaupt<sup>2</sup> lösbar ist, ob man es selber schafft (oder ob jemand anders schneller ist), oder ob der gewählte Weg nicht vielleicht eine Sackgasse ist und man am Ende der mühseligen Plackerei mit leeren Händen

---

<sup>1</sup>aus Gründen der sprachlichen Übersichtlichkeit sind nur die maskulinen Personenbezeichnungen angeführt ...

<sup>2</sup>Ausnahme: Spielzeugprobleme, bei denen jeder ahnt, was herauskommt, und die nur Wenige interessieren

dasteht. In der Schule ist das bekanntlich anders, denn dort bekommt man nur Aufgaben, die garantiert mit dem vorher vermittelten Wissen lösbar sind.

Kreative und phantasievolle Menschen sind klar im Vorteil, und es lohnt sich, „um die Ecke zu denken“!

Ein weiteres wichtiges Kennzeichen mathematischer Arbeit ist die *Abstraktion*. Das bedeutet, daß man bei einer Situation die Eigenschaften von allen beteiligten Objekten entsprechend der Kategorien *wichtig* / *unwichtig* sortiert, die unwichtigen wegwirft, damit die wichtigen klarer zu sehen sind (und damit das Problem überhaupt erst einmal handhabbar wird). Ein Physiker macht genau das gleiche, wenn er sagt „hierbei betrachten wir nur Punktmassen“. Bei Vektoren kann es zum Beispiel sein, daß ihre Pfeil-Gestalt wichtig ist (wenn man z.B. die Kräfte auf einen Körper untersucht, dann ist die Veranschaulichung der Kräfte als Pfeile völlig natürlich). Man kann die Eigenschaft eines Vektors, ein „Pfeil“ zu sein, aber auch wegwerfen, und lediglich die beiden Eigenschaften übrigbelassen, daß man Vektoren addieren kann, und daß man Vektoren mit einer Zahl multiplizieren kann (sowie einiger Rechenregeln für diese Operationen). Genau dasselbe trifft aber auch auf Funktionen zu, denn diese lassen sich auch zueinander addieren bzw. mit einer Zahl multiplizieren.

Wir dürfen also Funktionen als Vektoren ansehen !

Das ist ein Beispiel für einen Abstraktionsschritt. Und weil wir spätestens in der Quantenmechanik sowieso dazu gezwungen sein werden, diesen Abstraktionsschritt (Funktionen als Vektoren anzusehen) zu vollziehen, wollen wir den Gedanken *Vektor = Pfeil* gar nicht weiter verfestigen (dann müßten wir später umlernen) und arbeiten in diesem Kurs fast von Anfang an mit abstrakten Vektoren (ab Kapitel 2).

Wir schauen uns noch ein Beispiel an für „Funktionen als Vektoren“:

- Wir betrachten eine Funktion, die aus einer reellen *Zahl* eine reelle Zahl macht. Bekanntlich kann man von einer solchen Funktion Minima suchen, indem man die Ableitung gleich Null setzt.
- Jetzt betrachten wir eine Funktion, die aus einer *Funktion* eine reelle Zahl macht. Aus Gründen der sprachlichen Klarheit redet man besser von einem Funktional, und das wichtigste Beispiel ist vielleicht das Wirkungsfunktional aus der Theoretischen Mechanik. Viele Bewegungsgesetze der Mechanik (z.B. alles vom Typ  $F = ma$ ) folgen aus dem *Prinzip der kleinsten Wirkung*. Und wie sucht man nun Minima des Wirkungsfunktionalen ? Durch Nullsetzen der Ableitung (was auch immer das ist), wobei die eingesetzten Argumente des Wirkungsfunktionalen jetzt keine Zahlen  $x \in \mathbb{R}$  sind, sondern Funktionen aus einem unendlichdimensionalen Funktionenvektorraum.

Ein wenig hochtrabend formuliert, benutzt die Theoretische Mechanik also die Differentialrechnung in unendlichdimensionalen Funktionenvektorräumen, und dafür wollen wir mit diesem Mathematikurs die Grundlagen zu legen versuchen.

Mit etwas Glück kann es dann auf dem Wege der Abstraktion gelingen, Gemeinsamkeiten zwischen Dingen zu entdecken, die weit entfernt voneinander scheinen, und schon hat man ein wenig besser erkannt, wie die Welt aussieht und was deren Zusammenhänge sind.

## Methodische Konsequenzen

Es ist praktisch nicht durchführbar, sich im Studium auf jede Eventualität des späteren Berufslebens vorzubereiten und sich für jede auftretbare Situation ein paßgenau zugeschnittenes Rechenschema zurechtzulegen. Dafür ist die uns umgebende Welt einfach zu vielgestaltig. Außerdem ist das Gedächtnis gar nicht in der Lage, mehrere hundert isolierte Einzelinformationen abzuspeichern, wenn diese nicht untereinander vernetzt sind. Viel wichtiger ist es, inhaltliche Zusammenhänge zu erkennen, auch damit der Lernstoff ein assoziatives Netz bildet, das sich erheblich einfacher einprägen wird. Die im Skript aufgeführten Beweise sind anhand des Kriteriums ausgewählt worden, ob sie inhaltliche Zusammenhänge beleuchten und somit ein tieferes Verständnis ermöglichen (und eine akzeptable Länge aufweisen).

Bei den Hausaufgaben treten also Rechenaufgaben mit Zahlen in den Hintergrund. Kreativität und Phantasie sind gewünscht, weshalb bei einigen Hausaufgaben nicht sofort ersichtlich ist, wie der Lösungsweg aussieht; noch dazu kann es mehrere Wege zur Lösung geben. Wenn man sich unglücklicherweise verirrt, kann die Lösung womöglich lang und etwas häßlich sein (aber das merkt man natürlich).



*Garantie: Die 4 Aufgaben eines Hausaufgabenblattes sind auf etwa 6 Seiten lösbar.*

Das hängt natürlich von der Schriftgröße ab, aber wenn Sie mehr als 8 Seiten brauchen, machen Sie *sicherlich* etwas falsch, und wenn Sie weniger als 4 Seiten benötigen, machen Sie *eventuell* etwas falsch.

Am Ende eines jeden Kapitels sind Schlüsselbegriffe aufgelistet. Diese heißen deshalb so, weil sie eine Schlüsselbedeutung haben für das Verständnis des Skriptinhalts. Demnach sollten Sie diese Begriffe verstanden haben, also: Definition wiedergeben können, Eigenschaften beschreiben und Querverbindungen zu anderen Begriffen benennen. Mit diesen Begriffen sollten Sie so vertraut sein, daß Sie die kurzen Beweise verstehen und bei den längeren Beweisen (es sind eher wenige) auf jeden Fall die zentralen Gedanken angeben können.

## Wie löst man Hausaufgaben ?

Weil das Wichtigste am Studium nicht etwa die Vorlesungen sind, sondern die eigenständige Auseinandersetzung mit dem Lernstoff, kommen hier noch einige Anmerkungen zu den Hausaufgaben.

Lassen Sie sich von Ihrem Unterbewußtsein helfen. Dieses braucht allerdings seine Aufwärmzeit, und es muß wissen, wobei es eigentlich helfen soll, und es ist auch nicht zuverlässiger als die Eisenbahn.

Sie sollten also so schnell wie möglich die Aufgabe verstehen. Dazu besorgen Sie sich von allen vorkommenden mathematischen Fachbegriffen die *exakte* Bedeutung, also die Definition. Diese finden Sie im Skript und in Büchern (meistens mit einem Index ausgestattet); nur eingeschränkt empfehlenswert ist Wikipedia. Tragen Sie alle Eigenschaften dieser Begriffe zusammen, die Sie finden. Stellen Sie sicher, daß Sie die zusammengetragenen Dinge auch verstanden haben (hierbei sind ggf. Skizzen oder evtl. Zahlenbeispiele brauchbar).

Nun müßten Sie ein solides Verständnis dessen haben, worum es in der Aufgabe eigentlich geht. Wenn es sich um eine Rechenaufgabe handelt, ist jetzt nicht mehr allzuviel zu tun (die abiturtypischen Rechentechniken setzen wir als gefestigt voraus).

Falls es um einen Beweis geht: drücken Sie Voraussetzungen und Behauptung mit Ihren herausgeschriebenen Definitionen aus (formulieren Sie also um). Da Sie die Aufgabenstellung jetzt verinnerlicht haben, können Sie also beim Schlangestehen im Supermarkt drüber nachdenken oder bei jeder anderen Gelegenheit. Echte Zahlenbeispiele können hilfreich sein zur Ideenfindung, Taschenrechnerexperimente ebenfalls. Werden Sie kreativ (ab dieser Stelle gibt es kein Kochrezept mehr, wie auch ? Es geht ja schließlich um Phantasie, und wenn es dafür ein Schema zum Abspulen gäbe, dann würden wir wahrscheinlich in einer langweiligen Welt leben).

Diese Suche nach einer Stelle, wo man ansetzen kann, ist vielleicht mühselig, aber für den Lerneffekt unverzichtbar. Eine abgeschriebene Lösung haben Sie binnen 10 Tagen vergessen, aber Sie werden sich monatelang an den Moment erinnern, als nach längerer Anstrengung endlich der Groschen fiel. Der Schwung dieses Erfolgs wird Sie durch das ganze Semester tragen. Es lohnt sich ! Eine selbstgefundene Lösung ist soviel wert wie 10 abgeschriebene. Wer abpinselt, tut sich selbst keinen Gefallen und studiert einfach nur ineffizient.

Nach einiger Zeit haben Sie also eine mutmaßliche Lösung erhalten. Womöglich ist der Gedankengang etwas zickzackig (das geht den professionellen Mathematikern und Physikern genauso), eventuell findet Ihr Unterbewußtsein dann am nächsten Tag<sup>3</sup> eine Idee, wie man ihn begradigen kann.

Wenn Sie schon im Team arbeiten, können Sie ja so vorgehen, daß jeder einzeln für sich die Lösung sucht und danach beide Varianten verglichen werden. Vielleicht kann man beide Wege zu einem zusammenfügen, der schöner ist. Oder Sie lesen bei Ihrem Kommilitonen Korrektur (es hinterläßt einen schiefen Eindruck, wenn in einer Woche auf einmal 7 Physikstudenten der Meinung sind, daß  $\frac{1}{4} + \frac{1}{4} = \frac{1}{8}$  !). Falls Sie im Team auf gemeinsamem Blatt abgeben, seien Sie nicht überrascht, wenn die Übungsleiter dann strengere Bewertungsmaßstäbe anlegen, als wenn Sie alleine abgegeben hätten.

Achten Sie auf die äußere Form Ihrer Lösung. Die Lösung soll so klar dargestellt werden, daß auch ein Leser, der weder Aufgabe noch einen korrekten Lösungsweg kennt, erkennen kann, was eigentlich hier los ist. Korrektoren sind dankbar, wenn sie nicht gedrängt sind, irgendwelche hingekritzeltten Gedankenfragmente mühsam zusammenzupuzzeln, sondern wenn die Lösung ein richtiger Text ist mit syntaktisch korrekten

---

<sup>3</sup>im Übrigen werden die meisten Texte besser, wenn man sie mit einigem zeitlichen Abstand noch mal neu schreibt

Sätzen der deutschen Sprache. Die Kunst, sich im ganzen Satz gut auszudrücken, ist bemerkenswert schwierig, sodaß man frühestmöglich mit dem Üben anfangen sollte. Es ist durchaus wahrscheinlich, daß Sie Ihre Hausaufgabenzettel später noch ein mal benötigen werden, und dafür wäre es dann nutzvoll, die Lösungen so klar und einleuchtend aufgeschrieben zu haben, daß Sie diese auch nach einem halben Jahr noch zügig verstehen.

## Rechentechische Hinweise

Jede Rechnung dient einem bestimmten Zweck.

Oft möchte man mit der Rechnung irgendwelche Ergebnisse ermitteln und diese dann woanders weiterverwerten. In einem solchen Fall ist es ein natürlicher Wunsch, die Rechnerei nicht länger zu haben als angemessen.

Es kann aber auch sein, daß der Rechenweg selbst aufschlußreich ist und einige Zusammenhänge beleuchten kann. Das funktioniert natürlich nur, wenn man beim Rechnen kein unentwirrbares Gleichungsspaghetti fabriziert hat.

In beiden Fällen lohnt es sich, *ökonomisch* zu rechnen.

Generell gilt dabei: wenn eine Gleichungsumformung keinen benennbaren Nutzen hat — weglassen oder zumindest so weit wie möglich hinauszögern. Spezieller heißt das zum Beispiel: Unterdrücken Sie den Reflex „Ich multipliziere erst mal alles aus“.

Als instruktives Beispiel suchen wir Wendepunkte von  $f = f(x) = \frac{3x+7}{(x+2)^2}$  durch Nullsetzen von  $f''$ . Fleißiges Ausmultiplizieren sämtlicher Klammern und die Regel  $(\frac{u}{v})' = \frac{u'v-uv'}{v^2}$  führen auf

$$f'(x) = \frac{-3x^2 - 14x - 16}{x^4 + 8x^3 + 24x^2 + 32x + 16},$$

$$f''(x) = \frac{6x^5 + 66x^4 + 288x^3 + 624x^2 + 672x + 288}{x^8 + 16x^7 + 112x^6 + 448x^5 + 1120x^4 + 1792x^3 + 1792x^2 + 1024x + 256},$$

und nun stecken wir fest, denn die Gleichung  $f''(x) = 0$  bekommen wir nicht gelöst.

Das ist einfach keine Mathematik mehr, denn der Verstoß gegen Prinzip (a) ist offensichtlich (kein Mathematiker oder Physiker mit ästhetischem Gespür rechnet freudvoll einen Monsterterm wie  $(-3x^2 - 14x - 16)' \cdot (x^4 + 8x^3 + 24x^2 + 32x + 16) - (-3x^2 - 14x - 16) \cdot (x^4 + 8x^3 + 24x^2 + 32x + 16)'$  aus. Abgesehen davon ist die Wahrscheinlichkeit von Rechenfehlern viel zu hoch.)

Es liegt aber auch ein Verstoß gegen Prinzip (b) vor, denn der Bruch für  $f''(x)$  enthält eine verborgene Struktur, die jedoch aufgrund der mißlungenen Darstellung unsichtbar ist. Oder ist Ihnen etwa aufgefallen, daß man diesen Bruch kürzen kann durch  $x^4 + 8x^3 + 24x^2 + 32x + 16$  ?!

Wenn man stattdessen das komplett überflüssige Ausmultiplizieren des Nenners sein läßt und jede Gelegenheit zum Kürzen sofort nutzt, bekommt man auf wesentlich schnellerem Wege

$$f'(x) = \frac{-3x - 8}{(x + 2)^3}, \quad f''(x) = \frac{6x + 18}{(x + 2)^4}.$$

Weitere Tips:

Nutzen Sie bei Zwischenergebnissen, wann immer es sich anbietet, die Möglichkeit zur Zwischenprobe, und zwar auf einem anderen Wege. Wenn Sie solche anderen Zugänge gezielt suchen, trainieren Sie gleichzeitig Ihren Durchblick durch die Mathematik, und auf den kommt es ja an.

Gewöhnen Sie sich Überschlagsrechnungen im Kopf an; Sie werden schneller sein als der Taschenrechner.

Formulieren Sie mit Worten, was Sie tun wollen, und schreiben Sie das dann auch hin. Es sollen keine Romane sein, eine Anmerkung der Form „Setze Gleichung (\*) in (\*\*) ein und löse nach  $y$  auf“ reicht völlig, tut Wunder und beglückt den Leser, weil die Struktur Ihrer Arbeit sichtbar wird.

**Viel Erfolg !**

# Kapitel 1

## Grundlagen

### 1.1 Wiederholung aus der Schulanalysis

#### 1.1.1 Reelle Zahlen

Wir listen einige bekannte Eigenschaften der reellen Zahlen auf:

Seien  $a, b, c, d \in \mathbb{R}$  beliebig. Dann haben wir:

**Kommutativität der Addition:**  $a + b = b + a$

**Assoziativität der Addition:**  $(a + b) + c = a + (b + c)$

**0 ist neutrales Element für die Addition:**  $0 + a = a + 0 = a$

**Subtraktion:** Jede Gleichung  $a + x = b$  (mit gegebenem  $a, b$ ) ist lösbar, die (eindeutige) Lösung  $x$  wird geschrieben als  $x = b - a$ .

**Kommutativität der Multiplikation:**  $a \cdot b = b \cdot a$

**Assoziativität der Multiplikation:**  $(a \cdot b) \cdot c = a \cdot (b \cdot c)$

**1 ist neutrales Element für die Multiplikation:**  $1 \cdot a = a \cdot 1 = a$

**Division:** Jede Gleichung  $a \cdot x = b$  (mit gegebenem  $a, b$ ; jedoch  $a \neq 0$ ) ist lösbar, die (eindeutige) Lösung  $x$  wird geschrieben als  $x = b/a$ .

**Distributivgesetz:** Addition und Multiplikation sind verzahnt gemäß der Regel  $(a + b) \cdot c = a \cdot c + b \cdot c$ .

**Bemerkung 1.1.** *Man sagt auch, daß die Menge der reellen Zahlen, gemeinsam mit den Operationen  $(+, \cdot)$ , einen Körper bildet. Weitere Beispiele für Körper sind die rationalen Zahlen oder die gebrochen-rationalen Funktionen (also Funktionen der Form  $f = f(x) = \frac{p(x)}{q(x)}$ , wobei  $p$  und  $q$  Polynome sind).*

Außerdem haben wir noch eine Ordnungsrelation  $\leq$ , mit folgenden Eigenschaften:

**Reflexivität:** Es ist  $a \leq a$ .

**Antisymmetrie:** Wenn  $a \leq b$  und  $b \leq a$ , dann ist  $a = b$ .

**Transitivität:** Wenn  $a \leq b$  und  $b \leq c$ , dann auch  $a \leq c$ .

Und diese Ordnungsstruktur ist verzahnt mit der Körperstruktur auf folgende Weise:

**Addition von Ungleichungen:** Wenn  $a \leq b$  und  $c \leq d$ , dann ist auch  $a + c \leq b + d$ .

**Multiplikation von Ungleichungen mit nichtnegativen Zahlen:** Wenn  $a \leq b$  und  $0 \leq c$ , dann ist  $ac \leq bc$ .

**Warnung 1.2.** *Man darf zwar 2 Ungleichungen addieren, aber nicht voneinander abziehen oder durcheinander dividieren, oder mit negativen Zahlen multiplizieren.*

Bekanntlich kann man die reellen Zahlen als Punkte auf der Zahlengeraden darstellen, und in dieser Form wird der Abstand zweier reeller Zahlen  $a, b$  auf der Zahlengeraden gegeben durch  $|a - b|$ . Insbesondere stellt  $|a|$  den Abstand des Punktes  $a$  vom Ursprung dar. Der Betrag hat folgende 3 Eigenschaften:

- Es ist stets  $|a| \geq 0$ ; und  $|a| = 0$  genau dann, wenn  $a = 0$ .
- Es ist  $|a \cdot b| = |a| \cdot |b|$ .
- Wir haben die Dreiecksungleichung:  $|a + b| \leq |a| + |b|$ .

**Bemerkung 1.3.** *Diese 3 Eigenschaften werden uns später in anderen Zusammenhängen noch öfters begegnen; wir werden dann sagen, daß eine Norm vorliegt.*

## 1.1.2 Funktionen

Zu einer Funktion  $f$  gehören zwei Dinge:

- ein Definitionsbereich  $D_f \subset \mathbb{R}$ . Wenn man den Definitionsbereich verändert (zum Beispiel verkleinert), dann erhält man eine andere Funktion.
- eine Vorschrift, wie die  $x$  aus dem Definitionsbereich in den Wertebereich abgebildet werden.

Beispiele für Funktionen sind:

- Polynome; der Definitionsbereich ist ganz  $\mathbb{R}$ ,
- gebrochen rationale Funktionen  $f = f(x) = g(x)/h(x)$  mit Polynomen  $g$  und  $h$ ; der Definitionsbereich ist  $\mathbb{R} \setminus \{\text{Nullstellen von } h\}$ ,
- die Winkelfunktionen  $\sin$  und  $\cos$ ,
- die Logarithmusfunktionen (definiert auf  $\mathbb{R}_+$ ).

Eine wichtige Eigenschaft von Funktionen ist die *Stetigkeit*. Wenn wir sagen, daß eine Funktion  $f$  im Punkte  $x_0$  stetig ist, dann meinen wir im wesentlichen folgendes:

Wir können den Abstand von  $f(x)$  und  $f(x_0)$ , das heißt den Wert  $|f(x) - f(x_0)|$ , beliebig klein bekommen, wenn wir dafür sorgen, daß der Abstand von  $x$  und  $x_0$ , also  $|x - x_0|$ , nur klein genug ist (eine genauere Definition kommt später).

Die oben genannten Funktionen sind stetig überall dort, wo sie definiert sind.

Beispiele für Unstetigkeiten sind Sprungstellen oder Polstellen. Es gibt aber noch weitere Typen von Unstetigkeiten.

Eine weitere wichtige Eigenschaft einer Funktion ist die *Differenzierbarkeit*. Anschaulich gesprochen, ist eine Funktion  $f$  im Punkte  $x_0 \in D_f$  differenzierbar, wenn der Graph der Funktion in  $(x_0 | f(x_0))$  eine Tangente hat, die nicht vertikal verläuft. Dann kann diese Tangente als Graph einer linearen Funktion

$$t = t(x) = f(x_0) + A(x - x_0), \quad x \in \mathbb{R},$$

interpretiert werden. Der Anstieg  $A$  dieser linearen Funktion heißt *Ableitung von  $f$  in  $x_0$*  und wird geschrieben als

$$A = \frac{df}{dx}(x_0) \quad \text{oder} \quad A = f'(x_0).$$

Etwas exakter formuliert, ist die Funktion  $f$  in  $x_0$  differenzierbar genau dann, wenn der Grenzwert des Differenzenquotienten vorhanden ist, das heißt

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} \quad \text{existiert.}$$

Der Wert dieses Grenzwertes ist dann gleich  $A$ .

Jede differenzierbare Funktion ist stetig.

Die meisten aus der Schule bekannten Funktionen (Polynome, gebrochen rationale Funktionen, Winkelfunktionen, Exponentialfunktionen, Logarithmusfunktionen) sind überall dort, wo man sie sinnvoll definieren kann, differenzierbar. Die Wurzelfunktionen und die Betragsfunktion  $f = f(x) = |x|$  sind im Nullpunkt *nicht* differenzierbar.

Die Rechenregeln für die Ableitung (Produktregel, Kettenregel usw.) sowie die Formeln für die Ableitungen der elementaren Funktionen wollen wir als bekannt voraussetzen.

Wenn die Ableitung  $f'$  einer Funktion  $f$  stetig ist, dann heißt  $f$  *stetig differenzierbar*. Wenn diese stetige Ableitung  $f'$  wiederum differenzierbar ist, dann nennt man  $f$  *zweimal differenzierbar*, und die Ableitung von  $f'$  wird mit  $f''$  bezeichnet. Wenn diese zweite Ableitung  $f''$  stetig sein sollte, dann nennen wir  $f$  *zweimal stetig differenzierbar*, und so weiter. Wenn sämtliche Ableitungen existieren und stetig sind, dann heißt  $f$  unendlich oft differenzierbar.

**Frage:** Was sind die Unterschiede (in grammatischer und in mathematischer Hinsicht) zwischen den beiden folgenden Formulierungen:

- Sei  $f$  eine stetig differenzierbare Funktion.
- Sei  $f$  eine stetige, differenzierbare Funktion.

Jede stetige Funktion ist auch integrierbar, das heißt, man kann bestimmte und unbestimmte Integrale bilden. Darunter verstehen wir das folgende:

Sei das Intervall  $[a, b]$  im Definitionsbereich der stetigen Funktion  $f$  enthalten. Dann wird der Flächeninhalt derjenigen Fläche, die von der  $x$ -Achse, dem Graphen der Funktion  $f$  und den Geraden  $x = a$  und  $x = b$  eingegrenzt wird, als *bestimmtes Integral von  $f$  über dem Intervall  $[a, b]$*  bezeichnet:

$$A = \int_{x=a}^b f(x) dx.$$

Dies ist lediglich eine anschauliche Beschreibung, die für die ersten Zwecke jedoch ausreicht. Später werden wir eine exakte Definition nachreichen; insbesondere werden wir dann umgekehrt vorgehen: der Flächeninhalt wird definiert mit Hilfe des bestimmten Integrales.

Das *unbestimmte Integral*  $\int f(x) dx$  bezeichnet die Menge aller derjenigen Funktionen  $F = F(x)$ , deren Ableitung  $F'(x)$  gleich  $f(x)$  ist. Solche Funktionen  $F$  heißen *Stammfunktionen*.

Die Beziehung zwischen Differenzieren und Integrieren wird durch den folgenden Satz hergestellt, der besagt, daß jedes bestimmte Integral mit variabler oberer Grenze eine Stammfunktion ist.

**Satz 1.4 (Hauptsatz der Differential- und Integralrechnung).** *Sei  $f$  eine stetige Funktion und  $[a, b] \subset D_f$ . Sei  $F = F(x)$  die Funktion, die auf  $[a, b]$  definiert ist und durch*

$$F(x) = \int_{t=a}^x f(t) dt$$

*gegeben wird. Dann ist  $F$  differenzierbar, und es ist  $F'(x) = f(x)$  für jedes  $x \in [a, b]$ .*

## 1.2 Die komplexen Zahlen

### 1.2.1 Zur Einstimmung: Die Cardanische Formel

Wir schauen uns die Gleichung

$$x^3 + px + q = 0 \tag{1.1}$$

an. Hierbei sind  $p$  und  $q$  gegebene reelle Zahlen, und wir suchen reelle Lösungen  $x$ . Für solche kubischen Gleichungen gibt es ein Lösungsverfahren, das nach GERONIMO CARDANO (1501–1576) benannt ist:

Sei  $D = (\frac{p}{3})^3 + (\frac{q}{2})^2$  und  $w = \sqrt[3]{-\frac{q}{2} + \sqrt{D}}$ . Dann wird eine Lösung gegeben durch

$$x_1 = w - \frac{p}{3w}.$$

Durch naiv-unbekümmertes Einsetzen kann man nachrechnen, daß diese Zahl  $x_1$  tatsächlich eine Lösung von (1.1) ist.

Wenn man den zu dieser Lösung gehörenden Linearfaktor  $(x - x_1)$  abdividiert, also die Polynomdivision  $(x^3 + px + q) : (x - x_1)$  durchführt, kommt man zu einem quadratischen Polynom mit bis zu zwei Nullstellen  $x_2$  und  $x_3$ . Insgesamt hat die Gleichung (1.1) dann maximal 3 Lösungen  $x_1$ ,  $x_2$  und  $x_3$ .

Wir probieren dieses Verfahren mal an einem Beispiel aus:

$$x^3 + \frac{2}{3}x - \frac{20}{27} = 0.$$

Wir kommen auf  $D = \frac{108}{729}$  und  $w = \frac{1}{3} \sqrt[3]{10 + \sqrt{108}}$ . Dies ist eine irrationale Zahl, aber wenn wir daraus dann  $x_1$  berechnen, dann heben sich die Irrationalitäten mysteriöserweise heraus und wir finden

$$x_1 = \frac{2}{3},$$

was auch tatsächlich eine Lösung der kubischen Gleichung ist, wie man durch eine Probe feststellt. Abdividieren des Linearfaktors  $(x - \frac{2}{3})$  führt uns zum quadratischen Polynom  $x^2 + \frac{2}{3}x + \frac{10}{9}$ , welches keine reellen Nullstellen hat.

Als weiteres Beispiel betrachten wir

$$x^3 - 63x + 162 = 0. \tag{1.2}$$

Dann ist  $p = -63$ ,  $q = 162$ , und für  $w$  erhalten wir

$$w = \sqrt[3]{-81 + \sqrt{-2700}}.$$

Und hier ergeben sich mindestens zwei Probleme:

- was soll  $\sqrt{-2700}$  sein,
- wenn wir eine Deutung von  $\sqrt{-2700}$  gefunden haben sollten, was ist dann die dritte Wurzel aus  $-81 + \sqrt{-2700}$ ?

Wir könnten uns jetzt auf den Standpunkt stellen, daß diese Formeln keinen Sinn haben, und die Gleichung (1.2) eben keine einzige Lösung hat. Aber verwirrenderweise gibt es Lösungen, und zwar gleich drei, nämlich die reellen Zahlen 3, 6 und  $-9$ , wie man leicht nachrechnet.

CARDANO und viele seiner Kollegen in den nachfolgenden Jahrhunderten gewannen den Eindruck, daß es nützlich sein kann, mit solchen Ausdrücken zu rechnen. Allerdings gelang es lange nicht, Terme der Form  $\sqrt{-2700}$  zu interpretieren. Man meinte, daß Quadratwurzeln aus negativen Zahlen „nur in der Einbildung“ existieren, und bezeichnete solche Zahlen demzufolge als *imaginäre* Zahlen, im Gegensatz zu *reellen* Zahlen, die eine Entsprechung in der Wirklichkeit haben.

Drei Jahrhunderte später, 1831 durch CARL FRIEDRICH GAUSS (1777–1855) und 1837 durch WILLIAM R. HAMILTON (1805–1865), gelang es endlich, diese imaginären Zahlen exakt und logisch sauber einzuführen. Das sollten wir jetzt auch tun.

## 1.2.2 Die komplexen Zahlen

**Definition 1.5.** <sup>1</sup> *Unter einer komplexen Zahl<sup>2</sup>  $z$  verstehen wir ein geordnetes Paar  $(x, y) \in \mathbb{R}^2$ . Wir bezeichnen  $x$  als Realteil<sup>3</sup> und  $y$  als Imaginärteil<sup>4</sup> der komplexen Zahl  $z = (x, y)$ ; dafür schreiben wir:*

$$x = \Re z, \quad y = \Im z.$$

*Die Menge aller komplexen Zahlen wird mit  $\mathbb{C}$  bezeichnet.*

<sup>1</sup>Einige Worte zur Typographie: Definitionen und Sätze sind in *italics* gesetzt. Innerhalb einer Definition steht der zu definierende Begriff in aufrechten Lettern. Definitionen und Sätze sind dort zu Ende, wo der Textsatz in *italics* aufhört. Das Ende eines Beweises markieren wir mit  $\square$ .

<sup>2</sup>auf Englisch: complex number

<sup>3</sup>real part

<sup>4</sup>imaginary part

Wir sagen, daß zwei komplexe Zahlen  $z = (x, y)$  und  $w = (u, v)$  gleich sind, wenn sie in ihren Realteilen bzw. Imaginärteilen übereinstimmen, das heißt  $x = u$  und  $y = v$ .

Komplexe Zahlen kann man addieren und multiplizieren, und zwar wie folgt:

**Definition 1.6.** Seien  $z = (x, y)$  und  $w = (u, v)$  komplexe Zahlen. Dann definieren wir die Summe  $z + w$  und das Produkt  $z \cdot w$  als

$$z + w := (x + u, y + v), \quad z \cdot w := (xu - yv, xv + yu).$$

Wir vermerken kurz, daß  $(0, 1) \cdot (0, 1) = (-1, 0)$ .

**Lemma 1.7.**<sup>5</sup> Addition und Multiplikation komplexer Zahlen sind kommutativ und assoziativ. Außerdem gilt das Distributivgesetz. Die komplexe Zahl  $(0, 0)$  ist neutrales Element für die Addition, und die komplexe Zahl  $(1, 0)$  ist neutrales Element für die Multiplikation.

Das bedeutet in Formeln: für beliebige  $(p, q), (x, y), (u, v) \in \mathbb{C}$  gilt:

$$\begin{aligned} (p, q) + (x, y) &= (x, y) + (p, q), \\ \left( (p, q) + (x, y) \right) + (u, v) &= (p, q) + \left( (x, y) + (u, v) \right), \\ (q, p) \cdot (x, y) &= (x, y) \cdot (p, q), \\ \left( (p, q) \cdot (x, y) \right) \cdot (u, v) &= (p, q) \cdot \left( (x, y) \cdot (u, v) \right), \\ (p, q) \cdot \left( (x, y) + (u, v) \right) &= (p, q) \cdot (x, y) + (p, q) \cdot (u, v), \\ (0, 0) + (x, y) &= (x, y), \\ (1, 0) \cdot (x, y) &= (x, y). \end{aligned}$$

Beweis als Übungsaufgabe.

Es ist klar, wie man die Subtraktion als Umkehrung der Addition einführt: nämlich komponentenweise.

Es fehlt uns bloß noch die Division, um zu zeigen, daß die komplexen Zahlen ebenfalls einen Körper bilden, siehe auch Bemerkung 1.1. Wir verschieben die Einführung der Division für einen Moment. Als Vorbereitung schauen wir uns vorher erst diejenigen komplexen Zahlen genauer an, deren zweite Komponente Null ist:

**Lemma 1.8.** Es ist

$$(x, 0) + (u, 0) = (x + u, 0) \quad \text{und} \quad (x, 0) \cdot (u, 0) = (x \cdot u, 0).$$

Beweis als Übungsaufgabe.

Wir beobachten, daß die komplexen Zahlen mit verschwindender<sup>6</sup> zweiter Komponente sich wie reelle Zahlen verhalten. Es ist z.B.  $5 + 7 = 12$  und  $(5, 0) + (7, 0) = (12, 0)$ . Anstelle von  $(x, 0)$  könnten wir in Zukunft einfach  $x$  schreiben und somit Tinte sparen.

*Jede reelle Zahl kann als komplexe Zahl interpretiert werden.*

Wir definieren:

**Definition 1.9.** Die komplexe Zahl  $(0, 1)$  wird als imaginäre Einheit<sup>7</sup>  $i$  bezeichnet,

$$i := (0, 1).$$

Weiterhin rechnet man nach, daß für  $z = (x, y)$  gilt:

$$z = (x, y) = (x, 0) + (0, y) = (x, 0) + (0, 1) \cdot (y, 0).$$

In unserer neuen Schreibweise lautet das  $z = x + i \cdot y$ .

<sup>5</sup> Ein Lemma ist ein kleiner Satz oder oft auch ein Hilfssatz. Die nach Bedeutung abgestufte Folge der beweisbaren mathematischen Aussagen lautet *Theorem, Satz, Lemma*.

<sup>6</sup> Wir sagen, daß eine Zahl *verschwindet*, wenn sie Null wird.

<sup>7</sup>imaginary unit

Die Formel  $(0, 1) \cdot (0, 1) = (-1, 0)$  kann man also auch schreiben als

$$i \cdot i = -1.$$

Und, allgemeiner, die Formeln für die Addition und Multiplikation lauten in der neuen Schreibweise:

$$\begin{aligned} z + w &= (x + iy) + (u + iv) = (x + u) + i(y + v), \\ z \cdot w &= (x + iy) \cdot (u + iv) = (xu - yv) + i(xv + yu). \end{aligned}$$

Wir rechnen also wie gewohnt und beachten dabei lediglich die Sonderregel  $i^2 = -1$ .

Bevor wir zur Division kommen, brauchen wir noch 2 Begriffe:

**Definition 1.10.** Sei  $z = (x, y) \in \mathbb{C}$  eine komplexe Zahl. Dann heißt

$$\bar{z} := (x, -y) = x - iy$$

die komplex konjugierte Zahl<sup>8</sup> zu  $z$ , und

$$|z| := \sqrt{x^2 + y^2}$$

heißt Betrag<sup>9</sup> von  $z$ .

Man zeigt schnell, daß die folgenden Rechenregeln gelten:

**Satz 1.11.** Seien  $z = (x, y)$  und  $w = (u, v)$  komplexe Zahlen. Dann gilt:

$$\begin{aligned} \overline{z + w} &= \bar{z} + \bar{w}, \\ \overline{z \cdot w} &= \bar{z} \cdot \bar{w}, \\ \overline{\bar{z}} &= z, \\ z \cdot \bar{z} &= (|z|^2, 0). \end{aligned}$$

In Worten zusammengefaßt: es ist egal, ob wir zuerst konjugieren und dann addieren (multiplizieren), oder umgekehrt. Diese beiden Regeln können wir auch als kommutative Diagramme veranschaulichen:



Das Fragezeichen ist dabei als Platz-Freihalter zum Einsetzen zu lesen. Wir bezeichnen ein Diagramm als *kommutativ*, wenn jeder Pfad von der Ecke links oben zur Ecke rechts unten das gleiche Ergebnis liefert.

Die letzte Regel in Satz 1.11 bedeutet, daß das Produkt aus einer komplexen Zahl und ihrer Konjugierten stets reell ist.

**Satz 1.12.** Die Betragsfunktion erfüllt die Eigenschaften einer Norm (siehe Bemerkung 1.3). Das heißt:

$$\begin{aligned} |z| &\geq 0; \quad |z| = 0 \text{ genau dann, wenn } z = (0, 0), \\ |z \cdot w| &= |z| \cdot |w|, \\ |z + w| &\leq |z| + |w|. \end{aligned}$$

*Beweis.* Übungsaufgabe. □

<sup>8</sup>complex conjugate

<sup>9</sup>absolute value



Die erste und die letzte dieser Eigenschaften sind geometrisch unmittelbar einsichtig, wenn man sich die komplexen Zahlen (die ja geordnete Paare reeller Zahlen sind) als Vektoren in der Ebene vorstellt. Der Betrag einer komplexen Zahl ist nichts anderes als die geometrische Länge des zugeordneten Vektors.

Nun zur Division. Gegeben sind zwei komplexe Zahlen  $w = u + iv$  und  $r = p + iq$ , und gesucht ist deren Quotient  $z = x + iy$ , also soll gelten

$$r \cdot z = w.$$

Wir multiplizieren mit  $\bar{r}$ :

$$\bar{r} \cdot w = \bar{r} \cdot (r \cdot z) = (\bar{r} \cdot r) \cdot z = |r|^2 \cdot z = |r|^2 \cdot (x + iy) = |r|^2 x + i(|r|^2 y).$$

Jetzt nutzen wir aus, daß wir auf der ganz linken Seite eine bekannte komplexe Zahl stehen haben:

$$\bar{r} \cdot w = \overline{(p + iq)} \cdot (u + iv) = (p - iq) \cdot (u + iv) = (pu + qv) + i(pv - qu).$$

Wir vergleichen Real- und Imaginärteil in den beiden Formelzeilen:

$$x = \frac{pu + qv}{|r|^2}, \quad y = \frac{pv - qu}{|r|^2}.$$

Das Ganze wird vielleicht etwas einsichtiger in folgender Schreibweise:

$$z = \frac{w}{r} = \frac{u + iv}{p + iq} = \frac{(u + iv)(p - iq)}{(p + iq)(p - iq)} = \frac{(pu + qv) + i(pv - qu)}{p^2 + q^2} = \frac{pu + qv}{p^2 + q^2} + i \frac{pv - qu}{p^2 + q^2}.$$

Damit haben wir gezeigt, daß die Menge der komplexen Zahlen, zusammen mit der Addition und der Multiplikation, einen Körper bildet, was nichts anderes heißt, daß wir mit den 4 Grundrechenarten umgehen können, wie wir es gewohnt sind.

Leider ist es nicht möglich, auf  $\mathbb{C}$  eine Ordnungsrelation zu definieren, die sich mit den Rechenoperationen verträgt (warum?).

Wir wenden uns nun nochmal der geometrischen Veranschaulichung der komplexen Zahlen als Vektoren (bzw. Punkte) der zweidimensionalen Ebene zu. Wir bezeichnen die horizontale Achse als *reelle Achse*, und die vertikale Achse als *imaginäre Achse*. Durch Hinschauen stellen wir dann fest, daß für jede komplexe Zahl  $z \neq 0$  gilt:

$$\Re z = |z| \cos \varphi, \quad \Im z = |z| \sin \varphi.$$

Hierbei bezeichnet  $\varphi$  den Winkel zwischen dem positiven Teil der reellen Achse und dem Vektor, der zu  $z$  gehört. (Ab jetzt wollen wir (aus sprachlichen Gründen) meistens nicht mehr unterscheiden zwischen einer komplexen Zahl und dem Vektor, der diese komplexe Zahl geometrisch veranschaulicht.) Dieser Winkel  $\varphi$  wird auch als *Argument von  $z$*  bezeichnet.

Was passiert bei Addition und Multiplikation?

Man sieht schnell, daß die Addition komplexer Zahlen der gewöhnlichen Vektoraddition entspricht.

Für die Multiplikation betrachten wir zwei komplexe Zahlen

$$z = x + iy = |z|(\cos \varphi + i \sin \varphi), \quad w = u + iv = |w|(\cos \psi + i \sin \psi),$$

und erhalten aus den schulbekannten Additionstheoremen, daß

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

Das Produkt von  $z$  und  $w$  ist also eine komplexe Zahl, deren Betrag gleich dem Produkt der Beträge von  $z$  und  $w$  ist; und das Argument von  $z \cdot w$  ist gleich der Summe der Argumente von  $z$  und  $w$ .

*Die Multiplikation komplexer Zahlen entspricht einer Drehstreckung.*

Interessant ist noch der Fall  $z = w$ , dann haben wir

$$z^2 = |z|^2(\cos(2\varphi) + i \sin(2\varphi)), \quad \text{wenn } z = |z|(\cos \varphi + i \sin \varphi).$$

Mittels vollständiger Induktion bekommen wir die sogenannte Formel von Moivre (ABRAHAM DE MOIVRE, 1667–1754):

$$z^n = |z|^n(\cos(n\varphi) + i \sin(n\varphi)), \quad \text{wenn } z = |z|(\cos \varphi + i \sin \varphi) \text{ und } n \in \mathbb{N}.$$

### 1.2.3 Funktionen komplexer Zahlen

Genauso wie man Funktionen von reellen Variablen definieren und untersuchen kann, kann man dies auch mit Funktionen von komplexen Zahlen tun.

Einfache Beispiele für solche Funktionen sind Polynome,

$$P(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0, \quad n \in \mathbb{N},$$

mit Koeffizienten  $a_j \in \mathbb{C}$  und Definitionsbereich  $\mathbb{C}$ .

Ein anderes — und sehr wichtiges — Beispiel ist die komplexe Exponentialfunktion. Bekanntlich erfüllt die herkömmliche Exponentialfunktion die Beziehungen

$$e^{x_1+x_2} = e^{x_1} \cdot e^{x_2}, \quad e^{-x_1} = \frac{1}{e^{x_1}}, \quad e^0 = 1$$

für beliebige reelle  $x_1, x_2 \in \mathbb{R}$ .

**Definition 1.13.** Sei  $z = x + iy \in \mathbb{C}$  eine komplexe Zahl. Dann definieren wir

$$e^z = e^{x+iy} := e^x (\cos y + i \sin y).$$

Diese komplexwertige Funktion mit Definitionsbereich  $\mathbb{C}$  heißt komplexe Exponentialfunktion.

Später werden wir eine andere Definition bevorzugen, vgl. Definition 5.70. Weiterhin ist  $e^{x+iy}$  im Moment einfach nur eine Schreibweise. Wir denken ausdrücklich nicht daran, eine Zahl  $e \approx 2.71828\dots$  in die  $(x + iy)$ -te Potenz zu erheben, denn eine solche Potenz ist anschaulich kaum vorstellbar.

**Satz 1.14.** Die komplexe Exponentialfunktion hat die folgenden Eigenschaften:<sup>10</sup>

$$\begin{aligned} e^0 &= 1, \\ e^{z+w} &= e^z \cdot e^w, \quad z, w \in \mathbb{C}, \\ e^{-z} &= \frac{1}{e^z}, \quad z \in \mathbb{C}. \end{aligned}$$

*Beweis.* Wir haben  $e^0 = e^{0+i0} = e^0(\cos 0 + i \sin 0) = 1(1 + 0) = 1$ .

Seien nun  $z = x + iy$  und  $w = u + iv$ . Dann ist einerseits

$$e^{z+w} = e^{(x+u)+i(y+v)} = e^{x+u}(\cos(y+v) + i \sin(y+v)),$$

andererseits, wegen der Additionstheoreme für die Sinus- und Kosinusfunktionen,

$$\begin{aligned} e^z \cdot e^w &= e^x(\cos y + i \sin y)e^u(\cos v + i \sin v) \\ &= e^{x+u}((\cos y \cos v - \sin y \sin v) + i(\cos y \sin v + \sin y \cos v)) \\ &= e^{x+u}(\cos(y+v) + i \sin(y+v)). \end{aligned}$$

Also ist  $e^{z+w} = e^z \cdot e^w$ .

Und die dritte Beziehung beweist sich jetzt fast von selbst:

$$1 = e^0 = e^{z+(-z)} = e^z \cdot e^{-z}.$$

□

Wenn wir annehmen, daß  $z$  den Realteil Null hat (also eine sogenannte *rein imaginäre* Zahl ist), dann folgt die Beziehung

$$e^{iy} = \cos y + i \sin y, \quad y \in \mathbb{R}.$$

Falls wir hier  $y = \pi$  setzen, entsteht die berühmte Gleichung

$$e^{i\pi} + 1 = 0,$$

<sup>10</sup> die Schreibweise „...“,  $z, w \in \mathbb{C}$  am Ende der zweiten Formelzeile soll ausdrücken, daß die vorher geschriebene Aussage für *alle*  $z, w \in \mathbb{C}$  gelten möge. In Zukunft werden wir sehr oft diese abgekürzte Ausdrucksweise gebrauchen.

die sämtliche fünf für den Physiker unverzichtbaren Zahlen in einer Zeile vereinigt.

Und wenn  $z$  den Imaginärteil Null hat (also reell ist), dann erhalten wir aus  $\cos 0 = 1$  und  $\sin 0 = 0$ , daß

$$e^z = e^x.$$

Also stimmt die neu definierte komplexe Exponentialfunktion mit der herkömmlichen reellen Exponentialfunktion überein, wenn das Argument der komplexen Exponentialfunktion zufälligerweise reell ist.

Für eine komplexe Zahl  $z = x + iy$  mit Betrag  $|z|$  und Argument  $\varphi$  haben wir jetzt drei äquivalente Schreibweisen

$$\begin{aligned} z &= x + iy, \\ z &= |z|(\cos \varphi + i \sin \varphi), \\ z &= |z|e^{i\varphi}, \end{aligned}$$

die wir im Folgenden gleichberechtigt verwenden werden.

Nun zu den Winkelfunktionen. Wir können die Winkelfunktionen im Reellen mittels der Exponentialfunktion ausdrücken:

**Satz 1.15.** *Sei  $\varphi \in \mathbb{R}$ . Dann gilt*

$$\cos \varphi = \frac{1}{2}(e^{i\varphi} + e^{-i\varphi}), \quad \sin \varphi = \frac{1}{2i}(e^{i\varphi} - e^{-i\varphi}).$$

*Beweis.* Wir haben

$$e^{i\varphi} = \cos \varphi + i \sin \varphi, \quad e^{-i\varphi} = \cos(-\varphi) + i \sin(-\varphi) = \cos \varphi - i \sin \varphi,$$

denn die Kosinusfunktion ist gerade<sup>11</sup>, und die Sinusfunktion ist ungerade<sup>12</sup>. Jetzt brauchen wir diese beiden Gleichungen bloß zueinander addieren bzw. voneinander abziehen, und der Beweis ist vollendet.  $\square$

Damit können wir jetzt auch Winkelfunktionen für komplexe Zahlen erklären:

**Definition 1.16.** *Für  $z \in \mathbb{C}$  definieren wir  $\sin z$  und  $\cos z$  durch*

$$\sin z := \frac{1}{2i}(e^{iz} - e^{-iz}), \quad \cos z := \frac{1}{2}(e^{iz} + e^{-iz}).$$

**Frage:** Für  $x \in \mathbb{R}$  gilt bekanntlich  $|\sin x| \leq 1$  sowie  $|\cos x| \leq 1$  und  $\cos^2(x) + \sin^2(x) = 1$ . Welche dieser Beziehungen bleiben gültig, wenn man  $x \in \mathbb{R}$  ersetzt durch  $z \in \mathbb{C}$ ?

Falls  $z$  in dieser Definition reell sein sollte, erhalten wir genau die herkömmlichen Winkelfunktionen. Wir haben also diese Winkelfunktionen von der Menge der reellen Zahlen in die Menge der komplexen Zahlen *fortgesetzt*, im Sinne einer Vergrößerung des Definitionsbereiches. Wir könnten noch weitere Funktionen von den reellen Zahlen in die komplexen Zahlen fortsetzen, verschieben das aber auf ein späteres Kapitel. Stattdessen wollen wir uns überlegen, welche Eigenschaften Funktionen haben könnten.

Zum Beispiel könnte eine Funktion  $w : \mathbb{C} \rightarrow \mathbb{C}$  stetig<sup>13</sup> sein. Was heißt das? Wir können die Stetigkeit genauso definieren wie bei Funktionen  $w : \mathbb{R} \rightarrow \mathbb{R}$ . In Worten ausgedrückt, ist die Stetigkeit der Funktion  $w : \mathbb{C} \rightarrow \mathbb{C}$  im Punkte  $z_0 \in \mathbb{C}$  folgendermaßen definiert:

Wir können erzwingen, daß der Wert  $|w(z) - w(z_0)|$  beliebig klein wird, wenn wir nur dafür sorgen, daß  $|z - z_0|$  klein genug ist.

Hierbei bedeuten die Betragsstriche natürlich den komplexen Betrag.

Die bisher betrachteten Funktionen (Polynome, komplexe Exponentialfunktion, komplexe Winkelfunktionen) sind allesamt in ganz  $\mathbb{C}$  stetig.

Eine weitere interessante Eigenschaft einer Funktion ist die Differenzierbarkeit. Hierbei müssen wir allerdings etwas aufpassen und zwei Fälle unterscheiden.

<sup>11</sup>in der Schule sagt man gelegentlich „achsensymmetrisch“

<sup>12</sup>„punktsymmetrisch“

<sup>13</sup> Es lohnt sich, für die Schreibweise „ $w : \mathbb{C} \rightarrow \mathbb{C}$ “ eine Lesevorschrift mitzuteilen. Der Doppelpunkt bedeutet „bewirkt, daß“. Das erste  $\mathbb{C}$  gibt den Definitionsbereich der Funktion  $w$  an, und die Passage „ $\rightarrow \mathbb{C}$ “ lesen wir als „abgebildet wird nach  $\mathbb{C}$ “. Insgesamt also „ $w$  bewirkt, daß  $\mathbb{C}$  abgebildet wird nach  $\mathbb{C}$ “. Das zweite  $\mathbb{C}$  enthält (als Teilmenge) die Menge der von  $w$  tatsächlich angenommenen Werte, wobei es erlaubt ist, daß nicht ganz  $\mathbb{C}$  von den tatsächlich angenommenen Funktionswerten von  $w$  ausgeschöpft wird.

- Funktionen  $w : \mathbb{R} \rightarrow \mathbb{C}$ ,
- Funktionen  $w : \mathbb{C} \rightarrow \mathbb{C}$ .

Im ersten Fall können wir uns vorstellen, daß die Funktion  $w$  auf einem Intervall  $[a, b] \subset \mathbb{R}$  definiert ist, wir können also schreiben

$$w = w(t) = u(t) + iv(t), \quad t \in [a, b].$$

Naheliegenderweise wird man die Ableitung  $w'(t)$  komponentenweise definieren wollen:

$$w'(t) := u'(t) + iv'(t), \quad t \in [a, b].$$

Hierbei haben wir stillschweigend angenommen, daß die reellwertigen Funktion  $u, v : [a, b] \rightarrow \mathbb{R}$  im herkömmlichen Sinne differenzierbar sind.

Im zweiten Fall haben wir es mit einer Funktion  $w = w(z)$  zu tun, wobei jetzt  $z$  aus ganz  $\mathbb{C}$  stammen darf. Wenn wir  $w = u + iv$  und  $z = x + iy$  schreiben, haben wir

$$w(z) = u(x + iy) + iv(x + iy), \quad x + iy \in \mathbb{C},$$

und wir definieren die Ableitung als „Grenzwert der Differenzenquotienten“. Und zwar:

Wenn der Grenzwert

$$\lim_{z \rightarrow z_0} \frac{w(z) - w(z_0)}{z - z_0}, \quad z \in \mathbb{C},$$

existiert, dann sagen wir, daß die Funktion  $w : \mathbb{C} \rightarrow \mathbb{C}$  im Punkte  $z_0 \in \mathbb{C}$  differenzierbar ist, und setzen  $w'(z_0)$  gleich diesem Grenzwert. Wir verlangen dabei, daß  $z$  auf *beliebigem* Wege gegen  $z_0$  laufen darf (z.B. entlang einer heftig verknöteten Kurve, oder auch punktweise wild herumhopsend); und jedesmal soll der Grenzwert des Differenzenquotienten denselben Wert haben. Eine saubere Definition kommt später, siehe Definition 6.39.

**Warnung 1.17.** Die Ableitung ist jetzt eine komplexe Zahl, und es ist nur schwer möglich, den Wert der Ableitung in gewohnter Form als „Anstieg der Tangente“ zu interpretieren.

Diese Ableitung hat genau die gleichen Eigenschaften wie die herkömmliche Ableitung reellwertiger Funktionen:

**Satz 1.18.** Seien  $w = w(z)$  und  $r = r(z)$  differenzierbare Funktionen von  $\mathbb{C}$  nach  $\mathbb{C}$ . Dann gilt:

$$\begin{aligned} (w + r)'(z) &= w'(z) + r'(z), \quad z \in \mathbb{C}, \\ (cw)'(z) &= c \cdot w'(z), \quad z \in \mathbb{C}, \quad c \in \mathbb{C}, \\ (w \cdot r)'(z) &= w'(z) \cdot r(z) + w(z) \cdot r'(z), \quad z \in \mathbb{C}, \\ \left(\frac{w}{r}\right)'(z) &= \frac{w'(z) \cdot r(z) - w(z) \cdot r'(z)}{r(z)^2}, \quad z \in \mathbb{C}, \quad \text{falls } r(z) \neq 0, \\ w(r(z))' &= w'(r(z)) \cdot r'(z), \quad z \in \mathbb{C}. \end{aligned}$$

Die Beweise dieser Formeln verlaufen genauso wie im reellen Fall. Dies ist deshalb möglich, weil die reellen Zahlen genauso wie die komplexen Zahlen einen Körper bilden; mit anderen Worten, daß die Grundrechenarten denselben Regeln genügen.

Genauso wie im reellen Fall kann man nachweisen, daß jede differenzierbare Funktion  $w : \mathbb{C} \rightarrow \mathbb{C}$  stetig ist. Außerdem sind die oben vorgestellten Funktionen differenzierbar, und es ist

$$\begin{aligned} (z^n)' &= nz^{n-1}, \quad z \in \mathbb{C}, \quad n \in \mathbb{N} = \{1, 2, \dots\}, \\ (e^z)' &= e^z, \quad z \in \mathbb{C}, \\ \sin'(z) &= \cos(z), \quad z \in \mathbb{C}, \\ \cos'(z) &= -\sin(z), \quad z \in \mathbb{C}. \end{aligned}$$

**Warnung 1.19.** Leider gibt es Funktionen, die wunderbar harmlos ausschauen, aber nicht differenzierbar sind. Ein Beispiel einer solchen Funktion ist  $w = w(z) = |z|^2$ . Diese Funktion ist lediglich im Ursprung differenzierbar, und sonst nirgendwo. Sei z.B.  $z_0 = 1 + 2i$ . Wenn wir  $z$  einmal horizontal gegen  $z_0$  schicken und einmal vertikal, bekommen wir für den Grenzwert des Differenzenquotienten zwei verschiedene Werte, wie man schnell nachrechnet. Deshalb ist  $w$  im Punkte  $z_0 = 1 + 2i$  nicht differenzierbar. Wir müssen eine genauere Betrachtung der komplexen Differentialrechnung auf später verschieben.

### 1.2.4 Ausblick: Elektrotechnik

Hier ergibt sich ein kleines Problem mit den Schreibweisen: in der Elektrotechnik bezeichnet man die Stromstärke mit  $i$  bzw.  $I$ , weshalb wir für die imaginäre Einheit in diesem Abschnitt  $j$  schreiben,  $j^2 = -1$ .

Einen elektrischen Wechselstrom mit Amplitude  $i_m$ , Kreisfrequenz  $\omega = 2\pi f$  und Phasenverschiebung  $\varphi_i$  können wir ausdrücken als

$$i(t) = i_m \cos(\omega t + \varphi_i), \quad i_m > 0, \quad \omega > 0, \quad \varphi_i \in \mathbb{R}, \quad t \in \mathbb{R}.$$

Wenn wir für den sogenannten *Effektivwert* der Stromstärke  $I$  schreiben,  $I := i_m/\sqrt{2}$ , dann erhalten wir

$$i(t) = \Re\left(\sqrt{2}I e^{j\varphi_i} e^{j\omega t}\right).$$

Der Ausdruck  $I e^{j\varphi_i}$  ist eine komplexe Zahl, die nicht von der Zeit  $t$  abhängt. Um die Schreibweise zu vereinfachen, führen wir den *komplexen Effektivwert* ein:

$$\underline{I} := I e^{j\varphi_i}$$

und es folgt

$$i(t) = \sqrt{2}\Re\left(\underline{I} e^{j\omega t}\right).$$

Die Unterstriche sollen ausdrücken, daß die betreffende Größe sich wie ein „Zeiger“ benimmt. Genauso verfahren wir mit der Spannung:

$$u(t) = u_m \cos(\omega t + \varphi_u), \quad u_m > 0, \quad \omega > 0, \quad \varphi_u \in \mathbb{R}, \quad t \in \mathbb{R},$$

$$U := u_m/\sqrt{2}, \quad \underline{U} := U e^{j\varphi_u},$$

$$u(t) = \sqrt{2}\Re\left(\underline{U} e^{j\omega t}\right).$$

Die komplexen Zahlen  $\underline{I} e^{j\omega t}$  und  $\underline{U} e^{j\omega t}$  drehen sich in der komplexen Ebene im Gegenuhrzeigersinn, wenn  $t$  wächst; und die Radien der Kreisbahnen sind  $\underline{I}$  bzw.  $\underline{U}$ . Für einen Umlauf brauchen diese Vektoren die Zeit  $1/f$ , so wie man es erwartet.

Der Winkel zwischen den Vektoren  $\underline{I} e^{j\omega t}$  und  $\underline{U} e^{j\omega t}$  ist immer gleich, nämlich  $\varphi_i - \varphi_u$ . Dies ist genau die Phasenverschiebung zwischen Stromstärke und Spannung.

Spannung und Stromstärke stehen zueinander in Beziehung über den Widerstand. Wir haben die folgenden 3 Fälle:

**Ohmscher Widerstand:**  $\underline{U} = R\underline{I}$ ,  $R \in \mathbb{R}$ ,

**Spule (induktiver Widerstand):**  $\underline{U} = j\omega L\underline{I}$ ,  $L \in \mathbb{R}$ ,

**Kondensator (kapazitiver Widerstand):**  $\underline{U} = -j\frac{1}{\omega C}\underline{I}$ ,  $C \in \mathbb{R}$ .

Nun führen wir den *komplexen Scheinwiderstand* ein:

$$\underline{Z} = \frac{\underline{U}}{\underline{I}}.$$

Mit diesem Scheinwiderstand können wir genauso rechnen wie mit ohmschen Widerständen: bei einer Reihenschaltung addieren sich die Scheinwiderstände, bei einer Parallelschaltung addieren sich die Reziproken der Scheinwiderstände.

Für die Schaltung in Abbildung 1.1 haben wir zum Beispiel

$$\underline{Z} = \frac{1}{\frac{1}{R_1 - j/(\omega C_1)} + \frac{1}{R_2 + j\omega L}} - j\frac{1}{\omega C_2}.$$

Wenn man den Scheinwiderstand  $\underline{Z}$  in der Form

$$\underline{Z} = R + jX, \quad R, X \in \mathbb{R},$$

schreibt, dann heißt  $R = \Re\underline{Z}$  der *Wirkwiderstand* und  $X = \Im\underline{Z}$  der *Blindwiderstand* der Schaltung. Es ist eine interessante Aufgabe, mathematisch präzise zu beweisen, daß bei „allen typischen“ Schaltungen der Wirkwiderstand niemals negativ werden kann. Äquivalent dazu wäre zu zeigen, daß Strom und Spannung nicht zueinander um mehr als  $\pi/2$  phasenverschoben sein können.

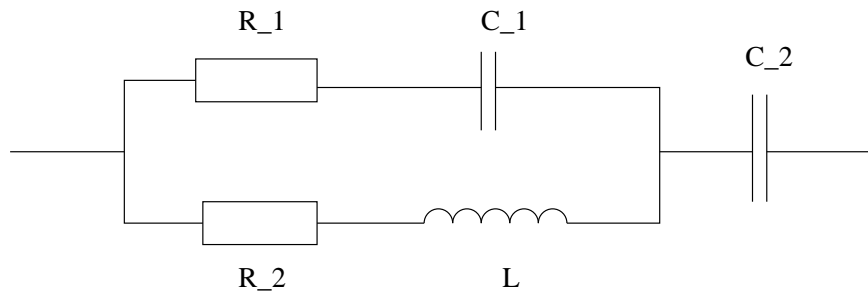


Abbildung 1.1: Elektrische Schaltung

### 1.2.5 Ausblick: Mechanik

Eine ebene Welle, die sich im Ortsraum  $\mathbb{R}^3$  ausbreitet, wird häufig modelliert mittels Funktionen der Form

$$(t, x) \mapsto e^{i(kx - \omega t)}.$$

Der Pfeil  $\mapsto$  besagt, daß das Paar  $(t, x)$  abgebildet wird auf  $e^{i(kx - \omega t)} \in \mathbb{C}$ . Hierbei benennen  $t \in \mathbb{R}$  und  $x \in \mathbb{R}^3$  die Variablen für Zeit und Ort, ihre Einheiten sind natürlich *Sekunde* und *Meter*. Weiterhin ist  $k = (k_1, k_2, k_3)$  ein Vektor, der senkrecht auf den *Wellenfronten* steht und in der Physik *Wellenzahlvektor* genannt wird. Seine Einheit ist  $\frac{1}{\text{Meter}}$ . Das Produkt  $kx$  ist zu verstehen als  $k_1x_1 + k_2x_2 + k_3x_3$ . Und  $\omega$  (omega) heißt oft *Kreisfrequenz* mit der Einheit  $\frac{1}{\text{Sekunde}}$ . Schließlich ist natürlich  $i^2 = -1$ . Der Zusammenhang zwischen diesen Parametern ist  $c = \frac{\omega}{|k|} = \frac{\omega}{\sqrt{k_1^2 + k_2^2 + k_3^2}}$ , mit  $c$  als *Ausbreitungsgeschwindigkeit* der Welle.

**Frage:** Warum liefert dies tatsächlich eine Beschreibung von ebenen Wellen im  $\mathbb{R}^3$  ?

**Frage:** Sei  $k \in \mathbb{R}^3$  gegeben. Wie sieht die Menge aller  $x \in \mathbb{R}^3$  geometrisch aus, für die  $kx = \text{const.}$  ?

## 1.3 Die Ebene und der $\mathbb{R}^2$

### 1.3.1 Allgemeine Eigenschaften

Wir starten mit einigen Begriffen, die bekannt sein sollten:

**Ebene:** genauer gesagt, die Menge der Ortsvektoren der (zweidimensionalen) Ebene. *Achtung:* dies ist nicht dasselbe wie der  $\mathbb{R}^2$  !

$\mathbb{R}^2$ : die Menge aller geordneten Paare<sup>14</sup>  $(x, y)$  mit  $x, y \in \mathbb{R}$ .

**Ortsvektor:** ein Vektor („Pfeil“), der im Ursprung anfängt.

**Addition von Ortsvektoren:** ergibt wieder einen Ortsvektor.

**Multiplikation eines Ortsvektors mit einer reellen Zahl:** ergibt wieder einen Ortsvektor.

**Kartesisches Koordinatensystem:** besteht aus zwei Koordinatenachsen, die einander im Ursprung schneiden (und zwar rechtwinklig), und je einem Einheitsvektor pro Achse. Diese beiden Einheitsvektoren sind gleichlang.

**Bemerkung 1.20.** In manchen Büchern (auch in der Schule) werden Vektoren definiert als Äquivalenzklassen von Pfeilen, wobei zwei Pfeile dann als äquivalent gelten, wenn sie durch Parallelverschiebung auseinander hervorgehen.

Von einer solchen Definition hält der Autor sehr wenig. Zum einen gehen diese Äquivalenzklassenbetrachtungen am eigentlichen Wesensgehalt von Vektoren vorbei, und zum zweiten sind physikalisch relevante Vektoren eben keine Äquivalenzklassen (man denke an die Kraftvektoren an einer Wippe).

<sup>14</sup> Die Formulierung „geordnetes Paar“ soll lediglich ausdrücken, daß die Reihenfolge der beiden Einträge im Paar wichtig ist. Es ist also  $(2, 3) \neq (3, 2)$ , wie bereits aus dem schulischen Umgang mit Koordinatensystemen bekannt.

Wichtig ist, sich folgendes klar zu machen:

Die Addition zweier Vektoren und die Multiplikation eines Vektors mit einer reellen Zahl kann man rein geometrisch definieren, völlig ohne den Begriff eines Koordinatensystems! Es ergibt sich z.B.  $\vec{a} + \vec{b}$  als der Diagonalen Ortsvektor des von den Ortsvektoren  $\vec{a}$  und  $\vec{b}$  aufgespannten Parallelogramms. Und die Multiplikation eines Ortsvektors mit einer Zahl kann man sich so vorstellen, daß an der „Pfeilspitze“ in „Pfeilrichtung“ gezogen wird.

Der folgende Satz kann dann bewiesen werden. Beim Beweis ist zu beachten, daß das Konzept der Koordinaten eines Vektors nicht zur Verfügung steht (man hat also nur die Elementargeometrie der Mittelstufe zur Hand), was für die Assoziativität der Addition eine gewisse Erschwernis darstellt.

**Satz 1.21.** Sei  $V$  die Menge der Ortsvektoren der Ebene. Weiterhin seien<sup>15</sup>

$$\begin{aligned} + & : V \times V \rightarrow V, \\ \cdot & : \mathbb{R} \times V \rightarrow V \end{aligned}$$

die geometrisch definierten Operationen „Addition zweier Vektoren“ und „Multiplikation eines Vektors mit einer reellen Zahl“. Dann haben wir die folgenden Eigenschaften für alle Vektoren  $x, y, z \in V$  und alle reellen Zahlen  $\lambda, \mu \in \mathbb{R}$ :

**Kommutativität der Addition:**  $x + y = y + x$

**Assoziativität der Addition:**  $(x + y) + z = x + (y + z)$

**$\vec{0}$  ist neutrales Element der Addition:**  $\vec{0} + x = x + \vec{0} = x$

**Subtraktion:** Jede Gleichung  $a + x = b$  (mit gegebenem  $a, b \in V$ ) ist lösbar, die (eindeutige) Lösung  $x$  wird geschrieben als  $x = b - a$ .

**Assoziativität der Multiplikation:**  $(\lambda \cdot \mu) \cdot x = \lambda \cdot (\mu \cdot x)$

**1 ist neutrales Element der Multiplikation:**  $1 \cdot x = x$

**Zwei Distributivgesetze:** Addition und Multiplikation sind verzahnt gemäß der Regeln  $(\lambda + \mu) \cdot x = \lambda \cdot x + \mu \cdot x$  und  $\lambda \cdot (x + y) = \lambda \cdot x + \lambda \cdot y$ .

**Frage:** Können Sie an einem Beispiel zeigen, daß die Regel  $1 \cdot x = x$  nicht aus den übrigen Regeln geschlußfolgert werden kann? Bei diesem Beispiel wären die Operationen  $+$  und  $\cdot$  anders definiert, und die genannten Eigenschaften würden immer noch gelten, abgesehen von  $1 \cdot x = x$ .

**Bemerkung 1.22.** Wir sagen auch, daß  $(V, +, \cdot)$  einen Vektorraum<sup>16</sup> über  $\mathbb{R}$  bildet.

Es gibt einen Zusammenhang zwischen der zweidimensionalen Ebene und dem  $\mathbb{R}^2$ , der durch ein (kartesisches) Koordinatensystem vermittelt wird: Bekanntlich kann man jeden Ortsvektor der Ebene als Linearkombination der Basisvektoren des Koordinatensystems darstellen. Die Koeffizienten dieser Linearkombination sind reelle Zahlen, bilden also gerade ein geordnetes Paar, also ein Element aus dem  $\mathbb{R}^2$ .

Allerdings haben wir einige Freiheiten bei der Wahl des Koordinatensystems. Es ist zwar üblich, daß die eine Koordinatenachse „nach rechts“ zeigt und die andere „nach oben“, aber das ist keineswegs zwingend. Die Koordinatenachsen könnten auch gespiegelt sein oder „schief liegen“. Und wenn das Koordinatensystem (oder, präziser, das geordnete Paar der Basisvektoren) anders gewählt werden, dann ändern sich auch die Koeffizienten eines Vektors bezüglich des jeweiligen Koordinatensystems.

*Die Koeffizienten eines Vektors hängen von der gewählten Basis ab!*

Im Folgenden haben wir uns für ein Koordinatensystem (also eine Orthonormalbasis) entschieden, das wir vorerst nicht mehr ändern, und dessen Basisvektoren  $e_1 \in V$  und  $e_2 \in V$  seien. Dann können wir jeden Vektor  $x \in V$  schreiben als

$$x = \xi_1 e_1 + \xi_2 e_2, \quad \xi_i \in \mathbb{R}.$$

<sup>15</sup> Für eine Lesehilfe zu den beiden Formelzeilen verweisen wir auf Fußnote 13.

<sup>16</sup>vector space

Auf diesem Wege bekommen wir eine Abbildung von  $V$  in den  $\mathbb{R}^2$ , die jedes  $x \in V$  abbildet auf seine Koordinaten  $\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \in \mathbb{R}^2$ . (Wenn wir eine andere Basis  $(e_1, e_2)$  gewählt hätten, dann sähe diese Abbildung anders aus.) Die Koordinaten von  $e_1$  bezüglich der Basis  $(e_1, e_2)$  sind zum Beispiel  $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ , die Koordinaten von  $e_2$  sind  $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ .

Wenn wir die geometrisch definierten Rechenoperationen auf den  $\mathbb{R}^2$  übertragen, kommen wir zur folgenden Definition:

**Definition 1.23.** Der  $\mathbb{R}^2$  ist definiert als Menge geordneter Paare reeller Zahlen:

$$\mathbb{R}^2 = \left\{ \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} : \xi_i \in \mathbb{R} \right\}.$$

Das Paar  $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$  wird mit  $0$  bezeichnet; außerdem setzen wir  $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$  und  $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ . Dann definieren wir 2 Operationen:

$$\begin{aligned} + : \mathbb{R}^2 \times \mathbb{R}^2 &\rightarrow \mathbb{R}^2, & \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} &:= \begin{pmatrix} \xi_1 + \eta_1 \\ \xi_2 + \eta_2 \end{pmatrix}, \\ \cdot : \mathbb{R} \times \mathbb{R}^2 &\rightarrow \mathbb{R}^2, & \lambda \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} &:= \begin{pmatrix} \lambda \xi_1 \\ \lambda \xi_2 \end{pmatrix}. \end{aligned}$$

Wir stellen schnell fest:

**Satz 1.24.** Der  $\mathbb{R}^2$  hat zusammen mit den Operationen  $+$  und  $\cdot$  die Eigenschaften aus Satz 1.21, ist also ein Vektorraum über  $\mathbb{R}$ .

Ab jetzt werden wir den geometrisch definierten Vektorraum der Ortsvektoren in der Ebene einerseits und den  $\mathbb{R}^2$  andererseits synonym gebrauchen, wobei wir im Hinterkopf behalten, daß diejenige Abbildung zwischen  $V$  und  $\mathbb{R}^2$ , die jedem Vektor seine Koordinaten zuordnet, von der gewählten (Orthonormal-)Basis in  $V$  abhängt.

Geometrisch ist glaubhaft, daß es zwischen 2 Ortsvektoren einen Winkel gibt (es sei denn, einer der Vektoren wäre der Nullvektor), und daß jeder Ortsvektor eine Länge hat. Wenn ein Ortsvektor  $x$  gegeben wird durch  $x = \xi_1 e_1 + \xi_2 e_2$ , dann beträgt seine Länge

$$|x| = \sqrt{\xi_1^2 + \xi_2^2}.$$

Diese Rechnung ist aber nur gültig, weil  $e_1$  und  $e_2$  eine Orthonormalbasis bilden. Wenn wir den Winkel zwischen dem Vektor  $x$  und der  $e_1$ -Achse  $\varphi$  taufen, dann ist

$$\xi_1 = |x| \cos \varphi, \quad \xi_2 = |x| \sin \varphi.$$

Nun betrachten wir noch einen weiteren Vektor  $y$  mit Koordinaten  $\begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}$  und Winkel  $\psi$  zur  $e_1$ -Achse. Der Winkel  $\angle(x, y)$  ist dann gerade  $\alpha = \psi - \varphi$ , und aus dem Additionstheorem für den Cosinus erhalten wir

$$\cos \alpha = \cos(\psi - \varphi) = \cos \varphi \cos \psi + \sin \varphi \sin \psi = \frac{\xi_1}{|x|} \cdot \frac{\eta_1}{|y|} + \frac{\xi_2}{|x|} \cdot \frac{\eta_2}{|y|}.$$

In den Zählern der Brüche rechts entdecken wir das bekannte *Skalarprodukt* im  $\mathbb{R}^2$ :

**Definition 1.25.** Für  $x = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \in \mathbb{R}^2$  und  $y = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} \in \mathbb{R}^2$  definieren wir das Skalarprodukt  $x \cdot y \in \mathbb{R}$  als

$$x \cdot y := \xi_1 \cdot \eta_1 + \xi_2 \cdot \eta_2.$$

Dafür schreiben wir in Zukunft auch  $\langle x, y \rangle$ .

Damit haben wir dann

$$\cos \angle(x, y) = \frac{\langle x, y \rangle}{|x| \cdot |y|},$$

falls keiner der Vektoren  $x, y$  gleich dem Nullvektor ist (ansonsten würden wir durch Null dividieren). Wenn aber einer der Vektoren  $x, y$  gleich  $0$  wäre, dann gäbe es auch keinen Winkel zwischen  $x$  und  $y$ .



**Bemerkung 1.26.** In die Berechnung des Skalarprodukts  $\langle x, y \rangle$  und der Längen  $|x|$ ,  $|y|$  gehen die Koordinaten  $\xi_1, \xi_2, \eta_1, \eta_2$  ein. Diese Koordinaten hängen vom gewählten Koordinatensystem ab, also von der gewählten Basis  $(e_1, e_2)$ . Wenn wir eine andere Basis gewählt hätten, zum Beispiel  $(e'_1, e'_2)$ , dann hätten wir andere Koordinaten  $\xi'_1, \xi'_2, \eta'_1, \eta'_2$  bekommen. Obwohl dies andere Koordinaten sind, bekommen wir für  $\cos \angle(x, y)$  am Ende denselben Wert! Das ist geometrisch auch glaubhaft. Hierbei haben wir stillschweigend vorausgesetzt, daß die Vektoren  $e_1, e_2$  aufeinander senkrecht stehen und jeweils die Länge 1 haben; und daß entsprechendes auch für  $e'_1, e'_2$  gilt. Das bedeutet gerade, daß  $(e_1, e_2)$  eine Orthonormalbasis (ONB) sein soll, und  $(e'_1, e'_2)$  ebenfalls.

**Satz 1.27.** Das Skalarprodukt

$$\langle \cdot, \cdot \rangle : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$$

besitzt die folgenden Eigenschaften für alle Vektoren  $x, y, z \in \mathbb{R}^2$  und Skalare  $\lambda \in \mathbb{R}$ :

$$\begin{aligned} \langle x, y \rangle &= \langle y, x \rangle, \\ \langle x, y + z \rangle &= \langle x, y \rangle + \langle x, z \rangle, \\ \langle \lambda \cdot x, y \rangle &= \lambda \langle x, y \rangle, \\ \langle x, x \rangle &\geq 0 \\ \langle x, x \rangle &= 0 \text{ genau dann, wenn } x = 0. \end{aligned}$$

*Beweis.* Beweis durch Nachrechnen aus Definition 1.25. □

Weil das Skalarprodukt  $\langle x, x \rangle$  eines Vektors  $x$  mit sich selbst nie negativ ist, läßt sich die Wurzel ziehen:

**Definition 1.28.** Für  $x \in \mathbb{R}^2$  erklären wir die euklidische Länge (Betrag, Norm) vermöge

$$|x| := \sqrt{\langle x, x \rangle}.$$

**Satz 1.29 (Ungleichung von Cauchy–Schwarz<sup>17</sup>).** Seien  $x, y \in \mathbb{R}^2$  zwei Vektoren. Dann gilt

$$|\langle x, y \rangle| \leq |x| \cdot |y|.$$

Das Gleichheitszeichen gilt genau dann, wenn die Vektoren  $x$  und  $y$  parallel sind.

*Beweis.* Geometrisch ist die Sache plausibel: wir haben  $-1 \leq \cos \angle(x, y) \leq 1$ , und deshalb  $-|x| \cdot |y| \leq \langle x, y \rangle \leq |x| \cdot |y|$ . Aber für spätere Zwecke ist es nützlich, einen Beweis zu haben, der ohne Appelle an das geometrische Vorstellungsvermögen auskommt und stattdessen nur Satz 1.27 ausnutzt.

Seien nun also  $x, y \in \mathbb{R}^2$  fest gewählt, und sei weiterhin  $\lambda \in \mathbb{R}$  beliebig. Dann haben wir

$$0 \leq \langle x + \lambda y, x + \lambda y \rangle,$$

was sich ausmultiplizieren läßt wie folgt:

$$\begin{aligned} 0 &\leq \langle x + \lambda y, x \rangle + \langle x + \lambda y, \lambda y \rangle \\ &= \langle x, x + \lambda y \rangle + \langle \lambda y, x + \lambda y \rangle = \langle x, x \rangle + \langle x, \lambda y \rangle + \langle \lambda y, x \rangle + \langle \lambda y, \lambda y \rangle \\ &= \langle x, x \rangle + \langle \lambda y, x \rangle + \lambda \langle y, x \rangle + \lambda \langle y, \lambda y \rangle = \langle x, x \rangle + \lambda \langle y, x \rangle + \lambda \langle y, x \rangle + \lambda \langle \lambda y, y \rangle \\ &= \langle x, x \rangle + 2\lambda \langle y, x \rangle + \lambda^2 \langle y, y \rangle = |x|^2 + 2\lambda \langle x, y \rangle + \lambda^2 |y|^2. \end{aligned}$$

Wir können annehmen, daß  $y \neq \vec{0}$  ist, weil ansonsten die Behauptung zu der Banalität  $0 \leq 0$  zusammenschrumpft. Dann ist aber  $|y|^2 \neq 0$ , und wir dürfen dividieren:

$$0 \leq \frac{|x|^2}{|y|^2} + \frac{2 \langle x, y \rangle}{|y|^2} \lambda + \lambda^2.$$

Nun sind  $x$  und  $y$  fest, aber  $\lambda$  variabel. Die rechte Seite ist also eine quadratische Funktion in  $\lambda$ :

$$P(\lambda) = \frac{|x|^2}{|y|^2} + \frac{2 \langle x, y \rangle}{|y|^2} \lambda + \lambda^2,$$

<sup>17</sup> AUGUSTIN LOUIS CAUCHY (1789–1857), HERMANN AMANDUS SCHWARZ (1843–1921)

von der wir wissen, daß sie niemals negative Werte annehmen kann. Also ist die Diskriminante der zugehörigen quadratischen Gleichung  $\leq 0$ :

$$\left(\frac{\langle x, y \rangle}{|y|^2}\right)^2 - \frac{|x|^2}{|y|^2} \leq 0,$$

woraus die gewünschte Ungleichung sofort folgt.

Wenn in der Cauchy–Schwarz–Ungleichung das Gleichheitszeichen eintritt, dann ist die Diskriminante gleich 0, also gibt es ein  $\lambda \in \mathbb{R}$  mit  $\langle x + \lambda y, x + \lambda y \rangle = 0$ , also gilt für dieses  $\lambda$  die Beziehung  $x + \lambda y = \vec{0}$ . Dann sind aber  $x$  und  $y$  parallel. Die Umkehrung dieser Aussage gilt auch, wie man schnell sieht.  $\square$

Man beachte, daß wir nirgends benutzt hatten, daß  $x$  und  $y$  Vektoren aus dem  $\mathbb{R}^2$  sind. Derselbe Beweis läuft auch im  $\mathbb{R}^3$  oder  $\mathbb{R}^n$ .

**Satz 1.30.** Die Norm aus Definition 1.28 erfüllt die folgenden drei Eigenschaften, für alle Vektoren  $x, y \in \mathbb{R}^2$  und alle Skalare  $\lambda \in \mathbb{R}$ :

$$\begin{aligned} |x| &\geq 0 \\ |x| &= 0 \text{ genau dann, wenn } x = 0, \\ |\lambda x| &= |\lambda| \cdot |x|, \\ |x + y| &\leq |x| + |y|. \end{aligned}$$

*Beweis.* Wir benutzen nur die 4 Eigenschaften aus Satz 1.27, aber keinerlei geometrische Interpretation: Wegen  $\sqrt{\cdot} \geq 0$  ist  $|x| \geq 0$ . Wenn  $|x| = 0$  sein sollte, dann muß  $\langle x, x \rangle = 0$  sein, und das geht nur für  $x = 0$ . Weiterhin ist

$$|\lambda x| = \sqrt{\langle \lambda x, \lambda x \rangle} = \sqrt{\lambda \langle x, \lambda x \rangle} = \sqrt{\lambda \langle \lambda x, x \rangle} = \sqrt{\lambda \lambda \langle x, x \rangle} = |\lambda| \sqrt{\langle x, x \rangle} = |\lambda| \cdot |x|.$$

Und schließlich haben wir, wenn wir die Ungleichung von Cauchy–Schwarz zitieren,

$$\begin{aligned} |x + y|^2 &= \langle x + y, x + y \rangle = \langle x + y, x \rangle + \langle x + y, y \rangle = \langle x, x + y \rangle + \langle y, x + y \rangle \\ &= \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle = |x|^2 + 2 \langle x, y \rangle + |y|^2 \\ &\leq |x|^2 + 2|x| \cdot |y| + |y|^2 = (|x| + |y|)^2. \end{aligned}$$

Nun ist  $|x + y|$  nie negativ, also dürfen wir die Wurzel ziehen und erhalten  $|x + y| \leq |x| + |y|$ .  $\square$

Die folgenden Eigenschaften können auf ähnlichem Wege leicht gezeigt werden:

**Satz 1.31.** Seien  $x, y \in \mathbb{R}^2$  beliebige Vektoren. Dann gilt für die Norm aus Definition 1.28, daß

$$\begin{aligned} |x - y|^2 &= |x|^2 + |y|^2 - 2 \langle x, y \rangle, \\ |x + y|^2 + |x - y|^2 &= 2(|x|^2 + |y|^2), \\ \langle x, y \rangle = 0 &\iff |x + y|^2 = |x|^2 + |y|^2. \end{aligned}$$

Geometrische Interpretationen davon zu suchen ist eine lehrreiche Übungsaufgabe.

**Bemerkung 1.32.** Seien  $a \in \mathbb{R}^2$  und  $\beta \in \mathbb{R}$  gegeben. Wir suchen einen Vektor  $x \in \mathbb{R}^2$  mit  $\langle a, x \rangle = \beta$ . Man stellt schnell fest, daß es keine eindeutige Lösung geben kann: Denn ausgeschrieben haben wir  $a_1 x_1 + a_2 x_2 = \beta$  und somit zwei Unbekannte, aber lediglich eine Gleichung. Und tatsächlich können wir zu einer Lösung  $x$  einen beliebigen Vektor addieren, der auf  $a$  senkrecht steht und bekommen so eine weitere Lösung.

*Es gibt keine Division als Umkehroperation des Skalarprodukts.*

### 1.3.2 Drehungen

**Literatur:** Greiner und Müller: *Quantenmechanik. Symmetrien*. Kapitel I.8: Rotationen und ihre Gruppeneigenschaften

Wir stellen uns folgendes Problem:

In einer Ebene sei durch die Einheitsvektoren  $e_1$  und  $e_2$  ein kartesisches Koordinatensystem gegeben. Ein Vektor  $x$  habe bezüglich dieses Koordinatensystems die (gegebenen) Koordinaten  $\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$ , also

$$x = \xi_1 e_1 + \xi_2 e_2.$$

Wir wollen den Vektor  $x$  um den Ursprung drehen, und zwar um den Winkel  $\varphi$  im Gegenuhrzeigersinn. Das Bild von  $x$  sei  $x'$ ; was sind die Koordinaten von  $x'$ ?

Offensichtlich können wir mit gesuchten  $\xi'_j$  schreiben:

$$x' = \xi'_1 e_1 + \xi'_2 e_2.$$

Um diese  $\xi'_j$  zu bestimmen, gehen wir einen Umweg: Wir drehen die Vektoren  $e_1$  und  $e_2$  um den Ursprung, mit dem Winkel  $\varphi$ . Die Bildvektoren seien  $e'_1$  und  $e'_2$ . Dann ist geometrisch glaubbar, daß

$$x' = \xi_1 e'_1 + \xi_2 e'_2,$$

das heißt, der Bildvektor  $x'$  hat in der gedrehten Basis  $(e'_1, e'_2)$  gerade die alten Koordinaten  $\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$ .

Nun ist anhand einer Skizze plausibel, daß

$$e'_1 = \cos \varphi e_1 + \sin \varphi e_2, \quad e'_2 = -\sin \varphi e_1 + \cos \varphi e_2.$$

Insgesamt ergibt sich damit

$$\begin{aligned} x' &= \xi'_1 e_1 + \xi'_2 e_2 \\ &= \xi_1 (\cos \varphi e_1 + \sin \varphi e_2) + \xi_2 (-\sin \varphi e_1 + \cos \varphi e_2) \\ &= (\xi_1 \cos \varphi - \xi_2 \sin \varphi) e_1 + (\xi_1 \sin \varphi + \xi_2 \cos \varphi) e_2, \end{aligned}$$

nach einem Koeffizientenvergleich (den wir später rechtfertigen werden) also

$$\xi'_1 = \cos \varphi \xi_1 - \sin \varphi \xi_2, \quad \xi'_2 = \sin \varphi \xi_1 + \cos \varphi \xi_2.$$

Wir schreiben diese zwei Gleichungen als eine Vektorgleichung:

$$\begin{pmatrix} \xi'_1 \\ \xi'_2 \end{pmatrix} = \xi_1 \begin{pmatrix} \cos \varphi \\ \sin \varphi \end{pmatrix} + \xi_2 \begin{pmatrix} -\sin \varphi \\ \cos \varphi \end{pmatrix}.$$

Um dies noch etwas anders zu schreiben, führen wir eine neue Notation ein: Ein Schema der Form

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad a_{ij} \in \mathbb{R},$$

heißt  $2 \times 2$ -Matrix reeller Zahlen, für die wir eine Operation

*Matrix mal Spaltenvektor ergibt Spaltenvektor*

gemäß

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} := \begin{pmatrix} a_{11}\xi_1 + a_{12}\xi_2 \\ a_{21}\xi_1 + a_{22}\xi_2 \end{pmatrix} = \xi_1 \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} + \xi_2 \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix}$$

definieren. Damit können wir insgesamt schreiben:

$$x' = \begin{pmatrix} \xi'_1 \\ \xi'_2 \end{pmatrix} = \begin{pmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = R(\varphi)x.$$

**Definition 1.33.** *Die Matrix*

$$R(\varphi) = \begin{pmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{pmatrix}, \quad \varphi \in \mathbb{R},$$

heißt Drehmatrix zum Winkel  $\varphi$ .

Was passiert, wenn wir für  $x$  einen der Basisvektoren  $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ ,  $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$  einsetzen? Als Bildvektoren erhalten wir dann

$$\begin{pmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos \varphi \\ \sin \varphi \end{pmatrix} = e'_1$$

beziehungsweise

$$\begin{pmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -\sin \varphi \\ \cos \varphi \end{pmatrix} = e'_2,$$

und das sind gerade die Spalten der Drehmatrix  $R(\varphi)$ .

*Die Spalten der Drehmatrix sind die Koordinaten der Bilder der Basisvektoren.*

Als Nächstes betrachten wir 2 Drehungen: zuerst drehen wir  $x$  um den Winkel  $\varphi$  und erhalten  $x'$ . Dann drehen wir  $x'$  um den Winkel  $\psi$  und erhalten  $x''$ . Wie können wir  $x''$  direkt aus  $x$  bestimmen?

Wir haben also

$$x' = R(\varphi)x, \quad x'' = R(\psi)x', \tag{1.3}$$

und geometrisch ist auch klar, daß

$$x'' = R(\varphi + \psi)x.$$

Die Additionstheoreme der Winkelfunktionen liefern uns

$$R(\varphi + \psi) = \begin{pmatrix} \cos \varphi \cos \psi - \sin \varphi \sin \psi & -\sin \varphi \cos \psi - \cos \varphi \sin \psi \\ \sin \varphi \cos \psi + \cos \varphi \sin \psi & \cos \varphi \cos \psi - \sin \varphi \sin \psi \end{pmatrix}.$$

Andererseits können wir die beiden Gleichungen (1.3) auch ineinander einsetzen:

$$x'' = R(\psi)x' = R(\psi)(R(\varphi)x).$$

Man beachte die Reihenfolge der Multiplikationen auf der rechten Seite: erst wird die Matrix  $R(\varphi)$  mit dem Spaltenvektor  $x$  multipliziert, was wieder einen Spaltenvektor liefert. Dieser wird dann in einem zweiten Schritt von links mit  $R(\psi)$  multipliziert, und wir erhalten  $x''$ .

Im Sinne eines Assoziativgesetzes wollen wir nun die Klammern umsetzen:

$$x'' \stackrel{???}{=} (R(\psi)R(\varphi))x.$$

Hierbei müßten wir aber noch erklären, was das Produkt zweier Matrizen  $R(\psi)$  und  $R(\varphi)$  sein soll.

**Definition 1.34.** *Seien  $A$  und  $B$  zwei Matrizen des Formats  $2 \times 2$ ,*

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}.$$

*Dann definieren wir das Matrixprodukt  $AB$  gemäß*

$$AB := \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{pmatrix}.$$

*Das heißt: vom linken Faktor  $A$  nehmen wir jeweils eine Zeile, und vom rechten Faktor  $B$  jeweils eine Spalte. Das Skalarprodukt dieser 2 Vektoren schreiben wir dort hin, wo Zeile (vom linken Faktor) und Spalte (vom rechten Faktor) einander kreuzen.*

**Bemerkung 1.35.** Die Multiplikation ist nicht kommutativ. Das heißt, meistens ist  $AB \neq BA$ .

**Satz 1.36.** Für die Drehmatrizen gilt allerdings

$$R(\varphi)R(\psi) = R(\varphi + \psi) = R(\psi)R(\varphi), \quad \varphi, \psi \in \mathbb{R}.$$

Von besonderer Bedeutung ist der Drehwinkel 0. Dann haben wir als Drehmatrix

$$R(0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

und man rechnet schnell nach, daß  $x' = R(0)x$  gleich dem Ausgangsvektor  $x$  ist.

**Definition 1.37.** Die  $2 \times 2$  Einheitsmatrix<sup>18</sup> ist

$$I_2 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Der Name erklärt sich daraus, daß Multiplikationen mit dieser Matrix den Ausgangsvektor unverändert lassen; genauso wie Multiplikationen mit der reellen Zahl 1 eine reelle Zahl nicht ändern.

Klar ist auch: wenn wir einen Vektor  $x$  erst um  $\varphi$  nach links drehen, und anschließend um  $-\varphi$  nach links drehen, erhalten wir wieder den Ausgangsvektor  $x$ . Es ist also

$$R(-\varphi)R(\varphi) = I_2.$$

**Definition 1.38.** Sei  $A$  eine Matrix vom Format  $2 \times 2$ . Wenn es eine Matrix  $B$  gibt mit

$$BA = I_2,$$

dann heißt die Matrix  $A$  invertierbar<sup>19</sup>, und die Matrix  $B$  heißt inverse Matrix<sup>20</sup> zu  $A$ . Wir schreiben auch  $B = A^{-1}$ .

**Bemerkung 1.39.** Die Multiplikation ist zwar nicht kommutativ. Aber wenn  $BA = I_2$  ist, dann ist auch  $AB = I_2$ , wie wir im nächsten Abschnitt beweisen werden. Wenn dem nicht so wäre, dann müßte man zwischen linksinversen und rechtsinversen Matrizen unterscheiden, was uns zum Glück erspart bleibt.

Offensichtlich ist dann:

**Satz 1.40.** Jede Drehmatrix  $R(\varphi)$  ist invertierbar, und ihre Inverse lautet

$$R(\varphi)^{-1} = R(-\varphi) = \begin{pmatrix} \cos(-\varphi) & -\sin(-\varphi) \\ \sin(-\varphi) & \cos(-\varphi) \end{pmatrix} = \begin{pmatrix} \cos \varphi & \sin \varphi \\ -\sin \varphi & \cos \varphi \end{pmatrix}.$$

## 1.4 Gruppentheorie

**Literatur:** Greiner und Müller: *Quantenmechanik. Symmetrien*. Kapitel I.7: Definition einer Gruppe

### 1.4.1 Einführung

Wenn wir mit einigem Abstand auf die reellen Zahlen, komplexen Zahlen, Vektoren und Matrizen schauen, dann entdecken wir einige Gemeinsamkeiten, die uns zu den folgenden Definitionen führen:

**Definition 1.41.** Sei  $G$  eine beliebige (nichtleere) Menge. Sei weiterhin  $\circ$  eine Operation mit 2 Argumenten auf  $G$ :<sup>21</sup>

$$\begin{aligned} \circ &: G \times G \rightarrow G, \\ \circ &: (x, y) \mapsto x \circ y. \end{aligned}$$

Wenn diese Operation auf ganz  $G \times G$  definiert ist und die folgenden Bedingungen erfüllt, dann heißt  $(G, \circ)$  eine Halbgruppe<sup>22</sup>.

<sup>18</sup>identity matrix

<sup>19</sup>invertible

<sup>20</sup>inverse matrix

<sup>21</sup> Aus Fußnote 13 wissen wir schon, wie die erste Formelzeile zu lesen ist. Und die zweite Formelzeile, in der der Pfeil  $\mapsto$  jetzt einen „Fuß“ bekommen hat, drückt das Verhalten der Abbildung  $\circ$  ein weiteres Mal aus, diesmal allerdings auf der Ebene von konkreten Elementen  $x$  und  $y$  anstatt auf der Ebene der Menge  $G \times G$ .

<sup>22</sup>semi-group

$\circ$  **ist assoziativ:**  $(x \circ y) \circ z = x \circ (y \circ z)$  für jegliche  $x, y, z \in G$ ,

**neutrales Element:** Es gibt genau ein  $e \in G$ , sodaß für jedes  $x \in G$  gilt:  $e \circ x = x \circ e = x$ .

**Definition 1.42.** Sei  $(G, \circ)$  eine Halbgruppe, die außerdem noch folgende Bedingung erfüllt:

**inverse Elemente:** Zu jedem  $x \in G$  gibt es genau ein  $y \in G$  mit  $x \circ y = y \circ x = e$ .

Dann nennen wir  $(G, \circ)$  eine Gruppe<sup>23</sup>.

**Definition 1.43.** Sei  $(G, \circ)$  eine Halbgruppe bzw. Gruppe. Wenn die Operation  $\circ$  kommutativ ist, also  $x \circ y = y \circ x$  für jegliche  $x, y \in G$  gilt, dann nennen wir die Halbgruppe/Gruppe abelsch<sup>24</sup>, nach NIELS HENRIK ABEL (1802–1829).

Für Halbgruppen haben wir die folgenden Beispiele:

1.  $(\mathbb{N}_0, +)$
2.  $(\mathbb{N}_0, \cdot)$
3.  $(\mathbb{Z}, \cdot)$
4.  $(M_2, \cdot)$ , wobei  $M_2$  die Menge der  $2 \times 2$ -Matrizen bezeichnet.

Die ersten drei Halbgruppen sind abelsch. Für abelsche Gruppen kennen wir unter anderem die folgenden Beispiele:

1.  $(\mathbb{Z}, +)$
2.  $(\mathbb{Q}, +)$
3.  $(\mathbb{C}, +)$
4.  $(\mathbb{R}^2, +)$
5.  $(\mathbb{R} \setminus \{0\}, \cdot)$
6.  $(\mathbb{C} \setminus \{0\}, \cdot)$
7.  $(\{R(\varphi) : \varphi \in \mathbb{R}\}, \cdot)$  — die Gruppe der Drehmatrizen.

Und nichtabelsche Gruppe sind zum Beispiel

1.  $(\{A \in M_2 : A \text{ invertierbar}\}, \cdot)$
2. die Menge aller Abbildungen, die ein Quadrat auf sich abbilden (Ecke auf Ecke), mit der Nacheinanderausführung als Operation.

Wir klassifizieren die Gruppen:

Die Operation  $\circ$  kann zum Beispiel sein die

**Addition:** von Zahlen, Vektoren, Pfeilen, Matrizen, ... ,

**Multiplikation:** von Zahlen oder Matrizen,

**Nacheinanderausführung:** von umkehrbaren Abbildungen einer Menge auf sich.

Das neutrale Element  $e$  ist dann in diesen drei Fällen die

**Null:** Null als Zahl, Null als Nullvektor, Null als Nullmatrix, ...

---

<sup>23</sup>group

<sup>24</sup>abelian

**Eins:** Eins als Zahl oder Einheitsmatrix,

**identische Abbildung:** bildet jedes Objekt auf sich selbst ab.

Das inverse Element  $a^{-1}$  zu einem Element  $a$  der Gruppe ist dann in diesen drei Fällen

**das entgegengesetzte Element:** beispielsweise zu  $5 \in \mathbb{R}$  also  $-5$ ,

**das „reziproke“ Element:** beispielsweise zu  $5 \in \mathbb{R}$  also  $0.2$ ; oder die inverse Matrix,

**die Umkehrabbildung:** diejenige Abbildung, die die Abbildung  $a$  ungeschehen macht.

**Frage:** Eine Gruppe  $G$  habe genau 1492 Elemente. Wieviele neutrale Elemente hat sie? Wieviele inverse Elemente?

Definition 1.42 ist übermäßig streng. Man kann einiges weglassen und hat trotzdem den selben Inhalt:

**Satz 1.44.** *Sei  $G$  eine nichtleere Menge und  $\circ$  eine assoziative Operation mit 2 Argumenten, die auf ganz  $G \times G$  definiert ist, mit Werten in  $G$ . Wir setzen weiterhin voraus:*

*Es gibt mindestens ein linksneutrales Element  $e \in G$ , das heißt  $e \circ x = x$  für jedes  $x \in G$ .*

*Zu jedem  $x \in G$  gibt es mindestens ein linksinverses Element  $\bar{x} \in G$ , das heißt  $\bar{x} \circ x = e$ .*

*Dann ist  $(G, \circ)$  eine Gruppe.*

*Beweis.* Übungsaufgabe. □

Eine Folgerung aus diesem Satz ist: Die Matrix  $I_2$  ist das einzige neutrale Element für die Matrizenmultiplikation. Sei die Matrix  $A$  invertierbar. Dann gibt es also (mindestens) eine Matrix  $B$  mit  $BA = I_2$ . Laut dem vorigen Satz ist dieses  $B$  dann die einzige linksinverse Matrix. Und obendrein ist dieses  $B$  auch noch rechtsinvers, d.h.  $AB = I_2$ . Und dies, obwohl die Multiplikation von Matrizen im Allgemeinen nicht kommutativ ist.

Man kann Gruppen auch anders definieren: anstatt zu fordern, daß zu jedem Element ein Inverses existiert, kann man auch verlangen, daß jede Gleichung lösbar ist:

**Satz 1.45.** *Sei  $(G, \circ)$  eine Gruppe und  $a, b \in G$  beliebig. Dann gibt es genau ein  $x \in G$  bzw.  $y \in G$  mit*

$$a \circ x = b, \quad y \circ a = b.$$

*Beweis.* Wir probieren unser Glück mit  $x = a^{-1} \circ b$ :

$$a \circ x = a \circ (a^{-1} \circ b) = (a \circ a^{-1}) \circ b = e \circ b = b.$$

Also ist dieses  $x$  eine Lösung. Um zu zeigen, daß es keine weitere Lösung gibt, nehmen wir das Gegenteil an. Sei also

$$a \circ x = b, \quad a \circ z = b.$$

Dann haben wir  $a \circ x = a \circ z$ . Wir setzen von links ein  $a^{-1}$  dran und erhalten  $a^{-1} \circ a \circ x = a^{-1} \circ a \circ z$ , woraus  $x = z$  folgt. Die Aussagen über  $y$  lassen sich genauso beweisen. □

**Bemerkung 1.46.** *Es gilt auch die Umkehrung: Sei  $G$  eine Menge mit einer zwei-argumentigen Abbildung  $\circ$  darauf, die aus  $G$  nicht herausführt. Wenn dann noch  $\circ$  assoziativ ist, und wenn jede Gleichung  $a \circ x = b$  und jede Gleichung  $y \circ a = b$  jeweils mindestens eine Lösung  $x$  bzw.  $y$  haben, dann ist  $(G, \circ)$  eine Gruppe.<sup>25</sup>*

**Warnung 1.47.** *Es ist zwar jedes linksinverse Element auch rechtsinvers, aber die Lösungen  $x$  und  $y$  zu  $a \circ x = b$  und  $y \circ a = b$  sind im Allgemeinen verschieden.*

Dieser Satz wirkt sich auf die Verknüpfungstafeln aus wie folgt:

*In jeder Zeile einer Gruppentafel taucht jedes Element genau einmal auf. Analog für jede Spalte.*

<sup>25</sup>Ein Beweis kann im schülerfreundlich geschriebenen Buch HERBERT KÄSTNER, PETER GÖTHNER, *Algebra — Aller Anfang ist leicht*, (Mathematische Schülerbücherei 107, Teubner Leipzig 1989) nachgelesen werden.

Wenn dies einmal nicht der Fall sein sollte, hat man sich entweder verrechnet, oder es ist keine Gruppe.

Jetzt, wo wir über den Begriff der Gruppe verfügen, können wir die Definition des Körpers kürzer formulieren. Am Beispiel von  $(\mathbb{R}, +, \cdot)$  würde die äquivalente Definition lauten:

- $(\mathbb{R}, +)$  ist eine abelsche Gruppe mit neutralem Element 0,
- $(\mathbb{R}, \cdot)$  ist eine abelsche Halbgruppe mit neutralem Element 1,
- $(\mathbb{R} \setminus \{0\}, \cdot)$  ist eine Gruppe,
- Addition und Multiplikation sind verzahnt über das Distributivgesetz.

**Frage:** Können Sie an einem Beispiel zeigen, daß ohne den zweiten • tatsächlich die Äquivalenz nicht gilt ?

**Lemma 1.48.** *Seien  $a$  und  $b$  Elemente einer Gruppe, und  $a^{-1}$ ,  $b^{-1}$  ihre inversen Elemente. Dann wird das inverse Element zu  $a \circ b$  gegeben durch*

$$(a \circ b)^{-1} = b^{-1} \circ a^{-1}.$$

*Beweis.* Beweis durch Einsetzen. □

Man beachte die geänderte Reihenfolge ! Andererseits ist diese neue Reihenfolge auch einleuchtend, wenn man sich z.B.  $a$  vorstellt als Drehung um den Ursprung der Ebene um einen Winkel  $\varphi$  nach links, und  $b$  als Spiegelung an einer gegebenen Geraden in derselben Ebene. Dann wäre natürlich  $a \circ b$  die Kombination beider Bewegungen, und  $(a \circ b)^{-1}$  kann sich jeder selbst überlegen.

Wenn zum Beispiel die Matrizen  $A$  und  $B$  invertierbar sind, dann ist auch ihr Produkt  $AB$  invertierbar, und die inverse Matrix des Produkts ist  $(AB)^{-1} = B^{-1}A^{-1}$ .

Die Formel  $(a \circ b)^{-1} = b^{-1} \circ a^{-1}$  kann man sich auch als kommutatives Diagramm veranschaulichen:

$$\begin{array}{ccc} \boxed{(a, b)} & \xrightarrow{\circ} & \boxed{a \circ b} \\ \text{Tausch} \downarrow \text{und } (\cdot)^{-1} & & \downarrow (\cdot)^{-1} \\ \boxed{(b^{-1}, a^{-1})} & \xrightarrow{\circ} & \boxed{\begin{array}{l} (a \circ b)^{-1} \\ = b^{-1} \circ a^{-1} \end{array}} \end{array}$$

Der Punkt  $\cdot$  ist jeweils als Platz-Freihalter zum Einsetzen zu verstehen.

Das Assoziativgesetz hat auch ein kommutatives Diagramm:

$$\begin{array}{ccc} \boxed{b} & \xrightarrow{a \circ \cdot} & \boxed{a \circ b} \\ \cdot \circ c \downarrow & & \downarrow \cdot \circ c \\ \boxed{b \circ c} & \xrightarrow{a \circ \cdot} & \boxed{a \circ b \circ c} \end{array}$$

Entsprechendes gilt für Distributivgesetze bei Körpern und Vektorräumen.

Wir tragen einige uns bekannte Strukturen zusammen:

**Halbgruppen** besitzen **eine** Operation  $\circ$ , und ein Beispiel ist  $(\mathbb{N}_0, +)$ .

**Gruppen** besitzen eine Operation  $\circ$ , und gemäß Satz 1.45 ist jede Gleichung mit dieser Operation lösbar. In diesem Sinne besitzt eine Gruppe **zwei** Operationen, wobei die eine die Umkehrung der anderen ist. Ein Beispiel ist  $(\mathbb{Z}, +)$  mit der Subtraktion als zweiter Operation.

**Körper** besitzen zwei Operationen  $+$  und  $\cdot$ , die beide umgekehrt werden können. In diesem Sinne besitzen Körper **vier** Operationen. Beispiele sind  $\mathbb{R}$ ,  $\mathbb{Q}$ ,  $\mathbb{C}$ .

Es fällt auf, daß wir keine algebraische Struktur angegeben haben mit genau drei Operationen. Tatsächlich kennt die Mathematik aber solche Strukturen:



**Ringe** besitzen **drei** Operationen. Nämlich eine Addition mit deren Umkehrung (Subtraktion), und eine Multiplikation. Über die Umkehrbarkeit der Multiplikation wird nichts ausgesagt. Die Addition ist kommutativ und assoziativ, die Multiplikation ist assoziativ (aber womöglich nicht kommutativ), und das Distributivgesetz gilt. Ein Beispiel ist  $(\mathbb{Z}, +, \cdot)$ . Bekanntlich kann die Division aus  $\mathbb{Z}$  herausführen.

Ein weiteres Beispiel für Ringe ist  $(\mathbb{R}^{2 \times 2}, +, \cdot)$ , also die Menge der Matrizen vom Format  $2 \times 2$ , für die man die Addition komponentenweise definiert, und  $\cdot$  ist die Multiplikation der Matrizen gemäß Definition 1.34.

{ Es reicht, wenn die folgenden Betrachtungen dieses Abschnitts zum passiven Wissen der Leserschaft gehören. Diese theoretischen Untersuchungen brauchen wir lediglich für das Unterkapitel über die Signalkodierung beim UMTS-Verfahren. }

Ein weiterer Ring wird von den durch 5 teilbaren Zahlen gebildet. Diesen Ring schreiben wir naheliegenderweise als  $(5\mathbb{Z}, +, \cdot)$ . Und ein weiterer Ring besteht aus den Resten von ganzen Zahlen bei Division durch 5. Offenkundig kann eine ganze Zahl bei Division durch 5 nur die Reste 0,1,2,3 oder 4 hinterlassen. Damit wir die Reste besser unterscheiden können von den Zahlen, setzen wir Klammern. Es gibt also die Reste [0], [1], [2], [3], [4], und wir vereinbaren die Schreibweise  $\{[0], [1], [2], [3], [4]\} =: \mathbb{Z}/5\mathbb{Z}$ . Für die Addition erhalten wir dann die Verknüpfungstafel 1.1:

+	[0]	[1]	[2]	[3]	[4]
[0]	[0]	[1]	[2]	[3]	[4]
[1]	[1]	[2]	[3]	[4]	[0]
[2]	[2]	[3]	[4]	[0]	[1]
[3]	[3]	[4]	[0]	[1]	[2]
[4]	[4]	[0]	[1]	[2]	[3]

Tabelle 1.1: Verknüpfungstafel von  $(\mathbb{Z}/5\mathbb{Z}, +)$

Es ist lohnenswert, auch noch die Verknüpfungstafel für die Multiplikation aufzustellen. Dabei wollen wir den Rest [0] ignorieren, da sein Verhalten bei der Multiplikation äußerst vorhersehbar ist. Als Lesehilfe verweisen wir auf  $2 \cdot 4 = 8$ , was den Rest 3 läßt, also  $[2] \cdot [4] = [3]$ .

·	[1]	[2]	[3]	[4]
[1]	[1]	[2]	[3]	[4]
[2]	[2]	[4]	[1]	[3]
[3]	[3]	[1]	[4]	[2]
[4]	[4]	[3]	[2]	[1]

Wir vermerken, daß kein Körper entsteht, wenn die die Zahl 5 durch die Zahl 6 ersetzen (man schaue sich die entsprechende Verknüpfungstafel für die Multiplikation an); und der tiefere Grund ist, daß 6 keine Primzahl ist. Man kann beweisen, daß  $\mathbb{Z}/q\mathbb{Z}$  ein Körper ist für jede Primzahl  $q$ . Für diesen Körper schreibt man auch  $\mathbb{F}_q$ .

## 1.4.2 Ausblick: Gruppen in der Physik

Wir listen einige Gruppen auf, die in der Physik öfters auftreten:

**Symmetriegruppen eines Kristalls:** gegeben sei ein Kristallgitter. Diese gibt es in verschiedenen Ausprägungen (kubisch, tetragonal, rhombisch, hexagonal, trigonal, monoklin, triklin, mit noch zusätzlichen Unterscheidungen, ob weitere Atome raumzentriert oder flächenzentriert oder auf den Mittelpunkten der Basisflächen angebracht sind). Die Elemente der Symmetriegruppe dieses Gesamtgitters bestehen aus allen Bewegungen, die das Gitter auf sich abbilden. Die Verknüpfung ist natürlich die Nacheinanderausführung. Dazu zählen alle Verschiebungen (sofern sie Gitterplätze aufeinander abbilden) und ggf. noch diverse Spiegelungen an Symmetrieebenen oder einige Drehungen. Die das Verhalten dieses Kristalls beschreibenden Differentialgleichungen können sehr kompliziert sein, aber wenn man die Symmetrien „herausfaktoriert“, wird es ein wenig einfacher.

**Verschiebungsgruppen im  $\mathbb{R}^n$ :** das sind alle Verschiebungen, die den  $\mathbb{R}^n$  auf sich abbilden. Die Verknüpfung ist wieder die Nacheinanderausführung (das schreiben wir in Zukunft nicht mehr mit).

**Orthogonale Gruppe im  $\mathbb{R}^3$ :** das sind alle längentreuen Abbildungen im  $\mathbb{R}^3$ , die den Ursprung unverändert lassen, also Drehungen, Spiegelungen, Drehspiegelungen. Diese Gruppe hat unendlich viele Elemente, man nennt sie  $O(3)$ .

**Spezielle orthogonale Gruppe im  $\mathbb{R}^3$ :** das sind alle längentreuen Abbildungen im  $\mathbb{R}^3$ , die den Ursprung und die Orientierung unverändert lassen, also alle Drehungen. Schreibweise:  $SO(3)$ .

**Lineare Gruppe des  $\mathbb{R}^3$ :** das sind alle linearen Abbildungen des  $\mathbb{R}^3$  auf sich, die invertiert werden können. Spate werden auf Spate abgebildet. Schreibweise:  $GL(3)$ .

**Spezielle lineare Gruppe des  $\mathbb{R}^3$ :** das sind alle Elemente der  $GL(3)$ , die die Orientierung des  $\mathbb{R}^3$  erhalten und Spate auf Spate mit gleichem Volumen abbilden.

**Heisenberggruppe:** die obigen Gruppen bestanden aus Abbildungen (und die Verknüpfung war jeweils die Nacheinanderausführung), die Punkte im  $\mathbb{R}^3$  auf Punkte im  $\mathbb{R}^3$  abbildeten. Im Gegensatz dazu besteht die Heisenberggruppe<sup>26</sup> aus Abbildungen, die Elemente eines gewissen *Zustandsraums* auf Elemente dieses Zustandsraums abbilden. Dieser Zustandsraum wiederum besteht aus allen Funktionen von  $\mathbb{R}^3$  nach  $\mathbb{C}$ , deren Betragsquadrat noch dazu integrierbar ist auf dem  $\mathbb{R}^3$ . Sei  $\psi$  ein solches Element des Zustandsraums, also eine Funktion

$$\psi = \psi(x): \mathbb{R}^3 \rightarrow \mathbb{C}.$$

Elemente der Heisenberggruppe wirken auf  $\psi$  wie folgt:

**Verschiebung um  $p \in \mathbb{R}^3$ :**  $\psi = \psi(x)$  wird abgebildet auf  $\psi_p = \psi_p(x) = \psi(x + p)$

**Phasenfaktor:**  $\psi = \psi(x)$  wird abgebildet auf  $\psi_q = \psi_q(x) = e^{iqx}\psi(x)$

**Drehung in  $\mathbb{C}$ :**  $\psi = \psi(x)$  wird abgebildet auf  $\psi_t = \psi_t(x) = e^{it}\psi(x)$ .

Außerdem natürlich noch alle Kompositionen dieser Operationen.

**Lorentzgruppe:**<sup>27</sup> diese ist in der speziellen Relativitätstheorie ganz wichtig. Sie bildet den  $\mathbb{R}^{1+3}$  in sich ab, und zwar so, daß *Eigenzeitabstände* gleich bleiben. Der Raum  $\mathbb{R}^{1+3}$  verwaltet die Variablen für Zeit und Ort. Er heißt *Minkowski*<sup>28</sup>-*Raum* und hat anstelle des gewöhnlichen Skalarprodukts ein Pseudoskalarprodukt:

$$(t, x_1, x_2, x_3) \cdot (s, y_1, y_2, y_3) = -ts + x_1y_1 + x_2y_2 + x_3y_3.$$

Man beachte das Minus vor dem Produkt der Zeiten. Das Pseudoskalarprodukt eines Zeit-Raum-Vektors mit sich selbst kann also negativ werden, und ein Zeit-Raum-Vektor gehört zum Lichtkegel, wenn dieses Pseudoskalarprodukt dieses Vektors mit sich selbst gleich null ist. Die Eigenzeitabstände werden über dieses Pseudoskalarprodukt bestimmt.

Die Lorentz-Gruppe hat unendliche viele Elemente, man schreibt (für die Gruppe) auch  $O(3, 1)$ .

Von diesen Gruppen gibt es noch einen weiteren Zusammenhang zur klassischen Mechanik: typischerweise wird die zeitliche Entwicklung eines physikalischen Systems beschrieben durch ein System von Differentialgleichungen. Diese Differentialgleichungen enthalten Ableitungen bezüglich der Orts- und Zeitvariablen (und eventuell noch bezüglich weiterer Variablen). Nun haben wir:

Wenn dieses Gleichungssystem sich nicht ändert unter einer Verschiebung aus der Verschiebungsgruppe des Ortsraumes  $\mathbb{R}^3$ , dann gilt für dieses physikalische System der Impulserhaltungssatz.

Wenn dieses Gleichungssystem sich nicht ändert unter einer Drehung aus der  $SO(3)$  des Ortsraumes  $\mathbb{R}^3$ , dann gilt für dieses physikalische System der Drehimpulserhaltungssatz.

Wenn dieses Gleichungssystem sich nicht ändert unter einer Verschiebung aus der Verschiebungsgruppe des Zeitraumes  $\mathbb{R}^1$ , dann gilt für dieses physikalische System der Energieerhaltungssatz.

Es versteht sich also von selbst, daß es nützlich ist, nach weiteren Gruppen zu suchen, deren Elemente das Gleichungssystem nicht ändern !

**Literatur:** Greiner und Müller: *Quantenmechanik. Symmetrien*. Kapitel VII: Die  $SU(3)$ -Symmetrie. Kapitel VIII: Quarks und die Gruppe  $SU(3)$ . Kapitel XI: Charm und  $SU(4)$ .

<sup>26</sup> WERNER KARL HEISENBERG (1901–1976), Nobelpreis für Physik 1932

<sup>27</sup> HENDRIK ANTOON LORENTZ (1853–1928), Namensgeber für die Lorentz-Kraft und die Lorentz-Transformation, Nobelpreis für Physik 1902

<sup>28</sup> HERMANN MINKOWSKI (1864–1909)

### 1.4.3 Ausblick: Mobilfunk

Beim Mobiltelefonieren überlagern sich die Funkwellen der einzelnen Teilnehmer — eben weil die Kommunikation nicht entlang von Drähten erfolgt. Da nun auf einer Antenne (am Mobiltelefon oder Funkmast) verschiedene Signale auftreffen, stellt sich die Frage, wie man diese Signale voneinander trennt.

Die naheliegendste Möglichkeit wäre, jedem Teilnehmer eine eigene Frequenz (bzw. ein Frequenzband) zu geben. Allerdings scheitert dies daran, daß es nicht genug Frequenzbänder gibt.

Im GSM-Verfahren wird daher das Zeitschlitzverfahren eingesetzt: jedes Frequenzband mit einer Breite von 200kHz wird zeitlich auf 8 Teilnehmer aufgeteilt. Das heißt, jedes Mobiltelefon auf diesem Band innerhalb einer Funkzelle bekommt einen von 8 Zeitschlitzten. Während eines solchen Zeitschlitzes kommunizieren Mobiltelefon und Basisstation miteinander, während der anderen 7 Zeitschlitzte ist die Sende- und Empfangselektronik des Mobiltelefons abgeschaltet und die anderen Teilnehmer sind an der Reihe.

Aus verschiedenen Gründen ist man bei UMTS vom Zeitschlitzverfahren abgegangen. Stattdessen funken jetzt mehrere Telefone innerhalb einer Funkzelle gleichzeitig auf demselben Frequenzband ihre Signale zur Basisstation. Das Frequenzband ist jetzt breiter, nämlich 5MHz, und die Funksignale eines Telefons sind über die gesamte Breite von 5MHz verschmiert, was eine höhere Sicherheit gegenüber schmalbandigen Störungen erlaubt. Andererseits funken jetzt viel mehr Teilnehmer gleichzeitig auf diesem Band, sodaß diese Signale irgendwie getrennt werden müssen.

Es gibt noch ein weiteres Problem: die Funkwellen, die z.B. ein Mobiltelefon an seine Basisstation sendet, hören ja nicht an der Grenze der Funkzelle plötzlich auf. Stattdessen strahlen sie auch in benachbarte Funkzellen hinein. Deshalb braucht man noch einen Mechanismus, um die Signale von/zu verschiedenen Basisstationen zu unterscheiden.

Aus diesem Grunde werden 2 Verfahren gleichzeitig verwendet:

**orthogonale variable Spreizfaktoren:** diese OVSF haben die Aufgabe, das schmalbandige Signal auf die gesamte Breite des Frequenzbandes zu spreizen; und zwar auf eine solche Art und Weise, daß man die Teilnehmer einer Funkzelle unterscheiden kann.

**scrambling codes:** diese trennen verschiedene Basisstationen (Funkmasten) voneinander.

Der Ablauf vom Mikrophon bis zur Sendeantenne ist in etwa wie folgt:

- Die Sprachsignale werden digitalisiert (also abgetastet) und liegen danach als Folge von Nullen und Einsen vor.
- Anschließend werden überflüssige Informationen weggeschnitten, das Signal wird komprimiert. Die Bandbreite des Signales ist jetzt etwa 12kHz.
- Das komprimierte Signal wird jetzt kanalkodiert, es wird also Redundanz hinzugefügt. Die sich ergebende Bandbreite ist  $(3.84/\text{Spreizfaktor})\text{MHz}$ .
- Das Signal wird gespreizt. Dabei hat jeder Teilnehmer innerhalb einer Basisstation seinen eigenen Spreizkode. Die Bandbreite ist jetzt 3.84MHz.
- Der Scrambling Code wird angewandt, um Signale unterscheiden zu können, die zu verschiedenen Basisstationen gehören. (Bandbreite bleibt gleich).
- Das Signal wird auf eine Trägerfrequenz von 2GHz aufmoduliert (Phasenmodulation). Die Bandbreite ist jetzt 5MHz, da man an den Bandgrenzen nicht exakt abschneiden kann.

Zu den Spreizfaktoren/Spreizkodes:

Ein Spreizfaktor  $F$  ist eine der Zahlen 4, 8, ..., 128, 256. Ein Spreizkode ist eine Folge von  $F$  Zahlen, jede Zahl ist 0 oder 1. Der Spreizfaktor  $F$  hängt von der Anzahl der Teilnehmer in einer Zelle ab: je mehr Teilnehmer, umso höher muß  $F$  sein. Jeder Teilnehmer bekommt von seiner Basisstation einen Spreizkode zugewiesen. Spreizkodes von verschiedenen Teilnehmern sind „orthogonal“ zueinander<sup>29</sup>. Bei einem Spreizfaktor  $F$  gibt es genau  $F$  mögliche Spreizkodes, die paarweise aufeinander „orthogonal“ stehen.

<sup>29</sup> Zwei Spreizkodes  $\vec{a}$  und  $\vec{b}$  aus  $\{0, 1\}^F$  stehen „orthogonal“ aufeinander, wenn das Produkt  $\sum_{j=1}^F (2a_j - 1)(2b_j - 1)$  gleich null ist. Dieses Produkt verhält sich *nicht* wie ein Skalarprodukt, deshalb jedesmal die Gänsefüßchen.

Die Spreizung verläuft so: sei z.B.  $F = 4$ , und der Spreizcode eines Teilnehmers sei z.B.  $(0, 1, 1, 0)$ . Die Sprachsignale dieses Teilnehmers seien z.B.

$$0, 1, 0, 1, 1, \dots$$

In einem ersten Schritt werden diese Signale mit dem Faktor  $F$  verlängert:

$$0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, \dots$$

Da diese viermal so lange Bitfolge in derselben Zeit gesendet wird, hat sich die Bandbreite vervierfacht.

Anschließend wird diese Folge mit dem Spreizcode  $(0, 1, 1, 0)$  geXORt:

$$0, 1, 1, 0, 1, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 0, 1, \dots$$

Wenn man jetzt noch mal mit dem Spreizcode XORen würde, käme man wieder zum Ausgangssignal.

Auf der Antenne der Basisstation kommt nun ein ganzes Gemisch von Signalen an. Wenn die Basisstation das Signal eines bestimmten Teilnehmers herausfiltern soll, dann wird dieses Signalgemisch mit dem Spreizcode geXORt. Weil Spreizcodes verschiedener Teilnehmer aber „orthogonal“ aufeinander stehen, führt dies zu einer Abschwächung der Signale der anderen Teilnehmer, und das gewünschte Signal ist also das stärkste.

Nun kommen die Scrambling Codes ins Spiel:

Da es nur  $F$  paarweise „orthogonale“ Spreizcodes gibt mit der Länge  $F$ , braucht man ein neues Verfahren, um Signale verschiedener Basisstationen voneinander abzugrenzen. Man nimmt dazu weitere Codes (scrambling codes), die „beinahe orthogonal“ zueinander sind und auf folgendem Wege konstruiert werden:

Sei  $\mathbb{F}_2 = \mathbb{Z}/2\mathbb{Z}$  der Körper der Reste bei Division durch 2. Dieser Körper besteht nur aus den Elementen 0 und 1, und es gelten die Regeln  $0 + 0 = 0$ ,  $0 + 1 = 1$ ,  $1 + 1 = 0$  sowie  $0 \cdot 0 = 0 \cdot 1 = 0$ ,  $1 \cdot 1 = 1$ .

Sei  $\mathbb{F}_2[X]$  die Menge aller Polynome in der Variablen  $X$  mit Koeffizienten aus  $\mathbb{F}_2$ . Man kann solche Polynome addieren, subtrahieren und miteinander multiplizieren (aber nicht durcheinander dividieren!) und erhält jedesmal wieder ein Ergebnis in  $\mathbb{F}_2[X]$ . Es gelten die üblichen Gesetze (Kommutativität, Assoziativität, Distributivität, 2 neutrale Elemente). Wir sagen auch, daß  $(\mathbb{F}_2[X], +, \cdot)$  einen *Ring* bildet.

Nun sei  $p \in \mathbb{F}_2[X]$  ein festes Polynom aus diesem Ring mit dem Grade  $N$ . Dieses Polynom ist für die Festlegung der scrambling codes entscheidend. Genauer gesagt, gibt es zwei solche Polynome  $p$ : eines für die Kommunikation vom Mobiltelefon zur Basisstation, und ein anderes für die andere Richtung. Die Grade diese Polynome sind 24 und 18; und diese beiden Polynome sind im UMTS-Standard festgelegt, also europaweit einheitlich (in Amerika und Asien gibt es verschiedene Abweichungen).

Dieses Polynom  $p$  ist ein ganz besonderes Polynom: denn es muß *irreduzibel* sein. Das bedeutet, daß man es nicht in Faktoren aus  $\mathbb{F}_2[X]$  zerlegen kann. Nun nimmt man sich alle diejenigen Polynome aus  $\mathbb{F}_2[X]$  her, die  $p$  als Faktor enthalten. Diese bilden ein sogenanntes *Ideal*, man schreibt für dieses Ideal auch  $(p)$ .

Dann „faktoriert“ man  $\mathbb{F}_2[X]$  nach diesem Ideal und bekommt  $\mathbb{F}_2[X]/(p)$ . Dies ist definiert als die Menge aller Restpolynome, die bei Division mit Rest eines Polynoms aus  $\mathbb{F}_2[X]$  durch  $p$  entstehen können, also diejenigen Polynome aus  $\mathbb{F}_2[X]$  mit Grad  $\leq N - 1$ . Jedes Element  $g$  von  $\mathbb{F}_2[X]/(p)$  hat die Form

$$g(X) \equiv a_{N-1}X^{N-1} + \dots + a_1X + a_0 \pmod{p}, \quad a_j \in \{0, 1\}.$$

Diese Restklassen bilden einen Ring, wie man leicht nachprüft. Weil jedoch das Polynom  $p$  irreduzibel ist, bilden diese Restklassen nicht bloß einen Ring, sondern etwas viel Besseres: nämlich einen Körper. (Zur Erinnerung: es ist  $\mathbb{F}_q$  ein Körper genau dann, wenn  $q$  eine Primzahl ist). Mit anderen Worten, man hat sogar noch die Division. Das heißt, zu jedem Polynom  $g = g(X)$  aus  $\mathbb{F}_2[X]/(p)$  gibt es genau ein Polynom  $h = h(X)$  aus  $\mathbb{F}_2[X]/(p)$  mit

$$g(X) \cdot h(X) \equiv 1 \pmod{p}.$$

Nun definiert man sich eine Folge  $g_0(X), g_1(X), \dots$  von Polynomen aus  $\mathbb{F}_2[X]/(p)$  gemäß der Vorschrift

$$g_k(X) \equiv X^k \pmod{p}, \quad k = 0, 1, 2, \dots$$

Es ist also  $g_k$  nichts anderes als der Rest des Polynomes  $X^k$  bei Division durch  $p(X)$ . Jedes dieser Polynome wird gegeben durch seine Koeffizienten  $a_{N-1}, \dots, a_1, a_0$ , wobei  $a_{N-1} \in \{0, 1\}$  der Koeffizient der höchsten  $X$ -Potenz ist und  $a_0$  das Absolutglied. Wir schauen uns jetzt die Folge der Absolutglieder der  $g_k$  an:

$$(a_0(g_0), a_0(g_1), a_0(g_2), \dots).$$

Dies ist eine unendlich lange Folge von Nullen und Einsen; sie ist periodisch mit der Periodenlänge  $2^N - 1$ . Die Periodenlänge ist deshalb so hoch, weil  $p$  irreduzibel ist.

Und in dieser Folge stecken die scrambling codes: jede Basisstation in Europa bekommt einen Startindex zugewiesen, und die Folgeelemente ab diesem Startindex bilden gerade den scrambling code dieser Basisstation. Verschiedene scrambling codes sind dabei „im wesentlichen orthogonal“.

Das mit dem Spreizcode geXORte Signal wird jetzt mit dem passenden scrambling code ein weiteres Mal geXORt. Anschließend wird das Signal auf die Trägerfrequenz aufmoduliert und gesendet.

An der Empfangsantenne kommt dann ein Wellensalat an, aus dem durch XORen mit dem scrambling code und Spreizcode das gewünschte Signal herausgefiltert wird.

## 1.5 Der Raum und der $\mathbb{R}^3$

### 1.5.1 Allgemeines

Die Begriffe „Ortsvektor“, „Addition von Vektoren“, „Vervielfachung von Vektoren“, „Kartesisches Koordinatensystem“ definieren wir im Falle des dreidimensionalen Raumes analog wie im Falle der Ebene, bloß daß jetzt noch eine weitere Dimension hinzukommt.

Die Menge der Ortsvektoren im dreidimensionalen Raum bildet, gemeinsam mit geometrisch definierter Addition und Multiplikation, einen Vektorraum.

**Definition 1.49.** *Der  $\mathbb{R}^3$  ist definiert als Menge geordneter Tripel reeller Zahlen:*

$$\mathbb{R}^3 := \left\{ \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} : \xi_i \in \mathbb{R} \right\}.$$

Um Vertikalplatz zu sparen, schreiben wir diese Tripel auch als  $(\xi_1, \xi_2, \xi_3)^\top$ . Wir definieren 2 Operationen:

$$\begin{aligned} + : \mathbb{R}^3 \times \mathbb{R}^3 &\rightarrow \mathbb{R}^3, & \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} + \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} &:= \begin{pmatrix} \xi_1 + \eta_1 \\ \xi_2 + \eta_2 \\ \xi_3 + \eta_3 \end{pmatrix}, \\ \cdot : \mathbb{R} \times \mathbb{R}^3 &\rightarrow \mathbb{R}^3, & \lambda \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} &:= \begin{pmatrix} \lambda \xi_1 \\ \lambda \xi_2 \\ \lambda \xi_3 \end{pmatrix} \end{aligned}$$

Weiterhin schreiben wir  $0 = (0, 0, 0)^\top$ ,  $e_1 = (1, 0, 0)^\top$ ,  $e_2 = (0, 1, 0)^\top$ ,  $e_3 = (0, 0, 1)^\top$ .

Wir definieren das Skalarprodukt zweier Vektoren  $x$  und  $y$  als

$$\langle x, y \rangle := \xi_1 \cdot \eta_1 + \xi_2 \cdot \eta_2 + \xi_3 \cdot \eta_3,$$

und die euklidische Norm (den Betrag, die Länge) eines Vektors  $x \in \mathbb{R}^3$  als

$$|x| := \sqrt{\langle x, x \rangle}.$$

Dann folgt schnell, daß dieser  $\mathbb{R}^3$  mit diesen beiden Rechenoperationen einen Vektorraum bildet, und daß das Skalarprodukt die Eigenschaften aus Satz 1.27 besitzt. Dann gilt automatisch auch die Ungleichung von Cauchy–Schwarz, und die Betragsfunktion hat die Eigenschaften, die man von ihr erwartet.

### 1.5.2 Vektorprodukt und Spatprodukt

Der dreidimensionale Raum ist ein Sonderfall. Denn dort (und ausschließlich dort) können wir ein weiteres Produkt von Vektoren definieren:

**Definition 1.50 (Vektorprodukt oder Kreuzprodukt im  $\mathbb{R}^3$ ).** *Seien  $a = (\alpha_1, \alpha_2, \alpha_3)^\top$  und  $b = (\beta_1, \beta_2, \beta_3)^\top$  zwei Vektoren. Dann wird das Vektorprodukt bzw. Kreuzprodukt<sup>30</sup> gegeben durch*

$$a \times b := \begin{pmatrix} \alpha_2 \beta_3 - \beta_2 \alpha_3 \\ \alpha_3 \beta_1 - \beta_3 \alpha_1 \\ \alpha_1 \beta_2 - \beta_1 \alpha_2 \end{pmatrix}.$$

<sup>30</sup> cross product, vector product, outer product

Das sieht zunächst erst mal kompliziert aus; aber später werden wir einfache geometrische Interpretationen und schönere Formeln zur Berechnung finden.

**Satz 1.51.** *Seien  $a, b, c, d \in \mathbb{R}^3$  und  $\lambda \in \mathbb{R}$ . Dann gelten die folgenden Eigenschaften:*

$\times$  **ist bilinear:** *Es ist  $(\lambda a) \times b = \lambda(a \times b)$  und  $(a + b) \times c = (a \times c) + (b \times c)$ ; analog für den zweiten Faktor.*

$\times$  **ist antikommutativ:**  $a \times b = -b \times a$

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

$\times$  **ist nicht assoziativ:** *aber es gilt  $a \times (b \times c) + b \times (c \times a) + c \times (a \times b) = \vec{0}$ .*

**Identität von Lagrange:** <sup>31</sup>  $\langle a \times b, c \times d \rangle = \langle a, c \rangle \langle b, d \rangle - \langle a, d \rangle \langle b, c \rangle$

$a \times b$  **steht senkrecht auf  $a$  und  $b$ :**  $\langle a, a \times b \rangle = \langle b, a \times b \rangle = 0$ .

**Länge des Produktvektors:** *Wenn  $\varphi$  einen der beiden Winkel zwischen  $a$  und  $b$  bezeichnet, ist*

$$|a \times b|^2 = |a|^2 \cdot |b|^2 - \langle a, b \rangle^2 = |a|^2 \cdot |b|^2 \sin^2(\varphi).$$

**Multiplikationstafel:** *Wir haben folgende Tabelle:*

$\times$	$e_1$	$e_2$	$e_3$
$e_1$	0	$e_3$	$-e_2$
$e_2$	$-e_3$	0	$e_1$
$e_3$	$e_2$	$-e_1$	0

(Erläuterung:  $e_1 \times e_2 = e_3$  usw.)

**Linear abhängige Vektoren:** *Es ist  $a \times b = 0$  genau dann, wenn  $a$  und  $b$  linear abhängig sind.*

*Beweis.* Die meisten Eigenschaften lassen sich simples Rechnen zeigen. Zum Beispiel ist

$$\begin{aligned} |a \times b|^2 &= (\alpha_2\beta_3 - \beta_2\alpha_3)^2 + (\alpha_3\beta_1 - \beta_3\alpha_1)^2 + (\alpha_1\beta_2 - \beta_1\alpha_2)^2 \\ &= (\alpha_2\beta_3)^2 + (\beta_2\alpha_3)^2 + (\alpha_3\beta_1)^2 + (\beta_3\alpha_1)^2 + (\alpha_1\beta_2)^2 + (\beta_1\alpha_2)^2 \\ &\quad - 2(\alpha_2\beta_3\beta_2\alpha_3 + \alpha_3\beta_1\beta_3\alpha_1 + \alpha_1\beta_2\beta_1\alpha_2) \\ &= (\alpha_1^2 + \alpha_2^2 + \alpha_3^2)(\beta_1^2 + \beta_2^2 + \beta_3^2) - (\alpha_1\beta_1 + \alpha_2\beta_2 + \alpha_3\beta_3)^2 \\ &= |a|^2 \cdot |b|^2 - \langle a, b \rangle^2 = |a|^2 \cdot |b|^2 (1 - (\cos \angle(a, b))^2). \end{aligned}$$

Daraus folgt sofort die letzte Eigenschaft: Denn wenn  $a \times b = \vec{0}$  ist, dann hat entweder einer der Vektoren  $a, b$  die Länge 0, oder die beiden Vektoren sind parallel oder antiparallel; also sind  $a$  und  $b$  linear abhängig. Und umgekehrt.  $\square$

**Frage:** Wie kann man die Formel für die Produktvektorenlänge aus der Lagrange-Identität herleiten ?

*Das Kreuzprodukt gibt es nur im  $\mathbb{R}^3$ .*

Für die geometrische Interpretation nehmen wir an, daß  $e_1, e_2, e_3$  ein Rechtssystem sind, also im Raum liegen wie Daumen, Zeigefinger, Mittelfinger einer gesunden rechten Hand. Dann gilt:

**Satz 1.52.** *Seien  $a, b \in \mathbb{R}^3$ , und  $c = a \times b$ . Dann gilt:*

1.  $c$  steht senkrecht auf  $a$  und auf  $b$ .
2. Die Länge  $|c|$  ist gleich der Fläche des Parallelogramms, das von  $a$  und  $b$  aufgespannt wird.
3.  $a, b$  und  $c$  bilden in dieser Reihenfolge ein Rechtssystem.

Die ersten beiden Aussagen sind schon bewiesen, den Beweis der letzten Aussage müssen wir aufschieben (siehe Satz 1.62).

<sup>31</sup>JOSEPH LOUIS LAGRANGE, 1736–1813

**Bemerkung 1.53.** Gegeben seien zwei Vektoren  $a$  und  $b$ . Gesucht sei ein Vektor  $x$  mit  $a \times x = b$ . Aus dem vorigen Satz ergibt sich, daß  $a$  und  $b$  aufeinander senkrecht stehen müssen. Ansonsten kann es eine Lösung  $x$  gar nicht geben.

Aber selbst wenn  $a$  und  $b$  aufeinander senkrecht stehen sollten, ist die Lösung  $x$  immer noch nicht eindeutig bestimmt (auch wenn wir jetzt drei Gleichungen für drei Unbekannte haben).

Seien zum Beispiel  $a = (1, 0, 0)^\top$  und  $b = (\beta_1, \beta_2, \beta_3)^\top$  gegeben, und  $x = (\xi_1, \xi_2, \xi_3)^\top$  gesucht. Dann ist

$$\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = b = a \times x = \begin{pmatrix} 0 \\ -\xi_3 \\ \xi_2 \end{pmatrix}.$$

Wir bekommen  $\beta_1 = 0$  (was nichts anderes als  $a \perp b$  bedeutet) sowie  $\xi_2 = \beta_3$  und  $\xi_3 = -\beta_2$ . Und offensichtlich können wir  $\xi_1$  nicht ermitteln. Es gibt mehrere Lösungen: wenn  $x$  eine Lösung ist, dann ist auch  $x + \lambda a$  eine Lösung, für jedes  $\lambda \in \mathbb{R}$  (natürlich nur wenn  $a \perp b$ ).

Es gibt keine Division als Umkehroperation des Kreuzproduktes.

Als Kombination von Skalarprodukt und Vektorprodukt bekommen wir (nur im  $\mathbb{R}^3$ !) das Spatprodukt:

**Definition 1.54 (Spatprodukt bzw. Determinante).** Seien  $x, y, z \in \mathbb{R}^3$ . Dann heißt der Wert  $\langle x, y \times z \rangle$  Spatprodukt<sup>32</sup> oder Determinante<sup>33</sup> der Vektoren  $x, y, z$ . Eine andere Schreibweise ist  $\det(x, y, z)$ .

Der Namen ergibt sich daraus, daß drei Vektoren im  $\mathbb{R}^3$  einen sogenannten Spat aufspannen; und das Spatprodukt ist gerade das (orientierte) Volumen des Spats. Wenn die Vektoren  $x, y, z$  ein Rechtssystem bilden, dann ist das Spatprodukt positiv, ansonsten negativ.

Seien nun  $x = (\xi_1, \xi_2, \xi_3)^\top$ ,  $y = (\eta_1, \eta_2, \eta_3)^\top$  und  $z = (\zeta_1, \zeta_2, \zeta_3)^\top$ . Dann erhalten wir nach Einsetzen und Ausrechnen die Formel

$$\det(x, y, z) = \xi_1 \eta_2 \zeta_3 + \eta_1 \zeta_2 \xi_3 + \zeta_1 \xi_2 \eta_3 - \zeta_1 \eta_2 \xi_3 - \xi_1 \zeta_2 \eta_3 - \eta_1 \xi_2 \zeta_3,$$

die man sich wie folgt merken kann: Man schreibt die Vektoren  $x, y$  und  $z$  spaltenweise nebeneinander, und anschließend die Vektoren  $x$  und  $y$  noch einmal rechts daneben:

$$\begin{array}{ccccc} \xi_1 & \eta_1 & \zeta_1 & \xi_1 & \eta_1 \\ \xi_2 & \eta_2 & \zeta_2 & \xi_2 & \eta_2 \\ \xi_3 & \eta_3 & \zeta_3 & \xi_3 & \eta_3 \end{array}$$

Anschließend bildet man drei diagonale Produkte von links oben nach rechts unten; diese Produkte werden positiv gezählt. Und entsprechend bildet man drei diagonale Produkte von rechts oben nach links unten, die negativ gezählt werden.

**Satz 1.55.** Das Spatprodukt hat die folgenden Eigenschaften:

- Es ist linear in jedem der drei Argumente (trilinear), das heißt zum Beispiel für das erste Argument

$$\det(x + y, z, w) = \det(x, z, w) + \det(y, z, w), \quad \det(\lambda x, y, z) = \lambda \det(x, y, z).$$

- Wenn zwei der Vektoren  $x, y, z$  gleich sind, dann verschwindet das Spatprodukt:  $\det(x, x, z) = 0$  usw.
- Für die kanonischen Basisvektoren gilt  $\det(e_1, e_2, e_3) = 1$ .

*Beweis.* Wir wissen von früher, daß sowohl das Skalarprodukt als auch das Vektorprodukt bilinear sind. Daraus ergibt sich sofort, daß das Spatprodukt linear in jedem seiner Faktoren ist (Distributivgesetze).

Die zweite Aussage folgt daraus, daß einerseits das Vektorprodukt senkrecht steht auf jedem seiner beiden Faktoren, andererseits das Vektorprodukt zweier gleicher Vektoren gleich dem Nullvektor ist.

Und die dritte Aussage ergibt sich aus elementarem Rechnen. □

<sup>32</sup>parallelepipedal product, triple product

<sup>33</sup>determinant

**Satz 1.56.** Für das Spatprodukt gelten die folgenden Rechenregeln:

- Wenn man ein Vielfaches eines Vektors zu einem anderen Vektor addiert, bleibt das Spatprodukt gleich:  $\det(x + \lambda y, y, z) = \det(x, y, z)$ .
- Wenn man zwei Vektoren im Spatprodukt tauscht, ändert sich das Vorzeichen:  $\det(y, x, z) = -\det(x, y, z)$ .
- Die drei Vektoren im Spatprodukt kann man zyklisch tauschen:  $\det(x, y, z) = \det(y, z, x) = \det(z, x, y)$ .

*Beweis.* Mit den Eigenschaften des vorigen Satzes haben wir

$$\det(x + \lambda y, y, z) = \det(x, y, z) + \det(\lambda y, y, z) = \det(x, y, z) + \lambda \det(y, y, z) = \det(x, y, z) + 0,$$

womit wir die zweite Aussage beweisen können:

$$\begin{aligned} \det(y, x, z) &= \det(y, x + y, z) = \det(y - (x + y), x + y, z) \\ &= \det(-x, x + y, z) = \det(-x, y, z) = -\det(x, y, z). \end{aligned}$$

Und wenn wir dies zweimal anwenden, erhalten wir den dritten Teil:

$$\det(x, y, z) = -\det(x, z, y) = +\det(y, z, x).$$

□

**Definition 1.57.** Seien  $x_1, x_2, \dots, x_n \in \mathbb{R}^3$  beliebige Vektoren.

Wenn es reelle Zahlen  $\lambda_1, \dots, \lambda_n$  gibt, von denen wenigstens eine nicht 0 ist, sodaß  $\lambda_1 x_1 + \dots + \lambda_n x_n = \vec{0}$  ist, dann heißen die Vektoren  $x_1, \dots, x_n$  linear abhängig<sup>34</sup>.

Ansonsten (wenn es also nur eine einzige Möglichkeit gibt, den Nullvektor mit den Vektoren  $x_j$  darzustellen: nämlich alle  $\lambda_j$  gleich 0 zu wählen) heißen die Vektoren  $x_1, \dots, x_n$  linear unabhängig<sup>35</sup>.

Anders formuliert: Seien die Vektoren  $x_1, \dots, x_n$  linear unabhängig. Wenn dann eine Gleichung der Form  $\lambda_1 x_1 + \dots + \lambda_n x_n = \vec{0}$  wahr ist, dann müssen sämtliche  $\lambda_j = 0$  sein.

Wenn die Vektoren  $x_1, \dots, x_n$  hingegen linear abhängig sind, dann kann man reelle Zahlen  $\lambda_1, \dots, \lambda_n$  finden mit  $\lambda_1 x_1 + \dots + \lambda_n x_n = \vec{0}$ , sodaß wenigstens ein  $\lambda_j \neq 0$  ist. Zum Beispiel sei  $\lambda_1 \neq 0$ . Dann dürfen wir durch  $\lambda_1$  dividieren, und bekommen

$$x_1 = -\left(\frac{\lambda_2}{\lambda_1} x_2 + \dots + \frac{\lambda_n}{\lambda_1} x_n\right) = \alpha_2 x_2 + \dots + \alpha_n x_n, \quad \alpha_j := -\frac{\lambda_j}{\lambda_1}.$$

Wir können also (weil  $\lambda_1 \neq 0$  ist)  $x_1$  als Linearkombination der anderen Vektoren schreiben. (Aber es kann sein, daß wir z.B.  $x_2$  nicht mittels  $x_1, x_3, x_4, \dots, x_n$  darstellen können.)

**Frage:** Finden Sie ein Beispiel für Vektoren  $x_1, x_2, x_3$  in der Ebene, sodaß man  $x_1$  als Linearkombination von  $x_2$  und  $x_3$  darstellen kann, aber  $x_2$  nicht als Linearkombination von  $x_1$  und  $x_3$ .

*Wenn eine Familie<sup>36</sup> von Vektoren linear abhängig ist, dann gibt es einen Vektor aus dieser Familie, den man als Linearkombination der anderen ausdrücken kann.*

Sobald eine Familie von Vektoren den Nullvektor enthält, ist sie linear abhängig.

**Satz 1.58.** Seien  $x, y, z \in \mathbb{R}^3$  beliebige Vektoren.

- Folgende drei Aussagen sind äquivalent:<sup>37</sup>

1.  $x \times y = 0$ ,

---

<sup>34</sup>linearly dependent

<sup>35</sup>linearly independent

<sup>36</sup> Wir reden von einer *Familie* von Vektoren, weil in einer Familie ein Vektor auch mehrfach vorkommen darf. Im Gegensatz dazu darf eine *Menge* kein Element doppelt enthalten.

<sup>37</sup> Das bedeutet nur, daß jede der drei Aussagen aus jeder der beiden anderen geschlußfolgert werden kann. Die drei Aussagen sind entweder alle gleichzeitig wahr, oder sie sind alle gleichzeitig falsch.



2.  $x$  und  $y$  sind linear abhängig,
3.  $x$  und  $y$  sind parallel (oder antiparallel).

• Folgende drei Aussagen sind äquivalent:

1.  $\det(x, y, z) \neq 0$ ,
2.  $x, y$  und  $z$  sind linear unabhängig,
3. jedes  $w \in \mathbb{R}^3$  kann auf eindeutige Weise geschrieben werden als  $w = \alpha x + \beta y + \gamma z$ .

• Je vier Vektoren im  $\mathbb{R}^3$  sind linear abhängig.

*Beweisskizze.* Der erste Teil wurde in Satz 1.51 bewiesen.

Für den zweiten Teil beschränken wir uns auf einige anschauliche geometrische Überlegungen: Wenn die Vektoren  $x, y, z$  linear unabhängig sind, dann liegen sie nicht in einer gemeinsamen Ebene. Also spannen sie einen Spat auf, der ein Volumen hat, das nicht 0 ist. Und umgekehrt, was die Äquivalenz  $1 \iff 2$  zeigt. Wenn die Vektoren  $x, y, z$  linear unabhängig sind und  $w$  ein beliebiger Vektor, dann existiert genau ein Spat, der die Strecke zwischen dem Ursprung 0 und  $w$  als Raumdiagonale hat und dessen Kanten parallel zu den Vektoren  $x, y, z$  sind. Die Koordinaten  $\alpha, \beta, \gamma$  kann man dann von den Kantenlängen ablesen.

Für den Beweis des dritten Teils fehlen uns im Moment die Mittel, sodaß wir ihn auf später verschieben.  $\square$

**Definition 1.59.** Wenn die Vektoren  $x_1, x_2, \dots, x_n \in \mathbb{R}^3$  paarweise orthogonal aufeinander stehen,

$$\langle x_i, x_j \rangle = 0, \quad i \neq j,$$

dann bilden diese Vektoren ein Orthogonalsystem<sup>38</sup>.

Wenn diese Vektoren außerdem noch jeweils die Länge 1 haben, also

$$\langle x_i, x_j \rangle = \delta_{ij} = \begin{cases} 0 & : i \neq j, \\ 1 & : i = j, \end{cases}$$

dann reden wir von einem Orthonormalsystem<sup>39</sup>. Der Ausdruck  $\delta_{ij}$  heißt Kroneckersymbol (LEOPOLD KRONECKER, 1823–1891).

**Satz 1.60.** • Wenn die Vektoren  $a_1, \dots, a_n \in \mathbb{R}^3$  ein OGS bilden und jeweils nicht  $\vec{0}$  sind, dann sind sie linear unabhängig.

• Wenn die Vektoren  $a_1, a_2, a_3 \in \mathbb{R}^3$  ein ONS bilden, dann ist ihre Determinante  $\det(a_1, a_2, a_3)$  entweder +1 oder -1.

*Beweis.* Im Beweis des ersten Teils gehen wir indirekt vor: wir setzen voraus, daß die Vektoren  $a_1, \dots, a_n$  jeweils nicht  $\vec{0}$  sind, ein Orthogonalsystem bilden, aber linear *abhängig* sind. Dann können wir zumindest einen Vektor als Linearkombination der anderen darstellen. Ohne Beschränkung der Allgemeinheit (o.B.d.A.) sei dieser Vektor  $a_1$  (ansonsten numerieren wir die Vektoren um). Dann ist also

$$a_1 = \lambda_2 a_2 + \dots + \lambda_n a_n.$$

Wir bilden das Skalarprodukt mit  $a_1$ :

$$\langle a_1, a_1 \rangle = \lambda_2 \langle a_2, a_1 \rangle + \dots + \lambda_n \langle a_n, a_1 \rangle.$$

Die linke Seite ist ungleich 0, denn  $a_1 \neq \vec{0}$ . Aber die rechte Seite verschwindet, denn es ist  $\langle a_j, a_1 \rangle = 0$  für  $j \geq 2$ . Das ist ein Widerspruch. Also müssen die Vektoren linear unabhängig sein.

Zum zweiten Teil: wir argumentieren geometrisch. Es ist  $\det(a_1, a_2, a_3) = \langle a_1, a_2 \times a_3 \rangle$ , und die Vektoren  $a_2, a_3$  spannen ein Quadrat der Seitenlänge 1 auf, also ist  $a_2 \times a_3$  ein Vektor mit der Länge 1, der auf  $a_2$  und  $a_3$  senkrecht steht. Damit ist  $a_2 \times a_3 = +a_1$  oder  $-a_1$ , und somit folgt  $\det(a_1, a_2, a_3) = \langle a_1, \pm a_1 \rangle = \pm |a_1|^2 = \pm 1$ .  $\square$

<sup>38</sup>orthogonal system

<sup>39</sup>orthonormal system

**Definition 1.61.** Ein Orthonormalsystem mit Determinante +1 bzw. -1 heißt positiv orientiert<sup>40</sup> bzw. negativ orientiert<sup>41</sup>.

Ein geordnetes Tripel  $(x_1, x_2, x_3)$  von Vektoren heißt Rechtssystem bzw. Linkssystem, wenn seine Determinante  $\det(x_1, x_2, x_3)$  positiv bzw. negativ ist.

**Satz 1.62.** Die Tripel  $(e_1, e_2, e_3)$  und  $(a, b, a \times b)$  sind Rechtssysteme (wenn  $a \times b \neq \vec{0}$ ).

*Beweis.* Man rechnet schnell nach, daß  $\det(e_1, e_2, e_3) = 1$  ist. Weiterhin ist

$$\det(a, b, a \times b) = \det(b, a \times b, a) = \det(a \times b, a, b) = \langle a \times b, a \times b \rangle = |a \times b|^2 > 0.$$

□

Damit ist der Beweis von Satz 1.52 vervollständigt.

**Bemerkung 1.63.** Ein anderer Zugang zum Kreuzprodukt führt über den LEVI-CIVITA<sup>42</sup> -Tensor  $\varepsilon$ . Dieser ist ein Tensor dritter Stufe (also ein würfelförmiges Zahlenschema, analog zu einer Matrix als einem quadratischen Zahlenschema und einem Vektor als einem eindimensionalen Zahlenschema). Er ist definiert als

$$\varepsilon_{jkl} := \begin{cases} 1 & : \text{wenn } (j, k, l) \text{ eine gerade Permutation von } (1, 2, 3) \text{ ist,} \\ -1 & : \text{wenn } (j, k, l) \text{ eine ungerade Permutation von } (1, 2, 3) \text{ ist,} \\ 0 & : \text{sonst.} \end{cases}$$

Dann ist  $\vec{a} \times \vec{b} = -(\varepsilon \cdot \vec{a}) \cdot \vec{b}$ . Hierbei ist das Produkt  $\varepsilon \cdot \vec{a}$  (ergibt eine Matrix) definiert analog zum Produkt von Matrix mal Vektor über das Summieren benachbarter identischer Indizes:

$$(\varepsilon \cdot \vec{a})_{jk} := \sum_{l=1}^3 \varepsilon_{jkl} a_l.$$

Wir haben auch  $(a \times b)_j = \sum_{k,l=1}^3 \varepsilon_{jkl} a_k b_l$ .

### 1.5.3 Drehungen im $\mathbb{R}^3$

Genauso wie im Falle des  $\mathbb{R}^2$  stellen wir uns folgende Frage:

Gegeben ist ein Vektor  $x \in \mathbb{R}^3$  und eine Drehachse durch den Ursprung. Wir wollen den Vektor  $x$  um diese Achse drehen mit einem Drehwinkel  $\varphi$ . Wie können wir den Bildvektor  $x'$  bestimmen?

**Der einfache Fall: die Drehachse ist eine Koordinatenachse**

Die Drehachse sei also die Achse entlang des Vektors  $e_1$ , und der Drehwinkel sei  $\varphi$ . Es bietet sich an, den Originalvektor  $x$  in zwei Teile zu zerlegen:

$$x = \xi_1 e_1 + \xi_2 e_2 + \xi_3 e_3 = x_{\parallel} + x_{\perp},$$

wobei  $x_{\parallel} = \xi_1 e_1$  parallel zur Drehachse ist, und  $x_{\perp} = \xi_2 e_2 + \xi_3 e_3$  in einer Ebene lebt, die senkrecht auf der Drehachse steht.

Man überlegt sich leicht, daß der Bildvektor  $x'$  sich zerlegt als

$$x' = x'_{\parallel} + x'_{\perp},$$

wobei  $x'_{\parallel} = x_{\parallel}$  gilt, denn dieser Anteil liegt genau auf der Drehachse. Und der senkrechte Anteil  $x'_{\perp}$  kann wie bei einer Drehung in der Ebene ausgerechnet werden:

$$\begin{pmatrix} \xi'_2 \\ \xi'_3 \end{pmatrix} = \begin{pmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{pmatrix} \begin{pmatrix} \xi_2 \\ \xi_3 \end{pmatrix}.$$

---

<sup>40</sup>positively oriented

<sup>41</sup>negatively oriented

<sup>42</sup>TULLIO LEVI-CIVITA (1873–1941)

Um die beiden Teilvektoren  $x_{\parallel}$  und  $x_{\perp}$  einheitlich zu behandeln, führen wir  $3 \times 3$ -Matrizen ein:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \quad a_{ij} \in \mathbb{R}.$$

Außerdem definieren wir eine Operation

*Matrix mal Spaltenvektor ergibt Spaltenvektor*

gemäß

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} := \begin{pmatrix} a_{11}\xi_1 + a_{12}\xi_2 + a_{13}\xi_3 \\ a_{21}\xi_1 + a_{22}\xi_2 + a_{23}\xi_3 \\ a_{31}\xi_1 + a_{32}\xi_2 + a_{33}\xi_3 \end{pmatrix} = \xi_1 \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix} + \xi_2 \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix} + \xi_3 \begin{pmatrix} a_{13} \\ a_{23} \\ a_{33} \end{pmatrix},$$

mit der wir die Abbildungsgleichung schreiben können als

$$\begin{pmatrix} \xi'_1 \\ \xi'_2 \\ \xi'_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \varphi & -\sin \varphi \\ 0 & \sin \varphi & \cos \varphi \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix}.$$

Damit wissen wir jetzt, wie wir einen Vektor um die erste Koordinatenachse drehen können.

### Der schwierige Fall: die Drehachse ist verschieden von allen Koordinatenachsen

Also ist die Drehachse entlang des Vektors  $r$  eine beliebige Gerade durch den Ursprung. Die Strategie ist:

1. wir wählen ein neues Koordinatensystem (also neue Basisvektoren  $r_1, r_2, r_3$ ), sodaß der Drehachsenvektor  $r$  gleich einem Koordinatensystemsachsenvektor wird,
2. wir rechnen die Koordinaten von  $x$  auf das neue Koordinatensystem um,
3. wir führen die Drehung im neuen Koordinatensystem aus,
4. wir rechnen die Koordinaten des Bildvektors  $x'$  wieder auf das alte Koordinatensystem zurück.

Es ist also ein Vektor  $x$  gegeben und eine Drehachse, die durch einen Vektor  $r_1 \in \mathbb{R}^3$  beschrieben wird. Wir wollen  $x$  um die Achse durch  $r_1$  mit einem Winkel  $\varphi$  drehen.

#### Zu Schritt 1:

Wir dürfen annehmen, daß der Vektor  $r_1$  die Länge 1 hat. Wir suchen uns zwei weitere Vektoren  $r_2$  und  $r_3$ , die in der zu  $r_1$  senkrechten Ebene liegen und gemeinsam mit  $r_1$  ein ONS bilden (später werden wir ein Verfahren angeben, solche Vektoren  $r_2, r_3$  zu bestimmen (Orthogonalisierungsverfahren von Gram-Schmidt)). Dabei ist zu beachten, daß  $r_1, r_2$  und  $r_3$  in dieser Reihenfolge ein Rechts-System bilden; anderenfalls muß das Vorzeichen des Drehwinkels geändert werden. Diese 3 Spaltenvektoren stellen wir nebeneinander und bekommen eine Matrix  $A$ :

$$A = \begin{pmatrix} | & | & | \\ r_1 & r_2 & r_3 \\ | & | & | \end{pmatrix}.$$

#### Zu Schritt 2:

Wir können diese 3 Vektoren als Basis des  $\mathbb{R}^3$  ansehen und haben dann

$$x = \varrho_1 r_1 + \varrho_2 r_2 + \varrho_3 r_3,$$

wobei die  $\varrho_j$  zunächst unbekannt sind. Außerdem verfügen wir über die Darstellung

$$x = \xi_1 e_1 + \xi_2 e_2 + \xi_3 e_3,$$

wobei die Zahlen  $\xi_j$  gegeben sind, und die  $e_j$  sind die kanonischen Basisvektoren:  $e_1 = (1, 0, 0)^\top$  usw.

Die Zahlen  $\varrho_j$  können wir folgendermaßen berechnen:

$$\langle x, r_j \rangle = \langle \varrho_1 r_1 + \varrho_2 r_2 + \varrho_3 r_3, r_j \rangle = \langle \varrho_j r_j, r_j \rangle = \varrho_j |r_j|^2 = \varrho_j, \quad j = 1, 2, 3,$$

denn die Vektoren  $r_1, r_2$  und  $r_3$  bilden ein ONS, und beim Bilden des Skalarproduktes fallen immer alle Summanden bis auf einen heraus.

Außerdem benötigen wir noch den Begriff der transponierten Matrix:

**Definition 1.64.** Sei  $A$  eine  $3 \times 3$ -Matrix, und  $A^\top$  diejenige Matrix, die durch Spiegelung der Matrixeinträge an der Hauptdiagonalen entsteht. Dann heißt  $A^\top$  transponierte Matrix<sup>43</sup> zu  $A$ . Also:

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

**Satz 1.65.** Seien  $A$  und  $B$   $3 \times 3$ -Matrizen. Dann ist

$$(AB)^\top = B^\top A^\top.$$

*Beweis.* Übungsaufgabe. □

Man beachte die geänderte Reihenfolge der Faktoren !

Etwas geschickter hingeschrieben, lauten die Gleichungen  $\langle x, r_j \rangle = \varrho_j$ :

$$\begin{pmatrix} \varrho_1 \\ \varrho_2 \\ \varrho_3 \end{pmatrix} = \begin{pmatrix} \text{---} & r_1^\top & \text{---} \\ \text{---} & r_2^\top & \text{---} \\ \text{---} & r_3^\top & \text{---} \end{pmatrix} \begin{pmatrix} | \\ x \\ | \end{pmatrix},$$

also

$$\begin{pmatrix} \varrho_1 \\ \varrho_2 \\ \varrho_3 \end{pmatrix} = A^\top x. \tag{1.4}$$

**Zu Schritt 3:**

Im Koordinatensystem der Vektoren  $(r_1, r_2, r_3)$  läßt sich die Drehung mühelos durchführen:

$$x' = \varrho'_1 r_1 + \varrho'_2 r_2 + \varrho'_3 r_3$$

mit unbekanntenen Koordinaten  $\varrho'_j$ , die sich bestimmen gemäß

$$\begin{pmatrix} \varrho'_1 \\ \varrho'_2 \\ \varrho'_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \varphi & -\sin \varphi \\ 0 & \sin \varphi & \cos \varphi \end{pmatrix} \begin{pmatrix} \varrho_1 \\ \varrho_2 \\ \varrho_3 \end{pmatrix}. \tag{1.5}$$

**Zu Schritt 4:**

Aus dem vorigen Schritt kennen wir bereits die Zahlenspalte  $(\varrho'_1, \varrho'_2, \varrho'_3)^\top$ , und jetzt suchen wir die Zahlenspalte  $(\xi'_1, \xi'_2, \xi'_3)^\top$  zum Bildvektor  $x'$ . Analog zu (1.4) haben wir auch jetzt die Beziehung

$$\begin{pmatrix} \varrho'_1 \\ \varrho'_2 \\ \varrho'_3 \end{pmatrix} = A^\top x',$$

und wenn wir von links die inverse Matrix  $(A^\top)^{-1}$  drannmultiplizieren, folgt

$$(A^\top)^{-1} \begin{pmatrix} \varrho'_1 \\ \varrho'_2 \\ \varrho'_3 \end{pmatrix} = (A^\top)^{-1} A^\top x' = I_3 x' = x',$$

und somit sind wir fertig, denn insgesamt haben wir jetzt (aus (1.4) und (1.5))

$$\begin{aligned} \begin{pmatrix} \xi'_1 \\ \xi'_2 \\ \xi'_3 \end{pmatrix} &= (A^\top)^{-1} \begin{pmatrix} \varrho'_1 \\ \varrho'_2 \\ \varrho'_3 \end{pmatrix} = (A^\top)^{-1} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \varphi & -\sin \varphi \\ 0 & \sin \varphi & \cos \varphi \end{pmatrix} \begin{pmatrix} \varrho_1 \\ \varrho_2 \\ \varrho_3 \end{pmatrix} \\ &= (A^\top)^{-1} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \varphi & -\sin \varphi \\ 0 & \sin \varphi & \cos \varphi \end{pmatrix} A^\top \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix}. \end{aligned}$$

Leider sind wir noch nicht ganz fertig, denn wir haben zwei offene Fragen:

---

<sup>43</sup>transposed matrix

- gibt es  $(A^\top)^{-1}$  überhaupt ?
- wenn ja, wie finden wir  $(A^\top)^{-1}$  ? Was wäre der Rechenweg ?

Zur Antwort brauchen wir ein neues Produkt, nämlich

*Matrix mal Matrix ergibt Matrix.*

Wir beobachten, daß man sich eine Matrix als nebeneinandergestellte Spaltenvektoren vorstellen kann:

$$A = \left( \begin{array}{c|c|c} | & | & | \\ a_1 & a_2 & a_3 \\ | & | & | \end{array} \right), \quad a_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ a_{3j} \end{pmatrix}, \quad j = 1, 2, 3.$$

Wenn nun eine  $3 \times 3$ -Matrix  $B$  entsteht aus 3 Spaltenvektoren  $b_1$ ,  $b_2$  und  $b_3$ , dann definieren wir

$$AB := A \left( \begin{array}{c|c|c} | & | & | \\ b_1 & b_2 & b_3 \\ | & | & | \end{array} \right) := \left( \begin{array}{c|c|c} | & | & | \\ Ab_1 & Ab_2 & Ab_3 \\ | & | & | \end{array} \right).$$

Oder anders formuliert: Man nimmt eine Zeile von  $A$  und eine Spalte von  $B$ , bildet das Skalarprodukt dieser beiden Vektoren, und das Ergebnis schreibt man dorthin, wo diese Zeile/Spalte einander kreuzen.

In unserem konkreten Fall sind die Spalten von  $A$  aber gerade die Vektoren  $r_1, r_2, r_3$ , die ein Orthonormalsystem bilden. Das heißt aber gerade  $A^\top A = I$ . Jetzt schauen wir uns die Definition der inversen Matrix an und erkennen daraus sofort:

- Die Matrix  $A$  ist invertierbar.
- Die Inverse zu  $A$  ist in diesem Falle gleich  $A^\top$ .

Wie wir im Abschnitt 1.4 gelernt haben, ist nicht nur  $A^\top A = I$ , sondern auch  $AA^\top = I$ . Das heißt:

Wenn die Spalten von  $A$  ein Orthonormalsystem bilden, dann bilden die Zeilen von  $A$  ebenfalls ein Orthonormalsystem!<sup>44</sup> Es folgt dann auch:  $(A^\top)^{-1} = A$ . Wir können also  $(A^\top)^{-1}$  mit einem Rechenaufwand nahe Null bestimmen !

Damit haben wir die Drehmatrix gefunden:

**Satz 1.66.** *Die Drehung um eine Drehachse durch den Vektor  $r_1$  mit dem Winkel  $\varphi$  wird beschrieben durch eine Drehmatrix  $R(r_1, \varphi)$ , die gegeben wird durch das Produkt*

$$R(r_1, \varphi) = A \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \varphi & -\sin \varphi \\ 0 & \sin \varphi & \cos \varphi \end{pmatrix} A^\top.$$

Hierbei ist  $A$  eine  $3 \times 3$ -Matrix, deren Spalten ein ONS bilden, und die erste Spalte ist gleich dem Vektor  $r_1$  (ggf. normiert auf Länge 1). Die beiden weiteren Spalten  $r_2$  und  $r_3$  von  $A$  sind beliebig wählbar, solange sie gemeinsam mit der ersten Spalte ein ONS bilden. Es ergibt sich bei anderer Wahl dieser Spalten jedesmal dieselbe Matrix  $R(r_1, \varphi)$ .

Es ist am besten, dieses Dreierprodukt von Matrizen stehenzulassen und nicht auszumultiplizieren.

**Satz 1.67.** *Die Spalten von  $R$  bilden ein Orthonormalsystem.*

*Beweis.* Wir müssen nur nachprüfen, daß  $R^\top R = I$  gilt. Zur Abkürzung sei die mittlere Matrix  $R_0$  getauft, also  $R = AR_0A^\top$ . Dann haben wir wegen Satz 1.65

$$\begin{aligned} R^\top R &= (AR_0A^\top)^\top (AR_0A^\top) = ((A^\top)^\top R_0^\top A^\top)(AR_0A^\top) \\ &= AR_0^\top (A^\top A) R_0 A^\top = AR_0^\top R_0 A^\top = A(R_0^\top R_0)A^\top = AA^\top \\ &= I. \end{aligned}$$

□

<sup>44</sup> Wer nicht an die Kraft der Gruppentheorie (insbesondere Satz 1.44) glauben will, möge versuchen, dieses Ergebnis zu Fuß zu beweisen, d.h. zum Beispiel mit den Methoden der Vektorrechnung ...

Solche Matrizen tauchen in Mathematik und Physik so oft auf, daß sie eines eigenen Namens würdig sind:

**Definition 1.68.** Eine Matrix  $A$  mit  $A^T A = I$  heißt orthogonale Matrix<sup>45</sup>.

Die Bedeutung der orthogonalen Matrizen liegt darin, daß sie die Längen von Vektoren nicht ändern: wenn  $x$  irgendein Vektor ist und  $A$  eine orthogonale Matrix, dann haben  $x$  und  $Ax$  dieselbe Länge. Die Umkehrung gilt auch: wenn eine Matrix die Länge von Vektoren nicht ändert, dann ist sie eine orthogonale Matrix.

Jede Bewegung des Raumes, die den Ursprung festhält (also Drehungen, Spiegelungen und jede Komposition davon, **aber keine Verschiebung**) kann durch eine orthogonale Matrix beschrieben werden.

### 1.5.4 Der affine Raum

Wir sollten unbedingt zwei Typen von Vektoren unterscheiden:

- Ortsvektoren,
- Vektoren im engeren Sinne.

Ortsvektoren bezeichnen einen Ort, also einen Punkt im Raum. Ein Ortsvektor startet stets im Ursprung. Ein Beispiel wäre ein Vektor, der zum aktuellen Aufenthaltsort eines Teilchens zeigt.

Die Vektoren im engeren Sinne beschreiben alles weitere: Geschwindigkeiten, Beschleunigungen, Kräfte, Strömungen, elektrische Felder, magnetische Felder usw.

Wenn Ortsvektoren benutzt werden, um einen „geographischen Ort“ anzuzeigen, dann kann man sie nicht addieren: die „Summe zweier geographischen Punkte“ ist in einer solchen Situation sinnlos.

Aber man kann einen Vektor (i.e.S.) an einen Ortsvektor anhängen, um zum Beispiel auszudrücken, wo das Teilchen nach dem Verstreichen eines Zeitintervalles  $\Delta t$  wäre, wenn es seine jetzige Geschwindigkeit beibehielte. In diesem Sinne hätte man einen Ortsvektor und einen Vektor i.e.S. addiert.

Mathematisch drückt man dies durch den Begriff des *affinen Raums* aus. Grob gesprochen besteht der affine Raum aus zwei Dingen:

- einer Menge von Punkten,
- und einer Menge von Vektoren.

Die Vektoren bilden einen Vektorraum wie schon früher definiert.

Zwischen beiden Mengen gelten dabei die folgenden Regeln:

- wenn man an einen Punkt einen Vektor anhängt, landet man wieder bei einem Punkt,
- die gerichtete Verbindungsstrecke zwischen zwei Punkten ergibt einen Vektor.

## 1.6 Schlüsselbegriffe

Die hier aufgeführten Begriffe sind von höchster Bedeutung. Ohne eine sichere Beherrschung dieser Begriffe und Konzepte ist ein Verständnis späterer Vorlesungskapitel nicht möglich.

Sie sollten auch in der Lage sein, die in das Skript eingestreuten Fragen überzeugend zu beantworten.

- algebraische Grundkonzepte: Gruppe, Körper, Vektorraum,
- $\mathbb{R}$  und  $\mathbb{C}$  als Körper, Definition von  $\mathbb{C}$ , geometrische Interpretation der Rechenoperationen in  $\mathbb{C}$ ,
- zwei Sichtweisen auf den Raum der Ortsvektoren in der Ebene: geometrisch und analytisch,
- Produkte Matrix  $\cdot$  Vektor und Matrix  $\cdot$  Matrix,
- Drehungen im  $\mathbb{R}^2$  und  $\mathbb{R}^3$ ,
- Vektorprodukt, Spatprodukt, Determinante,
- affiner Raum im Gegensatz zum Vektorraum.

---

<sup>45</sup>orthogonal matrix

# Kapitel 2

## Vektorräume

Wir haben bisher 4 Vektorräume näher untersucht:

- den geometrisch definierten Raum der Ortsvektoren in der Ebene,
- den  $\mathbb{R}^2$ ,
- den dreidimensionalen Raum der Ortsvektoren,
- und den  $\mathbb{R}^3$ .

Diese Räume waren geometrisch bzw. analytisch definiert. Es gibt noch viele andere wichtige Vektorräume, wie z.B. Funktionenvektorräume, die nicht nur in der Quantenmechanik unverzichtbar sind. Da wir aber nicht jeden einzeln behandeln können, untersuchen wir in diesem Kapitel Eigenschaften, die *jeder* Vektorraum besitzt.

Es gibt verschiedene Typen von Vektorräumen. Eine Unterscheidung beruht darauf, ob der Vektorraum ein Skalarprodukt besitzt oder nicht. Eine zweite Unterscheidung bezieht sich auf die Dimension: sie könnte endlich sein oder auch unendlich. Funktionenvektorräume sind praktisch immer unendlichdimensional, und sie treten auf bei der Modellierung von vielen Prozessen aus der Natur und Technik, z.B. bei der Untersuchung der Ausbreitung von elektrischen Wellen, oder bei der Wettervorhersage, oder bei der Simulation des elastischen Verhaltens von Bauteilen unter mechanischer Belastung. Eine präzise Untersuchung solcher unendlichdimensionaler Vektorräume ist sehr anspruchsvoll (sie findet z.B. im vierten Semester des Mathematikstudiums in der Funktionalanalysisvorlesung statt), und deshalb verzichten wir darauf weitgehend. In der numerischen Simulation zu Anwendungsaufgaben verwendet man stattdessen endlichdimensionale Vektorräume mit sehr hoher Dimension als Näherungsverfahren; z.B. sind bei der Wettervorhersage Gleichungssysteme mit vielen Millionen Unbekannten regelmäßig zu lösen.

### 2.1 Allgemeine Eigenschaften

Sei  $K$  ab jetzt ein Körper. In praktisch sämtlichen Fällen ist  $K = \mathbb{R}$  oder  $K = \mathbb{C}$ .

**Definition 2.1.** *Wir sagen, daß  $V$  einen Vektorraum über  $K$  bildet, wenn zwei Operationen*

$$\begin{aligned} + & : V \times V \rightarrow V, \\ \cdot & : K \times V \rightarrow V \end{aligned}$$

*gegeben sind mit folgenden Eigenschaften:*

- $(V, +)$  ist eine abelsche Gruppe mit neutralem Element  $0 = \vec{0} \in V$ ,
- Wir haben die beiden Distributivgesetze  $(\alpha + \beta) \cdot u = \alpha \cdot u + \beta \cdot u$  und  $\alpha \cdot (u + v) = \alpha \cdot u + \alpha \cdot v$ , für  $\alpha, \beta \in K$  und  $u, v \in V$ ,
- Es gilt das „Assoziativgesetz“  $(\alpha \cdot \beta) \cdot u = \alpha \cdot (\beta \cdot u)$  für  $\alpha, \beta \in K$  und  $u \in V$ ,
- Die Zahl 1 ist auch neutrales Element für die zweite Multiplikation:  $1 \cdot u = u$  für  $u \in V$ .

Ab jetzt werden wir mit 0 sowohl die reelle/komplexe Zahl 0 bezeichnen als auch den Nullvektor in  $V$ . Die Schreibweise  $\vec{0}$  werden wir nur noch in Notfällen anwenden, wenn Mißverständnisse drohen.

Die Multiplikationspunkte (auf die wir im Folgenden auch verzichten wollen) im „Assoziativgesetz“ bezeichnen zwei verschiedene Multiplikationen: einerseits die Multiplikation zweier Zahlen (Körpererelemente), andererseits die Multiplikation einer Zahl mit einem Vektor (und das später betrachtete Skalarprodukt zweier Vektoren wird ein dritter Typ von Multiplikation sein). Deshalb die „.“.

Entsprechend bezeichnet auch das Pluszeichen zwei verschiedene Additionen: von Körpererelementen (also Zahlen) bzw. Vektoren.

Außerdem sei darauf hingewiesen, daß auch das Wort „Vektor“ mindestens zwei verschiedene Bedeutungen besitzt: einerseits ein *konkretes* geometrisches Objekt, welches man sich üblicherweise als Pfeil in der Ebene oder im 3D vorstellt; andererseits ein *abstraktes* mathematisches Objekt, über dessen Beschaffenheit nichts gesagt wird (solange die Menge der Vektoren nur die Bedingungen aus Definition 2.1 erfüllt).

Zur Notation: Ab jetzt bezeichnen lateinische Großbuchstaben (meistens) Vektorräume, lateinische Kleinbuchstaben  $e, f, g, h, u, v, w, \dots$  Vektoren, und griechische Buchstaben Elemente des Körpers  $K$ .

### Beispiel 2.2.

- $\mathbb{R}^n$  als Vektorraum über dem Körper  $\mathbb{R}$ ,
- $\mathbb{C}^n$  als Vektorraum über  $\mathbb{C}$ ,
- $\mathbb{C}^n$  als Vektorraum über  $\mathbb{R}$ ,
- $M^{n \times m}(\mathbb{R})$  — der Vektorraum der Matrizen des Formats  $n \times m$  mit reellen Einträgen (über dem Körper  $\mathbb{R}$ ),
- $\mathbb{R}[X]$  — der Raum der Polynome in der Variablen  $X$  mit Koeffizienten aus  $\mathbb{R}$ ,
- $C([a, b] \rightarrow \mathbb{R})$  — der Raum der auf dem Intervall  $[a, b]$  stetigen und reellwertigen Funktionen,
- $C^k([a, b] \rightarrow \mathbb{R})$  — der Raum der auf dem Intervall  $[a, b]$   $k$ -mal stetig differenzierbaren reellwertigen Funktionen.

Hierbei definieren wir die Addition zweier Funktionen auf naheliegende Weise: seien  $f$  und  $g$  zwei Funktionen, dann wird die Summe  $(f + g) = (f + g)(x)$  definiert als  $f(x) + g(x)$ . Analog definiert man die Vervielfachung einer Funktion gemäß  $(\alpha f) = (\alpha f)(x) = \alpha \cdot (f(x))$ . Der „Nullvektor“  $0 \in V$  ist diejenige Funktion, die auf dem gesamten Intervall überall den Wert  $0 \in \mathbb{R}$  hat.

**Frage:** Warum kann man nicht  $\mathbb{R}^n$  als Vektorraum über dem Körper  $\mathbb{C}$  betrachten ?

Wie man sieht, kann alles mögliche einen Vektor darstellen. Deshalb ist es gar nicht so selbstverständlich, daß die Aussagen des nächsten Satzes stimmen, auch wenn sie offensichtlich erscheinen:

**Satz 2.3.** Für beliebige Vektoren  $u, v \in V$  und eine beliebige Zahl  $\alpha \in K$  gelten die Regeln

1.  $0 \cdot u = \vec{0}$ ,
2.  $\alpha \cdot \vec{0} = \vec{0}$ ,
3.  $(-1) \cdot u = -u$ .

Hierbei bezeichnet  $(-u)$  denjenigen Vektor aus  $V$ , der zu  $u$  addiert den Nullvektor ergibt:  $(-u) + u = \vec{0}$ .

*Beweis.* Wie benutzen nichts weiter als die Eigenschaften aus Definition 2.1:

1. Es ist  $u = 1 \cdot u = (1 + 0) \cdot u = 1 \cdot u + 0 \cdot u = u + 0 \cdot u$ . Weil  $(V, +)$  eine additive Gruppe ist und in additiven Gruppen subtrahiert werden darf, muß  $0 \cdot u$  gleich dem neutralen Element der Addition sein, also gleich dem Nullvektor.
2. Nach demselben Muster wie eben ist  $\alpha \cdot u = \alpha(u + \vec{0}) = \alpha \cdot u + \alpha \cdot \vec{0}$ , und  $\alpha \cdot \vec{0}$  muß gleich dem Nullvektor sein.
3. Wir benutzen das erste Ergebnis:  $\vec{0} = 0 \cdot u = (1 + (-1)) \cdot u = 1 \cdot u + (-1) \cdot u = u + (-1) \cdot u$ . Andererseits ist  $\vec{0} = u + (-u)$ , und es muß also  $(-1) \cdot u = -u$  sein.

□



## 2.2 Linearkombinationen, Erzeugendensysteme, Lineare Unabhängigkeit, Basen und Dimensionen

### 2.2.1 Linearkombinationen und Unterräume

**Definition 2.4 (Linearkombination, Span, endlich erzeugt).** Seien  $u_1, \dots, u_n \in V$  Vektoren und  $\alpha_1, \dots, \alpha_n \in K$  Zahlen. Dann heißt der Vektor

$$\alpha_1 u_1 + \dots + \alpha_n u_n \in V$$

Linearkombination<sup>1</sup> der Vektoren  $u_1, \dots, u_n$ . Wenn wir die  $u_j$  festhalten und die  $\alpha_j$  durch ganz  $K$  laufen lassen, erhalten wir den Span der Vektoren  $u_1, \dots, u_n$ :

$$\text{span}(u_1, \dots, u_n) := \left\{ \sum_{j=1}^n \alpha_j u_j : \alpha_1, \dots, \alpha_n \in K \right\}.$$

Diese Menge wird auch der von den  $u_j$  aufgespannte Unterraum oder lineare Hülle<sup>2</sup> der  $u_j$  genannt. Wir sagen auch, daß die Vektoren  $u_1, \dots, u_n$  die Menge  $\text{span}(u_1, \dots, u_n)$  erzeugen. Diese Vektoren bilden ein sogenanntes Erzeugendensystem.

Als Sonderfall legen wir fest, daß die leere Menge denjenigen Raum aufspannt, der nur aus dem Nullvektor besteht:

$$\text{span}(\emptyset) := \{\vec{0}\}.$$

Sei  $\mathfrak{U} \subset V$  eine Menge unendlich vieler Vektoren aus  $V$ . Dann definieren wir  $\text{span}(\mathfrak{U})$  als

$$\text{span}(\mathfrak{U}) = \{ \text{alle Linearkombinationen aus endlich vielen Elementen aus } \mathfrak{U} \}.$$

Man beachte, daß  $\mathfrak{U}$  auch überabzählbar<sup>3</sup> viele Elemente enthalten darf.

Sei  $V$  ein Vektorraum. Wenn es endlich viele Elemente  $u_1, \dots, u_n \in V$  gibt mit  $V = \text{span}(u_1, \dots, u_n)$ , dann heißt  $V$  endlich erzeugt.

**Definition 2.5 (Unterraum).** Sei  $V$  ein Vektorraum über dem Körper  $K$ , und sei  $U$  eine Teilmenge von  $V$ . Wenn  $U$  den Nullvektor von  $V$  enthält und wenn  $U$  gemeinsam mit den von  $V$  auf  $U$  eingeschränkten<sup>4</sup> Operationen „+“ und „·“ wieder einen Vektorraum über dem Körper  $K$  bildet, dann heißt  $U$  Untervektorraum von  $V$ <sup>5</sup>. Man sagt auch Unterraum.

Wenn jetzt eine solche Teilmenge  $U \subset V$  gegeben ist und wir wissen wollen, ob ein Unterraum vorliegt, dann sind sämtliche Eigenschaften aus Definition 2.1 nachzuprüfen. Zum Glück läßt sich diese Rechnerei abkürzen, denn wir haben den folgenden Satz:

**Satz 2.6 (Untervektorraumkriterium).** Sei  $V$  ein Vektorraum über dem Körper  $K$  und  $U \subset V$  mit folgenden Eigenschaften:

1.  $U$  enthält wenigstens ein Element,
2. wenn  $u \in U$  und  $v \in U$ , dann ist auch  $u + v \in U$ ,
3. wenn  $u \in U$  und  $\alpha \in K$ , dann ist auch  $\alpha u \in U$ .

<sup>1</sup>linear combination

<sup>2</sup>linear hull, span

<sup>3</sup>Eine Menge  $M$  heißt *abzählbar*, wenn ihre Elemente durchnumeriert werden können, wobei man zum Numerieren nur natürliche Zahlen verwenden darf. In mathematischer Formulierung: es ist  $M$  abzählbar, wenn es eine bijektive Abbildung von  $\mathbb{N}$  nach  $M$  gibt.

Wenn  $M$  so viele Elemente hat, daß eine solche Numerierung niemals möglich ist, dann heißt  $M$  *überabzählbar*. Es ist  $\mathbb{Q}$  abzählbar, aber  $\mathbb{R}$  überabzählbar.

Ein Skalarproduktvektorraum, der eine abzählbare Teilmenge besitzt, die jedes Element des Vektorraums beliebig genau annähert, wird in der Quantenmechanik (also im IK4) als *separabel* bezeichnet. Solche Vektorräume sind „schön“.

<sup>4</sup>im Sinne eines verkleinerten Definitionsbereiches

<sup>5</sup>linear subspace of  $V$

Dann ist  $U$  ein Unterraum von  $V$ .

Wir sagen auch, daß  $U$  abgeschlossen unter den Operationen  $+$  und  $\cdot$  ist.

*Beweis.* Zunächst ist zu prüfen, ob für die Operationen  $+$  und  $\cdot$  die Definitionsbereiche und Wertebereiche passen: weil  $V$  ein Vektorraum ist, haben wir  $+$  und  $\cdot$  als Abbildungen  $+: V \times V \rightarrow V$  und  $\cdot: K \times V \rightarrow V$ . Wenn wir diese Operationen einschränken, bekommen wir  $+: U \times U \rightarrow V$  und  $\cdot: K \times U \rightarrow V$ . Die Voraussetzungen 2. und 3. besagen gerade, daß sogar  $+: U \times U \rightarrow U$  und  $\cdot: K \times U \rightarrow U$  gelten.

Die üblichen Gesetze (Kommutativität, Assoziativität und Distributivität) vererben sich von  $V$  auf  $U$ . Es gilt auch  $1 \cdot u = u$  für jedes  $u \in U$ , weil diese Beziehung sogar für jedes  $u \in V$  gilt, und  $U \subset V$ .

Wir müssen jetzt bloß noch zeigen, daß wir auch in  $U$  ein neutrales Element zur Addition haben, und daß es in  $U$  zu jedem Element auch ein additiv inverses Element gibt. Nach Voraussetzung hat  $U$  wenigstens ein Element, nennen wir es  $u^*$ . Nach Voraussetzung 3. ist dann auch  $0_K \cdot u^* \in U$ . Aber nach Satz 2.3 ist  $0_K \cdot u^* = 0_V$ , also ist  $0_V$  ein Element von  $U$ .

Sei nun  $u \in U$  gegeben und ein additiv inverses Element dazu in  $U$  gesucht. Dieses ist  $-u \in V$ , aber nach Satz 2.3 ist  $-u = (-1) \cdot u$ , was jedoch nach Voraussetzung 3. ein Element von  $U$  ist und nicht bloß ein Element von  $V$ .  $\square$

**Frage:** Seien  $x_1, x_2 \in \mathbb{R}^3$  beliebige Vektoren. Ist dann  $\text{span}(x_1, x_2)$  eine Ebene im  $\mathbb{R}^3$  durch den Ursprung?

**Satz 2.7.** Sei  $V$  ein Vektorraum,  $u_1, \dots, u_n \in V$ , und sei  $U = \text{span}(u_1, \dots, u_n)$ . Dann haben wir folgende Regeln:

1.  $U$  ist ein Untervektorraum von  $V$ .
2. Wenn  $u \in V$  ein weiterer Vektor ist, dann ist  $U \subset \text{span}(u_1, \dots, u_n, u)$ .
3. Wenn  $u$  sogar in  $U$  liegt, dann ist  $U = \text{span}(u_1, \dots, u_n, u)$ .

*Beweis.* 1. Man benutze das Untervektorraumkriterium.

2. Fast trivial: wenn man einen weiteren Vektor (nämlich  $u$ ) beim Bilden von Linearkombinationen erlaubt, dann kann man mindestens die gleichen Vektoren erzeugen wie zuvor (indem man z.B. diesen Vektor  $u$  mit dem Koeffizienten  $0 \in K$  wichtet).
3. Wenn  $u \in \text{span}(u_1, \dots, u_n)$ , dann haben wir Koeffizienten  $\beta_j \in K$  mit  $u = \sum_{j=1}^n \beta_j u_j$ . Dann können wir jede Linearkombination von  $u_1, \dots, u_n, u$  als Linearkombination von  $u_1, \dots, u_n$  schreiben:

$$\alpha_1 u_1 + \dots + \alpha_n u_n + \alpha u = (\alpha_1 + \alpha \beta_1) u_1 + \dots + (\alpha_n + \alpha \beta_n) u_n.$$

Das beweist  $\text{span}(u_1, \dots, u_n, u) \subset \text{span}(u_1, \dots, u_n) = U$ . Andererseits haben wir aus 2., daß  $\text{span}(u_1, \dots, u_n, u) \supset \text{span}(u_1, \dots, u_n) = U$ . Damit ist der Beweis beendet.  $\square$

## 2.2.2 Lineare Unabhängigkeit und Basen

**Definition 2.8 (Linear abhängig, linear unabhängig, frei, Basis).** Sei  $V$  ein Vektorraum über  $K$ , und sei  $(u_1, \dots, u_n)$  eine Familie von Vektoren aus  $V$ .

1. Die Familie  $(u_1, \dots, u_n)$  heißt linear abhängig<sup>6</sup>, wenn es Zahlen  $\alpha_1, \dots, \alpha_n \in K$  gibt, von denen wenigstens eine nicht 0 ist, mit

$$\alpha_1 u_1 + \dots + \alpha_n u_n = \vec{0}.$$

2. Wenn die Familie  $(u_1, \dots, u_n)$  nicht linear abhängig ist, dann heißt sie linear unabhängig<sup>7</sup> oder frei. Das heißt, die einzige Möglichkeit, Zahlen  $\alpha_1, \dots, \alpha_n \in K$  zu wählen, sodaß  $\alpha_1 u_1 + \dots + \alpha_n u_n = \vec{0}$ , ist  $\alpha_1 = \dots = \alpha_n = 0$ .

<sup>6</sup>linearly dependent

<sup>7</sup>linearly independent

3. Sei  $\mathfrak{U}$  eine Familie von unendlich vielen Vektoren aus  $V$ . Diese Familie heißt linear unabhängig, wenn jede endliche Teilmenge von  $\mathfrak{U}$  linear unabhängig ist. Wenn  $\mathfrak{U}$  nicht linear unabhängig ist, dann heißt  $\mathfrak{U}$  linear abhängig.
4. Sei  $\mathfrak{U}$  eine Familie von Vektoren aus  $V$  (endlich viele oder unendlich viele). Wenn  $\mathfrak{U}$  ein Erzeugendensystem von  $V$  ist und linear unabhängig ist, dann heißt  $\mathfrak{U}$  Basis<sup>8</sup> von  $V$ .

**Bemerkung 2.9.** Wir vermerken zwei Spezialfälle:

- Die Familie  $(\vec{0})$ , die nur den Nullvektor aus  $V$  enthält, ist linear abhängig.
- Die leere Familie, die nichts enthält (nicht einmal den Nullvektor), ist linear unabhängig und spannt den Nullvektorraum  $\{\vec{0}\}$  auf. Die leere Familie ist Basis des Nullvektorraums.

Diese Sprechweisen werden uns später erlauben, Aussagen einfacher zu formulieren.

**Frage:** Was bedeutet die lineare Unabhängigkeit in Funktionenräumen ?

**Frage:** Sei  $V = \mathbb{R}[X]$  und  $\mathfrak{U} = \{1, x, x^2, x^3, \dots\}$ . Ist diese Familie linear abhängig ?

Damit eine Familie  $\mathfrak{U}$  eine Basis von  $V$  sein kann, muß sie zwei Eigenschaften erfüllen:

- es muß  $\mathfrak{U}$  den ganzen Raum  $V$  aufspannen,
- $\mathfrak{U}$  muß linear unabhängig sein.

Die erste Forderung besagt anschaulich, daß  $\mathfrak{U}$  „nicht zu wenige“ Elemente hat; und die zweite Forderung besagt anschaulich, daß  $\mathfrak{U}$  „nicht zu viele“ Elemente besitzt.

Als schnell beweisbare Merkgeregeln haben wir dann:

*Eine Familie  $\mathfrak{U}$  ist Erzeugendensystem eines Vektorraums  $V$  genau dann,  
wenn jeder Vektor aus  $V$  darstellbar ist als Linearkombination von Elementen aus  $\mathfrak{U}$ .*

*Eine Familie  $\mathfrak{U}$  ist linear unabhängig in  $V$  genau dann,  
wenn der Nullvektor  $0_V$  genau auf eine einzige Weise dargestellt werden kann  
als Linearkombination von Elementen aus  $\mathfrak{U}$ .*

**Satz 2.10.** Eine Familie  $(u_1, \dots, u_n)$  ist eine Basis von  $V$  genau dann, wenn jeder Vektor  $u \in V$  auf genau eine Art und Weise als Linearkombination der  $u_1, \dots, u_n$  dargestellt werden kann.

*Beweis.* Wir haben zwei Richtungen zu beweisen.

„ $\implies$ “:

Sei  $(u_1, \dots, u_n)$  eine Basis, und sei  $u \in V$  ein beliebiger Vektor. Wegen der ersten Merkgregel können wir den Vektor  $u$  auf mindestens eine Weise als Linearkombination der  $u_j$  darstellen:

$$u = \alpha_1 u_1 + \dots + \alpha_n u_n.$$

Wir müssen noch zeigen, daß dies die einzige Möglichkeit ist. Sei also außerdem noch  $u = \beta_1 u_1 + \dots + \beta_n u_n$ . Wenn wir beide Darstellungen abziehen, bekommen wir

$$\vec{0} = (\alpha_1 - \beta_1)u_1 + \dots + (\alpha_n - \beta_n)u_n.$$

Weil nun allerdings die Vektoren  $u_1, \dots, u_n$  eine linear unabhängige Familie bilden, läßt sich der Nullvektor nur darstellen, wenn  $(\alpha_j - \beta_j) = 0$  ist für jedes  $j$ . Es ist also  $\beta_j = \alpha_j$ .

„ $\impliedby$ “:

Wenn jeder Vektor genau eine Darstellung hat, dann hat er mindestens eine Darstellung. Wegen der ersten Merkgregel ist dann die Familie  $(u_1, \dots, u_n)$  also ein Erzeugendensystem. Und wegen der zweiten Merkgregel bilden die  $u_j$  eine linear unabhängige Familie, und somit auch eine Basis.  $\square$

Dieser Satz gilt auch dann, wenn diese Basis unendlich viele Elemente enthalten sollte; und auch der Beweis ist im wesentlichen unverändert.

---

<sup>8</sup>basis

**Lemma 2.11.** *Seien  $u_1, \dots, u_n \in V$  linear unabhängig, und sei  $v \notin \text{span}(u_1, \dots, u_n)$ . Dann ist auch  $(u_1, \dots, u_n, v)$  eine linear unabhängige Familie.*

**Frage:** Was bedeutet dies geometrisch ?

*Beweis.* Um die lineare Unabhängigkeit zu zeigen, setzen wir

$$\alpha_1 u_1 + \dots + \alpha_n u_n + \alpha v = 0$$

und müßten als nächstes zeigen, daß  $\alpha_1 = \dots = \alpha_n = \alpha = 0$  gilt.

Angenommen,  $\alpha \neq 0$ . Dann hätten wir

$$v = \left(-\frac{\alpha_1}{\alpha}\right) u_1 + \dots + \left(-\frac{\alpha_n}{\alpha}\right) u_n,$$

also  $v \in \text{span}(u_1, \dots, u_n)$ . Wir hatten aber gerade das Gegenteil vorausgesetzt. Also muß  $\alpha = 0$  sein. Damit bekommen wir  $\alpha_1 u_1 + \dots + \alpha_n u_n = 0$ . Weil die Vektoren  $u_j$  aber eine linear unabhängige Familie bilden, müssen die  $\alpha_j$  sämtlich gleich Null sein.  $\square$

**Satz 2.12 (Basisergänzungssatz).** *Sei  $(v_1, \dots, v_n)$  ein Erzeugendensystem des Vektorraumes  $V$ , und sei  $(u_1, \dots, u_m)$  eine linear unabhängige Familie in  $V$ .*

*Dann kann man die Familie  $(u_1, \dots, u_m)$  mit passend gewählten Vektoren  $v_j$  zu einer Basis von  $V$  ergänzen. Insbesondere existiert ein Index  $p \in \mathbb{N}$ , sodaß (bei geschickter Numerierung der  $v_j$ ) die Familie  $(u_1, \dots, u_m, v_{p+1}, \dots, v_n)$  eine Basis von  $V$  ist.*

*Beweis.* Wenn die Familie  $(u_1, \dots, u_m)$  schon den Vektorraum  $V$  aufspannt, dann sind wir fertig. Wir brauchen nichts ergänzen und setzen  $p = n$ . Anderenfalls existiert ein Vektor  $v \in V$ , der nicht durch die  $u_j$  dargestellt werden kann. Dann muß ein  $v_k$  aus dem Erzeugendensystem existieren, das nicht durch die  $u_j$  dargestellt werden kann (Begründung bitte selbst finden). Mit diesem  $v_k$  benutzen wir Lemma 2.11 und erhalten eine linear unabhängige Familie  $(u_1, \dots, u_m, v_k)$ . Wir numerieren die  $v_j$  so um, daß  $v_k$  jetzt  $v_n$  heißt. Mit dieser neuen Familie beginnen wir wieder von vorn. Nach  $n$  Schritten ist das Erzeugendensystem  $(v_1, \dots, v_n)$  in die linear unabhängige Familie der  $u_j$  eingefügt worden, und der Beweis ist spätestens dann fertig.  $\square$

Eine Familie  $(u_1)$  ist linear unabhängig, wenn  $u_1 \neq 0$ . Auf diesem Wege erhalten wir:

**Satz 2.13 (Basis-Satz).** *Jeder endlich erzeugte Vektorraum besitzt eine Basis.*

Die Voraussetzung „endlich erzeugt“ ist übrigens nicht nötig: *Jeder* Vektorraum hat eine Basis. Auf einen Beweis müssen wir verzichten, da die erforderlichen Hilfsmittel aus der Mengenlehre hier nicht bereitgestellt werden können.

Weiterhin gilt:

**Satz 2.14.** *Jedes Erzeugendensystem enthält eine Basis.*

*Beweis.* Man wende den Basisergänzungssatz geschickt an.  $\square$

**Frage:** Im  $\mathbb{R}^{1077}$  seien drei linear unabhängige Vektoren gegeben. Man entscheide, ob diese drei Vektoren den  $\mathbb{R}^3$  aufspannen.

{ Es folgen fakultative Betrachtungen, was alles in unendlichdimensionalen Räumen schiefgehen kann. }

Ein Vektorraum heißt unendlichdimensional, wenn es in ihm beliebig große linear unabhängige Familien gibt. Unser Basisbegriff für solche Vektorräume verlangt, daß jeder Vektor dargestellt werden kann durch eine *endliche* Linearkombination der Basisvektoren. Der Grund für diese Beschränkung auf *endliche* Linearkombinationen liegt darin, daß wir gar nicht wissen, wie man unendlich viele Summanden addieren würde. Und ohne eine Norm ist eine solche Summation auch kaum definierbar.

Wir schauen uns mal für ein Beispiel, bei dem eine Summation von unendlich vielen Summanden möglich zu sein scheint, an, was dann trotzdem noch passieren kann. Es sei  $V$  der Vektorraum aller stetigen Funktionen von  $\mathbb{R}$  nach  $\mathbb{R}$ , die im Unendlichen abklingen. Diesen Vektorraum statten wir mit der Norm  $\|f\|_V := \max_{x \in \mathbb{R}} |f(x)|$  aus, was dann eine Norm gemäß Satz 2.26 ergibt.

Für  $k \in \mathbb{N}_{-1} := \{-1, 0, 1, 2, 3, 4, \dots\}$  definieren wir Funktionen  $v_k$  gemäß

$$v_k(x) := \max(1 - 2^{k+1}|x|, 0), \quad x \in \mathbb{R}.$$

Diese stetigen Funktionen sind identisch Null für  $|x| \geq 2^{-k-1}$ , und auf dem Intervall  $[-2^{-k-1}, 2^{-k-1}]$  sind sie jeweils stückweise linear. Für große  $k$  erhalten wir also Funktionen, die nur auf einem sehr kurzen Intervall ungleich Null sind. Anschließend schieben wir solche Funktionen auf der reellen Achse hin und her. Das heißt, für  $m \in \mathbb{Z}$  setzen wir

$$v_{-1,m}(x) := v_{-1}(x - m), \quad x \in \mathbb{R},$$

was eine verschobene Kopie von  $v_{-1}$  ist. Wir beobachten, daß die Dreiecksteile der Graphen von zwei benachbarten Funktionen  $v_{-1,m}$  einander überlappen.

Weiterhin setzen wir, für  $k \in \mathbb{N}_0 := \{0, 1, 2, 3, \dots\}$  und  $m \in \mathbb{Z}$ ,

$$v_{k,m}(x) := v_k(x - 2^{-k-1} - 2^{-k}m), \quad x \in \mathbb{R},$$

also eine verschobene Kopie von  $v_k$  (jetzt überlappen die Graphen einander nicht mehr).

Das sollen unsere „Basis“-Funktionen sein. Wir verwenden deshalb Anführungszeichen, weil wir absichtlich jetzt unendliche Linearkombinationen erlauben wollen, die bisher ja unzulässig waren. Diese „Basis“-Funktionen werden für große  $k$  sehr „schmal“, sodaß die Hoffnung besteht, daß man wirklich jede Funktion aus dem Vektorraum  $V$  damit zusammenbauen kann. Das funktioniert tatsächlich. Denn sei  $f$  eine Funktion aus  $V$ . Für  $K \in \mathbb{N}_0$  setzen wir

$$\begin{aligned} f_K(x) := & \sum_{|m| \leq K} f(m)v_{-1,m}(x) \\ & + \sum_{0 \leq k \leq K; |m| \leq K} \left( f(2^{-k}m + 2^{-k-1}) - \frac{1}{2}f(2^{-k}m) - \frac{1}{2}f(2^{-k}m + 2^{-k}) \right) v_{k,m}(x), \end{aligned}$$

was tatsächlich eine endliche Linearkombination von Funktionen  $v_{l,n}$  ist. Man ahnt oder glaubt, daß  $f_K$  eine Annäherung von  $f$  ist, wenn  $K$  sehr groß ist. Damit ist gemeint, daß die Normdifferenz  $|f_K - f|_V$  beliebig klein wird, wenn man nur  $K$  genügend groß nimmt. Tatsächlich läßt sich sogar beweisen, daß

$$f(x) = \sum_{K=0}^{\infty} \left( f_{K+1}(x) - f_K(x) \right) + f_0(x), \quad (2.1)$$

und die Konvergenz der Reihe geschieht sogar im Sinne von Definition 5.25. Das ist eine sehr schöne Konvergenz.

Eigentlich könnten wir uns jetzt freuen, es gibt nur einen Haken. Durch die Klammersetzung in (2.1) wird erreicht, daß die „Basis“-Elemente  $v_{k,m}$  in einer ganz bestimmten Reihenfolge aufsummiert werden. Es stellt sich die Frage, ob man denn nicht die Anordnung dieser Summanden ändern könnte (im Sinne eines Assoziativgesetzes bzw. Kommutativgesetzes der Addition, bloß jetzt mit unendlich vielen Summanden). Eigentlich sollten solche Gesetze ja gelten. Die schlechte Nachricht ist aber: wenn man die „Basis“-Elemente clever anders anordnet, dann kann man erzielen, daß die Reihensumme nicht mehr  $f$  ist, sondern etwas anderes. *Ein Summenwert kann von der Reihenfolge der Summanden abhängen!* Das ist seltsam, aber wahr. Man sagt, daß die  $v_{k,m}$  eine *bedingte Basis* des Vektorraums  $V$  bilden.

Nun gut, das legt die Vermutung nahe, daß wir uns bei der Wahl einer „Basis“ für  $V$  einfach ungeschickt angestellt haben; und wenn man besser wählt, dann kommt bei beliebiger Anordnung der Summanden der „Linearkombination“ immer derselbe Reihenwert heraus.

Aber dem ist eben nicht so. Egal, wie man in diesem Raum  $V$  eine „Basis“ wählt: es wird stets so sein, daß verschiedene Anordnungen der Summanden der „Linearkombination“ auf unterschiedliche Summenwerte führen werden, obwohl jede Reihe konvergiert. Dieser Vektorraum  $V$  ist ein schrecklicher Raum, denn er besitzt keine *unbedingte Basis*. Leider.

{ Ende der fakultativen Betrachtungen }

### 2.2.3 Dimension eines Vektorraumes

**Frage:** Man gebe jeweils mehrere Basen an für

- $\mathbb{R}^3$ ,
- $\mathbb{C}^3$  als Vektorraum über  $\mathbb{C}$ ,
- $\mathbb{C}^3$  als Vektorraum über  $\mathbb{R}$ ,
- $\mathbb{R}[X]$ .

Wir haben also für einen Vektorraum verschiedene Basen gefunden, die gleichviel Vektoren enthalten. Es stellt sich die Frage, ob dies immer so ist. Denn nur wenn jede Basis eines Vektorraums gleichviel Elemente enthält, hat es überhaupt Sinn, von einer Dimension (als Anzahl der Basisvektoren) zu reden.

Der Beweis dieser Aussage verteilt sich über mehrere Sätze und Lemmata. Wir beginnen:

**Lemma 2.15 (Austauschlemma).** *Sei  $(v_1, \dots, v_n)$  eine Basis für  $V$ , und sei  $u$  ein Vektor mit der Darstellung  $u = \alpha_1 v_1 + \dots + \alpha_n v_n$ . Wenn für ein  $k$  der Koeffizient  $\alpha_k \neq 0$  ist, dann können wir den Basisvektor  $v_k$  gegen  $u$  austauschen und erhalten wieder eine Basis  $(v_1, \dots, v_{k-1}, u, v_{k+1}, \dots, v_n)$ .*

Es gibt immer ein solches  $k$ ,  $1 \leq k \leq n$ , mit  $\alpha_k \neq 0$ ; es sei denn  $u$  ist der Nullvektor.

*Beweis.* Wir nehmen an, daß die Basisvektoren so numeriert sind, daß  $k = 1$ . Zu zeigen sind zwei Dinge:

$(u, v_2, \dots, v_n)$  **erzeugt**  $V$ : Wir haben  $u = \alpha_1 v_1 + \dots + \alpha_n v_n$  mit  $\alpha_1 \neq 0$ , also

$$v_1 = \frac{1}{\alpha_1} u + \left( -\frac{\alpha_2}{\alpha_1} \right) v_2 + \dots + \left( -\frac{\alpha_n}{\alpha_1} \right) v_n \in \text{span}(u, v_2, \dots, v_n).$$

Nun benutzen wir Satz 2.7 mehrfach:

$$V \supset \text{span}(u, v_2, \dots, v_n) = \text{span}(u, v_1, v_2, \dots, v_n) \supset \text{span}(v_1, v_2, \dots, v_n) = V.$$

Nach dem Sandwichprinzip ist dann  $V = \text{span}(u, v_2, \dots, v_n)$ .

$(u, v_2, \dots, v_n)$  **ist linear unabhängig**: Wegen  $\alpha_1 \neq 0$  ist  $u \notin \text{span}(v_2, \dots, v_n)$  (Begründung bitte selbst suchen!). Nun brauchen wir bloß Lemma 2.11 anwenden, und der Beweis ist komplett. □

Als nächsten Schritt auf dem Weg zur Definition der Dimension wollen wir nicht nur einen Vektor austauschen, sondern mehrere:

**Satz 2.16 (Austauschsatz von Steinitz (ERNST STEINITZ, 1871–1928)).** *Sei  $(v_1, \dots, v_n)$  eine Basis, und sei  $(u_1, \dots, u_k)$  eine linear unabhängige Familie von  $V$ . Dann gilt:*

- $k \leq n$ ,
- bei geeigneter Numerierung der  $v_j$  ist  $(u_1, \dots, u_k, v_{k+1}, \dots, v_n)$  eine Basis von  $V$ .

Wir entfernen also aus der Familie der  $v_j$   $k$  Elemente und fügen stattdessen die  $u_j$  ein. Für  $k = 1$  erhalten wir gerade das Austauschlemma.

*Beweis.* Wir führen den Beweis mittels vollständiger Induktion<sup>9</sup> über  $k$ .

<sup>9</sup> Die vollständige Induktion ist ein Beweisverfahren der Mathematik.

Das Ziel ist stets, eine unendliche Kette  $A(1), A(2), A(3), \dots$  von Aussagen zu beweisen. Der Weg zum Beweis führt über die Methode der vollständigen Induktion, die aus zwei Schritten besteht.

In einem ersten Schritt (Induktionsanfang) beweist man die Aussage  $A(1)$ .

Im zweiten Schritt (Induktionsschritt) zeigt man (für jedes  $k$ ): Wenn  $A(k)$  wahr ist (Induktionsvoraussetzung), dann ist auch  $A(k+1)$  wahr (Induktionsbehauptung). Beim Beweis der I.B. darf dabei die I.V. benutzt werden.

Zum Beispiel soll gezeigt werden: es seien  $n$  Geraden (mit  $n \geq 1$ ) in der Ebene gegeben, wobei keine zwei Geraden identisch sein sollen (davon abgesehen ist über die Lage der Geraden nichts bekannt). Die Geraden zerlegen die Ebene in viele geradlinig berandete Stücke. Zu zeigen ist, daß diese Stücke so mit den Farben schwarz und weiß eingefärbt werden können, daß entlang einer Kante benachbarte Stücke immer unterschiedliche Farben besitzen. Diese Behauptung (formuliert für  $n$  Geraden) nennen wir  $A(n)$ .

Der Beweis von  $A(1)$  ist einfach: es gibt nur eine Gerade, und diese zerlegt die Ebene in zwei Hälften, von denen wir die eine weiß färben und die andere schwarz.

Sei nun  $n \geq 1$ , und seien  $n+1$  Geraden gegeben. Zu zeigen ist, daß eine Färbung möglich ist. Wir setzen also die Aussage  $A(n)$  voraus und wollen die Aussage  $A(n+1)$  zeigen. Dazu wählen wir eine Gerade aus und entfernen diese (merken uns aber, wo sie lag). Übrig bleiben  $n$  Geraden, für die eine zulässige Färbung möglich ist (denn das ist gerade die Induktionsvoraussetzung). Diese Färbung nehmen wir. Jetzt fügen wir die  $(n+1)$ -te Gerade wieder hinzu. Diese zerlegt die Ebene in zwei Hälften, und in einer von diesen Hälften tauschen wir überall schwarz und weiß gegeneinander aus. Man sieht, daß auf diese Weise eine zulässige Färbung für den Fall von  $n+1$  Geraden entsteht.

Damit ist die Aussage  $A(n)$  für jedes  $n \in \mathbb{N}$  bewiesen.

Für den Induktionsanfang sei  $k = 0$ . Dann gibt es die Familie der  $u_j$  gar nicht, und es ist auch nichts auszutauschen.

Sei nun  $k > 0$ . Wir setzen voraus, daß der Austauschsatz von Steinitz für  $k - 1$  schon bewiesen ist, und jetzt wollen wir ihn für  $k$  zeigen. Wir wissen damit, daß

- $k - 1 \leq n$ ,
- und daß bei geeigneter Numerierung der  $v_j$  die Familie  $(u_1, \dots, u_{k-1}, v_k, \dots, v_n)$  eine Basis von  $V$  ist.

Wir zeigen als erstes, daß  $k - 1 = n$  nicht sein kann. Denn dann gäbe es in der Familie  $(u_1, \dots, u_{k-1}, v_k, \dots, v_n)$  die  $v$ -Vektoren gar nicht, und die Vektoren  $(u_1, \dots, u_{k-1})$  würden eine Basis von  $V$  bilden. Dann könnte man  $u_k$  als Linearkombination der Vektoren  $u_1, \dots, u_{k-1}$  schreiben. Wir hatten aber vorausgesetzt, daß die Familie  $(u_1, \dots, u_k)$  linear unabhängig ist. Das ist ein Widerspruch, also ist  $k - 1 \neq n$ . Zusammen mit  $k - 1 \leq n$  ergibt das  $k \leq n$ .

Um den zweiten • zu beweisen, stellen wir  $u_k$  in der neuen Basis dar:

$$u_k = \alpha_1 u_1 + \dots + \alpha_{k-1} u_{k-1} + \beta_k v_k + \dots + \beta_n v_n.$$

Wenn jedes  $\beta_j = 0$  wäre, dann hätten wir  $u_k$  als Linearkombination von  $u_1, \dots, u_{k-1}$  dargestellt, was nicht möglich ist, wie wir eben sahen. Also muß mindestens ein  $\beta_j \neq 0$  sein. Wir numerieren die  $v_i$  so um, daß  $\beta_k \neq 0$  ist (wir tauschen also  $j$  und  $k$ ). Dann gibt uns das Austauschlemma die Möglichkeit, in der Basis  $(u_1, \dots, u_{k-1}, v_k, \dots, v_n)$  den Vektor  $v_k$  gegen  $u_k$  auszutauschen, was beweist, daß  $(u_1, \dots, u_{k-1}, u_k, v_{k+1}, \dots, v_n)$  eine Basis ist.

Nach dem Prinzip der vollständigen Induktion ist somit der Beweis vollendet.  $\square$

Daraus folgt fast sofort, daß die Basen eines Vektorraumes gleichlang sind:

**Satz 2.17.** 1. Jeder endlich erzeugte Vektorraum hat eine Basis endlicher Länge.

2. Wenn ein Vektorraum eine endliche Basis besitzt, dann hat jede Basis dieses Vektorraumes dieselbe Länge.
3. Wenn ein Vektorraum zu jedem  $n \in \mathbb{N}$  eine linear unabhängige Familie mit  $n$  Elementen besitzt, dann hat dieser Vektorraum keine endliche Basis.

*Beweis.* 1. Folgt aus dem Basissatz.

2. Seien  $(u_1, \dots, u_k)$  und  $(v_1, \dots, v_n)$  zwei Basen eines Vektorraumes  $V$ . Aus dem Austauschsatz von Steinitz folgt dann  $k \leq n$ . Wenn wir die Rollen der  $u_j$  und  $v_i$  vertauschen und den Satz von Steinitz ein weiteres Mal anwenden, erhalten wir  $n \leq k$ . Das ergibt  $k = n$ .
3. Folgt ebenfalls aus dem Satz von Steinitz.

$\square$

**Definition 2.18 (Dimension).** Für einen Vektorraum  $V$  definieren wir seine Dimension<sup>10</sup> als

$$\dim V = \begin{cases} 0 & : V = \{\vec{0}\}, \\ n & : V \text{ hat eine Basis } (u_1, \dots, u_n), \\ \infty & : V \text{ enthält beliebig große linear unabhängige Familien.} \end{cases}$$

Wegen Bemerkung 2.9 kann der erste Fall der Klammer auf der rechten Seite auch dem zweiten Fall zugeschlagen werden, worauf wir aus didaktischen Gründen aber verzichtet haben.

Um die Begriffe einzuüben, empfiehlt es sich folgenden Satz zu beweisen, und denkbare Hilfsmittel wären der Austauschsatz und der Basisergänzungssatz:

**Satz 2.19.** Sei  $U$  Unterraum eines  $n$ -dimensionalen Vektorraumes  $V$ .

1. Dann ist  $U$  endlich erzeugt und hat eine Dimension  $\leq n$ .

<sup>10</sup>dimension

2. Jede Basis  $(u_1, \dots, u_m)$  von  $U$  kann zu einer Basis  $(u_1, \dots, u_m, u_{m+1}, \dots, u_n)$  von  $V$  ergänzt werden.

*Beweis.* Übungsaufgabe. □

Wir brauchen noch einige Operationen mit Unterräumen.

Seien  $U$  und  $V$  Unterräume eines Vektorraumes  $W$ , so definieren wir

- $U \cap V$  als Schnittmenge von  $U$  und  $V$ ,
- $U + V := \{w = u + v : u \in U, v \in V\}$ .

Dies sind wieder Unterräume von  $W$ , genannt „Durchschnitt“ und „Summe“.

**Frage:** Wie kann man  $U + V$  beschreiben mithilfe von Basen für  $U$  und für  $V$  ?

**Frage:** Seien  $A$  und  $B$  Mengen mit jeweils endlich vielen Elementen ( $a$  und  $b$  Stück). Wieviele Elemente hat  $A \cup B$  ?

**Satz 2.20 (Dimensionsformel).** *Seien  $U$  und  $V$  Untervektorräume eines endlichdimensionalen Raumes  $W$ , dann gilt*

$$\dim U + \dim V = \dim(U \cap V) + \dim(U + V).$$

*Beweis.* Sei  $\dim U = k$ ,  $\dim V = m$ ,  $\dim(U \cap V) = r$ . Dann müßten wir zeigen, daß  $\dim(U + V) = k + m - r$ .

Wir starten von einer Basis  $(d_1, \dots, d_r)$  von  $U \cap V$ . Gemäß Satz 2.19 ist  $r \leq k$  sowie  $r \leq m$ , und wir können diese Basis ergänzen zu einer Basis von  $U$ ,

$$(d_1, \dots, d_r, u_{r+1}, \dots, u_k),$$

und zu einer Basis von  $V$ ,

$$(d_1, \dots, d_r, v_{r+1}, \dots, v_m).$$

Denn  $U \cap V$  ist Unterraum von  $U$ , und Unterraum von  $V$ .

Wenn wir diese beiden Basen zusammenfügen, erhalten wir die Familie von  $k + m - r$  Vektoren

$$B = (d_1, \dots, d_r, u_{r+1}, \dots, u_k, v_{r+1}, \dots, v_m).$$

Es ist klar, daß die Familie  $B$  die Summe  $U + V$  erzeugt.

Zu beweisen bleibt noch, daß die Familie  $B$  linear unabhängig ist. Wenn wir das gezeigt hätten, dann wäre  $B$  eine Basis von  $U + V$ , und die Dimensionsformel wäre bewiesen.

Sei nun also

$$\sum_{j=1}^r \alpha_j d_j + \sum_{j=r+1}^k \beta_j u_j + \sum_{j=r+1}^m \gamma_j v_j = \vec{0}, \quad (2.2)$$

und zu zeigen ist  $\alpha_j = 0$ ,  $\beta_j = 0$ ,  $\gamma_j = 0$  für jedes  $j$ .

Zu diesem Zwecke setzen wir

$$u := \sum_{j=r+1}^k (-\beta_j u_j) = \sum_{j=1}^r \alpha_j d_j + \sum_{j=r+1}^m \gamma_j v_j.$$

Die linke Summe enthält nur Summanden aus  $U$ , also ist  $u \in U$ . Andererseits enthalten die rechten Summen nur Summanden aus  $V$ , also ist  $u \in V$ , und somit muß  $u \in U \cap V$  sein. Weil  $U \cap V$  als Basis gerade  $(d_1, \dots, d_r)$  hat, haben wir eine Darstellung

$$u = \sum_{j=1}^r \delta_j d_j.$$

Dieser Vektor gehört zu  $U \cap V$ , insbesondere also zu  $V$ . Für den Vektor  $u$  im Raum  $V$  haben wir also zwei Darstellungen: einmal  $\sum_j \delta_j d_j$ , andererseits  $\sum_j \alpha_j d_j + \sum_j \gamma_j v_j$ . Gemäß Satz 2.10 darf es aber nur eine



solche Darstellung geben. Also muß zwangsläufig  $\alpha_j = \delta_j$  sein und  $\gamma_j = 0$  für jedes  $j$ . Mit  $\gamma_j = 0$  gehen wir in (2.2) und erhalten

$$\sum_{j=1}^r \alpha_j d_j + \sum_{j=r+1}^k \beta_j u_j = \vec{0}.$$

Nun ist aber die Familie  $(d_1, \dots, d_r, u_{r+1}, \dots, u_k)$  eine Basis für  $U$ , weshalb die  $\alpha_j = 0$  und  $\beta_j = 0$  sein müssen.

Damit ist bewiesen, daß  $B$  eine linear unabhängige Familie ist.  $\square$

Von besonderer Bedeutung ist der Fall, daß  $U \cap V = \{0\}$  ist. In diesem Fall kann man jeden Vektor  $w \in U + V$  auf genau eine einzige Weise als Summe  $w = u + v$  mit  $u \in U$  und  $v \in V$  schreiben, was die folgende Definition motiviert:

**Definition 2.21.** Seien  $U_1, \dots, U_k$  Untervektorräume eines Vektorraumes  $W$  mit der Eigenschaft, daß jedes Element  $w \in W$  auf genau eine Weise

$$w = u_1 + \dots + u_k, \quad u_i \in U_i, \quad \forall i,$$

dargestellt werden kann. Dann heißt  $W$  die direkte Summe<sup>11</sup> der  $U_i$ , und man schreibt

$$W = U_1 \oplus \dots \oplus U_k = \bigoplus_{j=1}^k U_j.$$

## 2.3 Vektorräume mit Skalarprodukt

**Literatur:** Greiner: *Quantenmechanik. Einführung*. Kapitel XVII: Das formale Schema der Quantenmechanik.

In den bisherigen Untersuchungen haben wir uns allgemeine Vektorräume angeschaut und dabei lediglich ihre beiden Operationen  $+: V \times V \rightarrow V$  und  $\cdot: K \times V \rightarrow V$  verwendet. Jetzt wollen wir Vektorräume studieren, die zusätzlich noch über ein Skalarprodukt verfügen.

### 2.3.1 Skalarprodukte, Normen, Orthogonalsysteme

**Definition 2.22.** Sei  $V$  ein Vektorraum über  $K$ , wobei  $K = \mathbb{R}$  oder  $K = \mathbb{C}$ . Eine Abbildung  $\langle \cdot, \cdot \rangle: V \times V \rightarrow K$  heißt Skalarprodukt, wenn die folgenden Eigenschaften gelten für alle  $u, v, w \in V$  und alle  $\alpha \in K$ :

- $\langle u, v \rangle = \overline{\langle v, u \rangle}$  (hermitesch<sup>12</sup>),
- $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$ ,
- $\langle \alpha u, v \rangle = \alpha \langle u, v \rangle$ ,
- $\langle u, u \rangle \geq 0$ , und  $\langle u, u \rangle = 0$  genau dann wenn  $u = 0$ .

Für  $K = \mathbb{R}$  erhalten wir genau die Eigenschaften aus Satz 1.27.

Aus der ersten Bedingung folgt  $\langle u, u \rangle \in \mathbb{R}$  für jedes  $u \in U$ , auch wenn  $K = \mathbb{C}$  sein sollte. Deshalb ist die vierte Bedingung sinnvoll.

Man beachte, daß ein Skalar im zweiten Faktor nur in konjugierter Form herausgezogen werden kann:

$$\langle u, \alpha v \rangle = \bar{\alpha} \langle u, v \rangle.$$

Standardbeispiele sind für die Fälle von  $V = \mathbb{R}^n$  bzw.  $V = \mathbb{C}^n$

$$\langle x, y \rangle = \sum_{j=1}^n \xi_j \eta_j \quad \text{bzw.} \quad \langle x, y \rangle = \sum_{j=1}^n \xi_j \bar{\eta}_j,$$

es gibt aber noch viele weitere Vektorräume und zugehörige Skalarprodukte.

**Frage:** Welchen Nutzen bringt uns der Konjugationsstrich über dem zweiten Faktor  $\bar{\eta}_j$  ?

<sup>11</sup>direct sum

<sup>12</sup>CHARLES HERMITE, 1822–1901

**Definition 2.23.** Ein Vektorraum über  $\mathbb{R}$  bzw.  $\mathbb{C}$  mit Skalarprodukt heißt euklidischer<sup>13</sup> Vektorraum<sup>14</sup> bzw. unitärer Vektorraum<sup>15</sup>.

Weil  $\langle u, u \rangle$  reell und nichtnegativ ist, können wir die Wurzel ziehen:

**Definition 2.24.** Sei  $V$  ein Vektorraum mit Skalarprodukt  $\langle \cdot, \cdot \rangle$ . Die Abbildung

$$\begin{aligned} |\cdot| : V &\rightarrow \mathbb{R}, \\ |\cdot| : u &\mapsto |u| := \sqrt{\langle u, u \rangle} \end{aligned}$$

heißt Norm von  $u$ . Manchmal schreibt man auch  $\|u\|$  oder  $\|u\|_V$  anstelle von  $|u|$ .

Eine weitere Beziehung zwischen Norm und Skalarprodukt wird gegeben durch die Ungleichung von CAUCHY–SCHWARZ:

**Satz 2.25.** Sei  $V$  ein unitärer oder euklidischer Vektorraum. Dann gilt für alle  $u, v \in V$  die Ungleichung

$$|\langle u, v \rangle| \leq |u| \cdot |v|.$$

Der Beweis verläuft im Prinzip genauso wie im ersten Kapitel, lediglich im Falle eines unitären Vektorraumes sind einige kleinere Änderungen erforderlich.

Für euklidische und unitäre Vektorräume gilt die Parallelogrammgleichung:

$$|u + v|^2 + |u - v|^2 = 2(|u|^2 + |v|^2).$$

**Satz 2.26.** Die Normfunktion hat folgende Eigenschaften, für alle  $u, v \in V$  und alle  $\alpha \in K$ :

1. Es ist  $|u| \geq 0$ , und  $|u| = 0$  genau dann wenn  $u = 0$ .
2. Es ist  $|\alpha u| = |\alpha| \cdot |u|$ .
3. Die Dreiecksungleichung<sup>16</sup> gilt:  $|u + v| \leq |u| + |v|$ .

*Beweis.* 1. Folgt aus der vierten Eigenschaft des Skalarprodukts.

$$2. \text{ Folgt aus } |\alpha u| = \sqrt{\langle \alpha u, \alpha u \rangle} = \sqrt{\alpha \overline{\alpha} \langle u, u \rangle} = \sqrt{|\alpha|^2 \langle u, u \rangle} = |\alpha| \cdot |u|.$$

3. Ergibt sich aus der Ungleichung von Cauchy–Schwarz auf demselben Wege wie im ersten Kapitel. □

**Definition 2.27.** Sei  $V$  ein reeller oder komplexer Vektorraum (nicht notwendig euklidisch oder unitär), und sei  $|\cdot| : V \rightarrow \mathbb{R}$  eine Abbildung, die den Bedingungen aus Satz 2.26 genügt. Dann heißt  $|\cdot|$  eine Norm auf  $V$ , und das Paar  $(V, |\cdot|)$  heißt normierter Raum<sup>17</sup>.

Beispiele für Normen auf dem  $\mathbb{R}^n$  sind

$$|x| := \sum_{j=1}^n |\xi_j|, \quad |x| := \max_{j=1, \dots, n} |\xi_j|, \quad \text{wenn } x = (\xi_1, \dots, \xi_n) \in \mathbb{R}^n.$$

**Frage:** Wie sehen die „Einheitskreise“ zu diesen Normen aus ?

**Definition 2.28.** Sei  $V$  ein unitärer Raum.

1. Zwei Vektoren  $u, v \in V$  heißen orthogonal zueinander bzw. aufeinander senkrecht<sup>18</sup>,  $u \perp v$ , wenn  $\langle u, v \rangle = 0$ .

<sup>13</sup>EUKLID VON ALEXANDRIA, 3. JH. V. CHR.

<sup>14</sup>euclidean (vector) space

<sup>15</sup>unitary space

<sup>16</sup>triangle inequality

<sup>17</sup>normed space

<sup>18</sup>perpendicular to each other

2. Eine Familie  $(u_j : j \in J)$  (mit endlich oder unendlich vielen Elementen) heißt Orthonormalsystem (ONS), wenn

$$\langle u_i, u_j \rangle = \delta_{ij} = \begin{cases} 1 & : i = j, \\ 0 & : i \neq j. \end{cases}$$

3. Eine Basis, die gleichzeitig ein ONS ist, heißt Orthonormalbasis (ONB).  
4. Sei  $U \subset V$  ein Unterraum. Dann heißt

$$U^\perp := \{v \in V : \langle v, u \rangle = 0 \quad \forall u \in U\}$$

orthogonales Komplement von  $U$ .

5. Für zwei Untervektorräume  $U_1$  und  $U_2$  von  $V$  bedeutet die Schreibweise  $U_1 \perp U_2$ , daß jeder Vektor aus  $U_1$  auf jedem Vektor aus  $U_2$  senkrecht steht.

Wegen  $U \cap U^\perp = \{0\}$  haben wir in den meisten Fällen  $V = U \oplus U^\perp$ . Ein Beweis dazu steht unten, falls der Unterraum  $U$  endlichdimensional sein sollte. Damit erklärt sich dann auch die Bezeichnung *Komplement*, denn dies bedeutet hier „Ergänzungsstück“<sup>19</sup>.

Der große Vorteil von Orthonormalbasen ist, daß sie die Rechnungen vereinfachen (im Vergleich zu einer herkömmlichen Basis):

**Satz 2.29.** Sei  $V$  ein euklidischer oder unitärer Raum, und sei  $U = \text{span}(u_1, \dots, u_n)$ , wobei die  $u_j$  eine ONB bilden. Dann hat jeder Vektor  $u \in U$  die Darstellung

$$u = \sum_{j=1}^n \langle u, u_j \rangle u_j,$$

und für die Norm haben wir die Formel

$$|u|^2 = \sum_{j=1}^n |\langle u, u_j \rangle|^2.$$

*Beweis.* Wegen  $u \in \text{span}(u_1, \dots, u_n)$  haben wir eine Darstellung  $u = \sum_{j=1}^n \alpha_j u_j$  mit eindeutig bestimmten  $\alpha_j \in K$ . Wenn wir das Skalarprodukt von  $u$  mit einem Basisvektor bilden, erhalten wir wegen der Linearität des Skalarprodukts und der Orthonormalität

$$\langle u, u_i \rangle = \left\langle \sum_{j=1}^n \alpha_j u_j, u_i \right\rangle = \sum_{j=1}^n \alpha_j \langle u_j, u_i \rangle = \sum_{j=1}^n \alpha_j \delta_{ji} = \alpha_i.$$

Das beweist die erste Formel. Analog erhalten wir

$$|u|^2 = \langle u, u \rangle = \left\langle \sum_{j=1}^n \alpha_j u_j, \sum_{k=1}^n \alpha_k u_k \right\rangle = \sum_{j=1}^n \sum_{k=1}^n \alpha_j \overline{\alpha_k} \langle u_j, u_k \rangle = \sum_{j=1}^n \sum_{k=1}^n \alpha_j \overline{\alpha_k} \delta_{jk} = \sum_{j=1}^n |\alpha_j|^2.$$

Die gemischten Produkte fallen raus wegen der Orthogonalität. □

**Frage:** Im  $\mathbb{R}^{1756}$  sei ein Untervektorraum  $U$  gegeben. Welche Dimension hat  $U^\perp$  ?

**Frage:** Im  $\mathbb{R}^{1517}$  sind zwei Vektoren  $u_1$  und  $u_2$  gegeben, und ein dritter Vektor ist gesucht, der senkrecht auf  $u_1$  und auf  $u_2$  stehen soll. Kann man diesen dritten Vektor mittels Kreuzprodukt ermitteln ?

Der nächste Satz sagt uns, wie wir von einer gewöhnlichen Basis übergehen können zu einer Orthonormalbasis:

<sup>19</sup>aus dem Lateinischen: *complementum* = Ergänzung, Vervollständigung

**Satz 2.30 (Orthonormalisierungsverfahren von Gram–Schmidt<sup>20</sup>).** Sei  $V$  ein euklidischer oder unitärer Raum, und sei  $(v_1, v_2, \dots)$  eine endliche oder unendliche linear unabhängige Familie. Wir definieren dann rekursiv

$$w_1 := v_1, \quad u_1 := \frac{1}{|w_1|} w_1,$$

$$w_{n+1} := v_{n+1} - \sum_{j=1}^n \langle v_{n+1}, u_j \rangle u_j, \quad u_{n+1} := \frac{1}{|w_{n+1}|} w_{n+1}, \quad \forall n \geq 1.$$

Dann gilt:

1. die Vektoren  $w_{n+1}$  sind nie  $\vec{0}$ , also sind die  $u_{n+1}$  definiert,
2. die Familie  $(u_1, u_2, \dots)$  ist linear unabhängig und ein Orthonormalsystem,
3.  $\text{span}(v_1, \dots, v_n) = \text{span}(u_1, \dots, u_n)$  für jedes  $n$ .

*Beweis.* Übungsaufgabe. □

**Frage:** Warum ist kein  $w_{n+1} = 0$  ?

**Frage:** Warum ist  $\langle u_{n+1}, u_k \rangle = 0$  für  $1 \leq k \leq n$  ?

### 2.3.2 Approximationsprobleme

**Literatur:** Greiner: *Klassische Elektrodynamik*. Kapitel I.3.1: Entwicklung beliebiger Funktionen in vollständige Funktionssysteme

Nun wollen wir Lote fällen auf einen Unterraum. Genauer gesagt, geht es um folgendes Approximationsproblem:

Sei  $V$  ein euklidischer oder unitärer Raum, und sei  $U$  ein Unterraum von  $V$ . Gegeben sei ein  $v \in V$ . Gesucht ist dasjenige  $u \in U$ , das von  $v$  den geringsten Abstand hat. Wenn es ein solches  $u \in U$  gibt, dann heißt  $u$  *Projektion von  $v$  auf  $U$*  oder auch *Proximum*.

Wenn  $U$  unendlichdimensional sein sollte (was häufiger vorkommt als man glaubt), dann ist es keineswegs selbstverständlich, daß so ein  $u$  überhaupt existiert, es gibt sogar Gegenbeispiele. Deshalb beschränken wir uns auf den Fall eines endlichdimensionalen  $U$ .

**Satz 2.31.** Sei  $V$  ein euklidischer oder unitärer Raum, und sei  $U = \text{span}(u_1, \dots, u_n)$  ein Unterraum von  $V$ , mit Orthonormalbasis  $(u_1, \dots, u_n)$ . Sei  $v \in V$  ein beliebiger Punkt außerhalb von  $U$ .

1. Dann existiert genau ein  $u \in U$ , sodaß  $v - u$  senkrecht auf  $U$  steht.
2. Dieses  $u$  wird gegeben durch  $u = \sum_{j=1}^n \langle v, u_j \rangle u_j$ .
3. Für jedes andere  $u' \in U$  gilt  $|v - u'| > |v - u|$ .
4. Für den Projektionsfehler  $|v - u|$  gilt  $|v - u|^2 = |v|^2 - |u|^2$ .

*Beweis.* 1. Für  $u$  machen wir den Ansatz  $u = \sum_{j=1}^n \alpha_j u_j$ . Der Vektor  $v - u$  steht senkrecht auf  $U$  genau dann, wenn er senkrecht auf jedem Basisvektor  $u_j$  steht. Nun ist<sup>21</sup>

$$0 \stackrel{!}{=} \langle v - u, u_j \rangle = \langle v, u_j \rangle - \left\langle \sum_{k=1}^n \alpha_k u_k, u_j \right\rangle = \langle v, u_j \rangle - \alpha_j,$$

also gibt es genau ein solches  $u \in U$ .

2. Ist soeben bewiesen worden.

<sup>20</sup> JØRGAN PEDERSEN GRAM (1850–1916) und ERHARD SCHMIDT (1876–1959)

<sup>21</sup>Das Ausrufezeichen über dem = soll bedeuten, daß wir uns wünschen, daß die genannte Gleichung gelten möge.

3. Wenn  $u' \in U$ , dann ist wegen  $u \in U$  auch  $u - u' \in U$ . Nun steht  $v - u$  senkrecht auf  $U$ , also ist insbesondere  $\langle v - u, u - u' \rangle = 0$ . Wegen  $u' \neq u$  und des Satzes von Pythagoras haben wir dann

$$|v - u'|^2 = |(v - u) + (u - u')|^2 = |v - u|^2 + |u - u'|^2 > |v - u|^2.$$

4. Wir können schreiben  $v = (v - u) + u$ . Weil  $v - u$  und  $u$  senkrecht aufeinander stehen (wegen 1.), gilt aufgrund des Satzes von Pythagoras

$$|v|^2 = |v - u|^2 + |u|^2,$$

woraus die Behauptung sofort folgt. □

Jetzt haben wir alle Werkzeuge beisammen, um den Beweis von  $V = U \oplus U^\perp$  nachzuholen:

**Satz 2.32.** Sei  $V$  ein unitärer oder euklidischer Raum, und sei  $U$  ein endlichdimensionaler Unterraum von  $V$ . Dann ist  $V = U \oplus U^\perp$ .

*Beweis.* Sei  $v \in V$  beliebig. Dann gibt es wegen Satz 2.31 genau ein  $u \in U$  mit  $v - u \perp U$ . Wir können also jeden Vektor  $v \in V$  in zwei Summanden  $v - u \in U^\perp$  und  $u \in U$  zerlegen, und es gibt nur diese eine Zerlegung. Also ist  $V = U \oplus U^\perp$ . □

**Beispiel 2.33.** Sei  $V = C([-1, 1] \rightarrow \mathbb{R})$  der Raum der auf dem Intervall  $[-1, 1]$  definierten stetigen und reellwertigen Funktionen, ausgestattet mit dem Skalarprodukt

$$\langle f, g \rangle = \int_{-1}^1 f(x)g(x)dx.$$

Dann ist  $|f| = \sqrt{\int_{-1}^1 |f(x)|^2 dx}$ . Wir wollen eine  $\cos$ -Funktion durch eine quadratische Parabel annähern. Sei also  $U = \text{span}(1, x^2)$  und  $v = v(x) = \cos(\frac{\pi}{2}x)$ . Gesucht ist ein  $u \in U$  mit  $|v - u| \rightarrow \min$ .

**Lösung:** Wir ermitteln mithilfe des GRAM-SCHMIDT-Verfahrens eine Orthonormalbasis  $(u_1, u_2)$  für  $U$ :

$$v_1 = v_1(x) = 1, \quad v_2 = v_2(x) = x^2,$$

$$u_1 = u_1(x) = \frac{1}{\sqrt{\int_{-1}^1 1^2 dx}} \cdot 1 = \frac{1}{\sqrt{2}},$$

$$w_2 = w_2(x) = x^2 - \left\langle x^2, \frac{1}{\sqrt{2}} \right\rangle \cdot \frac{1}{\sqrt{2}} = x^2 - \frac{1}{3},$$

$$|w_2| = \sqrt{\int_{-1}^1 \left(x^2 - \frac{1}{3}\right)^2 dx} = \sqrt{\frac{8}{45}},$$

$$u_2 = u_2(x) = \sqrt{\frac{45}{8}} \left(x^2 - \frac{1}{3}\right).$$

Dann ermittelt sich das Proximum  $u = u(x)$  wegen Satz 2.31, Teil 2, nach der Formel

$$u = u(x) = \langle v, u_1 \rangle u_1(x) + \langle v, u_2 \rangle u_2(x),$$

und für die Norm haben wir (wegen Satz 2.29) die einfache Darstellung

$$|u| = \sqrt{|\langle v, u_1 \rangle|^2 + |\langle v, u_2 \rangle|^2}.$$

Die Koordinaten von  $u$  in Bezug auf die Basis  $(u_1, u_2)$  sind dann

$$\langle v, u_1 \rangle = \int_{-1}^1 \cos\left(\frac{\pi}{2}x\right) \cdot \frac{1}{\sqrt{2}} dx = \frac{2\sqrt{2}}{\pi}$$

und

$$\langle v, u_2 \rangle = \sqrt{\frac{45}{8}} \int_{-1}^1 \cos\left(\frac{\pi}{2}x\right) \cdot \left(x^2 - \frac{1}{3}\right) dx.$$

Im BRONSTEIN finden wir die Formel

$$\int x^2 \cos(\alpha x) dx = \frac{2x}{\alpha^2} \cos(\alpha x) + \left( \frac{x^2}{\alpha} - \frac{2}{\alpha^3} \right) \sin(\alpha x) + \text{const.},$$

mit deren Hilfe wir ausrechnen können, daß

$$\langle v, u_2 \rangle = -\frac{\sqrt{10}}{\pi} + \sqrt{\frac{45}{2}} \left( \frac{2}{\pi} \right)^3 \left( \left( \frac{\pi}{2} \right)^2 - 2 \right).$$

Damit sind die Koordinaten von  $u$  bezüglich der Orthonormalbasis  $(u_1, u_2)$  von  $U$  bestimmt, und das gesuchte quadratische Polynom ist gefunden.

Zur Berechnung des Projektionsfehlers  $|v - u|$  benutzen wir die Formel  $|v - u|^2 = |v|^2 - |u|^2$ , siehe auch Satz 2.31, Teil 4. Wir haben

$$|v|^2 = \int_{-1}^1 \left( \cos\left(\frac{\pi}{2}x\right) \right)^2 dx.$$

Aus dem BRONSTEIN entnehmen wir die Formel

$$\int (\cos(\alpha x))^2 dx = \frac{x}{2} + \frac{1}{4\alpha} \sin(2\alpha x) + \text{const.},$$

die uns auf  $|v|^2 = 1$  führt. Schließlich ist nach einiger Rechnerei

$$|u|^2 = |\langle v, u_1 \rangle|^2 + |\langle v, u_2 \rangle|^2 = 0.810569468 + 0.18883446738 = 0.9994039,$$

woraus  $|v - u|^2 = 0.000596$  und  $|v - u| = 0.0244$  folgen. Wir haben also einen relativen Approximationsfehler  $\frac{|v-u|}{|v|}$  von weniger als 2.5%.

**Bemerkung 2.34.** Der Projektion  $v \mapsto u$  gemäß  $u = \sum_{j=1}^n \langle v, u_j \rangle u_j$  werden wir im 2. Semester bei der Fourierreihendarstellung von periodischen Funktionen wieder begegnen.

## 2.4 Ausblick: Vektorräume in der Physik 1

Wir bringen einige weitere Beispiele.

**Räume für Ortsvariablen:** das ist praktisch immer der  $\mathbb{R}^3$ , die Elemente davon beziehen sich auf Orte (oder sind Verbindungsvektoren zweier Orte). Die Maßeinheit ist zu interpretieren als „Meter“. (Das Gegenstück, die *Wellenzahlvektoren*, behandeln wir absichtlich hier nicht, sondern erst im übernächsten Kapitel.) Dieser Raum hat die Dimension 3.

**Räume für Vektorfelder:** an jeden Punkt  $x$  des  $\mathbb{R}^3$  heften wir einen Vektor  $\vec{u}(x)$  an, den wir als *Flußvektor* interpretieren. Man denke z.B. an einen strömendes Fluid oder an das elektrische Feld (was im Prinzip dasselbe ist). Die Überlagerung verschiedener Strömungen ergibt eine Addition in einem Vektorraum. Wenn wir annehmen, daß für  $|x| \rightarrow \infty$  die Strömung abklingt, dann ist es sinnvoll zu verlangen, daß

$$\int_{x \in \mathbb{R}^3} |\vec{u}(x)|^2 dx < \infty,$$

wobei  $|\vec{u}(x)|^2 = (u_1(x))^2 + (u_2(x))^2 + (u_3(x))^2$ . Integrale dieses Typs können häufig als Energie interpretiert werden, womit sie für die Physik automatisch interessant sind. Den Vektorraum aller solcher Vektorfelder schreiben wir als  $L^2(\mathbb{R}^3 \rightarrow \mathbb{R}^3)$ , er hat die Dimension unendlich und ein Skalarprodukt gemäß

$$\langle \vec{u}, \vec{v} \rangle_{L^2} := \int_{x \in \mathbb{R}^3} u_1(x)v_1(x) + u_2(x)v_2(x) + u_3(x)v_3(x) dx. \quad (2.3)$$

**Vektorräume in der Quantenmechanik:** wir betrachten ein einzelnes spinloses Teilchen in einem Potential. Der Zustand dieses Teilchens wird quantenmechanisch beschrieben durch eine Wellenfunktion  $\psi = \psi(x)$ , die vom Ortsraum  $\mathbb{R}^3$  nach  $\mathbb{C}$  abbildet. Wir interpretieren das Betragsquadrat  $|\psi(x)|^2$  als Wahrscheinlichkeitsdichte, das Teilchen an Position  $x$  anzutreffen. Sinnvollerweise ist

$$\int_{x \in \mathbb{R}^3} |\psi(x)|^2 dx < \infty.$$

Alle solchen Wellenfunktionen bilden einen Vektorraum, nämlich den  $L^2(\mathbb{R}^3 \rightarrow \mathbb{C})$ . Dieser ist unendlichdimensional, und er hat folgendes Skalarprodukt:

$$\langle \varphi, \psi \rangle_{L^2} := \int_{x \in \mathbb{R}^3} \varphi(x) \overline{\psi(x)} dx. \quad (2.4)$$

Für  $M$  Teilchen jeweils aus dem  $\mathbb{R}^3$  liegt die Ortsvariable  $x$  im  $\mathbb{R}^{3M}$ . Bei realistischen Werten von  $M$  ist dies meist keine handhabbare Situation mehr, aber mittels Gruppentheorie (Vertauschungen der Teilchen untereinander sind Elemente einer Permutationsgruppe!) kann man den Aufwand zur Beschreibung des Systems um einiges reduzieren.

**Literatur:** Greiner und Müller: *Quantenmechanik. Symmetrien*. Kapitel IX: Darstellungen der Permutationsgruppe und Young-Tableaux.

## 2.5 Ausblick: die Helmholtz-Projektion

Skalare Felder sind Funktionen  $\varphi$  von  $\mathbb{R}^3$  nach  $\mathbb{R}$ .

Vektorfelder sind Funktionen  $\vec{u}$  von  $\mathbb{R}^3$  nach  $\mathbb{R}^3$ .

Heute wird der  $\mathbb{R}^3$  jedesmal ausgestattet mit einem kartesischen Koordinatensystem (keine Polarkoordinaten oder ähnliches).

Der *Gradient* eines skalaren Feldes ist

$$\nabla \varphi(x) = \left( \frac{\partial \varphi}{\partial x_1}, \frac{\partial \varphi}{\partial x_2}, \frac{\partial \varphi}{\partial x_3} \right).$$

Die *Divergenz* eines Vektorfeldes  $\vec{u}$  ist

$$\operatorname{div} \vec{u}(x) = \frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3}.$$

Ein berühmter Satz von Helmholtz<sup>22</sup> besagt: jedes Vektorfeld  $\vec{u} \in L^2(\mathbb{R}^3 \rightarrow \mathbb{R}^3)$  kann auf eindeutige Weise zerlegt werden als  $\vec{u} = \vec{v} + \vec{w}$ , wobei  $\vec{v}, \vec{w} \in L^2(\mathbb{R}^3 \rightarrow \mathbb{R}^3)$  mit

$$\begin{aligned} \operatorname{div} \vec{v}(x) &= 0, & \forall x \in \mathbb{R}^3, \\ \vec{w}(x) &= \nabla \varphi(x), & \forall x \in \mathbb{R}^3, \end{aligned}$$

mit einem geeigneten Skalarfeld  $\varphi$ . Man sagt auch:  $\vec{v}$  ist divergenzfrei bzw. quellenfrei,  $\vec{w}$  ist ein Gradientenfeld bzw. hat ein Potential. Wir haben also die Helmholtz-Zerlegung

$$L^2(\mathbb{R}^3 \rightarrow \mathbb{R}^3) = L^2_{\operatorname{div}}(\mathbb{R}^3 \rightarrow \mathbb{R}^3) \oplus L^2_{\operatorname{pot}}(\mathbb{R}^3 \rightarrow \mathbb{R}^3).$$

Es kommt noch schöner: beide Summanden auf der rechten Seite bilden nicht nur eine direkte Summe, sondern sie stehen auch senkrecht aufeinander im Sinne des Skalarprodukts (2.3), wie wir im 2. Semester erkennen werden.

Es ist  $L^2_{\operatorname{div}}(\mathbb{R}^3 \rightarrow \mathbb{R}^3)$  ein (abgeschlossener) Untervektorraum des Vektorraums  $L^2(\mathbb{R}^3 \rightarrow \mathbb{R}^3)$ . Wenn wir von einem  $\vec{u} \in L^2(\mathbb{R}^3 \rightarrow \mathbb{R}^3)$  das Lot fällen auf  $L^2_{\operatorname{div}}(\mathbb{R}^3 \rightarrow \mathbb{R}^3)$ , dann lösen wir genau eine Approximationsaufgabe im Sinne von Abschnitt 2.3.2. Diese Abbildung heißt *Helmholtz-Projektion*.

<sup>22</sup> HERMANN VON HELMHOLTZ (1821–1894), Mediziner und Physiker, Professor für Physiologie und Pathologie in Königsberg, später Professor für Anatomie und Physiologie in Bonn. Formulierte den Energieerhaltungssatz, erfand den Augenspiegel, maß die Fortpflanzungsgeschwindigkeit von Nervenregungen, lieferte wichtige Beiträge zur Hydrodynamik und Elektrodynamik, begründete die wissenschaftliche Meteorologie.

**Anwendungen** gibt es dafür in der Elektrodynamik (denn elektrische und magnetische Felder  $\vec{E}$ ,  $\vec{B}$  gehören gerade zu  $L^2(\mathbb{R}^3 \rightarrow \mathbb{R}^3)$ ) und in der Festkörperphysik. Man stelle sich einen Festkörper vor, der elastische Schwingungen ausführt. Diese können longitudinal oder transversal verlaufen, und die dazugehörigen Verschiebungsvektorfelder sind gerade Potentialfelder bzw. divergenzfreie Felder. Die Helmholtz-Projektion erlaubt es, diese beiden Schwingungstypen getrennt zu betrachten, und man erhält z.B., daß longitudinale bzw. transversale Wellen unterschiedlich schnell laufen können.

## 2.6 Schlüsselbegriffe

- Definition eines Vektorraumes,
- Funktionenräume als Vektorräume,
- lineare Unabhängigkeit, Basis, Dimension,
- direkte Summe von Vektorräumen, Dimensionsformel,
- Definition von Skalarprodukt (im euklidischen Fall und im unitären Fall) und Norm,
- Verfahren von Gram-Schmidt,
- Lösungsverfahren zu Approximationsproblemen — auch im abstrakten Fall.



# Kapitel 3

## Matrizen

### 3.1 Operationen mit Matrizen

Wie immer, sei  $K$  auch jetzt ein Körper, zum Beispiel  $K = \mathbb{R}$  oder  $K = \mathbb{C}$ . Unter  $K^n$ ,  $K^m$  usw. verstehen wir den Vektorraum der Spaltenvektoren mit  $n$  bzw.  $m$  Einträgen. In diesem Kapitel soll es um Matrizen gehen, also Zahlenschemata der Form

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix}, \quad a_{ij} \in K.$$

Die Menge aller solchen Matrizen bezeichnen wir mit  $K^{n \times m}$ . Für  $m = 1$  erhalten wir die Spaltenvektoren

**Definition 3.1.** Für Matrizen  $A, B \in K^{n \times m}$  und Zahlen  $\alpha \in K$  definieren wir eine Addition zweier Matrizen sowie eine Multiplikation einer Zahl mit einer Matrix gemäß

$$\begin{aligned} (A + B)_{ij} &:= a_{ij} + b_{ij}, & 1 \leq i \leq n, & \quad 1 \leq j \leq m, \\ (\alpha A)_{ij} &:= \alpha a_{ij}, & 1 \leq i \leq n, & \quad 1 \leq j \leq m. \end{aligned}$$

Die Menge der  $n \times m$ -Matrizen bildet mit diesen beiden Operationen einen Vektorraum über  $K$ .

**Frage:** Was ist das Nullelement dieses Vektorraumes? Welche Dimension hat er?

Genauso wie im ersten Kapitel definieren wir eine Multiplikation von Matrizen und Vektoren:

**Definition 3.2.** Für eine Matrix  $A \in K^{n \times m}$  und einen Vektor  $x = (\xi_1, \dots, \xi_m)^\top \in K^m$  definieren wir das Produkt  $Ax \in K^n$  durch

$$Ax = \begin{pmatrix} a_{11} & \dots & a_{1m} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nm} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_m \end{pmatrix} := \begin{pmatrix} \sum_{j=1}^m a_{1j} \xi_j \\ \vdots \\ \sum_{j=1}^m a_{nj} \xi_j \end{pmatrix}.$$

Der Vektor  $x$  muß genau so viele Komponenten haben wie  $A$  Spalten hat. Das Ergebnis  $Ax$  hat so viele Komponenten wie  $A$  Zeilen hat.

**Satz 3.3.** Sei  $A \in K^{n \times m}$ . Dann erzeugt  $A$  eine Abbildung  $f_A$ ,

$$\begin{aligned} f_A: K^m &\rightarrow K^n, \\ f_A: x &\mapsto f_A(x) := Ax. \end{aligned}$$

Diese Abbildung ist linear (bzw. ein Homomorphismus), das heißt:

$$\begin{aligned} f_A(x + y) &= f_A(x) + f_A(y), & x, y &\in K^m, \\ f_A(\alpha x) &= \alpha f_A(x), & \alpha \in K, & \quad x \in K^m. \end{aligned}$$

*Beweis.* Übungsaufgabe. □

Später werden wir sehen, daß *jede* lineare Abbildung von  $K^m$  auf  $K^n$  in Form einer Matrixmultiplikation dargestellt werden kann.

Wir stellen die Matrix als nebeneinandergestellte Spaltenvektoren dar:

$$A = \left( \begin{array}{c|c|c|c} | & | & \cdots & | \\ a_1 & a_2 & & a_m \\ | & | & & | \end{array} \right).$$

Dann sieht man schnell, daß  $a_j = Ae_j$ , das heißt,

*die Spalten der Abbildungsmatrix sind die Koordinaten der Bilder der Einheitsvektoren.*

Daraus ergibt sich für einen Vektor  $x = (\xi_1, \dots, \xi_m)^\top = \sum_{j=1}^m \xi_j e_j$  aufgrund der Linearität:

$$f_A(x) = Ax = A \left( \sum_{j=1}^m \xi_j e_j \right) = \sum_{j=1}^m \xi_j Ae_j = \sum_{j=1}^m \xi_j a_j.$$

Das heißt,  $Ax$  ist nichts anderes als eine Linearkombination der Spalten von  $A$ . Die Komponenten von  $x$  sind gerade die Koeffizienten dieser Linearkombination.

Wir verallgemeinern die Matrix–Vektor–Multiplikation zu einer Matrix–Matrix–Multiplikation:

**Definition 3.4.** Sei  $A \in K^{n \times m}$  und  $B \in K^{m \times l}$ . Dann definieren wir das Produkt  $AB \in K^{n \times l}$  durch

$$(AB)_{ij} := \sum_{k=1}^m a_{ik} b_{kj}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq l.$$

Man beachte, daß das Produkt nur dann definiert ist, wenn  $B$  soviele Zeilen hat wie  $A$  Spalten hat.

Wenn  $B$  nur eine Spalte hat ( $l = 1$ ), dann erhalten wir die zuvor schon definierte Matrix–Vektor–Multiplikation.

Durch Hinschauen erkennen wir:

*die  $k$ -te Spalte von  $AB$  ist gleich  $A \cdot (k$ -te Spalte von  $B$ )*

Diese Multiplikation unterscheidet sich in mindestens zwei Punkten von der herkömmlichen Multiplikation reeller oder komplexer Zahlen: sie ist im Allgemeinen nicht kommutativ (also  $AB \neq BA$ ); und Nullteiler sind möglich. Das heißt, ein Produkt kann gleich der Nullmatrix sein, obwohl keiner seiner Faktoren gleich der Nullmatrix ist. Ein Beispiel dafür ist

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

**Satz 3.5.** Seien  $A \in K^{n \times m}$ ,  $B \in K^{m \times l}$ , und seien  $f_A: K^m \rightarrow K^n$  sowie  $f_B: K^l \rightarrow K^m$  die dazugehörigen Abbildungen. Sei weiterhin  $f_{AB}: K^l \rightarrow K^n$  die der Produktmatrix  $AB \in K^{n \times l}$  zugeordnete Abbildung. Dann ist

$$f_{AB} = f_A \circ f_B,$$

das heißt  $f_{AB}$  ist die Nacheinanderausführung der Abbildungen  $f_B$  und  $f_A$ .

*Beweis.* Es seien  $b_1, \dots, b_l$  die Spalten von  $B$ ; und sei  $x = (\xi_1, \dots, \xi_l)^\top = \sum_{j=1}^l \xi_j e_j$  ein beliebiger Vektor. Dann ist (wenn wir an die beiden obigen Merkgeregeln erinnern)

$$\begin{aligned} (f_A \circ f_B)(x) &= f_A(f_B(x)) = f_A(Bx) = f_A \left( \sum_{k=1}^l \xi_k b_k \right) = A \left( \sum_{k=1}^l \xi_k b_k \right) \\ &= \sum_{k=1}^l \xi_k Ab_k = \sum_{k=1}^l \xi_k (AB)e_k = (AB) \sum_{k=1}^l \xi_k e_k = (AB)x = f_{AB}(x). \end{aligned}$$

Das ist genau die gewünschte Aussage. □

Damit können wir jetzt zeigen, daß die Matrixmultiplikation assoziativ ist:

**Satz 3.6.** Seien  $A \in K^{n \times m}$ ,  $B \in K^{m \times l}$  und  $C \in K^{l \times r}$ . Dann gilt  $(AB)C = A(BC)$ .

*Beweis.* Wir beachten die Regel  $(PQ)_{ij} = \sum_k p_{ik}q_{kj}$  für kompatible Matrizen  $P$  und  $Q$ . Dann ergibt sich für die linke Seite gerade

$$\left((AB)C\right)_{ij} = \sum_{k=1}^l (AB)_{ik}c_{kj} = \sum_{k=1}^l \left(\sum_{s=1}^m a_{is}b_{sk}\right)c_{kj} = \sum_{k=1}^l \sum_{s=1}^m a_{is}b_{sk}c_{kj},$$

und die rechte Seite ist gleich

$$\left(A(BC)\right)_{ij} = \sum_{u=1}^m a_{iu}(BC)_{uj} = \sum_{u=1}^m a_{iu} \left(\sum_{v=1}^l b_{uv}c_{vj}\right) = \sum_{u=1}^m \sum_{v=1}^l a_{iu}b_{uv}c_{vj},$$

und wenn man hier  $u$  zu einem  $s$  umbauft und  $v$  zu einem  $k$ , dann erhält man genau den Ausdruck der anderen Seite. Die Summationszeichen dürfen getauscht werden wegen der Kommutativität der Addition in  $K$ .  $\square$

**Definition 3.7.** Sei  $A \in K^{n \times m}$  eine Matrix mit Einträgen  $a_{ij}$ . Dann definieren wir eine transponierte Matrix<sup>1</sup>  $A^\top \in K^{m \times n}$  und eine adjungierte Matrix<sup>2</sup>  $A^* \in K^{m \times n}$  durch

$$(A^\top)_{ij} = a_{ji}, \quad (A^*)_{ij} = \overline{a_{ji}}.$$

Wenn  $K = \mathbb{R}$ , dann sind die Matrizen  $A^\top$  und  $A^*$  identisch.

**Satz 3.8.** Wenn die Formate der Matrizen  $A$  und  $B$  zueinander passen, dann ist

$$(AB)^\top = B^\top \cdot A^\top, \quad (AB)^* = B^* \cdot A^*.$$

Die Standardskalarprodukte können geschrieben werden als

$$\langle x, y \rangle_{\mathbb{R}} = x^\top y, \quad \langle x, y \rangle_{\mathbb{C}} = \overline{x^\top y} = y^* x.$$

*Beweis.* Übungsaufgabe.  $\square$

## 3.2 Gleichungssysteme

Wir versuchen jetzt, die Matrix-Multiplikation umzukehren, also eine „Division“ zu finden. Konkret geht es um Folgendes:

**Problem:** Gegeben sind zwei Matrizen  $A \in K^{n \times m}$ ,  $B \in K^{n \times l}$ . Gesucht ist eine Matrix  $X \in K^{m \times l}$  mit  $AX = B$ .

Es stellen sich dabei fast von selbst einige Fragen:

- Gibt es überhaupt Lösungen  $X$  ?
- Falls die Lösbarkeit von  $A$  und  $B$  abhängen sollte: welche Bedingungen müssen  $A$  und  $B$  erfüllen, damit es Lösungen  $X$  gibt ?
- Wieviele Lösungen  $X$  gibt es ?
- Wie findet man diese Lösungen ?

Der Fall  $l = 1$  ist noch am einfachsten: dann ist  $B$  ein Spaltenvektor  $B = b = (\beta_1, \dots, \beta_n)^\top$ , und  $X$  ist ein Spaltenvektor  $X = x = (\xi_1, \dots, \xi_m)^\top$ . Wir erhalten in diesem Fall ein lineares Gleichungssystem

$$\begin{aligned} a_{11}\xi_1 + \dots + a_{1m}\xi_m &= \beta_1, \\ &\vdots \\ a_{n1}\xi_1 + \dots + a_{nm}\xi_m &= \beta_n. \end{aligned}$$

Dieses System wird noch etwas einfacher im Fall  $n = m$ :

<sup>1</sup>transposed matrix

<sup>2</sup>adjoint matrix

**Definition 3.9.** Eine Matrix  $A \in K^{n \times n}$  heißt invertierbar<sup>3</sup>, wenn es eine Matrix  $B \in K^{n \times n}$  gibt mit

$$BA = I_n,$$

wobei  $I_n$  die  $n \times n$ -Einheitsmatrix bezeichnet: Man schreibt  $B = A^{-1}$ , und nennt  $A^{-1}$  die zu  $A$  inverse<sup>4</sup> Matrix. Wenn  $A$  invertierbar bzw. nicht invertierbar ist, dann heißt  $A$  auch regulär<sup>5</sup> bzw. singular<sup>6</sup>.

Wenn  $A$  invertierbar ist, dann wird eine Lösung  $x$  des Gleichungssystems  $Ax = b$  gegeben durch  $x = A^{-1}b$ .

**Frage:** Kann es sein, daß das Gleichungssystem  $Ax = b$  bei invertierbarem  $A$  mehrere Lösungen  $x$  hat? Begründung?

**Satz 3.10.** Es seien  $A, B \in K^{n \times n}$ . Dann gilt:

1. Wenn  $A$  invertierbar ist, so gibt es genau eine inverse Matrix.
2. Wenn  $A$  invertierbar ist und  $AB = I$ , so ist  $B = A^{-1}$ .
3. Wenn  $A$  invertierbar ist, so ist auch  $A^{-1}$  invertierbar, und  $(A^{-1})^{-1} = A$ .
4. Wenn  $A$  invertierbar ist, so ist auch  $A^\top$  invertierbar, und  $(A^\top)^{-1} = (A^{-1})^\top$ .
5. Wenn  $A$  und  $B$  beide invertierbar sind, so ist auch  $AB$  invertierbar, und es ist  $(AB)^{-1} = B^{-1}A^{-1}$ .

*Beweis.* 1. Folgt direkt aus Satz 1.44.

2. Folgt aus Satz 1.44.

3. Folgt aus Satz 1.44.

4. Wir haben  $AA^{-1} = I$ . Wenn wir darauf Satz 3.8 anwenden, haben wir  $(A^{-1})^\top A^\top = I^\top = I$ , also ist  $(A^{-1})^\top$  die Inverse zu  $A^\top$ .

5. Folgt direkt aus Lemma 1.48. □

Gleichungssysteme der Form  $AX = B$  behandeln wir, indem wir versuchen, die Matrix  $A$  auf eine einfachere Form zu bringen. Dabei benutzen wir folgenden Satz.

**Satz 3.11.** Seien  $A \in K^{n \times m}$ ,  $B \in K^{n \times l}$ ,  $X \in K^{m \times l}$  und  $C \in K^{n \times n}$ , wobei  $C$  regulär ist. Dann ist  $X$  eine Lösung von  $AX = B$  genau dann, wenn  $X$  eine Lösung von  $CAX = CB$  ist.

*Beweis.* Sei  $AX = B$ . Dann ist auch  $CAX = CB$ .

Sei andererseits  $CAX = CB$ . Weil  $C$  invertierbar ist, existiert  $C^{-1}$ , und es ist  $C^{-1}CAX = C^{-1}CB$ , also  $AX = B$ . □

Wir wollen also eine Matrix  $C$  finden, sodaß  $CA$  einfacher zu behandeln ist als  $A$ . Wir wählen  $C$  als Produkt von sogenannten Eliminationsmatrizen und Permutationsmatrizen, und die angestrebte Form von  $CA$  ist die sogenannte GAUSS–JORDAN–Form bzw. Zeilenstufenform.

Ein Beispiel einer Matrix in Zeilenstufenform, mit beliebigen Einträgen an dem mit  $\times$  markierten Stellen, ist

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

---

<sup>3</sup>invertible

<sup>4</sup>inverse

<sup>5</sup>regular

<sup>6</sup>singular





Hierbei stehen die  $\times$  für beliebige Einträge.

Die Anzahl der Einsen in den Stufenknicken heißt auch *Rang* der Matrix.

Exakt formuliert:

**Definition 3.12.** Eine Matrix  $D \in K^{n \times m}$  ist in Zeilenstufenform, wenn es ein  $0 \leq r \leq n$  und Zahlen  $1 \leq j_1 < j_2 < \dots < j_r \leq m$  gibt mit folgenden Eigenschaften:

- $d_{ij_i} = 1$  für  $1 \leq i \leq r$ ,
- $d_{il} = 0$  für  $1 \leq l < j_i$ ,
- $d_{il} \in K$  beliebig für  $j_i < l \leq m$ ,
- $d_{kl} = 0$  für  $r < k \leq n$  und  $1 \leq l \leq m$ .

Im oben gezeichneten Beispiel ist  $r = 5$  und  $(j_1, j_2, \dots, j_5) = (2, 4, 7, 8, 9)$ .

Die zweiten bis vierten Bedingungen der Definition bedeuten:

- links von den Einsen, aber in derselben Zeile, stehen Nullen,
- rechts von den Einsen, aber in derselben Zeile, stehen  $\times$ ,
- ab der Zeile  $r + 1$  stehen nur noch Nullen.

Das Gauß–Jordan<sup>7</sup>–Verfahren verläuft wie folgt:

1. Man setze einen Spaltenzähler  $s$  und einen Zeilenzähler  $z$  jeweils auf 1.
2. Man suche in der Spalte  $s$  unterhalb des  $z$ -ten Eintrags von oben (einschließlich) nach einem Matrixeintrag  $\neq 0$ .
3. (a) Wenn es einen solchen nicht gibt, dann erhöhe man den Spaltenzähler  $s$  um 1 und gehe zu Schritt 2.  
(b) Wenn es einen Matrixeintrag  $\neq 0$  gibt<sup>8</sup>, dann tausche man die dazugehörige Langzeile mit der  $z$ -ten Langzeile. Jetzt ist der Eintrag  $a_{zs} \neq 0$ .
4. Man dividiere die  $z$ -te Langzeile durch  $a_{zs}$ . Jetzt ist der Eintrag  $a_{zs} = 1$ .
5. Man addiere geeignete Vielfache der  $z$ -ten Langzeile zu den tieferen Langzeilen, um die Einträge unterhalb von  $a_{zs}$  zu annullieren („Ausräumen“). Schräg links unterhalb von  $a_{zs}$  und links von  $a_{zs}$  passiert nichts, weil dort sowieso nur Nullen stehen.
6. Wenn unterhalb von  $a_{zs}$  die Nullen erzeugt worden sind, erhöhe man  $z$  und  $s$  jeweils um 1.
7. (a) Wenn diese Indizes  $z$  und  $s$  den Bereich der Matrix  $A$  verlassen, ist das Verfahren beendet.  
(b) Ansonsten gehe man zu Schritt 2.

Auf diesem Wege kommt man immer zu einer Gauß–Jordan–Form. Man kann das Verfahren aber variieren. Zum Beispiel kann man den Schritt 4 auch später ausführen. Man kann auch oberhalb der  $a_{zs}$  ausräumen.

Als Beispiel wollen wir die Inverse von

$$A = \begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix}$$

bestimmen, also das System  $AX = I_3$  lösen. Wir starten also mit

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

<sup>7</sup>WILHELM JORDAN (1842–1899), nicht zu verwechseln mit MARIE ENNEMOND CAMILLE JORDAN (1838–1922), französischer Mathematiker, dem Schöpfer der JORDANSchen Normalform. Und die JORDAN–Algebren der Physik stammen von PASCUAL JORDAN (1902–1980).

<sup>8</sup>Man nennt ihn auch Pivotelement.

Wir sehen, daß wir unterhalb der 3 links oben ausräumen können. Weil wir Bruchrechnung vermeiden wollen aus Bequemlichkeit, multiplizieren wir stattdessen die 2. und 3. Langzeile mit passenden Faktoren:

$$\left( \begin{array}{ccc|ccc} 3 & 1 & 1 & 1 & 0 & 0 \\ -3 & -9 & -3 & 0 & -3 & 0 \\ -3 & -3 & -9 & 0 & 0 & -3 \end{array} \right) \rightarrow \left( \begin{array}{ccc|ccc} 3 & 1 & 1 & 1 & 0 & 0 \\ 0 & -8 & -2 & 1 & -3 & 0 \\ 0 & -2 & -8 & 1 & 0 & -3 \end{array} \right).$$

Ausgehend von der  $-8$  an der Position  $(2,2)$  räumen wir nach oben und unten aus:

$$\left( \begin{array}{ccc|ccc} 24 & 8 & 8 & 8 & 0 & 0 \\ 0 & -8 & -2 & 1 & -3 & 0 \\ 0 & 8 & 32 & -4 & 0 & 12 \end{array} \right) \rightarrow \left( \begin{array}{ccc|ccc} 24 & 0 & 6 & 9 & -3 & 0 \\ 0 & -8 & -2 & 1 & -3 & 0 \\ 0 & 0 & 30 & -3 & -3 & 12 \end{array} \right)$$

Und jetzt starten wir von der 30 an der Position  $(3,3)$  und räumen nach oben aus:

$$\left( \begin{array}{ccc|ccc} -120 & 0 & -30 & -45 & 15 & 0 \\ 0 & -120 & -30 & 15 & -45 & 0 \\ 0 & 0 & 30 & -3 & -3 & 12 \end{array} \right) \rightarrow \left( \begin{array}{ccc|ccc} -120 & 0 & 0 & -48 & 12 & 12 \\ 0 & -120 & 0 & 12 & -48 & 12 \\ 0 & 0 & 30 & -3 & -3 & 12 \end{array} \right).$$

Anschließend kürzen wir die Langzeilen durch  $-120$ ,  $-120$  und  $30$ :

$$\left( \begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{48}{120} & -\frac{12}{120} & -\frac{12}{120} \\ 0 & 1 & 0 & -\frac{12}{120} & \frac{48}{120} & -\frac{12}{120} \\ 0 & 0 & 1 & -\frac{3}{30} & -\frac{3}{30} & \frac{12}{30} \end{array} \right) = \left( \begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{2}{5} & -\frac{1}{10} & -\frac{1}{10} \\ 0 & 1 & 0 & -\frac{1}{10} & \frac{2}{5} & -\frac{1}{10} \\ 0 & 0 & 1 & -\frac{1}{10} & -\frac{1}{10} & \frac{2}{5} \end{array} \right).$$

Und rechts vom Strich steht die Inverse von  $A$ . Man rechnet leicht nach, daß tatsächlich  $AX = I_3$  gilt.

**Frage:** Warum ist das so ?

Falls man stattdessen das System

$$Ax = \begin{pmatrix} 0 \\ 7 \\ 0 \end{pmatrix}$$

hätte lösen wollen, dann hätte man mit dem Schema

$$\left( \begin{array}{ccc|c} 3 & 1 & 1 & 0 \\ 1 & 3 & 1 & 7 \\ 1 & 1 & 3 & 0 \end{array} \right)$$

gestartet, und mit genau denselben Langzeilenumformungen wie oben wäre man auf das Schema

$$\left( \begin{array}{ccc|c} 1 & 0 & 0 & -\frac{7}{10} \\ 0 & 1 & 0 & \frac{14}{5} \\ 0 & 0 & 1 & -\frac{7}{10} \end{array} \right)$$

gekommen, von dem man die Lösung  $x = (-\frac{7}{10}, \frac{14}{5}, -\frac{7}{10})^\top$  abliest.

### 3.3 Schlüsselbegriffe

- Matrizen als zentrales Beispiel linearer Abbildungen,
- Beziehung zwischen Matrixprodukt einerseits und Nacheinanderausführung der zugeordneten linearen Abbildungen andererseits,
- Gauß–Jordan–Verfahren.



# Kapitel 4

## Homomorphismen

Wir setzen jetzt unsere Betrachtungen zur Linearen Algebra fort, und wir werden einigen Gedanken aus dem Einführungskapitel erneut begegnen, diesmal allerdings auf einer abstrakteren Ebene (didaktisches Spiralprinzip). Es ist also empfehlenswert, gelegentlich nochmal zurückzublättern und den früheren Stoff mit neuem Blick zu betrachten.

### 4.1 Allgemeine Eigenschaften

Jede Matrix  $A \in K^{n \times m}$  begründet eine Abbildung  $f_A: K^m \rightarrow K^n$ , gemäß der Vorschrift  $f_A(x) := Ax$ . Diese Abbildung genügt den Regeln

$$\begin{aligned} f_A(x + y) &= f_A(x) + f_A(y), & \forall x, y \in K^m, \\ f_A(\lambda x) &= \lambda f_A(x), & \forall \lambda \in K, \quad \forall x \in K^m. \end{aligned}$$

Wir benutzen dabei aber gar nicht, daß der Ausgangsvektorraum der  $K^m$  ist. Weil diese beiden Eigenschaften so wichtig sind, haben Abbildungen dieses Typs einen eigenen Namen bekommen:

**Definition 4.1.** Seien  $U$  und  $V$  Vektorräume über  $K$ , und sei  $f: U \rightarrow V$  eine Abbildung mit den Eigenschaften

$$f(x + y) = f(x) + f(y), \quad \forall x, y \in U, \tag{4.1}$$

$$f(\lambda x) = \lambda f(x), \quad \forall \lambda \in K, \quad \forall x \in U. \tag{4.2}$$

Dann sagen wir, daß  $f$  eine lineare Abbildung bzw. ein Homomorphismus<sup>1</sup> ist. Die Menge aller Homomorphismen von  $U$  nach  $V$  bezeichnen wir mit  $\text{Hom}(U \rightarrow V)$  bzw.  $\mathcal{L}(U \rightarrow V)$ . Wenn  $U = V$ , dann reden wir auch von Endomorphismen<sup>2</sup> und schreiben  $\mathcal{L}(U)$  anstatt von  $\mathcal{L}(U \rightarrow U)$ . Falls  $V = K$  ist, bezeichnet man die Homomorphismen auch als lineare Funktionale<sup>3</sup>.

Anstelle von Homomorphismen ist auch die Bezeichnung lineare Operatoren<sup>4</sup> gebräuchlich. In der Literatur sind die Schreibweisen  $\mathcal{L}(U, V)$  oder  $\mathcal{L}(U; V)$  weit verbreitet anstelle unserer Notation  $\mathcal{L}(U \rightarrow V)$ , die hoffentlich besser erklärt, wie hier eigentlich abgebildet wird.

Die beiden zentralen Eigenschaften (4.1) und (4.2) einer linearen Abbildung drücken wir wieder als kommutative Diagramme aus:

$$\begin{array}{ccc} \boxed{(x, y)} & \xrightarrow{f} & \boxed{(f(x), f(y))} \\ + \downarrow (\text{in } U) & & + \downarrow (\text{in } V) \\ \boxed{x + y} & \xrightarrow{f} & \boxed{\begin{array}{l} f(x) + f(y) \\ = f(x + y) \end{array}} \end{array} \qquad \begin{array}{ccc} \boxed{x} & \xrightarrow{f} & \boxed{f(x)} \\ \lambda \cdot \downarrow (\text{in } U) & & \lambda \cdot \downarrow (\text{in } V) \\ \boxed{\lambda x} & \xrightarrow{f} & \boxed{\begin{array}{l} \lambda f(x) \\ = f(\lambda x) \end{array}} \end{array}$$

Einige Eigenschaften ergeben sich direkt aus der Definition:

---

<sup>1</sup>homomorphism  
<sup>2</sup>endomorphism  
<sup>3</sup>linear functionals  
<sup>4</sup>linear operators

**Satz 4.2.** Sei  $f \in \mathcal{L}(U \rightarrow V)$ . Dann gilt

- $f(0) = 0$ ,
- $f(\sum_{j=1}^n \lambda_j x_j) = \sum_{j=1}^n \lambda_j f(x_j)$ .

*Beweis.* Sollten Sie selber können. □

**Beispiel 4.3.** 1. Sei  $A \in K^{n \times m}$ , dann definieren wir eine Abbildung  $f: K^m \rightarrow K^n$  gemäß  $f(x) = Ax$ ; und es ist dann  $f \in \text{Hom}(K^m \rightarrow K^n)$ .

2. Sei  $U = C^1([0, 1] \rightarrow \mathbb{R})$  der Raum der auf  $[0, 1]$  einmal stetig differenzierbaren Funktionen, und entsprechend  $V = C([0, 1] \rightarrow \mathbb{R})$ . Dann ist die Abbildung  $\frac{d}{dx}: U \rightarrow V$  eine lineare Abbildung.
3. Sei  $U = C^2(\mathbb{R} \rightarrow \mathbb{R})$ ,  $V = C(\mathbb{R} \rightarrow \mathbb{R})$  und  $L \in \mathcal{L}(U \rightarrow V)$  der zur Schwingungsdifferentialgleichung gehörige Differentialoperator  $L = \frac{d^2}{dt^2} + k \frac{d}{dt} + \omega^2$ , wobei  $k$  und  $\omega$  gewisse positive Konstanten sind. Die äquivalente Sprechweise in der Physik lautet: „für die Schwingungsdifferentialgleichung gilt das Superpositionsprinzip“.
4. Ähnliches gilt für die Partiellen Differentialoperatoren  $\Delta$  und  $\nabla$ .
5. Sei  $U = C(\mathbb{R} \rightarrow \mathbb{R})$  und  $V = C^1(\mathbb{R} \rightarrow \mathbb{R})$ . Dann ist der Integrationsoperator, der jedem  $f \in U$  eine Stammfunktion  $x \mapsto \int_0^x f(t)dt$  zuordnet, eine lineare Abbildung.
6. Sei  $U = C(\mathbb{R} \rightarrow \mathbb{R})$ ,  $V = \mathbb{R}$ , und  $\delta: U \rightarrow V$  die Abbildung  $f \mapsto f(0)$ , die jeder Funktion ihren Wert im Nullpunkt zuordnet. Diese sogenannte DIRACsche<sup>5</sup> Delta-Distribution ist ein lineares Funktional, das z.B. für die Beschreibung von Punktmassen verwendet wird.
7. Sei  $U = \mathbb{R}^3$  der Raum der Ortsvariablen und  $V = \mathbb{R}$ . Lineare Abbildungen  $k: U \rightarrow V$  werden in der Physik häufig als Wellenzahlvektoren bezeichnet.

Im Folgenden werden wir allgemeine Homomorphismen als Abbildungen zwischen zwei Vektorräumen  $U$  und  $V$  studieren.

**Satz 4.4.** Es seien  $U$  und  $V$  zwei Vektorräume. Sei  $(b_1, \dots, b_m)$  eine Basis für  $U$ , und sei  $(v_1, \dots, v_m)$  eine beliebige Familie von Vektoren in  $V$ .

Dann gibt es genau einen Homomorphismus  $f \in \mathcal{L}(U \rightarrow V)$  mit  $f(b_j) = v_j$  für jedes  $j$ . Dieser Homomorphismus wird gegeben durch die Formel

$$f(\alpha_1 b_1 + \dots + \alpha_m b_m) = \alpha_1 v_1 + \dots + \alpha_m v_m.$$

*Beweis.* Weil  $(b_1, \dots, b_m)$  eine Basis ist, gibt es zu jedem  $u \in U$  genau einen Satz von Koeffizienten  $(\alpha_1, \dots, \alpha_m)$  mit  $u = \alpha_1 b_1 + \dots + \alpha_m b_m$ . Die obige Formel beschreibt einen Homomorphismus (Beweis selbst vervollständigen!) und es ist  $f(b_j) = v_j$ , wie verlangt.

Es bleibt noch zu zeigen, daß dieses  $f$  der einzige Homomorphismus mit  $f(b_j) = v_j$  ist. Sei nun  $g$  ein weiterer Homomorphismus mit  $g(b_j) = v_j$  für jedes  $j$ . Dann ist wegen Satz 4.2

$$g(u) = g(\alpha_1 b_1 + \dots + \alpha_m b_m) = \alpha_1 g(b_1) + \dots + \alpha_m g(b_m) = \alpha_1 v_1 + \dots + \alpha_m v_m = f(u),$$

also ist  $g = f$ . □

Damit können wir jetzt eine offengebliebene Frage des vorigen Kapitels beantworten. Damals wurde behauptet, daß jede lineare Abbildung  $f: K^m \rightarrow K^n$  in Form einer Matrix dargestellt werden kann.

Gegeben sei eine lineare Abbildung  $f \in \mathcal{L}(K^m \rightarrow K^n)$ .

Gesucht ist eine Matrix  $A \in K^{n \times m}$  mit  $Ax = f(x)$  für jedes  $x \in K^m$ .

Um eine solche Matrix zu finden, wählen wir eine Basis in  $K^m$ , zum Beispiel die Standardbasis  $(e_1, \dots, e_m)$ . Seien  $v_j$  die Bilder der Einheitsvektoren,  $v_j = f(e_j)$ ,  $j = 1, \dots, m$ . Wenn wir diese Spaltenvektoren nebeneinanderstellen, erhalten wir eine Matrix  $A$ . Diese ist genau die gesuchte Matrix, denn gemäß der Regel

*die Spalten sind die Koordinaten der Bilder der Basisvektoren*

<sup>5</sup>PAUL ADRIEN MAURICE DIRAC (1902–1984), Nobelpreis 1933 zusammen mit Erwin Schrödinger

ist  $Ae_j = v_j$ , also  $f(e_j) = v_j$ . Und der vorige Satz sagt uns dann, daß die Gleichung  $f(u) = Au$  nicht nur für  $u \in \{e_1, \dots, e_m\}$  gilt, sondern für jeden Vektor  $u \in U$ .

**Frage:** Was ist der Unterschied zwischen den folgenden beiden Formulierungen:

„ $f$  ist eine lineare Abbildungen vom  $\mathbb{R}^7$  **in** den  $\mathbb{R}^7$ “

„ $f$  ist eine lineare Abbildungen vom  $\mathbb{R}^7$  **auf** den  $\mathbb{R}^7$ “ ?

**Definition 4.5.** Seien  $\mathcal{A}$  und  $\mathcal{B}$  Mengen, und  $f: \mathcal{A} \rightarrow \mathcal{B}$  eine Abbildung.

- $f$  heißt injektiv<sup>6</sup>, wenn aus  $f(x) = f(x')$  folgt:  $x = x'$ .
- $f$  heißt surjektiv<sup>7</sup>, wenn es zu jedem  $y \in \mathcal{B}$  mindestens ein  $x \in \mathcal{A}$  mit  $y = f(x)$  gibt.
- $f$  heißt bijektiv<sup>8</sup>, wenn  $f$  injektiv und surjektiv ist.
- Die identische Abbildung  $\text{id}_{\mathcal{A}}: \mathcal{A} \rightarrow \mathcal{A}$  wird definiert durch  $\text{id}_{\mathcal{A}}(x) := x$  für jedes  $x \in \mathcal{A}$ .

Wir betonen, daß  $\mathcal{A}$  und  $\mathcal{B}$  einfach nur Mengen sind; sie brauchen keine Vektorräume zu sein.

**Satz 4.6.** Identische Abbildungen sind bijektiv.

*Beweis.* ist recht einfach und soll deshalb eine Übungsaufgabe sein. □

**Satz 4.7.** Seien  $\mathcal{A}, \mathcal{B}, \mathcal{C}$  Mengen,  $f$  eine Abbildung von  $\mathcal{A}$  in  $\mathcal{B}$ ,  $g$  eine Abbildung von  $\mathcal{B}$  in  $\mathcal{C}$ . Dann gilt:

1. Wenn  $g \circ f$  surjektiv ist, dann auch  $g$ .
2. Wenn  $g \circ f$  injektiv ist, dann auch  $f$ .

*Beweis.* Wir benutzen die Beweismethode der *Kontraposition*:<sup>9</sup>

Zu Teil 1: Wenn  $g$  die Menge  $\mathcal{C}$  nicht ausschöpft, dann kann  $g \circ f$  die Menge  $\mathcal{C}$  erst recht nicht ausschöpfen.

Zu Teil 2: Wenn  $f$  nicht injektiv ist, dann bildet  $f$  zwei verschiedene Elemente von  $\mathcal{A}$  auf dasselbe Element von  $\mathcal{B}$  ab. Dann kann auch  $g \circ f$  nicht injektiv sein. □

**Satz 4.8.** Seien  $\mathcal{A}, \mathcal{B}$  Mengen, und sei  $f: \mathcal{A} \rightarrow \mathcal{B}$  eine Abbildung. Dann gilt:

$f$  ist genau dann bijektiv, wenn es eine Abbildung  $g: \mathcal{B} \rightarrow \mathcal{A}$  gibt mit

$$f \circ g = \text{id}_{\mathcal{B}}, \quad g \circ f = \text{id}_{\mathcal{A}}.$$

*Beweis.* „ $\implies$ “:

Sei  $f$  bijektiv. Dann ist  $f$  surjektiv, also gibt es für jedes  $b \in \mathcal{B}$  mindestens ein  $a \in \mathcal{A}$  mit  $f(a) = b$ . Und weil  $f$  auch injektiv ist, ist dieses Element  $a$  eindeutig. Die gesuchte Abbildung  $g$  ist gerade diejenige, die  $b$  auf  $a$  sendet.

„ $\impliedby$ “:

Wegen Satz 4.6 ist  $\text{id}_{\mathcal{B}}$  bijektiv, also ist auch  $f \circ g = \text{id}_{\mathcal{B}}$  bijektiv, also ist  $f \circ g$  surjektiv, und wegen Satz 4.7 ist dann auch  $f$  surjektiv.

Wegen Satz 4.6 ist  $\text{id}_{\mathcal{A}}$  bijektiv, also ist auch  $g \circ f = \text{id}_{\mathcal{A}}$  bijektiv, also ist  $g \circ f$  injektiv, und wegen Satz 4.7 ist dann auch  $f$  injektiv. □

**Satz 4.9.** Seien  $U, V, W$  Vektorräume über einem Körper  $K$ , und sei  $f \in \mathcal{L}(U \rightarrow V)$ ,  $g \in \mathcal{L}(V \rightarrow W)$ .

1. Die identische Abbildung  $\text{id}_U: u \mapsto u$  ist ein bijektiver Homomorphismus  $\text{id}_U \in \mathcal{L}(U \rightarrow U)$ .
2. Durch  $g \circ f: u \mapsto g(f(u))$  wird ein Homomorphismus  $g \circ f \in \mathcal{L}(U \rightarrow W)$  definiert.

<sup>6</sup>injective

<sup>7</sup>surjective

<sup>8</sup>bijjective

<sup>9</sup> Das bedeutet: Seien  $\mathfrak{A}$  und  $\mathfrak{B}$  Aussagen, und wir wollen beweisen: „Wenn  $\mathfrak{A}$ , dann  $\mathfrak{B}$ “. Äquivalent dazu ist die Aussage „Wenn  $\mathfrak{B}$  falsch ist, dann ist  $\mathfrak{A}$  auch falsch“. Diese zweite Variante ist manchmal einfacher zu beweisen.

*Beweis.* Das Ergebnis 1. ist ziemlich leicht und deshalb dem Leser überlassen.

Zum Beweis von 2. benutzen wir (4.1) und (4.2) für  $f$  und  $g$ :

$$(g \circ f)(x + y) = g(f(x + y)) = g(f(x) + f(y)) = g(f(x)) + g(f(y)) = (g \circ f)(x) + (g \circ f)(y).$$

Analog

$$(g \circ f)(\lambda x) = g(f(\lambda x)) = g(\lambda f(x)) = \lambda g(f(x)) = \lambda(g \circ f)(x).$$

□

**Definition 4.10.** *Bijektive Homomorphismen zwischen Vektorräumen heißen Isomorphismen<sup>10</sup>.*

*Wenn es zwischen zwei Vektorräumen  $U$  und  $V$  einen Isomorphismus gibt, dann nennt man beide Räume zueinander isomorph<sup>11</sup>.*

Im ersten Kapitel hatten wir bereits zwei zueinander isomorphe Vektorräume behandelt: nämlich einen Vektorraum der Verschiebungspfeile in der Ebene (nennen wir ihn  $U$ ), und den Raum  $V = \mathbb{R}^2$ , der aus Zahlenpaaren besteht. Der Zusammenhang zwischen  $\vec{u} \in U$  und einem dazugehörigen Zahlenpaar wird durch eine Basis  $(\vec{b}_1, \vec{b}_2)$  für  $U$  vermittelt. Wenn eine solche Basis gewählt ist, können wir jedes  $\vec{u}$  zerlegen im Sinne von  $\vec{u} = \xi_1 \vec{b}_1 + \xi_2 \vec{b}_2$ . Die Abbildung  $\vec{u} \mapsto \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$  ist eine Abbildung von  $U$  nach  $\mathbb{R}^2$ , und diese Abbildung ist linear und bijektiv, also ein Isomorphismus. Wir hätten auch eine andere Basis  $(\vec{b}'_1, \vec{b}'_2)$  anstatt  $(\vec{b}_1, \vec{b}_2)$  wählen können, und das hätte uns einen anderen Isomorphismus geliefert. Es gibt also unendlich viele Isomorphismen zwischen  $U$  und  $\mathbb{R}^2$ . In der Schule ist vermutlich der Eindruck entstanden, daß  $U$  als Vektorraum der Verschiebungspfeile in der Ebene und der  $\mathbb{R}^2$  „schon irgendwie derselbe Vektorraum“ sind. Dieser Eindruck ist in dem Sinne richtig, daß beide Vektorräume isomorph zueinander sind, und *isomorph* bedeutet eben *gleichgestaltig*.

**Satz 4.11.** *Seien  $A, B \in K^{n \times n}$  Matrizen, und  $f_A, f_B \in \mathcal{L}(K^n)$  die zugehörigen linearen Abbildungen. Dann ist die Komposition  $f_A \circ f_B: K^n \rightarrow K^n$  gegeben durch  $(f_A \circ f_B)(x) = ABx$ .*

*Zueinander inverse Isomorphismen  $f_A$  und  $f_B$  werden durch zueinander inverse Matrizen  $A$  und  $B$  beschrieben.*

*Beweis.* Übungsaufgabe. Man nutze Satz 3.5.

□

## 4.2 Geometrische Aspekte

Wir brauchen noch einige weitere Begriffe.

**Definition 4.12.** *Sei  $f \in \text{Hom}(U \rightarrow V)$ . Dann heißt*

- $\ker f := \{u \in U : f(u) = 0\}$  der Nullraum bzw. Kern<sup>12</sup> von  $f$ ,
- $\text{img } f := \{f(u) \in V : u \in U\}$  der Bildraum bzw. das Bild<sup>13</sup> von  $f$ .

*Analog definieren wir für eine Matrix  $A \in K^{n \times m}$*

- $\ker A := \{x \in K^m : Ax = 0\}$ ,
- $\text{img } A := \{Ax \in K^n : x \in K^m\}$ .

**Satz 4.13.** *Wenn  $f \in \text{Hom}(U \rightarrow V)$ , dann ist  $\ker f$  ein Unterraum von  $U$  und  $\text{img } f$  Unterraum von  $V$ .*

*Beweis.* Es ist  $f(0) = 0$ , also ist  $0 \in \ker f$  und  $0 \in \text{img } f$ . Wenn nun  $u_1, u_2 \in \ker f$  sind, dann ist  $f(u_1 + u_2) = f(u_1) + f(u_2) = 0$ , also auch  $u_1 + u_2 \in \ker f$ . Analog beweist man  $\alpha u \in \ker f$ , falls  $\alpha \in K$  und  $u \in \ker f$ . Gemäß Satz 2.6 ist dann  $\ker f$  ein Unterraum von  $U$ .

Seien nun  $v_1, v_2 \in \text{img } f$ . Dann gibt es  $u_1, u_2 \in U$  mit  $f(u_1) = v_1$ ,  $f(u_2) = v_2$ . Dann gilt  $f(u_1 + u_2) = f(u_1) + f(u_2) = v_1 + v_2$ , also ist  $v_1 + v_2 \in \text{img } f$ . Analog zeigt man  $\alpha v \in \text{img } f$ , wenn  $v \in \text{img } f$  und  $\alpha \in K$ . Nun verwendet man nochmal den Satz 2.6, und der Beweis ist fertig.

□

<sup>10</sup>isomorphism

<sup>11</sup>isomorph to each other

<sup>12</sup>kernel

<sup>13</sup>image

**Definition 4.14.** Die Dimension des Bildes von  $f$  heißt  $\text{Rang}^{14}$  von  $f$ :

$$\text{rang } f := \dim \text{img } f.$$

Der Rang einer Matrix  $A$  wird definiert durch

$$\text{rang } A := \dim \text{img } A.$$

Die Dimension des Kernes von  $f$  heißt  $\text{Defekt}^{15}$  von  $f$ ,

$$\text{def } f := \dim \ker f.$$

**Satz 4.15.** Sei  $f \in \text{Hom}(U \rightarrow V)$  und sei  $(u_1, \dots, u_m)$  eine Familie in  $U$ . Dann haben wir die Aussagen:

1.  $f$  injektiv  $\iff \ker f = \{0\}$ ,
2.  $f$  surjektiv  $\iff \text{img } f = V$ ,
3.  $(f(u_1), \dots, f(u_m))$  ist unabhängig in  $V \implies (u_1, \dots, u_m)$  sind unabhängig in  $U$ ,
4.  $(u_1, \dots, u_m)$  sind unabhängig in  $U$  und  $f$  ist injektiv  $\implies (f(u_1), \dots, f(u_m))$  sind in  $V$  unabhängig,
5.  $(u_1, \dots, u_m)$  erzeugen  $U \implies (f(u_1), \dots, f(u_m))$  erzeugen  $\text{img } f$ ,
6.  $(u_1, \dots, u_m)$  erzeugen  $U$  und  $f$  ist surjektiv  $\implies (f(u_1), \dots, f(u_m))$  erzeugen  $V$ .

*Beweis.* 1. Wenn  $f$  injektiv ist, dann gibt es nur ein  $u \in U$  mit  $f(u) = 0$ , nämlich  $u = 0$ . Also ist  $\ker f = \{0\}$ .

Sei nun andererseits  $\ker f = \{0\}$ . Wenn  $f(u) = f(u')$  ist, dann ist  $f(u - u') = 0$ , also  $u - u' \in \ker f = \{0\}$ , also  $u = u'$ .

2. So ist Surjektivität definiert.

3. Sei  $\sum_{j=1}^m \alpha_j u_j = 0$ . Zu zeigen wäre  $\alpha_j = 0$  für jedes  $j$ . Wir haben  $0 = f(0) = f(\sum_{j=1}^m \alpha_j u_j) = \sum_{j=1}^m \alpha_j f(u_j)$ . Nun sind aber die  $f(u_j)$  linear unabhängig, also sind alle  $\alpha_j = 0$ .

4. Sei  $\sum_{j=1}^m \alpha_j f(u_j) = 0$ , zu zeigen wäre  $\alpha_j = 0$  für jedes  $j$ . Nun ist wegen der Linearität  $f(\sum_{j=1}^m \alpha_j u_j) = 0$ . Weil  $f$  injektiv ist, haben wir mit 1.  $\ker f = \{0\}$ , also muß  $\sum_{j=1}^m \alpha_j u_j = 0$  sein. Da nun die  $(u_1, \dots, u_m)$  linear unabhängig sind, ist  $\alpha_j = 0$  für jedes  $j$ .

5. Sei  $v \in \text{img } f$ . Dann gibt es ein  $u \in U$  mit  $f(u) = v$ . Weil die  $u_j$  den Originalraum  $U$  erzeugen, gibt es Konstanten  $\alpha_1, \dots, \alpha_m$  mit  $u = \sum_{j=1}^m \alpha_j u_j$ . Dann ist  $v = f(u) = f(\sum_{j=1}^m \alpha_j u_j) = \sum_{j=1}^m \alpha_j f(u_j)$ . Also ist  $v$  als Linearkombination der  $(f(u_1), \dots, f(u_m))$  dargestellt.

6. Folgt direkt aus 2. und 5.

□

Ein Teilergebnis dieses Satzes ist so wichtig, daß wir es nochmal hinschreiben sollten:

**Folgerung 4.16.** Sei  $f \in \text{Hom}(U \rightarrow V)$  ein injektiver Homomorphismus. Dann gilt

$$(u_1, \dots, u_m) \text{ linear unabhängig in } U \iff (f(u_1), \dots, f(u_m)) \text{ linear unabhängig in } V.$$

Die Injektivität von  $f$  ist hierbei die entscheidende Voraussetzung !

Als weitere fast triviale Folgerung ergibt sich:

**Satz 4.17.** 1. Die Spalten einer Matrix sind ein Erzeugendensystem für das Bild.

2. Der Rang einer Matrix ist die Maximalzahl linear unabhängiger Spalten dieser Matrix.

Rang und Defekt einer Abbildung sind nicht unabhängig voneinander:

---

<sup>14</sup>rank  
<sup>15</sup>defect

**Satz 4.18 (Dimensionsformel).** Sei  $f \in \text{Hom}(U \rightarrow V)$ . Dann ist

$$\text{def } f + \text{rang } f = \dim U.$$

Für Matrizen  $A \in K^{n \times m}$  bedeutet das

$$\text{def } A + \text{rang } A = m.$$

*Beweis.* Sei  $(u_1, \dots, u_s)$  eine Basis für  $\ker f$ . Nach dem Basisergänzungssatz können wir diese Basis von  $\ker f$  zu einer Basis von ganz  $U$  ergänzen:  $(u_1, \dots, u_s, u_{s+1}, \dots, u_m)$ . Dann erzeugen die Vektoren  $(f(u_1), \dots, f(u_m))$  den Raum  $\text{img } f$ , wegen Satz 4.15, Teil 5. Weil  $f(u_1) = \dots = f(u_s) = 0$ , kann man diese Vektoren im Erzeugendensystem weglassen, und kommt zu einem Erzeugendensystem  $(f(u_{s+1}), \dots, f(u_m))$  für  $\text{img } f$ .

Als nächstes wollen wir zeigen, daß diese Vektoren  $(f(u_{s+1}), \dots, f(u_m))$  linear unabhängig sind.

Sei nun also  $\sum_{j=s+1}^m \alpha_j f(u_j) = 0$ . Dann ist  $f(\sum_{j=s+1}^m \alpha_j u_j) = 0$ , also ist  $\sum_{j=s+1}^m \alpha_j u_j \in \ker f$ . Aber  $\ker f$  hatte gerade die Basis  $(u_1, \dots, u_s)$ , und die Vektoren  $(u_1, \dots, u_s, u_{s+1}, \dots, u_m)$  sind linear unabhängig. Das kann nur sein, wenn alle  $\alpha_j = 0$  sind, für  $j = s+1, \dots, m$ .

Damit ist  $(f(u_{s+1}), \dots, f(u_m))$  eine Basis für  $\text{img } f$ .

Die Dimensionsformel ergibt sich nun durch Abzählen. □

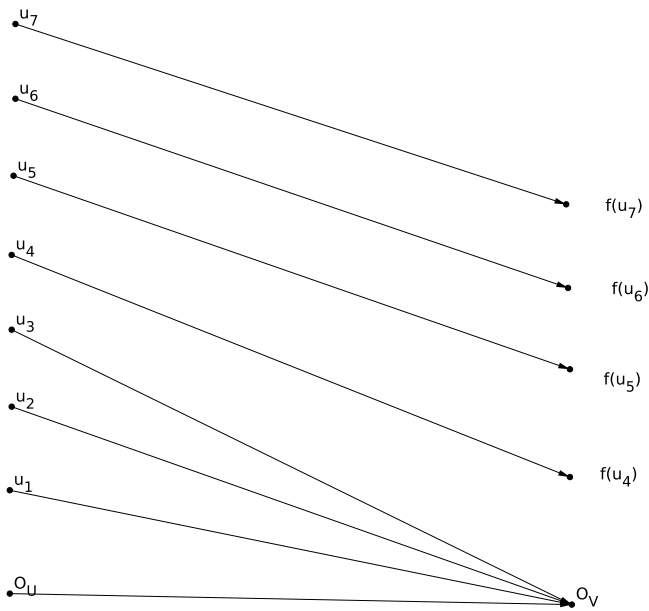


Abbildung 4.1: Pseudo-schematische Veranschaulichung der Dimensionsformel, mit  $\dim U = m = 7$  und  $\dim \ker f = s = 3$

Damit können wir nun die Abbildungseigenschaften der einer invertierbaren Matrix zugehörigen Abbildung beschreiben:

**Satz 4.19.** Seien  $U, V$  Vektorräume über  $K$ , und sei  $\dim U = \dim V$ . Sei  $f \in \text{Hom}(U \rightarrow V)$ . Dann sind die folgenden Aussagen äquivalent:

1.  $f$  ist Isomorphismus,
2.  $f$  ist injektiv,
3.  $f$  ist surjektiv,
4.  $f$  bildet eine Basis  $(u_1, \dots, u_m)$  von  $U$  auf eine Basis  $(f(u_1), \dots, f(u_m))$  von  $V$  ab.

*Beweis.* 2  $\iff$  3:

Wir haben  $\dim \ker f + \dim \operatorname{img} f = \dim U = \dim V$ . Wenn  $f$  injektiv ist, dann ist  $\ker f = \{0\}$ , also  $\dim \ker f = 0$ , also  $\dim \operatorname{img} f = \dim V$ , also  $\operatorname{img} f = V$ , also ist  $f$  surjektiv. Wenn  $f$  surjektiv ist, dann ist  $\operatorname{img} f = V$ , also  $\dim \operatorname{img} f = \dim V$ , also muß  $\dim \ker f = 0$  sein, also ist  $\ker f = \{0\}$ , also ist  $f$  injektiv.

1  $\iff$  2, 3:

Das ist genau Definition 4.10.

2, 3  $\implies$  4:

Es ist  $(u_1, \dots, u_m)$  unabhängig in  $U$ , und  $f$  ist injektiv: also ist wegen Satz 4.15, Teil 4 auch  $(f(u_1), \dots, f(u_m))$  eine linear unabhängige Familie in  $V$ .

Weiterhin erzeugt  $(u_1, \dots, u_m)$  den Raum  $U$ , und  $f$  ist surjektiv: also erzeugt wegen Satz 4.15, Teil 6 auch  $(f(u_1), \dots, f(u_m))$  den Raum  $V$ .

Also ist  $(f(u_1), \dots, f(u_m))$  eine Basis von  $V$ .

4  $\implies$  3:

Es liegen  $f(u_1), \dots, f(u_m)$  in  $\operatorname{img} f$ , aber diese  $m$  Vektoren sind eine linear unabhängige Familie, denn sie bilden eine Basis von  $V$ . Also ist  $\dim \operatorname{img} f \geq m$ . Wegen  $\dim V = m$  folgt dann  $\operatorname{img} f = V$ , also ist  $f$  surjektiv.  $\square$

Wir übertragen die Ergebnisse dieses Satzes auf Matrizen:

**Satz 4.20.** Sei  $A \in K^{n \times n}$  und  $f \in \operatorname{Hom}(K^n \rightarrow K^n)$  die von  $A$  erzeugte lineare Abbildung. Dann sind die folgenden Aussagen äquivalent:

1.  $f$  ist Isomorphismus,
2.  $A$  ist invertierbar,
3.  $\operatorname{rang} A = n$ ,
4.  $\operatorname{def} A = 0$ ,
5. Die Spalten von  $A$  erzeugen den  $K^n$ ,
6. Die Spalten von  $A$  sind linear unabhängig,
7. Die Spalten von  $A$  sind eine Basis des  $K^n$ .

*Beweis.* 1  $\iff$  2 :

Siehe Satz 4.11.

1  $\iff$  3  $\iff$  4 :

Die Aussage  $\operatorname{rang} A = n$  ist gleichbedeutend mit  $\operatorname{img} A = K^n$ , also mit der Surjektivität von  $f$ , gemäß Satz 4.15. Und  $\operatorname{def} A = 0$  bedeutet  $\ker A = \{0\}$ , also die Injektivität von  $f$ , ebenfalls wegen Satz 4.15.

Die Äquivalenz von 1, 3 und 4 folgt nun aus Satz 4.19.

1  $\iff$  7 :

Die Spalten sind gerade die Bilder  $(f(e_1), \dots, f(e_m))$  der Standardbasisvektoren  $(e_1, \dots, e_m)$ . Die Behauptung folgt aus Satz 4.19, von Aussage 1 zu Aussage 4.

5  $\implies$  7 :

Nach Satz 2.14 enthält das System der  $n$  Spaltenvektoren eine Basis des  $K^n$ . Wegen  $\dim K^n = n$  muß jede Basis aber aus genau  $n$  Vektoren bestehen, also ist das System der  $n$  Spaltenvektoren schon eine Basis.

7  $\implies$  5 und 7  $\implies$  6 :

So sind Basen definiert !

6  $\implies$  7 :

Nach dem Basisergänzungssatz kann man das linear unabhängige System der  $n$  Spaltenvektoren zu einer Basis des  $K^n$  ergänzen. Diese hat aber genau  $n$  Vektoren, also müssen die  $n$  Spaltenvektoren von  $A$  schon eine Basis sein. (Oder man nimmt die zweite Behauptung des Austauschsatzes von Steinitz.)  $\square$

**Folgerung 4.21.** Sei  $A \in K^{n \times m}$  und  $C \in K^{n \times n}$ , wobei  $C$  invertierbar sei.

Dann haben  $A$  und  $CA$  denselben Rang.

*Beweis.* Das folgt aus Folgerung 4.16: wenn die Matrix  $A$  einige linear unabhängige Spalten enthält, dann sind die entsprechenden Spalten von  $CA$  auch linear unabhängig, und umgekehrt. Die Maximalzahl linear unabhängiger Spalten ist aber gerade gleich dem Rang.  $\square$

**Satz 4.22.** Sei  $A \in K^{n \times m}$ , und sei  $\tilde{A}$  eine GAUSS–JORDAN–Transformierte von  $A$ . Das heißt,  $\tilde{A}$  ergibt sich aus  $A$  durch Zeilenumformungen und hat Zeilenstufenform.

Dann haben  $A$  und  $\tilde{A}$  denselben Rang, und  $A$  ist invertierbar genau dann, wenn  $n = m$  und  $\tilde{A}$  die Gestalt einer oberen Dreiecksmatrix hat.

*Beweis.* Wir erhalten  $\tilde{A}$  als Produkt  $LA$ , wobei  $L$  eine reguläre Matrix ist, die wir als Produkt von Eliminationsmatrizen und Permutationsmatrizen schreiben können. Dann haben  $\tilde{A}$  und  $A$  denselben Rang.

Wir wissen: wenn  $\tilde{A}$  eine obere Dreiecksmatrix ist, dann können wir das GAUSS–JORDAN–Verfahren weiterführen, indem wir noch nach oben ausräumen, solange bis eine Matrix  $\tilde{\tilde{A}} = I_n$  entsteht, also  $LA = I_n$ . Dann ist  $A$  invertierbar.

Andererseits, wenn nun  $\tilde{A}$  auf Zeilenstufenform gebracht wurde, aber keine Dreiecksgestalt hat, dann ist die unterste Zeile von  $\tilde{A}$  eine Nullzeile. Die Spalten von  $\tilde{A}$  spannen dann nicht den  $K^n$  auf, also ist  $\tilde{A}$  nicht invertierbar. Dann ist  $\text{rang } \tilde{A} < n$ , also auch  $\text{rang } A < n$ , also ist  $A$  nicht invertierbar.  $\square$

**12 Fragen:** Sei  $U = \mathbb{R}^7$  und  $V = \mathbb{R}^8$ . Wieviele injektive (surjektive) (bijektive) (identische) Abbildungen gibt es von  $U$  nach  $V$ ? Von  $V$  nach  $U$ ? Von  $V$  nach  $V$ ? (Alle Abbildungen seien linear.)

### 4.3 Lineare Gleichungen

Es geht um Gleichungen  $f(u) = v$ , wobei  $f \in \text{Hom}(U \rightarrow V)$  eine lineare Abbildung zwischen zwei Vektorräumen  $U$  und  $V$  ist. Dieses  $f$  und ein Vektor  $v \in V$  sind gegeben, und  $u \in U$  ist gesucht. Folgende Fragen ergeben sich naheliegenderweise:

- Gibt es überhaupt Lösungen?
- Wenn ja, wieviele sind es?
- Manchmal gibt es für gewisse  $v \in V$  eine Lösung  $u$ , für andere  $v \in V$  aber keine Lösung  $u$ . Welche Bedingungen muß  $v$  erfüllen, damit es (mindestens) eine Lösung gibt?
- Wie findet man die Lösungen?

**Satz 4.23.** Seien  $U$  und  $V$  Vektorräume über  $K$ ,  $f \in \text{Hom}(U \rightarrow V)$  und  $v \in V$  gegeben.

- Wenn  $u_1, u_2 \in U$  Lösungen von  $f(u) = v$  sind, dann ist  $u_1 - u_2$  eine Lösung von  $f(u) = 0$ .
- Wenn  $u_1$  eine Lösung von  $f(u) = v$  und  $u_2$  eine Lösung von  $f(u) = 0$  sind, dann ist  $u_1 + u_2$  eine Lösung von  $f(u) = v$ .

*Beweis.* Das sollten Sie selber können.  $\square$

Wir erhalten die folgende Merkregel:

$$\begin{aligned}
 & \text{Die allgemeine Lösung des inhomogenen Problems} \\
 & \quad = \\
 & \quad \text{eine spezielle Lösung des inhomogenen Problems} \\
 & \quad + \\
 & \quad \text{die allgemeine Lösung des homogenen Problems.}
 \end{aligned}$$



Die allgemeine Lösung des homogenen Problems wird gerade beschrieben durch den Kern von  $f$ . Demnach besitzt die Gleichung  $f(u) = v$  also Lösungen in einer linearen Mannigfaltigkeit mit der Dimension  $\text{def } f = \dim U - \text{rang } f$ ; aber nur, wenn sie überhaupt eine Lösung besitzt.

**Satz 4.24.** *Seien  $U$  und  $V$  Vektorräume über  $K$ , und zwar endlichdimensional. Sei  $f \in \text{Hom}(U \rightarrow V)$ . Dann gilt:*

1. Die Gleichung  $f(u) = v$  hat für jedes  $v \in V$  mindestens eine Lösung  $u \in U \iff \text{rang } f = \dim V$ .
2. Die Gleichung  $f(u) = v$  hat für jedes  $v \in V$  maximal eine Lösung  $u \in U \iff \text{rang } f = \dim U$ .
3. Die Gleichung  $f(u) = v$  hat für jedes  $v \in V$  genau eine Lösung  $u \in U \iff \text{rang } f = \dim U = \dim V$ .

*Beweis.* 1. Trivial wegen  $\text{img } f = V$ .

2. Wenn es maximal eine Lösung gibt, dann ist  $\ker f = \{0\}$ , also  $0 = \text{def } f = \dim U - \text{rang } f$  und umgekehrt.
3. Folgt sofort aus 1. und 2.

□

Wenn wir diesen Satz auf den speziellen Fall von Matrizen beziehen, erhalten wir:

**Satz 4.25.** *Sei  $A \in K^{n \times m}$ . Dann gilt für Gleichungssysteme  $Ax = y$  mit gegebenem  $y \in K^n$  und gesuchtem  $x \in K^m$ :*

1. Das Gleichungssystem  $Ax = y$  ist lösbar  $\iff y \in \text{img } A \iff \text{rang } A = \text{rang}(A|y)$ , wobei  $(A|y)$  die Matrix bezeichnet, die durch Nebeneinanderstellen von  $A$  und  $y$  entsteht. Es gibt dann  $\text{def } A = m - \text{rang } A$  linear unabhängige Lösungen für das homogene Problem  $Ax = 0$ .
2.  $Ax = y$  hat für jedes  $y \in K^n$  mindestens eine Lösung  $x \in K^m \iff \text{rang } A = n \iff$  die Spalten von  $A$  erzeugen den  $K^n$ .
3.  $Ax = y$  hat für jedes  $y \in K^n$  maximal eine Lösung  $x \in K^m \iff \text{rang } A = m \iff$  die Spalten von  $A$  sind linear unabhängig.
4.  $Ax = y$  hat für jedes  $y \in K^n$  genau eine Lösung  $x \in K^m \iff \text{rang } A = m = n \iff$  die Spalten von  $A$  sind eine Basis des  $K^n$ .

*Beweis.* 1. Hinschauen. Die Spalten von  $A$  spannen  $\text{img } A$  auf.

2. Siehe 4.24, Teil 1.
3. Siehe 4.24, Teil 2. Der Rang von  $A$  ist die Anzahl der linear unabhängigen Spalten von  $A$ , aber  $A$  hat genau  $m$  Spalten.
4. Folgt aus 2. und 3.

□

## 4.4 Basistransformationen

**Problem:** Gegeben sei  $U = K^m$  mit 2 Basen:  $(b_1, \dots, b_m)$  und  $(b'_1, \dots, b'_m)$ . Wir stellen uns  $B$  als alte Basis vor, und  $B'$  als neue Basis. Für einen Vektor  $u \in U$  haben wir dann zwei Darstellungen:

$$u = \sum_{j=1}^m \xi_j b_j = \sum_{j=1}^m \xi'_j b'_j,$$

mit Koeffizienten  $\xi_j, \xi'_j \in K$ . Wir fassen diese Koeffizienten zu Spaltenvektoren zusammen:

$$x = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_m \end{pmatrix}_B, \quad x' = \begin{pmatrix} \xi'_1 \\ \vdots \\ \xi'_m \end{pmatrix}_{B'}.$$

Wie kann man die neue Koordinatenspalte  $x'$  aus der alten Koordinatenspalte  $x$  bestimmen ?

**Lösung:** Weil die Basisvektoren aus dem  $K^m$  stammen, kann man sie als Spaltenvektoren interpretieren. Wenn man diese Spaltenvektoren nebeneinanderstellt, erhält man quadratische Matrizen  $B$  und  $B'$ . Diese sind regulär, weil die Spaltenvektoren linear unabhängig sind. Dann läßt sich die Darstellungsformel für  $u$  gerade schreiben als

$$u = Bx = B'x',$$

woraus wir sofort

$$x' = (B')^{-1}Bx$$

erhalten. Wir ziehen daraus eine Merkregel (und stellen uns dabei vor, daß  $x$  bezüglich der Standardbasis  $(e_1, e_2, \dots, e_m)$  dargestellt wird, also  $B = I_m$  die Einheitsmatrix ist):

*Im  $K^m$  kann man die Multiplikation einer Zahlenspalte  $x$  mit einer invertierbaren Matrix ansehen als Umrechnung der **Koordinaten** des von  $x$  bezüglich der Standardbasis dargestellten Vektors auf eine neue Basis.*

Einerseits: die *alte* Basis war jetzt  $(e_1, \dots, e_m)$ , und der  $j$ -te *neue* Basisvektor lautet  $b'_j = B'e_j$ .

Andererseits: die *alten* Koordinaten (bezüglich  $(e_1, \dots, e_m)$ ) eines Punktes hatten wir  $x$  genannt, und dann lauten die *neuen* Koordinaten desselben Punktes jetzt  $x' = (B')^{-1}x$ .

Wir bekommen also zwei verschiedene Umrechnungsregeln für die Basisvektoren und für die Koordinaten.

**Problem:** Seien  $U = K^m$ ,  $V = K^n$  und  $f \in \text{Hom}(U \rightarrow V)$ . Bezogen auf die Standardbasen  $(e_1, \dots, e_m)$  und  $(e_1, \dots, e_n)$  von  $U$  bzw.  $V$  wird die lineare Abbildung  $f$  durch eine Matrix  $A \in K^{n \times m}$  dargestellt:  $f(u) = Au$ , wenn  $u$  als Spaltenvektor interpretiert wird.

Wir geben uns noch jeweils eine weitere Basis von  $U$  bzw.  $V$  vor:  $(b_1^U, \dots, b_m^U)$  für  $U$  und  $(b_1^V, \dots, b_n^V)$  für  $V$ . Wenn wir diese Spaltenvektoren in der angegebenen Reihenfolge nebeneinanderstellen, erhalten wir reguläre Matrizen  $B_U \in K^{m \times m}$  und  $B_V \in K^{n \times n}$ . Sei nun  $v = f(u)$ , also  $v = Au$ . Die Vektoren  $u$  und  $v$  können wir auch in den neuen Basen darstellen:

$$u = \sum_{j=1}^m \xi_j^U b_j^U, \quad v = \sum_{j=1}^n \xi_j^V b_j^V,$$

wobei  $\xi_j^U, \xi_j^V \in K$ . Diese Koeffizienten können wir wieder in Spaltenform anordnen:

$$x^U = \begin{pmatrix} \xi_1^U \\ \vdots \\ \xi_m^U \end{pmatrix}_{B_U}, \quad x^V = \begin{pmatrix} \xi_1^V \\ \vdots \\ \xi_n^V \end{pmatrix}_{B_V}.$$

Wie sieht nun die Abbildungsmatrix bezüglich der neuen Basen aus ? Gesucht ist also eine Matrix  $\tilde{A} \in K^{n \times m}$  mit

$$x^V = \tilde{A}x^U.$$

Und diese Beziehung soll natürlich für jeden Koeffizientenvektor  $x^U$  gelten, wenn  $x^V$  den Koeffizientenvektor des zugehörigen Bildvektors  $v$  darstellt.

**Lösung:** Die Darstellung der Vektoren  $u$  und  $v$  als Linearkombinationen der Basisvektoren  $b_j^U$  und  $b_j^V$  kann man auch schreiben als

$$u = B_U x^U, \quad v = B_V x^V.$$

Zusammen mit  $v = Au$  erhalten wir damit

$$x^V = B_V^{-1}v = B_V^{-1}Au = (B_V^{-1}AB_U)x^U = \tilde{A}x^U.$$

Im Hinblick auf die obige Merkregel ist dieses Dreierprodukt von Matrizen nicht überraschend: die beiden äußeren Matrizen dienen dem Umrechnen zwischen verschiedenen Basen, und  $A$  erzeugt die Abbildung.

Wir fassen die Ergebnisse zusammen:

**Satz 4.26.** *Mit den obigen Bezeichnungen ist  $\tilde{A} = B_V^{-1}AB_U$ . Die Bilder der neuen Basisvektoren  $b_j^U$  werden gegeben durch*

$$f(b_j^U) = Ab_j^U = \sum_{k=1}^n \tilde{a}_{kj} b_k^V,$$

wobei  $\tilde{a}_{kj}$  gerade die Einträge der Matrix  $\tilde{A}$  sind.

*Beweis.* Es ist lediglich der zweite Teil noch offen. Nun ist aber

$$\begin{aligned} f(b_j^U) &= Ab_j^U = A \cdot (j\text{-te Spalte von } B_U) = j\text{-te Spalte von } (AB_U) = (AB_U)e_j \\ &= B_V B_V^{-1} AB_U e_j = B_V \tilde{A} e_j = B_V \cdot (j\text{-te Spalte von } \tilde{A}) \\ &= \sum_{k=1}^n \tilde{a}_{kj} b_k^V. \end{aligned}$$

□

Im wichtigen Sonderfall  $U = V = K^n$  haben wir die Freiheit,  $B_U$  und  $B_V$  unabhängig voneinander zu wählen, normalerweise nicht.

**Definition 4.27.** *Zwei Matrizen  $A, \tilde{A} \in K^{n \times n}$  heißen ähnlich<sup>16</sup>, wenn es eine invertierbare Matrix  $B \in K^{n \times n}$  gibt mit  $\tilde{A} = B^{-1}AB$ .*

Es stellt sich die Frage, wie man die Matrix  $B$  wählen kann/soll, damit  $\tilde{A}$  möglichst einfach oder möglichst schön wird. Die Antwort darauf ist mit unseren jetzigen Kenntnissen noch nicht möglich und muß auf ein späteres Kapitel vertagt werden.

**Satz 4.28.** *Sei  $A \in K^{n \times m}$ ,  $B_U \in K^{m \times m}$ ,  $B_V \in K^{n \times n}$ , wobei  $B_U$  und  $B_V$  invertierbar seien. Dann haben  $A$  und  $\tilde{A} = B_V^{-1}AB_U$  denselben Rang.*

*Beweis.* Wir definieren eine lineare Abbildung  $f \in \text{Hom}(K^m \rightarrow K^n)$  durch  $f(u) := Au$ . Dann ist  $\text{rang } A = \text{rang } f = \dim \text{img } f$ . Nun beschreiben  $A$  und  $\tilde{A}$  die gleiche Abbildung, lediglich in verschiedenen Basen. Da der Rang einer Matrix sich interpretieren läßt als Dimension des Bildraumes, müssen die Ränge von  $A$  und  $\tilde{A}$  gleich sein. □

Der Rang einer Matrix ist gleich der Anzahl der linear unabhängigen Spalten, man redet auch vom *Spaltenrang*.

**Definition 4.29.** *Der Zeilenrang einer Matrix ist gleich der Anzahl der linear unabhängigen Zeilen.*

**Satz 4.30.** *Für jede Matrix stimmen Zeilenrang und Spaltenrang überein.*

*Beweis.* Für den Zeilenrang bzw. Spaltenrang einer Matrix  $A$  schreiben wir  $\text{Zeilenrang}(A)$  bzw.  $\text{Spaltenrang}(A)$ . Sei  $A$  die fragliche Matrix mit Rang (Spaltenrang)  $r$ . Wir wählen geeignete Matrizen  $B_U$  und  $B_V$  (zusammenggebaut aus Eliminationsmatrizen und Permutationsmatrizen) so, daß

$$B_V^{-1}AB_U = \tilde{A} = \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \in K^{n \times m}.$$

Nun gilt für die beiden Rangtypen, durch zweimaliges Anwenden von Satz 4.28:

$$\begin{aligned} r &= \text{Spaltenrang}(A) = \text{Spaltenrang}(\tilde{A}) = \text{Zeilenrang}(\tilde{A}) = \text{Spaltenrang}((\tilde{A})^\top) \\ &= \text{Spaltenrang}(B_U^\top A^\top (B_V^{-1})^\top) = \text{Spaltenrang}(B_U^\top A^\top (B_V^\top)^{-1}) \\ &= \text{Spaltenrang}(A^\top) \\ &= \text{Zeilenrang}(A). \end{aligned}$$

□

**Bemerkung 4.31.** *Es empfiehlt sich (Spiralprinzip der Didaktik), den Abschnitt 1.5.3 erneut zu lesen, und dabei insbesondere das Dreierprodukt  $R = AR_0A^\top$  (in der Notation von Satz 1.67) mit dem jetzigen Dreierprodukt  $\tilde{A} = B^{-1}AB$  wie in Definition 4.27 zu vergleichen. Welche Gemeinsamkeiten finden Sie ?*

<sup>16</sup>similar

## 4.5 Differentialgleichungen

Ziel dieses einführenden Abschnittes soll es sein, die Lösungsmenge von Differentialgleichungen mithilfe der Homomorphismentheorie zu beschreiben. Dabei orientieren wir uns an 2 typischen Beispielen:

**Beispiel 4.32.** Eine Zerfalls/Wachstumsgleichung mit Zufuhr und Anfangsbedingung wird gegeben durch

$$u'(t) + \alpha u(t) = f(t), \quad u(0) = u_0, \quad \alpha \in \mathbb{R}.$$

Eine Schwingungsgleichung mit Dämpfung und Anregung und 2 Anfangsbedingungen ist zum Beispiel

$$u''(t) + ku'(t) + \omega^2 u(t) = f(t), \quad u(0) = u_0, \quad u'(0) = u_1, \quad k, \omega > 0.$$

**Definition 4.33.** Sei  $U = C^n(\mathbb{R} \rightarrow \mathbb{R})$  und  $V = C(\mathbb{R} \rightarrow \mathbb{R})$ . Hierbei ist  $u \in U$  genau dann, wenn  $u$   $n$ -mal differenzierbar ist, und die  $n$ -te Ableitung ist noch eine stetige Funktion. Seien  $a_0, a_1, \dots, a_{n-1}$  stetige Funktionen von  $\mathbb{R}$  nach  $\mathbb{R}$ .

Eine Abbildung

$$L: U \rightarrow V,$$

$$L: u \mapsto Lu = Lu(t) := \frac{d^n}{dt^n}u(t) + a_{n-1}(t)\frac{d^{n-1}}{dt^{n-1}}u(t) + \dots + a_1(t)\frac{d}{dt}u(t) + a_0(t)u(t)$$

heißt Differentialoperator der Ordnung  $n$ .

Wenn  $f \in V$  gegeben ist, dann heißt

$$Lu = f$$

lineare Differentialgleichung der Ordnung  $n$ , und typischerweise ist die Funktion  $u$  gesucht. Falls  $f \equiv 0$  (also  $f(t) = 0$  für jedes  $t \in \mathbb{R}$ ), dann reden wir von einer homogenen Gleichung, ansonsten von einer inhomogenen Gleichung.

**Frage:** Was sind  $\dim U$  und  $\dim V$  ?

Im allgemeinen Fall haben wir es mit einem Anfangswertproblem

$$Lu = f, \quad (u(0), u'(0), \dots, u^{(n-1)}(0)) = (u_0, u_1, \dots, u_{n-1}) \quad (4.3)$$

zu tun, wobei  $f$  und die  $u_j$  gegeben sind, und das  $u$  ist gesucht.

**Satz 4.34.** Wenn  $u_*$  eine spezielle Lösung des inhomogenen Problems  $Lu = f$  ist, dann liegen alle Lösungen zu  $Lu = f$  in der Menge

$$u_* + \ker L := \{u \in U : u - u_* \in \ker L\}.$$

*Beweis.* Siehe Abschnitt 4.3. □

**Satz 4.35.** Es ist  $\dim \ker L = n$ . Die Anfangswertabbildung, die eine Funktion  $u$  auf ihre verallgemeinerten Anfangswerte zur Zeit 0 abbildet,

$$\text{AWA: } \ker L \rightarrow \mathbb{R}^n,$$

$$\text{AWA: } u \mapsto (u(0), u'(0), u''(0), \dots, u^{(n-1)}(0)),$$

ist bijektiv.

Der Beweis dazu verteilt sich auf mehrere Schritte. Leider können wir ihn mit unseren Kenntnissen nicht in der hier behaupteten Allgemeinheit führen, sodaß wir uns stattdessen beim Beweis der Surjektivität auf die Schwingungsgleichung beschränken und den Beweis des allgemeinen Falls auf das dritte Semester vertagen.

Auf jeden Fall erhalten wir aus diesem Satz ein allgemeines Verfahren zur Bestimmung der Lösung des Anfangswertproblems (4.3):

1. Finde eine spezielle Lösung  $u_* = u_*(t)$  zu  $Lu = f$ , zum Beispiel durch geschicktes Raten. Die Anfangswerte für  $t = 0$  sind im Moment egal.

2. Bestimme  $\ker L$ , also alle Funktionen  $u \in U$  mit  $Lu = 0$ .
3. Wähle dann aus  $\ker L$  ein Element  $u_h = u_h(t)$  mit besonderen Anfangswerten, nämlich mit  $(u_*(0), u'_*(0), \dots, u_*^{(n-1)}(0)) + (u_h(0), u'_h(0), \dots, u_h^{(n-1)}(0)) = (u_0, u_1, \dots, u_{n-1})$ .
4. Setze  $u(t) = u_*(t) + u_h(t)$ .

Aus Satz 4.35 ergibt sich, daß es immer eine Lösung  $u$  gibt, und daß die so konstruierte Lösung tatsächlich die einzige Lösung ist.

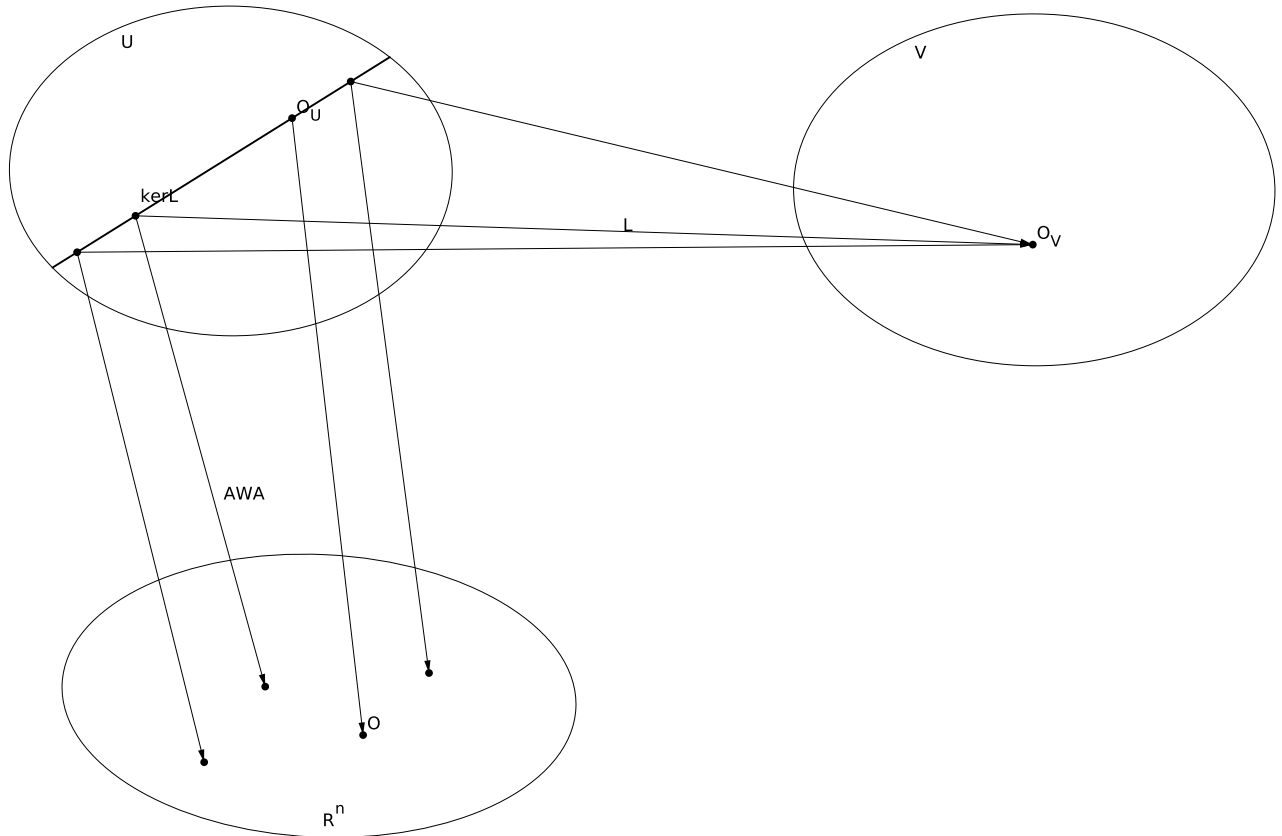


Abbildung 4.2: Die drei Vektorräume  $U = C^n(\mathbb{R} \rightarrow \mathbb{R})$ ,  $V = C(\mathbb{R} \rightarrow \mathbb{R})$  und  $\mathbb{R}^n$ , sowie die beiden linearen Abbildungen  $L$  und AWA.

Für den Beweis der Injektivität der Anfangswertabbildung brauchen wir ein Hilfsresultat:

**Lemma 4.36.** Sei  $u \in C^n(\mathbb{R} \rightarrow \mathbb{R})$  Lösung zu  $Lu = 0$  mit den verallgemeinerten Anfangsbedingungen  $(u(0), u'(0), \dots, u^{(n-1)}(0)) = (0, 0, \dots, 0)$ . Dann ist  $u(t) = 0$  für jedes  $t \in \mathbb{R}$ .

*Beweis.* Wir definieren uns eine (mathematische) Energie

$$E(t) := \frac{1}{2} \left( (u(t))^2 + (u'(t))^2 + \dots + (u^{(n-1)}(t))^2 \right)$$

und untersuchen deren Zeit-Ableitung:

$$\begin{aligned} E' &= uu' + u'u'' + \dots + u^{(n-1)}u^{(n)} \\ &\leq |u| \cdot |u'| + |u'| \cdot |u''| + \dots + |u^{(n-1)}| \cdot |u^{(n)}|. \end{aligned} \quad (4.4)$$

Für sämtliche reellen Zahlen  $p$  und  $q$  gilt die YOUNG'sche Ungleichung  $|p| \cdot |q| \leq \frac{1}{2}(p^2 + q^2)$ , was wir bald sehr oft gebrauchen können. Wir dürfen auch noch annehmen, daß die Koeffizientenfunktionen  $a_0, a_1, \dots, a_{n-1}$  im Betrag durch eine (notfalls große) Zahl  $A$  beschränkt sind:

$$|a_j(t)| \leq A \quad \forall j, \quad \forall t.$$

Dann ist (wegen  $Lu = 0$ ) auch  $|u^{(n)}(t)| = |\sum_{j=0}^{n-1} a_j(t)u^{(j)}(t)| \leq A \sum_{j=0}^{n-1} |u^{(j)}(t)|$ . Wir können hier (Spiralprinzip) die Ungleichung von Cauchy-Schwarz im  $\mathbb{R}^n$  verwenden:

$$\sum_{j=0}^{n-1} |u^{(j)}(t)| = \left\langle \begin{pmatrix} |u(t)| \\ \vdots \\ |u^{(n-1)}(t)| \end{pmatrix}, \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \right\rangle \leq \sqrt{2E(t)} \cdot \sqrt{1^2 + \dots + 1^2} = \sqrt{2E(t)} \cdot \sqrt{n},$$

also  $|u^{(n)}(t)| \leq \sqrt{n}A\sqrt{2E(t)}$ . Weiterhin ist  $|u^{(n-1)}(t)| \leq \sqrt{2E(t)}$ . Somit erhalten wir aus (4.4) und der Ungleichung von YOUNG

$$\begin{aligned} E' &\leq \frac{1}{2} (u^2 + (u')^2) + \frac{1}{2} ((u')^2 + (u'')^2) + \dots + \frac{1}{2} ((u^{(n-2)})^2 + (u^{(n-1)})^2) + \sqrt{2E} \cdot \sqrt{n}A\sqrt{2E} \\ &\leq 2E + 2\sqrt{n}AE. \end{aligned}$$

Zur einfacheren Schreibweise setzen wir  $\alpha := 2 + 2\sqrt{n}A$  und bekommen also  $E'(t) \leq \alpha E(t)$  für alle  $t \in \mathbb{R}$ .

Ein freundlicher Geist erscheint uns im Traum und gibt uns den Hinweis, die Ableitung der Funktion  $t \mapsto E(t)e^{-\alpha t}$  zu untersuchen:

$$(E(t)e^{-\alpha t})' = E'(t)e^{-\alpha t} + E(t) \cdot (-\alpha)e^{-\alpha t} \leq \alpha E(t)e^{-\alpha t} - \alpha E(t)e^{-\alpha t} = 0 \leq 0.$$

Also ist diese Funktion schwach monoton fallend, und somit ist, für  $t \geq 0$ ,

$$E(t)e^{-\alpha t} \leq E(0)e^{-\alpha \cdot 0}.$$

Nun ist allerdings  $E(0) = 0$ , also auch  $E(t) \leq 0$  für  $t \geq 0$ . Weiterhin zeigt uns die Definition der Funktion  $E$ , daß  $E$  gar nicht negativ werden kann. Damit ist also  $E(t) = 0$  für jedes  $t \geq 0$ .

Das erzwingt dann aber  $u(t) = 0$  für alle  $t \geq 0$ , nach Definition von  $E$ . Analog argumentiert man für negative  $t$ .  $\square$

Damit ist die Injektivität der verallgemeinerten Anfangswertabbildung fast schon gezeigt:

**Satz 4.37.** *Die Abbildung*

$$\begin{aligned} \text{AWA: } \ker L &\rightarrow \mathbb{R}^n, \\ \text{AWA: } u &\mapsto (u(0), u'(0), \dots, u^{(n-1)}(0)) \end{aligned}$$

ist injektiv und es ist  $\dim \ker L \leq n$ .

*Beweis.* Es seien  $u$  und  $\tilde{u}$  aus  $\ker L$  mit  $\text{AWA } u = \text{AWA } \tilde{u}$ . Sei  $z = u - \tilde{u}$ . Dann ist  $z \in \ker L$  und  $(z(0), z'(0), \dots, z^{(n-1)}(0)) = (0, 0, \dots, 0)$ . Nach Lemma 4.36 haben wir  $z(t) = 0$  für jedes  $t \in \mathbb{R}$ , also  $u = \tilde{u}$ . Also ist die Anfangswertabbildung injektiv.

Die Dimensionsaussage folgt aus Satz 4.15, Teil 4, denn  $\dim \mathbb{R}^n = n$ .  $\square$

Unsere jetzigen Kenntnisse reichen nicht aus, um die Surjektivität der Anfangswertabbildung im allgemeinen Fall zu zeigen. Wir beschränken uns stattdessen auf den Spezialfall der Schwingungsgleichung.

**Satz 4.38.** *Sei  $U = C^2(\mathbb{R} \rightarrow \mathbb{R})$ ,  $V = C(\mathbb{R} \rightarrow \mathbb{R})$  und  $\omega, k \in \mathbb{R}$ . Dann ist die Abbildung*

$$\begin{aligned} \text{AWA: } \ker \left( \frac{d^2}{dt^2} + k \frac{d}{dt} + \omega^2 \right) &\rightarrow \mathbb{R}^2, \\ \text{AWA: } u &\mapsto (u(0), u'(0)) \end{aligned}$$

surjektiv.

*Beweis.* Wir zeigen  $\dim \ker L = 2$ . Dann ist die Anfangswertabbildung eine injektive Abbildung von einem zweidimensionalen Raum in einen zweidimensionalen Raum, nach Satz 4.19 also auch surjektiv.

Weil  $u \in \ker L$ , erfüllt  $u$  die Differentialgleichung

$$u''(t) + ku'(t) + \omega^2 u(t) = 0.$$

Im Rest des Beweises konstruieren wir zwei linear unabhängige Lösungen  $u_+, u_- \in C^2(\mathbb{R} \rightarrow \mathbb{R})$ , also zwei linear unabhängige Elemente von  $\ker L$ . Dann ist die Dimension von  $\ker L$  also mindestens gleich 2, nach Satz 4.37 also genau gleich 2.

Wir machen für die Funktion  $u$  den Ansatz

$$u(t) = e^{\lambda t}, \quad t \in \mathbb{R},$$

mit wählbarem Parameter  $\lambda$ . Wenn wir diesen Ansatz in die homogene Gleichung einsetzen, bekommen wir

$$e^{\lambda t} (\lambda^2 + k\lambda + \omega^2) = 0.$$

Weil die Exponentialfunktion nie Null wird, müßte  $\lambda^2 + k\lambda + \omega^2 = 0$  sein, also

$$\lambda = \lambda_{1,2} = -\frac{k}{2} \pm \sqrt{\frac{k^2}{4} - \omega^2}.$$

**Fall 1:**  $\frac{k^2}{4} - \omega^2 > 0$

Dann sind  $\lambda_1, \lambda_2 \in \mathbb{R}$  und verschieden, also sind die Funktionen  $u_+(t) := e^{\lambda_1 t}$  und  $u_-(t) := e^{\lambda_2 t}$  linear unabhängig. Wir haben also zwei linear unabhängige Lösungen  $u$  der homogenen Schwingungsgleichung gefunden.

**Fall 2:**  $\frac{k^2}{4} - \omega^2 = 0$

Dann ist  $\lambda_1 = \lambda_2$  und die beiden Funktionen  $t \mapsto e^{\lambda_1 t}$  und  $t \mapsto e^{\lambda_2 t}$  sind linear abhängig.

Man rechnet aber schnell nach, daß eine weitere Lösung gegeben wird durch  $t \mapsto te^{\lambda_1 t}$ . Damit sind erneut zwei linear unabhängige Lösungen der Schwingungsgleichung gefunden.

**Fall 3:**  $\frac{k^2}{4} - \omega^2 < 0$

In diesem Fall sind  $\lambda_1$  und  $\lambda_2$  nicht reell. Wir haben stattdessen

$$\lambda_{1,2} = \mu \pm i\nu, \quad \mu = -\frac{k}{2}, \quad \nu = \sqrt{\left| \frac{k^2}{4} - \omega^2 \right|},$$

und dann gilt

$$e^{\lambda_1 t} = e^{\mu t} (\cos(\nu t) + i \sin(\nu t)), \quad e^{\lambda_2 t} = e^{\mu t} (\cos(\nu t) - i \sin(\nu t)).$$

Man rechnet nach, daß zwei linear unabhängige Lösungen der Schwingungsgleichung gegeben werden durch

$$u_+(t) := e^{\mu t} \cos(\nu t), \quad u_-(t) := e^{\mu t} \sin(\nu t).$$

In allen drei Fällen kamen wir auf  $\dim \ker L \geq 2$ , und der Beweis ist damit beendet.  $\square$

## 4.6 Ausblick: Lineare Abbildungen in der Physik

Die üblichen Differentialoperatoren sind typische Beispiele für lineare Abbildungen zwischen Funktionenvektorräumen:

**Gradient:** für ein Skalarfeld  $\varphi: \mathbb{R}^3 \rightarrow \mathbb{R}$  ist der Gradient definiert als

$$\nabla \varphi = \left( \frac{\partial \varphi}{\partial x_1}, \frac{\partial \varphi}{\partial x_2}, \frac{\partial \varphi}{\partial x_3} \right).$$

Die Abbildung  $\varphi \mapsto \nabla \varphi$  ist linear. Wir haben z.B.

$$\nabla \stackrel{?}{\in} \text{Hom} (L^2(\mathbb{R}^3 \rightarrow \mathbb{R}) \rightarrow L^2(\mathbb{R}^3 \rightarrow \mathbb{R}^3)),$$

dabei sollten wir aber aufpassen wegen des Definitionsbereiches von  $\nabla$ : dieser Operator kann nur auf solche Funktionen  $\varphi$  aus dem  $L^2(\mathbb{R}^3 \rightarrow \mathbb{R})$  angewandt werden, für die tatsächlich  $\nabla \varphi$  im Bildraum  $L^2(\mathbb{R}^3 \rightarrow \mathbb{R}^3)$  liegt. (Welche Funktionen  $\varphi$  das sind, wollen wir hier nicht weiter erörtern. Wir deuten lediglich diese Baustelle (bzw. dieses Minenfeld) durch ein Fragezeichen an.) Mit den Bezeichnungen der Helmholtz-Zerlegung haben wir dann

$$\text{img } \nabla = L^2_{\text{pot}}(\mathbb{R}^3 \rightarrow \mathbb{R}^3).$$

**Divergenz:** für ein Vektorfeld  $\vec{u}: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  ist die Divergenz definiert als

$$\operatorname{div} \vec{u} = \frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3}.$$

Die Abbildung  $\vec{u} \mapsto \operatorname{div} \vec{u}$  ist linear. Wir haben z.B.

$$\operatorname{div} \in \overset{?}{\operatorname{Hom}}(L^2(\mathbb{R}^3 \rightarrow \mathbb{R}^3) \rightarrow L^2(\mathbb{R}^3 \rightarrow \mathbb{R})),$$

und auch hier sollten wir aufpassen wegen des Definitionsbereiches von  $\operatorname{div}$ . Mit den Bezeichnungen der Helmholtz-Zerlegung haben wir dann

$$\ker \operatorname{div} = L^2_{\operatorname{div}}(\mathbb{R}^3 \rightarrow \mathbb{R}^3).$$

**Laplace-Operator:** Dieser ist definiert als  $\Delta = \operatorname{div} \circ \nabla$ . Aufgrund von Satz 4.9, Teil 2, ist er dann automatisch linear.

Lineare Differentialoperatoren haben die wunderschöne physikalische Eigenschaft, daß für ihre Lösungen das Superpositionsprinzip gilt.

**P und Q:** Wir betrachten ein einzelnes spinloses Teilchen. Sein quantenmechanischer Zustand wird beschrieben durch eine Wellenfunktion  $\psi = \psi(t, x): \mathbb{R}_t \times \mathbb{R}_x^3 \rightarrow \mathbb{C}$ . Diese liegt (für jede Zeit  $t$ ) im Vektorraum  $L^2(\mathbb{R}^3 \rightarrow \mathbb{C})$ , der ausgestattet ist mit dem Skalarprodukt  $\langle \cdot, \cdot \rangle_{L^2}$  wie in (2.4).

Wenn ein Teilchen einen Spin hat (mit Wert  $+\frac{1}{2}$  oder  $-\frac{1}{2}$ ), dann brauchen wir zwei Wellenfunktionen zur Beschreibung dieses Teilchens, also haben wir  $\psi = (\psi_+, \psi_-) \in (L^2(\mathbb{R}^3 \rightarrow \mathbb{C}))^2$ , und das Skalarprodukt ist dann  $\langle \varphi, \psi \rangle := \langle \varphi_+, \psi_+ \rangle_{L^2} + \langle \varphi_-, \psi_- \rangle_{L^2}$ .

Bei  $M$  Teilchen (ohne Spin) nehmen wir als Vektorraum den  $L^2(\mathbb{R}^{3M} \rightarrow \mathbb{C})$  mit naheliegenderem Skalarprodukt, und wenn diese Teilchen miteinander wechselwirken oder Spins haben können, dann wird es richtig kompliziert.

Alle diese Vektorräume sind dem jeweiligen physikalischen System zugeordnet und heißen *Systemhilberträume*. Ihre Dimension ist praktisch immer gleich unendlich.

Nun schauen wir uns Meßgrößen an, wie z.B. Ort, Impuls, Energie. Jeder solchen Meßgröße (Observable) entspricht eine lineare Abbildung, man sagt auch *linearer Operator*. Diese Operatoren bilden den Systemhilbertraum  $\mathcal{H}$  (bzw. einen Untervektorraum davon) in  $\mathcal{H}$  ab (bzw. in den  $\mathcal{H} \times \mathcal{H} \times \mathcal{H}$ , denn wir leben ja in einer dreidimensionalen Welt).

**Ortsoperator Q:** Es ist  $Q: \psi \mapsto Q\psi$ , wobei  $(Q\psi)(x) = x\psi(x)$ . Wegen  $x = (x_1, x_2, x_3)^T \in \mathbb{R}^3$  zerlegen wir  $Q = (Q_1, Q_2, Q_3)$  mit  $(Q_j\psi)(x) = x_j\psi(x)$ .

**Impulsoperator P:** Es ist  $P: \psi \mapsto P\psi$ , wobei

$$(P\psi)(x) = \frac{\hbar}{i} \nabla \psi(x),$$

und  $\hbar = \frac{h}{2\pi}$  ist das Planck-Wirkungsquantum, mit  $h = 6.6260755 \cdot 10^{-34}$  Js. Wir zerlegen analog zu  $Q$  auch  $P$  als  $P = (P_1, P_2, P_3)$ .

**Operator zur potentiellen Energie:** Sei  $V = V(x)$  ein Potential (für das Wasserstoffatom z.B. das Coulomb-Potential), dann gehört dazu ein Multiplikationsoperator  $M_V: \psi \mapsto M_V\psi$ , wobei  $(M_V\psi)(x) = V(x)\psi(x)$ .

**Operator zur Gesamtenergie:** Wir ersetzen  $E$  durch  $i\hbar \frac{\partial}{\partial t}$ .

Die Energiegleichung  $E = E_{\text{kin}} + E_{\text{pot}} = \frac{p^2}{2m} + V$  wird dann zu

$$i\hbar \frac{\partial}{\partial t} \psi(t, x) = -\frac{\hbar^2}{2m} \Delta \psi(t, x) + V(x)\psi(t, x),$$

was die berühmte Schrödingergleichung ist.

Die Wellenfunktion  $\psi$  ist als solche meßtechnisch nicht auffindbar. Messen kann man nur die üblichen Größen für ein Teilchen: Aufenthaltsort, Impuls, Energie, Drehimpuls usw., und Quantenmechanisch sind diese Meßwerte dann *Eigenwerte* der entsprechenden Operatoren. Für den Ortsoperator  $Q_1$  wäre ein solcher Eigenwert also eine Zahl  $\hat{x}_1 \in \mathbb{R}$  mit  $Q_1\psi = \hat{x}_1\psi$ . Diese Eigenwerte zu Operatoren



entsprechen genau den Eigenwerten zu selbstadjungierten Matrizen, wie wir sie im zweiten Semester behandeln werden.

Zum Abschluß erinnern wir daran, daß die Matrizenmultiplikation nicht kommutativ ist. Etwas ähnliches beobachten wir auch hier: die Operatoren  $P$  und  $Q$  vertauschen nicht miteinander, denn es ist  $P_j \circ Q_j \neq Q_j \circ P_j$ , für  $j = 1, 2, 3$ . Die physikalische Konsequenz davon ist die *Unschärferelation von Heisenberg*, die besagt, daß man für ein Teilchen seinen Aufenthaltsort und seinen Impuls nicht gleichzeitig beliebig genau messen kann.

Wir verlassen den Zoo der Differentialoperatoren und schauen uns einige physikalische Größen an:

**Wärmeleitungstensor:** Wir haben ein unbewegliches Medium, in dem Wärme strömen kann, z.B. einen Festkörper (aber kein Gas oder Fluid). Die Temperatur am Ort  $x$  nennen wir  $T(x)$ , was ein skalares Feld ergibt. Der Gradient davon, also  $\nabla T(x)$ , ist ein Vektor, der am Punkt  $x$  angeheftet ist und in Richtung des steilsten Temperaturanstiegs zeigt. Dieser Temperaturunterschied führt normalerweise zu einem Wärmefluß, der durch ein Vektorfeld beschrieben wird, das wir  $j$  nennen. Im Allgemeinen gilt dann

$$j = -\lambda \nabla T,$$

wobei  $\lambda$  meist eine Zahl ist, die die Wärmeleitfähigkeit des Stoffes beschreibt. Es ist aber auch möglich, daß der Stoff in unterschiedlichen Richtungen die Wärme unterschiedlich gut leitet. Zum Beispiel ist es denkbar, daß Holz die Wärme besser parallel zur Faser leitet als quer dazu. Oder vielleicht liegt ein Kristall vor, bei dem das Verhalten entlang der verschiedenen Achsen unterschiedlich ist. Dann brauchen  $j$  und  $\nabla T$  nicht mehr parallel zueinander sein, und  $\lambda$  ist eine Matrix, die auch *Wärmeleitfähigkeitstensor* genannt wird. Wir stellen uns vor, daß dieser Tensor die „Ursache“  $\nabla T$  auf die „Wirkung“  $j$  abbildet. Der isotrope Fall (daß alle Richtungen gleichberechtigt sind) ist mit enthalten, wenn wir für  $\lambda$  ein Vielfaches der Einheitsmatrix erlauben.

**Verzerrungstensor:** Ein Festkörper wird irgendwie belastet und verformt sich ein wenig. Das kann man sich so vorstellen, daß das Teilchen, was im Ruhezustand an der Position  $x$  war, durch die Verformung gewandert ist an die Stelle  $x + u(x)$ . Der Vektor  $u(x)$  heißt *Verschiebungsvektor*; wir stellen ihn uns als kurz vor. Bei einer Verschiebung des gesamten Körpers ist  $u$  konstant, was uns nicht interessiert. Bei einer Drehung des gesamten Körpers ist (nach einer geeigneten Wahl des Koordinatensystems)  $u(x) = Ax$ , wobei  $A$  eine Drehmatrix aus der Gruppe  $SO(3)$  ist. Das interessiert uns auch nicht, denn wir wollen nur die Verzerrungen beschreiben, weshalb wir voraussetzen, daß unser Körper solche Bewegungen nicht ausführt. (Das typische Beispiel ist ein Stab, an dessen beiden Enden gezogen wird, sodaß er länger wird, gleichzeitig aber dünner.) Unter dieser Voraussetzung bekommen wir, daß ein kurzer Verbindungsvektor  $\vec{z}$  zweier Punkte abgebildet wird auf einen Vektor  $\varepsilon \vec{z}$ , wobei  $\varepsilon$  der sogenannte *Verzerrungstensor* ist, der den alten Verbindungsvektor vor der Deformation zum neuen Verbindungsvektor nach der Deformation sendet.

**Spannungstensor:** Wir haben einen verformten Festkörper und wählen in seinem Innern ein kleines Flächenstück, und auf einer Seite davon denken wir uns das Material gedanklich markiert. Dieses gedanklich markierte Material kann auf das übrige Material eine Kraft ausüben, die über das kleine Flächenstück vermittelt wird. Diese Kraft kann senkrecht zur Fläche drücken (Druckspannung), oder senkrecht zur Fläche ziehen (Zugspannung) oder tangential zur Fläche schieben (Schubspannung), oder alles vermischt bewirken. Der Spannungstensor  $\sigma$  ist derjenige Tensor (also diejenige Matrix), der den Außennormalenvektor des kleinen Flächenstücks auf den Kraftvektor abbildet.

**Elastizitätstensor:** Wir stellen uns vor, daß zwischen Verzerrung und Spannung ein linearer Zusammenhang besteht, der dann wie folgt aussieht:

$$\sigma = C\varepsilon.$$

Hierbei sind  $\sigma$  und  $\varepsilon$  die obigen Tensoren zweiter Stufe mit jeweils  $3 \times 3 = 9$  Einträgen, also ist  $C$  ein Tensor vierter Stufe mit  $9 \times 9 = 81$  Einträgen, der *Elastizitätstensor* genannt wird. Aus Symmetriegründen sind viele von diesen Einträgen gleich, sodaß nur 21 Materialkonstanten übrigbleiben. Und für einen isotropen Festkörper sind es nur zwei Parameter, und es entsteht:

$$C_{ijkl} = K\delta_{ij}\delta_{kl} + G\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk} - \frac{2}{3}\delta_{ij}\delta_{kl}\right).$$

Wir nennen  $K$  den *Kompressionsmodul* und  $G$  den *Schermodul*.

Wenn dieser Festkörper jetzt beginnt zu schwingen, dann bekommen wir die Differentialgleichung für die Verschiebungsvektorfunktion  $\vec{u} = \vec{u}(t, x)$ :

$$\varrho \frac{\partial^2 \vec{u}}{\partial t^2} = G \Delta \vec{u} + \left( K + \frac{G}{3} \right) \nabla \operatorname{div} \vec{u},$$

wobei  $\varrho$  die Materialdichte bezeichnet.

**Knobelaufgabe:** Man zerlege mittels Helmholtz–Projektion das Verschiebungsvektorfeld  $\vec{u}$  in einen longitudinalen und einen transversalen Anteil. Man beobachte, wie diese Zerlegung die obige komplizierte Differentialgleichung erheblich vereinfacht. Man errate die Ausbreitungsgeschwindigkeiten der beiden Wellentypen.

## 4.7 Ausblick: Vektorräume in der Physik 2

Sei  $U$  ein Vektorraum über dem Körper  $K$ . Die Menge aller linearen Abbildungen von  $U$  in den Vektorraum  $K^1$  heißt *Dualraum von  $U$* . Dieser Dualraum ist seinerseits wiederum ein Vektorraum über dem Körper  $K$ . Man stelle sich vor, daß ein Element des Dualraumes nichts anderes tut, als auf Vektoren aus  $U$  zu warten, sie aufzufressen und Zahlen daraus zu machen.

Als Beispiel für  $U$  betrachten wir den  $\mathbb{R}^3$ , dann ist  $K = \mathbb{R}$ . Die Elemente von  $U$  nennen wir  $x = (x_1, x_2, x_3)^\top$  als Spaltenvektoren. Wir stellen uns diese Vektoren als Ortsvektoren vor, mit Maßeinheit *Meter*. Der Dualraum dazu besteht aus Zeilenvektoren  $k = (k_1, k_2, k_3)$  mit Maßeinheit  $\frac{1}{\text{Meter}}$ . Jedes solche  $k \in \mathbb{R}_k^3$  erzeugt eine Abbildung von  $\mathbb{R}_x^3$  in den einheitenlosen  $\mathbb{R}$  gemäß  $kx = k_1x_1 + k_2x_2 + k_3x_3$ . Diese Vektoren  $k$  nennt man gelegentlich *Wellenzahlvektoren*.

Als typisches Beispiel denken wir an einen triklinen Kristall. Dieser hat 3 Achsen, deren Winkel zueinander keine rechten Winkel sind, und die charakteristischen Kantenlängen sind unterschiedlich lang. Auf diesem Wege bekommen wir eine Gitterstruktur, und an jedem Gitterknotenpunkt sitzt eine Elementarzelle. Wir führen ein Koordinatensystem ein, wobei die Basisvektoren entlang der Gitterlinien zeigen. Die Basisvektorenlänge ist genau gleich dem Abstand zweier benachbarter Gitterpunkte entlang einer Gitterlinie.

Wir bekommen auf diesem Wege drei Basisvektoren  $b_1, b_2, b_3$ , die *keine* ONB des  $\mathbb{R}^3$  bilden. Jeder Vektor  $x$  kann dann geschrieben werden als  $x = \xi^1 b_1 + \xi^2 b_2 + \xi^3 b_3$  (hierbei hat es sich eingebürgert, die Indizes an die Koordinaten oben zu schreiben, nicht unten). Wenn wir noch einen weiteren Vektor  $y = \eta^1 b_1 + \eta^2 b_2 + \eta^3 b_3$  haben und uns für das Skalarprodukt  $x \cdot y$  interessieren, dann ist dieses leider nicht gleich  $\xi^1 \eta^1 + \xi^2 \eta^2 + \xi^3 \eta^3$ . Stattdessen entstehen weitere Ärgersummanden aus gemischten Produkten, eben weil keine ONB vorliegt.

Als Ausweg betrachten wir eine weitere Basis  $(b^1, b^2, b^3)$ , diesmal für den Dualraum, mit Indizes oben statt unten. Wir verlangen, daß für die Skalarprodukte mit den Elementen der alten Basis gilt:

$$b^j \cdot b_k = \delta_k^j, \quad j, k = 1, 2, 3,$$

wobei rechts das KRONECKERSymbol steht. Wenn wir jetzt  $y$  in der dualen Basis (man nennt sie auch *reziproke Basis*) entwickeln,  $y = \eta_1 b^1 + \eta_2 b^2 + \eta_3 b^3$ , dann ist tatsächlich  $x \cdot y = \xi^1 \eta_1 + \xi^2 \eta_2 + \xi^3 \eta_3$ .

Die reziproke Basis erzeugt in natürlicher Weise ein neues Gitter, das in Bezug auf das Kristallgitter das *reziproke Gitter* genannt wird. Für die Untersuchung von z.B. Gitterschwingungen spielt dieses reziproke Gitter eine ganz wichtige Rolle.

Wir rechnen damit noch ein bißchen. Als Hauptregeln halten wir dabei fest:

- Punkte  $\cdot$  bezeichnen das Skalarprodukt von Vektoren,
- Größen in lateinischen Buchstaben sind Vektoren (abgesehen von Indizes),
- Größen in griechischen Buchstaben sind reelle Zahlen,
- alle Summationen laufen von 1 bis 3 (die Summenkonvention von Einstein wird hier nicht praktiziert, da im ersten Semester vielleicht zu verwirrend),
- Größen mit unteren Indizes heißen *kovariant*, mit oberen Indizes heißen sie *kontravariant* (das ist im allgemeinen Fall nicht ganz korrekt, in der hier vorliegenden Situation aber richtig. In Wirklichkeit beziehen sich die Begriffe „kovariant“ und „kontravariant“ darauf, nach welcher Formel sich die betreffenden Größen transformieren bei einem Basiswechsel).

Gegeben sei jetzt eine kovariante Basis  $(b_1, b_2, b_3)$  für den  $\mathbb{R}^3$ , und wir suchen die dazugehörige kontravariante Basis  $(b^1, b^2, b^3)$ . Das definierende Gleichungssystem ist eindeutig lösbar, sodaß wir schon wissen, daß es diese kontravariante Basis tatsächlich gibt, es fehlt bloß noch der Rechenweg. Wir setzen

$$\beta_{jk} := b_j \cdot b_k, \quad \beta^{jk} := b^j \cdot b^k, \quad j, k = 1, 2, 3.$$

Die *kovarianten Metrik-Koeffizienten*  $\beta_{jk}$  sind bekannt, die *kontravarianten Metrik-Koeffizienten*  $\beta^{jk}$  sind es noch nicht.

Wir basteln uns einige Vektoren als Spielzeug, nämlich  $c^j := \sum_l \beta^{jl} b_l$ . Dann ist

$$c^j \cdot b^k = \sum_l \beta^{jl} b_l \cdot b^k = \sum_l \beta^{jl} \delta_l^k = \beta^{jk} = b^j \cdot b^k.$$

Wir haben also  $(c^j - b^j) \perp b^k$  für jedes  $k$ , demnach muß  $c^j - b^j = 0 \in \mathbb{R}^3$  sein, also  $c^j = b^j$ . Das bedeutet

$$b^j = \sum_l \beta^{jl} b_l,$$

und jetzt brauchen wir die  $c^j$  nicht mehr. Wenn es uns gelingt, die  $\beta^{jl}$  auszurechnen, dann haben wir die gesuchte kontravariante Basis  $(b^1, b^2, b^3)$ . Dazu beobachten wir, daß

$$\sum_l \beta^{jl} \beta_{lk} = \sum_l \beta^{jl} b_l \cdot b_k = b^j \cdot b_k = \delta_k^j,$$

was nichts anderes bedeutet, als daß die Matrix der  $\beta^{jl}$  und die Matrix der  $\beta_{jl}$  invers zueinander sind. Und Matrizen invertieren können wir ja.

Als nächstes Projekt bestimmen wir die Koordinaten eines Vektors. Sei  $x$  gegeben (die beiden Basen natürlich auch), und die Koordinaten  $\xi^j$  und  $\xi_j$  gemäß  $x = \sum_j \xi^j b_j$  sowie  $x = \sum_j \xi_j b^j$  seien gesucht. Man verifiziert schnell, daß die Lösung gegeben wird durch folgende Formeln:

$$\xi^j = x \cdot b^j, \quad \xi_j = x \cdot b_j, \quad j = 1, 2, 3.$$

Und als Schlußprojekt betrachten wir Basistransformationen: sei mit  $(\bar{b}_1, \bar{b}_2, \bar{b}_3)$  eine weitere kovariante Basis gegeben:

$$\bar{b}_j := \sum_k \underline{\alpha}_j^k b_k, \tag{4.5}$$

und wir wollen wissen, wie die Transformationsformeln für die anderen Größen sind. Die Unter- und Oberstriche bedeuten keinerlei Konjugationen oder ähnliches, sondern sollen lediglich das Einführen weiterer Buchstaben vermeiden.

Die umgekehrte Transformation ist dann

$$b_k = \sum_l \bar{\alpha}_k^l \bar{b}_l$$

mit noch unbekanntem  $\bar{\alpha}_k^l$ , die sich aber bestimmen lassen durch Einsetzen und Koeffizientenvergleich:

$$b_k = \sum_{l,m} \bar{\alpha}_k^l \underline{\alpha}_l^m b_m,$$

also muß  $\sum_l \bar{\alpha}_k^l \underline{\alpha}_l^m = \delta_k^m$  sein, und demnach sind die Matrizen der  $\underline{\alpha}_k^l$  und  $\bar{\alpha}_k^l$  invers zueinander. Dann muß aber auch die Gleichung  $\sum_l \underline{\alpha}_k^l \bar{\alpha}_l^m = \delta_k^m$  gelten, denn aus  $AA^{-1} = I$  folgt  $A^{-1}A = I$ .

Nun suchen wir Formeln für die reziproke Basis  $(\bar{b}^1, \bar{b}^2, \bar{b}^3)$ . Mit Phantasie, gestützt durch Probieren anhand echter Zahlen (oder durch einen Ansatz), kommt man zur Vermutung, daß  $\bar{b}^j = \sum_k \bar{\alpha}_k^j b^k$  sein könnte. Wir testen dies anhand folgender Rechnung:

$$\bar{b}_j \cdot \sum_k \bar{\alpha}_k^j b^k = \left( \sum_m \underline{\alpha}_j^m b_m \right) \cdot \left( \sum_k \bar{\alpha}_k^j b^k \right) = \sum_{m,k} \underline{\alpha}_j^m \bar{\alpha}_k^j b_m \cdot b^k = \sum_{m,k} \underline{\alpha}_j^m \bar{\alpha}_k^j \delta_m^k = \sum_m \underline{\alpha}_j^m \bar{\alpha}_m^j = \delta_j^j,$$

also gilt tatsächlich

$$\bar{b}^l = \sum_k \bar{\alpha}_k^l b^k. \quad (4.6)$$

Für die Transformation von Koordinaten eines Vektors  $x$  haben wir  $x = \sum_l \bar{\xi}^l \bar{b}_l$  mit

$$\bar{\xi}^l = x \cdot \bar{b}^l = x \cdot \sum_k \bar{\alpha}_k^l b^k,$$

woraus wir

$$\bar{\xi}^l = \sum_k \bar{\alpha}_k^l \xi^k \quad (4.7)$$

bekommen. Und für die kovarianten Koordinaten von  $x$  haben wir  $x = \sum_j \bar{\xi}_j \bar{b}^j$  mit

$$\bar{\xi}_j = x \cdot \bar{b}_j = x \cdot \sum_k \underline{\alpha}_j^k b_k$$

mit der Konsequenz

$$\bar{\xi}_j = \sum_k \underline{\alpha}_j^k \xi_k. \quad (4.8)$$

Wir erkennen, daß (4.5) und (4.8) sehr ähnlich aussehen: die Transformation erfolgt mittels der  $\underline{\alpha}$ . Sämtliche Größen, die sich so transformieren, bezeichnet man als *kovariante Tensoren*.

Und wir erkennen weiterhin, daß auch (4.6) und (4.7) einander sehr ähneln: die Transformation erfolgt mittels der  $\bar{\alpha}$ . Sämtliche Größen, die sich nach diesem Schema transformieren bei Basiswechsel, nennt man *kontravariante Tensoren*.

Das waren die Anfangsgründe der Tensorrechnung, mehr Informationen findet man z.B. in KLINGBEIL, *Tensorrechnung für Ingenieure*.

## 4.8 Schlüsselbegriffe

- Definition linearer Abbildungen, Differentialoperatoren als lineare Abbildungen,
- Begriffe injektiv, surjektiv, bijektiv, Isomorphismus,
- Kern, Bild, Rang,
- Dimensionsformel und Folgerungen daraus,
- Struktur der Lösungsmenge linearer Gleichungssysteme,
- ähnliche Matrizen und deren Beziehung zu Basistransformationen,
- Lösungsalgorithmus für lineare gewöhnliche Differentialgleichungen mit konstanten Koeffizienten, anhand der Schwingungsdifferentialgleichung.

# Kapitel 5

## Normierte Räume, Reelle Zahlen, Folgen, Reihen

Die bisherigen Kapitel 2,3,4 behandelten die Lineare Algebra, deren Leitbegriffe *Vektorraum* und *lineare Abbildung* lauten. In diesem und dem nächsten Kapitel soll es nun um die Analysis gehen, insbesondere wollen wir uns dann mit dem Differenzieren und Integrieren vertraut machen. Diese Techniken beruhen auf dem Begriff des *Grenzwerts*, der also die Leitidee der Analysis ist. Wir entfernen uns aber nicht übermäßig weit von der linearen Algebra, denn auch in Vektorräumen kann man Folgen und deren Grenzwerte anschauen.

### 5.1 Folgen im $\mathbb{R}^d$ und $\mathbb{C}^d$

Aus der Schule ist (so ungefähr) bekannt, was darunter zu verstehen ist, wenn man sagt, daß eine Folge  $(a_1, a_2, \dots)$  reeller Zahlen gegen einen Grenzwert  $a^*$  konvergiert. Wir wollen den Konvergenzbegriff auf die Räume  $\mathbb{R}^d$  und  $\mathbb{C}^d$  übertragen; und deshalb führen wir in diesen Räumen eine Norm ein, um Abstände messen zu können.

**Definition 5.1.** Im  $\mathbb{R}^d$  und im  $\mathbb{C}^d$  definieren wir die pythagoräischen Normen  $\|\cdot\|_{\mathbb{R}^d}$  und  $\|\cdot\|_{\mathbb{C}^d}$  mittels

$$\|x\|_{\mathbb{R}^d} := \sqrt{x_1^2 + \dots + x_d^2}, \quad \|z\|_{\mathbb{C}^d} := \sqrt{|z_1|^2 + \dots + |z_d|^2}.$$

Im Folgenden schreiben wir immer  $K^d$  mit  $K = \mathbb{R}$  oder  $K = \mathbb{C}$ .

Diese Normen besitzen die gleichen Eigenschaften wie in Satz 2.26 beschrieben. Unter dem Abstand zweier Punkte verstehen wir die Norm von deren Differenzvektor.

Unter einer Folge im Raum  $K^d$  verstehen wir einen Ausdruck der Form  $(a_1, a_2, a_3, \dots) =: (a_n)_{n \in \mathbb{N}}$  mit  $a_n \in K^d$ . Mathematisch präzise werden wir den Begriff dann in Definition 5.16 festlegen.

**Definition 5.2.** Eine Folge  $(a_n)_{n \in \mathbb{N}} \subset K^d$  heißt konvergent mit Grenzwert  $a^*$ <sup>1</sup>, wenn

$$\lim_{n \rightarrow \infty} \|a_n - a^*\|_{K^d} = 0.$$

Das ist mathematisch definiert als:

$$\forall \varepsilon > 0 \exists N_0(\varepsilon) : \forall n \geq N_0(\varepsilon) : \|a_n - a^*\|_{K^d} < \varepsilon.$$

Wir lesen diese Formelzeile folgendermaßen: „Für jedes<sup>2</sup> positive  $\varepsilon$  existiert ein  $N_0$  (das von  $\varepsilon$  abhängen darf), sodaß für jedes  $n \geq N_0$  gilt, daß  $\|a_n - a^*\|_{K^d}$  kleiner als  $\varepsilon$  ist“.

Zu verstehen ist diese Formelzeile wie folgt:

Wir können den Abstand von  $a_n$  und  $a^*$  beliebig klein bekommen ( $\|a_n - a^*\|_{K^d} < \varepsilon$ )  
— und zwar beliebig klein ( $\forall \varepsilon > 0$ ) —  
wenn wir nur vorher sicherstellen, daß  $n$  genügend groß ( $\exists N_0(\varepsilon) : \forall n \geq N_0(\varepsilon)$ ) ist.  
Dabei sind alle  $a_n$  mit  $n \geq N_0(\varepsilon)$  höchstens  $\varepsilon$  von  $a^*$  entfernt, nicht bloß einige  $a_n$  ( $\forall n \geq N_0(\varepsilon)$ ).

<sup>1</sup>convergent with limit  $a^*$

<sup>2</sup> die gelegentliche Lesung für alle anstelle für jedes ist grammatisch und inhaltlich falsch.

Eine Merkgel für den Konvergenzbegriff ist:

*Ein Element  $a^*$  ist Grenzwert einer Folge  $(a_n)_{n \in \mathbb{N}}$ ,  
wenn es zu jeder Umgebung von  $a^*$  ein  $N_0$  gibt,  
sodaß das Folgenendstück ab  $N_0$  in dieser Umgebung liegt.*

Eine „Umgebung von  $a^*$ “ ist ein Intervall mit Mittelpunkt  $a^*$ . Ein „Folgenendstück ab  $N_0$ “ sind alle Folgenglieder ab  $a_{N_0}$ .

Der Zweck des folgenden Satzes besteht auch darin, die Argumentationstechnik der sogenannten „Epsilon-Logik“ einzüben. Wir erkennen, daß man durch konsequentes Anwenden der  $\varepsilon$ - $N_0$ -Sprache tatsächlich nützliche Aussagen zeigen kann. Wenn man sich an diese Ausdrucksweise erstmal gewöhnt hat, dann ist es gar nicht mehr so schwierig:

**Satz 5.3.** *Konvergente Folgen haben genau einen Grenzwert.*

*Beweis.* Wir nehmen das Gegenteil an und machen uns auf die Suche nach einem Widerspruch: Sei  $(a_n)_{n \in \mathbb{N}}$  eine konvergente Folge mit 2 verschiedenen Grenzwerten  $a^*$  und  $a^{**}$ . Dann ist laut Definition

$$\begin{aligned} \forall \varepsilon > 0 \exists N_{0,*}(\varepsilon) : \forall n \geq N_{0,*}(\varepsilon) : \|a_n - a^*\|_{K^d} < \varepsilon, \\ \forall \varepsilon > 0 \exists N_{0,**}(\varepsilon) : \forall n \geq N_{0,**}(\varepsilon) : \|a_n - a^{**}\|_{K^d} < \varepsilon. \end{aligned}$$

Nun wählen wir  $\varepsilon := \frac{1}{17} \|a^* - a^{**}\|_{K^d} > 0$ . Sei nun  $n \geq N_{0,*}(\varepsilon)$  und  $n \geq N_{0,**}(\varepsilon)$ . Dann haben wir

$$\|a^* - a^{**}\|_{K^d} = \|a^* - a_n + a_n - a^{**}\|_{K^d} \leq \|a^* - a_n\|_{K^d} + \|a_n - a^{**}\|_{K^d} < \varepsilon + \varepsilon = \frac{2}{17} \|a^* - a^{**}\|_{K^d},$$

und jetzt dürfen wir durch die positive Zahl  $\|a^* - a^{**}\|_{K^d}$  kürzen: also  $1 < \frac{2}{17}$ , was absurd ist.  $\square$

**Definition 5.4.** *Sei  $(a_n)_{n \in \mathbb{N}}$  eine Folge von Elementen einer Menge  $M$ , und sei  $(k(n))_{n \in \mathbb{N}}$  eine streng monoton wachsende Folge natürlicher Zahlen. Dann heißt die Folge*

$$(a_{k(n)})_{n \in \mathbb{N}}$$

Teilfolge<sup>3</sup> der Folge  $(a_n)_{n \in \mathbb{N}}$ .

**Satz 5.5.** *Wenn eine Folge gegen einen Grenzwert konvergiert, dann konvergiert auch jede ihrer Teilfolgen gegen denselben Grenzwert.*

*Beweis.* Übungsaufgabe.  $\square$

**Satz 5.6.** *Seien  $(a_n)_{n \in \mathbb{N}}$  und  $(b_n)_{n \in \mathbb{N}}$  zwei konvergente Folgen in  $K^d$  mit Grenzwerten  $a^*$  und  $b^*$ , und seien  $\alpha, \beta \in K$ . Dann konvergiert auch die Folge  $(\alpha a_n + \beta b_n)_{n \in \mathbb{N}}$ , und zwar gegen  $\alpha a^* + \beta b^*$ .*

*Beweis.* Zunächst haben wir für sämtliche  $n \geq 1$ , daß

$$\|(\alpha a_n + \beta b_n) - (\alpha a^* + \beta b^*)\|_{K^d} \leq \|\alpha(a_n - a^*)\|_{K^d} + \|\beta(b_n - b^*)\|_{K^d} = |\alpha| \|a_n - a^*\|_{K^d} + |\beta| \|b_n - b^*\|_{K^d}.$$

Laut Voraussetzung und Definition ist

$$\begin{aligned} \forall \varepsilon > 0 \exists N_{0,a}(\varepsilon) : \forall n \geq N_{0,a}(\varepsilon) : \|a_n - a^*\|_{K^d} < \varepsilon, \\ \forall \varepsilon > 0 \exists N_{0,b}(\varepsilon) : \forall n \geq N_{0,b}(\varepsilon) : \|b_n - b^*\|_{K^d} < \varepsilon. \end{aligned}$$

Sei nun ein positives  $\varepsilon$  gegeben, und wir suchen ein  $N_{0,+}(\varepsilon)$ , sodaß für jedes  $n \geq N_{0,+}(\varepsilon)$  gilt, daß

$$\|(\alpha a_n + \beta b_n) - (\alpha a^* + \beta b^*)\|_{K^d} \stackrel{!}{<} \varepsilon.$$

Wenn wir dieses  $N_{0,+}(\varepsilon)$  erfolgreich gebaut haben, dann ist der Beweis komplett. Mit der obigen Zerlegung erkennen wir, daß folgendes  $N_{0,+}(\varepsilon)$  unsere Wünsche erfüllt:

$$N_{0,+}(\varepsilon) := \max \left\{ N_{0,a} \left( \frac{\varepsilon}{2|\alpha|} \right), N_{0,b} \left( \frac{\varepsilon}{2|\beta|} \right) \right\},$$

zumindest in dem Fall, daß weder  $\alpha$  noch  $\beta$  gleich Null ist. Dieser Sonderfall  $\alpha \cdot \beta = 0$  sei den Studierenden als Übungsaufgabe ans Herz gelegt.  $\square$

<sup>3</sup>sub-sequence

**Warnung 5.7.** Die Regel

$$\lim_{n \rightarrow \infty} (\alpha a_n + \beta b_n) = \alpha \lim_{n \rightarrow \infty} a_n + \beta \lim_{n \rightarrow \infty} b_n$$

gilt nur dann, wenn beide Grenzwerte auf der rechten Seite existieren! Als Beispiel betrachte man  $\alpha = \beta = 1$  sowie  $(a_1, a_2, a_3, \dots) = (1, 2, 1, 2, 1, 2, 1, 2, \dots)$  und  $(b_1, b_2, b_3, \dots) = (6, 5, 6, 5, 6, 5, 6, 5, \dots)$ .

**Definition 5.8.** Sei  $M \subset K^d$ . Ein Punkt  $a^* \in K^d$  heißt Häufungspunkt von  $M$ <sup>4</sup>, wenn es eine Folge

$$(a_n)_{n \in \mathbb{N}} \subset M \setminus \{a^*\}$$

gibt mit  $a^* = \lim_{n \rightarrow \infty} a_n$ .

Die folgende Vereinigungsmenge heißt Abschluß<sup>5</sup> von  $M$ :

$$\overline{M} = M \cup \{\text{alle Häufungspunkte von } M\}.$$

**Beispiele:**

- $K^d = \mathbb{R}$  und  $M = (0, 1]$ ,
- $K^d = \mathbb{R}^2$  und  $M = \{x : \|x\|_{K^d} < R\}$ .

**Definition 5.9.** Eine Menge  $M$  ist abgeschlossen<sup>6</sup>, wenn jeder Häufungspunkt von  $M$  in  $M$  liegt.

**Beispiele:** Für  $K^d = \mathbb{R}$  betrachte man  $M = [0, 1]$  bzw.  $M = [0, 1)$  bzw.  $M = [0, \infty)$ .

**Definition 5.10.** Sei  $x_0 \in K^d$  und  $\varepsilon \in \mathbb{R}_+$ . Die Menge

$$B_\varepsilon(x_0) = \{x \in K^d : \|x - x_0\|_{K^d} < \varepsilon\}$$

heißt  $\varepsilon$ -Umgebung<sup>7</sup> von  $x_0 \in K^d$  oder auch  $\varepsilon$ -Ball um  $x_0$ .

**Definition 5.11.** Sei  $M \subset K^d$  eine Menge. Ein Punkt  $x_0 \in K^d$  heißt innerer Punkt<sup>8</sup> von  $M$ , wenn es eine Umgebung  $B_\varepsilon(x_0) \subset M$  gibt. Ein Punkt  $x_0 \in K^d$  heißt Randpunkt<sup>9</sup> von  $M$ , wenn in jeder  $\varepsilon$ -Umgebung  $B_\varepsilon(x_0)$  ein Punkt in  $M$  und ein Punkt in  $K^d \setminus M$  liegen. Die Menge aller Randpunkte wird mit  $\partial M$  bezeichnet.

**Definition 5.12.** Eine Menge  $M \subset K^d$  ist offen<sup>10</sup>, wenn jeder Punkt von  $M$  ein innerer Punkt von  $M$  ist.

**Satz 5.13.** Eine Menge  $M$  ist abgeschlossen genau dann, wenn  $\partial M \subset M$ .

Eine Menge  $M \subset K^d$  ist offen genau dann, wenn  $K^d \setminus M$  abgeschlossen ist.

*Beweis.* Das schaffen Sie selbst. □

Man sagt auch umgangssprachlich:

*Bei einer abgeschlossenen Menge gehört der Rand dazu.  
Bei einer offenen Menge gehört der Rand nicht dazu.*

Als Sonderfälle haben wir  $M = \emptyset$  und  $M = K^d$ . Diese Mengen sind gleichzeitig offen und abgeschlossen. Es gibt auch Mengen in  $\mathbb{R}$ , die weder offen noch abgeschlossen sind, z.B.  $M = (0, 1]$ .

**Definition 5.14.** Eine Menge  $M \subset K^d$  heißt beschränkt<sup>11</sup>, wenn es eine Konstante  $C \in \mathbb{R}$  gibt, sodaß  $\|x\|_{K^d} \leq C$  für jedes  $x \in M$  gilt. Also

$$\exists C \in \mathbb{R} : \forall x \in M : \|x\|_{K^d} \leq C.$$

Das bedeutet anschaulich, daß jede beschränkte Menge in einer passend groß gewählten Kugel untergebracht werden kann. Der Kugelradius  $C$  darf von der betreffenden Menge  $M$  abhängen.

**Frage:** Sei  $M = \mathbb{Q} \subset \mathbb{R}^1$  die Menge der rationalen Zahlen im Vektorraum  $\mathbb{R}^1$ . Was sind die Randpunkte von  $M$ ? die Häufungspunkte? die inneren Punkte?

<sup>4</sup>cluster point of  $M$

<sup>5</sup>closure

<sup>6</sup>closed

<sup>7</sup> $\varepsilon$ -neighbourhood

<sup>8</sup>inner point

<sup>9</sup>boundary point

<sup>10</sup>open

<sup>11</sup>bounded

## 5.2 Folgen und Reihen in normierten Räumen

Das charakteristische Merkmal der Vektorräume  $\mathbb{R}^d$  und  $\mathbb{C}^d$  ist, daß sie endlichdimensional sind. In den Betrachtungen des vorigen Abschnitts haben wir aber diese Endlichdimensionalität nirgendwo benutzt. Deshalb werden wir jetzt allgemeinere Vektorräume verwenden, die auch unendlichdimensional sein können. Dabei lassen wir uns von den Anwendungen in der Physik leiten, denn der für die Quantenmechanik höchst wichtige Vektorraum  $L^2(\mathbb{R}^3 \rightarrow \mathbb{C})$  ist ja unendlichdimensional.

Um Abstände in Vektorräumen messen zu können, ist es hilfreich, eine Norm zu haben, was uns dann zum Konzept der normierten Räume führen wird. Insbesondere betrachten wir folgende normierten Räume:

$$\mathbb{R}^d, \mathbb{C}^d, K^{p \times q}, C^k(\mathbb{R} \rightarrow \mathbb{R}^d), \dots$$

Da ab jetzt  $U$  ein beliebiger Vektorraum sein wird, über den nichts bekannt ist (abgesehen von den Eigenschaften, die in der Definition von abstrakten Vektorräumen gefordert werden), können wir auch den Begriff einer Norm nur abstrakt definieren:

**Definition 5.15.** Sei  $U$  ein Vektorraum über einem Körper  $K$ ,  $K = \mathbb{R}$  oder  $K = \mathbb{C}$ . Eine Abbildung

$$\begin{aligned} \|\cdot\| : U &\rightarrow \mathbb{R}, \\ \|\cdot\| : u &\mapsto \|u\| \end{aligned}$$

heißt Norm von  $U$ , wenn gilt:

- $\|u\| \geq 0$  für jedes  $u \in U$ ; und  $\|u\| = 0$  genau dann, wenn  $u = 0$ ,
- $\|\lambda u\| = |\lambda| \cdot \|u\|$ , für jedes  $u \in U$  und jedes  $\lambda \in K$ ,
- $\|u + v\| \leq \|u\| + \|v\|$  für jegliche  $u, v \in U$ .

Das Paar  $(U, \|\cdot\|)$  heißt normierter Raum<sup>12</sup>.

Jeder normierte Raum ist ein Vektorraum. Die Umkehrung gilt nicht. Anschaulich bedeutet  $\|u\|$  die „Länge“ des Vektors  $u$ .

**Beispiele:** Für  $U = \mathbb{R}^d$  sind folgende Normen gebräuchlich:

$$\begin{aligned} \|x\|_2 &:= \sqrt{x_1^2 + \dots + x_d^2}, \\ \|x\|_1 &:= |x_1| + \dots + |x_d|, \\ \|x\|_\infty &:= \max_{j=1, \dots, d} |x_j|. \end{aligned}$$

Analog für  $U = \mathbb{C}^d$ . Für Matrizen  $A \in K^{p \times q}$  benutzt man unter anderem

$$\|A\| := \sqrt{\sum_{i,j} |a_{ij}|^2}, \quad \text{oder auch} \quad \|A\| := \max_{i,j} |a_{ij}|.$$

Für den Funktionenraum  $U = C([a, b] \rightarrow \mathbb{R})$  erscheinen folgende Normen plausibel:

$$\begin{aligned} \|f\|_\infty &:= \max_{x \in [a, b]} |f(x)|, \\ \|f\|_2 &:= \sqrt{\int_a^b f(x)^2 dx}, \\ \|f\|_1 &:= \int_a^b |f(x)| dx. \end{aligned}$$

In Wirklichkeit ist die  $\|\cdot\|_\infty$ -Norm etwas anders definiert; die obige Formel ist aber nicht falsch, wenn die Funktion  $f$  auf  $[a, b]$  stetig ist (wie wir es ja vorausgesetzt hatten).

Und für den Raum  $U = C^1([a, b] \rightarrow \mathbb{R})$  verwendet man im Allgemeinen  $\|f\|_{C^1} := \|f\|_\infty + \|f'\|_\infty$ .

**Definition 5.16.** Sei  $M$  eine Menge. Eine Abbildung

$$\begin{aligned} \mathbb{N} &\rightarrow M, \\ n &\mapsto a_n \end{aligned}$$

heißt Folge<sup>13</sup> von Elementen aus  $M$ .

<sup>12</sup>normed space

<sup>13</sup>sequence



Mit Hilfe der jetzt bereitgestellten Begriffe „Norm“ und „Folge“ definieren wir anschließend die Begriffe „Konvergenz“, „Teilfolge“, „Häufungspunkt“, „Abschluß“, „abgeschlossene Menge“, „ $\varepsilon$ -Umgebung“, „innerer Punkt“, „Randpunkt“, „offene Menge“, „beschränkte Menge“ genauso wie im vorigen Abschnitt (wir ersetzen einfach überall „ $K^d$ “ durch „ $U$ “). Die Sätze 5.3, 5.5, 5.6 und 5.13 gelten auch jetzt.

**Beispiel 5.17.** Sei  $U$  der Vektorraum der auf  $[0, 1]$  „vernünftig integrierbaren“ Funktionen (diese ungenaue Formulierung ist unvermeidlich), sei  $f_n = f_n(x) := x^n$ , und sei  $f_*(x) := 0$  für  $0 \leq x < 1$  sowie  $f_*(1) := 1$ .

- für jedes feste  $x \in [0, 1]$  ist  $\lim_{n \rightarrow \infty} f_n(x) = f_*(x)$ , mit Konvergenz im normierten Raum  $\mathbb{R}$ ,
- wir haben auch  $\lim_{n \rightarrow \infty} f_n = f_*$  mit Konvergenz in der  $\|\cdot\|_1$ -Norm für  $U$ , denn  $\|f_n - f_*\|_1 = \frac{1}{n+1}$ ,
- sei  $g = g(x) := 0$  die Nullfunktion auf  $[0, 1]$ , also  $g(x) = 0$  insbesondere auch für  $x = 1$ . Dann ist  $\lim_{n \rightarrow \infty} f_n = g$  mit Konvergenz in der  $\|\cdot\|_1$ -Norm für  $U$ , denn  $\|f_n - g\|_1 = \frac{1}{n+1}$ . (Nun ist aber  $g \neq f_*$ . Wie versöhnen Sie das mit dem zweiten • und mit Satz 5.3 ?)
- gemessen in der  $\|\cdot\|_\infty$ -Norm konvergiert die Folge der  $f_n$  nirgendwohin, auch nicht nach  $f_*$  oder  $g$ , denn  $\|f_n - f_*\|_\infty = \|f_n - g\|_\infty = 1$  für alle  $n$ .

### 5.2.1 Vollständigkeit

Aus Beispiel 5.17 lernen wir, daß wir im Falle eines unendlichdimensionalen Vektorraumes  $U$  bei der Wahl der Norm von  $U$  Vorsicht walten lassen sollten. Ansonsten könnte es passieren, daß der entstehende normierte Raum nicht vollständig ist, was einige unangenehme Konsequenzen mit sich brächte.

**Definition 5.18.** Eine Folge  $(a_n)_{n \in \mathbb{N}} \subset U$  heißt CAUCHY-Folge<sup>14</sup> genau dann, wenn:

$$\forall \varepsilon > 0 \quad \exists N_0(\varepsilon) : \forall n, m \geq N_0(\varepsilon) : \|a_n - a_m\|_U < \varepsilon.$$

**Definition 5.19.** Ein normierter Raum  $(U, \|\cdot\|_U)$  heißt vollständig<sup>15</sup>, wenn jede Cauchy-Folge  $(a_n)_{n \in \mathbb{N}} \subset U$  einen Grenzwert  $a^* \in U$  hat. Ein vollständiger normierter Raum heißt auch BANACHRAUM<sup>16</sup>. Ein vollständiger euklidischer bzw. unitärer Raum heißt HILBERTRAUM<sup>17</sup>.

Jeder Hilbertraum ist ein Banachraum. Jeder Banachraum ist ein normierter Raum. Die Umkehrungen gelten beide nicht.

**Beispiel:** Der Raum  $(\mathbb{Q}, |\cdot|)$  ist ein normierter Vektorraum über dem Körper  $\mathbb{Q}$ , aber nicht vollständig. Als Gegenbeispiel betrachten wir

$$(a_1, a_2, \dots) = (3, 3.1, 3.14, 3.141, 3.1415, 3.14159, \dots),$$

definiert als Folge der abgehackten Dezimalbruchentwicklung von  $\pi$ . Der „Grenzwert“ dieser Folge ist  $\pi \notin \mathbb{Q}$ . Diese Cauchyfolge hat also keinen Grenzwert in  $\mathbb{Q}$ ! Wir stellen uns dies so vor, daß genau an der Stelle, wo der Grenzwert dieser Folge sein müßte (wenn es ihn gäbe), der Vektorraum  $\mathbb{Q}$  leider ein Loch hat.

**Satz 5.20.** Der Raum  $(\mathbb{R}, |\cdot|)$  als normierter Raum über dem Körper  $\mathbb{R}$  ist vollständig. Er ist sogar der kleinste vollständige Raum, der als Erweiterung von  $\mathbb{Q}$  gewonnen werden kann.

*Beweis.* Wir können ihn hier nicht führen, da wir die reellen Zahlen nirgends definiert hatten. Stattdessen begnügen wir uns mit der Bemerkung, daß  $\mathbb{R}$  genau als Vervollständigung von  $\mathbb{Q}$  definiert wurde.  $\square$

Die linearen Räume  $\mathbb{R}^d$ ,  $\mathbb{C}^d$  sind mit jeder der Normen  $\|\cdot\|_1$ ,  $\|\cdot\|_2$  und  $\|\cdot\|_\infty$  vollständig, genauso wie der Raum der Matrizen mit jeder der obigen Normen. Dies folgt alles aus der Vollständigkeit von  $\mathbb{R}$ .

**Satz 5.21.** Der Raum der auf dem Intervall  $[a, b]$  stetigen Funktionen,  $C([a, b] \rightarrow \mathbb{R})$ , versehen mit der  $\|\cdot\|_\infty$ -Norm, ist vollständig.

*Beweis.* Kommt später.  $\square$

<sup>14</sup>Cauchy-sequence

<sup>15</sup>complete

<sup>16</sup>STEFAN BANACH, 1892–1945, polnischer Mathematiker

<sup>17</sup>DAVID HILBERT, 1862–1943, deutscher Mathematiker

**Frage:** Man zeige durch ein Beispiel, daß der Raum  $C([a, b] \rightarrow \mathbb{R})$ , versehen mit der  $\|\cdot\|_1$ -Norm, *nicht* vollständig ist. Analog für die  $\|\cdot\|_2$ -Norm.

Die Unvollständigkeit des Raumes der stetigen Funktionen unter der  $\|\cdot\|_1$ - bzw.  $\|\cdot\|_2$ -Norm ist unerwünscht. Andererseits sind diese Normen so wichtig, daß man nicht auf sie verzichten kann. Eine solche Funktion könnte z.B. ein Geschwindigkeitsfeld sein, und eine physikalisch naheliegende Norm wäre dann die kinetische Energie (oder vielmehr die Wurzel daraus), die diesem Geschwindigkeitsfeld innewohnt. Das ist aber exakt eine  $L^2$ -Norm. Der Ausweg ist, den Raum der stetigen Funktionen solange zu ergänzen, bis ein vollständiger Raum entsteht. Diese Räume bezeichnet man dann als  $L^1([a, b] \rightarrow \mathbb{R})$  bzw.  $L^2([a, b] \rightarrow \mathbb{R})$ ; und sie bestehen aus all jenen Funktionen, für die die Integrale (im Lebesgue<sup>18</sup>-Sinne)

$$\int_{x=a}^b |f(x)| dx \quad \text{bzw.} \quad \int_{x=a}^b |f(x)|^2 dx$$

endlich sind. Für die Theorie des Lebesgue-Integrales ist bei den Mathematikstudierenden die gesamte Maßtheorie-Vorlesung des dritten Semesters reserviert, und wir verzichten hier auf sämtliche Einzelheiten.

**Satz 5.22.** *Konvergente Folgen sind Cauchy-Folgen.*

*Beweis.* Sei  $(a_n)_{n \in \mathbb{N}}$  eine konvergente Folge mit Grenzwert  $a^*$ . Wir haben:

$$\forall \varepsilon > 0 \exists N_0(\varepsilon) : \forall n \geq N_0(\varepsilon) : \|a_n - a^*\| < \varepsilon.$$

Wir wollen zeigen, daß:

$$\forall \varepsilon > 0 \exists N_1(\varepsilon) : \forall n, m \geq N_1(\varepsilon) : \|a_n - a_m\| < \varepsilon.$$

Gesucht ist ein  $N_1(\varepsilon)$ , sodaß die Aussage in der vorigen Zeile gilt. Wir wählen  $N_1(\varepsilon) := N_0(\varepsilon/2)$ . Dieses  $N_1$  hat die gewünschten Eigenschaften, denn: Seien nun  $n, m \geq N_1(\varepsilon)$ . Dann haben wir

$$\|a_n - a_m\| = \|a_n - a^* + a^* - a_m\| \leq \|a_n - a^*\| + \|a^* - a_m\| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Also ist die Folge  $(a_n)_{n \in \mathbb{N}}$  tatsächlich eine Cauchy-Folge. □

Die Umkehrung dieses Satzes gilt bekanntlich nicht.

**Satz 5.23.** *Cauchy-Folgen sind beschränkt.*

*Beweis.* Man wähle in der Definition der Cauchy-Folgen  $\varepsilon := 1$ ,  $m := N_0(1)$ , schreibe sich die Definition des Cauchy-Folgen-Begriffs genau hin, und schon steht es da. □

**Satz 5.24.** *Sei  $(a_n)_{n \in \mathbb{N}}$  eine konvergente Folge in einem normierten Raum  $U$ . Dann gilt*

$$\left\| \lim_{n \rightarrow \infty} a_n \right\|_U = \lim_{n \rightarrow \infty} \|a_n\|_U.$$

*Beweisskizze.* In jedem normierten Raum  $U$  gilt die Ungleichung

$$\left| \|x\|_U - \|y\|_U \right| \leq \|x - y\|_U. \tag{5.1}$$

Wir setzen  $x := \lim_{n \rightarrow \infty} a_n$  und  $y := a_n$ , und jetzt ist es nicht mehr schwierig. □

**Frage:** Warum gilt (5.1) ?

---

<sup>18</sup>HENRI LÉON LEBESGUE, 1875–1941

### 5.2.2 Reihen in normierten Räumen

**Definition 5.25.** Sei  $(U, \|\cdot\|)$  ein normierter Raum und  $(a_n)_{n \in \mathbb{N}}$  eine Folge in  $U$ . Unter einer Reihe<sup>19</sup>

$$\sum_{n=0}^{\infty} a_n$$

verstehen wir zwei Dinge (gleichzeitig):

- einerseits die Folge der Partialsummen  $(S_N)_{N \in \mathbb{N}}$ , wobei  $S_N = \sum_{n=0}^N a_n$ ,
- andererseits den Grenzwert  $S = \lim_{N \rightarrow \infty} S_N$ , **aber nur, falls dieser existiert.**

Wenn die Folge der Partialsummen  $(S_N)_{N \in \mathbb{N}}$  gegen  $S$  konvergiert, dann sagt man, daß die Reihe  $\sum_{n=0}^{\infty} a_n$  konvergiert und den Wert  $S$  hat. Ansonsten sagt man, daß die Reihe  $\sum_{n=0}^{\infty} a_n$  divergiert.

**Definition 5.26.** Sei  $\sum_{n=0}^{\infty} a_n$  eine Reihe von Elementen aus einem normierten Raum  $U$ . Wenn die Reihe von reellen Zahlen

$$\sum_{n=0}^{\infty} \|a_n\|_U$$

konvergiert, dann nennt man die Reihe  $\sum_{n=0}^{\infty} a_n$  absolut konvergent<sup>20</sup>.

**Beispiel 5.27.** Sei  $U = \mathbb{R}$ .

1. Die geometrische Reihe  $\sum_{k=0}^{\infty} q^k$  konvergiert genau dann, wenn  $|q| < 1$ . Dann ist ihr Grenzwert gleich  $\frac{1}{1-q}$ , denn für die Partialsumme  $S_N$  gilt die Formel  $S_N = \frac{1-q^{N+1}}{1-q}$ , und somit ist  $\lim_{N \rightarrow \infty} S_N = \frac{1}{1-q}$ , aber nur wenn  $|q| < 1$ .
2. Die harmonische Reihe  $\sum_{k=1}^{\infty} \frac{1}{k}$  divergiert, denn wenn  $N = 2^k$  ist mit  $k \in \mathbb{N}_+$ , dann ist  $S_N > \frac{k}{2}$ , und somit ist  $\lim_{N \rightarrow \infty} S_N = +\infty$ .
3. Die Reihe  $\sum_{k=1}^{\infty} \frac{1}{k^2}$  konvergiert (warum ?), und zwar gegen  $\frac{1}{6}\pi^2$  (Beweis im 2. Semester mittels Fourierreihen).
4. Die Reihe  $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} \mp \dots$  konvergiert (und zwar gegen  $\ln 2$ , vgl. Satz 6.72), aber nicht absolut (warum ?).

**Frage:** Vollenden Sie die Argumentation dieses Beispiels in den Unterpunkten 1., 2. und 3. Die Ungleichung  $\frac{1}{k^2} < \frac{1}{k(k-1)}$  könnte nützlich sein.

Alle wichtigen Funktionen (Winkelfunktionen, Arcusfunktionen, Exponentialfunktionen, Logarithmusfunktionen) lassen sich mit Potenzreihen definieren (wie man das genau macht, schauen wir uns später an):

**Beispiel 5.28 (Potenzreihe).** Sei  $U = C([-1, 1] \rightarrow \mathbb{R})$  oder  $U = C(B_1(0) \rightarrow \mathbb{C})$ , versehen mit der  $\|\cdot\|_{\infty}$ -Norm. Die Reihe

$$\sum_{k=0}^{\infty} \frac{1}{k!} x^k, \quad |x| \leq 1,$$

konvergiert in  $U$ , und zwar gegen die Exponentialfunktion (Beweis folgt). Konvergenz gilt auch in der  $\|\cdot\|_1$ - bzw.  $\|\cdot\|_2$ -Norm.

In der Signaltechnik sind Reihen von Winkelfunktionen wichtig:

**Beispiel 5.29 (Fourierreihe).** Sei  $U = L^2([-\pi, \pi] \rightarrow \mathbb{R})$ . Im 2. Semester werden wir erkennen: Die Reihe

$$\frac{4}{\pi} \left( \sin(x) + \frac{\sin(3x)}{3} + \frac{\sin(5x)}{5} + \frac{\sin(7x)}{7} + \dots \right)$$

konvergiert in der  $\|\cdot\|_2$ -Norm gegen die  $2\pi$ -periodisch fortgesetzte Funktion („Rechtecksignal“)

$$x \mapsto S(x) := \begin{cases} 1 & : 0 < x < \pi, \\ -1 & : -\pi < x < 0. \end{cases}$$

<sup>19</sup>series

<sup>20</sup>absolutely convergent

**Frage:** Warum konvergiert diese Reihe *nicht* in der  $\|\cdot\|_\infty$ -Norm?

**Satz 5.30.** Sei  $\sum_{n=0}^\infty a_n$  eine Reihe, die im normierten Raum  $U$  konvergiert. Dann ist  $\lim_{n \rightarrow \infty} a_n = \vec{0}_U$ .

*Beweis.* Die Folge  $(S_N)_{N \in \mathbb{N}}$  der Partialsummen konvergiert, ist also eine Cauchy-Folge, wegen Satz 5.22. Die Definition des Begriffes „Cauchy-Folge“ sagt uns jetzt:

$$\forall \varepsilon > 0 \exists N_0(\varepsilon) : \forall N, M \geq N_0(\varepsilon) : \|S_N - S_M\|_U < \varepsilon.$$

Wir wählen nun  $M := N + 1$  und erhalten

$$\forall \varepsilon > 0 \exists N_0(\varepsilon) : \forall N \geq N_0(\varepsilon) : \|a_{N+1}\|_U < \varepsilon.$$

Also streben die  $a_N$  gegen  $\vec{0}_U$ , denn  $\|a_{N+1}\|_U = \|a_{N+1} - \vec{0}_U\|_U$ . □

**Warnung 5.31.** Die Umkehrung dieses Satzes ist falsch (und ein beliebter Fehler).

Absolute Konvergenz ist besser als Konvergenz — das ist die nächste Aussage. Siehe auch Abschnitt 5.3.3.

**Satz 5.32.** Wenn die Reihe  $\sum_{n=0}^\infty a_n$  in einem Banach-Raum  $U$  absolut konvergiert, dann konvergiert sie dort auch (im Sinne von Definition 5.25).

*Beweis.* Die Folge  $(\|a_1\|, \|a_1\| + \|a_2\|, \|a_1\| + \|a_2\| + \|a_3\|, \dots)$  konvergiert, ist also eine Cauchy-Folge. Das bedeutet:

$$\forall \varepsilon > 0 \exists N_0(\varepsilon) : \forall m, n \geq N_0(\varepsilon) (m \geq n) : \|a_n\| + \|a_{n+1}\| + \dots + \|a_m\| < \varepsilon.$$

Sei nun  $(S_N)_{N \in \mathbb{N}}$  die Partialsummenfolge der Reihe  $\sum_{n=0}^\infty a_n$ . Dann ist für  $m > n$

$$\|S_m - S_n\| = \|a_{n+1} + a_{n+2} + \dots + a_m\| \leq \|a_{n+1}\| + \|a_{n+2}\| + \dots + \|a_m\|.$$

Also ist  $(S_N)_{N \in \mathbb{N}}$  eine Cauchy-Folge und folglich konvergent, denn  $U$  ist ein Banachraum. □

*Eine Reihe in einem Banachraum konvergiert genau dann,  
wenn beliebig lange Reihenendteilstücke beliebig klein werden.*

Ein Reihenendstück wäre  $\sum_{n=N}^\infty a_n$ , und ein Reihenendteilstück ist  $\sum_{n=N}^M a_n = S_M - S_{N-1}$ . Hierbei darf die Anzahl der Summanden (also  $M - N + 1$ ) beliebig groß sein. Die Konvergenz der Reihe ist ja gerade gleichbedeutend damit, daß die Partialsummenfolge eine Cauchy-Folge ist.

**Satz 5.33.** Seien  $\sum_{n=0}^\infty a_n$  und  $\sum_{n=0}^\infty b_n$  zwei konvergente Reihen in einem normierten Raum mit Reihenwerten  $s_a$  und  $s_b$ . Seien weiterhin  $\alpha, \beta \in K$ . Dann konvergiert auch die Reihe  $\sum_{n=0}^\infty (\alpha a_n + \beta b_n)$ , und zwar gegen den Wert  $\alpha s_a + \beta s_b$ .

*Beweis.* Folgt sofort aus dem entsprechenden Satz für Folgen, also Satz 5.6. □

### 5.2.3 Konvergenzkriterien

Um zu entscheiden, ob Reihen konvergieren, benutzt man üblicherweise Vergleichskriterien:

**Satz 5.34 (Allgemeines Majorantenkriterium).** Sei  $\sum_{n=0}^\infty a_n$  eine Reihe in einem Banachraum  $U$  (konvergent oder divergent). Wenn es eine Folge  $(\gamma_n)_{n \in \mathbb{N}}$  reeller Zahlen gibt mit den zwei Eigenschaften:

- es existiert ein  $N$ , sodaß  $\|a_n\|_U \leq \gamma_n$  für jedes  $n \geq N$ ,
- die Reihe  $\sum_{n=0}^\infty \gamma_n$  konvergiert,

dann konvergiert auch die Reihe  $\sum_{n=0}^\infty a_n$ , sogar absolut.

*Beweis.* Das Konvergenzverhalten einer Reihe (divergent/konvergent/absolut konvergent) ändert sich nicht, wenn man am Anfang einige  $a_n$  wegläßt. Wir dürfen also annehmen, daß die Reihe erst mit  $a_N$  beginnt. Sei nun  $(S_n)_{n \geq N}$  die Partialsummenfolge zu  $\sum_{n=N}^{\infty} \|a_n\|_U$ , und  $(\sigma_n)_{n \geq N}$  die Partialsummenfolge zu  $\sum_{n=N}^{\infty} \gamma_n$ . Weil die Reihe über die  $\gamma_n$  konvergiert, wissen wir, daß die  $\sigma_n$  eine Cauchy-Folge bilden:

$$\forall \varepsilon > 0 \exists N_0(\varepsilon) : \forall n, m \geq N_0(\varepsilon) : |\sigma_n - \sigma_m| < \varepsilon.$$

Nun ist (für  $n \geq m$ )

$$|S_n - S_m| = \|a_{m+1}\|_U + \cdots + \|a_n\|_U \leq \gamma_{m+1} + \cdots + \gamma_n = |\sigma_n - \sigma_m|.$$

Also ist auch  $(S_n)_{n \geq N}$  eine Cauchy-Folge in  $\mathbb{R}$ . □

**Beispiel 5.35.** Wir betrachten im Raum  $U = C((-\infty, \infty) \rightarrow \mathbb{R})$  die Fourier-Reihe

$$\sin(x) + \frac{\sin(2x)}{4} + \frac{\sin(3x)}{9} + \frac{\sin(4x)}{16} + \dots$$

und fragen nach ihrer Konvergenz in der  $\|\cdot\|_{\infty}$ -Norm. Wir haben

$$a_n = \frac{\sin(nx)}{n^2}$$

und  $\|a_n\|_U = \frac{1}{n^2}$ . Wir können jetzt  $\gamma_n := \frac{1}{n^2}$  wählen, was eine konvergente Reihe ergibt. Also konvergiert obige Fourierreihe.

**Beispiel 5.36.** Die Reihe  $\sum_{n=1}^{\infty} \frac{1}{n!}$  konvergiert wegen  $n! = 1 \cdot 2 \cdot 3 \cdots n \geq 2^{n-1}$ .

**Folgerung 5.37.** Wir vergleichen eine beliebige Reihe mit der geometrischen Reihe:

- Wenn es  $C > 0$ ,  $N \in \mathbb{N}$  und  $p < 1$  gibt mit  $\|a_n\|_U \leq Cp^n$  für alle  $n \geq N$ , dann konvergiert  $\sum_{n=0}^{\infty} a_n$  absolut in  $U$ ,
- Wenn es  $C > 0$ ,  $N \in \mathbb{N}$  und  $p > 1$  gibt mit  $\|a_n\|_U \geq Cp^n$  für alle  $n \geq N$ , dann divergiert  $\sum_{n=0}^{\infty} a_n$ .

Der erste Teil folgt aus Satz 5.34, der zweite folgt aus Satz 5.30.

**Satz 5.38 (Wurzelkriterium).** Sei  $\sum_{n=0}^{\infty} a_n$  eine Reihe in einem Banachraum  $U$ . Sei  $q \in \mathbb{R}$  definiert durch

$$q = \lim_{n \rightarrow \infty} \sqrt[n]{\|a_n\|_U}$$

wobei wir voraussetzen, daß dieser Limes existiert. Dann gilt:

$q < 1$ : die Reihe  $\sum_{n=0}^{\infty} a_n$  konvergiert absolut,

$q > 1$ : die Reihe  $\sum_{n=0}^{\infty} a_n$  divergiert,

$q = 1$ : das Konvergenzverhalten ist unbekannt.

*Beweis.* Wir setzen  $p := \frac{1}{2}(1 + q)$ . Für  $q < 1$  haben wir ab einem  $N_0$  die Ungleichung  $\sqrt[n]{\|a_n\|_U} \leq p < 1$ , und für  $q > 1$  haben wir ab einem  $N_0$  die Ungleichung  $\sqrt[n]{\|a_n\|_U} \geq p > 1$ . Nun liften wir das in die  $n$ -te Potenz und verwenden Folgerung 5.37. □

Das folgende Quotientenkriterium ist gelegentlich einfacher anzuwenden. Andererseits gibt es Beispiele, in denen das Quotientenkriterium versagt, aber das Wurzelkriterium noch eine Aussage liefert.

**Satz 5.39 (Quotientenkriterium).** Sei  $\sum_{n=0}^{\infty} a_n$  eine Reihe in einem Banachraum  $U$ , und keiner der Summanden sei  $\vec{0}$ . Sei  $q \in \mathbb{R}$  definiert durch

$$q = \lim_{n \rightarrow \infty} \frac{\|a_{n+1}\|_U}{\|a_n\|_U},$$

wobei wir voraussetzen, daß dieser Limes existiert. Dann gilt:

$q < 1$ : die Reihe  $\sum_{n=0}^{\infty} a_n$  konvergiert absolut,

$q > 1$ : die Reihe  $\sum_{n=0}^{\infty} a_n$  divergiert,

$q = 1$ : das Konvergenzverhalten ist unbekannt.

*Beweis.* Sei  $q < 1$ . Wir setzen  $p := \frac{1}{2}(1+q) < 1$ , also  $q < p < 1$ . Dann gibt es ein  $N_0$ , sodaß für  $m \geq N_0$  die Quotienten  $\|a_{m+1}\|_U / \|a_m\|_U$  unterhalb von  $p$  liegen. Multiplikation passend vieler Ungleichungen diesen Typs liefert

$$\|a_{N_0+n}\|_U \leq \|a_{N_0}\|_U p^n, \quad n \geq 0.$$

Wir benutzen nun Folgerung 5.37 mit  $C = \|a_{N_0}\|_U$ , und die absolute Konvergenz der Reihe  $\sum_{n=0}^{\infty} a_n$  folgt sofort.

Für den Fall  $q > 1$  verwende man den zweiten Teil von Folgerung 5.37.  $\square$

**Beispiel 5.40.** Sei  $U = \mathbb{R}$ . Für jedes  $x \in (0, \infty)$  betrachten wir die Reihe reeller Zahlen

$$\sum_{k=0}^{\infty} \frac{1}{k!} (-2x)^k$$

und fragen, für welche  $x$  Konvergenz vorliegt. Wir haben (für fixiertes  $x$ )

$$a_n = \frac{1}{n!} (-2x)^n, \quad \text{also } |a_n| = \frac{1}{n!} (2x)^n,$$

und bekommen demnach

$$q = \lim_{n \rightarrow \infty} \frac{|a_{n+1}|}{|a_n|} = \lim_{n \rightarrow \infty} \frac{\frac{1}{(n+1)!} (2x)^{n+1}}{\frac{1}{n!} (2x)^n} = \lim_{n \rightarrow \infty} \frac{2x}{n+1} = 0,$$

also Konvergenz. Damit konvergiert diese Potenzreihe für jedes  $x \in (0, \infty)$  absolut.

### 5.3 Folgen und Reihen reeller Zahlen

Die Menge der reellen Zahlen bildet zusammen mit der üblichen Betragsfunktion einen vollständigen normierten Raum. Das bedeutet, dass alle Ergebnisse des vorigen Abschnitts auch auf Folgen und Reihen von reellen Zahlen angewandt werden können.

Andererseits haben die reellen Zahlen auch Eigenschaften, die man in einem beliebigen normierten Raum meist nicht zur Verfügung hat. Insbesondere:

- reelle Zahlen kann man ordnen,
- reelle Zahlen kann man multiplizieren.

Aus diesem Grunde ist die Theorie der Folgen und Reihen reeller Zahlen etwas reichhaltiger.

Wir erinnern an ein einfaches Ergebnis, das zum Beispiel mittels vollständiger Induktion gezeigt werden kann (und übrigens auch in  $\mathbb{C}$  gilt):

**Satz 5.41 (Binomische Formel).** Für reelle Zahlen  $a, b$  und  $n \in \mathbb{N}_0$  gilt

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}.$$

Hierbei steht  $\binom{n}{k}$  für den Binomialkoeffizienten,

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{1 \cdot 2 \cdot \dots \cdot k}, \quad 0! = 1.$$

### 5.3.1 Schranken und Grenzen

In allgemeinen normierten Räumen hatten wir beschränkte Mengen definiert. Im  $\mathbb{R}$  haben wir zusätzlich noch einseitig beschränkte Mengen:

**Definition 5.42.** Sei  $M \subset \mathbb{R}$  eine Menge reeller Zahlen. Wenn es eine Zahl  $S \in \mathbb{R}$  gibt mit der Eigenschaft, daß  $x \leq S$  für jedes  $x \in M$ , dann heißt  $S$  obere Schranke<sup>21</sup> von  $M$ . In diesem Falle sagt man, daß die Menge  $M$  nach oben beschränkt<sup>22</sup> ist. Entsprechend definiert man untere Schranken<sup>23</sup> und den Begriff nach unten beschränkt<sup>24</sup>.

Jede Zahl, die größer ist als eine obere Schranke einer Menge, ist ebenfalls eine obere Schranke dieser Menge.

**Definition 5.43.** Sei  $M$  eine Menge reeller Zahlen. Wenn es eine Zahl  $S \in \mathbb{R}$  gibt, die gleichzeitig obere Schranke von  $M$  und Element von  $M$  ist, dann heißt  $S$  Maximum<sup>25</sup> von  $M$ . Analog definiert man das Minimum<sup>26</sup>.

Eine Menge kann zwar unendlich viele verschiedene obere oder untere Schranken haben, aber höchstens ein Maximum bzw. Minimum.

**Beispiel 5.44.** Die Menge  $M = \{x \in \mathbb{R} : 0 < x < 1\} =: (0, 1)$  hat weder Minimum noch Maximum. Sie ist nach oben beschränkt durch  $S = 1$  und nach unten beschränkt durch  $S = 0$ .

**Definition 5.45.** Sei  $M \subset \mathbb{R}$  eine nach oben beschränkte Menge. Eine Zahl  $S \in \mathbb{R}$  heißt Supremum bzw. kleinste obere Schranke bzw. obere Grenze<sup>27</sup> wenn folgendes gilt:

1.  $S$  ist obere Schranke von  $M$ ,
2. jede kleinere Zahl als  $S$  ist keine obere Schranke von  $M$ .

Entsprechend definiert man die Begriffe Infimum bzw. größte untere Schranke bzw. untere Grenze<sup>28</sup>.

**Beispiel 5.46.** Die Menge  $M = (0, 1)$  hat Infimum 0 und Supremum 1.

Die Menge  $M = \{x \in \mathbb{Q} : x^2 < 2\}$  hat in den rationalen Zahlen weder Supremum noch Infimum, obwohl sie nach oben und unten beschränkt ist.

**Satz 5.47.** Jede nichtleere nach oben beschränkte Menge reeller Zahlen hat genau ein Supremum in  $\mathbb{R}$ . Jede nichtleere nach unten beschränkte Menge reeller Zahlen hat genau ein Infimum in  $\mathbb{R}$ .

Bevor wir zum Beweis dieses zentralen Satzes kommen, schieben wir ein nützliches Ergebnis ein:

**Satz 5.48.** Sei  $(a_n)_{n \in \mathbb{N}}$  eine monoton wachsende Folge reeller Zahlen, die nach oben beschränkt ist. Dann hat diese Folge einen Grenzwert, und dieser Grenzwert ist gleich dem Supremum der Menge  $\{a_n : n \in \mathbb{N}\}$ .

Hierbei verstehen wir unter *monoton wachsend*<sup>29</sup>, daß  $a_n \leq a_{n+1}$  für jedes  $n$  gilt. Ein entsprechender Satz gilt für monoton fallende Folgen, die nach unten beschränkt sind.

*Beweis.* Sei  $a_n \leq a_{n+1}$  und  $a_n \leq S$  für jedes  $n \in \mathbb{N}$ . Wir wollen zeigen, daß die Folge  $(a_n)_{n \in \mathbb{N}}$  eine Cauchy-Folge ist, also

$$\forall \varepsilon > 0 \exists N_0(\varepsilon) : \forall n, m \geq N_0(\varepsilon) : |a_n - a_m| < \varepsilon.$$

Wir nehmen das Gegenteil an und suchen einen Widerspruch. Das Gegenteil ist

$$\exists \varepsilon_0 > 0 : \forall N \exists n = n(N), m = m(N) \geq N : |a_n - a_m| \geq \varepsilon_0.$$

---

<sup>21</sup>upper bound

<sup>22</sup>bounded from above

<sup>23</sup>lower bounds

<sup>24</sup> bounded from below

<sup>25</sup>maximum

<sup>26</sup>minimum

<sup>27</sup>supremum, least upper bound

<sup>28</sup>infimum, greatest lower bound

<sup>29</sup>monotonically increasing

Wir kommen auf diesem Wege zu einer steigenden Folge von Indizes

$$k_1 < k_2 < k_3 < k_4 < k_5 < k_6 \dots$$

mit  $\lim_{j \rightarrow \infty} k_j = \infty$ , sodaß

$$|a_{k_2} - a_{k_1}| \geq \varepsilon_0, \quad |a_{k_4} - a_{k_3}| \geq \varepsilon_0, \quad |a_{k_6} - a_{k_5}| \geq \varepsilon_0, \dots$$

Andererseits ist  $a_{k_1} \leq a_{k_2} \leq a_{k_3} \leq a_{k_4} \leq \dots$ . Dann muß zwangsläufig die Folge der  $a_n$  die Schranke  $S$  irgendwann überschreiten. Das ist ein Widerspruch. Also ist die Folge  $(a_n)_{n \in \mathbb{N}}$  eine Cauchy-Folge und somit konvergent. Die Aussage über das Supremum sieht man schnell.  $\square$

*Beweis des Satzes 5.47.* Wir zeigen nur die Aussage über das Supremum.

Man überlegt sich schnell, daß eine Menge keine zwei verschiedenen Suprema haben kann.

Die Menge hat mindestens ein Element und mindestens eine obere Schranke; diese taufen wir  $x_1$  und  $S_1$ . Es gilt  $x_1 \leq S_1$ . Wenn  $x_1 = S_1$  sein sollte, dann ist dies gerade das Supremum, und wir wären fertig. Sei also jetzt  $x_1 < S_1$ . Wir taufen das arithmetische Mittel dieser beiden Zahlen  $Z$ .

Es gibt 2 Fälle:

**$Z$  ist obere Schranke von  $M$ :** dann setzen wir  $x_2 := x_1$  und  $S_2 := Z$ .

**$Z$  ist keine obere Schranke von  $M$ :** dann gibt es ein Element von  $M$ , das zwischen  $Z$  und  $S_1$  liegt. Wir nennen dieses Element  $x_2$ , und setzen  $S_2 := S_1$ .

Auf jeden Fall haben wir jetzt 2 Zahlen  $x_1, x_2$ , die Element von  $M$  sind, und zwei obere Schranken  $S_1$  und  $S_1$ . Es gilt

$$x_1 \leq x_2 \leq S_2 \leq S_1, \quad |x_2 - S_2| \leq \frac{1}{2}|x_1 - S_1|.$$

Mit dem Intervall  $[x_2, S_2]$  können wir verfahren wie eben. Wir finden ein Element  $x_3 \in M$  und eine obere Schranke  $S_3$ , sodaß gilt

$$x_1 \leq x_2 \leq x_3 \leq S_3 \leq S_2 \leq S_1, \quad |x_3 - S_3| \leq \frac{1}{4}|x_1 - S_1|.$$

Induktiv setzen wir dieses Verfahren fort und erhalten eine monoton wachsende Folge  $(x_n)_{n \in \mathbb{N}}$  und eine monoton fallende Folge  $(S_n)_{n \in \mathbb{N}}$ . Da die wachsende Folge  $(x_n)_{n \in \mathbb{N}}$  nach oben beschränkt ist durch  $S_1$ , hat diese Folge einen Grenzwert  $x^*$ . Analog hat die monoton fallende Folge  $(S_n)_{n \in \mathbb{N}}$  einen Grenzwert  $S^*$ .

Da aber zusätzlich die Folge der Differenzen  $(x_n - S_n)_{n \in \mathbb{N}}$  nach 0 strebt, muß  $x^* = S^*$  sein.

Für jedes  $\varepsilon > 0$  gilt dann:

- im Intervall  $(x^* - \varepsilon, x^*)$  ist mindestens ein Element von  $M$ ,
- im Intervall  $(x^*, x^* + \varepsilon)$  ist mindestens eine obere Schranke von  $M$ .

Die erste Aussage bedeutet, daß keine Zahl unterhalb von  $x^*$  jemals obere Schranke von  $M$  sein kann. Die zweite Aussage bedeutet, daß oberhalb von  $x^*$  kein Element von  $M$  sein kann. Insbesondere ist dann  $x^*$  eine obere Schranke von  $M$ ; und nach der ersten Aussage ist dies die kleinste obere Schranke von  $M$ .  $\square$

Das folgende Ergebnis ist für spätere Zwecke gedacht:

**Satz 5.49 (Bolzano<sup>30</sup>–Weierstraß<sup>31</sup>).** Sei  $M \subset \mathbb{R}$  ein Intervall mit folgenden beiden Eigenschaften:

- $M$  ist beschränkt,
- $M$  ist abgeschlossen.

<sup>30</sup> BERNARD PLACIDUS JOHANN NEPOMUK BOLZANO, 1781–1848

<sup>31</sup> KARL THEODOR WILHELM WEIERSTRASS, 1815–1897



Sei  $(a_n)_{n \in \mathbb{N}}$  eine Folge, die in  $M$  liegt; also  $a_n \in M$  für jedes  $n \in \mathbb{N}$ . Dann besitzt diese Folge eine Teilfolge, die gegen einen Grenzwert **in**  $M$  strebt.

*Beweis.* Wir teilen das Intervall  $M =: M_1$  in zwei gleichgroße Teilintervalle. In einem davon liegen unendlich viele Glieder der Folge  $(a_n)_{n \in \mathbb{N}}$ . Dieses Teilintervall nennen wir  $M_2$ . Nun teilen wir  $M_2$  in zwei gleichgroße Teilintervalle. In einem davon liegen unendlich viele Folgenglieder, dieses nennen wir  $M_3$ . In diesem Stile fahren wir fort und erhalten eine Folge von Intervallen

$$M_1 \supset M_2 \supset M_3 \supset \dots,$$

von denen jedes halb so lang ist wie das vorhergehende. In jedem dieser Intervalle liegen unendlich viele Elemente der Folge.

Aus  $M_1$  wählen wir ein Element  $a_{k_1}$  der Folge. Aus  $M_2$  wählen wir ein weiteres Element  $a_{k_2}$  der Folge, das von  $a_{k_1}$  verschieden sein soll. Aus  $M_3$  wählen wir ein Folgenglied  $a_{k_3}$ , das von den beiden vorigen verschieden sein soll. Wir bekommen auf diesem Wege eine Teilfolge  $(a_{k_j})_{j \in \mathbb{N}}$ . Weil alle Glieder dieser Teilfolge ab  $j = N_0$  im Teilintervall  $M_{N_0}$  enthalten sind, ist diese Folge offensichtlich eine Cauchy-Folge. Diese hat einen reellen Grenzwert, der nach Konstruktion ein Häufungspunkt des Intervalles ist. Weil das Intervall abgeschlossen ist, liegt der Grenzwert im Intervall.  $\square$

**Bemerkung 5.50.** Dieser Satz kann verallgemeinert werden. Es reicht, daß  $M \subset \mathbb{R}^d$  eine beschränkte und abgeschlossene Menge ist, damit jede Folge in  $M$  eine in  $M$  konvergente Teilfolge enthält.

**Definition 5.51.** Sei  $U$  ein Banachraum. Eine Menge  $M \subset U$  heißt kompakt<sup>32</sup>, wenn jede Folge in  $M$  eine Teilfolge enthält, die einen Grenzwert in  $M$  hat.

Wir fassen zusammen:

- in jedem Banachraum gilt: wenn eine Teilmenge kompakt ist, dann ist sie auch beschränkt und abgeschlossen (das überlegt man sich schnell).
- im  $\mathbb{R}^d$  gilt: wenn eine Teilmenge beschränkt und abgeschlossen ist, dann ist sie auch kompakt (das haben wir in Bemerkung 5.50 schon erwähnt).
- in **jedem** unendlichdimensionalen Banachraum gilt: es gibt Teilmengen, die beschränkt und abgeschlossen sind, aber leider nicht kompakt.

Relevant für die Physik ist diese Betrachtung aus folgendem Grund: in gewisser Hinsicht ist die Quantenmechanik eine Form von *linearer Algebra in unendlichdimensionalen Räumen*. Jetzt beobachten wir aber, daß in solchen Räumen die Begriffe *kompakt* sowie *beschränkt* & *abgeschlossen* logisch nicht äquivalent sind. Das ist unangenehm, aber unvermeidlich. Für Mathematikstudierende gibt es im vierten Semester eine Funktionalanalysisvorlesung, in der (mit hohem Aufwand!) solche Fragen erörtert und beantwortet werden. Auf die Einzelheiten müssen wir aus Zeitgründen verzichten.

### 5.3.2 Beispiele für konvergente Folgen

Bei Folgen und Reihen gibt es prinzipiell (mindestens) zwei Fragen zu beantworten:

**Konvergiert die Folge/Reihe überhaupt?** Antworten dazu werden gegeben durch Majorantenkriterien oder durch Satz 5.48.

**Wohin konvergiert sie? Oder: wie lautet der Grenzwert genau?** Ein Werkzeug für die Antwort zu dieser Frage kommt jetzt.

**Satz 5.52 (Sandwichprinzip).** Es seien  $(a_n)_{n \in \mathbb{N}}$ ,  $(b_n)_{n \in \mathbb{N}}$ ,  $(c_n)_{n \in \mathbb{N}}$  Folgen reeller Zahlen mit den Eigenschaften, daß

- $a_n \leq b_n \leq c_n$  für jedes  $n \in \mathbb{N}$ ,
- die Limites  $\lim_{n \rightarrow \infty} a_n$  und  $\lim_{n \rightarrow \infty} c_n$  existieren und sind beide gleich  $g^* \in \mathbb{R}$ .

<sup>32</sup>compact

Dann existiert auch der Grenzwert  $\lim_{n \rightarrow \infty} b_n$ ; und er ist ebenfalls gleich  $g^*$ .

*Beweis.* Für jedes  $\varepsilon > 0$  suchen wir ein  $N_0(\varepsilon)$ , sodaß  $|b_n - g^*| < \varepsilon$  gilt für jedes  $n \geq N_0(\varepsilon)$ .

Wir wissen aus der Voraussetzung, daß

$$\begin{aligned} \forall \varepsilon > 0 \exists N_1(\varepsilon) : \forall n \geq N_1(\varepsilon) : |a_n - g^*| < \varepsilon, \\ \forall \varepsilon > 0 \exists N_2(\varepsilon) : \forall n \geq N_2(\varepsilon) : |c_n - g^*| < \varepsilon. \end{aligned}$$

Wir wählen  $N_0(\varepsilon) := \max(N_1(\varepsilon), N_2(\varepsilon))$ . Dann haben wir für  $n \geq N_0(\varepsilon)$ :

$$g^* - \varepsilon < a_n \leq b_n \leq c_n < g^* + \varepsilon,$$

was den Beweis vollendet. □

Als Anwendung des Sandwichprinzips haben wir folgendes Ergebnis:

**Lemma 5.53.** *Es ist  $\lim_{n \rightarrow \infty} \sqrt[n]{n} = 1$ .*

*Beweis.* Offensichtlich ist  $\sqrt[n]{n} \geq 1$  für  $n \geq 17$ . Also können wir für jedes solche  $n$  schreiben:

$$\sqrt[n]{n} = 1 + x_n, \quad \text{wobei } x_n \geq 0.$$

Wir wenden die binomische Formel an:

$$n = (1 + x_n)^n = \sum_{k=0}^n \binom{n}{k} x_n^k \geq \binom{n}{0} x_n^0 + \binom{n}{2} x_n^2 = 1 + \frac{n(n-1)}{2} x_n^2.$$

Wenn wir diese Ungleichung  $n \geq 1 + \frac{n(n-1)}{2} x_n^2$  nach  $x_n$  umstellen, kommen wir auf

$$0 \leq x_n \leq \sqrt{\frac{2}{n}}, \quad \text{für jegliches } n \geq 17.$$

Damit sind die  $x_n$  eingeklemmt zwischen zwei Folgen, die beide nach Null streben. Also ist  $\lim_{n \rightarrow \infty} x_n = 0$  und somit  $\lim_{n \rightarrow \infty} \sqrt[n]{n} = 1$ . □

**Satz 5.54.** *Seien  $(a_n)_{n \in \mathbb{N}}$  und  $(b_n)_{n \in \mathbb{N}}$  zwei konvergente Folgen reeller Zahlen mit Grenzwerten  $a^*$  und  $b^*$ . Dann konvergiert auch die Folge  $(a_n b_n)_{n \in \mathbb{N}}$ , und ihr Grenzwert ist  $a^* b^*$ .*

*Wenn zusätzlich noch  $b_n \neq 0$  für jedes  $n \in \mathbb{N}$  und  $b^* \neq 0$ , dann konvergiert auch die Folge  $(a_n/b_n)_{n \in \mathbb{N}}$ , und ihr Grenzwert ist  $a^*/b^*$ .*

*Beweis.* Wir beweisen nur die Aussage über das Produkt, die zweite Aussage läßt sich ähnlich zeigen. Die entscheidende Idee ist das Einfügen einer *fruchtbaren Null*:

$$|a_n b_n - a^* b^*| = |a_n b_n - a_n b^* + a_n b^* - a^* b^*| \leq |(a_n - a^*) b^*| + |a_n (b_n - b^*)|.$$

Nun sind die Folgen  $(a_n)_{n \in \mathbb{N}}$  und  $(b_n)_{n \in \mathbb{N}}$  beschränkt (wegen Satz 5.22 und Satz 5.23), also

$$|a_n| \leq A, \quad |b_n| \leq B, \quad n \in \mathbb{N}.$$

Dieselbe Beschränkung gilt dann auch für die Grenzwerte. Wir haben somit

$$|a_n b_n - a^* b^*| \leq B |a_n - a^*| + A |b_n - b^*|,$$

woraus sich die Behauptung unmittelbar ergibt. □

**Bemerkung 5.55.** *Mit einem ähnlichen Beweis kann man zeigen: die Produktfolge einer Nullfolge und einer beschränkten Folge ist wieder eine Nullfolge. Hierbei verstehen wir unter einer Nullfolge eine solche Folge, die nach Null strebt.*

**Beispiel 5.56.** *Seien  $k, x \in \mathbb{R}$ . Die Folge  $(n^k x^n)_{n \in \mathbb{N}}$  konvergiert gegen 0 für  $|x| < 1$  und divergiert für  $|x| > 1$ . Wir fassen dies zusammen in folgender Merkregel (wobei wir sprachlich etwas unscharf sind):*

*Im Konfliktfall ist ein Exponentialterm stärker als ein Potenzterm.*

### 5.3.3 Nichtabsolute Konvergenz und Umordnungen

Wir kennen aus dem Majorantenkriterium bereits einen Mechanismus, der die Konvergenz einer Reihe bewirkt: die Folge der Summanden klingt *schnell genug* ab.

Es gibt noch einen weiteren Mechanismus, der die Konvergenz einer Reihe nach sich zieht: die *Auslöschung*.

**Satz 5.57 (Leibniz<sup>33</sup>-Kriterium).** Sei  $\sum_{n=0}^{\infty} a_n$  eine Reihe reeller Zahlen. Wenn die  $a_n$  die folgenden Bedingungen erfüllen:

- die Folge der  $|a_n|$  strebt streng monoton nach Null,
- $a_n a_{n+1} < 0$  für jedes  $n \in \mathbb{N}$ ,

dann konvergiert die Reihe  $\sum_{n=0}^{\infty} a_n$ .

Man beachte, daß die Folge der Summanden  $a_n$  auch extrem langsam nach Null streben darf, ohne die Reihenkonvergenz in Gefahr zu bringen.

*Beweis.* Die zweite Bedingung besagt, daß die Summanden  $a_n$  alternierende Vorzeichen haben. Zum Beweis betrachten wir die Partialsummen  $s_n$ . Seien die  $a_{2n}$  positiv, und die  $a_{2n+1}$  negativ. Dann haben wir

$$s_{2n+2} - s_{2n} = a_{2n+2} + a_{2n+1} < 0,$$

also ist die Folge  $(s_0, s_2, s_4, \dots)$  streng monoton fallend. Analog zeigt man, daß die Folge  $(s_1, s_3, s_5, \dots)$  streng monoton steigend ist. Weiterhin ist  $s_{n+2} - s_{n+1} > 0$ , sodaß die fallende Folge  $(s_0, s_2, \dots)$  nach unten beschränkt ist durch  $s_1$ . Also konvergiert sie. Analog ist die wachsende Folge  $(s_1, s_3, \dots)$  nach oben beschränkt durch  $s_0$ , also konvergent. Beide Grenzwerte müssen übereinstimmen wegen  $a_n \rightarrow 0$ .  $\square$

Wir bleiben noch etwas bei solchen Reihen mit Summanden wechselnden Vorzeichens. Es gilt (wie wir später sehen werden)

$$\ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + \frac{1}{9} - \frac{1}{10} \mp \dots$$

Durch Verdoppelung ist dann

$$\begin{aligned} 2 \ln 2 &= 2 - \frac{2}{2} + \frac{2}{3} - \frac{2}{4} + \frac{2}{5} - \frac{2}{6} + \frac{2}{7} - \frac{2}{8} + \frac{2}{9} - \frac{2}{10} \mp \dots \\ &= 2 - 1 + \frac{2}{3} - \frac{1}{2} + \frac{2}{5} - \frac{1}{3} + \frac{2}{7} - \frac{1}{4} + \frac{2}{9} - \frac{1}{5} \mp \dots \end{aligned}$$

Wenn man die Brüche mit gleichem Nenner zusammenfaßt, folgt

$$2 \ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + \frac{1}{9} - \frac{1}{10} \mp \dots = \ln 2.$$

Wo steckt der Fehler?

**Satz 5.58.** Sei  $S$  eine beliebige reelle Zahl. Dann gibt es eine Reihe  $\sum_{n=0}^{\infty} a_n$ , die sich durch Umsortieren der Summanden der Reihe  $\sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n}$  ergibt und den Reihenwert  $S$  hat.

Die Umsortierung hängt natürlich von  $S$  ab. Der Satz besagt also, daß zumindest bei dieser Reihe das Kommutativgesetz und das Assoziativgesetz leider nicht gelten.

*Beweisskizze.* Die Teilreihen  $1 + \frac{1}{3} + \frac{1}{5} + \dots$  und  $-\frac{1}{2} - \frac{1}{4} - \frac{1}{6} - \dots$  der positiven und negativen Summanden divergieren beide.  $\square$

Dieses Phänomen kann nicht auftreten, wenn die Ausgangsreihe *absolut* konvergiert.

**Satz 5.59.** Sei  $\sum_{n=0}^{\infty} a_n$  eine absolut konvergierende Reihe. Dann konvergiert auch jede Umordnung dieser Reihe absolut und hat denselben Reihenwert. Es gilt auch die Umkehrung: wenn jede Umordnung der Reihe denselben Reihenwert hat, dann ist die Ausgangsreihe absolut konvergent.

<sup>33</sup> GOTTFRIED WILHELM VON LEIBNIZ, 1646–1716

*Beweis.* Lassen wir weg. □

Die Frage der Konvergenz von umgeordneten Reihen ist keineswegs akademisch. Wenn man zum Beispiel zwei Reihen  $\sum_{n=0}^{\infty} a_n$  und  $\sum_{k=0}^{\infty} b_k$  multiplizieren will, dann möchte man die einzelnen Produkte  $a_n b_k$  bilden, in eine geeignete Reihenfolge bringen, und in dieser Reihenfolge dann aufsummieren. Nun gibt es aber keine Reihenfolge dieser Teilprodukte, die „richtiger“ wäre als eine andere Reihenfolge. Falls bei einer anderen Anordnung der Teilprodukte sich ein anderer Reihenwert ergäbe, wäre dies natürlich schlecht.

**Satz 5.60.** *Es seien  $\sum_{n=0}^{\infty} a_n$  und  $\sum_{k=0}^{\infty} b_k$  zwei absolut konvergente Reihen. Dann konvergiert die „Reihe“  $\sum_{n,k=0}^{\infty} a_n b_k$  bei beliebiger Anordnung der Summanden  $a_n b_k$  gegen denselben Wert, nämlich  $(\sum_{n=0}^{\infty} a_n)(\sum_{k=0}^{\infty} b_k)$ .*

*Beweis.* Lassen wir weg. □

Es hat sich eine spezielle Anordnung eingebürgert, die auf die sogenannte CAUCHY–Produktreihe führt. Das Konzept der Cauchy–Produktreihe wird sich als ganz natürlich herausstellen, wenn wir uns im nächsten Abschnitt Potenzreihen anschauen.

**Definition 5.61.** *Es seien  $\sum_{n=0}^{\infty} a_n$  und  $\sum_{k=0}^{\infty} b_k$  zwei absolut konvergente Reihen. Wir definieren eine Folge  $(c_m)_{m \in \mathbb{N}_0}$  gemäß*

$$c_m := \sum_{k+n=m} a_n b_k = a_0 b_m + a_1 b_{m-1} + \cdots + a_{m-1} b_1 + a_m b_0, \quad m \in \mathbb{N}_0.$$

Dann heißt  $\sum_{m=0}^{\infty} c_m$  die Cauchy–Produktreihe der beiden Reihen  $\sum_{n=0}^{\infty} a_n$  und  $\sum_{k=0}^{\infty} b_k$ .

## 5.4 Potenzreihen

**Definition 5.62.** *Unter einer Potenzreihe<sup>34</sup> verstehen wir eine Reihe der Form*

$$\sum_{n=0}^{\infty} a_n (z - z_0)^n, \quad a_n, z, z_0 \in \mathbb{C}.$$

Hierbei denken wir uns  $z_0$  als fixiert und  $z$  als variabel. Für den Summanden mit  $n = 0$  vereinbaren wir, daß  $0^0 := 1$ .

Ein Beispiel ist die geometrische Reihe

$$\sum_{n=0}^{\infty} z^n, \quad z \in \mathbb{C}.$$

Diese konvergiert genau für  $|z| < 1$ , wie man zum Beispiel mithilfe des Wurzelkriteriums sieht. Das Konvergenzgebiet ist in diesem Falle also eine Kreisscheibe mit Radius 1 um  $z_0 = 0$ . Gemäß Definition 5.10 verwenden wir für diese Kreisscheibe die Schreibweise  $B_1(0)$ .

Diese Potenzreihe kann man von 2 Standpunkten aus betrachten.

- Einerseits kann man jeden Summanden  $z^n$  als stetige Funktion ansehen, also zum Beispiel als Element eines Vektorraumes  $U = C(B_R(0) \rightarrow \mathbb{C})$  (wobei  $R$  eine passend gewählte positive Zahl sei), diesen Raum mit der  $\|\cdot\|_{\infty}$ -Norm ausstatten (wodurch er zu einem Banachraum wird), und dann nach der Konvergenz der Reihe in diesem Banachraum fragen.
- Andererseits könnte man  $z \in \mathbb{C}$  festhalten. Dann erhält man eine Reihe von komplexen Zahlen, die konvergieren kann oder auch nicht.

Wir vermerken nur kurz, daß die geometrische Reihe genau dann im Banachraum  $U = C(B_R(0) \rightarrow \mathbb{C})$  konvergiert, wenn  $R < 1$ , und werden uns ab jetzt auf den zweiten Aspekt konzentrieren.

---

<sup>34</sup>power series

**Definition 5.63.** Sei  $\sum_{n=0}^{\infty} a_n(z-z_0)^n$  eine Potenzreihe, und sei  $M \subset \mathbb{C}$  eine beliebige Menge. Wir sagen, daß die Potenzreihe auf der Menge  $M$  gleichmäßig gegen  $P(z)$  konvergiert<sup>35</sup>, wenn gilt:

$$\forall \varepsilon > 0 \exists N_0(\varepsilon) : \forall n \geq N_0(\varepsilon), \forall z \in M : \left| \sum_{k=0}^n a_k(z-z_0)^k - P(z) \right| < \varepsilon.$$

Entscheidend ist dabei, daß die Schranke  $N_0(\varepsilon)$  nicht von  $z \in M$  abhängt, sondern für alle solchen  $z$  dasselbe  $N_0(\varepsilon)$  verwendet werden kann. Dann kann man die letzte Formelzeile umschreiben zu

$$\forall \varepsilon > 0 \exists N_0(\varepsilon) : \forall n \geq N_0(\varepsilon) : \sup_{z \in M} \left| \sum_{k=0}^n a_k(z-z_0)^k - P(z) \right| < \varepsilon,$$

und das wiederum kann logisch äquivalent umformuliert werden zu

$$\forall \varepsilon > 0 \exists N_0(\varepsilon) : \forall n \geq N_0(\varepsilon) : \left\| \sum_{k=0}^n a_k(\cdot - z_0)^k - P(\cdot) \right\|_{\infty} < \varepsilon,$$

denn genauso ist die  $\|\cdot\|_{\infty}$ -Norm in Bezug auf die Menge  $M$  definiert. Weil wir es jetzt mit Funktionen zu tun haben, ist das  $z$  in der Notation verschwunden und ein Joker-Zeichen  $\cdot$  an seine Stelle getreten.

*Die gleichmäßige Konvergenz ist gleichbedeutend mit der Konvergenz in der  $\|\cdot\|_{\infty}$ -Norm.*

**Beispiel 5.64.** Wir betrachten nochmal die geometrische Reihe  $\sum_{n=0}^{\infty} z^n$ . Diese konvergiert auf der Menge  $M_1 := \{z \in \mathbb{C} : |z| < 1\}$ , aber dort nicht gleichmäßig. Die Konvergenz wird beliebig langsam, wenn  $z$  nach 1 strebt. Wir haben aber gleichmäßige Konvergenz, wenn wir die Kreisscheibe  $M_1$  durch eine kleinere Kreisscheibe  $M_r := \{z \in \mathbb{C} : |z| < r\}$  mit  $r < 1$  ersetzen (warum?).

**Satz 5.65.** Sei  $\sum_{n=0}^{\infty} a_n(z-z_0)^n$  eine Potenzreihe. Sei eine reelle Zahl  $t$  definiert durch

$$t := \lim_{n \rightarrow \infty} \sqrt[n]{|a_n|},$$

wobei wir voraussetzen, daß dieser Limes existiert oder daß die Folge der  $\sqrt[n]{|a_n|}$  gegen  $+\infty$  divergiert. Dann gilt:

1. die Potenzreihe konvergiert für jedes  $z \in \mathbb{C}$  mit  $|z-z_0| < \frac{1}{t}$  absolut (aber nicht unbedingt gleichmäßig),
2. die Potenzreihe konvergiert in jeder kleineren Kreisscheibe  $M_r := \{z \in \mathbb{C} : |z-z_0| < r\}$  ( $r < \frac{1}{t}$ ) gleichmäßig,
3. die Potenzreihe divergiert für jedes  $z \in \mathbb{C}$  mit  $|z-z_0| > \frac{1}{t}$ ,
4. wenn  $t = 0$  ist, dann konvergiert die Potenzreihe absolut in ganz  $\mathbb{C}$ , und auf jeder kompakten Teilmenge von  $\mathbb{C}$  konvergiert sie gleichmäßig,
5. wenn  $t = \infty$ , dann konvergiert die Potenzreihe nur für  $z = z_0$ ,
6. im Konvergenzgebiet stellt die Potenzreihe eine stetige Funktion dar.

Man braucht nicht voraussetzen, daß die Folge  $\sqrt[n]{|a_n|}$  einen Grenzwert hat. Es reicht,  $t$  als größten Häufungspunkt der Menge  $\{\sqrt[n]{|a_n|} : n \in \mathbb{N}\}$  zu definieren, und die restlichen Aussagen des Satzes gelten nach wie vor.

Wenn jeder Koeffizient  $a_n \neq 0$  ist, dann kann man die Zahl  $t$  auch über die (gelegentlich einfacher handhabbare) Formel

$$t := \lim_{n \rightarrow \infty} \frac{|a_{n+1}|}{|a_n|}$$

bestimmen. Falls dieser Limes existieren sollte, dann existiert auch der Grenzwert  $\lim_{n \rightarrow \infty} \sqrt[n]{|a_n|}$  und beide sind gleich. Es kann übrigens passieren, daß  $\lim_{n \rightarrow \infty} \frac{|a_{n+1}|}{|a_n|}$  nicht existiert, aber  $\lim_{n \rightarrow \infty} \sqrt[n]{|a_n|}$  gibt es doch. Ein Beispiel ist  $(a_0, a_1, a_2, a_3, \dots) = (1, 2, 1, 2, 1, 2, \dots)$ .

Die Zahl  $\frac{1}{t}$  heißt auch *Konvergenzradius*.

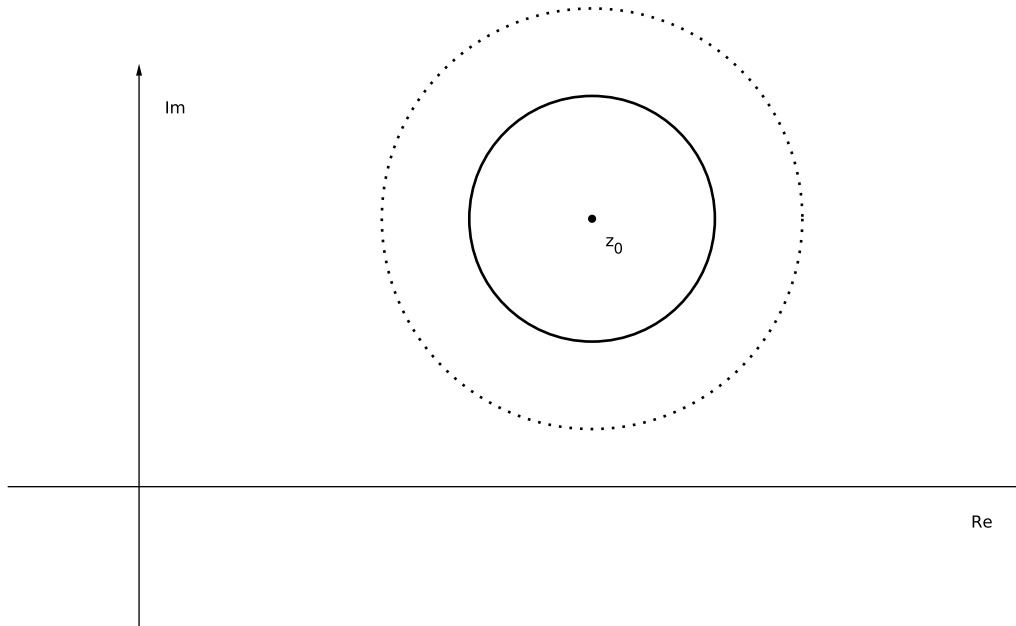


Abbildung 5.1: Zum Konvergenzverhalten von Potenzreihen. Der gepunktete Kreis umschließt die offene Kreisscheibe  $B_{1/t}(z_0)$ , der durchgezogene Kreis umschließt die abgeschlossene Kreisscheibe  $\overline{B_r(z_0)}$ , mit  $0 < r < 1/t$ .

*Beweis.* Für die Punkte 1 bis 5 benutzt man das Wurzelkriterium bzw. Quotientenkriterium. Punkt 6 wird später bewiesen, siehe auch Satz 5.21.  $\square$

Insgesamt erhalten wir folgendes Verhalten (vgl. Abbildung 5.1). Sei hierbei eine positive Zahl  $r$  mit  $r < 1/t$  fest gewählt.

**für**  $|z - z_0| > 1/t$ : Divergenz der Potenzreihe,

**für**  $|z - z_0| = 1/t$ : Konvergenzverhalten unbekannt,

**für**  $|z - z_0| < 1/t$ : Konvergenz ist absolut, jedoch nicht gleichmäßig,

**für**  $|z - z_0| < r$ : Konvergenz ist absolut und gleichmäßig.

**Satz 5.66.** Sei  $\sum_{n=0}^{\infty} a_n(z - z_0)^n$  eine Potenzreihe mit Konvergenzradius  $R$ . Dann hat die durch termweise Differentiation entstehende Reihe  $\sum_{n=0}^{\infty} n a_n(z - z_0)^{n-1}$  denselben Konvergenzradius.

*Beweis.* Übungsaufgabe.  $\square$

**Bemerkung 5.67.** Im zweiten Semester werden wir sehen: sei  $P = P(z)$  eine Funktion, die durch die Potenzreihe  $\sum_{n=0}^{\infty} a_n(z - z_0)^n$  dargestellt wird, für  $z \in B_{1/t}(z_0)$ . Dann ist  $P$  dort sogar differenzierbar, und wir haben die Formel

$$P'(z) = \left( \sum_{n=0}^{\infty} a_n(z - z_0)^n \right)' \stackrel{*}{=} \sum_{n=0}^{\infty} (a_n(z - z_0)^n)' = \sum_{n=0}^{\infty} n a_n(z - z_0)^{n-1}, \quad \text{falls } z \in B_{1/t}(z_0).$$

Die durch \* markierte Umformung versteht sich nicht von selbst, sondern wird noch zu beweisen sein. Die gleichmäßige Konvergenz von Potenzreihen ist hierbei ein wichtiges Hilfsmittel (aber noch nicht ganz ausreichend).

<sup>35</sup>converges uniformly to  $P(z)$

Wie multipliziert man zwei Potenzreihen? Der folgende Satz sagt uns, daß die Multiplikation in naheliegender Weise (Zusammenfassen aller gleichen Potenzen) tatsächlich das richtige Ergebnis liefert:

**Satz 5.68.** Seien  $\sum_{n=0}^{\infty} a_n(z-z_0)^n$  und  $\sum_{k=0}^{\infty} b_k(z-z_0)^k$  zwei Potenzreihen mit Konvergenzradien  $r_a$  und  $r_b$ . Wir definieren eine Folge  $(c_m)_{m \in \mathbb{N}_0}$  von Koeffizienten durch

$$c_m := \sum_{n+k=m} a_n b_k$$

und setzen  $r_c := \min(r_a, r_b)$ . Dann konvergiert die Potenzreihe  $\sum_{m=0}^{\infty} c_m(z-z_0)^m$  für  $|z-z_0| < r_c$ , und dort gilt

$$\sum_{m=0}^{\infty} c_m(z-z_0)^m = \left( \sum_{n=0}^{\infty} a_n(z-z_0)^n \right) \left( \sum_{k=0}^{\infty} b_k(z-z_0)^k \right).$$

*Beweis.* Die links stehende Reihe ist gerade die Cauchy-Produktreihe der beiden rechts stehenden Reihen, aufgrund der speziellen Definition der  $c_m$ .  $\square$

## 5.5 Beispiel: Die Exponentialfunktion

Die wichtigste Funktion der Physikerinnen und Physiker hat folgende zentrale Eigenschaften:

$$\exp(z) = \sum_{k=0}^{\infty} \frac{1}{k!} z^k, \quad z \in \mathbb{C}, \quad (5.2)$$

$$\exp(u+v) = \exp(u) \exp(v), \quad u, v \in \mathbb{C}, \quad (5.3)$$

$$\exp(z) = \lim_{n \rightarrow \infty} \left( 1 + \frac{z}{n} \right)^n, \quad z \in \mathbb{C}, \quad (5.4)$$

$$\text{wenn } x = \frac{p}{q} \in \mathbb{Q}_+, \quad p, q \in \mathbb{N}_+, \quad \text{dann } \exp(x) = \sqrt[q]{e^p}, \quad \text{wobei } e = 2.718281828459\dots, \quad (5.5)$$

$$\text{wenn } z = x + iy \in \mathbb{C}, \quad x, y \in \mathbb{R}, \quad \text{dann } \exp(z) = e^x (\cos y + i \sin y), \quad (5.6)$$

die wir jetzt zeigen werden. Gleichung (5.2) gilt, weil genau so die Exponentialfunktion definiert wird. Die (absolute und auf Kompakta gleichmäßige) Konvergenz der Reihe auf der rechten Seite ergibt sich schnell aus dem Quotientenkriterium, analog zu Beispiel 5.40.

Gleichung (5.4) stellt die Beziehung her zur stetigen Verzinsung eines Kapitals (das ist der historisch traditionsreichste Zugang zur Exponentialfunktion), und der Beweis davon wird die Hauptarbeit in diesem Abschnitt sein. Das Additionstheorem (5.3) zeigt man entweder durch banales Rechnen (Definition (5.2) einsetzen, maximal mögliches Ausmultiplizieren aller Klammern, gleiche Terme wegstreichen bis zur Gleichung  $0=0$ , Nachweis der Äquivalenz der Umformungen), was aber eine lange Rechnung werden kann, weshalb wir diesen Weg hier nicht verfolgen. Oder man beweist anstatt (5.4) gleich eine etwas allgemeinere Aussage (nämlich Lemma 5.69), und dann folgt (5.3) innerhalb weniger Zeilen. Aus dem Additionstheorem (5.3) folgt dann (5.5) sofort. Und (5.6) „zeigen“ wir später dadurch, daß wir die Funktionen  $\sin$  und  $\cos$  genauso wie in (5.6) definieren, und hinterher beweisen, daß die so definierten analytischen Winkelfunktionen die schulbekannten geometrischen Eigenschaften besitzen.

Wie angekündigt, beginnen wir mit einer Bonustrack-Version von (5.4). Der Beweis der Originalgleichung (5.4) hätte übrigens im Prinzip denselben Arbeitsaufwand erfordert.

**Lemma 5.69.** Sei  $(z_n)_{n \in \mathbb{N}}$  eine Folge komplexer Zahlen mit Grenzwert  $z_* \in \mathbb{C}$ . Dann gilt

$$\lim_{n \rightarrow \infty} \left( 1 + \frac{z_n}{n} \right)^n = \sum_{k=0}^{\infty} \frac{1}{k!} z_*^k.$$

*Beweis.* Zur Vereinfachung der Schreibweisen setzen wir

$$a_n := \left( 1 + \frac{z_n}{n} \right)^n, \quad a_* := \sum_{k=0}^{\infty} \frac{1}{k!} z_*^k.$$

Diese Reihe für  $a_*$  konvergiert (Begründung analog zu Beispiel 5.40). Für  $a_n$  haben wir aus der binomischen Formel die Darstellung

$$a_n = \sum_{k=0}^n \binom{n}{k} \left(\frac{z_n}{n}\right)^k.$$

Hier würden wir jetzt gern in jedem Summanden separat zur Grenze  $n \rightarrow \infty$  übergehen wollen. Wir halten also ein  $k$  fest und untersuchen die Folge der  $k$ -ten Summanden auf ihr Verhalten für  $n \rightarrow \infty$ :

$$\begin{aligned} \lim_{n \rightarrow \infty} \binom{n}{k} \left(\frac{z_n}{n}\right)^k &= \lim_{n \rightarrow \infty} \frac{n(n-1) \cdot \dots \cdot (n-k+1)}{k!} \cdot \frac{z_n^k}{n^k} \\ &= \lim_{n \rightarrow \infty} \frac{n(n-1) \cdot \dots \cdot (n-k+1)}{n^k} \cdot \frac{z_n^k}{k!} \\ &= \left( \lim_{n \rightarrow \infty} \frac{n(n-1) \cdot \dots \cdot (n-k+1)}{n^k} \right) \cdot \left( \lim_{n \rightarrow \infty} \frac{z_n^k}{k!} \right) \\ &= 1 \cdot \frac{z_*^k}{k!}, \end{aligned}$$

aber jetzt haben wir noch die Schwierigkeit zu bewältigen, daß die Summanden in der Summe  $a_n = \sum_{k=0}^n \dots$  immer zahlreicher werden für  $n \rightarrow \infty$ . Diesen Beweisversuch brechen wir ab.

Für einen zweiten Beweisversuch gehen wir auf die Definition der Konvergenz zurück: für jedes  $\varepsilon > 0$  suchen wir ein  $N_0(\varepsilon)$  mit

$$|a_* - a_n| < \varepsilon \quad \forall n \geq N_0(\varepsilon).$$

Sei  $N \in \mathbb{N}$  groß, und sei  $n > N$ . Dann haben wir die Zerlegung

$$|a_* - a_n| \leq \left| \sum_{k=0}^N \frac{1}{k!} z_*^k - \sum_{k=0}^n \binom{n}{k} \left(\frac{z_n}{n}\right)^k \right| + \left| \sum_{k=N+1}^{\infty} \frac{1}{k!} z_*^k \right| + \left| \sum_{k=N+1}^n \binom{n}{k} \left(\frac{z_n}{n}\right)^k \right|.$$

Nun ist  $\sum_{k=0}^{\infty} \frac{1}{k!} z_*^k$  konvergent (mit Reihensumme  $a_*$ ), wie schon eingangs geschrieben. Also muß das Reihenendstück  $|\sum_{k=N+1}^{\infty} \dots|$  (also der mittlere Summand auf der rechten Seite) kleiner sein als  $\frac{1}{3}\varepsilon$ , wenn wir  $N$  groß genug wählen.

Weiterhin können wir annehmen, daß  $|z_n| \leq |z_*| + 1$  für alle  $n > N$ , denn die Folge der  $z_n$  konvergiert nach  $z_*$ , siehe auch Satz 5.23. Also ist

$$\begin{aligned} \left| \sum_{k=N+1}^n \binom{n}{k} \left(\frac{z_n}{n}\right)^k \right| &\leq \sum_{k=N+1}^n \frac{n(n-1) \cdot \dots \cdot (n-k+1)}{n^k} \cdot \frac{|z_n|^k}{k!} \leq \sum_{k=N+1}^n \frac{n^k}{n^k} \cdot \frac{(|z_*| + 1)^k}{k!} \\ &< \sum_{k=N+1}^{\infty} \frac{(|z_*| + 1)^k}{k!} \leq \frac{\varepsilon}{3}, \end{aligned}$$

wenn  $N$  groß ist. Nun halten wir  $\varepsilon$  und  $N$  fest und kümmern uns um den ersten Summanden:

$$\left| \sum_{k=0}^N \frac{1}{k!} z_*^k - \sum_{k=0}^n \binom{n}{k} \left(\frac{z_n}{n}\right)^k \right| \leq \sum_{k=0}^N \left| \frac{1}{k!} z_*^k - \binom{n}{k} \left(\frac{z_n}{n}\right)^k \right|$$

(wegen der Dreiecksungleichung), und solche Differenzen hatten wir schon bei unserem ersten Beweisversuch erforscht. Wir finden also ein  $N_0(\varepsilon) \gg N$ , sodaß für  $n \geq N_0(\varepsilon)$  gilt:

$$\sum_{k=0}^N \left| \frac{1}{k!} z_*^k - \binom{n}{k} \left(\frac{z_n}{n}\right)^k \right| \leq \frac{\varepsilon}{3}.$$

Damit ist der Beweis komplett. □

**Definition 5.70 (Endgültige Definition der Exponentialfunktion).** Die EULER<sup>36</sup>sche Zahl  $e = 2.718281828459045 \dots$  wird definiert durch

$$e := \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = \sum_{k=0}^{\infty} \frac{1}{k!}.$$

<sup>36</sup>LEONHARD EULER, 1707–1783



Die Exponentialfunktion  $z \mapsto \exp(z)$  definieren wir (anders als im ersten Kapitel) durch

$$\exp(z) := \sum_{k=0}^{\infty} \frac{1}{k!} z^k, \quad z \in \mathbb{C}.$$

Nun beweisen wir (5.3).

**Satz 5.71 (Additionstheorem).** Die Exponentialfunktion erfüllt

$$\exp(u+v) = \exp(u) \exp(v), \quad u, v \in \mathbb{C}.$$

*Beweis.* Es ist

$$\begin{aligned} \exp(u) \cdot \exp(v) &= \left( \lim_{n \rightarrow \infty} \left(1 + \frac{u}{n}\right)^n \right) \cdot \left( \lim_{n \rightarrow \infty} \left(1 + \frac{v}{n}\right)^n \right) && \left| \text{wegen Lemma 5.69} \right. \\ &= \lim_{n \rightarrow \infty} \left( \left(1 + \frac{u}{n}\right)^n \left(1 + \frac{v}{n}\right)^n \right) && \left| \text{wegen Satz 5.54} \right. \\ &= \lim_{n \rightarrow \infty} \left(1 + \frac{u}{n} + \frac{v}{n} + \frac{uv}{n^2}\right)^n \\ &= \lim_{n \rightarrow \infty} \left(1 + \frac{u+v+\frac{uv}{n}}{n}\right)^n. \end{aligned}$$

Wir setzen  $z_n := u + v + \frac{uv}{n} \in \mathbb{C}$ . Dann ist  $\lim_{n \rightarrow \infty} z_n = z_* := u + v$ , und nach Lemma 5.69 haben wir

$$\lim_{n \rightarrow \infty} \left(1 + \frac{u+v+\frac{uv}{n}}{n}\right)^n = \sum_{k=0}^{\infty} \frac{1}{k!} (u+v)^k = \exp(u+v),$$

was zu beweisen war. □

Wir kommen jetzt zu (5.5).

Aus dem Additionstheorem haben wir zum Beispiel  $\exp(2z) = \exp(z) \exp(z) = (\exp(z))^2$ , und nach dem Prinzip der vollständigen Induktion können wir zeigen, daß

$$\exp(mz) = (\exp(z))^m, \quad z \in \mathbb{C}, \quad m \in \mathbb{N}_+. \quad (5.7)$$

Wenn wir hierin  $z := \frac{1}{q}$  und  $m := q$  setzen, bekommen wir  $e = \exp(1) = (\exp(1/q))^q$ , also

$$\exp(1/q) = \sqrt[q]{e} = e^{1/q}, \quad q \in \mathbb{N}_+.$$

Wenn wir (5.7) nocheinmal verwenden, mit  $m = p$  und  $z = 1/q$ , dann folgt

$$\exp(p/q) = (\exp(1/q))^p = e^{\frac{p}{q}}, \quad p, q \in \mathbb{N}_+.$$

Wir haben also die Gleichung  $\exp(x) = e^x$  gezeigt für alle positiven rationalen Zahlen. Analog zeigt man diese Gleichung für negative rationale Zahlen. Nun ist  $z \mapsto \exp(z)$  eine stetige Funktion (wie wir im Satz 5.65 vorhin noch nicht bewiesen haben), und man definiert  $x \mapsto e^x$  für irrationale  $x$  als Grenzwert der Folge  $x_n \mapsto e^{x_n}$ , wobei die  $x_n$  eine Folge rationaler Zahlen sind, die gegen  $x$  streben.

Damit haben wir gezeigt:

**Satz 5.72.** Für  $x \in \mathbb{R}$  gilt:

$$e^x = \sum_{n=0}^{\infty} \frac{1}{n!} x^n = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n.$$

Wenn wir in Zukunft  $e^z$  schreiben mit  $z \in \mathbb{C}$ , dann meinen wir damit stets  $\exp(z)$ . Später werden wir dafür sorgen, daß

$$\exp(i\varphi) = \cos(\varphi) + i \sin(\varphi), \quad \varphi \in \mathbb{R},$$

gilt. Diese Formel hatten wir bisher einige Male benutzt, aber noch nicht bewiesen.

## 5.6 Schlüsselbegriffe

- Definition und Beispiele für Normen,
- Begriffe Konvergenz, Häufungspunkt, offene und abgeschlossene Mengen,
- Definition Cauchy-Folge,
- Definition Reihe, Summenformel für die geometrische Reihe,
- Quotientenkriterium, Wurzelkriterium, Leibnizkriterium,
- Supremum/Infimum im Gegensatz zu Maximum/Minimum,
- Definition absolute und gleichmäßige Konvergenz,
- Konvergenzradius von Potenzreihen,
- Definition der Exponentialfunktion.

# Kapitel 6

## Funktionen

**Definition 6.1.** Eine eindeutige Abbildung  $f$  von einer Menge  $D_f$  in eine Menge  $V$  heißt Funktion<sup>1</sup>. Die Menge  $D_f$  heißt Definitionsbereich<sup>2</sup> von  $f$ . Wenn  $f$  ein  $x \in D_f$  auf  $y \in V$  abbildet, dann heißt  $y$  der Wert<sup>3</sup> der Funktion  $f$  an der Stelle  $x$ . Der Wert  $y$  heißt auch Bild von  $x$ .

Die Menge  $W_f \subset V$  sämtlicher tatsächlich angenommener Funktionswerte heißt Wertebereich<sup>4</sup> von  $f$ .

Sei  $W \subset V$  eine beliebige Menge. Die Menge aller  $x \in D_f$  mit  $f(x) \in W$  heißt Urbild<sup>5</sup> von  $W$ .

Hierbei verstehen wir unter *eindeutig*, daß jedem  $x \in D_f$  genau ein Wert  $y = f(x)$  zugeordnet wird. Der populären Schreibweise  $\sqrt{9} = \pm 3$  wollen wir uns hier *nicht* anschließen. In diesem Skript ist  $\sqrt{9} = 3$ .

Die Mengen  $D_f$  und  $V$  können ganz beliebige Mengen sein. In den meisten Fällen wird es sich in dieser Vorlesung um Teilmengen von  $\mathbb{R}$  oder  $\mathbb{C}$  handeln, aber man kann auch beliebige Banachräume betrachten.

**Beispiel 6.2.** Eine eindeutige Abbildung von  $\mathbb{N}$  nach  $\mathbb{R}$  ist eine Funktion. Diese Abbildung könnte man schreiben als  $n \mapsto a_n$ . Man sagt dazu auch „Folge reeller Zahlen“. Das ist uns eine eigene Merkregel wert:

*Jede Folge kann als eine Funktion angesehen werden.*

Ab jetzt bezeichnen wir mit  $z$  und  $w$  stets komplexe Zahlen und mit  $x, y, u, v$  stets reelle Zahlen.

### 6.1 Grenzwerte von Funktionen

Wie in Beispiel 6.2 schon dargelegt, können wir mit ein wenig Phantasie Folgen als Funktionen interpretieren. Und dann liegt es nahe, die Definition des Grenzwertes einer Folge (Definition 5.2) als Anleitung zu benutzen, um eine Definition des Grenzwertes einer Funktion zu gewinnen. Wir gehen erst sprachlich vor, und dann wird die formelhafte Definition uns fast entgegenfallen.

Zuerst erinnern wir an die Bedeutung der Schreibweise  $\lim_{n \rightarrow \infty} a_n = a^*$ :

*Ein Element  $a^*$  ist Grenzwert einer Folge  $(a_n)_{n \in \mathbb{N}}$ ,  
wenn es zu jeder Umgebung von  $a^*$  ein  $N_0$  gibt,  
sodaß das Folgenendstück ab  $N_0$  in dieser Umgebung liegt.*

Wir werden kreativ und vereinbaren die Sprechweise, daß die Menge  $\{N_0, N_0 + 1, \dots\}$  von natürlichen Zahlen eine „Umgebung von  $\infty$ “ ist. (Selbstverständlich ist  $\infty \notin \mathbb{N}$ .) Weiterhin erinnern wir daran, daß die Folge beschrieben wird durch die Vorschrift  $n \mapsto a_n$ . Damit bekommen wir die Neuformulierung:

---

<sup>1</sup>function  
<sup>2</sup>domain  
<sup>3</sup>value  
<sup>4</sup>range  
<sup>5</sup>pre-image

Ein Element  $a^*$  ist Grenzwert einer Folge  $(a_n)_{n \in \mathbb{N}}$ ,  
wenn es zu jeder Umgebung von  $a^*$  eine Umgebung von  $\infty$  gibt,  
sodaß die Folgenreihe  $n \mapsto a_n$  diese Umgebung von  $\infty$  in jene Umgebung von  $a^*$  abbildet.

Und nun ersetzen wir sprachlich:

$$f^* \leftrightarrow a^*, \quad \text{Funktion} \leftrightarrow \text{Folge}, \quad z \leftrightarrow n, \quad f \leftrightarrow a, \quad z^* \leftrightarrow \infty,$$

und erhalten die Bedeutung der Schreibweise  $\lim_{z \rightarrow z^*} f(z) = f^*$ :

Ein Element  $f^*$  ist Grenzwert einer Funktion  $f$  im Punkt  $z^*$ ,  
wenn es zu jeder Umgebung von  $f^*$  eine Umgebung von  $z^*$  gibt,  
sodaß die Funktionsvorschrift  $z \mapsto f(z)$  diese Umgebung von  $z^*$  in jene Umgebung von  $f^*$  abbildet.

Um nun zur Formeldefinition zu kommen, denken wir nur noch daran, daß die genannte Umgebung von  $f^*$  ein  $\varepsilon$ -Ball um  $f^* \in \mathbb{C}$  sein kann, und die genannte Umgebung von  $z^*$  wird ein  $\delta_0$ -Ball um  $z^* \in \mathbb{C}$  sein. Und um es ganz exakt zu machen, denken wir noch über die Definitionsbereiche nach: offensichtlich muß  $z \in D_f$  sein, ansonsten gäbe es  $f(z)$  gar nicht. Andererseits braucht  $z^*$  nicht in  $D_f$  liegen; aber ein Häufungspunkt von  $D_f$  muß  $z^*$  schon sein, ansonsten können die  $z$  nicht gegen  $z^*$  konvergieren und gleichzeitig im Definitionsbereich von  $f$  verbleiben.

**Definition 6.3.** Sei  $D_f \subset \mathbb{C}$  eine Menge, und  $f: D_f \rightarrow \mathbb{C}$  eine Funktion. Sei  $z^* \in \mathbb{C}$  ein Häufungspunkt von  $D_f$ . Wir sagen, daß die Funktion  $f$  im Punkt  $z^*$  den Grenzwert<sup>6</sup>  $f^* \in \mathbb{C}$  hat, wenn gilt:

$$\forall \varepsilon > 0 \exists \delta_0(\varepsilon) > 0 : \forall z \in D_f \text{ mit } z \neq z^* \text{ und } |z - z^*| < \delta_0(\varepsilon) : |f(z) - f^*| < \varepsilon.$$

Wir vergleichen diese Definition mit der des Grenzwerts einer Folge:

$$\begin{aligned} \lim_{n \rightarrow \infty} a_n = a^* &\iff \forall \varepsilon > 0 \exists N_0(\varepsilon) > 0 : \forall n \quad \text{mit } n \geq N_0(\varepsilon) : |a_n - a^*| < \varepsilon, \\ \lim_{z \rightarrow z^*} f(z) = f^* &\iff \forall \varepsilon > 0 \exists \delta_0(\varepsilon) > 0 : \forall z \in D_f \setminus \{z^*\} \quad \text{mit } |z - z^*| < \delta_0(\varepsilon) : |f(z) - f^*| < \varepsilon. \end{aligned}$$

Mit den oben angeführten Ersetzungen gehen diese beiden Definitionen ineinander über. Die Menge  $\{z \in D_f : z \neq z^*, |z - z^*| < \delta_0(\varepsilon)\}$  beschreibt eine (punktierte) Umgebung von  $z^*$ , während die Menge  $\{n : n \geq N_0(\varepsilon)\}$  eine Umgebung von  $\infty$  beschreibt.

**Bemerkung 6.4.** Weil isolierte Punkte (auch Einsiedlerpunkte genannt) keine Häufungspunkte des Definitionsbereiches sind, ist ein Grenzwert dort nicht definiert (und auch nicht sinnvoll definierbar).

In Definition 6.3 tauchte in der Formelzeile die Einschränkung  $z \neq z^*$  auf, die vermutlich überraschend ist. Wir betrachten zu ihrer Erklärung die Funktion  $f: \mathbb{R} \rightarrow \mathbb{C}$ , die durch

$$f(z) := \begin{cases} 1 & : z \neq 13, \\ 37 & : z = 13 \end{cases}$$

gegeben ist. Sei  $z^* := 13$ . Der Graph von  $f$  legt den Wunsch nahe,  $\lim_{z \rightarrow z^*} f(z) = 1$  zu haben, und dies wird durch die Einschränkung  $z \neq z^*$  überhaupt erst möglich. Wenn man auf die Passage „mit  $z \neq z^*$ “ in Definition 6.3 verzichtete, dann würde  $\lim_{z \rightarrow z^*} f(z)$  gar nicht existieren.

**Bemerkung 6.5.** Die Definition eines Grenzwertes einer Funktion läßt sich direkt auf Abbildungen zwischen zwei Banachräumen  $U$  und  $V$  übertragen. Man muß nur die Betragsstriche  $|\cdot|$  durch Normstriche  $\|\cdot\|_U$  und  $\|\cdot\|_V$  ersetzen.

Den folgenden Satz beweist man genauso wie den entsprechenden Satz für Folgen (Satz 5.3).

**Satz 6.6.** Wenn eine Funktion in einem Punkt einen Grenzwert hat, dann ist dieser eindeutig.

Das folgende Ergebnis ist nützlich, wenn man durch Funktionen dividieren will:

**Lemma 6.7.** Es sei  $\lim_{z \rightarrow z^*} f(z) = f^*$ , wobei  $f^* \neq 0$  ist. Dann finden wir eine gelochte Umgebung  $\{z : z \in D_f, 0 < |z - z^*| < \delta\}$ , sodaß in dieser Umgebung  $f(z) \neq 0$  gilt.

<sup>6</sup>limit

*Beweis.* Man wähle  $\varepsilon = \frac{1}{2}|f^*|$  und beachte die „Dreiecksungleichung nach unten“ in (5.1).  $\square$

Es gibt noch eine zweite Definition des Grenzwertes; aber wir werden hier nicht beweisen, daß beide Definitionen äquivalent sind:

**Satz 6.8.** Die Funktion  $f$  hat in einem Häufungspunkt  $z^*$  des Definitionsbereiches den Grenzwert  $f^*$  genau dann, wenn gilt:

Für jede Folge  $(z_n)_{n \in \mathbb{N}}$  mit  $z_n \in D_f \setminus \{z^*\}$  (für jedes  $n$ ) und  $\lim_{n \rightarrow \infty} z_n = z^*$  ist  $\lim_{n \rightarrow \infty} f(z_n) = f^*$ .

Die folgenden Definitionen sogenannter *uneigentlicher Grenzwerte* und *einseitiger Grenzwerte* beziehen sich nur auf reelle Funktionen.

**Definition 6.9.** Sei  $f: \mathbb{R} \rightarrow \mathbb{R}$  eine Funktion und  $f^* \in \mathbb{R}$ . Wir sagen, daß  $f$  gegen  $f^*$  strebt für  $x \rightarrow +\infty$ ,  $\lim_{x \rightarrow +\infty} f(x) = f^*$ , wenn:

$$\forall \varepsilon > 0 \exists R_0(\varepsilon) > 0 : \forall x \geq R_0(\varepsilon) : |f(x) - f^*| < \varepsilon.$$

Analog schreiben wir  $\lim_{x \rightarrow -\infty} f(x) = f^*$ , wenn

$$\forall \varepsilon > 0 \exists R_0(\varepsilon) > 0 : \forall x \leq -R_0(\varepsilon) : |f(x) - f^*| < \varepsilon.$$

Unter *rechtsseitigen* bzw. *linksseitigen* Grenzwerten  $\lim_{x \rightarrow x^*+} f(x)$  bzw.  $\lim_{x \rightarrow x^*-} f(x)$  verstehen wir die üblichen Grenzwerte einer Funktion an einer Stelle  $x^*$ , wobei die Laufvariable sich nur von rechts bzw. nur von links an  $x^*$  annähern darf.

Schließlich definieren wir noch die *bestimmte Divergenz einer Funktion gegen  $+\infty$  oder  $-\infty$* :

**Definition 6.10.** Sei  $f: D_f \rightarrow \mathbb{R}$  eine Funktion und  $x^*$  ein Häufungspunkt von  $D_f$ . Wir sagen, daß  $f$  bestimmt gegen  $+\infty$  divergiert für  $x \rightarrow x^*$ , wenn gilt:

$$\forall R > 0 \exists \delta_0(R) > 0 : \forall x \in D_f \setminus \{x^*\}, |x - x^*| < \delta_0(R) : f(x) > R.$$

Analog definieren wir bestimmte Divergenz gegen  $-\infty$ .

Sinngemäß definiert man einseitige bestimmte Divergenz.

**Warnung 6.11.** Die populäre Redeweise „ $1/x^2$  strebt/konvergiert nach  $+\infty$  für  $x \rightarrow 0$ “ ist unsinnig. Präziser ist: „ $1/x^2$  divergiert (bestimmt) nach  $+\infty$  für  $x \rightarrow 0$ “.

Für das Rechnen mit Grenzwerten haben wir das folgende Ergebnis.

**Satz 6.12.** In jeder der folgenden Formeln steht  $\lim$  überall für eines der Symbole  $\lim_{x \rightarrow x^*}$ ,  $\lim_{x \rightarrow x^*+}$ ,  $\lim_{x \rightarrow x^*-}$ ,  $\lim_{x \rightarrow -\infty}$ ,  $\lim_{x \rightarrow +\infty}$  im Sinne der obigen Definitionen, jedoch **nicht** Definition 6.10.

Wenn  $\lim f(x)$  und  $\lim g(x)$  existieren, dann existieren auch

1.  $\lim(\alpha f(x) + \beta g(x)) = \alpha \lim f(x) + \beta \lim g(x)$ ,
2.  $\lim(f(x)g(x)) = (\lim f(x))(\lim g(x))$ ,
3.  $\lim \frac{f(x)}{g(x)} = \frac{\lim f(x)}{\lim g(x)}$ , falls  $g(x) \neq 0$ .

*Beweis.* Wegen Satz 6.8 dürfen wir die entsprechenden Sätze für Zahlenfolgen verwenden, also Satz 5.6 und Satz 5.54.  $\square$

Für Limites im Sinne von Definition 6.10 nutzen wir später die Regel von BERNOULLI–L'HOSPITAL.

Für komplexe Funktionen können einseitige Grenzwerte nicht definiert werden, weil Begriffe wie „links von  $z^*$ “ keinen Sinn haben, denn die komplexen Zahlen kann man nicht ordnen. Weiterhin hat es keinen Sinn,  $+\infty$  und  $-\infty$  zu unterscheiden (denn sie sind gleich).

Stattdessen definiert man die bestimmte Divergenz allgemein:

**Definition 6.13.** Sei  $f: D_f \rightarrow \mathbb{C}$  eine Funktion und  $z^* \in \mathbb{C}$  ein Häufungspunkt von  $D_f$ . Wir sagen, daß  $f$  bestimmt gegen  $\infty$  divergiert für  $z \rightarrow z^*$ , wenn gilt:

$$\forall R > 0 \exists \delta_0(R) > 0 : \forall z \in D_f \setminus \{z^*\}, |z - z^*| < \delta_0(R) : |f(z)| > R.$$

## 6.2 Stetigkeit

Ab jetzt betrachten wir bis auf weiteres komplexe Funktionen.

**Definition 6.14.** Sei  $f: D_f \rightarrow \mathbb{C}$  eine Funktion und  $z^* \in \mathbb{C}$  nicht nur ein Häufungspunkt des Definitionsbereiches  $D_f$ , sondern zusätzlich noch ein Element von  $D_f$ . Wir sagen, daß die Funktion  $f$  an der Stelle  $z^*$  stetig<sup>7</sup> ist, wenn folgendes gilt:

- $\lim_{z \rightarrow z^*} f(z)$  existiert,
- und  $\lim_{z \rightarrow z^*} f(z) = f(z^*)$ .

Wenn eine dieser beiden Bedingungen verletzt ist, heißt die Funktion  $f$  unstetig an der Stelle  $z^*$ <sup>8</sup>.

**Bemerkung 6.15.** Die Forderung, daß  $z^*$  ein Element von  $D_f$  und ein Häufungspunkt von  $D_f$  sein soll, klingt zunächst doppelt gemoppelt; sie ist es aber nicht: der Punkt  $z^*$  könnte ja ein Einsiedlerpunkt von  $D_f$  sein, und dann wird der Stetigkeitsbegriff sinnlos.

Wir können die Stetigkeit von  $f$  an der Stelle  $z^*$  auch als kommutatives Diagramm ausdrücken, im Sinne von  $\lim_n f(z_n) = f(\lim_n z_n)$ :

$$\left. \begin{array}{ccc} \boxed{(z_1, z_2, \dots)} & \xrightarrow{f} & \boxed{(f(z_1), f(z_2), \dots)} \\ \lim_n \downarrow & & \downarrow \lim_n \\ \boxed{z^*} & \xrightarrow{f} & \boxed{\begin{array}{l} f(\lim_n z_n) \\ = \lim_n f(z_n) \end{array}} \end{array} \right\} \quad (6.1)$$

**Beispiel 6.16.** Die Funktion  $f = f(z) = 1/z$ ,  $D_f = \mathbb{C} \setminus \{0\}$  ist im Punkt  $z^* = 0$  undefiniert und unstetig. Die Funktion  $g = g(x) = \ln x$  mit  $D_g = \mathbb{R}_+$  ist für negative Zahlen undefiniert, und es hat auch nicht viel Sinn, dort nach Stetigkeit/Unstetigkeit zu fragen.

**Definition 6.17.** Wenn  $f: D_f \rightarrow \mathbb{C}$  in jedem Punkt  $z^* \in D_f$  stetig ist, dann heißt  $f$  stetig in  $D_f$ <sup>9</sup>. Den Vektorraum aller in  $D$  stetigen Funktionen bezeichnen wir mit  $C(D)$ ,  $C(D \rightarrow \mathbb{C})$  oder  $C(D \rightarrow \mathbb{R})$ .

**Satz 6.18.** Sei  $f: D_f \rightarrow \mathbb{C}$ , und sei  $z^* \in D_f$  ein Häufungspunkt von  $D_f$ . Dann sind äquivalent:

1.  $f$  ist stetig in  $z^*$ ,
2.  $\lim_{z \rightarrow z^*} f(z) = f(z^*)$ ,
3.  $\lim_{n \rightarrow \infty} f(z_n) = f(z^*)$  für jede Folge  $(z_n)_{n \in \mathbb{N}} \subset D_f \setminus \{z^*\}$ , die gegen  $z^*$  konvergiert,
4. jedes  $\varepsilon > 0$  hat ein  $\delta_0(\varepsilon) > 0$ , sodaß für jedes  $z \in D_f$  mit  $|z - z^*| < \delta_0(\varepsilon)$  gilt:  $|f(z) - f(z^*)| < \varepsilon$ .

*Beweis.* Ergibt sich aus Satz 6.8 und Definition 6.14. □

Im weiteren Verlauf des Unterkapitels 6.2 wollen wir auch folgende Frage studieren:

*Sei  $A \subset D_f$  eine Menge. Welche Eigenschaften von  $A$  vererben sich auf  $f(A)$ , wenn  $f$  stetig ist?*

Wir wissen bereits: wenn  $A$  eine in  $D_f$  konvergente Folge ist, dann ist (bei entsprechender Numerierung der Elemente) auch die Bildmenge  $f(A)$  eine in  $W_f$  konvergente Folge.

Es gibt Eigenschaften, die sich nicht vererben, nämlich Offenheit und Abgeschlossenheit:

**Frage:** Man gebe ein Beispiel einer stetigen Funktion von  $\mathbb{R}$  nach  $\mathbb{R}$  an, bei der das Bild einer offenen Menge nicht unbedingt wieder eine offene Menge ist.

Man gebe ein weiteres Beispiel einer stetigen Funktion von  $\mathbb{R}$  nach  $\mathbb{R}$  an, bei der das Bild einer abgeschlossenen Menge nicht unbedingt wieder eine abgeschlossene Menge ist.

Andererseits gilt für Offenheit und Abgeschlossenheit die „Vererblichkeit rückwärts“:

<sup>7</sup> continuous at  $z^*$

<sup>8</sup> discontinuous at  $z^*$

<sup>9</sup> continuous in  $D_f$

**Satz 6.19.** Sei  $f: \mathbb{C} \rightarrow \mathbb{C}$  eine Funktion. Dann sind äquivalent:

1.  $f$  ist stetig auf  $\mathbb{C}$ ,
2. das Urbild jeder offenen Menge ist offen,
3. das Urbild jeder abgeschlossenen Menge ist abgeschlossen.

*Beweis.* Lassen wir weg. Die Äquivalenz von 2. und 3. folgt übrigens sofort aus dem zweiten Teil von Satz 5.13.  $\square$

**Bemerkung 6.20.** Analog wie bei Bemerkung 6.5 halten wir auch hier fest, daß man die punktweise Stetigkeit im Sinne von Definition 6.14 auch für Abbildungen  $f: U \rightarrow V$  zwischen Banachräumen definieren kann. Man muß nur überall die Betragsstriche  $|\cdot|$  passend durch Normstriche  $\|\cdot\|_U$  und  $\|\cdot\|_V$  ersetzen, vor allem in Satz 6.18, Teil 4. Und die Stetigkeit in einem Gebiet im Sinne von Definition 6.17 läßt sich auch für solche allgemeineren Abbildungen definieren. Satz 6.19 gilt dann ebenfalls.

**Satz 6.21.** Die folgenden Aussagen gelten sowohl für „Stetigkeit in einem Punkt“ als auch für „Stetigkeit in einem Gebiet“.

- Wenn  $f$  und  $g$  stetig sind, dann auch  $\alpha f + \beta g$  für jegliche  $\alpha, \beta \in \mathbb{C}$ .
- Wenn  $f$  und  $g$  stetig sind, dann auch  $f \cdot g$ .
- Wenn  $f$  und  $g$  stetig sind und zusätzlich noch  $g(x) \neq 0$  gilt, dann ist auch  $\frac{f}{g}$  stetig.

*Beweis.* Folgt beinahe sofort aus Satz 6.12.  $\square$

**Satz 6.22.** Seien  $f: D_f \rightarrow W_f$  und  $g: D_g \rightarrow W_g$  stetige Funktionen mit  $W_g \subset D_f$ . Dann ist die zusammengesetzte Funktion (Komposition<sup>10</sup>)  $f \circ g: D_g \rightarrow W_f$  definiert und wieder stetig.

*Beweis.* Sei  $x^* \in D_g$  und  $y^* = g(x^*) \in D_f$ . Die Stetigkeiten von  $g$  in  $x^*$  und von  $f$  in  $y^*$  bedeuten: wenn  $(x_n)_{n \in \mathbb{N}}$  und  $(y_n)_{n \in \mathbb{N}}$  Folgen sind mit Grenzwert  $x^*$  und  $y^*$ , dann ist

$$\begin{aligned} \lim_{n \rightarrow \infty} g(x_n) &= g\left(\lim_{n \rightarrow \infty} x_n\right) = g(x^*), \\ \lim_{n \rightarrow \infty} f(y_n) &= f\left(\lim_{n \rightarrow \infty} y_n\right) = f(y^*). \end{aligned}$$

Nun sei  $(x_n)_{n \in \mathbb{N}}$  gewählt, und sei  $y_n := g(x_n)$ . Dann hat diese Folge den Grenzwert  $y^*$ , und es folgt

$$\begin{aligned} \lim_{n \rightarrow \infty} (f \circ g)(x_n) &= \lim_{n \rightarrow \infty} (f(g(x_n))) = \lim_{n \rightarrow \infty} f(y_n) = f\left(\lim_{n \rightarrow \infty} y_n\right) = f\left(\lim_{n \rightarrow \infty} g(x_n)\right) \\ &= f\left(g\left(\lim_{n \rightarrow \infty} x_n\right)\right) = (f \circ g)\left(\lim_{n \rightarrow \infty} x_n\right). \end{aligned}$$

Also kommutiert  $f \circ g$  mit  $\lim$ , und das wollten wir haben.  $\square$

**Beispiel 6.23.** • Konstante Funktionen sind trivialerweise stetig.

- Lineare Funktionen der Form  $x \mapsto x$  sind stetig.
- Polynome sind stetig (folgt aus den ersten beiden •).
- Gebrochen rationale Funktionen sind dort stetig, wo der Nenner nicht Null wird.

**Satz 6.24.** Sei  $f$  eine stetige Funktion mit Definitionsbereich  $D_f$  in einem Banachraum  $U$  und Wertebereich  $W_f$  in einem Banachraum  $V$ . Dann ist das Bild einer jeden kompakten Menge wieder kompakt.

*Beweis.* Sei  $M \subset D_f$  eine in  $U$  kompakte Menge. Wir wollen zeigen, daß die Bildmenge  $f(M)$  in  $V$  kompakt ist.

Sei also  $(v_n)_{n \in \mathbb{N}}$  eine Folge in  $f(M)$ . Gesucht ist eine Teilfolge davon, die einen Grenzwert in  $f(M)$  hat. Zu jedem  $v_n$  gibt es ein  $u_n \in M$  mit  $f(u_n) = v_n$ . Die Folge  $(u_n)_{n \in \mathbb{N}}$  ist eine Folge in  $M$ , und  $M$  ist kompakt. Also gibt es eine Teilfolge  $(u_{n_k})_{k \in \mathbb{N}}$ , die gegen einen Grenzwert  $u^* \in M$  konvergiert. Weil  $f$  in  $u^*$  stetig ist, muß  $\lim_{k \rightarrow \infty} f(u_{n_k}) = f(u^*)$  sein. Nun ist aber  $f(u_{n_k}) = v_{n_k}$ , und diese  $v_{n_k}$  bilden eine Teilfolge der Folge  $(v_n)_{n \in \mathbb{N}}$ . Wegen  $u^* \in M$  ist  $f(u^*) \in f(M)$ . Also strebt die Teilfolge  $(v_{n_k})_{k \in \mathbb{N}}$  gegen ein Element aus  $f(M)$ . Folglich ist  $f(M)$  kompakt.  $\square$

<sup>10</sup>composed function

Die Kompaktheit einer Menge in  $D_f$  vererbt sich also auf die entsprechende Bildmenge. Dies gilt auch für stetige Funktionen zwischen unendlichdimensionalen Banachräumen.

**Folgerung 6.25.** *Jede auf einer kompakten Menge stetige Funktion ist dort beschränkt.*

*Das Bild einer kompakten Menge unter einer stetigen Funktion ist abgeschlossen.*

*Beweis.* Hinter Definition 5.51 hatten wir vermerkt: Kompakte Mengen sind beschränkt. Kompakte Mengen sind abgeschlossen. Fertig.  $\square$

Jetzt untersuchen wir für reelle Funktionen  $f: D_f \rightarrow W_f \subset \mathbb{R}$ , welche Bedingungen der Definitionsbereich  $D_f$  erfüllen sollte, damit die Menge  $W_f$  der tatsächlich angenommenen Werte ein Maximum oder Minimum enthält. Wir beginnen mit zwei Beispielen.

**Beispiel:** Sei  $f = f(x) = \exp(x)$  mit Definitionsbereich  $D_f = \mathbb{R}$ . Es ist  $D_f$  abgeschlossen, aber nicht beschränkt. Und die Menge der tatsächlich angenommenen Werte ist  $W_f = (0, \infty)$ , die kein kleinstes Element enthält. Es existiert also  $\min W_f$  nicht.

**Beispiel:** Sei  $f = f(x) = \tan(x)$  mit Definitionsbereich  $D_f = (-\pi/2, \pi/2)$ . Der Definitionsbereich  $D_f$  ist beschränkt, aber nicht abgeschlossen. Und tatsächlich ist  $W_f = (-\infty, \infty) = \mathbb{R}$ , was ebenfalls kein kleinstes Element besitzt.

Jeweils für sich alleine reichen die Beschränktheit und die Abgeschlossenheit des Definitionsbereiches also nicht aus, um die Existenz von  $\min W_f$  (oder analog  $\max W_f$ ) zu sichern. Aber beide zusammen bringen uns zum gewünschten Ziel, wie der folgende Satz vom Maximum zeigt (einen analogen Satz vom Minimum kann jeder selbst formulieren):

**Satz 6.26 (Satz vom Maximum).** *Sei  $f: D_f \rightarrow \mathbb{R}$  eine stetige Funktion, und sei  $D_f \neq \emptyset$  kompakt. Dann nimmt  $f$  auf  $D_f$  sein Supremum und sein Infimum an. Das heißt, es gibt  $x_*, x^* \in D_f$  sodaß*

$$f(x_*) \leq f(x) \leq f(x^*) \quad \forall x \in D_f.$$

*Beweis.* Der Wertebereich  $W_f$  ist kompakt (wegen Satz 6.24) und nichtleer (wegen  $D_f \neq \emptyset$ ). Er ist also nach oben beschränkt, hat wegen Satz 5.47 demnach ein Supremum. Weil  $W_f$  abgeschlossen ist (wegen Folgerung 6.25), ist dieses Supremum von  $W_f$  sogar ein Maximum, also besitzt  $W_f$  ein größtes Element  $f(x^*)$ . Analog argumentiert man für die Aussage über das Minimum.  $\square$

**Bemerkung 6.27.** *Sei  $U = C([a, b] \rightarrow \mathbb{R})$  der Raum der auf  $[a, b]$  stetigen und reellwertigen Funktionen. Wir hatten die  $\|\cdot\|_\infty$ -Norm bisher als*

$$\|f\|_\infty := \max_{x \in [a, b]} |f(x)|$$

*definiert. Dabei hatten wir den (denkbaren) Fall ignoriert, daß es dieses Maximum gar nicht gäbe: sei es, weil  $|f|$  auf  $[a, b]$  unbeschränkt wäre, sei es, weil  $|f|$  zwar beschränkt wäre und ein Supremum hätte, aber kein Maximum. Der Satz vom Maximum sagt uns jetzt, daß das Maximum von  $|f|$  immer existiert. Die eigentlich korrekte Definition der Norm ist*

$$\|f\|_\infty := \sup_{x \in [a, b]} |f(x)|.$$

Reellwertige stetige Funktionen springen nicht — das ist das Ergebnis des folgenden Satzes.

**Satz 6.28 (Zwischenwertsatz<sup>11</sup>).** *Sei  $f: D_f \rightarrow \mathbb{R}$  stetig auf dem kompakten Intervall  $[a, b]$ . Dann nimmt  $f$  auf  $[a, b]$  jeden Wert zwischen  $f(a)$  und  $f(b)$  an. Wenn insbesondere  $f(a)$  und  $f(b)$  verschiedene Vorzeichen haben, dann gibt es eine Nullstelle zwischen  $a$  und  $b$ .*

*Beweis.* Es reicht, die Aussage über die Nullstelle zu beweisen (Warum?)

Wir setzen  $a_0 = a$  und  $b_0 = b$ . OBdA sei  $f(a_0) < 0$  und  $f(b_0) > 0$ . Nimm  $Z := \frac{1}{2}(a_0 + b_0)$ .

**Fall 1: angenommen, daß  $f(Z) = 0$ :** Dann sind wir fertig.

**Fall 2: angenommen, daß  $f(Z) < 0$ :** Dann setzen wir  $a_1 = Z$  und  $b_1 = b_0$ .

<sup>11</sup>intermediate value theorem



**Fall 3: angenommen, daß  $f(Z) > 0$ :** Dann setzen wir  $a_1 = a$  und  $b_1 = Z$ .

Das Ergebnis ist in Fall 2 und Fall 3 jeweils  $f(a_1) < 0 < f(b_1)$  und  $|a_1 - b_1| = \frac{1}{2}|a_0 - b_0|$ .

In diesem Stil verfahren wir induktiv weiter und finden zwei Folgen  $(a_n)_{n \in \mathbb{N}}$ ,  $(b_n)_{n \in \mathbb{N}}$  so daß:

$$\begin{aligned} a_0 &\leq a_1 \leq a_2 \leq a_3 \leq \dots, \\ b_0 &\geq b_1 \geq b_2 \geq b_3 \geq \dots, \\ a_n &< b_n \quad \forall n, \\ |b_{n+1} - a_{n+1}| &= \frac{1}{2}|b_n - a_n| \quad \forall n, \\ f(a_n) &< 0 < f(b_n) \quad \forall n. \end{aligned}$$

Die Folgen  $(a_n)_{n \in \mathbb{N}}$  und  $(b_n)_{n \in \mathbb{N}}$  streben gegen einen gemeinsamen Grenzwert  $x^*$ , siehe Satz 5.48. Wegen  $f(a_n) < 0$  und der Stetigkeit (in der Interpretation als kommutatives Diagramm gemäß (6.1)) ist

$$f(x^*) = f(\lim_{n \rightarrow \infty} a_n) = \lim_{n \rightarrow \infty} f(a_n) \leq 0,$$

analog zeigt man  $f(x^*) \geq 0$ . Also ist  $f(x^*) = 0$ . □

**Frage:** Kennen Sie ein Beispiel für eine stetige Funktion  $f: \mathbb{C} \rightarrow \mathbb{C}$ , für die dieser Satz falsch ist ?

Wir übertragen die Begriffe „injektiv“ und „monoton“ auf Funktionen:

**Definition 6.29.** Eine Funktion  $f: D_f \rightarrow W_f$  heißt injektiv<sup>12</sup>, wenn aus  $y_1 = f(x_1), y_2 = f(x_2) \in W_f$  und  $y_1 = y_2$  immer  $x_1 = x_2$  folgt.

Sei  $I \subset \mathbb{R}$  ein Intervall. Eine Funktion  $f: I \rightarrow \mathbb{R}$  heißt monoton wachsend<sup>13</sup>, wenn aus  $x_1 \leq x_2$  stets  $f(x_1) \leq f(x_2)$  folgt.

Eine Funktion  $f: I \rightarrow \mathbb{R}$  heißt streng monoton wachsend<sup>14</sup>, wenn aus  $x_1 < x_2$  stets  $f(x_1) < f(x_2)$  folgt.

Analog werden (streng) monoton fallende Funktionen definiert.

**Satz 6.30.** Sei  $f: D_f \rightarrow W_f$  eine Funktion, wobei  $D_f, W_f$  Intervalle aus  $\mathbb{R}$  seien.

1. Die Funktion  $f$  ist surjektiv auf  $W_f$ . Wenn  $f$  injektiv ist, dann existiert eine Umkehrfunktion  $g: W_f \rightarrow D_f$  mit  $g \circ f = \text{id}_{D_f}$  und  $f \circ g = \text{id}_{W_f}$ .
2. Wenn  $f$  streng monoton ist, dann ist  $f$  injektiv. Die Umkehrfunktion ist wieder streng monoton.
3. Wenn  $f$  streng monoton und stetig ist, dann ist auch  $g$  stetig.

*Beweis.* 1. und 2. sind sehr leicht, siehe auch Satz 4.8.

Bei 3. nehmen wir oBdA an, daß  $f$  wächst. Sei  $y_0 = f(x_0) \in W_f$ . Zu zeigen wäre, daß die Umkehrfunktion  $g$  in  $y_0$  stetig ist. Angenommen,  $y_0$  sei kein Randpunkt von  $W_f$ . Dann ist auch  $x_0$  kein Randpunkt von  $D_f$ , und für kleines  $\varepsilon > 0$  ist  $[x_0 - \varepsilon, x_0 + \varepsilon] \subset D_f$ . Wir suchen zu diesem  $\varepsilon$  ein  $\delta_0(\varepsilon)$ , sodaß aus  $|y - y_0| < \delta_0(\varepsilon)$  die Ungleichung  $|g(y) - g(y_0)| < \varepsilon$  folgt.

Nun ist das Bild des Intervalles  $[x_0 - \varepsilon, x_0 + \varepsilon]$  unter der Abbildung  $f$  aber gerade gleich dem Intervall  $[f(x_0 - \varepsilon), f(x_0 + \varepsilon)]$ . Wir wählen  $\delta_0(\varepsilon) = \min(f(x_0) - f(x_0 - \varepsilon), f(x_0 + \varepsilon) - f(x_0))$  und sind fertig.

Ein anderer Beweis zu 3. beruht darauf zu begründen, daß man bei dem kommutativen Diagramm für die Definition der Stetigkeit von  $f$  die  $f$ -Pfeile umdrehen darf (es ist nicht so trivial wie es aussieht). □

**Definition 6.31.** Die Umkehrfunktion<sup>15</sup> einer injektiven Funktion  $f$  bezeichnet man mit  $f^{-1}$ .

Diese Bezeichnung ist mißverständlich (denn  $f^{-1}$  könnte ja auch die Division  $\frac{1}{f}$  ausdrücken wollen), aber trotzdem allgemein gebräuchlich. Wenn  $y$  im Wertebereich von  $f$  liegt, dann wird die Urbildmenge von  $y$  oft als  $f^{-1}(y)$  geschrieben, und zwar auch für nicht-injektive  $f$ . Das ist dann eine dritte Bedeutung von  $f^{-1}$ .

<sup>12</sup>injective

<sup>13</sup>monotonically increasing

<sup>14</sup>strictly monotonically increasing

<sup>15</sup>inverse function

**Beispiel 6.32.** Wir setzen  $D_f = \{x \in \mathbb{R} : x \geq 0\} = \mathbb{R}_{\geq 0}$  und  $f = f(x) = x^n$  mit  $n \in \mathbb{N}$ . Dann ist  $W_f = \mathbb{R}_{\geq 0}$ . Die Funktion  $x \mapsto x$  ist offensichtlich stetig, nichtnegativ-wertig und streng monoton wachsend, also ist auch die Funktion  $f$  stetig und streng monoton wachsend, denn sie ist das mehrfache Produkt der Funktion  $x \mapsto x$  mit sich selbst. Nach Satz 6.30 hat dann  $f$  eine Umkehrfunktion  $g$ , die wir Wurzelfunktion nennen und es ist

$$\begin{aligned} g: \mathbb{R}_{\geq 0} &\rightarrow \mathbb{R}_{\geq 0}, \\ g: x &\mapsto \sqrt[n]{x}. \end{aligned}$$

Man beachte, daß die Wurzelfunktion lediglich für nichtnegative reelle Zahlen definiert ist. Die manchmal anzutreffenden Aussagen der Form

$$\sqrt[3]{-8} = -2$$

sind verwegener Unsinn. Man meditiere zum Beispiel über der Zeile

$$-2 = \sqrt[3]{-8} = (-8)^{\frac{1}{3}} = (-8)^{\frac{2}{6}} = ((-8)^2)^{\frac{1}{6}} = 64^{\frac{1}{6}} = \sqrt[6]{64} = +2.$$

**Satz 6.33.** Die Funktion  $z \mapsto \exp(z)$  ist stetig auf  $\mathbb{C}$ .

*Beweis.* Wir könnten auf das entsprechende Ergebnis aus Satz 5.65 verweisen, das allerdings immer noch seines Beweises harret.

Oder wir rechnen es schnell aus, wobei wir mit der Stetigkeit in  $z_0 = 0$  anfangen. Wir wollen zeigen, daß

$$\lim_{z \rightarrow 0} |\exp(z) - \exp(0)| = 0$$

ist, weshalb wir  $|z| \leq 1$  voraussetzen können. Nun ist

$$\begin{aligned} 0 \leq |\exp(z) - \exp(0)| &= \left| \sum_{k=0}^{\infty} \frac{z^k}{k!} - 1 \right| \leq \sum_{k=1}^{\infty} \frac{|z|^k}{k!} = |z| \sum_{k=0}^{\infty} \frac{|z|^k}{(k+1)!} \leq |z| \sum_{k=0}^{\infty} \frac{|z|^k}{k!} = |z| \exp(|z|) \\ &\leq |z| \exp(1); \end{aligned}$$

und wenn wir jetzt das Sandwichprinzip anwenden, ist die Stetigkeit im Nullpunkt gezeigt.

Sei nun  $z_0 \in \mathbb{C}$  beliebig, und wir wollen beweisen, daß

$$\lim_{z \rightarrow z_0} |\exp(z) - \exp(z_0)| = 0.$$

Nun ist aber

$$\lim_{z \rightarrow z_0} |\exp(z) - \exp(z_0)| = \lim_{z \rightarrow z_0} |\exp(z_0)| \cdot |\exp(z - z_0) - \exp(0)| = |\exp(z_0)| \cdot \lim_{z \rightarrow z_0} |\exp(z - z_0) - 1| = 0.$$

Also ist die Exponentialfunktion auch in  $z_0 \in \mathbb{C}$  stetig.  $\square$

Weil die Exponentialfunktion  $x \mapsto e^x$  für reelle  $x$  noch dazu streng monoton wachsend ist, können wir eine Umkehrfunktion dazu betrachten:

**Definition 6.34.** Die Umkehrfunktion der Funktion

$$\begin{aligned} \exp: \mathbb{R} &\rightarrow \mathbb{R}_+, \\ \exp: x &\mapsto \exp(x) = e^x \end{aligned}$$

heißt Logarithmus<sup>16</sup> (zur Basis  $e$ ) und wird geschrieben als

$$\begin{aligned} \ln: \mathbb{R}_+ &\rightarrow \mathbb{R}, \\ \ln: x &\mapsto \ln(x). \end{aligned}$$

Aus den Eigenschaften der Exponentialfunktion erhalten wir schnell die folgenden Eigenschaften der Logarithmusfunktion:

---

<sup>16</sup>logarithm

**Lemma 6.35.** Die Logarithmusfunktion ist stetig und streng monoton auf  $\mathbb{R}_+$ . Für  $x, y \in \mathbb{R}_+$  gilt dann

$$\begin{aligned}\ln(x \cdot y) &= \ln(x) + \ln(y), \\ \ln(x/y) &= \ln(x) - \ln(y), \\ \ln(1) &= 0.\end{aligned}$$

Wir wollen als nächstes beliebige Potenzen  $a^x$  definieren, wobei  $a \in \mathbb{R}_+$  und  $x \in \mathbb{R}$  ist.

**Definition 6.36 (Allgemeine Potenz).** Sei  $a \in \mathbb{R}_+$  und  $x \in \mathbb{R}$ . Dann definieren wir

$$a^x := \exp(x \ln a).$$

Diese Funktion ist stetig, denn sie ist eine Komposition von stetigen Funktionen.

Wenn jetzt  $x = n \in \mathbb{N}$ , dann hat die Schreibweise  $a^x$  zwei Definitionen: einerseits  $a^n = a \cdot \dots \cdot a$  ( $n$  Faktoren), andererseits  $a^n = \exp(n \ln a)$ . Diese widersprechen einander nicht, denn es ist

$$\exp(n \ln a) = \exp(\ln a + \dots + \ln a) = \exp(\ln a) \cdot \dots \cdot \exp(\ln a) = a \cdot \dots \cdot a.$$

**Satz 6.37.** Seien  $a, b \in \mathbb{R}_+$  und  $x, y \in \mathbb{R}$ . Dann gilt

$$\begin{aligned}\ln(a^x) &= x \ln a, \\ (a \cdot b)^x &= a^x \cdot b^x, \\ a^{x+y} &= a^x \cdot a^y, \\ a^{-x} &= \frac{1}{a^x}, \\ (a^x)^y &= a^{(x \cdot y)}.\end{aligned}$$

*Beweis.* Die erste Eigenschaft folgt sofort aus der Definition von  $a^x$ , und die weiteren ergeben sich dann unmittelbar.  $\square$

Für  $a^{1/n}$  erhalten wir gerade die Wurzelfunktion  $\sqrt[n]{a}$ , denn  $(a^{1/n})^n = a^{(1/n) \cdot n} = a$ .

**Warnung 6.38.** Die obigen Regeln für Exponentialfunktionen der Form  $x \mapsto a^x$  gelten nur für reelle  $x$  und positive  $a$ . Analog gelten die Regeln für die Logarithmusfunktion nur für positive Argumente von  $\ln$ .

Gedankenlose Ausdehnung auf negative oder gar komplexe  $a$  führt in den meisten Fällen zu schweren Fehlern. Auf dieselben Probleme wird man stoßen, wenn man die Logarithmusfunktion für negative oder komplexe Zahlen definieren will.

## 6.3 Differenzierbarkeit

**Definition 6.39.** Sei  $f: D_f \rightarrow W_f$  eine Funktion, wobei  $D_f, W_f \subset \mathbb{R}$  oder  $D_f, W_f \subset \mathbb{C}$ .

Sei  $z_0 \in D_f$  ein Punkt im Inneren von  $D_f$ , d.h. nicht auf dem Rand  $\partial D_f$ . Wir sagen, daß  $f$  im Punkte  $z_0$  differenzierbar<sup>17</sup> ist, wenn der Grenzwert

$$\lim_{z \rightarrow z_0} \frac{f(z) - f(z_0)}{z - z_0} \quad (\text{im Sinne von Definition 6.3})$$

existiert. In diesem Falle nennen wir den Grenzwert Ableitung von  $f$  im Punkt  $z_0$ <sup>18</sup> und schreiben dafür

$$f'(z_0) = \frac{df}{dz}(z_0) = \frac{d}{dz} f(z_0) = \left. \frac{d}{dz} f(z) \right|_{z=z_0}.$$

Sei  $G \subset D_f$  eine Menge. Wenn  $f$  in jedem (inneren) Punkt von  $G$  differenzierbar ist, dann sagen wir, daß  $f$  differenzierbar in  $G$  ist. In diesem Falle ist die Ableitung  $f'$  eine Funktion, die in  $G$  definiert ist. Wenn diese Ableitung zusätzlich stetig sein sollte, dann nennt man  $f$  stetig in  $G$  differenzierbar<sup>19</sup>.

<sup>17</sup> differentiable at  $z_0$

<sup>18</sup> derivative of  $f$  at  $z_0$

<sup>19</sup> continuously differentiable in  $G$

Wenn  $f'$  wieder differenzierbar sein sollte, dann heißt  $f$  zweimal differenzierbar<sup>20</sup>. Analog definiert man höhere Ableitungen.

Für die Menge der auf  $G$   $k$ -mal stetig differenzierbaren Funktionen mit reellen Werten schreibt man  $C^k(G \rightarrow \mathbb{R})$ .

Es gibt zwei verschiedene Theorien differenzierbarer Funktionen. Einerseits diejenige, wo  $D_f$  und  $W_f$  Teilmengen von  $\mathbb{R}$  sind, andererseits diejenige mit  $D_f, W_f \subset \mathbb{C}$ .

Diese zwei Theorien sind total verschieden voneinander. So ist zum Beispiel eine komplex differenzierbare Funktion automatisch unendlich oft differenzierbar. Man redet in diesem Zusammenhang auch von *holomorphen* oder *analytischen*<sup>21</sup> Funktionen. Weiterhin gilt zum Beispiel: wenn eine komplex differenzierbare Funktion in einer kleinen Kreisscheibe überall den Wert 0 hat, dann hat sie überall in  $\mathbb{C}$  den Wert 0. Dieses Verhalten findet man bei reell differenzierbaren Funktionen bekanntlich nicht.

Außerdem ist zum Beispiel die Funktion  $z \mapsto |z|^4$  lediglich im Nullpunkt komplex differenzierbar und sonst nirgendwo.

Holomorphe Funktionen werden im 3. Semester eingehend studiert werden; ein großer Teil der jetzigen Untersuchungen beschränkt sich deshalb auf reell differenzierbare Funktionen.

**Definition 6.40.** Sei  $f: D_f \rightarrow W_f$  differenzierbar in  $D_f$  und  $x_0 \in D_f$ . Sei  $\Delta x = dx \in \mathbb{R}$  (bzw.  $\mathbb{C}$ ) eine beliebige Zahl. Wir setzen voraus, daß der Ball  $B_{|\Delta x|}(x_0)$  in  $D_f$  enthalten ist. Dann definieren wir:

$$\begin{aligned}\Delta f &= f(x_0 + \Delta x) - f(x_0), \\ df &= f'(x_0) \cdot \Delta x = f'(x_0) \cdot dx.\end{aligned}$$

Der Ausdruck  $df$  heißt Differential von  $f$ <sup>22</sup>.

Die Differenz  $\Delta f - df$  beschreibt den Fehler, den man begeht, wenn man  $f$  in der Nähe des Punktes  $x_0$  durch eine lineare Funktion annähert. Die Differenzierbarkeit von  $f$  in  $x_0$  bedeutet gerade, daß dieser Fehler von höherer als erster Ordnung nach 0 konvergiert, wenn  $\Delta x$  nach 0 strebt.

**Definition 6.41.** Seien  $u$  und  $v$  zwei Funktionen, definiert nahe  $x = x_0$ . Wir sagen, daß

$$u(x) = \mathfrak{O}(v(x)), \quad x \rightarrow x_0,$$

wenn es eine Konstante  $C > 0$  gibt, sodaß in einer Umgebung von  $x = x_0$  gilt:

$$|u(x)| \leq C|v(x)|.$$

Wir sagen, daß

$$u(x) = \mathfrak{o}(v(x)), \quad x \rightarrow x_0$$

wenn  $u(x) = \mathfrak{O}(v(x))$  für  $x \rightarrow x_0$  und wenn zusätzlich  $u$  in  $x_0$  eine stärkere Nullstelle hat als  $v$ , also

$$\lim_{x \rightarrow x_0} \frac{u(x)}{v(x)} = 0.$$

Diese Ausdrücke  $\mathfrak{O}$  und  $\mathfrak{o}$  heißen LANDAU<sup>23</sup>-Symbole. Sinngemäß genauso definiert man Landau-Symbole für den Fall  $x \rightarrow \infty$ .

Wenn wir an einer Stelle  $\mathfrak{O}(v(x))$  schreiben, und an einer anderen Stelle des Textes nochmal  $\mathfrak{O}(v(x))$  schreiben, dann soll das nicht bedeuten, daß beide Terme identisch sind, sondern nur, daß sie sich gleich verhalten. Das bedeutet, daß der erste Term abgeschätzt werden kann als  $\leq C_1|v(x)|$ , und daß der zweite Term abgeschätzt werden kann als  $\leq C_2|v(x)|$ . Hierbei sind  $C_1$  und  $C_2$  positive Konstanten, deren genauer Wert oft nicht übermäßig interessant ist.

<sup>20</sup>twice differentiable

<sup>21</sup>holomorph, analytic

<sup>22</sup> Es gibt noch eine andere Bedeutung des Begriffes *Differential*. Gelegentlich wird diejenige lineare *Abbildung*, die  $\Delta x$  auf  $f'(x_0) \cdot \Delta x$  abbildet, als Differential bezeichnet. Der obige Differentialbegriff versteht unter  $df$  also den Ausdruck  $f'(x_0) \cdot \Delta x$ , der andere Differentialbegriff versteht unter  $df$  aber den Term  $f'(x_0)$ . An dieser Stelle ist es vielleicht interessant zu vermerken, daß mit der umgangssprachlichen Bezeichnung „Bronstein“ gelegentlich zwei verschiedene Werke gemeint sind: eine Ausgabe des B.G.Teubner-Verlags, und eine Ausgabe des Verlags Harri Deutsch. Beide haben ihre Wurzeln in der sowjetischen Erstauflage von 1937, aber aus juristischen Gründen darf nur die letztere „Bronstein“ genannt werden, und beide vertreten unterschiedliche Auffassungen, was ein Differential denn nun eigentlich ist.

<sup>23</sup>EDMUND GEORG HERMANN LANDAU, 1877–1938, nicht zu verwechseln mit LEV DAVIDOVICH LANDAU, 1908–1968, bekannt für den gemeinsam mit Evgenni Mikhailovich Lifshitz geschriebenen Kurs zur Theoretischen Physik

**Beispiel 6.42.** *Es gilt*

$$\begin{aligned}\sin(x) &= \mathfrak{O}(x), & x \rightarrow 0, \\ \sin(x) &= x + \mathfrak{O}(x^3), & x \rightarrow 0, \\ \cos(x) &= 1 + \mathfrak{O}(x^2) = 1 + \mathfrak{o}(x), & x \rightarrow 0, \\ \frac{2x+5}{x+1} &= \mathfrak{O}(1), & x \rightarrow +\infty.\end{aligned}$$

**Lemma 6.43.** *Für  $x \rightarrow x_0$  haben wir die folgenden Rechenregeln:*

- wenn  $u$  in einer Umgebung von  $x_0$  beschränkt ist, dann ist  $u(x) = \mathfrak{O}(1)$ ,
- $\mathfrak{O}(v(x)) + \mathfrak{O}(v(x)) = \mathfrak{O}(v(x))$ ,
- $\mathfrak{O}(v_1(x)) \cdot \mathfrak{O}(v_2(x)) = \mathfrak{O}(v_1(x) \cdot v_2(x))$ ,
- $\mathfrak{O}((x - x_0)^2) = \mathfrak{o}(x - x_0)$ .

Hierbei sind die letzten drei • von links nach rechts zu lesen, sonst begibt man sich evtl. in die Schwierigkeiten, die in Warnung 5.7 beschrieben wurden. Die Beweise dieser Regeln folgen schnell aus der Definition der Landau-Symbole.

**Satz 6.44 (Weierstraß-Zerlegung).** *Sei  $f$  eine Funktion, definiert in einer Umgebung von  $x_0$ .*

1. *Sei  $f$  differenzierbar in  $x_0$ . Dann ist für  $x$  in der Nähe von  $x_0$*

$$f(x) = f(x_0) + f'(x_0) \cdot (x - x_0) + R(x; x_0)$$

*mit  $R(x; x_0) = \mathfrak{o}(|x - x_0|)$  für  $x \rightarrow x_0$ .*

2. *Wenn es eine Zahl  $A \in \mathbb{R}$  gibt (die nicht von  $x$  abhängt) mit*

$$f(x) = f(x_0) + A \cdot (x - x_0) + \mathfrak{o}(|x - x_0|), \quad x \rightarrow x_0,$$

*dann ist  $f$  in  $x_0$  differenzierbar, und es ist  $f'(x_0) = A$ .*

*Beweis.* 1. Wir haben

$$\frac{R(x; x_0)}{x - x_0} = \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0),$$

und die rechte Seite strebt nach 0 falls  $x \rightarrow x_0$ .

2. Es ist

$$\frac{f(x) - f(x_0)}{x - x_0} = A + \frac{\mathfrak{o}(|x - x_0|)}{x - x_0},$$

und die rechte Seite hat einen Limes für  $x \rightarrow x_0$ , nämlich  $A$ . Also ist  $f$  differenzierbar in  $x_0$ . □

Später werden wir sehen, daß sogar  $R(x; x_0) = \mathfrak{O}(|x - x_0|^2)$  gilt, wenn  $f$  in einer Umgebung von  $x_0$  *zweimal* stetig differenzierbar ist.

Als Merkregel halten wir fest:

$$\begin{aligned}\text{Wenn } f(x) &= f(x_0) + A(x - x_0) + \mathfrak{o}(x - x_0) \text{ für } x \rightarrow x_0, \\ \text{dann muß } f &\text{ in } x_0 \text{ differenzierbar sein,} \\ \text{und } f'(x_0) &\text{ muß gleich } A \text{ sein.}\end{aligned}$$

Damit werden wir demnächst die Ableitungsformeln für Summen, Produkte, Verkettungen und Umkehrfunktionen zeigen, wobei wir ausgiebigen Gebrauch von den Rechenregeln für die Landau-Symbole machen werden.

**Satz 6.45.** *Differenzierbare Funktionen sind stetig.*

*Beweis.* Sieht man sofort aus Satz 6.44. □

**Beispiel 6.46.** *Die Ableitungen von  $f = f(x) = 1$  und  $g = g(x) = x$  lauten  $f' = f'(x) = 0$  und  $g' = g'(x) = 1$ .*

**Satz 6.47.** *Die Exponentialfunktion  $z_0 \mapsto \exp(z_0)$  ist für komplexe  $z_0$  differenzierbar, und es ist  $\exp'(z_0) = \exp(z_0)$ .*

*Beweis.* Wir zeigen zunächst, daß  $\exp'(0) = 1$ . Dazu müssen wir nur beweisen, daß

$$\exp(z) - \exp(0) - 1 \cdot (z - 0) = \mathfrak{o}(|z - 0|), \quad z \rightarrow 0.$$

Nun ist aber

$$\left| \frac{\exp(z) - 1 - z}{z - 0} \right| = \left| \frac{\sum_{k=2}^{\infty} \frac{z^k}{k!}}{z} \right| \leq \sum_{k=1}^{\infty} \frac{|z|^k}{(k+1)!} = |z| \sum_{k=0}^{\infty} \frac{|z|^k}{(k+2)!} \leq |z| \sum_{k=0}^{\infty} \frac{|z|^k}{k!} = |z| \exp(|z|)$$

und dies strebt nach 0 für  $z \rightarrow 0$ , denn die Exponentialfunktion ist stetig, siehe Satz 6.33.

Sei nun  $z_0 \in \mathbb{C}$  beliebig. Wir wollen zeigen, daß

$$\exp(z) - \exp(z_0) - \exp(z_0) \cdot (z - z_0) = \mathfrak{o}(|z - z_0|), \quad z \rightarrow z_0.$$

Offensichtlich ist nun wegen des Additionstheorems der Exponentialfunktion

$$\left| \frac{\exp(z) - \exp(z_0) - \exp(z_0) \cdot (z - z_0)}{z - z_0} \right| = |\exp(z_0)| \cdot \left| \frac{\exp(z - z_0) - 1 - 1 \cdot (z - z_0)}{(z - z_0) - 0} \right|,$$

und die rechte Seite geht nach 0 für  $(z - z_0) \rightarrow 0$ , denn die Exponentialfunktion ist im Ursprung differenzierbar. □

**Satz 6.48.** *Seien  $f$  und  $g$  in  $z_0$  differenzierbar, und  $\alpha, \beta \in \mathbb{C}$ . Dann ist auch  $\alpha f + \beta g$  in  $z_0$  differenzierbar, und es ist*

$$(\alpha f + \beta g)'(z_0) = \alpha f'(z_0) + \beta g'(z_0).$$

*Beweis.* Wir haben für  $z \rightarrow z_0$  die Entwicklungen

$$\begin{aligned} f(z) &= f(z_0) + f'(z_0) \cdot (z - z_0) + R_f(z; z_0), \\ g(z) &= g(z_0) + g'(z_0) \cdot (z - z_0) + R_g(z; z_0). \end{aligned}$$

Dann folgt sofort

$$(\alpha f + \beta g)(z) = (\alpha f + \beta g)(z_0) + (\alpha f'(z_0) + \beta g'(z_0)) \cdot (z - z_0) + (\alpha R_f(z; z_0) + \beta R_g(z; z_0)).$$

□

Also ist die Menge der auf einem Gebiet differenzierbaren Funktionen ein Vektorraum.

**Satz 6.49 (Produktregel).** *Seien  $f$  und  $g$  in  $z_0$  differenzierbar. Dann ist auch  $f \cdot g$  in  $z_0$  differenzierbar, und es ist*

$$(f \cdot g)'(z_0) = f'(z_0)g(z_0) + f(z_0)g'(z_0).$$

*Beweis.* Wir haben

$$\begin{aligned} f(z) &= f(z_0) + f'(z_0) \cdot (z - z_0) + R_f(z; z_0), \\ g(z) &= g(z_0) + g'(z_0) \cdot (z - z_0) + R_g(z; z_0). \end{aligned}$$

Dann liefert eine simple Multiplikation

$$\begin{aligned} (f \cdot g)(z) &= (f(z_0) + f'(z_0) \cdot (z - z_0) + R_f(z; z_0)) \cdot (g(z_0) + g'(z_0) \cdot (z - z_0) + R_g(z; z_0)) \\ &= (f \cdot g)(z_0) + (f'(z_0)g(z_0) + f(z_0)g'(z_0)) \cdot (z - z_0) + R_f(z; z_0) \cdot \mathfrak{O}(1) + \mathfrak{O}(1) \cdot R_g(z; z_0), \end{aligned}$$

und der Beweis ist komplett. □

Bevor wir jetzt zur Regel für die Division kommen, betrachten wir zuerst die Funktion  $w \mapsto \frac{1}{w}$  und die Kettenregel.

**Lemma 6.50.** Die Funktion  $w_0 \mapsto \frac{1}{w_0}$  ist differenzierbar für  $w_0 \neq 0$ , und die Ableitung ist

$$\left(\frac{1}{w_0}\right)' = -\frac{1}{w_0^2}.$$

*Beweis.* Sei  $w$  nahe  $w_0$ . Dann können wir deshalb  $w \neq 0$  annehmen. Nun ist

$$\begin{aligned} \frac{1}{w} &= \frac{1}{w_0} + \left(\frac{1}{w} - \frac{1}{w_0}\right) = \frac{1}{w_0} + \frac{w_0 - w}{w_0 w} = \frac{1}{w_0} + \frac{w_0 - w}{w_0^2} + \left(\frac{w_0 - w}{w_0 w} - \frac{w_0 - w}{w_0^2}\right) \\ &= \frac{1}{w_0} - \frac{1}{w_0^2} \cdot (w - w_0) + \frac{w_0 - w}{w_0} \left(\frac{1}{w} - \frac{1}{w_0}\right) \\ &= \frac{1}{w_0} - \frac{1}{w_0^2} \cdot (w - w_0) + \frac{w_0 - w}{w_0} \cdot \frac{w_0 - w}{w w_0} \\ &= \frac{1}{w_0} - \frac{1}{w_0^2} \cdot (w - w_0) + \mathfrak{O}(|w - w_0|^2). \end{aligned}$$

Also muß der Wert der Ableitung an der Stelle  $w_0$  gerade  $-\frac{1}{w_0^2}$  sein.  $\square$

**Satz 6.51 (Kettenregel).** Seien  $f$  und  $g$  zwei Funktionen mit  $W_g \subset D_f$ . Sei  $g$  differenzierbar in  $z_0$  und  $f$  differenzierbar in  $w_0 = g(z_0)$ . Dann ist die Komposition  $f \circ g$  differenzierbar in  $z_0$ , und es ist

$$(f \circ g)'(z_0) = f'(w_0) \cdot g'(z_0).$$

*Beweis.* Wir haben

$$\begin{aligned} g(z) &= g(z_0) + g'(z_0) \cdot (z - z_0) + R_g(z; z_0), \\ f(w) &= f(w_0) + f'(w_0) \cdot (w - w_0) + R_f(w; w_0). \end{aligned}$$

Wir setzen die erste Gleichung in die zweite ein, und wählen dabei  $w = g(z)$ :

$$\begin{aligned} (f \circ g)(z) &= (f \circ g)(z_0) + f'(w_0) \cdot (g'(z_0) \cdot (z - z_0) + R_g(z; z_0)) + R_f(w, w_0) \\ &= (f \circ g)(z_0) + f'(w_0) \cdot g'(z_0) \cdot (z - z_0) + \mathfrak{O}(1)R_g(z; z_0) + R_f(w; w_0). \end{aligned}$$

Es bleibt noch zu zeigen, daß  $R_f(w; w_0) = \mathfrak{o}(|z - z_0|)$ . Wenn  $w = w_0$  sein sollte, dann ist  $R_f(w; w_0) = 0$  und wir wären in diesem Falle fertig. Ansonsten dürfen wir durch  $w - w_0$  dividieren und haben

$$\frac{R_f(w; w_0)}{z - z_0} = \frac{R_f(w; w_0)}{w - w_0} \cdot \frac{w - w_0}{z - z_0}$$

Für  $z \rightarrow z_0$  strebt  $w$  nach  $w_0$ . Dann geht der erste Faktor nach 0 und der zweite strebt zu  $g'(z_0)$ , also

$$\lim_{\substack{z \rightarrow z_0, \\ w(z) \neq w_0}} \frac{R_f(w; w_0)}{z - z_0} = 0,$$

was den Beweis der Kettenregel vollendet.  $\square$

**Satz 6.52 (Divisionsregel).** Seien die Funktionen  $f$  und  $g$  in  $z_0$  differenzierbar, und sei  $g \neq 0$  in einer Umgebung von  $z_0$ . Dann ist auch die Funktion  $f/g$  in  $z_0$  differenzierbar, und es gilt

$$\left(\frac{f}{g}\right)'(z_0) = f'(z_0) \cdot \frac{1}{g(z_0)} - f(z_0) \cdot \frac{1}{g(z_0)^2} \cdot g'(z_0).$$

*Beweis.* Folgt sofort aus der Produktregel, Lemma 6.50 und der Kettenregel.  $\square$

Man kann sich die Divisionsregel auch folgendermaßen plausibel machen: seien  $f$  und  $g$  differenzierbare Funktionen, wobei  $g \neq 0$  ist in einer Umgebung von  $z_0$ . Angenommen, eine gute Fee erzählt uns, daß auch die Funktion  $w := \frac{f}{g}$  differenzierbar wäre. Wir können ja immer schreiben  $f = w \cdot g$ , und die Auskunft der Fee gestattet es uns, hier die Produktregel anzuwenden:  $f' = w' \cdot g + w \cdot g'$ . Das braucht man dann nur noch nach  $w'$  umzustellen.

Jetzt wollen wir Umkehrfunktionen ableiten. Sei eine differenzierbare Funktion  $f$  gegeben. Angenommen, eine gute Fee sagt uns, daß  $f$  eine Umkehrfunktion  $g$  hat, und daß  $g$  sogar differenzierbar ist. Dann haben wir für alle  $z$  in der Nähe von  $z_0$  die Identität

$$g(f(z)) = z,$$

und wenn wir das gemäß Kettenregel ableiten auf beiden Seiten, bekommen wir an der Stelle  $z_0$  dann

$$g'(f(z_0)) \cdot f'(z_0) = 1.$$

Mit der Notation  $w_0 := f(z_0)$  schreibt sich das dann als  $g'(w_0) = \frac{1}{f'(z_0)}$ , unter der stillen Voraussetzung, daß das Dividieren erlaubt wäre.

Nun stellen wir uns auf den Standpunkt, daß uns heute noch keine gute Fee besucht hat:

**Satz 6.53 (Existenz der Umkehrfunktion).** *Sei  $f: D_f \rightarrow W_f$  eine stetige Funktion,  $z_0 \in D_f$  und  $w_0 = f(z_0)$ . Wenn  $f$  in einer Umgebung von  $z_0$  stetig differenzierbar ist und  $f'(z_0) \neq 0$  ist, dann existiert in der Nähe von  $w_0$  eine Umkehrfunktion  $g = g(w)$  zur Funktion  $f$ .*

Achtung: über die Differenzierbarkeit von  $g$  wird jetzt noch nichts behauptet (erst im Satz 6.54)!

*Beweisskizze.* Wir schauen uns zuerst die reelle Situation an, also  $x$  und  $x_0$  anstatt  $z$  und  $z_0$ . Nach Voraussetzung existiert  $f'$  nahe  $x_0$  und ist stetig. Wegen dieser Stetigkeit,  $f'(x_0) \neq 0$  und Lemma 6.7 hat dann die Ableitung  $f'$  in einer kompletten Umgebung von  $x_0$  immer dasselbe Vorzeichen. Sei nun  $\tilde{x}$  ein Punkt in dieser Umgebung von  $x_0$ , sodaß  $f'(\tilde{x})$  dasselbe Vorzeichen hat wie  $f'(x_0)$ . Für  $x$  nahe  $\tilde{x}$  haben wir aus der Weierstraß-Zerlegung, daß

$$\frac{f(x) - f(\tilde{x})}{x - \tilde{x}} = f'(\tilde{x}) + \frac{R_f(x; \tilde{x})}{x - \tilde{x}}.$$

Hier ist der erste Summand auf der rechten Seite nicht Null, und der Betrag des zweiten ist kleiner als  $|f'(\tilde{x})|$ , wenn  $x$  nahe genug an  $\tilde{x}$  ist. Also hat die rechte Seite das gleiche Vorzeichen wie  $f'(\tilde{x})$ , also wie  $f'(x_0)$ . Das heißt: wenn  $f'(x_0) > 0$  ist, dann gilt für  $x > \tilde{x}$ , daß  $f(x) > f(\tilde{x})$  ist; und für  $x < \tilde{x}$  gilt entsprechend  $f(x) < f(\tilde{x})$ . Also ist dann  $f$  in dieser kompletten Umgebung streng monoton. Dann können wir Satz 6.30 anwenden und erhalten, daß es eine stetige und streng monotone Umkehrfunktion  $g$  gibt.

In der komplexen Situation, also  $z, z_0 \in \mathbb{C}$ , können wir mit Monotonie nicht argumentieren, weil man in  $\mathbb{C}$  keine Ordnung findet. Ein Beweis kommt dann im zweiten Semester (Satz über implizite Funktionen).  $\square$

Wir vermerken noch, daß man auf die geforderte Stetigkeit von  $f'$  nicht verzichten kann, wie das Beispiel

$$y = f(x) = \begin{cases} 2 + x + x^2 \sin(x^{-2}) & : x \neq 0, \\ 2 & : x = 0 \end{cases}$$

zeigt. Die Ableitung  $f'$  existiert auf ganz  $\mathbb{R}$ ; und es ist  $f'(0) = 1$ . Trotzdem gibt es keine Umkehrfunktion in der Nähe von  $y = 2$ , denn die Funktion  $f = f(x)$  ist abwechselnd monoton wachsend und monoton fallend, wenn  $x$  nach 0 strebt. Nahe  $x = 0$  wechselt das Monotonieverhalten von  $f$  unendlich oft.

**Satz 6.54 (Ableitung der Umkehrfunktion).** *Sei  $f: D_f \rightarrow W_f$  stetig,  $z_0 \in D_f$ ,  $w_0 = f(z_0)$ . Sei weiterhin  $f$  in  $z_0$  differenzierbar, und sei  $f'(z_0) \neq 0$ . Wir setzen außerdem voraus, daß die Funktion  $f$  in einer Umgebung von  $z_0$  eine Umkehrfunktion  $g$  hat. Dann ist  $g$  in  $w_0$  differenzierbar, und es ist*

$$g'(w_0) = \frac{1}{f'(z_0)}.$$

*Beweis.* Wir wissen, daß

$$f(z) = f(z_0) + f'(z_0) \cdot (z - z_0) + R_f(z; z_0) \tag{6.2}$$



gilt, falls  $|z - z_0|$  klein genug ist. Hierbei ist  $\lim_{z \rightarrow z_0} \frac{R_f(z; z_0)}{z - z_0} = 0$ . Das bedeutet: wenn  $z$  nahe genug an  $z_0$  ist, dann ist der letzte Summand in (6.2) 1000 mal kleiner als der mittlere (es sei denn, der mittlere wäre identisch gleich Null; aber das ist unmöglich wegen  $f'(z_0) \neq 0$ ). Dann haben wir für solche  $z$

$$0.999|f'(z_0)| \leq \frac{|f(z) - f(z_0)|}{|z - z_0|} \leq 1.001|f'(z_0)|. \quad (6.3)$$

Wir können also  $|f(z) - f(z_0)|$  und  $|z - z_0|$  gegeneinander vergleichen. Nun setzen wir  $w = f(z)$  und  $w_0 = f(z_0)$ , also  $z = g(w)$  und  $z_0 = g(w_0)$ . Wenn  $w \neq w_0$ , dann ist nach dieser Formel also auch  $z \neq z_0$ . Wir entwickeln  $g = g(w)$  in eine Reihe:

$$\begin{aligned} g(w) &= z \\ &= z_0 + (z - z_0) = g(w_0) + (z - z_0) \\ &= g(w_0) + \frac{f(z) - f(z_0) - R_f(z; z_0)}{f'(z_0)} = g(w_0) + \frac{w - w_0}{f'(z_0)} - \frac{R_f(z; z_0)}{f'(z_0)}. \end{aligned}$$

Es bleibt zu zeigen, daß der Restterm schneller als von erster Ordnung (in  $|w - w_0|$ ) nach 0 strebt, wenn  $w \rightarrow w_0$ . Das ergibt sich aus

$$\frac{\frac{R_f(z; z_0)}{f'(z_0)}}{w - w_0} = \frac{1}{f'(z_0)} \cdot \frac{R_f(z(w); z_0)}{z(w) - z_0} \cdot \frac{z(w) - z_0}{w - w_0} = \frac{1}{f'(z_0)} \cdot \frac{z - z_0}{f(z) - f(z_0)} \cdot \frac{R_f(z(w); z_0)}{z(w) - z_0}.$$

Der erste Faktor ist beschränkt, der zweite wegen (6.3) auch, und der dritte strebt nach 0 für  $w \rightarrow w_0$ .  $\square$

Mit diesem Satz können wir die Ableitungen von Logarithmus und allgemeinen Potenzen bestimmen:

**Satz 6.55.** Für  $x > 0$  ist

$$(\ln x)' = \frac{1}{x}.$$

Für  $a \in \mathbb{R}_+$  und  $x \in \mathbb{R}$  ist

$$\begin{aligned} \frac{d}{dx} a^x &= (\ln a) a^x, \\ \frac{d}{da} a^x &= x a^{x-1}. \end{aligned}$$

*Beweis.* Sei  $y = \ln x$ , dann ist  $x = \exp(y)$  und schließlich

$$(\ln x)' = \frac{1}{(\exp(y))'} = \frac{1}{\exp(y)} = \frac{1}{x}.$$

Der zweite Teil ergibt sich aus  $a^x = \exp(x \ln a)$ , der Kettenregel und den Regeln für die Ableitung der Exponentialfunktion und Logarithmusfunktion.  $\square$

## 6.4 Mittelwertsatz und Taylorscher Satz

Die Betrachtungen dieses Abschnitts setzen *reelle* Funktionen voraus, und die meisten Ergebnisse werden falsch, wenn man stattdessen komplexe Funktionen betrachtet.

**Definition 6.56.** Sei  $f: I \rightarrow \mathbb{R}$  eine Funktion, definiert auf einem Intervall  $I \subset \mathbb{R}$ . Sei  $x_0 \in I$  ein innerer Punkt oder ein Randpunkt. Wir sagen, daß  $f$  im Punkt  $x_0$  ein lokales Maximum<sup>24</sup> hat, wenn in einer Umgebung von  $x_0$  die Ungleichung  $f(x) \leq f(x_0)$  gilt.

Analog definiert man ein lokales Minimum<sup>25</sup>.

Wenn  $f$  in einem inneren Punkt  $x_0 \in I$  ein lokales Maximum oder Minimum hat, dann redet man auch von einem lokalen Extremum<sup>26</sup>.

<sup>24</sup>local maximum

<sup>25</sup>local minimum

<sup>26</sup>local extremum

**Satz 6.57.** Sei  $f: I \rightarrow \mathbb{R}$  eine Funktion mit einem lokalen Extremum in einem inneren Punkt  $x_0 \in I$ . Wenn  $f$  in  $x_0$  differenzierbar ist, dann ist  $f'(x_0) = 0$ .

**Warnung 6.58.** Die Aussage des Satzes wird falsch, wenn  $x_0$  ein Randpunkt des Intervalles sein darf!

*Beweis.* Sei das Extremum ein Maximum (der Fall des Minimums geht ähnlich). In der Nähe von  $x_0$  gilt dann  $f(x) \leq f(x_0)$ , und somit gilt

$$\frac{f(x) - f(x_0)}{x - x_0} \begin{cases} \leq 0 & : x \geq x_0, \\ \geq 0 & : x \leq x_0. \end{cases}$$

Wenn  $x \rightarrow x_0$ , dann muß der Grenzwert des Differenzenquotienten existieren. Der linksseitige Grenzwert ist  $\geq 0$ , der rechtsseitige ist  $\leq 0$ . Also ist  $f'(x_0) = 0$ .  $\square$

**Satz 6.59 (Satz von Rolle<sup>27</sup>).** Die Funktion  $f: I \rightarrow \mathbb{R}$  sei auf dem abgeschlossenen Intervall  $I = [a, b]$  stetig und auf dem offenen Intervall  $(a, b)$  differenzierbar. Sei weiterhin  $f(a) = f(b) = 0$ . Dann existiert ein  $\xi \in (a, b)$  mit  $f'(\xi) = 0$ .

*Beweis.* Wir unterscheiden drei Fälle.

**$f$  hat in  $(a, b)$  wenigstens einen positiven Wert:** Nach dem Satz vom Maximum nimmt dann  $f$  irgendwo im Intervall  $(a, b)$  ein positives Maximum  $f(\xi)$  an. Nach dem vorigen Satz ist dann  $f'(\xi) = 0$ .

**$f$  hat in  $(a, b)$  keinen positiven, jedoch mindestens einen negativen Wert:** Dann kann man ähnlich wie im ersten Fall argumentieren.

**$f$  ist identisch 0:** Dann ist  $f'(\xi) = 0$  für jedes  $\xi \in (a, b)$ .

In jedem der drei Fälle haben wir ein  $\xi$  im Intervallinneren mit der Eigenschaft  $f'(\xi) = 0$  gefunden.  $\square$

**Satz 6.60 (Verallgemeinerter Mittelwertsatz).** Seien die Funktionen  $f, g: I \rightarrow \mathbb{R}$  auf dem abgeschlossenen Intervall  $I = [a, b]$  stetig und auf dem offenen Intervall  $(a, b)$  differenzierbar. Weiterhin sei  $g'(x) \neq 0$  auf  $(a, b)$  und  $g(a) \neq g(b)$ . Dann gibt es ein  $\xi \in (a, b)$  mit

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(\xi)}{g'(\xi)}.$$

*Beweis.* Man setze

$$h(x) = f(x) - f(a) - \frac{f(b) - f(a)}{g(b) - g(a)}(g(x) - g(a))$$

und wende den Satz von ROLLE auf die Funktion  $h$  an.  $\square$

**Satz 6.61 (Mittelwertsatz<sup>28</sup>).** Sei die Funktion  $f: I \rightarrow \mathbb{R}$  auf dem abgeschlossenen Intervall  $I = [a, b]$  stetig und auf dem offenen Intervall  $(a, b)$  differenzierbar. Dann gibt es ein  $\xi \in (a, b)$  mit

$$\frac{f(b) - f(a)}{b - a} = f'(\xi).$$

*Beweis.* Man benutze den verallgemeinerten Mittelwertsatz mit  $g(x) = x$ .  $\square$

Diese beiden unscheinbaren Mittelwertsätzchen haben viele wichtige und tiefgreifende Anwendungen. Die ersten sind zusammengefaßt im folgenden Satz:

**Satz 6.62.** Sei  $f: I \rightarrow \mathbb{R}$  stetig auf  $I = [a, b]$  und differenzierbar auf  $(a, b)$ .

<sup>27</sup> MICHEL ROLLE, 1652–1719

<sup>28</sup>mean value theorem

1. Wenn  $|f'(x)| \leq K$  für jedes  $x \in (a, b)$ , dann ist  $|f(x_1) - f(x_2)| \leq K|x_1 - x_2|$  für beliebige  $x_1, x_2 \in [a, b]$ .<sup>29</sup>
2. Wenn  $f'(x) = 0$  für jedes  $x \in (a, b)$ , dann ist  $f$  auf  $[a, b]$  konstant.
3. Wenn  $\lim_{x \rightarrow b-} f'(x)$  existiert, dann ist  $f$  in  $x = b$  linksseitig differenzierbar und  $f'$  ist in  $b$  linksseitig stetig. Analog für  $a$  anstelle von  $b$ .
4. Die Funktion  $f$  ist monoton wachsend auf  $[a, b]$  genau dann, wenn  $f'(x) \geq 0$  für jedes  $x \in (a, b)$ .
5. Wenn  $f'(x) > 0$  für jedes  $x \in (a, b)$ , dann ist  $f$  streng monoton wachsend auf  $[a, b]$ .

Die Umkehrung der letzten Aussage ist falsch.

*Beweis.*

1. Wir haben  $f(x_1) - f(x_2) = f'(\xi) \cdot (x_1 - x_2)$  für ein gewisses  $\xi$  zwischen  $x_1$  und  $x_2$ .
2. Ergibt sich sofort aus 1. mit  $K = 0$ .
3. Für  $x < b$  haben wir wegen des Mittelwertsatzes mit einem unbekanntem  $\xi \in (x, b)$

$$\frac{f(x) - f(b)}{x - b} = f'(\xi).$$

Wenn  $x$  von links nach  $b$  strebt, dann strebt auch  $\xi = \xi(x)$  nach  $b$  (Sandwichprinzip). Also existiert der Grenzwert der rechten Seite (für  $x \rightarrow b-$ ), und somit existiert auch der Grenzwert der linken Seite.

4. Wenn  $f$  monoton wachsend ist, dann hat die Differenz  $f(x_1) - f(x_2)$  dasselbe Vorzeichen wie  $x_1 - x_2$ . Also ist der Quotient dieser beiden Differenzen  $\geq 0$ . Übergang zum Grenzwert liefert  $f'(x) \geq 0$ .  
Wenn nun  $f'(x) \geq 0$  ist für jedes  $x \in (a, b)$  und  $x_1 > x_2$ , so ist auch  $f(x_1) \geq f(x_2)$  wegen des Mittelwertsatzes.
5. Der Beweis verläuft analog zum Beweis von 4. Ein Gegenbeispiel für die Umkehrung ist z.B.  $[a, b] = [-1, 1]$  und  $f = f(x) = x^7$ .

□

Eine weitere wichtige Anwendung des verallgemeinerten Mittelwertsatzes ist die Regel von BERNOULLI<sup>30</sup>–DE L'HOSPITAL<sup>31</sup>:

**Satz 6.63.** Seien  $f, g: [a, b] \rightarrow \mathbb{R}$  stetige Funktionen mit folgenden Eigenschaften:

1.  $\lim_{x \rightarrow b-} f(x) = \lim_{x \rightarrow b-} g(x) = 0$ ,
2.  $g'(x) \neq 0$  für  $x \in (a, b)$ ,
3.  $f$  und  $g$  sind differenzierbar in  $(a, b)$ ,
4. der Grenzwert

$$\lim_{x \rightarrow b-} \frac{f'(x)}{g'(x)}$$

existiert und hat den Wert  $S$ .

<sup>29</sup>Solche Funktionen heißen DEHNUNGSBESCHRÄNKT oder LIPSCHITZSTETIG, nach RUDOLF OTTO SIGISMUND LIPSCHITZ, 1832–1903

<sup>30</sup> JOHANN I BERNOULLI, 1667–1748, nicht zu verwechseln mit seinem Sohn JOHANN II BERNOULLI, 1710–1790, oder seinem Enkel JOHANN III BERNOULLI, 1744–1807, oder 5 weiteren berühmten Bernoullis

<sup>31</sup> GUILLAUME FRANCOIS ANTOINE MARQUIS DE L'HOSPITAL, 1661–1704

Dann existiert auch der Grenzwert

$$\lim_{x \rightarrow b^-} \frac{f(x)}{g(x)}$$

und hat ebenfalls den Wert  $S$ . Dies gilt auch für bestimmte Divergenz gegen  $S = \infty$ .

Eine entsprechende Aussage gilt natürlich auch für rechtsseitige Grenzwerte oder beidseitige Grenzwerte.

**Warnung 6.64.** Die Voraussetzung 4. ist entscheidend. Es gibt Beispiele, in denen  $\lim_{x \rightarrow b^-} \frac{f(x)}{g(x)}$  existiert, aber  $\lim_{x \rightarrow b^-} \frac{f'(x)}{g'(x)}$  gibt es nicht.

*Beweis von Satz 6.63.* Sei  $x < b$ . Dann gibt es ein  $\xi \in (x, b)$  mit

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(b)}{g(x) - g(b)} = \frac{f'(\xi)}{g'(\xi)}.$$

Wenn  $x$  nach  $b$  strebt, dann strebt auch  $\xi = \xi(x)$  nach  $b$ . Die rechte Seite konvergiert nach Voraussetzung 4., also auch die linke Seite.  $\square$

**Frage:** Welchen Wert hat

$$\lim_{x \rightarrow 0} \frac{e^x - 1}{\ln(1+x)} ?$$

Es gibt wichtige Variationen der Regel von Bernoulli–de l’Hospital:

**Satz 6.65.** Seien  $f, g: (a, b) \rightarrow \mathbb{R}$  stetige Funktionen mit folgenden Eigenschaften:

1.  $\lim_{x \rightarrow b^-} f(x) = \lim_{x \rightarrow b^-} g(x) = \infty$ ,
2.  $g'(x) \neq 0$  für  $x \in (a, b)$ ,
3.  $f$  und  $g$  sind differenzierbar in  $(a, b)$ ,
4. der Grenzwert

$$\lim_{x \rightarrow b^-} \frac{f'(x)}{g'(x)}$$

existiert und hat den Wert  $S$ .

Dann existiert auch der Grenzwert

$$\lim_{x \rightarrow b^-} \frac{f(x)}{g(x)}$$

und hat ebenfalls den Wert  $S$ . Dies gilt auch für bestimmte Divergenz gegen  $S = \infty$ .

*Beweis.* Lassen wir weg.  $\square$

Die beiden letzten Sätze gelten sinngemäß auch für  $b = +\infty$ . Zum Beweis betrachtet man dann

$$\lim_{x \rightarrow +\infty} \frac{f(x)}{g(x)} = \lim_{t \rightarrow 0^+} \frac{f(t^{-1})}{g(t^{-1})}$$

und wendet die herkömmliche Regel von Bernoulli–de l’Hospital an.

Und wir haben noch eine weitere Folgerung aus dem Verallgemeinerten Mittelwertsatz:

**Satz 6.66 (Taylor<sup>32</sup>scher Satz).** Die Funktion  $f: I \rightarrow \mathbb{R}$  sei auf dem abgeschlossenen Intervall  $I = [a, b]$  stetig und auf dem offenen Intervall  $(a, b)$   $n + 1$ -mal differenzierbar. Seien  $x, x_0 \in (a, b)$ . Dann gibt es ein  $\xi$  zwischen  $x$  und  $x_0$  sodaß

$$f(x) = f(x_0) + f'(x_0) \cdot (x - x_0) + \frac{1}{2!} f''(x_0) \cdot (x - x_0)^2 + \cdots + \frac{1}{n!} f^{(n)}(x_0) \cdot (x - x_0)^n + R_n(x; x_0)$$

mit einem Restterm der Form

$$R_n(x; x_0) = \frac{1}{(n+1)!} f^{(n+1)}(\xi) \cdot (x - x_0)^{n+1}. \quad (6.4)$$

Der Anteil  $f(x_0) + \cdots + \frac{1}{n!} f^{(n)}(x_0)(x - x_0)^n$  heißt auch  $n$ -tes Taylor-Polynom von  $f$  an der Stelle  $x_0$ <sup>33</sup>.

*Beweis.* Seien  $x$  und  $x_0$  fixiert, und  $t \in (a, b)$  variabel. Wir basteln uns mit Phantasie zwei Funktionen:

$$F(t) = f(x) - f(t) - f'(t)(x - t) - \frac{1}{2!} f''(t)(x - t)^2 - \cdots - \frac{1}{n!} f^{(n)}(t)(x - t)^n,$$

$$G(t) = \frac{(x - t)^{n+1}}{(n+1)!}.$$

Die Funktion  $F$  wurde genau so gewählt, daß  $F(x_0) = R_n(x; x_0)$  ist. Die Funktion  $G$  ist monoton im Intervall zwischen  $x$  und  $x_0$ , also dürfen wir den verallgemeinerten Mittelwertsatz anwenden:

$$\frac{F(x) - F(x_0)}{G(x) - G(x_0)} = \frac{F'(\xi)}{G'(\xi)},$$

für ein gewisses unbekanntes  $\xi$  zwischen  $x$  und  $x_0$ .

Wir rechnen im Folgenden die einzelnen Terme aus. Zunächst ist  $F(x) = 0$  und  $G(x) = 0$ . Weiterhin ist

$$\frac{d}{dt} \left( \frac{1}{k!} f^{(k)}(t)(x - t)^k \right) = \frac{1}{k!} f^{(k+1)}(t)(x - t)^k - \frac{1}{(k-1)!} f^{(k)}(t)(x - t)^{k-1},$$

$$F'(t) = -\frac{1}{n!} f^{(n+1)}(t)(x - t)^n,$$

$$G'(t) = -\frac{(x - t)^n}{n!}.$$

Insgesamt ist dann

$$F(x_0) = \frac{F'(\xi)}{G'(\xi)} G(x_0) = f^{(n+1)}(\xi) \frac{(x - x_0)^{n+1}}{(n+1)!},$$

und genau das war unser Ziel. □

**Beispiel 6.67.** Sei  $f = f(x) = (1 + x)^\alpha$  mit  $x \in (-1, 1) = I$  und  $\alpha \in \mathbb{R}$ . Mit  $x_0 = 0$  haben wir

$$f(x_0) = 1, \quad f'(x_0) = \alpha, \quad f''(x_0) = \alpha(\alpha - 1),$$

und somit ergibt sich die nützliche Näherung

$$(1 + x)^\alpha = 1 + \alpha \cdot x + \frac{\alpha(\alpha - 1)}{2} x^2 + \mathcal{O}(|x|^3), \quad x \rightarrow 0.$$

An dieser Stelle können wir eine in der Physik übliche Sprechweise betrachten: für eine genügend oft differenzierbare Funktion  $f$  und  $x$  nahe genug bei  $x_0$  wollen wir  $f(x)$  annähern. Dabei stellen wir uns vor, daß  $f$  eine hochkomplizierte Funktion ist, insbesondere viel komplizierter als Polynome. Wir wollen  $f$  durch das Taylorpolynom mit Entwicklungspunkt  $x_0$  annähern, und bei der Wahl von  $x_0$  lassen wir uns von zwei Gesichtspunkten leiten:

- es sollten  $x$  und  $x_0$  nahe beieinander sein, damit das Restglied  $R_n$  klein ist,

<sup>32</sup> BROOK TAYLOR, 1685–1731

<sup>33</sup>  $n$ th order Taylor polynomial of  $f$  at  $x_0$

- es sollte  $x_0$  so gewählt sein, daß wir tatsächlich die Ableitungen  $f^{(j)}(x_0)$  mit geringem Arbeitsaufwand bestimmen können.

Für die gesuchten Näherungen gibt es dann mehrere Varianten:

**Nullte Näherung:**  $f(x) \approx f(x_0)$

**Erste Näherung:**  $f(x) \approx f(x_0) + f'(x_0)(x - x_0)$

**Zweite Näherung:**  $f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2$ ,

und so weiter. Der Unterschied zwischen linker und rechter Seite ist jeweils gleich  $R_0(x; x_0)$ ,  $R_1(x; x_0)$  bzw.  $R_2(x; x_0)$ . Sinnvoll sind diese Näherungen, wenn dieser Fehler  $R_n$  deutlich kleiner ist als der letzte noch mitgenommene Summand; und wünschenswerterweise sollte die  $(k+1)$ -te Näherung besser sein als die  $k$ -te Näherung. Manchmal tut uns aber die Natur diesen Gefallen nicht, z.B. dann nicht, wenn der Graph von  $f$  „heftige Ausschläge“ aufweist, aber trotzdem noch oft genug differenzierbar ist (dann kann  $f^{(n+1)}(\xi)$  für große  $n$  riesig werden).

Wir tragen einige Formeln zusammen. Sei dazu  $f \in C([a, b] \rightarrow \mathbb{R})$  und  $x, x_0 \in (a, b)$ . Dann ist

$$f(x) = f(x_0) + f'(x_0) \cdot (x - x_0) + \mathfrak{o}(|x - x_0|), \quad x \rightarrow x_0, \quad (\text{Weierstraß-Zerlegung})$$

$$f(x) = f(x_0) + f'(\xi) \cdot (x - x_0), \quad (\text{Mittelwertsatz})$$

$$f(x) = f(x_0) + f'(\xi) \cdot (x - x_0), \quad (\text{Taylorischer Satz mit } n = 0)$$

$$f(x) = f(x_0) + f'(x_0) \cdot (x - x_0) + \frac{1}{2}f''(\xi) \cdot (x - x_0)^2 \quad (\text{Taylorischer Satz mit } n = 1).$$

Für die ersten drei Formeln benötigen wir noch  $f \in C^1([a, b] \rightarrow \mathbb{R})$  als Voraussetzung, und für die vierte Formel sogar  $f \in C^2([a, b] \rightarrow \mathbb{R})$ . Die beiden Zwischenstellen  $\xi$  in den beiden Taylorschen Sätzen sind im Allgemeinen nicht gleich. Wir beobachten, daß der Taylorsche Satz mit  $n = 0$  nichts anderes ist als der Mittelwertsatz, und der Taylorsche Satz mit  $n = 1$  erlaubt eine präzisere Beschreibung des Restterms in der Weierstraß-Zerlegung, falls  $f$  zweimal stetig differenzierbar ist.

Als eine weitere Anwendung des TAYLORSchen Satzes haben wir zum Beispiel das folgende Kriterium für das Vorliegen eines Extremums:

**Satz 6.68.** *Sei die Funktion  $f$  auf dem Intervall  $I$   $n + 1$ -mal stetig differenzierbar. In einem inneren Punkte  $x_0 \in I$  sei*

$$f'(x_0) = f''(x_0) = \dots = f^{(n)}(x_0), \quad f^{(n+1)}(x_0) \neq 0.$$

Dann gilt:

$n$  **ungerade** und  $f^{(n+1)}(x_0) > 0$ :  $f$  hat in  $x_0$  ein Minimum,

$n$  **ungerade** und  $f^{(n+1)}(x_0) < 0$ :  $f$  hat in  $x_0$  ein Maximum,

$n$  **gerade**:  $f$  hat in  $x_0$  kein Extremum, sondern die Funktion  $x \mapsto f(x) - f(x_0)$  wechselt im Punkt  $x_0$  das Vorzeichen.

*Beweis.* Wir haben

$$f(x) = f(x_0) + \frac{1}{(n+1)!} f^{(n+1)}(\xi)(x - x_0)^{n+1}.$$

Weil  $f^{(n+1)}$  stetig ist, haben  $f^{(n+1)}(\xi)$  und  $f^{(n+1)}(x_0)$  das gleiche Vorzeichen, falls  $\xi$  nahe genug bei  $x_0$  ist, siehe Lemma 6.7.  $\square$

Wenn die Funktion  $f$  unendlich oft differenzierbar ist, dann können wir einen Versuch wagen, in der TAYLORSchen Formel den Grenzübergang  $n \rightarrow \infty$  zu vollziehen. Das klappt aber nur, wenn das Restglied  $R_n$  nach 0 strebt für  $n \rightarrow \infty$ .

**Satz 6.69.** Die Funktion  $f$  sei auf dem Intervall  $I$  unendlich oft differenzierbar. Wir setzen voraus, daß es einen inneren Punkt  $x_0 \in I$  und ein kleines positives  $\delta$  gibt, sodaß

$$\lim_{n \rightarrow \infty} \left( \max_{|\xi - x_0| \leq \delta} \frac{1}{n!} |f^{(n)}(\xi)| \delta^n \right) = 0.$$

Dann gilt: für jedes  $x \in I$  mit  $|x - x_0| \leq \delta$  konvergiert die Restgliedfolge der  $R_n(x; x_0)$  für  $n \rightarrow \infty$  nach 0, und es ist

$$f(x) = \sum_{k=0}^{\infty} \frac{1}{k!} f^{(k)}(x_0)(x - x_0)^k,$$

wobei diese Reihe für  $|x - x_0| \leq \delta$  absolut konvergiert.

Diese Reihe heißt auch *Taylorreihe*<sup>34</sup>.

*Beweis.* Wir haben für jedes  $n \in \mathbb{N}$  die Zerlegung

$$\begin{aligned} f(x) &= p_n(x; x_0) + R_n(x; x_0) \\ &= \sum_{k=0}^n \frac{1}{k!} f^{(k)}(x_0)(x - x_0)^k + \frac{1}{(n+1)!} f^{(n+1)}(\xi)(x - x_0)^{n+1}, \end{aligned}$$

wobei das unbekannte  $\xi$  von  $x$  und  $n$  abhängt. Falls  $|x - x_0| \leq \delta$ , dann ist

$$\lim_{n \rightarrow \infty} R_n(x; x_0) = 0,$$

also muß der Grenzwert  $\lim_{n \rightarrow \infty} p_n(x; x_0)$  auch existieren und den Wert  $f(x)$  annehmen.  $\square$

Wir kehren noch einmal zurück zur Funktion  $f = f(x) = (1+x)^\alpha$  mit  $\alpha \in \mathbb{R}$  und  $-1 < x < 1$ . Wir setzen  $x_0 = 0$  und haben

$$f^{(n)}(x) = \alpha(\alpha-1) \cdot \dots \cdot (\alpha-n+1)(1+x)^{\alpha-n}, \quad -1 < x < 1.$$

**Definition 6.70.** Für  $\alpha \in \mathbb{R}$  und  $n \in \mathbb{N}$  definieren wir den Binomialkoeffizienten  $\binom{\alpha}{n}$  als

$$\binom{\alpha}{n} := \frac{\alpha(\alpha-1) \cdot \dots \cdot (\alpha-n+1)}{n!}.$$

Dann können wir die Funktion  $f(x) = (1+x)^\alpha$  schreiben als

$$\begin{aligned} (1+x)^\alpha &= p_n(x; 0) + R_n(x; 0) \\ &= \sum_{k=0}^n \binom{\alpha}{k} x^k + \binom{\alpha}{n+1} (1+\xi)^{\alpha-(n+1)} x^{n+1}, \quad -1 < x < 1, \end{aligned}$$

wobei das unbekannte  $\xi$  zwischen 0 und  $x$  liegt und von  $x$  und  $n$  abhängt.

Wir hätten gerne, daß

$$(1+x)^\alpha \stackrel{?}{=} \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n, \quad -1 < x < 1.$$

Um dies zu beweisen, reicht es aber nicht aus zu zeigen, daß der Grenzwert  $\lim_{n \rightarrow \infty} p_n(x; 0)$  existiert (das tut er für  $-1 < x < 1$ , wie man mit dem Quotientenkriterium schnell sieht).

Wir brauchen mehr. Nämlich, daß dieser Grenzwert nicht bloß schön existiert, sondern auch noch den Wert  $f(x) = (1+x)^\alpha$  hat. Dies ist aber gleichbedeutend mit

$$\lim_{n \rightarrow \infty} R_n(x; 0) \stackrel{?}{=} 0.$$

---

<sup>34</sup>Taylor series

Sei  $x \in (-1, 1)$  fixiert und  $n$  so groß, daß  $\alpha - n < 0$ . Dann ist

$$\begin{aligned} |R_n(x; 0)| &\leq \frac{|\alpha \cdot (\alpha - 1) \cdot \dots \cdot (\alpha - n)|}{(n+1)!} (1 - |x|)^{\alpha - (n+1)} |x|^{n+1} \\ &= (1 - |x|)^\alpha \frac{|\alpha \cdot (\alpha - 1) \cdot \dots \cdot (\alpha - n)|}{(n+1)!} \left( \frac{|x|}{1 - |x|} \right)^{n+1} \\ &=: S_n(x), \end{aligned}$$

wobei  $S_n$  für „Schranke“ steht. Der Quotient  $S_{n+1}/S_n$  strebt nach  $\frac{|x|}{1-|x|}$ , und dies ist  $< 1$  für  $|x| < \frac{1}{2}$ . Damit können wir zeigen <sup>35</sup>, daß

$$\lim_{n \rightarrow \infty} R_n(x; 0) = 0 \quad \text{für } |x| < \frac{1}{2}.$$

Auf diesem Wege haben wir bewiesen, daß

$$(1+x)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n, \quad \alpha \in \mathbb{R}, \quad -\frac{1}{2} < x < \frac{1}{2}. \quad (6.5)$$

Dies ist die berühmte *binomische Reihe*. Im dritten Semester werden wir stärkere Werkzeuge kennenlernen, die uns garantieren, daß die Formel (6.5) auch für  $-1 < x < 1$  gilt.

**Warnung 6.71.** Die Funktion

$$f(x) = \begin{cases} e^{-1/x^2} & : x \neq 0, \\ 0 & : x = 0 \end{cases}$$

ist auf ganz  $\mathbb{R}$  definiert, dort stetig und überall unendlich oft differenzierbar. Sämtliche Ableitungen  $f^{(n)}(0)$  sind gleich 0. Die Folge der Werte der Taylorpolynome  $p_n(x; 0)$  konvergiert für  $n \rightarrow \infty$ , nämlich nach 0. Aber offensichtlich ist

$$0 = \lim_{n \rightarrow \infty} p_n(x; 0) \neq f(x)$$

falls  $x \neq 0$ , denn für solche  $x$  ist  $f(x) \neq 0$ .

**Frage:** Wie sieht der Graph dieser Funktion  $f$  aus ?

**Satz 6.72.** Für  $-1 < x \leq 1$  gilt

$$\ln(1+x) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n}.$$

Drei Beweise für Neugierige.

Ein **erster Beweis** verläuft so: wir besorgen uns die Potenzreihe für die Funktion  $x \mapsto \frac{1}{1+x}$ , wobei die Summenformel der geometrischen Reihe nützlich ist. Dann integrieren wir diese Potenzreihe summandenweise (und im zweiten Semester zeigen wir dann, daß diese summandenweise Integration tatsächlich zulässig ist).

Ein **zweiter Beweis** funktioniert sehr ähnlich zum Beweis von (6.5), aber er hat den Nachteil, daß wir die Konvergenz der Restgliedfolge nach Null nur für  $-\frac{1}{2} < x < 1$  hinbekommen, da die Restgliedformel (6.4) (auch bekannt als die *Lagrange-Form des Restglieds*) hier nicht kräftig genug ist.

Als Alternative verschaffen wir uns im **dritten Beweis** eine andere Restglieddarstellung, die zwar etwas häßlicher ist, aber zumindest bei der Logarithmusfunktion leistungsfähiger. Dafür schauen wir uns den Beweis von Satz 6.66 nochmal an, jetzt aber wählen wir  $G = G(t) = x - t$ . Dann lassen wir den dort vorgeführten Beweis nochmal durchlaufen, und als Ergebnis ergibt sich dann

$$R_n(x; x_0) = F(x_0) = \frac{1}{n!} f^{(n+1)}(\xi) \cdot \left( \frac{x - \xi}{x - x_0} \right)^n \cdot (x - x_0)^{n+1},$$

<sup>35</sup> Wenn wir etwas sorgfältiger arbeiten und uns überlegen, wo  $\xi$  liegen kann, dann bekommen wir  $\lim_{n \rightarrow \infty} R_n(x; 0) = 0$  sogar für  $-\frac{1}{2} < x < 1$ .



mit einem  $\xi$  zwischen  $x$  und  $x_0$  (mehr ist von  $\xi$  nicht bekannt). Dies ist die *Cauchy-Form des Restglieds*, und sicherlich hat  $\xi$  jetzt nicht mehr denselben Wert wie in (6.4). Für unsere Logarithmusfunktion ist  $f(x) = \ln(1+x)$  sowie

$$f^{(n)}(x) = (-1)^{n-1}(n-1)! \frac{1}{(1+x)^n},$$

und daraus ergibt sich dann (mit  $x_0 = 0$ )

$$|R_n(x; 0)| \leq \frac{1}{n!} \cdot n! \cdot \frac{1}{|1+\xi|^{n+1}} \cdot \left| \frac{x-\xi}{x-0} \right|^n \cdot |x-0|^{n+1} = \frac{|x|}{|1+\xi|} \cdot \left| \frac{x-\xi}{1+\xi} \right|^n.$$

Wegen  $-1 < x < 1$  ist auf jeden Fall  $1 + \xi > 0$ .

Sei nun  $0 < x < 1$ . Dann haben wir  $0 < \xi < x < 1$ , also folgt  $|x - \xi| < |x|$  und  $1 + \xi > 1$ ; somit ist dann  $|R_n(x; 0)| < |x|^{n+1}$ .

Sei nun  $-1 < x < 0$ . Dann haben wir  $-1 < x < \xi < 0$ , also folgt  $|x - \xi| = |x| - |\xi|$  und  $1 + \xi = 1 - |\xi|$ ; somit ist dann

$$\begin{aligned} |R_n(x; 0)| &= \frac{|x|}{1-|\xi|} \cdot \left( \frac{|x| - |\xi|}{1-|\xi|} \right)^n \\ &= \frac{|x|}{1-|\xi|} \cdot \left( \frac{|x| \cdot (1-|\xi|) + |x| \cdot |\xi| - |\xi|}{1-|\xi|} \right)^n \\ &= \frac{|x|}{1-|\xi|} \cdot \left( |x| - |\xi| \cdot \frac{1-|x|}{1-|\xi|} \right)^n \quad \left| \quad 1-|x| < 1-|\xi| \right. \\ &\leq \frac{|x|}{1-|\xi|} \cdot |x|^n \\ &\leq \frac{|x|^{n+1}}{1-|x|}, \end{aligned}$$

also gilt wegen  $|x| < 1$  tatsächlich  $\lim_{n \rightarrow \infty} R_n(x; 0) = 0$ .

Offen bleibt (in allen drei Beweisen) der Fall  $x = 1$ , den wir nicht weiter verfolgen wollen.  $\square$

Insbesondere erhalten wir die früher schon benutzte Identität

$$\ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} \pm \dots$$

Wir fassen die beiden Stolperfallen bei der Konvergenz von Taylorreihen zusammen:

- Es kann geschehen, daß bei unpassender Wahl von  $x$  die Taylorreihe divergiert. Zum Beispiel hat die Funktion  $f = f(x) = \frac{1}{1-x}$ , entwickelt am Punkt  $x_0 = 0$ , die Taylorreihe

$$f(x) = \sum_{n=0}^{\infty} x^n,$$

die nun leider für  $x = 13$  divergiert. Die Taylorreihe konvergiert nur für  $-1 < x < 1$ .

- Es kann geschehen, daß die Taylorreihe zwar konvergiert, aber nicht zur Funktion  $f$ . Siehe Warnung 6.71 für ein Beispiel dafür. Dieses zweite Problem tritt in der Praxis allerdings seltener auf.

## 6.5 Elementare Funktionen

### 6.5.1 Der geometrische Zugang zu den Winkelfunktionen

Die Winkelfunktion  $\sin$ ,  $\cos$  und  $\tan$ <sup>36</sup> für Winkel zwischen  $0$  und  $\frac{\pi}{2}$  werden wie üblich am rechtwinkligen Dreieck definiert:

$$\sin = \frac{\text{Gegenkathete}}{\text{Hypotenuse}}, \quad \cos = \frac{\text{Ankathete}}{\text{Hypotenuse}}, \quad \tan = \frac{\sin}{\cos} = \frac{\text{Gegenkathete}}{\text{Ankathete}}.$$

<sup>36</sup>sine, cosine, tangent

Wenn nun  $\varphi \in \mathbb{R}$  beliebig ist, dann definiert man  $\sin \varphi$  und  $\cos \varphi$  wie folgt:

In einem kartesischen Koordinatensystem startet man im Punkt  $(1, 0)$  und läuft auf dem Einheitskreis im Gegenuhrzeigersinn den Winkel  $\varphi$  ab. Die kartesischen Koordinaten desjenigen Punktes, an dem man dann ankommt, definieren die Zahlen  $\cos \varphi$  (als  $x$ -Koordinate) und  $\sin \varphi$  (als  $y$ -Koordinate). Falls  $\cos \varphi \neq 0$ , dann definiert man  $\tan \varphi := \frac{\sin \varphi}{\cos \varphi}$ . Die Länge des auf dem Einheitskreis zurückgelegten Weges ist gerade gleich  $\varphi$ .

Man sieht dann durch Hinschauen:

**Satz 6.73.** *Die geometrisch definierten Winkelfunktionen  $\sin$  und  $\cos$  sind stetig und haben folgende Eigenschaften für jedes  $\varphi \in \mathbb{R}$ :*

$$\begin{array}{ll} \sin(\varphi + 2\pi) = \sin \varphi, & \cos(\varphi + 2\pi) = \cos \varphi, \\ \sin(\varphi + \pi) = -\sin \varphi, & \cos(\varphi + \pi) = -\cos \varphi, \\ \sin\left(\frac{\pi}{2} - \varphi\right) = \cos \varphi, & \cos\left(\frac{\pi}{2} - \varphi\right) = \sin \varphi, \\ \sin(\pi - \varphi) = \sin \varphi, & \cos(\pi - \varphi) = -\cos \varphi, \\ \sin(-\varphi) = -\sin \varphi, & \cos(-\varphi) = \cos \varphi, \end{array}$$

sowie

$$\begin{array}{l} \sin\left(\varphi + \frac{\pi}{2}\right) = \cos \varphi, \\ \sin^2 \varphi + \cos^2 \varphi = 1. \end{array}$$

Die Funktionen  $\sin$  und  $\cos$  haben die Periodenlänge  $2\pi$ , während  $\tan$  die halbe Periodenlänge  $\pi$  hat.

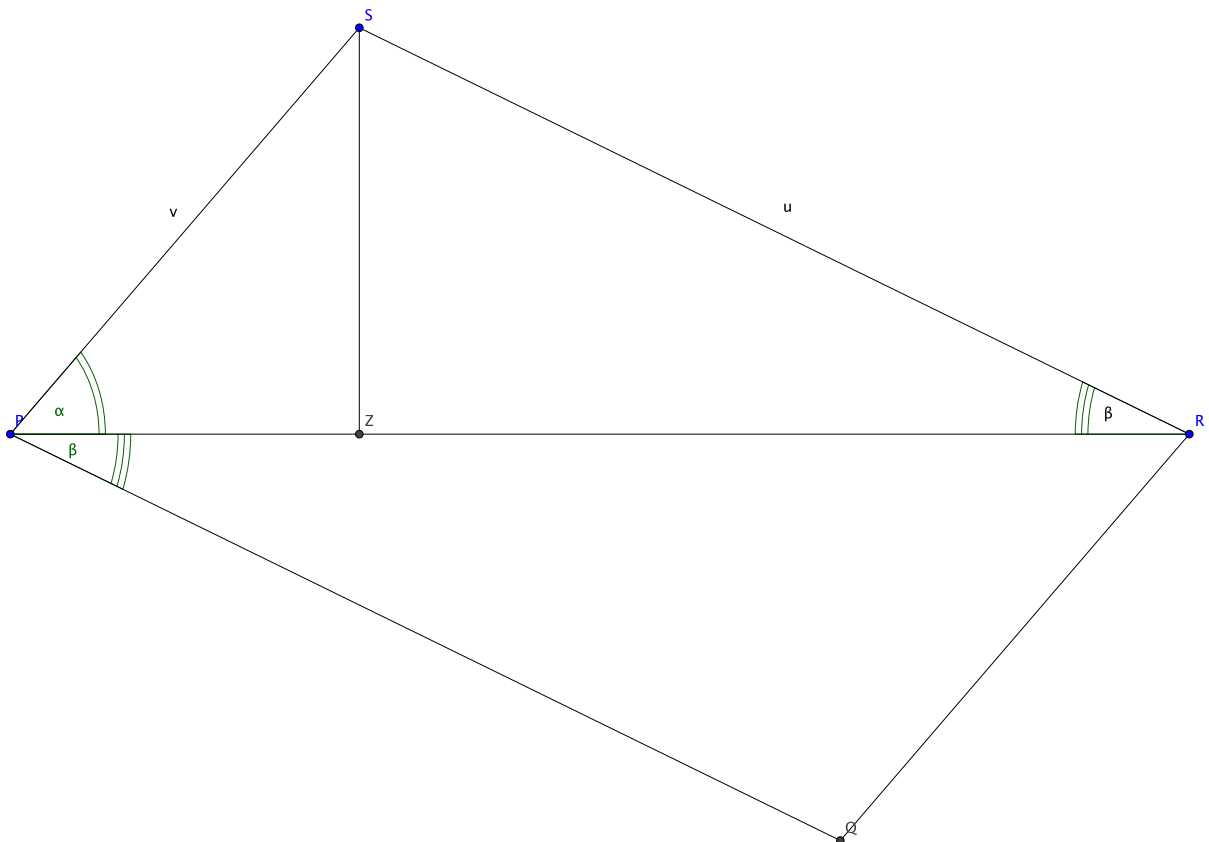


Abbildung 6.1: Zum Additionstheorem des Sinus

**Satz 6.74 (Additionstheoreme).** Für  $\alpha, \beta \in \mathbb{R}$  gilt

$$\begin{aligned}\sin(\alpha + \beta) &= \sin \alpha \cos \beta + \cos \alpha \sin \beta, \\ \cos(\alpha + \beta) &= \cos \alpha \cos \beta - \sin \alpha \sin \beta, \\ \sin \alpha - \sin \beta &= 2 \cos \left( \frac{\alpha + \beta}{2} \right) \sin \left( \frac{\alpha - \beta}{2} \right), \\ \cos \alpha - \cos \beta &= -2 \sin \left( \frac{\alpha + \beta}{2} \right) \sin \left( \frac{\alpha - \beta}{2} \right).\end{aligned}$$

*Beweis.* Die letzten beiden Formeln folgen aus den ersten beiden (wie ?)

Wir beginnen mit dem Beweis des Additionstheorems für den Sinus im Falle von  $0 < \alpha < \pi/2$  und  $0 < \beta < \pi/2$ , vgl. Abbildung 6.1. Wir zeichnen ein Dreieck  $PRS$  mit Innenwinkel  $\alpha$  bei  $P$  und Innenwinkel  $\beta$  bei  $R$ . Wir nennen die Streckenlängen  $u := RS$  und  $v := SP$ . Dann drehen wir das Dreieck  $PRS$  um den Mittelpunkt der Strecke  $PR$  um  $\pi$ , wobei  $S$  auf einen Punkt  $Q$  abgebildet wird. Es entsteht ein Parallelogramm  $PQRS$  mit Innenwinkel  $\alpha + \beta$  bei  $P$ . Der Flächeninhalt dieses Parallelogramms ist

$$A = uv \cdot \sin(\alpha + \beta). \quad (6.6)$$

Der Lotfußpunkt von  $S$  auf  $PR$  sei getauft auf den Namen  $Z$ . Für den Flächeninhalt des rechtwinkligen Dreiecks  $PZS$  haben wir dann

$$A_{PZS} = \frac{1}{2} PZ \cdot ZS = \frac{1}{2} \cdot (v \cos \alpha) \cdot (u \sin \beta) = \frac{uv}{2} \cos \alpha \sin \beta.$$

Für den Flächeninhalt des rechtwinkligen Dreiecks  $ZRS$  folgt analog

$$A_{ZRS} = \frac{1}{2} ZR \cdot ZS = \frac{1}{2} \cdot (u \cos \beta) \cdot (v \sin \alpha) = \frac{uv}{2} \sin \alpha \cos \beta.$$

Der Gesamtflächeninhalt des Parallelogramms ist dann

$$A = 2(A_{PZS} + A_{ZRS}) = uv \cdot (\cos \alpha \sin \beta + \sin \alpha \cos \beta).$$

Wir vergleichen dies mit (6.6), und das gewünschte Additionstheorem für den Sinus steht da, zumindest für den Fall  $0 < \alpha, \beta < \pi/2$ . Weil Sinus/Cosinus ungerade/gerade Funktionen sind, bekommen wir das Additionstheorem für den Sinus auch im Fall  $-\pi/2 < \alpha, \beta < 0$ .

Als nächstes wollen wir das Additionstheorem des Sinus zeigen im Falle, daß einer der beiden Winkel im Intervall  $(0, \pi/2)$  liegt, und der andere im Intervall  $(-\pi/2, 0)$ . Mit ein wenig Überlegung erkennt man, daß es genügt zu zeigen:

$$\sin(\beta - \alpha) = \sin \beta \cos \alpha - \cos \beta \sin \alpha, \quad 0 < \alpha < \beta < \pi/2.$$

Dazu zeichnen wir ein rechtwinkliges Dreieck  $PQR$  mit rechtem Winkel bei  $Q$  und Winkel  $\beta$  bei  $P$ , und darin enthalten ein weiteres rechtwinkliges Dreieck  $PQT$  mit rechtem Winkel bei  $Q$  und Winkel  $\alpha$  bei  $P$ . Wir taufen Streckenlängen:  $w := PR$ ,  $v := PT$ ,  $u := PQ$ . Siehe Abbildung 6.2. Dann haben wir

$$u = w \cos \beta, \quad u = v \cos \alpha \quad \implies \quad v = w \frac{\cos \beta}{\cos \alpha}.$$

Der Flächeninhalt des Dreiecks  $PTR$  ist einerseits

$$A_{PTR} = \frac{1}{2} wv \sin(\beta - \alpha) = \frac{w^2 \cos \beta}{2 \cos \alpha} \sin(\beta - \alpha).$$

Andererseits haben wir  $PTR$  als „Differenzdreieck“ zweier rechtwinkliger Dreiecke, also

$$\begin{aligned}A_{PTR} &= A_{PQR} - A_{PQT} = \frac{1}{2} wu \sin \beta - \frac{1}{2} wv \sin \alpha = \frac{w^2}{2} \left( \cos \beta \sin \beta - \cos \beta \frac{\cos \beta}{\cos \alpha} \sin \alpha \right) \\ &= \frac{w^2 \cos \beta}{2 \cos \alpha} \left( \cos \alpha \sin \beta - \cos \beta \sin \alpha \right).\end{aligned}$$

Das wollten wir zeigen. Also gilt das Additionstheorem des Sinus jetzt sogar im Falle von  $-\pi/2 \leq \alpha, \beta \leq \pi/2$ , denn diese Winkelfunktionen sind stetig, also können wir zum abgeschlossenen Intervall übergehen.

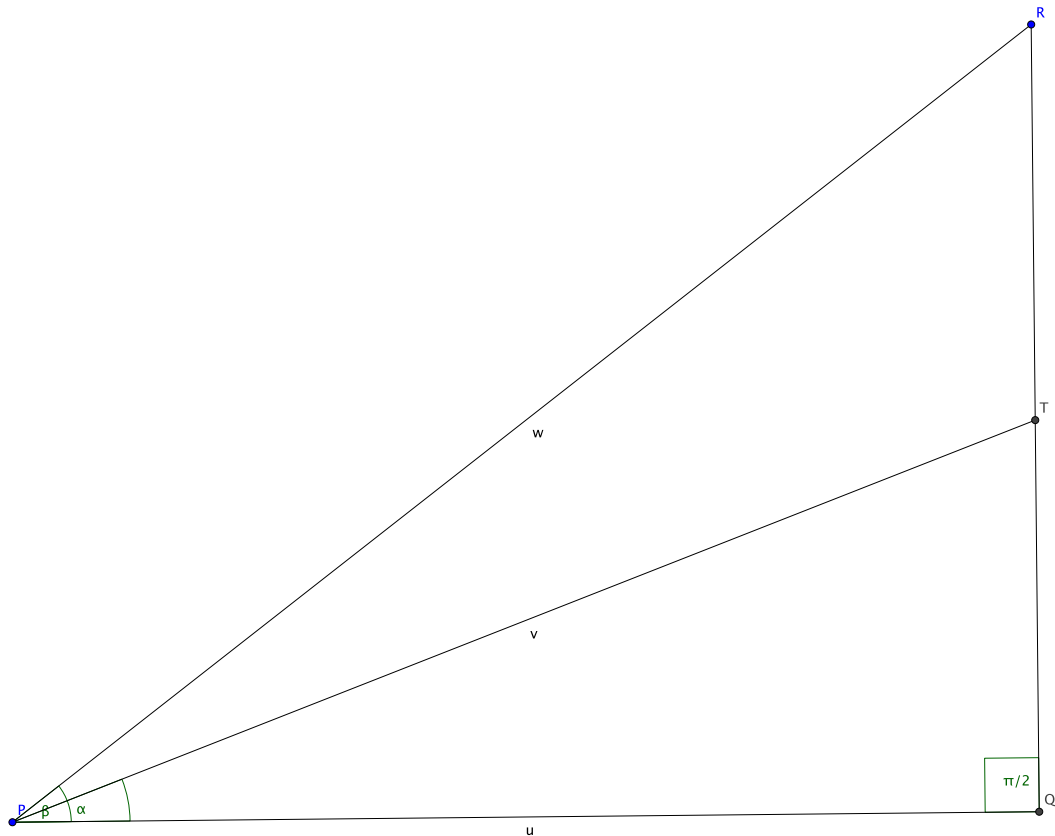


Abbildung 6.2: Zum Additionstheorem des Sinus

Aufgrund von  $\sin(\gamma + \pi) = -\sin \gamma$  und  $\cos(\gamma + \pi) = -\cos \gamma$  haben wir das Additionstheorem dann endlich für alle  $\alpha, \beta \in \mathbb{R}$ .

Und das Additionstheorem für den Cosinus ergibt sich dann sofort gemäß

$$\begin{aligned} \cos(\alpha + \beta) &= \sin\left(\frac{\pi}{2} - (\alpha + \beta)\right) = \sin\left(\left(\frac{\pi}{2} - \alpha\right) + (-\beta)\right) \\ &= \sin\left(\frac{\pi}{2} - \alpha\right) \cos(-\beta) + \cos\left(\frac{\pi}{2} - \alpha\right) \sin(-\beta) \\ &= \cos \alpha \cos \beta - \sin \alpha \sin \beta, \end{aligned}$$

wobei wir hier Satz 6.73 benutzt haben. □

Aus den Additionstheoremen und einfachen geometrischen Überlegungen ergeben sich dann die Formeln für die Ableitungen:

**Satz 6.75.** Für alle  $\varphi \in \mathbb{R}$  gilt

$$(\sin \varphi)' = \cos \varphi, \quad (\cos \varphi)' = -\sin \varphi.$$

*Beweis.* Geometrisch ist klar, daß  $\sin 0 = 0$  und  $\cos 0 = 1$ .

Wir zeigen als erstes, daß  $(\sin')(0) = 1$ .

Sei  $0 < \varphi < \frac{\pi}{2}$ . Einfache Flächeninhaltsüberlegungen am Kreis zeigen, daß

$$\sin \varphi < \varphi < \tan \varphi \iff 1 < \frac{\varphi}{\sin \varphi} < \frac{\tan \varphi}{\sin \varphi} = \frac{1}{\cos \varphi}.$$

Das Sandwichprinzip liefert dann  $\lim_{\varphi \rightarrow 0} \frac{\sin \varphi}{\varphi} = 1$ . Wegen  $\sin 0 = 0$  ist das aber gerade die Formel für  $(\sin')(0) = 1$ .

Sei nun  $\varphi_0 \in \mathbb{R}$  beliebig und  $\varphi \neq \varphi_0$ . Aus Satz 6.74 erhalten wir dann

$$\frac{\sin \varphi - \sin \varphi_0}{\varphi - \varphi_0} = 2 \cos \left( \frac{\varphi + \varphi_0}{2} \right) \frac{\sin \left( \frac{\varphi - \varphi_0}{2} \right)}{\varphi - \varphi_0} = \cos \left( \frac{\varphi + \varphi_0}{2} \right) \frac{\sin \left( \frac{\varphi - \varphi_0}{2} \right)}{\frac{\varphi - \varphi_0}{2}}.$$

Für  $\varphi \rightarrow \varphi_0$  strebt die rechte Seite dann gegen  $\cos \varphi_0$ .

Die Ableitungsformel für den Cosinus ergibt sich aus der Ableitungsformel für den Sinus und  $\cos \varphi = \sin(\pi/2 - \varphi)$ .  $\square$

Weil (modulo Vorzeichen) eine Winkelfunktion die Ableitung der anderen ist und umgekehrt, sind die Winkelfunktionen  $\sin$  und  $\cos$  unendlich oft differenzierbar. Dann stellt sich naturgemäß die Frage nach der Taylorreihe, und ob sie konvergiert, und wenn ja, gegen welchen Grenzwert. Diese Frage werden wir im nächsten Abschnitt beantworten (obwohl wir es auch jetzt schon könnten).

Wir vermerken nur kurz zum Schluß, daß die Sinusfunktion die (einzige) Lösung des Anfangswertproblems

$$y'' + y = 0, \quad y(0) = 0, \quad y'(0) = 1$$

ist; und die Kosinusfunktion ist die (wiederum einzige) Lösung des Anfangswertproblems

$$y'' + y = 0, \quad y(0) = 1, \quad y'(0) = 0.$$

Die Einzigkeit der Lösungen der genannten Anfangswertproblem war gerade Inhalt von Lemma 4.36.

### 6.5.2 Der analytische Zugang zu den Winkelfunktionen

Die Ausführungen des vorigen Abschnitts haben einen kleinen Nachteil: sie sind logisch etwas fragwürdig angeordnet. Dazu sollten wir kurz über das Konzept der Definition nachdenken: eine Definition bedeutet, daß der betreffende Begriff „geboren“ wird. Bekanntlich kann niemand mehrfach geboren werden, und genauso kann auch kein Begriff mehrfach definiert werden. Weiterhin verwendet man beim Definieren neuer Begriffe andere Begriffe, die schon vorher definiert worden waren, sodaß sich im Laufe eines dreisemestrigen Kurses ein großer Stammbaum von Begriffen ergibt.

Im vorigen Abschnitt hatten wir allerdings einige Begriffe benutzt, von denen wir gar nicht wissen, wie wir sie definieren sollen:

- Winkel,
- Länge eines Kreisbogens,
- Flächeninhalt eines Dreiecks bzw. Kreissektors,

und sicherlich noch einige andere mehr.

Es hat sich deshalb in der Mathematik eingebürgert, die Winkelfunktionen auf dem Wege der Analysis einzuführen und, darauf aufbauend, die geometrischen Begriffe hinterher zu definieren. Zum Beispiel definiert man Flächeninhalte und Kurvenlängen als bestimmte Integrale.

Wir stellen uns jetzt auf den Standpunkt, von einer Zahl namens  $\pi$  noch nie etwas gehört zu haben. Wir werden ihr erst zum Schluß wiederbegegnen. Wir wissen auch gar nicht, was ein Kreis ist.

Wir beginnen mit einem kleinen Satz zur Exponentialfunktion.

**Satz 6.76.** Für  $z \in \mathbb{C}$  ist  $\overline{\exp(z)} = \exp(\bar{z})$ . Für  $x \in \mathbb{R}$  ist  $|\exp(ix)| = 1$ .

*Beweis.* Die erste Behauptung erhalten wir aus

$$\overline{\exp(z)} = \overline{\sum_{k=0}^{\infty} \frac{z^k}{k!}} = \sum_{k=0}^{\infty} \frac{\overline{z^k}}{k!} = \sum_{k=0}^{\infty} \frac{(\bar{z})^k}{k!} = \exp(\bar{z}).$$

Nun ist einerseits

$$\exp(ix) \exp(-ix) = \exp(ix - ix) = \exp(0) = 1,$$

andererseits nach der ersten Behauptung

$$\overline{\exp(ix)} = \exp(\overline{ix}) = \exp(-ix).$$

Insgesamt ergibt sich  $\exp(ix)\overline{\exp(ix)} = 1$ , also  $|\exp(ix)| = 1$ . □

**Definition 6.77.** Für  $x \in \mathbb{R}$  definieren wir zwei Funktionen  $c = c(x)$  und  $s = s(x)$  durch

$$c(x) = \Re \exp(ix), \quad s(x) = \Im \exp(ix).$$

**Satz 6.78.** Diese Funktionen haben die folgenden Eigenschaften für beliebige  $x, y \in \mathbb{R}$ :

$$\begin{aligned} c(x) &= \frac{1}{2}(\exp(ix) + \exp(-ix)), & s(x) &= \frac{1}{2i}(\exp(ix) - \exp(-ix)), \\ c(0) &= 1, & s(0) &= 0, \\ c(-x) &= c(x), & s(-x) &= -s(x), \\ c^2(x) + s^2(x) &= 1, \\ c(x+y) &= c(x)c(y) - s(x)s(y), & s(x+y) &= s(x)c(y) + s(y)c(x), \\ c'(x) &= -s(x), & s'(x) &= c(x). \end{aligned}$$

*Beweis.* Die erste Zeile folgt aus  $\overline{\exp(ix)} = \exp(-ix)$ .

Aus  $\exp(0) = 1$  folgt  $c(0) = 1$  und  $s(0) = 0$ .

Die  $c$ -Funktion ist gerade, denn wegen des Satzes 6.76 haben wir

$$c(-x) = \Re \exp(-ix) = \Re \overline{\exp(ix)} = \Re \exp(ix) = c(x),$$

und analog argumentiert man für die  $s$ -Funktion. Aus Satz 6.76 bekommt man auch  $c^2(x) + s^2(x) = 1$ .

Die Additionstheoreme für  $c$  und  $s$  folgen aus

$$\begin{aligned} c(x+y) + is(x+y) &= \exp(i(x+y)) = \exp(ix) \exp(iy) \\ &= (c(x) + is(x))(c(y) + is(y)) \\ &= (c(x)c(y) - s(x)s(y)) + i(c(x)s(y) + s(x)c(y)) \end{aligned}$$

und Vergleich von Realteil und Imaginärteil auf beiden Seiten.

Die Beziehung für die Ableitungen läßt sich ähnlich zeigen. Wir starten mit den Formeln

$$c(x) = \Re \exp(ix) = \frac{1}{2}(\exp(ix) + \exp(-ix)), \quad s(x) = \Im \exp(ix) = \frac{1}{2i}(\exp(ix) - \exp(-ix)).$$

Die rechten Seiten sind differenzierbare Funktionen, also müssen auch die linken Seiten differenzierbar sein. Dann erhalten wir nach kurzer Rechnung  $c'(x) = -s(x)$  und  $s'(x) = c(x)$ , was den Beweis komplettiert. □

**Satz 6.79.** Für jedes  $x \in \mathbb{R}$  gilt

$$c(x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!}, \quad s(x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!}.$$

Diese Reihen konvergieren für jedes  $x \in \mathbb{R}$  absolut.

*Beweis.* Wir haben  $c(x) = \frac{1}{2}(\exp(ix) + \exp(-ix))$  und  $s(x) = \frac{1}{2i}(\exp(ix) - \exp(-ix))$  sowie

$$\exp(ix) = \sum_{n=0}^{\infty} \frac{i^n x^n}{n!}, \quad \exp(-ix) = \sum_{n=0}^{\infty} \frac{(-1)^n i^n x^n}{n!}.$$

Der Rest ist Arithmetik. □

Als nächstes untersuchen wir die Funktionen  $c$  und  $s$  auf Monotonie. Wir haben für  $c$  die Potenzreihe

$$c(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} \mp \dots,$$

und es gilt die Abschätzung

$$c(x) \leq 1 - \frac{x^2}{2!} + \frac{x^4}{4!}, \quad \text{wenn } x^2 \leq 56,$$

denn für solche  $x$  ist der Reihenrest  $-\frac{1}{6!}x^6 + \frac{1}{8!}x^8 \mp$  eine Leibnizreihe, also nicht positiv, denn der erste Summand dieser Leibnizreihe ist negativ.

Für  $x = 2$  haben wir also  $c(2) \leq -\frac{1}{3}$ . Andererseits ist die  $c$ -Funktion stetig und  $c(0) = 1$ . Wegen des Zwischenwertsatzes hat die Funktion  $c$  also mindestens eine Nullstelle zwischen  $x = 0$  und  $x = 2$ . Könnte vielleicht  $c$  dort mehr als eine Nullstelle besitzen? Dazu beobachten wir, daß im Intervall  $(0, 2]$  gilt, daß

$$s(x) = x \left( 1 - \frac{x^2}{6} + \frac{x^4}{120} - + \dots \right) = x \left( 1 - \frac{x^2}{6} \right) + (\text{positiv}) > x \left( 1 - \frac{x^2}{6} \right),$$

und dies ist positiv für  $0 < x \leq 2$ . Wegen  $c' = -s$  ist dann aber  $c$  auf dem Intervall  $[0, 2]$  streng monoton fallend, und demnach besitzt  $c$  auf diesem Intervall genau eine Nullstelle. Man kann errechnen, daß sie folgenden Wert hat:

$$p_2 = 1.570796326794897 \dots$$

**Definition 6.80.** Wir definieren

$$\pi := 2p_2.$$

Das ist einfach nur eine Schreibweise, aber Assoziationen zu schulischen Bildungserlebnissen sind nicht unbeabsichtigt.

Wegen  $c(p_2) = 0$  und  $c^2(x) + s^2(x) = 1$  muß  $|s(p_2)| = 1$  sein. Weil aber  $s'(x) = c(x)$  und  $c > 0$  auf dem Intervall  $[0, p_2)$ , gilt demnach  $s(p_2) = 1$ . Mit der neuen Schreibweise  $\frac{\pi}{2} = p_2$  erhalten wir so

$$\exp\left(i\frac{\pi}{2}\right) = i.$$

Potenzieren gibt  $\exp(i\pi) = -1$  und  $\exp(2i\pi) = 1$ , also auch  $c(\pi) = -1$ ,  $s(\pi) = 0$ ,  $c(2\pi) = 1$ ,  $s(2\pi) = 0$ .

Die Additionstheoreme für die Funktionen  $c$  und  $s$  liefern dann

$$\begin{aligned} c\left(x + \frac{\pi}{2}\right) &= c(x) \cdot 0 - s(x) \cdot 1 = -s(x), \\ c(x + \pi) &= c(x) \cdot (-1) - s(x) \cdot 0 = -c(x), \\ c(x + 2\pi) &= c(x) \cdot 1 - s(x) \cdot 0 = c(x), \\ s\left(x + \frac{\pi}{2}\right) &= s(x) \cdot 0 + c(x) \cdot 1 = c(x), \\ s(x + \pi) &= s(x) \cdot (-1) + c(x) \cdot 0 = -s(x), \\ s(x + 2\pi) &= s(x) \cdot 1 + c(x) \cdot 0 = s(x). \end{aligned}$$

**Satz 6.81.** Für jedes  $x \in \mathbb{R}$  gilt  $c(x) = \cos(x)$ ,  $s(x) = \sin(x)$  sowie

$$\exp(ix) = \cos(x) + i \sin(x). \tag{6.7}$$

Hierbei sind  $\cos$  und  $\sin$  geometrisch definiert.

*Beweis.* Die Funktionen  $c$  und  $\cos$  sind Lösung des Anfangswertproblems

$$y'' + y = 0, \quad y(0) = 1, \quad y'(0) = 0,$$

und die Funktionen  $s$  und  $\sin$  sind Lösung des Anfangswertproblems

$$y'' + y = 0, \quad y(0) = 0, \quad y'(0) = 1.$$

Da es aber (wegen Lemma 4.36) nur jeweils eine einzige Lösung geben kann, muß  $c(x) = \cos(x)$  und  $s(x) = \sin(x)$  gelten. Die Gleichung (6.7) folgt dann aus der Definition von  $c$  und  $s$ .  $\square$

**Definition 6.82 (Komplexe Winkelfunktionen).** Für  $z \in \mathbb{C}$  definieren wir die Winkelfunktionen

$$\cos(z) = \frac{1}{2}(\exp(iz) + \exp(-iz)), \quad \sin(z) = \frac{1}{2i}(\exp(iz) - \exp(-iz)).$$

Offensichtlich sind diese Funktionen in ganz  $\mathbb{C}$  unendlich oft differenzierbar, und Satz 6.74 sowie Satz 6.79 gelten auch für komplexe Argumente.

**Frage:** Man bestimme  $\sup_{z \in \mathbb{C}} |\cos(z)|$  und  $\sup_{z \in \mathbb{C}} |\sin(z)|$ .

**Satz 6.83.** Die einzigen Nullstellen der komplexen Winkelfunktionen sind  $z_{\sin, n} = n\pi$  für den Sinus und  $z_{\cos, n} = \frac{\pi}{2} + n\pi$  für den Cosinus, wobei  $n \in \mathbb{Z}$ .

Wenn  $\exp(iz) = 1$ , dann ist  $z = 2n\pi$  für ein geeignetes  $n \in \mathbb{Z}$ .

*Beweis.* Sei  $z = x + iy$  mit  $x, y \in \mathbb{R}$ . Für den Cosinus haben wir zum Beispiel

$$\begin{aligned} \Re(\cos z) &= \frac{1}{2} \Re(\exp(iz) + \exp(-iz)) = \frac{1}{2} \Re(e^{ix} e^{-y} + e^{-ix} e^y) \\ &= \frac{1}{2} (e^{-y} \cos x + e^y \cos x) = \frac{1}{2} \cos x (e^{-y} + e^y). \end{aligned}$$

Die Aussage über die Nullstellen der Sinusfunktion bekommt man auf ähnlichem Wege.

Wenn nun  $\exp(iz) = 1$ , dann ist auch  $\exp(-iz) = 1$ , und somit ist  $\sin(z) = 0$ , also  $z = k\pi$  für ein  $k \in \mathbb{Z}$ . Die ungeraden  $k$  kommen nicht in Frage (warum?), also bleiben nur die  $z = 2n\pi$  übrig, und für diese gilt  $\exp(iz) = 1$  tatsächlich.  $\square$

**Definition 6.84.** Für  $z \in \mathbb{C}$  mit  $z \notin \frac{\pi}{2} + \pi\mathbb{Z}$  definieren wir die Tangensfunktion als

$$\tan z = \frac{\sin z}{\cos z}.$$

**Satz 6.85.** Die Tangensfunktion besitzt die folgenden Eigenschaften (überall dort, wo sie definiert ist):

$$\begin{aligned} \tan(z + \pi) &= \tan z, \\ \tan(-z) &= -\tan z, \\ \tan(z + w) &= \frac{\tan z + \tan w}{1 - \tan z \tan w}, \\ \tan'(z) &= 1 + \tan^2 z = \frac{1}{\cos^2 z}. \end{aligned}$$

*Beweis.* Diese Eigenschaften ergeben sich direkt aus den analogen Eigenschaften von sin und cos.  $\square$

Durch Division der Potenzreihen von sin und cos bekommt man die folgende Potenzreihe des Tangens im Ursprung mit Konvergenzradius  $\frac{\pi}{2}$ . Weitere Terme der Potenzreihe ergeben sich durch Rechnen.

$$\begin{aligned} \tan z &= z + \frac{1}{3}z^3 + \frac{2}{15}z^5 + \frac{17}{315}z^7 + \frac{62}{2835}z^9 + \frac{1382}{155925}z^{11} + \frac{21844}{6081075}z^{13} + \frac{929569}{638512875}z^{15} \\ &\quad + \mathcal{O}(z^{17}), \quad z \rightarrow 0. \end{aligned}$$

**Satz 6.86 (Arcusfunktionen).**

1. Die Sinusfunktion bildet das Intervall  $[-\frac{\pi}{2}, \frac{\pi}{2}]$  bijektiv auf  $[-1, 1]$  ab. Die Umkehrfunktion

$$\arcsin: [-1, 1] \rightarrow \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$$

heißt Arcussinus und hat die Ableitung  $(\arcsin y)' = \frac{1}{\sqrt{1-y^2}}$ .

2. Die Cosinusfunktion bildet das Intervall  $[0, \pi]$  bijektiv auf  $[-1, 1]$  ab. Die Umkehrfunktion

$$\arccos: [-1, 1] \rightarrow [0, \pi]$$

heißt Arcuscosinus und hat die Ableitung  $(\arccos y)' = -\frac{1}{\sqrt{1-y^2}}$ .



3. Die Tangensfunktion bildet das Intervall  $(-\frac{\pi}{2}, \frac{\pi}{2})$  bijektiv auf  $\mathbb{R}$  ab. Die Umkehrfunktion

$$\arctan: \mathbb{R} \rightarrow \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$$

heißt Arcustangens und hat die Ableitung  $(\arctan y)' = \frac{1}{1+y^2}$ .

*Beweis.* Das schaffen Sie alleine. □

**Frage:** Wie lassen sich die Ableitungsformeln für die Arcusfunktionen und die Formel für die Allgemeine binomische Reihe (siehe (6.5)) so kombinieren, daß mit sehr geringem Arbeitsaufwand Potenzreihen für die Arcusfunktionen erhalten werden können? Insbesondere sollten Sie folgende Formel gefunden haben:

$$\arctan x = \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} x^{2n+1}, \quad -1 \leq x \leq 1.$$

Wir beschließen den Abschnitt mit einigen kulturell-historischen Betrachtungen. Wenn wir in der Arcustangens-Potenzreihe  $x = 1$  einsetzen, dann bekommen wir die erstaunliche Formel

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} \mp \dots,$$

die rein theoretisch verwendet werden könnte, um  $\pi$  auszurechnen; aber in Wirklichkeit ist sie weitgehend unbrauchbar, denn man benötigt mindestens 10.000 Summanden, um  $\pi$  auf 4 Stellen nach dem Komma zu ermitteln. Diese Formel wurde in Kontinentaluropa 1682 gefunden von Leibniz, aber sie war im 14. Jahrhundert schon in Indien bekannt.

Bemerkenswert ist, daß man nach einigen Transformationen (elementar, aber nur mit Gespür zu finden) deutlich schneller konvergierende Reihen angeben kann, die tatsächlich für die Bestimmung von  $\pi$  genutzt wurden. Wir folgen hierbei den Darlegungen von JOHN MACHIN aus dem Jahr 1706. Und zwar haben wir

$$\frac{\tan x + \tan y}{1 - \tan x \cdot \tan y} = \tan(x + y),$$

also auch (wenn wir  $t = \tan x$  und  $s = \tan y$  setzen)

$$\arctan\left(\frac{t+s}{1-ts}\right) = x+y = \arctan t + \arctan s.$$

Sei  $a \notin \{1, 1/2\}$  eine positive Zahl, und seien  $t := \frac{a}{a-1}$  sowie  $s := \frac{-1}{2a-1}$ . Dann haben wir<sup>37</sup>

$$\begin{aligned} t+s &= 1 + \frac{1}{a-1} - \frac{1}{2a-1} = 1 + \frac{(2a-1) - (a-1)}{(a-1)(2a-1)} = 1 + \frac{a}{(a-1)(2a-1)} \\ &= 1 - \frac{a}{a-1} \cdot \frac{-1}{2a-1} = 1 - ts, \end{aligned}$$

also bekommen wir dann

$$\frac{\pi}{4} = \arctan\left(\frac{a}{a-1}\right) - \arctan\left(\frac{1}{2a-1}\right).$$

Jetzt wählen wir  $a$  clever, nämlich  $a = 120$ , und es ergibt sich dann

$$\frac{\pi}{4} = \arctan \frac{120}{119} - \arctan \frac{1}{239}.$$

Den ersten Arcustangens können wir deutlich verschönern. Es ist nämlich

$$\frac{120}{119} = \frac{120}{144-25} = \frac{2 \cdot 5 \cdot 12}{12 \cdot 12 - 5 \cdot 5} = \frac{2 \cdot \frac{5}{12}}{1 - \frac{5}{12} \cdot \frac{5}{12}} = \frac{\frac{5}{12} + \frac{5}{12}}{1 - \frac{5}{12} \cdot \frac{5}{12}},$$

also entsteht

$$\arctan \frac{120}{119} = \arctan \frac{5}{12} + \arctan \frac{5}{12} = 2 \arctan \frac{5}{12}.$$

<sup>37</sup>dies ist eine sehr schöne Übung darin, Klammern möglichst **nicht** auszumultiplizieren

Das war so entzückend, daß wir diesen Umformungstrick nochmal wiederholen:

$$\frac{5}{12} = \frac{10}{24} = \frac{2 \cdot 1 \cdot 5}{5 \cdot 5 - 1 \cdot 1} = \frac{2 \cdot \frac{1}{5}}{1 - \frac{1}{5} \cdot \frac{1}{5}} = \frac{\frac{1}{5} + \frac{1}{5}}{1 - \frac{1}{5} \cdot \frac{1}{5}},$$

und das liefert uns dann

$$\arctan \frac{5}{12} = \arctan \frac{1}{5} + \arctan \frac{1}{5} = 2 \arctan \frac{1}{5},$$

und insgesamt entsteht somit die Formel von JOHN MACHIN:

$$\frac{\pi}{4} = 4 \arctan \frac{1}{5} - \arctan \frac{1}{239}.$$

Der große Vorteil dieser Darstellung gegenüber der Leibniz-Reihe aus dem Jahr 1682 ist jetzt, daß die Argumente der Arcustangensfunktionen deutlich näher an der Null sind, sodaß die Potenzreihen jetzt wesentlich schneller konvergieren. Jeder neue Summand ist offenkundig mindestens 25 mal kleiner als sein Vorgänger, und tatsächlich hat Machin im Jahr 1706 mit dieser Formel die Zahl  $\pi$  ohne technische Hilfsmittel auf 100 Dezimalstellen berechnet.

### 6.5.3 Die Hyperbelfunktionen

**Definition 6.87.** *Die Funktionen*

$$\begin{aligned} \sinh: \mathbb{C} &\rightarrow \mathbb{C}, & \sinh: z &\mapsto \sinh(z) = \frac{1}{2}(\exp(z) - \exp(-z)), \\ \cosh: \mathbb{C} &\rightarrow \mathbb{C}, & \cosh: z &\mapsto \cosh(z) = \frac{1}{2}(\exp(z) + \exp(-z)), \\ \tanh: \mathbb{C} \setminus \left(\frac{\pi i}{2} + \pi i \mathbb{Z}\right) &\rightarrow \mathbb{C}, & \tanh: z &\mapsto \tanh(z) = \frac{\sinh(z)}{\cosh(z)} = \frac{\exp(z) - \exp(-z)}{\exp(z) + \exp(-z)}, \end{aligned}$$

heißen Sinus hyperbolicus, Cosinus hyperbolicus, Tangens hyperbolicus<sup>38</sup>.

Der Name kommt von der Formel  $\cosh^2(x) - \sinh^2(x) = 1$  für  $x \in \mathbb{R}$ , die besagt, daß die Punkte der Form  $(\cosh x, \sinh x)$  auf dem Ast einer Hyperbel liegen.

Nach diesen Definitionen untersuchen wir die üblichen Eigenschaften

- Additionstheoreme,
- Symmetrieeigenschaften (z.B. Invarianz der Funktionen unter der Spiegelung  $x \mapsto -x$  oder unter einer Verschiebung  $x \mapsto x + p$ ),
- Ableitungen,
- Potenzreihendarstellungen,
- Umkehrfunktionen (aber nur in  $\mathbb{R}$ , nicht in  $\mathbb{C}$ !).

**Satz 6.88.** *Im Definitionsbereich dieser Funktionen gelten die folgenden Eigenschaften:*

$$\begin{aligned} \sinh(z+w) &= \sinh z \cosh w + \cosh z \sinh w, & \cosh(-z) &= \cosh z, & \tanh(-z) &= -\tanh z, \\ \cosh(z+w) &= \cosh z \cosh w + \sinh z \sinh w, & \sinh'(z) &= \cosh z, & \tanh'(z) &= \frac{1}{\cosh^2(z)}, \\ \tanh(z+w) &= \frac{\tanh z + \tanh w}{1 + \tanh z \tanh w}, & \sinh(z+2\pi i) &= \sinh z, & \cosh(z+2\pi i) &= \cosh z, \\ \sinh(-z) &= -\sinh z, & \sinh z &= -i \sin(iz), & \cosh z &= \cos(iz), \\ \sinh'(z) &= \cosh z, & \sinh z &= -i \sin(iz), & \tanh(z+\pi i) &= \tanh z, \\ \sinh(z+2\pi i) &= \sinh z, & \cosh z &= \cos(iz), & \tanh z &= -i \tan(iz), \\ \sinh z &= \sum_{k=0}^{\infty} \frac{z^{2k+1}}{(2k+1)!}, & \cosh z &= \sum_{k=0}^{\infty} \frac{z^{2k}}{(2k)!}. \end{aligned}$$

<sup>38</sup>hyperbolic sine, hyperbolic cosine, hyperbolic tangent

*Beweis.* Eine wunderbare Gelegenheit, den Umgang mit elementaren Funktionen zu üben ! □

Für den Tangens hyperbolicus bekommen wir auf bekanntem Wege die Reihendarstellung

$$\tanh z = z - \frac{1}{3}z^3 + \frac{2}{15}z^5 - \frac{17}{315}z^7 + \frac{62}{2835}z^9 - \frac{1382}{155925}z^{11} + \frac{21844}{6081075}z^{13} - \frac{929569}{638512875}z^{15} + \mathcal{O}(z^{17}), \quad z \rightarrow 0.$$

Der Konvergenzradius ist  $\frac{\pi}{2}$ .

**Satz 6.89 (Areafunktionen).** Die Funktion  $y = \sinh(x)$  bildet das Intervall  $\mathbb{R} = (-\infty, \infty)$  bijektiv auf sich ab und hat dort die Umkehrfunktion

$$x = \operatorname{Arsinh}(y) = \ln \left( y + \sqrt{y^2 + 1} \right).$$

Die Funktion  $y = \cosh(x)$  bildet  $[0, \infty)$  bijektiv auf  $[1, \infty)$  ab und hat dort die Umkehrfunktion

$$x = \operatorname{Arcosh}(y) = \ln \left( y + \sqrt{y^2 - 1} \right).$$

Die Funktion  $y = \tanh(x)$  bildet  $\mathbb{R} = (-\infty, \infty)$  bijektiv auf  $(-1, 1)$  ab und hat dort die Umkehrfunktion

$$x = \operatorname{Artanh}(y) = \frac{1}{2} \ln \frac{1+y}{1-y}.$$

Diese Funktionen heißen Area sinus hyperbolicus, Area cosinus hyperbolicus, Area tangens hyperbolicus<sup>39</sup>.

*Beweis.* Übungsaufgabe. □

Die Namen cosh und sinh erklären sich wie folgt.

Am Einheitskreis (beschrieben durch die Gleichung  $x^2 + y^2 = 1$ ) betrachten wir den Sektor mit den Ecken  $(1, 0)$  und  $(x, y)$ . Dann gilt: falls dieser Sektor den Flächeninhalt  $\phi/2$  besitzt, so ist  $(x, y) = (\cos \phi, \sin \phi)$ . Siehe Abbildung 6.3.

An der Einheitshyperbel (beschrieben durch die Gleichung  $x^2 - y^2 = 1$ ) betrachten wir den Sektor mit den Ecken  $(1, 0)$  und  $(x, y)$ . Dann gilt: falls dieser Sektor den Flächeninhalt  $\phi/2$  besitzt, so ist  $(x, y) = (\cosh \phi, \sinh \phi)$ . Siehe Abbildung 6.4.

### 6.5.4 Wurzeln aus komplexen Zahlen

In diesem Abschnitt ziehen wir Wurzeln aus komplexen Zahlen. Diese bilden allerdings **keine Wurzelfunktion**, denn die Bedingung der Eindeutigkeit ist verletzt.

**Satz 6.90.** Sei  $n \in \mathbb{N}$ .

1. Dann gibt es genau  $n$  verschiedene komplexe Zahlen  $\zeta_0, \dots, \zeta_{n-1}$  als Lösungen der Gleichung

$$z^n = 1.$$

Dies sind die Zahlen

$$\zeta_j = \exp \left( 2\pi i \frac{j}{n} \right), \quad j = 0, 1, \dots, n-1.$$

2. Sei  $w = |w| \exp(i\varphi) \in \mathbb{C}$ . Dann gibt es genau  $n$  verschiedene komplexe Zahlen  $z_0, \dots, z_{n-1}$  als Lösungen der Gleichung

$$z^n = w.$$

Diese werden gegeben durch  $z_j = z_0 \zeta_j$ ,  $j = 0, 1, \dots, n-1$ , wobei

$$z_0 = \sqrt[n]{|w|} \exp \left( i \frac{\varphi}{n} \right).$$

Hierbei bedeutet  $\sqrt[n]{|w|}$  die gewöhnliche nichtnegative Wurzel aus einer nichtnegativen reellen Zahl.

---

<sup>39</sup> Nicht Arcus ...

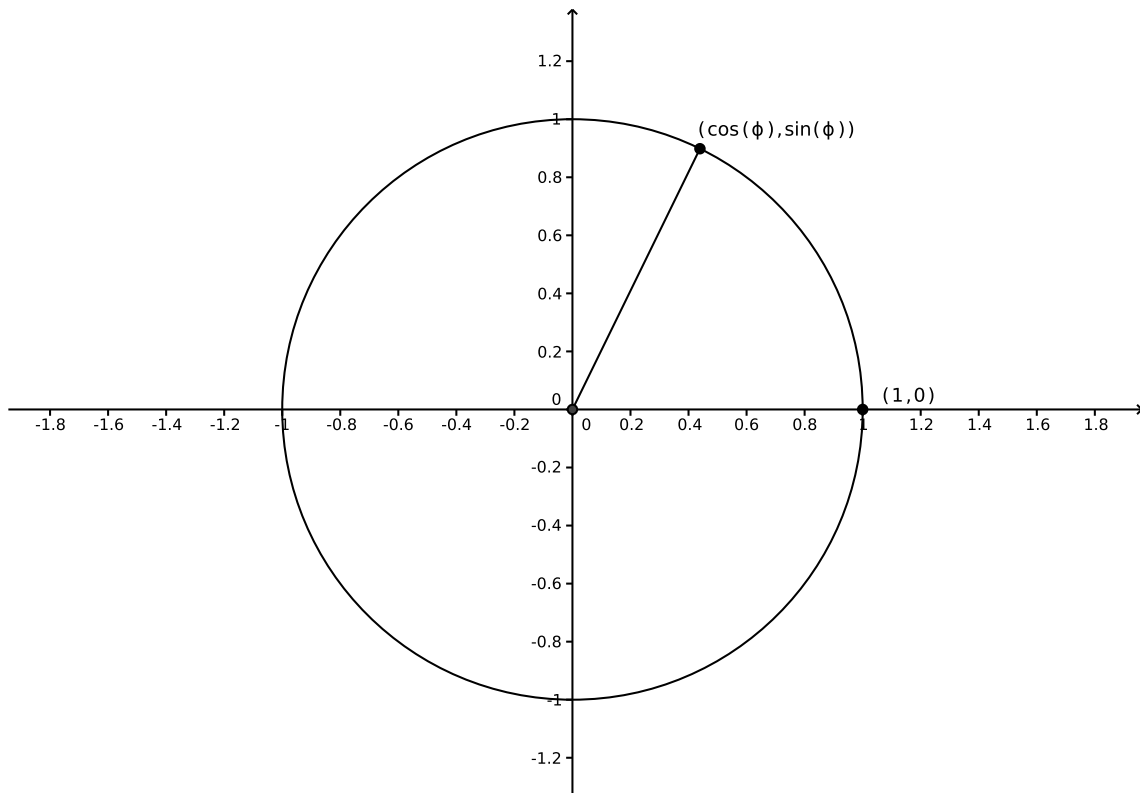


Abbildung 6.3: Ein Sektor am Einheitskreis

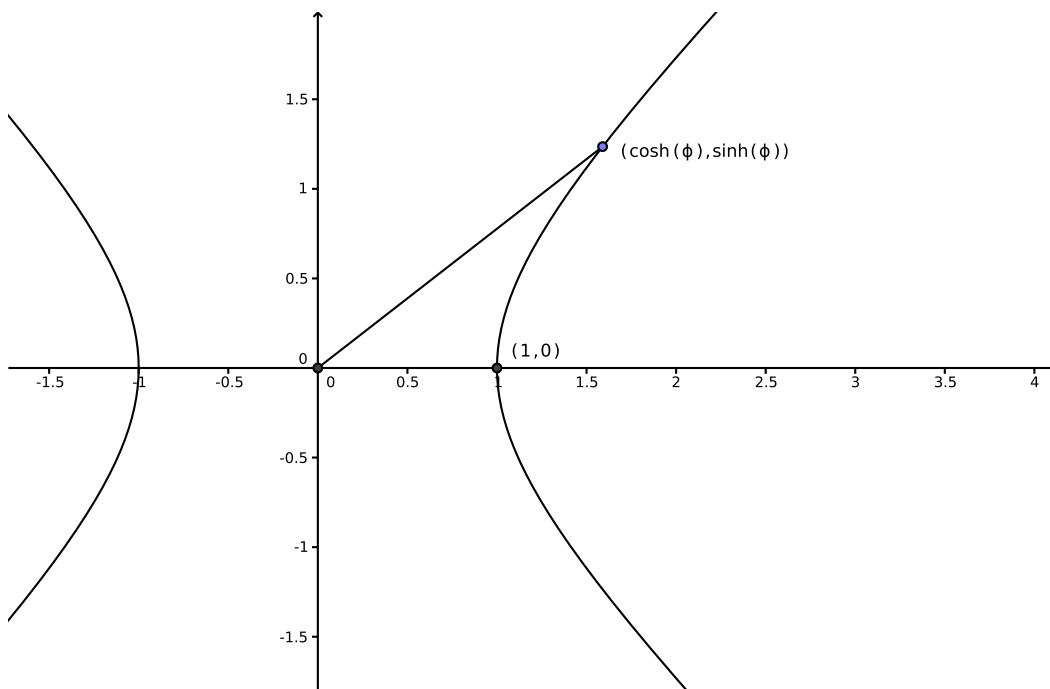


Abbildung 6.4: Ein Sektor an der Einheitshyperbel

*Beweis.* 1. Sei  $z^n = 1$ . Wir können  $z$  schreiben als  $z = |z| \exp(i\varphi)$  mit  $\varphi \in \mathbb{R}$ . Dann ist

$$1 = z^n = |z|^n \exp(in\varphi),$$

also  $|z| = 1$  und somit  $\exp(in\varphi) = 1$ .

Wir wissen aus Satz 6.83: wenn  $\exp(i\psi) = 1$ , dann ist  $\psi = 2j\pi$  für ein  $j \in \mathbb{Z}$ . Die Relation  $n\varphi = 2j\pi$  führt gerade auf die Darstellung der  $\zeta_j$ .

2. Man rechnet nach, daß diese  $z_j$  tatsächlich Lösungen sind. Sei nun  $z^*$  eine weitere Lösung von  $z^n = w$ . Dann ist der Quotient  $\frac{z^*}{z_0}$  aber eine Lösung von  $z^n = 1$ . Und diese Gleichung hat lediglich die Lösungen  $\zeta_0, \dots, \zeta_{n-1}$ . Also kann es außer den  $z_0, \dots, z_{n-1}$  keine weiteren Lösungen der Gleichung  $z^n = w$  geben. □

Die Zahlen  $\zeta_0, \dots, \zeta_{n-1}$  heißen *n-te Einheitswurzeln*, und die Zahl Eins wird auch gern als *Einheit* bezeichnet.

## 6.6 Verfahren zur numerischen Lösung nichtlinearer Gleichungen

Lineare Gleichungen und Systeme von linearen Gleichungen können mittels Standardmethoden gelöst werden (und zwar *exakt*); und die Theorie der Matrizen stellt viele Hilfsmittel bereit, die Lösungen näher zu beschreiben.

Demgegenüber ist die Situation bei nichtlinearen Gleichungen viel schwieriger:

- Man kann manchmal nur hoffen, daß es überhaupt Lösungen gibt.
- Oft weiß man nicht, wieviele Lösungen es sind.
- Es ist keineswegs einfach, diese Lösungen dann auch noch zu finden.
- Und schließlich kann man nur in wenigen Fällen die Lösungen exakt ermitteln; meist muß man sich mit Näherungswerten begnügen.

**Beispiel 6.91.** *Wir versuchen die Gleichungen*

$$x = \cos(x),$$

$$y = e^y - 2$$

zu lösen, wobei  $x, y > 0$ . Wir erhalten die Wertetabellen

$x$	$\cos(x)$
0	1
$\pi/2$	0

$y$	$e^y - 2$
0	-1
1	0.718
2	5.39

Das führt uns auf die Idee einer Iteration,

$$x_0 := 0, \quad x_{n+1} := \cos(x_n), \quad \text{bzw.} \quad y_0 := 1.5, \quad y_{n+1} := \exp(y_n) - 2.$$

Damit erhalten wir dann

$n$	$x_n$
0	0
1	1
2	0.54
3	0.86
4	0.65
5	0.79
6	0.70
7	0.76
8	0.72

*bzw.*

$n$	$y_n$
0	1.5
1	2.48
2	9.96
3	21191.5
4	$E$

Im ersten Fall strebt die Folge gegen 0.739085133, im zweiten Fall gegen  $\infty$ . Wir unternehmen einen zweiten Versuch für den mißglückten zweiten Fall:  $y + 2 = e^y$ , also sei jetzt

$$y_0 := 1.5, \quad y_{n+1} := \ln(y_n + 2).$$

Das führt uns auf

$n$	$y_n$
0	1.5
1	1.25
2	1.18
3	1.157
4	1.149

mit Grenzwert 1.146193. Dieser ist eine Lösung zu  $y = e^y - 2$ .

Warum versagt das eine Verfahren, aber das andere nicht ?

### 6.6.1 Das Halbierungsverfahren

Zunächst erinnern wir an ein einfaches Verfahren, das in solchen Situationen *immer* funktioniert. Wir hatten es schon beim Beweis des Zwischenwertsatzes benutzt.

Wir suchen zum Beispiel die Nullstelle von  $f = f(y) = e^y - 2 - y$ . Dazu starten wir von zwei Werten  $y_0$  und  $y_1$ , für die  $f(y_0)$  und  $f(y_1)$  verschiedenes Vorzeichen haben. Dazwischen muß eine Nullstelle von  $f$  liegen, denn  $f$  ist stetig. Als nächstes probieren wir  $y_2 := \frac{1}{2}(y_0 + y_1)$ , und so weiter:

$n$	$y_n$	$f(y_n)$
0	1	-0.28
1	2	3.39
2	1.5	0.98
3	1.25	0.24
4	1.125	-0.045
5	1.1875	...

Die Fehlerschranke für die Nullstelle verbessert sich mit jedem Schritt um den Faktor  $\frac{1}{2}$ . Das Verfahren konvergiert langsam, dafür aber auf jeden Fall.

### 6.6.2 Funktionaliteration und der Banachsche Fixpunktsatz

**Definition 6.92.** Sei  $F$  eine Abbildung einer Menge  $M$  in sich. Wir sagen, daß  $x^* \in M$  ein Fixpunkt<sup>40</sup> von  $F$  ist, wenn  $F(x^*) = x^*$ .

**Satz 6.93 (Banachscher Fixpunktsatz<sup>41</sup>).** Sei  $M \subset U$  abgeschlossen,  $U$  ein Banachraum, und  $F: M \rightarrow M$  mit  $\|F(x) - F(y)\|_U \leq \alpha \|x - y\|_U$  für alle  $x, y \in M$ , wobei  $\alpha < 1$ .

Dann hat  $F$  genau einen Fixpunkt in  $M$ ,  $F(x^*) = x^*$ , der gefunden werden kann durch den Algorithmus

$$\begin{aligned} x_0 &\in M \text{ beliebig,} \\ x_{n+1} &:= F(x_n), \quad n \in \mathbb{N}_0. \end{aligned}$$

Wir haben die Fehlerschätzung

$$\|x^* - x_n\|_U \leq \frac{\alpha^n}{1 - \alpha} \|x_1 - x_0\|_U. \quad (6.8)$$

*Beweis.* Zunächst haben wir

$$\begin{aligned} \|x_2 - x_1\|_U &= \|F(x_1) - F(x_0)\|_U \leq \alpha \|x_1 - x_0\|_U, \\ \|x_3 - x_2\|_U &= \|F(x_2) - F(x_1)\|_U \leq \alpha \|x_2 - x_1\|_U \leq \alpha^2 \|x_1 - x_0\|_U, \\ &\vdots \\ \|x_{n+1} - x_n\|_U &\leq \alpha^n \|x_1 - x_0\|_U. \end{aligned}$$

<sup>40</sup>fixed point

<sup>41</sup>Banach fixed point theorem

Sei nun  $n < m$ . Dann ergibt sich

$$\begin{aligned} \|x_m - x_n\|_U &= \|(x_m - x_{m-1}) + (x_{m-1} - x_{m-2}) + \cdots + (x_{n+1} - x_n)\|_U \\ &\leq \|x_m - x_{m-1}\|_U + \|x_{m-1} - x_{m-2}\|_U + \cdots + \|x_{n+1} - x_n\|_U \\ &\leq (\alpha^{m-1} + \alpha^{m-2} + \cdots + \alpha^n) \|x_1 - x_0\|_U \\ &\leq \sum_{k=n}^{\infty} \alpha^k \|x_1 - x_0\|_U \\ &= \frac{\alpha^n}{1 - \alpha} \|x_1 - x_0\|_U \rightarrow 0, \end{aligned}$$

falls  $n \rightarrow \infty$ . Also ist  $(x_n)_{n \in \mathbb{N}}$  eine Cauchyfolge. Weil der Raum  $U$  als Banachraum keine Löcher hat, besitzt diese Folge einen Grenzwert  $x^*$ ,  $\lim_{n \rightarrow \infty} \|x_n - x^*\|_U = 0$ . Und weil  $M$  abgeschlossen ist, liegt  $x^*$  in  $M$ .

Wenn wir in der eben gezeigten Ungleichung

$$\|x_m - x_n\|_U \leq \frac{\alpha^n}{1 - \alpha} \|x_1 - x_0\|_U, \quad m > n,$$

den Index  $m$  nach Unendlich schicken und Satz 5.24 clever einsetzen, dann bekommen wir genau die Fehlerschätzung (6.8).

Dieser Grenzwert  $x^*$  ist Fixpunkt der Abbildung  $F$ , denn

$$x^* = \lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} x_{n+1} = \lim_{n \rightarrow \infty} F(x_n) = F(\lim_{n \rightarrow \infty} x_n) = F(x^*).$$

Hierbei haben wir im vorletzten Schritt die Stetigkeit von  $F$  (als kommutatives Diagramm) benutzt. Und die Stetigkeit von  $F$  wiederum folgt aus Satz 6.18 (Teil 4) mit  $\delta_0(\varepsilon) = \varepsilon/\alpha$ , in Verbindung mit Bemerkung 6.20.

Es kann keinen zweiten Fixpunkt geben, denn falls  $x^{**} = F(x^{**})$ , dann gilt

$$\|x^{**} - x^*\|_U = \|F(x^{**}) - F(x^*)\|_U \leq \alpha \|x^{**} - x^*\|_U,$$

was nur für  $x^* = x^{**}$  möglich ist.

Damit ist der Banachsche Fixpunktsatz bewiesen.  $\square$

**Beispiel 6.94.** Sei  $F: [a, b] \rightarrow [a, b]$ , und  $|F'(x)| \leq \alpha < 1$  auf  $[a, b]$ . Dann gibt es genau ein  $x \in [a, b]$  mit  $F(x) = x$ .

Warum ist das so ?

**Beispiel 6.95.** Bezogen auf das Beispiel der Gleichung  $x = \cos(x)$  heißt das:  $[a, b] = [0.1, 1]$ ,  $F(x) = \cos(x)$ , also  $|F'(x)| = |\sin(x)|$ . Auf dem Intervall  $[0.1, 1]$  ist die Sinusfunktion echt kleiner als 1, also können wir argumentieren wie eben. (Die Wahl der Intervallgrenzen 0.1 und 1 hat sich dabei nach Probieren als zweckmäßig herausgestellt.)

**Frage:** Warum kam es für die Iteration  $y_{n+1} := \exp(y_n) - 2$  im obigen Beispiel zur Divergenz ?

### Ein Beispiel aus der Physik

Wir betrachten ein Elektron in einem Potentialtopf. Die Ortsvariable  $x$  ist aus dem  $\mathbb{R}^1$ , und das Potential  $V$  ist

$$V(x) = \begin{cases} V_0 & : -a < x < a, \\ 0 & : |x| \geq a, \end{cases}$$

wobei  $a > 0$  und  $V_0 < 0$  gegeben sind. Wenn  $E$  der Energiewert des Elektrons ist und  $m$  seine Masse, dann erfüllt die Wellenfunktion  $\varphi: \mathbb{R} \rightarrow \mathbb{C}$  des Elektrons im stationären Fall die Differentialgleichung

$$\left( -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \right) \varphi(x) = E\varphi(x), \quad x \in \mathbb{R}, \quad \int_{x=-\infty}^{\infty} |\varphi(x)|^2 dx = 1.$$

Wir betrachten nur gebundene Zustände, also ist  $V_0 < E < 0$ . Zuerst skalieren wir die Konstanten weg, also führen wir reelle Parameter  $\kappa$ ,  $\kappa_0$  und  $K$  ein mit

$$E =: -\frac{\hbar^2}{2m} \kappa^2, \quad V_0 =: -\frac{\hbar^2}{2m} \kappa_0^2, \quad E - V_0 =: \frac{\hbar^2}{2m} K^2,$$

und es ergeben sich dann die Differentialgleichungen

$$\begin{aligned} \left(\frac{d^2}{dx^2} - \kappa^2\right) \varphi(x) &= 0, & -\infty < x < -a, \\ \left(\frac{d^2}{dx^2} + K^2\right) \varphi(x) &= 0, & -a < x < a, \\ \left(\frac{d^2}{dx^2} - \kappa^2\right) \varphi(x) &= 0, & a < x < +\infty. \end{aligned}$$

Die Parameter erfüllen die Nebenbedingung  $K^2 = \kappa_0^2 - \kappa^2$ , und  $\kappa_0$  ist gegeben, während  $K$  gesucht ist.

Die Lösungen der Differentialgleichungen sind

$$\varphi(x) = \begin{cases} a_1 e^{\kappa x} + a_2 e^{-\kappa x} & : -\infty < x < -a, \\ b_1 e^{iKx} + b_2 e^{-iKx} & : -a < x < a, \\ c_1 e^{\kappa x} + c_2 e^{-\kappa x} & : a < x < \infty. \end{cases}$$

Die Bedingung  $\int_{x=-\infty}^{\infty} |\varphi(x)|^2 dx = 1$  erzwingt dann  $c_1 = 0$  und  $a_2 = 0$ . Es bleiben also noch vier Parameter  $a_1, b_1, b_2$  und  $c_2$  übrig. An den Übergangsstellen  $x = \pm a$  soll  $\varphi$  stetig sein und  $\varphi'$  ebenfalls. Dies ergibt vier Gleichungen für die vier Parameter. Dieses Gleichungssystem ist linear und homogen. Damit wir allerdings eine physikalisch relevante Lösung bekommen, muß der Rang der Matrix weniger als vier sein (ansonsten hätten wir nur die Lösung  $a_1 = b_1 = b_2 = c_2 = 0$ , die keinen physikalischen Zustand beschreibt).

Aus physikalischen Überlegungen ergibt sich: weil  $V$  symmetrisch ist (denn  $V(-x) = V(x)$ ), ist die Wellenfunktion  $\varphi$  entweder symmetrisch (also  $\varphi(-x) = \varphi(x)$ ) oder antisymmetrisch (also  $\varphi(-x) = -\varphi(x)$ ).

Im symmetrischen Fall ist  $b_1 = b_2$  und  $a_1 = c_2$ , und wir erhalten (nach einiger Rechnung im Selbststudium) aus der Rangbedingung an die Matrix die notwendige Bedingung

$$\tan(Ka) = \frac{\kappa}{K}.$$

Und im antisymmetrischen Fall ist  $b_1 = -b_2$  und  $a_1 = -c_2$ , und die Rangbedingung an die Matrix liefert uns jetzt

$$\tan(Ka) = -\frac{K}{\kappa}.$$

Wir sehen  $K$  als gesucht an und werfen jetzt  $\kappa$  heraus: es ist  $0 \leq \kappa \leq \kappa_0$  und  $\kappa = \sqrt{\kappa_0^2 - K^2}$ , also sind folgende hochgradig nichtlineare Gleichungen nach  $K$  aufzulösen:

$$\tan(Ka) \stackrel{!}{=} \frac{\sqrt{\kappa_0^2 - K^2}}{K}, \quad \tan(Ka) \stackrel{!}{=} \frac{-K}{\sqrt{\kappa_0^2 - K^2}}.$$

Eine Arbeitsstrategie könnte die folgende sein:

- man plote jeweils die linke Gleichungsseite als Funktion von  $K$ , und die rechte Gleichungsseite auch,
- man bestimme geometrisch die Anzahl der Kreuzungspunkte der Graphen (diese Anzahl gibt uns an, wieviele Energieniveaus in diesem Potentialtopf möglich sind),
- die tatsächlichen Werte für  $K$  könnte man mit dem Halbierungsverfahren bestimmen, oder mit einer behutsam eingesetzten Funktionaliteration. Aus diesen Werten für  $K$  ermitteln wir dann die Energieniveaus, die ein Elektron haben kann. Genau an diesen Energieniveaus waren wir interessiert.

### 6.6.3 Das Newtonverfahren

**Definition 6.96.** Wir sagen, daß eine Folge  $(x_n)_{n \in \mathbb{N}}$  von Elementen eines Banachraums linear gegen einen Grenzwert  $x^*$  konvergiert, wenn

$$\|x_{n+1} - x^*\|_U \leq \alpha \|x_n - x^*\|_U$$



gilt für alle  $n$ , mit einem gewissen  $\alpha < 1$ , das nicht von  $n$  abhängt.

Wir sagen, daß die Folge  $(x_n)_{n \in \mathbb{N}}$  quadratisch konvergiert, wenn

$$\|x_{n+1} - x^*\|_U \leq C \|x_n - x^*\|_U^2$$

gilt, falls  $x_n$  nahe bei  $x^*$  ist. Hierbei ist  $C$  eine positive Konstante, die nicht von  $n$  abhängt, aber auch größer als eins sein darf.

Das Newtonverfahren ist ein Beispiel für ein quadratisch konvergentes Iterationsverfahren.

Noch ein Wort zur Schreibweise: für Abbildungen mit gesuchtem Fixpunkt  $x^*$  schreiben wir Groß  $F$ , und für Funktionen mit gesuchter Nullstelle  $x^*$  schreiben wir klein  $f$ .

### Das Newtonverfahren im $\mathbb{R}^1$

Sei  $f: \mathbb{R} \rightarrow \mathbb{R}$  eine zweimal stetig differenzierbare Funktion mit Nullstelle  $x^*$ , und sei  $f'(x^*) \neq 0$ . Dann ist das Newtonverfahren wie folgt definiert:

$$\begin{aligned} x_0 &\text{ nahe genug an } x^*, \\ x_{n+1} &:= x_n - \frac{f(x_n)}{f'(x_n)}, \quad n = 0, 1, \dots \end{aligned}$$

Wenn die Folge der  $x_n$  konvergieren sollte gegen einen Grenzwert, dann strebt die Folge  $(x_{n+1} - x_n)_{n \in \mathbb{N}}$  der Differenzen benachbarter Folgenglieder nach 0, also geht die Folge  $(f(x_n))_{n \in \mathbb{N}}$  ebenfalls nach 0.

**Satz 6.97.** *Es gibt ein  $\delta > 0$ , sodaß für das Intervall  $M = [x^* - \delta, x^* + \delta]$  gilt:*

1. *Die Iterationsvorschrift des Newtonverfahrens bildet  $M$  in sich ab.*
2. *Das Newtonverfahren konvergiert quadratisch gegen  $x^*$ .*

*Beweis.* Zunächst ist  $f'(x^*) \neq 0$ . Wir wählen nun  $\delta$  so klein, daß im Intervall  $M$  überall  $f'(x) \neq 0$  ist. Das geht wegen Lemma 6.7 und der Stetigkeit von  $f'$ . Als nächstes basteln wir uns eine Hilfsfunktion

$$F(x) := x - \frac{f(x)}{f'(x)}, \quad x \in M.$$

Eine Division durch Null kann für  $x \in M$  nicht eintreten. Für diese Hilfsfunktion rechnet man schnell nach, daß

$$F(x^*) = x^*, \quad F'(x) = \frac{f(x)f''(x)}{(f'(x))^2}, \quad F'(x^*) = 0.$$

Wir dürfen fordern, daß  $\delta$  so klein ist, daß im Intervall  $M$  gilt:  $|F'(x)| \leq \frac{1}{2}$ , wegen der Stetigkeit von  $F'$ .

Das Newtonverfahren wird nun beschrieben durch die Vorschrift  $x_{n+1} := F(x_n)$ . Sei  $x_n \in M$ , also  $|x_n - x^*| \leq \delta$ . Der Mittelwertsatz liefert dann, daß  $|x_{n+1} - x^*| \leq \frac{1}{2}|x_n - x^*| \leq \frac{1}{2}\delta$ , also  $x_{n+1} \in M$ . Damit kann das Banach-Fixpunkt-Verfahren für die Funktion  $F$  durchgeführt werden und wir bekommen sofort die Konvergenz gegen  $\lim_{n \rightarrow \infty} x_n = x^*$  sowie eine Fehlerschranke:

$$|x_{n+1} - x_n| \leq \frac{1}{2^{n-1}} \delta.$$

Diese Schranke können wir aber gewaltig verbessern. Dazu gehen wir zurück zur Funktion  $f$ , für die wir eine Taylorentwicklung im Entwicklungspunkt  $x_n$  veranstalten:

$$0 = f(x^*) = f(x_n + (x^* - x_n)) = f(x_n) + f'(x_n)(x^* - x_n) + \frac{1}{2}f''(\xi_n)(x^* - x_n)^2.$$

Hierbei liegt  $\xi_n$  zwischen  $x_n$  und  $x^*$ ; mehr ist vom  $\xi_n$  nicht bekannt. Wir dividieren durch  $f'(x_n)$  und sortieren um:

$$\begin{aligned} 0 &= \frac{f(x_n)}{f'(x_n)} + x^* - x_n + \frac{f''(\xi_n)}{2f'(x_n)}(x^* - x_n)^2, \\ x_{n+1} - x^* &= \frac{f''(\xi_n)}{2f'(x_n)}(x^* - x_n)^2. \end{aligned}$$

Weil dieser Bruch beschränkt auf  $M$  ist, ist dies genau die gewünschte quadratische Konvergenz.  $\square$

**Beispiel 6.98 (Babylonisches Wurzelziehen).** Wir wollen die Wurzel aus 2 ziehen. Sei also  $f(x) = x^2 - 2$ , und  $x_0 = 1$ . Dann haben wir die Iterationsvorschrift

$$x_{n+1} := x_n - \frac{x_n^2 - 2}{2x_n} = \frac{1}{2} \left( x_n + \frac{2}{x_n} \right).$$

Diese Iterationsvorschrift war schon den alten Sumerern in Babylon bekannt, vor etwa 4000 Jahren. Eine entsprechende Formel gibt es auch für höhere Wurzeln; dann ist  $f(x) = x^n - a$ . Diese Formel geht auf HERON <sup>42</sup> zurück. Es ergibt sich folgende Tabelle. Man beobachte die Verdopplung der Anzahl korrekter Dezimalstellen, worin sich die quadratische Konvergenz widerspiegelt.

$n$	$x_n$	Anzahl richtiger Stellen
0	1	1
1	1.5	1
2	1.416666666666667	3
3	1.41421568627451	6
4	1.41421356237469	12
5	1.414213562373095	$\geq 16$
6	1.414213562373095	$\geq 16$

## 6.7 Schlüsselbegriffe

- Definition Grenzwert und Stetigkeit einer Funktion,
- Satz vom Maximum, Zwischenwertsatz,
- Definition von Ableitung und Landau-Symbolen,
- Produktregel, Kettenregel, Ableitung der Umkehrfunktion,
- Mittelwertsatz, Regel von Bernoulli und de l'Hospital,
- Taylorsatz und seine Beziehung zu Extremwerten,
- komplexe Winkelfunktionen, Arcusfunktionen, Hyperbelfunktionen, Areafunktionen,
- Wurzeln aus komplexen Zahlen im Gegensatz zur Wurzelfunktion,
- Banachscher Fixpunktsatz und Newtonverfahren.

---

<sup>42</sup>HERON VON ALEXANDRIA, lebte irgendwann zwischen 150 vor Christus und 250 nach Christus.

# Anhang A

## Algebraische Strukturen

Name	Operationen	Beispiele und Anmerkungen
Halbgruppe	$\circ$	Bsp: $(\mathbb{N}_0, +)$ Bsp: Evolutionsoperator eines Diffusionsprozesses Anm: $\circ$ braucht nicht kommutativ zu sein
Gruppe	$\circ$ jede Gleichung $a \circ x = b$ bzw. $y \circ a = b$ ist lösbar	Bsp: $(\mathbb{Z}, +)$ Bsp: Verschiebungsgruppe Bsp: Gruppe der Drehungen Bsp: alle invertierbaren Matrizen des $\mathbb{R}^{n \times n}$
Ring	$+$ , $-$ und $\cdot$	Bsp: $(\mathbb{Z}, +, \cdot)$ Bsp: Restklassenring bei Division durch $k \in \mathbb{Z}_{\neq 0}$ Bsp: Polynomring Bsp: Matrizenring $\mathbb{R}^{n \times n}$ Anm: $+$ ist kommutativ, aber $\cdot$ nicht unbedingt
Körper	$+$ , $-$ , $\cdot$ und $/$	Bsp: $\mathbb{Q}$ , $\mathbb{R}$ und $\mathbb{C}$ Bsp: Restklassenring bei Division durch Primzahl Bsp: gebrochen rationale Funktionen
Vektorraum	Vektor $+$ Vektor = Vektor Zahl $\cdot$ Vektor = Vektor	Bsp: $\mathbb{R}^n$ als Menge der Verschiebungsvektoren Bsp: $\mathbb{R}^n$ als Zahlenspalten Bsp: Funktionenraum $L^2([a, b] \rightarrow \mathbb{R})$ Bsp: Matrizen $\mathbb{R}^{n \times m}$ Bsp: Menge der linearen Abbildungen $\text{Hom}(U \rightarrow V)$
normierter Raum	wie Vektorraum $U$ , aber zusätzlich noch Norm $\ \cdot\  : U \rightarrow \mathbb{R}_{\geq 0}$	Bsp: $\mathbb{R}^n$ mit $\ x\ _1 =  x_1  + \dots +  x_n $ Bsp: $\mathbb{R}^n$ mit $\ x\ _2 = (x_1^2 + \dots + x_n^2)^{1/2}$ Bsp: $\mathbb{R}^n$ mit $\ x\ _\infty = \max\{ x_1 , \dots,  x_n \}$ Anm: Norm entstammt genau dann einem Skalarprodukt, wenn die Parallelogrammgleichung gilt
euklidischer/ unitärer Raum	wie Vektorraum $U$ , aber zusätzlich noch reelles / komplexes Skalarprodukt $\langle \cdot, \cdot \rangle$ , und Norm $\ u\ _U := \sqrt{\langle u, u \rangle}$	Bsp: $\mathbb{R}^n$ mit $\langle x, y \rangle = \sum_{j=1}^n x_j y_j$ Bsp: $\mathbb{C}^n$ mit $\langle x, y \rangle = \sum_{j=1}^n x_j \overline{y_j}$ Bsp: $C([a, b] \rightarrow \mathbb{R})$ mit $\langle f, g \rangle = \int_a^b f(x)g(x)dx$ Bsp: $C([a, b] \rightarrow \mathbb{C})$ mit $\langle f, g \rangle = \int_a^b f(x)\overline{g(x)}dx$
Banachraum/ vollständig. normierter Raum	wie normierter Raum	Anm: jede Cauchyfolge konvergiert Bsp: $\mathbb{R}^n$ und $\mathbb{C}^n$ mit jeder Norm Bsp: $L^2([a, b] \rightarrow \mathbb{R})$ Bsp: $C([a, b] \rightarrow \mathbb{R})$ mit $\ \cdot\ _\infty$
Hilbertraum	wie euklidischer bzw. unitärer Raum	Anm: jede Cauchyfolge konvergiert Anm: per Definition ist jeder Hilbertraum gleichzeitig euklidisch/unitär und Banachraum Bsp: $L^2([a, b] \rightarrow \mathbb{R})$

Es gibt noch weitere physikalisch relevante algebraische Strukturen, die im Skript zwar gelegentlich verwendet wurden, ohne aber definiert worden zu sein. Insbesondere sind *Algebren* interessant. Grob gesprochen, ist eine (assoziative) Algebra ein Vektorraum, der gleichzeitig ein Ring ist. Das heißt, zusätzlich zu den beiden Vektorraumoperationen „Vektor + Vektor = Vektor“ und „Zahl · Vektor = Vektor“ gibt es noch eine Ringoperation „Vektor mal Vektor = Vektor“, die in jedem Faktor linear ist und teilweise noch weitere Eigenschaften hat. Für diese weitere Multiplikation verwenden wir das Symbol  $\diamond$ .

Name	Operationen	Beispiele und Anmerkungen
Algebra	wie Vektorraum, und zusätzlich Vektor $\diamond$ Vektor = Vektor	Anm: $\diamond$ braucht weder kommutativ noch assoziativ sein, und ein neutrales Element braucht es auch nicht geben Bsp: $\mathbb{R}^3$ mit Kreuzprodukt
assoziative Algebra	wie Algebra	Anm: $\diamond$ ist jetzt assoziativ Bsp: Matrizen aus $\mathbb{R}^{n \times n}$ mit $\diamond =$ Matrizenmultipl. Bsp: $\text{Hom}(U \rightarrow U)$ mit $\diamond =$ Nacheinanderausführung Bsp: $L^1(\mathbb{R} \rightarrow \mathbb{R})$ mit $\diamond = *$ , und $*$ ist das Faltungsprodukt $(f * g)(x) = \int_{y=-\infty}^{\infty} f(x-y)g(y)dy$
LIE-Algebra	wie Algebra, schreibe $[x, y]$ statt $x \diamond y$	Anm: per Definition ist $[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0$ und $[x, x] = 0$ Bsp: $\mathbb{R}^3$ mit Kreuzprodukt Bsp: $\text{Hom}(U \rightarrow U)$ mit $[A, B] := A \circ B - B \circ A$ , hierbei ist $\circ =$ Nacheinanderausführung Bsp: Drehimpulsoperatoren der Quantenmechanik

Ein Buch, das über mehrere hundert Seiten hinweg Lie-Algebren diskutiert, ist:

**Literatur:** Greiner und Müller: *Quantenmechanik. Symmetrien*, 2005

# Index

- Abbildung
  - lineare, 65, 73
- Abel, 30
- abgeschlossene Menge, 95, 97
- Ableitung, 12, 123
- Abschluß, 95, 97
- absolut konvergent, 99
- absolute Konvergenz, 107
- Additionstheorem, 113, 138
- adjungierte Matrix, 67
- affiner Raum, 46
- ähnliche Matrizen, 83
- allgemeine Potenz, 123
- analytisch, 124
- Anfangswertabbildung, 84
- Approximationsproblem, 60
- Arcusfunktionen, 144
- Argument einer komplexen Zahl, 17
- assoziativ, 29
- Ausbreitungsgeschwindigkeit, 22
- Austauschlemma, 54
- Austauschsatz, 54
  
- Banach, 97
- Banach-scher Fixpunktsatz, 150
- Banachraum, 97
- Basis, 23, 50
  - duale, 90
  - reziproke, 90
- Basisergänzungssatz, 52
- Basissatz, 52
- Basistransformation, 81
- Bernoulli, 131
- beschränkte Menge, 95, 97
- bestimmte Divergenz, 117
- bestimmtes Integral, 13
- Betrag, 25
- Betrag einer komplexen Zahl, 16
- bijektiv, 75
- Bild, 76, 115
- Bildraum, 76
- Binomialkoeffizient, 102, 135
- binomische Reihe, 136
- Blindwiderstand, 21
- Bolzano, 104
  
- $\mathbb{C}$ , 14
- Cardano, 13
- Cauchy, 25
- Cauchy-Folge, 97
- Cauchy-Produktreihe, 108, 111
  
- cos, 144
- cosh, 146
  
- Defekt, 77
- Definitionsbereich, 115
- Delta-Distribution, 74
- Determinante, 39
- Differential, 124
- Differentialgleichungssystem, 84
- Differentialoperator, 84
- differenzierbar, 12, 19, 123
  - stetig, 123
- Dimension, 55
- Dimensionsformel für Abbildungen, 77
- Dimensionsformel für Unterräume, 56
- Dirac, 74
- direkte Summe, 57
- Divergenz, 63
- Divisionsregel, 127
- Drehmatrix, 28, 45
- Drehstreckung, 17
- Dreiecksungleichung, 58
- duale Basis, 90
- Dualraum, 90
  
- ebene Welle, 22
- Effektivwert, 21
  - komplexer, 21
- Einheitsmatrix, 29
- einseitiger Grenzwert, 117
- Element
  - inverses, 30
  - neutrales, 30
- Eliminationsmatrix, 69
- endlich erzeugt, 49
- $\varepsilon$ -Umgebung, 95, 97
- Erzeugendensystem, 49
- Euklid, 57
- euklidischer Raum, 57
- Euler, 112
- Euler-sche Zahl  $e$ , 112
- Exponentialfunktion, 18, 113
  
- Familie, 40
- Folge, 96
  - konvergente, 93, 97
- Fourierreihe, 99
- frei, 50
- Funktion, 12, 115
  - Grenzwert, 116
  - Konvergenz, 116

- Funktionale
  - lineare, 73
- Gauß, 14
- Gauß-Jordan-Form, 68
- geometrische Reihe, 99, 109
- gleichmäßige Konvergenz, 109
- Gleichungssystem, 67
- Gradient, 63
- Gram, 59
- Grenze
  - obere, 103
  - untere, 103
- Grenzwert
  - einseitiger, 117
  - uneigentlicher, 117
- Grenzwert einer Folge, 93, 97
- Grenzwert einer Funktion, 116
- Gruppe, 30
  - lineare, 34
  - orthogonale, 34
  - spezielle lineare, 34
  - spezielle orthogonale, 34
- Halbgruppe, 29
- Hamilton, 14
- harmonische Reihe, 99
- Hauptsatz der Differential- und Integralrechnung, 13
- Heisenberggruppe, 34
- Helmholtz-Projektion, 63
- Hermite, 57
- Hilbertraum, 97
- holomorph, 124
- homogenes System, 84
- Homomorphismus, 65, 73
- de l'Hospital, 131
- Hyperbelfunktionen, 146
- Häufungspunkt, 95, 97
- Imaginärteil, 14
- Induktion, 54
- Infimum, 103
- inhomogenes System, 84
- injektiv, 75, 121
- innerer Punkt, 95, 97
- Integral
  - bestimmtes, 13
  - unbestimmtes, 13
- inverse Matrix, 68
- invertierbare Matrix, 68
- isomorph, 76
- Isomorphismus, 76
- Jordan
  - Camille, 71
  - Pascual, 71
  - Wilhelm, 71
- Kern, 76
- Kettenregel, 127
- kommutatives Diagramm, 16
- kompakte Menge, 105
- Komplement
  - orthogonales, 58
- Komplementmenge, 95
- komplexe Winkelfunktionen, 144
- komplexe Zahl, 14
- Komposition, 119
- konjugiert komplexe Zahl, 16
- Kontraosition, 75
- kontravariant, 90
- konvergent, 93, 97
- Konvergenz
  - absolute, 99, 107
  - einer Reihe, 99
  - gleichmäßige, 109
- Konvergenz einer Funktion, 116
- Konvergenzradius, 109
- Koordinatensystem
  - kartesische, 22
- kovariant, 90
- Kreisfrequenz, 22
- Kreuzprodukt, 37
- Kristall, 33
- Kronecker, 41
- Körper, 11, 17
- $L^2$ , 98
- Lagrange, 38
- Landau-Symbole, 124
- Lebesgue, 98
- Leibniz, 107
- Leibniz-Kriterium, 107
- Levi-Civita-Tensor, 42
- Limes, 93, 97
- linear abhängig, 38, 40, 50
- linear unabhängig, 40, 50
- lineare Abbildung, 65, 73
- lineare Funktionale, 73
- lineare Gruppe, 34
- lineare Hülle, 49
- lineare Operatoren, 73
- Linearkombination, 49
- Linkssystem, 42
- Logarithmus, 122, 136
- Lorentz-Gruppe, 34
- Länge, 25
- Majorantenkriterium, 100
- Matrix, 27
  - adjungierte, 67
  - inverse, 29, 68
  - invertierbare, 29, 68
  - reguläre, 68
  - singuläre, 68
  - transponierte, 44, 67
- Matrix-Differentialoperator, 84
- Matrixprodukt, 28, 45, 66
- Matrizen

- ähnliche, 83
- Maximum, 103
  - lokales, 129
- Menge
  - abgeschlossene, 95, 97
  - beschränkt, 95, 97
  - kompakte, 105
  - offene, 95, 97
- Metrik-Koeffizienten, 91
- Minimum, 103
  - lokales, 129
- Minkowski-Raum, 34
- Mittelwertsatz, 130
  - Verallgemeinerter, 130
- Moivre, 17
- monoton wachsend, 103, 121
  
- Newtonverfahren, 152
- Norm, 12, 16, 25, 58, 96
- normierter Raum, 58, 96
- Nullraum, 76
  
- obere Grenze, 103
- obere Schranke, 103
- offene Menge, 95, 97
- Operatoren
  - lineare, 73
- Ordnungsrelation, 11
- Orientierung, 42
  - negative, 42
  - positive, 42
- orthogonal, 58
- orthogonale Gruppe, 34
- Orthogonalisierungsverfahren
  - von Gram-Schmidt, 59
- Orthogonalsystem, 41
- Orthonormalbasis, 58
- Orthonormalsystem, 41, 45, 58
- Ortsvektor, 22
  
- Parallelogrammgleichung, 58
- Partialsomme, 99
- Permutationsmatrix, 70
- $\pi$ , 143
- Pivotelement, 71
- Potenz
  - allgemeine, 123
- Potenzreihe, 99, 108
  - der Logarithmusfunktion, 136
  - der Winkelfunktionen, 142
  - Produkt zweier P., 111
- Produktregel, 126
- Proximum, 60
- Punkt
  - innerer, 95, 97
  
- Quotientenkriterium, 101
  
- Randpunkt, 95, 97
- Rang, 77
  
- Raum
  - affiner, 46
  - euklidischer, 57
  - normierter, 58, 96
  - unitärer, 57
  - vollständiger normierter, 97
- Realteil, 14
- Rechtssystem, 42
- reelle Zahlen, 97
- reguläre Matrix, 68
- Reihe, 99
  - binomische, 136
  - geometrische, 99, 109
  - harmonische, 99
  - Umordnung, 107
- reziproke Basis, 90
  
- Sandwichprinzip, 105
- Satz vom Maximum, 120
- Satz von Bolzano-Weierstraß, 104
- Satz von Rolle, 130
- Scheinwiderstand
  - komplexer, 21
- Schmidt, 59
- Schranke
  - obere, 103
  - untere, 103
- Schwarz, 25
- Schwingungsgleichung, 86
- sin, 144
- singuläre Matrix, 68
- sinh, 146
- Skalarprodukt, 24, 37, 57
- Spaltenrang, 83
- Span, 49
- Spatprodukt, 39
- spezielle lineare Gruppe, 34
- spezielle orthogonale Gruppe, 34
- Stammfunktion, 13
- Steinitz, 54
- stetig, 12, 19, 118
- stetig differenzierbar, 13
- streng monoton wachsend, 121
- Summe von Vektorräumen, 56
- Supremum, 103
- surjektiv, 75
- Symmetriegruppe, 33
- System
  - homogenes, 84
  - inhomogenes, 84
  
- tan, 144
- tanh, 146
- Taylor, 132
- Teilfolge, 94, 97
- Teilsumme, 99
- transponierte Matrix, 44, 67
  
- Umgebung, 95, 97
- Umkehrfunktion, 121, 128

- der Hyperbelfunktionen, 147
- der Winkelfunktionen, 144
- unbestimmtes Integral, 13
- uneigentlicher Grenzwert, 117
- Ungleichung von Cauchy-Schwarz, 25, 58
- unitärer Raum, 57
- untere Grenze, 103
- untere Schranke, 103
- Unterraum, 49
- Untervektorraum, 49
- Urbild, 115
  
- Vektorprodukt, 37
- Vektorraum, 23, 47
  - euklidischer, 57
  - unitärer, 57
- Vergleichskriterium, 100
- Verknüpfungstafel, 33
- Verschiebungsgruppe, 33
- vollständige Induktion, 54
- vollständiger normierter Raum, 97
  
- Weierstraß, 104
- Wellenfront, 22
- Wellenzahlvektor, 22, 90
- Wert einer Reihe, 99
- Wertebereich, 115
- Widerstand
  - induktiver, 21
  - kapazitiver, 21
  - Ohmscher, 21
- Winkelfunktionen, 19, 137, 141
  - komplexe, 144
- Wirkwiderstand, 21
- Wurzel
  - aus komplexer Zahl, 147
- Wurzelfunktion, 121
- Wurzelkriterium, 101
  
- Zahlen
  - reelle, 97
- Zeilenrang, 83
- Zeilenstufenform, 68, 71
- Zwischenwertsatz, 120



# Literaturhinweise

Die im Folgenden genannten Bücher sind mehrheitlich in der Lehrbuchsammlung der Bibliothek vorhanden. Die empfohlene Einteilung in Lesergruppen ist rein subjektiv vorgenommen worden, was Sie aber keineswegs davon abhalten soll, mit dem Buch zu arbeiten, mit dem Sie am Besten zurecht kommen.

## Eher für Ingenieure

Papula, *Mathematik für Ingenieure 1, 2*

de Boer, *Vektor- und Tensorrechnung für Ingenieure*

Burg, Haf, Wille, *Höhere Mathematik für Ingenieure 1, 2, 3, 4, 5*

## Eher für Physiker

Fischer, Kaul, *Mathematik für Physiker 1, 2, 3*

Jänich, *Mathematik 1, 2. Geschrieben für Physiker*, (Sehr zu empfehlen)

Jänich, *Analysis für Physiker und Ingenieure*, (Sehr zu empfehlen)

Kuscer, Kodre, *Mathematik in Physik und Technik*

Fischer, *Lineare Algebra*

Barner, Flohr, *Analysis 1, 2*

Endl, Luh, *Analysis 1, 2, 3*

Madelung, *Die mathematischen Hilfsmittel des Physikers*, (Eher historisch interessant)

## Eher für Mathematiker

Heuser, *Analysis 1, 2*

Walter, *Analysis 1, 2*

Königsberger, *Analysis 1, 2*

Hildebrandt, *Analysis 1, 2*

Blatter, *Analysis 1, 2, 3*

Koecher, *Lineare Algebra und Analytische Geometrie*

Courant, *Vorlesungen über Differential- und Integralrechnung*, (Traditioneller Zugang)