Mathematik für Physiker ... und Mathematiker

Vorlesungsskript 1999/2015

Klaus Wirthmüller

Vorwort

Das Skript ist als Textversion meiner Vorlesung “Mathematik für Physiker” in den Studienjahren 1998/1999 und 1999/2000 entstanden. Bei der Stoffauswahl und der Darstellung habe ich mich konsequent darum bemüht, die Substanz der ja ganz klassischen Themen herauszuarbeiten. Daß dabei so manches gewohnheitsmäßig zum Stoff gezählte Detail auf der Strecke geblieben ist, bedaure ich nicht: die Vorlesung ist dadurch entrümpelt.

Der Substanz den Vorrang vor den Ausschmückungen zu geben, dieses Prinzip hat sich auch bewährt, als ich später die entsprechende einjährige Grundvorlesung für Mathematikstudenten gehalten habe; daher jetzt der Zusatz im Titel. Der nicht der Physik zugetane Student mag sich zwar an der Vielzahl von physikalischen Beispielen stören, aber diese sind zumeist ohnehin nicht durch vergleichbare aus anderen Anwendungsfächern zu ersetzen.

In einem Anhang zum Skript habe ich Material zusammengestellt, das alternative Versionen des Kurses erlaubt. Der Abschnitt $15\frac{1}{2}$ besteht aus einer kurzen Darstellung des eindimensionalen Integrals für den Fall, daß man damit nicht so lange warten mag wie im Hauptteil des Skriptes. Dieser Abschnitt nimmt dann einen Teil von Abschnitt 31 vorweg. Es folgen die Abschnitte $30\frac{1}{3}$ und $30\frac{2}{3}$ zur Maßtheorie. Die Entscheidung, das lebesguesche Maß und Integral in der einführenden Vorlesung axiomatisch zu begründen und die Konstruktion einer Vorlesung über Maßtheorie zu überlassen, halte ich auch für Mathematikstudenten für richtig. Je nach äußeren Vorgaben habe ich das aber auch anders gemacht; die genannten Abschnitte ersetzen dann, entsprechend der Numerierung eingeschoben, Teile der Abschnitte 31 und 33.

Mathematikstudenten sollten sich nicht nur mit dem Satz von der lokalen Umkehrabbildung befassen, sondern auch mit dessen Beweis in Abschnitt $35\frac{1}{2}$, einem schönen Lehrbeispiel für eine schon etwas komplexere Beweisaufgabe. Noch einige Ergänzungen findet im Anhang der sehr knapp gehaltenene Abschnitt 43 über höhere Ableitungen. — Natürlich muß aus Zeitgründen jede Erweiterung durch Kürzung an anderer Stelle kompensiert werden. Eine solche Kürzungsmöglichkeit besteht darin, auf Abschnitt 44 mit seinen Morse-Punkten zu verzichten und sich mit der hausbackeneren Version 44E über lokale Extrema zu begnügen. Ein Blick auf diese Version empfiehlt sich freilich für alle, weil das bei der letzten Aktualisierung eingefügte Lemma $44\text{E}.5\frac{1}{3}$ die Bestimmung der Hesse-Form ganz einfach macht.

Zahlreiche Schreib- und andere Fehler im Skript konnte ich aufgrund von Hinweisen aufmerksamer Leser berichtigen, von denen ich Thorsten Fütterer und Martin Busley nennen möchte.

Anhang

Einleitung

Der für den beginnenden Physikstudenten wohl wichtigste und zugleich am schwersten zu erlernende Teil der Mathematik ist die sogenannte Vektoranalysis, konkret die Differential- und Integralrechnung mehrerer Veränderlicher. Nicht einer, sondern eben mehrerer Veränderlicher deshalb, weil die physikalische Welt nicht ein-, sondern zumindest dreidimensional ist. So ist es das Hauptanliegen dieser zweisemestrigen Vorlesung, eine Einführung in die Analysis mehrerer Veränderlicher zu geben.

Die typischen Denkweisen und Techniken der Analysis erlernen sich freilich leichter im Eindimensionalen, und ich beginne die Vorlesung deshalb mit der Differential- und Integralrechnung einer Veränderlichen, damit unmittelbar an Gegenstände anknüpfend, die aus der Schule vertraut sein sollten.

Es ist eine der Grundideen der Analysis, nichtlineare Objekte durch lineare zu approximieren; sie stützt sich deshalb auch auf die sogenannte lineare Algebra, die, wie der Name sagt, die Untersuchung der linearen Strukturen zum Inhalt hat. Diese lineare Algebra ist auch für sich genommen in der Physik unentbehrlich, und sie wird im zweiten Drittel der Vorlesung behandelt. (Daß die eindimensionale Analysis scheinbar ohne lineare Algebra auskommt, liegt bloß daran, daß die Theorie der linearen Strukturen in einer einzigen Variablen so einfach ist, daß man sie nicht als solche wahrnimmt).

Im letzten Drittel der Vorlesung wird dann die Vektoranalysis selbst zu Wort kommen. Damit ergibt sich im Groben die folgende Dreiteilung:

- Analysis einer Veränderlichen (Abschnitte 1 bis 16)
- Lineare Algebra (Abschnitte 17 bis 29)
- Vektoranalysis (Abschnitte 30 bis 44)

1 Mengen und Abbildungen

Obwohl Mathematik und Physik inhaltlich eng miteinander verflochten sind, sprechen sie nicht geradezu dieselbe Sprache. Der augenfälligste Unterschied besteht darin, daß Mathematiker sich heute konsequent in der Sprache der Cantorschen *Mengenlehre* ausdrücken, anders als Physiker das tun, wenn sie einen an sich mathematischen Sachverhalt beschreiben. Der Zweck dieses Abschnittes ist es, diese Sprache zu erklären; von der eigentlichen Mengenlehre wird uns das Allerelementarste genügen.

1.1 Definition Eine Menge ist eine (gedankliche) Zusammenfassung bestimmter wohlunterschiedener Objekte; diese heißen die Elemente der Menge.

Eine Menge M zu kennen, bedeutet also, von jedem (wie auch immer gearteten) Objekt x (in der ganzen Welt) zu wissen, ob x ein Element von M ist oder nicht.

1.2 Beispiele (1) Die Menge aller Physik-Studenten,

(2) die Menge aller Himmelskörper,

(3) die Menge $\mathbb{N} = \{0, 1, 2, \ldots\}$ der natürlichen Zahlen,

(4) $\mathbb{Z}$, die Menge der ganzen Zahlen,

(5) $\mathbb{Q}$, die Menge der rationalen Zahlen und

(6) $\mathbb{R}$, die Menge der reellen Zahlen.

Der Begriff der Menge erlaubt sich kein Urteil darüber, ob eine konkrete Menge interessant sein mag oder nicht. So ist etwa die Menge

(7) M, deren Elemente erstens Sie (die Sie hier im Hörsaal vor mir sitzen), zweitens das Stück Kreide, das ich in der Hand halte, und drittens das Newtonsche Gravitationsgesetz sind,

eine gemäß der Definition sinnvolle Menge: Schließlich sind alle aufgezählten Elemente Objekte, wie es die Definition 1.1 verlangt. Daß es schwerfällt, sich eine interessante Aussage vorzustellen, die man über diese konkrete Zusammenfassung von Objekten machen könnte, und daß einem M daher eher sinnlos vorkommt, ist nur ein subjektives Urteil, auf das es hier nicht ankommt.

In der Definition wird von einer Menge nicht verlangt, daß sie überhaupt ein Element enthält, vielmehr ist die aus gar keinem Element bestehende Menge

leere Menge $\emptyset$

eine durchaus akzeptable Menge.

Im Umgang mit Mengen verwendet man folgende

1.3 Symbole (a) Das Elementsymbol $\in$

(b) die Mengenklammern $\{\ldots\}$

(c) das Teilmengenzeichen $\subset$

(d) das Durchschnittzeichen $\cap$

(e) das Vereinigungszeichen $\cup$ und

(f) das Mengendifferenzzeichen $\setminus$

Ich erläutere diese Zeichen eines nach dem anderen.

Zu (a): Ist M eine Menge, so bedeutet $x \in M$, daß x ein Element von M ist (man sagt auch: x zu M gehört, in M liegt). Mit dem Durchstreichen eines Symbols bringt man immer das Gegenteil zum Ausdruck: hier heißt $y \notin M$ also, daß y kein Element von M ist. Etwa ist

$$-1 \in \mathbb{Z}, \text{ aber } -1 \notin \mathbb{N}.$$

Zu (b): Die Mengenklammern dienen in verschiedener Weise dazu, eine bestimmte Menge anzugeben. Am einfachsten dadurch, daß man alle Elemente der Menge zwischen den Klammern auflistet, zum Beispiel $\{1, 2, 3\}$. Beachten Sie, daß ich dieselbe Menge auch anders hinschreiben könnte, etwa

$$\{1, 2, 3\} = \{3, 1, 2\} = \{1, 2, 2, 3, 3, 3\} :$$

eine Menge ist allein dadurch bestimmt, *welche* Elemente sie enthält, und es gibt keinen Sinn, vom ersten, zweiten oder dritten Element einer Menge zu reden, und schon gar nicht davon, wie oft ein Element der Menge in ihr enthalten ist.

Diese einfachste Verwendung der Mengenklammern ist grundsätzlich natürlich nur bei endlichen Mengen (solchen mit endlich vielen Elementen) möglich und auch dann nicht immer praktikabel. Andererseits genügt es oft, von der darzustellenden Menge nur einige der Elemente wirklich hinzuschreiben und die übrigen durch "Pünktchen" anzudeuten. Etwa würde

$$\{1, 2, \ldots, 10\}$$

von jedermann so verstanden, wie es gemeint ist, nämlich als $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, und die Schreibweise

$$\mathbb{N} = \{0, 1, 2, 3, \ldots\},$$

die ich oben schon verwendet habe, haben Sie sicher auch richtig interpretiert.

Aber natürlich darf man diese Methode nur anwenden, wenn wirklich unmißverständlich klar ist, für welche Elemente die Punkte stehen. Deswegen wären etwa

$$\{1, 2, 4, \ldots\} \text{ oder } \{1, 5, 0, 13, \ldots\}$$

keine zulässigen Bezeichnungen einer Menge, denn im ersten Fall sind $\{1, 2, 4, 8, 16, \ldots\}$ und $\{1, 2, 4, 7, 11, \ldots\}$ zwei verschiedene naheliegende Interpretationen, während im zweiten überhaupt nicht erkennbar ist, wie es nach der 13 weitergehen soll.

Die dritte, wichtigste und immer korrekte Methode, eine Menge M mittels der Klammern anzugeben, besteht darin, zunächst ein willkürlich gewähltes Symbol (meist einen Buchstaben) als eine Art Platzhalter für die Elemente von M in die Klammer zu schreiben und dann hinter einem senkrechten Strich die Elemente durch eine oder mehrere Eigenschaften zu charakterisieren. Zum Beispiel ist

$$\{1, 2, \ldots, 10\} = \{x \,|\, x \text{ ist ganze Zahl und } 1 \leq x \leq 10\} = \{y \,|\, y \text{ ist ganze Zahl und } 1 \leq y \leq 10\},$$

worin ich die letzte Version nur hinzugefügt habe, um zu illustrieren, daß die Wahl des Platzhaltersymbols wirklich keine Rolle spielt. Häufig gehören die in Betracht kommenden Elemente von vornherein einer Menge an, für die man schon einen Namen hat (so wie hier); diese Zugehörigkeit kann man dann bequemer links vom Strich notieren:

$$\{1, 2, \ldots, 10\} = \{x \in \mathbb{Z} \,|\, 1 \leq x \leq 10\} = \{x \in \mathbb{N} \,|\, 1 \leq x \leq 10\}$$

Zu (c): Sind A und B Mengen, so bedeutet

$$A \subset B,$$

daß jedes Element von A auch Element von B ist. Man sagt: A ist eine Teilmenge von B. Zum Beispiel ist

$$\emptyset \subset \{1, 2, 3\} \subset \{x \in \mathbb{Z} \,|\, 1 \leq x \leq 10\} \subset \mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R},$$

während für jede Menge M trivialerweise

$$\emptyset \subset M \text{ und } M \subset M$$

gelten. Mit

$$A \not\subset B$$

ist natürlich gemeint, daß A *keine* Teilmenge von B ist. Das bedeutet nicht etwa, daß kein Element von A zu B gehört, sondern nur, daß es *mindestens ein* Element $a \in A$ gibt, das nicht zu B gehört:

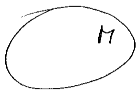
$$\{1, 2, 3\} \not\subset \{1, 2, 4\}$$

wegen $3 \notin \{1, 2, 4\}$.

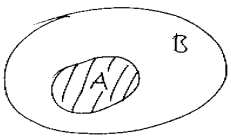
Wohl von selbst versteht sich:

$$A = B \text{ gilt genau dann, wenn } A \subset B \text{ und } B \subset A \text{ ist.}$$

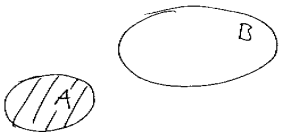
Manchmal ist es praktisch, sich die Begriffe der elementaren Mengenlehre durch naive Skizzen der folgenden Art zu veranschaulichen: In



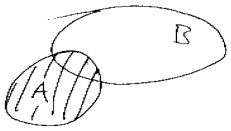
soll M in der Regel durch die Menge aller von der geschlossenen Linie eingezäunten Punkte repräsentiert werden. Also etwa



als Illustration für $A \subset B$, und



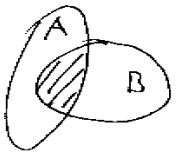
oder



als Möglichkeiten für $A \not\subset B$.

Zu (d) bis (f): Aus zwei gegebenen Mengen A und B lassen sich auf die verschiedensten Arten weitere konstruieren, insbesondere der Durchschnitt von A und B

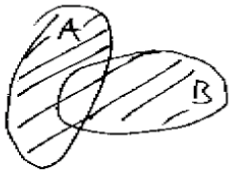
$$A \cap B := \{x \mid x \in A \text{ und } x \in B\},$$



$A \cap B$

die Vereinigung

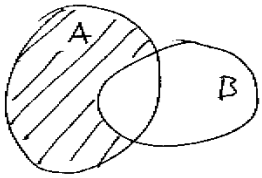
$$A \cup B := \{x \mid x \in A \text{ oder } x \in B\}$$



$A \cup B$

und die Differenz

$$A \setminus B := \{x \,|\, x \in A,\ x \notin B\}.$$



$A \setminus B$

Bei der Gelegenheit habe ich einige Feinheiten der Notation eingeführt: Der Doppelpunkt vor dem Gleichheitszeichen weist darauf hin, daß die linke Seite als die rechte definiert wird. Das hat natürlich nur deswegen einen Sinn, weil links etwas Neues, eben bisher noch nicht definiertes steht, während rechts schon Bekanntes steht. (Logischerweise darf also in der ganzen Vorlesung nie wieder ein "$A \setminus B :=$" auftauchen, jedenfalls nicht, wenn A und B Mengen sind.) Die andere Feinheit, auf die ich Sie aufmerksam machen möchte, ist, daß ein Komma in einer Aufzählung von Eigenschaften einfach die Bedeutung des logischen "und" haben soll.

Mengen A, B mit $A \cap B = \emptyset$ nennt man übrigens (zueinander) disjunkt. Beachten Sie noch, daß die Mengendifferenz $A \setminus B$ auch dann erklärt ist, wenn B keine Teilmenge von A ist: die Skizze illustriert das ja schon.

Eine weitere wichtige Konstruktion mit Mengen ist das kartesische Produkt. Dazu müssen wir klären, was wir unter einem *Paar* von Objekten verstehen:

1.4 Definition Ein Paar (a, b) von Objekten besteht aus der Angabe eines ersten Objekts a und eines zweiten Objekts b. Diese heißen auch erste bzw. zweite Komponente von (a, b).

Die Gleichheit zweier Paare

$$(a, b) = (a', b')$$

bedeutet demgemäß dasselbe wie

$$a = a' \text{ und } b = b'.$$

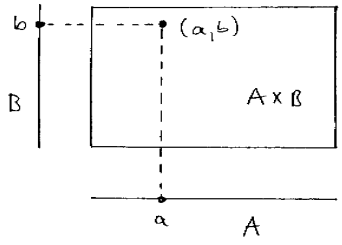
Verwechseln Sie das Paar (a, b) nicht mit der Menge $\{a, b\}$: während immer $\{a, b\} = \{b, a\}$ ist, gilt $(a, b) = (b, a)$ nur in dem speziellen Fall $a = b$. Dann aber ist (a, a) immer noch ein richtiges Paar (dessen beide Komponenten gleich sind), während die Menge $\{a, a\} = \{a\}$ nur *ein* Element hat.

1.5 Definition A und B seien Mengen. Dann heißt die Menge

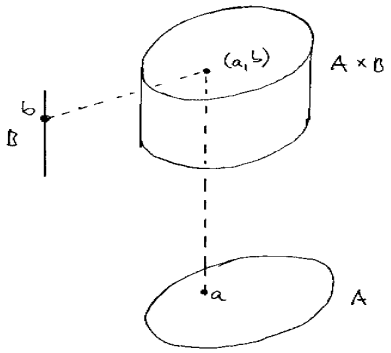
$$A \times B := \{(a, b) \,|\, a \in A,\ b \in B\}$$

das (kartesische) Produkt von A und B.

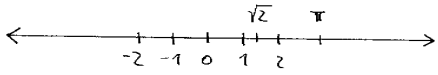
Dieses Produkt kann man sich am einfachsten veranschaulichen, wenn man sich A und B als Strecken vorstellt:



oder auch eine der beiden Mengen als Scheibe:

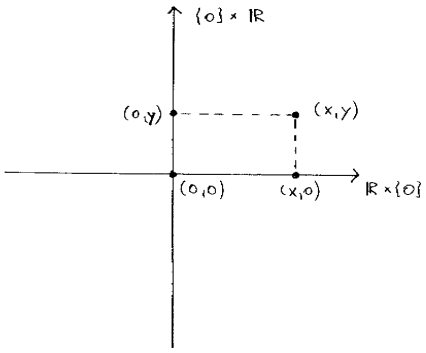


In der Mathematik besteht der — nicht zu verachtende — Wert solcher Skizzen zunächst nur darin, daß sie eine anschauliche Merkhilfe für einen abstrakten Sachverhalt bieten. In Wirklichkeit werden A und B natürlich nicht so aussehen wie in den Bildern, einfach deshalb, weil es ganz sinnlos ist, über das "Aussehen" von abstrakten Mengen zu reden. Es gibt andererseits Situationen, in denen die Skizzen über das rein Symbolische hinaus interpretiert werden dürfen, und in dem für die Physik relevanten Teil der Mathematik sind diese Situationen häufig. Beispielsweise wird Ihnen vertraut sein, daß man sich die Menge $\mathbb{R}$ der reellen Zahlen oft mit Vorteil als die "Zahlengerade" vorstellt:



Diese Zahlengerade ist sicher mehr als ein nur symbolisches Bild der Menge $\mathbb{R}$; etwa kann man auf ihr reelle Zahlen miteinander vergleichen: die größeren Zahlen liegen rechts, die kleineren links. Physikalisch wird zum Beispiel die Zeit nach Wahl eines Zeitpunktes als Gegenwart und einer Zeiteinheit zweckmäßig durch die Zahlengerade dargestellt: positive Zahlen stehen für Zeitpunkte in der Zukunft, negative für solche in der Vergangenheit.

Wenn wir die reellen Zahlen als Punkte auf der Zahlengeraden auffassen, dann werden die Elemente des kartesisches Produktes $\mathbb{R} \times \mathbb{R}$ konsequenterweise durch die Punkte der Zeichenebene im Sinne der Skizze



dargestellt. Wir können noch weiter gehen, wenn wir wie Paare auch Tripel, Quadrupel etc. betrachten:

1.6 Definition Analog zu Paaren (a, b) sind Tripel (a, b, c) und für beliebiges $p \in \mathbb{N}$ allgemein p-tupel $(a_1, a_2, \ldots, a_p)$ erklärt. Sind $A_1, A_2, \ldots, A_p$ Mengen, so heißt

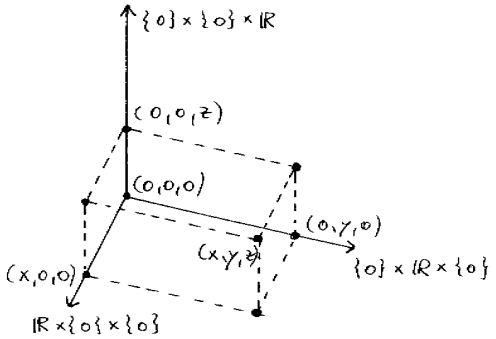
$$A_1 \times A_2 \times \cdots \times A_p := \{(a_1, a_2, \ldots, a_p) \,|\, a_1 \in A_1,\ a_2 \in A_2,\ \ldots,\ a_p \in A_p\}$$

das kartesische Produkt der Mengen $A_1, A_2, \ldots, A_p$. Letztere brauchen natürlich nicht alle verschieden zu sein; sind sie sogar alle gleich, so schreibt man

$$A^p = A \times A \times \cdots \times A \quad (p \text{ Faktoren}).$$

Für $p = 1$ wollen wir das vernünftigerweise einfach als $A^1 = A$ lesen, indem wir das 1-tupel (a) mit dem Element $a \in A$ identifizieren. Im Fall $p = 0$ soll A^0 nicht etwa die leere Menge bedeuten, sondern die, die als einziges Element das 0-tupel () enthält, das eben keine Komponente hat.

Unter den "Potenzen" $\mathbb{R}^p$ spielt $\mathbb{R}^3$ in der Physik eine besondere Rolle, weil dieses dreifache Produkt ein gutes mathematisches Modell für den dreidimensionalen physikalischen Raum unserer Anschauung ist — zumindest dann, wenn wir in letzterem einen Nullpunkt, die Richtungen "vorn, rechts, oben" und einen Maßstab gewählt haben. Auf die Abhängigkeit von diesen mehr oder weniger willkürlichen Wahlen und die Grenzen der Tauglichkeit dieses Modells werden wir später übrigens noch ausführlich zu sprechen kommen. Im Augenblick wollen wir uns daran aber erst mal erfreuen und einen "Punkt" des Raumes, alias ein Tripel $(x, y, z) \in \mathbb{R}^3$ in ein Bild von $\mathbb{R}^3$ einzeichnen:



Wie Sie sehen, ist es übersichtlicher, von den Achsen $\mathbb{R} \times \{0\} \times \{0\}$, $\{0\} \times \mathbb{R} \times \{0\}$ und $\{0\} \times \{0\} \times \mathbb{R}$ nur noch die "positiven" Hälften einzuzeichnen. Freilich, anders als bei $\mathbb{R}^2$ läßt sich (x, y, z) aus dem so bezeichneten Punkt auf der Tafel nicht rekonstruieren. Das dürfte Sie kaum überraschen, und es ist auch ganz typisch: Man darf von den Skizzen nicht mehr erwarten, als daß sie den einen oder anderen Teilaspekt veranschaulichen: darin liegt ihr Nutzen. Vom streng logischen Standpunkt aus sind sie geradezu überflüssig: mit der Definition "$\mathbb{R}^3$ =Menge aller Tripel reeller Zahlen" ist schon alles gesagt. Mathematiker, die gern geometrisch denken (zu denen zähle ich mich), mögen diese Anschauungshilfen trotzdem nicht missen, und verwenden geometrische Vokabeln selbst dann, wenn es sich eigentlich um abstrakte Objekte handelt. Stören Sie sich also nicht daran, daß ich gelegentlich von "Punkten" in einer (abstrakten) Menge A rede, wenn ich die Elemente von A meine.

In der zweiten Hälfte dieses Abschnitts geht es um den Begriff der Abbildung.

1.7 Definition X und Y seien Mengen. Eine Abbildung (oder Funktion) f von X nach Y ist eine Vorschrift, die jedem $x \in X$ genau ein Element $f(x) \in Y$ zuordnet; dieses nennt man den Wert von f bei x (oder an der Stelle x), oder auch den Wert oder Bildpunkt von x unter f.

Es gibt verschiedene Schreibweisen dafür, daß f eine solche Abbildung von X nach Y ist:

$$f\colon X \longrightarrow Y$$

oder

$$X \xrightarrow{f} Y$$

oder, wenn man für jedes $x \in X$ seinen Bildpunkt $f(x) \in Y$ wirklich hinschreiben will oder muß:

$$\begin{aligned} X &\xrightarrow{f} Y \\ x &\mapsto f(x) \end{aligned}$$

oder

$$X \ni x \mapsto f(x) \in Y$$

(es dürfte klar sein, was mit dem 'rumgedrehten Elementsymbol gemeint ist).

1.8 Beispiele (1) Die Abbildung $\mathbb{Z} \ni x \mapsto x^2 \in \mathbb{N}$ (hier ist es gar nicht nötig, der Abbildung selbst einen Namen zu geben)

(2) Die Abbildung $\mathbb{Z} \ni x \mapsto x^2 \in \mathbb{Z}$. Ist das nicht dieselbe? Nun, wir wollen die Definition so verstehen, daß zur vollständigen Angabe einer Abbildung $f\colon X \longrightarrow Y$ auch ihr Definitionsbereich X und ihre Zielmenge (oder Zielbereich) Y gehören, unabhängig davon, welche $y \in Y$ tatsächlich als Werte, d.h. als $f(x)$ für mindestens ein $x \in X$ vorkommen. In diesem Sinne sind die Beispiele (1) und (2) verschiedene Abbildungen, auch wenn Ihnen das jetzt pedantisch vorkommen mag.

(3) $\mathbb{N} \ni x \mapsto \pm\sqrt{x} \in \mathbb{R}$ definiert *keine* Abbildung, denn diese Vorschrift verstößt gegen die Forderung, daß jedem $x \in \mathbb{N}$ *genau* ein $f(x) \in \mathbb{R}$ zuzuordnen ist. Dagegen ist

$$\mathbb{N} \ni x \mapsto \sqrt{x} \in \mathbb{R}$$

eine ordentliche Abbildung, wenn mit $\sqrt{x}$ wie üblich diejenige *positive* (genauer: nicht negative) reelle Zahl gemeint ist, deren Quadrat x ergibt.

(4) Eine ausgefallenere Abbildung $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ wird durch die Vorschrift

$$f(x) = \begin{cases} 1 & \text{falls } x \in \mathbb{Q} \\ 0 & \text{falls } x \notin \mathbb{Q} \end{cases}$$

definiert: nirgends wird verlangt, daß die Zuordnung durch eine "glatte" Formel bewerkstelligt werden müßte. Dagegen wäre das Weglassen einer der beiden Alternativen wieder unzulässig, weil dann ja nicht mehr *jedem* $x \in \mathbb{R}$ ein Wert $f(x)$ zugeordnet würde.

(5) Eine geradezu verrückte Abbildung $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ könnte man durch

$$f(x) = \begin{cases} 1 & \text{falls die Dezimalbruchdarstellung von } x \text{ die Ziffernfolge 4711 enthält} \\ 0 & \text{sonst} \end{cases}$$

definieren: Diese abwegige und vermutlich ganz nutzlose Vorschrift liefert nichtsdestoweniger eine richtige Abbildung im Sinne der Definition.

(6) Die Addition reeller Zahlen ist auch eine Abbildung, nämlich

$$\begin{aligned} \mathbb{R} \times \mathbb{R} &\longrightarrow \mathbb{R} \\ (x, y) &\mapsto x+y. \end{aligned}$$

(7) Für jede Menge A hat man die identische Abbildung

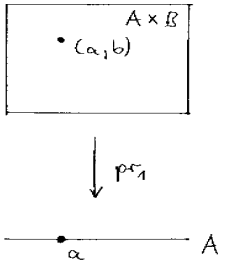
$$\mathrm{id}_A\colon A \longrightarrow A; \quad x \mapsto x.$$

Man schreibt sie auch einfach id, wenn aus dem Zusammenhang klar ist, welche Menge A gemeint ist.

(8) Für beliebige Mengen A, B nennt man die Abbildung

$$\mathrm{pr}_1\colon A \times B \longrightarrow A; \quad (x, y) \mapsto x$$

die Projektion auf den ersten Faktor:



Analog ist natürlich pr_2 erklärt, und wenn A und B verschieden sind, schreibt man statt pr_1 und pr_2 auch gern pr_A beziehungsweise pr_B.

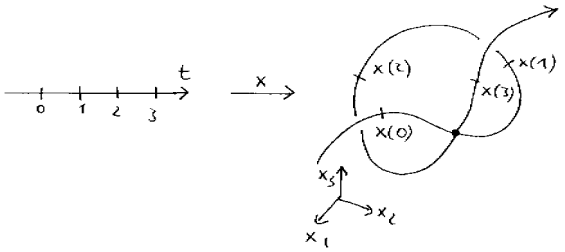
(9) Sind X, Y Mengen, X nicht leer und $c \in Y$ ein Element, so nennt man die Abbildung

$$X \longrightarrow Y; \quad x \mapsto c$$

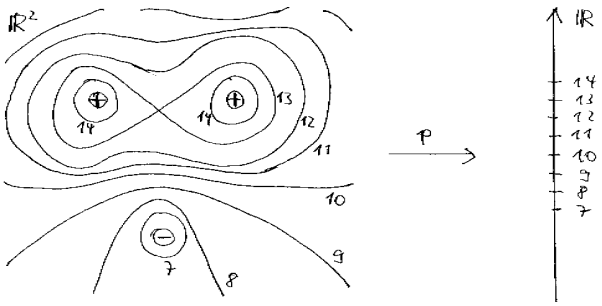
konstante Abbildung (mit Wert c).

(10) In der Mechanik idealisiert man im Vergleich zu anderen Abmessungen kleine Körper zweckmäßig zu einem *Massenpunkt*, vernachlässigt also die räumliche Ausdehnung des Körpers selbst. Die Bahn eines solchen Massenpunktes wird dann durch eine Abbildung der Zeitachse in den Raum gegeben, in einer der in der Physik üblichen Notationen also durch

$$x\colon \mathbb{R} \longrightarrow \mathbb{R}^3; \quad t \mapsto x(t) = \big(x_1(t), x_2(t), x_3(t)\big).$$



(11) Größen, die man in der Physik als *skalare Felder* bezeichnet, sind mathematisch gesehen Abbildungen von $f\colon \mathbb{R}^3 \longrightarrow \mathbb{R}$. Nehmen wir als Beispiel den Luftdruck p in der Atmosphäre, wobei ich der einfachen Darstellbarkeit halber nur den Druck am Boden, also eine Abbildung $p\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ skizziere:



Wie? Nun, indem ich einige Linien konstanten Drucks, also einige Isobaren eingezeichnet habe.

Übrigens verwendet man das Wort "Funktion" vorzugsweise für $\mathbb{R}$-wertige Abbildungen, sagt dann auch gern, daß $f\colon X \longrightarrow \mathbb{R}$ eine Funktion auf X ist. Wir wollen aber nicht so weit gehen, die Bedeutung des Wortes nur auf diesen Fall zu beschränken: für uns sind "Abbildung" und "Funktion" also grundsätzlich austauschbare Vokabeln.

Ein allgemeiner Sachverhalt, der Ihnen vielleicht bei einem der beiden letzten Beispiele aufgefallen ist: In der Mathematik werden Buchstaben als Symbole in einer etwas anderen Weise gebraucht als in der Physik. Dort haben ja viele Buchstaben von vornherein eine feste (wenn auch nicht immer einheitlich vereinbarte) Bedeutung: m für die Masse, t für die Zeit, T für die absolute Temperatur, U (manchmal φ) für elektrische Spannungen und p für den Druck.... Der Mathematiker ist dagegen in der Wahl von Buchstaben als Symbolen fast völlig frei, solange er bloß Kollisionen mit den wenigen Doppelstrichbuchstaben $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ vermeidet und auch ansonsten nicht böswillig Mißverständnisse etwa dadurch provoziert, daß er ein Objekt π nennt, wenn im Zusammenhang auch die reelle Zahl π eine Rolle spielt. Für die Verständigung zwischen Mathematikern und Physikern ist dieser Unterschied Anlaß zu einigen Problemen. Schauen wir uns noch mal die Zeile

$$x\colon \mathbb{R} \longrightarrow \mathbb{R}^3; \quad t \mapsto x(t) = \big(x_1(t), x_2(t), x_3(t)\big)$$

an. Ein Physiker hätte dafür viel eher

$$\underline{x} = \underline{x}(t) \text{ oder in der ausführlichen Version } \underline{x} = \big(x_1, x_2, x_3\big) = \underline{x}(t) = \big(x_1(t), x_2(t), x_3(t)\big)$$

geschrieben. Darin soll die Unterstreichung von x zum Ausdruck bringen, daß es sich nicht um eine skalare Größe (eine Zahl), sondern einen *Vektor* (bestehend aus drei Zahlkomponenten) handelt; unter dem nicht-unterschrichenen x wird dann stillschweigend der *Betrag* dieses Vektors verstanden. Statt der Unterstreichung sind auch Fettdruck: $\mathbf{x}$, Überpfeilung: $\vec{x}$ oder andere Gags in Gebrauch. In einem mathematischen Text kann man darauf verzichten, die Notation so stark zu beladen, weil man dieselbe Information schon aus der Angabe $x \in \mathbb{R}^3$ entnimmt. Die wieder kann der Physiker oft guten Gewissens weglassen, weil ja schon die Wahl des Buchstabens x auf die Bedeutung "Ort" hinweist. Jedenfalls ist das *eine* Konvention; eine andere verwendet dafür ein $\mathbf{r}$, um x für dessen erste Komponente freizuhalten, also $\mathbf{r} = (x, y, z)$ (das nicht-fette r bedeutet dann den Abstand von $\mathbf{r}$ vom Nullpunkt). Zumindest (x, y, z) für die Tripel in $\mathbb{R}^3$ mag ich auch ganz gern, weil man damit die schwerfälligen Indizes vermeidet.

Das eigentliche Problem ist aber dies: Die in der Physik so geläufige und dort auch vernünftige Schreibweise $x = x(t)$ (x ist der Ort des Massenpunktes, und der hängt eben von der Zeit ab) ist als mathematische Formel ganz sinnlos. Denn x ist die ganze Abbildung $\mathbb{R} \longrightarrow \mathbb{R}^3$, während $x(t) \in \mathbb{R}^3$ bloß ein einzelner Wert dieser Abbildung, also ein Element von $\mathbb{R}^3$ ist. Das gilt natürlich entsprechend auch für die Physikergleichungen $(x_1, x_2, x_3) = \big(x_1(t), x_2(t), x_3(t)\big)$ und $p = p(x, y, z)$, die deshalb in der Mathematik ebenfalls zu vermeiden sind.

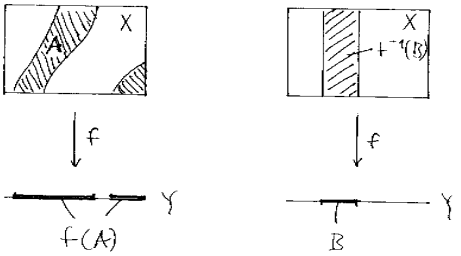
1.9 Definition Sei $f\colon X \longrightarrow Y$ eine Abbildung, und seien $A \subset X$, $B \subset Y$ Teilmengen. Dann heißt

$$f(A) := \{f(x) \,|\, x \in A\} \subset Y$$

die Bildmenge oder kurz das Bild von A unter f. Die Menge

$$f^{-1}B := \{x \in X \,|\, f(x) \in B\} \subset X$$

dagegen heißt das Urbild von B unter f.



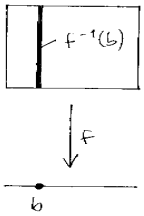
Wahrscheinlich werden Sie von diesen beiden Bildungen die erste als die einfachere empfinden. In Wirklichkeit ist es aber umgekehrt. Das sehen Sie daran, daß die Schreibweise $\{f(x) \,|\, x \in A\}$ gar nicht die unter 1.3(b) vereinbarte Form hat, sondern eine erklärungsbedürftige Abkürzung für die Menge

$$\{y \in Y \,|\, \text{es gibt ein } x \in A \text{ mit } f(x) = y\}$$

ist: jetzt sieht man deutlich, daß 'Bild' komplizierter als 'Urbild' ist. Ich versuche Sie durch die Notation ein wenig an diesen Sachverhalt zu erinnern, indem ich beim Bild $f(A)$ die Klammern setze, beim Urbild $f^{-1}B$ dagegen nur dann, wenn es zur Vermeidung von Mißverständnissen erforderlich ist.

Speziell wenn $B = \{b\}$ aus einem einzigen Element besteht, nennt man $f^{-1}\{b\}$ gern die Faser von f über b:

$$f^{-1}\{b\} = \{x \in X \,|\, f(x) = b\}$$

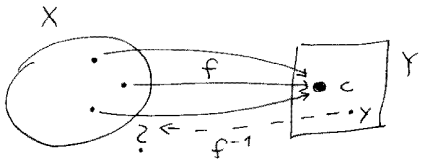


Um an eine Ihnen sicher vertrautere Sprechweise anzuschließen: Die Faser $f^{-1}\{b\}$ ist die Menge aller Lösungen $x \in X$ der Gleichung

$$f(x) = b.$$

In Beispiel (10) sind die Fasern gerade die Isobaren.

Die Schreibweise $f^{-1}B$ verführt den Anfänger immer wieder dazu, an eine *Umkehrabbildung* f^{-1} von f zu denken, die die Wirkung von f wieder rückgängig macht. Die gibt es im allgemeinen aber nicht: ein extremes Beispiel dafür sind die konstanten Abbildungen $f: X \longrightarrow Y$: Wenn $f(x) = c$ für *jedes* $x \in X$ ist, wie wollen Sie dann x aus $f(x)$ rekonstruieren (falls nicht gerade $X = \{x\}$ nur aus einem Element besteht)?



Und was sollte erst eine solche Umkehrabbildung mit einem $y \in Y \setminus \{c\}$ machen (falls es solche gibt, d.h. wenn Y mehr als ein Element hat)? Dagegen ist das Urbild $f^{-1}B$ einer jeden Menge $B \subset Y$ auch hier ohne weiteres definiert (was kommt heraus?).

Bild und Urbild haben einige einfache und auch leicht zu beweisende

1.10 Eigenschaften Sei $f: X \longrightarrow Y$ eine Abbildung. Für beliebige Teilmengen $A, A' \subset X$ gilt

$$f(A \cup A') = f(A) \cup f(A') \text{ sowie } f(\emptyset) = \emptyset;$$

für beliebige Teilmengen $B, B' \subset Y$ gilt

$$f^{-1}(B \cup B') = f^{-1}B \cup f^{-1}B', \quad f^{-1}(B \cap B') = f^{-1}B \cap f^{-1}B' \text{ und } f^{-1}(B \setminus B') = f^{-1}B \setminus f^{-1}B'$$

sowie

$$f^{-1}\emptyset = \emptyset \text{ und } f^{-1}Y = X.$$

1.11 Definition Eine Abbildung $f: X \longrightarrow Y$ heißt surjektiv (oder eine Abbildung von X auf Y), wenn jedes Element von Y als Wert von f vorkommt, wenn es also zu jedem $y \in Y$ ein (d.h. mindestens ein) $x \in X$ gibt mit

$$f(x) = y.$$

Eine Abbildung $f: X \longrightarrow Y$ heißt injektiv, wenn sie zwei verschiedene Elemente von X niemals auf dasselbe Element von Y abbildet, wenn also aus

$$x, y \in X \text{ und } f(x) = f(y)$$

stets

$$x = y$$

folgt.

Schließlich heißt die Abbildung $f: X \longrightarrow Y$ bijektiv, wenn sie sowohl injektiv als auch surjektiv ist.

Bemerkung Machen Sie sich klar, daß die verbale und die formelmäßige Beschreibung des Begriffes "injektiv" wirklich dasselbe bedeuten. Warum überhaupt zwei Formulierungen? Nun, die erste scheint mir die plastischere, um sich Injektivität vorzustellen: was in X verschieden ist, bleibt auch beim Abbilden mit f voneinander verschieden. Andererseits rechnet und argumentiert es sich mit "voneinander verschieden" nicht besonders gut; für den praktischen Gebrauch ist deshalb die zweite Version, die stattdessen ganz mit der Gleichheit von Elementen arbeitet, in aller Regel vorzuziehen.

Übrigens kann man die drei jetzt eingeführten Eigenschaften einer Abbildung f auch mittels der Fasern ausdrücken: f ist surjektiv genau dann, wenn alle Fasern nicht-leer sind; f ist injektiv genau dann, wenn alle Fasern von f höchstens einpunktig sind; f ist bijektiv, wenn jede Faser aus genau einem Punkt besteht.

1.12 Beispiele (1) Die Abbildung

$$\begin{aligned} \{1,2,3,\dots,26\} &= \{a,b,c,\dots,z\} \\ 1 &\mapsto a \\ 2 &\mapsto b \\ &\vdots \\ 26 &\mapsto z \end{aligned}$$

ist bijektiv.

(2) Die Abbildung

$$\begin{aligned} \mathbb{N} &\longrightarrow \mathbb{N} \\ x &\mapsto x+1 \end{aligned}$$

ist injektiv, aber nicht surjektiv.

(3) Die Abbildung

$$\begin{aligned} \mathbb{N} &\longrightarrow \mathbb{N} \\ 0 &\mapsto 0 \\ x &\mapsto x-1 \text{ für } x>0 \end{aligned}$$

ist surjektiv, aber nicht injektiv.

1.13 Definition Sind $X \xrightarrow{f} Y$ und $Y \xrightarrow{g} Z$ Abbildungen, so erklärt man die zusammengesetzte Abbildung oder Komposition $g \circ f$ (oder kurz gf) durch

$$\begin{aligned} X &\xrightarrow{g\circ f} Z \\ x &\mapsto g\big(f(x)\big). \end{aligned}$$

Die Gewohnheit, daß wir $f(x)$ und nicht $(x)f$ schreiben, erzwingt, jedenfalls fast, die Reihenfolge in $g \circ f$, obwohl zuerst f und dann g angewendet wird. Das mag man als störend empfinden, aber man muß damit leben.

1.14 Lemma Für beliebige Abbildungen $W \xrightarrow{f} X$, $X \xrightarrow{g} Y$ und $Y \xrightarrow{h} Z$ gilt die Regel

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

weswegen wir in Zukunft die Klammern weglassen dürfen.

Ein Lemma bezeichnet in der Mathematik übrigens in der Regel einen (Lehr-)satz, also ein mathematisches Resultat, dem man wegen seiner Einfachheit aber nicht den Namen eines richtigen Satzes zubilligen möchte. Ob einfach oder nicht, in der Mathematik muß jede Behauptung bewiesen werden, hier also der

Beweis Zu zeigen ist, daß

$$\big(h \circ (g \circ f)\big)(w) = \big((h \circ g) \circ f\big)(w) \text{ für jedes } w \in W$$

gilt. Sei also $w \in W$ beliebig. Dann ist einerseits

$$\big(h \circ (g \circ f)\big)(w) = h\big((g \circ f)(w)\big) = h\Big(g\big(f(w)\big)\Big)$$

und andererseits

$$\big((h\circ g)\circ f\big)(w) = (h\circ g)\big(f(w)\big) = h\Big(g\big(f(w)\big)\Big).$$

Damit ist unsere Behauptung schon bewiesen.

Wie gesagt muß in der Mathematik jede nicht unmittelbar evidente Behauptung bewiesen werden. Daß ich nun schon gut zwei Vorlesungen lang ohne Beweise ausgekommen bin, liegt bloß daran, daß ich auch noch nichts behauptet habe, sondern bloß Begriffe, also Vokabeln definiert habe. Am Anfang ist das ja zwangsläufig so; ohne Begriffe kann man auch nichts behaupten. Allmählich werden Sätze (oder Lemmata) und damit auch deren Beweise immer häufiger werden; in ihnen liegt die eigentliche Substanz der Mathematik.

Muß man, vor allem müssen Sie als Physikstudenten die Beweise überhaupt lernen? Grundsätzlich ja. So wie die definierten Begriffe nur dadurch verständlich werden, was man mit ihnen machen kann, also durch die Sätze, so versteht man diese erst richtig, wenn man weiß, *warum* sie gelten. Warum gilt ein Satz? Na ja, das erklärt gerade der Beweis. Mit Lernen ist übrigens nicht Auswendiglernen gemeint. Es geht vielmehr darum, die in jedem Beweis steckende *Idee* zu verstehen: die kann man oft ganz leicht behalten, und nach einiger Übung werden Sie dann die Ausführung des Beweises im einzelnen bei Bedarf allein anhand dieser Idee ergänzen können. Nicht immer freilich ist die einem Beweis zugrundeliegende Idee interessant, und ich werde Ihnen im Laufe der Zeit auch schon mal mathematische Sätze ohne Beweis präsentieren, nämlich dann, wenn der Beweis nach meiner Einschätzung weniger zum Verständnis und dem Erlernen des sicheren Umgangs mit dem Satzes beiträgt.

Ich habe schon darauf hingewiesen, daß es zu einer Abbildung $X \xrightarrow{f} Y$ im allgemeinen keine Umkehrung $f^{-1}\colon Y \longrightarrow X$ gibt, die die Wirkung von f rückgängig macht. In besonderen Fällen kann es aber eine solche Umkehrung schon geben, deshalb die

1.15 Definition $f\colon X \longrightarrow Y$ sei eine Abbildung. Eine Abbildung $g\colon Y \longrightarrow X$ heißt eine Umkehrabbildung von f (oder kurz Umkehrung von f), wenn

$$g\circ f = \mathrm{id}_X \text{ und } f\circ g = \mathrm{id}_Y$$

gilt.

Auskunft über solche Umkehrungen gibt der

1.16 Satz (und **Bezeichnung**) Sei $f\colon X \longrightarrow Y$ eine Abbildung. Eine Umkehrung von f existiert genau dann, wenn f bijektiv ist. Ist das der Fall, dann gibt es auch nur eine (und nicht mehrere) Umkehrungen von f, und man bezeichnet sie mit

$$f^{-1}\colon Y \longrightarrow X.$$

Beweis Das "genau" bedeutet, daß zwei Richtungen zu beweisen sind. In der ersten setzen wir voraus, daß eine Umkehrabbildung $g\colon Y \longrightarrow X$ von f existiert: zu zeigen ist, daß f dann bijektiv sein muß. Zum Beweis der Injektivität betrachten wir zwei Elemente $x, x' \in X$ mit $f(x) = f(x')$. Anwenden von g liefert

$$x = \mathrm{id}_X(x) = (g\circ f)(x) = g\big(f(x)\big) = g\big(f(x')\big) = (g\circ f)(x') = \mathrm{id}_X(x') = x',$$

und das beweist, daß f injektiv ist. Um zu sehen, daß f auch surjektiv ist, betrachten wir ein beliebiges Element $y \in Y$. Dann ist

$$f\big(g(y)\big) = (f\circ g)(y) = \mathrm{id}_Y(y) = y,$$

insbesondere kommt y als Wert von f vor. Also ist f auch surjektiv und damit bijektiv.

Im zweiten Teil des Beweises zeigen wir die umgekehrte Richtung: Daß f bijektiv ist, wird jetzt vorausgesetzt, und zu beweisen ist, daß dann eine Umkehrung g von f existiert. Sei dazu $y \in Y$ beliebig. Weil f surjektiv ist, gibt es ein $x \in X$ mit $f(x) = y$. Weil f injektiv ist, ist dieses x eindeutig bestimmt: aus $f(x) = y = f(x')$ folgt ja $x = x'$. Also definiert die Zuordnung

$$Y \ni y \mapsto x \in X$$

eine Abbildung $g\colon Y \longrightarrow X$. Diese erfüllt nach Konstruktion $g \circ f = \mathrm{id}_X$ und $f \circ g = \mathrm{id}_Y$, ist also eine Umkehrung von f.

Schließlich (dritter Teil) müssen wir beweisen, daß die Umkehrung von f im Falle ihrer Existenz eindeutig bestimmt ist. Dazu brauchen wir aber bloß zu bemerken, daß die im zweiten Teil gewählte Definition von g auch die einzig mögliche war: Wenn wir das Element $y \in Y$ als $y = f(x)$ schreiben und $g \circ f = \mathrm{id}$ gelten soll, so muß zwangsläufig

$$g(y) = g\big(f(x)\big) = (g \circ f)(x) = x$$

werden. Deshalb ist das so erklärte g auch die einzige Umkehrung von f.

Das war der vollständige Beweis des Satzes. Weil dessen logische Struktur aber schon etwas verwickelter ist, will ich sie hier rückblickend noch analysieren. Dazu wollen wir die vorkommenden Aussagen mit

(A)	f besitzt (mindestens) eine Umkehrung
(B)	f ist bijektiv
(C)	f besitzt genau eine Umkehrung

benennen. Der Satz verspricht dann zunächst, daß A genau dann richtig ist, wenn B richtig ist. Alternative Ausdrucksweisen dafür: A gilt *dann und nur dann*, wenn B gilt, oder A und B sind *äquivalente Aussagen*. Eine solche Behauptung ist ihrerseits immer gleichbedeutend mit zwei Teilbehauptungen, nämlich erstens der, daß aus A die Aussage B folgt, und zweitens, daß umgekehrt aus B wieder A folgt. Diese beiden Behauptungen entsprechen genau den beiden ersten Teilen des Beweises.

Der Satz behauptet darüber hinaus aber noch mehr: daß nämlich A (oder, nach dem ersten Teil gleichwertig B) die dritte Eigenschaft C zur Folge hat. Diese Behauptung wird im dritten Teil des Beweises gezeigt. Beachten Sie, daß es dabei völlig egal ist, ob wir mit A oder B oder beiden als Voraussetzung beginnen, da deren Äquivalenz ja schon durch die beiden ersten Beweisteile gesichert ist.

Für die Aussage, daß B eine Folgerung aus A ist, gibt es auch ein Zeichen: $A \Rightarrow B$, und für die Äquivalenz von A und B dann naheliegenderweise $A \Longleftrightarrow B$. Mit den eingeführten Abkürzungen ließe sich die Aussage von Satz 1.16 dann als

$$A \iff B \Longrightarrow C$$

formulieren. Ich schwärme nicht sehr für diese Symbole, weil ich keinen Grund sehe, die ohnehin eher karge Sprache der Mathematik noch weiter auszutrocknen. Kompliziertere Sachverhalte werden in einer solchen Notation schnell undurchschaubar, und nach meinem Geschmack ist es schon an der Grenze des Erträglichen, den oben beschriebenen Sachverhalt als

$$(A \Longleftrightarrow B) \iff (A \Rightarrow B\,,\, B \Rightarrow A)$$

zu formulieren. In jedem Fall unzulässig ist es, Symbole als Abkürzungen für grammatische Satzteile zu mißbrauchen, etwa dies zu schreiben: A ist äquivalent zu $B \iff$ aus A folgt B und $B \Rightarrow A$.

Noch zum Inhalt des Satzes 1.16: Beachten Sie, daß erst die Gültigkeit des Satzes es erlaubt, für die Umkehrung einer bijektiven Abbildung f eine Bezeichnung einzuführen, und daß es dabei neben der Existenz- auch auf die Eindeutigkeitsaussage ankommt.

Eine kleine Problematik liegt noch in der neuen Bezeichnung f^{-1} für die Umkehrabbildung von $f\colon X \longrightarrow Y$. Wenn man für eine Teilmenge $B \subset Y$ jetzt $f^{-1}(B)$ hinschreibt — mit Klammern, etwa weil $B = B' \cap B''$ ist, so können damit zwei a priori verschiedene Mengen gemeint sein. Nämlich erst mal wie bisher das Urbild von B unter f, aber für bijektives f auch das Bild von B unter der Abbildung f^{-1}. Tatsächlich ist das aber beidesmal dieselbe Teilmenge von A, wie man aus den Definitionen sofort abliest. Die neue Bezeichnung ist also nicht nur logisch einwandfrei, sondern auch zweckmäßig gewählt.

Wichtig ist noch die folgende häufig benutzte Sprechweise:

1.17 Definition Sei X eine beliebige Menge und $A \subset X$ eine Teilmenge. Die durch den Pfeil mit rundem Schwanz bezeichnete Abbildung

$$A \hookrightarrow X; \quad a \mapsto a$$

heißt die Inklusionsabbildung von A nach X (oder kurz Inklusion von A in X). Ist $f\colon X \longrightarrow Y$ eine beliebige weitere Abbildung, so nennt man

$$f|A\colon A \longrightarrow Y; \quad a \mapsto f(a)$$

die Einschränkung von f auf A.

Notiz Selbstverständlich ist eine Inklusionsabbildung immer injektiv, und im Fall $A = X$ ist sie die identische Abbildung von X. Klar auch, daß die Einschränkung $f|A$ gerade die Komposition von f mit der Inklusion $A \hookrightarrow X$ ist. — Kann man das wohl auch machen, wenn A die leere Menge ist? Nun, wenn's nicht so wäre, dann hätte ich diesen Fall ausdrücklich ausschließen müssen. Denken Sie selbst mal darüber nach, was für Abbildungen $\emptyset \longrightarrow Y$, $X \longrightarrow \emptyset$ und $\emptyset \longrightarrow \emptyset$ es wohl gibt.

Sie werden von der Schule daran gewöhnt sein, Funktionen von $\mathbb{R}$ nach $\mathbb{R}$ graphisch darzustellen. Die dabei benutzte Konstruktion gibt in Wirklichkeit schon auf der abstrakten Ebene der Mengenlehre Sinn:

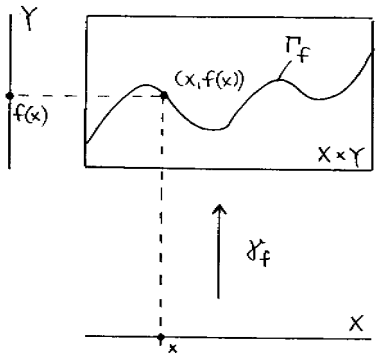
1.18 Definition Sei $f\colon X \longrightarrow Y$ eine Abbildung. Dann heißt

$$\gamma_f\colon X \longrightarrow X \times Y; \quad x \mapsto \big(x, f(x)\big)$$

die Graphenabbildung von f. Ihre Wertemenge, also

$$\Gamma_f := \gamma_f(X) = \big\{\big(x, f(x)\big) \,\big|\, x \in X\big\} = \big\{(x,y) \in X \times Y \,\big|\, f(x) = y\big\} \subset X \times Y$$

heißt der Graph von f.



Während im allgemeinen die Wertemenge einer Abbildung weniger Information enthält als die Abbildung selbst, kann man aus der Kenntnis des Graphen Γ_f die Abbildung f rekonstruieren. Einzelheiten dazu in Aufgabe 1.10.

Übungsaufgaben

1.1 Unter den folgenden sechs Aussagen:

(1) $\{x\} \subset M$
(2) $\{x\} \in M$
(3) $x \in M$
(4) $\{x\} \cap M = \emptyset$
(5) $\{x\} \setminus M = \emptyset$
(6) $M \setminus \{x\} = \emptyset$

sind einige nur verschiedene Beschreibungen ein und desselben Sachverhalts. Finden Sie heraus, welche das sind, und begründen Sie Ihre Antwort.

1.2 A, B, A' und B' seien Mengen. Beweisen Sie, daß

$$(A \times B) \setminus (A' \times B')$$

im allgemeinen nicht dasselbe ist wie

$$(A \setminus A') \times (B \setminus B').$$

Überlegen Sie sich dazu erst etwas Grundsätzliches: was muß man logischerweise tun, um zu zeigen, daß eine Aussage "im allgemeinen" falsch ist, d.h. hier: nicht für *jede* Wahl der Mengen $A, B \ldots$ zutrifft? — Wie besprochen kann eine auch noch so schöne Skizze nicht als Lösung der Aufgabe gelten, aber sie kann sehr nützlich sein, um auf die Lösung zu kommen.

Zeigen Sie weiter, daß man $(A \times B) \setminus (A' \times B')$ immer als Vereinigung zweier Mengen der Form $C \times D$ schreiben kann.

1.3 A, B, A' und B' seien Mengen. Untersuchen Sie, welche der Formeln

(1) $$(A \times B) \cap (A' \times B') = (A \cap A') \times (B \cap B')$$

(2) $$(A \times B) \cup (A' \times B') = (A \cup A') \times (B \cup B')$$

allgemein richtig sind: Die Richtigkeit jeder dieser Formeln ist also entweder für beliebige A, B, A', B' zu beweisen oder durch ein Gegenbeispiel zu widerlegen.

1.4 Sei $f\colon X \longrightarrow Y$ eine Abbildung und $A \subset X$ eine Teilmenge. Untersuchen Sie, ob $f^{-1}\big(f(A)\big)$ etwas mit A zu tun hat.

Es mag Sie stören, wenn hier nicht genau gesagt ist, was Sie eigentlich machen sollen. Aber diese Art der Fragestellung ist in der Wissenschaft durchaus praxisnah: man weiß ja in der Regel nicht im voraus, was herauskommt. Im übrigen werden Sie sicher eine Vermutung zu dieser Aufgabe haben; versuchen Sie diese zu beweisen (dann sind Sie fertig) oder zu widerlegen (was dann Anlaß zu einer neuen Vermutung wäre) ...

1.5 Sei $f\colon X \longrightarrow Y$ eine Abbildung; $A, A' \subset X$ und $B \subset Y$ seien Teilmengen. Zeigen Sie:

(1) Es gilt stets $f(f^{-1}B) \subset B$, im allgemeinen aber nicht $f(f^{-1}B) = B$.

(2) Es gilt stets $f(A \cap A') \subset f(A) \cap f(A')$, im allgemeinen aber nicht $f(A \cap A') = f(A) \cap f(A')$.

1.6 $f\colon X \longrightarrow Y$ sei eine Abbildung. Beweisen Sie: Jedes $x \in X$ liegt in einer Faser von f, und je zwei verschiedene Fasern von f sind disjunkt.

1.7 Geben Sie zwei Abbildungen $f\colon X \longrightarrow Y$ und $g\colon Y \longrightarrow X$ an, für die zwar $g \circ f = \mathrm{id}_X$, nicht aber $f \circ g = \mathrm{id}_Y$ gilt (natürlich sollen Sie das auch begründen, d.h. beweisen). Gibt es auch solche Beispiele, in denen außerdem $X = Y$ ist?

1.8 $f\colon X \longrightarrow Y$ und $g\colon Y \longrightarrow Z$ seien Abbildungen. Untersuchen Sie, welche der folgenden Aussagen allgemein richtig sind:

(1) Sind f und g injektiv, so ist $g \circ f$ injektiv.

(2) Ist $g \circ f$ injektiv, so ist f injektiv.

(3) Ist $g \circ f$ injektiv, so ist g injektiv.

1.9 $f\colon X \longrightarrow Y$ und $g\colon Y \longrightarrow Z$ seien Abbildungen. Untersuchen Sie, welche der folgenden Aussagen allgemein richtig sind:

(1) Sind f und g surjektiv, so ist $g \circ f$ surjektiv.

(2) Ist $g \circ f$ surjektiv, so ist f surjektiv.

(3) Ist $g \circ f$ surjektiv, so ist g surjektiv.

1.10 $f\colon X \longrightarrow Y$ sei eine beliebige Abbildung. Beweisen Sie: Die zugehörige Graphenabbildung γ_f ist injektiv, und die Abbildung $\pi := \mathrm{pr}_1 | \Gamma_f\colon \Gamma_f \longrightarrow X$ ist bijektiv. Ist $f'\colon X \longrightarrow Y$ eine weitere Abbildung mit $\Gamma_f = \Gamma_{f'}$, so ist $f = f'$.

2 Zahlen

Wir wollen uns jetzt mit den schon erwähnten Mengen von Zahlen $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ etwas näher befassen. Zahlen sind zum Rechnen da, und ich darf mich wohl darauf verlassen, daß Sie dieses Rechnen beherrschen. Ein paar grundsätzliche, teils auch in anderem Rahmen wichtige Tatsachen dazu möchte ich aber trotzdem hier kurz darstellen.

Wie ordnet sich das Rechnen mit Zahlen in die Wissenschaft von den Mengen und Abbildungen ein? Nun, dadurch daß auf den vier Mengen $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ je zwei grundlegende "Verknüpfungen"

$$\mathbb{N} \times \mathbb{N} \xrightarrow{+} \mathbb{N},\ (x, y) \mapsto x + y \quad \text{und} \quad \mathbb{N} \times \mathbb{N} \xrightarrow{\cdot} \mathbb{N},\ (x, y) \mapsto xy,$$

entsprechend

$$\mathbb{Z} \times \mathbb{Z} \xrightarrow{+,\cdot} \mathbb{Z},\ \mathbb{Q} \times \mathbb{Q} \xrightarrow{+,\cdot} \mathbb{Q},\ \mathbb{R} \times \mathbb{R} \xrightarrow{+,\cdot} \mathbb{R},$$

erklärt sind, die *Addition* bzw. *Multiplikation* heißen. Dabei ist "Verknüpfung" bloß das in diesem Zusammenhang gebräuchliche Wort für "Abbildung". Allerdings sind die Eigenschaften dieser Verknüpfungen von Fall zu Fall ganz verschieden.

2.1 Definition Eine Gruppe ist ein Paar aus einer Menge G und einer Verknüpfung

$$G \times G \longrightarrow G; \quad (x, y) \mapsto x \cdot y = xy$$

mit den Eigenschaften

(a) Assoziativität: $(xy)z = x(yz)$ für alle $x, y, z \in G$

(b) Existenz der Eins: es gibt ein Element $1 \in G$ mit

$$1x = x1 = x \quad \text{für jedes } x \in G$$

(c) Existenz des Inversen: zu jedem $x \in G$ gibt es ein Element $x^{-1} \in G$ mit

$$xx^{-1} = x^{-1}x = 1$$

Bemerkungen Das Element $1 \in G$, dessen Existenz in (b) verlangt wird, ist automatisch eindeutig bestimmt: sind $1'$ und $1''$ zwei solche Einselemente, so folgt sofort

$$1' = 1' \cdot 1'' = 1''.$$

Beachten Sie, daß erst diese Tatsache es erlaubt, ein festes Symbol, eben 1 dafür zu verwenden. Entsprechend sieht man leicht, daß zu gegebenem $x \in G$ nur *ein* Inverses existiert, was die Bezeichnung x^{-1} dafür rechtfertigt.

Von einer Gruppe G wird nicht allgemein auch das

(d) Kommutativgesetz: $xy = yx$ für alle $x, y \in G$

verlangt. Gilt es aber, so nennt man die Gruppe eben kommutativ oder abelsch. In einer abelschen Gruppe schreibt man statt x^{-1} auch $\frac{1}{x}$, und nur bei solchen Gruppen erlaubt man sich als Alternative, sie auch additiv zu schreiben, d.h. mit "+" als Verknüpfung und dann mit einem Nullelement "0" statt "1" und "$-x$" statt "x^{-1}". Natürlich sind dann alle Eigenschaften (a) bis (d) entsprechend zu lesen: $(x + y) + z = x + (y + z)$ usw., ohne daß das an ihrem Inhalt etwas ändern würde.

Bequemerweise schreibt man für eine Gruppe meist nur G statt $(G,+)$ oder $(G,\cdot)$, wenn klar ist, welche Verknüpfung gemeint ist.

2.2 Beispiele (1) $(\mathbb{Z},+)$ ist eine abelsche Gruppe, nicht aber $(\mathbb{N},+)$, denn zu $1\in\mathbb{N}$ gibt es kein $y\in\mathbb{N}$ mit $1+y=0$.

(2) $(\mathbb{Q},+)$ und $(\mathbb{R},+)$ sind abelsche Gruppen.

(3) $(\mathbb{Q}\backslash\{0\},\cdot)$ ist eine abelsche Gruppe, nicht aber $(\mathbb{Q},\cdot)$ (denn $0\in\mathbb{Q}$ besitzt kein Inverses) oder $(\mathbb{Z}\backslash\{0\},\cdot)$ (denn $2\in\mathbb{Z}\backslash\{0\}$ besitzt kein Inverses in $\mathbb{Z}\backslash\{0\}$).

(4) $(\mathbb{R}\backslash\{0\},\cdot)$ ist eine abelsche Gruppe.

(5) $(\{1,-1\},\cdot)$ ist eine abelsche Gruppe mit zwei Elementen.

2.3 Definition Sei G eine Gruppe. Eine Teilmenge $G'\subset G$ heißt eine Untergruppe von G, wenn die Verknüpfung $G\times G\longrightarrow G$ sich zu $G'\times G'\longrightarrow G'$ einschränken läßt und diese Einschränkung G' selbst zu einer Gruppe macht.

Bemerkungen Die erste, etwas salopp formulierte Bedingung verlangt, daß das Produkt zweier Elemente von G' wieder zu G' gehört.

Es ist leicht zu sehen, daß eine Untergruppe $G'\subset G$ zwangsläufig das Einselement von G enthalten muß und das dieses auch das Einselement von G' ist, so daß es sich erübrigt, die beiden durch eine besondere Notation auseinanderzuhalten. Entsprechendes gilt für das zu einem Element $x\in G'$ inverse. Einzelheiten dazu in Aufgabe 2.3.

Weitere Beispiele:

(6) Jede Gruppe $(G,\cdot)$ enthält als Untergruppen die *triviale* Untergruppe $\{1\}\subset G$ und natürlich auch $G\subset G$.

(7) $\mathbb{Z}\subset\mathbb{Q}\subset\mathbb{R}$ und $\{1,-1\}\subset\mathbb{Q}\backslash\{0\}\subset\mathbb{R}\backslash\{0\}$ sind Ketten von Untergruppen.

Gruppen sind die wichtigsten algebraischen Objekte mit *einer* Rechenoperation: Die Subtraktion in $(G,+)$ und die Division in $(G,\cdot)$ sieht man nicht als selbständige Operationen an, weil sie sich vermöge

$$x-y:=x+(-y)\quad\text{bzw.}\quad x/y:=xy^{-1}$$

aus der Addition bzw. Multiplikation von selbst ergeben. Nun zu der wichtigsten algebraischen Strukur mit zwei Verknüpfungen:

2.4 Definition Ein Ring besteht aus einer Menge R mit zwei Verknüpfungen

$$R\times R\longrightarrow R;\quad (x,y)\mapsto x+y;\quad (x,y)\mapsto x\cdot y=xy$$

mit folgenden Eigenschaften:

(a) $(R,+)$ ist eine abelsche Gruppe

(b) Assoziativität auch der Multiplikation: $(xy)z=x(yz)$ für alle $x,y,z\in R$

(c) Existenz der Eins: es gibt ein Element $1\in R$ mit

$$1x=x1=x\quad\text{für jedes } x\in R$$

(d) Distributivität: für alle $x,y,z\in R$ gilt

$$x(y+z)=xy+xz$$
$$(x+y)z=xz+yz$$

Ringe, in denen nicht nur die Addition, sondern auch die Multiplikation kommutativ ist, heißen kommutativ (bei Ringen sagt man aber nicht "abelsch"). Von einem Unterring $R'\subset R$ verlangt man, daß er das Einselement von R enthält (es folgt hier nicht automatisch).

Als Beispiel bietet sich an:

(8) $\mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$ sind nicht nur Untergruppen, sondern sogar Unterringe voneinander; alle drei Ringe sind außerdem kommutativ.

Was in einem Ring — kommutativ oder nicht — im allgemeinen nur eingeschränkt möglich ist, ist die Division. Weil für jedes Element x eines Ringes stets $0 \cdot x = 0$ gilt (warum?), wird man ohnehin nicht erwarten, daß man durch 0 teilen kann. Aber etwa im Ring $\mathbb{Z}$ kann man auch sonst nicht durch alle Elemente teilen, vielmehr nur durch ± 1. Die folgende Definition zeichnet eine spezielle Klasse von Ringen aus, die fast uneingeschränktes Teilen erlauben:

2.5 Definition Einen kommutativen Ring K nennt man einen Körper, wenn die Ringmultiplikation sich zu $K\setminus\{0\} \times K\setminus\{0\} \longrightarrow K\setminus\{0\}$ einschränken läßt und $K\setminus\{0\}$ so zu einer (abelschen) Gruppe macht.

Bemerkungen Konkret bedeutet das über die Ringeigenschaften hinaus, daß das Produkt zweier von Null verschiedener Körperelemente wieder von Null verschieden ist, daß $1 \neq 0$ gilt (was nicht schon aus den Ringaxiomen folgt) und daß jedes von Null verschiedene Element ein multiplikatives Inverses besitzt (das seinerseits zwangsläufig von Null verschieden ist).

Es besteht eigentlich kein logischer Grund, nicht-kommutative Ringe hier von vornherein auszuschließen. Das doch zu tun ist bloß eine Bequemlichkeit für diejenigen Bereiche der Mathematik, in denen die sogenannten *Schiefkörper*, die man sonst erhält, keine Rolle spielen.

(9) $\mathbb{Q}$ und $\mathbb{R}$ sind die naheliegenden Beispiele.

Gut, was ist daran für Sie überhaupt neu? Vielleicht am ehesten, daß ich grundlegende Eigenschaften der Ihnen an sich vertrauten Zahlbereiche $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ zu sogenannten *Axiomen* gemacht habe, mit denen die abstrakten Begriffe "Gruppe, Ring, Körper" erklärt sind. Es liegt im Wesen solcher Axiome, daß sie nicht etwa zu beweisen sind; sie dienen ja nur dazu, zu sagen, was eine Gruppe, ein Ring, ein Körper *ist*. Beweisbedürftig freilich wäre, daß die aus den Zahlbereichen $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ gebildeten Beispiele tatsächlich den Axiomen genügen, und das läuft letztlich darauf hinaus, diese Bereiche erst mal zu konstruieren. Das ist aber ein Punkt, aus dem für die Zwecke der Physik kein besonderer Gewinn zu ziehen und der übrigens auch für Mathematiker nicht so spannend ist. Wie gesagt gehe ich ja ohnehin davon aus, daß Sie das beherrschen, was man traditonell die "vier Grundrechnungsarten" nennt.

Ganz stiefmütterlich habe ich soweit die natürlichen Zahlen behandelt; zwar sind für sie die Verknüpfungen Addition und Multiplikation erklärt, aber $\mathbb{N}$ wird dadurch in keiner Weise zu einer Gruppe oder einem Ring, geschweige einem Körper. Dafür hat $\mathbb{N}$ eine andere interessante Besonderheit, es gilt nämlich das (offensichtliche)

2.6 Prinzip der vollständigen Induktion Es sei eine Folge

$$A_0,\ A_1,\ A_2,\ \ldots$$

von Aussagen gegeben. Wenn

$$A_0 \text{ wahr ist}$$

und

$$\text{für jedes } n \in \mathbb{N} \text{ aus } A_n \text{ die Aussage } A_{n+1} \text{ folgt,}$$

dann sind alle Aussagen $A_0,\ A_1,\ A_2,\ \ldots$ wahr.

2.7 Beispiel Wir wollen die Formel

$$1 + 2 + \cdots + n = \frac{n(n+1)}{2}$$

für alle $n \in \mathbb{N}$ beweisen (für $n = 1$ steht links natürlich die "Summe" mit dem einzigen Summanden 1, für $n = 0$ hört die Summation auf, bevor sie anfängt: damit ist die "leere" Summe 0 gemeint).

Wir wenden das Induktionsprinzip auf die Aussagen

$$A_n: \qquad 1+2+\cdots+n=\frac{n(n+1)}{2}$$

($n\in\mathbb{N}$) an: Zunächst ist aufgrund der Interpretation der leeren Summe

$$A_0: \qquad 0=\frac{0\cdot 1}{2}$$

eine offensichtlich wahre Aussage.

Für den noch fehlenden sogenannten Induktions*schluß* müssen wir aus der Induktions*annahme*, der Richtigkeit von A_n, die Richtigkeit von A_{n+1} folgern. Dazu schreiben wir

$$1+2+\cdots+n+(n+1)=(1+2+\cdots+n)+(n+1);$$

aufgrund der Induktionsannahme A_n ist die erste Klammer rechts

$$1+2+\cdots+n=\frac{n(n+1)}{2},$$

also

$$\begin{aligned}1+2+\cdots+n+(n+1)&=\frac{n(n+1)}{2}+(n+1)\\&=\frac{n(n+1)+2(n+1)}{2}\\&=\frac{(n+1)(n+2)}{2}\\&=\frac{(n+1)\big((n+1)+1\big)}{2},\end{aligned}$$

worin die letzte Umformung nur klarstellen soll, daß wir jetzt die Formel vor uns haben, deren Richtigkeit gerade der Inhalt der Aussage A_{n+1} ist. Beachten Sie, daß wir mit dem Induktionsschritt allein nicht etwa A_{n+1} bewiesen haben. Wir haben nur bewiesen, daß A_{n+1} eine Folgerung aus A_n ist. Erst das Prinzip der vollständigen Induktion erlaubt es, daraus in Verbindung mit dem Induktions*anfang* die Richtigkeit aller A_n zu schließen.

Übrigens könnte jemand auf die Idee kommen, unsere Summenformel auf ganz andere Art zu beweisen: In der offenbar richtigen Gleichung

$$\left.\begin{array}{cccccc}1 & +2 & +\cdots & +(n-1) & +n\\ +n & +(n-1) & +\cdots & +2 & +1\end{array}\right\}=2\cdot(1+\cdots+n)$$

addieren sich auf der linke Seite die untereinanderstehenden Zahlen zu $n+1$, also ist

$$n\cdot(n+1)=2\cdot(1+\cdots+n)$$

und damit

$$1+2+\cdots+n=\frac{n(n+1)}{2}.$$

Für einen mathematischen Satz reicht es natürlich völlig, *einen* richtigen Beweis zu haben. Hier haben wir nun zwei ganz verschiedene davon. Welcher ist besser? Keiner, aber jeder der beiden hat seine Vor- und Nachteile.

Die Stärke des Induktionsbeweises liegt darin, daß er eigentlich ganz automatisch abläuft, ohne daß man einen besonderen Trick hineinstecken müßte, und daß er in ähnlicher Form auch auf kompliziertere Probleme, etwa die Berechnung von $1^2+2^2+\cdots n^2$ anwendbar ist. Andererseits liefert er über den nackten Beweis hinaus wenig Einsicht darin, *warum* die bewiesene Formel richtig ist, ganz zu schweigen davon, daß man diese schon kennen, zumindest geraten haben muß, bevor man mit dem Beweis loslegen kann.

Beim zweiten Beweis ist es gerade umgekehrt: Die Formel ergibt sich aus dem Beweis ganz von selbst, aber der Beweis beginnt mit einer kleinen Idee, eben dem Trick, die Summe zweimal hinzuschreiben und die Summanden geschickt zusammenzufassen. Dadurch wird die Formel auch unmittelbar einleuchtend, und tatsächlich kann man sich diesen lustigen Trick viel leichter merken als die Formel selbst. Freilich: Schon bei der Auswertung von $1^2 + 2^2 + \cdots n^2$ scheitert die Methode, würde zumindest einen neuen, raffinierteren Trick erfordern.

Einer mathematischen Fragestellung sieht man in der Regel nicht ohne weiteres an, mit was für einer Methode man sie angreifen kann, und es gibt dafür kein Patentrezept. Doch, vielleicht dieses: sich nicht von vornherein auf einen Weg festzulegen, sondern geduldig die verschiedensten Ansätze zu verfolgen.

Noch zwei ganz einfache Anmerkungen zur vollständigen Induktion: Erstens kann man den Induktionsschritt, also den Schluß von A_n auf A_{n+1}, selbstverständlich auch als Schluß von A_{n-1} auf A_n für alle $n > 0$ formulieren, was manchmal — je nach verwendeter Notation — angenehmer hinzuschreiben ist. Die andere Anmerkung: Wenn man es mit Aussagen

$$A_m,\ A_{m+1},\ A_{m+2}, \ldots$$

für eine feste ganze (nicht unbedingt positive) Zahl m zu tun hat, kann man das Induktionsprinzip natürlich ganz analog verwenden, wobei der Induktionsanfang dann im Nachweis von A_m besteht.

Interessanter ist die folgende

2.8 Variante (des Prinzips der vollständigen Induktion) Seien A_0, A_1, A_2, ... Aussagen. Wenn für jedes $n \in \mathbb{N}$ aus der Richtigkeit von A_0, A_1, ..., A_{n-1} die von A_n folgt, dann sind alle Aussagen A_n ($n \in \mathbb{N}$) richtig.

Erläuterung Der scheinbar fehlende Induktionsanfang besteht in der Bemerkung, daß A_0 laut Voraussetzung aus der leeren Aussage folgt, d.h. aber eben, daß A_0 wahr ist. Die Gültigkeit der weiteren Aussagen A_1, A_2, ... liegt dann wieder auf der Hand.

Anläßlich unserer kleinen Summenformel wollen wir noch einige vereinfachende Notationen vereinbaren, die Sie wahrscheinlich auch schon kennen. Auf die Dauer kann es ja recht unpraktisch werden, für eine Summe mit einer großen oder nicht expliziten Anzahl von Summanden immer

$$x_1 + x_2 + \cdots + x_n$$

schreiben zu müssen, und wir führen stattdessen die Schreibweise mit dem *Summenzeichen*

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

ein. Sind allgemeiner $m, n \in \mathbb{Z}$ zwei ganze Zahlen mit $m \le n$, so vereinbaren wir

$$\sum_{i=m}^{n} x_i := x_m + x_{m+1} + \cdots + x_{n-1} + x_n.$$

Für $m = n$ ist damit einfach x_m gemeint, und wir ergänzen die Definition zweckmäßigerweise noch durch

$$\sum_{i=m}^{m-1} x_i = 0 \quad \text{(leere Summe)}.$$

Für $n < m-1$ wollen wir dem Symbol $\displaystyle\sum_{i=m}^{n}$ aber keinen Sinn geben.

Mit dem "i" in $\displaystyle\sum_{i=m}^{n} x_i$ verhält es sich genau wie mit dem "x" in $\{x \in X \mid x \ldots\}$. Es handelt sich nur um einen Platzhalter für den Zählindex, dessen Benennung ganz willkürlich ist:

$$\sum_{i=m}^{n} x_i = \sum_{j=m}^{n} x_j.$$

Wenn aus dem Zusammenhang unmißverständlich klar ist, über welche Indizes die Summe zu bilden ist, erlaubt man sich schon mal, bloß $\sum_i x_i$ oder gar (inkonsequenterweise) $\sum x_i$ zu schreiben.

Auf der Möglichkeit, den Summationsindex bei Bedarf umzubenennen, beruht ein häufig angewandter und manchmal Indexverschiebung genannter kleiner Trick.

2.9 Beispiel Wir wollen $\sum_{i=1}^{n} \frac{1}{i(i+1)}$ berechnen. Dazu bemerken wir

$$\frac{1}{i(i+1)} = \frac{(i+1)-i}{i(i+1)} = \frac{1}{i} - \frac{1}{i+1},$$

so daß wir

$$\sum_{i=1}^{n} \frac{1}{i(i+1)} = \sum_{i=1}^{n} \left(\frac{1}{i} - \frac{1}{i+1} \right) = \sum_{i=1}^{n} \frac{1}{i} - \sum_{i=1}^{n} \frac{1}{i+1}$$

zu berechnen haben. In der zweiten Summe können wir $i+1 =: j$ setzen, wobei j natürlich von 2 bis $n+1$ zu laufen hat, und damit ergibt sich schließlich

$$\sum_{i=1}^{n} \frac{1}{i(i+1)} = \sum_{i=1}^{n} \frac{1}{i} - \sum_{j=2}^{n+1} \frac{1}{j} = \sum_{i=1}^{n} \frac{1}{i} - \sum_{i=2}^{n+1} \frac{1}{i} = \frac{1}{1} - \frac{1}{n+1} = 1 - \frac{1}{n+1}.$$

Völlig analog zum Summenzeichen verwendet man das Produktzeichen $\prod$, nämlich

$$\prod_{i=m}^{n} x_i = x_m x_{m+1} \cdots x_{n-1} x_n$$

für $m \leq n$, während unter dem leeren Produkt zweckmäßigerweise

$$\prod_{i=m}^{m-1} x_i = 1$$

verstanden wird. Speziell sind die Potenzen einer beliebigen Zahl x oder allgemeiner eines beliebigen Elements eines Ringes R (mit nicht-negativem Exponenten $n \in \mathbb{N}$)

$$x^n = \prod_{i=1}^{n} x \in R,$$

insbesondere

$$x^0 = 1 \quad \text{für jedes } x$$

(auch für $x = 0$).

Für reelle Zahlen $x \neq 0$ (allgemeiner Elemente $x \neq 0$ eines Körpers K) ist es praktisch, auch Potenzen mit negativen (ganzen) Exponenten n einzuführen: sie sind dann durch

$$x^n := \left(\frac{1}{x} \right)^{-n}$$

erklärt (beachten Sie, daß dann ja $-n > 0$ und die rechte Seite deshalb schon definiert ist). Man überzeugt sich mit etwas Geduld (Fallunterscheidungen), aber ohne Schwierigkeiten davon, daß in jedem Fall die bekannten Formeln für die Potenzen

$$x^m \cdot x^n = x^{m+n}, \quad (x^m)^n = x^{mn} \quad \text{und} \quad x^n y^n = (xy)^n$$

für beliebige $m, n \in \mathbb{Z}$ und alle x, y gelten, für die beide Seiten überhaupt erklärt sind.

Anders als beim Rechnen mit Zahlen, also der Manipulation von Gleichungen, will ich mich nicht darauf verlassen, daß Sie im Umgang mit *Ungleichungen* so sicher sind, daß Ihnen dabei keine Fehler passieren. Ungleichungen beruhen auf der Möglichkeit, reelle Zahlen der Größe nach miteinander zu vergleichen; die Basis dafür sind die

2.10 Anordnungsaxiome Gewisse reelle Zahlen x heißen positiv ($x > 0$), und es gilt:

(a) Auf jede Zahl $x \in \mathbb{R}$ trifft genau eine der Aussagen

$$x > 0, \quad x = 0 \quad \text{oder} \quad -x > 0$$

zu.

(b) Aus $x, y \in \mathbb{R}, x > 0, y > 0$ folgt $x + y > 0$

(c) Aus $x, y \in \mathbb{R}, x > 0, y > 0$ folgt $xy > 0$

(d) Zu jedem $x \in \mathbb{R}$ gibt es ein $n \in \mathbb{N}$ mit $n - x > 0$ (sogenanntes *archimedisches Axiom*)

Wie jeder weiß, nennt man die Zahlen $x \in \mathbb{R}$ mit $-x > 0$ negativ. Außerdem erweitert man die Bedeutung des Zeichens "$>$" durch die Festsetzung

$$x > y :\Longleftrightarrow x - y > 0$$

und erlaubt sich, diesen Sachverhalt alternativ als $y < x$ zu schreiben.

Ebenso wie sich aus den Körperaxiomen die Regeln der Bruchrechnung ergeben, so folgen aus den Anordnungsaxiomen die Regeln für den Umgang mit der Anordnung:

2.11 Regeln Für alle $a, b, x, y, z \in \mathbb{R}$ gilt:

(e) Aus $x < y,\ y < z$ folgt $x < z$ (die Relation "$<$" ist *transitiv*, wie man sagt).

(f) Aus $a < b,\ x < y$ folgt $a + x < b + y$

(g) Aus $a > 0$ und $x < y$ folgt $ax < ay$, aus $a < 0$ und $x < y$ dagegen $ax > ay$

(h) Für jedes $x \neq 0$ ist $x^2 > 0$ (insbesondere $1 > 0$)

(i) Ist $x > 0$, so auch $x^{-1} > 0$; ist $x < 0$, so $x^{-1} < 0$

(j) Aus $0 < x < y$ folgt $x^{-1} > y^{-1}$

(k) Zu jedem $\varepsilon \in \mathbb{R}$ mit $\varepsilon > 0$ gibt es ein positives $n \in \mathbb{N}$ mit $\frac{1}{n} < \varepsilon$

Es ist keine Kunst, die Regeln (z.B. in dieser Reihenfolge) aus (a), (b), (c) und (d) abzuleiten. Eingehen möchte ich nur auf einige der multiplikativen Regeln, bei deren Anwendung man leicht Fehler macht. Etwa bedeutet (g), daß die Ungleichung

$$x < y$$

nach Multiplikation mit einer positiven Zahl erhalten bleibt:

$$ax < ay$$

sich für $a < 0$ dagegen umkehrt:

$$ax > ay$$

In der Tat ist in letzterem Fall ja $-a > 0$, also durch

$$\begin{aligned} (-a)x &< (-a)y && \mid \text{addiere } ax + ay \\ ay &< ax \end{aligned}$$

der zweite auf den ersten Fall zurückgeführt.

In diesem Sinne empfehle ich auch die Regel (j) Ihrer besonderen Beachtung, wo nämlich mit gutem Grund nicht nur $x < y$, sondern außerdem $x > 0$ vorausgesetzt ist. Wegen (j) ist (k) im wesentlichen bloß eine Umformulierung des archimedischen Axioms: Während jenes sicherstellt, daß jede (noch so große) reelle Zahl durch eine natürliche übertroffen wird, garantiert (k), daß jede (noch so kleine) positive reelle Zahl durch den Kehrwert einer natürlichen unterboten wird.

Obwohl man mit "$<$" und "$>$" an sich auskäme, ist es bequem, auch

$$x \le y \quad :\Longleftrightarrow \quad x < y \text{ oder } x = y$$

einzuführen (im Jargon der Mathematiker "kleinergleich" gesprochen), analog natürlich "$\ge$". Für die reellen Zahlen x mit $x \ge 0$ ($x \le 0$) hat man leider keinen farbigeren Namen als nicht-negativ bzw. -positiv.

Es mag Ihnen pedantisch vorkommen, auf dem Unterschied zwischen "$<$" und "$\le$" herumzureiten: Das Stück Kreide, das ich hier auf den Tisch lege, hat von der Tischplatte den Abstand 0 — jedenfalls makroskopisch gesehen. Aber unter einem (gedachten) Mikroskop? Da sehen wir Atome, die zum Tisch gehören und solche, die zur Kreide gehören, und dazwischen leeren Raum. Ist "Abstand null" physikalisch nicht dasselbe wie ein winzig kleiner, aber positiver Abstand?

Wie auch immer, Sie werden bald sehen, daß es in der Mathematik, die Sie als Physiker ja schließlich anwenden wollen, viele Situationen gibt, in der dieser kleine Unterschied ganz fundamental ist und geradezu den Clou der Sache ausmacht. Sie tun deshalb gut daran, es mit "$<$" und "$\le$" ebenso genau zu nehmen wie ich. Als kleine Übung könnten Sie sich etwa überlegen, welche der obigen Regeln auch mit "$\le$" statt "$<$" gelten (eventuell passend abgeändert).

Eine für die Arbeit mit reellen Zahlen sehr wichtige Folgerung aus den Regeln ist die

2.12 Bernoullische Ungleichung Für jedes $x \in \mathbb{R}$ mit $x \ge -1$ und jedes $n \in \mathbb{N}$ gilt

$$(1+x)^n \ge 1 + nx$$

Beweis, durch (vollständige) Induktion nach $n \ge 0$. Für $n = 0$ wird $1 \ge 1$ versprochen: stimmt. Zum Schluß von n auf $n+1$ wenden wir die Regeln an:

$$\begin{aligned}(1+x)^{n+1} &= (1+x)(1+x)^n \\ &\ge (1+x)(1+nx)\end{aligned}$$

wegen $1 + x \ge 0$ nach der "$\le$"-Version von (f) und der Induktionsannahme. Weiter ist nun

$$\begin{aligned}(1+x)(1+nx) &= 1 + (n+1)x + \underbrace{nx^2}_{\ge 0} \\ &\ge 1 + (n+1)x\end{aligned}$$

woraus mittels der Transitivität (e) die Induktionsbehauptung

$$(1+x)^{n+1} \ge 1 + (n+1)x$$

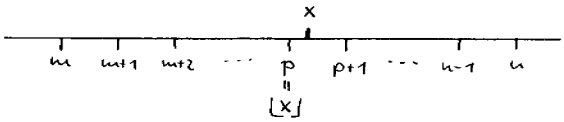
für $n+1$ folgt.

Bemerkungen Das, was wir eben gemacht haben, nämlich

$$(1+x)^{n+1} \ge \cdots \ge \cdots \ge \cdots \ge 1 + (n+1)x$$

heißt im Jargon der Analysis eine *Abschätzung*, mit der nämlich die Zahl $(1+x)^{n+1}$ durch die Zahl $1+(n+1)x$ *nach unten abgeschätzt* wird. Man meint also nicht etwa eine Näherung, wenn man sich so ausdrückt, sondern die Herleitung einer exakten Ungleichung. Meist geht es darum, einen verhältnismäßig komplizierten Ausdruck wie $(1+x)^{n+1}$ durch einen einfacheren, hier $1 + (n+1)x$, eben "abzuschätzen".

Übrigens beruht der bekannte Begriff des *größten Ganzen* $\lfloor x \rfloor$ einer reellen Zahl x, also derjenigen ganzen Zahl p mit $p \le x < p+1$, auf dem archimedischen Axiom: Dieses garantiert nämlich, daß x überhaupt zwischen zwei ganzen Zahlen m und n liegt;



man bestimmt $\lfloor x \rfloor$ dann leicht als eine der endlich vielen ganzen Zahlen m, $m+1$, ..., n.

Als Anwendung überlegen wir uns noch:

2.13 Satz Zu je zwei verschiedenen reellen Zahlen x, y, etwa $x < y$, gibt es eine rationale Zahl $r \in \mathbb{Q}$ mit:

$$x < r < y$$

Beweis Nach (k) wählen wir eine positive ganze Zahl q mit

$$\frac{1}{q} < y - x$$

und setzen

$$p := \lfloor qx \rfloor + 1.$$

Wir rechnen dann:

$$\begin{aligned} p - 1 &\le qx < p \\ \frac{p}{q} - \frac{1}{q} &\le x < \frac{p}{q}, \end{aligned}$$

also einerseits $x < \frac{p}{q}$, andererseits

$$\frac{p}{q} \le x + \frac{1}{q} < y,$$

und deshalb löst $r := \frac{p}{q}$ unser Problem.

In der Analysis, der Wissenschaft von den Grenzwerten, auf die wir jetzt zusteuern, hat man es häufig mit symmetrischen Ungleichungen vom Typ $-a < x < a$ zu tun. Diese drückt man bequemer mittels des Absolutbetrags einer reellen Zahl aus, der ganz primitiv erklärt ist:

2.14 Definition Der Absolutbetrag oder Betrag von $x \in \mathbb{R}$ ist als die reelle Zahl

$$|x| := \begin{cases} x & \text{für } x \ge 0 \\ -x & \text{für } x < 0 \end{cases}$$

definiert, die offenbar nicht-negativ ist.

Der Betrag genügt der einfachen, aber sehr wichtigen

2.15 Dreiecksungleichung Für beliebige $x, y \in \mathbb{R}$ gilt:

$$|x \pm y| \le |x| + |y|$$

Beweis Man kann die möglichen Fälle (je acht) direkt nachprüfen. Der Name "Dreiecksungleichung" kommt übrigens von einer allgemeineren, mehrdimensionalen Version und läßt sich auch erst an dieser überzeugend erklären.

Übungsaufgaben

2.1 Die Menge

$$\big\{f\colon\{1,2,3\}\longrightarrow\{1,2,3\}\,\big|\,f\text{ ist bijektive Abbildung}\big\}$$

wird durch die Komposition als Verknüpfung zu einer Gruppe (warum?), die wir mit Sym_3 bezeichnen; sie ist ein Beispiel einer sogenannten *symmetrischen Gruppe*. Zeigen Sie, daß Sym_3 nicht abelsch ist. Wieviele Elemente hat Sym_3?

2.2 Beweisen Sie, daß jede nicht-abelsche Gruppe G mindestens fünf Elemente enthalten muß.

2.3 Sei G eine Gruppe, $U\subset G$ eine Teilmenge. Beweisen Sie: U ist genau dann eine Untergruppe von G, wenn sie die folgenden beiden Eigenschaften hat:

- $U\neq\emptyset$,
- aus $x\in U$, $y\in U$ folgt $xy^{-1}\in U$.

Beachten Sie: Es ist nicht a priori klar, daß das Einselement $1\in G$ zugleich das Einselement einer jeden Untergruppe ist und daß das in U gebildete Inverse von $x\in U$ mit dem Inversen x^{-1} bezüglich G übereinstimmt. Jedoch ergeben sich diese beiden Tatsachen bei sorgfältiger Argumentation von selbst.

2.4 Beweisen Sie die für alle natürlichen Zahlen $m\leq n$ und alle reellen $x\neq 1$ gültige Formel

$$\sum_{j=m}^{n-1}x^j=\frac{x^m-x^n}{1-x}$$

durch vollständige Induktion. Finden Sie einen anderen Beweis, der ohne vollständige Induktion auskommt?

2.5 Im Beispiel 2.7 wurde durch vollständige Induktion die Formel $\sum_{i=1}^{n}i=\frac{n(n+1)}{2}\quad(n\in\mathbb{N})$ bewiesen. Was halten Sie von den folgenden drei Alternativvorschlägen für den Induktionsschluß von n auf $n+1$?

(1) Man rechnet

$$\sum_{i=1}^{n+1}i=1+\sum_{i=2}^{n+1}i=1+\sum_{j=1}^{n}(j+1)=1+\sum_{j=1}^{n}j+\sum_{j=1}^{n}1=1+\sum_{j=1}^{n}j+n$$

(Indexverschiebung) und drückt $\sum_{j=1}^{n}j$ mittels der Induktionsvoraussetzung aus:

$$\sum_{i=1}^{n+1}i=1+\frac{n(n+1)}{2}+n=\frac{2+n(n+1)+2n}{2}=\frac{(n+1)(n+2)}{2}$$

Damit hat man

$$\sum_{i=1}^{n+1}i=\frac{(n+1)\big((n+1)+1\big)}{2}$$

gezeigt und ist fertig.

(2) Die Induktionsvoraussetzung ist $\displaystyle\sum_{i=1}^{n}i=\frac{n(n+1)}{2}$, oder (Indexverschiebung):

$$\sum_{j=0}^{n-1}(j+1)=\frac{n(n+1)}{2}$$

Wir benennen nun n in m um:

$$\sum_{j=0}^{m-1}(j+1) = \frac{m(m+1)}{2}$$

und setzen dann $j+1 = i$ und $m-1 = n$:

$$\sum_{i=1}^{m} i = \frac{m(m+1)}{2}$$
$$\sum_{i=1}^{n+1} i = \frac{(n+1)(n+2)}{2}$$

Damit folgt wie oben:

$$\sum_{i=1}^{n+1} i = \frac{(n+1)\big((n+1)+1\big)}{2}$$

(3) Nach Induktionsannahme gilt $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$. Zum Schluß von n auf $n+1$ rechnet man nun

$$\begin{aligned}\sum_{i=1}^{n+1} i = \sum_{j=0}^{n}(j+1) &= \frac{1}{2}\sum_{j=0}^{n}(j+1) + \frac{1}{2}\sum_{j=0}^{n}(j+1)\\ &= \frac{1}{2}\sum_{j=0}^{n}(j+1) + \frac{1}{2}\sum_{k=0}^{n}(n-k+1)\\ &= \frac{1}{2}\sum_{j=0}^{n}(j+1) + \frac{1}{2}\sum_{j=0}^{n}(n-j+1)\end{aligned}$$

(Indexspiegelung $j = n-k$ und Umbenennung) und erhält durch Zusammenfassen wieder

$$\sum_{i=1}^{n+1} i = \frac{1}{2}\sum_{j=0}^{n}(j+1+n-j+1) = \frac{1}{2}\sum_{j=0}^{n}(n+2) = \frac{1}{2}(n+1)(n+2) = \frac{(n+1)\big((n+1)+1\big)}{2}$$

wie gewünscht.

2.6 In dieser Aufgabe sind $x_1, x_2, \ldots, x_n$ und x, y, a, b reelle Zahlen.

(a) Aus der Transitivität der Anordnung von $\mathbb{R}$ folgt:

$$x_1 \le x_2 \le \cdots \le x_n \Rightarrow x_1 \le x_n$$

Wann kann man sogar $x_1 < x_n$ schließen?

(b) Darf man

$$x < y \Rightarrow x^2 < y^2$$

schließen? Unter welcher zusätzlichen Voraussetzung doch?

(c) Darf man umgekehrt

$$x^2 < y^2 \Rightarrow x < y$$

schließen, oder muß man auch hier etwas Zusätzliches über x und/oder y wissen? Wie ist es mit dem Schluß

$$a < b \Rightarrow \sqrt{a} < \sqrt{b}$$

(für $a \ge 0$; sonst sind die Wurzeln nicht erklärt)?

(d) Warum darf man für $x \neq 0 \neq y$ nicht

$$x < y \Rightarrow \frac{1}{y} < \frac{1}{x}$$

schließen?

2.7 Bei der Menge

$$C := \left\{(x, y) \in \mathbb{R}^2 \,\middle|\, y^2 = x^3 + x^2\right\}$$

handelt es sich — anschaulich gesprochen — um eine Kurve in der Ebene. Konstruieren Sie eine "Parametrisierung" von C; damit soll hier eine Abbildung $\varphi\colon \mathbb{R} \longrightarrow \mathbb{R}^2$ gemeint sein, deren Bildmenge C ist und die vielleicht nicht ganz, aber doch fast injektiv ist (letzteres soll präzisiert, und beides genau bewiesen werden).

Tip: Die meisten Geraden durch den Nullpunkt treffen C in genau einem weiteren Punkt. Natürlich ist es hilfreich, sich zuerst eine Skizze von C zu machen.

2.8 Sie kennen die Definition von $\lfloor x \rfloor$ für eine reelle Zahl x. Ganz analog erklärt man $\lceil x \rceil \in \mathbb{Z}$ als diejenige ganze Zahl q mit $q - 1 < x \leq q$.

Wie kann man $\lceil x \rceil$ mit Hilfe von $\lfloor \quad \rfloor$ ausdrücken? Skizzieren Sie den Graphen Γ_f der durch $f(x) = 1 + \lfloor x \rfloor - \lceil x \rceil$ definierten Funktion $f\colon \mathbb{R} \longrightarrow \mathbb{R}$.

2.9 Beweisen Sie, daß für je zwei Zahlen $x, y \in \mathbb{R}$ die "umgekehrte" Dreiecksungleichung

$$|x + y| \geq \big||x| - |y|\big|$$

gilt. (Sie können wie beim Beweis der gewöhnlichen Dreiecksungleichung mit Fallunterscheidungen arbeiten; eleganter ist es, die neue Ungleichung auf die gewöhnliche zurückzuführen.)

3 Konvergente Zahlenfolgen

Mit diesem Abschnitt beginnen wir das Studium der Analysis. Den zentralen Begriff der Analysis, den Grenzwert oder Limes, gibt es in ungezählten, untereinander aber sehr ähnlichen Variationen. Wir lernen ihn zuerst als Grenzwert von reellen Zahlenfolgen kennen. Was ist überhaupt eine Zahlenfolge? Nun, dazu ganz allgemein die

3.1 Definition Sei Y eine Menge. Unter einer Folge in Y versteht man eine Abbildung

$$x\colon \mathbb{N} \longrightarrow Y,$$

die man aber meist nicht als $\mathbb{N} \ni n \mapsto x(n) \in Y$, sondern in einer der gleichwertigen Formen

$$(x_0, x_1, x_2, \ldots) = (x_n)_{n\in\mathbb{N}} = (x_n)_{n=0}^{\infty} = (x_n)_n$$

notiert, in letzterer (inkonsequenter) natürlich nur, wenn n im Zusammenhang keine weitere Bedeutung hat. Allgemeiner zieht man oft auch Folgen vom Typ $(x_n)_{n=m}^{\infty}$ in Betracht, wobei m eine feste ganze Zahl ist.

3.2 Definition Sei $(x_n)_{n=0}^{\infty}$ eine Folge reeller Zahlen (d.h. eine Folge in $\mathbb{R}$), und sei $a \in \mathbb{R}$. Man sagt, diese Folge konvergiert gegen a, wenn es zu jeder reellen Zahl $\varepsilon > 0$ ein $D \in \mathbb{N}$ gibt mit

$$|x_n - a| < \varepsilon \quad \text{für alle } n \in \mathbb{N} \text{ mit } n > D.$$

Schreibweise:

$$x_n \xrightarrow[n\to\infty]{} a \quad \text{oder} \quad \lim_{n\to\infty} x_n = a$$

oder kurz (aber inkonsequent)

$$\lim x_n = a.$$

Die Folge $(x_n)_n$ heißt dann konvergent (gegen a); sie heißt divergent, wenn es kein $a \in \mathbb{R}$ gibt, gegen das sie konvergiert.

3.3 Beispiele (1) Die Folge $(x_n)_{n=1}^{\infty} = \left(\frac{1}{n}\right)_{n=1}^{\infty}$ konvergiert gegen 0. Zu jedem $\varepsilon > 0$ gibt es nämlich nach Archimedes, genauer nach Regel 2.11(k), ein $D \in \mathbb{N},\ D > 0$ mit:

$$\frac{1}{D} < \varepsilon$$

Für alle $n \in \mathbb{N}$ mit $n > D$ gilt dann in der Tat

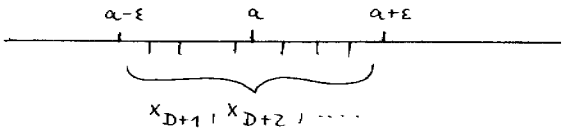
$$|x_n - 0| = \left|\frac{1}{n}\right| = \frac{1}{n} < \frac{1}{D} < \varepsilon,$$

wie es die Definition verlangt.

Zum besseren Verständnis der Definition 3.2 wollen wir für den Augenblick ein *festes* $\varepsilon > 0$ und ein *festes* $D \in \mathbb{N}$ nehmen. Die Forderung

$$|x_n - a| < \varepsilon \quad \text{für alle } n \in \mathbb{N} \text{ mit } n > D$$

besagt dann, daß alle Folgenglieder, angefangen mit x_{D+1}, nahe bei a liegen in dem Sinne, daß sie von a um weniger als ε entfernt sind:



(von $x_0, x_1, \dots, x_D$ wird nichts verlangt)

Die Raffinesse der Definition liegt nun darin, daß eben weder ε noch D fest sind, sondern daß für *jedes* $\varepsilon > 0$ die Existenz eines $D \in \mathbb{N}$ verlangt wird, derart, daß von x_{D+1} an alle Folgenglieder von a einen Abstand kleiner als ε haben. Typischerweise passiert dabei folgendes: Je kleiner $\varepsilon > 0$ vorgegeben ist, je schwieriger $|x_n - a| < \varepsilon$ also zu erfüllen ist, desto größer muß D ausfallen, weil dann ja weniger Folgenglieder bleiben, von denen etwas verlangt wird. Übrigens werden immer nur endlich viele von der Forderung ausgenommen, so groß D auch sein mag. In unserem Beispiel, wo wir D so gewählt haben, daß $\dfrac{1}{D} < \varepsilon$ ist, kann man den Effekt schön sehen.

In der Beispielsammlung nun zu

(2) Die Folge $(x_n)_{n=0}^{\infty} = \big((-1)^n\big)_{n=0}^{\infty}$, deren Glieder abwechseld 1 und -1 sind, divergiert. Konvergierte sie nämlich, etwa gegen $a \in \mathbb{R}$, so gäbe es $(\varepsilon := 1)$ ein $D \in \mathbb{N}$ mit

$$|x_n - a| < 1 \quad \text{für alle } n \in \mathbb{N} \text{ mit } n > D,$$

insbesondere (Dreiecksungleichung 2.15)

$$2 = |x_{n+1} - x_{n+2}| \le |x_{n+1} - a| + |a - x_{n+2}| < 1 + 1 = 2$$

— ein offensichtlicher Widerspruch.

Bemerkung Ein solcher *Widerspruchsbeweis* ist hier sehr zweckmäßig, weil das, was wir behaupten, nämlich die Divergenz der Folge, beweistechnisch schlecht zu packen ist. Deswegen nehmen wir im Gegenteil an, daß die Folge konvergiert, und beweisen, daß diese Annahme zu einem Widerspruch führt: Dann muß diese Annahme natürlich falsch gewesen sein!

Viele weitere Beispiele ergeben sich aus dem gleich folgenden Lemma. Dazu die

3.4 Definition Eine Menge reeller Zahlen $Y \subset \mathbb{R}$ heißt beschränkt, wenn es Zahlen $a, b \in \mathbb{R}$ gibt mit

$$a \le y \le b \quad \text{für alle } y \in Y.$$

Eine auf einer beliebigen Menge X definierte Funktion $f\colon X \longrightarrow \mathbb{R}$ heißt beschränkt, wenn ihre Wertemenge $f(X) \subset \mathbb{R}$ beschränkt ist; für eine Folge $(x_n)_{n=0}^{\infty}$ bedeutet das eben, daß die Menge $\{x_n \mid n \in \mathbb{N}\}$ der Folgenglieder beschränkt ist.

Bemerkungen Manchmal verfeinert man zu nach unten bzw. oben beschränkt, wenn eben statt $a \le y \le b$ nur die erste oder zweite Ungleichung verlangt wird. Die (natürlich nicht eindeutig bestimmten) Zahlen a und b nennt man untere und obere Schranken für Y bzw. f. Beispielsweise ist die Menge $\mathbb{N}$ der natürlichen Zahlen nach unten beschränkt (0, aber auch jede negative reelle Zahl ist eine untere Schranke). Daß $\mathbb{N}$ nicht auch nach oben beschränkt ist, ist gerade die Aussage des archimedischen Axioms 2.10(d): Kein $b \in \mathbb{R}$ kann eine obere Schranke für $\mathbb{N}$ sein, denn es gibt stets ein $n \in \mathbb{N}$ mit $n > b$. Die (gewöhnliche) Beschränktheit kann man auch symmetrisch formulieren:

3.4½ Notiz $Y \subset \mathbb{R}$ ist genau dann beschränkt, wenn es ein $c \in \mathbb{R}$ mit

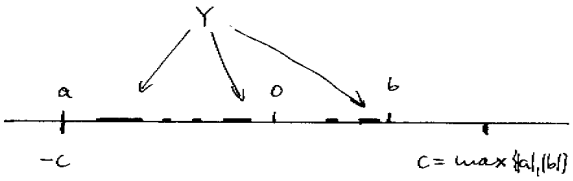
$$|y| \le c \quad \text{für alle } y \in Y$$

gibt.

Beweis Gibt es ein solches c, so ist $-c \le y \le c$ für alle $y \in Y$, und Y damit beschränkt. Ist umgekehrt $a \le y \le b$ für alle $y \in Y$ bekannt, so setze man $c := \max\{|a|, |b|\}$: Dann gilt

$$-c \le -|a| \le a \le y \le b \le |b| \le c,$$

also $|y| \le c$ für alle $y \in Y$.



Nun zu dem angekündigten

3.5 Lemma Jede konvergente Folge ist beschränkt.

Beweis Sei $(x_n)_{n=0}^{\infty}$ konvergent gegen a. Dann ($\varepsilon := 1$) können wir ein $D \in \mathbb{N}$ wählen mit

$$|x_n - a| < 1 \quad \text{für alle } n > D.$$

Wir setzen

$$c := \max\{|x_0|, |x_1|, \dots, |x_D|, |a|+1\};$$

dann ist $|x_n| \le c$ klar für $n \le D$, und für $n > D$ gilt es wegen

$$|x_n| = \big|(x_n - a) + a\big| \le |x_n - a| + |a| < 1 + |a| \le c$$

auch.

3.6 Satz Sei $x \in \mathbb{R}$ fest. Die *geometrische* Folge

$$(x_n)_{n=0}^{\infty} = (x^n)_{n=0}^{\infty}$$

hat das folgende Konvergenzverhalten:

$$\lim_{n\to\infty} x^n = 0 \quad \text{falls } |x| < 1,$$
$$\lim_{n\to\infty} x^n = 1 \quad \text{falls } x = 1;$$

für alle übrigen x divergiert die Folge.

Beweis Nach der Bernoullischen Ungleichung 2.12 ist

$$\bullet \qquad |x^n| = |x|^n = \big(1 + (|x| - 1)\big)^n \ge 1 + n(|x| - 1) > n(|x| - 1).$$

Ist $|x| > 1$ und etwa $c \in \mathbb{R}$ vorgegeben, so folgt

$$|x^n| > c,$$

sobald $n \ge \dfrac{c}{|x| - 1}$ ist: die Folge (x^n) ist dann also unbeschränkt und nach Lemma 3.5 erst recht divergent.

Die Divergenz der Folge $\big((-1)^n\big)$ haben wir schon erkannt, während $x = 1$ die konstante Folge $(1)_{n=0}^{\infty}$ ergibt, die in trivialer Weise gegen 1 konvergiert. Ebenso einfach ist der Fall $x = 0$: die Glieder der Folge $(0^n)_{n=0}^{\infty}$ sind bis auf das erste alle null.

Bleibt $0 < |x| < 1$ zu untersuchen. In unserer Abschätzung $\bullet$ ersetzen wir dazu x durch x^{-1} und erhalten

$$\frac{1}{|x^n|} = \left|\left(\frac{1}{x}\right)^n\right| > n\left(\left|\frac{1}{x}\right| - 1\right) = n\Big(\underbrace{\frac{1}{|x|} - 1}_{>0}\Big),$$

für $n > 0$ also:

$$|x^n| < \frac{1}{n} \cdot \left(\frac{1}{|x|} - 1\right)^{-1}$$

Ist nun $\varepsilon > 0$ vorgegeben, so gilt

$$|x^n| < \frac{1}{n} \cdot \left(\frac{1}{|x|} - 1\right)^{-1} \leq \varepsilon,$$

sobald $n \geq \left(\frac{1}{|x|} - 1\right)^{-1} \varepsilon^{-1}$ ist. Als D im Sinne der Konvergenzdefinition können wir also etwa das größte Ganze dieser letzten Zahl nehmen.

Der Konvergenzbegriff schließt nicht a priori aus, daß eine Folge gegen zwei verschiedene Grenzwerte konvergieren könnte. Tatsächlich ist das aber nicht möglich.

3.7 Lemma Der Grenzwert einer konvergenten Folge ist durch diese eindeutig bestimmt.

Beweis (x_n) konvergiere sowohl gegen a als auch gegen b. Für jedes $\varepsilon > 0$ gibt es also ein D mit

$$|x_n - a| < \frac{\varepsilon}{2} \quad \text{für alle } n > D$$

und ein E mit

$$|x_n - b| < \frac{\varepsilon}{2} \quad \text{für alle } n > E.$$

Indem wir D und E durch die größere der beiden Zahlen ersetzen, erreichen wir $D = E$. Für alle $n > D$ folgt dann

$$|a - b| \leq |a - x_n| + |x_n - b| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Da es solche n tatsächlich gibt (z.B. $n = D + 1$), schließen wir

$$|a - b| < \varepsilon \quad \text{für jedes } \varepsilon > 0,$$

was nur für $|a - b| = 0$, d.h. für $a = b$ möglich ist (wäre $|a - b| > 0$, könnten wir diese Zahl als ε wählen, und es ergäbe sich $\varepsilon < \varepsilon$).

Dieser letzte ganz putzige Schluß mag Sie erst mal irritieren: Kann man dann in der Konvergenzdefinition aus

$$|x_n - a| < \varepsilon \ldots$$

nicht auch $|x_n - a| = 0$ und damit $x_n = a$ für alle $n > D$ schließen? Nein, kann man nicht, weil es von D und damit letztlich von dem vorgegebenen ε abhängt, für *welche* n die Ungleichung $|x_n - a| < \varepsilon$ richtig ist.

So wichtig es ist, die Definition der Konvergenz genau zu verstehen, in den meisten Fällen wird man Konvergenz und Limes einer Zahlenfolge doch nicht direkt mit ε und D etablieren, sondern man wird sich möglichst auf fertige Regeln berufen, die einige häufig vorkommende Situationen schon abdecken.

3.8 Regeln $(x_n)_{n=0}^{\infty}$ und $(y_n)_{n=0}^{\infty}$ seien konvergente Zahlenfolgen mit Grenzwert a bzw. b. Dann gilt:

(a) Die Summenfolge $(x_n + y_n)_{n=0}^{\infty}$ konvergiert gegen $a + b$

(b) Die Produktfolge $(x_n y_n)_{n=0}^{\infty}$ konvergiert gegen ab

(c) Sei $b \neq 0$. Dann gibt es ein $m \in \mathbb{N}$ mit $y_n \neq 0$ für alle $n \geq m$, und die deshalb definierte Folge $\left(\frac{x_n}{y_n}\right)_{n=m}^{\infty}$ konvergiert gegen $\frac{a}{b}$

(d) Ist $x_n \leq y_n$ für alle $n \in \mathbb{N}$, so folgt $a \leq b$

Bemerkungen Wenn Sie sich z.B. (a) einfach als

$$\lim_{n\to\infty} (x_n + y_n) = \lim_{n\to\infty} x_n + \lim_{n\to\infty} y_n$$

merken wollen, ist das ganz in Ordnung, aber Sie sollten sich bei der Anwendung dieser Formel immer darüber im Klaren sein, daß die Existenz der Grenzwerte rechts vorausgesetzt wird (und die des linken dann folgt). Plumpes Beispiel dazu: Aus der verständnislos hingeschriebenen Formel

$$\lim_{n\to\infty}(n+1)=\lim_{n\to\infty}n+\lim_{n\to\infty}1$$

würde sofort

$$\lim_{n\to\infty}n=\lim_{n\to\infty}n+\lim_{n\to\infty}1,$$

also $0=1$ folgen. (Freilich ist der Einwand gegen die Limesformel nicht, daß etwas Absurdes herauskommt, sondern daß Regel (a) hier gar nicht anwendbar ist, weil es sich nicht um die Summe zweier konvergenter Folgen handelt.)

Die Pingeligkeit, mit der ich Regel (c) formuliert habe, geht in der saloppen Formulierung

$$\lim\frac{x_n}{y_n}=\frac{\lim x_n}{\lim y_n}\quad\text{falls }\lim y_n\neq 0$$

einfach unter. Das darf sie ruhig, da im Zusammenhang mit dem Limes ohnehin ein endliches Anfangsstück der Folge weggelassen oder auch beliebig abgeändert werden darf.

Schließlich sei davor gewarnt, analog zu (d) aus

$$x_n<y_n\quad\text{für alle }n\in\mathbb{N}$$

auf die *strenge* Ungleichung $a<b$ schließen zu wollen: Das Folgenpaar

$$(x_n)=(0)_{n=1}^{\infty}\quad\text{und}\quad(y_n)=\left(\frac{1}{n}\right)_{n=1}^{\infty}$$

illustriert, warum.

Beweis der Regeln Zu (a): Sei $\varepsilon>0$. Wie im Beweis von Lemma 3.7 finden wir ein D mit

$$|x_n-a|<\varepsilon\quad\text{und}\quad|y_n-b|<\varepsilon\quad\text{für alle }n>D.$$

Die Dreiecksungleichung gibt daraus

$$|(x_n+y_n)-(a+b)|=|(x_n-a)+(y_n-b)|\le|x_n-a|+|y_n-b|<2\varepsilon$$

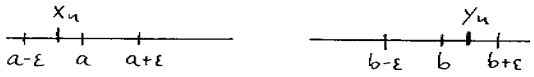
für alle $n>D$.

Damit wären wir fertig, wenn nicht der dumme Faktor 2 vor dem ε stünde! Aber den hätten wir vermeiden können, indem wir bei unserer Abschätzung gleich mit $\varepsilon/2$ statt ε anfangen. Statt das wirklich so zu machen, merken wir uns lieber, daß ein *fester* Faktor vor dem ε beim Nachweis der Konvergenz nicht stört. "Fest" bedeutet hier, daß der Faktor feststehen muß, bevor es im Beweis "Sei $\varepsilon>0$ ..." heißt.

Die Beweise zu (b) und (c) sind ähnlich.

Den Beweis von (d) führe ich aber noch vor: Sei $\varepsilon>0$. Wir finden ein D mit

$$|x_n-a|<\varepsilon\text{ und }|y_n-b|<\varepsilon\quad\text{für alle }n>D,$$



schließen daraus (für diese n)

$$a<x_n+\varepsilon\le y_n+\varepsilon<(b+\varepsilon)+\varepsilon$$

und weiter:

$$a-b<2\varepsilon$$

Nach dem schon in Lemma 3.7 geübten Schluß folgt $a - b \leq 0$, d.h. $a \leq b$.

Wir wollen die Limesregeln systematisch anwenden, um die Grenzwerte einer ganzen Klasse von Zahlenfolgen zu bestimmen. Die folgenden Begriffe dienen dazu, diese Klasse zu beschreiben; Sie werden sicher schon einmal von ihnen gehört haben.

3.9 Definition Eine Funktion $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ der Form

$$\mathbb{R} \ni t \mapsto \sum_{k=0}^{d} a_k t^k \in \mathbb{R}$$

mit Konstanten $a_0, a_1, \ldots, a_d \in \mathbb{R}$ nennt man ein (reelles) Polynom, diese Konstanten selbst die Koeffizienten des Polynoms. Wenn $d \in \mathbb{N}$ so gewählt ist, daß $a_d \neq 0$ ist, dann sagt man, f habe den Grad $\deg f := d$, nennt a_d den Leitkoeffizienten und das Polynom $t \mapsto a_d t^d$ den Leitterm von f.

Erläuterung Diese Definitionen machen implizit davon Gebrauch, daß f (zu gegebenem $d \in \mathbb{N}$) *nur eine* Darstellung der Form $\mathbb{R} \ni t \mapsto f(t) = \sum_{k=0}^{d} a_k t^k \in \mathbb{R}$ besitzt, eine Tatsache, auf die ich nach der Formulierung des folgenden Satzes zurückkomme. Übrigens ist die Abbildung $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ mit dem konstanten Wert 0 auch ein Polynom; für dieses, selbst kurz mit 0 bezeichnete *Nullpolynom* ist weder ein Grad noch ein Leitkoeffizient definiert.

3.10 Satz $f \neq 0$ und $g \neq 0$ seien zwei Polynome vom Grad d bzw. e:

$$f(t) = \sum_{k=0}^{d} a_k t^k \quad \text{und} \quad g(t) = \sum_{l=0}^{e} b_l t^l$$

mit $a_d \neq 0 \neq b_e$. Dann gilt:

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = \frac{a_d}{b_e} \quad \text{falls } d = e$$

und

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0 \quad \text{falls } d < e,$$

während die Quotientenfolge für $d > e$ divergiert.

Beweis In jedem Fall ziehen wir aus $f(n)$ und $g(n)$ für $n > 0$ erst mal den Faktor n^d bzw. n^e heraus:

$$f(n) = n^d \cdot \sum_{k=0}^{d} a_k n^{k-d} = n^d \left(a_d + a_{d-1}\frac{1}{n} + \cdots + a_0 \frac{1}{n^d} \right)$$
$$g(n) = n^e \cdot \sum_{l=0}^{e} b_l n^{l-e} = n^e \left(b_e + b_{e-1}\frac{1}{n} + \cdots + b_0 \frac{1}{n^e} \right)$$

Mit $\lim\limits_{n\to\infty} \frac{1}{n} = 0$ (Beispiel 3.3(1)) folgt durch wiederholte Anwendung der Regeln (beachten Sie $b_e \neq 0$) zunächst, daß $g(n) \neq 0$ für alle genügend großen n gilt und damit für diese n der Quotient

$$\frac{f(n)}{g(n)} = n^{d-e} \cdot \frac{a_d + a_{d-1}\frac{1}{n} + \cdots + a_0 \frac{1}{n^d}}{b_e + b_{e-1}\frac{1}{n} + \cdots + b_0 \frac{1}{n^e}}$$

überhaupt definiert ist. Es folgt weiter, daß der Bruch rechts den Limes $\frac{a_d}{b_e}$ hat.

Für $d = e$ ist das schon die Behauptung des Satzes, und für $d < e$ folgt sie mit

$$\lim n^{d-e} = \lim \frac{1}{n^{e-d}} = 0$$

aus der Produktregel (b).

Im Fall $d > e$ schließlich gilt das für den Kehrwert:

$$\lim_{n\to\infty} \frac{g(n)}{f(n)} = 0$$

Wäre in diesem Fall auch $\left(\frac{f(n)}{g(n)}\right)_n$ konvergent, etwa gegen $c \in \mathbb{R}$, so müßte nach der Produktregel

$$1 = \lim 1 = \lim \frac{f(n)g(n)}{g(n)f(n)} = \lim \frac{f(n)}{g(n)} \cdot \lim \frac{g(n)}{f(n)} = c \cdot 0 = 0$$

sein, was natürlich nicht stimmt. Also ist $\left(\frac{f(n)}{g(n)}\right)_n$ divergent wie behauptet.

Wie versprochen zurück zur Frage, warum die Koeffizienten eines Polynoms durch diese wohlbestimmt sind. Wir bemerken, daß Satz 3.10 davon nur zwecks bequemer Formulierung, nicht aber inhaltlich Gebrauch macht: er betrachtet eben die beiden durch

$$f(t) = \sum_{k=0}^{d} a_k t^k \quad \text{und} \quad g(t) = \sum_{l=0}^{e} b_l t^l \quad \text{mit } a_d \neq 0 \neq b_e$$

gegebenen Funktionen. Falls diese nun übereinstimmen, ist natürlich $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$, wir sind also im mittleren Fall der Satzaussage und schließen $d = e$ und $a_d = b_e$. Zwei Darstellungen desselben Polynoms müssen also denselben Leitterm haben, und indem wir diesen von beiden abziehen und vollständige Induktion anwenden, folgt weiter, daß auch die anderen einander entsprechenden Koeffizienten von f und g übereinstimmen. Das beweist die Behauptung und rechtfertigt damit alle in Definition 3.9 eingeführten Begriffe.

Zum Abschluß ein konkretes Beispiel dazu:

(3) Ohne überhaupt groß zu rechnen sieht man mittels Satz 3.10:

$$\lim_{n\to\infty} \frac{2n^3 + 4n^2 - n - 1}{\frac{1}{6}n(n+1)(2n+1)} = \frac{2}{2/6} = 6$$

Übungsaufgaben

3.1 $(x_n)_{n\in\mathbb{N}}$ sei eine reelle Zahlenfolge mit $\lim_{n\to\infty} x_n = a \in \mathbb{R}$. Zeigen Sie, daß dann $\lim_{n\to\infty} |x_n| = |a|$ gilt.

3.2 Berechnen Sie:

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{k=1}^{n} k \quad \text{und} \quad \lim_{n\to\infty} \left(n - \sqrt{n^2 - 1}\right)$$

Das ist im ersten Fall eher eine Routineangelegenheit, erfordert im zweiten dagegen einen kleinen Trick. Übrigens soll die Berechnung eines Grenzwertes immer auch einen Beweis der Konvergenz einschließen; der ergibt sich bei korrekter Anwendung der Regeln meist von selbst.

3.3 Beweisen Sie: Sind $(x_n)_{n=0}^{\infty}$ und $(z_n)_{n=0}^{\infty}$ zwei konvergente Zahlenfolgen mit demselben Grenzwert a, und ist $(y_n)_{n=0}^{\infty}$ eine weitere Zahlenfolge mit

$$x_n \le y_n \le z_n \text{ für alle } n \in \mathbb{N},$$

so ist auch $(y_n)_{n=0}^{\infty}$ konvergent, und es gilt $\lim\limits_{n\to\infty} y_n = a$. (Beachten Sie, daß die Konvergenz der Folge (y_n) ausdrücklich *nicht* vorausgesetzt wird.)

3.4 Benutzen Sie die das Ergebnis der Aufgabe 3.3, um die Grenzwerte

$$\lim_{n\to\infty} \sum_{j=1}^{n} \frac{n}{n^2+j} \quad \text{und} \quad \lim_{n\to\infty} \sum_{j=1}^{n} \frac{1}{n+\frac{1}{j}}$$

zu finden.

3.5. Definitionsgemäß bedeutet $\lim\limits_{n\to\infty} x_n = a$, daß es zu jedem $\varepsilon > 0$ ein $D \in \mathbb{N}$ gibt mit

$$|x_n - a| < \varepsilon \quad \text{für alle } n > D.$$

An drei Stellen kommt ein echtes Kleiner- oder Größerzeichen vor: Welche dieser Zeichen darf man durch "$\le$" bzw. durch "$\ge$" ersezten, ohne daß sich der Inhalt der Definition ändert?

3.6 Die Aussage $\lim\limits_{n\to\infty} x_n = a$ wird gern auf die folgende bequeme Art formuliert:

> Für jedes $\varepsilon > 0$ gilt $|x_n - a| < \varepsilon$ für alle $n \in \mathbb{N}$ bis auf endlich viele Ausnahmen.

Warum ist das tatsächlich korrekt? Worauf muß man aber achten?

4 Cauchy-Folgen

Mit den im letzten Abschnitt beschriebenen Methoden sind Sie in der Lage, konvergente Folgen in vielen Fällen als solche nachzuweisen und ihre Grenzwerte zu berechnen. Obwohl das sehr befriedigend klingt, liegt darin auch eine Beschränkung. Wir werden nämlich nicht damit zufrieden sein, Grenzwerte von Folgen als schon bekannte Zahlen zu erkennen, sondern wir wollen Folgen vor allem dazu verwenden, um neue, nicht auf einfachere Weise darstellbare Zahlen als Grenzwerte von Folgen erst zu *definieren*. Zu diesen neuen Zahlen gehören im Grunde genommen alle irrationalen, konkret etwa Werte der Exponential- und Logarithmus-, aber auch der trigonometrischen Funktionen Cosinus und Sinus, und viele mehr. Um den Folgenlimes in diesem Sinne einzusetzen, fehlt uns ein entscheidendes Werkzeug: Wir brauchen eine Methode, eine Folge als konvergent zu erkennen, ohne gleichzeitig von ihrem Grenzwert zu reden. Werfen Sie doch noch mal einen Blick auf die Definition des Limes (3.2): Wie sollte man die umformulieren, ohne auf den dort a genannten Limes Bezug zu nehmen? Dieses Problem werden wir jetzt lösen.

4.1 Lemma Sei $a \in \mathbb{R}$, und sei $(x_n)_{n=0}^{\infty}$ eine Zahlenfolge mit $\lim x_n = a$. Zu jedem $\varepsilon > 0$ gibt es dann ein $D \in \mathbb{N}$ mit

$$|x_m - x_n| < \varepsilon \quad \text{für alle } m, n \in \mathbb{N} \text{ mit } m > D,\ n > D.$$

Beweis Zu gegebenem $\varepsilon > 0$ finden wir wegen $\lim x_n = a$ ein $D \in \mathbb{N}$ mit $|x_n - a| < \varepsilon$ für alle $n > D$. Für alle $m, n \in \mathbb{N}$ mit $m > D,\ n > D$ folgt nach der Dreiecksungleichung:

$$|x_m - x_n| \le |x_m - a| + |a - x_n| < 2\varepsilon$$

Auch hier hätten wir den Faktor 2 offenbar vermeiden können, und mit dieser Bemerkung sind wir schon fertig.

Während die Konvergenz gegen a die Vorstellung präzisiert, daß die Folgenglieder x_n mit wachsendem n immer dichter an a rücken, ist die Aussage des Lemmas, daß die Folgenglieder mit wachsendem n *untereinander* immer dichter zusammenrücken.

4.2 Definition Eine Zahlenfolge $(x_n)_{n=0}^{\infty}$ heißt eine Cauchy-Folge, wenn es zu jedem $\varepsilon > 0$ ein $D \in \mathbb{N}$ gibt mit

$$|x_{n+k} - x_n| < \varepsilon \quad \text{für alle } k \in \mathbb{N} \text{ und alle } n > D.$$

Gegenüber der Schlußfolgerung von Lemma 4.1 habe ich hier nur die Formulierung etwas geändert: Es ist klar, daß die Rollen von m und n dort vertauschbar sind; deshalb darf ich die größere der beiden m nennen und dann als $n + k$ mit $k \in \mathbb{N}$ schreiben.

Das Lemma verspricht also, daß jede konvergente Folge eine Cauchy-Folge ist. Die Umkehrung ist nun auch richtig, aber nicht aus dem bisher Besprochenen beweisbar, sondern ein weiteres Axiom, das sogenannte

4.3 Vollständigkeitsaxiom Jede Cauchy-Folge reeller Zahlen konvergiert gegen einen reellen Grenzwert.

Bemerkung Zu beweisen gibt es da, wie gesagt, nichts. Interessant ist aber sich zu vergegenwärtigen, daß in allem, was ich Ihnen bis zu dieser Stelle vorgetragen habe, anstelle des Körpers der reellen Zahlen ebensogut der der rationalen hätte stehen können. Erst am Vollständigkeitsaxiom, das für $\mathbb{Q}$ nicht gilt, scheiden sich die beiden Körper $\mathbb{Q}$ und $\mathbb{R}$. Man kann zeigen, daß der Körper der reellen Zahlen dadurch charakterisiert ist, daß in ihm über die Anordnungsaxiome 2.10 hinaus auch das Vollständigkeitsaxiom gilt. Übrigens kann man auch beweisen, daß es den Körper $\mathbb{R}$ überhaupt gibt, was ja nicht selbstverständlich ist. Das ist aber eine

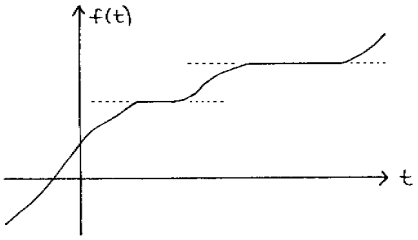
eher für Mathematiker interessante Angelegenheit, und ich denke, daß Sie an dieser Stelle zufrieden sind, wenn ich Ihnen hiermit die Existenz der reellen Zahlen einfach versichere.

Als Fazit merken Sie sich vor allem, daß konvergente Zahlenfolgen und Cauchy-Folgen ein und dasselbe sind. Damit ist genau das verfügbar geworden, was wir uns gewünscht hatten: Die in der Definition 4.2 formulierte *Cauchy-Eigenschaft* hat ja mit dem Grenzwert der Folge gar nichts zu tun.

Um das neue Werkzeug wirksam einsetzen zu können, sind zwei weitere Begriffe wichtig, die der Monotonie und der Teilfolge.

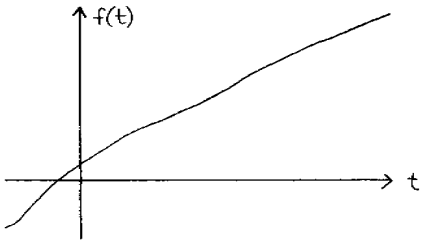
4.4 Definition Seien $X, Y \subset \mathbb{R}$ Teilmengen. Eine Funktion $f\colon X \longrightarrow Y$ heißt (streng) monoton wachsend, wenn

$$f(s) \le f(t) \quad \text{für alle } s,t \in X \text{ mit } s < t$$



bzw.

$$f(s) < f(t) \quad \text{für alle } s,t \in X \text{ mit } s < t$$



gilt. Speziell im Fall $X = \mathbb{N}$, also dem einer Folge $(x_n)_{n=0}^{\infty}$ genügt dafür natürlich

$$x_n \le x_{n+1} \quad \text{bzw.} \quad x_n < x_{n+1} \quad \text{für alle } n \in \mathbb{N}.$$

Analog fallende Monotonie.

4.5 Definition Seien $(x_n)_{n=0}^{\infty}$ eine Folge und $(n_k)_{k=0}^{\infty}$ eine streng monoton wachsende Folge natürlicher Zahlen ("Indizes"). Die Folge

$$(x_{n_k})_{k=0}^{\infty}$$

heißt dann eine Teilfolge von $(x_n)_{n=0}^{\infty}$.

Beachten Sie: Die Schreibweise mit dem "indizierten" Index x_{n_k} bedeutet $x_{(n_k)}$ und nicht $(x_n)_k$.

4.6 Beispiele (1) Ist $m \in \mathbb{N}$ eine feste Zahl und $n_k = m + k$ für alle k, so ist

$$(x_{n_k})_{k=0}^{\infty} = (x_{m+k})_{k=0}^{\infty}$$

die Teilfolge, die durch Weglassen der Anfangsglieder $x_0, x_1, \ldots, x_{m-1}$ entsteht.

(2) Für $n_k = 2k$ oder $n_k = 2k+1$ erhält man mit

$$(x_{2k})_{k=0}^{\infty} \quad \text{bzw.} \quad (x_{2k+1})_{k=0}^{\infty}$$

die Teilfolgen der geraden bzw. ungeraden Folgenglieder.

Es ist klar, daß man eine Teilfolge von $(x_n)_{n=0}^{\infty}$ auch dadurch beschreiben kann, daß man sagt, welche Indizes in der Teilfolge "vorkommen" sollen: Dazu gibt man eine beliebige unendliche Teilmenge

$$T \subset \mathbb{N}$$

an und definiert $n_k \in T$ als die der Größe nach k-te Zahl in T, beginnend mit der 0-ten. In den beiden Beispielen ist

$$T = \{n \in \mathbb{N} \,|\, n \geq m\}$$

bzw.

$$T = \{n \in \mathbb{N} \,|\, n \text{ (un-)gerade}\};$$

ein weiteres, bei dem diese Methode die praktischere ist, wäre

$$T = \{n \in \mathbb{N} \,|\, n \text{ Primzahl}\}. \tag{3}$$

Wenn man eine Teilfolge so beschreiben will, muß man selbstverständlich darauf achten, daß die Menge T wirklich unendlich ist.

4.7 Satz Jede reelle Zahlenfolge enthält eine monotone Teilfolge.

Beweis Ein kleines Juwel, weil sehr scharfsinnig und eigentlich doch ganz einfach:

Sei $(x_n)_{n=0}^{\infty}$ die gegebene Folge. Ad hoc wollen wir einen Index $m \in \mathbb{N}$ *bequem* nennen, wenn

$$x_m \geq x_n \quad \text{für alle } n \geq m$$

ist: das Folgenglied x_m wird dann von keinem Nachfolger übertroffen. Wir unterscheiden die beiden Fälle:

Fall 1: Es gibt unendlich viele bequeme Indizes. Diese definieren dann eine monoton fallende Teilfolge.

Fall 2: Es gibt nur endlich viele bequeme Indizes; diese seien alle kleiner als $m \in \mathbb{N}$. Dann definieren wir durch vollständige Induktion eine (streng) monoton wachsende Teilfolge $(x_{n_k})_{k=0}^{\infty}$ so:

$$\begin{aligned} n_0 &:= m; \\ n_{k+1} &:= \text{kleinste Zahl } n \in \mathbb{N} \text{ mit } n > n_k \text{ und } x_n > x_{n_k} \quad (k \in \mathbb{N}) \end{aligned}$$

Wegen $n_k \geq m$ ist n_k nämlich unbequem und n_{k+1} deshalb definiert.

Damit ist der Satz bewiesen.

Bemerkungen Der Beweis illustriert zugleich, wie man Folgen $(x_k)_{k=0}^{\infty}$ mittels vollständiger Induktion definieren kann, indem man nur das Anfangsglied x_0 direkt angibt, bei der Beschreibung von x_{k+1} aber die von x_k (oder auch aller vorangehenden Glieder) schon verwendet.

Für monotone Folgen erweist sich die Untersuchung auf Konvergenz nun als besonders einfach:

4.8 Satz Eine monotone Folge konvergiert genau dann, wenn sie beschränkt ist.

Beweis Wir wissen schon (aus Lemma 3.5), daß jede konvergente Folge (monoton oder nicht) beschränkt ist. Neu und nicht ganz einfach zu beweisen ist die umgekehrte Richtung. Wir setzen voraus, daß $(x_n)_{n=0}^{\infty}$ eine monotone Zahlenfolge ist, sagen wir eine monoton wachsende. Wir werden zeigen: Wenn diese Folge divergiert, d.h. wenn sie keine Cauchy-Folge ist, dann ist sie nicht nach oben beschränkt.

Zuerst wählen wir ein $\varepsilon > 0$, zu dem es *kein* $D \in \mathbb{N}$ gibt mit

$$x_{n+k} - x_n < \varepsilon \quad \text{für alle } k \in \mathbb{N} \text{ und alle } n > D;$$

die Betragstriche können wegen der Monotonie ja entfallen. Anders gesagt gibt es zu jedem $D \in \mathbb{N}$ natürliche Zahlen $n > D$ und k mit

$$x_{n+k} - x_n \geq \varepsilon.$$

Wegen $x_n \geq x_D$ impliziert die letzte Ungleichung, daß auch $x_{n+k} - x_D \geq \varepsilon$ ist, und wenn wir $l := n+k$ schreiben, ergibt sich insbesondere: Zu jedem $D \in \mathbb{N}$ existiert eine natürliche Zahl $l > D$ mit

$$x_l - x_D \geq \varepsilon.$$

Aufgrund dieser Tatsache können wir durch vollständige Induktion leicht eine Teilfolge $\left(x_{n_j}\right)_{j=0}^{\infty}$ mit der Eigenschaft

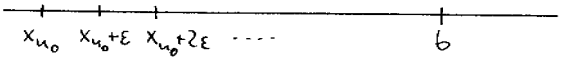
$$x_{n_{j+1}} - x_{n_j} \geq \varepsilon \quad \text{für jedes } j \in \mathbb{N}$$

konstruieren: Die Definition $n_0 = 0$ ist der Induktionsanfang. Im Induktionsschritt setzen wir n_j als schon definiert voraus. Die Wahl $D = n_j$ liefert dann ein $l > n_j$ mit $x_l - x_{n_j} \geq \varepsilon$, und wir setzen einfach $n_{j+1} := l$.

Es liegt auf der Hand, daß die so konstruierte Teilfolge $\left(x_{n_j}\right)_{j=0}^{\infty}$ nicht nach oben beschränkt ist. Addition der Ungleichungen liefert nämlich

$$x_{n_j} = \sum_{i=0}^{j-1} \left(x_{n_{i+1}} - x_{n_i}\right) + x_{n_0} \geq \sum_{i=0}^{j-1} \varepsilon + x_{n_0} = j\varepsilon + x_{n_0},$$

und wegen $\varepsilon > 0$ wird jede feste Zahl $b \in \mathbb{R}$ (nach Archimedes) von $j\varepsilon + x_{n_0}$ übertroffen, wenn $j \in \mathbb{N}$ nur genügend groß ist.



Wenn aber eine Teilfolge von $(x_n)_{n=0}^{\infty}$ nicht nach oben beschränkt ist, kann die Folge selbst es erst recht nicht sein.

Zur Vervollständigung des Beweises bleibt nur anzumerken, daß der Fall einer monoton fallenden Folge entweder analog behandelt oder (geschickter) auf den anderen zurückgeführt werden kann, indem man die Folge $(-x_n)_{n=0}^{\infty}$ betrachtet.

Nach dieser Anstrengung erholen wir uns bei einem

4.9 Beispiel Durch

$$x_0 = 1$$
$$x_{n+1} = \frac{x_n}{2} + \frac{1}{x_n} \quad \text{für } n \in \mathbb{N}$$

ist induktiv eine Folge positiver Zahlen $(x_n)_{n=0}^{\infty}$ definiert. Für jedes $n \in \mathbb{N}$ gilt nun $x_{n+1}^2 \geq 2$:

$$x_{n+1}^2 = \left(\frac{x_n}{2} + \frac{1}{x_n}\right)^2 = \left(\frac{x_n}{2} - \frac{1}{x_n}\right)^2 + 2 \geq 2,$$

und daraus ergibt sich weiter, daß die Folge $(x_n)_{n=1}^{\infty}$ monoton fällt: Für $n \geq 1$ ist

$$x_n - x_{n+1} = x_n - \left(\frac{x_n}{2} + \frac{1}{x_n}\right) = \frac{x_n}{2} - \frac{1}{x_n} = \frac{1}{2x_n}\left(x_n^2 - 2\right) \geq 0.$$

Satz 4.8 garantiert also, daß die reelle Zahl

$$w := \lim_{n\to\infty} x_n$$

existiert, und nach Regel 3.8(d) ist $w \geq 0$. Tatsächlich ist sogar $w > 0$, denn ebenfalls aufgrund der Limesregeln gilt ja

$$w^2 = \lim x_n^2 \geq \lim 2 = 2 > 0.$$

Können wir w auch berechnen? Nach dem, was ich eingangs dieses Abschnitts gesagt habe, vielleicht eher nicht, nämlich wenn wir mit w wirklich eine "neue" Zahl konstruiert haben. Aber dieser Zahl kommen wir leicht auf die Spur, wenn wir $\lim\limits_{n\to\infty} x_n = \lim\limits_{n\to\infty} x_{n+1}$ beachten: Abermals nach den Limesregeln folgt

$$w = \lim x_{n+1} = \lim\left(\frac{x_n}{2} + \frac{1}{x_n}\right) = \frac{w}{2} + \frac{1}{w}$$

und damit $\dfrac{w}{2} = \dfrac{1}{w}$, oder $w^2 = 2$. Wir haben in diesem Beispiel also die Zahl $\sqrt{2}$ als Grenzwert einer reellen Zahlenfolge dargestellt und damit insbesondere die Existenz dieser Quadratwurzel bewiesen.

Aus den beiden vorigen Sätzen ergibt sich nun ohne jede weitere Mühe der sehr wichtige

4.10 Satz von **Bolzano** und **Weierstraß** Jede beschränkte Folge enthält eine konvergente Teilfolge.

Beweis Die gegebene Folge enthält nach Satz 4.7 eine monotone Teilfolge; weil diese natürlich ebenfalls beschränkt ist, konvergiert sie nach Satz 4.8.

Es leuchtet ein, daß nicht etwa jede beschränkte Folge schon selbst konvergiert; denken Sie an die Folge $\big((-1)^n\big)_{n=0}^{\infty}$ aus Beispiel 3.3(2). Obwohl der Satz von Bolzano und Weierstraß nicht explizit verrät, *welche* Teilfolgen einer beschränkten Folge konvergieren, wird er sich noch als außerordentlich nützlich erweisen. In dem eben genannten Beispiel freilich kann man sofort sagen, welches die konvergenten Teilfolgen sind, oder?

Zum Schluß dieses Abschnitts will ich Ihnen noch eine andere Anwendung der Cauchy-Folgen vorstellen. Betrachten wir etwa die Menge

$$X := \{t \in \mathbb{R} \,|\, 0 \le t \le 1\}$$

(ein sogenanntes *Invervall*). Es ist offensichtlich, daß X ein kleinstes und ein größtes Element enthält, nämlich

$$\min X = 0 \quad \text{und} \quad \max X = 1.$$

Die ebenfalls beschränkte Menge

$$Y := \{t \in \mathbb{R} \,|\, 0 < t < 1\}$$

dagegen enthält weder ein kleinstes noch ein größtes Element: Wäre etwa $t \in Y$ das kleinste, so würde es wegen $0 < \frac{t}{2} < t$ sofort durch $\frac{t}{2}$ unterboten! Immerhin ist klar, daß die Zahlen 0 und 1 die größtmögliche untere bzw. die kleinstmögliche obere Schranke der Menge Y sind. Für beliebige beschränkte Mengen $Y \subset \mathbb{R}$ ist die Existenz solcher bestmöglicher Schranken aber eine beweisbedürftige Behauptung. Wir präzisieren:

4.11 Satz und **Definition** Sei $Y \subset \mathbb{R}$ eine nicht-leere nach oben beschränkte Menge. Dann gibt es unter allen oberen Schranken für Y eine kleinste, die (natürlich eindeutig bestimmt ist und) das Supremum

$$\sup Y \in \mathbb{R}$$

von Y heißt. Entsprechend ist das Infimum

$$\inf X \in \mathbb{R}$$

einer nach unten beschränkten nicht-leeren Menge $X \subset \mathbb{R}$ als deren größte untere Schranke erklärt.

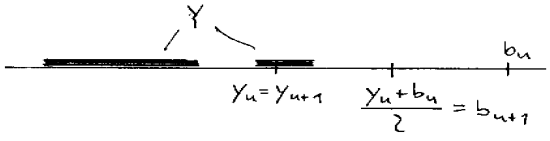
Beweis Es genügt, den Fall einer nach oben beschränkten Menge $Y \neq \emptyset$ anzusehen. Wir wählen zuerst willkürlich ein $y_0 \in Y$ und eine obere Schranke $b_0 \in \mathbb{R}$ für Y. Jetzt konstruieren wir durch vollständige Induktion eine monoton wachsende Folge $(y_n)_{n=0}^{\infty}$ und eine monoton fallende Folge $(b_n)_{n=0}^{\infty}$ mit den drei Eigenschaften

$$y_n \in Y,$$
$$b_n \text{ ist obere Schranke von } Y$$
$$b_n - y_n \le \frac{b_0 - y_0}{2^n}$$

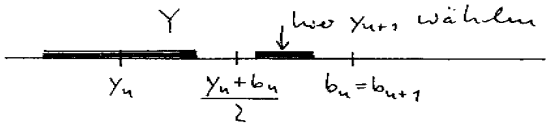
für alle $n \in \mathbb{N}$.

Der Induktionsanfang ist schon gemacht, und aus y_n und b_n gewinnen wir y_{n+1} und b_{n+1} durch die folgende Vorschrift:

- wenn $\dfrac{y_n + b_n}{2}$ noch obere Schranke von y ist, dann sei $y_{n+1} := y_n$, $b_{n+1} := \dfrac{y_n + b_n}{2}$



- wenn nicht, dann sei $b_{n+1} := b_n$, und y_{n+1} sei ein Element von Y mit $\dfrac{y_n + b_n}{2} < y_{n+1}$



Die Abschätzung für $b_{n+1} - y_{n+1}$ ergibt sich auf je nach Fall verschiedene Weise aus der Induktionsannahme:

$$b_{n+1} - y_{n+1} = \frac{y_n + b_n}{2} - y_n = \frac{b_n - y_n}{2} \le \frac{1}{2}\frac{b_0 - y_0}{2^n} = \frac{b_0 - y_0}{2^{n+1}}$$

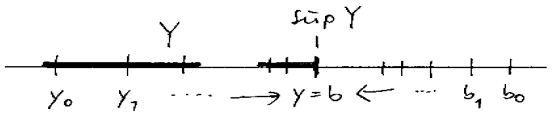
bzw.

$$b_{n+1} - y_{n+1} < b_n - \frac{y_n + b_n}{2} = \frac{b_n - y_n}{2} \le \frac{1}{2}\frac{b_0 - y_0}{2^n} = \frac{b_0 - y_0}{2^{n+1}}$$

Nun ist die Folge $(y_n)_{n=0}^{\infty}$ nach oben beschränkt (durch b_0); ebenso ist $(b_n)_{n=0}^{\infty}$ nach unten beschränkt (durch y_0). Nach Satz 4.8 sind also beide konvergent, etwa mit

$$\lim y_n = y \in \mathbb{R} \quad \text{und} \quad \lim b_n = b \in \mathbb{R}.$$

Tatsächlich müssen beide Grenzwerte gleich sein, denn wegen $y_n \le b_n$ für alle $n \in \mathbb{N}$ ist einerseits $y \le b$, wegen $b_n - y_n \le \dfrac{b_0 - y_0}{2^n}$ andererseits $b - y \le \lim\limits_{n\to\infty} \dfrac{b_0 - y_0}{2^n} = 0$, d.h. $y \ge b$.



Ich behaupte, daß $y = b$ die kleinste obere Schranke, also das Supremum von Y ist. Erstens gilt für jedes $t \in Y$

$$t \le b_n \quad \text{für alle } n \in \mathbb{N}$$

und damit auch $t \le b$; also ist b jedenfalls eine obere Schranke von Y. Ist aber $c \in \mathbb{R}$ eine weitere obere Schranke, so drehen wir den Spieß einfach herum: Aus

$$y_n \le c \quad \text{für alle } n \in \mathbb{N}$$

folgt dann $b = y \le c$, und damit ist bewiesen, daß b die kleinste unter den oberen Schranken ist.

Aus dem Beweis halten wir noch fest den

4.12 Zusatz Ist $Y \subset \mathbb{R}$ nicht-leer und nach oben beschränkt, so gibt es eine monoton wachsende Folge in Y, die gegen $\sup Y$ konvergiert.

Verwechseln Sie trotzdem nicht: Den Limes bildet man von Folgen, Supremum und Infimum dagegen von Mengen reeller Zahlen.

Übungsaufgaben

4.1 Zeigen Sie, daß durch

$$x_0 := 2 \quad \text{und} \quad x_{n+1} := 2 - \frac{1}{x_n} \quad (n \in \mathbb{N})$$

eine Folge reeller Zahlen $(x_n)_n$ mit Werten zwischen 1 und 2 erklärt wird; zeigen Sie weiter, daß diese Folge konvergiert, und berechnen Sie $\lim\limits_{n\to\infty} x_n$.

4.2 Sei $I := \{t \in \mathbb{R} \,|\, 0 < t < 1\}$. Zeigen Sie: Durch die Formel

$$x_{n+1} := 1 - (1-x_n)^2 \quad (n \in \mathbb{N})$$

wird für jede Wahl des Anfangswertes $x_0 \in I$ eine Zahlenfolge in I erklärt; zeigen Sie weiter, daß diese Folge konvergiert, und berechnen Sie $\lim\limits_{n\to\infty} x_n$.

4.3 Ist $T \subset \mathbb{N}$ eine endliche Menge, so ist $\mathbb{N} \setminus T$ natürlich unendlich. Wenn $T \subset \mathbb{N}$ aber unendlich ist, kann $\mathbb{N} \setminus T$ endlich oder unendlich sein: Belegen Sie das durch Beispiele. Zeigen Sie, daß man $\mathbb{N}$ sogar als Vereinigung *unendlich vieler* paarweise disjunkter unendlicher Teilmengen schreiben kann; in Formeln

$$\mathbb{N} = \bigcup_{\lambda\in\Lambda} T_\lambda = \{t \,|\, \text{es gibt ein } \lambda \in \Lambda \text{ mit } t \in T_\lambda\} \quad \text{und} \quad T_\lambda \cap T_\mu = \emptyset \text{ für } \lambda \neq \mu$$

mit Mengen Λ und T_λ $(\lambda \in \Lambda)$, die alle unendlich sind.

4.4 Sei $(x_n)_{n=0}^\infty$ eine Folge mit $\lim\limits_{n\to\infty} x_n = a$, und sei $\mathbb{N} \ni k \mapsto n_k \in \mathbb{N}$ eine Bijektion (*Permutation*, wie man auch sagt). Beweisen Sie, daß dann auch die entsprechend "umgeordnete" Folge $(x_{n_k})_{k=0}^\infty$ gegen a konvergiert:

$$\lim_{k\to\infty} x_{n_k} = a$$

Tip: Denken Sie an Aufgabe 3.6.

4.5 Man könnte auf die Idee kommen, die Cauchy-Eigenschaft analog zu Aufgabe 3.6 durch die Forderung

Für jedes $\varepsilon > 0$ gilt $|x_m - x_n| < \varepsilon$ für alle Paare $(m, n) \in \mathbb{N} \times \mathbb{N}$ bis auf endlich viele Ausnahmen

zu beschreiben. Zeigen Sie, daß dadurch aber in Wirklichkeit die konstanten Zahlenfolgen charakterisiert werden.

4.6 $A, B \subset \mathbb{R}$ seien zwei nicht-leere nach oben beschränkte Teilmengen; es sei $s = \sup A$ und $t = \sup B$. Zeigen Sie, daß auch die Menge

$$A + B := \{x+y \,|\, x \in A \text{ und } y \in B\}$$

nach oben beschränkt ist und daß $\sup(A + B) = s + t = \sup A + \sup B$ gilt.

5 Reihen

An sich ist mit den Ergebnissen des vorigen Abschnitts der Weg frei, um interessante reelle Zahlen und Funktionen als Grenzwerte von konvergenten Folgen zu konstruieren. In der Praxis schreibt man die in Betracht kommenden Folgen allerdings meist lieber in einer etwas anderen, wenn auch grundsätzlich völlig gleichwertigen Form, nämlich als *Reihen*.

5.1 Definition Eine (unendliche) Reihe reeller Zahlen ist logisch gesehen dasselbe wie eine reelle Zahlenfolge; man schreibt statt $(x_n)_{n=0}^\infty$ aber eine symbolische Summe:

$$\sum_{n=0}^{\infty} x_n \quad \text{oder auch} \quad x_0 + x_1 + \cdots$$

Damit deutet man an, daß mit der Konvergenz dieser Reihe etwas Neues gemeint ist, nämlich die Konvergenz der zugehörigen Partialsummenfolge

$$\left(\sum_{n=0}^{m} x_n\right)_{m=0}^{\infty}$$

Im Falle der Konvergenz spricht man dann auch nicht vom Limes der Reihe, sondern von der Reihensumme

$$\sum_{n=0}^{\infty} x_n := \lim_{m\to\infty} \sum_{n=0}^{m} x_n \in \mathbb{R}$$

Daß man in diesem Fall das Symbol $\sum_{n=0}^\infty$ in zwei wesentlich verschiedenen Bedeutungen (als Bezeichnung für die Reihe selbst und für ihre Summe) verwendet, ist ein durch die Tradition gerechtfertigter Mißbrauch.

5.2 Beispiel Sei $x \in \mathbb{R}$. Die *geometrische Reihe*

$$\sum_{n=0}^{\infty} x^n$$

konvergiert genau dann, wenn $|x| < 1$ ist, und hat dann die Summe

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}.$$

Beweis Hier läßt sich die zugehörige Partialsummenfolge direkt ausrechnen:

$$\sum_{n=0}^{m} x^n = \frac{1-x^{m+1}}{1-x} \quad \text{für } x \neq 1$$

Für $|x| < 1$ hat die geometrische Folge nach Satz 3.6 den Grenzwert $\lim\limits_{m\to\infty} x^{m+1} = 0$, und wie behauptet folgt die Konvergenz der Reihe:

$$\sum_{n=0}^{\infty} x^n = \lim_{m\to\infty} \sum_{n=0}^{m} x^n = \lim_{m\to\infty} \frac{1-x^{m+1}}{1-x} = \frac{1}{1-x}$$

Die Divergenz der Reihe in allen anderen Fällen ergibt sich (wieder mit Satz 3.6) sofort aus dem gleich folgenden Lemma. Zuerst wollen wir aber die Tatsache, daß die konvergenten Folgen gerade die Cauchy-Folgen sind, in die für Reihen passende Formulierung bringen.

5.3 Cauchy-Kriterium für Reihen Die Reihe $\sum_{n=0}^{\infty} x_n$ konvergiert genau dann, wenn gilt:

Zu jedem $\varepsilon > 0$ gibt es ein $D \in \mathbb{N}$ mit:

$$\left| \sum_{n=m+1}^{m+k} x_n \right| < \varepsilon \quad \text{für alle } m > D \text{ und alle } k \in \mathbb{N}$$

Beweis $\sum_{n=0}^{m+k} x_n - \sum_{n=0}^{m} x_n = \sum_{n=m+1}^{m+k} x_n$

5.4 Lemma Die Reihe $\sum_{n=0}^{\infty} x_n$ kann nur dann konvergieren, wenn

$$\lim_{n\to\infty} x_n = 0$$

ist.

Beweis Die Reihe $\sum_n x_n$ sei konvergent. Nach dem Cauchy-Kriterium gibt es zu jedem $\varepsilon > 0$ ein D mit:

$$\left| \sum_{n=m+1}^{m+k} x_n \right| < \varepsilon \quad \text{für alle } m > D \text{ und alle } k \in \mathbb{N}$$

Wenn wir speziell $k = 1$ setzen, bleibt davon bloß

$$|x_{m+1}| < \varepsilon \quad \text{für alle } m > D$$

und damit $\lim x_m = 0$.

Daß man die Logik des Lemmas nicht umdrehen darf, sieht man an dem zweiten wichtigen

5.5 Beispiel Die *harmonische Reihe*

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

divergiert, obwohl ihre Glieder gegen 0 konvergieren.

Beweis Die für jedes $m > 0$ gültige Abschätzung

$$\sum_{n=m+1}^{2m} \frac{1}{n} \geq \sum_{n=m+1}^{2m} \frac{1}{2m} = m \cdot \frac{1}{2m} = \frac{1}{2}$$

zeigt, daß das Cauchy-Kriterium verletzt ist ($\varepsilon = \frac{1}{2}$ und $k = m$).

Bemerkungen Wie bei jeder Reihe mit nicht-negativen Gliedern wächst hier die Partialsummenfolge monoton. Nach dem Satz 4.8 über die Konvergenz monotoner Folgen bedeutet die Divergenz einer solchen Reihe also, daß ihre Partialsummenfolge nicht nach oben beschränkt ist. Die harmonische Reihe ist ein schönes

Beispiel dafür, daß Konvergenz oder Divergenz selbst einer einfach gebauten Reihe nicht immer schon durch ein paar Tests mit dem Taschenrechner zu erkennen sind:

$$\sum_{n=1}^{100} \frac{1}{n} \approx 5.19 \quad \text{und selbst} \quad \sum_{n=1}^{10000} \frac{1}{n} \approx 9.79$$

Die Reihe divergiert, aber sehr langsam oder "schwach".

Natürlich möchte man mit den Summen konvergenter Reihen auch rechnen können. Ich stelle Ihnen dazu wieder einige einfache Regeln zusammen.

5.6 Regeln (a) Wie bei Folgen zu interpretieren sind:

$$\sum_{n=0}^{\infty}(x_n + y_n) = \sum_{n=0}^{\infty} x_n + \sum_{n=0}^{\infty} y_n$$
$$\sum_{n=0}^{\infty} \lambda x_n = \lambda \sum_{n=0}^{\infty} x_n$$

(b) Hinzufügen, Weglassen oder Ändern endlich vieler Reihenglieder hat keinen Einfluß auf das Konvergenzverhalten einer Reihe (im allgemeinen aber auf die Reihensumme).

(c) Einfügen oder Weglassen beliebig vieler Nullen ändert nichts.

(d) In einer konvergenten Reihe darf man Glieder durch "Klammernsetzung" beliebig zusammenfassen, ohne daß Kovergenz oder Reihensumme dadurch gestört würden:

$$\sum_{n=0}^{\infty} x_n = \underbrace{x_0 + \cdots + x_{n_0-1}}_{y_0} + \underbrace{x_{n_0} + \cdots + x_{n_1-1}}_{y_1} + \underbrace{x_{n_1} + \cdots + x_{n_2-1}}_{y_2} + \cdots$$
$$= \sum_{k=0}^{\infty} y_k$$

Beweis (a) folgt aus den Regeln 3.8 über konvergente Folgen.

Die in (b) genannten Prozesse bewirken nur eine Verschiebung der Partialsummenfolge um die Bilanz der Änderungen — abgesehen von endlich vielen Gliedern am Anfang.

In (c) bewirkt das Einfügen von Nullen bloß, daß die entsprechenden Glieder der Partialsummenfolge wiederholt werden, und das Entfernen von Nullen das Gegenteil, daß nämlich aus einer "stotternden" Partialsummenfolge Wiederholungen entfernt werden.

Schließlich ist in (d) die neue Partialsummenfolge eine Teilfolge der alten.

Angesichts der Regeln drängt sich die Frage auf, inwieweit man mit der Reihensumme $\sum_{n=0}^{\infty}$ überhaupt so rechnen darf, als wäre es eine wirkliche Summe, d.h. Summe endlich vieler reeller Zahlen: zweifellos soll die Schreibweise das ja nahelegen. Nun, sicher kann man etwa die Regel (d) nicht einfach umkehren: Die trivialerweise konvergente Reihe

$$\sum_{n=0}^{\infty} 0 = \sum_{n=0}^{\infty}(1-1) = (1-1) + (1-1) + \cdots$$

wird durch Weglassen der Klammern zu der ebenso offensichtlich divergenten Reihe

$$\sum_{n=0}^{\infty}(-1)^n = 1 - 1 + 1 - 1 \pm \cdots$$

Die durch $x_0 = 1$ und $x_n = 0$ $(n > 0)$ definierte Reihe mit Summe 1

$$\sum_{n=0}^{\infty} x_n = 1 + (-1 + 1) + (-1 + 1) + \cdots$$

aber auch! Es besteht aber kein Anlaß, daraus auf $0 = 1$ zu schließen, denn es gibt keine Regel, die das Rechnen mit diesen divergenten Reihen erlauben würde (beachten Sie, daß die Gleichheitszeichen in diesen Zeilen für die Gleichheit der Reihen selber, nicht für die von Reihensummen stehen).

Tiefer liegt die Tatsache, daß für konvergente Reihen $\sum_{n=0}^{\infty} x_n$ das Kommutativgesetz nicht immer gilt. Das wollen wir uns genauer ansehen. Sei

$$\mathbb{N} \ni k \mapsto n_k \in \mathbb{N}$$

eine bijektive Abbildung; man nennt so etwas auch eine *Vertauschung* oder *Permutation*. Gilt dann

$$\sum_{n=0}^{\infty} x_n = \sum_{k=0}^{\infty} x_{n_k}$$

wenigstens dann, sagen wir, wenn beide Reihen konvergieren? Wer spontan meint, das müsse immer so sein, hat wahrscheinlich eine zu enge Vorstellung von den Permutationen $k \mapsto n_k$: Gewiß gibt es deren solche, die nur endlich viele Indizes bewegen, zu denen also ein $D \in \mathbb{N}$ mit

$$n_k = k \text{ für alle } k > D$$

existiert, und als Umordnung einer konvergenten Reihe sind diese Permutationen nach Regel 5.6(b) harmlos. Aber nicht jede Permutation ist von dieser Art; betrachten wir dazu das

5.7 Beispiel Die sogenannte *alternierende harmonische Reihe*

$$\sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{5} \pm \cdots$$

ist im Gegensatz zur gewöhnlichen harmonischen Reihe konvergent (nach dem in der Übungsstunde zu besprechenden Lemma von Leibniz), und für ihre Reihensumme s gilt sicher

$$s = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n} = 1 - \frac{1}{2} + \sum_{k=2}^{\infty} \left(\frac{1}{2k-1} - \frac{1}{2k} \right) \geq \frac{1}{2} > 0.$$

Durch erlaubte Operationen ergibt sich:

$$\begin{array}{rcccccccccccccc}
s & = & 1 & -\frac{1}{2} & +\frac{1}{3} & -\frac{1}{4} & +\frac{1}{5} & -\frac{1}{6} & +\frac{1}{7} & -\frac{1}{8} & +\frac{1}{9} & -\frac{1}{10} & +\frac{1}{11} & -\frac{1}{12} & \pm\cdots \\
\frac{1}{2}s & = & 0 & +\frac{1}{2} & +0 & -\frac{1}{4} & +0 & +\frac{1}{6} & +0 & -\frac{1}{8} & +0 & +\frac{1}{10} & +0 & -\frac{1}{12} & \pm\cdots \\
\hline
\frac{3}{2}s & = & 1 & & +\frac{1}{3} & -\frac{1}{2} & +\frac{1}{5} & & +\frac{1}{7} & -\frac{1}{4} & +\frac{1}{9} & & +\frac{1}{11} & -\frac{1}{6} & \pm\cdots
\end{array}$$

In der ersten Zeile steht die alternierende harmonische Reihe; die Reihe in der zweiten ist aus ihr durch Multiplikation mit $1/2$ nach Regel (a) und Einschieben von Nullen nach Regel (c) gebildet, während die dritte Reihe durch Addition konvergenter Reihen nach Regel (a) entsteht, wobei die auftretenden Nullen nach Regel (c) gleich wieder weggelassen sind. Wenn Sie einen Moment überlegen, stellen Sie fest, daß verblüffenderweise in dieser neuen Reihe genau dieselben Glieder vorkommen wie in der alternierenden harmonischen Reihe selbst, daß diese beiden Reihen also durch eine Permutation der Reihenglieder auseinander hervorgehen oder, wie man sagt, Umordnungen voneinander sind. In diesem Fall gilt das Kommutativgesetz also nicht, wegen $s > 0$ ist ja $\frac{3}{2}s \neq s$.

Man kann übrigens zu jeder vorgegebenen reellen Zahl eine Umordnung der alternierenden harmonischen Reihe konstruieren, die genau diese Zahl als Reihensumme hat, und es gibt auch divergente Umordnungen.

Auf der anderen Seite ist die Möglichkeit, unendliche Reihen beliebig umordnen zu können, von großer theoretischer wie praktischer Bedeutung; man *möchte* das einfach machen. Was tun, wenn man aber doch

nicht darf, wie wir eben gesehen haben? Man versucht, herauszufinden, *welche* Reihen man straflos umordnen darf. Die Konvergenz dieser Reihen ist dann "stärker" oder "robuster" als die gewöhnliche eben in dem Sinne, daß sie sich von Umordnungen nicht stören läßt.

Erfreulicherweise ist ein solcher Konvergenzbegriff ganz leicht zu beschreiben. Woran hat es eigentlich gelegen, daß wir die harmonische Reihe so umordnen konnten, daß die Reihensumme dabei größer wurde? Daran, daß wir die negativen Reihenglieder immer mehr nach hinten geschoben haben, so daß die Partialsummen schneller wachsen konnten als bei der ursprünglichen Reihe! Wenn man diesen Effekt verhindern will, muß man bei der Fassung des neuen Konvergenzbegriffs jedenfalls zu verhindern suchen, daß Konvergenz nur deswegen eintritt, weil relativ große positive Beiträge zu den Partialsummen durch vergleichbar große negative kompensiert werden. Genau das leistet der folgende Begriff.

5.8 Definition Eine Reihe $\sum_{n=0}^{\infty} x_n$ heißt absolut konvergent, wenn die Reihe

$$\sum_{n=0}^{\infty} |x_n|$$

konvergiert.

5.9 Notizen (a) Die Partialsummenfolge der Reihe $\sum_{n=0}^{\infty} |x_n|$ wächst monoton. Nach dem, was wir aus Satz 4.8 über solche Folgen wissen, ist die absolute Konvergenz von $\sum_{n=0}^{\infty} x_n$ gleichbedeutend damit, daß die Folge

$$\left(\sum_{n=0}^{m} |x_n|\right)_{m=0}^{\infty}$$

nach oben beschränkt ist.

(b) Summen und Vielfache absolut konvergenter Reihen sind wieder absolut konvergent, ersteres wegen der Abschätzung

$$\sum_{n=0}^{m} |x_n + y_n| \le \sum_{n=0}^{m} |x_n| + \sum_{n=0}^{m} |y_n|$$

nach der Dreiecksungleichung.

(c) Für Reihen mit nicht-negativen Gliedern bringt der Begriff gegenüber der gewöhnlichen Konvergenz nichts Neues.

(d) Für beliebige Reihen folgt aus der absoluten Konvergenz die gewöhnliche. Sei nämlich $\sum_{n=0}^{\infty} x_n$ absolut konvergent, also $\sum_{n=0}^{\infty} |x_n|$ konvergent. Nach dem Cauchy-Kriterium 5.3 gibt es dann zu jedem $\varepsilon > 0$ ein D mit

$$\left|\sum_{n=m+1}^{m+k} |x_n|\right| < \varepsilon \quad \text{für alle } m > D \text{ und alle } k \in \mathbb{N},$$

erst recht also

$$\left|\sum x_n\right| \le \sum |x_n| < \varepsilon,$$

und wieder nach dem Cauchy-Kriterium folgt die Konvergenz von $\sum_{n=0}^{\infty} x_n$.

Wir kennen schon ein paar Beispiele: Die alternierende harmonische Reihe konvergiert, aber nicht absolut. Die geometrische Reihe $\sum_{n=0}^{\infty} q^n$ konvergiert für $|q| < 1$ wegen $|q^n| = |q|^n$ sogar absolut.

Im Umgang mit Reihen ist die absolute Konvergenz von ungleich größerer Bedeutung als die gewöhnliche. Wie findet man heraus, ob eine Reihe absolut konvergiert? Das folgende Kriterium ist ganz simpel, aber extrem wichtig, weil es ohne große Mühe eine Unmenge absolut konvergenter Reihen liefert.

5.10 Majoranten- und **Minorantenkriterium** Sei $\sum_{n=0}^{\infty} \mu_n$ eine Reihe mit nicht-negativen Gliedern, und sei $\sum_{n=0}^{\infty} x_n$ eine beliebige Reihe. Dann gilt:

Konvergiert $\sum \mu_n$ und gilt

$$|x_n| \le \mu_n \quad \text{für alle } n \in \mathbb{N}$$

(ist also, wie man sagt, $\sum \mu_n$ eine konvergente *Majorante* von $\sum x_n$), so konvergiert die Reihe $\sum x_n$ absolut.

Divergiert dagegen $\sum \mu_n$ und gilt

$$|x_n| \ge \mu_n \quad \text{für alle } n \in \mathbb{N}$$

(ist $\sum \mu_n$ eine divergente *Minorante* von $\sum x_n$), so konvergiert die Reihe $\sum x_n$ nicht absolut (während sie im gewöhnlichen Sinne konvergieren oder divergieren kann).

Beweis Ganz einfach: Im ersten Fall ist für jedes $m \in \mathbb{N}$

$$\sum_{n=0}^{m} |x_n| \le \sum_{n=0}^{m} \mu_n,$$

und rechts, also auch links stehen die Glieder einer nach oben beschränkten Folge. Im zweiten Fall ist es gerade umgekehrt:

$$\sum_{n=0}^{m} |x_n| \ge \sum_{n=0}^{m} \mu_n,$$

und rechts, also auch links stehen die Glieder einer nach oben nicht beschränkten Folge.

5.11 Beispiel Sei $q \in \mathbb{R}$ fest mit $0 < q < 1$. Die geometrische Reihe $\sum_n q^n$ konvergiert dann, und für jede positive reelle Zahl c konvergiert die etwas allgemeinere Reihe

$$c \sum_{n=0}^{\infty} q^n = \sum_{n=0}^{\infty} cq^n$$

deshalb nach Notiz 5.9(b) auch. Sei nun $(\lambda_n)_{n=0}^{\infty}$ eine beschränkte, ansonsten aber ganz beliebige reelle Zahlenfolge. Dann konvergiert die Reihe

$$\sum_{n=0}^{\infty} \lambda_n q^n$$

absolut, denn wenn etwa $|\lambda_n| \le c$ für alle $n \in \mathbb{N}$ ist, kann man die Reihe $\sum cq^n$ als Majorante nehmen:

$$|\lambda_n q^n| \le cq^n \quad \text{für alle } n \in \mathbb{N}$$

Auf dieser Tatsache beruht übrigens die Darstellung von reellen Zahlen als (im allgemeinen nicht abbrechenden) Dezimalbrüchen. Wenn wir in Dezimaldarstellung der — sagen wir nicht-negativen — reellen Zahl x

$$m.\lambda_1\lambda_2\lambda_3\ldots \quad \text{mit } m \in \mathbb{N} \text{ und } 0 \le \lambda_n < 10 \text{ für } n = 1, 2, 3, \ldots$$

schreiben, so meinen wir damit, daß x die Summe der nach dem eben gesagten konvergenten Reihe

$$m + \sum_{n=1}^{\infty} \lambda_n \left(\frac{1}{10}\right)^n$$

ist (setze $q = 1/10$ und $c = 9$).

Das Instrument Majoranten-/Minorantenkriterium ist natürlich um so schärfer, je mehr Vergleichsreihen mit bekanntem Konvergenzverhalten (eben Majoranten und Minoranten) zur Verfügung stehen. Wenn man es nicht gerade mit ganz esoterischen Reihen zu tun hat, kommt man für die Untersuchung einer vorgelegten

Reihe auf Konvergenz mit den folgenden paar Schritten in der Regel aus; diese sollten Sie aber auch wirklich beherrschen.

(1) Wenn Sie die absolute Konvergenz der Reihe $\sum x_n$ vermuten, versuchen Sie eine geometrische Reihe

$$\sum_n q^n \quad \text{oder auch} \quad \sum_n cq^n \ \ (c > 0 \text{ fest})$$

mit $0 < q < 1$ als Majorante. Weil die Majorantenbedingung

$$|x_n| \le q^n$$

dasselbe bedeutet wie $\sqrt[n]{|x_n|} \le q$, ist diese Methode bekannt unter dem irreführenden Namen

5.12 Wurzelkriterium Sei $\sum_{n=0}^{\infty} x_n$ eine Reihe, und q eine reelle Zahl mit

$$0 \le q < 1.$$

Gilt dann

$$\sqrt[n]{|x_n|} \le q \quad \text{für alle } n \in \mathbb{N},$$

so konvergiert $\sum x_n$ absolut.

Bemerkung Wie gut Sie Ihr Gefühl für die Begriffe der Analysis inzwischen geschult haben, können Sie daran testen, ob Ihnen unmittelbar einleuchtet, daß die hier formulierte Bedingung *nicht* dasselbe wie einfach

$$\sqrt[n]{|x_n|} < 1 \quad \text{für alle } n \in \mathbb{N}$$

ist.

Oft genügt statt dem Wurzelkriterium in der Form 5.12 schon die folgende etwas speziellere, aber bequemere

5.13 Variante Sei $\sum_{n=0}^{\infty} x_n$ eine Reihe. Gilt dann

$$\lim_{n\to\infty} \sqrt[n]{|x_n|} < 1,$$

so konvergiert $\sum x_n$ absolut.

Beweis Sei $a := \lim_{n\to\infty} \sqrt[n]{|x_n|}$, und sei

$$q := \frac{a+1}{2}$$

der Mittelwert; dann ist $0 \le a < q < 1$. Nach Definition des Limes gibt es ($\varepsilon := q - a$) ein D mit

$$\left| \sqrt[n]{|x_n|} - a \right| < q - a,$$

insbesondere

$$\sqrt[n]{|x_n|} \le \left| \sqrt[n]{|x_n|} - a \right| + a < q \quad \text{für alle } n > D.$$

Jetzt kann man sich auf 5.12 berufen: Daß die Ungleichung für $n \le D$ vielleicht nicht erfüllt ist, macht aus dem bekannten Grunde nichts.

(2) Wenn Ihnen die Anwendung des Wurzelkriterium eben wegen der darin auftretenden n-ten Wurzel zu umständlich erscheint, prüfen Sie stattdessen nach dem etwas schwächeren, aber noch einfacheren

5.14 Quotientenkriterium Sei $\sum_{n=0}^{\infty} x_n$ eine Reihe mit $x_n \ne 0$ für alle $n \in \mathbb{N}$. Wenn es eine reelle Zahl q mit $0 \le q < 1$ und

$$\frac{|x_{n+1}|}{|x_n|} \le q \quad \text{für alle } n \in \mathbb{N}$$

gibt, dann konvergiert $\sum_{n=0}^{\infty} x_n$ absolut.

Ein solches q gibt es insbesondere dann, wenn

$$\lim_{n\to\infty} \frac{|x_{n+1}|}{|x_n|} < 1$$

ist.

Beweis Ähnlich wie 5.12/5.13

(3) Wenn Sie den Verdacht haben, daß die Reihe $\sum x_n$ divergiert, untersuchen Sie zweckmäßig zuerst, ob überhaupt $\lim\limits_{n\to\infty} x_n = 0$ gilt: Ist das nicht der Fall, so kann die Reihe nach Lemma 5.4 ja nicht konvergieren. Sollten Sie bei (1) oder (2) ohnehin schon

$$\lim_{n\to\infty} \sqrt[n]{|x_n|} > 1 \quad \text{bzw.} \quad \lim_{n\to\infty} \frac{|x_{n+1}|}{|x_n|} > 1$$

festgestellt haben, brauchen Sie überhaupt nichts weiter zu tun, denn wie man sofort einsieht, ist keine dieser beiden Aussagen mit $\lim\limits_{n\to\infty} x_n = 0$ vereinbar.

(4) Schließlich kommt als (ziemlich subtile) Minorante die harmonische Reihe (Beispiel 5.5) in Betracht, mit der man weitere Reihen als divergent (oder zumindest nicht absolut konvergent) nachweisen kann.

Bemerkung Wurzel- und Quotientenkriterium finden Sie oft in einer Weise formuliert, mit der man auch Divergenz (genauer: diese oder nicht-absolute Konvergenz) nachweisen kann. Abgesehen von der in (3) erwähnten Situation, wo man die relevanten Limites ohnehin schon bestimmt hat, sind diese Erweiterungen eher eine Spielerei und ohne praktischen Nutzen; sie beruhen nämlich in jedem Fall darauf, daß man eine divergente geometrische Reihe als Minorante benutzt, und können deshalb nur dann zum Ziel führen, wenn schon der grundsätzlich einfachere Ansatz (3) greift. Dem Lernenden können solche Formulierungen gefährlich werden, denn sie verstärken den Eindruck, es handele sich bei den beiden Methoden tatsächlich um *Kriterien*, mit denen man die Konvergenzfrage in jedem Fall entscheiden könne. Das sind sie aber nicht: Aus

$$\lim_{n\to\infty} \frac{|x_{n+1}|}{|x_n|} = 1$$

kann man weder auf Konvergenz noch auf Divergenz schließen, und selbst noch so verfeinerte Varianten lassen stets Fälle unentscheidbar. Das ist auch nicht anders zu erwarten; schließlich beruhen all diese sogenannten Kriterien bloß auf dem Vergleich mit sehr speziellen Reihen, nämlich den geometrischen.

5.15 Beispiele Für jedes feste $x \in \mathbb{R}$ sind die Reihen

$$\sum_n \frac{x^n}{n!}, \quad \sum_n (-1)^n \frac{x^{2n}}{(2n)!} \quad \text{und} \quad \sum_n (-1)^n \frac{x^{2n+1}}{(2n+1)!}$$

absolut konvergent; sie definieren die wichtigen Funktionen (von $\mathbb{R}$ nach $\mathbb{R}$) mit den Namen Exponential-, Cosinus- und Sinusfunktion:

$$\exp x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

$$\cos x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}$$

$$\sin x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}$$

Beweis Für $x = 0$ ist die Konvergenz klar und für $x \neq 0$ folgt sie nach dem Quotientenkriterium 5.14, denn etwa bei der Exponentialreihe kommt es auf den Quotienten

$$\frac{\left|x^{n+1}/(n+1)!\right|}{|x^n/n!|} = \frac{|x|^{n+1}}{|x|^n} \cdot \frac{1 \cdot 2 \cdot \cdots \cdot n}{1 \cdot 2 \cdot \cdots \cdot (n+1)} = \frac{|x|}{n+1}$$

an, und wie wir wissen, ist $\lim\limits_n \dfrac{|x|}{n+1} = 0$.

Mehr über diese Funktionen demnächst. Im Augenblick aber schulde ich Ihnen noch etwas Anderes: Den Begriff der absoluten Konvergenz habe ich ja mit dem Argument angepriesen, daß diese anders als die gewöhnliche unempfindlich gegen Reihenumordnungen sei. Freilich habe ich das bisher bloß in Aussicht gestellt, aber nicht genau formuliert und schon gar nicht bewiesen. Das hole ich jetzt nach.

5.16 Umordnungssatz Sei $\sum\limits_{n=0}^{\infty} x_n$ eine absolut konvergente Reihe, und sei

$$\mathbb{N} \ni k \mapsto n_k \in \mathbb{N}$$

eine Permutation. Dann konvergiert auch die umgeordnete Reihe $\sum\limits_{k=0}^{\infty} x_{n_k}$ absolut und mit derselben Reihensumme.

Beweis Wir setzen zunächst zusätzlich $x_n \geq 0$ für alle $n \in \mathbb{N}$ voraus und bilden die Menge

$$S := \left\{ \sum_{n \in N} x_n \;\middle|\; N \subset \mathbb{N} \text{ endlich} \right\} \subset \mathbb{R}.$$

Die Notation $\sum\limits_{n \in N}$ gibt ohne weiteres Sinn, weil für diese Summen das Kommutativgesetz natürlich gilt.

Ich behaupte nun: Die Reihe $\sum_n x_n$ konvergiert genau dann, wenn S (nach oben) beschränkt ist, und dann ist

$$\sum_{n=0}^{\infty} x_n = \sup S.$$

Beweis Sei $\sum_n x_n$ konvergent. Für eine nicht-leere endliche Teilmenge $N \subset \mathbb{N}$ sei $m \in N$ die größte in N enthaltene Zahl; dann ist sicher

$$\sum_{n \in N} x_n \leq \sum_{n=0}^{m} x_n \leq \sum_{n=0}^{\infty} x_n.$$

Insbesondere ist S durch die Zahl $\sum_{n=0}^{\infty} x_n$ nach oben beschränkt, folglich $\sup S \leq \sum_{n=0}^{\infty} x_n$.

Umgekehrt sei jetzt S als beschränkt vorausgesetzt. Für jedes $m \in \mathbb{N}$ gilt dann

$$\sum_{n=0}^{m} x_n = \sum_{n \in \{0,1,\ldots,m\}} x_n \in S$$

und deshalb $\sum_{n=0}^{m} x_n \leq \sup S$. Die Partialsummenfolge $\left(\sum_{n=0}^{m} x_n\right)_m$ ist daher beschränkt, d.h. die Reihe $\sum_n x_n$ konvergiert, mit $\sum_{n=0}^{\infty} x_n \leq \sup S$.

Damit folgen die behauptete Äquivalenz und auch die Gleichung.

Die eben bewiesene Behauptung stellt nun eine neue Beschreibung von Konvergenz und Reihensumme dar, die von dem Umordnungsprozeß offenbar gar nichts merkt: Schon die Menge S ist ja für die umgeordnete

Reihe dieselbe wie für die ursprüngliche. Deshalb ist der Satz für Reihen mit nicht-negativen Gliedern jetzt bewiesen.

Für eine beliebige Reihe $\sum_n x_n$ bilden wir zwei Hilfsreihen:

$$\sum_n \frac{|x_n| + x_n}{2} \quad \text{und} \quad \sum_n \frac{|x_n| - x_n}{2}$$

Wegen

$$0 \le \frac{|x_n| \pm x_n}{2} \le |x_n|$$

sind das konvergente Reihen mit nicht-negativen Gliedern (Majorantenkriterium!). Diese sind also gegen Umordnung unempfindlich, und die ursprüngliche Reihe ist es auch, weil man sie nach der Formel

$$x_n = \frac{|x_n| + x_n}{2} - \frac{|x_n| - x_n}{2}$$

als Differenz der beiden Hilfsreihen zurückerhält.

Übungsaufgaben

5.1 Begründen Sie:

(a) Für jede konvergente Reihe $\sum_{n=0}^{\infty} x_n$ gilt:

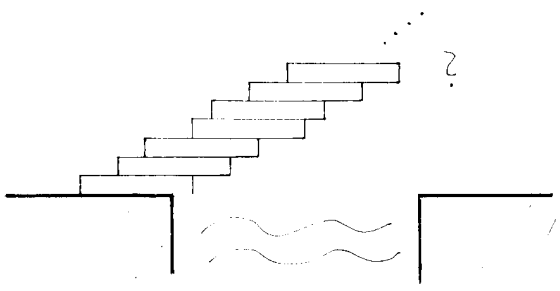
$$\lim_{m\to\infty} \sum_{n=m}^{\infty} x_n = 0$$

(b) Für jede absolut konvergente Reihe $\sum_{n=0}^{\infty} x_n$ gilt die Dreiecksungleichung

$$\left| \sum_{n=0}^{\infty} x_n \right| \le \sum_{n=0}^{\infty} |x_n|.$$

5.2 Zeigen Sie, daß die Reihe $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^2}$ und deshalb überhaupt die Reihen $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^d}$ für jedes $d \in \mathbb{N}$ mit $d \ge 2$ konvergieren. (Werfen Sie einen Blick auf Beispiel 2.9)

5.3 Für den Bau einer "Brücke" über einen Fluß der Breite b stehe an einem Ufer ein unbegrenzter Vorrat an gleichartigen Balken der Länge 1 zur Verfügung. Die Balken dürfen aber nur lose aufeinander gestapelt werden:



Kann man das Bauwerk so konstruieren, daß es bis über das andere Ufer reicht? Kann man dabei auch stabiles Gleichgewicht erreichen?

5.4 Beweisen Sie das folgende, oft als Konvergenzkriterium von Leibniz bezeichnete Lemma: Sei $(x_n)_{n=0}^\infty$ eine monoton fallende Folge mit $\lim\limits_{n\to\infty} x_n = 0$. Dann konvergiert die Reihe

$$\sum_{n=0}^{\infty}(-1)^n x_n.$$

Den Namen "Kriterium" verdient dieses Lemma freilich ebensowenig wie das Wurzel- oder Quotientenkriterium. In der Praxis ist es auch von geringem Nutzen, weil es keine absolute Konvergenz liefert. Es ist aber hübsch, einfach anzuwenden, deshalb populär, und immerhin liefert es natürlich die Konvergenz der alternierenden harmonischen Reihe:

$$\sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} \pm \cdots$$

Anleitung: Die Teilfolgen

$$\left(g_m = \sum_{n=0}^{2m}(-1)^n x_n\right)_m \quad \text{und} \quad \left(u_m = \sum_{n=0}^{2m+1}(-1)^n x_n\right)_m$$

der geraden und der ungeraden Partialsummen sind monoton und beschränkt. Was kann man über die Differenzen $g_m - u_m$ sagen?

5.5 Beweisen Sie das Konvergenzkriterium von Leibniz durch direkte Anwendung des Cauchy-Kriteriums.

Tip: Man sieht zum Beispiel leicht, daß

$$x_{2m+2k} \le \sum_{n=2m}^{2m+2k}(-1)^n x_n \le x_{2m}$$

für alle $m, k \in \mathbb{N}$ gilt.

5.6 Ebenso wie Teilfolgen kann man auch Teilreihen bilden. Sei dazu $(n_k)_{k=0}^\infty$ eine streng monoton wachsende Folge natürlicher Zahlen. Beweisen Sie: Wenn die Reihe $\sum_{n=0}^\infty x_n$ absolut konvergiert, dann konvergiert auch die Teilreihe $\sum_{k=0}^\infty x_{n_k}$ absolut; aber man darf in dieser Aussage nicht absolute durch gewöhnliche Konvergenz ersetzen.

5.7 Sei $\sum_{k=0}^\infty x_k$ eine Reihe reeller Zahlen. Beweisen Sie: Ist diese Reihe absolut konvergent, dann ist auch die Reihe $\sum_{k=0}^\infty (x_k)^2$ absolut konvergent. Diese Aussage wird aber falsch, wenn man absolute Konvergenz durchweg durch gewöhnliche ersetzt.

5.8 $f, g\colon \mathbb{R} \longrightarrow \mathbb{R}$ seien zwei Polynome, beide vom Nullpolynom verschieden. Beweisen Sie, daß die Reihe

$$\sum_n \frac{f(n)}{g(n)}x^n$$

für $|x| < 1$ absolut konvergiert, für $|x| > 1$ dagegen divergiert.

6 Abzählbare Mengen

Bei den bisherigen Überlegungen haben Sie ganz nebenbei eine ganze Menge konkreter, durchweg aus den üblichen Zahlbereichen $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ abgeleiteter Mengen kennengelernt. An verschiedenen Stellen ist klargeworden, daß vor allem der Unterschied zwischen endlichen und unendlichen Mengen ein ganz wesentlicher ist. Vor allem der Begriff der Konvergenz ist ja gerade deswegen so raffiniert, weil es *unendlich* viele $\varepsilon > 0$ gibt, für die da etwas verlangt wird.

In diesem kurzen Abschnitt möchte ich Ihnen etwas mehr über Größenbegriffe von Mengen vortragen. Wann sind zwei Mengen gleich groß? Nun, für zwei endliche Mengen X und Y liegt die Antwort auf der Hand: eben dann, wenn sie aus gleich vielen Elementen bestehen. Alternativ kann man sagen, daß X und Y gleich groß sind, wenn es eine bijektive Abbildung $X \longrightarrow Y$ gibt. Da wir für alle endlichen Menge eine Art Standardmodell kennen, können wir auch folgendes sagen: Eine Menge X ist genau dann endlich, wenn es eine Bijektion

$$\{0, 1, \ldots, n-1\} \longrightarrow X$$

gibt, wobei $n = |X|$ eben die Zahl der Elemente von X ist.

Daß X und Y gleich groß sind, wenn es eine Bijektion $X \longrightarrow Y$ gibt, kann man als Definition ohne weiteres für den nicht-endlichen Fall übernehmen; gelehrter spricht man übrigens von gleich mächtigen Mengen. Sind nun je zwei unendliche Mengen in diesem Sinne gleich mächtig? Das würde bedeuten, daß es außer den Mengen mit $0, 1, 2, 3, \ldots$ Elementen nur noch eine weitere Mengengröße "unendlich" gibt. Das ist aber keineswegs so, wie wir gleich sehen werden. In Wirklichkeit ist "unendlich" ein Sammelbegriff für Mengen der verschiedensten Größen (Mächtigkeiten), deren einzige Gemeinsamkeit die ist, daß sie eben nicht endlich sind.

Es ist hier nicht der Platz, das detailliert darzustellen, wir wollen uns vielmehr mit den Anfängen begnügen.

6.1 Definition Eine Menge X heißt abzählbar, wenn entweder $X = \emptyset$ ist oder es eine surjektive Abbildung $\mathbb{N} \xrightarrow{f} X$ gibt (eine Abzählung von X).

6.2 Beispiele (1) Jede endliche Menge ist abzählbar. Manchmal möchte man die von der Betrachtung ausschließen und spricht dann von abzählbar unendlichen Mengen. (In der Literatur ist mit abzählbar manchmal das gemeint.) Aus einer Abzählung $f\colon \mathbb{N} \longrightarrow X$ einer unendlichen Menge X kann man immer eine solche herstellen, die sogar bijektiv ist: man wähle aus f, das ja nichts Anderes ist als eine Folge in X, durch vollständige Induktion eine passende Teilfolge aus. Insbesondere haben alle abzählbar unendlichen Mengen dieselbe Mächtigkeit, nämlich die der Menge $\mathbb{N}$.

(2) Ist X abzählbar und gibt es eine surjektive Abbildung

$$X \xrightarrow{g} Y,$$

so ist auch Y eine abzählbare Menge, denn ist $f\colon \mathbb{N} \longrightarrow X$ eine Abzählung für X, so ist $g \circ f\colon \mathbb{N} \longrightarrow Y$ eine für Y. Insbesondere ist jede zu einer abzählbaren Menge X gleich mächtige Menge selbst abzählbar.

(3) Jede Menge von natürlichen Zahlen ist abzählbar: Für $\emptyset \neq X \subset \mathbb{N}$ definieren wir eine Abzählung $f\colon \mathbb{N} \longrightarrow X$ etwa dadurch, daß wir ein festes Element $a \in X$ wählen und f durch die Vorschrift

$$f(n) = \begin{cases} n & \text{falls } n \in X \\ a & \text{sonst} \end{cases}$$

festlegen.

(4) Allgemeiner ist jede Teilmenge einer abzählbaren Menge selbst abzählbar.

Pfiffiger sind die beiden

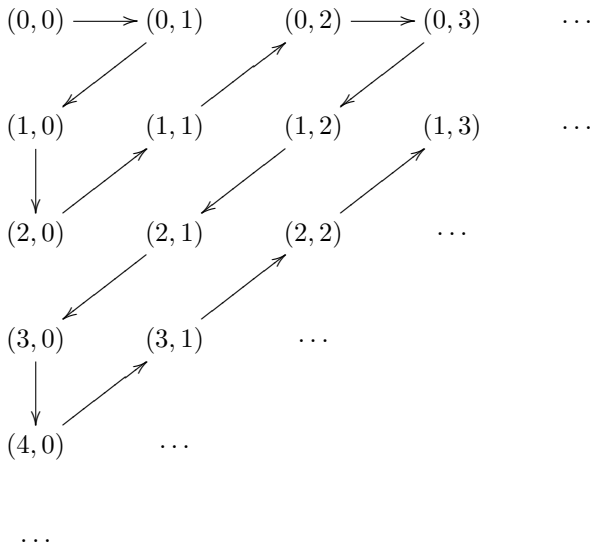
6.3 Regeln (a) Das kartesische Produkt zweier abzählbarer Mengen (und damit auch das endlich vieler solcher Mengen) ist wieder abzählbar.

(b) Die Vereinigung abzählbar vieler abzählbarer Mengen ist abzählbar.

Beweis Das folgende Schema beschreibt eine surjektive (sogar bijektive) Abbildung

$$\mathbb{N} \xrightarrow{\varphi} \mathbb{N} \times \mathbb{N},$$

also eine Abzählung von $\mathbb{N} \times \mathbb{N}$:



Sind nun $f\colon \mathbb{N} \longrightarrow X$ und $g\colon \mathbb{N} \longrightarrow Y$ Abzählungen von X und Y, so ist die Abbildung

$$\mathbb{N} \xrightarrow{\varphi} \mathbb{N} \times \mathbb{N} \xrightarrow{f \times g} X \times Y$$

surjektiv, also eine Abzählung von $X \times Y$. (Mit $f \times g$ ist natürlich die Abbildung gemeint, die dem Paar (m, n) das Paar $\big(f(m), g(n)\big)$ zuordnet.) Das beweist (a).

Die Aussage von (b) ist erst mal zu präzisieren: Gegeben sind eine Menge Λ und für jedes $\lambda \in \Lambda$ eine weitere Menge X_λ. Eine solche Vorgabe nennt man übrigens eine (mit Λ indizierte) Familie von Mengen und schreibt

$$(X_\lambda)_{\lambda \in \Lambda}$$

in Verallgemeinerung der Ihnen schon bekannten Begriffe

$$\begin{aligned} &\text{Paar } (X_1, X_2) \\ &n\text{-tupel } (X_1, X_2, \ldots, X_n) \\ &\text{Folge } (X_n)_{n \in \mathbb{N}} \end{aligned}$$

von Mengen. Regel (b) verspricht nun: Sind sowohl Λ als auch alle Mengen X_λ abzählbar, so gilt das gleiche für die Vereinigung

$$\bigcup_{\lambda \in \Lambda} X_\lambda := \{x \,|\, \text{es gibt ein } \lambda \in \Lambda \text{ mit } x \in X_\lambda\}.$$

Zum Beweis dürfen wir annehmen, daß alle auftretenden Mengen nicht-leer sind, und wir wählen Abzählungen $f\colon \mathbb{N} \longrightarrow \Lambda$ und $g_\lambda\colon \mathbb{N} \longrightarrow X_\lambda$ $(\lambda \in \Lambda)$. Dann ist die Abbildung

$$\begin{array}{ccc} \mathbb{N} \times \mathbb{N} & \longrightarrow & \displaystyle\bigcup_{\lambda \in \Lambda} X_\lambda \\ (m, n) & \mapsto & g_{f(m)}(n) \in X_{f(m)} \end{array}$$

surjektiv, $\bigcup_{\lambda\in\Lambda} X_\lambda$ also abzählbar nach (a) und (2).

Jetzt können wir interessantere Beispiele bilden:

(5) Weil $\mathbb{Z} = \mathbb{N} \cup (-\mathbb{N})$ abzählbar ist, ist nach Regel (a) für jedes $n \in \mathbb{N}$ auch $\mathbb{Z}^n$ abzählbar. (Die Menge $\mathbb{Z}^3$ zum Beispiel kann man sich gut als die Menge aller Punkte im Raum mit ganzzahligen Koordinaten vorstellen.)

(6) Weil die Abbildung

$$\begin{array}{ccc} \mathbb{Z} \times \mathbb{N}\backslash\{0\} & \longrightarrow & \mathbb{Q} \\ (p,q) & \mapsto & p/q \end{array}$$

surjektiv ist, ist auch $\mathbb{Q}$, und damit $\mathbb{Q}^n$ für jedes $n \in \mathbb{N}$ abzählbar.

(7) Die Menge aller Polynome mit rationalen Koeffizienten ist abzählbar: Sei für $d \in \mathbb{N}$

$$P_d := \left\{ f\colon x \mapsto \sum_{k=0}^{d-1} a_k x^k \,\middle|\, a_k \in \mathbb{Q} \right\}$$

die Menge aller solcher Polynome, deren Grad kleiner als d ist (einschließlich des Nullpolynoms). Durch die d Koeffizienten ist sofort eine Bijektion zwischen $\mathbb{Q}^d$ und P_d gegeben; alle Mengen P_d sind also abzählbar. Die Menge *aller* rationalen Polynome ist aber

$$\bigcup_{d=0}^{\infty} P_d$$

und deshalb auch abzählbar nach Regel (b).

Es ist an der Zeit zu sehen, daß nicht alle Mengen abzählbar sind:

6.4 Satz $\mathbb{R}$ ist nicht abzählbar.

Beweis Ich stütze mich hier auf die bekannte Tatsache, daß reelle Zahlen auf eindeutige Weise den (im allgemeinen) unendlichen Dezimalbrüchen entsprechen.

Wir nehmen an, $\mathbb{R}$ sei abzählbar; erst recht ist dann die Teilmenge $I := \{t \in \mathbb{R} \,|\, 0 \le t < 1\}$ abzählbar. Wir wählen eine Abzählung von I und denken uns die Dezimalbruchentwicklungen aller Zahlen aus I in der Reihenfolge der Abzählung untereinander aufgeschrieben:

$$\begin{aligned} t_1 &= 0.t_{11}t_{12}t_{13}t_{14}\ldots \\ t_2 &= 0.t_{21}t_{22}t_{23}t_{24}\ldots \\ t_3 &= 0.t_{31}t_{32}t_{33}t_{34}\ldots \\ t_4 &= 0.t_{41}t_{42}t_{43}t_{44}\ldots \\ &\ldots\ldots\ldots\ldots\ldots \end{aligned}$$

Jetzt bilden wir einen neuen Dezimalbruch

$$u = 0.u_1u_1u_3u_4\ldots$$

nach der Regel

$$u_n := \begin{cases} 0 & \text{falls } t_{nn} > 0, \\ 1 & \text{falls } t_{nn} = 0. \end{cases}$$

Die Definition ist so eingerichtet (und nur darauf kommt es an), daß dieser Dezimalbruch sich von jedem der Dezimalbrüche t_n unterscheidet (zumindest nämlich an der n-ten Stelle) und außerdem der Konvention genügt, nicht die Periode $\overline{9}$ zu haben. u kommt also einerseits nicht in der obigen Liste vor, ist andererseits aber die Dezimalbruchentwicklung einer reellen Zahl aus I: ein klarer Widerspruch. $\mathbb{R}$ kann also nicht abzählbar sein.

Alles in allem ist das doch erstaunlich: Wir haben nicht nur die Existenz irrationaler reeller Zahlen erneut bewiesen, sondern sogar gezeigt, daß es "mehr" irrationale als rationale Zahlen gibt. Demgegenüber sind die rationalen Zahlen, ja sogar die Punkte von $\mathbb{Q}^3 \subset \mathbb{R}^3$, obwohl im Raum $\mathbb{R}^3$ dicht, in die Nähe der natürlichen Zahlen gerückt: Etwas provozierend könnte man sagen, es gibt nicht mehr solche rationalen Punkte im Raum als natürliche Zahlen $0, 1, 2, 3 \ldots$

Die Tatsache, daß die Menge $\mathbb{N} \times \mathbb{N}$ abzählbar ist, ist für das Rechnen mit Reihen von großer Bedeutung, weil sie die Bildung von Mehrfachreihen erlaubt.

6.5 Definition Mit einer Doppelreihe

$$\sum_{m,n=0}^{\infty} x_{mn}$$

ist natürlich eine Abbildung $\mathbb{N} \times \mathbb{N} \longrightarrow \mathbb{R}$; $(m, n) \mapsto x_{mn}$ gemeint. Diese Reihe heißt absolut konvergent, wenn es eine bijektive Abzählung $\mathbb{N} \ni k \mapsto \varphi(k) = \big(\varphi_1(k), \varphi_2(k)\big) \in \mathbb{N} \times \mathbb{N}$ gibt, so daß die Reihe

$$\sum_{k=0}^{\infty} x_{\varphi_1(k)\varphi_2(k)}$$

absolut konvergiert. Deren Reihensumme ist dann per definitionem die Summe der Doppelreihe.

Erläuterung Die Doppelreihe wird also mittels einer bijektiven Abzählung von $\mathbb{N} \times \mathbb{N}$ in eine gewöhnliche Reihe umgewandelt. Der springende Punkt: Entscheidet man sich für eine andere bijektive Abzählung, so erhält man eine Umordnung dieser gewöhnlichen Reihe, was nach dem Umordnungssatz 5.16 auf deren Konvergenz und Reihensumme keinen Einfluß hat. Tatsächlich gilt die in der Definition geforderte Konvergenz also unabhängig von der Wahl der bijektiven Abzählung, und die Summe einer konvergenten Doppelreihe ist damit wohldefiniert. Das ist auch der Grund dafür, daß man bei Doppelreihen von vornherein nur den absoluten Konvergenzbegriff ins Auge faßt.

Warum sollte man überhaupt so was betrachten? Nun, Doppelreihen treten beim Rechnen mit Reihen ganz automatisch auf, wenn man nämlich Reihen miteinander multipliziert. Der folgende Satz ist gewissermaßen das Distributivgesetz für absolut konvergente Reihen.

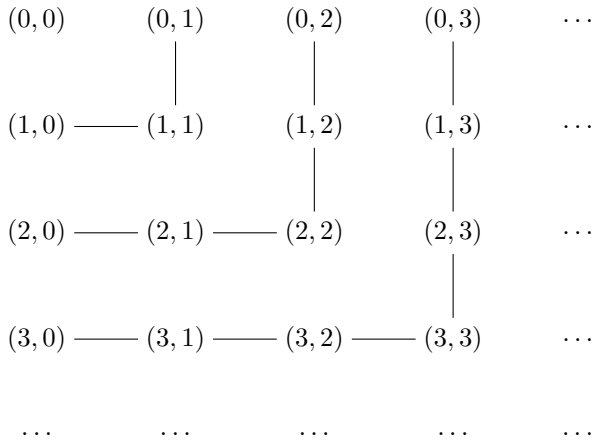
6.6 Satz $\sum_{j=0}^{\infty} x_j$ und $\sum_{k=0}^{\infty} y_k$ seien absolut konvergente Reihen. Dann konvergiert auch die Doppelreihe

$$\sum_{j,k=0}^{\infty} x_j y_k$$

absolut, und für die Reihensummen gilt

$$\sum_{j,k=0}^{\infty} x_j y_k = \left(\sum_{j=0}^{\infty} x_j\right)\left(\sum_{k=0}^{\infty} y_k\right).$$

Beweisidee Man verwendet eine Abzählung von $\mathbb{N} \times \mathbb{N}$, die die "Quadrate" $\{0, 1, \ldots, m\} \times \{0, 1, \ldots, m\}$ nacheinander ausschöpft:



So entstehen unter anderem die Partialsummen

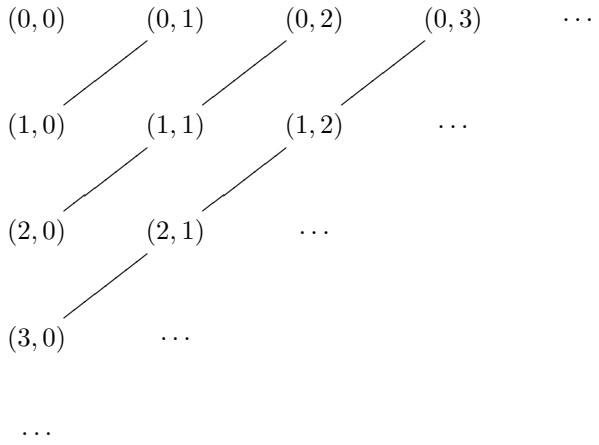
$$\sum_{j,k=0}^{m} x_j y_k = \left(\sum_{j=0}^{m} x_j\right)\left(\sum_{k=0}^{m} y_k\right)$$

(nämlich wenn man mit dem m-ten Quadrat gerade fertig ist), und das soll als Beweisandeutung genügen.

6.7 Beispiel Für alle $x, y \in \mathbb{R}$ gilt:

$$\exp(x+y) = (\exp x)(\exp y)$$

Beweis Wir wissen, daß die Exponentialreihen absolut konvergieren, und können Satz 6.6 deshalb auf das Produkt $(\exp x)(\exp y)$ anwenden. Wir sind aber berechtigt, zur Auswertung der entstehenden Doppelreihe eine andere Abzählung von $\mathbb{N}\times\mathbb{N}$ zu benutzen, und wählen diesmal eine, die die Diagonalen $\{(j,k)\,|\,j+k=n\}$ der Reihe nach abarbeitet:



Damit ergibt sich in der Tat

$$(\exp x)(\exp y) = \lim_{m\to\infty} \sum_{n=0}^{m} \sum_{j+k=n} \frac{x^j}{j!}\frac{y^k}{k!} = \lim_{m\to\infty} \sum_{n=0}^{m} \frac{1}{n!} \sum_{j+k=n} \underbrace{\frac{n!}{j!\cdot k!}}_{\binom{n}{j}} x^j y^k = \lim_{m\to\infty} \sum_{n=0}^{m} \frac{1}{n!}(x+y)^n = \exp(x+y)$$

unter Verwendung des binomischen Satzes

$$(x+y)^n = \sum_{j=0}^{n} \binom{n}{j} x^j y^{n-j}.$$

Ich denke, daß die Idee, absolute Konvergenz einzuführen, nun schon reiche Früchte getragen hat: Mit absolut konvergenten Reihen darf man wirklich recht unbefangen fast wie mit endlichen Summen rechnen. Dazu gehört auch noch folgender Aspekt der Doppelreihen, den ich hier ohne Beweis nur mitteilen möchte:

6.8 Satz $\sum_{m,n=0}^{\infty} x_{mn}$ sei eine absolut konvergente Doppelreihe. Für jedes $m \in \mathbb{N}$ konvergiert dann die Reihe $\sum_n x_{mn}$ absolut, und die Summe der Doppelreihe läßt sich als

$$\sum_{m,n=0}^{\infty} x_{mn} = \sum_{m=0}^{\infty} \left(\sum_{n=0}^{\infty} x_{mn}\right)$$

berechnen.

Bemerkungen Beachten Sie, daß diese Tatsache nicht schon durch den Umordnungssatz gedeckt ist: Keine Abzählung von $\mathbb{N}\times\mathbb{N}$ kann ja alle Paare (m,n) für ein festes m nacheinander durchlaufen. — Natürlich kann man die Rollen von m und n auch vertauschen:

$$\sum_{m,n=0}^{\infty} x_{mn} = \sum_{n=0}^{\infty}\left(\sum_{m=0}^{\infty} x_{mn}\right)$$

Zum Schluß sei noch erwähnt, daß der Begriff der absoluten Konvergenz sich nicht nur ohne weiteres auf Drei- und Mehrfachreihen ausdehnen läßt, sondern es sogar erlaubt, Summen

$$\sum_{\lambda\in\Lambda} x_\lambda \quad \text{mit } x_\lambda \in \mathbb{R} \text{ für alle } \lambda \in \Lambda$$

einen Sinn zu geben, in denen Λ eine ganz beliebige abzählbare Menge ist. Im Falle der absoluten Konvergenz nennt man $(x_\lambda)_{\lambda\in\Lambda}$ eine summierbare Familie.

Übungsaufgabe

6.1 Die Menge aller Teilmengen einer Menge X nennt man die *Potenzmenge* $\mathbf{P}X$ von X. Sind die folgenden Mengen abzählbar?

(a) $\mathbf{P}\mathbb{N}$

(b) $\{N\in\mathbf{P}\mathbb{N}\,|\,N \text{ endlich}\}$

7 Stetige Funktionen

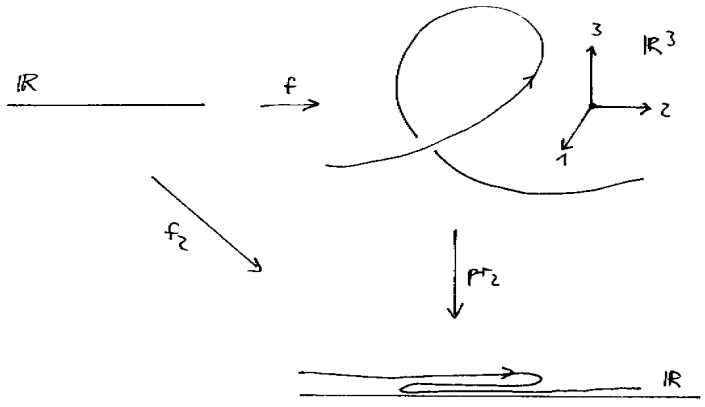
In diesem Abschnitt beginnen wir das Studium reeller Funktionen, also von Abbildungen

$$X \xrightarrow{f} Y$$

mit $X, Y \subset \mathbb{R}$. Damit kommen wir endlich auch der konkreten mathematischen Beschreibung physikalischer Vorgänge ein deutliches Stück näher. Wenn auch die in der Physik relevanten Funktionen $f\colon X \longrightarrow Y$ eigentlich eher solche sind, bei denen X oder Y oder beide mehrdimensional, also Teilmengen von $\mathbb{R}^n$ etwa für $n = 3$ oder $n = 4$ sind. Aber es gibt schon interessante Situationen, in denen von den drei Raumkoordinaten zwei unwesentlich sind und damit nur ein "Freiheitsgrad" bleibt, sei es, daß mit diesen Koordinaten nichts passiert (freier Fall) oder sie zwangsweise festgelegt sind (ein Punkt auf dem Umfang eines Rades). Abgesehen davon lassen sich etwa Bahnkurven eines Massenpunktes, also Abbildungen

$$\mathbb{R} \xrightarrow{f} \mathbb{R}^3$$

ebensogut durch die drei Komponentenfunktionen $\mathbb{R} \xrightarrow{f_j} \mathbb{R}$ $(j = 1, 2, 3)$ beschreiben, die in gelehrter Ausdrucksweise durch $f_j = \mathrm{pr}_j \circ f$ erklärt sind:



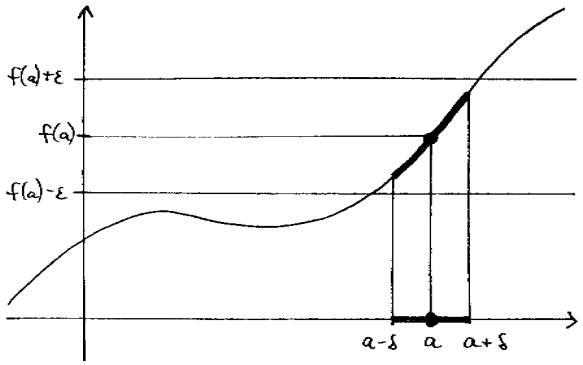
Bleiben wir gleich bei Bahnkurven. Es ist unmittelbar anschaulich, daß nicht jede mathematisch mögliche Funktion als Komponente einer solchen Bahn auftreten wird; vielmehr kommen dafür nur *stetige* Funktionen in Frage. Jetzt machen wir uns daran, diesen Begriff, von dem Sie wahrscheinlich schon eine Vorstellung haben, mathematisch zu präzisieren.

7.1 Definition Seien $X, Y \subset \mathbb{R}$,

$$f\colon X \longrightarrow Y$$

eine Funktion und $a \in X$. Dann heißt f an der Stelle a (oder kurz: bei a) stetig, wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt mit

$$|f(x) - f(a)| < \varepsilon \quad \text{für alle } x \in X \text{ mit } |x - a| < \delta.$$



f heißt stetig schlechthin, wenn f an jeder Stelle $a \in X$ stetig ist.

Sie sehen sofort, daß die Definition genau wie die der Konvergenz einer Folge strukturiert ist: $\delta > 0$ tritt an die Stelle von $D \in \mathbb{N}$, und die Forderung

$$\text{für alle } x \in X \text{ mit } |x - a| < \delta$$

(d.h. für alle $x \in X$, die genügend nahe an a sind) an die Stelle von

$$\text{für alle } n \in \mathbb{N} \text{ mit } n > D$$

(d.h. für alle genügend großen n). Wieder ist der eigentliche Pfiff, daß für *jede* Vorgabe von $\varepsilon > 0$ etwas zu erfüllen ist.

7.2 Beispiele (1) Für jedes feste $c \in \mathbb{R}$ ist die ebenfalls einfach mit c bezeichnete konstante Funktion

$$c : \mathbb{R} \longrightarrow \mathbb{R}$$

stetig. Sei nämlich $a \in \mathbb{R}$ und $\varepsilon > 0$. Dann gilt

$$|c(x) - c(a)| = |c - c| = 0 < \varepsilon$$

sogar für alle $x \in \mathbb{R}$; hier tut's also jedes $\delta > 0$, etwa $\delta = 1$.

(2) Die identische Abbildung

$$\mathrm{id} : \mathbb{R} \longrightarrow \mathbb{R}$$

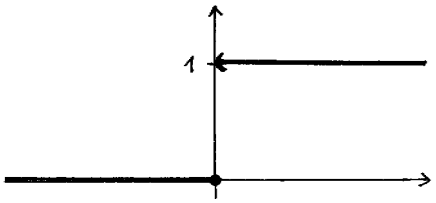
ist stetig: Seien $a \in \mathbb{R}$ und $\varepsilon > 0$. Dann gilt

$$|\mathrm{id}(x) - \mathrm{id}(a)| = |x - a| < \varepsilon$$

für welche $x \in \mathbb{R}$? Na eben für alle mit $|x - a| < \varepsilon$, wir können einfach $\delta = \varepsilon$ nehmen.

(3) Die Funktion $f : \mathbb{R} \longrightarrow \mathbb{R}$ mit

$$f(x) = \begin{cases} 0 & \text{falls } x \leq 0 \\ 1 & \text{falls } x > 0 \end{cases}$$



ist an der Stelle 0 unstetig. Zu $\varepsilon := 1 > 0$ kann es nämlich kein $\delta > 0$ mit

$$|f(x) - f(0)| < \varepsilon \quad \text{für alle } x \text{ mit } |x - 0| < \delta$$

geben: Explizit hieße das ja

$$|f(x)| < 1 \quad \text{für alle } x \text{ mit } |x| < \delta,$$

im Widerspruch zu $f(\frac{\delta}{2}) = 1$. An allen anderen Stellen ist f aber stetig, denn für $a \neq 0$ gilt

$$|f(x) - f(a)| = 0 \quad \text{für alle } x \text{ mit } |x - a| < |a|,$$

und wir kommen also mit $\delta = |a| > 0$ zurecht. Wenn man will, kann man die Idee dieses kleinen Beweises prägnanter so formulieren: Für jedes $a \neq 0$ wird f, eingeschränkt auf die Menge $\{x \in \mathbb{R} \mid |x - a| < |a|\}$ zu einer konstanten, insbesondere stetigen Funktion,

und wir können uns darauf berufen, daß Stetigkeit ohnehin eine — wie man sagt — *lokale* Eigenschaft ist, die nur von den Werten von f in der Nähe des betrachteten Punktes abhängt. Hier die genaue Formulierung:

7.3 Lemma Seien $X \subset \mathbb{R}$ eine Teilmenge, $f: X \longrightarrow \mathbb{R}$ eine Funktion und $a \in X$. Wenn es ein $\delta > 0$ gibt, so daß die Einschränkung

$$f|\{x \in X \mid |x - a| < \delta\}$$

bei a stetig ist, dann ist auch f selbst bei a stetig.

Beweis Zu $\varepsilon > 0$ finden wir ein $\delta' > 0$ mit

$$|f(x) - f(a)| < \varepsilon \quad \text{für alle } x \in X \text{ mit } |x - a| < \delta \text{ und } |x - a| < \delta'.$$

Es genügt jetzt, δ' durch die kleinere der beiden Zahlen δ, δ' zu ersetzen.

Bemerkung Umgekehrt ist klar, daß aus der Stetigkeit einer Funktion auch die jeder Einschränkung folgt: es wird ja dann einfach weniger verlangt.

(4) Die schon als Beispiel 1.8(4) vorgestellte abstruse Funktion $f: \mathbb{R} \longrightarrow \mathbb{R}$ mit

$$f(x) = \begin{cases} 1 & \text{falls } x \in \mathbb{Q} \\ 0 & \text{falls } x \notin \mathbb{Q} \end{cases}$$

ist eine Funktion, die an keiner Stelle stetig ist. Der Grund ist, daß für jedes $\delta > 0$ zwischen $a - \delta$ und $a + \delta$ sowohl rationale als auch irrationale Zahlen liegen.

Ebensowenig wie die Konvergenz von Zahlenfolgen muß man die Stetigkeit einer Funktion jedesmal mit Epsilon (und Delta) beweisen. Vielmehr erledigen sich die meisten Fälle mühelos durch Anwendung der

7.4 Regeln Summen, Produkte und Quotienten von (an einer Stelle a) stetigen Funktionen $X \longrightarrow \mathbb{R}$ sind stetig.

Erläuterung All diese Bildungen sind punktweise gemeint, d.h. man addiert, multipliziert, dividiert die Funktionswerte von f und g an derselben Stelle:

$$(f + g)(x) = f(x) + g(x)$$
$$(f \cdot g)(x) = f(x) \cdot g(x)$$
$$(f/g)(x) = f(x)/g(x)$$

Es versteht sich von selbst, daß im Fall der Quotientenbildung die Nennerfunktion $g: X \longrightarrow \mathbb{R}$ keine Nullstellen haben darf. Nötigenfalls muß man das erzwingen, indem man auf

$$X' := \{x \in X \mid g(x) \neq 0\}$$

einschränkt.

Die Beweise dieser Regeln sind Routine, wie bei konvergenten Folgen.

Während man im Zusammenhang mit der Stetigkeit wie immer sorgfältig auf die Definitionsbereiche der Funktionen achten muß, kann man bei deren Zielmenge großzügiger sein. Tatsächlich macht es in der Definition 7.1 nicht den geringsten Unterschied, wenn Sie die Zielmenge $Y \subset \mathbb{R}$ gleich durch $\mathbb{R}$ ersetzen. Wenn es bequem ist, identifiziert man für diese Zwecke $f\colon X \longrightarrow Y$ schon mal mit der Funktion $X \xrightarrow{f} Y \hookrightarrow \mathbb{R}$, die ja gemäß unseren Vereinbarungen eine andere ist.

Noch eine wichtige

7.5 Regel Sind $X \xrightarrow{f} Y$ und $Y \xrightarrow{g} Z$ stetig, so ist auch die Komposition

$$g \circ f\colon X \xrightarrow{f} Y \xrightarrow{g} Z$$

stetig.

Beweis, etwa an der Stelle $a \in X$. Sei $\varepsilon > 0$. Dann finden wir ein $\delta > 0$ mit

$$|g(y) - g\big(f(a)\big)| < \varepsilon \quad \text{für alle } y \in Y \text{ mit } |y - f(a)| < \delta,$$

denn g ist bei $f(a)$ stetig. Jetzt die Stetigkeit von f bei a mit der "Eingabe" $\delta > 0$ ausnutzend, finden wir ein $\gamma > 0$ mit

$$|f(x) - f(a)| < \delta \quad \text{für alle } x \in X \text{ mit } |x - a| < \gamma.$$

Insbesondere gilt dann

$$|(g \circ f)(x) - (g \circ f)(a)| = |(g\big(f(x)\big) - g\big(f(a)\big)| < \varepsilon \quad \text{für alle } x \in X \text{ mit } |x - a| < \gamma.$$

Durch fleißiges Anwenden dieser Regeln erhalten wir unter anderem:

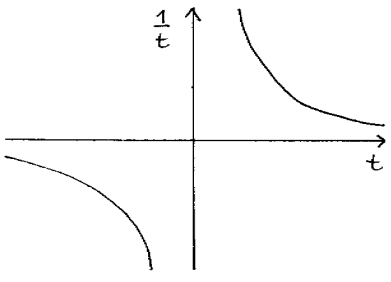
7.6 Satz Jedes reelle Polynom

$$\mathbb{R} \ni t \mapsto \sum_{j=0}^{d} a_j t^j \in \mathbb{R}$$

ist stetig. Allgemeiner sind alle rationale Funktionen stetig: das sind die durch den Quotienten zweier Polynome f und $g \neq 0$ gegebenen Funktionen:

$$\frac{f}{g}\colon \{t \in \mathbb{R} \,|\, g(t) \neq 0\} \longrightarrow \mathbb{R}, \quad t \mapsto \frac{f(t)}{g(t)}$$

Beachten Sie in diesem Zusammenhang: Die rationale Funktion $t \mapsto \frac{1}{t}$ ist an der Stelle $0 \in \mathbb{R}$ nicht etwa unstetig; sie ist dort gar nicht definiert, so daß die Frage nach ihrer Stetigkeit bei 0 keinen Sinn gibt.



Die Begriffe der Stetigkeit und der Folgenkonvergenz sind nicht nur, wie schon erwähnt, gleich strukturiert, sie sind auch inhaltlich miteinander verbunden:

7.7 Satz $\mathbb{R} \supset X \xrightarrow{f} \mathbb{R}$ sei eine Funktion, $a \in X$. Dann sind äquivalent:

(a) f ist bei a stetig

(b) Für jede Folge $(x_n)_{n=0}^{\infty}$ in X mit $\lim\limits_{n\to\infty} x_n = a$ gilt $\lim\limits_{n\to\infty} f(x_n) = f(a)$

Beweis Sei Stetigkeit bei a vorausgesetzt und (x_n) eine Folge in X mit $\lim x_n = a$. Sei außerdem $\varepsilon > 0$ gegeben. Dann finden wir ein $\delta > 0$ mit

$$|f(x) - f(a)| < \varepsilon \quad \text{für alle } x \in X \text{ mit } |x - a| < \delta.$$

Wegen $\lim x_n = a$ gibt es zu diesem δ ein D mit

$$|x_n - a| < \delta \quad \text{für alle } n > D.$$

Für diese n gilt dann aber auch

$$|f(x_n) - f(a)| < \varepsilon,$$

womit $\lim f(x_n) = f(a)$ bewiesen ist.

Jetzt setzen wir das Gegenteil von (a) voraus und beweisen, daß dann auch (b) nicht gilt. f ist also bei a unstetig: Es gibt ein $\varepsilon > 0$ mit der Eigenschaft, daß zu jedem $\delta > 0$ ein "Verbrecher"-$x \in X$ existiert, also eines mit $|x - a| < \delta$, aber $|f(x) - f(a)| \geq \varepsilon$. Insbesondere können wir zu $\delta := 1/n$ je ein solches $x = x_n$ wählen, haben also:

$$|f(x_n) - f(a)| \geq \varepsilon \quad \text{für jedes positive } n \in \mathbb{N}.$$

Die Folge $\big(f(x_n)\big)_{n=1}^{\infty}$ konvergiert dann sicher nicht gegen $f(a)$ (wobei offen bleibt, ob sie überhaupt konvergiert). Andererseits folgt aus $|x_n - a| < 1/n$ sofort $\lim x_n = a$. Damit ist (b) verletzt, und der Beweis geführt.

Bemerkungen Der Satz wird naheliegenderweise gern als Folgenkriterium für Stetigkeit bezeichnet. Die Richtung (a)⇒(b) ist von praktischem Nutzen bei der Arbeit mit Folgen, z.B. folgt so aus $\lim \frac{1}{n} = 0$ und der bald zu besprechenden Stetigkeit der Wurzelfunktion

$$\lim_{n\to\infty} \sqrt{1 + \frac{1}{n}} = \sqrt{1 + 0} = 1.$$

Dagegen ist die umgekehrte Richtung ein wichtiges theoretisches Hilfsmittel, mit dem man einen Stetigkeitsbeweis auf einen Konvergenzbeweis zurückführen kann.

Übungsaufgaben

7.1 Führen Sie den Beweis der Behauptung in Beispiel 7.2(4) aus.

7.2 Die Funktion $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ sei durch

$$f(x) = \begin{cases} 1/q & \text{falls } x \in \mathbb{Q} \text{ und } x = p/q \text{ mit } p \in \mathbb{Z} \text{ und dem kleinstmöglichen positiven } q \in \mathbb{N} \\ 0 & \text{wenn } x \text{ irrational ist} \end{cases}$$

definiert. Beweisen Sie, daß f an jeder rationalen Stelle unstetig, an jeder irrationalen aber stetig ist.

7.3 Sei $X \subset \mathbb{R}$ und seien $f, g\colon X \longrightarrow \mathbb{R}$ zwei stetige Funktionen. Beweisen Sie, daß die Funktion

$$|f|\colon X \longrightarrow \mathbb{R}, \quad |f|(x) := |f(x)|$$

und (als Anwendung davon) auch die Funktion

$$\max(f, g)\colon X \longrightarrow \mathbb{R}, \quad \max(f, g)(x) := \max\{f(x), g(x)\}$$

stetig ist.

8 Stetige Funktionen auf Intervallen

Während für den Begriff und die formalen Eigenschaften der Stetigkeit einer Funktion $f: X \longrightarrow Y$ die Natur des Definitionsbereiches $X \subset \mathbb{R}$ überhaupt keine Rolle spielt, sieht das ganz anders aus, wenn es um handfeste Resultate über stetige Funktion geht. Hier beschränke ich mich gleich auf den Fall eines sogenannten Intervalls, das ist der praktisch wichtigste.

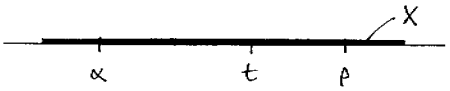
8.1 Definition Die Mengen der Form

Intervalltyp		offen	abgeschlossen	kompakt
$[a,b]$	$= \{x \in \mathbb{R} \mid a \le x \le b\}$		$\times$	$\times$
$(a,b]$	$= \{x \in \mathbb{R} \mid a < x \le b\}$	$\emptyset$	$\emptyset$	$\emptyset$
$[a,b)$	$= \{x \in \mathbb{R} \mid a \le x < b\}$	$\emptyset$	$\emptyset$	$\emptyset$
(a,b)	$= \{x \in \mathbb{R} \mid a < x < b\}$	$\times$	$\emptyset$	$\emptyset$
$[a,\infty)$	$= \{x \in \mathbb{R} \mid a \le x\}$		$\times$	
(a,∞)	$= \{x \in \mathbb{R} \mid a < x\}$	$\times$		
$(-\infty,b]$	$= \{x \in \mathbb{R} \mid x \le b\}$		$\times$	
$(-\infty,b)$	$= \{x \in \mathbb{R} \mid x < b\}$	$\times$		
$(-\infty,\infty)$	$= \mathbb{R}$	$\times$	$\times$	

worin stets $a, b \in \mathbb{R}$ und $a \le b$ sei, heißen Intervalle. Die Prädikate offen, abgeschlossen und kompakt sind durch die Tabelle erklärt. Ein Kreuzchen bei einem Intervalltyp bedeutet, daß dieser Typ die darüberstehende Eigenschaft immer hat, während ein $\emptyset$ eingetragen ist, wenn das nur in dem Ausnahmefall $a = b$ gilt, d.h. dann, wenn es sich um das leere Intervall handelt. Daß das leere Intervall alle drei Eigenschaften hat, wollen wir hiermit noch einmal ausdrücklich feststellen. Im nicht-leeren Fall sind die auftretenden Zahlen a und b durch das Intervall eindeutig bestimmt, und sie heißen die Randpunkte des Intervalls. Anschaulich bedeutet die Offenheit eines Intervalls, daß es keinen seiner Randpunkte enthält, und die Abgeschlossenheit, daß es alle seine Randpunkte enthält. "Kompakt" ist offenbar dasselbe wie "abgeschlossen und beschränkt".

Ich habe diese Definition in Tabellenform gewählt, weil es wichtig ist, alle Möglichkeiten vor Augen zu haben. Es gibt aber auch eine gemeinsame begriffliche Eigenschaft, an denen man die Intervalle erkennen kann:

8.2 Lemma Die Intervalle sind unter allen Teilmengen $X \subset \mathbb{R}$ durch die folgende Zwischenpunkteigenschaft charakterisiert:



Ist $\alpha, \beta \in X$ und $\alpha < \beta$, so gehört auch jede Zahl $t \in \mathbb{R}$ mit $\alpha < t < \beta$ zu X

Beweisskizze Klar, daß jedes Intervall diese Eigenschaft hat. Umgekehrt sei sie für eine Menge $X \subset \mathbb{R}$ vorausgesetzt. Wenn X nicht-leer und beschränkt ist, existieren

$$a := \inf X \quad \text{und} \quad b := \sup X.$$

Aus deren Definition folgt sofort

$$X \subset [a,b].$$

Ist andererseits $t \in (a,b)$, so ist t weder untere noch obere Schranke für X, also gibt es $\alpha, \beta \in X$ mit $\alpha < t < \beta$, und nach der Zwischenpunkteigenschaft ist $t \in X$. Es folgt

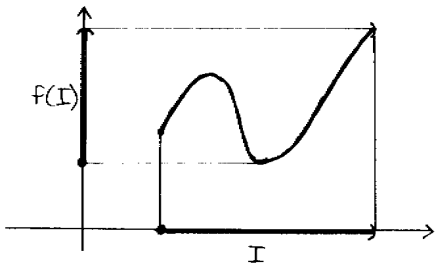
$$(a,b) \subset X,$$

und das läßt für X nur vier Möglichkeiten, je nachdem, welche der beiden Randpunkt zu X gehören.

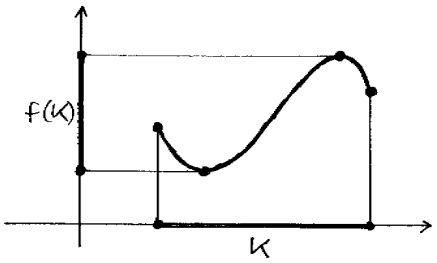
In den nicht-beschränkten Fällen kann man ganz ähnlich schließen.

Für stetige Funktionen, die auf Intervallen definiert sind, hatte ich "handfeste" Resultate in Aussicht gestellt. Es sind deren drei, und ich präsentiere sie gleich zusammen:

8.3 Zwischenwertsatz Sei $I \subset \mathbb{R}$ ein Intervall und $f\colon I \longrightarrow \mathbb{R}$ eine stetige Funktion. Dann ist auch die Bildmenge $f(I) \subset \mathbb{R}$ ein Intervall.



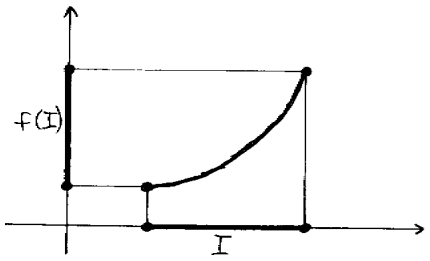
8.4 Satz von der **Annahme** des **Maximums** Sei $K \subset \mathbb{R}$ ein kompaktes Intervall und $f\colon K \longrightarrow \mathbb{R}$ stetig. Dann ist $f(K) \subset \mathbb{R}$ ein kompaktes Intervall.



8.5 Satz von der **Umkehrfunktion** Sei $I \subset \mathbb{R}$ ein Intervall und $f\colon I \longrightarrow \mathbb{R}$ eine injektive stetige Funktion. Dann ist f streng monoton, und die Umkehrfunktion

$$f^{-1}\colon f(I) \longrightarrow I \subset \mathbb{R}$$

ist ebenfalls stetig.



Zu den einzelnen Sätzen nun Erläuterungen, Beweise, Anwendungen.

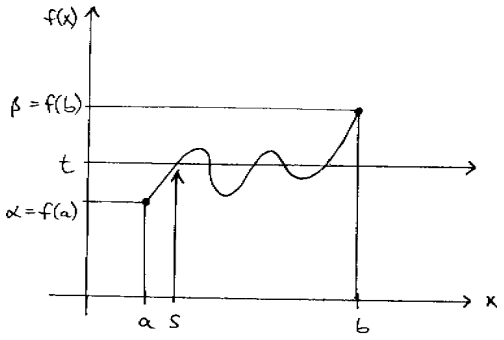
Der Name "Zwischenwertsatz" wird vor dem Hintergrund von Lemma 8.2 sofort klar. Die Schlußfolgerung dieses Satzes ist ja, daß die Wertemenge $f(I)$ die Zwischenpunkteigenschaft hat: Sind $\alpha, \beta \in \mathbb{R}$ mit $\alpha < \beta$ Werte von f, so ist auch jedes $t \in (\alpha, \beta)$ ein Wert von f. Das bringt uns auch gleich zum

Beweis (Zwischenwertsatz) Zu α, β wie eben wählen wir $a, b \in I$ mit

$$f(a) = \alpha, \quad f(b) = \beta.$$

Wir dürfen $a < b$ annehmen, indem wir den Beweis sonst für die Funktion $-I \ni x \mapsto F(-x) \in \mathbb{R}$ führen.

Sei jetzt also $t \in (\alpha, \beta)$. Gesucht ist ein $s \in I$ mit $f(s) = t$. Wie viele Beweise, in denen die Existenz eines nicht unbedingt eindeutig bestimmten Objektes zu etablieren ist, hat auch dieser zwei Teile: im ersten wird erst mal ein möglicher Kandidat für s bestimmt, im zweiten dann bewiesen, daß dieses s das Gewünschte leistet.



Wie könnte man ein passendes s finden? Eine guter Kandidat wäre sicher dort in $[a, b]$ zu suchen, wo die Werte von f zum erstenmal die Schwelle t überschreiten. Eine solche Stelle können wir mittels des Infimums wie folgt festnageln. Die wegen $[a, b] \subset I$ sinnvoll erklärte und offensichtlich beschränkte Menge

$$S := \{x \in [a, b] \mid f(x) \geq t\}$$

ist wegen $f(b) = \beta > t$, also $b \in S$ nicht-leer, besitzt deshalb ein Infimum $s \in \mathbb{R}$. Neben $s \leq b$ gilt auch $a \leq s$, denn a ist untere Schranke für S. Insbesondere ist also $s \in I$.

Jetzt beweisen wir, daß dieses s wirklich gut ist. Nach dem Satz über das Infimum (genau: nach dem Zusatz 4.12) gibt es in S eine Folge $(x_n)_{n=0}^{\infty}$, die (monoton fällt und) gegen s konvergiert. Weil f bei s stetig ist und weil

$$f(x_n) \geq t \quad \text{für alle } n$$

gilt, folgt nach Satz 7.7 und Regel 3.8(d)

$$f(s) = \lim_{n\to\infty} f(x_n) \geq t.$$

Insbesondere ist $s \neq a$, also $s > a$. Damit wird $\left(s - \frac{1}{n}\right)_n$ zu einer Folge in $[a, b]$, wenn man ein genügend großes Anfangsstück wegläßt. Wegen $s - \frac{1}{n} \notin S$ gilt

$$f\left(s - \frac{1}{n}\right) < t \quad \text{für alle } n,$$

woraus wie oben nun auch

$$f(s) = \lim_{n\to\infty} f\left(s - \frac{1}{n}\right) \leq t$$

folgt. Also ist $f(s) = t$ und damit der Beweis geführt.

Anwendungen Sei $I \subset \mathbb{R}$ ein Intervall, $f\colon I \longrightarrow \mathbb{R}$ stetig. Der Zwischenwertsatz liefert den meist entscheidenden Beitrag dazu, die Bildmenge $f(I)$ zu bestimmen: Ein Intervall ist ja schon eine sehr spezielle Teilmenge von $\mathbb{R}$, allein durch die Daten Typ und Randpunkte festgelegt, und diese sind in vielen konkreten Fällen ganz mühelos zu bestimmen.

8.6 Beispiel Sei $0 < n \in \mathbb{N}$. Die Funktion

$$f\colon [0, \infty) \longrightarrow \mathbb{R}, \quad f(x) = x^n$$

genügt dem Zwischenwertsatz. Um welches Intervall kann es sich bei $J := f\big([0, \infty)\big)$ wohl handeln? Nun, wegen $f(0) = 0$ und $f(x) \geq 0$ für alle $x \in [0, \infty)$ ist jedenfalls

$$\{0\} \subset J \subset [0, \infty).$$

Weil J offensichtlich (wegen $f(x) \geq x$ für $x \geq 1$) nicht nach oben beschränkt ist, kann nur

$$f\big([0,\infty)\big) = J = [0,\infty)$$

sein. Das bedeutet, daß jede Zahl $y \in [0,\infty)$ eine n-te Wurzel besitzt! Weil f außerdem streng monoton wächst, also injektiv ist, ist diese Wurzel eindeutig bestimmt; sie darf deswegen bedenkenlos mit $\sqrt[n]{y}$ bezeichnet werden.

Für ungerades n erkennt man auch die gleiche Weise, daß die Funktion

$$f\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad f(x) = x^n$$

bijektiv ist, und erhält für diese n die n-ten Wurzeln beliebiger reeller Zahlen. Allgemeiner zeigt man so den

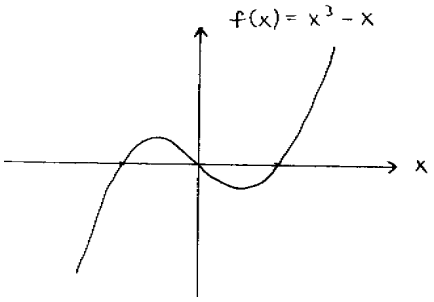
8.7 Satz Jedes Polynom $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ von ungeradem Grad besitzt mindestens eine Nullstelle in $\mathbb{R}$.

Beweisskizze Man braucht nur den Fall eines positiven Leitkoeffizienten zu betrachten und zeigt dann, daß

$$\begin{aligned} &f\big([0,\infty)\big) \quad \text{nicht nach oben, und} \\ &f\big((-\infty,0]\big) \quad \text{nicht nach unten} \end{aligned}$$

beschränkt ist.

Übrigens kann man hier natürlich nicht auf die Eindeutigkeit der Nullstelle schließen, denn ein Polynom, dessen Grad ungerade ist, braucht deswegen noch lange nicht monoton zu sein:



Nun zum Satz von der Annahme des Maximums einer stetigen Funktion

$$f\colon K \longrightarrow \mathbb{R} \quad (K \text{ kompaktes Intervall}).$$

Dieser (klassische) Name, der auf den ersten Blick gar nichts mit der Aussage zu tun hat, erklärt sich so: Abgesehen von dem trivialen Fall $K = \emptyset$ sagt der Satz, daß es Zahlen $c, d \in \mathbb{R}$ mit

$$f(K) = [c,d]$$

gibt. Über den Zwischenwertsatz hinaus wird also versprochen, daß f einen kleinsten Wert c und einen größten Wert d annimmt; explizit: daß es $s, t \in K$ gibt mit

$$c = f(s) \leq f(x) \leq f(t) = d \quad \text{für alle } x \in K.$$

Man sagt treffenderweise, daß die Funktion bei s und s ihr Minimum bzw. Maximum annimmt. Die gelegentlich anzutreffende Ausdrucksweise, daß die Funktion dort ein Minimum bzw. Maximum hat, ist schlechter, weil sie suggeriert, es könne mehr als ein Minimum geben, während es in Wirklichkeit nur mehrere Stellen geben kann, an denen das einzige Minimum angenommen wird. Ganz falsch ist es zu sagen, $s \in K$ sei ein Minimum von f, das versteht sich wohl von selbst.

Beweis (Annahme des Maximums) Nachdem der Fall $K = \emptyset$ trivial ist, sei etwa $K = [a, b]$ mit $a \le b$. Wie wir eben gesehen haben, kommt es darauf an, ein $t \in [a, b]$ mit

$$f(x) \le f(t) \quad \text{für alle } x \in [a, b]$$

zu finden (die Annahme des Minimums erledigt man dann durch Übergang von f zu $-f$).

Als erstes zeigen wir, daß die Menge $f\big([a, b]\big)$ nach oben beschränkt ist. Wäre sie das nicht, so gäbe es eine Folge $(x_n)_{n=0}^{\infty}$ in $[a, b]$ mit

$$f(x_n) > n \quad \text{für jedes } n \in \mathbb{N}.$$

Weil diese Folge selbst beschränkt ist, enthält sie nach dem Satz von Bolzano und Weierstraß (4.10) eine konvergente Teilfolge $(x_{n_k})_{k=0}^{\infty}$. Wegen

$$f\left(x_{n_k}\right) > n_k \ge k$$

hat sie noch die gleiche Eigenschaft, nach der die ursprüngliche Folge ausgesucht war: Wir dürfen diese also einfach durch die Teilfolge ersetzen, d.h. annehmen, daß schon $(x_n)_{n=0}^{\infty}$ konvergiert. Sei etwa

$$\lim_{n\to\infty} x_n = t.$$

Wegen

$$a \le x_n \le b \quad \text{für alle } n$$

ist $t \in [a, b]$. Nun ist aber f an der Stelle t stetig, und nach dem Folgenkriterium 7.7 ergibt sich

$$\lim_{n\to\infty} f(x_n) = f(t),$$

was mit $f(x_n) > n$ für alle $n \in \mathbb{N}$ unvereinbar ist.

Jetzt wissen wir, daß die Wertemenge $f\big([a, b]\big)$ nach oben beschränkt ist. Wegen $[a, b] \ne \emptyset$ ist sie auch nicht-leer, und wir können deshalb

$$d := \sup f\big([a, b]\big) \in \mathbb{R}$$

ins Auge fassen. Wieder mal nach dem Zusatz 4.12 zum Satz über das Infimum gibt es eine Folge in $f\big([a, b]\big)$, die gegen d konvergiert, d.h. eine Folge $(x_n)_{n=0}^{\infty}$ in $[a, b]$ selbst, so daß

$$\lim_{n\to\infty} f(x_n) = d.$$

Jetzt geht alles wie vorhin im Beweis der Beschränktheit: Wir dürfen (x_n) durch eine konvergente Teilfolge ersetzen; sei etwa

$$\lim_{n\to\infty} x_n = t.$$

Wieder folgt $t \in [a, b]$, und die Stetigkeit von f bei t erzwingt

$$d = \lim f(x_n) = f(\lim x_n) = f(t).$$

Damit ist der Beweis geführt, denn natürlich gilt

$$f(x) \le d = f(t) \quad \text{für alle } x \in [a, b].$$

Bemerkung Beachten Sie, wie in beide Beweisteile sowohl die Beschränktheit als auch die Abgeschlossenheit von $K = [a, b]$ eingeflossen sind. Die Gültigkeit dieses Satzes ist eine wichtige Besonderheit speziell der kompakten Intervalle; die analoge Aussage für nur beschränkte oder nur abgeschlossene Intervalle wäre falsch.

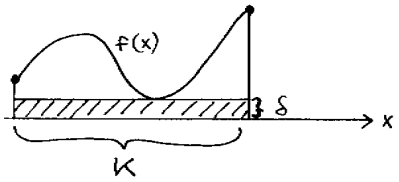
Eine typische Anwendung des Satzes:

8.8 Lemma Sei K ein kompaktes Intervall,

$$f\colon K \longrightarrow (0, \infty)$$

eine überall positive stetige Funktion. Dann gibt es ein $\delta > 0$ mit

$$f(x) \geq \delta \quad \text{für alle } x \in K.$$



Beweis $f(K)$ ist ein ganz in $(0, \infty)$ enthaltenes kompaktes Intervall, also (jedenfalls für $K \neq \emptyset$) von der Form

$$f(K) = [\delta, d] \quad \text{mit } \delta > 0.$$

Fertig.

Auf "konkrete" Anwendungen des Satzes müssen wir allerdings noch etwas warten, nämlich bis uns Methoden aus der Differentialrechnung erlauben, die versprochenen Stellen, an denen Minimum und Maximum angenommen werden, auch zu berechnen.

Schließlich zum Satz von der Umkehrfunktion. Es ist klar, daß eine streng monotone Funktion immer injektiv sein muß. Der erste Teil des Satzes sagt, daß für stetige Funktionen auf einem Intervall auch die Umkehrung gilt: ist eine solche Funktion injektiv, so ist sie zwangsläufig streng monoton. Den Beweis übergehe ich aus verschiedenen Gründen: erstens läßt er sich viel natürlicher im Kontext von Funktionen zweier statt nur einer formulieren, und zweitens ist dieser Teil des Satzes eher von theoretischem Interesse, weil die Differentialrechnung eine einfache praktische Methode liefert, Monotonieeigenschaften von Funktionen direkt zu erkennen. Der auch praktisch wichtige Teil des Satzes ist der zweite, der die Stetigkeit der Umkehrfunktion garantiert und damit eine neue Quelle stetiger Funktionen ist.

Beweis dieses Teiles Wir dürfen von einer auf dem Intervall I definierten streng wachsenden Funktion $f\colon I \to \mathbb{R}$ ausgehen und wollen die Stetigkeit von $f^{-1}\colon f(I) \to I$ beweisen — verblüffenderweise spielt es dafür gar keine Rolle, ob f selbst stetig ist.

Wir fixieren ein $a \in I$; ich schreibe erst mal in den auf f bezogenen Standardnotationen aus, was die zu beweisende Stetigkeit von f^{-1} an der Stelle $f(a)$ bedeutet:

Zu jedem $\delta > 0$ gibt es ein $\varepsilon > 0$, so daß $|x - a| < \delta$ für alle $x \in I$ mit $|f(x) - f(a)| < \varepsilon$ oder, wenn ich die Beträge auflöse:

$$a-\delta < x < a+\delta \quad \text{für alle } x \in I \text{ mit } f(a)-\varepsilon < f(x) < f(a)+\varepsilon$$

Wir unterstellen erst mal, daß die Punkte $a \pm \delta$ noch zu I gehören. Wegen

$$f(a-\delta) < f(a) < f(a+\delta)$$

ist dann $\varepsilon := \min\{f(a) - f(a-\delta),\, f(a+\delta) - f(a)\}$ eine zulässige und auch erfolgreiche Wahl, denn aus $f(a)-\varepsilon < f(x) < f(a)+\varepsilon$ folgt $f(a-\delta) < f(x) < f(a+\delta)$ und damit $a-\delta < x < a+\delta$. Wenn aber eine oder beide Zahlen $a \pm \delta$ außerhalb von I liegen, ist die entsprechende Ungleichung für x automatisch erfüllt, und die Wahl von ε vereinfacht sich nur.

Den Nutzen des Satzes illustriert die naheliegende

8.9 Anwendung Für jedes $n > 0$ ist die Wurzelfunktion

$$[0, \infty) \longrightarrow [0, \infty), \quad y \mapsto \sqrt[n]{y}$$

eine stetige Funktion. Bei ungeradem n gilt das auch für

$$\mathbb{R} \longrightarrow \mathbb{R}, \quad y \mapsto \sqrt[n]{y}.$$

Übungsaufgaben

8.1 Denken Sie sich Beispiele von auf Intervallen I definierten stetigen Funktionen $f: I \longrightarrow \mathbb{R}$ aus, die jeweils die folgenden Eigenschaften haben:

(a) I beschränkt, aber nicht abgeschlossen, und f ist beschränkt, nimmt aber kein Maximum an;

(b) I beschränkt, aber nicht abgeschlossen, und f ist nicht beschränkt;

(c) I abgeschlossen, aber nicht beschränkt, und f nimmt kein Maximum an.

Fangen Sie damit an, solche Beispiele nur zu skizzieren, und realisieren Sie die Skizzen dann durch konkrete Formeln.

8.2 Überlegen Sie sich entsprechend Beispiele von stetigen Funktionen $f: X \longrightarrow \mathbb{R}$, wobei $X \subset \mathbb{R}$ kein Intervall ist, und

(a) f injektiv, aber nicht (streng) monoton ist;

(b) f injektiv und monoton, $f^{-1}: f(X) \longrightarrow X$ aber unstetig ist.

8.3 $f: \mathbb{R} \longrightarrow \mathbb{R}$ sei eine Funktion derart, daß die Funktion $|f|: \mathbb{R} \longrightarrow \mathbb{R}$ eine stetige Funktion ist. Natürlich kann man daraus noch nicht auf Stetigkeit von f selbst schließen: man kann die Vorzeichen der Werte $f(x)$ ja beliebig ändern, ohne daß $|f|$ davon etwas merkt. Beweisen Sie aber: Wenn man zusätzlich weiß, daß für jedes Intervall $I \subset \mathbb{R}$ auch $f(I)$ ein Intervall ist, dann folgt, daß f stetig ist.

8.4 Skizzieren Sie eine stetige Funktion $f: \mathbb{R} \longrightarrow \mathbb{R}$, die jede reelle Zahl mindestens zweimal als Wert annimmt. Beweisen Sie: Es gibt aber keine stetige Funktion $f: \mathbb{R} \longrightarrow \mathbb{R}$, die jede reelle Zahl *genau* zweimal als Wert annimmt. (Wer Spaß daran hat, kann noch etwas mehr beweisen: es gibt auch keine stetige Funktion $f: \mathbb{R} \longrightarrow \mathbb{R}$, die jeden *ihrer Werte* genau zweimal annimmt.)

8.5 Im Leben hat man es manchmal mit auf einem Intervall $I \subset \mathbb{R}$ erklärten sogenannten *stückweise linearen* Funktionen $f: I \longrightarrow \mathbb{R}$ zu tun. Diese sind durch folgende Forderung definiert:

Zu jedem $a \in I$ gibt es ein $\delta > 0$ und Zahlen $A, B, C, D \in \mathbb{R}$ mit

$$f(x) = A + Bx \text{ für } x \in I \text{ mit } a - \delta < x \leq a, \quad \text{und} \quad f(x) = C + Dx \text{ für } x \in I \text{ mit } a \leq x < a + \delta$$

(wobei natürlich $A + Ba = C + Da$ sein muß).

Zeigen Sie, daß die Koeffizienten A, B, C, D durch f und a eindeutig bestimmt sind, solange a kein Randpunkt des Intervalls I ist. Beweisen Sie aber vor allem, daß stückweise lineare Funktionen stetig sind. (Dazu ist es bequem, wenn auch nicht unbedingt nötig, schon Begriffe aus Abschnitt 9 zu verwenden.)

8.6 Sei f stückweise linear wie in der vorigen Aufgabe. Einen inneren Punkt $a \in I$ wird man einen "Knick" von f nennen, wenn für die zugehörigen Koeffizienten $B \neq D$ gilt. Beweisen Sie, daß eine stückweise lineare Funktion f auf einem *kompakten* Intervall $[u, v]$ nur endlich viele Knicke haben kann.

Tip: Im Beweis von Satz 8.4 haben Sie gelernt, wie man die Kompaktheit eines Intervalls ausnutzen kann.

8.7 $f: [0, \infty) \longrightarrow \mathbb{R}$ sei eine stetige Funktion, die nur endlich viele Nullstellen hat. Zeigen Sie, daß f dann nach oben beschränkt oder nach unten beschränkt sein muß.

9 Grenzwerte von Funktionen

Die im Titel genannten Grenzwerte sind eng mit dem Begriff der Stetigkeit verwandt.

9.1 Definition Sei $I \subset \mathbb{R}$ ein Intervall. Wir machen die im folgenden häufig vorkommende Voraussetzung, daß I mindestens zwei, und damit unendlich viele Zahlen enthält — künftig wollen wir solche Intervalle echt nennen. Weiter sei $a \in I$ ein fester Punkt, $I' := I \setminus \{a\}$ und

$$f\colon I' \longrightarrow \mathbb{R}$$

eine Funktion. Man schreibt

$$\lim_{x \to a} f(x) = b \in \mathbb{R}$$

und sagt, f habe bei der Annäherung $x \to a$ (kurz: für $x \to a$) den Grenzwert oder Limes b, wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt mit:

$$|f(x) - b| < \varepsilon \quad \text{für alle } x \in I' \text{ mit } |x - a| < \delta$$

Bemerkungen (1) Wie beim Grenzwert einer Zahlenfolge beweist man, daß der Limes b im Falle seiner Existenz durch die übrigen Daten eindeutig bestimmt ist; damit ist die Schreibweise gerechtfertigt.

(2) Die Aussage (in den Bezeichungen der Definition)

$$\lim_{x \to a} f(x) = b$$

ist gleichbedeutend damit, daß die Funktion

$$F\colon I \longrightarrow \mathbb{R}, \quad F(x) := \begin{cases} f(x) & \text{für } x \in I' \\ b & \text{für } x = a \end{cases}$$

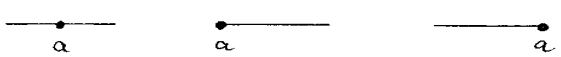
an der Stelle a stetig ist. Das sieht man sofort durch Vergleich der Definitionen.

(3) Setzt man für f eine konkrete Funktion ein, etwa $f(x) = x^3 + 2x$, so entsteht ein Notationsproblem: Der (in jedem Fall korrekten) Formel

$$\lim_{x \to 1} (x^3 + 2x) = 1^3 + 2 \cdot 1 = 3$$

kann man das Intervall I, auf das sich der Limes bezieht, nicht mehr ansehen. Dem helfen wir durch eine genauere Bezeichnung ab. Im Grunde genommen müssen ja nur drei Fälle unterschieden werden: $a \in I$ ist

ein "innerer", oder der linke oder rechte Randpunkt



von I. In der Tat ist der Limes ja offensichtlich eine bei a lokale Bildung, insbesondere merkt er nichts davon, wie lang das Intervall I ist und welche weiteren Randpunkte zu I gehören. Die drei Fälle unterscheidet man nun nötigenfalls als

$$\lim_{x \to a} f(x) \qquad \lim_{x \searrow a} f(x) \qquad \lim_{x \nearrow a} f(x)$$

(Limes von rechts bzw. von links). Mit dem "nötigenfalls" ist gemeint, daß zum Beispiel neben $\lim\limits_{x \searrow 0} \sqrt{x}$ auch

$$\lim_{x \to 0} \sqrt{x}$$

zugelassen sein und dasselbe bedeuten soll, weil hier ja schon der Definitionsbereich der Funktion $x \mapsto \sqrt{x}$ eine linksseitige Annäherung ausschließt. Übrigens kann (wie hier) der Ausdruck für $f(x)$ auch an der Stelle $x=a$ selbst einen Sinn haben. Das muß aber nicht so sein und ist für die Bildung von $\lim\limits_{x\to a} f(x)$ in jedem Fall irrelevant.

Mittels des neuen Limesbegriffs kann man einige einfache Arten von Unstetigkeit einer Funktion genauer beschreiben. Zunächst dürfte nach den Bemerkungen klar sein:

9.2 Notiz Seien $I \subset \mathbb{R}$ ein Intervall, $f\colon I \longrightarrow \mathbb{R}$ eine Funktion, $a \in I$. Dann sind äquivalent:

(a) f ist bei a stetig

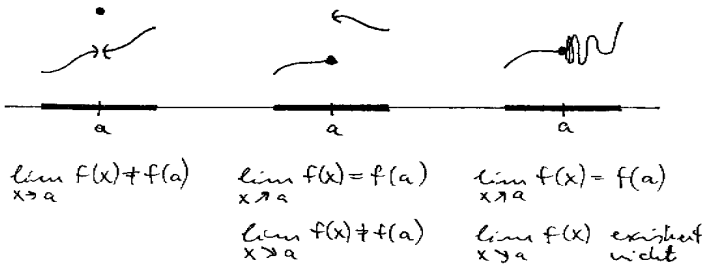
(b) $f|\{x \in I \mid x \le a\}$ und $f|\{x \in I \mid x \ge a\}$ sind bei a stetig

(c) $\lim\limits_{x\to a} f(x) = f(a)$

(d) $\lim\limits_{x\nearrow a} f(x) = f(a) = \lim\limits_{x\searrow a} f(x)$

Dabei sind, wenn a ein Randpunkt von I ist, aus (c) und (d) die nicht sinnvollen Aussagen zu streichen.

Für einen inneren Punkt a von I können wir jetzt beispielsweise folgende Arten der Unstetigkeit unterscheiden:



Natürlich gibt es noch etliche weitere Möglichkeiten.

Auch das Konzept des Grenzwertes selbst ist es für manche Zwecke vorteilhaft zu erweitern, indem man "unendlich", sei es als a oder als Wert $\lim_{x\to a} f(x)$ zuläßt. Dazu die

9.3 Definition Durch

$$[-\infty,\infty] := \{-\infty\} \cup \mathbb{R} \cup \{\infty\},$$

worin $-\infty$ und ∞ zwei zusätzliche abstrakte Elemente sind, erklärt man eine neue Menge $[-\infty,\infty]$, die um eben diese beiden Elemente größer als $\mathbb{R} = (-\infty,\infty)$ ist. Für $\pm\infty$ werden *keine* Rechenoperationen erklärt (weil das nicht sinnvoll möglich ist), wohl aber die Vergleichsrelationen

$$-\infty < x < \infty \quad \text{für jedes } x \in \mathbb{R}.$$

Die grundlegenden Eigenschaften dieser Relation bleiben damit erhalten: Für je zwei Elemente $x, y \in [-\infty,\infty]$ trifft genau eine der Aussagen

$$x < y, \quad x = y, \quad x > y$$

zu, und das Transitivitätsgesetz

$$x < y < z \implies x < z$$

gilt auch für alle $x, y, z \in [-\infty,\infty]$.

Die Definition der Teilmengen

$$(-\infty,\infty], \quad (a,\infty] \quad \text{usw.}$$

liegt jetzt auf der Hand; wir wollen die aber alle nicht als "richtige" Intervalle gelten lassen und lieber von uneigentlichen Intervallen reden.

9.4 Definition I sei ein solches uneigentliches Intervall, und zwar ein echtes, etwa mit $\infty \in I$. Auf $I' := I \setminus \{\infty\} \neq \emptyset$ sei eine Funktion f gegeben. Mit

$$\lim_{x\to\infty} f(x) = b \in \mathbb{R}$$

meint man dann: Zu jedem $\varepsilon > 0$ gibt es ein $D \in \mathbb{R}$ mit

$$|f(x) - b| < \varepsilon \quad \text{für alle } x \in I' \text{ mit } x > D.$$

Bemerkungen Erst mal ist diese Definition das logische Analogon zu 9.1: So wie "nahe an a" durch

$$|x - a| < \delta$$

präzisiert wird, wird "nahe an ∞" durch

$$x > D$$

zum Ausdruck gebracht. Die zweite Beobachtung ist, daß der neue Begriff inhaltlich fast identisch mit dem der Folgenkonvergenz ist. In der Tat ist eine Folge $(x_n)_{n=0}^{\infty}$ dasselbe wie eine auf $\mathbb{N}$ definierte reelle Funktion, und die Definitionen gehen ineinander über, wenn man bloß durchweg $\mathbb{N}$ durch I' ersetzt.

In konkreten Situationen, etwa mit $f(x) = \frac{1}{x}$, gerät man leicht in ein neues Bezeichnungsdilemma. Da in der Formel

$$\lim_{x\to\infty} \frac{1}{x} = 0$$

(zutreffend, setze $D := 1/\varepsilon$) die Wahl des Buchstabens x ganz willkürlich ist, könnte ich genausogut

$$\lim_{n\to\infty} \frac{1}{n} = 0$$

schreiben, was aber eine andere Bedeutung suggeriert, nämlich den Limes der Folge $\left(\frac{1}{n}\right)_n$. Hier hilft man sich mit einer Art impliziter Typzuweisung: Die typischen "Integerbuchstaben" $i, j, \ldots, m$ sollen für natürliche, andere Buchstaben für reelle Zahlen stehen, es sei denn, etwas Anderes ist ausdrücklich vereinbart oder aus dem Zusammenhang evident.

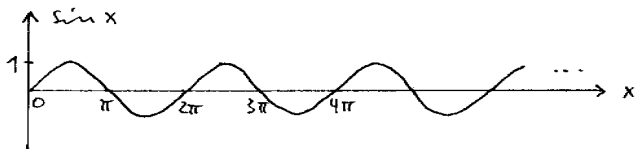
Wie hängen die beiden Limesbegriffe miteinander zusammen? Nun, aus

$$\lim_{x\to\infty} f(x) = b$$

folgt

$$\lim_{n\to\infty} f(n) = b,$$

weil im Folgenfall für weniger Zahlen x (nämlich nur $x = n \in \mathbb{N}$) etwas verlangt wird. Umgekehrt ist das nicht so. Beispiel: Sie werden natürlich schon eine Vorstellung vom Verlauf der Sinusfunktion haben und insbesondere wissen, daß deren Nullstellen die ganzen Vielfachen von $\pi = 180°$ sind:



Ersichtlich ist dann

$$\lim_{n\to\infty} \sin \pi n = \lim_{n\to\infty} 0 = 0,$$

während $\lim_{x\to\infty} \sin \pi x$ nicht existieren kann, da es zu jedem $D \in \mathbb{R}$ sowohl reelle $x > D$ mit $\sin \pi x = 0$ als auch solche mit $\sin \pi x = 1$ gibt.

Die neuen Objekte $\pm\infty$ kommen auch als Grenzwerte in Betracht:

9.5 Definition Seien I ein echtes Intervall, $a \in I$ ein Punkt, $I' := I \setminus \{a\}$ und $f\colon I' \longrightarrow \mathbb{R}$ eine Funktion.

$$\lim_{x \to a} f(x) = \infty$$

bedeutet dann: Zu jedem $E \in \mathbb{R}$ gibt es ein $\delta > 0$ mit

$$f(x) > E \quad \text{für alle } x \in I' \text{ mit } |x - a| < \delta.$$

Natürlich geht das alles entsprechend mit $-\infty$ statt ∞, und schließlich kann man mit der in 9.4 betrachteten Situation kombinieren und

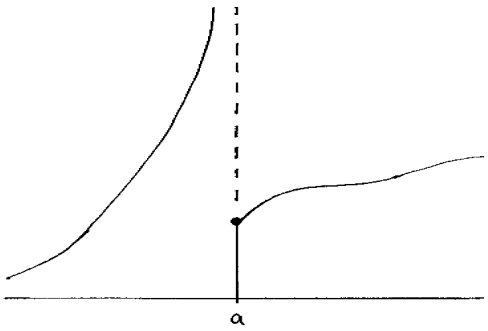
$$\lim_{x \to \infty} f(x) = \infty \quad \text{etc.}$$

erklären, ebenso wie

$$\lim_{n \to \infty} x_n = \infty$$

für eine Folge $(x_n)_{n=0}^{\infty}$. Folgen mit dieser Eigenschaft bezeichnet man aber nach wie vor als *divergent*: Es handelt sich hier, etwa bei $(x_n) = (n)$, nur um eine besondere Art der Divergenz, anders etwa als die Divergenz von $(x_n) = \big((-1)^n\big)$.

Übrigens können wir jetzt noch weitere interessante Typen der Unstetigkeit einer Funktion beschreiben, wie:



Für den Umgang mit all diesen Grenzwerten gibt es wieder tausend Regeln, die man sich nicht alle einzeln merken, aber bei Bedarf leicht überlegen kann. Lohnend sind ohnehin nur die, in denen $\pm\infty$ vorkommt, weil man andernfalls gemäß der Notiz 9.2 die Regeln für stetige Funktionen anwenden kann. Hier nur eine Auswahl, ohne die ganz einfachen Beweise:

9.6 Ein paar **Limesregeln**

(a)
$$\lim_{x \searrow 0} f\left(\frac{1}{x}\right) = \lim_{y \to \infty} f(y)$$

ist so zu lesen: Für ein $a \in \mathbb{R}$ sei $f\colon (a, \infty) \longrightarrow \mathbb{R}$ eine Funktion, dann existiert der eine Limes genau dann, wenn der andere existiert (im "eigentlichen" Sinne als reelle Zahl oder im "uneigentlichen" als $\pm\infty$), und beide sind gleich.

(b) Aus $\lim_{x \to a} f(x) = \infty$ folgt $\lim_{x \to a} \dfrac{1}{f(x)} = 0$. Darf man sich das in der Form

$$\frac{1}{\infty} = 0$$

merken? Meinetwegen, solange Sie nicht glauben, diese isolierte Formel sei nun doch der Anfang des Rechnens mit ∞. Das ist sie nämlich nicht, man darf sie schon nicht von rechts nach links lesen, wie die Formulierung der folgenden Regel deutlich macht.

(c) Aus $\lim\limits_{x\to a} f(x) = 0$ folgt $\lim\limits_{x\to a} \dfrac{1}{f(x)} = \pm\infty$, falls $f(x) > 0$ bzw. $f(x) < 0$ für alle x gilt. Hier entscheidet die Zusatzvoraussetzung nicht nur über das Vorzeichen des Grenzwerts; ohne sie wird der Limes im allgemeinen überhaupt nicht existieren. Sich diese Regel als $\dfrac{1}{0} = \pm\infty$ zu merken zu wollen, wäre schon arg fragwürdig und auch nicht mehr nützlich.

Es sei auch illustriert, daß es schon gar keine Formel vom Typ

$$0 \cdot \infty = ?$$

geben kann: Für jedes vorgegebene $c \in \mathbb{R}$ ist

$$\lim_{x\to 0} cx = 0 \quad \text{sowie} \quad \lim_{x\searrow 0} \frac{1}{x} = \infty,$$

andererseits

$$\lim_{x\searrow 0} \left(cx \cdot \frac{1}{x} \right) = \lim c = c.$$

Im Umgang mit Grenzwerten von Funktionen sind manchmal die sogenannten *Landauschen Symbole* ganz praktisch, die so erklärt sind:

9.6$\frac{1}{3}$ Schreibweise Sei I ein echtes eigentliches oder uneigentliches Intervall, $a \in I$ ein Punkt und wie üblich $I' = I\backslash\{a\}$. Bezogen auf den Grenzübergang $x \to a$ schreibt man für auf I' erklärte reelle Funktionen f und h

$$f(x) = o\big(h(x)\big) \quad \text{oder} \quad f(x) = O\big(h(x)\big)$$

(gesprochen *klein-oh* und *groß-oh*), wenn $\lim\limits_{x\to a} \dfrac{f(x)}{h(x)} = 0$ ist bzw. wenn $x \mapsto \dfrac{f(x)}{h(x)}$ bei der Annäherung $x \to a$ immerhin (definiert ist und) beschränkt bleibt. Mit letzterem ist, etwa für endliches $a \in \mathbb{R}$, natürlich gemeint, daß es ein $\delta > 0$ und ein $E \in \mathbb{R}$ gibt, so daß

$$h(x) \neq 0 \text{ und } \left| \frac{f(x)}{h(x)} \right| < E \quad \text{für alle } x \in I' \text{ mit } |x - a| < \delta$$

gilt.

Ferner nennt man f und h *asymptotisch gleich* (für $x \to a$) und schreibt

$$f \sim h,$$

wenn $\lim\limits_{x\to a} \dfrac{f(x)}{h(x)} = 1$ ist.

Klar, daß diese traditionelle Schreibweise mit den Landauschen Symbolen das Gleichheitszeichen vergewaltigt, denn aus den korrekten Aussagen $x^2 = o(x)$ und $x^3 = o(x)$ für $x \to 0$ folgt ja in keiner Weise $x^2 = x^3$. Eine modernere, logisch einwandfreie und in der Informatik vorgezogene Variante sieht deshalb o und O als Bezeichung für durch die Funktion h definierte Mengen von Funktionen an und schreibt

$$f(x) \in o\big(h(x)\big) \quad \text{oder} \quad f(x) \in O\big(h(x)\big)\,.$$

Beim traditionellen Gebrauch der Symbole hilft das aber nicht weiter, weil man die Symbole in der Regel zum Vergleich von f mit einer weiteren Funktion $g\colon I \longrightarrow \mathbb{R}$ verwendet und statt $f(x) - g(x) = o\big(h(x)\big)$ dann

$$f(x) = g(x) + o\big(h(x)\big)$$

schreibt. Etwa

$$(a+x)^3 = a^3 + 3a^2x + O(x^2) \quad (x \to 0)$$
$$(a+x)^3 = x^3 + 3ax^2 + o(x^2) \quad (x \to \infty)$$

sind zwei typische Beispiele. Stellen Sie sich am besten vor, daß der Zusatz "$+o\big(h(x)\big)$" das voranstehende Gleichheitszeichen abschwächt, nämlich zu Gleichheit bis auf einen Fehlerterm, über dessen Verhalten bei Annäherung an a eine Aussage gemacht wird. Trotz der fragwürdigen Verwendung des Gleichheitszeichens kann man mit den Landauschen Symbolen ebenso wie mit dem "$\sim$" sicher und vorteilhaft rechnen, und es gelten viele leicht einzusehende Regeln, von denen hier nur einige zitiert seien:

9.6$\frac{2}{3}$ Auswahl von **Regeln** (a) Für $m, n \in \mathbb{N}$ folgt aus $f(x) = O(x^m)$ und $g(x) = O(x^n)$ für $x \to 0$

$$f(x) + g(x) = O\left(x^{\min(m,n)}\right) \text{ für } x \to 0;$$

wenn es um dagegen durchweg um die Annäherung $x \to \infty$ geht, folgt

$$f(x) + g(x) = O\left(x^{\max(m,n)}\right) \text{ für } x \to \infty$$

(nur die jeweils schlechtere Abschätzung überträgt sich auf die Summe).

(b) Aus $f(x) = O\big(h_1(x)\big)$ und $g(x) = O\big(h_2(x)\big)$ folgt

$$(f \cdot g)(x) = O\big(h_1 h_2(x)\big);$$

um o statt O zu schließen, genügt es, wenn o für einen der Faktoren gilt.

(c) $f \sim g$ impliziert $f = O(g)$ und $g = O(f)$ (aber nicht umgekehrt). Für die asymptotische Äquivalenz gilt neben $f \sim f$ sowie $f \sim g \iff g \sim f$ auch das Transitivitätsgesetz: Aus $f \sim g$ und $g \sim h$ folgt $f \sim h$.

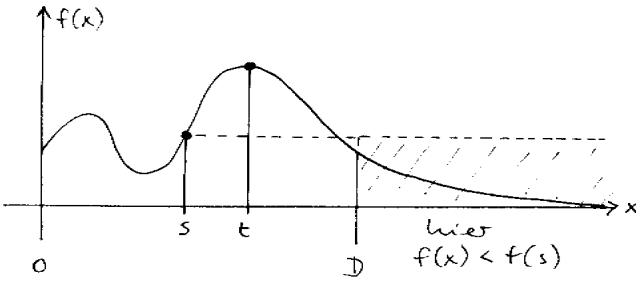
Zum Schluß dieses Abschnittes will ich Ihnen an einem Beispiel illustrieren, wie man die Wissenschaft von den Grenzwerten manchmal geschickt einsetzen kann, um den Anwendungsbereich der Sätze über stetige Funktionen auf Intervallen zu erweitern.

9.7 Lemma $f\colon [0,\infty) \longrightarrow \mathbb{R}$ sei eine stetige Funktion mit

$$f(x) \geq 0 \quad \text{für alle } x \in [0,\infty), \quad \text{und}$$
$$\lim_{x\to\infty} f(x) = 0.$$

Dann hat die Funktion f ein Maximum.

Beweis Satz 8.4 läßt sich nicht direkt anwenden, weil $[0,\infty)$ nicht kompakt ist. Mit einem kleinen Trick geht's aber doch: Den trivialen Fall $f = 0$ beiseite lassend, wählen wir irgendein $s \in [0,\infty)$ mit $f(s) > 0$. Dann finden wir ein $D \geq 0$ mit $f(x) < f(s)$ für alle $x > D$, und wir dürfen natürlich $D \geq s$ annehmen.



Nun nimmt die Einschränkung $f|[0,D]$ nach Satz 8.4 ihr Maximum an, etwa bei $t \in [0,D]$. Es gilt also $f(x) \leq f(t)$ für alle $x \in [0,D]$. Aber dank unserer Vorarbeit gilt es für die übrigen x, also $x \in (D,\infty)$ auch:

$$f(x) < f(s) \leq f(t), \quad \text{letzteres wegen } s \in [0,D]$$

Damit ist $f(t)$ der größte Wert von f überhaupt.

Übungsaufgaben

9.1 Sei $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ eine monotone Funktion, und $a \in \mathbb{R}$ beliebig. Beweisen Sie: Wenn

$$\lim_{n\to\infty} f\left(a-\tfrac{1}{n}\right) = f(a) \quad \text{und} \quad \lim_{n\to\infty} f\left(a+\tfrac{1}{n}\right) = f(a)$$

gilt, dann ist f an der Stelle a stetig.

9.2 $f\colon [0,\infty) \longrightarrow \mathbb{R}$ sei eine stetige Funktion ohne Nullstellen mit

$$\lim_{x\to\infty} |f(x)| = \infty.$$

Zeigen Sie, daß der Kehrwert $g\colon [0,\infty) \longrightarrow \mathbb{R},\ g(x) = \frac{1}{f(x)}$ eine beschränkte Funktion ist.

9.3 Untersuchen Sie analog zu (und in den Bezeichnungen von) Satz 3.10 die Grenzwerte

$$\lim_{x\to\pm\infty} \frac{f(x)}{g(x)}$$

einer rationalen Funktion f/g. Beachten Sie, daß man dank der inzwischen eingeführten Objekte $\pm\infty$ jetzt auch im Fall $d > e$ eine positive Konvergenzaussage machen kann. Wie könnte man Resultat und Beweis dieser Überlegungen mittels der Symbole o, O und $\sim$ fassen?

10 Komplexe Zahlen, Grenzwerte und Funktionen

An sich ist die Analysis komplexer Zahlen, traditionell Funktionentheorie genannt, Thema einer eigenen Vorlesung. In manchen Bereichen hat die Funktionentheorie jedoch einen so starken Einfluß auf die reelle Analysis, daß es ganz dumm wäre, sich strikt auf diese zu beschränken. So würde allein schon die Vereinfachung beim Umgang mit Cosinus und Sinus den Aufwand rechtfertigen, sich mit den komplexen Zahlen zu befassen. Was sind also komplexe Zahlen? Leider hat das erste, was man über komplexe Zahlen zu hören bekommt, meist mit dem famosen $i = \sqrt{-1}$ zu tun. Daß man damit schnell auf schwankenden Grund gerät — einerseits sollte ja

$$i^2 = \left(\sqrt{-1}\right)^2 = -1,$$

andererseits

$$i^2 = \sqrt{-1} \cdot \sqrt{-1} = \sqrt{(-1)^2} = \sqrt{1} = +1$$

sein — hat viel zum schlechten Ruf der komplexen Zahlen beigetragen. Der ist aber ganz unverdient; komplexe Zahlen haben zunächst gar nichts mit einer Wurzel aus -1 zu tun und sind auch ganz einfach zu erklären, wie Sie jetzt sehen werden. Ebenso wie die rationalen und die reellen bilden die komplexen Zahlen einen Körper; er wird mit dem Sondersymbol $\mathbb{C}$ bezeichnet. Es folgt eine Konstruktion von $\mathbb{C}$ auf der Basis der uns ja hinreichend vertrauten reellen Zahlen.

10.1 Konstruktion des Körpers der **komplexen Zahlen** Als Menge ist schlicht und einfach

$$\mathbb{C} = \mathbb{R}^2 :$$

eine komplexe Zahl ist ein Paar $z = (x, y)$ reeller Zahlen. Die beiden Komponenten heißen Real- und Imaginärteil

$$x = \mathrm{Re} z \quad \text{und} \quad y = \mathrm{Im} z$$

von z. Um $\mathbb{C}$ zu einem Körper zu machen, müssen wir eine Addition und eine Multiplikation erklären. Die Addition geschieht komponentenweise: für z wie eben und $z' = (x', y')$ ist

$$z + z' = (x, y) + (x', y') := (x + x', y + y').$$

Man sieht sofort, daß diese Addition $\mathbb{C}$ zu einer abelschen Gruppe mit der Zahl $(0, 0) \in \mathbb{C}$ als Null macht, einfach weil $(\mathbb{R}, +)$ eine solche ist.

Die Multiplikation komplexer Zahlen wird nicht komponentenweise, sondern durch

$$zz' = (x, y) \cdot (x', y') := (xx' - yy' \,,\, xy' + yx')$$

erklärt. Zu verifizieren, daß $\mathbb{C}$ damit zu einem kommutativen Ring wird, macht etwas mehr Mühe, aber keine Schwierigkeiten. Als Einselement fungiert jedenfalls $(1, 0) \in \mathbb{C}$:

$$(1, 0) \cdot z = (1, 0) \cdot (x, y) := (1x - 0y \,,\, 1y + 0x) = z \quad \text{für jedes } z = (x, y) \in \mathbb{C}.$$

So weit hätten übrigens auch viele andere denkbare Definitionen der Multiplikation geführt. Die Besonderheit derer, die wir tatsächlich genommen haben, ist, daß der so erklärte Ring sogar ein Körper ist. Um das zu zeigen, müssen wir zu jeder komplexen Zahl $z \in \mathbb{C} \setminus \{(0, 0)\}$ eine zu ihr inverse vorweisen. Das geht so: Die für beliebiges $z = (x, y) \in \mathbb{C}$ definierte reelle Zahl $x^2 + y^2 \in [0, \infty)$ ist im Fall $z \neq (0, 0)$ positiv, so daß wir mit ihrer Hilfe

$$\left(\frac{x}{x^2 + y^2} \,,\, \frac{-y}{x^2 + y^2} \right) \in \mathbb{C}$$

bilden können, und diese komplexe Zahl ist tatsächlich die Inverse z^{-1}:

$$(x,y)\cdot\left(\frac{x}{x^2+y^2}\,,\,\frac{-y}{x^2+y^2}\right)=\left(\frac{x\cdot x-y\cdot(-y)}{x^2+y^2}\,,\,\frac{x\cdot(-y)-y\cdot x}{x^2+y^2}\right)=(1,0)$$

Damit ist die Konstruktion des Körpers $\mathbb{C}$ abgeschlossen.

Die komplexen Zahlen mit verschwindendem Imaginärteil bilden, wie die Formeln

$$(x,0)+(x',0)=(x+x',0)$$
$$(x,0)\cdot(x',0)=(xx',0)$$

zeigen, einen Unterring von $\mathbb{C}$, den wir vermöge

$$\mathbb{R}\ni x\mapsto(x,0)\in\mathbb{C}$$

mit dem Körper der reellen Zahlen identifizieren dürfen; insbesondere schreiben wir jetzt einfach $0\in\mathbb{C}$ und $1\in\mathbb{C}$ statt $(0,0)$ und $(1,0)$. Um bei allen komplexen Zahlen auf die schwerfälligen Klammern verzichten zu können, führt man die Abkürzung

$$i:=(0,1)\in\mathbb{C}$$

ein; so läßt sich jede komplexe Zahl $z\in\mathbb{C}$ klammerfrei durch Real- und Imaginärteil ausdrücken:

$$z=x+iy\quad\text{mit } x=\mathrm{Re}z,\ \ y=\mathrm{Im}z$$

(So ist es in der Regel gemeint, wenn Sie irgendwo "sei $z=x+iy$ eine komplexe Zahl..." lesen, auch wenn nicht ausdrücklich gesagt wird, daß x und y *reell* und nicht komplex sind.) Die Multiplikationsregel für komplexe Zahlen, die damit ja

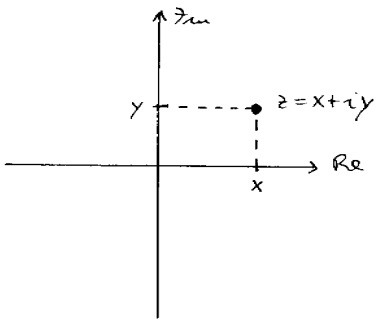
$$(x+iy)\,(x'+iy')=xx'-yy'+i\,(xy'+yx')$$

lautet, gibt insbesondere die berühmte Formel

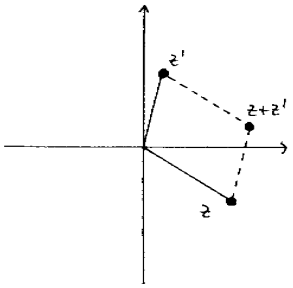
$$i^2=-1$$

und folgt umgekehrt aus ihr nach dem Distributivgesetz. Die eingangs erwähnten Halbwahrheiten betreffend die komplexen Zahlen beruhen auf dem Versuch, diese Formel gewaltsam als $i=\sqrt{-1}$ umzuschreiben. Gerade das läßt man aber besser bleiben; in $\mathbb{C}$ gibt es eben nicht ohne weiteres (wohldefinierte) Wurzeln.

Die komplexen Zahlen, die ja Paare reeller Zahlen sind, veranschaulicht man sich geometrisch natürlich am besten als Punkte der Ebene, die in dieser Bedeutung die *Gaußsche Zahlenebene* heißt.



Die Addition komplexer Zahlen entspricht dann dem, was man üblicherweise Vektoraddition nennt:



Auf eine Veranschaulichung der Multiplikation kommen wir noch zu sprechen.

Für das Rechnen in $\mathbb{C}$ sind noch zwei zusätzliche Bildungen wichtig:

10.2 Definition Sei $z = x + iy \in \mathbb{C}$. Dann heißt

$$\overline{z} = x - iy \in \mathbb{C}$$

die zu z (komplex-)konjugierte Zahl, und

$$|z| = \sqrt{x^2 + y^2} \in [0, \infty)$$

der Absolutbetrag oder kurz Betrag von z.

Dazu die

10.3 Regeln Real- und Imaginärteil der komplexen Zahl z lassen sich vermöge

$$\mathrm{Re} z = \frac{1}{2}(z + \overline{z}) \quad \text{und} \quad \mathrm{Im} z = \frac{1}{2i}(z - \overline{z}) \quad \text{für alle } z \in \mathbb{C},$$

mittels der Konjugation ausdrücken; insbesondere ist z genau dann reell, wenn $z = \overline{z}$ ist. Für $w, z \in \mathbb{C}$ gilt

$$\overline{\overline{z}} = z, \quad \overline{w + z} = \overline{w} + \overline{z} \quad \text{und} \quad \overline{wz} = \overline{w}\,\overline{z}.$$

Es ist $|z| = 0$ genau dann, wenn $z = 0$ ist. Der Absolutbetrag erfüllt

$$|z|^2 = |\overline{z}|^2 = \overline{z}z,$$

woraus sich im Fall $z \neq 0$ mit

$$\frac{1}{z} = \frac{\overline{z}}{|z|^2}$$

eine Darstellung des Kehrwertes ergibt, die einfacher zu handhaben ist als die oben angegebene. Für $w, z \in \mathbb{C}$ gilt

$$|wz| = |w||z|$$

sowie die Dreiecksungleichung

$$|w \pm z| \ \leq \ |w| + |z|.$$

Für reelle $z = x + i0$ schließlich stimmt $|z| = \sqrt{x^2} = \pm z$ mit dem reellen Absolutbetrag überein.

Beweis Folgt alles sofort aus den Definitionen, mit Ausnahme der Dreiecksungleichung. Für diese bemerken wir vorweg, daß für jedes $z = x + iy \in \mathbb{C}$ die Ungleichung

$$\mathrm{Re} z = x \leq |x| = \sqrt{x^2} \leq \sqrt{x^2 + y^2} = |z|,$$

und für die gegebenen w, z deshalb

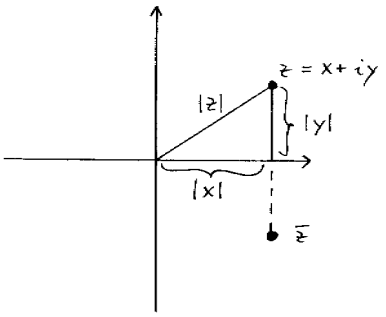
$$\mathrm{Re}(\overline{w}z) \leq |\overline{w}z| = |\overline{w}||z| = |w||z|$$

gilt. Daraus ergibt sich

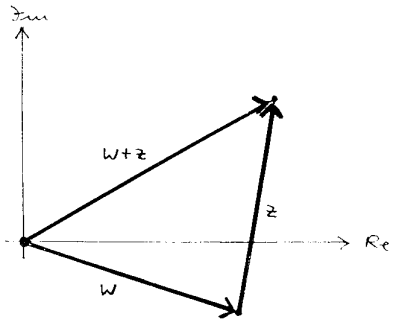
$$\begin{aligned}|w+z|^2 &= (\overline{w}+\overline{z})(w+z)\\ &= \overline{w}w + \overline{w}z + \overline{z}w + \overline{z}z\\ &= |w|^2 + 2\mathrm{Re}(\overline{w}z) + |z|^2\\ &\le |w|^2 + 2|w||z| + |z|^2 = (|w|+|z|)^2,\end{aligned}$$

und es bleibt bloß die (reelle) Wurzel zu ziehen.

Geometrisch gesehen, spiegelt die komplexe Konjugation $z \mapsto \overline{z}$ die Gaußsche Zahlenebene an der reellen Achse. Der Betrag von $z \in \mathbb{C}$ ist nach dem Satz von Pythagoras der Abstand des Punktes z vom Nullpunkt,



und der Name "Dreiecksungleichung" wird jetzt geometrisch verständlich: die Länge einer Dreiecksseite ist höchstens die Summe der Längen der beiden anderen Seiten (wobei Gleichheit nur bei einem entarteten Dreieck eintreten kann):



Was fehlt dem Körper $\mathbb{C}$ im Vergleich zu $\mathbb{R}$? Es fehlt die Anordnung, also die Möglichkeit, komplexe Zahlen der Größe nach miteinander zu vergleichen. Auf $\mathbb{C}$ kann man keine Anordnung erklären, die Axiomen analog 2.10 genügen würden: In einer solchen Anordnung müßten gemäß Regel 2.11(d) alle Quadrate nicht-negativ sein, insbesondere $1 = 1^2 > 0$, aber auch $-1 = i^2 > 0$ sein, was ja nicht beides sein kann.

Auf der positiven Seite hat der Körper der komplexen Zahlen aber etwas Gewichtiges vorzuweisen; für komplexe Polynome

$$\mathbb{C} \ni z \mapsto \sum_{k=0}^{d} a_k z^k \in \mathbb{C} \quad \text{(mit Konstanten } a_0, a_1, \ldots, a_d \in \mathbb{C})$$

gilt, anders als für reelle, der sogenannte Fundamentalsatz der Algebra. Es liegt weitgehend an diesem Satz, daß "komplex" oft viel einfacher als "reell" ist.

10.4 Fundamentalsatz der Algebra Jedes normierte komplexe Polynom

$$z \mapsto f(z) = \sum_{k=0}^{d} a_k z^k \quad (a_d = 1)$$

zerfällt in komplexe Linearfaktoren: Es gibt paarweise verschiedene Zahlen $c_1, c_2, \dots, c_r \in \mathbb{C}$ und Exponenten $e_1, e_2, \dots, e_r \in \mathbb{N} \setminus \{0\}$, so daß

$$f(z) = (z - c_1)^{e_1}(z - c_2)^{e_2} \cdots (z - c_r)^{e_r}$$

für alle $z \in \mathbb{C}$ ist.

Bemerkungen Der Name stammt aus einer Zeit, als man unter Algebra allein die Wissenschaft des Auflösens von Polynomgleichungen verstand. Aus heutiger Sicht ist der Satz zwar wichtig, aber nicht fundamental. Er ist ironischerweise auch gar kein Satz der Algebra, sondern der Geometrie, und der populärste Beweis benutzt eine analytische Methode; dieser Beweis kommt in jeder einführenden Vorlesung über Funktionentheorie vor. Wir wollen den Satz hier einfach als wahr akzeptieren und uns mit einigen Erläuterungen zufriedengeben.

Die in der Zerlegung von f vorkommenden komplexen Zahlen c_j $(j = 1, \dots, r)$ sind natürlich genau die Nullstellen von f, deren jede aber noch eine Vielfachheit (oder Ordnung) $e_j > 0$ hat. Deswegen ist auch nicht unbedingt $r = d$, vielmehr gilt offenbar

$$e_1 + e_2 + \cdots + e_r = d.$$

Es ist ganz leicht zu sehen, daß die Vielfachheiten ebenso wie die Nullstellen selbst durch f eindeutig bestimmt sind; damit ist auch die gesamte Linearfaktorzerlegung abgesehen von der Reihenfolge der Faktoren eindeutig.

Die wichtigste Botschaft dieses Abschnitts ist die: In $\mathbb{C}$ kann man praktisch genauso Analysis treiben wie in $\mathbb{R}$. Auf den ersten Blick mag das verwundern, beruhen doch alle analytischen Begriffe, die wir eingeführt haben, wesentlich auf der Anordnung der reellen Zahlen. Es geht aber doch, beginnend mit:

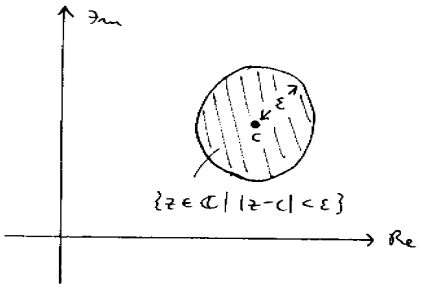
10.5 Definition Sei $(z_n)_{n=0}^{\infty}$ eine Folge komplexer Zahlen, und sei $c \in \mathbb{C}$. Mit

$$\lim_{n\to\infty} z_n = c$$

meint man dann: Zu jeder reellen Zahl $\varepsilon > 0$ gibt es ein $D \in \mathbb{N}$ mit

$$|z_n - c| < \varepsilon \quad \text{für alle } n \in \mathbb{N} \text{ mit } n > D.$$

Was ist im Vergleich zur reellen Version neu? Formal gar nichts! Zur Beschreibung der Konvergenz braucht man nicht die Zahlen $z_n - c$ selbst, sondern nur ihre Absolutbeträge mit der reellen Zahl ε zu vergleichen, und die sind ja wie bisher reell. Inhaltlich besteht die Neuerung in der erweiterten Bedeutung des Betrags.



Während man $|x_n - a| < \varepsilon$ im reellen Fall als eine Kurzform von $-\varepsilon < x_n - a < \varepsilon$ auffassen konnte, ist das komplexe $|z_n - c| < \varepsilon$ mehr geometrischer Natur: z_n hat von c einen Abstand kleiner als ε, liegt also im Inneren des Kreises um c mit Radius ε. Der präzisen Verständigung halber legen wir fest:

10.6 Definition Sei $c \in \mathbb{C}$ und $\varepsilon > 0$ (letzteres soll automatisch die Forderung $\varepsilon \in \mathbb{R}$ einschließen). Dann heißt

$$U_\varepsilon(c) := \big\{ z \in \mathbb{C} \,\big|\, |z - c| < \varepsilon \big\}$$

die offene Kreisscheibe um c vom Radius ε, und

$$D_\varepsilon(c) := \{z \in \mathbb{C} \mid |z - c| \leq \varepsilon\}$$

entsprechend die abgeschlossene Kreisscheibe.

Bei der Übertragung von Resultaten aus der reellen auf die komplexe Analysis gibt es zwei Standardmethoden. Eine besteht darin, die Begriffe und Beweise einfach zu inspizieren und festzustellen, daß alles mehr oder weniger wörtlich genau so geht. Das ist, wie Sie gerade gesehen haben, bei der Definition des Folgenlimes so, und tatsächlich häufig der Fall (wenn auch nicht immer). Die andere Methode besteht darin, eine komplexe Fragestellung durch Zerlegung in Real- und Imaginärteil auf die entsprechende reelle zurückzuführen. Das funktioniert beispielsweise beim

10.7 Lemma Sei $(z_n = x_n + iy_n)_{n=0}^\infty$ eine komplexe Folge, $c = a + ib \in \mathbb{C}$. Dann gilt:

$$\lim z_n = c \iff \lim x_n = a \text{ und } \lim y_n = b$$

Beweis Gelte $\lim z_n = c$. Zu jedem $\varepsilon > 0$ gibt es also ein D mit

$$|z_n - c| < \varepsilon \quad \text{für alle } n > D.$$

Wegen $|x_n - a|^2 + |y_n - b|^2 = |z_n - c|^2$ gilt für diese n erst recht

$$|x_n - a| \leq |z_n - c| < \varepsilon \quad \text{und} \quad |y_n - b| \leq |z_n - c| < \varepsilon,$$

woraus sich die Konvergenz der beiden reellen Folgen sofort ergibt.

In der umgekehrten Richtung impliziert

$$|x_n - a| < \varepsilon \quad \text{und} \quad |y_n - b| < \varepsilon$$

die Abschätzung

$$|z_n - c|^2 = |x_n - a|^2 + |y_n - b|^2 < 2\varepsilon^2,$$

d.h.

$$|z_n - c| < \sqrt{2}\,\varepsilon,$$

und das ist für die Konvergenz von (z_n) gut genug.

Mit diesem bescheidenen Werkzeug gewappnet, sieht man ohne große Mühe folgendes. Von dem, was wir über Konvergenz von reellen Folgen und Reihen gesagt haben, bleibt für komplexe Folgen und Reihen das richtig, was man überhaupt sinnvoll formulieren kann. Dazu gehören insbesondere *nicht* Aussagen über monotone Folgen, auch über Supremum und Infimum, sowie die uneigentlichen Grenzwerte $\pm\infty$. Problemlos übertragen sich dagegen die Begriffe und zugehörigen Sätze "beschränkte Mengen, Folgen (und auch Funktionen)": zwar verliert beispielsweise für eine komplexwertige Funktion f

$$a \leq f(z) \leq b \quad \text{für alle } z$$

seinen Sinn, nicht aber

$$|f(z)| \leq c \quad \text{für alle } z.$$

Mittels und analog zu Lemma 10.7 gelten für komplexe Folgen das Konvergenzkriterium von Cauchy ebenso wie der Satz von Bolzano und Weierstraß; daß wir letzteren mittels monotoner (reeller) Folgen als Hilfsmittel bewiesen hatten, tut dem keinen Abbruch.

Aufgrund der Gültigkeit des Cauchy-Kriteriums übertragen sich auch die wichtigen Resultate über Reihen auf den komplexen Fall. Insbesondere impliziert absolute Konvergenz wieder die gewöhnliche, wobei erstere ja eine Eigenschaft einer *reellen* Reihe ist:

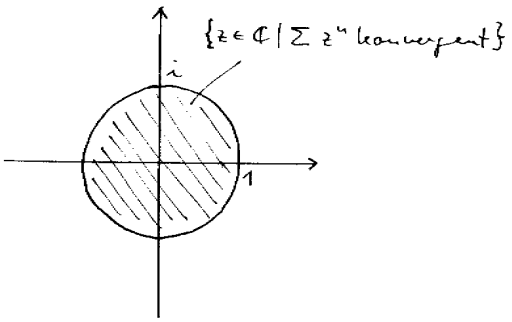
$$\sum_{n=0}^\infty |z_n| \text{ konvergent} \implies \sum_{n=0}^\infty z_n \text{ konvergent}$$

Richtig bleiben auch die Kriterien für absolute Konvergenz (Majoranten/Minoranten, Quotientenkriterium) sowie die Resultate, die mit der Umordnung absolut konvergenter Reihen und Mehrfachreihen zu tun haben.

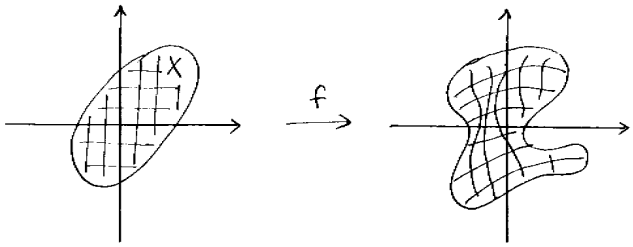
Das wieder wichtigste Beispiel überhaupt ist die geometrische Reihe $\sum z^n$, die für $|z| < 1$ mit Summe

$$\sum_{n=0}^{\infty} z^n = \frac{1}{1-z}$$

konvergiert, für $|z| \geq 1$ dagegen divergieren muß, weil dann nicht $\lim z^n = 0$ ist.



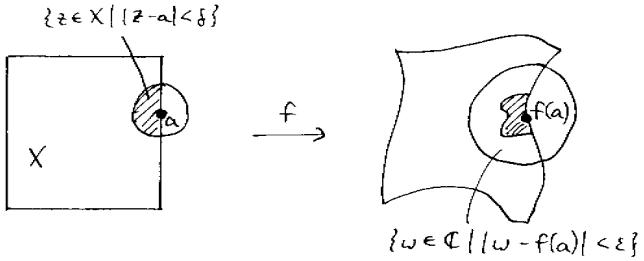
Nun zu komplexen Funktionen:



Die Definition des Begriffs "stetig" überträgt sich ganz automatisch.

10.8 Definition Seien $X \subset \mathbb{C}$ eine Teilmenge, $f\colon X \longrightarrow \mathbb{C}$ eine Funktion und $a \in X$ ein Punkt. Man nennt f bei a <u>stetig</u>, wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt mit

$$|f(z) - f(a)| < \varepsilon \quad \text{für alle } z \in X \text{ mit } |z - a| < \delta.$$

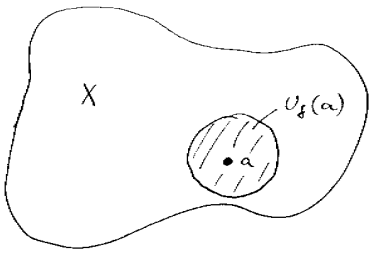


Wenn Sie wollen, können Sie das auch ganz geometrisch fassen: Zu jeder offenen Kreisscheibe E um $f(a)$ gibt es eine offene Kreisscheibe D um a mit $f(X \cap D) \subset E$. Beachten Sie übrigens, daß die Definition sich für den ja nicht verbotenen Fall, daß $X \subset \mathbb{R}$ und auch $f(X) \subset \mathbb{R}$ ist, auf die alte, reelle Definition reduziert.

Die Übertragung unserer Regeln für stetige Funktionen ins Komplexe ist evident. Ernsthaft nachdenken muß man erst, wenn man sich die drei Sätze über stetige Funktionen auf Intervallen vornimmt; das stellen wir bis zu dem Zeitpunkt zurück, wo wir ohnehin über stetige Funktionen *mehrerer* reeller Veränderlicher

reden. Trotzdem: Was entspricht wohl im Komplexen einem Intervall? Man könnte an die offenen und abgeschlossenen Kreisscheiben als die Analoga der entsprechenden Intervalle denken. In der Praxis erweist sich das aber als zu speziell, und nützlicher sind die folgenden Begriffe.

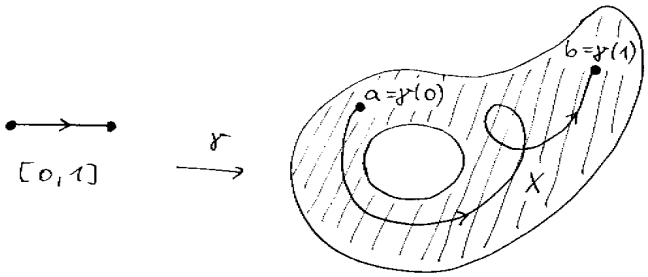
10.9 Definition Eine Teilmenge $X \subset \mathbb{C}$ heißt offen, wenn es zu jedem $a \in X$ ein $\delta > 0$ gibt mit $U_\delta(a) \subset X$.



Eine Teilmenge $X \subset \mathbb{C}$ heißt zusammenhängend, wenn es zu je zwei Punkten $a, b \in X$ einen Weg in X von a nach b gibt, nämlich eine stetige Funktion

$$\gamma\colon [0,1] \longrightarrow X$$

mit $\gamma(0) = a$ und $\gamma(1) = b$.



Eine Menge $X \subset \mathbb{C}$, die offen ist und zusammenhängt, nennt man ein Gebiet.

10.10 Beispiel Jede offene Kreisscheibe $U_r(c)$ ist ein Gebiet: Die Dreiecksungleichung $|z - c| \le |z - a| + |a - c|$ beweist die Inklusion

$$U_{r-|a-c|}(a) \subset U_r(c) \quad \text{für jedes } a \in U_r(c)$$

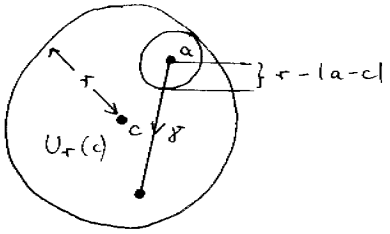
und damit die Offenheit von $U_r(c)$. Zum Beweis des Zusammenhangs verbindet man zwei Punkte $a, b \in U_r(c)$ in $U_r(c)$ einfach durch die Strecke zwischen ihnen, nämlich den Weg

$$\gamma\colon [0,1] \longrightarrow U_r(c); \quad t \mapsto (1-t)\,a + tb.$$

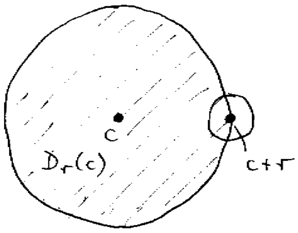
Daß dieser Weg tatsächlich ganz in $U_r(c)$ verläuft, ist elementare Geometrie; wenn Sie einen Beweis wünschen, lesen Sie einfach die Zeile

$$|\gamma(t)-c| = |(1-t)a + tb - c| = |(1-t)(a-c) + t(b-c)| \le (1-t)|a-c| + t|b-c| < (1-t)r + tr = r$$

für alle $t \in [0,1]$.



Aus dem gleichen Grund hängt auch die abgeschlossene Kreissscheibe $D_r(c)$ zusammen; sie ist aber nicht offen, weil zum Beispiel der Punkt $c + r \in D_r(c)$ offenbar kein $\delta > 0$ mit $U_\delta(c + r) \subset D_r(c)$ zuläßt.



Der Begriff des Gebiets ist jedenfalls weit genug gefaßt, um die natürlichen Definitionsbereiche der komplexen rationalen Funktionen zu umfassen.

10.11 Satz Jede komplexe rationale Funktion

$$z \mapsto \frac{f(z)}{g(z)} \quad (f \text{ und } g \neq 0 \text{ komplexe Polynome})$$

ist auf einem solchen Gebiet erklärt (und stetig, nach der komplexen Version des Satzes 7.6), nämlich auf

$$G := \{z \in \mathbb{C} \mid g(z) \neq 0\},$$

also der komplexen Ebene, aus der man die endlich vielen Nennernullstellen entfernt hat.

Beweis Eine lustige Übung (Aufgabe 10.7 oder 10.8)

Natürlich läßt sich jedes reelle Polynom und jede reelle rationale Funktion auch komplex lesen; wenn es auf den Unterschied ankommt, macht man ihn häufig nur durch die Wahl des Variablennamens x oder z deutlich. Jedenfalls hat der komplexe Standpunkt oft Vorteile, zum Beispiel weil man jede *komplexe* rationale Funktion mittels der sogenannten Partialbruchzerlegung in eine besonders gut zu handhabende Form bringen kann. Dazu zwei Vorbemerkungen. Zunächst kann es passieren, daß in einer rationalen Funktion

$$z \mapsto h(z) = \frac{f(z)}{g(z)}$$

die Polynome f und $g \neq 0$ eine gemeinsame Nullstelle c haben. Dann läßt sich der Bruch f/g durch den Linearfaktor $z - c$ kürzen, wobei sich der Definitionsbereich von r um den Punkt c vergrößern (und damit streng genommen eine neue Funktion entstehen) kann. Wenn man das — nötigenfalls wiederholt — gemacht hat, haben f und g keine gemeinsamen Nullstellen mehr, und in diesem Fall nennt man die verbleibenden Nullstellen von g die Polstellen von r. Sie gehören natürlich nicht zum Definitionsbereich von r, sind nach den Limesregeln vielmehr durch

$$\lim_{z \to c} |h(z)| = \infty \quad \text{für jede Polstelle } c \text{ von } r$$

charakterisiert. Genauer kann man von Polstellen der Ordnung e sprechen, wenn c eine e-fache Nullstelle von g ist.

In der zweiten Vorbemerkung möchte ich an die bekannte Tatsache erinnern, daß man Polynome durcheinander mit Rest teilen kann: Zu gegebenen Polynomen f und $g \neq 0$ gibt es eindeutig bestimmte Polynome q und r mit

$$f = g \cdot q + r$$

und $\deg r < \deg g$ (wir wollen spitzfindigerweise vereinbaren, daß eine Abschätzung von $\deg r$ nach oben auch dann als wahr gelten soll, wenn r das Nullpolynom ist). Quotient q und Rest r lassen sich nach dem gleichen Schema berechnen, nach dem man auch Dezimalzahlen teilt, zum Beispiel:

$$\begin{array}{l}
(z^3 + z^2 \;\; - z + 1)/(z^2 + 1) \;\longrightarrow\; z + 1 = q(z) \\
\underline{z^3 \qquad\;\; + z} \\
\qquad z^2 - 2z + 1 \\
\qquad \underline{z^2 \qquad + 1} \\
\qquad\quad -2z = r(z)
\end{array}$$

Die Polynomdivision erlaubt es, jede gegebene rationale Funktion $h = \dfrac{f}{g}$ als

$$h = \frac{g\,q + r}{g} = q + \frac{r}{g}$$

zu schreiben, d.h. durch Abspalten eines Polynoms kann man den Grad des Zählers immer kleiner als den des Nenners machen. Für Funktionen dieses letzten Typs gilt im komplexen Fall nun der

10.12 Satz von der **Partialbruchzerlegung** Seien f und $g \neq 0$ komplexe Polynome mit $\deg f < \deg g$. Seien

$$c_1, \dots, c_r \in \mathbb{C}$$

die Nullstellen von g, und

$$e_1, \dots, e_r \in \mathbb{N} \setminus \{0\}$$

ihre Vielfachheiten. Dann gibt es Konstanten $\alpha_{jk} \in \mathbb{C}$ mit:

$$\begin{aligned}\frac{f(z)}{g(z)} = {} & \frac{\alpha_{11}}{z-c_1} + \frac{\alpha_{12}}{(z-c_1)^2} + \dots + \frac{\alpha_{1,e_1}}{(z-c_1)^{e_1}} \\ & + \frac{\alpha_{21}}{z-c_2} + \dots \\ & \vdots \\ & + \frac{\alpha_{r1}}{z-c_r} + \dots \qquad\qquad \dots + \frac{\alpha_{r,e_r}}{(z-c_r)^{e_r}}\end{aligned}$$

Merkhilfe Wenn alle Nullstellen einfach sind, treten in der Partialbruchzerlegung nur die Terme der ersten Spalte auf.

Ich möchte den Satz hier zwar nicht beweisen, wohl aber eine Methode erklären, die Koeffizienten α_{jk} aus f und g zu berechnen. Dazu muß man die Nullstellen von g und ihre Vielfachheiten schon kennen; das sei jetzt vorausgesetzt.

Was passiert, wenn man die Partialbruchdarstellung von $\dfrac{f(z)}{g(z)}$ mit $(z-c_j)^{e_j}$ multipliziert und dann $z = c_j$ setzt? Nun, nach Definition der Vielfachheit e_j ist

$$g(z) = (z-c_j)^{e_j} \cdot g_j(z)$$

mit einem Polynom g_j und $g_j(c_j) \neq 0$. Auf der linken Seite der Partialbruchzerlegung kürzt sich $(z-c_j)^{e_j}$ heraus, und es bleibt die Zahl

$$\frac{f(c_j)}{g_j(c_j)} \in \mathbb{C}.$$

Rechts läßt das Auswerten bei c_j freundlicherweise alle Terme verschwinden, bis auf den letzten in der j-ten Zeile, und der ist α_{j,e_j}. Wir können diesen Koeffizienten

$$\alpha_{j,e_j} = \frac{f(c_j)}{g_j(c_j)}$$

also einfach ablesen.

10.13 Beispiel In der Zerlegung von

$$\begin{aligned}\frac{z^2+z+1}{(z-1)^2(z+2)} = {} & \frac{\alpha_{11}}{z-1} + \frac{\alpha_{12}}{(z-1)^2} \\ & + \frac{\alpha_{21}}{z+2}\end{aligned}$$

ergibt sich so

$$\frac{1^2+1+1}{1+2} = \alpha_{12}, \quad \text{d.h. } \alpha_{12} = 1$$

und

$$\frac{(-2)^2+(-2)+1}{(-2-1)^2} = \alpha_{21}, \quad \text{also } \alpha_{21} = \frac{1}{3}.$$

Die übrigen Koeffizienten bekommt man schrittweise, indem man den Teil der Partialbruchzerlegung abzieht, den man jeweils schon hat: In

$$\frac{f(z)}{g(z)} - \frac{\alpha_{1,e_1}}{(z-c_1)^{e_1}} - \frac{\alpha_{2,e_2}}{(z-c_2)^{e_2}} - \cdots - \frac{\alpha_{r,e_r}}{(z-c_r)^{e_r}}$$

muß sich jeder Linearfaktor von g mindestens einmal wegkürzen lassen. Im Beispiel:

$$\frac{z^2+z+1}{(z-1)^2(z+2)} - \frac{1}{(z-1)^2} - \frac{1/3}{z+2} = \frac{z^2+z+1-(z+2)-\frac{1}{3}(z-1)^2}{(z-1)^2(z+2)} = \frac{2}{3}\cdot\frac{z^2+z-2}{(z-1)^2(z+2)} = \frac{2}{3}\cdot\frac{1}{z-1};$$

hier ergibt sich das fehlende $\alpha_{11} = \dfrac{2}{3}$ ohne weitere Rechnung.

Anmerkung Das Verfahren beweist zugleich, daß die Koeffizienten der Partialbruchzerlegung eindeutig bestimmt sind.

Schließlich sei erwähnt, daß Gebiete $G \subset \mathbb{C}$ auch den passenden Rahmen für die Bildung von Grenzwerten

$$\lim_{z\to c} f(z)$$

abgeben, worin f eine auf $G \setminus \{c\}$ erklärte Funktion ist. Die Regeln für solche Grenzwerte sind formaler Natur und übertragen sich deshalb auf die komplexe Situation, soweit sie dort überhaupt Sinn haben; freilich schließt diese letzte Einschränkung vieles Interessante aus, wie $\lim\limits_{x\nearrow a}$, $\lim\limits_{x\searrow a}$ und $\lim\limits_{x\to\pm\infty}$.

Übungsaufgaben

10.1 Berechnen Sie $\dfrac{2+3i}{4+5i}$ sowie die Potenzen $(1+i)^n$ für jedes $n \in \mathbb{Z}$.

10.2 Seien $\alpha \in \mathbb{R}$, $c \in \mathbb{C}$. Was für Teilmengen der Zahlenebene werden durch die folgenden Gleichungen für $z \in \mathbb{C}$ jeweils beschrieben (skizzieren Sie je einen typischen Fall):

(a) $\mathrm{Re}(\bar{c}z) + \alpha = 0$;

(b) $|z|^2 + 2\mathrm{Re}(\bar{c}z) + \alpha = 0$?

10.3 Zeigen Sie, daß die durch $f(z) = \dfrac{z-i}{z+i}$ gegebene rationale Funktion ihren natürlichen Definitionsbereich $\mathbb{C} \setminus \{-i\}$ bijektiv auf $\mathbb{C} \setminus \{1\}$ abbildet, und berechnen Sie f^{-1}. Berechnen Sie außerdem die Menge $f(\mathbb{R}) \subset \mathbb{C}$.

Tip: Teilmengen einer Menge X werden häufig entweder durch Gleichungen oder durch eine Parametrisierung gegeben: Aufgabe 2.7 illustriert, was gemeint ist. Im zweiten Fall ist es leicht, für eine Abbildung $h\colon X \to Y$ die

Bildmenge $h(X) \subset Y$ zu beschreiben, im zweiten dagegen besonders einfach, für eine Abbildung $h\colon W \to X$ das Urbild $h^{-1}(X) \subset W$ darzustellen.

10.4 Berechnen und skizzieren Sie die Bilder der horizontalen und der vertikalen Geraden unter der Abbildung f aus Aufgabe 10.3.

10.5 Zeigen Sie, daß die durch $f(z) = \dfrac{1-z}{1+z}$ gegebene rationale Funktion ihren natürlichen Definitionsbereich $\mathbb{C} \setminus \{-1\}$ bijektiv in sich abbildet, und berechnen Sie $g := f^{-1}$. Berechnen Sie außerdem das Bild der rechten Halbebene

$$H := \{z \in \mathbb{C} \mid \mathrm{Re} z > 0\}$$

unter f.

10.6 Sind $X, Y \subset \mathbb{C}$ zwei Gebiete und ist $X \cap Y \neq \emptyset$, so ist auch $X \cup Y$ ein Gebiet.

10.7 Sei $E \subset \mathbb{C}$ eine endliche Menge. Beweisen Sie, daß die Menge $G := \mathbb{C} \setminus E$ ein Gebiet in $\mathbb{C}$ ist. (Sehr anschaulich ist das ja, aber es kommt darauf an, einen schlüssigen Beweis zu formulieren. Keine Ahnung, wie Sie das machen sollen? Vielleicht gibt Ihnen ein Blick auf Aufgabe 8.5 eine Idee.)

10.8 Ist $G \subset \mathbb{C}$ ein Gebiet, so ist für jedes $e \in G$ auch $G \setminus \{e\}$ ein Gebiet. (Dieses Resultat umfaßt das der Aufgabe 10.7.)

Tip: Schauen Sie sich den Beweis des Zwischenwertsatzes an.

11 Potenzreihen

So wie wir neue reelle — inzwischen auch komplexe — Zahlen als Grenzwerte von konvergenten Zahlenfolgen oder Reihen erhalten konnten, wollen wir in diesem Abschnitt neue reelle oder komplexe Funktionen als Limites von Funktionenfolgen oder -reihen konstruieren.

11.1 Sprechweise Sei $X \subset \mathbb{C}$, und sei $(f_n)_{n=0}^{\infty}$ eine Folge von Funktionen

$$f_n \colon X \longrightarrow \mathbb{C}.$$

Für jedes $x \in X$ existiere

$$f(x) := \lim_{n\to\infty} f_n(x) \in \mathbb{C}.$$

Dann sagt man, die Folge $(f_n)_{n=0}^{\infty}$ konvergiere punktweise gegen die so definierte Grenzfunktion $f \colon X \longrightarrow \mathbb{C}$. In Formeln also:

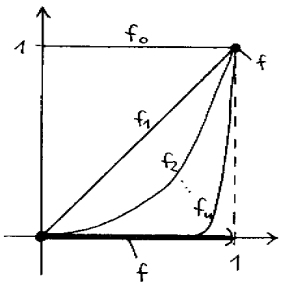
$$f = \lim f_n \text{ (punktweise)} \iff f(x) = \lim f_n(x) \text{ für alle } x \in X$$

11.2 Beispiel Sei $X = [0,1]$, $f_n(x) = x^n$. Dann ist, wie wir wissen,

$$\lim_{n\to\infty} f_n(x) = 0 \ \ (x<1) \quad \text{und} \quad \lim_{n\to\infty} f_n(1) = 1,$$

also konvergiert die Folge $(f_n)_{n=0}^{\infty}$ punktweise gegen die Grenzfunktion

$$f \colon [0,1] \longrightarrow \mathbb{R}, \quad f(x) = \begin{cases} 0 & (0 \le x < 1) \\ 1 & (x = 1). \end{cases}$$



Das Beispiel zeigt nicht nur, daß die Grenzfunktion einer konvergenten Folge stetiger Funktionen nicht stetig zu sein braucht; es läßt auch erkennen, warum: Zwar gilt $\lim f_n(x) = 0$ für jedes $x < 1$, aber diese Konvergenz wird um so langsamer, je näher x bei 1 liegt. Das legt den folgenden schärferen Konvergenzbegriff für Funktionenfolgen nahe:

11.3 Definition Sei $X \subset \mathbb{C}$, sei $(f_n)_{n=0}^{\infty}$ eine Folge von Funktionen

$$f_n \colon X \longrightarrow \mathbb{C},$$

und schließlich

$$f \colon X \longrightarrow \mathbb{C}$$

eine weitere Funktion. Man sagt, die Folge $(f_n)_{n=0}^{\infty}$ konvergiere gleichmäßig gegen f, wenn es zu jedem $\varepsilon > 0$ ein D gibt mit

$$|f_n(x) - f(x)| < \varepsilon \quad \text{für alle } x \in X \text{ und alle } n \in \mathbb{N} \text{ mit } n > D.$$

Bemerkungen (1) Der Witz der Sache ist, daß man unabhängig von x ein D finden muß, so daß die Abschätzung *gleichzeitig* für alle $x \in X$ gilt. Im Gegensatz zur "Sprechweise" 11.1 habe ich hier bewußt wieder "Definition" und nicht "Sprechweise" geschrieben. Denn gleichmäßige Konvergenz ist ein wirklich neuer Begriff, der sich nicht durch die Konvergenz der Zahlenfolgen $(f_n(x))_{n=0}^{\infty}$ für alle $x \in X$ ausdrücken läßt, während die punktweise Konvergenz gerade darin besteht.

(2) Häufig ist in derselben Situation zusätzlich eine Teilmenge $Y \subset X$ gegeben, so daß zwar nicht die Folge (f_n) selbst, wohl aber die der Einschränkungen $(f_n|Y)_{n=0}^{\infty}$ gleichmäßig gegen eine Grenzfunktion $f\colon Y \longrightarrow \mathbb{C}$ konvergiert. Man sagt dann gern, die Folge (f_n) konvergiere auf Y gleichmäßig.

Daß wir mit dem neuen Begriff den im Beispiel beobachteten Effekt überlistet haben, zeigt sich sofort:

11.4 Satz Sei $X \subset \mathbb{C}$ und sei $(f_n)_{n=0}^{\infty}$ eine gleichmäßig konvergente Folge stetiger Funktionen

$$f_n\colon X \longrightarrow \mathbb{C}.$$

Dann ist auch die Grenzfunktion

$$f := \lim_{n\to\infty} f_n\colon X \longrightarrow \mathbb{C}$$

stetig.

Beweis Ein schöner 3ε-Beweis: Sei $a \in X$ und $\varepsilon > 0$. Dann wählen wir ein $n \in \mathbb{N}$ mit

$$|f_n(x) - f(x)| < \varepsilon \quad \text{für alle } x \in X, \text{ speziell also auch für } x = a:$$
$$|f_n(a) - f(a)| < \varepsilon.$$

Weil f_n bei a stetig ist, finden wir außerdem ein $\delta > 0$ mit

$$|f_n(x) - f_n(a)| < \varepsilon \quad \text{für alle } x \in X \text{ mit } |x - a| < \delta.$$

Aufaddieren gibt

$$|f(x) - f(a)| \le |f(x) - f_n(x)| + |f_n(x) - f_n(a)| + |f_n(a) - f(a)| < 3\varepsilon \quad \text{für alle } x \in X \text{ mit } |x - a| < \delta,$$

und damit sind wir fertig.

Daß die Konvergenz der Folge in Beispiel 11.2 nicht gleichmäßig sein kann, folgt nun, ohne daß wir groß rechnen müßten, weil die Grenzfunktion offensichtlich unstetig ist.

Natürlich kann man auch von gleichmäßiger Konvergenz einer Funktionen*reihe* reden. Wichtig ist, daß das Cauchy-Kriterium auch in einer gleichmäßigen Ausgabe lieferbar ist, die — gleich für Reihen umgeschrieben — so lautet:

11.5 Cauchy-Kriterium für gleichmäßige Konvergenz Sei $X \subset \mathbb{C}$, und sei

$$\sum_{n=0}^{\infty} f_n$$

eine Reihe von Funktionen $f_n\colon X \longrightarrow \mathbb{C}$. Diese Reihe ist genau dann gleichmäßig konvergent, wenn es zu jedem $\varepsilon > 0$ ein D gibt, so daß

$$\left| \sum_{n=m+1}^{m+k} f_n(x) \right| < \varepsilon \quad \text{für alle } x \in X, \text{ alle } m > D \text{ und alle } k \in \mathbb{N}.$$

Beweis Daß gleichmäßig konvergente Reihen diese Eigenschaft haben, folgt wie früher (Lemma 4.1). Sei umgekehrt (das ist die wichtige Richtung) die gleichmäßige Cauchy-Eigenschaft vorausgesetzt. Weil diese natürlich für jedes $x \in X$ die Cauchy-Eigenschaft der Zahlenreihe $\sum_n f_n(x)$ impliziert, konvergiert die Funktionenreihe wenigstens punktweise, womit $\sum_{n=0}^{\infty} f_n$ jetzt auch als Funktion $X \longrightarrow \mathbb{C}$ einen Sinn hat. Sei nun $\varepsilon > 0$ beliebig und $D \in \mathbb{N}$ dazu passend gewählt. Wenn wir in

$$\left| \sum_{n=m+1}^{m+k} f_n(x) \right| < \varepsilon \quad \text{für alle } x \in X, \text{ alle } m > D \text{ und alle } k \in \mathbb{N}$$

zum Limes für $k \to \infty$ übergehen, ergibt sich nach der altbekannten Regel

$$\left| \sum_{n=m+1}^{\infty} f_n(x) \right| = \lim_{k\to\infty} \left| \sum_{n=m+1}^{m+k} f_n(x) \right| \leq \varepsilon < 2\varepsilon \quad \text{für alle } x \in X \text{und alle } m > D,$$

und weil links gerade

$$\left| \sum_{n=0}^{\infty} f_n(x) - \sum_{n=0}^{m} f_n(x) \right|$$

steht, beweist das die gleichmäßige Konvergenz der Partialsummenfolge und damit der Reihe.

Aus dem Cauchy-Kriterium folgt vor allem wieder, daß die gleichmäßig-absolute Konvergenz der Reihe $\sum_n f_n$ (d.h. die gleichmäßige Konvergenz von $\sum_n |f_n|$) die gleichmäßige Konvergenz von $\sum_n f_n$ impliziert. Deshalb können zur Untersuchung auf gleichmäßige Konvergenz Majoranten und Minoranten eingesetzt werden. Als solche kommen vor allem Reihen konstanter Funktionen in Betracht; für die ist punktweise und gleichmäßige Konvergenz natürlich dasselbe. Und wieder ist die praktisch wichtigste Vergleichsreihe die geometrische: Aus

$$|f_n(x)| \leq cq^n \quad \text{für alle } x \in X,\ n \in \mathbb{N}$$

mit festem $c \in [0, \infty)$ und festem $q \in [0, 1)$ folgt die gleichmäßig-absolute Konvergenz von $\sum_n f_n$.

Von größter Bedeutung ist die Realisierung dieser neuen Möglichkeiten in Gestalt der sogenannten Potenzreihen; sie spielen in der Analysis eine Rolle, die der der Polynome in der Algebra ähnlich ist.

11.6 Definition Sei $a \in \mathbb{C}$. Eine Funktionenreihe der Form

$$\sum_{n=0}^{\infty} a_n(z-a)^n$$

nennt man eine Potenzreihe um a. Genauer sind die Glieder der Reihe die Funktionen

$$\mathbb{C} \ni z \mapsto a_n(z-a)^n \in \mathbb{C};$$

und letzten Endes handelt es sich bei einer Potenzreihe (um den gegebenen Punkt a) schlicht um die Folge ihrer Koeffizienten $(a_n)_{n=0}^{\infty}$ in $\mathbb{C}$, die bloß auf eine besondere Art interpretiert wird.

Wie schon die geometrische Reihe

$$\sum_{n=0}^{\infty} z^n$$

(Potenzreihe um 0, alle Koeffizienten sind 1) zeigt, braucht eine Potenzreihe nicht für alle $z \in \mathbb{C}$ zu konvergieren. Man kann aber allgemein recht genaue Aussagen über die Menge der z machen, für die das der Fall ist.

11.7 Satz über die **Konvergenz** von **Potenzreihen** Sei $a \in \mathbb{C}$, und sei

$$\sum_{n=0}^{\infty} a_n(z-a)^n$$

eine Potenzreihe um a.

$$K := \left\{ z \in \mathbb{C} \;\middle|\; \sum a_n (z-a)^n \text{ konvergiert} \right\}$$

sei die Menge ihrer Konvergenzpunkte. Dann gilt:

(a) Es gibt ein $r \in [0, \infty]$ mit

$$U_r(a) \subset K \subset D_r(a)$$

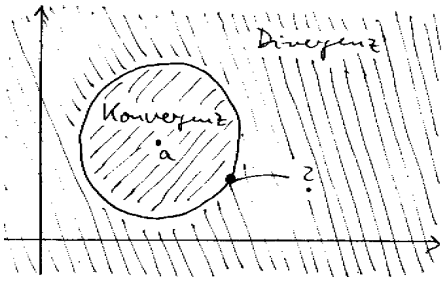
(was für $r = 0$ als $K = \{a\}$ und für $r = \infty$ als $K = \mathbb{C}$ interpretiert werden soll). Dieses natürlich eindeutig bestimmte r heißt der Konvergenzradius der Potenzreihe, die offene Kreisscheibe $U_r(a)$ (bzw. $\emptyset$ im Fall $r=0$, ganz $\mathbb{C}$ im Fall $r=\infty$) ihr Konvergenzkreis.

(b) Für jedes $\rho \in [0, r)$ konvergiert die Reihe $\sum a_n (z-a)^n$, genauer gesagt, die Funktionenreihe

$$\sum_n \left(z \mapsto a_n (z-a)^n \right)$$

auf der abgeschlossenen Kreisscheibe $D_\rho(a)$ gleichmäßig-absolut.

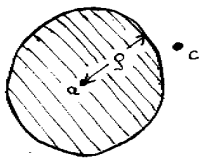
Erläuterung Die Namen in (a) erklären sich von selbst; der Satz sagt ja im Groben, daß K eine Kreisscheibe um a vom Radius r ist, wobei nur offen bleibt, in welchen Punkten der Kreislinie die Reihe noch konvergiert.



Durch (b) wird die Aussage verschärft: Zwar braucht die Konvergenz nicht auf ganz K und auch nicht auf $U_r(a)$ gleichmäßig zu sein, wohl aber auf jeder in $U_r(a)$ ganz enthaltenen abgeschlossenen Kreisscheibe. Wenn Ihnen das paradox vorkommt, erinnern Sie sich bitte daran, daß die Gleichmäßigkeit der Konvergenz sich auf eine Reihe von *Funktionen* auf einem gegebenen Definitionsbereich bezieht und nicht auf lauter Zahlenreihen an verschiedenen Stellen.

Beweis des Satzes Als Kernpunkt zeigen wir folgende etwas technischere Aussage:

> Sei $c \in K$ ein Konvergenzpunkt, und sei $0 \le \rho < |c-a|$. Dann konvergiert die Reihe
>
> $$\sum_n |a_n| \, |z-a|^n$$
>
> auf $D_\rho(a)$ gleichmäßig.



Das ist ganz einfach: Wir vergleichen mit einer geometrischen Reihe, nämlich:

$$|a_n| \, |z-a|^n \le |a_n| \, \rho^n = |a_n| \, |c-a|^n \left(\frac{\rho}{|c-a|} \right)^n$$

für $|z-a| \le \rho$. Weil $\sum a_n (c-a)^n$ konvergiert, gilt (Lemma 5.4)

$$\lim_{n \to \infty} a_n (c-a)^n = 0,$$

abgesehen von endlich vielen Anfangsgliedern also $|a_n|\,|c-a|^n \le 1$ und damit

$$|a_n|\,|z-a|^n \le \left(\frac{\rho}{|c-a|}\right)^n.$$

Da $\rho < |c-a|$ war, folgt die gleichmäßige Konvergenz der Reihe $\sum |a_n|\,|z-a|^n$ wie behauptet.

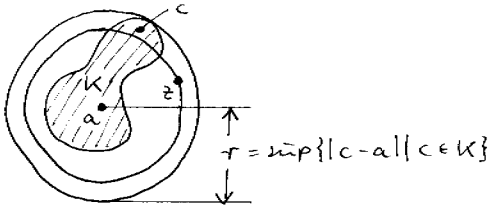
Jetzt folgt der Satz schnell. Entweder ist K, die Menge der Konvergenzpunkte, unbeschränkt: dann muß $K = \mathbb{C}$ sein, und wir haben $r = \infty$. Oder K ist beschränkt (wegen $a \in K$ auch nicht-leer), und wir setzen

$$r := \sup\left\{|c-a| \,\middle|\, c \in K\right\}.$$

Von der Behauptung

$$U_r(a) \subset K \subset D_r(a)$$

ist die zweite Inklusion dann klar (weil das Supremum eine obere Schranke ist), und die erste folgt aus unserer Vorüberlegung: Ist $z \in U_r(a)$, so setzen wir $\rho := |z-a| < r$ und finden ein $c \in K$ mit $\rho < |c-a|$, damit ist $D_\rho(a) \subset K$ und insbesondere $z \in K$.



(hypothetisches K; wir beweisen ja gerade, daß K in Wirklichkeit *nicht* so aussieht)

Schließlich ergibt sich auch die in (b) behauptete Gleichmäßigkeit der Konvergenz unmittelbar aus der Vorüberlegung.

11.8 Beispiele (1) Wir wissen schon, daß die geometrische Reihe $\sum z^n$ den Konvergenzradius 1 hat. Auf dem Rand des Konvergenzkreises konvergiert sie nirgends.

(2) Die Reihe $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n} z^n$ ist für $z = \pm 1$ divergent bzw. konvergent (harmonische Reihen). Allein daraus folgt nach Satz 11.7 schon, daß der Konvergenzradius 1 ist. Diese Potenzreihe konvergiert also in manchen, aber nicht allen Randpunkten des Konvergenzkreises (wie eine genauere Untersuchung zeigt, in allen außer $+1$). Natürlich ist die Konvergenz in diesen Randpunkten nicht absolut.

(3) Die Reihe $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^2} z^n$ hat ebenfalls den Konvergenzradius 1, wie man mittels der Methode von Aufgabe 11.5 mühelos findet. Für $z = 1$ erhält man die nach Aufgabe 5.2 konvergente Reihe $\sum \frac{1}{n^2}$, daher konvergiert die Potenzreihe auch überall auf dem Rand des Konvergenzkreises (absolut), dort gilt ja $\sum \left|\frac{1}{n^2} z^n\right| = \sum \frac{1}{n^2}$.

(4) Sei allgemeiner $h \neq 0$ eine ansonsten beliebige komplexe rationale Funktion. Die Reihe

$$\sum_n h(n) z^n$$

hat dann den Konvergenzradius 1, so lautet in neuer Sprechweise das Ergebnis der Aufgabe 5.8. Diese bemerkenswerte Tatsache lohnt es sich auch für den praktischen Gebrauch zu merken: Solange die Koeffizienten einer Potenzreihe oder ihre Kehrwerte nicht schneller als polynomial wachsen, bleibt der Konvergenzradius der Reihe der der geometrischen.

(5) Die Exponential-, Cosinus- und Sinusreihen

$$\exp z = \sum_{n=0}^{\infty} \frac{z^n}{n!}$$
$$\cos z = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!}$$
$$\sin z = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!}$$

haben den Konvergenzradius ∞; das haben wir uns unter 5.15 schon überlegt.

(6) Schreibt man z.B. die Fakultät in den Zähler, so erhält man mit

$$\sum_n n! z^n$$

eine Potenzreihe vom Konvergenzradius 0, denn für jedes $z \neq 0$ ist $\lim\limits_{n\to\infty} n!|z|^n = \infty$.

Zurück zum Allgemeinen: Mit den Summenfunktionen konvergenter Potenzreihen wollten wir uns eine neue Quelle stetiger Funktionen erschließen. Tatsächlich erlaubt Satz 11.7 die

11.9 Folgerung Die durch die Potenzreihe $\sum\limits_{n=0}^{\infty} a_n(z-a)^n$ in ihrem Konvergenzkreis (also der *offenen* Kreisscheibe) dargestellte Funktion

$$f(z) = \sum_{n=0}^{\infty} a_n (z-a)^n$$

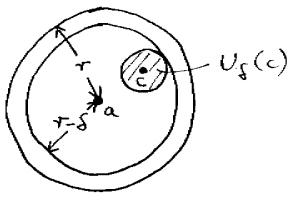
ist stetig.

Beweis Es ist die kleine Schwierigkeit zu umschiffen, daß die Konvergenz der Reihe im allgemeinen nicht auf dem ganzen Konvergenzkreis gleichmäßig ist. Sei r der Konvergenzradius; wir prüfen die Stetigkeit an einer Stelle $c \in U_r(a)$. Dazu setzen wir

$$\delta := \frac{1}{2}(r - |c-a|) > 0$$

und bemerken

$$U_\delta(c) \subset D_{r-\delta}(a) \subset U_r(a)$$



(denn aus $|z-c| < \delta$ folgt $|z-a| \leq |z-c| + |c-a| < \delta + |c-a| = 2\delta + |c-a| - \delta = r - \delta$). Nach Satz 11.7 konvergiert die Potenzreihe gleichmäßig auf $D_{r-\delta}(a) \subset U_r(a)$, also erst recht auf der kleineren Menge $U_\delta(c)$, deshalb ist die Einschränkung

$$f|U_\delta(a) = f\big|\,\{z \in \mathbb{C} \,\big|\, |z-c| < \delta\}$$

stetig (Satz 11.4). Nach dem früheren Lemma 7.3 ist daher auch f bei c stetig.

Bemerkung Delikater ist die Frage, ob f noch stetig bleibt, wenn man eventuelle Konvergenzpunkte auf dem Rand des Konvergenzkreises mit in den Definitionsbereich aufnimmt. Sehen Sie dazu bei Bedarf in der Literatur unter dem Stichwort "Abelscher Grenzwertsatz" nach.

Es liegt auf der Hand, daß die (gliedweise) Summe zweier Potenzreihen um denselben Punkt $a \in \mathbb{C}$ wieder eine Potenzreihe um a ist; nach den vertrauten Regeln konvergiert die Summe mindestens dort, wo die Ausgangsreihen konvergieren, und die durch die Reihen dargestellten stetigen Funktionen addieren sich. Entsprechendes gilt auch für die Multiplikation:

11.10 Satz $\sum_{j=0}^{\infty} a_j(z-a)^j$ und $\sum_{k=0}^{\infty} b_k(z-a)^k$ seien zwei Potenzreihen um a, deren Konvergenzradius jeweils mindestens $r \in [0,\infty]$ ist. Dann hat das Produkt dieser beiden Potenzreihen, die Potenzreihe

$$\sum_{n=0}^{\infty}\left(\sum_{j+k=n} a_j b_k\right)(z-a)^n$$

ebenfalls mindestens den Konvergenzradius r, und die durch die Reihen auf $U_r(a)$ dargestellten Funktionen multiplizieren sich:

$$\sum_{n=0}^{\infty}\left(\sum_{j+k=n} a_j b_k\right)(z-a)^n = \left(\sum_{j=0}^{\infty} a_j(z-a)^j\right)\cdot\left(\sum_{k=0}^{\infty} b_k(z-a)^k\right)$$

Beweis Das im Satz erklärte Produkt ist wirklich das Produkt $\sum_{j,k=0}^{\infty} a_j b_k (z-a)^{j+k}$ der beiden Reihen, nach Diagonalen abgezählt. Wir haben damit bloß einen Spezialfall von Satz 6.6 vor uns.

Übungsaufgaben

11.1 Die Konvergenz der Funktionenfolge $(f_n)_{n=0}^{\infty}$ mit $f_n\colon [0,1] \longrightarrow \mathbb{R};\ f_n(x) = x^n$ ist, wie wir im Beispiel 11.2 gesehen haben, nicht gleichmäßig. Zeigen Sie, daß die Konvergenz auch auf dem Intervall $[0,1)$ nicht gleichmäßig ist, wohl aber auf jedem Intervall $[0,b]$ mit $b < 1$.

11.2 Die Exponentialreihe $\sum_{n=0}^{\infty} \frac{z^n}{n!}$ ist ein Beispiel einer konvergenten Reihe von Funktionen $\mathbb{C} \longrightarrow \mathbb{C}$. Ist die Konvergenz gleichmäßig?

11.3 Aus einer Potenzreihe $\sum_{j=n}^{\infty} a_j(z-a)^j$ mit $n \in \mathbb{N}$ und bekanntem Konvergenzradius $r \in [0,\infty]$ wird man oft den Faktor $(z-a)^n$ herausziehen und

$$\sum_{j=n}^{\infty} a_j(z-a)^j = (z-a)^n \sum_{k=0}^{\infty} a_{n+k}(z-a)^k$$

schreiben wollen. Ist das in jedem Fall zulässig?

11.4 In der Potenzreihe $\sum_{n=0}^{\infty} a_n(z-a)^n$ seien a sowie alle Koeffizienten a_n ($n \in \mathbb{N}$) reell. Beweisen Sie, daß für die durch die Reihe im Konvergenzkreis dargestellte Funktion f dann

$$f(\overline{z}) = \overline{f(z)} \text{ für alle } z$$

gilt.

11.5 In der Potenzreihe $\sum_{j=0}^{\infty} a_j(z-a)^j$ seien alle Koeffizienten (vielleicht mit endlich vielen Ausnahmen) von null verschieden. Zeigen Sie: Wenn

$$r := \lim_{j\to\infty} \frac{|a_j|}{|a_{j+1}|} \in [0,\infty]$$

existiert, dann ist r der Konvergenzradius dieser Potenzreihe. Überzeugen Sie sich davon, daß man mit dieser Formel den Konvergenzradius der Potenzreihen aus den Beispielen 11.8 tatsächlich ganz bequem bestimmen kann.

11.6 Wenden Sie Satz 11.10 an, um die Funktion $z \mapsto \dfrac{1}{(1-z)^2}$ in der Kreisscheibe $U_1(0) \subset \mathbb{C}$ durch eine konvergente Potenzreihe darzustellen.

11.7 Sei $g(z) = \sum_{k=0}^{\infty} b_k z^k$ eine Potenzreihe um 0. Welche der Koeffizienten b_k muß man kennen, um den Koeffizienten von z^{12} in der Produktreihe

$$(z - \sin z)^3 \cdot g(z)$$

zu bestimmen (worin für $\sin z$ natürlich die Sinusreihe einzusetzen ist)?

12 Die Exponentialfunktion

Wir wollen uns jetzt endlich die schon mehrfach erwähnte Exponentialfunktion systematisch vornehmen.

12.1 Definition Die durch die überall konvergente Potenzreihe (um 0)

$$\sum_{n=0}^{\infty} \frac{1}{n!} z^n = 1 + z + \frac{1}{2} z^2 + \frac{1}{6} z^3 + \cdots$$

dargestellte stetige Funktion

$$\exp\colon \mathbb{C} \longrightarrow \mathbb{C}; \quad \exp z := \sum_{n=0}^{\infty} \frac{1}{n!} z^n$$

heißt, ebenso wie ihre reelle Einschränkung $\exp\colon \mathbb{R} \longrightarrow \mathbb{R}$, die Exponentialfunktion.

Ohne weiteres aus dem Beispiel 6.7 ins Komplexe übertragen kann man den Beweis der fundamentalen

12.2 Formel $\exp(w+z) = (\exp w)(\exp z)$ für alle $w, z \in \mathbb{C}$.

Aus dieser Formel und $\exp 0 = \sum_{n=0}^{\infty} \frac{1}{n!} 0^n = 1$ ergibt sich insbesondere $\big(\exp z\big)\big(\exp(-z)\big) = 1$, also

$$\exp z \neq 0 \quad \text{und} \quad \frac{1}{\exp z} = \exp(-z) \quad \text{für jedes } z \in \mathbb{C}.$$

Es hat sich eingebürgert, die sogenannte *eulersche Zahl* $\exp 1$ mit e zu bezeichnen und statt $\exp z$ auch e^z zu schreiben. Offenbar ist das mit der schon vorhandenen Bedeutung von e^k für $k \in \mathbb{Z}$ verträglich. Übrigens ist es nicht schwer zu sehen, daß $e \approx 2.72$ eine irrationale Zahl ist.

Besonders aufschlußreich ist das Studium von e^z einerseits für reelle, andererseits für rein imaginäre z. Zu ersteren:

12.3 Satz und **Definition** Es gilt $e^x > 0$ für jedes $x \in \mathbb{R}$, und

$$\mathbb{R} \xrightarrow{\exp} (0, \infty)$$

ist streng monoton wachsend und bijektiv. Die nach unserer Kenntnis über stetige Funktionen (Satz 8.5) ihrerseits stetige Umkehrfunktion

$$(0, \infty) \xrightarrow{\log} \mathbb{R}$$

heißt die Logarithmusfunktion.

Beweis $e^x = \sum_{n=0}^{\infty} \frac{1}{n!} x^n \geq 1 > 0$ für $x \geq 0$ ist klar. Dann folgt aber auch

$$e^{-x} = \frac{1}{\exp x} > 0 \quad \text{für } x \geq 0,$$

also $e^x > 0$ für alle $x \in \mathbb{R}$. Die strenge Monotonie ergibt sich so: Sei $x < y$; dann ist

$$e^{y-x} e^x = e^y$$

und wegen $y-x > 0$

$$e^{y-x} = \sum_{n=0}^{\infty} \frac{1}{n!}(y-x)^n \geq 1 + (y-x) > 1,$$

also $e^y > e^x$. Jetzt bleibt nur noch das Bildintervall von exp zu bestimmen. Aber aus

$$\exp x \geq 1 + x \quad \text{für } x \geq 0$$

folgt sofort

$$\lim_{x\to\infty} e^x = \infty$$

und damit

$$\lim_{x\to-\infty} e^x = \lim_{y\to\infty} \frac{1}{e^y} = 0,$$

folglich ist $\exp(\mathbb{R}) = (0, \infty)$.

Bemerkung Vor allem in der technischen Literatur wird log oft der "natürliche" Logarithmus genannt und mit ln bezeichnet, wie es auch irgendwelche Normen von uns wollen.

Die Funktionen exp und log erlauben es, auch Potenzen mit nicht-ganzen Exponenten zu erklären:

12.4 Definition Für $a \in (0, \infty)$ und $z \in \mathbb{C}$ definiert man die Potenz $a^z \in \mathbb{C}$ durch

$$a^z = e^{z \cdot \log a}$$

(für $x \in \mathbb{R}$ ist natürlich auch $a^x \in \mathbb{R}$).

Beachten Sie, daß diese Definition wegen der Einschränkung $a \in (0, \infty)$ keineswegs alle bisherigen Fälle umfaßt. Dort, wo sie das aber tut, stimmt sie mit der alten Definition überein. Das rechnet man mit einiger Geduld nach, ebenso die zahlreichen Eigenschaften, die man von einer Potenz erwartet. Hier nur eine

12.5 Auswahl von **Regeln**

(a) $a^{w+z} = a^w a^z \quad (a \in (0, \infty);\ w, z \in \mathbb{C})$

(b) $a^{1/n} = \sqrt[n]{a} \quad (a \in (0, \infty);\ 0 < n \in \mathbb{N})$

(c) Die Funktion $\mathbb{R} \longrightarrow (0, \infty),\ x \mapsto a^x$ ist für $a < 1$ und für $a > 1$ streng monoton fallend bzw. wachsend und surjektiv; ihre Umkehrung ist $y \mapsto \dfrac{1}{\log a} \log y$ $(:= \log_a y$, Logarithmus zur Basis $a)$.

Häufig gebraucht werden die folgenden Regeln über das Verhalten von exp und log bei Annäherung an die "Enden" ihres Definitionsintervalls.

12.6 Lemma (a) Für jedes $b \in \mathbb{R}$ gilt:

$$\lim_{x\to\infty} \frac{e^x}{x^b} = \infty$$

Wenn x^b auch für negative x Sinn gibt (zum Beispiel für $b \in \mathbb{N}$), kann man dual dazu auch

$$\lim_{x\to-\infty} x^b e^x = 0$$

notieren.

(b) Für jedes $b > 0$ gilt

$$\lim_{x\to\infty} \frac{\log x}{x^b} = 0$$

und

$$\lim_{x\to 0} x^b \log x = 0.$$

Man kann zum Beispiel (b) auch als

$$\log x = o(x^b) \quad (x \to \infty) \qquad \text{bzw.} \qquad \log x = o(x^{-b}) \quad (x \to 0)$$

schreiben, und sich überhaupt merken: Für $x \to \infty$ wächst e^x schneller als jede (noch so große) Potenz von x, dagegen $\log x$ langsamer als jede (noch so kleine positive) Potenz etc.

Beweis (a) Für $x \geq 1$ und $n := \lceil b+1 \rceil$ gilt

$$\frac{1}{n!}\, x^b \cdot x \leq \frac{1}{n!} x^n \leq e^x$$

und damit

$$\frac{1}{n!} x \leq \frac{e^x}{x^b},$$

woraus die erste Behauptung schon folgt. Unter der für die zweite genannten Voraussetzung ist dann auch

$$\lim_{x \to -\infty} \left| \frac{1}{x^b e^x} \right| = \lim_{y \to \infty} \left| \frac{1}{y^b \cdot e^{-y}} \right| = \lim_{y \to \infty} \left| \frac{e^y}{y^b} \right| = \infty$$

und folglich

$$\lim_{x \to -\infty} x^b e^x = 0.$$

(b) Nach (a) gibt es ein $E > 0$ mit

$$y^{2/b} < e^y \quad \text{für alle } y > E,$$

folglich ($y := x^{b/2}$)

$$x < e^{\left(x^{b/2}\right)} \quad \text{für alle } x > E^{2/b} =: D,$$

also (weil log monoton wächst) $\log x < x^{b/2}$ oder

$$\frac{\log x}{x^b} < x^{-b/2} \quad \text{für alle } x > D.$$

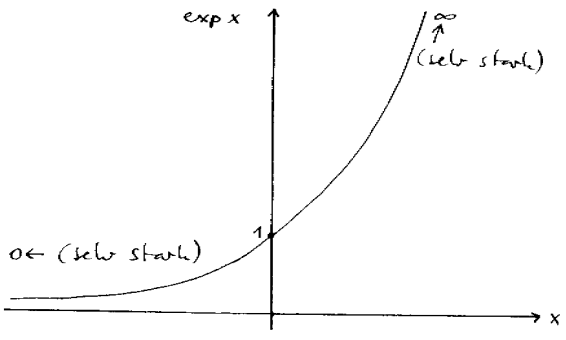
Wegen $\log x \geq 0$ für $x \geq 1$ und $\lim\limits_{x \to \infty} x^{-b/2} = 0$ folgt die Behauptung: $\lim\limits_{x \to \infty} \dfrac{\log x}{x^b} = 0$.

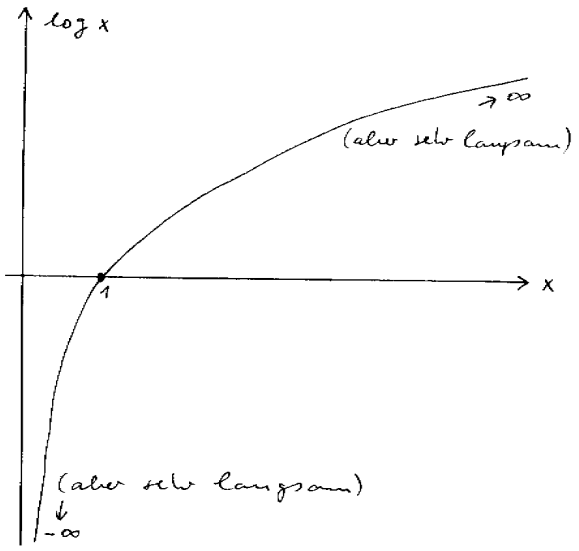
Mit $x := 1/y$ wird daraus schließlich

$$0 = \lim_{y \searrow 0} \frac{\log(1/y)}{(1/y)^b} = \lim_{y \searrow 0} \left(-y^b \log y\right),$$

was zu $\lim\limits_{x \to 0} x^b \log x = 0$ äquivalent ist.

Insgesamt haben wir damit eine recht gute Vorstellung von der reellen Exponential- und der Logarithmusfunktion gewonnen.





Ungleich witziger ist e^z für rein imaginäre $z \in \mathbb{C}$. Für zunächst noch beliebige $z \in \mathbb{C}$ sehen wir uns erst mal die Reihe

$$e^{iz} = \sum_{n=0}^{\infty} \frac{1}{n!}(iz)^n$$

an: Weil i^n nur die vier Werte $1, i, -1, -i$ annimmt — je nach dem Rest, den n beim Teilen durch 4 läßt — können wir e^{iz} in

$$\begin{aligned} & \sum_{n \text{ gerade}} \frac{i^n}{n!} z^n \quad + \quad i \sum_{n \text{ gerade}} \frac{i^n}{(n+1)!} z^{n+1} \\ = \quad & \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} z^{2k} \quad + \quad i \quad \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} z^{2k+1} \end{aligned}$$

aufspalten. Die beiden — natürlich ebenfalls überall konvergenten — Teil(potenz)reihen

$$\cos z = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} z^{2k} = 1 - \frac{1}{2}z^2 + \frac{1}{24}z^4 - \cdots$$

und

$$\sin z = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} z^{2k+1} = z - \frac{1}{6}z^3 + \frac{1}{120}z^5 - \cdots$$

gehören wie die Exponentialreihe zu den wenigen, die man auswendig wissen soll; sie definieren die Cosinus- bzw. Sinusfunktion

$$\cos\colon \mathbb{C} \longrightarrow \mathbb{C}, \qquad \sin\colon \mathbb{C} \longrightarrow \mathbb{C}.$$

Mit der Exponentialfunktion sind die beiden definitionsgemäß durch die

12.7 Formel $e^{iz} = \cos z + i \sin z$ für alle $z \in \mathbb{C}$

verbunden. Nun ist cos offensichtlich eine gerade, und sin eine ungerade Funktion:

$$\cos(-z) = \cos z \quad \text{und} \quad \sin(-z) = -\sin z$$

Deshalb ist $e^{-iz} = \cos z - i \sin z$, und diese beiden Gleichungen zusammen erlauben es, Cosinus und Sinus "direkt" durch die Exponentialfunktion auszudrücken, nämlich:

$$\begin{aligned} \cos z &= \frac{1}{2}(e^{iz} + e^{-iz}) \\ \sin z &= \frac{1}{2i}(e^{iz} - e^{-iz}) \end{aligned} \qquad (z \in \mathbb{C})$$

Wo wir uns aber speziell für rein imaginäre iz, d.h. für reelle z interessieren wollten, wird die Sache noch einfacher: Ersichtlich nehmen cos und sin für reelle Argumente auch reelle Werte an, und deshalb stellt die Formel

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

gerade die Zerlegung von $e^{iy} \in \mathbb{C}$ in Realteil — nämlich $\cos y$ — und Imaginärteil — $\sin y$ — dar. Diese Darstellung ist für das Rechnen mit Cosinus und Sinus außerordentlich nützlich. Das wird überzeugend belegt durch die großspurig "Additionstheoreme" genannten

12.8 Formeln Für beliebige $x, y \in \mathbb{R}$ (tatsächlich sogar aus $\mathbb{C}$) gilt:

$$\cos(x+y) = \cos x \, \cos y - \sin x \, \sin y$$
$$\sin(x+y) = \sin x \, \cos y + \cos x \, \sin y$$

Beweis Man zerlegt einfach beide Seiten von $e^{i(x+y)} = e^{ix}e^{iy}$ in Real- und Imaginärteil.

Die eigentliche Empfehlung für das Rechnen mit Cosinus und Sinus besteht aber nicht darin, auf diese Art jede benötigte Formel einzeln aus der Formel 12.2. herauszuziehen, sondern von vornherein statt mit $\cos y$ und/oder $\sin y$ möglichst mit e^{iy} zu arbeiten. Physiker und Elektrotechniker tun das dann auch gerne mit der entschuldigenden Bemerkung "wir rechnen mit komplexen Zahlen, aber nur der Realteil hat physikalische Bedeutung". Warum nicht, wenn man dafür mit viel durchsichtigeren Formeln belohnt wird?

Als Analogon zu Satz 12.3 wollen wir uns überlegen:

12.9 Satz Für jedes $y \in \mathbb{R}$ ist $|e^{iy}| = 1$. Die Abbildung

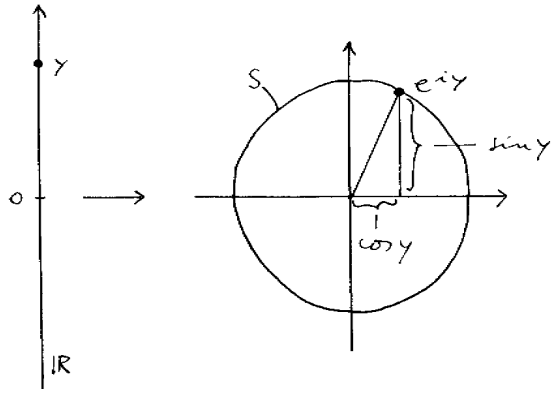
$$\mathbb{R} \longrightarrow S := \{w \in \mathbb{C} \mid |w| = 1\}, \quad y \mapsto e^{iy}$$

ist surjektiv, und für $x, y \in \mathbb{R}$ gilt:

$$e^{ix} = e^{iy} \iff x - y = k \cdot 2\pi \text{ für ein } k \in \mathbb{Z}$$

Dabei ist $\pi \in (0, 4)$ die berühmte Zahl, die im folgenden noch genau erklärt wird.

Geometrische Deutung Der Satz (zusammen mit einigen Details aus dem Beweis) besagt im wesentlichen, daß die Abbildung $y \mapsto e^{iy}$ die reelle Gerade auf die Kreislinie S wickelt, von $e^{i0} = 1$ aus entgegen dem Uhrzeigersinn fortschreitend, oder, wie man sagt, im *mathematisch positiven Sinn*. Es liegt nahe, die reelle Zahl y wie in der Skizze als den *orientierten* Winkel von 1 nach e^{iy} zu interpretieren; beachten Sie, daß dieser nur bis auf die Addition von (positiven oder negativen) Vielfachen von $2\pi = 360°$ definiert ist, wie es ja auch im Wesen eines Winkels liegt. Damit ist auch der Anschluß an die Ihnen aus der Schule vertraute Rolle von Cosinus und Sinus in der Geometrie rechtwinkliger Dreiecke hergestellt.



Zum Beweis brauchen wir drei kleine Hilfssätze:

Erster Hilfssatz $\cos 2 < 0$

Beweis Wir schreiben

$$\cos 2 = 1 - \frac{2^2}{2!} + \frac{2^4}{4!} \underbrace{- \frac{2^6}{6!} + \frac{2^8}{8!}}_{<0} \underbrace{- \frac{2^{10}}{10!} + \frac{2^{12}}{12!}}_{<0} \underbrace{- \quad +}_{<0} \cdots$$

und lesen sogar $\cos 2 < 1 - 2 + \frac{16}{24} = -\frac{1}{3}$ ab.

Zweiter Hilfssatz $\sin x > 0$ für alle $x \in (0, 2]$

Beweis Schreibe

$$\begin{aligned} \sin x &= x - \frac{x^3}{3!} \underbrace{+ \frac{x^5}{5!} - \frac{x^7}{7!}}_{>0} \underbrace{+ \frac{x^9}{9!} - \frac{x^{11}}{11!}}_{>0} \underbrace{+ \quad -}_{>0} \cdots \\ &> x - \frac{x^3}{3!} = x \cdot \left(1 - \frac{x^2}{6}\right) > 0. \end{aligned}$$

Dritter Hilfssatz Die eingeschränkte Funktion $\cos|[0, 2]$ fällt streng monoton.

Beweis Sei $0 \le x < y \le 2$. Wir setzen

$$u = \frac{y + x}{2} > 0 \quad \text{und} \quad v = \frac{y - x}{2} > 0.$$

Nach den Formeln 12.8 ist

$$\begin{array}{rr} \cos x = \cos u \, \cos v + & \sin u \, \sin v \\ \cos y = \cos u \, \cos v - & \sin u \, \sin v \\ \hline \cos x - \cos y = & 2 \sin u \, \sin v, \end{array}$$

und nach dem zweiten Hilfssatz ist die Zahl $2 \sin u \, \sin v$ positiv. Fertig.

Jetzt können wir auf $\cos|[0, 2]$ die vertraute Theorie anwenden: Diese Funktion ist stetig, streng monoton fallend, mit $\cos 0 = 1$ und $\cos 2 < 0$; nach dem Zwischenwertsatz existiert also genau eine Nullstelle.

12.10 Definition $\pi/2$ wird als die einzige Nullstelle von cos im Intervall $(0, 2)$ definiert.

Damit zum eigentlichen

Beweis von 12.9 Für jedes $y \in \mathbb{R}$ ist

$$|e^{iy}|^2 = e^{iy} \cdot \overline{e^{iy}} = e^{iy} \cdot e^{\overline{iy}} = e^{iy} \cdot e^{-iy} = e^0 = 1,$$

damit $y \mapsto e^{iy}$ tatsächlich eine Abbildung mit Werten in der Kreislinie S. In Komponenten zerlegt ist das übrigens die bekannte Identität $(\cos y)^2 + (\sin y)^2 = 1$, und ganz nebenbei folgt daraus

$$|\cos y| \le 1 \text{ und } |\sin y| \le 1 \text{ für alle } y \in \mathbb{R}.$$

Nach Definition von π bildet cos das Intervall $[0, \frac{\pi}{2}]$ bijektiv (und monoton fallend) auf $[0, 1]$ ab. Weil der Sinus nach dem zweiten Hilfssatz auf $[0, \frac{\pi}{2}]$ nicht-negativ ist, haben wir damit eine zumindest injektive Abbildung

$$\left[0, \frac{\pi}{2}\right] \longrightarrow \{w \in S \mid \mathrm{Re}\, w \ge 0, \mathrm{Im}\, w \ge 0\}; \quad y \mapsto e^{iy}$$

in den Viertelkreis. Die ist aber auch surjektiv: Zu gegebenem $w = u + iv$ rechts gibt es genau ein $y \in [0, \frac{\pi}{2}]$ mit $\cos y = u$, und es ist zwangsläufig

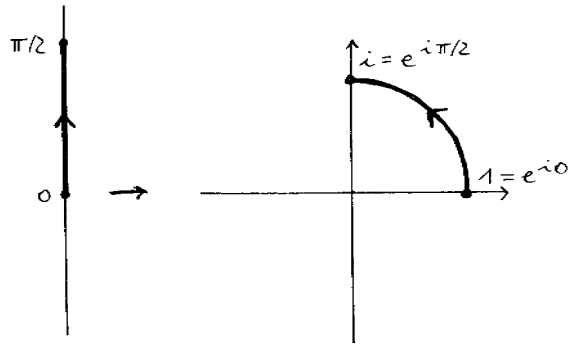
$$\sin y = \sqrt{1 - (\cos y)^2} \quad \text{ebenso wie} \quad v = \sqrt{1 - u^2},$$

also auch $\sin y = v$.

Speziell ergibt sich $\sin \frac{\pi}{2} = 1$ und damit die wichtige Tatsache:

$$e^{i\frac{\pi}{2}} = i$$

Geometrisch gesehen scheinen wir den Satz nun zu einem Viertel bewiesen zu haben:



Tatsächlich ergibt sich der Rest aber jetzt ganz schnell: Aus

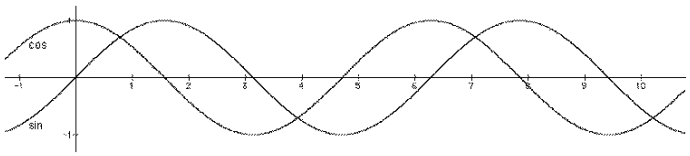
$$e^{i(y+\frac{\pi}{2})} = e^{iy}e^{i\frac{\pi}{2}} = i \cdot e^{iy} \quad \text{und der allgemeinen Formel} \quad i \cdot (u + iv) = -v + iu$$

liest man ab, daß e^{iy} für y in den weiteren Intervallen $[\frac{\pi}{2}, \pi]$, $[\pi, \frac{3\pi}{2}]$, $[\frac{3\pi}{2}, 2\pi]$ die anderen drei Viertelkreisbögen jeweils bijektiv durchläuft, und daß

$$e^{i(y+2\pi)} = e^{iy} \quad \text{für alle } y \in \mathbb{R}$$

(sogar alle $y \in \mathbb{C}$) ist. Damit folgt der Satz unmittelbar.

Im Satz 12.9 und den im Beweis abgeleiteten Formeln sind all die bekannten Eigenschaften der Cosinus- und der Sinusfunktion enthalten, wie sie ja in



andeutungsweise zum Ausdruck kommen. Das gleiche gilt für die beiden weiteren trigonometrischen Funktionen

$$\tan = \frac{\sin}{\cos} \quad \text{und} \quad \cot = \frac{\cos}{\sin},$$

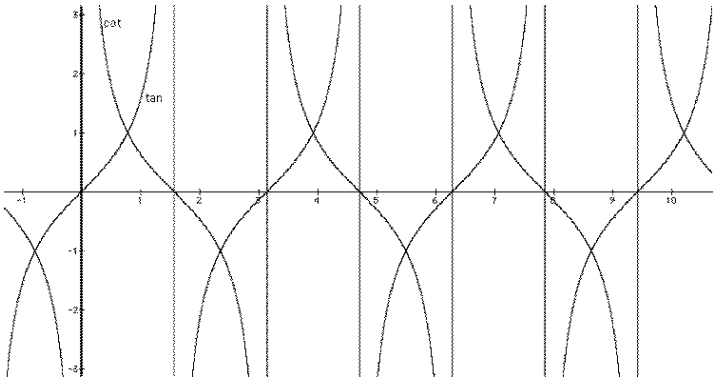
die natürlich nur dort definiert sind, wo ihr Nenner nicht null wird:

$$\mathbb{C} \setminus \left\{ \left(k+\tfrac{1}{2}\right)\pi \,\middle|\, k \in \mathbb{Z} \right\} \xrightarrow{\tan} \mathbb{C}$$
$$\mathbb{C} \setminus \{k\pi \mid k \in \mathbb{Z}\} \xrightarrow{\cot} \mathbb{C}$$

(die bekannten reellen Nullstellen von cos: $z \mapsto (e^{iz} + e^{-iz})/2$ und sin: $z \mapsto (e^{iz} - e^{-iz})/2i$ sind auch in $\mathbb{C}$ die einzigen). An reellen Stellen haben Tangens und Cotangens auch reelle Werte, und am häufigsten hat man es mit den reellen Versionen

$$\mathbb{R} \setminus \left\{ \left(k+\tfrac{1}{2}\right)\pi \,\middle|\, k \in \mathbb{Z} \right\} \xrightarrow{\tan} \mathbb{R}$$
$$\mathbb{R} \setminus \{k\pi \mid k \in \mathbb{Z}\} \xrightarrow{\cot} \mathbb{R}$$

zu tun.



Als stetige reelle Funktionen besitzen alle trigonometrischen Funktionen stetige Umkehrungen, wenn man sie auf Intervalle einschränkt, auf denen sie streng monoton sind. In der Wahl solcher Intervalle liegt eine gewisse Willkür; geeinigt hat man sich auf folgende: Die Einschränkungen

$$\begin{aligned}
\cos\colon\ & [0,\pi] \longrightarrow [-1,1] \\
\sin\colon\ & \left[-\frac{\pi}{2},\frac{\pi}{2}\right] \longrightarrow [-1,1] \\
\tan\colon\ & \left(-\frac{\pi}{2},\frac{\pi}{2}\right) \longrightarrow \mathbb{R} \\
\cot\colon\ & (0,\pi) \longrightarrow \mathbb{R}
\end{aligned}$$

haben als Umkehrungen die sogenannten *Arcusfunktionen*

$$\begin{aligned}
\arccos\colon\ & [-1,1] \longrightarrow [0,\pi] \\
\arcsin\colon\ & [-1,1] \longrightarrow \left[-\frac{\pi}{2},\frac{\pi}{2}\right] \\
\arctan\colon\ & \mathbb{R} \longrightarrow \left(-\frac{\pi}{2},\frac{\pi}{2}\right) \\
\operatorname{arccot}\colon\ & \mathbb{R} \longrightarrow (0,\pi).
\end{aligned}$$

Arcusfunktionen braucht man im Beweis von

12.11 Satz *$I \subset \mathbb{R}$ sei ein offenes Intervall, dessen Länge höchstens 2π ist. Dann ist die Funktion*

$$I \longrightarrow S = \left\{ w \in \mathbb{C} \,\middle|\, |w| = 1 \right\}, \quad y \mapsto e^{iy}$$

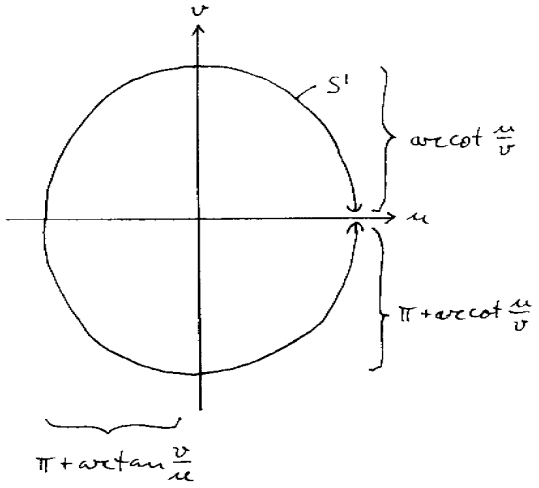
injektiv, und ihre Umkehrfunktion ist stetig.

Beweis Die Injektivität ist nach Satz 12.9 klar. Daß die Umkehrung stetig ist, beweisen wir der Einfachheit halber nur für den repräsentativen Fall $I = (0, 2\pi)$; die Bildmenge ist dann

$$S' = \left\{ w \in \mathbb{C} \,\middle|\, |w| = 1,\ w \neq 1 \right\}.$$

Ich behaupte, die Umkehrung ist durch die Formeln

$$S' \longrightarrow I, \quad w = u + iv \mapsto \begin{cases} \operatorname{arccot} \frac{u}{v} & (v > 0) \\ \pi + \arctan \frac{v}{u} & (u < 0) \\ \pi + \operatorname{arccot} \frac{u}{v} & (v < 0) \end{cases}$$



gegeben. Da $I \ni y \mapsto e^{iy} \in S'$ surjektiv ist, brauchen wir uns bloß davon zu überzeugen, daß Einsetzen von $u = \cos y$ und $v = \sin y$ in jede der drei Formeln genau $y \in (0, 2\pi)$ zurückgibt: dazu muß man nur beachten, daß die Kompositionen $\operatorname{arccot} \circ \cot$ und $\arctan \circ \tan$ ja nicht automatisch die Identität, sondern im allgemeinen Verschiebungen um ganze Vielfache von π sind — welche, das wird erst durch die Vorzeichen von u und v festgelegt. Insbesondere ergibt sich so, daß die Formeln dort, wo mehrere anwendbar sind, denselben Wert liefern.

Die Stetigkeit der Umkehrung folgt jetzt daraus, daß sie nach Lemma 7.3 eine in S' lokale Eigenschaft ist.

Bemerkung Es liegt nahe, durch Einschränken etwa auf $[0, 2\pi)$ eine stetige Abbildung

$$[0, 2\pi) \ni y \mapsto e^{iy} \in S = \{w \in \mathbb{C} \mid |w| = 1\}$$

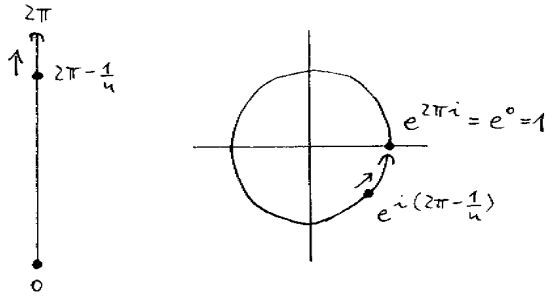
herzustellen, die sogar bijektiv ist. Deren Umkehrung ist aber bei $1 \in S$ unstetig: Die Folge in S

$$\left(e^{i\left(2\pi - \frac{1}{n}\right)}\right)_{n=1}^{\infty}$$

konvergiert gegen $e^{i \cdot 2\pi} = e^0 = 1$, während die Folge

$$\left(2\pi - \tfrac{1}{n}\right)_{n=1}^{\infty}$$

nicht gegen $0 \in [0, 2\pi)$ konvergiert (Satz 7.7).



Dem Satz 8.5 über die Stetigkeit der Umkehrfunktion widerspricht das deswegen nicht, weil es sich ja auch nicht um eine Abbildung zwischen Intervallen handelt.

Unentbehrlich ist die Exponentialfunktion bei der Beschreibung periodischer Vorgänge.

12.12 Definition Sei $T > 0$. Eine Funktion

$$f\colon \mathbb{R} \longrightarrow \mathbb{C}$$

heißt T-periodisch, wenn

$$f(t+T) = f(t) \quad \text{für alle } t \in \mathbb{R}$$

gilt.

Offenbar ist

$$\mathbb{R} \ni t \mapsto e^{2\pi i \frac{t}{T}} \in S$$

T-periodisch. Ist nun $g\colon S \longrightarrow \mathbb{C}$ eine beliebige Funktion, so ist die Komposition

$$\mathbb{R} \ni t \mapsto g\left(e^{2\pi i \frac{t}{T}}\right) \in \mathbb{C}$$

ebenfalls T-periodisch: $g\left(e^{2\pi i \frac{t+T}{T}}\right) = g\left(e^{2\pi i \frac{t}{T}}\right)$.

Ist umgekehrt eine beliebige T-periodische Funktion

$$\mathbb{R} \xrightarrow{f} \mathbb{C}$$

gegeben, so wird durch

$$S \ni e^{iy} \mapsto f\left(\tfrac{T}{2\pi} y\right) \in \mathbb{R}$$

eine Funktion g erklärt: tatsächlich folgt aus $e^{ix} = e^{iy}$ nach Satz 12.9 ja $x = y + k \cdot 2\pi$ mit $k \in \mathbb{Z}$ und damit

$$f\left(\frac{T}{2\pi} x\right) = f\left(\frac{T}{2\pi} y + kT\right) = f\left(\frac{T}{2\pi} y\right).$$

Das halten wir fest als

12.13 Notiz T-periodische Funktionen $f\colon \mathbb{R} \longrightarrow \mathbb{C}$ entsprechen vermöge der Formel

$$f(t) = g\left(e^{2\pi i \frac{t}{T}}\right)$$

umkehrbar eindeutig Funktionen $g\colon S \longrightarrow \mathbb{C}$.

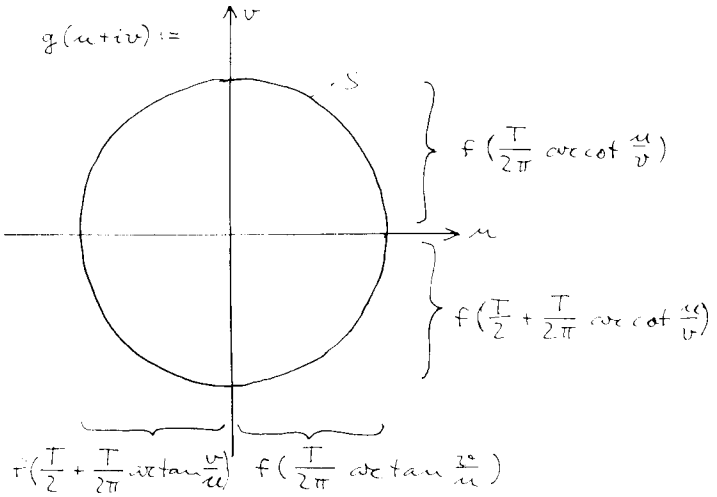
Im wesentlichen aus Satz 12.11 ziehen wir die

12.14 Folgerung Eine Funktion $g\colon S \longrightarrow \mathbb{C}$ ist genau dann stetig, wenn die zugehörige T-periodische Funktion $f\colon \mathbb{R} \longrightarrow \mathbb{C}$ stetig ist.

Beweis Wenn g stetig ist, dann ist die Komposition

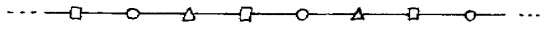
$$\begin{array}{rccccc} f\colon & \mathbb{R} & \longrightarrow & S & \longrightarrow & \mathbb{C} \\ & t & \mapsto & e^{2\pi i \frac{t}{T}} & \mapsto & g\left(e^{2\pi i \frac{t}{T}}\right) \end{array}$$

auch stetig. Umgekehrt sei f als stetig vorausgesetzt. Dann können wir g zwar nicht durch eine einzige "glatte" Formel aus f gewinnen, wohl aber durch vier auf den verschiedenen Halbkreisen (ohne Endpunkte) gültige, nämlich:

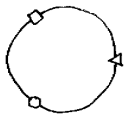


Wieder weil Stetigkeit eine lokale Eigenschaft ist, genügt das.

Bemerkungen Die Periode $T > 0$ wird im allgemeinen natürlich nicht die kleinstmögliche Periode von f sein. Man braucht übrigens nicht unbedingt an zeitliche Periodizität zu denken, auch wenn ich das durch die Buchstabenwahl vielleicht suggeriert habe. Ein grundlegender Ansatz der Festkörperphysik besteht darin, sich Kristallgitter allseits unbegrenzt fortgesetzt zu denken; Objekte, die makroskopisch beobachtbaren Größen entsprechen, sollten dann bezüglich dieses Gitters periodisch sein.



Nach unseren Überlegungen sind Funktionen mit makroskopischer Bedeutung auf einem eindimensionalen Kristall also im wesentlichen Funktionen auf der Kreislinie S (bei einem richtigen, dreidimensionalen Kristallgitter dann Funktionen auf $S \times S \times S$).



Schließlich sind wir jetzt auch in der Lage, uns von der komplexen Exponentialfunktion insgesamt ein gutes Bild zu machen. Schreiben wir

$$z = x + iy,$$

so hängt wegen $e^z = e^x e^{iy}$ der Betrag

$$|e^z| = |e^x| \cdot |e^{iy}| = e^x \in (0, \infty)$$

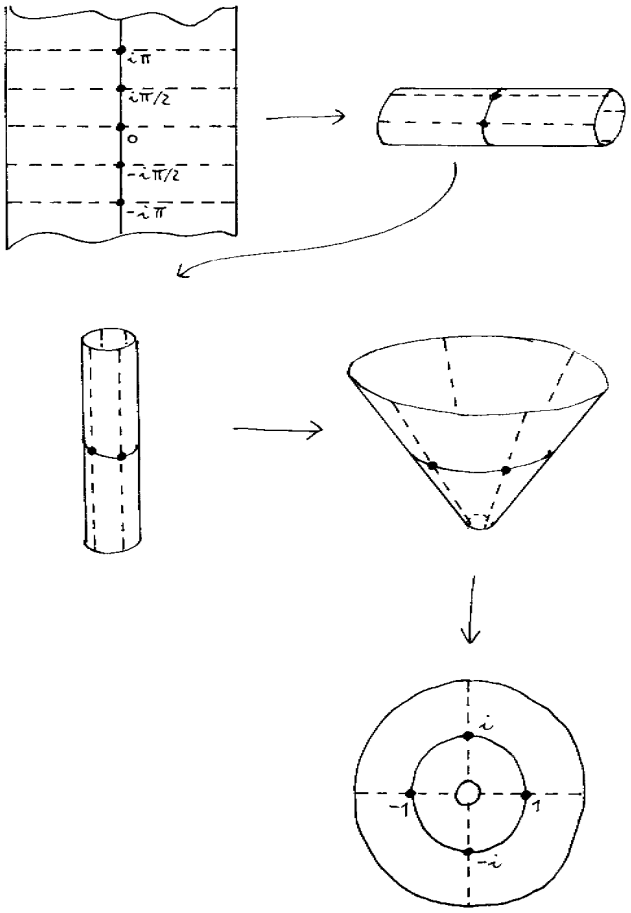
nur von x, und der "Winkelanteil"

$$\frac{e^z}{|e^z|} = e^{iy} \in S$$

nur von y ab. Die komplexe Exponentialfunktion ist daher eng mit der Abbildung durch ebene Polarkoordinaten

$$(0, \infty) \times \mathbb{R} \ni (r, \varphi) \mapsto (r \cos \varphi, r \sin \varphi) \in \mathbb{R}^2$$

verwandt: Bei der üblichen Identifikation von $\mathbb{C}$ mit $\mathbb{R}^2$ spielt $r \in (0, \infty)$ die Rolle von e^x, also x die von $\log r$, und y unmittelbar die des Winkels φ. Der folgende Comic strip illustriert die Wirkung von exp auf einen vertikalen Streifen $[-a, a] \times i\,\mathbb{R} = \{z \in \mathbb{C} \mid |\mathrm{Re}\, z| \leq a\}$:



Jetzt ist auch klar, was die Multiplikation komplexer Zahlen geometrisch bedeutet: Das Multiplizieren mit einer festen komplexen Zahl $e^{iy} \in S$ wirkt auf die Gaußsche Zahlenebene als Drehung um den (gerichteten) Winkel y, zum Beispiel dreht die Multiplikation mit $i = e^{i\pi/2}$ um einen rechten Winkel. Für eine beliebige komplexe Zahl $z \neq 0$ kommt dann noch eine Streckung um den reellen Faktor $|z|$ hinzu.

Analog zu den Sätzen 12.9 und 12.11 wollen wir noch festhalten:

12.15 Satz Die Abbildung $\mathbb{C} \xrightarrow{\exp} \mathbb{C} \setminus \{0\}$ ist surjektiv, und es gilt $e^w = e^z$ genau dann, wenn $w - z = k \cdot 2\pi i$ für ein $k \in \mathbb{Z}$ ist.

12.16 Satz Sei $I \subset \mathbb{R}$ sei ein offenes Intervall der Länge höchstens 2π. Dann besitzt die Einschränkung von exp auf den horizontalen Streifen

$$\mathbb{R} \times i\,I = \{z \in \mathbb{C} \mid \operatorname{Im} z \in I\} \xrightarrow{\exp} \mathbb{C} \setminus \{0\},$$

die ja injektiv (aber nicht surjektiv) ist, auf ihrem Bild eine stetige Umkehrung. Zum Beispiel ist für die Wahl $I = \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ das Bild

$$\exp\Big(\mathbb{R} \times i\left(-\tfrac{\pi}{2}, \tfrac{\pi}{2}\right)\Big) = \{w \in \mathbb{C} \mid \operatorname{Re} w > 0\}$$

die rechte Halbebene, und exp wird dort durch

$$w = u + iv \mapsto \log|w| + i \arctan \frac{v}{u}$$

umgekehrt.

Natürlich nennt man eine solche Umkehrung einen komplexen Logarithmus, aber eben nur *einen*. Anders als im Reellen gibt es viele komplexe Logarithmusfunktionen, zu deren genauer Festlegung man exp erst genügend einschränken muß — wie, das hängt vom Kontext ab. Darauf muß man auch dann sorgfältig achten, wenn man für $1 < n \in \mathbb{N}$ durch

$$z \longmapsto \sqrt[n]{z} := e^{\frac{1}{n}\log z}$$

komplexe Wurzelfunktionen definiert: diese Formel ist durch eine genaue Festlegung des verwendeten Logarithmus zu ergänzen (insbesondere niemals für alle $z \in \mathbb{C}$ simultan gültig). Hat man das getan, so liefern

$$e^{\frac{1}{n}(2\pi i+\log z)}, e^{\frac{1}{n}(2\cdot 2\pi i+\log z)}, \dots, e^{\frac{1}{n}((n-1)\cdot 2\pi i+\log z)}$$

$n-1$ weitere n-te Wurzelfunktionen mit demselben Definitionsbereich, wie man durch Potenzieren sofort verifizieren kann.

Übungsaufgaben

12.1 Beweisen Sie die beiden folgenden Abschätzungen:

(a) $\left(1+\frac{1}{n}\right)^n \le \sum_{k=0}^{n} \frac{1}{k!}$ und damit $\left(1+\frac{1}{n}\right)^n < e$ für jedes $n \ge 1$

(b) $e < 3$

12.2 Beweisen Sie, daß die eulersche Zahl e irrational ist.

Tip: Wenn Sie vorher Teil (b) der vorigen Aufgabe lösen, hilft Ihnen zwar nicht das Resultat, wohl aber das dabei erworbene Know-how weiter.

12.3 Beweisen Sie für jedes $n \in \mathbb{N}$ und jedes $y \in \mathbb{R}$, das kein Vielfaches von 2π ist:

$$\sum_{k=0}^{n} \sin ky = \frac{\sin \frac{n}{2}y \cdot \sin \frac{n+1}{2}y}{\sin \frac{1}{2}y}$$

Fällt beim Beweis auch eine Formel für $\sum_{k=0}^{n} \cos ky$ ab?

12.4 Berechnen Sie für jedes positive $n \in \mathbb{N}$ die n Nullstellen des komplexen Polynoms $z \mapsto z^n - 1$ (die sogenannten n-ten Einheitswurzeln). Zeigen Sie, daß diese eine Untergruppe der Kreislinie $S = \{z \in \mathbb{C} \mid |z| = 1\}$ bilden.

12.5 Beweisen Sie darüber hinaus (schwieriger): Für jedes $z \in S$ ist

$$\langle z \rangle := \{z^n \mid n \in \mathbb{Z}\}$$

eine Untergruppe von S. Wenn z *keine* Einheitswurzel ist (wenn also $z^n \ne 1$ für jedes $n > 0$ ist), dann ist $\langle z \rangle$ in dem Sinne dicht in S, daß es zu jedem $c \in S$ und jedem $\delta > 0$ ein $w \in \langle z \rangle$ mit $|w - c| < \delta$ gibt.

12.6 Skizzieren Sie die Menge

$$B := \{w \in \mathbb{C} \mid |w| \le 1 \text{ und } \mathrm{Im} w \ge |\mathrm{Re} w|\}$$

und berechnen und skizzieren Sie ihr Urbild $\exp^{-1}(B)$ unter der Exponentialabbildung $\exp\colon \mathbb{C} \longrightarrow \mathbb{C}$.

12.7 Berechnen Sie alle Lösungen $z \in \mathbb{C}$ der Gleichung

$$\cos z + \sin z = 2.$$

12.8 $g\colon S \longrightarrow \mathbb{R}$ sei eine stetige Funktion auf der Kreislinie $S = \{w \in \mathbb{C} \,|\, |w| = 1\}$. Beweisen Sie, daß die Wertemenge von g ein kompaktes Intervall $[c, d] \subset \mathbb{R}$ ist und daß jede Zahl aus (c, d) mindestens zweimal von g getroffen wird.

13 Differenzieren

Jetzt sind wir bei einem Kernthema dieses Kurses angelangt. Ohne Zweifel hat jeder von Ihnen schon mal vom Differenzieren gehört; und ich vertraue darauf, daß Sie in der Tätigkeit des Differenzierens schon so geübt sind, wie ein angehender Physiker das sein muß. Hier in der Vorlesung werde ich die Akzente auf Punkte setzen, die bei einer ersten Bekanntschaft mit dem Thema oft untergehen, die ich aber doch für wichtig halte. Das beginnt damit, daß man nicht alle Funktionen differenzieren kann, sondern eben nur die differenzierbaren.

13.1 Definition $I \subset \mathbb{R}$ sei ein echtes Intervall und $a \in I$. Eine Funktion $f\colon I \longrightarrow \mathbb{R}$ heißt bei a differenzierbar, wenn

$$f'(a) := \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \in \mathbb{R}$$

existiert. f heißt differenzierbar schlechthin, wenn f an jeder Stelle $a \in I$ differenzierbar ist; die damit definierte Funktion

$$f'\colon I \longrightarrow \mathbb{R}$$

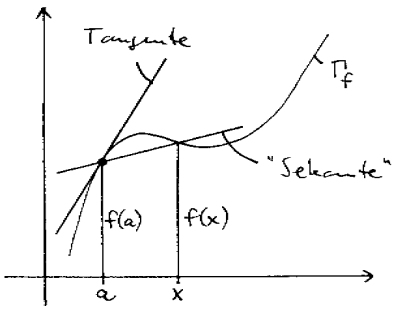
heißt die (erste) Ableitung von f.

Alternative Schreibweisen sind

$$\frac{df}{dx}\bigg|_{x=a} = \frac{d}{dx} f\bigg|_{x=a} = \frac{d}{dx} f(x)\bigg|_{x=a} = \frac{df}{dx}(a)$$

(aber nicht $\frac{df}{da}$) anstelle von $f'(a)$.

Die geometrische Deutung ist wohlbekannt: Für festes $x \neq a$ ist der Differenzenquotient $\dfrac{f(x) - f(a)}{x - a}$ die Steigung der eingezeichneten "Sekanten",



und im Falle der Differenzierbarkeit "konvergiert" letztere für $x \to a$ gegen die Tangente an den Graphen von f im Punkt $\big(a, f(a)\big)$; deren Steigung ist also $f'(a)$.

Bemerkungen Es kann bequem sein, eine Hilfsvariable h einzuführen, die in einem 0 enthaltenden Intervall lebt, und die Definition in der Form

$$f'(a) = \lim_{h \to 0} \frac{f(a + h) - f(a)}{h}$$

zu lesen. — Ist a ein Randpunkt von I und möchte man das betonen, so spricht man von rechts- bzw. linksseitiger Differenzierbarkeit und Ableitung bei a. — Oft ist die folgende bruchstrichfreie Beschreibung der ersten Ableitung von Vorteil:

13.1½ Notiz Die Funktion $f\colon I \longrightarrow \mathbb{R}$ sei auf dem echten Intervall $I \subset \mathbb{R}$ definiert, es sei $a \in I$ und $b \in \mathbb{R}$. Dann sind äquivalent:

(a) f ist bei a differenzierbar mit $f'(a) = b$

(b) $f(a+h) = f(a) + b{\cdot}h + o(h)$ für $h \to 0$

Beweis Teilt man die Formel in (b) durch h, so erhält man gleichwertig

$$\frac{f(a+h) - f(a)}{h} = b + o(1) \quad \text{für } h \to 0,$$

was bloß eine andere Formulierung von (a) ist.

13.2 Folgerung Nur eine bei a stetige Funktion kann dort differenzierbar sein.

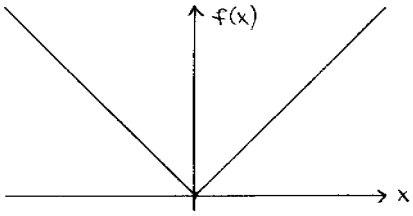
Beweis Für eine bei a differenzierbare Funktion f gilt

$$f(a+h) = f(a) + f'(a){\cdot}h + o(h) = f(a) + o(1) \quad \text{für } h \to 0,$$

d.h. $\lim\limits_{h\to 0} f(a+h) = f(a)$. Nach der Notiz 9.2 bedeutet das die Stetigkeit von f bei a.

13.3 Beispiele (1) Nicht jede stetige Funktion ist auch differenzierbar, wie die Funktion

$$f\colon \mathbb{R} \longrightarrow \mathbb{R},\ f(x) = |x|$$



belegt: f kann bei 0 nicht differenzierbar sein, weil der Differenzenquotient

$$\frac{f(x) - f(0)}{x - 0} = \frac{|x|}{x} = \begin{cases} -1 & (x < 0) \\ 1 & (x > 0), \end{cases}$$

für $x \to 0$ offenbar keinen Limes hat. Freilich ist das Beispiel ziemlich plump, die Lage nämlich sofort zu durchschauen: f ist bei 0 sowohl links- als auch rechtsseitig differenzierbar, bloß stimmen die beiden einseitigen Ableitungen nicht überein. — Es sei aber erwähnt, daß es stetige Funktionen gibt, die an keiner Stelle differenzierbar sind; tatsächlich gilt das sogar für die "meisten" stetigen Funktionen (in einem ähnlichen Sinne, wie die meisten reellen Zahlen nicht rational sind).

(2) Für jedes $n \in \mathbb{N}$ ist die Funktion $x \mapsto x^n$ differenzierbar, mit Ableitung $x \mapsto nx^{n-1}$: Man rechne nach dem binomischen Satz

$$(x+h)^n = x^n + nx^{n-1}h + \sum_{k=2}^{n} \binom{n}{k} x^{n-k}h^k$$

und bemerke $\sum_{k=2}^{n} \binom{n}{k} x^{n-k}h^k = o(h)$.

Zum systematischen Aufbau eines Vorrats an differenzierbaren Funktionen und zur Berechnung ihrer Ableitungen braucht man wieder die üblichen

13.4 Regeln Summen, Vielfache, Produkte, Quotienten und Kompositionen differenzierbarer Funktionen f, g sind differenzierbar, und ihre Ableitungen berechnen sich so:

$$\begin{aligned}
(f+g)' &= f' + g' \\
(\lambda f)' &= \lambda f' \qquad \text{für } \lambda \in \mathbb{R} \\
(fg)' &= f'g + fg' \\
\left(\frac{f}{g}\right)' &= \frac{f'g - fg'}{g^2} \\
(g \circ f)' &= (g' \circ f) \cdot f' \quad \text{(sogenannte Kettenregel)}
\end{aligned}$$

Beweis der Produkt- und der Kettenregel Man verwendet systematisch die Notiz $13.1\frac{1}{2}$, schreibt bei der Produktregel also für $h \to 0$

$$f(x+h) = f(x) + f'(x)h + o(h) \quad \text{sowie} \quad g(x+h) = g(x) + g'(x)h + o(h)$$

und multipliziert aus:

$$\begin{aligned}(fg)(x+h) &= f(x+h)g(x+h)\\ &= \big(f(x) + f'(x)h + o(h)\big)\big(g(x) + g'(x)h + o(h)\big)\\ &= f(x)g(x) + \big(f(x)g'(x) + f'(x)g(x)\big)h + o(h)\\ &= (fg)(x) + (fg'+f'g)(x)\cdot h + o(h)\end{aligned}$$

Zur Kettenregel: Sei $f(x) = y$. Ausgehend von

$$f(x+h) = f(x) + f'(x)h + o(h) \quad (h \to 0) \quad \text{und} \quad g(y+k) = g(y) + g'(y)k + o(k) \quad (k \to 0)$$

rechnen wir

$$\begin{aligned}(g\circ f)(x+h) &= g\big(f(x+h)\big)\\ &= g\big(f(x) + f'(x)h + o(h)\big)\\ &= g(y) + g'(y)\big(f'(x)h + o(h)\big) + o\big(f'(x)h + o(h)\big)\\ &= g(y) + g'(y)f'(x)h + g'(y)o(h) + o\big(O(h)\big)\\ &= g(y) + g'(y)f'(x)h + o(h)\\ &= (g\circ f)(x) + (g'\circ f)(x)f'(x)h + o(h)\end{aligned}$$

und lesen mittels $13.1\frac{1}{2}$ wieder alles ab.

Mit diesen Regeln kann man zum Beispiel nachrechnen, daß die Formel $\frac{d}{dx}x^n = nx^{n-1}$ auch mit negativen ganzen Exponenten n gilt:

$$\frac{d}{dx}\frac{1}{x^n} = -\frac{n}{x^{n+1}} \quad \text{für } x \neq 0.$$

Überdies sind wir jetzt in der Lage, jede rationale Funktion zu differenzieren (die Ableitung ist wieder rational), und damit wollen wir im Augenblick zufrieden sein.

Kann man eigentlich auch komplexe Funktionen differenzieren? Gewiß. Dabei wollen wir aber zwei wesentlich verschiedene Situationen unterscheiden. Ist $I \subset \mathbb{R}$ ein echtes Intervall wie gehabt, $f\colon I \longrightarrow \mathbb{C}$ aber jetzt eine komplex*wertige* Funktion, so ist für $a \in I$ eben auch

$$f'(a) = \lim_{x\to a}\frac{f(x)-f(a)}{x-a} \in \mathbb{C}$$

eine komplexe Zahl und deshalb $f'\colon I \longrightarrow \mathbb{C}$ wieder komplexwertig. Weil $\dfrac{1}{x-a}$ reell ist, sieht man sofort, daß die Zerlegung

$$f = u + iv \quad (\text{d.h. } f(x) = u(x) + iv(x) \text{ für } x \in I)$$

von f in Real- und Imaginärteil die Zerlegung von

$$f' = u' + iv'$$

in die Ableitung u' von u als Realteil und die von v als Imaginärteil nach sich zieht.

Die andere denkbare und interessante Situation ist die, daß $f\colon G \longrightarrow \mathbb{C}$ auf einem Gebiet $G \subset \mathbb{C}$ erklärt ist. Auch hier gibt der Begriff der Differenzierbarkeit und der Ableitung bei $a \in G$

$$f'(a) = \left.\frac{df}{dz}\right|_{z=a} = \lim_{z\to a}\frac{f(z)-f(a)}{z-a} \in \mathbb{C}$$

durchaus einen Sinn, und auch die Ableitungsregeln bleiben, da formal hergeleitet, richtig, zum Beispiel

$$\frac{d}{dz}z^n = nz^{n-1} \quad \text{für alle } n \in \mathbb{Z},\ z \in \mathbb{C}\ (z \neq 0 \text{ für } n < 0).$$

Die Differenzierbarkeit einer solchen komplexen Funktion hat aber eine tiefere Bedeutung als im reellen Fall, wo sie nur für "Glattheit" der Funktion steht. Zum Beispiel ist eine so harmlose Funktion wie die Konjugation

$$\mathbb{C} \ni z \mapsto \overline{z} \in \mathbb{C}$$

erstaunlicherweise an keiner Stelle differenzierbar: Der mit komplexem $h \neq 0$ zu bildende Differenzenquotient bei $z \in \mathbb{C}$

$$\frac{\overline{z+h} - \overline{z}}{h} = \frac{\overline{h}}{h} = \frac{\overline{h}^2}{|h|^2}$$

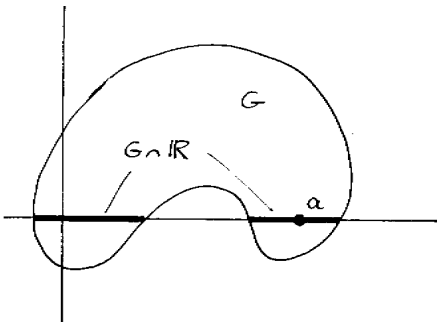
kann für $h \to 0$ keinen Limes haben, weil

$$\lim_{\mathbb{R} \ni h \to 0} \frac{\overline{h}^2}{|h|^2} = 1,$$

aber

$$\lim_{i\mathbb{R} \ni ih \to 0} \frac{\left(\overline{ih}\right)^2}{|ih|^2} = \lim_{\mathbb{R} \ni h \to 0} \frac{-h^2}{|h|^2} = -1$$

ist. — Immerhin können wir sicher sein, daß alle komplexen rationalen Funktionen differenzierbar sind. Außerdem stimmt die Ableitung einer auf einem Gebiet G erklärten komplexen Funktion in einem reellen Punkt $a \in G$ im Falle ihrer Existenz natürlich mit der reell, d.h. nach Einschränkung auf ein Intervall gebildeten überein.



Reell oder komplex, in jedem Fall ist die erste Ableitung einer differenzierbaren Funktion wieder eine Funktion auf demselben Definitionsbereich. Diese braucht nicht wieder differenzierbar, ja nicht einmal stetig zu sein. In der Praxis ist sie es aber oft, und deshalb ist wichtig:

13.5 Definition Ist die Ableitung einer differenzierbaren Funktion f selbst differenzierbar, so nennt man f zweimal differenzierbar und

$$f'' := (f')'$$

die zweite Ableitung von f. Entsprechend redet man für $k \in \mathbb{N}$ von der k-ten Ableitung

$$f^{(k)} := \left(f^{(k-1)}\right)'$$

von f, falls f eben k-mal differenzierbar ist. Für den Fall, daß $f^{(k)}$ stetig ist, hat sich noch die Sprechweise "f ist k-mal stetig differenzierbar, oder eine C^k-Funktion" eingebürgert. Insbesondere also:

$$\begin{aligned} C^0 &= \text{ stetig} \\ C^1 &= \text{ differenzierbar + Ableitung ist stetig} \end{aligned}$$

Alternative Schreibweisen für die k-te Ableitung an der Stelle a:

$$\left.\frac{d^k f}{dx^k}\right|_{x=a} = \left.\frac{d^k}{dx^k} f\right|_{x=a} = \left.\left(\frac{d}{dx}\right)^k f(x)\right|_{x=a} = \frac{d^k f}{dx^k}(a)$$

Konkret hat man es am häufigsten mit Funktionen zu tun, die dort, wo sie überhaupt differenzierbar sind, das auch gleich beliebig oft und damit sogenannte C^∞-Funktionen sind. — Nur erwähnt sei an dieser Stelle, daß auf einem *Gebiet* erklärte differenzierbare *komplexe* Funktionen automatisch beliebig oft differenzierbar sind, weswegen man für die Prädikate C^k bei solchen Funktionen keinen Bedarf hat.

Übungsaufgaben

13.1 Die Funktion $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ sei durch

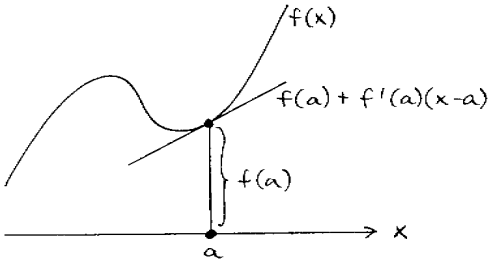
$$f(x) = \begin{cases} x^2 \cdot \cos(1/x) & (x \neq 0) \\ 0 & (x = 0) \end{cases}$$

erklärt. Zeigen Sie, daß f differenzierbar, die Ableitung f' an der Stelle 0 aber unstetig ist.

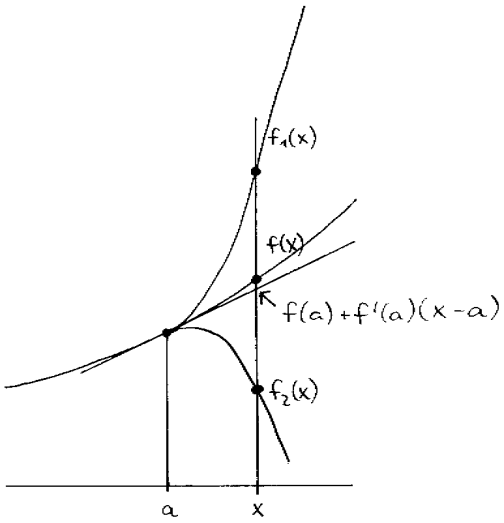
13.2 Zeigen Sie, daß die durch $f(z) = z + |z|^2$ gegebene Funktion $f\colon \mathbb{C} \to \mathbb{C}$ im Nullpunkt, aber nirgends sonst komplex differenzierbar ist.

14 Der Mittelwertsatz

In diesem Abschnitt geht es ausschließlich um reelle Funktionen auf Intervallen. Welche Anwendungen könnte die Differentialrechnung für eine solche Funktion f haben? Nun, geometrisch gesehen bedeutet Differenzieren von f bei a, die Tangente an den Graphen Γ_f im Punkt $(a, f(a))$ zu berechnen.



Man approximiert die Funktion f also in der Nähe von a — in einem bestimmten Sinne bestmöglich — durch eine sehr einfache, nämlich ein Polynom vom Grad (höchstens) 1. Als eine Anwendung könnte man sich vorstellen, daß man aus $f(a)$ und $f'(a)$ eben die Werte $f(x)$ für x nahe bei a näherungsweise berechnen kann. Die Skizze zeigt aber, daß das so einfach nicht geht:



An der Stelle x wird zwar f durch $f(a) + f'(a)\cdot(x-a)$ vernünftig approximiert, nicht aber die alternative Funktion f_1 oder f_2, obwohl beide bei a denselben Wert und dieselbe Ableitung wie f haben. Man hat einfach keine Kontrolle darüber, *wie gut* Γ_f durch die Tangente angenähert wird.

Für die geplante und für viele weitere Anwendungen genügt es deshalb nicht, f' nur an einer einzelnen Stelle a zu kennen, vielmehr muß man die Ableitung in einem ganzen Intervall verwenden. Den Schlüssel dazu bildet der sogenannte

14.1 Mittelwertsatz Es seien $a < b$ reelle Zahlen. Die Funktion

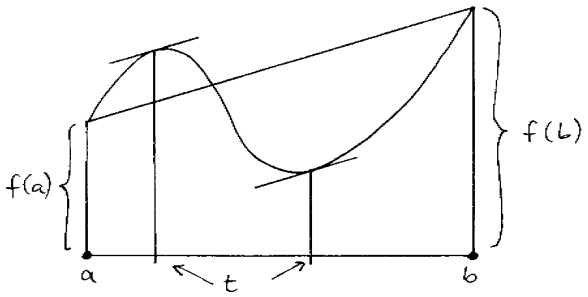
$$f\colon [a,b] \longrightarrow \mathbb{R}$$

sei stetig und im offenen Intervall (a, b) auch differenzierbar. Dann gibt es eine Stelle

$$t \in (a,b)$$

mit:

$$\frac{f(b)-f(a)}{b-a} = f'(t)$$



In der Skizze ist $\frac{f(b)-f(a)}{b-a}$ die Steigung der eingezeichneten Strecke; man könnte sie die mittlere Steigung von f zwischen a und b nennen (denken Sie an eine Straße). Der Mittelwertsatz verspricht, daß diese mittlere Steigung ingendwo im Inneren des Intervalls als tatsächliche Steigung vorkommt.

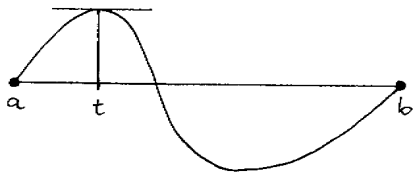
Um den Satz zu beweisen, behandelt man zuerst den für sich schon nützlichen Spezialfall $f(a) = f(b) = 0$, bekannt als

14.2 Satz von **Rolle** Gelte $a < b$,

$$f\colon [a,b] \longrightarrow \mathbb{R}$$

stetig, in (a,b) sogar differenzierbar, $f(a) = f(b) = 0$. Dann gibt es ein $t \in (a,b)$ mit $f'(t) = 0$.

Beweis Für $f = 0$ tut's jedes $t \in (a,b)$. Im anderen Fall dürfen wir annehmen, daß $f(x) > 0$ für mindestens ein $x \in [a,b]$ gilt, sonst können wir mit $-f$ weiterarbeiten. Nun nimmt f auf dem kompakten Intervall $[a,b]$ nach Satz 8.4 ein Maximum an, etwa bei t. Notwendigerweise ist dann $f(t) > 0$, also $t \in (a,b)$. Wir zeigen $f'(t) = 0$:



Es gilt

$$\frac{f(x)-f(t)}{x-t} \begin{cases} \geq 0 & \text{für } x \in (a,t) \\ \leq 0 & \text{für } x \in (t,b), \end{cases}$$

daher

$$\lim_{x \nearrow t} \frac{f(x)-f(t)}{x-t} \geq 0$$
$$\lim_{x \searrow t} \frac{f(x)-f(t)}{x-t} \leq 0$$

und damit $f'(t) = 0$, weil $f'(t)$ beiden Limites gleich ist.

Beweis des Mittelwertsatzes Wir bilden aus dem gegebenen $f\colon [a,b] \longrightarrow \mathbb{R}$ die Hilfsfunktion

$$h\colon [a,b] \longrightarrow \mathbb{R}, \quad h(x) = f(x) - \alpha - \beta x,$$

worin wir $\alpha, \beta \in \mathbb{R}$ so bestimmen, daß

$$h(a) = f(a) - \alpha - \beta a = 0$$
$$h(b) = f(b) - \alpha - \beta b = 0$$

wird. Wegen $a \neq b$ ist das möglich, mit

$$\beta = \frac{f(b) - f(a)}{b - a}.$$

Der Satz von Rolle serviert uns ein $t \in (a, b)$ mit

$$h'(t) = f'(t) - \beta = 0,$$

d.h. mit

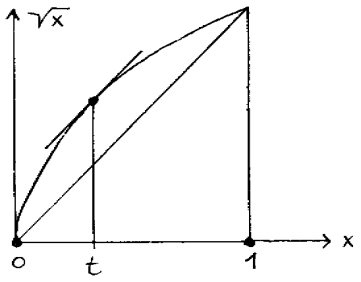
$$f'(t) = \beta = \frac{f(b) - f(a)}{b - a}.$$

Fertig.

Anmerkungen (1) Bitte achten Sie bei diesen und den folgenden Sätzen auf die *genauen* Voraussetzungen; diese sind auch praktisch wichtig. Sie erlauben es zum Beispiel, den Mittelwertsatz auf die Funktion

$$[0, 1] \longrightarrow \mathbb{R}, \quad x \mapsto \sqrt{x}$$

anzuwenden, die an der Stelle 0 nicht differenzierbar ist (näheres dazu gleich):



(2) Man kann die Aussage des Mittelwertsatzes schon als eine einfache Approximationsformel im Sinne der Einleitung zu diesem Abschnitt ansehen: Kennt man den Wert $f(a)$, sowie eine Näherung für f' zwischen a und b, präzise ausgedrückt also Schranken $\alpha, \beta \in \mathbb{R}$ mit

$$\alpha \leq f'(t) \leq \beta \quad \text{für alle } t \in (a, b),$$

so erhält man für $f(b) = f(a) + f'(t) \cdot (b - a)$ die Näherung

$$f(a) + \alpha \cdot (b - a) \leq f(b) \leq f(a) + \beta \cdot (b - a).$$

Wenn man will, kann man den folgenden hübschen Satz als einen Extremfall dieser Näherungsformel auffassen.

14.3 Satz Sei $f: I \longrightarrow \mathbb{R}$ eine differenzierbare Funktion auf einem echten Intervall. Dann gilt:

$$f' = 0 \iff f \text{ ist konstant}$$

Beweis Daß die Ableitung einer konstanten Funktion überall verschwindet, ist klar. Sei umgekehrt $f' = 0$ vorausgesetzt. Für je zwei Zahlen $a < b$ in I ist dann $[a, b] \subset I$, und nach dem Mittelwertsatz gibt es ein $t \in [a, b]$ mit

$$f(b) - f(a) = f'(t) \cdot (b - a) = 0 \cdot (b - a) = 0.$$

Also ist f konstant.

Von großer Bedeutung sind die Anwendungen des Mittelwertsatzes auf Monotonie und lokale Extrema:

14.4 Satz Sei $f: [a, b] \longrightarrow \mathbb{R}$ stetig, auf (a, b) sogar differenzierbar. Dann gilt:

(a) $f'(x) \geq 0$ für alle $x \in (a, b) \iff f$ wächst monoton

(b) $f'(x) > 0$ für alle $x \in (a, b) \implies f$ wächst streng monoton

Ist (a) erfüllt und f auch bei a oder b differenzierbar, so ist auch die einseitige Ableitung dort nicht-negativ.

Beweis Sei $f'(x) \geq 0$ bzw. $f'(x) > 0$ für alle $x \in (a, b)$. Zu beliebigen $x, y \in \mathbb{R}$ mit

$$a \leq x < y \leq b$$

gibt es nach dem Mittelwertsatz ein $t \in (x, y) \subset (a, b)$ mit

$$f(y) - f(x) = f'(t) \cdot (y - x),$$

und es folgt $f(x) \leq f(y)$ bzw. $f(x) < f(y)$.

Weiß man umgekehrt, daß f monoton wächst, so sind für jedes $x \in [a, b]$ alle Quotienten

$$\frac{f(y) - f(x)}{y - x} \quad (y \neq x)$$

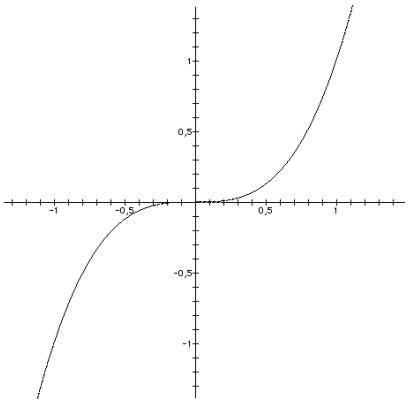
nicht-negativ, und $f'(x)$ als deren Limes also auch.

Bemerkungen (1) Achten Sie auch hier darauf: Zum Nachweis der strengen Monotonie im *gesamten* Intervall $[a, b]$ genügt es nach (b) schon, $f' > 0$ im *offenen* Intervall (a, b) zu prüfen — ja f' braucht in den Randpunkten nicht mal zu existieren.

(2) Der Satz gilt natürlich auch für andere Intervalltypen (wobei die vorige Bemerkung ganz oder teilweise gegenstandslos wird).

(3) Die Aussage (b) ist nicht umkehrbar, wie das Beispiel

$$f\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad x \mapsto x^3$$



mit $f'(x) = 3x^2$ zeigt. Trotzdem könnten wir (b) benutzen, um die strenge Monotonie dieser Funktion zu beweisen (sie ist freilich ohnehin klar): nach (b) wachsen die Einschränkungen $f|(-\infty, 0]$ und $f|[0, \infty)$ streng monoton, und daraus folgt natürlich die strenge Monotonie von f selbst. Ich hoffe, daß Sie in solchen Fällen (die häufig sind) den gesunden Menschenverstand walten lassen und sich entsprechend zu helfen wissen.

Erst mit dem Satz 14.4, speziell (b), wird der frühere Satz über die Umkehrfunktion stetiger Funktionen auch praktisch leicht anwendbar. Die Schwierigkeit lag bisher ja vor allem darin, die Injektivität — de facto die strenge Monotonie — der umzukehrenden Funktion zu erkennnen.

Eine etwas andere Frage ist die, ob eine differenzierbare Funktion eine Umkehrung besitzt, die selbst differenzierbar ist. Hier gibt die Differentialrechnung sogar vollständige Auskunft:

14.5 Satz $I \subset \mathbb{R}$ sei ein echtes Intervall, und $f\colon I \longrightarrow \mathbb{R}$ sei differenzierbar. Dann sind äquivalent:

(a) $f'(x) > 0$ für alle $x \in I$

(b) f wächst streng monoton *und* die Umkehrfunktion $f^{-1}\colon f(I) \longrightarrow I \subset \mathbb{R}$ ist differenzierbar.

Die Ableitung von f^{-1} berechnet sich dann zu $(f^{-1})' = \dfrac{1}{f' \circ f^{-1}}$, also

$$\big(f^{-1}\big)'(y) = \frac{1}{f'\big(f^{-1}(y)\big)} = \frac{1}{\left.\dfrac{df}{dx}\right|_{x=f^{-1}(y)}}$$

Beweis (a) $\Rightarrow$ (b) Die Existenz und Stetigkeit von f^{-1} wissen wir nach den Sätzen 14.4(b) und 8.5 schon; bleibt nur die Differenzierbarkeit zu untersuchen, etwa bei

$$b = f(a) \in f(I).$$

Dazu schreiben wir den Diffenzenquotienten als

$$\frac{f^{-1}(y) - f^{-1}(b)}{y - b} = \left(\frac{f\left(f^{-1}(y)\right) - f(a)}{f^{-1}(y) - a}\right)^{-1}$$

um; weil f^{-1} an der Stelle b stetig ist, gilt $\lim\limits_{y\to b} f^{-1}(y) = f^{-1}(b) = a$, folglich

$$\lim_{y\to b} \frac{f\left(f^{-1}(y)\right) - f(a)}{f^{-1}(y) - a} = f'(a) \neq 0$$

und damit

$$\lim_{y\to b} \frac{f^{-1}(y) - f^{-1}(b)}{y - b} = \frac{1}{f'(a)} = \frac{1}{f'\big(f^{-1}(b)\big)}$$

wie behauptet.

(b) $\Rightarrow$ (a) Aus der wachsenden Monotonie von f folgt $f'(x) \geq 0$ für alle $x \in I$ nach Teil (a) von Satz 14.4. Andererseits folgt aus

$$f^{-1} \circ f = \mathrm{id}$$

nach der Kettenregel

$$\big(f^{-1}\big)'\big(f(x)\big) \cdot f'(x) = \mathrm{id}'(x) = 1$$

für jedes $x \in I$. Insbesondere muß $f'(x) \neq 0$, also $f'(x) > 0$ sein.

Bemerkung Selbstverständlich gibt es denselben Satz in einer fallenden Version, mit derselben Formel für die Ableitung von f^{-1}.

14.6 Beispiel Die Potenzfunktionen $x \mapsto x^n$ mit $0 < n \in \mathbb{N}$ sind auf $[0,\infty)$ und für ungerades n sogar auf ganz $\mathbb{R}$ streng monoton wachsend, ihre Umkehrungen

$$[0,\infty) \longrightarrow [0,\infty) \text{ bzw. } \mathbb{R} \longrightarrow \mathbb{R}, \quad y \mapsto \sqrt[n]{y}$$

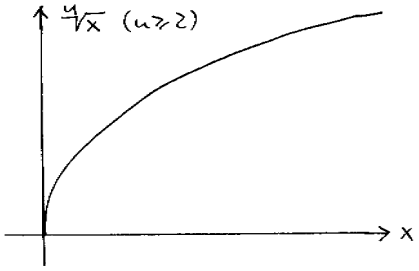
stetig. Aber differenzierbar sind die Wurzelfunktionen nur außerhalb des Nullpunktes:

$$\frac{d}{dy}\sqrt[n]{y} = \frac{1}{\left.\dfrac{d}{dx}x^n\right|_{x=\sqrt[n]{y}}} = \frac{1}{n \cdot x^{n-1}|_{x=\sqrt[n]{y}}} = \frac{1}{n \cdot \sqrt[n]{y^{n-1}}} \quad (y \neq 0),$$

denn für $n \geq 2$ verschwindet die Ableitung der Potenzfunktion bei 0:

$$\left.\frac{d}{dx}x^n\right|_{x=0} = n \cdot 0^{n-1} = 0$$

Anschaulich-geometrisch wird das das darin sichtbar, daß der Graph der Wurzelfunktion im Punkt 0 eine vertikale Tangente hat.



Die Ableitungsformel merkt man sich übrigens am einfachsten in der zum Fall eines ganzzahligen Exponenten analogen Schreibweise:

$$\frac{d}{dx} x^{\frac{1}{n}} = \frac{1}{n} x^{\frac{1}{n}-1}$$

Eng mit den Monotoniefragen zusammen hängt die Suche nach Extrema einer reellwertigen Funktion; dies ist die sicher populärste Anwendung der Differentialrechnung überhaupt. Wir haben im Umfeld von Satz 8.4 schon darüber gesprochen, was es heißt, daß eine Funktion $f\colon X \longrightarrow \mathbb{R}$ an einer Stelle $a \in X$ ein Extremum (ihr Minimum oder Maximum) annimmt. Den Wert $f(a)$ nennt man in diesem Fall auch das globale Minimum bzw. Maximum von f, um es von dem folgenden subtileren Begriff zu unterscheiden.

14.7 Definition Sei $X \subset \mathbb{R}$, und sei $f\colon X \longrightarrow \mathbb{R}$ eine Funktion. Man sagt, f hat an der Stelle $a \in X$ ein lokales Minimum, wenn es ein $\delta > 0$ gibt mit

$$f(x) \geq f(a) \quad \text{für alle } x \in X \text{ mit } |x-a| < \delta.$$

Von einem strengen lokalen Minimum spricht man, wenn sogar

$$f(x) > f(a) \quad \text{für alle } x \in X \text{ mit } 0 < |x-a| < \delta$$

ist. Analog natürlich lokale Maxima.

14.8 Satz $I \subset \mathbb{R}$ sei ein *offenes* Intervall, und $f\colon I \longrightarrow \mathbb{R}$ eine differenzierbare Funktion. Dann gilt:

(a) Hat f bei $a \in I$ ein lokales Extremum, so ist $f'(a) = 0$.

(b) Ist $f'(a) = 0$ und f bei a sogar zweimal differenzierbar mit $f''(a) > 0$, so hat f bei a ein strenges lokales Minimum.

Beweis (a) beweist man wie beim Satz von Rolle: Wenn $f(x) \geq f(a)$ für alle $x \in I$ mit $|x-a| < \delta$ gilt, verkleinern wir $\delta > 0$ so weit, daß $(a-\delta, a+\delta) \subset I$ wird, haben dann

$$\frac{f(x)-f(a)}{x-a} \begin{cases} \leq 0 & \text{für } x \in (a-\delta, a) \\ \geq 0 & \text{für } x \in (a, a+\delta) \end{cases}$$

und schließen $f'(a) = 0$.

(b) Gelte $f'(a) = 0$ und $f''(a) > 0$. Dann gibt es ein $\delta > 0$ mit $(a-\delta, a+\delta) \subset I$ und

$$\frac{f'(x)-f'(a)}{x-a} > 0 \quad \text{für alle } x \in I \text{ mit } 0 < |x-a| < \delta,$$

denn der Limes dieses Quotienten für $x \to a$ ist ja $f''(a)$. Nun ist $f'(a) = 0$, also folgt

$$f'(x) \begin{cases} < 0 & \text{für } x \in (a-\delta, a) \\ > 0 & \text{für } x \in (a, a+\delta). \end{cases}$$

Nach Satz 14.4(b) fällt/wächst f auf $(a-\delta, a]$ bzw. $[a, a+\delta)$ streng monoton; insbesondere gilt:

$$f(x) > f(a) \quad \text{für } x \in (a-\delta, a) \cup (a, a+\delta)$$

Fertig.

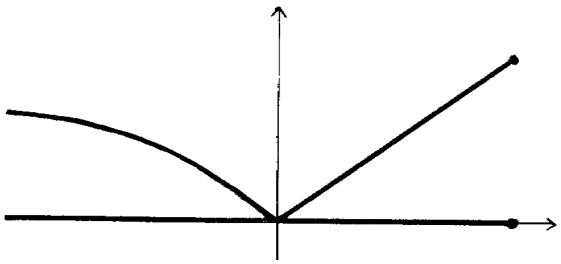
Bemerkungen (1) Natürlich liefert Teil (b) für $f''(a) < 0$ ein strenges lokales Maximum.

(2) Wie schon erwähnt, ist der Satz sehr beliebt; fast ebenso beliebt ist es aber auch, ihn falsch anzuwenden. Häufig wird geglaubt, (b) lasse sich umkehren. Aber schon an der einfachen Funktion $\mathbb{R} \ni x \mapsto x^4 \in \mathbb{R}$ sieht man, daß das falsch ist:

$$\frac{d}{dx} x^4 \Big|_{x=0} = 4x^3\big|_{x=0} = 0$$
$$\frac{d^2}{dx^2} x^4 \Big|_{x=0} = \frac{d}{dx} 4x^3 \Big|_{x=0} = 12x^2\big|_{x=0} = 0,$$

trotzdem hat diese Funktion bei 0 ein strenges (sogar globales) Minimum.

(3) Ein anderer beliebter Fehler besteht in dem Versuch, die Methode des Satzes gedankenlos auf Funktionen anzuwenden, deren Definitionsintervall nicht offen ist oder die nicht überall differenzierbar sind. Man darf sich nicht wundern, daß man dann etwa bei der Funktion



sowohl das Minimum bei 0 als auch das Maximum im Endpunkt des Intervalls übersieht. Solche Punkte müssen immer gesondert untersucht werden.

(4) Wenn man aus irgendeinem Grunde ohnehin alle Intervalle bestimmt, auf denen die untersuchte Funktion monoton ist, kann man sich die Anwendung von Satz 14.8(b) und damit die Berechnung der zweiten Ableitung in der Regel sparen: Hat man zum Beispiel für $a < b < c$ streng wachsende Monotonie auf $(a, b]$ und streng fallende auf $[b, c)$ nachgewiesen, so liegt bei b natürlich ein strenges lokales Maximum, während in den offenen Intervallen (a, b) und (b, c) kein lokales Extremum angenommen werden kann.

(5) Mit der Differentialrechnung lassen sich manche Ungleichungen routinemäßig herleiten, zu deren Beweis man sonst erst einen speziellen Ansatz finden müßte. Bei der Ungleichung

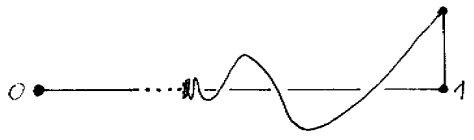
$$x(1-x) \le \frac{1}{4} \quad \text{für jedes } x \in \mathbb{R}$$

bestünde dieser Ansatz darin, links die quadratische Ergänzung durchzuführen. Zum Beweis mittels Differentialrechnung dagegen bestimmt man man nach Satz 14.8 alle lokalen Extrema der Funktion

$$f\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad x \mapsto x(1-x),$$

sieht sofort, daß das einzige, das man findet, nämlich bei $1/2$, auch das globale Maximum von f ist, und braucht nur noch $f(1/2) = 1/4$ auszurechnen.

(6) Auch noch so differenzierbare, also "glatte" Funktionen können überraschende Eigenschaften haben. So gibt es — um nur zwei Beispiele zu nennnen — C^∞-Funktionen auf $[0, 1]$, die unendlich viele lokale Minima und Maxima besitzen,



und andere, die streng monoton wachsen, deren erste Ableitung aber mehr als abzählbar viele Nullstellen besitzt. Man darf die in diesem Abschnitt erklärten Methoden daher nicht eigenmächtig um Argumente "anreichern", die bloß auf die Anschauung gestützt sind. Angesichts der Reichhaltigkeit und leichten Anwendbarkeit dieser Methoden sollte die Versuchung, das zu tun, aber auch nicht besonders groß sein.

Zum Schluß dieses Abschnitts stelle ich Ihnen ohne Beweis zwei bekannte Regeln vor, mit der man viele Grenzwerte von Funktionen bequem berechnen kann; sie beruhen auf einer Verallgemeinerung des Mittelwertsatzes. Ich fasse mich dabei kurz; ausführliche Darstellungen der Beweise finden Sie in der Standardliteratur.

14.9 $\frac{0}{0}$**-Regel** von **de l'Hospital** Sei $a < b$ (hier ist $b = \infty$ zugelassen).

$$f\colon (a,b) \longrightarrow \mathbb{R} \quad \text{und} \quad g\colon (a,b) \longrightarrow \mathbb{R}$$

seien differenzierbare Funktionen mit

$$\lim_{x\to b} f(x) = \lim_{x\to b} g(x) = 0$$

und $g'(x) \neq 0$ für alle $x \in (a,b)$. Wenn dann

$$\lim_{x\to b} \frac{f'(x)}{g'(x)} \in [-\infty, \infty]$$

existiert, dann existiert

$$\lim_{x\to b} \frac{f(x)}{g(x)} \in [-\infty, \infty]$$

auch, mit demselben Wert.

14.10 $\frac{\infty}{\infty}$**-Regel** von **de l'Hospital** Gleicher Satz, statt $\lim f(x) = \lim g(x) = 0$ aber die Voraussetzung

$$\lim_{x\to b} f(x) = \lim_{x\to b} g(x) = \infty.$$

Beide Regeln gelten analog natürlich auch für rechts- und für beidseitige Grenzwerte.

Bemerkungen Die Bezeichnungen $\frac{0}{0}$ und $\frac{\infty}{\infty}$ sind natürlich nur als Merkhilfe gemeint. — Bei der Einfachheit dieser sehr beliebten Regeln sollte man meinen, daß man in ihrer Anwendung nichts verkehrt machen kann. Trotzdem werden sie oft falsch (und dann meist auch mit falschem Resultat) eingesetzt, indem entweder nicht überprüft wird, daß tatsächlich eine $\frac{0}{0}$- oder $\frac{\infty}{\infty}$-Situation vorliegt, oder indem bei der Existenz der beiden Limites die logische Schlußrichtung mißachtet wird.

14.11 Beispiel Mit einigem Einfallsreichtum kann man direkt mit den früheren Methoden beweisen, daß für beliebiges $t \in \mathbb{R}$

$$\lim_{n\to\infty} \left(1 + \frac{t}{n}\right)^n = e^t$$

gilt. Mit Regel 14.9 kommen wir dagegen fast ohne Nachdenken zum Ziel: Wir ziehen gleich $\left(1 + \frac{t}{x}\right)^x$ für reelle $x > 0$ in Betracht und schreiben

$$\left(1 + \frac{t}{x}\right)^x = \exp\left(x \cdot \log\left(1 + \frac{t}{x}\right)\right) = \exp \frac{\log\left(1 + \frac{t}{x}\right)}{1/x}\ .$$

Wegen $\lim\limits_{x\to\infty} \log\left(1 + \frac{t}{x}\right) = 0$ hat man im Exponenten eine $\frac{0}{0}$-Situation mit, in den Bezeichnungen der Regel 14.9,

$$f(x) = \log\left(1 + \frac{t}{x}\right) \quad \text{und} \quad g(x) = \frac{1}{x}$$

und damit (die Ableitung des Logarithmus aus dem nächsten Abschnitt vorwegnehmend)

$$f'(x) = \frac{1}{1+\frac{t}{x}} \cdot \left(-\frac{t}{x^2}\right) \quad \text{und} \quad g'(x) = -\frac{1}{x^2} \neq 0.$$

Da $\lim\limits_{x\to\infty} \dfrac{f'(x)}{g'(x)} = \lim\limits_{x\to\infty} \dfrac{t}{1+\frac{t}{x}} = t$ existiert, folgt auch

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = t$$

und wegen der Stetigkeit von exp bei t schließlich

$$\lim_{x\to\infty} \left(1+\frac{t}{x}\right)^x = e^t.$$

Übungsaufgaben

14.1 Seien $I \subset \mathbb{R}$ ein Intervall, $b > 1$ und $c \geq 0$ reelle Konstanten und $f\colon I \longrightarrow \mathbb{R}$ eine Funktion mit

$$|f(x) - f(y)| \leq c \cdot |x-y|^b \quad \text{für alle } x, y \in I.$$

Beweisen Sie, daß f eine konstante Funktion ist.

14.2 Sei $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ eine differenzierbare Funktion mit den Eigenschaften

$$f(0) = 0 \quad \text{und} \quad |f'(x)| \leq \frac{1}{2} \text{ für alle } x \in \mathbb{R}.$$

Beweisen Sie

$$\lim_{n\to\infty} f^n(x) = 0 \text{ für jedes } x \in \mathbb{R},$$

wobei mit f^n hier nicht die Potenz, sondern die n-fache Komposition $f^n = f \circ f \circ \cdots \circ f$ gemeint ist.

14.3 Lassen Sie in der Situation der vorigen Aufgabe die Forderung $f(0) = 0$ fallen und beweisen Sie allgemeiner, daß es (unter ansonsten unveränderten Voraussetzungen) genau ein $a \in \mathbb{R}$ mit $f(a) = a$ gibt (einen sogenannten *Fixpunkt* von f), und daß

$$\lim_{n\to\infty} f^n(x) = a \text{ für jedes } x \in \mathbb{R}$$

gilt.

14.4 Bestimmen Sie alle Intervalle, auf denen die Funktion $f\colon \mathbb{R} \longrightarrow \mathbb{R};\ f(x) = x^3 e^{-x^2}$ streng monoton ist.

14.5 Bestimmen Sie alle Stellen, an denen die Funktion $f\colon \mathbb{R} \longrightarrow \mathbb{R};\ f(x) = (1-x)\sqrt{|x|}$ ein lokales Extremum hat.

14.6 Für festes reelles $T > 0$ sei die Funktion $f\colon (0,\infty) \longrightarrow \mathbb{R}$ durch

$$f(\omega) = \frac{\omega^3}{\exp\frac{\omega}{T} - 1}$$

erklärt (wenn T und ω die in der Physik übliche Bedeutung haben, ist f die Strahlungsdichte gemäß der Planckschen Formel). Berechnen Sie, soweit diese Grenzwerte existieren,

$$\lim_{\omega\to 0} f(\omega) \quad \text{und} \quad \lim_{\omega\to\infty} f(\omega).$$

Beweisen Sie, daß f an genau einer Stelle ω_0 ein lokales Extremum hat, und daß dieses das globale Maximum von f ist. Zeigen Sie, daß ω_0 zu T proportional ist.

14.7 Seien $\omega > 0$ und $\gamma \geq 0$ reelle Zahlen. Bestimmen Sie alle lokalen Extrema der Funktion $f\colon [0,\infty) \longrightarrow \mathbb{R}$ mit

$$f(t) = e^{-\gamma t}\cos\omega t.$$

Bemerkung: Diese ganz reelle Aufgabe kann man auch ganz reell lösen, aber man muß nicht.

14.8 Berechnen Sie einige der folgenden Grenzwerte:

(a) $\lim\limits_{x\to 0} \dfrac{3^x - 2^x}{x}$ (b) $\lim\limits_{x\to 0} \dfrac{(x-\sin x)^8}{(1-\cos x)^{12}}$ (c) $\lim\limits_{x\searrow 0} \dfrac{\log\tan 2x}{\log\tan 3x}$

(d) $\lim\limits_{x\to 1} \dfrac{x^3+x^2-x-1}{x^2-1}$ (e) $\lim\limits_{x\to 1}\left(\dfrac{1}{x-1} - \dfrac{1}{\log x}\right)$ (f) $\lim\limits_{x\to\infty} \dfrac{e^x - e^{-x}}{e^x + e^{-x}}$

(g) $\lim\limits_{x\to\infty} \dfrac{\log x}{\log x + \sin x}$

Erinnerung: Die Formulierung der Aufgabe soll nicht schon die Behauptung enthalten, daß diese Grenzwerte existieren.

14.9 Sei $I \subset \mathbb{R}$ ein echtes Intervall, $a \in I$. Die Funktion $f\colon I \longrightarrow \mathbb{R}$ sei stetig und in $I\setminus\{a\}$ differenzierbar. Zeigen Sie: Wenn $\lim\limits_{x\to a} f'(x) \in \mathbb{R}$ existiert, dann ist f auch an der Stelle a differenzierbar und f' dort stetig.

14.10 Die Funktion $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ sei durch

$$f(x) = \begin{cases} 0 & (x \leq 0) \\ e^{-\frac{1}{x}} & (x > 0) \end{cases}$$

erklärt. Verschaffen Sie sich genügend viel Information über diese Funktion, um eine realistische Skizze des Graphen Γ_f anzufertigen.

14.11 Beweisen Sie, daß die Funktion f aus der vorigen Aufgabe eine C^∞-Funktion, also auch bei 0 beliebig oft differenzierbar ist.

Untersuchen Sie dazu (soweit nicht schon geschehen) $\lim\limits_{x\searrow 0} f(x)$ und $\lim\limits_{x\searrow 0} f'(x)$; beachten Sie, daß die Existenz des zweiten Limes noch nichts darüber sagt, ob f auch bei 0 differenzierbar ist, daß aber ...

Wenn Sie nun gezeigt haben, daß f eine C^1-Funktion ist, werden Sie den Beweis durch vollständige Induktion auf die höheren Ableitungen verallgemeinern wollen. Überlegen Sie sich, daß Sie dazu nicht unbedingt explizite Formeln für die Ableitungen von f brauchen (die sind kompliziert), sondern mit einer Aussage über deren allgemeine Gestalt gut zurechtkommen.

14.12 Sei $I \subset \mathbb{R}$ ein echtes Intervall, $a \in I$. $f\colon I \longrightarrow \mathbb{R}$ sei eine C^k-Funktion mit

$$f(a) = f'(a) = \cdots = f^{(k)}(a) = 0.$$

Beweisen Sie, daß dann

$$f(x) = o(|x-a|^k),$$

und wenn $f^{(k+1)}(a)$ existiert, sogar

$$f(x) = O(|x-a|^{k+1})$$

gilt.

15 Analytische Funktionen

Uns fehlt noch eine Regel, um die durch Potenzreihen dargestellten Funktionen zu differenzieren. Ich möchte diese Regel aber noch einen Moment zurückstellen, um vorher die so erklärten Funktionen in den ihnen gebührenden systematischen Rahmen stellen. Wir haben schon ein wenig mit Potenzreihen gerechnet, insbesondere über Addition und (Satz 11.10) Multiplikation von Potenzreihen gesprochen. Jetzt wollen wir darüber hinaus konvergente Potenzreihen ineinander einsetzen.

Dazu fixieren wir zwei Potenzreihen

$$\sum_{j=0}^{\infty} a_j(z-a)^j \quad \text{und} \quad \sum_{k=0}^{\infty} b_k(w-b)^k$$

um möglicherweise verschiedene komplexe Punkte a und b; wir wollen versuchen, die erste Reihe in die zweite einzusetzen, also

- $$\sum_{k=0}^{\infty} b_k \left(\sum_{j=0}^{\infty} a_j(z-a)^j - b \right)^k$$

durch Zusammenfassen nach Potenzen von $z-a$ zu einer neuen Potenzreihe

$$\sum_{n=0}^{\infty} c_n(z-a)^n$$

um a zu machen. Probieren wir doch mal, die ersten Koeffizienten c_0 und c_1 auszurechnen: In

$$\begin{aligned} c_0 + c_1(z-a) + \cdots &= \sum_{k=0}^{\infty} b_k \big(a_0 + a_1(z-a) - b\big)^k + \cdots \\ &= \sum_{k=0}^{\infty} b_k \left((a_0-b)^k + k \cdot (a_0-b)^{k-1} a_1(z-a) \right) + \cdots \\ &= \underbrace{\sum_{k=0}^{\infty} b_k (a_0-b)^k}_{c_0} + \underbrace{a_1 \sum_{k=1}^{\infty} k\, b_k \big(a_0-b\big)^{k-1}}_{c_1} (z-a) + \cdots \end{aligned}$$

stehen die Pünktchen für alle Terme, die mindestens $(z-a)^2$ als Faktor enthalten und deshalb auf c_0 und c_1 keinen Einfluß haben können. Wie Sie sehen, sind diese Koeffizienten selbst unendliche Reihen, in deren Glieder die a_j, die b_k und b eingehen. Wenn eine dieser Reihen divergiert, ist die durch • beschriebene "eingesetzte Reihe" nicht definiert.

Es zeigt sich aber wieder, daß die Konvergenz von Potenzreihen so robust ist, daß die eingesetzte Reihe doch existiert und in zu erwartendem Umfang konvergiert, sobald man entsprechende Konvergenzeigenschaften der Ausgangsreihen voraussetzt:

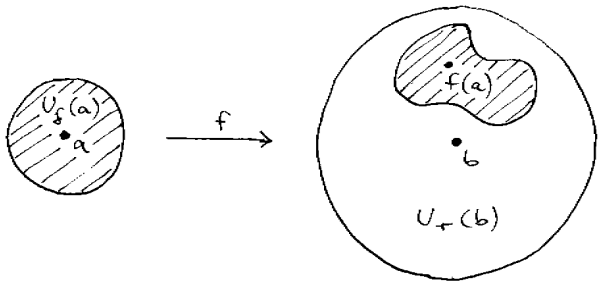
15.1 Satz Es seien

$$f(z) = \sum_{j=0}^{\infty} a_j(z-a)^j \quad \text{und} \quad g(z) = \sum_{k=0}^{\infty} b_k(w-b)^k$$

komplexe Potenzreihen, $r > 0$ der Konvergenzradius von g und $\delta > 0$ eine Zahl, so daß f in der Kreisscheibe $U_\delta(a)$ konvergiert und außerdem

$$f\big(U_\delta(a)\big) \subset U_r(b)$$

gilt (δ darf aber kleiner als der Konvergenzradius von f sein).



Dann ist die eingesetzte Reihe

$$\sum_{n=0}^{\infty} c_n(z-a)^n = \sum_{k=0}^{\infty} b_k \left(\sum_{j=0}^{\infty} a_j(z-a)^j - b \right)^k$$

definiert, und sie konvergiert auf $U_\delta(a)$ gegen die Komposition

$$g \circ f \colon U_\delta(a) \xrightarrow{f} U_r(b) \xrightarrow{g} \mathbb{C}.$$

Beweis Der Multiplikationssatz 11.10 folgt ja ziemlich einfach aus dem Satz 6.6 über Doppelreihen. Der Beweis hier ist im Prinzip ähnlich, aber doch komplizierter, und ihm liegt eine verallgemeinerte Version des Satzes 6.8 zugrunde.

Kompliziert ist im allgemeinen auch die Berechnung einzelner Koeffizienten c_n aus den Ausgangsdaten. In manchen Spezialfällen, die aber schon interessant sind, wird es einfacher:

15.2 Beispiele (1) Wenn wir in die geometrische Reihe

$$\frac{1}{1-w} = g(w) = \sum_{k=0}^{\infty} w^k \quad (|w| < 1)$$

die "Potenzreihe" (um 0)

$$f(z) = -z^2$$

einsetzen, ergibt sich

$$\frac{1}{1+z^2} = (g \circ f)(z) = \sum_{k=0}^{\infty} (-1)^k z^{2k} = 1 - z^2 + z^4 - z^6 + \cdots,$$

gültig für alle $z \in \mathbb{C}$ mit $f(z) \in U_1(0)$, d.h. mit $|z| < 1$. Hier braucht man Satz 15.1 natürlich gar nicht.

(2) Gleiches g, aber allgemeiner

$$f(z) = \sum_{j=1}^{\infty} a_j(z-a)^j$$

eine beliebige Potenzreihe um a mit $f(a) = 0$ und positivem Konvergenzradius. Weil f bei a stetig ist, finden wir ein $\delta > 0$, so daß

$$|f(z)| < 1 \quad \text{für alle } z \in U_\delta(a)$$

gilt. Nach dem Satz ist

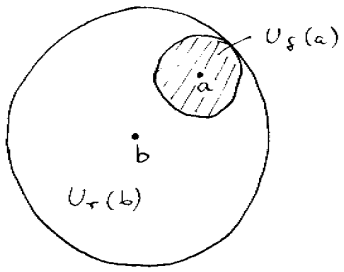
$$\frac{1}{1-f(z)} = (g \circ f)(z) = \sum_{k=0}^{\infty} \left(\sum_{j=1}^{\infty} a_j(z-a)^j \right)^k = \sum_{n=0}^{\infty} c_n(z-a)^n$$

eine in $U_\delta(a)$ konvergente Potenzreihe. Weil es hier kein a_0 gibt, tragen zum Koeffizienten von $(z-a)^n$ nur Summanden mit $k \le n$, und sowieso nur solche mit $j \le n$ bei, insgesamt also nur endlich viele. Deshalb sind die Koeffizienten c_n der neuen Reihe bloß algebraische Ausdrücke, keine unendlichen Reihen in den a_j.

(3) Jetzt sei $g(w) = \sum b_k(w-b)^k$ wieder beliebig mit Konvergenzradius $r > 0$. Wir wählen ein $a \in U_r(b)$ im Konvergenzkreis und nehmen als f die identische Funktion, aber als "Potenzreihe" um a geschrieben:

$$z = f(z) = a + (z-a)$$

Für $\delta := r - |a-b| > 0$ ist dann $f\big(U_\delta(a)\big) = U_\delta(a) \subset U_r(b)$,



und unter Benutzung von Satz 6.8 erhalten wir

$$\begin{aligned} g(z) = (g \circ f)(z) &= \sum_{k=0}^{\infty} b_k\big(a + (z-a) - b\big)^k \\ &= \sum_{k=0}^{\infty} b_k \sum_{n=0}^{k} \binom{k}{n} (a-b)^{k-n}(z-a)^n \\ &= \sum_{n \le k} b_k \binom{k}{n} (a-b)^{k-n}(z-a)^n \\ &= \sum_{n=0}^{\infty} \left(\sum_{k=n}^{\infty} b_k \binom{k}{n} (a-b)^{k-n} \right) (z-a)^n \quad \text{für } z \in U_\delta(a). \end{aligned}$$

Diese Formel beschreibt, wie man jede durch eine Potenzreihe um b definierte Funktion g als Potenzreihe um einen beliebigen anderen Punkt a ihres Konvergenzkreises darstellen kann, wobei man natürlich mit einer Schrumpfung des Konvergenzradius rechnen muß. Immerhin bleibt dieser positiv: konkret mindestens $\delta = r - |a-b|$.

Allgemein nennt man Funktionen, die man wenigstens lokal durch Potenzreihen darstellen kann, analytisch:

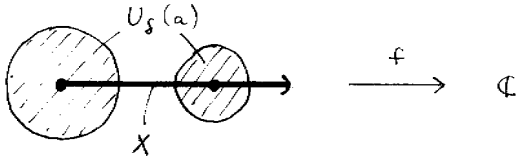
15.3 Definition Sei entweder $X \subset \mathbb{R}$ ein echtes Intervall oder $X \subset \mathbb{C}$ ein Gebiet. Eine Funktion

$$f\colon X \longrightarrow \mathbb{C}$$

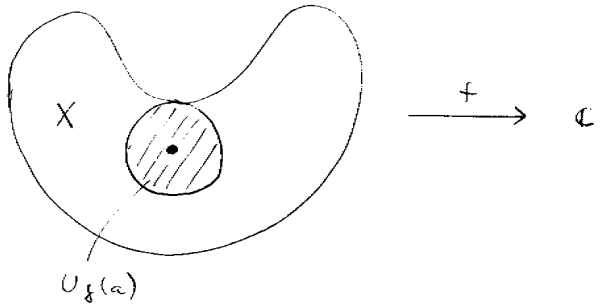
heißt analytisch, wenn es zu jedem $a \in X$ eine Potenzreihe $\sum a_n(z-a)^n$ um a und ein $\delta > 0$ gibt, so daß die Reihe auf $U_\delta(a)$ konvergiert und

$$f(z) = \sum_{n=0}^{\infty} a_n(z-a)^n \quad \text{für alle } z \in X \cap U_\delta(a)$$

gilt.



(reelle Version; f darf aber auch hier komplexe Werte annehmen)



(komplexe Version; hier kann man $\delta > 0$ so klein wählen, daß $U_\delta(a) \subset X$ ist)

Die Klasse der analytischen Funktionen ist zugegebenermaßen etwas kompliziert zu definieren; sie enthält aber viele wichtige Funktionen und erweist sich im Gebrauch als besonders gutmütig und nützlich. Dazu erst mal die

15.4 Notiz Jede Potenzreihe mit positivem Konvergenzradius definiert eine analytische Funktion auf ihrem Konvergenzkreis.

Beweis Das ist die Kernaussage von Beispiel (3).

15.5 Satz Nicht nur Summen und Produkte, sondern auch Quotienten und Kompositionen von analytischen Funktionen sind analytisch.

Beweis Für Summen und Produkte wissen wir das schon, und für die Kompositionen folgt es unmittelbar aus Satz 15.1. Statt des Quotienten zweier analytischen Funktionen brauchen wir nur noch den Kehrwert einer einzelnen, etwa h anzusehen, natürlich an einer Stelle $a \in \mathbb{C}$ mit $h(a) \neq 0$. Nach Definition gibt es ein $\delta > 0$, so daß h auf $U_\delta(a)$ durch eine Reihe

$$h(z) = \sum_{j=0}^{\infty} c_j (z-a)^j$$

dargestellt wird. Diese schreiben wir listig als

$$\begin{aligned} h(z) &= h(a) \cdot \sum_{j=0}^{\infty} \frac{c_j}{c_0} (z-a)^j \\ &= h(a) \cdot \left(1 - \sum_{j=1}^{\infty} \frac{-c_j}{c_0} (z-a)^j \right) \\ &= h(a) \cdot \big(1 - f(z)\big). \end{aligned}$$

Nun ist $f(a) = 0$, und nach Beispiel (2) also

$$\frac{1}{h(z)} = \frac{1}{h(a)} \cdot \frac{1}{1 - f(z)}$$

eine analytische Funktion.

Damit haben wir eine Menge uns vertrauter Funktionen als analytisch erkannt: die Exponentialfunktion und die trigonometrischen Funktionen, selbstverständlich die Polynome und damit auch alle rationalen Funktionen. — Jetzt werden wir die zu Beginn des Abschnitts angesprochene Lücke füllen und zeigen, daß man Potenzreihen gliedweise differenzieren darf. Zur Vorbereitung dient der folgende Satz, der auch für sich genommen interessant ist.

15.5$\frac{1}{2}$ Satz Sei $X \subset \mathbb{R}$ ein echtes Intervall oder $X \subset \mathbb{C}$ ein Gebiet. Die Folge $(f_n)_{n=0}^{\infty}$ differenzierbarer Funktionen $f_n \colon X \to \mathbb{C}$ konvergiere, etwa gegen die Grenzfunktion $f \colon X \to \mathbb{C}$, und die Folge ihrer Ableitungen konvergiere sogar gleichmäßig gegen $g \colon X \to \mathbb{C}$. Dann ist f differenzierbar und $f' = g$.

Beweis Wir beweisen die Differenzierbarkeit von f bei $a \in X$. Für jedes $n \in \mathbb{N}$ ist die Funktion

$$X \ni z \longmapsto F_n(z) := \begin{cases} \dfrac{f_n(z) - f_n(a)}{z - a} & \text{für } z \neq a \\ f_n'(a) & \text{für } z = a \end{cases}$$

stetig; wir zeigen zuerst, daß die Funktionenfolge $(F_n)_{n=0}^{\infty}$ eine gleichmäßige Cauchy-Folge ist. Von $(f_n')_{n=0}^{\infty}$ wissen wir das schon, und wählen zu gegebenem $\varepsilon > 0$ ein $D \in \mathbb{N}$, so daß

$$\left|f_{n+k}'(t) - f_n'(t)\right| < \varepsilon \quad \text{für alle } t \in X \text{ und } n, k \in \mathbb{N} \text{ mit } n > D$$

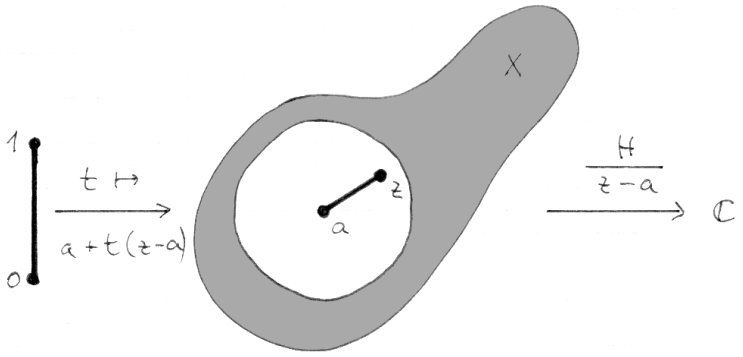
gilt; solche n und k halten wir vorübergehend fest. Im rein reellen Fall — X ein Intervall und f reellwertig — wenden wir den Mittelwertsatz 14.1 auf die Hilfsfunktion $H := f_{n+k} - f_n$ an, erhalten ein $t \in X$ mit

$$F_{n+k}(z) - F_n(z) = \frac{H(z) - H(a)}{z - a} = H'(t) = f_{n+k}'(t) - f_n'(t) \quad \text{für alle } z \neq a$$

und schließen $|F_{n+k}(z) - F_n(z)| < \varepsilon$ für alle $z \neq a$. Zusammen mit der uns bekannten Konvergenz an der einzelnen Stelle $z = a$ beweist das die gleichmäßige Cauchy-Eigenschaft der Folge $(F_n)_{n=0}^{\infty}$.

Wenn f auch nicht-reelle Werte hat, können wir den Mittelwertsatz nicht direkt auf die Hilfsfunktion H anwenden, wohl aber separat auf ihren Real- und Imaginärteil: statt $|F_{n+k}(z) - F_n(z)| < \varepsilon$ erhalten wir dieselbe Abschätzung für die Real- und Imaginärteile der Funktionen F_n, und das genügt uns auch. Ist schließlich X ein Gebiet, so dürfen wir annehmen, daß es sich um eine offene Kreisscheibe um den Punkt a etwa vom endlichen Radius r handelt, denn die Differenzierbarkeit bei a ist ein lokales Problem. Die Verbindungsstrecke von a nach z verläuft dann ganz in X, und wir verwenden statt H die neue Hilfsfunktion $h\colon [0,1] \to \mathbb{C}$ mit der Wirkung

$$t \longmapsto \frac{H\big(a + t\,(z-a)\big)}{z - a}\,,$$



die nach der Kettenregel die Ableitung $h'(t) = H'\big(a + t\,(z-a)\big)$ hat. Wegen

$$F_{n+k}(z) - F_n(z) = \frac{H(z) - H(a)}{z - a} = h(1) - h(0)$$

genügt es jetzt, den Mittelwertsatz auf Real- und Imaginärteil von h anzuwenden.

In jedem Fall hat sich die Folge $(F_n)_{n=0}^{\infty}$ als gleichmäßig konvergent erwiesen. Nach Satz 11.4 ist die Grenzfunktion F stetig, insbesondere folgt

$$\lim_{z \to a} \frac{f(z) - f(a)}{z - a} = \lim_{z \to a} F(z) = F(a) = g(a)\,,$$

und das war behauptet.

15.6 Satz Sei $a \in \mathbb{C}$, und sei

$$\sum_{n=0}^{\infty} a_n (z - a)^n$$

eine Potenzreihe um a mit Konvergenzradius $r \in [0,\infty]$. Dann hat die gliedweise differenzierte Reihe

$$\sum_{n=1}^{\infty} n\, a_n (z-a)^{n-1} = \sum_{m=0}^{\infty} (m+1)\, a_{m+1} (z-a)^m$$

denselben Konvergenzradius, und im Konvergenzkreis der Reihen, also für alle $z \in U_r(a)$ gilt:

$$\frac{d}{dz} \sum_{n=0}^{\infty} a_n (z-a)^n = \sum_{n=1}^{\infty} n\, a_n (z-a)^{n-1}$$

Insbesondere ist jede analytische Funktion beliebig oft differenzierbar, und ihre Ableitungen sind wieder analytisch.

Beweis Wo die Reihe $\sum_{n=1}^{\infty} n\, a_n (z-a)^{n-1}$ absolut konvergiert, da konvergiert $\sum_{n=0}^{\infty} a_n (z-a)^n$ auch, das folgt aus dem Majorantenkriterium:

$$\left| a_n (z-a)^n \right| = \frac{|z-a|}{n} \cdot \left| n\, a_n (z-a)^{n-1} \right| .$$

Der Konvergenzradius der differenzierten Reihe ist also sicher nicht größer. Um zu zeigen, daß er auch nicht kleiner ausfällt, betrachten wir wie im Beweis von 11.7 einen Konvergenzpunkt c der Ausgangsreihe und zeigen, daß die differenzierte Reihe an jeder Stelle z mit $|z-a| < |c-a|$ konvergiert. Für $z = a$ ist das trivial, und sonst setzen wir $q = |z-a|/|c-a| < 1$ und wissen

$$\left| n\, a_n (z-a)^{n-1} \right| = \frac{n}{|z-a|} \cdot \underbrace{|a_n||c-a|^n}_{\to 0} \cdot q^n \le n\, q^n \quad \text{für alle genügend großen } n .$$

Für die Quotienten zweier aufeinanderfolgender Terme rechts gilt nun

$$\lim_{n\to\infty} \frac{(n+1)\, q^{n+1}}{n\, q^n} = q ,$$

und die Konvergenz folgt nach dem Quotientenkriterium 5.14.

Zu beweisen bleibt, daß die Funktion $z \mapsto \sum_{n=0}^{\infty} a_n (z-a)^n$ in $U_r(a)$ differenzierbar ist und man die Ableitung wie behauptet durch gliedweises Differenzieren berechnen kann. Das ist aber eine lokale Frage, und weil die Potenzreihen nach 11.7(b) in jeder Kreisscheibe von echt kleinerem Radius gleichmäßig konvergieren, löst uns Satz $15.5\frac{1}{2}$ das Problem.

15.7 Beispiele (1) Die Exponentialreihe $\displaystyle\sum_{n=0}^{\infty} \frac{1}{n!} z^n$ reproduziert sich beim Differenzieren:

$$\sum_{n=1}^{\infty} n \frac{1}{n!} z^{n-1} = \sum_{n=1}^{\infty} \frac{1}{(n-1)!} z^{n-1} = \sum_{m=0}^{\infty} \frac{1}{m!} z^m$$

Also:

$$\exp' = \exp$$

Zusammen mit Satz 14.5 ergibt sich nun die Ableitung der Logarithmusfunktion $\log\colon (0,\infty) \longrightarrow \mathbb{R}$:

$$\frac{d}{dy} \log y = \frac{1}{\frac{d}{dx} e^x \big|_{x=\log y}} = \frac{1}{e^{\log y}} = \frac{1}{y}$$

Also:

$$\frac{d}{dx} \log x = \frac{1}{x} \quad (x > 0)$$

Damit ergibt sich für die allgemeine Potenz aus

$$\frac{d}{dx}x^b = \frac{d}{dx}e^{b\cdot\log x} = e^{b\cdot\log x}\cdot\frac{d}{dx}(b\cdot\log x) = x^b\cdot b\cdot\frac{1}{x} = b\cdot x^{b-1}$$

wieder die schon bekannte Formel

$$\frac{d}{dx}x^b = b\cdot x^{b-1},$$

die damit immer dann gilt, wenn x^{b-1} überhaupt Sinn hat (zum Beispiel für $x>0$, $b\in\mathbb{C}$).

(2) Analog kann man

$$\cos' = -\sin \quad \text{und} \quad \sin' = \cos$$

durch gliedweises Differenzieren der beiden Potenzreihen erhalten, wenn man es nicht vorzieht, die Darstellungen $\cos z = (e^{iz}+e^{-iz})/2$ usw. zu differenzieren. Mit Satz 14.5 kommt man schließlich an die Ableitungen der Arcusfunktionen

$$\frac{d}{dx}\arccos x = -\frac{1}{\sqrt{1-x^2}} \quad \text{für } x\in(-1,1)$$
$$\frac{d}{dx}\arcsin x = \frac{1}{\sqrt{1-x^2}} \quad \text{für } x\in(-1,1)$$
$$\frac{d}{dx}\arctan x = \frac{1}{1+x^2} \quad \text{für } x\in\mathbb{R}$$
$$\frac{d}{dx}\operatorname{arccot} x = -\frac{1}{1+x^2} \quad \text{für } x\in\mathbb{R}$$

Beachten Sie, daß Arcuscosinus und -sinus bei ±1 nicht differenzierbar sind, weil die Ableitung des Cosinus an den Stellen 0 und π, die des Sinus bei $\pm\frac{\pi}{2}$ verschwindet.

Es ist bemerkenswert, daß der Logarithmus und die Arcusfunktion Ableitungen haben, die "elementarer" sind als sie selbst. Man kann ganz allgemein nach differenzierbaren Funktionen fragen, die eine gegebene Funktion als Ableitung haben:

15.8 Definition Sei entweder $X\subset\mathbb{R}$ ein echtes Intervall oder $X\subset\mathbb{C}$ ein Gebiet, und sei $f\colon X\longrightarrow\mathbb{C}$ eine Funktion. Unter einer Stammfunktion von f versteht man eine differenzierbare Funktion $F\colon X\longrightarrow\mathbb{C}$ mit $F'=f$.

Es ist klar, daß mit F auch $F+c$ für jede Konstante $c\in\mathbb{C}$ eine Stammfunktion von f ist. Das erweist sich — zunächst nur für auf einem Intervall definierte Funktionen — als die einzige Freiheit bei der Wahl einer Stammfunktionen, denn Satz 14.3, gegebenenfalls getrennt auf Real- und Imaginärteil angewendet, liefert sofort die

15.9 Notiz Ist $I\subset\mathbb{R}$ ein Intervall und sind F und G zwei Stammfunktionen von $f\colon I\longrightarrow\mathbb{C}$, so ist $F-G$ eine konstante Funktion.

Stammfunktionen sind also bis auf eine additive Konstante eindeutig bestimmt. Wir werden später im Rahmen der Integralrechnung sehen, daß jede auf einem Intervall definierte stetige Funktion Stammfunktionen besitzt; wegen dieses Zusammenhangs hat es sich eingebürgert, das Aufsuchen einer Stammfunktion auch als *Integrieren* zu bezeichnen. Der Begriff der Stammfunktion hat aber zunächst mit Integralrechnung nichts zu tun.

In unserem Kontext können wir für jede durch eine konvergente Potenzreihe

$$f(z) = \sum_{n=0}^{\infty} a_n(z-a)^n$$

dargestellte Funktion $f\colon U_r(a)\longrightarrow\mathbb{C}$ eine Stammfunktion sofort aus Satz 15.6 ablesen: Die durch gliedweises Integrieren gebildete Potenzreihe

$$F(z) = \sum_{n=0}^{\infty}\frac{a_n}{n+1}(z-a)^{n+1} = \sum_{m=1}^{\infty}\frac{a_{m-1}}{m}(z-a)^m$$

hat ja als gliedweise Ableitung wieder die ursprüngliche Reihe. Wir merken uns das als

15.10 Notiz Potenzreihen dürfen gliedweise integriert werden.

Weitere Beispiele analytischer Funktionen beschert uns jetzt die

15.11 Folgerung Ist $I \subset \mathbb{R}$ ein echtes Intervall und $f\colon I \longrightarrow \mathbb{C}$ eine differenzierbare Funktion derart, daß f' analytisch ist, so ist f selbst analytisch.

Beweis Sei $a \in I$. Für genügend kleines $\delta > 0$ wird f' auf $I \cap U_\delta(a)$ durch eine konvergente Potenzreihe beschrieben:

$$f'(z) = \sum a_n (z-a)^n$$

Die gliedweise integrierte Reihe definiert eine Stammfunktion von f' auf dem Intervall $I \cap U_\delta(a)$. Nach Notiz 15.9 stimmt die Einschränkung $f|I \cap U_\delta(a)$ als weitere Stammfunktion bis auf eine additive Konstante mit der Summe dieser Reihe überein.

15.12 Beispiele (1) Auch die Funktionen log sowie arctan und arccot sind analytisch, als Stammfunktionen von

$$x \mapsto \frac{1}{x} \quad (x > 0), \qquad x \mapsto \pm\frac{1}{1+x^2} \quad (x \in \mathbb{R})$$

nämlich.

(2) Damit sind für alle $b \in \mathbb{C}$ die Funktionen

$$x \mapsto x^b = e^{b \cdot \log x} \quad (x > 0)$$

analytisch, und folglich auch arccos und arcsin auf dem *offenen* Intervall $(-1, 1)$, als Stammfunktionen von

$$x \mapsto \pm\frac{1}{\sqrt{1-x^2}} \quad (-1 < x < 1).$$

(3) Welche Funktionen sind nicht analytisch? Aufgrund von Satz 15.6 jedenfalls alle Funktionen, die nicht oder nicht beliebig oft differenzierbar sind. Abgesehen von abstrusen Funktionen fallen darunter zum Beispiel $x \mapsto |x|$ (bei 0), die Wurzeln $x \mapsto \sqrt[n]{x}$ für $n > 1$ (bei 0) und arccos, arcsin (bei ± 1). Wenn man mehr an komplexe Funktionen denkt, ist vor allem $z \mapsto \overline{z}$ (und überhaupt alles $\overline{z}$-haltige) zu erwähnen.

(4) Die reelle Funktion $f\colon \mathbb{R} \longrightarrow \mathbb{R}$,

$$f(x) = \begin{cases} 0 & (x \le 0) \\ e^{-\frac{1}{x}} & (x > 0) \end{cases}$$

ist eine C^∞-Funktion, das ist das Ergebnis der Aufgabe 14.9. Aber im folgenden Abschnitt wird sich erweisen, daß f nicht analytisch ist.

Übungsaufgaben

15.1 Der quantenmechanische sogenannte eindimensionale harmonische Oszillator führt bei gegebener Energie $E \in \mathbb{R}$ auf die Differentialgleichung

$$f''(x) - 2xf'(x) + (2E-1)f(x) = 0.$$

Bestimmen Sie alle Lösungen f dieser Gleichung, die auf einem nicht-leeren offenen Intervall $(-\delta, \delta)$ analytische Funktionen sind:

- Setzen Sie f als Potenzreihe $f(x) = \sum_{n=0}^{\infty} a_n x^n$ an und zeigen Sie, daß die Differentialgleichung zu einer Rekursionsformel für die Koeffizienten a_n äquivalent ist.
- Begründen Sie, daß es zu jeder Vorgabe von $a_0 \in \mathbb{R}$ genau eine gerade, und zu jeder Vorgabe von $a_1 \in \mathbb{R}$ genau eine ungerade Potenzreihe gibt, die die Differentialgleichung erfüllt.
- Beweisen Sie, daß jede dieser Potenzreihen überall konvergiert und damit eine sogar auf ganz $\mathbb{R}$ definierte analytische Lösung der Differentialgleichung ist.

15.2 Von den in der vorigen Aufgabe bestimmten Lösungsfunktionen f haben nicht alle eine physikalische Bedeutung, sondern nur diejenigen, für die zumindest

$$\lim_{x \to \pm\infty} e^{-x^2/2} f(x) = 0$$

ist. Bestimmen Sie, wann es solche Lösungen $f \neq 0$ gibt:

- Nur für spezielle Werte der Energie E (welche?) wird f ein Polynom.
- Für jeden anderen Wert von E gibt es ein $c > 0$ und ein $D \in \mathbb{R}$ mit

$$|f(x)| \geq c \cdot e^{x^2/2} \quad \text{für alle } x > D.$$

16 Taylor-Reihen

Konvergente Potenzreihen liefern analytische Funktionen. In diesem Abschnitt betrachten wir diesen Sachverhalt in umgekehrter Sicht: Wie kann man zu gegebener analytischer Funktion eine Potenzreihe finden, die diese Funktion lokal darstellt? Theoretisch wird diese Frage beantwortet durch das

16.1 Lemma Sei $f: X \longrightarrow \mathbb{C}$ analytisch, $a \in X$ ein Punkt. Wenn f auf einem a enthaltenden echten Intervall (reelle Version) bzw. auf einer offenen Kreisscheibe um a (komplexe Version) durch eine Potenzreihe

$$f(z) = \sum_{n=0}^{\infty} a_n (z-a)^n$$

dargestellt wird, dann gilt notwendig

$$a_n = \frac{f^{(n)}(a)}{n!} \quad \text{für alle } n \in \mathbb{N}.$$

Insbesondere ist die darstellende Reihe durch f und a eindeutig festgelegt.

Beweis Ganz einfach: k-maliges Differenzieren gibt:

$$f^{(k)}(z) = \sum_{n=k}^{\infty} a_n n(n-1)\cdots(n-k+1)(z-a)^{n-k}$$

Auswerten bei $z = a$ läßt davon bloß

$$f^{(k)}(a) = a_k k(k-1)\cdots 2 \cdot 1 = a_k k!$$

übrig, und das war die Behauptung.

In der Form

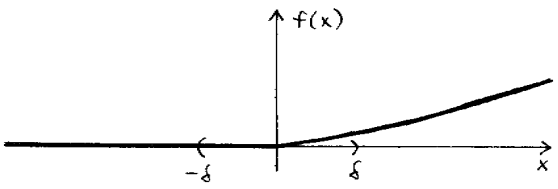
$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(z-a)^n$$

nennt man die f darstellende Potenzreihe gern die Taylor-Reihe der analytischen Funktion f um den Punkt a, und man sagt, f sei um den Punkt a in diese Taylor-Reihe "entwickelt".

Wir sehen jetzt unmittelbar, daß die Funktion f aus Beispiel 15.12(4) nicht analytisch sein kann. Als Potenzreihe um 0, die f in einem nicht-leeren Intervall $(-\delta, \delta)$ darstellt, kommt nach Lemma 16.1 ja nur die Taylor-Reihe, also

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} z^n$$

in Frage. Aber weil die Ableitungen $f^{(n)}(0)$ alle verschwinden, handelt es sich bei dieser Reihe einfach um die Nullreihe: das ist unvereinbar mit der Tatsache, daß $f(x) > 0$ für alle $x > 0$ ist.



Daß diese Funktion f nicht analytisch ist, zeigt in frappanter Weise auch der folgende schöne und überraschende Satz.

16.2 Identitätssatz Es sei $X \subset \mathbb{R}$ ein Intervall oder $X \subset \mathbb{C}$ ein Gebiet. $f, g\colon X \longrightarrow \mathbb{C}$ seien zwei analytische Funktionen. Wenn es ein $a \in X$ derart gibt, daß f und g bei a die gleiche Taylor-Reihe haben, dann gilt überhaupt $f = g$ (in ganz X!).

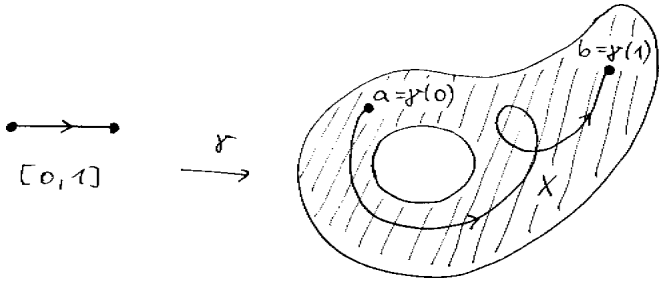
Beweis (nicht vorgetragen) Statt f und g können wir $h := f - g$ und 0 betrachten. Die Voraussetzung ist dann

$$h^{(k)}(a) = 0 \quad \text{für jedes } k \in \mathbb{N},$$

und zeigen wollen wir, daß $h = 0$ die Nullfunktion ist.

Wir nehmen im Gegenteil an, es gebe ein $b \in X$ mit $h(b) \neq 0$. Weil X ein Gebiet (oder ein Intervall) ist, gibt es in X einen Weg γ von a nach b, also eine stetige Funktion

$$\gamma\colon [0,1] \longrightarrow X \quad \text{mit } \gamma(0) = a,\ \gamma(1) = b.$$



Die Menge

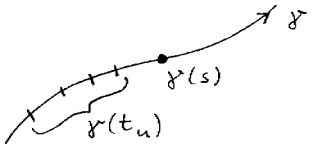
$$T := \Big\{ t \in [0,1] \,\Big|\, h^{(k)}\big(\gamma(t)\big) = 0 \text{ für alle } k \in \mathbb{N} \Big\}$$

ist nicht-leer ($0 \in T$), besitzt also ein Supremum

$$s := \sup T \in [0,1].$$

Wir wählen nach dem Satz über das Infimum (4.12) eine (monoton wachsende) Folge $(t_n)_{n=0}^{\infty}$ in T mit $\lim t_n = s$. Weil γ stetig ist, gilt

$$\lim_{n\to\infty} \gamma(t_n) = \gamma(s),$$



und weil jedes $h^{(k)}$ stetig ist, folgt

$$0 = \lim_{n\to\infty} h^{(k)}\big(\gamma(t_n)\big) = h^{(k)}\big(\gamma(s)\big) \quad \text{für jedes } k \in \mathbb{N},$$

d.h. $s \in T$. Insbesondere — wegen $h^{(0)}\big(\gamma(1)\big) = h(b) \neq 0$ — muß $s \in [0,1)$ sein.

Nun ist h analytisch. Es gibt also ein $\varepsilon > 0$, so daß h auf $U := X \cap U_\varepsilon\big(\gamma(s)\big)$ durch die Taylor-Reihe um $\gamma(s)$ dargestellt wird, das heißt aber, daß dort

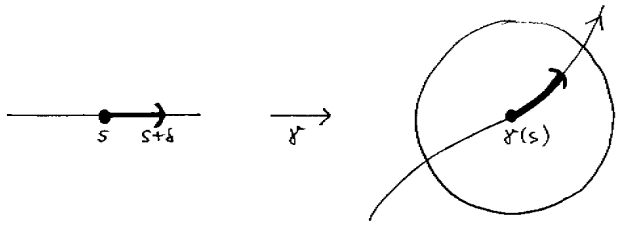
$$h(z) = \sum_{k=0}^{\infty} \frac{h^{(k)}\big(\gamma(s)\big)}{k!} \big(z - \gamma(s)\big)^k = 0$$

ist. Dann ist aber auch

$$h^{(k)}(z) = 0 \quad \text{für jedes } k \in \mathbb{N} \text{ und alle } z \in U.$$

Wegen der Stetigkeit von γ bei s finden wir schließlich ein $\delta > 0$ mit

$$[s, s+\delta) \subset [0,1) \quad \text{(beachte } s < 1\text{)} \quad \text{und} \quad \gamma(t) \in U \quad \text{für jedes } t \in [s, s+\delta).$$



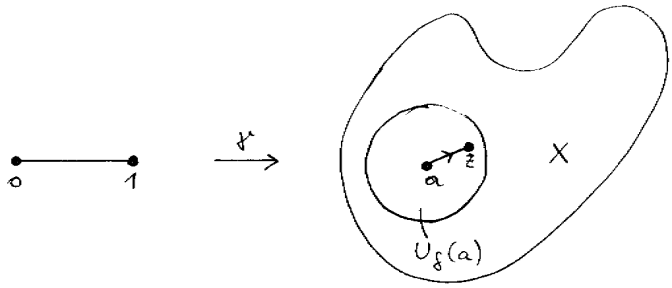
Für diese t gilt dann $h^{(k)}\big(\gamma(t)\big) = 0$ für jedes $k \in \mathbb{N}$; es ist also $t \in T$, in offensichtlichem Widerspruch zur Definition des Supremums s. Damit ist der Satz bewiesen.

Jetzt erkennen wir leicht, daß sich die Aussage der Notiz 15.9 (Eindeutigkeit von Stammfunktionen) auf den komplexen Fall überträgt:

16.3 Folgerung Ist $X \subset \mathbb{C}$ ein Gebiet und sind F und G zwei Stammfunktionen von $f\colon X \longrightarrow \mathbb{C}$, so ist $F - G$ eine konstante Funktion.

Beweis Sei $a \in X$, und sei $\delta > 0$ so klein, daß $U_\delta(a) \subset X$ ist. Für beliebiges $z \in U_\delta(a)$ ist dann

$$[0,1] \ni t \overset{\gamma}{\mapsto} (1-t)a + tz \in X$$



ein differenzierbarer Weg in X, und nach der Kettenregel hat die Komposition $(F-G)\circ\gamma$ identisch verschwindende Ableitung:

$$\frac{d}{dt}\big(F-G\big)\big(\gamma(t)\big) = \big(F-G\big)'\big(\gamma(t)\big)\cdot\gamma'(t) = 0$$

Nach 15.9 ist diese Komposition deshalb konstant, insbesondere

$$(F-G)(a) = \big((F-G)\circ\gamma\big)(0) = \big((F-G)\circ\gamma\big)(1) = (F-G)(z).$$

Die Funktion $F-G$ ist also auf einer offenen Kreisscheibe um a konstant, und weil $a \in X$ beliebig war, ist $F-G$ damit sicher eine analytische Funktion auf X. Nach dem Identitätssatz 16.2 kann es sich nur um eine konstante Funktion handeln.

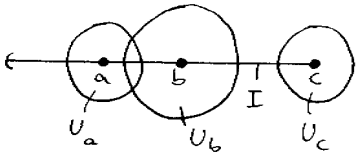
Jetzt überträgt sich auch der Beweis der Folgerung 15.11 wörtlich, und deshalb ist jede Stammfunktion einer auf einem Gebiet X erklärten analytischen Funktion selbst analytisch. Anders aber als im Reellen braucht eine auf X erklärte analytische Funktion gar keine Stammfunktion zu besitzen; das hängt damit zusammen, daß die Geometrie von Gebieten in der Ebene komplizierter sein kann als die von Intervallen, die ja völlig trivial ist. Zum Beispiel hat, wie wir später sehen werden, schon die einfache Funktion $\mathbb{C}\setminus\{0\} \ni z \mapsto \dfrac{1}{z} \in \mathbb{C}$ keine Stammfunktion.

Zu den bemerkenswerten Konsequenzen des Identitätssatzes zählt die Tatsache, daß "reell-analytisch" sich nun nicht als ein Analogon, sondern als Spezialfall von "komplex-analytisch" erweist. Man kann eine auf einem

Intervall gegebene analytische Funktion nämlich immer zu einer auf einem Gebiet analytischen Funktion (mit komplexen Werten) erweitern:

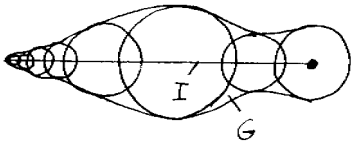
16.4 Satz Sei $I \subset \mathbb{R}$ ein Intervall, und sei $f\colon I \longrightarrow \mathbb{C}$ eine analytische Funktion. Dann gibt es ein Gebiet $G \subset \mathbb{C}$ mit $I \subset G \cap \mathbb{R}$ und eine analytische Funktion $F\colon G \longrightarrow \mathbb{C}$ mit $F|I = f$.

Beweis Zu jedem $a \in I$ wählen wir eine Kreisscheibe $U_a = U_\delta(a)$ und eine dort konvergente Potenzreihe F_a um a, deren Reihensumme auf $U_a \cap I$ mit f übereinstimmt.



Sind $a, b \in I$ Punkte derart, daß U_a und U_b sich treffen, so enthält $U_a \cap U_b$ einen Punkt von I, und nach dem Identitätssatz stimmen F_a und F_b auf dem Gebiet $U_a \cap U_b$ überein. Man sieht nun leicht, daß

$$G := \bigcup_{a \in I} U_a \subset \mathbb{C}$$



ein Gebiet ist, und nach dem eben Gesagten fügen sich die Funktionen $F_a\colon U_a \longrightarrow \mathbb{C}$ vermöge

$$F(z) := F_a(z) \quad \text{falls } z \in U_a$$

zu einer wohldefinierten Funktion $F\colon G \longrightarrow \mathbb{C}$ zusammen. Diese Funktion ist offenbar analytisch, und ihre Einschränkung auf I ist natürlich f.

Auf den Identitätssatz kann man sich auch berufen, wenn man die Gültigkeit mancher Identitäten zwischen komplex-analytischen Funktionen der Bequemlichkeit halber nur für reelle Argumente bewiesen hat. Das habe ich zum Beispiel bei den Additionstheoremen 12.8 gemacht. Die Formel

$$\cos(x + y) = \cos x \, \cos y - \sin x \, \sin y,$$

bewiesen für alle $x, y \in \mathbb{R}$, gilt in Wirklichkeit für beliebige *komplexe* x und y. Um das einzusehen, halten wir erst mal $x \in \mathbb{R}$ fest, dann stehen auf beiden Seiten der Gleichung analytische Funktionen der komplexen Variablen y. Weil diese Funktionen auf $\mathbb{R}$ übereinstimmen, haben sie sicher bei 0 dieselbe Taylor-Reihe, also sind sie nach dem Identitätssatz überhaupt gleich. Damit ist die Formel für $x \in \mathbb{R}$ und $y \in \mathbb{C}$ bewiesen. Dasselbe Argument mit festem $y \in \mathbb{C}$ und variablem x liefert dann die Gültigkeit für beliebige $x, y \in \mathbb{C}$.

Das praktische Problem, die Taylor-Reihe einer analytischen Funktion f um einen Punkt a in "geschlossener Form" zu berechnen, wird nicht immer lösbar sein. Daß wir da andererseits nicht ganz ohne Werkzeug dastehen, sollen die folgenden Beispiele illustrieren.

16.5 Beispiele (1) Aus Beispiel 15.2(1) kennen wir schon die Reihe

$$\frac{1}{1+z^2} = \sum_{n=0}^{\infty} (-1)^n z^{2n}$$

vom Konvergenzradius 1. Durch gliedweises Integrieren erhalten wir mit

$$\arctan z = \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} z^{2n+1} = z - \frac{z^3}{3} + \frac{z^5}{5} - \frac{z^7}{7} + \cdots$$

die Taylor-Reihe um 0 der Arcustangensfunktion, ebenfalls vom Konvergenzradius 1. Beachten Sie, daß die Folgerung 16.3 das Gleichheitszeichen nur bis auf Addition einer komplexen Konstanten liefert, die sich hier aber als null erweist. Ganz nebenbei erhält durch diese Taylor-Reihe $\arctan z$ auch für komplexe $z \in U_1(0)$ einen Sinn.

(2) Statt der Taylor-Reihe von $\log\colon (0,\infty) \longrightarrow \mathbb{R}$ um den Punkt 1 betrachten wir die von $x \mapsto \log(1+x)$ um 0, die ja offenbar dieselben Koeffizienten hat. Wir gewinnen diese Reihe aus

$$\frac{1}{1+z} = \sum_{n=0}^{\infty} (-1)^n z^n \quad (|z| < 1)$$

durch gliedweise Integration (wegen $\log 1 = 0$ ist die Integrationskonstante wieder null):

$$\log(1+z) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n+1} z^{n+1} = \sum_{m=1}^{\infty} \frac{(-1)^{m-1}}{m} z^m = z - \frac{z^2}{2} + \frac{z^3}{3} - \frac{z^4}{4} + \cdots \quad (|z| < 1)$$

Auch hier erhält $\log z$ automatisch für gewisse komplexe z einen Sinn.

(3) Bei $f(z) = (1+z)^b$ (für $b \in \mathbb{C}$) läßt sich die Taylor-Reihe um 0 direkt nach der Formel 16.1 berechnen. Offenbar ist

$$f^{(n)}(z) = b(b-1)\cdots(b-n+1)(1+z)^{b-n},$$

also $f^{(n)}(0) = b(b-1)\cdots(b-n+1)$, und damit wird

$$\frac{f^{(n)}(0)}{n!} = \frac{b(b-1)\cdots(b-n+1)}{n!} = \binom{b}{n}$$

ein verallgemeinerter Binomialkoeffizient. Die Potenzreihe

$$(1+z)^b = \sum_{n=0}^{\infty} \binom{b}{n} z^n \quad (b \in \mathbb{C})$$

heißt binomische Reihe. Für $b \in \mathbb{N}$ reduziert sie sich auf die binomische Formel; sie hat dann natürlich den Konvergenzradius ∞. Sonst ist sie aber eine "richtige" Reihe mit Konvergenzradius 1, wie wir gleich sehen werden. Beispielhaft hervorgehoben sei zuerst der Fall $b = -1$: Hier ist

$$\binom{-1}{n} = \frac{(-1)(-2)\cdots(-n)}{n!} = (-1)^n,$$

und die binomische Reihe geht in die geometrische

$$\frac{1}{1+z} = \sum_{n=0}^{\infty} (-1)^n z^n$$

über, wie es ja auch sein muß. Interessanter ist $b = \frac{1}{2}$:

$$\begin{aligned}
\binom{1/2}{n} &= \frac{\frac{1}{2}\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\left(-\frac{5}{2}\right)\cdots\left(-\frac{2n-3}{2}\right)}{n!} \\
&= \frac{(-1)^{n-1}}{2^n} \cdot \frac{1\cdot 3\cdot 5\cdots(2n-3)}{n!} \\
&= \frac{(-1)^{n-1}}{2^n} \cdot \frac{1\cdot 2\cdot 3\cdot 4\cdot 5\cdots(2n-4)\cdot(2n-3)}{2^{n-2}(n-2)!n!} \\
&= \frac{(-1)^{n-1}}{2^{2n-2}} \cdot \frac{(2n-3)!}{(n-2)!n!} \quad \text{für } n \geq 2,
\end{aligned}$$

damit ergibt sich

$$\sqrt{1+z} = 1 + \frac{1}{2}z + \sum_{n=2}^{\infty} \frac{(-1)^{n-1}}{2^{2n-2}} \cdot \frac{(2n-3)!}{(n-2)!n!} \cdot z^n = 1 + \frac{1}{2}z - \frac{1}{8}z^2 + \frac{1}{16}z^3 - \frac{5}{128}z^4 + \cdots$$

für die Wurzelreihe.

(4) Die Taylor-Reihe einer rationalen Funktion um einen beliebigen Punkt a ihres Definitionsbereiches läßt sich systematisch berechnen. Wir haben im Abschnitt 10 ja schon gesehen, wie man die Untersuchung einer solchen Funktion mittels Polynomdivision (mit Rest) und Partialbruchzerlegung (Satz 10.12) auf den Fall einer Funktion der Form

$$z \mapsto z^n \quad (n \in \mathbb{N})$$

einerseits (für den Polynomanteil) oder der Form

$$z \mapsto \frac{1}{(z-c)^n}$$

mit $0 < n \in \mathbb{N}$ und $a \neq c \in \mathbb{C}$ andererseits (für die Terme der Partialbruchzerlegung) reduzieren kann.

Für erstere liefert die binomische Formel

$$z^n = \big(a + (z-a)\big)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k}(z-a)^k$$

eine abbrechende Taylor-Reihe. Den gebrochenen Term entwickeln wir für $n = 1$ in eine geometrische Reihe

$$\frac{1}{z-c} = \frac{1}{(a-c)+(z-a)} = \frac{1}{a-c} \cdot \frac{1}{1-\frac{z-a}{c-a}} = \frac{1}{a-c} \cdot \sum_{k=0}^{\infty} \frac{1}{(c-a)^k}(z-a)^k,$$

die für $z \in U_{|c-a|}(a)$ konvergiert. Die Taylor-Entwicklung von $z \mapsto \dfrac{1}{(z-c)^n}$ für $n > 1$ gewinnt man daraus, indem man $(n-1)$-mal differenziert.

In dem schon früher betrachteten Beispiel 10.13 etwa ergibt sich als Taylor-Reihe um 0

$$\begin{aligned}
\frac{z^2+z+1}{(z-1)^2(z+2)} &= \frac{2}{3} \cdot \frac{1}{z-1} + \frac{1}{(z-1)^2} + \frac{1}{3} \cdot \frac{1}{z+2} \\
&= -\frac{2}{3} \cdot \frac{1}{1-z} + \frac{d}{dz}\frac{1}{1-z} + \frac{1}{3} \cdot \frac{1}{z+2} \\
&= -\frac{2}{3} \cdot \sum_{k=0}^{\infty} z^k + \sum_{l=1}^{\infty} l z^{l-1} + \frac{1}{3} \cdot \frac{1}{2} \cdot \sum_{k=0}^{\infty} \frac{1}{(-2)^k} z^k \\
&= \sum_{k=0}^{\infty} \left(-\frac{2}{3} + (k+1) + \frac{1}{3}\frac{(-1)^k}{2^{k+1}} \right) z^k \\
&= \sum_{k=0}^{\infty} \left(\frac{(-1)^k}{2^{k+1} \cdot 3} + k + \frac{1}{3} \right) z^k.
\end{aligned}$$

Die nächste praktische Aufgabe besteht darin, den Konvergenzradius der so gewonnenen Potenzreihen zu bestimmen. Häufig liefert die zur Berechnung der Reihe verwendete Methode den Konvergenzradius gleich mit, zum Beispiel sicher in den Beispielen (1) und (2). Wenn nicht, kann man die Konvergenz mit den Methoden aus Abschnitt 5 untersuchen, insbesondere probieren, ob das Quotientenkriterium greift, am einfachsten in der Formulierung von Aufgabe 11.5: in diesem Fall ergibt sich der Konvergenzradius von $\sum a_k(z-a)^k$ direkt zu

$$\lim_{k\to\infty} \frac{|a_k|}{|a_{k+1}|} \in [0, \infty].$$

Wenn man aber von einer bekannten analytischen Funktion (und nicht einer Potenzreihe) ausgeht, führt der folgende Satz oft ganz mühelos zum Ziel. Er gehört wie der Fundamentalsatz der Algebra zu den Resultaten, die mit den typischen, uns jetzt nicht verfügbaren Methoden der *komplexen* Analysis zu beweisen sind; wegen seiner Nützlichkeit sei er hier zitiert.

16.6 Satz Es seien $a \in \mathbb{C}$ und $r > 0$, sowie

$$f\colon U_r(a) \longrightarrow \mathbb{C}$$

eine analytische Funktion. Dann hat die Taylor-Reihe von f um den Punkt a mindestens den Konvergenzradius r.

Das Verblüffende ist, daß man mit diesem Satz etwas über die Konvergenz von Potenzreihen erfahren kann, die man explizit vielleicht gar nicht kennt! Wie einfach die Anwendung ist, zeigen die

16.7 Beispiele (1) $z \mapsto h(z) = \dfrac{f(z)}{g(z)}$ sei eine rationale Funktion, natürlicherweise definiert auf dem Gebiet

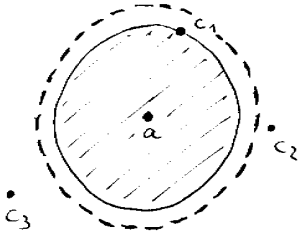
$$G = \mathbb{C} \setminus \{c_1, \dots, c_s\},$$

wobei $c_1, \dots, c_s$ die Nullstellen des Nennerpolynoms g sind. Wir nehmen darüber hinaus an, daß $f(c_j) \neq 0$ für $j = 1, \dots, s$ ist, die c_j also die Polstellen von h sind. Insbesondere gilt dann

$$\lim_{z \to c_j} |h(z)| = \infty \quad \text{für alle } j.$$

Ist nun $a \in G$ beliebig, so folgt aus Satz 16.6, daß die Taylor-Reihe von h um a mindestens in der größten offenen Kreisscheibe um a konvergiert, die keine der Polstellen c_j enthält. Mit anderen Worten ist der Konvergenzradius mindestens

$$\min\left\{|a - c_j| \,\middle|\, 1 \leq j \leq s\right\}.$$



Größer kann er aber nicht sein, sonst würde die Taylor-Reihe die Funktion h auf ein eine Polstelle enthaltendes Gebiet analytisch, insbesondere stetig fortsetzen, was wegen $\lim\limits_{z \to c_j} |h(z)| = \infty$ nicht möglich ist. Also ist die angegebene Zahl der genaue Konvergenzradius.

(2) Speziell lesen wir aus (1) die (uns schon bekannte) Tatsache ab, daß die Reihe

$$\frac{1}{1+z^2} = \sum_{n=0}^{\infty} (-1)^n z^{2n}$$

den Konvergenzradius 1 hat. Aber erst die Existenz der beiden nicht-reellen Polstellen $\pm i$ macht das auch verständlich! Vom rein reellen Standpunkt gesehen ist es ja verwunderlich, daß die Taylor-Reihe der auf ganz $\mathbb{R}$ erklärten und analytischen Funktion $x \mapsto \dfrac{1}{1+x^2}$ einen so kleinen Konvergenzradius hat.

(3) Die durch die binomische Reihe dargestellte Funktion

$$z \mapsto e^{b \cdot \log(1+z)} = (1+z)^b = \sum_{n=0}^{\infty} \binom{b}{n} z^n$$

ist auf $U_1(0)$ analytisch, weil die Logarithmusreihe dort konvergiert. Der Konvergenzradius der binomischen Reihe ist nach Satz 16.6 also mindestens 1. Außer wenn $b \in \mathbb{N}$ ist und die Reihe abbricht, ist er genau 1: Für reelle $b < 0$ folgt das aus

$$\lim_{x \searrow -1} (1+x)^b = \infty,$$

für $b > 0$ (aber $b \notin \mathbb{N}$) etwas aufwendiger aus

$$\lim_{x \searrow -1} \left(\frac{d}{dx}\right)^{\lceil b \rceil} (1+x)^b = \infty;$$

beides zeigt ja, daß $z \mapsto (1+z)^b$ nicht analytisch auf ein den Punkt -1 enthaltendes Gebiet fortgesetzt werden kann.

Warum ist es nun nützlich, wenn man Taylor-Reihen berechnen kann? Grob gesagt deshalb, weil man aus der Taylor-Reihe einer analytischen Funktion an einer Stelle die lokalen Eigenschaften der Funktion in der Umgebung dieser Stelle bequem ablesen kann. Sei $G \subset \mathbb{C}$ ein Gebiet, $a \in G$ und $f\colon G \longrightarrow \mathbb{C}$ analytisch. In einer Kreisscheibe $U_\delta(a)$ um a wird f dann durch seine Taylor-Reihe dargestellt:

$$f(z) = \sum_{k=0}^{\infty} a_k (z-a)^k.$$

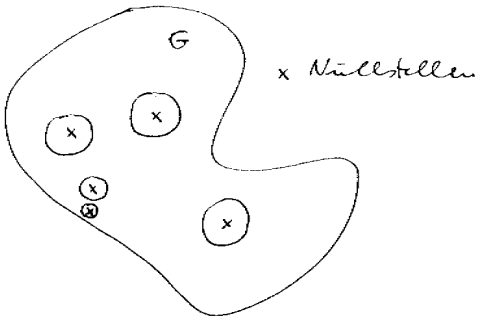
Wenn $a_k = 0$ für alle $k \in \mathbb{N}$ ist, dann ist f die Nullfunktion und damit alles gesagt. Wenn nicht, dann schreiben wir

$$\begin{aligned} f(z) &= \sum_{k=n}^{\infty} a_k (z-a)^k \quad \text{mit } a_n \neq 0 \\ &= (z-a)^n \sum_{k=n}^{\infty} a_k (z-a)^{k-n} \\ &= (z-a)^n \cdot h(z), \end{aligned}$$

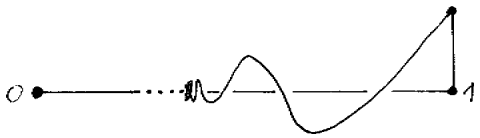
worin $h\colon U_\delta(a) \longrightarrow \mathbb{C}$ wieder eine analytische, insbesondere stetige Funktion mit $h(a) \neq 0$ ist. Wählen wir den Radius $\delta > 0$ klein genug, so ist sogar $h(z) \neq 0$ für alle $z \in U_\delta(a)$. Man sagt in dieser Situation, daß f bei a eine Nullstelle der *Ordnung* n hat; das verallgemeinert den schon bei den Polynomen eingeführten Begriff. (Wie dort ist eine Nullstelle der Ordnung null eben eine Nichtnullstelle.) Unabhängig von dieser Ordnung n sehen wir jedenfalls:

Es gibt ein $\delta > 0$ mit $f(z) \neq 0$ für alle $z \in G$ mit $0 < |z-a| < \delta$.

Das bedeutet, daß die Nullstellen einer analytischen Funktion (mit Ausnahme der Nullfunktion) sich im Definitionsgebiet G nicht "häufen" können, genauer: Eine Folge paarweise verschiedener Nullstellen kann keinen Limes in G besitzen.

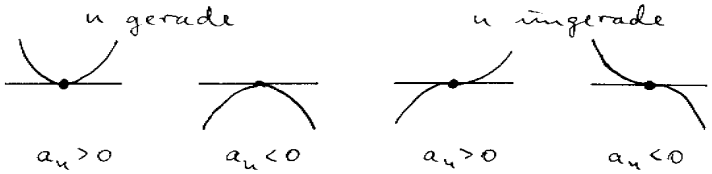


Diese Aussage gilt natürlich wörtlich auch für auf einem Intervall definierte analytische Funktionen. Insofern entsprechen solche Funktionen viel mehr der naiven Anschauung als bloß stetige, selbst C^∞-Funktionen, die — wie schon im Abschnitt 14 erwähnt — ganz "wilde" Nullstellenmengen haben können.



Wenn wir uns jetzt auf den reellen Fall, sagen wir den einer auf einem offenen Intervall I erklärten analytischen Funktion $f\colon I \longrightarrow \mathbb{R}$ beschränken, lesen wir gleich weiter ab:

> Ist die Nullstellenordnung n positiv und gerade, so hat f bei a ein strenges lokales Minimum oder Maximum, je nachdem, ob $a_n > 0$ oder $a_n < 0$ ist. Ist n ungerade, so hat f bei a kein lokales Extremum (ist vielmehr in einem offenen Intervall um a streng monoton).



Für analytische Funktionen haben wir damit ein notwendiges und hinreichendes Kriterium dafür gewonnen, ob bei a ein lokales Extremum vorliegt. Zwar läßt sich der Hauptteil der Aussage auch für C^∞-Funktionen retten, aber das läßt die Frage für all diejenigen Funktionen offen, deren sämtliche Ableitungen bei a verschwinden. Daß es tatsächlich nicht-triviale Funktionen mit dieser Eigenschaft gibt, belegt Beispiel 15.12(4).

Natürlich gibt der Begriff der Taylor-Reihe nicht nur für analytische Funktionen einen Sinn, und allgemeiner vereinbart man:

16.8 Definition Sei $I \subset \mathbb{R}$ ein echtes Intervall und $a \in I$. Für eine C^k-Funktion $f\colon I \longrightarrow \mathbb{C}$ heißt das Polynom $T_a^k f$ mit

$$\mathbb{R} \ni x \mapsto T_a^k f(x) := \sum_{j=0}^{k} \frac{f^{(j)}(a)}{j!}(x-a)^j \in \mathbb{C}$$

das k-te Taylor-Polynom von f an der Stelle a. Im Fall $k = \infty$ spricht man wie im analytischen Fall von der Taylor-Reihe:

$$T_a^\infty f(x) := \sum_{j=0}^{\infty} \frac{f^{(j)}(a)}{j!}(x-a)^j$$

(Üblicherweise setzt man in $T_a^k f(x)$ und $T_a f(x)$ keine weitere Klammer; natürlich ist $(T_a^k f)(x)$ und $(T_a f)(x)$ zu lesen.) Statt $T_a^\infty f$ schreibt man auch einfach $T_a f$, wenn keine Verwechslung mit dem später in 37.2 einzuführenden Differential einer Abbildung zu befürchten ist.

Im Gegensatz zur Taylor-Reihe einer analytischen Funktion ist die Taylor-Reihe einer C^∞-Funktion eine Potenzreihe, über deren Konvergenzradius a priori nichts gesagt werden kann; man kann sogar zeigen, daß *jede* Potenzreihe als Taylor-Reihe einer C^∞-Funktion auftritt. Aber selbst dort, wo die Taylor-Reihe konvergiert, braucht ihre Summe nicht mit der Ausgangsfunktion übereinzustimmen (abgesehen davon, daß natürlich $T_a f(a) = f(a)$ ist): bei der Funktion f aus Beispiel 15.12(4) war $T_0 f$ die Nullreihe, obwohl $f(x) \neq 0$ für alle $x > 0$ ist.

Bei den Taylor-*Polynomen* ist eine solche Übereinstimmung ohnehin nicht zu erwarten (wenn nicht gerade die Ausgangsfunktion selbst ein Polynom ist). Die Berechnung der Taylor-Koeffizienten im Beweis von Lemma 16.1 liefert aber sofort eine begriffliche Charakterisierung dieser Polynome:

16.9 Notiz Sei $I \subset \mathbb{R}$ ein echtes Intervall, $a \in I$, und sei $f\colon I \longrightarrow \mathbb{C}$ eine C^k-Funktion ($k \in \mathbb{N}$). Unter allen Polynomen vom Grad höchstens k ist dann $T_a^k f$ dasjenige, das an der Stelle a dieselben Ableitungen bis zur Ordnung k hat wie f:

$$\left(\frac{d}{dx}\right)^j T_a^k f(x)\bigg|_{x=a} = \left(\frac{d}{dx}\right)^j f(x)\bigg|_{x=a} \qquad \text{für } j = 0, 1, \ldots, k.$$

Insbesondere gilt

$$T_a^j(T_a^k f) = T_a^j f \quad \text{für } j \le k.$$

Mit den Taylor-Reihen oder -Polynomen von C^∞- und C^k-Funktionen kann man nun fast genau so rechnen wie mit den Taylor-Reihen analytischer Funktionen. Daß die Anwendung von T_a und T_a^k mit Summen und der Multiplikation mit Konstanten vertauschbar ist, liegt ja auf der Hand. Darüber hinaus gilt aber auch die Produktregel

$$T_a(f\cdot g) = (T_a f)\cdot(T_a g),$$

und die Reihe $T_a(g\circ f)$ ergibt sich durch Einsetzen der Potenzreihe $T_a f$ in $T_{f(a)}g$ im Sinne von Satz 15.1. Beides bedarf freilich einer neuen Begründung, denn möglicherweise haben all diese Potenzreihen ja den Konvergenzradius null!

Im Fall der Taylor-*Polynome* ist eine geringfügige Modifikation erforderlich. Etwa hat das Produkt der beiden Taylor-Polynome $(T_a^k f)\cdot(T_a^k g)$ im allgemeinen nicht den Grad k, sondern den Grad $2k$ und damit keine Aussicht, mit $T_a^k(f\cdot g)$ übereinzustimmen. Man erhält aber eine richtige Formel, wenn man die überzähligen Terme einfach abschneidet:

16.10 Regeln Sei $I\subset\mathbb{R}$ ein echtes Intervall, $a\in I$ ein Punkt, und seien $f, g\colon I\longrightarrow\mathbb{C}$ zwei C^k-Funktionen. Dann gilt

$$T_a^k(f\cdot g) = T_a^k\left(T_a^k f\cdot T_a^k g\right) \quad \text{für jedes } k\in\mathbb{N}.$$

Ist f wie vor, $J\subset\mathbb{R}$ ein Intervall mit $f(I)\subset J$ und g diesmal eine C^k-Funktion $g\colon J\longrightarrow\mathbb{C}$, so gilt

$$T_a^k(g\circ f) = T_a^k\big(T_{f(a)}^k g\circ T_a^k f\big) \quad \text{für jedes } k\in\mathbb{N}.$$

Bemerkung Die Physiker machen das eigentlich schon immer so, indem sie mit

$$f(x) = T_a^k f(x) + \text{höhere Terme}$$

rechnen. Weil das im nicht-analytischen Fall keine sinnvolle Gleichung zwischen Funktionswerten ist, haben sie dabei manchmal ein schlechtes Gewissen. Brauchen sie aber nicht, denn diese Art des Rechnens ist völlig korrekt; wie gesagt, bedarf sie nur besonderer (aber nicht schwieriger, bloß die Notiz 16.9 systematisch ausnutzender) Beweise, für die ich einen Blick in das Buch von Bröcker empfehle. Erstaunlicherweise dringt das Rechnen mit Taylor-Polynomen und -Reihen erst in jüngster Zeit in die mathematische Lehrbuchliteratur ein (auch das sonst so ausführliche Buch von Heuser läßt einen hier im Stich). Dabei handelt es sich um die in der Regel einzige intelligente Art, mit den höheren Ableitungen zu rechnen: Die Produkt- und die Kompositionsregel für die k-ten Taylor-Polynome enthalten ja die Regeln, nach denen sich die k-te Ableitung eines Produktes oder einer Komposition aus den dazu nötigen Ableitungen der Faktoren berechnen. Zwar kann man diese Regeln auch "Taylor-frei" formulieren, aber das bringt keinen Vorteil, vielmehr im Fall der Kompositionsregel einen unüberschaubaren Wust. Immerhin wollen wir uns noch davon überzeugen, daß sich für $k=1$ tatsächlich die gewöhnliche Kettenregel ergibt:

$$\begin{aligned}
T_a^1(g\circ f)(x) &= T_a^1\big(T_{f(a)}^1 g\circ T_a^1 f\big)\\
&= T_a^1\left(g\big(f(a)\big) + \frac{g'\big(f(a)\big)}{1!}\big(y-f(a)\big)\Big|_{y=f(a)+\frac{f'(a)}{1!}(x-a)}\right)\\
&= T_a^1\Big(g\big(f(a)\big) + g'\big(f(a)\big)f'(a)(x-a)\Big)\\
&= \underbrace{g\big(f(a)\big)}_{(g\circ f)(a)} + \underbrace{g'\big(f(a)\big)f'(a)}_{\frac{(g\circ f)'(a)}{1!}}(x-a)
\end{aligned}$$

(in diesem besonders einfachen Fall treten keine wegzulassenden "höheren Terme" auf).

Denken Sie nun noch einmal an die Definition der ersten Ableitung einer Funktion f zurück: Die Idee war, f nahe der Stelle a durch die lineare Funktion

$$T_a^1 f\colon x\mapsto f(a) + f'(x)\cdot(x-a)$$

zu approximieren, und nach der Notiz 13.1$\frac{1}{2}$ ist diese Approximation bestmöglich im Sinne von

$$f(x) = T_a^1 f(x) + o(|x-a|) \quad \text{für } x \to a.$$

Für größere $k \in \mathbb{N}$ ist nun $T_a^k f$ im allgemeinen ein Polynom vom Grad k, also komplizierter als $T_a^1 f$, und wie man hoffen darf, wohl auch eine bessere Approximation von f. Daß das ist tatsächlich so ist, sieht man an der folgenden, an die Notiz 16.9 erinnernden Charakterisierung der Taylor-Polynome:

16.11 Satz Sei $I \subset \mathbb{R}$ ein echtes Intervall, $a \in I$ ein Punkt, und sei $f\colon I \longrightarrow \mathbb{C}$ eine C^k-Funktion ($k \in \mathbb{N}$). Unter allen Polynomen vom Grad höchstens k hat $T_a^k f$ als einziges die Eigenschaft

$$f(x) = T_a^k f(x) + o(|x-a|^k) \quad \text{für } x \to a.$$

Beweis Der interessanteste Teil der Aussage, nämlich daß das Taylor-Polynom diese Eigenschaft besitzt, ist das Hauptergebnis der Aufgabe 14.10, denn nach der Notiz 16.9 hat die Differenz $f(x) - T_a^k f(x)$ ein verschwindendes k-tes Taylor-Polynom bei a.

Sei andererseits $p\colon \mathbb{R} \longrightarrow \mathbb{R}$ ein weiteres Polynom mit $\deg p \le k$ und

$$f(x) = p(x) + o(|x-a|^k) \quad \text{für } x \to a;$$

dann folgt auch $p(x) - T_a^k f(x) = o(|x-a|^k)$. Wir wollen zeigen, daß $p = T_a^k f$ ist, und nehmen dazu an, die Differenz $p - T_a^k f$ sei nicht das Nullpolynom, habe bei a etwa eine Nullstelle der Ordnung e mit $0 \le e \le k$. Wir können dann

$$p(x) - T_a^k f(x) = (x-a)^e \cdot q(x)$$

schreiben, worin q ein Polynom mit $q(a) \ne 0$ ist, und lesen aus $(x-a)^e \cdot q(x) = o(|x-a|^k)$ sofort den Widerspruch

$$q(x) = o(|x-a|^{k-e}) \quad \text{für } x \to a$$

heraus.

Die in der Experimentalphysik geläufige sogenannte Fehlerrechnung ist nichts Anderes als das Rechnen mit dem ersten Taylor-Polynom: Man setzt dort einen Meßwert a in eine "Formel", nämlich eine C^1-Funktion f ein und möchte wissen, wie sich ein in a enthaltener Meßfehler von a auf den Wert $f(a)$ auswirkt. Schreibt man den wahren physikalischen Wert für a als $a+h$, so ist natürlich $f(a+h)$ statt $f(a)$ der wahre Wert von f. Das nützt einem aber wenig, weil man h selbst nicht kennt, sondern nur die Größenordnung von h. Anstatt nun die Werte $f(a+h)$ für alle in Frage kommenden h zu berechnen oder abzuschätzen, bestimmt man nur $T_a^1 f$, d.h. zusätzlich zu $f(a)$ noch $f'(a)$. Satz 16.11 garantiert dann immerhin $f(a+h) = T_a^1 f(h) + o(h)$ für $h \to 0$ oder — anders geschrieben

$$f(a+h) - f(a) = f'(a) \cdot h + o(h) \quad \text{für } h \to 0.$$

Wenn man nun den Eingangsfehler h kontrollieren kann, also eine Schranke c mit $|h| \le c$ kennt, kann man dann den in $f(a)$ enthaltenen unbekannten Fehler durch

$$|f(a+h) - f(a)| \le |f'(a)| \cdot c$$

abschätzen, wie man das in der Fehlerrechnung tut? Natürlich nicht, denn zwar dominiert der Term $f'(a) \cdot h$ das $o(h)$ für kleine $|h|$ (jedenfalls solange $f'(a) \ne 0$ ist), aber man weiß nicht, für *wie kleine* $|h|$, während hier h ja eine ganz bestimmte, wenn auch unbekannte Zahl ist. Die Fehlerrechnung in dieser Form liefert deshalb nicht, wie der Name glauben machen will, eine wirkliche Abschätzung oder gar Berechnung des Fehlers, sie macht nur eine Aussage über die Größenordnung des (bei genauer werdenden Eingangsdaten, also kleiner werdendem $|h|$) zu erwartenden Fehlers. Das für sich genommen ist schon eine sehr nützliche Information; wenn man das Bedürfnis nach exakten Abschätzungen hätte, wäre sie aber erst der Ausgangspunkt, um etwa mit dem Mittelwertsatz argumentieren (siehe die Bemerkung (2) im Anschluß an den Satz von Rolle 14.2).

Bemerkungen Die Situation ändert sich nicht grundsätzlich, wenn man eine "höhere" Fehlerrechnung auf der Basis der zweiten, dritten... Taylorpolynome, ja gar der ganzen Taylor-Reihe anpeilt: Da der Limes, hier

in Gestalt des $o(h^k)$, für ein festes $h \neq 0$ nun einmal gar nichts aussagt, kann es sogar daurchaus vorkommen, daß der in zweiter und höherer Ordnung berechnete Fehler den tatsächlichen Fehler *schlechter* annähert als der in erster Ordnung bestimmte. Fehlerrechnung höherer Ordnung in dieser groben Art durchzuführen gibt eigentlich nur dann Sinn, wenn die Rechnung in erster Ordnung den Fehler null und damit überhaupt keine Vorhersage liefert. — Wie Ihnen aus der Praxis natürlich vertraut ist, hat man es in aller Regel mit Funktionen nicht eines, sondern mehrerer fehlerbehafteter Meßwerte zu tun, was die direkte Abschätzung des Fehlers sehr viel schwieriger macht. Dagegen läßt sich die Methode der Fehlerrechnung ohne weiteres auf diesen Fall ausdehnen; wenn wir im Sommersemester über Differentialrechnung und Taylor-Polynome mehrerer Veränderlicher reden, werden Sie das sofort erkennen.

Die Taylor-Entwicklungen einer (nicht notwendig analytischen) C^∞-Funktion kann man als Spezialfall sogenannter asymptotischer Entwicklungen auffassen. Diese spielen auch in der Physik eine Rolle; von den vielen möglichen Versionen sei hier nur eine erwähnt:

16.12 Definition Gegeben seien ein eigentliches oder uneigentliches echtes Intervall $I \subset [-\infty, \infty]$, ein Punkt $a \in I$, so daß also $I' := I \setminus \{a\}$ nicht-leer ist, außerdem eine Folge $(g_k)_{k=0}^\infty$ von Funktionen $g_k \colon I' \longrightarrow \mathbb{C}$, derart daß für jedes $k \in \mathbb{N}$

$$g_{k+1} = o(g_k) \quad \text{für } x \to a$$

gilt. Unter einer asymptotischen Entwicklung einer Funktion $f \colon I' \longrightarrow \mathbb{C}$ (nach der Folge $(g_k)_k$ und um den Punkt a) versteht man eine Funktionenreihe der Form

$$\sum_{k=0}^{\infty} \lambda_k g_k$$

mit konstanten Koeffizienten $\lambda_k \in \mathbb{C}$, die für jedes $n \in \mathbb{N}$ die Eigenschaft

$$f = \sum_{k=0}^{n} \lambda_k g_k + o(g_n) \quad \text{für } x \to a$$

hat.

Wie Notiz 16.9 zeigt, handelt es sich bei der Taylor-Reihe einer C^∞-Funktion f um den Spezialfall

$$g_k(x) = (x-a)^k \quad \text{und} \quad \lambda_k = \frac{f^{(k)}(a)}{k!}.$$

Als g_k kommen aber auch andere Funktionen in Frage: Zum Beispiel wenn $a \in \mathbb{R}$ der linke Randpunkt von I und damit $x > a$ ist, gebrochene Potenzen

$$g_k(x) = (x-a)^{\alpha k + \beta} \quad \text{mit festen Zahlen } \alpha > 0 \text{ und } \beta \in \mathbb{R},$$

oder, im Fall $a = \infty$, negative Potenzen $g_k(x) = x^{-k}$, oder auch logarithmushaltige Funktionen vom Typ

$$x \mapsto (x-a)^k \cdot (\log x)^\gamma$$

oder Kombinationen von all dem. Natürlich hängt es von der Wahl der g_k ab, welche Funktionen f überhaupt eine solche asymptotische Darstellung zulassen und welche Eigenschaften diese Entwicklungen haben. Bemerken Sie jedenfalls, daß nirgendwo verlangt oder behauptet wird, die Reihe $\sum \lambda_k g_k$ müsse in irgendeinem Sinne konvergieren. Tatsächlich arbeitet man mit solchen asymptotischen Entwicklungen gerade dann, wenn man über die Konvergenz der auftretenden Funktionenreihe keine Aussage machen kann oder will. In der Physik hilft man sich so zum Beispiel ganz gern bei Systemen, die quantentheoretisch zu kompliziert, klassisch aber einfach zu berechnen sind: Man sieht dann die Plancksche Naturkonstante h als eine Variable an und entwickelt die vorkommenden Funktionen in ihre Taylor-Reihe bezüglich h (um 0). Das so berechnete Resultat enthält dann einerseits das der klassischen Theorie als den konstanten Term, beschreibt andererseits auf eine ziemlich subtile Art, wie das quantentheoretische Resultat für klein werdendes Wirkungsquantum in das klassische "einmündet". Für das reale Wirkungsquantum freilich hat das so erhaltene Ergebnis wie bei der Fehlerrechnung nur den Charakter einer Vorhersage (die vom Experiment oft glänzend bestätigt wird).

In ganz analoger Weise kann man bei einem relativistischen Problem die Lichtgeschwindigkeit c als variabel ansehen und mit asymptotischen Entwicklungen "um den klassischen Fall $c = \infty$" arbeiten.

Übungsaufgaben

16.1 Beweisen Sie, daß die Funktion

$$\mathbb{C}\setminus\{0\} \ni z \longmapsto \frac{1}{z} \in \mathbb{C}$$

keine Stammfunktion besitzt.

16.2 Berechnen Sie die Taylor-Reihe der Funktion $f\colon (-\infty, 1) \longrightarrow \mathbb{R}$,

$$f(x) = \frac{\log(1-x)}{1-x}$$

um den Nullpunkt, sowie die Taylor-Reihe der Funktion $x \mapsto x^b$ um einen beliebigen Punkt $a > 0$. Welchen Konvergenzradius haben diese Reihen?

16.3 Berechnen Sie das 10-te Taylorpolynom der Funktion

$$x \mapsto (\log\cos x)^4$$

an der Stelle 0.

16.4 Berechnen Sie das 4-te Taylorpolynom der Funktion

$$x \mapsto \log(1+e^x)$$

an der Stelle 0. (Dies ist mal ein Beispiel, wo die bekannten Taylor-Reihen nicht ohne weiteres ineinanderpassen.)

16.5 Berechnen Sie die Taylor-Reihe der rationalen Funktion

$$z \mapsto \frac{z}{(z+1)(z-2)}$$

um den Nullpunkt, sowie deren Konvergenzkreis.

16.6 Sei $n \in \mathbb{N}$. Beweisen Sie, daß die durch die Reihe

$$\sum_{k=0}^{\infty} k^n z^k$$

auf der Kreisscheibe $U_1(0)$ dargestellte Funktion Einschränkung einer rationalen Funktion ist (die sich im Prinzip explizit, d.h. als Quotient von Polynomen berechnen läßt; zum Beispiel für $n=2$ erhält man welche Funktion?).

16.7 Sei $a < b$, und sei $f\colon [a,b] \longrightarrow \mathbb{R}$ eine nicht-konstante analytische Funktion. Beweisen Sie, daß es ein $n \in \mathbb{N}$ und Punkte

$$a = a_0 < a_1 < \cdots < a_{j-1} < a_j < \cdots < a_n = b$$

gibt, so daß jede der Einschränkungen $f|[a_{j-1}, a_j]$ (mit $1 \le j \le n$) streng monoton ist.

16.8 Berechnen Sie die Werte der zehnten und der elften Ableitung der Funktion

$$f\colon (-\infty, 1) \longrightarrow \mathbb{R}, \quad x \mapsto \big(x + \log(1-x)\big)^5$$

16.9 Sei $I \subset \mathbb{R}$ ein echtes Intervall, und seien $f, g\colon I \longrightarrow \mathbb{R}$ zwei C^k-Funktionen. Berechnen Sie $(f \cdot g)^{(k)}$ aus den nötigen Ableitungen von f und g.

17 Vektorräume

Aus der Sicht des Mathematikers haben wir den physikalischen Raum bisher kurzerhand mit $\mathbb{R}^3$ gleichgesetzt. Das wollen wir jetzt einer genaueren gedanklichen Analyse unterziehen.

Stellen Sie sich zwei Physiker in zwei voneinander und von der Außenwelt abgeschirmten Zimmern vor. Jeder von ihnen mag für sich den Raum mit $\mathbb{R}^3$ identifizieren, aber wahrscheinlich wird jeder der beiden das in einer anderen Weise tun. Erstens kann jeder den Punkt $0 = (0, 0, 0) \in \mathbb{R}^3$ an eine willkürliche Stelle des Raumes, z.B. in den Mittelpunkt oder eine Ecke *seines* Zimmers legen. Zweitens sind die Richtungen der drei Koordinatenachsen weitgehend willkürlich. Die beiden werden sich aufgrund der Schwerkraft, die sie beide verspüren, wahrscheinlich darüber einig sein, wo oben ist; aber die beiden horizontalen Achsen werden sie im allgemeinen ganz verschieden wählen, etwa parallel zu zwei Zimmerwänden (falls die Zimmer überhaupt quaderförmig sind). Sind die Zimmer nach irdischen oder gar außerirdischen Maßstäben weit voneinander entfernt, so entfällt die Einigkeit über die senkrechte Richtung auch. Drittens werden unsere beiden Physiker, wenn sie unbefangen genug und nicht durch Kenntnis von Einheitensystemen verdorben sind, auch ganz unabhängige Maßstäbe wählen.

Der Raum *ist* also nicht $\mathbb{R}^3$, erst die Wahl von Koordinaten erlaubt es, ihn mit $\mathbb{R}^3$ zu identifizieren. Gut, wählen wir ein für allemal, oder auch von Fall zu Fall solche Koordinaten, und man hat kein Problem. Stimmt — aber ich hatte ja nicht behauptet, daß das ein Problem ist. Wahr ist aber auch, daß man mit diesem Standpunkt etwas Wesentliches verschenkt hat. Wozu möchte man denn überhaupt den Raum mit $\mathbb{R}^3$ identifizieren? Um physikalische Gesetzmäßigkeiten durch mathematische Objekte wie Gleichungen in $\mathbb{R}^3$ auszudrücken! Wenn man dem Raum zu dieser Identifizierung aber etwas Künstliches, nämlich die Koordinaten, erst hinzufügen muß, dann müssen wahre physikalische Gesetze immer dieselben sein, ganz egal wie die Koordinaten gewählt wurden. Nehmen wir als Beispiel das Newtonsche Gravitationsgesetz in der ganz einfachen Form, die bloß sagt, daß die Gravitationskraft K, die zwei Punkte der Massen m_1 und m_2 aufeinander ausüben, ihrem Betrag nach

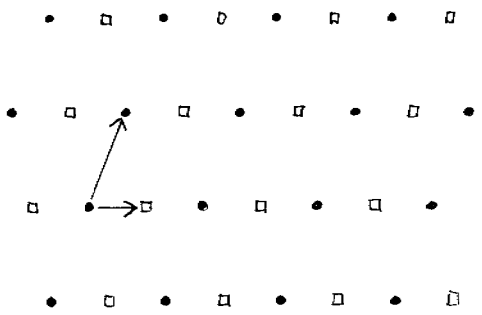
$$|K| = \gamma \frac{m_1 m_2}{|x_1 - x_2|^2}$$

ist. Darin sind $x_1 \in \mathbb{R}^3$ und $x_2 \in \mathbb{R}^3$ die Koordinatentripel oder -vektoren der beiden Massenpunkte. Die Formel nimmt also definitiv Bezug auf die gewählten Koordinaten. Aber sie tut das in einer speziellen Weise: Erstens geht nur die Differenz der beiden Tripel ein, was das Ergebnis von der Wahl des Nullpunkts unabhängig macht. Zweitens geht von dieser Differenz, die selbst wieder ein Vektor ist, nur der Betrag ein; das macht das Ergebnis auch von Koordinatendrehungen unabhängig. (Die Abhängigkeit von der Wahl der Einheiten wird dadurch kompensiert, daß die Gravitationskonstante γ keine reine Zahl ist, sondern selbst einen Einheitenfaktor enthält.) Allein diese Überlegung hätte genügt, um viele andere Ansätze für das Gravitationsgesetz wie

$$|K| = \gamma \frac{m_1 m_2}{|x_1 - x_2| \cdot |x_1 + x_2|}$$

zu verwerfen. In der Tat gehört es zum täglichen Brot des Physikers, bekannte oder vermutete physikalische Gesetzmäßigkeiten in mathematischer Formulierung auf ihr Verhalten unter Koodinatenwechsel abzuklopfen.

Bisher haben wir stillschweigend unterstellt, daß die drei Koordinatenachsen immerhin aufeinander senkrecht stehen und auf ihnen dieselbe Längeneinheit verwendet wird. So natürlich das einem vorkommen mag, es gibt Situationen, in denen das zumindest unpraktisch ist.



Das Bild zeigt einen Ausschnitt aus einem idealisierten ebenen Kristall: Ist man hier nicht am besten mit einem Koordinatensystem bedient, wie ich es eingezeichnet habe, in dem nämlich die Orte der Atome gerade diejenigen mit ganzzahligen Koordinaten sind? Natürlich muß man dann berücksichtigen, daß die *metrischen* Verhältnisse im Raum, also Winkel- und Längenmessungen sich in der Koordinatendarstellung nicht in der gewohnten, sondern in einer verzerrten Form widerspiegeln.

Eine viel radikalere Idee zeichnet sich in fortgeschrittenen Formulierungen der klassischen Mechanik ab, und in der Relativitätstheorie spielt sie dann eine ganz zentrale Rolle: Danach ist die Gesamtheit der metrischen Verhältnisse des Raumes, kurz dessen Metrik überhaupt keine Eigenschaft des Raumes selbst, sie wird ihm vielmehr erst durch die Materie, nämlich die Massen aufgeprägt, ist insbesondere von Ort zu Ort verschieden (und natürlich auch von Zeit zu Zeit; bekanntlich wird die Zeit in der Relativitätstheorie zu einer vierten, den anderen drei gleichberechtigten Ortskoordinate). Unter diesem Blickwinkel wäre es konsequent, nach einem mathematisches Modell für den Raum zu suchen, in dem die metrischen Begriffe zunächst gar nicht existieren, später aber als eine zusätzliche Struktur hinzugefügt werden können.

Es trifft sich nun gut, daß ein solches Modell zu den wichtigsten und geläufigsten mathematischen Objekten überhaupt gehört: den sogenannten Vektorräumen, um die es in diesem und den folgenden Abschnitten geht. Bevor ich zu deren formaler Definition komme, will ich gleich einen dann naheliegenden Einwand vorwegnehmen. Ebenso wie $\mathbb{R}^3$ wird jeder Vektorraum V ein ausgezeichnetes Element, den Nullvektor enthalten. Aber weil der physikalische Raum, für den V Modell stehen soll, von Natur aus keinen ausgezeichneten Punkt besitzt, kommt V nur als Modell für den Raum mit einem künstlich ausgezeichneten Punkt in Betracht; etwas, das ja gerade vermieden werden sollte. Trotzdem wollen wir diesen Schönheitsfehler vorerst einfach in Kauf nehmen, weil seine Auswirkungen leicht zu durchschauen sind und man unverhältnismäßig viel formalen Aufwand treiben müßte, um ihn von vornherein zu vermeiden. Im übrigen erheben die Vektorräume auch keineswegs den Anspruch, die "wahren" Modelle des physikalischen Raums zu sein; sie kommen diesem Ideal bloß einen Schritt näher als der Koordinatenraum $\mathbb{R}^3$ (oder $\mathbb{R}^4$).

Der Begriff des Vektorraums bezieht sich immer auf einen gegebenen Körper. Im folgenden sei stillschweigend ein solcher Körper K zugrundegelegt. Soll der Vektorraum Modell des physikalischen Raumes sein, so denkt man an $K = \mathbb{R}$; aber auch der Fall $K = \mathbb{C}$ ist in der Physik wichtig.

17.1 Definition Ein Vektorraum (genauer K-Vektorraum oder im konkreten Fall reeller bzw. komplexer Vektorraum) besteht aus

- einer Menge V,
- einer Verknüpfung $V \times V \xrightarrow{+} V$, der (Vektor)-Addition, und
- einer Abbildung $K \times V \xrightarrow{\cdot} V$, genannt skalare Multiplikation.

Diese Daten unterliegen den folgenden Axiomen:

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

(b) $\lambda(\mu x) = (\lambda\mu)x$ für alle $\lambda, \mu \in K$ und alle $x \in V$

(c) $1x = x$ für jedes $x \in V$

(d) für alle $\lambda, \mu \in K$ und alle $x, y \in V$ gelten die Distributivgesetze

$$\lambda(x+y) = \lambda x + \lambda y$$
$$(\lambda+\mu)x = \lambda x + \mu x$$

Die Elemente von V nennt man Vektoren, die Elemente von K zur Unterscheidung oft Skalare.

Bemerkungen Ein Vektorraum ist definitionsgemäß ein Tripel $(V, +, \cdot)$. Wenn klar ist, welche Addition und skalare Multiplikation gemeint sind, schreibt man bloß V hin, so wie wir das früher schon bei Gruppen und Ringen gehalten haben. — Das Nullelement der Gruppe $(V, +)$, eben den *Nullvektor*, schreibt man ebenso als 0 wie den Nullskalar; wenn dadurch in seltenen Fällen eine Verwechslungsgefahr entsteht, muß man anderweitig klarstellen, was gemeint ist. — Einige scheinbar selbstverständliche, in Wirklichkeit aus den Axiomen folgende

17.2 Regeln für das Rechnen in einem K-Vektorraum V:

(e) Für jeden Skalar $\lambda \in K$ ist $\lambda 0 = 0$, und für jeden Vektor $x \in V$ ist $0x = 0$.

(f) Umgekehrt folgt aus $\lambda x = 0$, daß $\lambda = 0$ oder $x = 0$ ist.

(g) Für jedes $x \in V$ ist $(-1)x = -x$.

Beweis (e) folgt aus

$$\lambda 0 = \lambda(0+0) = \lambda 0 + \lambda 0 \quad \text{und} \quad 0x = (0+0)x = 0x + 0x,$$

beides nach (d), durch Subtraktion von $\lambda 0$ bzw. von $0x$.

Ist $\lambda x = 0$ und $\lambda \neq 0$, so ist nach (c) und (b)

$$x = 1x = \frac{1}{\lambda}(\lambda x) = \frac{1}{\lambda} \cdot 0 = 0.$$

Schließlich folgt (g) aus

$$x + (-1)x = 1x + (-1)x = (1-1)x = 0x = 0$$

nach (c), (d) und (e).

Bevor wir uns Beispiele von Vektorräumen ansehen, ist es praktisch, noch den Begriff des Unterraums einzuführen:

17.3 Definition Sei V ein K-Vektorraum. Eine Teilmenge $T \subset U$ heißt ein Untervektorraum, linearer Teilraum, kurz auch *Unter-* oder *Teilraum* von V, wenn Addition und skalare Multiplikation von V sich zu Verknüpfungen

$$T \times T \xrightarrow{+} T \quad \text{und} \quad K \times T \xrightarrow{\cdot} T$$

einschränken lassen und diese T selbst zu einem Vektorraum machen.

Selbstverständlich sind $\{0\} \subset V$ und $V \subset V$ stets Unterräume von V. Zum Nachweis, daß eine Teilmenge $T \subset V$ ein Unterraum ist, braucht man nicht alle Axiome neu zu prüfen, wenn man sich auf das folgende einfache Lemma beruft:

17.4 Lemma Sei V ein K-Vektorraum. Eine Teilmenge $T \subset V$ ist T genau dann ein Unterraum von V, wenn sie die folgenden drei Eigenschaften hat:

- $T \neq \emptyset$
- $x, y \in T \Longrightarrow x + y \in T$
- $\lambda \in K,\ x \in T \Longrightarrow \lambda x \in T$

Insbesondere ist der Nullvektor von V zugleich der von T, und für jedes $x \in T$ ist $-x$ der in T ebenso wie in V additiv inverse Vektor.

Beweis Sei T ein Unterraum: als additive Gruppe kann T nicht leer sein, und die beiden anderen Eigenschaften sind klar.

Von der Teilmenge $T \subset V$ seien jetzt nur die drei genannten Eigenschaften vorausgesetzt. Addition und skalare Multiplikation schränken sich dann zu $T \times T \longrightarrow T$ und $K \times T \longrightarrow T$ ein. Wegen $T \neq \emptyset$ können wir ein $x \in T$ wählen, und es folgt

$$0 = 0x \in T.$$

Für jedes $x \in T$ ist aber auch

$$-x = (-1)x \in T,$$

und damit ist $(T, +)$ als Gruppe nachgewiesen. Die ist natürlich abelsch, und auch die übrigen Vektorraumaxiome gelten für T einfach deswegen, weil sie für V gelten.

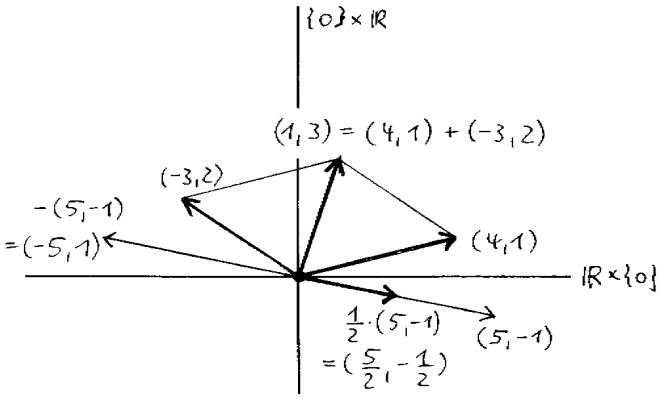
17.5 Beispiele (1) Sei $n \in \mathbb{N}$. Für jeden Körper K wird das kartesische Produkt

$$K^n = K \times K \times \cdots \times K$$

durch die komponentenweise Addition und skalare Multiplikation

$$\begin{aligned}(x_1, \dots, x_n) + (y_1, \dots, y_n) &:= (x_1 + y_1, \dots, x_n + y_n)\\ \lambda\,(x_1, \dots, x_n) &:= (\lambda x_1, \dots, \lambda x_n)\end{aligned}$$

zu einem K-Vektorraum. Das ist trivial nachzuprüfen: der Nullvektor ist $0 = (0, \dots, 0)$, der zu $x = (x_1, \dots, x_n)$ (additiv) inverse ist $-x = (-x_1, \dots, -x_n)$. Der wichtige Fall $K = \mathbb{R}$ ist Ihnen sicher bestens vertraut:



Naheliegende Unterräume von K^n sind die n "Koordinatenachsen"

$$\big\{(x_1, \dots, x_n) \in K^n \,\big|\, x_j = 0 \text{ für alle } j \neq k\big\}$$

$(k = 1, 2, \dots, n)$, oder auch für jede Zerlegung $n = k + l$ die Unterräume

$$K^k \times \{0\} := \big\{(x_1, \dots, x_n) \in K^n \,\big|\, x_j = 0 \text{ für alle } j > k\big\}$$

und $\{0\} \times K^l \subset K^n$.

Übrigens wollen wir nicht nur $K^1 = K$, sondern auch $K^0 = \{0\}$ zulassen; das einzige Element von K^0, das 0-tupel (das eben gar keinen Eintrag hat), ist darin zwangsläufig der Nullvektor.

(2) Die Menge aller reellen Zahlenfolgen

$$\big\{(x_n)_{n=0}^{\infty} \,\big|\, x_n \in \mathbb{R} \text{ für alle } n\big\}$$

ist auf die naheliegende Weise ein $\mathbb{R}$-Vektorraum, der als Beispiele interessanter Unterräume den Teilraum der beschränkten und den Teilraum der konvergenten Folgen enthält (hier kommen die Limesregeln 3.8 wieder zu Ehren).

(3) Für jede Menge I hat man den $\mathbb{R}$-Vektorraum

$$\{f \mid f\colon I \longrightarrow \mathbb{R}\}$$

der Funktionen von I nach $\mathbb{R}$; Addition und skalare Multiplikation darin sind punktweise erklärt. Ist speziell $I \subset \mathbb{R}$, so bilden nach den bekannten Regeln (7.4) die stetigen Funktionen einen Untervektorraum darin, und wenn I ein echtes Intervall ist, können wir gleich eine ganze Serie von Unterräumen angeben, nämlich für jedes $k \in \mathbb{N}$, und auch für $k = \infty$ den Raum

$$C^k(I, \mathbb{R}) := \{f\colon I \longrightarrow \mathbb{R} \mid f \text{ ist } C^k\text{-Funktion}\}.$$

Nach Lemma 17.4 ist der springende Punkt dabei nur, daß für je zwei C^k-Funktionen f, g auch $f+g$, und für $\lambda \in \mathbb{R}$ auch λf wieder eine solche Funktion ist — daß überdies auch das Produkt fg eine C^k-Funktion ist, ist zwar wahr, aber irrelevant. In all diesen Räumen enthalten ist noch der Vektorraum

$$\mathcal{O}(I, \mathbb{R}) := \{f\colon I \longrightarrow \mathbb{R} \mid f \text{ analytisch}\},$$

und man hat damit eine unendliche absteigende Kette von Teilräumen

$$C^0(I, \mathbb{R}) \supset C^1(I, \mathbb{R}) \supset \cdots \supset C^k(I, \mathbb{R}) \supset C^{k+1}(I, \mathbb{R}) \supset \cdots \supset C^\infty(I, \mathbb{R}) \supset \mathcal{O}(I, \mathbb{R})$$

vor sich.

(4) Oft noch wichtiger sind die entsprechenden Räume komplexwertiger Funktionen:

$$C^k(I, \mathbb{C}) := \{f\colon I \longrightarrow \mathbb{C} \mid f \text{ ist } C^k\text{-Funktion}\} \quad \text{und} \quad \mathcal{O}(I, \mathbb{C}) := \{f\colon I \longrightarrow \mathbb{C} \mid f \text{ analytisch}\}$$

Wenn aus dem Zusammenhang hervorgeht, ob reell- oder komplexwertige Funktionen gemeint sind, erlaubt man sich, einfach $C^k(I)$ zu schreiben. Für ein Gebiet $G \subset \mathbb{C}$ als Definitionsbereich ist auch der komplexe Vektorraum $\mathcal{O}(G)$ der analytischen Funktionen $G \longrightarrow \mathbb{C}$ von Interesse.

(5) Die Polynome mit Koeffizienten im Körper K bilden einen K-Vektorraum, den man mit

$$K[X] := \left\{ f(X) = \sum_{k=0}^{n} a_k X^k \;\middle|\; n \in \mathbb{N} \text{ und } a_k \in K \text{ für } k = 0, 1, \ldots, n \right\}$$

bezeichnet. Darin bedeutet das Symbol X strenggenommen gar nichts; es ist nur ein Platzhalter, den man einfügt, weil sich Polynome sonst nicht gut hinschreiben lassen. Natürlich kann man sich ebensogut für ein anderes Symbol entscheiden. Auch hier kommt es nur darauf an, daß man Polynome addieren und mit einem Skalar (aus K) multiplizieren kann; daß man sie auch *miteinander* multiplizieren kann (und $K[X]$ deshalb auch ein Ring ist), spielt für die Vektorraumeigenschaft keine Rolle.

Für jedes $d \in \mathbb{N}$ ist

$$\{f \in K[X] \mid \deg f \leq d\}$$

ein Untervektorraum (erinnern Sie sich an die spitzfindige Konvention, nach der das Nullpolynom zwar keinen Grad hat, aber doch $\deg 0 \leq d$ erfüllt?). Dagegen ist die Menge $\{f \in K[X] \mid \deg f = d\}$ *kein* Vektorraum, allein schon, weil sie das Nullpolynom als einzig möglichen Nullvektor nicht enthält.

An dieser Stelle sei angemerkt, daß es auch Körper gibt, die nur endlich viele Elemente haben, und daß für solche Körper die Definition des Polynombegriffs gegenüber der von 3.9 modifiziert werden muß. Solange unser Interesse ohnehin nur den Körpern $\mathbb{R}$ und $\mathbb{C}$ gilt, brauchen wir uns darum nicht zu sorgen.

Zum Thema Unterräume wollen wir noch einen ganz einfachen Sachverhalt festhalten:

17.6 Notiz Ist V ein Vektorraum und sind $T_1 \subset V$ und $T_2 \subset V$ Untervektorräume, so ist auch $T_1 \cap T_2$ ein Unterraum von V. (Ist allgemeiner $(T_\lambda)_{\lambda \in \Lambda}$ eine Familie von Unterräumen $T_\lambda \subset V$, so ist

$$\bigcap_{\lambda \in \Lambda} T_\lambda = \{x \in V \mid x \in T_\lambda \text{ für jedes } \lambda \in \Lambda\}$$

ein Unterraum von V.)

Unter den Abbildungen zwischen Vektorräumen sind naturgemäß diejenigen von besonderem Interesse, die auf die Vektorraumstrukturen Rücksicht nehmen.

17.7 Definition V und W seien K-Vektorräume. Eine Abbildung

$$V \xrightarrow{f} W$$

heißt linear, wenn

$$f(x+y) = f(x) + f(y)$$
$$f(\lambda x) = \lambda f(x)$$

für alle $x, y \in V$ und alle $\lambda \in K$ gilt.

17.8 Notiz Natürlich ist die identische Abbildung $\mathrm{id}_V : V \longrightarrow V$ stets linear; sind $V \xrightarrow{f} W$ und $W \xrightarrow{g} X$ linear, so auch $g \circ f : V \longrightarrow X$.

17.9 Lemma und **Definition** $f : V \longrightarrow W$ sei linear. Ist $T \subset W$ ein Unterraum, so ist auch das Urbild $f^{-1}T \subset V$ ein Unterraum. Für jeden Unterraum $S \subset V$ ist andererseits auch $f(S) \subset W$ ein Unterraum — insbesondere gilt $f(0) = 0$.

Speziell ist die Faser von f über $0 \in W$ ein Unterraum von V, den man den Kern von f nennt:

$$\operatorname{Kern} f = f^{-1}\{0\} = \{x \in V \mid f(x) = 0\} \subset V\,.$$

Andererseits ist die Bildmenge von f ein Unterraum

$$\operatorname{Bild} f = f(V) = \{f(x) \mid x \in V\}$$

von W.

Beweis Ganz leicht, hier zur Illustration für den Kern als besonders wichtigen Spezialfall: Wegen $f(0) = f(0+0) = f(0) + f(0)$ ist $f(0) = 0$, d.h $0 \in \operatorname{Kern} f$. Aus $x, y \in \operatorname{Kern} f$, also $f(x) = f(y) = 0$ folgt $f(x+y) = f(x) + f(y) = 0$, d.h. $x + y \in \operatorname{Kern} f$. Aus $\lambda \in K$ und $x \in \operatorname{Kern} f$ folgt schließlich $f(\lambda x) = \lambda f(x) = \lambda 0 = 0$, also $\lambda x \in \operatorname{Kern} f$. Damit liegen die drei in Lemma 17.4 genannten Eigenschaften vor.

Während $f : V \longrightarrow W$ natürlich genau dann surjektiv ist, wenn $\operatorname{Bild} f = W$ ist, kann man dem Kern ansehen, ob f injektiv ist:

17.10 Lemma Eine lineare Abbildung f ist genau dann injektiv, wenn $\operatorname{Kern} f = \{0\}$ ist.

Beweis f sei injektiv. Ist $x \in \operatorname{Kern} f$, so ist

$$f(x) = 0 = f(0)$$

und damit $x = 0$. Also ist $\operatorname{Kern} f = \{0\}$.

Sei $\operatorname{Kern} f = \{0\}$ vorausgesetzt, und seien $x, y \in V$ Vektoren mit $f(x) = f(y)$. Aus der Gleichung

$$f(x-y) = f(x) - f(y) = 0$$

folgt dann $x - y \in \operatorname{Kern} f = \{0\}$, also $x - y = 0$, d.h. $x = y$. Also ist f injektiv.

17.11 Lemma Ist die lineare Abbildung $f : V \longrightarrow W$ sogar bijektiv, so ist auch $f^{-1} : W \longrightarrow V$ linear.

Beweis Seien $u, v \in W$ sowie $\lambda \in K$. Wir können $u = f(x),\ v = f(y)$ mit eindeutig bestimmten $x, y \in V$ schreiben. Wenn wir nun auf beide Seiten von

$$u + v = f(x) + f(y) = f(x+y)$$
$$\lambda u = \lambda f(x) = f(\lambda x)$$

die Abbildung f^{-1} loslassen, ergeben sich die gewünschten Identitäten

$$f^{-1}(u+v) = x+y = f^{-1}(u) + f^{-1}(v)$$
$$f^{-1}(\lambda u) = \lambda x = \lambda f^{-1}(u).$$

Bemerkungen In der Analysis werden oft auch Abbildungen des Typs

$$\mathbb{R} \ni x \mapsto ax + b \in \mathbb{R}$$

als linear bezeichnet. Für $b \neq 0$ sind diese Abbildungen *nicht* linear in dem jetzt besprochenen Sinne (0 wird dabei ja nicht auf 0, sondern auf b abgebildet). In der linearen Algebra nennt man solche Abbildungen *affin* (oder auch *affin-linear*). — Anders auch als in der Analysis ist man in der linearen Algebra mit dem Begriff der Surjektivitiät sehr genau und redet von der Umkehrabbildung f^{-1} nur dann, wenn f wirklich bijektiv und nicht nur injektiv ist. — In der Physik wird das Wort "linear" gelegentlich im Sinne von "eindimensional" mißbraucht, zum Beispiel beim *linearen* harmonischen Oszillator.

Lineare Abbildungen nennt man alternativ auch lineare Homomorphismen. Das Attribut "linear" unterscheidet dabei von anderen, analog erklärten Homomophismen etwa zwischen Gruppen $G \xrightarrow{f} H$ (Forderung: $f(xy) = f(x)f(y)$ für alle $x, y \in G$) oder zwischen Ringen $R \xrightarrow{f} S$ (Forderungen: $f(x+y) = f(x) + f(y)$ und $f(xy) = f(x)f(y)$ für alle $x, y \in R$, sowie $f(1) = 1$) oder anderen, vielleicht noch nicht erfundenen algebraischen Strukturen. Wenn kein Mißverständnis zu befürchten ist, läßt man den Zusatz aber auch weg. In ebenso allgemeinem Rahmen verwenden kann man die

17.12 Definition Bijektive Homomorphismen $f: V \longrightarrow W$ heißen Isomorphismen; zwei K-Vektorräume V und W heißen isomorph, wenn es einen Isomorphismus $f: V \longrightarrow W$ gibt. In Zeichen:

$$f: V \xrightarrow{\simeq} W \quad \text{bzw.} \quad V \simeq W$$

Ein Isomorphismus $f: V \simeq W$ erlaubt es, zwischen den beiden Vektorräumen V und W beliebig hin- und herzugehen, wobei sich Vektoraddition und skalare Multiplikation in V und in W entsprechen. Wenn man einen solchen Isomorphismus f hat, kennt man V genau so gut (oder schlecht) wie W; und eine Beschreibung eines zunächst unbekannten Vektorraumes V besteht oft darin, einen Isomorphismus f zwischen V und einem Vektorraum W anzugeben, den man schon kennt.

Beispiele linearer Abbildungen zwischen den K-Vektorräumen K^n und K^p erhält man systematisch aus sogenannten Matrizen.

17.13 Definition Sei K ein Körper (oder allgemeiner ein Ring), und seien $p, n \in \mathbb{N}$. Eine $p \times n$-Matrix über K ist dann — formal gesehen — eine Abbildung

$$\{1, \ldots, p\} \times \{1, \ldots, n\} \longrightarrow K.$$

Ähnlich aber wie man ein n-tupel a normalerweise nicht als Funktion $j \mapsto a(j)$ schreibt, sondern als Zeile $a = (a_1, \ldots, a_n)$, notiert man eine solche $p \times n$-Matrix als ein rechteckiges Schema

$$a = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ldots & \vdots \\ a_{p1} & a_{p2} & \ldots & a_{pn} \end{pmatrix}$$

Die pn Skalare a_{ij} heißen die Komponenten oder Einträge der Matrix a. Naheliegenderweise nennt man die $1 \times n$-Matrix

$$\begin{pmatrix} a_{i1} & a_{i2} & \ldots & a_{in} \end{pmatrix}$$

die i-te Zeile von a, dagegen die $p\times 1$-Matrix

$$\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{pj} \end{pmatrix}$$

die j-te Spalte von a.

Die Menge aller $p\times n$-Matrizen über K bezeichen wir mit $\mathrm{Mat}(p\times n, K)$.

Eine $1\times n$-Matrix ist offenbar nichts wesentlich Anderes als ein n-tupel von Elementen aus K, und entsprechend eine $p\times 1$-Matrix im wesentlichen ein p-tupel; wir könnten also bedenkenlos $\mathrm{Mat}(1\times n, K)$ mit K^n und $\mathrm{Mat}(p\times 1, K)$ mit K^p identifizieren. Für das Rechnen mit Matrizen ist es wichtig, sich auf nur eine dieser beiden Möglichkeiten festzulegen, und aus Gründen der Verträglichkeit mit anderen etablierten Konventionen entscheidet man sich für die zweite: Wir treffen also hier die *Vereinbarung*, die Vektoren aus K^p künftig nur noch als Spalten

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} \in \mathrm{Mat}(p\times 1, K) = K^p$$

zu schreiben, auch wenn das schreibtechnisch manchmal etwas lästig sein kann. Weiter hat man nun eine Multiplikation

$$\begin{aligned} \mathrm{Mat}(p\times n, K)\times K^n &\longrightarrow K^p \\ (a, x) &\mapsto ax \end{aligned}$$

festgelegt, nämlich durch

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \dots & \vdots \\ a_{p1} & \dots & a_{pn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1+ & \cdots & +a_{1n}x_n \\ & \vdots & \\ a_{p1}x_1+ & \cdots & +a_{pn}x_n \end{pmatrix},$$

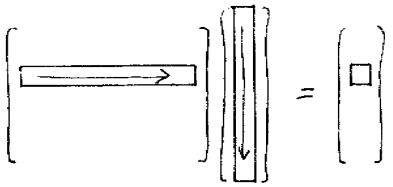
oder platzsparend geschrieben:

$$(ax)_i = \sum_{j=1}^{n} a_{ij}x_j \quad \text{für } i = 1, \dots, p.$$

Daß man das gerade so und nicht anders macht, auch das ist eine etwas willkürliche Übereinkunft, aber eine allgemein akzeptierte. Jedenfalls müssen Sie sich all diese Definitionen gut einprägen, damit Sie insbesondere nicht

- eine $p\times n$-Matrix mit einer $n\times p$-Matrix oder
- den Zeilenindex (den ersten) mit dem Spaltenindex (dem zweiten) verwechseln, oder
- das Matrixprodukt falsch berechnen.

Speziell letzteres dürfte kaum passieren, wenn man sich das leicht zu merkende Schema



vor Augen hält.

Der angekündigte Zusammenhang mit linearen Abbildungen ergibt sich nun so:

17.14 Lemma Für jede Matrix $a \in \mathrm{Mat}(p\times n, K)$ ist die Abbildung

$$\begin{aligned} K^n &\longrightarrow K^p \\ x &\mapsto ax \end{aligned}$$

linear.

Beweis Man muß bloß rechnen:

$$\begin{aligned} \big(a(x+y)\big)_i &= \sum_{j=1}^n a_{ij}(x+y)_j = \sum_{j=1}^n a_{ij}(x_j+y_j) = \sum_{j=1}^n a_{ij}x_j + \sum_{j=1}^n a_{ij}y_j = (ax)_i + (ay)_i \\ \big(a(\lambda x)\big)_i &= \sum_{j=1}^n a_{ij}(\lambda x)_j = \lambda\sum_{j=1}^n a_{ij}(x_j) = \lambda(ax)_i \end{aligned}$$

Jede Matrix über K liefert damit ein Beispiel einer linearen Abbildung.

Bemerkung Die Autoren physikalischer Texte neigen dazu, statt Matrizen deren Komponenten hinzuschreiben. So könnten Sie die Gleichung

$$ax = b \quad (\text{mit } a \in \mathrm{Mat}(p\times n, K),\ x \in K^n \text{ und } b \in K^p)$$

dort in einer der Formen

$$\sum_{j=1}^n a_{ij}x_j = b_i \quad \text{oder} \quad \sum_j a_{ij}x_j = b_i \quad \text{oder bloß} \quad a_{ij}x_j = b_i$$

lesen, wobei der Pfiff der dritten Varianten die stillschweigende Abmachung ist, daß über jeden doppelt vorkommenden Index (der einen aus dem Zusammenhang hervorgehenden Bereich durchläuft) summiert werden soll. Ich selbst kann mit dieser Art der Physiker ganz gut leben, empfehle Ihnen aber doch eher den vorgetragenen Standpunkt, nach dem eine Matrix ein eigenständiges Objekt ist und nicht bloß eine symbolische Schreibweise für die Gesamtheit ihrer Einträge. Wenn man sich daran konsequent hält, ist man fast automatisch vor vielen Fehlern geschützt, die sich beim Umgang mit den indizierten Komponenten leicht einschleichen.

Die folgenden Beispiele linearer Abbildungen sehen ganz anders aus:

17.15 Beispiele (1) Die Abbildung

$$\big\{(x_n)_{n=0}^\infty \,\big|\, x_n \in \mathbb{R} \text{ für alle } n,\ (x_n)_{n=0}^\infty \text{ konvergiert}\big\} \xrightarrow{\lim} \mathbb{R},$$

die jeder konvergenten Folge ihren Limes zuordnet, ist nach den Regeln 3.8 linear. Lineare Abbildungen, die auf einem so "großen" Vektorraum wie dem der konvergenten Folgen definiert sind und skalare Werte haben, heißen auch lineare *Funktionale*. Auch die auf einem Vektorraum von Funktionen definierten Integrale $\int_a^b : C^0[a,b] \to \mathbb{R}$ sind Beispiele von Funktionalen.

(2) Sei $I \subset \mathbb{R}$ ein echtes Intervall. Das Differenzieren $f \mapsto f'$ definiert für jedes $k \in \mathbb{N}$ einen sogenannten *Differentialoperator*

$$D: C^{k+1}(I) \longrightarrow C^k(I),$$

der nach den Ableitungsregeln 13.4 eine lineare Abbildung ist. Desgleichen

$$D: C^\infty(I) \longrightarrow C^\infty(I) \quad \text{und } D: \mathcal{O}(I) \longrightarrow \mathcal{O}(I).$$

In jedem Fall besteht der Kern von D nach 15.9 aus den konstanten Funktionen auf I. In gelehrter Sprechweise: Die lineare Abbildung $\mathbb{R} \longrightarrow \operatorname{Kern} D$, die $c \in \mathbb{R}$ die konstante Funktion $f: I \longrightarrow \mathbb{R}$ mit Wert c zuweist,

ist ein Isomorphismus. (Dies ist ein gutes Beispiel dafür, wie man den zunächst unbekannten Vektorraum Kern D dadurch beschreibt, daß man einen Isomorphismus zu dem leicht zu durchschauenden $\mathbb{R}$ herstellt.)

(3) Sei I wie vor, und sei $a \in I$ ein fester Punkt. Wenn wir den Vektorraum der reellen Polynome vom Grad höchstens k ad hoc mit P_k bezeichnen, dann definiert die Bildung des k-ten Taylor-Polynoms bei a eine lineare Abbildung

$$T_a^k \colon C^k(I) \longrightarrow P_k.$$

P_k selbst ist offenbar zu $\mathbb{R}^{k+1}$ isomorph, vermöge

$$\mathbb{R}^{k+1} \ni \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{pmatrix} \mapsto \sum_{j=0}^{k} a_j X^j \in P_k.$$

(4) Der quantenmechanische harmonische Oszillator, der der Gegenstand der Übungsaufgaben 15.1 und 15.2 war, liefert eine Fülle weiterer Beispiele. Bei gegebenem $E \in \mathbb{R}$ definiert die Differentialgleichung

$$f''(x) - 2xf'(x) + (2E-1)f(x) = 0$$

verschiedene lineare Differentialoperatoren, je nachdem, welche Art von Funktionen man als Lösungen ins Auge faßt. Wir hatten speziell nach analytischen Lösungen $f \colon (-\delta, \delta) \longrightarrow \mathbb{C}$ gesucht, für festes $\delta > 0$ also den Differentialoperator

$$D_\delta \colon \mathcal{O}(-\delta, \delta) \longrightarrow \mathcal{O}(-\delta, \delta), \quad (D_\delta f)(x) = f''(x) - 2xf'(x) + (2E-1)f(x)$$

betrachtet. Die Differentialgleichung zu lösen bedeutet, den Kern von D_δ zu bestimmen, und unser Resultat war, daß es unabhängig von der Wahl von δ zu jeder Vorgabe der beiden ersten Taylor-Koeffizienten bei 0 genau eine Lösung gibt, oder — gelehrt ausgedrückt — daß die lineare Abbildung

$$\operatorname{Kern} D_\delta \longrightarrow \mathbb{R}^2; \quad f \mapsto \begin{pmatrix} f(0) \\ f'(0) \end{pmatrix}$$

bijektiv, also ein Isomorphismus ist.

Das zweites Resultat besagte, daß all diese Lösungen der Gleichung in Wirklichkeit auf ganz $\mathbb{C}$ erklärte analytische Funktionen sind. In der Sprache der linearen Algebra läßt sich das mittels der Einschränkungsabbildung

$$\mathcal{O}(\mathbb{C}) \longrightarrow \mathcal{O}(-\delta, \delta), \quad f \mapsto f|(-\delta, \delta)$$

ausdrücken. Beachten Sie, daß diese Abbildung nicht nur linear, sondern auch injektiv ist: Eine analytische Funktion, die auf $(-\delta, \delta)$ verschwindet, hat bei 0 sicher das Talyor-Polynom 0 und muß nach dem Identitätssatz 16.2 deshalb selbst schon die Nullfunktion sein.

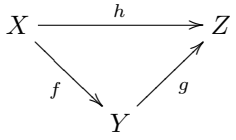
Schließlich (Aufgabe 15.2) hatten wir uns überlegt, daß die gefundenen Lösungen $f \neq 0$ nur für $E = k + \frac{1}{2}$ (mit $k \in \mathbb{N}$) der Wachstumsbedingung $\lim\limits_{x \to \pm\infty} e^{-x^2/2}|f(x)| = 0$ genügen, also daß für alle übrigen $E \in \mathbb{R}$ der Durchschnitt der beiden Unterräume von $\mathcal{O}(\mathbb{C})$

$$\operatorname{Kern} D \cap \left\{ f \in \mathcal{O}(\mathbb{C}) \,\middle|\, \lim_{x \to \pm\infty} e^{-x^2/2}|f(x)| = 0 \right\} = \{0\}$$

der Nullraum ist.

Wenn man es wie im vorstehenden Beispiel mit einer Vielzahl zusammengehöriger Mengen und Abbildungen zu tun hat, stellt man deren Zusammenwirken am besten in einem sogenannten kommutativen Diagramm dar. Das ist eine von der linearen Algebra ganz unabhängige, vielmehr schon zur Mengenlehre gehörige

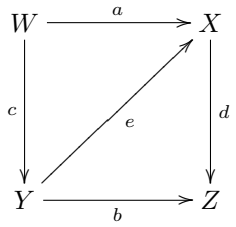
17.16 Sprechweise In einem Diagramm (von Mengen und Abbildungen) wird jede der beteiligten Abbildungen durch einen Pfeil vom jeweiligen Definitions- zum Zielbereich repräsentiert, so bringt man zum Beispiel mit dem Diagramm



zum Ausdruck, daß $f\colon X \longrightarrow Y$, $g\colon Y \longrightarrow Z$ und $h\colon X \longrightarrow Z$ Abbildungen sind. Aneinandersetzbaren Pfeilen wie

$$X \xrightarrow{f} Y \quad \text{und} \quad Y \xrightarrow{g} Z$$

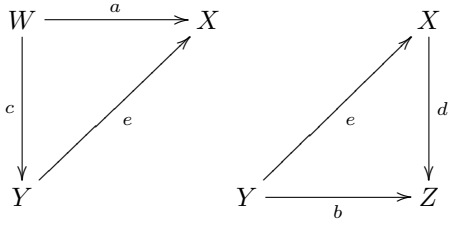
kann man die Komposition der entsprechenden Abbildungen zuordnen, hier also $g \circ f$. Allgemeiner bestimmt jede Kette aneinandersetzbarer Pfeile eine Komposition ebensovieler Abbildungen, und wenn man von dem Diagramm sagt, es sei kommutativ, heißt das, daß je zwei Ketten mit gemeinsamer Start- und Zielmenge dieselbe Komposition liefern. Im Beispieldiagramm gibt es für solche Ketten nicht viel Auswahl, und die Kommutativität bedeutet schlichtweg $g \circ f = h$. Bei dem schon interessanteren Diagramm



faßt Kommutativität die Aussagen

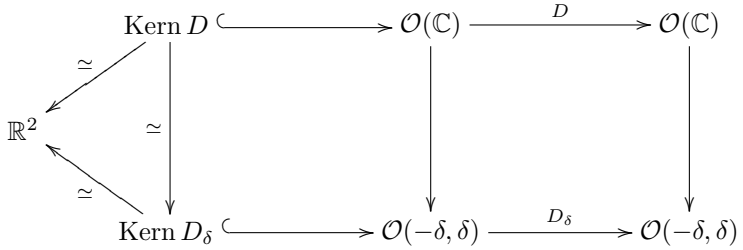
$$a = e \circ c,\ b = d \circ e \text{ und } d \circ a = b \circ c\ (= d \circ e \circ c)$$

zusammen. Man macht oft davon Gebrauch, daß man zwei kommutative Diagramme mit gemeinsamen Pfeilen zu einem größeren Diagramm zusammensetzen kann, das dann ebenfalls kommutiert. So mag man sich das obige Diagramm als aus den beiden kommutativen Dreiecken



entstanden vorstellen, und in der Tat ist $d \circ a = b \circ c$ eine Konsequenz der beiden schon aus den Dreiecken abzulesenden Gleichungen $a = e \circ c$ und $b = d \circ e$.

Die meisten der im Beipiel (4) zum harmonischen Oszillator eingeführten Vektorräume und Homomorphismen lassen sich jetzt ganz übersichtlich in dem kommutativen Diagramm



darstellen, in dem auch der Operator D durch die Differentialgleichung gegeben ist, die vertikalen Pfeile Einschränkungshomomorphismen sind und die beiden schrägen Pfeile wie beschrieben Funktion und erste Ableitung bei 0 auswerten.

Übungsaufgaben

17.1 Welche der folgenden Teilmengen T des Vektorraums V sind Untervektorräume?

(a) $T := \{f \mid f(0) = f(1) = 0\} \subset C^0[0,1] =: V$

(b) $T := \{f \mid f(0) = f(1) = 1\} \subset C^0[0,1] =: V$

(c) $T := \{f \mid f(0)f(1) = 0\} \subset C^0[0,1] =: V$

(d) $T := \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \;\middle|\; x_1^3 + x_1 x_2^2 = 0 \right\} \subset \mathbb{R}^2 =: V$

(e) $T := \left\{ \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \;\middle|\; z_1^3 + z_1 z_2^2 = 0 \right\} \subset \mathbb{C}^2 =: V$

17.2 V sei ein Vektorraum. Beweisen Sie: Die Vereinigung zweier Teilräume $T_1, T_2 \subset V$ ist nie ein Teilraum von V — na ja, *fast* nie, wann nämlich doch?

17.3 Jede Wahl einer Funktion $h \in C^0(\mathbb{R})$ definiert durch $H(f) := h \cdot f$, explizit also

$$\big(H(f)\big)(t) = h(t) \cdot f(t) \quad \text{für alle } t \in \mathbb{R}$$

eine lineare Abbildung $H\colon C^0(\mathbb{R}) \longrightarrow C^0(\mathbb{R})$. Für welche h ist H surjektiv, für welche injektiv?

18 Basen

Dieser Abschnitt ist grundlegend für das konkrete Rechnen in den sogenannten endlichdimensionalen Vektorräumen (die im Laufe des Abschnitts definiert werden und zu denen jedenfalls die K-Vektorräume K^n für alle $n \in \mathbb{N}$ zählen).

18.1 Sprechweisen und **Definitionen** Es sei V ein K-Vektorraum und $r \in \mathbb{N}$; weiter seien $v_1, \dots, v_r \in V$ Vektoren und $\lambda_1, \dots, \lambda_r \in K$ ebensoviele Skalare. Aus diesen Daten können wir den Vektor

$$v = \lambda_1 v_1 + \cdots + \lambda_r v_r = \sum_{i=1}^{r} \lambda_i v_i \in V$$

bilden. Man sagt, daß v eine Linearkombination der Vektoren $v_1, \dots, v_r \in V$ ist, nämlich diejenige mit den Koeffizienten $\lambda_1, \dots, \lambda_r$.

Die Menge aller Linearkombinationen von $v_1, \dots, v_r \in V$ heißt die lineare Hülle von $v_1, \dots, v_r \in V$, oder auch der von $v_1, \dots, v_r \in V$ aufgespannte Unterraum von V:

$$\mathrm{Lin}(v_1, \dots, v_r) := \Big\{ \sum_{i=1}^{r} \lambda_i v_i \;\Big|\; \lambda_i \in K \Big\} \subset V$$

Gleichwertig, aber gelehrter ausgedrückt ist $\mathrm{Lin}(v_1, \dots, v_r)$ die Bildmenge der durch

$$K^r \ni \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_r \end{pmatrix} \longmapsto \sum_{i=1}^{r} \lambda_i v_i \in V$$

definierten linearen Abbildung $K^r \longrightarrow V$, die wir ab jetzt durchweg mit $\Phi_{(v_1,\dots,v_r)} \colon K^r \longrightarrow V$ bezeichnen.

Es liegt nahe, wei die Definition im Fall $r = 0$ zu lesen ist; die leere Summe ergibt den Nullvektor, und es ist $\mathrm{Lin}(\,) = \{0\}$ und nicht etwa $\mathrm{Lin}(\,) = \emptyset$.

18.2 Lemma Für jede Wahl von $v_1, \dots, v_r \in V$ ist $\mathrm{Lin}(v_1, \dots, v_r) \subset V$ tatsächlich ein Untervektorraum, und zwar der kleinste Unterraum von V, der die Vektoren $v_1, \dots, v_r$ enthält.

Beweis Nach Lemma 17.9 ist $\mathrm{Lin}(v_1, \dots, v_r) \subset V$ ein Unterraum, nämlich der Bildraum der linearen Abbildung $\Phi_{(v_1,\dots,v_r)}$. Andererseits muß offenbar jeder $v_1, \dots, v_r$ enthaltende Unterraum von V auch alle Linearkombinationen dieser Vektoren enthalten, d.h. $\mathrm{Lin}(v_1, \dots, v_r)$ umfassen.

18.3 Definition Sei V ein K-Vektorraum. Ein r-tupel $(v_1, \dots, v_r)$ heißt linear unabhängig, wenn aus

$$0 = \sum_{i=1}^{r} \lambda_i v_i \quad \text{mit } \lambda_i \in K \text{ für } i = 1, \dots, r$$

stets

$$\lambda_1 = \lambda_2 = \cdots = \lambda_r = 0$$

folgt.

Mit anderen Worten: Der Nullvektor darf sich nur auf die triviale Weise als Linearkombination von $v_1, \dots, v_r$ schreiben lassen. Konsequenterweise ist auch das leere, also das 0-tupel als linear unabhängig einzustufen, denn dann sind ja gar keine Koeffizienten λ_i vorhanden und schon deshalb alle gleich null.

18.3½ Notiz Das r-tupel von Vektoren $(v_1, \dots, v_r)$ ist genau dann linear unabhängig, wenn die oben betrachtete Abbildung $\Phi_{(v_1,\dots,v_r)}\colon K^r \longrightarrow V$ injektiv ist.

Beweis Die lineare Unabhängigkeit besagt, daß der Kern dieser linearen Abbildung der Nullraum ist, und nach Lemma 17.10 ist das zur Injektivität gleichwertig.

Wieder zurückübersetzend, erweist sich damit $(v_1, \dots, v_r)$ genau dann als linear unabhängig, wenn die Koeffizienten λ_i einer jeden Linearkombination $v = \sum_{i=1}^r \lambda_i v_i$ durch v eindeutig bestimmt sind. Wieder eine andere Charakterisierung des Unabhängigkeitsbegriffs liefert

18.4 Lemma Das r-tupel von Vektoren $(v_1, \dots, v_r)$ ist genau dann linear abhängig, wenn es einen Index $k \in \{1, \dots, r\}$ gibt, so daß v_k eine Linearkombination der v_i mit $i \neq k$ ist.

Beweis Sei v_k eine solche Linearkombination. Es gibt also Skalare $\lambda_1, \dots, \lambda_{k-1}, \lambda_{k+1}, \dots, \lambda_r \in K$ mit

$$v_k = \sum_{i=1}^{k-1} \lambda_i v_i + \sum_{i=k+1}^{r} \lambda_i v_i.$$

Indem wir $\lambda_k := -1$ setzen, können wir diese Gleichung als

$$0 = \sum_{i=1}^{r} \lambda_i v_i$$

schreiben, und das zeigt, daß $(v_1, \dots, v_r)$ ein linear abhängiges r-tupel ist.

Jetzt setzen wir umgekehrt gerade das voraus: Es existieren also Skalare $\lambda_1, \dots, \lambda_r \in K$, nicht alle null, mit

$$0 = \sum_{i=1}^{r} \lambda_i v_i.$$

Sei etwa $\lambda_k \neq 0$. Multiplikation der vorstehenden Gleichung mit $\lambda_k^{-1} \in K$ liefert

$$0 = \lambda_k^{-1} \sum_{i=1}^{r} \lambda_i v_i = \sum_{i=1}^{r} (\lambda_k^{-1}\lambda_i) v_i = v_k + \sum_{\substack{i=1 \\ i\neq k}}^{r} (\lambda_k^{-1}\lambda_i) v_i,$$

also

$$v_k = \sum_{\substack{i=1 \\ i\neq k}}^{r} (-\lambda_k^{-1}\lambda_i) v_i.$$

Damit ist v_k in der Tat eine Linearkombination der übrigen Vektoren v_i.

Bleiben wir bei derselben Situation. Wenn man von vornherein schon weiß, daß das $(r-1)$-tupel $(v_1, \dots, v_{r-1})$ linear unabhängig ist, dann kann man sich auf ein bestimmtes k, nämlich $k = r$ festlegen. In etwas geänderten Bezeichnungen:

18.5 Lemma $(v_1, \dots, v_r)$ sei ein linear unabhängiges r-tupel von Vektoren, und $w \in V$ ein weiterer Vektor. Dann gilt:

$$(v_1, \dots, v_r, w) \text{ ist linear abhängig } \iff w \in \mathrm{Lin}(v_1, \dots, v_r)$$

Beweis Sei $(v_1, \dots, v_r, w)$ linear abhängig, etwa

$$0 = \sum_{i=1}^{r} \lambda_i v_i + \mu w$$

mit Skalaren λ_i und μ, die nicht alle null sind. Dann ist zwangsläufig $\mu \neq 0$, denn sonst bliebe von der Gleichung ja $0 = \sum_{i=1}^{r} \lambda_i v_i$, im Widerspruch zur linearen Unabhängigkeit von $(v_1, \ldots, v_r)$. Damit folgt aber wie vorhin $w \in \mathrm{Lin}(v_1, \ldots, v_k)$.

Die umgekehrte Schlußrichtung ergibt sich aus Lemma 18.4 ohnehin.

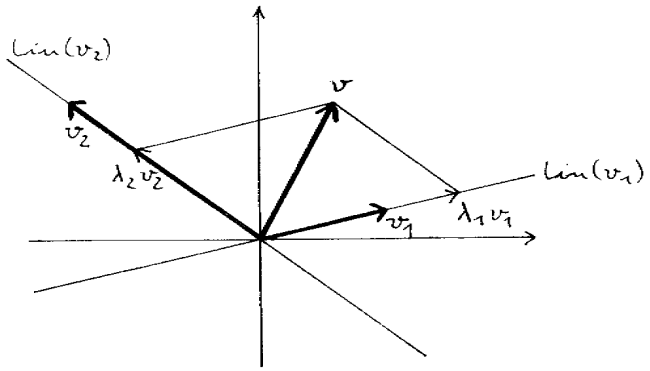
18.6 Definition Sei V ein Vektorraum. Ein n-tupel $(v_1, \ldots, v_n)$ von Vektoren aus V heißt eine Basis von V, wenn

- $\mathrm{Lin}(v_1, \ldots, v_n) = V$ ist — die Vektoren $v_1, \ldots, v_n$ also den ganzen Raum V aufspannen — und
- $(v_1, \ldots, v_n)$ linear unabhängig ist.

Das n-tupel $(v_1, \ldots, v_n)$ ist also genau dann eine Basis, wenn die in 18.1 definierte lineare Abbildung $\Phi_{(v_1,\ldots,v_n)} \colon K^n \longrightarrow V$ bijektiv, d.h. ein Isomorphismus von Vektorräumen ist. Ganz explizit heißt das, daß jeder Vektor aus v sich als Linearkombination $\sum_{i=1}^{n} \lambda_i v_i$ mit eindeutig bestimmten Koeffizienten $\lambda_i \in K$ schreiben läßt.

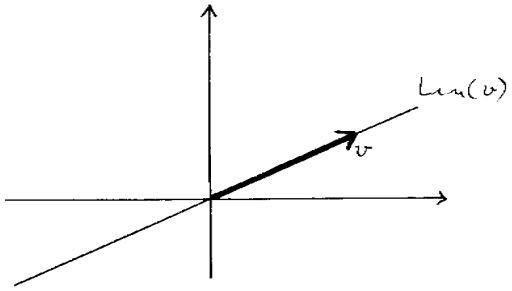
18.7 Beispiele Wir begeben uns in den reellen Vektorraum $\mathbb{R}^2$ und stützen uns zu Illustrationszwecken auf Argumente der ebenen Elementargeometrie.

(1)



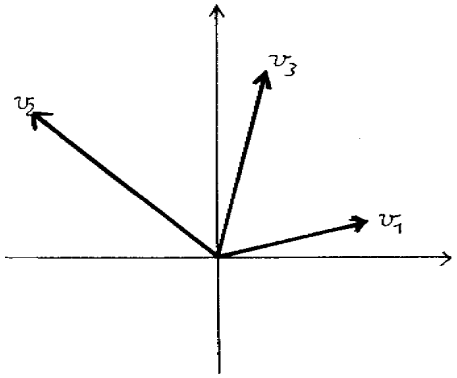
In diesem Bild ist (v_1, v_2) eine Basis von $\mathbb{R}^2$, denn jeder Vektor $v \in \mathbb{R}^2$ ist eine Linearkombination $v = \lambda_1 v_1 + \lambda_2 v_2$ mit reellen Zahlen λ_1 und λ_2, die sich eindeutig aus der eingezeichneten Parallelogrammkonstruktion ergeben; insbesondere ist der Nullvektor nur mit $\lambda_1 = \lambda_2 = 0$ darstellbar.

(2)



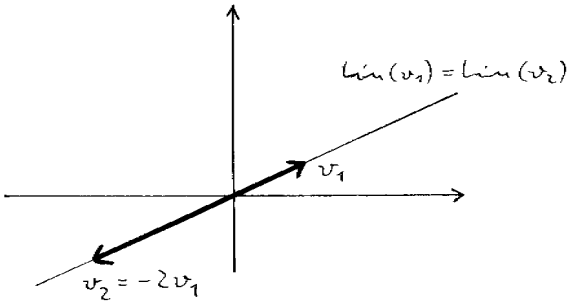
Ein einzelner Vektor $v \neq 0$ bildet keine Basis von $\mathbb{R}^2$: nur die Vektoren auf der von v aufgespannten Geraden sind Linearkombinationen, d.h. ja skalare Vielfache von v.

(3)



Auch (v_1, v_2, v_3) ist keine Basis von $\mathbb{R}^2$: Diese drei Vektoren spannen zwar $\mathbb{R}^2$ auf, aber es ist etwa v_3 selbst Linearkombination von v_1 und v_2, das Tripel (v_1, v_2, v_3) also linear abhängig.

(4)



Hier haben wir zwar die, wie uns das Gefühl sagt, "richtige" Anzahl von Vektoren, aber v_1 und v_2 spannen dieselbe Gerade auf, und deshalb ist auch $\mathrm{Lin}(v_1, v_2)$ nur diese Gerade und nicht ganz $\mathbb{R}^2$. Außerdem ist (v_1, v_2) linear abhängig: $2v_1 + v_2 = 0$.

Im Vorbeigehen noch die

18.8 Notiz Ein r-tupel, in dem

- der Nullvektor, oder
- zwei gleiche Vektoren

vorkommen, ist sicher linear abhängig.

Man sieht daran übrigens, daß lineare Abhängigkeit oder Unabhängigkeit wirklich eine Eigenschaft des r-tupels $(v_1, \ldots, v_r)$ ist und nicht bloß der daraus gebildeten Menge $\{v_1, \ldots, v_r\}$, der man das eventuell mehrfache Auftreten eines Vektors ja nicht mehr ansieht. Keinen Einfluß hat es freilich, wenn man die Komponenten des r-tupels untereinander vertauscht.

Die folgende Definition enthält auch ein Beispiel, aber ein besonders wichtiges:

18.9 Lemma und **Definition** Sei K ein Körper, $n \in \mathbb{N}$ beliebig. Der Vektorraum K^n besitzt dann die kanonische oder Standardbasis:

$$(e_1, e_2, \ldots, e_{n-1}, e_n) = \left(\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \right)$$

Beweis Die Identität

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \\ x_n \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \cdots + x_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

macht die Basiseigenschaft offensichtlich.

Diese Standardbasis von K^n ist so naheliegend und unverwechselbar — eben "kanonisch" — daß man sich fragen wird, warum man überhaupt jemals andere Basen in Betracht ziehen sollte. Nun, einen physikalischen Grund dafür habe ich Ihnen schon in Form eines Kristallgitters gezeigt, das ja die Wahl einer an die Kristallstruktur angepaßten Basis nahelegt. Sie werden aber bald sehen, daß es auch gute mathematische Gründe dafür gibt, in K^n nicht immer die Standardbasis zu verwenden. Viele Probleme der linearen Algebra lassen sich nämlich durch geschickte Wahl einer dem Problem angepaßten Basis erheblich vereinfachen, oft so weit, daß die Lösung allein in der Konstruktion einer solchen Basis besteht.

Das wichtigste Werkzeug für den Umgang mit Basen ist der sogenannte

18.10 Basisergänzungssatz V sei ein Vektorraum, und $v_1, \ldots, v_r, w_1, \ldots, w_s \in V$ seien Vektoren. Ist

$$(v_1, \ldots, v_r) \text{ linear unabhängig}$$

und

$$\mathrm{Lin}(v_1, \ldots, v_r, w_1, \ldots, w_s) = V,$$

so kann man das r-Tupel $(v_1, \ldots, v_r)$ zu einer Basis von V ergänzen, indem man geeignete der Vektoren $w_1, \ldots, w_s$ hinzufügt. Formaler gesagt: Es gibt dann ein $t \in \mathbb{N}$ und Indizes $j_1, \ldots, j_t \in \{1, \ldots, s\}$, so daß

$$(v_1, \ldots, v_r, w_{j_1}, \ldots, w_{j_t})$$

eine Basis von V ist.

Bemerkung Auch hier sind die Fälle $r = 0$ und/oder $s = 0$ nicht ausgeschlossen, wobei letzterer uninteressant ist, aber ersterer schon für sich Beachtung verdient: Aus einem s-Tupel von Vektoren, die V aufspannen, läßt sich stets eine Basis von V auswählen.

Beweis des Satzes Es kommt auf den richtigen Induktionsansatz an: Wir fixieren den Vektorraum V und beweisen durch vollständige Induktion nach s die folgende Aussage für jedes $s \in \mathbb{N}$.

> Gegeben seien $r \in \mathbb{N}$, ein linear unabhängiges r-tupel $(v_1, \ldots, v_r)$ von Vektoren aus V sowie s weitere Vektoren $w_1, \ldots, w_s \in V$ mit $\mathrm{Lin}(v_1, \ldots, v_r, w_1, \ldots, w_s) = V$. Dann wird $(v_1, \ldots, v_r)$ durch Hinzunahme geeigneter der Vektoren $w_1, \ldots, w_s$ zu einer Basis von V ergänzt.

Der Induktionsanfang ($s = 0$) ist trivial: Daß $(v_1, \ldots, v_r)$ linear unabhängig und $\mathrm{Lin}(v_1, \ldots, v_r) = V$ ist, bedeutet ja gerade, daß $(v_1, \ldots, v_r)$ selbst schon eine Basis von V ist.

Den Induktionsschritt formulieren wir als Schluß "von $s-1$ auf s" für $s \geq 1$: Gegeben sind dann das linear unabhängige r-tupel $(v_1, \ldots, v_r)$ und $w_1, \ldots, w_s$ mit $\mathrm{Lin}(v_1, \ldots, v_r, w_1, \ldots, w_s) = V$. Wir unterscheiden zwei Fälle:

$w_1 \in \mathrm{Lin}(v_1, \ldots, v_r)$: Nach der Charakterisierung 18.2 ist dann

$$\mathrm{Lin}(v_1, \ldots, v_r, w_1, \ldots, w_s) = \mathrm{Lin}(v_1, \ldots, v_r, w_2, \ldots, w_s),$$

weil der rechte Vektorraum auch w_1 und damit alle Vektoren $v_1, \ldots, v_r, w_1, \ldots, w_s$ enthält. Wir können w_1 also einfach ignorieren und die Induktionsannahme auf $(v_1, \ldots, v_r)$ und $w_2, \ldots, w_s$ anwenden.

$w_1 \notin \mathrm{Lin}(v_1, \ldots, v_r)$: Nach Lemma 18.5 ist dann $(v_1, \ldots, v_r, w_1)$ ein linear unabhängiges $(r+1)$-tupel, und wegen

$$\mathrm{Lin}(v_1, \ldots, v_r, w_1, \underbrace{w_2 \ldots, w_s}_{s-1 \text{ Vektoren}}) = V$$

verspricht die Induktionsannahme, daß wir dieses $(r+1)$-tupel zu einer Basis von V ergänzen können, indem wir gewisse der Vektoren $w_2, \dots, w_s$ hinzufügen.

Damit ist der Induktionsschluß geführt.

Als Folgerung aus diesem Satz zeigen wir das sogenannte

18.11 Austauschlemma Sind $(v_1, \dots, v_r)$ und $(w_1, \dots, w_s)$ zwei Basen ein und desselben Vektorraums V, so gibt es zu jedem $i \in \{1, \dots, r\}$ ein $j \in \{1, \dots, s\}$, so daß

$$(v_1, \dots, v_{i-1}, w_j, v_{i+1}, \dots, v_r)$$

ebenfalls eine Basis von V ist.

Beweis Weil v_i keine Linearkombination von $v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_r$ ist, gilt

$$\mathrm{Lin}(v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_r) \neq V.$$

Deshalb können die Vektoren $w_1, \dots, w_s$ nicht alle in $\mathrm{Lin}(v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_r)$ enthalten sein; sei etwa

$$w_j \notin \mathrm{Lin}(v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_r).$$

Nach Lemma 18.5 ist $(v_1, \dots, v_{i-1}, w_j, v_{i+1}, \dots, v_r)$ linear unabhängig, und wegen

$$\mathrm{Lin}(v_1, \dots, v_{i-1}, w_j, v_{i+1}, \dots, v_r, v_i) = V$$

ist nach dem Basisergänzungssatz (mit $s = 1$) entweder $(v_1, \dots, v_{i-1}, w_j, v_{i+1}, \dots, v_r)$ schon eine Basis von V (was wir gerade beweisen wollen), oder es ist $(v_1, \dots, v_{i-1}, w_j, v_{i+1}, \dots, v_r, v_i)$ eine Basis: das ist aber unmöglich, weil w_j eine Linearkombination von $v_1, \dots, v_r$ sein muß.

Damit kommen wir zu einem ersten Hauptresultat über Basen:

18.12 Satz Sind $(v_1, \dots, v_r)$ und $(w_1, \dots, w_s)$ Basen des Vektorraums V, so ist $r = s$.

Beweis Wir bearbeiten die Basis $(v_1, \dots, v_r)$ sukzessive nach dem Austauschlemma und erhalten eine Kette von Basen

$$\begin{gathered}(v_1, v_2, v_3, \dots, v_r)\\ (w_{j_1}, v_2, v_3, \dots, v_r)\\ (w_{j_1}, w_{j_2}, v_3, \dots, v_r)\\ \vdots\\ (w_{j_1}, \dots, w_{j_{r-2}}, w_{j_{r-1}}, v_r)\\ (w_{j_1}, \dots, w_{j_{r-2}}, w_{j_{r-1}}, w_{j_r})\end{gathered}$$

Wir enden mit einer Basis, die aus genau r (natürlich verschiedenen) der w_j besteht, so daß zwangsläufig $r \leq s$ gelten muß. Wegen der Symmetrie der Satzaussage ist dann auch $r \geq s$, mithin $r = s$.

18.13 Definition Ein Vektorraum V heißt endlichdimensional, wenn es eine Basis $(v_1, \dots, v_n)$ von V gibt. Die nach Satz 18.12 von der Wahl einer solchen Basis unabhängige Zahl $n \in \mathbb{N}$ heißt dann die Dimension $\dim V$ von V.

18.13½ Notiz Damit der Vektorraum V endlichdimensional ist, genügt es zu wissen, daß es Vektoren $w_1, \dots, w_s \in V$ mit $\mathrm{Lin}(w_1, \dots, w_s) = V$ gibt.

Denn nach dem Basisergänzungssatz 18.10 können wir dann aus $(w_1, \dots, w_s)$ eine Basis auswählen. — Das Vorhandensein der Standardbasis $(e_1, \dots, e_n)$ von K^n begründet sofort die

18.14 Notiz $\dim K^n = n$.

Zum Satz 18.12 beweisen wir noch gleich die

18.15 Folgerung Sei V ein n-dimensionaler Vektorraum, und seien $v_1, \dots, v_r$ Vektoren in V. Dann gilt:

(a) Ist $\mathrm{Lin}(v_1, \dots, v_r) = V$, so ist $r \geq n$.

(b) Ist $(v_1, \dots, v_r)$ ein linear unabhängiges r-tupel, dann ist $r \leq n$.

(Natürlich darf man keine der beiden Aussagen umkehren!)

Beweis (a) folgt direkt daraus, daß man aus den V aufspannenden Vektoren $v_1, \dots, v_r$ nach dem Ergänzungssatz 18.10 eine Basis auswählen kann. Zum Beweis von (b) wählen wir eine Basis $(w_1, \dots, w_n)$ von V; dann ist $\mathrm{Lin}(w_1, \dots, w_n) = V$ und erst recht

$$\mathrm{Lin}(v_1, \dots, v_r, w_1, \dots, w_n) = V.$$

Nach dem Basisergänzungssatz können wir $(v_1, \dots, v_r)$ also durch Hinzunahme gewisser w_j zu einer Basis machen: diese ist aber ein n-tupel, folglich ist $r \leq n$.

Das sollte man sich gut merken: Mit weniger als n Vektoren kann man einen n-dimensionalen Vektorraum nicht aufspannen, andererseits sind $n+1$ Vektoren $v_1, \dots, v_{n+1}$ darin stets linear abhängig (lässige Ausdrucksweise dafür, daß das $(n+1)$-tupel $(v_1, \dots, v_{n+1})$ linear abhängig ist). — Die Folgerung zeigt auch, daß die Vektorräume von Folgen und Funktionen aus den Beispielen 17.5, (2) bis (5) nicht endlichdimensional sind (abgesehen von einigen offensichtlichen Ausnahmefällen), denn in einem solchen Raum findet man leicht linear unabhängige r-tupel zu jedem vorgegebenen $r \in \mathbb{N}$ (Aufgabe 18.2).

Wir haben schon gesehen, wie häufig man Unterräume eines gegebenen Vektorraums zu betrachten hat; deshalb ist das folgende Resultat von naheliegendem Interesse:

18.16 Satz Sei V ein endlichdimensionaler Vektorraum. Dann ist auch jeder Unterraum U von V endlichdimensional, und es gilt

$$\dim U \leq \dim V.$$

Der Fall $\dim U = \dim V$ tritt nur dann ein, wenn $U = V$ ist.

Beweis Ist $(v_1, \dots, v_r)$ ein linear unabhängiges r-Tupel in U (also auch in V), so ist gemäß der Folgerung $r \leq \dim V$. Unter allen solchen r-tupeln können wir also eines mit größtmöglichem r wählen. Dann ist zwangsläufig $\mathrm{Lin}(v_1, \dots, v_r) = U$, denn für jedes $w \in U$ ist $(v_1, \dots, v_r, w)$ linear abhängig, nach Lemma 18.5 also $w \in \mathrm{Lin}(v_1, \dots, v_r)$. Damit wissen wir, daß $(v_1, \dots, v_r)$ eine Basis von U ist, also ist U endlichdimensional mit $\dim U = r \leq \dim V$.

Nach dem Basisergänzungssatz läßt sich $(v_1, \dots, v_r)$ in jedem Fall zu einer Basis von ganz V verlängern. Im Fall $r = \dim V$ bedeutet das aber, daß $(v_1, \dots, v_r)$ schon eine Basis von V ist, und dann folgt

$$U = \mathrm{Lin}(v_1, \dots, v_r) = V.$$

Wir wollen jetzt wieder beliebige Vektorräume betrachten und den Dimensionsbegriff an zwei einfachen neuen Bildungen üben.

18.17 Definition Sind V und W zwei K-Vektorräume, so wird das kartesische Produkt der Mengen V und W durch komponentenweise Addition und Skalarenmultiplikation

$$(v, w) + (v', w') = (v + v', w + w'), \quad \lambda(v, w) = (\lambda v, \lambda w)$$

zu einem K-Vektorraum, den man naheliegenderweise ebenfalls mit $V \times W$ bezeichnet und das kartesische oder direkte Produkt von V und W nennt.

Jeder Vektor $(v, w) \in V \times W$ zerlegt sich auf Wunsch in $(v, w) = (v, 0) + (0, w)$. Ist also $(v_1, \dots, v_n)$ eine Basis von V und $(w_1, \dots, w_p)$ eine Basis von W, dann ist das $(n+p)$-tupel

$$\big((v_1, 0), \dots, (v_n, 0), (0, w_1), \dots, (0, w_p)\big)$$

eine Basis von $V \times W$, das ist klar. Insbesondere gilt

$$\dim(V \times W) = \dim V + \dim W,$$

wenn V und W beide endlichdimensional sind.

In der zweiten Definition gehen wir von einer ganz anderen Situation aus:

18.18 Definition V sei ein Vektorraum; $S, T \subset V$ seien Unterräume. Dann heißt der Unterraum (!)

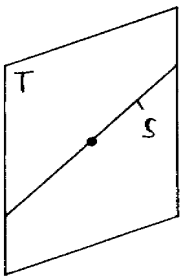
$$S + T := \{x + y \,|\, x \in S,\, y \in T\} \subset V$$

die Summe von S und T.

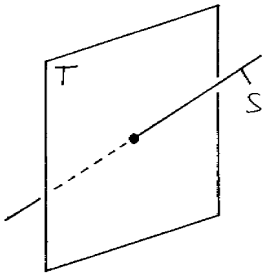
18.19 Beispiele (1) $K^n \times K^p$ kann man in der offensichtlichen Weise mit K^{n+p} identifizieren.

(2) Stets gilt $\mathrm{Lin}(v_1, \ldots, v_r) + \mathrm{Lin}(w_1, \ldots, w_s) = \mathrm{Lin}(v_1, \ldots, v_r, w_1, \ldots, w_s)$

(3) Sind S und T endlichdimensional, so darf man nicht erwarten, daß sich bei der Bildung von $S + T$ die Dimensionen einfach addieren. Es kommt vielmehr auf die Lage der beiden Unterräume zueinander an, wie man den folgenden Bildern mit $V = \mathbb{R}^3$, $\dim S = 1$ und $\dim T = 2$ ansieht.



$S \cap T = S$
$S + T = T$



$S \cap T = \{0\}$
$S + T = \mathbb{R}^3$

Jedoch sind die Dimensionen der vier auftretenden Teilräume nicht unabhängig voneinander, denn es gilt die

18.20 Dimensionsformel für **Unterräume** Sind $S, T \subset V$ endlichdimensionale Unterräume des Vektorraums V, so ist nicht nur $S \cap T$, sondern auch $S + T$ endlichdimensional, und es gilt

$$\dim S + \dim T = \dim(S \cap T) + \dim(S + T).$$

Beweis Daß $S \cap T$ endliche Dimension hat, wissen wir schon aus Satz 18.16. Sei nun $(u_1, \ldots, u_r)$ eine Basis von $S \cap T$. Nach dem Basisergänzungssatz 18.10 können wir diese Basis einerseits zu einer Basis

$$(u_1, \ldots, u_r, v_1, \ldots, v_s)$$

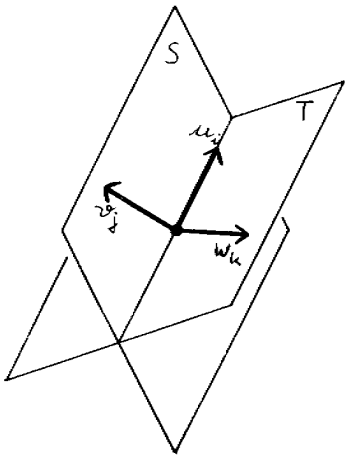
von S und andererseits zu einer Basis

$$(u_1, \ldots, u_r, w_1, \ldots, w_t)$$

von T ergänzen. Ich behaupte, daß

$$(u_1, \ldots, u_r, v_1, \ldots, v_s, w_1, \ldots, w_t)$$

dann eine Basis von $S + T$ ist.



Weil $\mathrm{Lin}(u_1, \dots, u_r, v_1, \dots, v_s, w_1, \dots, w_t) = S + T$ offensichtlich ist, müssen wir nur die lineare Unabhängigkeit beweisen. Wir setzen also

$$0 = \sum_{i=1}^{r} \lambda_i u_i + \sum_{j=1}^{s} \mu_j v_j + \sum_{k=1}^{t} \nu_k w_k$$

mit Skalaren λ_i, μ_j, ν_k an. Die beiden ersten Summen liegen in S — und die dritte damit auch. In T liegt diese aber sowieso, also gilt sogar $\sum_{k=1}^{t} \nu_k w_k \in S \cap T$, und es gibt Skalare $\lambda'_1, \dots, \lambda'_r$ mit

$$\sum_{k=1}^{t} \nu_k w_k = \sum_{i=1}^{r} \lambda'_i u_i.$$

Weil $(u_1, \dots, u_r, w_1, \dots, w_t)$ als Basis von T linear unabhängig ist, müssen alle ν_k (und alle λ'_i) null sein, und von der ursprünglich angesetzten Gleichung bleibt nur

$$0 = \sum_{i=1}^{r} \lambda_i u_i + \sum_{j=1}^{s} \mu_j v_j.$$

Mittels der linearen Unabhängigkeit von $(u_1, \dots, u_r, v_1, \dots, v_s)$ folgt jetzt das Verschwinden auch aller λ_i und μ_j. Damit ist die lineare Unabhängigkeit bewiesen.

Damit haben wir Basen

$(u_1, \dots, u_r)$	für $S \cap T$
$(u_1, \dots, u_r, v_1, \dots, v_s)$	für S
$(u_1, \dots, u_r, w_1, \dots, w_t)$	für T
$(u_1, \dots, u_r, v_1, \dots, v_s, w_1, \dots, w_t)$	für $S + T$

gewonnen, und wir können die Dimensionsformel direkt ablesen: $(r + s) + (r + t) = r + (r + s + t)$.

18.21 Definition V sei ein Vektorraum; $S, T \subset V$ seien Unterräume. Wenn $S \cap T = \{0\}$ ist, sagt man, daß die Summe $S + T$ direkt ist. Wenn außerdem $S + T = V$ gilt, dann nennt man die Teilräume S und T zueinander komplementär, oder S ein Komplement von T in V (und umgekehrt).

Eine solche Situation liegt zum Beispiel vor, wenn $V = S \times T$ das direkte Produkt zweier (abstrakter) Vektorräume S und T ist: die Teilräume

$$S \times \{0\} \subset S \times T \quad \text{und} \quad \{0\} \times T \subset S \times T$$

sind dann zueinander komplementär. In gewisser Weise, nämlich bis auf Isomorphie, ist das auch der typische Fall:

18.22 Lemma V sei ein Vektorraum; $S, T \subset V$ seien Unterräume. Genau dann, wenn S und T zueinander komplementär sind, ist die kanonische Abbildung

$$S \times T \longrightarrow V, \quad (x, y) \mapsto x + y$$

ein Isomorphismus von Vektorräumen.

Beweis Die Abbildung ist natürlich linear, und ihr Bildraum ist $S+T$. Andererseits ist der Kern der Abbildung, also $\{(x, y) \in S \times T \,|\, x+y = 0\}$, offenbar in $(S \cap T) \times (S \cap T)$ enthalten und ergibt sich damit zu

$$\big\{(x, y) \in (S \cap T) \times (S \cap T) \,\big|\, y = -x\big\}.$$

Aus diesen Beschreibungen von Bild und Kern liest man die Aussage direkt ab.

Bemerkungen Das heißt natürlich nicht, daß (S, T) das einzige Paar zueinander komplementärer Teilräume von V wäre; selbst zu einem gegebenen Unterraum $S \subset V$ gibt es im allgemeinen viele verschiedene Komplemente von S in V. — Die Dimensionsformel 18.20 reduziert sich für den Fall, daß die Summe zweier endlichdimensionaler Unterräume $S, T \subset V$ direkt ist, auf $\dim(S + T) = \dim S + \dim T$, und wenn T sogar ein Komplement von S in V ist, kann man natürlich auch $\dim S + \dim T = \dim V$ schreiben.

18.23 Beispiel Der Vektorraum $C^0(\mathbb{R})$ der stetigen Funktionen auf $\mathbb{R}$ enthält die beiden Untervektorräume

$$S = \big\{f \in C^0(\mathbb{R}) \,\big|\, f(-t) = f(t) \text{ für alle } t \in \mathbb{R}\big\}$$
$$T = \big\{f \in C^0(\mathbb{R}) \,\big|\, f(-t) = -f(t) \text{ für alle } t \in \mathbb{R}\big\}$$

der geraden bzw. ungeraden Funktionen. Nur die Nullfunktion ist zugleich gerade und ungerade: $S \cap T = \{0\}$. Andererseits läßt sich jede stetige Funktion f aufgrund der Identität

$$f(t) = \frac{f(t)+f(-t)}{2} + \frac{f(t)-f(-t)}{2} \;\in\; S + T$$

in einen (ebenfalls stetigen) geraden und einen ungeraden Anteil zerlegen. Also ist $C^0(\mathbb{R})$ direkte Summe der beiden Unterräume S und T.

Das Beispiel erläutert auch den Unterschied zum direkten Produkt: Zwar *entspricht* jede Funktion f aus $S + T = C^0(\mathbb{R})$ einem ganz bestimmten Paar aus einer geraden und einer ungeraden Funktion, aber man kann deswegen ja nicht sagen, daß f ein solches Paar *sei*.

Übungsaufgaben

18.1 V sei ein Vektorraum, und $v_1, v_2, \ldots, v_n \in V$ seien Vektoren. Beweisen Sie, daß die folgenden drei Aussagen äquivalent sind:

(a) $(v_1, \ldots, v_n)$ ist eine Basis von V

(b) $\mathrm{Lin}(v_1, \ldots, v_n) = V$, aber für jedes $j \in \{1, \ldots, n\}$ ist $\mathrm{Lin}(v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_n) \neq V$

(c) $(v_1, \ldots, v_n)$ ist linear unabhängig, aber für jeden Vektor $w \in V$ ist $(v_1, \ldots, v_n, w)$ linear abhängig

18.2 Konstruieren Sie im Vektorraum $V := C^0(\mathbb{R}, \mathbb{R})$ der stetigen Funktionen linear unabhängige r-tupel von vorgegebener Länge $r \in \mathbb{N}$:

- Warum ist das $(r+1)$-tupel reeller Polynome

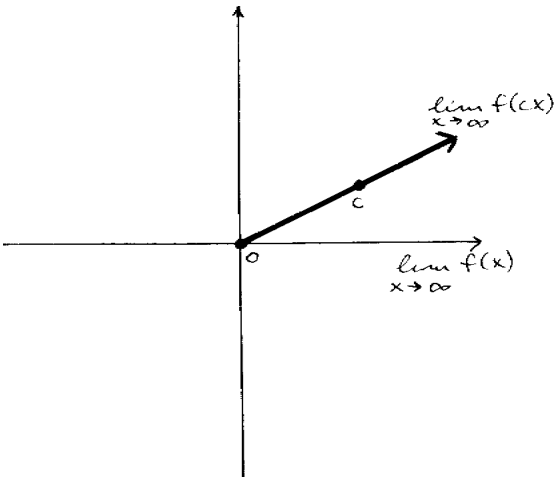
$$(1, X, X^2, \ldots, X^r)$$

linear unabhängig? (Hier ist die Schreibweise aus Beispiel 17.5(5) verwendet, mit X^j also das Polynom $\mathbb{R} \ni x \mapsto x^j \in \mathbb{R}$ gemeint.) Was bedeutet übrigens ganz allgemein die lineare Unabhängigkeit eines 1-tupels, also eines einzelnen Vektors?

- Beweisen Sie allgemeiner: Für jedes $j \in \mathbb{N}$ sei $f_j(X) \in \mathbb{R}[X]$ ein Polynom vom Grad j; dann ist das $(r+1)$-tupel $(f_0, f_1, \ldots, f_r)$ in V linear unabhängig. Tip: Wenn man vollständige Induktion nach r verwendet, hat man dabei praktisch nichts zu rechnen.

18.3 Es seien $c_1, \ldots, c_r$ paarweise verschiedene reelle Zahlen, und $f_j \in C^0(\mathbb{R}, \mathbb{R})$ durch $f_j(x) = e^{c_j x}$ definiert $(j = 1, \ldots, r)$. Beweisen Sie, daß $(f_1, \ldots, f_r)$ linear unabhängig ist. Tip: Aus $0 = \sum_{j=1}^{r} \lambda_j f_j$ muß man ja $\lambda_j = 0$ für alle j folgern. Wenn man die c_j der Größe nach geordnet hat, sieht man immerhin $\lambda_r = 0$, indem man das Verhalten der Funktionswerte $f_j(x)$ für $x \to \infty$ vergleicht.

Wer Lust hat, kann sich das entsprechende Resultat auch für den Fall paarweise verschiedener *komplexer* c_j überlegen; die Funktionen $f_j: z \mapsto e^{c_j z}$ leben dann natürlich in einem komplexen Vektorraum V, etwa $V = C^0(\mathbb{C}, \mathbb{C})$. Den reellen Beweis kann man dann imitieren, indem man statt Grenzwerten für $x \to \infty$ solche betrachtet, die längs eines klug gewählten Strahls in der komplexen Ebene gebildet sind:



Wem das zu kompliziert ist, der mag sich immer noch mit dem Fall rein imaginärer c_j zufrieden geben. In der Notation

$$f_j: t \mapsto e^{i\omega_j t}$$

erkennt man leicht dessen physikalische Bedeutung: (Endlich viele) Sinusschwingungen mit paarweise verschiedenen Frequenzen können sich nicht durch Überlagerung gegenseitig auslöschen.

18.4 V und W seien K-Vektorräume (nicht unbedingt endlicher Dimension),

$$\mathrm{pr}_1: V \times W \longrightarrow V$$

bezeichne wie üblich die Projektion auf den ersten Faktor, und $U \subset V \times W$ sei ein Unterraum des kartesischen Produkts. Beweisen Sie, daß folgende Aussagen äquivalent sind:

- Die Einschränkung $p := \mathrm{pr}_1 | U$ ist ein Isomorphismus von Vektorräumen.
- U und $\{0\} \times W$ sind komplementäre Teilräume von $V \times W$.
- U ist der Graph einer linearen Abbildung $f: V \longrightarrow W$.

19 Karten und Matrizen

Basen ermöglichen es, in endlichdimensionalen Vektorräumen explizite Rechnungen durchzuführen — ich erkläre jetzt, wie.

19.1 Definition Sei V ein K-Vektorraum und $\underline{v} = (v_1, \dots, v_n)$ eine Basis von V. Die in der Definition 18.1 eingeführte lineare Abbildung $\Phi_{\underline{v}} = \Phi_{(v_1,\dots,v_n)} \colon K^n \longrightarrow V$ mit $\Phi_{\underline{v}}(\lambda) = \sum_{i=1}^n \lambda_i v_i$ ist bijektiv, das haben wir in 18.6 bemerkt. Sie heißt der zu $\underline{v}$ gehörige Basisisomorphismus, und den inversen Isomorphismus

$$\Phi_{\underline{v}}^{-1} \colon V \longrightarrow K^n$$

nennen wir die durch $\underline{v}$ bestimmte (lineare) Karte von V.

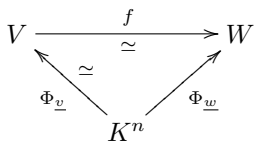
Von den beiden grundsätzlich gleichwertigen Objekten $\Phi_{\underline{v}}$ und $\Phi_{\underline{v}}^{-1}$ hat der Basisisomorphismus $\Phi_{\underline{v}}$ den technischen Vorteil, daß seine Wirkung $\lambda \mapsto \sum \lambda_j v_j$ sich direkt hinzuschreiben ist. Die Karte $\Phi_{\underline{v}}^{-1}$ gibt aber vielleicht eine bessere Vorstellung von ihrem Verwendungszweck: So wie eine Landkarte die unübersichtlichen Verhältnisse einer Landschaft auf ein überschaubares Stück Papier abbildet, so bildet die zu $\underline{v}$ gehörige Karte den abstrakten Vektorraum V auf den konkreten Spaltenraum K^n ab, in dem man unmittelbar rechnen kann: Jedem abstrakten Vektor wird durch $\Phi_{\underline{v}}^{-1}$ ja ein konkretes n-tupel von Skalaren zugordnet, das diesen Vektor charakterisiert. Natürlich ist die physikalische "Raum-Zeit-Welt" ein Beispiel dafür: Unter der Annahme, daß diese (nach Wahl eines Nullpunktes) ein vierdimensionaler reeller Vektorraum ist, bestimmt die Wahl einer Basis $\underline{v}$ dieses Vektorraumes eine Karte, in der jeder Weltpunkt durch vier reelle Zahlen repräsentiert ist. In physikalischer Sprache sind das die Orts- und Zeitkoordinaten des Weltpunkts, und überhaupt sagen die Physiker statt Karte lieber Bezugs- oder Koordinatensystem. "Karte" ist aber so schön kurz! Der Zusatz "linear" ist übrigens nur nötig, wenn man auf den Unterschied zu allgemeineren, noch zu besprechenden Kartentypen aufmerksam machen will.

19.2 Satz Zwei endlichdimensionale K-Vektorräume sind genau dann zueinander isomorph, wenn sie dieselbe Dimension haben.

Beweis Seien V, W zwei K-Vektorräume derselben Dimension $n \in \mathbb{N}$. Die Wahl von Basen $\underline{v}$ für V und $\underline{w}$ für W liefert gemäß der Definition 18.6 einen Isomorphismus

$$V \xrightarrow{\Phi_{\underline{v}}^{-1}} K^n \xrightarrow{\Phi_{\underline{w}}} W.$$

Umgekehrt seien endlichdimensionale K-Vektorräume V, W und ein Isomorphismus $f \colon V \longrightarrow W$ gegeben. Wir wählen eine Basis $\underline{v} = (v_1, \dots, v_n)$ für V und definieren das n-tupel $\underline{w} = (w_1, \dots, w_n)$ durch $w_j := f(v_j)$ für $j = 1, \dots, n$. Das Diagramm von Vektorräumen und linearen Abbildungen



ist dann kommutativ:

$$(f \circ \Phi_{\underline{v}})(\lambda) = f\left(\sum \lambda_j v_j\right) = \sum \lambda_j f(v_j) = \sum \lambda_j w_j = \Phi_{\underline{w}}(\lambda).$$

Insbesondere ist auch $\Phi_{\underline{w}}$ ein Isomorphismus, also $\underline{w}$ eine Basis von W gemäß Definition 18.6. Es folgt $\dim V = n = \dim W$.

Bemerkungen Der zweite Beweisteil zeigt allgemein, daß ein Isomorphismus Basen auf Basen abbildet, daß deshalb von zwei isomorphen Vektorräumen entweder beide endlichdimensional sind oder keiner. — Es gilt kein 19.2 entsprechender Satz für Vektorräume, die nicht endlichdimensional sind. Genauso, wie es bei Mengen verschiedene Arten der Unendlichkeit gibt (siehe Abschnitt 6), gibt es bei Vektorräumen unüberschaubar viele Arten der "Unendlichdimensionalität". Aus der (häufig anzutreffenden) nachlässigen Formulierung $\dim V = \infty = \dim W$ läßt sich deshalb nicht auf Isomorphie noch auf sonst eine Beziehung zwischen V und W schließen, außer daß eben weder V noch W eine Basis besitzt. — Über endlichdimensionale Vektorräume aber gibt uns Satz 19.2 eine sehr befriedigende Auskunft: Nach dem Prinzip, daß wir einen Vektorraum kennen, wenn wir einen Isomorphismus zu einem schon bekannten angeben können, kennen wir mit den Spaltenräumen K^n ($n \in \mathbb{N}$) grundsätzlich alle endlichdimensionalen K-Vektorräume.

Wir wollen die Karten jetzt benutzen, um lineare Abbildungen zwischen endlichdimensionalen Vektorräumen rechnerisch in den Griff zu bekommen, nämlich durch Matrizen zu beschreiben. Wie wir schon wissen, liefert jede $p \times n$-Matrix a über K eine lineare Abbildung $K^n \ni x \mapsto ax \in K^p$. Tatsächlich entsteht jede lineare Abbildung von K^n nach K^p auf diese Weise aus einer eindeutig bestimmten Matrix:

19.3 Satz Seien $n, p \in \mathbb{N}$ beliebig. Dann gibt es zu jeder linearen Abbildung $f\colon \mathbb{K}^n \longrightarrow \mathbb{K}^p$ genau eine Matrix $a \in \mathrm{Mat}(p \times n, K)$ mit

$$f(x) = ax \quad \text{für alle } x \in K^n.$$

Beweis Wir benutzen die Standardbasis $(e_1, \dots, e_n)$ von K^n und machen erst mal die fundamentale Beobachtung, daß für jede Matrix $a \in \mathrm{Mat}(p \times n, K)$ und jedes $j \in \{1, \dots, n\}$ der Vektor $ae_j \in K^p$ gerade die j-te Spalte von a ist:

$$\begin{pmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1n} \\ \vdots & \dots & \vdots & \dots & \vdots \\ a_{p1} & \dots & a_{pj} & \dots & a_{pn} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{pj} \end{pmatrix}$$

Um bei gegebenem f die Gleichung $f(x) = ax$ für alle $x \in K^n$ zu erfüllen, bleibt uns also gar nichts übrig, als die aus den Spalten $f(e_j) \in K^p$ gebildete Matrix a zu nehmen:

$$a = \begin{pmatrix} f(e_1) & f(e_2) & \dots & f(e_n) \end{pmatrix} \in \mathrm{Mat}(p \times n, K)$$

Das zeigt die Eindeutigkeit von a. Zum Existenzbeweis rechnen wir nun noch nach, daß bei dieser Wahl von a die Identität $f(x) = ax$ nicht nur für $x = e_1, \dots, e_n$, sondern für alle $x \in K^n$ gilt. Dazu schreiben wir

$$x = \sum_{j=1}^{n} x_j e_j$$

und erhalten, weil f und $x \mapsto ax$ linear sind, in der Tat

$$f(x) = \sum_{j=1}^{n} x_j f(e_j) = \sum_{j=1}^{n} x_j (ae_j) = a \sum_{j=1}^{n} x_j e_j = ax.$$

Weil der Zusammenhang zwischen den Matrizen $a \in \mathrm{Mat}(p \times n, K)$ und den linearen Abbildungen $K^n \xrightarrow{f} K^p$ völlig kanonisch ist, begehen wir keine große Sünde, wenn wir künftig die Matrix a mit der linearen Abbildung $x \mapsto ax$ identifizieren, also geradezu sagen: Die $p \times n$-Matrizen über K *sind* die linearen Abbildungen von K^n nach K^p. Was den Zusammenhang selber betrifft, sollte man sich den im Beweis von 19.3 festgestellten Sachverhalt dauerhaft einprägen:

19.4 Merkregel Die Spalten von $a \in \mathrm{Mat}(p \times n, K)$ sind die Bilder der (Standard-)Basisvektoren unter der linaren Abbildung $K^n \xrightarrow{a} K^p$.

Seien jetzt V ein n-dimensionaler und W ein p-dimensionaler K-Vektorraum, sowie $f\colon V \longrightarrow W$ eine lineare Abbildung. Natürlich können wir nicht einfach sagen, f sei eine Matrix: daß V zu K^n und W zu K^p isomorph ist, heißt ja nicht, daß V und W diese Spaltenräume *sind*. Wir können f aber durch eine Matrix beschreiben, sobald wir in V und W Basen gewählt haben. Sei nämlich $\underline{v} = (v_1, \dots, v_n)$ eine Basis von V, und $\underline{w} = (w_1, \dots, w_p)$ eine solche von W. Dann ist die Komposition

$$a\colon K^n \xrightarrow{\Phi_{\underline{v}}} V \xrightarrow{f} W \xrightarrow{\Phi_{\underline{w}}^{-1}} K^p$$

eine lineare Abbildung von K^n nach K^p, und wir wollten uns ja erlauben, diese Abbildung als eine Matrix $a \in \mathrm{Mat}(p\times n, K)$ anzusehen. Man redet von diesem a als der *Matrix der linearen Abbildung f bezüglich der Basen $\underline{v}$ und $\underline{w}$*, oder der linearen Abbildung f *geschrieben in den zu $\underline{v}$ und $\underline{w}$ gehörigen Karten*. Bei gegebenen Basen $\underline{v}$ und $\underline{w}$ bestimmt umgekehrt jede Matrix a eine lineare Abbildung f, nämlich die Komposition

$$f\colon V \xrightarrow{\Phi_{\underline{v}}^{-1}} K^n \xrightarrow{a} K^p \xrightarrow{\Phi_{\underline{w}}} W.$$

Dieser Zusammenhang zwischen f und a wird in jedem Fall unübertroffen klar durch die Forderung beschrieben, daß das Diagramm von K-Vektorräumen und linearen Abbildungen

$$\begin{array}{ccc} V & \xrightarrow{f} & W \\ \Phi_{\underline{v}} \uparrow \simeq & & \Phi_{\underline{w}} \uparrow \simeq \\ K^n & \xrightarrow{a} & K^p \end{array}$$

kommutieren soll. Dieses einfache Diagramm sollte man bei jeder theoretischen oder praktischen Frage, die mit der Beschreibung linearer Abbildungen durch Matrizen zu tun hat, zu Hilfe zu ziehen.

Wir wollen an dem Diagramm das Schicksal eines Basisvektors e_j verfolgen:

$$\begin{array}{ccc} v_j & \xmapsto{f} & f(v_j) \\ \Phi_{\underline{v}} \uparrow & & \Phi_{\underline{w}} \uparrow \\ e_j & \xmapsto{a} & ae_j \end{array}$$

Weil ae_j die j-te Spalte von a ist, sehen wir, daß die Merkregel 19.4 auch in dieser Situation im wesentlichen gültig bleibt: Die Spalten von a sind die Bilder der in V gewählten Basisvektoren, allerdings ausgedrückt in der zu $\underline{w}$ gehörigen Karte $\Phi_{\underline{w}}^{-1}$.

Wenn man nun speziell $V = K^n$ und $W = K^p$ nimmt, also f selbst schon eine Matrix ist? Selbstverständlich ist das zulässig. Hat man sich außerdem für die Standardbasen als $\underline{v}$ und $\underline{w}$ entschieden, so sind $\Phi_{\underline{v}}$ und $\Phi_{\underline{v}}$ die identische Abbildung von K^n bzw. K^p, und die f in den zugehörigen Karten beschreibende Matrix a ist eben f selbst. Sind dagegen $\underline{v}$ und/oder $\underline{w}$ andere Basen, braucht das nicht mehr der Fall zu sein; wie man zwischen f und a wechselt, muß man dann wie immer dem kommutativen Diagramm

$$\begin{array}{ccc} K^n & \xrightarrow{f} & K^p \\ \Phi_{\underline{v}} \uparrow \simeq & & \Phi_{\underline{w}} \uparrow \simeq \\ K^n & \xrightarrow{a} & K^p \end{array}$$

entnehmen, das jetzt in dem Sinne konkreter ist, als *alle* Pfeile für Matrizen stehen.

Das bringt die Frage auf, wie man überhaupt mit Matrizen rechnet. Erst mal liegt auf der Hand, daß man Matrizen gleichen Formats komponentenweise addieren oder auch mit einem Skalar multiplizieren kann. Diese beiden Operationen machen $\mathrm{Mat}(p\times n, K)$ für jede feste Wahl von n, $p \in \mathbb{N}$ zu einem Vektorraum. Auch dieser Vektorraum hat eine naheliegende Standardbasis, bestehend aus den pn Matrizen e_{kl} für $k \in \{1,\dots,p\}$ und $l \in \{1,\dots,n\}$, die einen einzigen Eintrag 1 und als übrige Einträge lauter Nullen haben:

$$e_{kl} = \begin{pmatrix} & & \\ & & \\ & 1 & \\ & & \end{pmatrix} \leftarrow k$$
$$\uparrow$$
$$l$$

(Bei Matrizen mit vielen Nulleinträgen ist es oft übersichtlicher, diese gar nicht hinzuschreiben.) Formal gesehen müßte man die e_{kl} eigentlich noch zu einem pn-tupel anordnen, um dieses als Basis von $\mathrm{Mat}(p\times n, K)$ ansprechen zu dürfen. Das vermeidet man aber möglichst, weil in der Wahl der Anordnung ja eine Willkür liegt. Wie auch immer — klar ist, daß $\mathrm{Mat}(p\times n, K)$ als K-Vektorraum die Dimension pn hat.

Wenn Sie es vorziehen, die Basismatrizen raumpsparend durch Angabe ihrer Komponenten zu beschreiben, ist das traditionelle durch

$$\delta_{ij} = \begin{cases} 1 & \text{für } i = j \\ 0 & \text{sonst} \end{cases}$$

definierte *Kroneckersymbol* δ_{ij} praktisch: Sie schreiben dann einfach

$$(e_{kl})_{ij} = \delta_{ik} \cdot \delta_{jl}\,.$$

Da man Matrizen als lineare Abbildungen interpretieren kann, wird man auch auf der Menge der linearen Homomorphismen zwischen zwei gegebenen K-Vektorräumen V und W eine Vektorraumstruktur vermuten. Diese ist auch leicht gefunden: Die üblicherweise mit

$$\mathrm{Hom}(V,W) := \{f\colon V \longrightarrow W \mid f \text{ ist linear}\}$$

bezeichnete Menge wird durch punktweise Addition und Skalarenmultiplikation, also

$$\left.\begin{aligned} (f+g)(v) &:= f(v) + g(v) \\ (\lambda f)(v) &:= \lambda f(v) \end{aligned}\right\} \text{ für } f,g \in V \text{ und } \lambda \in K$$

selbst zu einem K-Vektorraum, wie man ohne Schwierigkeiten nachprüft. Nicht überraschen dürfte nun das folgende Resultat, das zugleich die vorangehende Diskussion noch einmal zusammenfaßt:

19.5 Satz *V und W seien K-Vektorräume mit $\dim V = n$, $\dim W = p$. Für jede Wahl von Basen $\underline{v}$ und $\underline{w}$ ist die Abbildung*

$$\mathrm{Hom}(V,W) \longrightarrow \mathrm{Mat}(p\times n, K),$$

die der linearen Abbildung f ihre Matrix bezüglich $\underline{v}$ und $\underline{w}$ zuordnet, ein linearer Isomorphismus. Insbesondere gilt

$$\dim \mathrm{Hom}(V,W) = \dim V \cdot \dim W.$$

Beweis Daß diese Abbildung bijektiv ist, wissen wir schon, und nur die Linearität ist noch nachzurechnen: Seien $f, g \in \mathrm{Hom}(V,W)$ Homomorphismen mit den zugehörigen Matrizen $a, b \in \mathrm{Mat}(p\times n, K)$ bezüglich $\underline{v} = (v_1,\dots,v_n)$ und $\underline{w} = (w_1,\dots,w_p)$; es ist also

$$f(v_j) = \sum_{i=1}^{p} a_{ij} w_i \quad \text{und} \quad g(v_j) = \sum_{i=1}^{p} b_{ij} w_i$$

für $j = 1, \dots, n$. Schlichtes Addieren gibt

$$(f+g)(v_j) = f(v_j) + g(v_j) = \sum_{i=1}^{p} (a_{ij} + b_{ij}) w_i \,;$$

zu $f+g$ gehört also die Matrix $a+b$. Entsprechend für die skalaren Vielfachen λf.

19.6 Satz V und W seien Vektorräume, $(v_1, \dots, v_n)$ eine Basis von V. Für jede Wahl von Vektoren $w_1, \dots, w_n \in W$ gibt es dann genau eine lineare Abbildung $f\colon V \longrightarrow W$ mit

$$f(v_j) = w_j \quad \text{für } j = 1, \dots, n.$$

Beweis f muß allgemeiner $f\left(\sum \lambda_j v_j\right) = \sum_j \lambda_j w_j$ erfüllen, deshalb ist

$$f := \Phi_{(w_1,\dots,w_n)} \circ \Phi_{\underline{v}}^{-1}$$

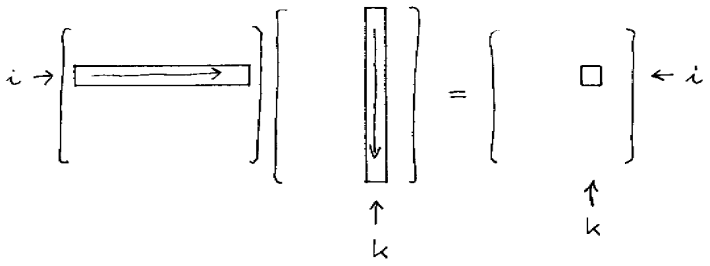
die eindeutig bestimmte Lösung des Problems. Alternativ läßt sich dieser Satz als eine witzige Folgerung aus Satz 19.5. auffassen, zumindest unter der zusätzlichen Annahme, daß neben V auch W endlichdimensional ist. Wenn wir nämlich auch in W eine Basis wählen, übersetzt 19.5 die Behauptung in eine, die völlig trivial ist: Zu je n Vektoren $a_1, \dots, a_n \in K^p$ gibt es genau eine Matrix $a \in \mathrm{Mat}(p \times n, K)$ mit den Spalten $a_1, \dots, a_n$.

Die folgende Definition verallgemeinert die schon im Abschnitt 17 eingeführte Multiplikation einer Matrix mit einer Spalte.

19.7 Definition n, p und q seien natürliche Zahlen. Das Produkt zweier Matrizen $b \in \mathrm{Mat}(q \times p, K)$ und $a \in \mathrm{Mat}(p \times n, K)$ ist die durch

$$(ba)_{ik} := \sum_{j=1}^{p} b_{ij} a_{jk}$$

erklärte Matrix $ba \in \mathrm{Mat}(q \times n, K)$.



Zum Beispiel gilt für die Matrizen aus der Standardbasis (passenden Formats)

$$e_{kl} e_{mn} = \begin{cases} e_{kn} & \text{wenn } l = m, \\ 0 & \text{sonst.} \end{cases}$$

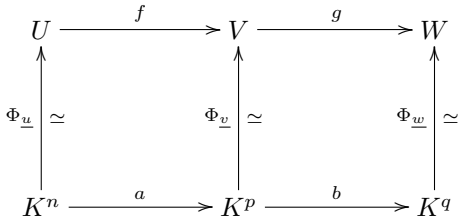
Das sieht man den Matrizen mit etwas Übung direkt an; sonst rechnet man tapfer:

$$(e_{kl} e_{mn})_{\alpha\gamma} = \sum_{\beta} (e_{kl})_{\alpha\beta} (e_{mn})_{\beta\gamma} = \sum_{\beta} \delta_{k\alpha} \delta_{l\beta} \delta_{m\beta} \delta_{n\gamma} = \delta_{k\alpha} \delta_{lm} \delta_{n\gamma} = \delta_{lm} \cdot (e_{kn})_{\alpha\gamma}$$

Multiplizieren kann man jede p-spaltige Matrix mit jeder p-zeiligen — ganz anders als bei der Matrizenaddition kommt es nicht etwa auf die Gleichheit der Matrizenformate an! Die Bedeutung dieser Multiplikation ergibt sich aus dem

19.8 Lemma Seien U, V, W, drei K-Vektorräume mit Basen $\underline{u}, \underline{v}, \underline{w}$. Die linearen Abbildungen $f\colon U \longrightarrow V$ und $g\colon V \longrightarrow W$ seien bezüglich dieser Basen durch die Matrizen $a \in \mathrm{Mat}(p \times n, K)$ und $b \in \mathrm{Mat}(q \times p, K)$ dargestellt. Dann repräsentiert das Matrizenprodukt ba die Komposition $g \circ f$ bezüglich der Basen $\underline{u}$ und $\underline{w}$.

Beweis Die beiden zur Situation gehörigen kommutativen Quadrate fügen sich zu dem Diagramm



zusammen, das natürlich ebenfalls kommutiert. Deshalb bleibt bloß zu verifizieren, daß das Matrizenprodukt $ba \in \mathrm{Mat}(q\times n, K)$ mit der Komposition $b\circ a\colon K^n \xrightarrow{a} K^p \xrightarrow{b} K^q$ übereinstimmt:

$$\big((ba)x\big)_i = \sum_k (ba)_{ik}x_k = \sum_{j,k} b_{ij}a_{jk}x_k,$$

und

$$\big(b(ax)\big)_i = \sum_j b_{ij}(ax)_j = \sum_{j,k} b_{ij}a_{jk}x_k$$

gibt dasselbe.

Beachten Sie auch den Fall, daß U, V, W Spaltenräume und $\underline{u}, \underline{v}, \underline{w}$ deren Standardbasen sind: Die repräsentierenden Matrizen sind dann einfach die linearen Abbildungen selbst, und die für diese gültigen Regeln, insbesondere die Tatsache, daß die Komposition assoziativ ist, übertragen sich unmittelbar auf Matrizen:

19.9 Notiz Für alle Matrizen a, b, c gilt, soweit die man die betrachteten Ausdrücke überhaupt bilden kann:

- $(ab)c = a(bc)$ (Assoziativgesetz)
- $(a+b)c = ac + bc$ und $a(b+c) = ab + ac$ (Distributivgesetze)
- $\lambda(ab) = (\lambda a)b = a(\lambda b)$ für jeden Skalar λ

Für jedes $p \in \mathbb{N}$ gibt es die sogenannte $p \times p$-Einheitsmatrix

$$1 = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix} \in \mathrm{Mat}(p\times p, K)$$

mit den Einträgen

$$1_{ij} = \delta_{ij} = \begin{cases} 1 & \text{für } i = j \\ 0 & \text{für } i \neq j. \end{cases}$$

— es tut nicht weh, die systematische Bezeichnung 1_{ij} und das Kroneckersymbol δ_{ij} als Alternativen nebeneinander zuzulassen. Als lineare Abbildung $K^p \longrightarrow K^p$ ist die Einheitsmatrix die identische Abbildung, und deshalb gilt

$$1a = a \quad \text{für alle } a \in \mathrm{Mat}(p\times n, K)$$
$$b1 = b \quad \text{für alle } b \in \mathrm{Mat}(q\times p, K),$$

wobei $n, q \in \mathbb{N}$ ganz beliebig sind.

Da die Matrizenmultiplikation nicht zwischen Matrizen gleichen Formates erfolgt, gibt es keinen Sinn, danach zu fragen, ob $\mathrm{Mat}(p\times n, K)$ ein Ring ist, außer im Fall $p = n$, auf den wir später noch zu sprechen kommen. Das Matrizenprodukt ist im allgemeinen nicht kommutativ:

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

aber

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = 0 \in \mathrm{Mat}(2\times 2, K),$$

und Sie sehen zugleich, daß das Produkt zweier von null verschiedener Matrizen durchaus die Nullmatrix sein kann. Deshalb kann es auch nicht zu jeder Matrix $a \neq 0$ eine Inverse geben: eine Matrix a^{-1} mit $aa^{-1} = 1$ und $a^{-1}a = 1$.

19.10 Definition Eine Matrix $u \in \mathrm{Mat}(p\times n, K)$ heißt umkehrbar oder invertierbar, wenn es eine Matrix $v \in \mathrm{Mat}(n\times p, K)$ mit $uv = 1 \in \mathrm{Mat}(p\times p, K)$ und $vu = 1 \in \mathrm{Mat}(n\times n, K)$ gibt. Diese heißt dann die zu u inverse Matrix u^{-1}.

Unmittelbar klar und erhellend ist die

19.11 Notiz $u \in \mathrm{Mat}(p\times n, K)$ ist genau dann invertierbar, wenn u, aufgefaßt als lineare Abbildung $K^n \xrightarrow{u} K^p$, ein Isomorphismus ist, und $u^{-1}\colon K^p \longrightarrow K^n$ ist dann die Umkehrabbildung.

19.12 Lemma und **Definition** (a) Ist $u \in \mathrm{Mat}(p\times n, K)$ invertierbar, so ist $p = n$: Nur quadratische Matrizen können invertierbar sein.

(b) Für festes $n \in \mathbb{N}$ macht die Matrizenmultiplikation die Menge

$$GL(n, K) := \{u \in \mathrm{Mat}(n\times n, K) \mid u \text{ invertierbar}\}$$

zu einer (für $n > 1$ nicht abelschen) Gruppe, die man allgemeine lineare Gruppe nennt (General Linear; alternative Schreibweise: $GL_n(K)$). Allgemeiner hat man für jeden Vektorraum V die Gruppe

$$GL(V) := \{g\colon V \longrightarrow V \mid g \text{ linearer Isomorphismus}\}$$

mit der Komposition als Verknüpfung, die allgemeine lineare Gruppe von V. Im Fall $\dim V = n$ liefert die Wahl einer Basis von V in der bekannten Weise einen Gruppenisomorphismus $GL(V) \simeq GL(n, K)$.

Beweis (a) folgt daraus, daß isomorphe Vektorräume nach Satz 19.2 dieselbe Dimension haben. Zum Beweis von (b) ist nur zu bemerken, daß mit $g, h \in GL(V)$ auch $g \circ h$ und g^{-1} Isomorphismen sind.

Bemerkung Beachten Sie, daß wegen $(g \circ h)^{-1} = h^{-1} \circ g^{-1}$ auch in $GL(n, K)$ (wie überhaupt in jeder Gruppe) $(uv)^{-1} = v^{-1}u^{-1}$ und im allgemeinen nicht $(uv)^{-1} = u^{-1}v^{-1}$ gilt.

Einer gegebenen quadratischen Matrix a ansehen, ob sie invertierbar ist — nun, das können wir im Augenblick noch nicht, und schon gar nicht können wir a^{-1} dann berechnen. Stattdessen wollen wir einige Beispiele invertierbarer Matrizen studieren, die sich bald als besonders wichtig erweisen werden.

19.13 Beispiele und **Definitionen** Sei K ein Körper, und sei $n \in \mathbb{N}$. Die im folgenden erklärten Matrizen p_{kl}, $d_{k\lambda}$ und $u_{kl\lambda}$ in $GL(n, K)$ heißen Elementarmatrizen (über K).

(1) Seien $k, l \in \{1, \dots, n\}$ zwei verschiedene Indizes: $k \neq l$. Die Matrix

$$p_{kl} := \begin{pmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & 0 & \cdots & 1 & & & \leftarrow k \\ & & & \vdots & 1 & \vdots & & & \\ & & & \vdots & \ddots & \vdots & & & \\ & & & 1 & \cdots & 0 & & & \leftarrow l \\ & & & & & & 1 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 1 \end{pmatrix}$$

$$\begin{matrix} \uparrow & & \uparrow \\ k & & l \end{matrix}$$

in $\mathrm{Mat}(n\times n, K)$, deren Einträge also durch

$$(p_{kl})_{ij} = \begin{cases} 1 & \text{falls } k \neq i = j \neq l \\ 1 & \text{falls } \{i,j\} = \{k,l\} \\ 0 & \text{sonst} \end{cases}$$

gegeben sind, ist sicher invertierbar, denn es gilt

$$(p_{kl})^2 = \begin{pmatrix} \ddots & & & & & & \\ & 0 & \cdots & & 1 & & \\ & \vdots & 1 & & \vdots & & \\ & \vdots & & \ddots & \vdots & & \\ & \vdots & & & 1 & & \\ & 1 & \cdots & & 0 & & \\ & & & & & \ddots & \end{pmatrix} \begin{pmatrix} \ddots & & & & & & \\ & 0 & \cdots & & 1 & & \\ & \vdots & 1 & & \vdots & & \\ & \vdots & & \ddots & \vdots & & \\ & \vdots & & & 1 & & \\ & 1 & \cdots & & 0 & & \\ & & & & & \ddots & \end{pmatrix} = 1$$

und damit $(p_{kl})^{-1} = p_{kl}$. — Daß das Produkt wirklich die Einheitsmatrix gibt, sieht man durch konzentriertes Hinschauen, und man sollte sich schon darin üben, Produkte von Matrizen, deren Einträge notgedrungen teils durch Pünktchen ersetzt sind, so zu berechnen. Wer beim Erklären nicht so viel mit den Armen rudern mag, kann die Rechnung aber auch mit $p_{kl} = 1 - e_{kk} - e_{ll} + e_{kl} + e_{lk}$ ausführen.

(2) Die Matrix

$$d_{k\lambda} := \begin{pmatrix} 1 & & & & & & & \\ & \ddots & & & & & & \\ & & \ddots & & & & & \\ & & & 1 & & & & \\ & & & & \lambda & & & \\ & & & & & 1 & & \\ & & & & & & \ddots & \\ & & & & & & & 1 \end{pmatrix} \leftarrow k$$

ist für $k \in \{1,\dots,n\}$ und $0 \neq \lambda \in K$ invertierbar, denn offensichtlich gilt

$$d_{k\lambda} d_{k\mu} = d_{k,\lambda\mu} \quad \text{für alle } \lambda, \mu \in K\setminus\{0\}$$

und damit $(d_{k\lambda})^{-1} = d_{k,\lambda^{-1}}$ (doppelte Indizes trennt man durch ein Komma, wenn die Lesbarkeit das erfordert). Zum Rechnen kann es praktisch sein, auch diese Matrix (etwas gewaltsam) durch die Basismatrizen auszudrücken:

$$d_{k\lambda} = 1 + (\lambda - 1)e_{kk}$$

(3) Für $k \neq l$ und beliebige $\lambda \in K$ ist auch

$$u_{kl\lambda} := 1 + \lambda e_{kl} = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & \ddots & & \lambda & & \\ & & & \ddots & & & \\ & & & & \ddots & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \leftarrow k$$

$$\begin{matrix} \\ \\ \\ \\ \\ \\ \uparrow \\ l \end{matrix}$$

eine Matrix in $GL(n,K)$; ihre Inverse $(u_{kl\lambda})^{-1} = u_{kl,-\lambda}$ ergibt sich aus der allgemeinen Formel

$$u_{kl\lambda} \cdot u_{kl\mu} = (1+\lambda e_{kl})(1+\mu e_{kl}) = 1 + (\lambda+\mu)e_{kl} = u_{kl,\lambda+\mu}.$$

Gelehrter ausgedrückt besagt diese Formel übrigens, daß für jede Wahl von $k \neq l$ die Abbildung

$$K \ni \lambda \mapsto u_{kl\lambda} \in GL(n,K)$$

ein Homomorphismus von Gruppen ist; genauer ein Homomorphismus zwischen $(K,+)$ und $\big(GL(n,K),\cdot\big)$.

Es ist interessant zu untersuchen, welche Wirkung die Multiplikation mit einer Elementarmatrix auf eine beliebige Matrix $a \in \mathrm{Mat}(p\times n, K)$ hat. Dabei muß man unterscheiden, ob wir mit einer Elementarmatrix $u \in \mathrm{Mat}(p\times p, K)$ von links, oder mit $v \in \mathrm{Mat}(n\times n, K)$ von rechts multiplizieren.

Zunächst, sagen wir, mit p_{kl} von links. Der Übersichtlichkeit halber schreiben wir a in der Form

$$a = \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix} \in \mathrm{Mat}(p\times n, K) \quad \text{mit den Zeilenmatrizen } a_1, \dots, a_p \in \mathrm{Mat}(1\times n, K);$$

dann wird $p_{kl}a$ zu :

$$\begin{array}{r} \\ k \to \\ \\ l \to \\ \\ \\ \\ \end{array} \begin{pmatrix} \ddots & & & & & \\ \dots & 0 & \dots & 1 & \dots & \\ & & \ddots & & & \\ \dots & 1 & \dots & 0 & \dots & \\ & & & & \ddots & \\ & \uparrow & & \uparrow & & \\ & k & & l & & \end{pmatrix} \begin{pmatrix} a_k \\ \\ a_l \\ \\ \\ \end{pmatrix} = \begin{pmatrix} a_l \\ \\ a_k \\ \\ \\ \end{pmatrix} \begin{array}{l} \\ \leftarrow k \\ \\ \leftarrow l \\ \\ \\ \end{array}$$

Wie man sieht, besteht der Effekt darin, daß die k-te Zeile von a mit der l-ten vertauscht wird. (Hier und im folgenden sind die von vornherein nicht betroffenen Zeilen von a der Übersichtlichkeit halber nicht eingetragen; damit ist nicht etwa gemeint, daß das Nullzeilen sein müßten.) Nun zu $d_{k\lambda}$ (von links):

$$\begin{pmatrix} \ddots & & & & \\ & 1 & & & \\ & & \lambda & & \\ & & & 1 & \\ & & & & \ddots \\ & & \uparrow & & \\ & & k & & \end{pmatrix} \begin{pmatrix} a_k \\ \\ \\ \\ \\ \end{pmatrix} = \begin{pmatrix} \lambda a_k \\ \\ \\ \\ \\ \end{pmatrix} \leftarrow k$$

Hier wird also die k-te Zeile von a mit $\lambda \neq 0$ multipliziert. Am interessantesten ist die Wirkung von $u_{kl\lambda}$:

$$k \to \begin{pmatrix} 1 & & & & \\ & \ddots & & \lambda & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & 1 \\ & & & \uparrow & \\ & & & l & \end{pmatrix} \begin{pmatrix} a_k \\ \\ a_l \\ \\ \\ \end{pmatrix} = \begin{pmatrix} a_k + \lambda a_l \\ \\ a_l \\ \\ \\ \end{pmatrix} \leftarrow k$$

Diesmal wird zur k-ten Zeile von a das λ-fache der l-ten hinzuaddiert, während die l-te selbst ebenso unverändert bleibt wie alle übrigen.

19.14 Sprechweise Die Multiplikation einer Matrix a mit einer Elementarmatrix von links nennt man — weil dabei jede Zeile von a als Ganzes verändert wird — eine elementare Zeilenumformung (nicht das Resultat,

sondern der Vorgang heißt so). Analog spricht man beim Multiplizieren mit einer Elementarmatrix von rechts von einer elementaren Spaltenumformung; diese hat natürlich die entsprechende Wirkung auf die Spalten von a. Beide Arten von Umformungen kann man auch wiederholt (mit wechselnden Elementarmatrizen) durchführen; wenn man das meint, läßt man den Zusatz "elementar" weg.

Bemerkung Multipliziert man mit $u_{kl\lambda}$ von rechts, so sind auch die Rollen von k und l vertauscht: In $au_{kl\lambda}$ ist zur l-ten Spalte von a das λ-fache der k-ten addiert, nicht umgekehrt.

Unter Zeilen- und Spaltenumformungen kann eine Matrix sich drastisch verändern, zum Beispiel sehen Sie hier eine Folge von zwei elementaren Zeilen- und zwei darauffolgenden elementaren Spaltenumformungen:

$$\begin{pmatrix} 1 & 2 & 3 \\ -2 & -4 & -6 \\ -2 & -4 & -6 \end{pmatrix} \xrightarrow{u_{21,2}} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ -2 & -4 & -6 \end{pmatrix} \xrightarrow{u_{31,2}} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\xrightarrow{u_{12,-2}} \begin{pmatrix} 1 & 0 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \xrightarrow{u_{13,-3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Das populärste Rechenschema der linearen Algebra, der sogenannte Gaußsche Algorithmus, beruht darauf, daß man eine beliebige Matrix durch gezielte Zeilen- und/oder Spaltenumformungen in eine besonders einfache und vor allem für den jeweiligen Zweck leicht durchschaubare Form bringen kann.

Übungsaufgaben

19.1 $P_k \subset \mathbb{R}[X]$ sei der Raum der reellen Polynome vom Grad höchstens k. Durch

- $Df = f'$
- $Df = T_0^k\big(X \cdot f(X)\big)$
- $Df(X) = f(X+1)$

sind drei lineare Abbildungen $D\colon P_k \longrightarrow P_k,\ f \mapsto Df$ (klammernsparende Schreibweise) gegeben. Bestimmen Sie deren Matrizen bezüglich der Basis $(1, X, X^2, \ldots, X^k)$ von P_k.

19.2 Es sei $V := \big\{f\colon \mathbb{R} \longrightarrow \mathbb{R} \,\big|\, f''+2f'+f = 0\big\}$ der Lösungsraum der Schwingungsgleichung (Kriechfall!). Für gegebenes $T \in \mathbb{R}$ bewirke $\Phi\colon V \longrightarrow V$ die Zeitverschiebung um T:

$$(\Phi f)(t) := f(t+T)$$

Bestimmen Sie die Matrix von Φ bezüglich der Basis $\underline{v} = (v_1, v_2)$ mit $v_1(t) = e^{-t}$ und $v_2(t) = te^{-t}$.

19.3 Sei V ein endlichdimensionaler, W ein beliebiger K-Vektorraum; $S, T \subset V$ seien Teilräume. Gegeben seien weiter zwei lineare Abbildungen $f\colon S \longrightarrow W$ und $g\colon T \longrightarrow W$ mit $f|(S\cap T) = g|(S\cap T)$. Zeigen Sie, daß es eine lineare Abbildung $F\colon V \longrightarrow W$ derart gibt, daß $F|S = f$ und $F|T = g$ gilt.

19.4 Beweisen Sie die folgenden Beziehungen zwischen den Elementarmatrizen (gleichen Formats):

(a) $p_{kl} \cdot d_{k\lambda} \cdot p_{kl} = d_{l\lambda}$

(b) $p_{lm} \cdot u_{kl\lambda} \cdot p_{lm} = u_{km\lambda}$ für alle paarweise verschiedenen k, l, m

(c) $u_{kl\lambda} \cdot u_{km\mu} = u_{km\mu} \cdot u_{kl\lambda}$ für $l \neq k \neq m$

Die folgenden Aufgaben behandeln praktische Fragen, die bei der Beschreibung von linearen Abbildungen durch Matrizen routinemäßig auftreten. Lösen Sie diese Aufgaben nicht durch Herumprobieren, sondern durch systematische Anwendung des Satzes 19.5, speziell der Merkregel 19.4. Fangen Sie in jedem Fall mit dem passenden kommutativen Diagramm an; denken Sie daran, daß man mehrere kommutative Diagramme unter Umständen zu einem größeren zusammenfügen kann.

19.5 Schreiben Sie die durch

$$K^3 \ni \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \mapsto \begin{pmatrix} x_1 + 2x_2 + 3x_3 \\ 4x_1 + 5x_2 + 6x_3 \end{pmatrix} \in K^2$$

gegebene lineare Abbildung f als Matrix. Welche Matrix a hat die Abbildung f bezüglich der Basis

$$\underline{v} = \left(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ 2 \end{pmatrix} \right)$$

von K^3 und der Standardbasis $\underline{w}$ von K^2?

19.6 Die lineare Abbildung $f: U \longrightarrow V$ sei bezüglich der Basen $\underline{u} = (u_1, u_2)$ von U und $\underline{v} = (v_1, v_2, v_3)$ von V durch die Matrix

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

gegeben. Nun sei $\underline{w} = (w_1, w_2, w_3)$ eine weitere Basis von V, und es gelte

$$\begin{array}{lcrrr} v_1 & = & 2w_1 & +3w_2 & +w_3 \\ v_2 & = & w_1 & & +w_3 \\ v_3 & = & & -2w_2 & -w_3. \end{array}$$

Berechnen Sie die Matrix b von f bezüglich der Basen $\underline{u}$ und $\underline{w}$.

19.7 Die lineare Abbildung $f: V \longrightarrow W$ sei bezüglich der Basen $\underline{v} = (v_1, v_2, v_3)$ von V und $\underline{w} = (w_1, w_2)$ von W durch die Matrix

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

gegeben. Nun sei $\underline{u} = (u_1, u_2, u_3)$ eine weitere Basis von V, und es gelte

$$\begin{array}{lcrrr} u_1 & = & 2v_1 & +3v_2 & +v_3 \\ u_2 & = & v_1 & & +v_3 \\ u_3 & = & & -2v_2 & -v_3. \end{array}$$

Berechnen Sie die Matrix b von f bezüglich der Basen $\underline{u}$ und $\underline{w}$.

19.8 Die durch die Matrix

$$a = \begin{pmatrix} 0 & 3 & 1 & 5 \\ -3 & 0 & -1 & -3 \\ 1 & 2 & 13 & -7 \\ 2 & -6 & -10 & 2 \end{pmatrix} \in \mathrm{Mat}(4 \times 4, \mathbb{R})$$

gegebene lineare Abbildung $\mathbb{R}^4 \longrightarrow \mathbb{R}^4$ bildet den Unterraum

$$V := \left\{ x \in \mathbb{R}^4 \;\middle|\; \begin{array}{rrrrr} x_1 & +2x_2 & +2x_3 & +3x_4 & = 0 \\ 2x_1 & -x_2 & +x_3 & +x_4 & = 0 \end{array} \right\}$$

in sich ab (das wird versprochen), man erhält also durch Einschränken einen Homomorphismus $f: V \longrightarrow V$. Die Basis $\underline{v}$ von V sei dadurch bestimmt, daß die zugehörige lineare Karte den Vektor $x \in V$ auf $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$ abbildet. Welche Matrix $b \in \mathrm{Mat}(2 \times 2, \mathbb{R})$ beschreibt f bezüglich $\underline{v}$?

20 Der Gaußsche Algorithmus

Er ist sehr viel älter als Gauß (lesen Sie das Literaturzitat dazu im Buch von Jänich), und das spricht für die Behauptung mancher Zyniker, die Person, nach der ein mathematisches Resultat benannt ist, sei wer auch immer, aber jedenfalls nicht dessen erster Entdecker. Der Algorithmus dient in seiner Grundform dazu, von einer gegebenen linearen Abbildung zwischen endlichdimensionalen Vektorräumen Kern und Bild zu berechnen. Für das Ergebnis spielt naturgemäß die Dimension dieser Räume eine zentrale Rolle, und man definiert ganz allgemein:

20.1 Definition Von einer linearen Abbildung $f: V \longrightarrow W$ sagt man, sie habe endlichen Rang, wenn der Vektorraum Bild f endlichdimensional ist, und man nennt dann

$$\operatorname{rk} f := \dim \operatorname{Bild} f \in \mathbb{N}$$

den Rang von f (englisch: rank).

Zumindest wenn alles endliche Dimension hat, kann man sich einen extra Namen für die Dimension des Kerns sparen, denn es gilt die wichtige

20.2 Dimensionsformel für **lineare Abbildungen** V und W seien K-Vektorräume, V endlichdimensional. Dann hat jede lineare Abbildung $f: V \longrightarrow W$ endlichen Rang, und es gilt:

$$\dim \operatorname{Kern} f + \operatorname{rk} f = \dim V$$

Beweis Als Teilraum von V ist Kern f sicher endlichdimensional. Wir können deshalb eine Basis $(v_1, \dots, v_k)$ von Kern f wählen und diese zu einer Basis $(v_1, \dots, v_k, \dots, v_n)$ von V vervollständigen. Dann ist die Einschränkung

$$\operatorname{Lin}(v_{k+1}, \dots, v_n) \longrightarrow \operatorname{Bild} f, \quad v \mapsto f(v)$$

ein Isomorphismus: Für beliebige Skalare $\lambda_i \in K$ gilt nämlich

$$f\left(\sum_{i=1}^{n} \lambda_i v_i\right) = \sum_{i=1}^{n} \lambda_i f(v_i) = \sum_{i=k+1}^{n} \lambda_i f(v_i) = f\left(\sum_{i=k+1}^{n} \lambda_i v_i\right);$$

das zeigt, daß die Einschränkung surjektiv ist. Andererseits ist der Kern dieser Einschränkung gerade

$$\operatorname{Kern} f \cap \operatorname{Lin}(v_{k+1}, \dots, v_n) = \operatorname{Lin}(v_1, \dots, v_k) \cap \operatorname{Lin}(v_{k+1}, \dots, v_n),$$

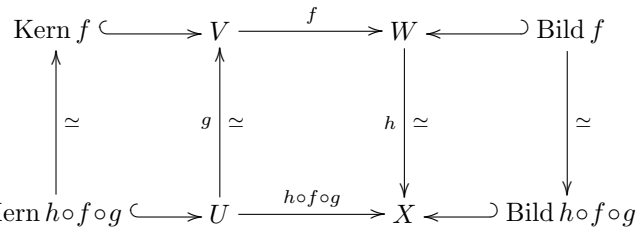
und wegen der linearen Unabhängigkeit von $(v_1, \dots, v_n)$ ist dieser Durchschnitt der Nullraum, womit auch die Injektivität bewiesen ist. Die Formel folgt durch Abzählen der Basisvektoren.

Bemerkung Die Idee des Beweises besteht darin, einen zu Kern f komplementären Unterraum von V zu konstruieren, nämlich $\operatorname{Lin}(v_{k+1}, \dots, v_n)$.

Der Rang ist eine recht robuste Invariante einer linearen Abbildung; er läßt sich durch Vor- oder Nachschalten von linearen Isomorphismen nicht stören:

20.3 Lemma U, V, W, X seien K-Vektorräume, $g: U \longrightarrow V$ und $h: W \longrightarrow X$ seien Isomorphismen. Wenn die lineare Abbildung $f: V \longrightarrow W$ endlichen Rang hat, so hat $h \circ f \circ g$ ebenfalls endlichen, und zwar denselben Rang; ist Kern f endlichdimensional, so auch Kern $h \circ f \circ g$ endlichdimensional, und beide haben dieselbe Dimension.

Beweis Das ist sofort aus dem kommutativen Diagramm



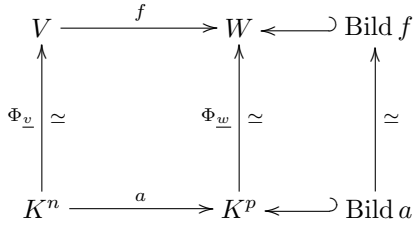
abzulesen, in dem sich freilich zwei selbst beweisbedürftige, wenn auch einfache Behauptungen verbergen. Wir zeigen, daß g den Unterraum $\operatorname{Kern} h\circ f\circ g$ auf $\operatorname{Kern} f$ bewegt: Sei $u\in\operatorname{Kern} h\circ f\circ g$, dann ist $h\big((f\circ g)(u)\big)=0$, also, weil h injektiv ist, schon $(f\circ g)(u)=0$, d.h. $f\big(g(u)\big)=0$ und damit $g(u)\in\operatorname{Kern} f$. Ist andererseits $v\in\operatorname{Kern} f$ gegeben, so können wir wegen der Surjektivität von g jedenfalls $v=g(u)$ mit $u\in U$ schreiben; wegen $(h\circ f\circ g)(u)=(h\circ f)(v)=h(0)=0$ ist dann $u\in\operatorname{Kern} h\circ f\circ g$. Ebenso einfach sieht man, daß h das Bild von f auf $\operatorname{Bild} h\circ f\circ g$ bewegt.

Da Matrizen $\operatorname{Mat}(p\times n,K)$ zugleich lineare Abbildungen von K^n nach K^p sind, können wir vom Rang einer Matrix reden. Aus dem Lemma ergibt sich unmittelbar die

20.4 Folgerung (a) Ist $f\colon V\longrightarrow W$ eine lineare Abbildung zwischen endlichdimensionalen Vektorräumen, so ist der Rang von f gleich dem Rang der f beschreibenden Matrix bezüglich beliebig gewählten Basen von V und W.

(b) Zeilen- oder Spaltenumformungen ändern den Rang einer Matrix nicht.

Beweis (a) Sind $\underline{v}$ und $\underline{w}$ Basen von V und W, so wählt man im Lemma $g=\Phi_{\underline{v}}$ und $h=\Phi_{\underline{w}}^{-1}$:



(b) Elementarmatrizen sind umkehrbar.

Einzeln auf durch Elementarmatrizen u und v bewirkte Zeilen- und Spaltenumformungen einer Matrix angewendet reduziert das Diagramm aus dem Beweis des Lemmas sich auf

und liefert über die Folgerung 20.4 hinaus die wichtige

20.5 Notiz Zeilenumformungen einer Matrix ändern ihren Kern nicht; Spaltenumformungen ändern ihr Bild nicht.

Nun zur Berechnung von Kern und Bild nach dem Gaußschen Algorithmus. Dazu muß die lineare Abbildung $f\colon V\longrightarrow W$ natürlich in einer Form gegeben sein, mit der man überhaupt rechnen kann: Wir setzen deshalb voraus, daß f durch die zugehörige Matrix a bezüglich Basen $\underline{v}$ und $\underline{w}$ gegeben ist. Ebenfalls klargestellt

werden muß, was es heißen soll, einen Unterraum von V oder W zu "berechnen". Bis auf weiteres wollen wir darunter verstehen, eine Basis dieses Unterraums zu konstruieren. Die Vektoren einer solchen Basis wird man dann in den durch $\underline{v}$ und $\underline{w}$ festgelegten Karten, d.h. als Linearkombinationen von $\underline{v}$ und $\underline{w}$ ausgeben wollen. Angesichts des Diagramms

$$\begin{array}{ccccccc}
\operatorname{Kern} f & \hookrightarrow & V & \xrightarrow{f} & W & \hookleftarrow & \operatorname{Bild} f \\
\uparrow \simeq & & \uparrow \Phi_{\underline{v}} \simeq & & \uparrow \Phi_{\underline{w}} \simeq & & \uparrow \simeq \\
\operatorname{Kern} a & \hookrightarrow & K^n & \xrightarrow{a} & K^p & \hookleftarrow & \operatorname{Bild} a
\end{array}$$

reduziert sich damit die Aufgabe auf den Spezialfall, Kern und Bild einer linearen Abbildung $a\colon K^n \longrightarrow K^p$, also einer Matrix $a \in \operatorname{Mat}(p\times n, K)$ zu finden. Als erstes erläutere ich die Berechnung des Kerns an einem konkreten

20.6 Beispiel Die Matrix

$$a = \begin{pmatrix} 1 & 1 & 2 & -2 \\ 1 & 0 & 0 & 1 \\ -1 & 1 & 2 & 0 \\ 0 & 1 & 2 & -1 \end{pmatrix} \in \operatorname{Mat}(4\times 4, K)$$

bearbeiten wir durch folgende Zeilenumformungen:

$$\begin{pmatrix} 1 & 1 & 2 & -2 \\ 1 & 0 & 0 & 1 \\ -1 & 1 & 2 & 0 \\ 0 & 1 & 2 & -1 \end{pmatrix} \xrightarrow{u_{31,1}\cdot u_{21,-1}} \begin{pmatrix} 1 & 1 & 2 & -2 \\ 0 & -1 & -2 & 3 \\ 0 & 2 & 4 & -2 \\ 0 & 1 & 2 & -1 \end{pmatrix} \xrightarrow{d_{2,-1}} \begin{pmatrix} 1 & 1 & 2 & -2 \\ 0 & 1 & 2 & -3 \\ 0 & 2 & 4 & -2 \\ 0 & 1 & 2 & -1 \end{pmatrix}$$

$$\xrightarrow{u_{32,-2}\cdot u_{42,-1}} \begin{pmatrix} 1 & 1 & 2 & -2 \\ 0 & 1 & 2 & -3 \\ 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 2 \end{pmatrix} \xrightarrow{d_{3,\frac{1}{4}}} \begin{pmatrix} 1 & 1 & 2 & -2 \\ 0 & 1 & 2 & -3 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

$$\xrightarrow{u_{43,-2}} \begin{pmatrix} 1 & 1 & 2 & -2 \\ 0 & 1 & 2 & -3 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \xrightarrow{u_{13,2}\cdot u_{23,3}} \begin{pmatrix} 1 & 1 & 2 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\xrightarrow{u_{12,-1}} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Nennen wir die umformende Matrix u, dann ist die Ergebnismatrix ua. Der sieht man ihren Kern aber sofort an: Die Forderung $x \in \operatorname{Kern} ua$ bedeutet ja

$$(ua)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

oder ausgeschrieben

$$\begin{array}{rrrr}
x_1 & & & = 0 \\
 & x_2 & +2x_3 & = 0 \\
 & & x_4 & = 0,
\end{array}$$

das heißt $x_1 = 0$, $x_2 = -2x_3$ und $x_4 = 0$ mit frei wählbarem $x_3 \in K$. Mit anderen Worten ist

$$\operatorname{Kern} a = \operatorname{Kern} ua = \operatorname{Lin}\left(\begin{pmatrix} 0 \\ -2 \\ 1 \\ 0 \end{pmatrix}\right) \subset K^4.$$

Zur Berechnung von Bild a arbeiten wir mit Spaltenumformungen:

$$\begin{pmatrix} 1 & 1 & 2 & -2 \\ 1 & 0 & 0 & 1 \\ -1 & 1 & 2 & 0 \\ 0 & 1 & 2 & -1 \end{pmatrix} \xrightarrow[\cdot u_{14,2}]{u_{12,-1}\cdot u_{13,-2}} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & -1 & -2 & 3 \\ -1 & 2 & 4 & -2 \\ 0 & 1 & 2 & -1 \end{pmatrix} \xrightarrow{d_{2,-1}} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & -2 & 3 \\ -1 & -2 & 4 & -2 \\ 0 & -1 & 2 & -1 \end{pmatrix}$$

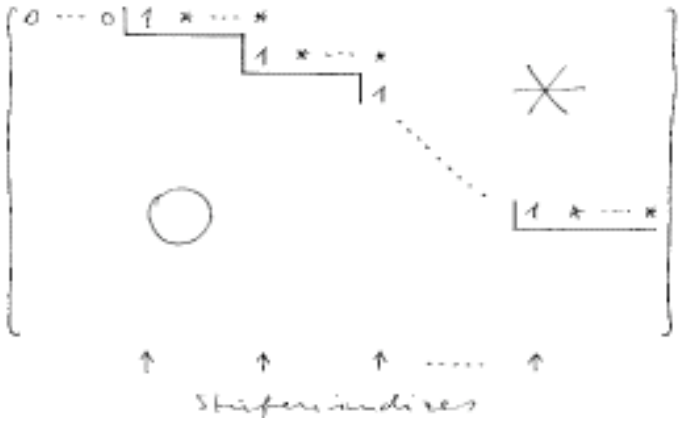
$$\xrightarrow{u_{23,2}\cdot u_{24,-3}} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ -1 & -2 & 0 & 4 \\ 0 & -1 & 0 & 2 \end{pmatrix} \xrightarrow{p_{34}} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ -1 & -2 & 4 & 0 \\ 0 & -1 & 2 & 0 \end{pmatrix}$$

$$\xrightarrow{d_{3,\frac{1}{4}}} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ -1 & -2 & 1 & 0 \\ 0 & -1 & \frac{1}{2} & 0 \end{pmatrix}$$

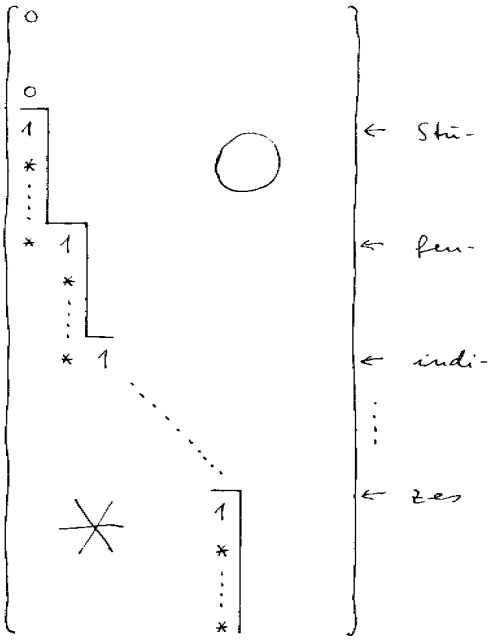
Es ist sofort klar, daß die von null verschiedenen Spalten der erhaltenen Matrix av linear unabhängig sind und Bild av = Bild a aufspannen, womit wir eine Basis für das Bild von a gefunden haben.

Jetzt bleibt das, was wir an diesem Beispiel gemacht haben, systematisch und allgemein zu beschreiben.

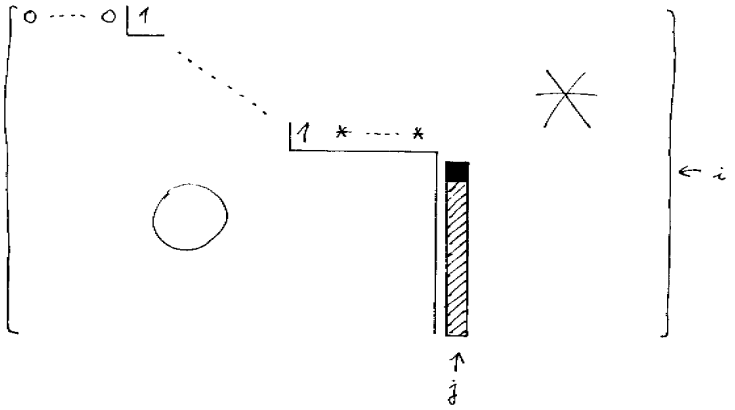
20.7 Definition Von einer Matrix $a \in \operatorname{Mat}(p \times n, K)$ der Gestalt



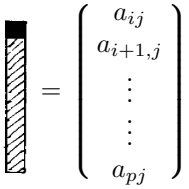
sagt man, sie habe Zeilenstufenform. Dabei bedeutet ein Stern, daß von den Matrixeinträgen an dieser Stelle nichts weiter verlangt ist. Der flotten Verständigung halber wollen wir diejenigen Spaltenindizes, an denen die Matrix eine "Stufe" hat, die Stufenindizes der Matrix nennen. Der Begriff Spaltenstufenform ist natürlich ganz entsprechend definiert:



20.8 Gaußscher Algorithmus Das folgende Verfahren macht aus einer beliebig vorgegebenen Matrix $a \in \mathrm{Mat}(p \times n, K)$ durch Zeilenumformung eine Matrix in Zeilenstufenform: Die Eigenschaften einer Zeilenstufenform werden sukzessive für die erste, zweite, dritte Spalte usw. hergestellt. In einem typischen Schritt hat man also eine Matrix der Form



und bearbeitet die hervorgehobene Teilspalte:



$$= \begin{pmatrix} a_{ij} \\ a_{i+1,j} \\ \vdots \\ \vdots \\ a_{pj} \end{pmatrix}$$

- Wenn diese die Nullspalte ist, ist nichts zu tun (j wird dann kein Stufenindex).
- Handelt es sich nicht um die Nullspalte, ist aber $a_{ij} = 0$, so sucht man das erste $h > i$ mit $a_{hj} \neq 0$ und vertauscht die h-te mit der i-ten Zeile. Danach ist in jedem Fall $a_{ij} \neq 0$ (wir verwenden a hier im Sinne einer Programmvariablen, nennen die neue Matrix also wieder a).
- Jetzt teilt man die gesamte i-te Zeile durch a_{ij}. Danach ist sogar $a_{ij} = 1$, und j ist ein neuer Stufenindex.
- Schließlich subtrahiert man für jedes $k > i$ von der k-ten Zeile das a_{kj}-fache der i-ten und erreicht damit $a_{kj} = 0$ für $k > i$.

Selbstverständlich gibt es auch eine Spaltenversion des Gaußschen Algorithmus, die völlig analog arbeitet und jede gegebene Matrix in eine Matrix in Spaltenstufenform überführt.

Wir müssen noch allgemein beschreiben, wie man einer Matrix in Stufenform Kern bzw. Bild ansieht. Einfacher ist das für das Bild:

20.9 Lemma Die Matrix $a \in \mathrm{Mat}(p \times n, K)$ habe Spaltenstufenform. Ihre von null verschiedenen Spalten bilden dann eine Basis für $\mathrm{Bild}\, a$.

Beweis Daß diese Spalten, etwa $a_1, \ldots, a_r$ das Bild aufspannen, ist klar. Sie sind aber auch linear unabhängig: aus dem Ansatz

$$0 = \sum_{j=1}^{r} \lambda_j a_j \quad \text{mit } \lambda_j \in K$$

folgt durch Betrachtung der Stufenkomponenten sukzessive $\lambda_1 = 0$, $\lambda_2 = 0, \ldots, \lambda_r = 0$.

Bemerkung Auch einer Matrix a in Zeilenstufenform sieht man ihren Bildraum sofort an, er hängt nur vom Rang r ab und ist

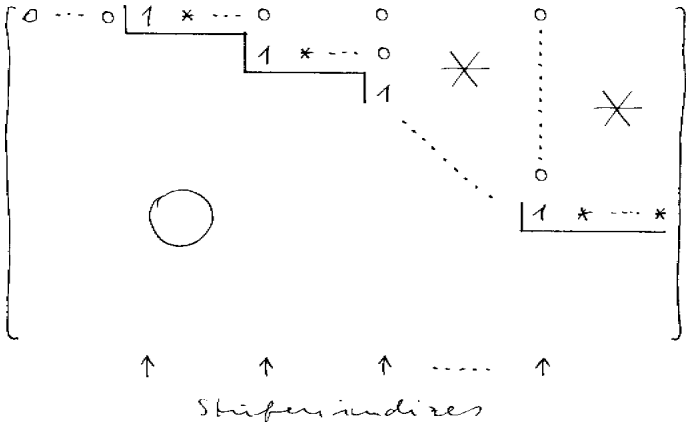
$$\mathrm{Bild}\, a = K^r \times \{0\} \subset K^r \times K^{n-r} = K^n\,.$$

Um die entsprechende Frage für den Kern zu beantworten, ist es günstig, den Gaußschen Algorithmus noch weiter zu führen, wie wir ja auch in unserem Beispiel schon getan haben. Dazu die

20.9$\frac{1}{3}$ Definition Die Matrix $a \in \mathrm{Mat}(p \times n, K)$ hat veredelte Zeilenstufenform, wenn sie Zeilenstufenform hat und für jeden Stufenindex, etwa den s-ten Stufenindex j_s

$$a_{ij_s} = 0 \quad \text{für } i = 1, \ldots, s-1$$

gilt, die j_s-te Spalte von a also der s-te Vektor der Standardbasis von K^p ist:



Matrizen in veredelter Zeilenstufenform produziert der

20.9$\frac{2}{3}$ Veredelungsalgorithmus $a \in \mathrm{Mat}(p \times n, K)$ sei eine Matrix in Zeilenstufenform. Die noch fehlenden Eigenschaften werden sukzessiv für die Stufenindizes $j_r > j_{r-1} > \cdots > j_1$ etabliert. Und zwar, etwa für den s-ten Stufenindex j_s, indem man für jedes $k < s$ von der k-ten Zeile das a_{kj_s}-fache der s-ten subtrahiert.

Bemerkungen Die Reihenfolge der Veredelungsschritte ist grundsätzlich unwesentlich, aber die angegebene ist günstig, um den Rechenaufwand klein zu halten. — Es ist wohl klar, was unter veredelter *Spalten*stufenform zu verstehen ist, und durch welchen Algorithmus man zu dieser gelangt, ausgehend von einer gewöhnlichen Spaltenstufenform.

20.10 Lemma Die Matrix $a \in \mathrm{Mat}(p \times n, K)$ habe veredelte Zeilenstufenform; $j_1 < j_2 < \cdots < j_r$ seien ihre Stufenindizes. Für jedes $k \in \{1, \ldots, n\}$, das *kein* Stufenindex ist, definieren wir einen Vektor $v_k \in K^n$ so: Sind $j_1, \ldots, j_t$ die Stufenindizes, die kleiner als k sind, dann ist

$$v_k = \begin{pmatrix} \\ -a_{1k} \\ \\ -a_{sk} \\ \\ -a_{tk} \\ \\ 1 \\ \\ \end{pmatrix} \begin{matrix} \\ \leftarrow j_1 \\ \\ \leftarrow j_s \quad (1 < s < t) \\ \\ \leftarrow j_t \\ \\ \leftarrow k \\ \\ \end{matrix}$$

(alle übrigen Komponenten sind null). Dann ist das $(n–r)$-tupel $\left(v_{k_1}, \ldots, v_{k_{n-r}}\right)$, in dem $k_1 < k_2 < \cdots < k_{n-r}$ die Nichtstufenindizes sind, eine Basis für Kern a.

Beweis Die angegebenen Vektoren liegen im Kern; so ist v_k gerade gemacht:

$$av_k = \begin{pmatrix} 1 & 0 & 0 & a_{1k} & 0 \\ & \vdots & \vdots & \vdots & \vdots \\ & 1 & 0 & a_{sk} & 0 \\ & & \vdots & \vdots & \vdots \\ & & 1 & a_{tk} & 0 \\ & & & & 1 \\ & & & & \\ & & & & \end{pmatrix} \begin{pmatrix} -a_{1k} \\ \\ -a_{sk} \\ \\ -a_{tk} \\ \\ 1 \\ \\ \end{pmatrix} = 0$$

Weil v_k an der k-ten Stelle eine Eins, als alle anderen Nichtstufenkomponenten aber Nullen hat, sind die Vektoren $v_{k_1}, \ldots, v_{k_{n-r}}$ linear unabhängig, spannen also einen $(n-r)$-dimensionalen Unterraum von Kern a auf. Wegen Bild $a = K^r \times \{0\}$ ist andererseits $\mathrm{rk}\, a = r$, und nach der Dimensionsformel 20.2 folgt, daß Kern a selbst die Dimension $n-r$ hat. Also ist $\left(v_{k_1}, \ldots, v_{k_{n-r}}\right)$ eine Basis von Kern a.

Bemerkung Der Kern einer Matrix a in Spaltenstufenform mit r nichttrivialen Spalten ist offensichtlich $\{0\} \times K^{n-r} \subset K^r \times K^{n-r} = K^n$.

Damit haben wir den Gaußschen Algorithmus zur Berechnung von Kern und Bild einer linearen Abbildung vollständig beschrieben. Übrigens ist das nur eine von etlichen möglichen Varianten. Manche Autoren verlangen zum Beispiel von den Stufenkomponenten einer Matrix in Zeilenstufenform nur, daß sie ungleich null sind, was im Gaußschen Algorithmus einen Schritt einspart und dafür andere komplizierter macht. Andererseits kann man auch zur Bestimmung des Kerns auf die Veredelung verzichten; das macht das Ablesen des Kerns dann komplizierter.

Übrigens wächst der Rechenaufwand für den Algorithmus bei größeren Matrizenformaten nur moderat (grob gesagt mit der dritten Potenz der Größe). Wenn man allerdings numerisch, d.h. mit Rundungsfehlern rechnen muß oder schon die Ausgangsmatrix Rundungsfehler enthält, ist der Algorithmus mit Vorsicht anzuwenden: Es kann dabei ja leicht passieren, daß von zwei relativ genauen, aber fast gleichen Zahlen die Differenz zu bilden ist, und deren Genauigkeit läßt sich dann nicht mehr kontrollieren. Man muß sich auch vor Augen halten, daß schon der qualitative Verlauf der Rechnung davon abhängt, ob bestimmte Zwischenresultate null werden: aber wie soll man "null" gegen "nicht null" abgrenzen, wenn man ohnehin mit Rundungsfehlern zu rechnen hat? Um den Einfluß dieser Fehlerquellen zu begrenzen, unterwirft man die Matrix für

numerische Rechnung in der Regel einigen vorbereitenden Umformungen. Diese Probleme aus dem Bereich der numerischen Mathematik können wir hier nicht besprechen, aber angesichts der großen Bedeutung des Gaußschen Algorithmus gibt es natürlich detaillierte Untersuchungen dazu.

Zum Gaußschen Algorithmus nun einige weitere

20.11 Anwendungen (1) Sind Vektoren $v_1, \dots, v_n$ eines K-Vektorraums V gegeben, so kann man eine Basis des aufgespannten Teilraums $\mathrm{Lin}(v_1, \dots, v_n)$ berechnen: Da Vorgabe und Ergebnis bezüglich einer Basis von V erwartet werden, ist man de facto im Fall $V = K^p$, und da ist $\mathrm{Lin}(v_1, \dots, v_n) \subset K^p$ einfach das Bild der durch Nebeneinandersetzen der Spalten v_j gebildeten Matrix

$$a := (\, v_1 \quad v_2 \quad \dots \quad v_n \,) \in \mathrm{Mat}(p \times n, K).$$

Dieses Bild bestimmt man nach dem Gaußschen Algorithmus in der Spaltenversion. Die Dimension von $\mathrm{Lin}(v_1, \dots, v_n)$ ist der Rang von a, und durch Vergleich von $\mathrm{rk}\, a$ mit n erkennt man auch, ob $(v_1, \dots, v_n)$ linear unabhängig ist. Ein anderes, die Zeilenversion benutzendes Verfahren ist in Aufgabe 20.7 beschrieben.

(2) Sind Teilräume S, T eines Vektorraums V gegeben, so läßt sich jetzt leicht eine Basis der Summe $S + T$ bestimmen: Ist etwa $S = \mathrm{Lin}(v_1, \dots, v_s)$ und $T = \mathrm{Lin}(w_1, \dots, w_t)$, so braucht man bloß Verfahren (1) auf $S + T = \mathrm{Lin}(v_1, \dots, v_s, w_1, \dots, w_t)$ anzuwenden. (Auf eine Methode, um auch $S \cap T$ zu berechnen, kommen wir später, im Abschnitt 27 zu sprechen.)

(3) Sind $S, T \subset V$ wieder durch aufspannende Vektoren gegebene Teilräume, so können wir entscheiden, ob $S \subset T$ gilt; damit natürlich auch, ob $S \supset T$ oder $S = T$ ist. Tatsächlich gilt offenbar

$$S \subset T \iff S + T = T \iff \dim(S + T) = \dim T,$$

und (2) und (3) erlauben es, die relevanten Dimensionen zu bestimmen. Insbesondere können wir von jedem Vektor $v \in V$ entscheiden, ob er zu dem durch aufspannende Vektoren gegebenen Unterraum $T \subset V$ gehört oder nicht. Freilich ist diese Methode etwas umständlich, und wir werden bald eine geschicktere finden.

(4) Wir können jedes *homogene lineare Gleichungssystem*

$$\begin{array}{ccccccc} a_{11}x_1 & + & \cdots & + a_{1n}x_n & = & 0 \\ \vdots & & & \vdots & & \vdots \\ a_{p1}x_1 & + & \cdots & + a_{pn}x_n & = & 0 \end{array}$$

für $x_1, \dots, x_n$ mit dem Gaußschen Algorithmus lösen. Denn wenn wir die gegebenen Koeffizienten a_{ij} zu einer Matrix $a \in \mathrm{Mat}(p \times n, K)$ zusammenfassen, lautet die Gleichung

$$ax = 0$$

für $x \in K^n$, es ist also gerade nach (einer Basis für) $\mathrm{Kern}\, a$ gefragt. Mit den allgemeineren Gleichungen vom Typ $ax = b$ befassen wir uns im nächsten Abschnitt.

Bei den meisten Anwendungen des Gaußschen Algorithmus möchte man entweder nur Zeilen- oder nur Spaltenumformungen durchführen (um entweder den Kern oder das Bild der Ausgangsmatrix zu erhalten). Im Beweis des folgenden Satzes kommt der Algorithmus aber in Zeilen- *und* Spaltenversion zum Zuge:

20.12 Satz Seien V und W endlichdimensionale K-Vektorräume, $\dim V = n$ und $\dim W = p$. Dann gibt es zu jeder linearen Abbildung $f\colon V \longrightarrow W$ Basen von V und W derart, daß f bezüglich dieser Basen die Matrix

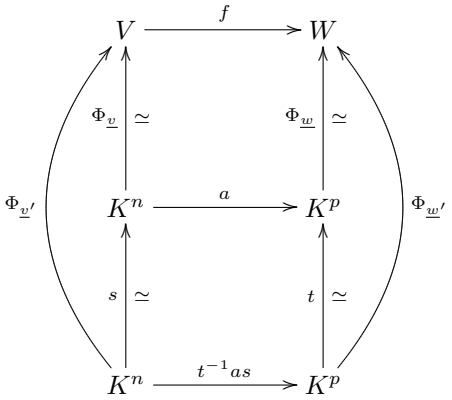
$$\left(\begin{array}{ccc|c} 1 & & & \\ & \ddots & & \\ & & 1 & \\ \hline & & & \end{array}\right) \in \mathrm{Mat}(p \times n, K)$$

(r Einsen auf der Diagonale)

hat (die leeren Blöcke sind Nullmatrizen). Dabei ist zwangsläufig $r = \operatorname{rk} f$ der Rang von f.

Beweis Der Satz ist nicht schwer zu beweisen, indem man den Beweis der Dimensionsformel 20.2 noch etwas weiterführt. Der folgende alternative Beweis hat aber den Vorzug, zugleich ein Verfahren anzugeben, wie man Basen mit der gewünschten Eigenschaft konstruieren kann. Wir wählen erst mal beliebige Basen $\underline{v} = (v_1, \ldots, v_n)$ und $\underline{w} = (w_1, \ldots, w_p)$. Die f bezüglich dieser Basen beschreibende Matrix a bearbeiten wir mit dem Gaußschen Algorithmus zuerst in Spalten- und dann in Zeilenversion. Man überlegt sich sofort, daß die zweite Anwendung zu einer Matrix der vom Satz versprochenen Form führt. Wir bezeichnen die die Umformungen bewirkenden Matrizen mit $s \in GL(n, K)$ und $t^{-1} \in GL(p, K)$

Wir wollen die Basen $\underline{v}'$ und $\underline{w}'$ jetzt so wählen, daß in dem Diagramm



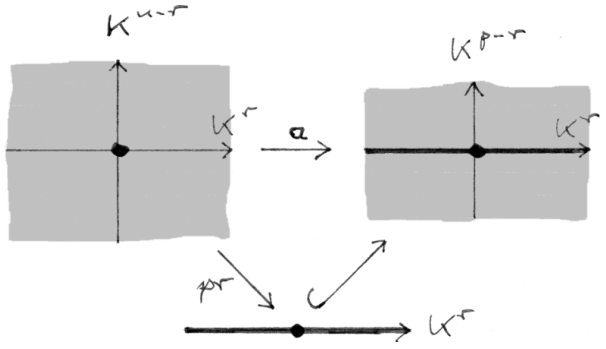
auch die angesetzten Dreiecke kommutativ werden. Das wird offenbar durch

$$\underline{v}' := \big(\Phi_{\underline{v}}(s_1), \ldots, \Phi_{\underline{v}}(s_n)\big)$$
$$\underline{w}' := \big(\Phi_{\underline{w}}(t_1), \ldots, \Phi_{\underline{w}}(t_p)\big)$$

geleistet, worin $s_1, \ldots, s_n$ die Spalten der Matrix s und $t_1, \ldots, t_p$ die von t sind.

Die Aussage über den Rang ist schon in der Folgerung 20.4 enthalten.

Der Satz illustriert eine früher gemachte Bemerkung: Wenn wir es mit einer linearen Abbildung $f\colon K^n \longrightarrow K^p$ zu tun haben, kann es durchaus zweckmäßig sein, in K^n und K^p nicht die Standardbasen zu verwenden, sondern Basen, wie sie eben Satz 20.12 verspricht. Die Matrix $a \in \operatorname{Mat}(p{\times}n, K)$, die f dann beschreibt, kann man sich einfacher ja kaum vorstellen; wenn wir $K^n = K^r \times K^{n-r}$ und $K^p = K^r \times K^{p-r}$ schreiben, wirkt sie als die kartesische Projektion gefolgt von der Inklusion:



$$K^n = K^r \times K^{n-r} \ni (x, y) \longmapsto x \longmapsto (x, 0) \in K^r \times K^{p-r} = K^p$$

Übungsaufgaben

20.1 $f\colon V \longrightarrow W$ und $g\colon W \longrightarrow X$ seien lineare Abbildungen von endlichem Rang. Beweisen Sie, daß dann

$$\operatorname{rk} g\circ f \le \operatorname{rk} f \quad \text{und} \quad \operatorname{rk} g\circ f \le \operatorname{rk} g,$$

andererseits aber, wenn W endlichdimensional ist,

$$\operatorname{rk} g\circ f \ge \operatorname{rk} f + \operatorname{rk} g - \dim W$$

gilt.

20.2 V sei ein K-Vektorraum, und $f\colon V \longrightarrow V$ sei eine lineare Abbildung mit endlichem Rang. Für die n-fache Komposition $f\circ f\circ\cdots\circ f$ werde kurz f^n geschrieben. Beweisen Sie: Immer gilt $\operatorname{rk} f^2 \le \operatorname{rk} f$; wenn aber $\operatorname{rk} f^2 = \operatorname{rk} f$ ist, dann folgt $\operatorname{rk} f^n = \operatorname{rk} f$ für alle $n > 0$.

20.3 Es sei K ein Körper; n und p seien natürliche Zahlen. Beweisen Sie: Zu jeder Matrix $a \in \operatorname{Mat}(p\times n, K)$ vom Rang eins gibt es eine Spaltenmatrix $s \in \operatorname{Mat}(p\times 1, K)$ und eine Zeilenmatrix $z \in \operatorname{Mat}(1\times n, K)$ mit $a = s\cdot z$.

20.4 Für festes $\lambda \in \mathbb{R}$ sei $S = \operatorname{Lin}(v_1, v_2) \subset \mathbb{R}^4$ und $T = \operatorname{Lin}(w_1, w_2, w_3) \subset \mathbb{R}^4$ mit

$$v_1 = \begin{pmatrix} 1\\0\\-4\\3 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0\\0\\1\\2 \end{pmatrix}, \quad w_1 = \begin{pmatrix} 1\\3\\0\\-1 \end{pmatrix}, \quad w_2 = \begin{pmatrix} 2\\5\\-1\\0 \end{pmatrix} \quad \text{und} \quad w_3 = \begin{pmatrix} 0\\1\\2\\\lambda \end{pmatrix}.$$

Bestimmen Sie alle $\lambda \in \mathbb{R}$, für die $S \subset T$ ist.

20.5 Bestimmen Sie alle $\lambda \in \mathbb{R}$, für die es eine lineare Abbildung $f\colon \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ mit den Eigenschaften

$$f(v_1) = w_1, \quad f(v_2) = w_2, \quad f(v_3) = w_3$$

gibt, wobei die Spalten $v_j, w_j \in \mathbb{R}^3$ durch

$$v_1 = \begin{pmatrix} 1\\0\\1 \end{pmatrix}, \quad w_1 = \begin{pmatrix} \lambda\\1\\2 \end{pmatrix}; \quad v_2 = \begin{pmatrix} 0\\1\\-2 \end{pmatrix}, \quad w_2 = \begin{pmatrix} -3\\0\\-1 \end{pmatrix}; \quad v_3 = \begin{pmatrix} 2\\\lambda\\0 \end{pmatrix}, \quad w_3 = \begin{pmatrix} -1\\2\\3 \end{pmatrix}$$

gegeben sind.

20.6 Die linearen Unterräume S und T_λ von $\mathbb{R}^4$ seien durch

$$S = \operatorname{Lin}\left(\begin{pmatrix} 1\\0\\1\\0 \end{pmatrix}, \begin{pmatrix} -2\\1\\4\\2 \end{pmatrix}, \begin{pmatrix} 0\\1\\6\\2 \end{pmatrix}\right) \quad \text{und} \quad T_\lambda = \operatorname{Lin}\left(\begin{pmatrix} 2\\1\\0\\-2 \end{pmatrix}, \begin{pmatrix} 1\\-1\\-1\\-\lambda \end{pmatrix}\right)$$

gegeben. Bestimmen Sie die Dimension von $S \cap T_\lambda$ in Abhängigkeit von $\lambda \in \mathbb{R}$.

20.7 Aus beliebig vorgegebenen Vektoren $a_1, \dots, a_n \in K^p$ kann man nach dem Basisergänzungssatz immer eine Basis für $\operatorname{Lin}(a_1, \dots, a_n)$ auswählen. Begründen Sie das folgende Verfahren dafür, und illustrieren Sie

es mit einem angemessenen Beispiel: Die Spalten $a_1, \ldots, a_n$ werden zu einer $p \times n$-Matrix a zusammengefaßt, diese wird durch Zeilenumformungen in Zeilenstufenform gebracht. Sind $j_1 < \ldots < j_r$ die Stufenindizes, so ist $(a_{j_1}, \ldots, a_{j_r})$ eine Basis von $\mathrm{Lin}(a_1, \ldots, a_n)$.

20.8 Zeigen Sie folgenden Eindeutigkeitssatz für die veredelte Spaltenstufenform: Sind $a, b \in \mathrm{Mat}(p \times n, K)$ Matrizen in veredelter Spaltenstufenform und läßt sich a durch Spaltenumformungen in b überführen, so ist zwangsläufig schon $a = b$.

Tip: Man kann a aus der Kenntnis von $\mathrm{Bild}\, a$ rekonstruieren: Wenn man für $k = 1, \ldots, p$ die Unterräume

$$B_k := \mathrm{Bild}\, a \cap \left(\{0\} \times K^k\right) \subset K^{p-k} \times K^k = K^p$$

bildet, kann man zunächst aus den Zahlen $\dim B_k$ die Stufenindizes ablesen und dann aus den Räumen B_k selbst die Spalten von a bestimmen.

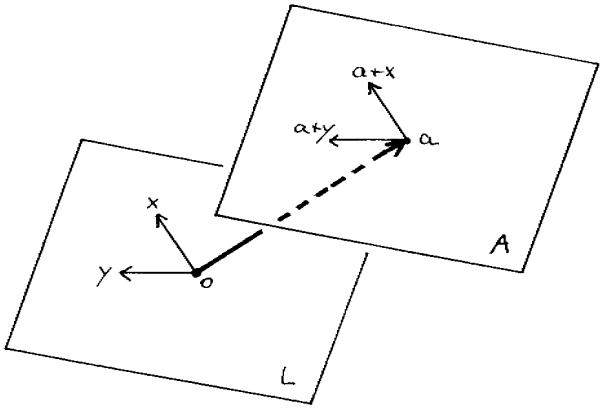
21 Lineare Gleichungen

Für die Behandlung dieses Themas ist es zweckmäßig, dem Begriff des Untervektorraums eine allgemeinere Variante an die Seite zu stellen.

21.1 Definition V sei ein Vektorraum. Eine Teilmenge $A \subset V$ der Form

$$A = a + L := \{a + x \,|\, x \in L\},$$

worin $a \in V$ ein Vektor und $L \subset V$ ein Untervektorraum ist, heißt ein affiner Unter- oder Teilraum von V. Außerdem gilt per Konvention die leere Menge als affiner Teilraum von V.



Wenn man der Kürze halber weiterhin bloß von Teil- oder Unterräumen von V redet, muß man künftig natürlich klarstellen, ob "affin" oder "linear" gemeint ist.

21.2 Lemma und **Definition** V sei ein Vektorraum, $A = a + L \subset V$ ein nicht-leerer affiner Teilraum.

(a) Es gilt $L = \{x - y \,|\, x, y \in A\}$; insbesondere ist L durch A eindeutig bestimmt, und man darf deshalb L den zu A parallelen linearen Unterraum von V nennen und die Dimension von A als die von L definieren:

$$\dim A := \dim L$$

(eine Dimension des leeren affinen Raumes wird nicht erklärt).

(b) Für jedes $b \in V$ gilt

$$A = b + L \iff b \in A.$$

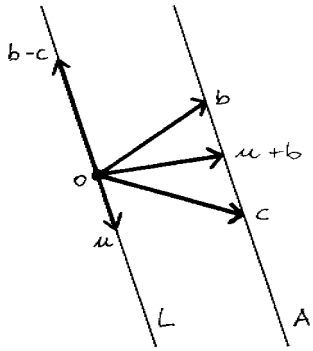
Insbesondere sind die linearen Teilräume von V genau diejenigen affinen Teilräume, die den Nullvektor enthalten.

Beweis (a) Aus $x, y \in a+L$, etwa $x = a+u$ und $y = a+v$ mit $u, v \in L$, folgt offenbar $x-y = (a+u)-(a+v) = u - v \in L$. Umgekehrt läßt sich jedes $u \in L$ als Differenz $u = (a + u) - a$ zweier Elemente aus A schreiben.

(b) Sei $A = b+L$, wegen $a \in A$ gibt es dann ein $u \in L$ mit $a = b+u$, und es folgt $b = a+(-u) \in a+L = A$. Sei umgekehrt $b \in A$ vorausgesetzt. Für jedes Element $a \in A$ ist nach (a) dann $a - b \in L$ und folglich $a = b+(a-b) \in b+L$, das zeigt $A \subset b+L$. Zum Beweis der umgekehrten Inklusion sei $b+v \in b+L$ (mit $v \in L$); wegen $b \in A$ können wir $b = a+u$ mit $u \in L$ schreiben, und es folgt $b+v = (a+u)+v = a+(u+v) \in a+L = A$.

Bemerkungen Zu den affinen Teilräumen von V zählen insbesondere die einpunktigen Mengen, die man ja in der Form $\{a\} = a + \{0\} \subset V$ schreiben kann. Abgesehen von diesem Fall ist in der Darstellung $A = a + L$ der Vektor a *nicht* eindeutig bestimmt: das geht aus Teil (b) des Lemmas hervor. Beachten Sie auch, daß ein affiner Teilraum von V, der nicht linear ist, von V keine der Vektorraumverknüpfungen erbt; es entstehen vielmehr zwei neuartige (einander gleichwertige) Verknüpfungen

$$L \times A \xrightarrow{+} A \quad \text{und} \quad A \times A \xrightarrow{-} L.$$



Affine Teilräume treten ganz natürlich als Lösungsräume linearer Gleichungen auf. Wie wir längst wissen, hat die zu einer linearen Abbildung $f\colon V \longrightarrow W$ gehörige sogenannte *homogene* lineare Gleichung für x

$$f(x) = 0$$

als Lösungsmenge einen Untervektorraum von V, nämlich den Kern von f. Ist zusätzlich ein Vektor $b \in W$ gegeben, so nennt man die Gleichung für x

$$f(x) = b$$

eine *inhomogene* lineare Gleichung (manchmal reserviert man diese Bezeichnung für den Fall, daß tatsächlich $b \neq 0$, die Gleichung also nicht homogen ist). Die Lösungsmenge einer solchen Gleichung ist im allgemeinen kein linearer, aber immer ein affiner Unterraum von V:

21.3 Lemma $f\colon V \longrightarrow W$ sei linear. Für jedes $b \in W$ ist die Faser

$$f^{-1}\{b\} \subset V$$

ein affiner Teilraum von V. Ist er nicht leer, so ist $\operatorname{Kern} f$ der zu ihm parallele lineare Unterraum.

Beweis Ist $f^{-1}\{b\} = \emptyset$, so ist nichts zu zeigen. Wenn nicht, wählen wir ein $a \in f^{-1}\{b\}$. Für jedes $x \in V$ gilt dann

$$x \in \operatorname{Kern} f \iff f(x) = 0 \iff f(a+x) = f(a) + f(x) = b \iff a + x \in f^{-1}\{b\},$$

also ist

$$f^{-1}\{b\} = a + \operatorname{Kern} f.$$

Im endlichdimensionalen Fall haben wir es im wesentlichen mit einer Matrix $a \in \operatorname{Mat}(p \times n, K)$ anstelle f, einer Spalte $b \in K^p$ und der Gleichung $ax = b$ für $x \in K^n$ zu tun. Aus dem vorigen Abschnitt wissen wir, wie man $\operatorname{Kern} a$ berechnet. Bleibt also noch, wenigstens *eine* Lösung x von $ax = b$ zu finden. Auch das geht mit dem Gaußschen Algorithmus. Wie? Nun, im homogenen Fall haben wir uns zunutze gemacht, daß für jede Elementarmatrix u

$$\operatorname{Kern} ua = \operatorname{Kern} a,$$

das heißt für jedes $x \in K^n$

$$uax = 0 \iff ax = 0$$

gilt. Im inhomogenen Fall haben wir allgemeiner

$$ax = b \iff uax = ub.$$

Wir dürfen also ruhig a durch Zeilenumformungen verändern, solange wir b auf die gleiche Weise mitverändern. Praktisch geschieht das so, daß wir die $p \times n$-Matrix a und die Spalte b zu einer $p \times (n+1)$-Matrix

$$(a \quad b) = \begin{pmatrix} a_{11} & \dots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{p1} & \dots & a_{1n} & b_p \end{pmatrix}$$

zusammenfassen und diese den Zeilenumformungen gemäß dem Gaußschen Algorithmus unterwerfen.

21.4 Beispiel Die Gleichung

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

für $x \in K$ führt auf die Umformungskette

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

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

(der Teilungsstrich soll nur an die Sonderrolle der letzten Spalte erinnern). Aus der derart umgeformten Matrix lesen wir sofort eine Lösung x ab, nämlich

$$x_1 = 2, \quad x_2 = -1, \quad x_3 = 0.$$

Am linken Teil der Matrix sehen wir außerdem, daß der Kern eindimensional ist und von der durch

$$x_1 = -3, \quad x_2 = 1, \quad x_3 = 1$$

bestimmten Spalte x aufgespannt wird: damit ist

$$\begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} + \operatorname{Lin}\left(\begin{pmatrix} -3 \\ 1 \\ 1 \end{pmatrix}\right) \subset K^3$$

der vollständige Lösungsraum.

Nun zum allgemeinen

21.5 Verfahren Gegeben sei die Gleichung $ax = b$ für $x \in K^n$. Wir behandeln die Matrix $(a \mid b)$ durch Zeilenumformung nach dem Gaußschen Algorithmus soweit, bis immerhin die Teilmatrix links vom Teilungsstrich Zeilenstufenform hat. Nennen wir die neue Gesamtmatrix wieder $(a \mid b)$, so ist in

$$(a \mid b) = \left(\begin{array}{ccccccc|c} 1 & \dots & * & \dots & * & \dots & a_{1n} & b_1 \\ & & 1 & \dots & * & \dots & a_{2n} & b_2 \\ & & & \ddots & \vdots & & \vdots & \vdots \\ & & & & 1 & \dots & a_{rn} & b_r \\ & & & & & & & b_{r+1} \\ & & & & & & & \vdots \\ & & & & & & & b_p \end{array}\right)$$

die Stufenzahl r der Rang von a. Weil jetzt offenbar $\operatorname{Bild} a = K^r \times \{0\}$ ist, kann die Gleichung $ax = b$ nur dann lösbar sein, wenn

$$b_i = 0 \quad \text{für alle } i > r$$

gilt. Ist das der Fall, so hat die Gesamtmatrix $(a \mid b)$ Zeilenstufenform, und man veredelt zu

$$(a \mid b) = \left(\begin{array}{ccccccc|c} 1 & \dots & 0 & \dots & 0 & \dots & a_{1n} & b_1 \\ & & 1 & \dots & 0 & \dots & a_{2n} & b_2 \\ & & & \ddots & \vdots & & \vdots & \vdots \\ & & & & 1 & \dots & a_{rn} & b_r \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \end{array}\right).$$

Jetzt liest man sofort

$$x = \begin{pmatrix} b_1 \\ b_s \\ b_r \\ \\ \end{pmatrix} \begin{array}{l} \leftarrow j_1 \\ \leftarrow j_s \quad (1 < s < r) \\ \leftarrow j_r \end{array}$$

als eine Lösung ab; hier sind $j_1 < \dots < j_r$ die Stufenindizes, und es ist $x_j = 0$ für alle anderen Indizes j.

Bemerkungen Vergessen Sie nicht, daß man so nur *eine* Lösung findet und daß zur Beschreibung des Lösungsraum auch $\operatorname{Kern} a$ noch zu berechen ist. Aber die Umformung von a in eine veredelte Zeilenstufenform hat man an diesem Punkt ohnehin schon durchgeführt, und es bleibt bloß Lemma 20.10 anzuwenden. — Das beschriebene Verfahren enthält eine zu 20.11(3) alternative und flottere Methode, um zu entscheiden, ob ein gegebener Vektor (nämlich b) in einem durch aufspannende Vektoren (die Spalten von a) gegebenen linearen Teilraum enthalten ist. Auch lesen wir sofort ab:

21.5½ Notiz Die Gleichung $ax = b$ hat genau dann eine Lösung x, wenn $\operatorname{rk} a = \operatorname{rk} (a \mid b)$ gilt.

Eine besonders wichtige spezielle Situation ist die, daß das lineare Gleichungssystem $ax = b$ für jede Wahl von b eine eindeutig bestimmte Lösung x hat. Das heißt natürlich nichts Anderes, als daß a als Abbildung bijektiv, als Matrix invertierbar ist. In diesem Fall erhält man durch Multiplikation der Gleichung mit a^{-1} von links die Formel

$$x = a^{-1}b,$$

die die Lösung in Abhängigkeit von b ausdrückt. Damit stoßen wir — nicht zum erstenmal — auf die Frage, wie man einer Matrix $a \in \operatorname{Mat}(n \times n, K)$ ansieht, ob sie invertierbar ist, und wie man die inverse Matrix berechnen kann. Im Prinzip reicht dazu das bisher Gesagte aus, denn die beiden Matrixgleichungen $ax = 1$ und $xa = 1$ stellen ja ein inhomogenes lineares Gleichungssystem für die n^2 Einträge der gesuchten Inversen x dar. Aber allein die Vorstellung, dieses Gleichungssytem in einzelnen Koeffizienten oder mittels einer $2n^2 \times n^2$-Matrix hinzuschreiben, läßt einem kalte Schauer über den Rücken laufen. Tatsächlich braucht man das auch gar nicht zu tun; erstens erweist sich nämlich die Hälfte der Gleichungen als überflüssig, und zweitens läßt das verbleibende Problem sich in einer ganz kompakten und den Rechenaufwand drastisch reduzierenden Form behandeln. Zum ersten Punkt formulieren wir den folgenden inzwischen fast trivialen, aber sehr wichtigen

21.6 Satz Für jede *quadratische* Matrix $a \in \operatorname{Mat}(n \times n, K)$ sind die folgenden Aussagen äquivalent:

(a) $\operatorname{rk} a = n$

(b) $\operatorname{Kern} a = \{0\}$

(c) $a \in GL(n, K)$

(d) es gibt ein $x \in \mathrm{Mat}(n \times n, K)$ mit $ax = 1$

(e) es gibt ein $x \in \mathrm{Mat}(n \times n, K)$ mit $xa = 1$

Treffen diese Aussagen zu, so ist in (d) und (e) zwangläufig $x = a^{-1}$.

Beweis Die Dimensionsformel für a (als lineare Abbildung $K^n \longrightarrow K^n$) besagt $\dim \mathrm{Kern}\, a + \mathrm{rk}\, a = n$. Aus der Surjektivität (a) oder der Injektivität (b) von a folgt also die jeweils andere Eigenschaft automatisch, deshalb sind (a), (b) und (c) äquivalent. Andererseits folgt aus (d), daß a surjektiv, und aus (e), daß a injektiv ist, jede der beiden impliziert also (c) und ist deshalb ebenfalls zu (c) äquivalent. Schließlich erhält man im Fall der Invertierbarkeit aus (d) oder (e) die Gleichung $x = a^{-1}$, indem man von links bzw. von rechts mit a^{-1} multipliziert.

Es geht also bei gegebenem $a \in \mathrm{Mat}(n \times n, K)$ darum, die Gleichung $ax = 1$ für $x \in \mathrm{Mat}(n \times n, K)$zu lösen, und das macht man genau so wie bei einer inhomogenen Gleichung für eine Spalte x. Wenn wir nämlich von links mit einem Produkt von Elementarmatrizen u multiplizieren, erhalten wir die äquivalente Gleichung $uax = u$, und wenn wir nach dem Gaußschen Algorithmus vorgehen, hat ua dann Zeilenstufenform. An der sehen wir sofort, ob $\mathrm{rk}\, a = n$ und damit a überhaupt invertierbar ist oder nicht. Wenn ja, können wir gleich weitermachen, bis ua veredelte Stufenform hat. Da es sich bei ua aber um eine quadratische Matrix von vollem Rang handelt, ist jeder Index ein Stufenindex, also $ua = 1$ die Einheitsmatrix. Die zu $ax = 1$ nach wie vor äquivalente Gleichung $uax = u$ reduziert sich damit auf $x = u$. Das bedeutet das folgende praktische

21.7 Verfahren Die gegebene Matrix $a \in \mathrm{Mat}(n \times n, K)$ wird durch die $n \times n$-Einheitsmatrix zu

$$(a \mid 1) \in \mathrm{Mat}(n \times 2n, K)$$

erweitert; diese Matrix wird dem Gaußschen Algorithmus in der Zeilenversion bis zu dem Punkt unterworfen, wo die linke quadratische Teilmatrix Zeilenstufenform hat. Ist die Zahl von deren Stufen kleiner als n, so ist a nicht invertierbar und nichts weiter zu tun. Gibt es aber n Stufen, dann führt man noch den Veredelungsalgorithmus $20.9\frac{2}{3}$ durch. Die resultierende Matrix ist dann $(1 \mid u)$ mit $u = a^{-1}$.

21.8 Beispiel $a = \begin{pmatrix} 1 & 1 & -1 \\ 2 & 3 & 1 \\ 1 & 0 & -3 \end{pmatrix} \in \mathrm{Mat}(3 \times 3, K)$

Herstellung der Zeilenstufenform:

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

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

An dieser Stelle weiß man, daß a^{-1} existiert, und nach Veredelung

$$\left(\begin{array}{ccc|ccc} 1 & 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & 3 & -2 & 1 & 0 \\ 0 & 0 & 1 & -3 & 1 & 1 \end{array}\right) \longrightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & -4 & 3 & -1 & 0 \\ 0 & 1 & 3 & -2 & 1 & 0 \\ 0 & 0 & 1 & -3 & 1 & 1 \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -9 & 3 & 4 \\ 0 & 1 & 0 & 7 & -2 & -3 \\ 0 & 0 & 1 & -3 & 1 & 1 \end{array}\right)$$

liest man $a^{-1} = \begin{pmatrix} -9 & 3 & 4 \\ 7 & -2 & -3 \\ -3 & 1 & 1 \end{pmatrix}$ bequem ab. Man kann sich den Spaß machen, nachzurechnen, daß es sich wirklich um die inverse Matrix handelt. Daß die gesamte Rechnung im Bereich der ganzen Zahlen abgelaufen ist, ist natürlich — auch bei ganzzahligem a und selbst dann, wenn neben a auch a^{-1} ganzzahlig ausfallen sollte — ganz untypisch und nur der freundlichen Wahl des Beispiels zu verdanken.

In einem kleinen Anhang zu diesem Abschnitt will ich noch ein paar Worte über Differentialgleichungen sagen, die ja in der Physik allgegenwärtig sind. Wir betrachten nur den recht speziellen, aber schon wichtigen Fall einer linearen Differentialgleichung für eine reell- oder komplexwertige Funktion auf einem Intervall I, das wie immer, wenn differenziert wird, mindestens zwei Punkte enthalten sollte: Etwa sind insgesamt $n+1$ Funktionen

$$a_j\colon I \longrightarrow \mathbb{C} \quad (j = 0, 1, \ldots, n-1) \quad \text{sowie} \quad b\colon I \longrightarrow \mathbb{C},$$

gegeben, und gesucht sind Funktionen $f\colon I \longrightarrow \mathbb{C}$, die der Gleichung

$$f^{(n)} + a_{n-1}f^{(n-1)} + \cdots + a_1 f' + a_0 f = b$$

genügen. Man spricht dann von einer *Differentialgleichung n-ter Ordnung*. Die Frage nach den Stammfunktionen einer gegebenen Funktion b ist die nach den Lösungen der speziellen linearen Differentialgleichung erster Ordnung $f' = b$, während Differentialgleichungen nullter Ordnung natürlich uninteressant sind. In der Physik dominieren Differentialgleichungen erster oder zweiter Ordnung.

In welchem Sinne handelt es sich tatsächlich um lineare Gleichungen? Man wird vernünftigerweise voraussetzen, daß alle Koeffizientenfunktionen a_j ebenso wie b entweder C^k-Funktionen für eine natürliche Zahl k oder C^∞- oder gar analytische Funktionen sind. Dann kann man den Differentialoperator

$$D\colon f \mapsto Df := f^{(n)} + a_{n-1}f^{(n-1)} + \cdots + a_1 f' + a_0 f$$

als eine lineare Abbildung

$$D\colon C^{k+n}(I) \longrightarrow C^k(I) \quad \text{bzw.} \quad D\colon C^\infty(I) \longrightarrow C^\infty(I) \quad \text{bzw.} \quad D\colon \mathcal{O}(I) \longrightarrow \mathcal{O}(I)$$

ansehen, und die Frage ist die nach der Faser $D^{-1}\{b\}$. Nach Lemma 21.3 ist diese Faser ein affiner Unterraum des entsprechenden Funktionenraums, der, soweit nicht leer, zu dem linearen Unterraum Kern D parallel ist.

Da all diese Funktionenräume nicht endlichdimensional sind, kann man nicht einmal davon träumen, hier mit dem Gaußschen Algorithmus etwas ausrichten zu wollen. Vielmehr ist die Theorie der linearen Differentialgleichungen eine eigene, überwiegend von der Analysis und weniger von der linearen Algebra geprägte Wissenschaft, die ihrerseits einen nicht trivialen Teil der Theorie der Differentialgleichungen überhaupt (linear oder nicht) bildet. Aus dieser Theorie möchte ich einen grundlegenden Satz in der hier relevanten Form zitieren:

21.9 Satz $I \subset \mathbb{R}$ sei ein Intervall mit mindestens zwei Punkten, $t_0 \in I$ ein fest gewählter Punkt. Außerdem seien stetige Funktionen a_j für $j = 0, \ldots, n-1$, sowie eine weitere stetige Funktion b auf I gegeben (sagen wir mit komplexen Werten). Dann gibt es zu jedem Vektor

$$x = \begin{pmatrix} x_0 \\ \vdots \\ x_{n-1} \end{pmatrix} \in \mathbb{C}^n$$

genau eine Funktion $f \in C^n(I)$, die die Differentialgleichung

$$Df = f^{(n)} + a_{n-1}f^{(n-1)} + \cdots + a_1 f' + a_0 f = b$$

und die *Anfangsbedingungen*

$$f(t_0) = x_0,\ f'(t_0) = x_1,\ \ldots,\ f^{(n-1)}(t_0) = x_{n-1}$$

erfüllt. Entsprechend für reellwertige Funktionen, und wenn die Ausgangsdaten C^∞-Funktionen oder auf dem Intervall I gar analytisch sind, so gilt dasselbe für die Lösungen.

Bemerkung Satz 21.9 überträgt sich nicht auf Differentialgleichungen für komplexe Funktionen, die auf einem Gebiet definiert sind: Nach Aufgabe 16.1 hat die lineare Differentialgleichung erster Ordnung $f'(z) = \frac{1}{z}$ auf dem Gebiet $\mathbb{C}\setminus\{0\}$ keine Lösung.

Um den Satz in der Sprache der linearen Algebra zu formulieren, faßt man den Differentialoperator D wie gehabt als lineare Abbildung $D\colon C^n(I) \longrightarrow C^0(I)$ auf und betrachtet die Abbildung T, die jeder Funktion f die Werte ihrer Ableitungen bis zur Ordnung $n-1$ an der Stelle t_0 zuordnet:

$$C^n(I) \ni f \mapsto Tf := \begin{pmatrix} f(t_0) \\ \vdots \\ f^{(n-1)}(t_0) \end{pmatrix} \in \mathbb{C}^n$$

(gleichwertig könnte man auch das $(n-1)$-te Taylor-Polynom bei t_0 bilden). Diese Abbildung ist natürlich linear, und der Satz besagt, daß für jedes $b \in C^0(I)$ die Einschränkung von D auf die Faser

$$T|D^{-1}\{b\}\colon D^{-1}\{b\} \longrightarrow \mathbb{C}^n$$

bijektiv ist. Insbesondere handelt es sich im homogenen Fall ($b = 0$) um einen Vektorraumisomorphismus $\operatorname{Kern} D \simeq \mathbb{C}^n$.

21.10 Beispiel Die in den Aufgaben 15.1 und 15.2 untersuchte Gleichung des quantenmechanischen harmonischen Oszillators

$$f''(x) - 2xf'(x) + (2E-1)f(x) = 0$$

ist eine homogene lineare Differentialgleichung zweiter Ordnung mit analytischen Koeffizientenfunktionen. Der eben zitierte Satz rechtfertigt es, daß wir damals von vornherein nur nach analytischen Lösungen gefragt haben: jede C^2-Lösung ist automatisch analytisch. Die Lösungen selbst hatten wir durch einen Potenzreihenansatz konstruiert, und wie man jetzt sieht, ist es kein Zufall, daß wir den nullten und den ersten Koeffizienten der Potenzreihe frei wählen konnten und daß dadurch alle weiteren Koeffizienten festgelegt waren; vielmehr ist das (im wesentlichen) die Existenz- und Eindeutigkeitsaussage des Satzes.

Übrigens kann man bei analytischen Koeffizientenfunktionen immer mit einem solchen Potenzreihenansatz arbeiten, und das ist eigentlich auch die einzige allgemein anwendbare Methode, die Lösungen einer linearen Differentialgleichung zu berechnen. Freilich muß man als Lösung dann eine Rekursionsformel für die Taylor-Koeffizienten akzeptieren, aber wie das Beispiel ja schön illustriert hat, läßt sich daraus oft mehr Information herausziehen als man zunächst meint. In speziell gelagerten Fällen gibt es aber auch explizite Lösungsformeln, und die folgende sollte jeder kennen.

21.11 Satz Zu gegebenen komplexen Zahlen $a_0, a_1, \dots, a_{n-1} \in \mathbb{C}$ werde die homogene lineare Differentialgleichung mit *konstanten* Koeffizienten

$$Df = f^{(n)} + a_{n-1}f^{(n-1)} + \cdots + a_1f' + a_0f = 0$$

betrachtet. Es seien

$$c_1, \dots, c_r \in \mathbb{C}$$

die Nullstellen des Polynoms $p(X) = X^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0 \in \mathbb{C}[X]$, und

$$e_1, \dots, e_r \in \mathbb{N} \setminus \{0\}$$

ihre Vielfachheiten. Dann bilden die Funktionen

$$\begin{array}{llll} t \mapsto \exp c_1t, & t \mapsto t\exp c_1t, & \dots, & t \mapsto t^{e_1-1}\exp c_1t; \\ t \mapsto \exp c_2t, & t \mapsto t\exp c_2t, & \dots, & t \mapsto t^{e_2-1}\exp c_2t; \\ \vdots & \vdots & & \vdots \\ t \mapsto \exp c_rt, & t \mapsto t\exp c_rt, & \dots, & t \mapsto t^{e_r-1}\exp c_rt; \end{array}$$

eine Basis des Lösungsraums Kern D.

Beweis Das Polynom p zerfällt voraussetzungsgemäß in

$$p(X) = (X - c_1)^{e_1}(X - c_2)^{e_2} \cdots (X - c_r)^{e_r},$$

und weil die $c_j \in \mathbb{C}$ konstant sind, spricht nichts dagegen, auch den Differentialoperator D entsprechend zu zerlegen, nämlich als Komposition

$$D = p\left(\frac{d}{dt}\right) = \left(\frac{d}{dt} - c_1\right)^{e_1} \left(\frac{d}{dt} - c_2\right)^{e_2} \cdots \left(\frac{d}{dt} - c_r\right)^{e_r}$$

zu schreiben; dabei dürfen wir uns die Reihenfolge der Faktoren nach Bedarf aussuchen. Nach bekannten Formeln gilt nun für jede differenzierbare Funktion h

$$\left(\frac{d}{dt} - c_j\right) h(t) \exp c_j t = h'(t) \exp c_j t + h(t) c_j \exp c_j t - c_j h(t) \exp c_j t = h'(t) \exp c_j t,$$

und daraus sieht man, daß die angegebenen Funktionen tatsächlich im Kern von D liegen. Weil dieser Kern nach Satz 21.9 die Dimension n hat, bleibt nur noch zu zeigen, daß diese n Funktionen linear unabhängig sind. Dazu kann man zum Beispiel ihr Wachstum längs geeigneter Strahlen in der komplexen Zahlenebene untersuchen (wie in Aufgabe 18.3 angedeutet).

21.12 Beispiel Der Fall $n = 2$ mit nicht-negativen reellen Koeffizienten ist den Physikern als gedämpfter (klassischer) harmonischer Oszillator vertraut: Die Bewegungsgleichung für die Observable f (zum Beispiel eine Ortskoordinate oder elektrische Spannung) ist

$$f'' + 2\gamma f' + \omega^2 f = 0,$$

worin γ der Dämpfungs- und ω^2 der Rückstellfaktor ist. Das zugehörige Polynom p hat die komplexe Zerlegung

$$p(X) = X^2 + 2\gamma X + \omega^2 = \left(X + \gamma - \sqrt{\gamma^2 - \omega^2}\right)\left(X + \gamma + \sqrt{\gamma^2 - \omega^2}\right),$$

und man hat drei wesentlich verschiedene Fälle:

- $\gamma < \omega$ (schwache Dämpfung): $\sqrt{\gamma^2 - \omega^2}$ ist genauer als $i\omega'$ mit $\omega' = \sqrt{\omega^2 - \gamma^2} \in (0, \infty)$ zu interpretieren, und die Basislösungen
$$t \mapsto e^{-\gamma t} e^{\pm i\omega' t}$$
beschreiben echte (für $\gamma > 0$ gedämpfte) Schwingungen, vergleiche Aufgabe 14.5.
- $\gamma > \omega$ (starke Dämpfung): Die Basislösungen
$$t \mapsto e^{\left(-\gamma \pm \sqrt{\gamma^2 - \omega^2}\right) t}$$
haben einen aperiodischen Verlauf.
- $\gamma = \omega$ (Kriechfall): Beide Linearfaktoren sind gleich, und die Basislösungen sind
$$t \mapsto e^{-\gamma t} \quad \text{und} \quad t \mapsto t e^{-\gamma t}.$$

Die Tatsache, daß der Lösungsraum in jedem Fall ein Vektorraum ist, hat übrigens die bekannte Bedeutung, daß man durch Überlagerung von (freien) Schwingungen wieder eine solche erhält.

Übungsaufgaben

21.1 V sei ein Vektorraum; $A, A' \subset V$ seien affine Teilräume. Beweisen Sie: $A \cap A'$ und

$$A + A' := \{x + y \,|\, x \in A,\, y \in A'\}$$

sind ebenfalls affine Teilräume von V. Ist $f\colon V \to W$ linear und $B \subset W$ ein weiterer affiner Teilraum, so sind auch $f(A) \subset W$ und $f^{-1}B \subset V$ affine Teilräume von V.

21.2 Der Vektorraum V sei direkte Summe der beiden linearen Unterräume L und M. Beweisen Sie: Für jede Wahl von $a, b \in V$ besteht der Durchschnitt der beiden affinen Teilräume $a + L$ und $b + M$ aus genau einem Punkt.

21.3 Berechnen Sie alle Lösungen x des linearen Gleichungssystems

$$\begin{array}{rrrrcr} 3x_1 & & +x_3 & +2x_4 & = & -1 \\ x_1 & +2x_2 & +x_3 & & = & 1\,, \end{array}$$

die in dem Unterraum

$$S = \mathrm{Lin}\left(\begin{pmatrix} 1 \\ 0 \\ -2 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 1 \\ -6 \end{pmatrix}\right) \subset \mathbb{R}^4$$

enthalten sind.

21.4 Die Matrix

$$a = \begin{pmatrix} 1 & 3 & 4 \\ 2 & 5 & 6 \end{pmatrix} \in \mathrm{Mat}(2\times 3, \mathbb{R})$$

wird als eine lineare Abbildung $a\colon \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ aufgefaßt. Bestimmen Sie eine Basis $\underline{v}$ von $\mathbb{R}^3$, so daß a bezüglich $\underline{v}$ und der Standardbasis von $\mathbb{R}^2$ durch die Matrix

$$c = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \in \mathrm{Mat}(2\times 3, \mathbb{R})$$

beschrieben wird. (Man kann von vornherein sicher sein, daß es eine solche Basis $\underline{v}$ gibt — warum?)

21.5 Sei K ein Körper. Für $n \in \mathbb{N}$ und $\lambda \in K$ werde die Matrix

$$a = \begin{pmatrix} 1 & -\lambda & & & & \\ & 1 & -\lambda & & & \\ & & \ddots & \ddots & & \\ & & & 1 & -\lambda & \\ & & & & 1 & -\lambda \\ & & & & & 1 \end{pmatrix} \in \mathrm{Mat}(n\times n, K)$$

betrachtet (also $a_{jj} = 1$, $a_{j,j+1} = -\lambda$ und $a_{jk} = 0$ für $j \neq k \neq j+1$). Begründen Sie, warum a invertierbar ist, und berechnen Sie a^{-1}.

22 Die Determinante

Jede quadratische Matrix $a \in \mathrm{Mat}(n \times n, K)$ hat eine sogenannte Determinante $\det a \in K$, und die wollen wir jetzt kennenlernen. Obwohl es möglich wäre, zur Definition einfach eine Formel hinzuschreiben, ist es aufschlußreicher, wenn wir indirekt vorgehen, indem wir nicht die einzelnen Zahlen $\det a$, sondern gleich die ganze Determinantenfunktion

$$\mathrm{Mat}(n \times n, K) \longrightarrow K, \quad a \mapsto \det a$$

ins Auge fassen. Wir werden sehen, daß diese Funktion sich durch ein paar einfache Axiome charakterisieren läßt. Weil diese Axiome speziell auf die Spalten der Matrix a Bezug nehmen, wollen wir für diese durch $a = (\, a_1 \quad a_2 \quad \dots \quad a_n \,)$, also

$$a_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{pmatrix} \in K^n$$

die naheliegende Bezeichnung fixieren.

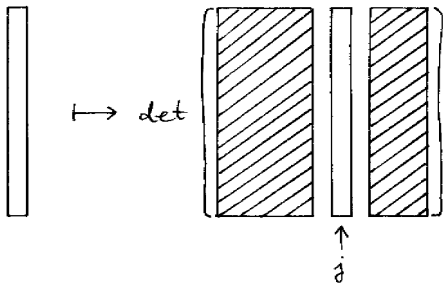
22.1 Satz und **Definition** Es sei K ein Körper und $n \in \mathbb{N}$. Es gibt genau eine Abbildung

$$\det\colon \mathrm{Mat}(n \times n, K) \longrightarrow K$$

mit den folgenden Eigenschaften:

(a) det ist linear in jeder Spalte: Für jede Wahl von $j \in \{1, \dots, n\}$ und $a_1, \dots, a_{j-1}, a_{j+1}, \dots, a_n \in K^n$ ist

$$K^n \ni x \mapsto \det (\, a_1 \quad \dots \quad a_{j-1} \quad x \quad a_{j+1} \quad \dots \quad a_n \,) \in K$$



eine lineare Funktion.

(b) Enthält a zwei gleiche Spalten, so ist $\det a = 0$.

(c) $\det 1 = 1$.

Diese Abbildung det heißt die Determinantenfunktion oder einfach Determinante (während man mit der Determinante *von* a natürlich den Wert $\det a$ meint).

Zum Beweis des Satzes ist es praktisch, jede Funktion $d\colon \mathrm{Mat}(n \times n, K) \longrightarrow K$ mit den beiden ersten Eigenschaften (a) und (b) vorübergehend eine *Prädeterminante* zu nennnen. Für diese zeigen wir zuerst:

22.2 Lemma Jede Prädeterminante $d\colon \mathrm{Mat}(n \times n, K) \longrightarrow K$ verhält sich unter elementaren Spaltenumformungen wie folgt:

$$d(a \cdot p_{kl}) = -d(a)$$
$$d(a \cdot d_{k\lambda}) = \lambda \cdot d(a)$$
$$d(a \cdot u_{kl\lambda}) = d(a)$$

Bemerkung Das Lemma ist von dauerhaftem Interesse, denn wenn Satz 22.1 einmal bewiesen ist, gilt die Aussage ja auch für die Determinante.

Beweis Die zweite Regel folgt unmittelbar daraus, daß d linear in der k-ten Spalte ist:

$$\begin{aligned} d(a\cdot d_{k\lambda}) &= d\,(\,a_1 \quad \dots \quad a_{k-1} \quad \lambda a_k \quad a_{k+1} \quad \dots \quad a_n\,) \\ &= \lambda\cdot d\,(\,a_1 \quad \dots \quad a_{k-1} \quad a_k \quad a_{k+1} \quad \dots \quad a_n\,) \\ &= \lambda\cdot d(a). \end{aligned}$$

Zum Beweis der dritten rechnen wir

$$\begin{aligned} d(a\cdot u_{kl\lambda}) &= d\,(\,a_1 \quad \dots \quad a_{l-1} \quad a_l+\lambda a_k \quad a_{l+1} \quad \dots \quad a_n\,) \\ &= d\,(\,a_1 \quad \dots \quad a_{l-1} \quad a_l \quad a_{l+1} \quad \dots \quad a_n\,) + \lambda\cdot \underbrace{d\,(\,a_1 \quad \dots \quad a_{l-1} \quad a_k \quad a_{l+1} \quad \dots \quad a_n\,)}_{=0 \text{ nach(b)}} \\ &= d\,(\,a_1 \quad \dots \quad a_{l-1} \quad a_l \quad a_{l+1} \quad \dots \quad a_n\,) \\ &= d(a). \end{aligned}$$

Schließlich führen wir die erste Regel durch Raffinesse auf die beiden anderen zurück. Die Wirkung von p_{kl} auf a, also die Vertauschung der k-ten mit der l-ten Spalte, läßt sich nämlich auch durch folgende Kette von Elementarumformungen erreichen:

$$\begin{aligned} &(\dots \quad a_k \quad \dots \quad a_l \quad \dots) \xrightarrow{u_{kl,1}} (\dots \quad a_k \quad \dots \quad a_k+a_l \quad \dots) \\ &\xrightarrow{d_{k,-1}} (\dots \quad -a_k \quad \dots \quad a_k+a_l \quad \dots) \xrightarrow{u_{lk,1}} (\dots \quad a_l \quad \dots \quad a_k+a_l \quad \dots) \\ &\xrightarrow{u_{kl,-1}} (\dots \quad a_l \quad \dots \quad a_k \quad \dots) \end{aligned}$$

(es gilt also $u_{kl,1}\cdot d_{k,-1}\cdot u_{lk,1}\cdot u_{kl,-1} = p_{kl}$). Der Wert von d ändert sich nur bei der zweiten Umformung, und zwar um den Faktor -1. Damit ist das Lemma bewiesen.

Beweis von Satz 22.1 Wir beweisen die Eindeutigkeit auf eine ganz praktische Art: Wir leiten allein aus den Eigenschaften (a), (b) und (c) ein Verfahren her, das es erlaubt, die Determinante einer gegebenen Matrix zu berechnen — wenn es sie gibt.

Sei also d *eine* Determinantenfunktion (so müssen wir uns ja ausdrücken, solange die Eindeutigkeit noch nicht bewiesen ist), und sei $a \in \mathrm{Mat}(n\times n, K)$ gegeben. Wir bearbeiten a nach dem Gaußschen Algorithmus (Spaltenversion) und erhalten eine Matrix av in Spaltenstufenform. Dabei führen wir über die durchgeführten Elementarumformungen in der Weise Buch, daß wir einen mit 1 initialisierten "Korrekturfaktor" bei der Anwendung von $d_{k\lambda}$ mit λ und bei Anwendung von p_{kl} mit -1 multiplizieren. Ist $\mu \in K\setminus\{0\}$ der Endwert dieses Faktors, so gilt nach Lemma 22.2

$$d(av) = \mu\cdot d(a).$$

Hat nun av weniger als n Stufen, ist also die letzte Spalte von av eine Nullspalte, so ist wegen der Linearität von d in dieser Spalte $d(av) = 0$ und damit auch $d(a) = \frac{1}{\mu}d(av) = 0$. Hat av dagegen n Stufen — was natürlich genau dann passiert, wenn a invertierbar ist — dann können wir av durch Veredelung zur Einheitsmatrix machen, wobei nur noch Spaltenumformungen vom Typ $u_{kl\lambda}$ verwendet werden. Bezeichnen wir das Produkt aller verwendeten Elementarmatrizen erneut mit v, so gilt daher immer noch $d(av) = \mu\cdot d(a)$, aber außerdem $av = 1$ und damit

$$d(a) = \frac{1}{\mu}d(av) = \frac{1}{\mu}d(1) = \frac{1}{\mu}$$

gemäß Eigenschaft (c).

Damit ist der Eindeutigkeitsbeweis geführt, und wir dürfen ab jetzt det statt d schreiben, selbst wenn die Existenz noch offen ist. Wir merken uns aus diesem Teil des Beweises die wichtige

22.3 Notiz Es ist $\det a \neq 0$ genau dann, wenn a invertierbar ist.

Bemerkung Der Beweis hat noch etwas mehr gezeigt als nur die Eindeutigkeit der Determinante, nämlich daß jede Prädeterminante d ein Vielfaches der Determinante sein muß:

$$d = d(1) \cdot \det .$$

Denn für die Erkenntnis $d(a) = \frac{1}{\mu} d(1)$ wurde von der Normierungseigenschaft (c) noch kein Gebrauch gemacht.

Bevor wir fortfahren, zu dem beschriebenen Verfahren ein konkretes

22.4 Beispiel Berechnung von $\det \begin{pmatrix} 1 & 3 & 4 & 5 \\ -1 & 0 & -1 & 2 \\ 0 & -1 & -1 & -2 \\ 1 & 1 & -2 & 2 \end{pmatrix}$ (unter dem Vorbehalt der Existenz):

$$\begin{pmatrix} 1 & 3 & 4 & 5 \\ -1 & 0 & -1 & 2 \\ 0 & -1 & -1 & -2 \\ 1 & 1 & -2 & 2 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 3 & 3 & 7 \\ 0 & -1 & -1 & -2 \\ 1 & -2 & -6 & -3 \end{pmatrix} \xrightarrow{\cdot\frac{1}{3}} \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 3 & 7 \\ 0 & -\frac{1}{3} & -1 & -2 \\ 1 & -\frac{2}{3} & -6 & -3 \end{pmatrix}$$

$$\longrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -\frac{1}{3} & 0 & \frac{1}{3} \\ 1 & -\frac{2}{3} & -4 & \frac{5}{3} \end{pmatrix} \xrightarrow{\cdot(-1)} \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -\frac{1}{3} & \frac{1}{3} & 0 \\ 1 & -\frac{2}{3} & \frac{5}{3} & -4 \end{pmatrix}$$

$$\xrightarrow{\cdot 3} \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -\frac{1}{3} & 1 & 0 \\ 1 & -\frac{2}{3} & 5 & -4 \end{pmatrix} \xrightarrow{\cdot\left(-\frac{1}{4}\right)} \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -\frac{1}{3} & 1 & 0 \\ 1 & -\frac{2}{3} & 5 & 1 \end{pmatrix}$$

Weiter brauchen wir nicht zu machen: Veredelung dieser letzten Matrix ändert ihre Determinante nicht mehr, sie ist also schon 1. Für die Ausgangsmatrix a ergibt sich damit

$$\det a = \left(\frac{1}{3} \cdot (-1) \cdot 3 \cdot \left(-\frac{1}{4} \right) \right)^{-1} = 4.$$

Beweis von Satz 22.1 (Fortsetzung) Wir führen den Existenzbeweis, indem wir für jedes $n \in \mathbb{N}$ eine Funktion $\det\colon \mathrm{Mat}(n \times n, K) \longrightarrow K$ mit den Eigenschaften (a), (b) und (c) konstruieren, und zwar durch vollständige Induktion nach n.

Den Induktionsanfang leistet die Funktion $\det\colon \mathrm{Mat}(0 \times 0, K) \longrightarrow K$, die der leeren Matrix die Zahl 1 zuordnet. Im Induktionsschritt (von $n-1$ auf n) dürfen wir von der bereits konstuierten Funktion

$$\det\colon \mathrm{Mat}\left((n-1) \times (n-1), K\right) \longrightarrow K$$

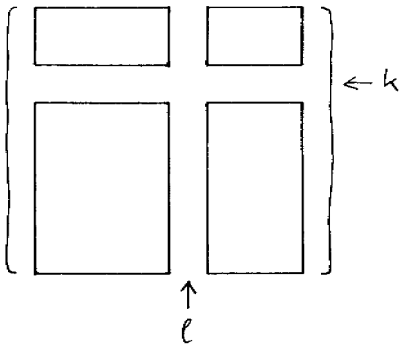
ausgehen, und wir definieren

$$\det\colon \mathrm{Mat}(n \times n, K) \longrightarrow K$$

wie folgt. Für zunächst ganz beliebige Matrizen $a \in \mathrm{Mat}(p \times n, K)$ bezeichne

$$\hat{a}_{kl} \in \mathrm{Mat}\left((p-1) \times (n-1), K\right)$$

die aus a durch Weglassen der k-ten Zeile und der l-ten Spalte entstehende Matrix:



In Formeln also

$$(\hat{a}_{kl})_{ij} = \begin{cases} a_{ij} & \text{für } i < k,\ j < l; \\ a_{i+1,j} & \text{für } i \geq k,\ j < l; \\ a_{i,j+1} & \text{für } i < k,\ j \geq l; \\ a_{i+1,j+1} & \text{für } i \geq k,\ j \geq l. \end{cases}$$

Zur Definition der Determinantenfunktion fixieren wir nun willkürlich einen Zeilenindex k und setzen

$$\det a = \sum_{l=1}^{n} (-1)^{k+l} a_{kl} \det \hat{a}_{kl};$$

die Verwendung der rechts stehenden Determinante einer $(n-1)\times(n-1)$-Matrix ist durch die Induktionsannahme gerechtfertigt. Wir verifizieren die Eigenschaften (a), (b) und (c).

(a) Linearität in den Spalten, etwa in der j-ten:

$$a_j \mapsto (-1)^{k+j} a_{kj} \det \hat{a}_{kj}$$

ist linear, weil in $\hat{a}_{kj}$ die j-te Spalte nicht vorkommt. Dagegen ist

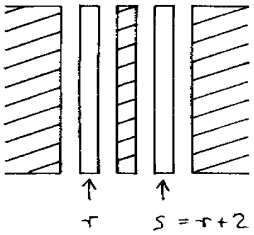
$$a_j \mapsto (-1)^{k+l} a_{kl} \det \hat{a}_{kl}$$

für $l \neq j$ linear, weil jetzt der Faktor $(-1)^{k+l} a_{kl}$ nicht von a_j abhängt und $a_j \mapsto \det \hat{a}_{kl}$ nach Induktionsannahme linear ist. Damit ist auch det als Summe dieser Ausdrücke in a_j linear.

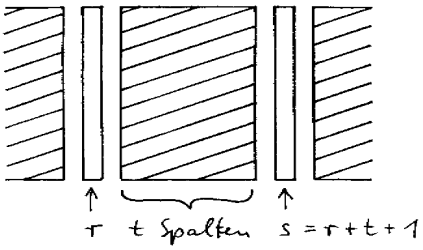
(b) a enthalte zwei gleiche Spalten, etwa $a_r = a_s$ mit $r < s$. Dann ist

$$\det a = (-1)^{k+r} a_{kr} \det \hat{a}_{kr} + (-1)^{k+s} a_{ks} \det \hat{a}_{ks},$$

denn in den übrigen Summanden enthält auch $\hat{a}_{kl}$ zwei gleiche Spalten, so daß nach Induktionsannahme $\det \hat{a}_{kl} = 0$ ist. Wir wollen $\det \hat{a}_{kr}$ und $\det \hat{a}_{ks}$ miteinander vergleichen. Sind r und s benachbart, ist also $s = r + 1$, so ist $\hat{a}_{kr} = \hat{a}_{ks}$ und folglich $\det \hat{a}_{kr} = \det \hat{a}_{ks}$. Ist dagegen $s = r + 2$, so läßt sich $\hat{a}_{kr}$ durch eine einzelne Spaltenvertauschung in $\hat{a}_{ks}$ überführen:



Nach Induktionsannahme gilt dann $\det \hat{a}_{ks} = -\det \hat{a}_{kr}$. Ist allgemein $s = r + t + 1$ mit beliebigem $t \geq 0$, so läßt $\hat{a}_{kr}$ sich durch t aufeinanderfolgende Spaltenvertauschungen in $\hat{a}_{ks}$ verwandeln,



und wir haben

$$\det \hat{a}_{ks} = (-1)^t \det \hat{a}_{kr}.$$

Jetzt brauchen wir bloß noch einzusetzen und erhalten

$$\det a = (-1)^{k+r} a_{kr} \det \hat{a}_{kr} + (-1)^{k+s}(-1)^t \det \hat{a}_{kr} = \big((-1)^{k+r} + (-1)^{k+r+t+1}(-1)^t\big) \det \hat{a}_{kr} = 0.$$

(c) Die Formel für die Determinante der Einheitsmatrix reduziert sich sofort auf

$$\det 1 = \sum_{l=1}^{n} (-1)^{k+l} 1_{kl} \det \hat{1}_{kl} = \det \hat{1}_{kk}.$$

Aber $\hat{1}_{kk}$ ist selbst die $(n-1)\times(n-1)$-Einheitsmatrix, und es folgt $\det 1 = 1$.

Damit ist Satz 22.1 vollständig bewiesen.

Auch der Existenzbeweis hat eine Methode zur Berechnung der Determinante beigetragen, nämlich die

22.5 Formel (Entwicklung nach der k-ten Zeile) $\det a = \sum_{l=1}^{n} (-1)^{k+l} a_{kl} \det \hat{a}_{kl}$.

Anders als der Gaußsche Algorithmus ist das eine explizite Formel für die Determinante. Zum Beispiel liefert Entwicklung nach der ersten Zeile außer dem trivialen Fall $n = 1$ schnell die bekannten Regeln für $n = 2$

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

und $n = 3$

$$\begin{aligned} \det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} &= a \cdot \det \begin{pmatrix} e & f \\ h & i \end{pmatrix} - b \cdot \det \begin{pmatrix} d & f \\ g & i \end{pmatrix} + c \cdot \det \begin{pmatrix} d & e \\ g & h \end{pmatrix} \\ &= aei + bfg + cdh - afh - bdi - ceg, \end{aligned}$$

die man sich mit einem Blick auf die Diagonalen der Matrix und ihre Parallelen merken mag. Mit wachsendem Matrizenformat steigt der Rechenaufwand im allgemeinen aber schnell an (man erhält insgesamt $n!$ Terme), so daß die Formel für die numerische Rechnung nur in speziellen Fällen geeignet ist. Dagegen ist die Entwicklungsformel für theoretische Anwendungen sehr nützlich.

Was wäre, wenn wir in der Definition der Determinante statt mit den Spalten konsequent mit den Zeilen gearbeitet hätten, wie es zum Beispiel Jänich in seinem Buch macht? Wir werden gleich sehen, daß genau dasselbe herausgekommen wäre. Der Übersichtlichkeit dabei dient die

22.6 Definition Sei $a \in \mathrm{Mat}(p\times n, K)$ eine Matrix. Dann heißt

$$a^{\mathrm{t}} \in \mathrm{Mat}(n\times p, K), \quad \big(a^{\mathrm{t}}\big)_{ij} := a_{ji}$$

die zu a transponierte Matrix.

$$a = \begin{pmatrix} a_{11} & \dots & \dots & a_{1n} \\ \vdots & & & \vdots \\ a_{p1} & \dots & \dots & a_{pn} \end{pmatrix} \qquad a^{\mathrm{t}} = \begin{pmatrix} a_{11} & \dots & a_{p1} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ a_{1n} & \dots & a_{pn} \end{pmatrix}$$

Die Matrix erscheint also an einer mit 45° fallenden Linie gespiegelt.

Da wir gewohnt sind, Matrizen als lineare Abbildungen zu interpretieren, werden wir uns natürlich fragen, was das Transponieren einer Matrix für die zugehörige lineare Abbildung bedeutet. Das ist merkwürdigerweise viel komplizierter zu erklären als das doch wirklich simple Transponieren selbst, und ich werde im Abschnitt 27 darauf zurückkommen. Bis dahin wollen wir das Transponieren als eine bloß formale Manipulation von Matrizen ansehen. — Evident sind die

22.7 Regeln

$$\begin{aligned} \left(a^{\mathrm{t}}\right)^{\mathrm{t}} &= a \\ (a+b)^{\mathrm{t}} &= a^{\mathrm{t}} + b^{\mathrm{t}} \\ (ab)^{\mathrm{t}} &= b^{\mathrm{t}} a^{\mathrm{t}} \\ \operatorname{rk} a &= \operatorname{rk} a^{\mathrm{t}} \end{aligned}$$

Insbesondere ist a^{t} genau dann invertierbar, wenn a das ist, und dann gilt

$$\left(a^{\mathrm{t}}\right)^{-1} = \left(a^{-1}\right)^{\mathrm{t}}.$$

Zurück zur Determinante:

22.8 Satz Für jede Matrix $a \in \mathrm{Mat}(n \times n, K)$ gilt

$$\det a = \det a^{\mathrm{t}}.$$

Beweis Die Funktion

$$\mathrm{Mat}(n \times n, K) \longrightarrow K, \quad a \mapsto \det a^{\mathrm{t}}$$

hat die Eigenschaften (a), (b), (c) der Determinantenfunktion. In der Tat sieht man der Entwicklungsformel

$$\det a^{\mathrm{t}} = \sum_{l=1}^{n} (-1)^{k+l} a_{lk} \det \left(\hat{a_{lk}}\right)^{\mathrm{t}}$$

direkt an, daß $\det a^{\mathrm{t}}$ linear von der k-ten Spalte von a abhängt, denn von dieser ist $\hat{a}_{lk}$ und damit $\det \hat{a}_{lk}$ ja unabhängig. Und wenn a zwei gleiche Spalten hat, ist sicher $\operatorname{rk} a < n$ und damit auch $\operatorname{rk} a^{\mathrm{t}} < n$, nach der Notiz 22.3 also $\det a^{\mathrm{t}} = 0$. Schließlich ist auch $\det 1^{\mathrm{t}} = \det 1 = 1$.

Nach Satz 22.1 bleibt der Funktion $a \mapsto \det a^{\mathrm{t}}$ nichts übrig, als mit $a \mapsto \det a$ übereinzustimmen.

22.9 Folgerung Die Determinante verhält sich unter Zeilenumformungen so, wie in Lemma 22.2 für die entsprechenden Spaltenumformungen beschrieben. Man kann die Determinante von a auch durch Entwicklung nach der l-ten Spalte berechnen:

$$\det a = \sum_{k=1}^{n} (-1)^{k+l} a_{kl} \det \hat{a}_{kl}$$

Besonders wichtig ist es, das Verhalten der Determinante unter Matrizenmultiplikation zu kennen.

22.10 Satz und **Definition** Für alle $a, b \in \mathrm{Mat}(n\times n, K)$ gilt

$$\det ab = \det a \cdot \det b.$$

Insbesondere definiert die Determinante einen Gruppenhomomorphismus

$$GL(n,K) \xrightarrow{\det} K\setminus\{0\}$$

in die multiplikative Gruppe des Körpers. Der Kern dieses Homomorphismus

$$SL(n,K) := \{u \in \mathrm{Mat}(n\times n, K) \mid \det u = 1\} \subset GL(n,K)$$

(die Faser über $1 \in K\setminus\{0\}$) heißt spezielle lineare Gruppe.

Beweis Wir greifen noch einmal auf das Konzept der Prädeterminanten zurück. Wenn wir die j-te Spalte der Matrix b mit b_j bezeichnen, ist ab_j die j-te Spalte von ab. Deshalb ist bei festem $a \in \mathrm{Mat}(n\times n, K)$ die Funktion

$$\mathrm{Mat}(n\times n, K) \xrightarrow{d} K, \quad d(b) = \det ab$$

eine Prädeterminante. Aufgrund der Bemerkung in Anschluß an 22.3 ist also $d = d(1)\cdot\det = \det a \cdot \det$, und das war's schon.

22.11 Folgerung Es gilt $\det a^{-1} = \dfrac{1}{\det a}$ für jedes $a \in GL(n,K)$.

Beweis $\det a \cdot \det a^{-1} = \det aa^{-1} = \det 1 = 1$

Mittels der Determinante läßt sich auch eine explizite Formel für die zu einer Matrix $a \in GL(n,K)$ inverse angeben.

22.12 Definition Sei $a \in \mathrm{Mat}(n\times n, K)$ eine quadratische Matrix. Die durch

$$\tilde{a}_{kl} := (-1)^{k+l} \det \hat{a}_{lk}$$

definierte Matrix $\tilde{a}$ desselben Formats heißt die Adjunkte von a (achten Sie auf die Vertauschung der Indizes).

22.13 Satz Für jedes $a \in \mathrm{Mat}(n\times n, K)$ gilt

$$a\tilde{a} = \tilde{a}a = \det a \cdot 1.$$

Ist a invertierbar, so ist also

$$a^{-1} = \frac{1}{\det a} \cdot \tilde{a}.$$

Beweis Es ist

$$(a\tilde{a})_{ik} = \sum_{j=1}^{n} a_{ij}\tilde{a}_{jk} = \sum_{j=1}^{n} a_{ij}(-1)^{j+k} \det \hat{a}_{kj} = \sum_{j=1}^{n} (-1)^{k+j} a_{ij} \det \hat{a}_{kj}.$$

Nach der Zeilenentwicklungsformel (für die k-te Zeile) ist das für $i = k$ gerade $\det a$, und für $i \neq k$ die Determinante derjenigen Matrix, die aus a dadurch entsteht, daß man die k-te Zeile durch die i-te ersetzt. Diese Matrix mit zwei gleichen Zeilen hat aber die Determinante null.

Es ist also $(a\tilde{a})_{ik} = \det a \cdot 1_{ik}$, das heißt $a\tilde{a} = \det a \cdot 1$. Genauso verifiziert man $\tilde{a}a = \det a \cdot 1$ (für invertierbares a folgt es nach Satz 21.6 automatisch).

22.14 Beispiel Die zu einer 2×2-Matrix inverse läßt sich auf diese Weise leicht hinschreiben:

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}^{-1} = \frac{1}{\alpha\delta - \beta\gamma} \begin{pmatrix} \delta & -\beta \\ -\gamma & \alpha \end{pmatrix}$$

Als allgemeine Methode zur Bestimmung der Inversen ist die Formel wegen der vielen darin zu berechnenden Determinanten nicht so geeignet. Ihre Bedeutung liegt mehr im theoretischen Bereich, unter anderem, weil sie von vornherein eine Auskunft über die zu erwarteten Nenner gibt. Übrigens enthält Satz 22.13 die in der Schule beliebte sogenannte Cramersche Regel zur Lösung quadratischer linearer Gleichungssysteme von vollem Rang: Ist $ax = b$ mit $a \in GL(n, K)$ ein solches System, so ist dessen eindeutige Lösung ja

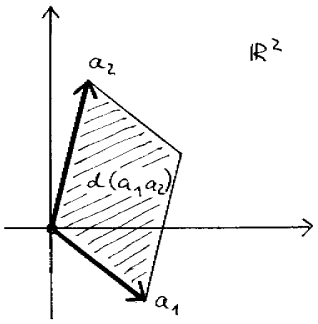
$$x = a^{-1}b = \frac{1}{\det a} \cdot \tilde{a}b,$$

für die Komponenten von x ergibt sich also

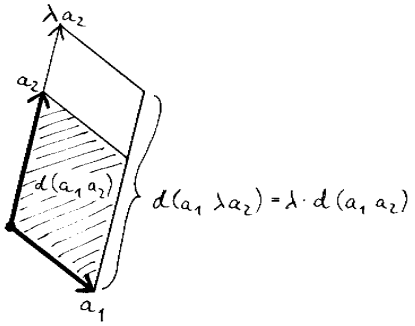
$$x_i = \frac{1}{\det a} \sum_{j=1}^{n} \tilde{a}_{ij} b_j = \frac{1}{\det a} \sum_{j=1}^{n} (-1)^{i+j} b_j \det \hat{a}_{ji} = \frac{\det c_i}{\det a},$$

wobei (Entwicklung nach der i-ten Spalte) die $n \times n$-Matrix c_i aus a dadurch entsteht, daß die i-te Spalte durch b ersetzt wird.

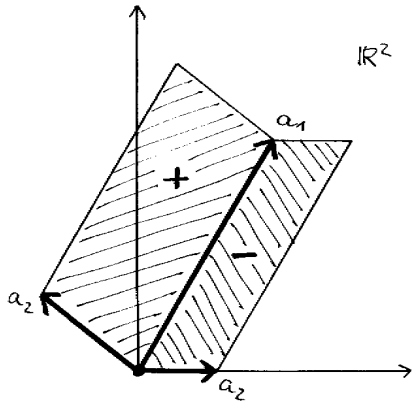
Soweit haben Sie die Determinante als ein rein algebraisches Objekt kennengelernt. Sie hat im reellen Fall aber auch eine sehr anschauliche geometrische Bedeutung, die ich zuerst in der zweidimensionalen, also ebenen Situation erklären will. Zwei Spalten $a_1, a_2 \in \mathbb{R}^2$ interpretieren wir wie üblich als Vektoren in der Ebene, und wir wollen die Funktion $d\colon \mathrm{Mat}(2\times 2, \mathbb{R}) \longrightarrow \mathbb{R}$ betrachten, die der Matrix $(a_1\ a_2)$ den (anschaulichen) Flächeninhalt zuordnet, den das von den Spalten a_1 und a_2 aufgespannte Parallelogramm hat.



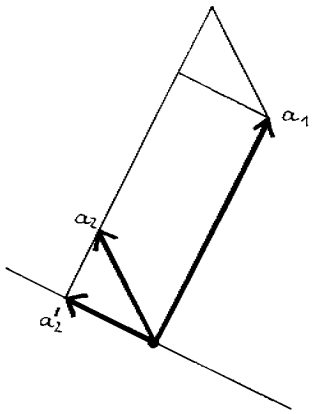
Inwieweit hat diese Funktion die Eigenschaften (a), (b), (c) der Determinantenfunktion? Nun, zweifellos hat sie (b) und (c), denn für $a_1 = a_2$ entartet das Parallelogramm zu einer Strecke, und für $a = 1$, das heißt $a_1 = e_1$ und $a_2 = e_2$ handelt es sich um ein Quadrat mit Seitenlänge 1. Dagegen trifft Linearität etwa in der zweiten Spalte nur teilweise zu: das von a_1 und λa_2 aufgespannte Parallelogramm hat nur dann wie erhofft den λ-fachen Flächeninhalt, wenn $\lambda \geq 0$ ist.



Ist ja auch klar, sonst käme man auf einen negativen Flächeninhalt! Ein geistreicher Trick besteht aber nun gerade darin, solche negativen Flächeninhalte zu akzeptieren, indem man den absoluten Flächeninhalt je nach der Orientierung des Vektorpaares (a_1, a_2) mit einem Vorzeichen versieht, nämlich einem negativen genau dann, wenn a_2 nach rechts weist, wenn man in Richtung von a_1 blickt



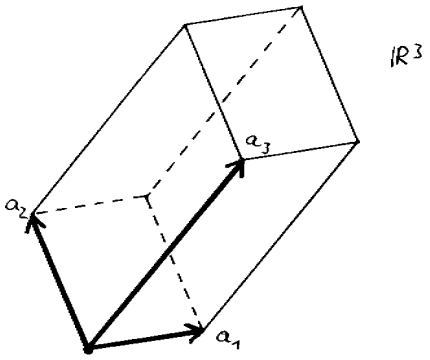
(in allen Fällen, wo das keinen Sinn gibt, zum Beispiel wenn a_2 ein Vielfaches von a_1 ist, wird der Flächeninhalt ohnehin null). Man spricht vom *orientierten Flächeninhalt*. Dieser verhält sich nun tatsächlich auch dann linear, wenn man die zweite Spalte mit einem negativen Skalar multipliziert. Außerdem zeigt eine einfache elementargeometrische Überlegung, daß man bei der Berechnung des Flächeninhaltes den Vektor a_2 durch seine zu a_1 senkrechte Projektion a_2' ersetzen kann,



und daß der orientierte Flächeninhalt deshalb überhaupt eine lineare Funktion der ersten und analog auch der zweiten Spalte ist. Diese Funktion muß aber dann mit der Determinantenfunktion identisch sein: Die zweidimensionale Determinante mißt also den orientierten Flächeninhalt eines Parallelogramms.

In höheren Dimensionen verhält es sich ganz analog. An die Stelle des Parallelogramms tritt das von den Spalten einer reellen $n \times n$-Matrix a aufgespannte *Parallelepiped*

$$\left\{ x \in \mathbb{R}^n \,\middle|\, x = \sum_{i=1}^{n} \lambda_i a_i,\ \lambda_i \in [0,1] \text{ für alle } i \right\}.$$



Die Zahl $\det a$ gibt dessen n-dimensionales Volumen an, versehen mit einem Orientierungsvorzeichen. Für $n = 3$ handelt es sich um das gewöhnliche Volumen, und die Orientierung ergibt sich aus der Rechte-Hand-Regel der Physiker: das von Daumen, Zeige- und Mittelfinger (in dieser Reihenfolge) aufgespannte Parallelepiped hat positives Volumen.

Natürlich bedürfen "Volumen" und "Orientierung" als mathematische Begriffe erst mal einer Präzisierung, um die wir uns an dieser Stelle aber nicht bemühen wollen. Jedenfalls spielt dabei die Determinante eine zentrale Rolle.

Bemerkung Wenn man eine Basis des physikalischen Raumes wählt, zeichnet man damit zugleich eine Orientierung aus. Eine solche Wahl ist auch physikalisch bedeutsam: Die für den β-Zerfall von Atomkernen verantwortliche schwache Wechselwirkung zeigt eine je nach Wahl der Orientierung verschiedene Gesetzmäßigkeit, was sich konkret darin äußert, daß es bezüglich der üblichen Wahl nur sogenannte *linkshändige* Neutrinos gibt (Impuls und Spin einander entgegengerichtet), aber keine rechtshändigen.

Zum Schluß noch eine Anwendung der Determinante auf die Theorie der Gruppen. Wie schon an früherer Stelle erwähnt, nennt man bijektive Abbildungen, vor allem solche zwischen den Mengen $\{1, 2, \dots, n\}$, häufig *Permutationen*. Sie bilden bei festem $n \in \mathbb{N}$ unter der Komposition eine Gruppe, wie man sofort einsieht.

22.15 Definition Sei $n \in \mathbb{N}$. Die Gruppe

$$\mathrm{Sym}_n := \big\{\sigma\colon \{1, 2, \dots, n\} \longrightarrow \{1, 2, \dots, n\} \,\big|\, \sigma \text{ ist bijektiv}\big\}$$

heißt symmetrische Gruppe (in n Ziffern oder Symbolen).

22.16 Beispiele Die Gruppen Sym_0 und Sym_1 bestehen nur aus dem Einselement, also der identischen Abbildung. Die Gruppe Sym_2 enthält darüber hinaus die Permutation (1 2), die 1 mit 2 vertauscht; Permutationen, die nur zwei Ziffern miteinander vertauschen, nennt man allgemein Transpositionen. Interessanter schon ist Sym_3 mit den sechs Elementen

$$\mathrm{Sym}_3 = \big\{1,\ (1\ 2),\ (1\ 3),\ (2\ 3),\ (1\ 2\ 3),\ (1\ 3\ 2)\big\};$$

dabei bezeichnet allgemein $(j_1\ j_2\ j_3\ \dots\ j_r)$ diejenige Permutation σ, die alle nicht aufgeführten Ziffern unverändert läßt und $j_1, \dots, j_r$ "im Kreis" herumschiebt:

$$j_1 \overset{\sigma}{\mapsto} j_2 \overset{\sigma}{\mapsto} j_3 \overset{\sigma}{\mapsto} \dots \overset{\sigma}{\mapsto} j_r \overset{\sigma}{\mapsto} j_1$$

(solche Permutationen nennt man zyklisch). Für $n \geq 4$ kommen unter den $n!$ Elementen von Sym_n auch nichtzyklische Permutationen vor, zum Beispiel das Produkt $(1\ 2)(3\ 4) = (1\ 2) \circ (3\ 4)$, und schon ab $n = 3$ ist Sym_n keine abelsche Gruppe:

$$(1\ 2)(1\ 3) = (1\ 3\ 2), \quad \text{aber} \quad (1\ 3)(1\ 2) = (1\ 2\ 3).$$

Anstatt auf die Ziffern $1, \dots, n$ kann man die Gruppe Sym_n auch auf den K-Vektorraum K^n durch Vertauschung der Standardbasisvektoren wirken lassen. Die Permutation $\sigma \in \mathrm{Sym}_n$ wird dann zu der invertierbaren Matrix

$$\begin{pmatrix} e_{\sigma 1} & e_{\sigma 2} & \dots & e_{\sigma n} \end{pmatrix} \in \mathrm{Mat}(n \times n, K),$$

und die symmetrische Gruppe so zu einer Untergruppe

$$\mathrm{Sym}_n \subset GL(n, K).$$

Die dabei auftretenden Matrizen nennt man übrigens Permutationsmatrizen. Sie sind dadurch charakterisiert, daß in jeder Zeile und in jeder Spalte genau eine 1 steht und alle übrigen Einträge null sind.

22.17 Beispiele Die zur Transposition $(k\ l)$ gehörige Permutationsmatrix ist gerade die Elementarmatrix p_{kl}. Als Untergruppe von $GL(3,K)$ aufgefaßt besteht Sym_3 aus den Matrizen

$$1 = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix},$$

$$(1\ 2) = \begin{pmatrix} & 1 & \\ 1 & & \\ & & 1 \end{pmatrix}, \quad (1\ 3) = \begin{pmatrix} & & 1 \\ & 1 & \\ 1 & & \end{pmatrix}, \quad (2\ 3) = \begin{pmatrix} 1 & & \\ & & 1 \\ & 1 & \end{pmatrix},$$

$$(1\ 2\ 3) = \begin{pmatrix} & & 1 \\ 1 & & \\ & 1 & \end{pmatrix}, \quad (1\ 3\ 2) = \begin{pmatrix} & 1 & \\ & & 1 \\ 1 & & \end{pmatrix}.$$

Es leuchtet unmittelbar ein, daß man jede Permutationsmatrix durch wiederholte elementare Spaltenvertauschungen in die Einheitsmatrix überführen kann. Als Determinante einer Permutationsmatrix kommt nach Lemma 22.2 also nur 1 oder -1 in Frage. Gelehrter ausgedrückt: Die Komposition

$$\mathrm{Sym}_n \hookrightarrow GL(n,K) \xrightarrow{\det} K\backslash\{0\}$$

ist ein Homomorphismus von Gruppen mit Werten in der Untergruppe $\{\pm 1\}$. Darauf beruht die

22.18 Definition Für jede Permutation $\sigma \in \mathrm{Sym}_n$ nennt man die Zahl $\det \sigma$ das Vorzeichen oder Signum von σ, und man schreibt dafür symbolisch

$$(-1)^\sigma \in \{\pm 1\}.$$

Das Vorzeichen von σ ist also genau dann $+1$, wenn σ Produkt einer geraden Anzahl von Transpositionen ist. Beachten Sie, daß es viele solche Zerlegungen von σ gibt und die Anzahl der Faktoren darin keineswegs eindeutig bestimmt ist, nur eben die Parität dieser Anzahl. — In Aufgabe 22.3(b) wird de facto das Signum derjenigen Permutation σ berechnet, die die natürliche Reihenfolge der Ziffern $1, 2, \ldots, n$ umkehrt:

$$(-1)^\sigma = (-1)^{n(n-1)/2}$$

Der Vorzeichenhomomorphismus erlaubt es seinerseits, eine weitere Formel für die Determinante hinzuschreiben:

22.19 Determinantenformel Für jede Matrix $a \in \mathrm{Mat}(n\times n, K)$ gilt

$$\det a = \sum_{\sigma\in\mathrm{Sym}_n} (-1)^\sigma a_{1,\sigma 1}\cdot a_{2,\sigma 2}\cdots a_{n,\sigma n} = \sum_{\sigma\in\mathrm{Sym}_n} (-1)^\sigma a_{\sigma 1,1}\cdot a_{\sigma 2,2}\cdots a_{\sigma n,n}.$$

Beweis Wie wohl? Natürlich verifiziert man, daß die durch die Formel definierte Funktion die Eigenschaften (a), (b) und (c) aus Satz 22.1 hat ...

Zwar ist die Formel 22.19 so explizit wie man nur wünschen kann, aber weil sie so viele Terme enthält, hat sie wie die Zeilen- und Spaltenentwicklungsformeln eher theoretische Bedeutung.

Übungsaufgaben

22.1 Ist $\det(a+b) = \det a + \det b$ eine richtige Formel? Dafür könnte sprechen, daß die Determinante in jeder Spalte linear ist. Wenn man andererseits $a = b = 1 \in \mathrm{Mat}(2\times 2, K)$ einsetzt ...

22.2 $a \in \mathrm{Mat}(p\times n, K)$ sei eine Matrix vom Rang r. Beweisen Sie: Es gibt eine $r \times r$-Teilmatrix a' von a mit $\det a' \neq 0$, aber für $s > r$ hat jede $s \times s$-Teilmatrix von a die Determinante null. (Mit einer *Teilmatrix* von a ist jede Matrix gemeint, die man aus a durch Wegstreichen von Zeilen und/oder Spalten herstellen kann.)

22.3 (a) Berechnen Sie die Determinante

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

- nach dem Gaußschen Algorithmus,
- durch Entwicklung nach der vierten Zeile.

(b) Berechnen Sie für beliebige $\lambda_1, \dots, \lambda_n \in K$ die Determinanten $\det a$ und $\det b$ der Matrizen

$$a = \begin{pmatrix} \lambda_1 & 0 & \dots & \dots & 0 \\ * & \lambda_2 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ * & \dots & * & \lambda_{n-1} & 0 \\ * & \dots & \dots & * & \lambda_n \end{pmatrix} \quad \text{und} \quad b = \begin{pmatrix} 0 & \dots & \dots & 0 & \lambda_1 \\ 0 & \dots & 0 & \lambda_2 & * \\ \vdots & \iddots & \iddots & \iddots & \vdots \\ 0 & \lambda_{n-1} & * & \dots & * \\ \lambda_n & * & \dots & \dots & * \end{pmatrix}.$$

22.4 Berechnen Sie

$$\det \begin{pmatrix} x & -1 & 0 & \dots & 0 & 0 \\ 0 & x & -1 & \ddots & & 0 \\ 0 & 0 & x & \ddots & 0 & \vdots \\ \vdots & & \ddots & \ddots & -1 & 0 \\ 0 & 0 & \dots & 0 & x & -1 \\ a_n & a_{n-1} & a_{n-2} & \dots & a_2 & x \end{pmatrix} \in K$$

für beliebige $a_2, \dots, a_n, x \in K$.

22.5 Je n Skalare $\lambda_1, \lambda_2, \dots, \lambda_n$ definieren die sogenannte *Vandermonde-Determinante*:

$$V(\lambda_1, \dots, \lambda_n) := \det \begin{pmatrix} 1 & \lambda_1 & \lambda_1^2 & \dots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \lambda_2^2 & \dots & \lambda_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_n & \lambda_n^2 & \dots & \lambda_n^{n-1} \end{pmatrix}$$

Zeigen Sie, daß $V(\lambda_1, \dots, \lambda_n) = \prod_{i<j} (\lambda_j - \lambda_i)$ gilt.

Tip: Es liegt nahe, vollständige Induktion zu Hilfe zu ziehen; trotzdem erfordert es einige Geduld, die Determinante durch direkte Anwendung der üblichen Methoden zu "knacken". Einfacher kann man es sich machen, wenn man vorweg zeigt, daß $V(\lambda_1, \dots, \lambda_n)$ nicht von den λ_i selbst abhängt, sondern nur von Differenzen $\lambda_i - \lambda_j$.

22.6 Die quadratische Matrix $A \in \mathrm{Mat}\big((r+s) \times (r+s), K\big)$ sei durch drei Teilmatrizen $a \in \mathrm{Mat}(r \times r, K)$, $c \in \mathrm{Mat}(s \times r, K)$ und $d \in \mathrm{Mat}(s \times s, K)$ in der Form

$$A = \left(\begin{array}{c|c} a & 0 \\ \hline c & d \end{array}\right)$$

gegeben, worin 0 für die Nullmatrix in $\mathrm{Mat}(r \times s, K)$ steht. Zeigen Sie, daß dann

$$\det A = \det a \cdot \det d$$

gilt. (Tip: Der bequemste Beweis besteht darin, den des Multiplikationssatzes 22.10 zu imitieren.)

22.7 Sei a eine quadratische Matrix mit ganzzahligen Einträgen: $a \in \mathrm{Mat}(n \times n, \mathbb{Z})$. Zeigen Sie, daß die beiden folgenden Eigenschaften von a zueinander äquivalent sind:

(a) a ist invertierbar und $a^{-1} \in \mathrm{Mat}(n \times n, \mathbb{Z})$

(b) $\det a = \pm 1$

22.8 Führen Sie den Beweis der Determinantenformel 22.19 aus.

23 Reelle und komplexe Vektorräume

Alles, was wir bisher in der linearen Algebra gemacht haben, ist von der Wahl des zugrundeliegenden Körpers K völlig unabhängig (lediglich in den Beispielen habe ich stillschweigend angenommen, daß K zumindest die ganzen und damit die rationalen Zahlen enthält, denn solche habe ich ja als Einträge der Matrizen etc. hingeschrieben). Insbesondere bei zwei Resultaten ist bemerkenswert, daß sie für jeden Körper K in gleicher Weise gelten: Da ist einmal die Tatsache, daß jeder n-dimensionale K-Vektorraum zu K^n isomorph ist (Satz 19.2), und andererseits die Möglichkeit, jede lineare Abbildung zwischen endlichdimensionalen K-Vektorräumen mittels geschickter Basiswahl durch eine Matrix

$$\left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & 0 \end{array}\right)$$

zu beschreiben, in der außer 0 und 1 überhaupt keine Elemente von K mehr vorkommen, die vielmehr allein vom Rang dieser Abbildung abhängt (Satz 20.12).

Das mag Sie nicht besonders beeindrucken, weil außer den Zahlkörpern $\mathbb{Q}$, $\mathbb{R}$ und $\mathbb{C}$ bisher ohnehin keine weiteren Körper vorgekommen sind; deshalb hier einer, dessen Elemente keine Zahlen sind: Die rationalen Funktionen im Sinne von 10.11 bilden unter der üblichen punktweisen Addition und Multiplikation einen Körper, den man in Anlehnung an den Polynomring $\mathbb{C}[X]$ mit $\mathbb{C}(X)$ bezeichnet. Der springende Punkt dabei ist offensichtlich: der Kehrwert einer von der Nullfunktion verschiedenen rationalen Funktion ist wieder eine solche (während der Kehrwert eines Polynoms von positivem Grad natürlich kein Polynom ist). $\mathbb{C}(X)$ enthält die Teilkörper $\mathbb{R}(X)$ und $\mathbb{Q}(X)$ der rationalen Funktionen mit reellen bzw rationalen Koeffizienten.

Dennoch sind $\mathbb{R}$ und $\mathbb{C}$ ohne Zweifel die in der Physik wichtigsten Körper. Sie sind außerdem eng miteinander verwandt, wie ja aus der Konstruktion von $\mathbb{C}$ als einer angereicherten Version des kartesischen Produkts $\mathbb{R}^2$ hervorgeht. Auch des $\mathbb{R}$-Vektorraums $\mathbb{R}^2$? Gewiß: Wenn man die komplexe Multiplikation $\mathbb{C}\times\mathbb{C}\longrightarrow\mathbb{C}$ zu $\mathbb{R}\times\mathbb{C}\longrightarrow\mathbb{C}$ einschränkt, erhält man die skalare Multiplikation des reellen Vektorraums $\mathbb{R}^2$:

$$\lambda(x+iy)=\lambda x+i\cdot\lambda y$$

$\mathbb{C}$ ist also ganz nebenbei auch ein zweidimensionaler $\mathbb{R}$-Vektorraum, und natürlich ist $(1,i)$ die der Standardbasis von $\mathbb{R}^2$ entsprechende Basis.

Wir wollen uns in diesem kleinen Abschnitt überlegen, wie man allgemeiner aus einem reellen einen komplexen und aus einem komplexen einen reellen Vektorraum machen kann. Beginnen wir mit dem zweiten Vorgang, der geradezu lächerlich erscheint: Ist V ein $\mathbb{C}$-Vektorraum, so können wir die skalare Multiplikation $\mathbb{C}\times V\longrightarrow V$ zu $\mathbb{R}\times V\longrightarrow V$ einschränken, und schon ist V zu einem $\mathbb{R}$-Vektorraum geworden. Man verzichtet also einfach darauf, mit nicht-reellen Skalaren zu multiplizieren. Weil man Vektorräume normalerweise ohnehin nur durch Angabe der zugrundeliegenden Menge bezeichnet, wird man für den so entstehenden reellen Vektorraum nicht mal unbedingt ein anderes Symbol als V verwenden wollen. Wenn aber doch, dann schreibt man $V_{\mathbb{R}}$, wie man überhaupt alle verwechslungsgefährdeten Symbole nötigenfalls durch ein angehängtes $\mathbb{R}$ bzw. $\mathbb{C}$ präzisiert. So unterscheidet etwa $\mathrm{Lin}_{\mathbb{R}}(\ldots)$ und $\mathrm{Lin}_{\mathbb{C}}(\ldots)$ zwischen den mit reellen und komplexen Skalaren gebildeten Linearkombinationen.

Wohl keiner weiteren Begründung bedarf die

23.1 Notiz *V sei ein komplexer Vektorraum, und $v_1,\ldots,v_r\in V$ seien Vektoren. Dann gilt:*

$$\begin{aligned}\mathrm{Lin}_{\mathbb{C}}(v_1,\ldots,v_r) &= \mathrm{Lin}_{\mathbb{R}}(v_1,\ldots,v_r,iv_1,\ldots,iv_r)\\ (v_1,\ldots,v_r)\text{ linear unabhängig in }V &\iff (v_1,\ldots,v_r,iv_1,\ldots,iv_r)\text{ linear unabhängig in }V_{\mathbb{R}}\end{aligned}$$

Insbesondere ist $(v_1, \dots, v_r)$ genau dann eine Basis von V, wenn $(v_1, \dots, v_r, iv_1, \dots, iv_r)$ eine Basis von $V_{\mathbb{R}}$ ist, und für endlichdimensionales V gilt deshalb

$$\dim V_{\mathbb{R}} = 2 \dim V \quad \text{(andere mögliche Ausdrucksweise: } \dim_{\mathbb{R}} V = 2 \dim_{\mathbb{C}} V\text{)}.$$

Übrigens hindert einen niemand daran, auch in $V_{\mathbb{R}}$ noch mit der komplexen Zahl i zu multiplizieren; dabei handelt es sich aber nicht mehr um eine skalare Multiplikation im Vektorraum $V_{\mathbb{R}}$, sondern eine lineare Abbildung $h: V_{\mathbb{R}} \longrightarrow V_{\mathbb{R}},\ h(v) = iv$, die der komplexe Vektorraum V als zusätzliche Struktur auf $V_{\mathbb{R}}$ vererbt hat. Natürlich gilt $h^2 = -\mathrm{id}$. Man überlegt sich auch leicht die Umkehrung: Ist U ein beliebiger reeller Vektorraum, und $h \in \mathrm{Hom}(U, U)$ eine lineare Abbildung mit $h^2 = -\mathrm{id}$, so kann man aus U einen komplexen Vektorraum V mit derselben zugrundeliegenden Menge machen, indem man die komplexe Skalarenmultiplikation $\mathbb{C} \times V \longrightarrow V$ durch

$$(\lambda + i\mu)v := \lambda v + \mu h(v)$$

erklärt; $V_{\mathbb{R}}$ ist dann wieder U. Wegen dieses Zusammenhangs wird ein Homomorphismus $h \in \mathrm{Hom}(U, U)$ mit $h^2 = -\mathrm{id}$ manchmal eine *komplexe Struktur* auf dem reellen Vektorraum U genannt.

Sind V und W zwei $\mathbb{C}$-Vektorräume, so ist jede lineare Abbildung $f: V \longrightarrow W$ natürlich auch $\mathbb{R}$-linear, das heißt eine lineare Abbildung $V_{\mathbb{R}} \longrightarrow W_{\mathbb{R}}$; als solche bezeichnet man sie mit $f_{\mathbb{R}}$. Ist im endlichdimensionalen Fall f durch eine Matrix beschrieben, so tritt die Frage auf, welche Matrix dann $f_{\mathbb{R}}$ beschreibt.

23.2 Lemma *V und W seien $\mathbb{C}$-Vektorräume mit Basen $(v_1, \dots, v_n)$ und $(w_1, \dots, w_p)$. Hat die lineare Abbildung $f: V \longrightarrow W$ bezüglich dieser Basen die Matrix $c = a + ib \in \mathrm{Mat}(p \times n, \mathbb{C})$, so ist*

$$c_{\mathbb{R}} = \left(\begin{array}{c|c} a & -b \\ \hline b & a \end{array}\right) \in \mathrm{Mat}(2p \times 2n, \mathbb{R})$$

die Matrix von $f_{\mathbb{R}}$ bezüglich der Basen $(v_1, \dots, v_n, iv_1, \dots, iv_n)$ und $(w_1, \dots, w_p, iw_1, \dots, iw_p)$.

Beweis "Die Spalten der Matrix sind die Bilder der Basisvektoren."

Bemerkung Eine andere vernünftige Anordnung der neuen Basisvektoren wäre $(v_1, iv_1, v_2, iv_2, \dots, v_n, iv_n)$ etc. Wenn man das so macht, ist $c_{\mathbb{R}}$ aus c dadurch zu bilden, daß man jeden Eintrag $c_{jk} = a_{jk} + ib_{jk}$ durch das "Kästchen"

$$\begin{pmatrix} a_{jk} & -b_{jk} \\ b_{jk} & a_{jk} \end{pmatrix}$$

ersetzt.

Jetzt machen wir umgekehrt aus einem beliebigen reellen Vektorraum einen komplexen. Diesmal nicht, indem wir etwas vergessen, sondern indem wir etwas Neues konstruieren.

23.3 Definition *V sei ein reeller Vektorraum. Die Komplexifizierung von V ist der folgendermaßen definierte komplexe Vektorraum $V_{\mathbb{C}}$. Als reeller Vektorraum ist $V_{\mathbb{C}} = V \times V$ das kartesische Produkt (komponentenweise Verknüpfungen), und die skalare Multiplikation wird durch die Formel*

$$(\lambda + i\mu) \cdot (u, v) := (\lambda u - \mu v,\ \lambda v + \mu u)$$

auf komplexe Skalare $\lambda + i\mu$ erweitert.

In der gleichen Weise, wie wir einmal die ursprünglich als Paar definierte komplexe Zahl $(a, 0)$ mit $a \in \mathbb{R}$ identifiziert haben, identifiziert man hier $(u, 0) \in V \times \{0\} \subset V \times V$ mit $u \in V$; das erlaubt es, $V \times V$ als direkte Summe der beiden *reellen* Untervektorräume V und iV zu schreiben:

$$V_{\mathbb{C}} = V + iV$$

Obwohl $V_{\mathbb{C}}$ wirklich größer als V ist, merkt man die Komplexifizierung den nackten Formeln oft nicht an. Das hat seinen Grund in dem

23.4 Lemma V sei ein reeller Vektorraum, und $v_1, \dots, v_r \in V$ seien Vektoren. Dann gilt

$$\mathrm{Lin}_{\mathbb{C}}(v_1, \dots, v_r) = \mathrm{Lin}_{\mathbb{R}}(v_1, \dots, v_r)_{\mathbb{C}}$$

(rechts steht die Komplexifizierung des reellen Vektorraums $\mathrm{Lin}_{\mathbb{R}}(v_1, \dots, v_r) \subset V$) sowie

$$(v_1, \dots, v_r) \text{ linear unabhängig in } V \iff (v_1, \dots, v_r) \text{ linear unabhängig in } V_{\mathbb{C}}.$$

Insbesondere ist jede Basis von V auch eine Basis von $V_{\mathbb{C}}$, und wenn V endlichdimensonal ist, gilt deshalb $\dim V_{\mathbb{C}} = \dim V$.

Beweis Es genügt, für $\lambda_j, \mu_j \in \mathbb{R}$ die Zerlegung

$$\sum_{j=1}^{r} (\lambda_j + i\mu_j) v_j = \underbrace{\sum_{j=1}^{r} \lambda_j v_j}_{\in V} + \underbrace{i \sum_{j=1}^{r} \mu_j v_j}_{\in iV}$$

zu betrachten und daran zu denken, daß die Summe $V_{\mathbb{C}} = V + iV$ direkt ist.

Ähnlich wie der Übergang von einem komplexen Vektorraum V zu $V_{\mathbb{R}}$ den reellen Vektorraum $V_{\mathbb{R}}$ mit einer zusätzlichen, nämlich einer komplexen Struktur ausstattet, erbt die Komplexifizierung eines reellen Vektorraums V von diesem eine Art reelle Struktur. In $V_{\mathbb{C}}$ gibt es nämlich Sinn, vom Real- und Imaginärteil eines Vektors sowie vom konjugiert-komplexen Vektor zu reden. Für $w = u + iv \in V + iV = V_{\mathbb{C}}$ sind das natürlich

$$\mathrm{Re} w = u \in V \quad \text{und} \quad \mathrm{Im} w = v \in V \quad \text{sowie} \quad \overline{w} = u - iv \in V_{\mathbb{C}}.$$

Beachten Sie dagegen, daß in einem beliebigen komplexen Vektorraum W a priori keine Konjugation definiert ist. (Zwar kann man in jedem n-dimensionalen Vektorraum eine Basis wählen und die komplexe Konjugation von $\mathbb{C}^n$ mittels der zugehörigen Karte nach W übertragen, aber dieser Konjugationsbegriff hängt dann von der willkürlichen Wahl dieser Basis ab.)

Ist $f\colon V \longrightarrow W$ eine lineare Abbildung zwischen reellen Vektorräumen, so definiert man deren Komplexifizierung naheliegenderweise als die $\mathbb{C}$-lineare Abbildung

$$f_{\mathbb{C}}\colon V_{\mathbb{C}} \longrightarrow W_{\mathbb{C}}, \quad x + iy \mapsto f(x) + if(y).$$

War f bezüglich Basen von V und W durch eine Matrix $a \in \mathrm{Mat}(p \times n, \mathbb{R})$ beschrieben, so können wir $f_{\mathbb{C}}$ gemäß Lemma 23.4 bezüglich derselben Basen ausdrücken, und das geschieht natürlich durch genau dieselbe Matrix, die jetzt bloß als $a \in \mathrm{Mat}(p \times n, \mathbb{C})$ aufgefaßt wird.

23.5 Beispiel Sei $I \subset \mathbb{R}$ ein Intervall mit mindestens zwei Punkten. Die Funktionenräume $C^k(I, \mathbb{C})$ und $\mathcal{O}(I, \mathbb{C})$ sind dann gerade die Komplexifizierungen der entsprechenden Räume $C^k(I, \mathbb{R})$ und $\mathcal{O}(I, \mathbb{R})$ reellwertiger Funktionen. Einen Differentialoperator D mit reellen Koeffizienten, zum Beispiel den des klassischen harmonischen Oszillators $f \mapsto Df = f'' + 2\gamma f' + \omega^2 f$, können wir als Abbildung zwischen den reellen Räumen

$$C^2(I, \mathbb{R}) \xrightarrow{D} C^0(I, \mathbb{R})$$

auffassen, aber das in Satz 21.11 beschriebene Lösungsverfahren legt es nahe, lieber gleich die (durch dieselbe Formel gegebene) Komplexifizierung

$$C^2(I, \mathbb{C}) \xrightarrow{D_{\mathbb{C}}} C^0(I, \mathbb{C})$$

zu bilden. Dann gilt $\mathrm{Kern}\, D_{\mathbb{C}} = (\mathrm{Kern}\, D)_{\mathbb{C}}$, und nach Lemma 23.4 müssen beide Vektorräume eine gemeinsame Basis besitzen. Eine solche hatten wir im stark gedämpften und im Kriechfall auch erhalten, nicht jedoch im schwach gedämpften: die Basislösungen

$$t \mapsto e^{-\gamma t} e^{\pm i\omega' t}$$

sind definitiv nicht reell, gehören also nicht zum Kern (nicht mal zum Definitionsbereich) von D. Aber sie sind zueinander komplex-konjugiert, und man gewinnt eine Basis von Kern D mittels der

23.6 Notiz Sei V ein $\mathbb{R}$-Vektorraum, und $w \in V_{\mathbb{C}}$. Dann gilt

$$\mathrm{Lin}_{\mathbb{C}}(w, \overline{w}) = \mathrm{Lin}_{\mathbb{C}}(\mathrm{Re}w, \mathrm{Im}w).$$

Hier erhalten wir also

$$t \mapsto e^{-\gamma t} \cos \omega' t \quad \text{und} \quad t \mapsto e^{-\gamma t} \sin \omega' t$$

als reelle Basislösungen. Daß das Verfahren immer funktionieren wird, sieht man schon daran, daß die komplexen Nullstellen eines reellen Polynoms in Paaren konjugiert-komplexer Zahlen auftreten und deshalb auch die Basislösungen solche Paare bilden.

Die Methode der Komplexifizierung ist aber noch in einem viel weiteren Rahmen merkenswert. Eine Unmenge von zunächst reellen Problemen kann man behandeln, indem man sie erst komplex und damit — im Widerspruch zur Semantik des Wortes — einfacher und lösbar macht und anschließend die komplexen Lösungen interpretiert oder weiterbehandelt. In den folgenden Abschnitten werden wir auf zahlreiche Beispiele dafür stoßen.

Übungsaufgabe

23.1 V sei ein komplexer Vektorraum. Die Komplexifizierung von $V_{\mathbb{R}}$ sollte man nur als kartesisches Produkt $(V_{\mathbb{R}})_{\mathbb{C}} = V_{\mathbb{R}} \times V_{\mathbb{R}}$ und nicht als $V_{\mathbb{R}} + iV_{\mathbb{R}}$ schreiben, weil man sonst leicht die bei der Komplexifizierung neu eingeführte Multiplikation mit i einerseits und die von V auf $V_{\mathbb{R}}$ vererbte komplexe Struktur h (mit $h^2 = -\mathrm{id}$) andererseits miteinander verwechselt.

Sei nun $(v_1, \dots, v_n)$ eine Basis von V; die lineare Abbildung $f\colon V \longrightarrow V$ habe bezüglich dieser Basis die Matrix $c = a + ib \in \mathrm{Mat}(n \times n, \mathbb{C})$. Bestimmen Sie die Matrizen von $h_{\mathbb{C}}$ und von $(f_{\mathbb{R}})_{\mathbb{C}}$ bezüglich der Basis

$$\big((v_1, iv_1), \dots, (v_n, iv_n), (v_1, -iv_1), \dots, (v_n, -iv_n)\big)$$

von $(V_{\mathbb{R}})_{\mathbb{C}}$.

Als Nebenprodukt erhält man die Formel

$$\det \left(\begin{array}{c|c} a & -b \\ \hline b & a \end{array}\right) = |\det(a + ib)|^2 \quad \text{für beliebige Matrizen } a, b \in \mathrm{Mat}(n \times n, \mathbb{R}).$$

24 Lineare Endomorphismen

24.1 Definition Einen Homomorphismus

$$V \xrightarrow{f} V$$

(von Vektorräumen, Gruppen, Ringen ...) nennt man einen Endomorphismus. Ist er zugleich bijektiv, nennt man ihn Automorphismus. In Tabellenform also:

	V, W beliebig	$V = W$
f beliebig	homo	endo
f bijektiv	iso	auto

Hier wollen wir uns für *lineare* Endomorphismen, also solche eines K-Vektorraums V interessieren. Wie wir wissen, bilden sie selbst einen K-Vektorraum

$$\operatorname{End} V := \operatorname{Hom}(V, V),$$

der bei n-dimensionalem V die Dimension n^2 hat. Mit der Komposition von Endomorphismen haben wir eine weitere Verknüpfung

$$\operatorname{End} V \times \operatorname{End} V \xrightarrow{\circ} \operatorname{End} V,$$

die — zusammen mit der Addition — den Vektorraum $\operatorname{End} V$ gleichzeitig zu einem (nicht kommutativen) Ring macht. Im Fall $V = K^n$ handelt es sich natürlich um den Vektorraum $\operatorname{Mat}(n \times n, K)$ der quadratischen Matrizen, der durch die Matrizenmultiplikation eben auch zu einem Ring wird.

Wenn bei einer linearen Abbildung $f: V \longrightarrow W$ Definitions- und Zielraum identisch sind, scheint das zuerst nicht besonders aufregend zu sein. Nirgends war bisher ja verboten, daß "zufällig" mal $V = W$ ist, und alle unsere Ergebnisse gelten da selbstverständlich auch. Das besondere Interesse an dem Fall $V = W$ liegt aber in Fragen, die ansonsten gar keinen Sinn geben, darunter:

- Sicher ist $f(0) = 0$. Gibt es weitere sogenannte Fixvektoren, nämlich Vektoren $v \in V$ mit $f(v) = v$?
- Gibt es lineare Unterräume $U \subset V$ mit $f(U) \subset U$, die also von f in sich abgebildet werden? Von solchen Unterräumen sagt man, sie seien unter f invariant, und trivialerweise haben zumindest $\{0\}$ und V diese Eigenschaft.

Wie wir wissen, kann f im endlichdimensionalen Fall mittels der Wahl zweier Basen von V durch eine Matrix beschrieben werden, und es liegt nahe, dabei zweimal *dieselbe* Basis $\underline{v} = (v_1, \dots, v_n)$ zu nehmen;

$$\begin{array}{ccc} V & \xrightarrow{f} & V \\ {\scriptstyle \Phi_{\underline{v}}}\uparrow{\scriptstyle \simeq} & & {\scriptstyle \Phi_{\underline{v}}}\uparrow{\scriptstyle \simeq} \\ K^n & \xrightarrow{a} & K^n \end{array}$$

dann sind zum Beispiel die Fixvektoren, in der Karte $\Phi_{\underline{v}}^{-1}$ gelesen, gerade die Fixvektoren $x \in K^n$ der Matrix a.

24.2 Definition Sei V ein K-Vektorraum, $f\colon V \longrightarrow V$ ein Endomorphismus. Ein Eigenwert von f ist ein Skalar $\lambda \in K$, zu dem es mindestens einen Vektor $v \in V$ gibt mit

$$v \neq 0 \quad \text{und} \quad f(v) = \lambda v,$$

jeder solche Vektor heißt ein Eigenvektor von f (zum Eigenwert λ). Für jedes $\lambda \in K$ nennt man den linearen Unterrraum

$$\{v \in V \mid f(v) = \lambda v\} \subset V$$

den zu λ gehörigen Eigenraum von f.

Bemerkungen Achten Sie gut auf die Feinheiten dieser etwas eigenwilligen, aber zweckmäßigen Definition: Der Nullvektor gilt nicht als Eigenvektor, ist jedoch in jedem Eigenraum enthalten. Die Eigenwerte sind diejenigen Skalare, für die der zugehörige Eigenraum nicht der Nullraum ist. — Der zu $0 \in K$ gehörige Eigenraum ist der Kern von f, und nicht-triviale Fixvektoren von f sind dasselbe wie Eigenvektoren zum Eigenwert 1.

Ist die Existenz von Eigenwerten und -vektoren nun der Normalfall oder etwas Besonderes? Das werden wir gleich sehen; die folgenden Überlegungen zielen nämlich darauf, alle Eigenwerte eines Endomorphismus, gegeben durch eine Matrix $a \in \mathrm{Mat}(n\times n, K)$, geradezu zu berechnen. Sie gelten übrigens für beliebige Körper K; ich nehme aber zur Vereinfachung der Argumentation schon mal stillschweigend an, daß K unendlich viele Elemente enthält, damit man Polynome aus $K[X]$ zugleich als Terme in X und als Funktionen von K nach K ansehen kann. Bei den uns interessierenden Körpern ist das ohnehin immer der Fall.

24.3 Notiz und **Definition** Sei $a \in \mathrm{Mat}(n\times n, K)$ eine quadratische Matrix. Dann ist

$$\chi_a(X) := \det(X - a)$$

ein normiertes Polynom vom Grad n in $K[X]$; es heißt das charakteristische Polynom von a.

Erklärung Es liegt nahe, die skalaren Vielfachen $\lambda \cdot 1$ der Einheitsmatrix einfach als $\lambda \in \mathrm{Mat}(n\times n, K)$ zu schreiben, und das ist hier kurzerhand auch mit dem X gemacht, das ja bloß ein Platzhalter für Skalare ist. Die aus 22.19 übernommene Formel

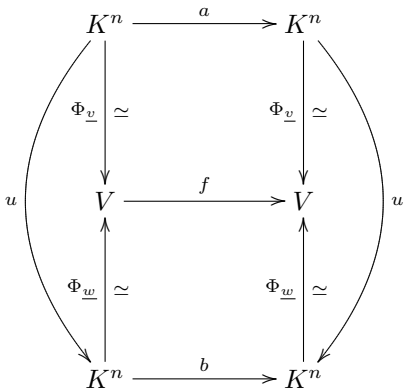
$$\det(X - a) = \sum_{\sigma\in \mathrm{Sym}_n} (-1)^\sigma (1_{1,\sigma 1}X - a_{1,\sigma 1})\cdot(1_{2,\sigma 2}X - a_{2,\sigma 2}) \cdots (1_{n,\sigma n}X - a_{n,\sigma n}) \in K[X]$$

präzisiert nun die Behauptung und beweist sie auch, denn zum Koeffizienten von X^n trägt offenbar nur der Summand mit $\sigma = \mathrm{id}$ bei.

Ist $f\colon V \longrightarrow V$ ein Endomorphismus und a die Matrix von f bezüglich einer Basis von V, so wird zwar a von der Wahl dieser Basis abhängen, nicht aber die Determinante von a. Wir beweisen gleich mehr:

24.4 Lemma $f\colon V \longrightarrow V$ sei ein Endomorphismus; $\underline{v}$ und $\underline{w}$ seien Basen von V. Sind $a, b \in \mathrm{Mat}(n\times n, K)$ die zugehörigen f bezüglich $\underline{v}$ und $\underline{w}$ beschreibenden Matrizen, so gilt $\chi_a = \chi_b$. Man darf dieses Polynom deshalb als das charakteristische Polynom von f bezeichnen.

Beweis Aus dem kommutativen Diagramm



liest man die mittels des "Kartenwechsels" $u := \Phi_{\underline{w}}^{-1} \circ \Phi_{\underline{v}} \in GL(n, K)$ auszudrückende Beziehung

$$a = u^{-1}bu$$

ab. Weil die Einheitsmatrix und damit auch X mit jeder Matrix vertauschbar ist, folgt

$$\chi_a(X) = \det(X - a) = \det(X - u^{-1}bu) = \det\left(u^{-1}(X - b)u\right)$$

und weiter

$$\det\left(u^{-1}(X - b)u\right) = \frac{1}{\det u} \det(X - b) \det u = \det(X - b) = \chi_b(X)$$

nach dem Determinantenmultiplikationssatz.

Zwei Koeffizienten des charakteristischen Polynoms sind besonders leicht zu durchschauen:

24.5 Lemma und **Definition** Sei $a \in \mathrm{Mat}(n \times n, K)$ eine Matrix mit charakteristischem Polynom

$$\chi_a(X) = X^n + c_{n-1}X^{n-1} + \cdots + c_1X + c_0 \in K[X].$$

Darin ist

$$c_0 = \chi_a(0) = (-1)^n \det a$$

bis auf das Vorzeichen die Determinante, und

$$-c_{n-1} = \sum_{j=1}^{n} a_{jj} =: \operatorname{tr} a$$

die Summe der Diagonaleinträge, die sogenannte Spur von a (englisch: trace).

Beweis Die erste Formel ist klar: $\chi_a(0) = \det(0 - a) = (-1)^n \det a$. Für die andere bemerken wir, daß in

$$\det(X - a) = \sum_{\sigma \in \mathrm{Sym}_n} (-1)^\sigma (1_{1,\sigma 1}X - a_{1,\sigma 1}) \cdot (1_{2,\sigma 2}X - a_{2,\sigma 2}) \cdots (1_{n,\sigma n}X - a_{n,\sigma n})$$

ein Summand nur dann zu c_{n-1} beitragen kann, wenn σ mindestens $n-1$, damit überhaupt alle n Ziffern fest läßt. Alle Beiträge kommen also von

$$(X - a_{11}) \cdot (X - a_{22}) \cdots (X - a_{nn});$$

ihre Summe ist $-\sum_{j=1}^{n} a_{jj}$.

24.6 Folgerung Jeder Endomorphismus f eines endlichdimensionalen K-Vektorraums V hat eine wohldefinierte Determinante $\det f \in K$ und eine wohldefinierte Spur $\operatorname{tr} f \in K$, nämlich $\det a$ bzw. $\operatorname{tr} a$, wenn a die Matrix von f bezüglich einer willkürlich gewählten Basis von V ist.

Bemerkung Im Fall der Determinante leuchtet das auch anschaulich ein: Jedes gedachte Volumen im Vektorraum V wird durch Anwenden von f mit einem gewissen Faktor multipliziert (eben der Determinante), und dieser Faktor ist unabhängig davon, welchen Maßstab man für die Volumenmessung wählt.

Das charakteristische Polynom hängt nun eng mit den Eigenwerten zusammen:

24.7 Satz $f: V \longrightarrow V$ sei Endomorphismus des endlichdimensionalen K-Vektorraums V. Dann gilt für jedes $\lambda \in K$:

$$\lambda \text{ ist Eigenwert von } f \iff \chi_f(\lambda) = 0$$

Bemerkung Die Eigenwerte sind also genau die Nullstellen des charakteristischen Polynoms, das ist ein unbedingt zu merkender Sachverhalt. Vergessen Sie darüber aber nicht, was ein Eigenwert eigentlich *ist*: ein Skalar, zu dem ein Eigenvektor existiert.

Beweis Definitionsgemäß ist λ ein Eigenwert genau dann, wenn die Gleichung

$$f(v) = \lambda v$$

für $v \in V$ eine nicht-triviale Lösung hat. Drücken wir f mittels einer Basis von V durch eine Matrix $a \in \mathrm{Mat}(n \times n, K)$ aus, so bedeutet dies, daß die Gleichung $ax = \lambda x$ oder

$$(\lambda - a)x = 0$$

eine Lösung $x \in K^n \setminus \{0\}$ hat. Wie wir wissen, ist das genau dann der Fall, wenn $\mathrm{rk}(\lambda - a) < n$ oder, gleichwertig,

$$\chi_a(\lambda) = \det(\lambda - a) = 0$$

ist.

Der Beweis ist zugleich eine Anleitung zur Berechnung der Eigenwerte und -vektoren, die wir sofort auskosten wollen:

24.8 Beispiele (1) Die Matrix $a = \begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix} \in \mathrm{Mat}(2 \times 2, \mathbb{R})$ hat das charakteristische Polynom

$$\begin{aligned} \chi_a(X) = \det(X - a) &= \det \begin{pmatrix} X - 1 & -2 \\ -4 & X - 3 \end{pmatrix} \\ &= (X - 1)(X - 3) - 8 = X^2 - 4X - 5 \\ &= (X + 1)(X - 5), \end{aligned}$$

also hat a die beiden Eigenwerte

$$\lambda = -1 \quad \text{und} \quad \mu = 5.$$

Zur Bestimmung der zugehörigen Eigenräume sind die Gleichungen

$$(\lambda - a)x = 0 \quad \text{und} \quad (\mu - a)x = 0$$

zu lösen, explizit

$$\begin{pmatrix} -2 & -2 \\ -4 & -4 \end{pmatrix} x = 0 \quad \text{und} \quad \begin{pmatrix} 4 & -2 \\ -4 & 2 \end{pmatrix} x = 0$$

oder

$$\begin{pmatrix} 1 & 1 \end{pmatrix} x = 0 \quad \text{und} \quad \begin{pmatrix} 1 & -1/2 \end{pmatrix} x = 0.$$

Also sind

$$\mathrm{Lin}\left(\begin{pmatrix} -1 \\ 1 \end{pmatrix} \right) \quad \text{und} \quad \mathrm{Lin}\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix} \right)$$

die Eigenräume zu den Eigenwerten -1 und 5.

(2) Den 2×2-Matrizen $u = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ und $1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ sieht man sofort das beiden gemeinsame charakteristische Polynom

$$\chi_u(X) = \chi_1(X) = (X - 1)^2$$

mit dem einzigen ("doppelten") Eigenwert 1 an. Der zugehörige Eigenraum — hier also der Fixraum — von u ist

$$\mathrm{Kern}(1 - u) = \mathrm{Kern} \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix} = \mathrm{Lin}\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix} \right),$$

während der Eigenraum der Einheitsmatrix natürlich ganz K^2 ist.

(3) Bei Vektorräumen, die nicht endlichdimensional sind, ist kein charakteristisches Polynom erklärt. Die beiden folgenden Beispiele zeigen, daß die Verhältnisse dort auch komplizierter liegen. Wir wählen als Vektorraum $C^\infty(\mathbb{R})$, und als Endomorphismen die Differentialoperatoren

$$D, Q \colon C^\infty(\mathbb{R}) \longrightarrow C^\infty(\mathbb{R}); \quad Df(t) := f'(t), \quad Qf(t) = t \cdot f(t).$$

Die Eigenwertgleichung $Df = \lambda f$ hat nach Satz 21.11 für jedes $\lambda \in \mathbb{R}$ einen eindimensionalen Lösungsraum (aufgespannt von der Funktion $t \mapsto e^{\lambda t}$), also ist jede reelle Zahl Eigenwert von D. Dagegen besitzt der Endomorphismus Q überhaupt keine Eigenwerte, denn aus $Qf = \lambda f$, also

$$t \cdot f(t) = \lambda f(t) \text{ oder } (t - \lambda) f(t) = 0 \quad \text{für alle } t \in \mathbb{R}$$

folgt sofort, daß $f(t) = 0$ für jedes $t \neq \lambda$ gilt, und weil f stetig ist, impliziert das $f(t) = 0$ für alle $t \in \mathbb{R}$.

Es geht direkt aus der Definition hervor, daß Eigenräume zu verschiedenen Eigenwerten nur den Nullvektor gemeinsam haben können. Tatsächlich liegen mehrere Eigenräume immer so "quer" zueinander wie nur möglich. Um das zu präzisieren, verallgemeinern wir zweckmäßig den Begriff der direkten Summe (Definition 18.21) auf eine größere Anzahl von Summanden:

24.9 Definition $U_1, \ldots, U_r$ seien Unterräume des K-Vektorraums V. Die Summe dieser Unterräume

$$U_1 + \cdots + U_r := \{u_1 + \cdots + u_r \,|\, u_1 \in U_1, \ldots, u_r \in U_r\}$$

nennt man direkt, wenn aus

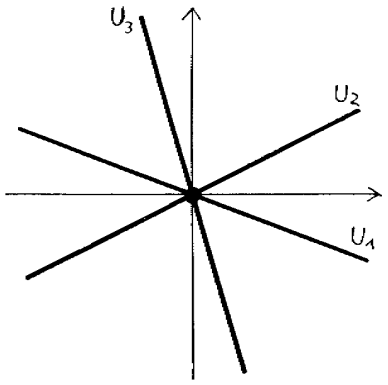
$$0 = \sum_{j=1}^{r} u_j \quad \text{mit } u_j \in U_j \text{ für alle } j$$

das Verschwinden sämtlicher u_j folgt.

Bemerkung Die Direktheit der Summe besagt gleichwertig, daß die Abbildung

$$\begin{aligned} U_1 \times \cdots \times U_r &\longrightarrow U_1 + \cdots + U_r \\ (u_1, \ldots, u_r) &\mapsto u_1 + \ldots + u_r \end{aligned}$$

ein linearer Isomorphismus des kartesischen Produktes auf die Summe ist. Insbesondere sind in einer Darstellung $v = u_1 + \ldots + u_r$ die Summanden durch v eindeutig bestimmt, und durch Zusammenwerfen von Basen der U_j erhält man eine Basis für $U_1 + \cdots + U_r$. Beachten Sie, daß die Direktheit mehr bedeutet als nur, daß die Summe von je zwei der Unterräume direkt ist:



In diesem Bildbeispiel in $\mathbb{R}^2$ ist die Summe von je zwei der Geraden direkt, aber die Summe $U_1 + U_2 + U_3$ kann es schon aus Dimensionsgründen nicht sein.

24.10 Satz $f: V \longrightarrow V$ sei ein Endomorphismus des K-Vektorraums V. Die Skalare $\lambda_1, \ldots, \lambda_r \in K$ seien paarweise verschieden, und $E_1, \ldots, E_r \subset V$ seien die zugehörigen Eigenräume. Dann ist die Summe

$$E_1 + \cdots + E_r \subset V$$

direkt. Eigenvektoren zu paarweise verschiedenen Eigenwerten sind also stets linear unabhängig.

Beweis, durch Induktion nach $r \in \mathbb{N}$. Der Induktionsanfang ($r = 0$) ist klar. Sei also $r > 0$, und

$$0 = \sum_{i=1}^{r} v_i$$

mit $v_i \in E_i$ für $i = 1, \dots, r$. Anwenden von f liefert

$$0 = f\left(\sum_{i=1}^{r} v_i\right) = \sum_{i=1}^{r} f(v_i) = \sum_{i=1}^{r} \lambda_i v_i,$$

und wenn wir davon das λ_r-fache der ursprünglichen Gleichung substrahieren, bleibt

$$0 = \sum_{i=1}^{r} (\lambda_i - \lambda_r) v_i = \sum_{i=1}^{r-1} (\lambda_i - \lambda_r) v_i.$$

Nach Induktionsannahme folgt

$$(\lambda_i - \lambda_r) v_i = 0,$$

wegen $\lambda_i \neq \lambda_r$ also

$$v_i = 0 \quad \text{für } i = 1, \dots, r-1.$$

Vom ursprünglichen Ansatz bleibt dann die noch fehlende Gleichung $v_r = 0$.

Was bedeutet es, wenn im Fall endlicher Dimension die Summe der Eigenräume gleich V ist? Nach Satz 24.10 können wir dann eine Basis $\underline{v} = (v_1, \dots, v_n)$ von V wählen, die aus lauter Eigenvektoren von f besteht, sagen wir zu den (jetzt nicht unbedingt verschiedenen) Eigenwerten $\lambda_1, \dots, \lambda_n$. Die durch das kommutative Diagramm

$$\begin{array}{ccc} V & \xrightarrow{f} & W \\ {\scriptstyle \Phi_{\underline{v}}}\uparrow\,{\scriptstyle \simeq} & & {\scriptstyle \Phi_{\underline{v}}}\uparrow\,{\scriptstyle \simeq} \\ K^n & \xrightarrow{a} & K^n \end{array}$$

bestimmte Matrix a ist dann offenbar

$$a = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_r \end{pmatrix}$$

(mit Nullen außerhalb der Diagonalen). Eine solche Matrix nennt man naheliegenderweise eine Diagonalmatrix, und für f selbst rechtfertigt das die folgende

24.11 Sprechweise Ein Endomorphismus f von V mit nur endlich vielen Eigenwerten heißt diagonalisierbar, wenn V die (nach 24.10 direkte) Summe der Eigenräume von f ist.

Bemerkungen Man muß sich vor Augen halten, daß eine solche Diagonalisierung (durch Wahl einer *Eigenbasis*, wie man kurz sagt) den Endomorphismus auf so einfache und durchsichtige Weise beschreibt, wie man sich nur wünschen kann: jeder Basisvektor wird von f bloß mit einem skaleren Faktor multipliziert! Übrigens ist klar, daß die Diagonaleeinträge durch f bis auf die Reihenfolge eindeutig bestimmt sind, denn das sind ja die Eigenwerte, genauer die Nullstellen von χ_f mit ihren Vielfachheiten. Jedenfalls ist Diagonalisierung ein außerordentlich nützliches und beliebtes Hilfsmittel. Zur Illustration gleich zwei

24.12 Anwendungen (1) Eine eher theoretische: Es sei V ein endlichdimensionaler $\mathbb{R}$-Vektorraum und $f\colon V \longrightarrow V$ ein Endomorphismus. Wenn f diagonalisierbar ist und nur nicht-negative Eigenwerte hat, dann gibt es eine Art Quadratwurzel aus f, nämlich einen Endomorphismus $h\colon V \longrightarrow V$ mit $h \circ h = f$, der überdies

ebenfalls diagonalisierbar mit nicht-negativen Eigenwerten und mit f vertauschbar ist. Denn die Wahl einer Eigenbasis reduziert die Frage auf den Spezialfall

$$V = \mathbb{R}^n \quad \text{und} \quad f = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \in \mathrm{Mat}(n\times n, \mathbb{R})$$

mit $\lambda_j \geq 0$ für $j = 1, \ldots, n$, und dann leistet

$$h := \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{pmatrix} \in \mathrm{Mat}(n\times n, \mathbb{R})$$

das Gewünschte.

(2) Eine praktische Anwendung: Beispiel 24.8(1) hat gezeigt, daß $a = \begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix} \in \mathrm{Mat}(2\times 2, \mathbb{R})$ eine diagonalisierbare Matrix ist. Wenn wir nun wissen wollen, wie a^n für große $n \in \mathbb{N}$ aussieht, empfiehlt es sich, nicht mit a selbst zu rechnen, sondern einer Diagonalisierung von a, das heißt der Matrix d, die a bezüglich einer Eigenbasis beschreibt. Als eine solche Basis hatten wir schon

$$\underline{v} = \left(\begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right)$$

bestimmt, und aus dem kommutativen Diagramm

$$\begin{array}{ccc} \mathbb{R}^2 & \xrightarrow{\ a\ } & \mathbb{R}^2 \\ {\scriptstyle \Phi_{\underline{v}}}\uparrow\ \simeq & & \Phi_{\underline{v}}\uparrow\ \simeq \\ \mathbb{R}^2 & \xrightarrow{\ d\ } & \mathbb{R}^2 \end{array}$$

ergibt sich die Beziehung

$$u^{-1}au = d = \begin{pmatrix} \lambda & \\ & \mu \end{pmatrix} = \begin{pmatrix} -1 & \\ & 5 \end{pmatrix} \quad \text{mit } u = \Phi_{\underline{v}} = \begin{pmatrix} -1 & 1 \\ 1 & 2 \end{pmatrix}.$$

Also ist

$$\begin{aligned} a^n = \left(udu^{-1}\right)^n = ud^nu^{-1} &= \begin{pmatrix} -1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} -1 & \\ & 5 \end{pmatrix}^n \begin{pmatrix} -1 & 1 \\ 1 & 2 \end{pmatrix}^{-1} \\ &= \begin{pmatrix} -1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} (-1)^n & \\ & 5^n \end{pmatrix} \left(-\frac{1}{3}\right) \begin{pmatrix} 2 & -1 \\ -1 & -1 \end{pmatrix} \end{aligned}$$

eine geschlossene Formel für a^n, aus der man zum Beispiel den (komponentenweise gebildeten) Grenzwert

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{5^n} a^n &= \frac{1}{3} \begin{pmatrix} -1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} \lim(-\frac{1}{5})^n & \\ & 1 \end{pmatrix} \begin{pmatrix} -2 & 1 \\ 1 & 1 \end{pmatrix} \\ &= -\frac{1}{3} \begin{pmatrix} -1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 0 & \\ & 1 \end{pmatrix} \begin{pmatrix} -2 & 1 \\ 1 & 1 \end{pmatrix} \\ &= \frac{1}{3} \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} \end{aligned}$$

bequem abliest.

Für diagonalisierbare Endomorphismen läßt sich jetzt auch die eingangs gestellte Frage nach den invarianten Unterräumen erschöpfend beantworten:

24.13 Satz f sei ein diagonalisierbarer Endomorphismus des endlichdimensionalen K-Vektorraums V. Es seien $\lambda_1, \dots, \lambda_r \in K$ die verschiedenen Eigenwerte von f, und $E_1, \dots, E_r \subset V$ die zugehörigen Eigenräume. Für jede beliebige Wahl von linearen Teilräumen $U_i \subset E_i$ ist dann

$$U_1 + \cdots + U_r \ \subset \ E_1 + \cdots + E_r = V$$

ein invarianter Unterraum von f, und jeder invariante Unterraum ist von dieser Art.

Beweis Daß diese Unterräume invariant sind, ist klar: f wirkt auf U_i durch Multiplikation mit dem Skalar λ_i. Ist umgekehrt ein beliebiger invarianter Unterraum $U \subset V$ gegeben, so setzen wir $U_i := U \cap E_i$ für $i = 1, \dots, r$ und haben damit sicher

$$U_1 + \cdots + U_r \subset U.$$

Zum Beweis der umgekehrten Inklusion betrachten wir einen beliebigen Vektor $v \in U$; wegen $E_1 + \cdots + E_r = V$ besitzt er eine Darstellung

$$v = \sum_{i=1}^{r} v_i \quad \text{mit } v_i \in E_i.$$

Wenden wir auf diese Summe die Komposition

$$(f - \lambda_2) \circ (f - \lambda_3) \circ \cdots \circ (f - \lambda_r)$$

an, so wird jeder Summand außer v_1 mit null multipliziert, und es bleibt nur

$$(f - \lambda_2) \circ (f - \lambda_3) \circ \cdots \circ (f - \lambda_r)(v) = (\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3) \cdots (\lambda_1 - \lambda_r) v_1.$$

Weil U auch unter dieser Komposition invariant und der Faktor vor v_1 von null verschieden ist, folgt $v_1 \in U$, also $v_1 \in U \cap E_1 = U_1$.

Aus Symmetriegründen gilt ebenso $v_i \in U_i$ für alle i, und damit $v = v_1 + \cdots + v_r \in U_1 + \cdots + U_r$.

Ob ein gegebener Endomorphismus eines endlichdimensionalen Vektorraums diagonalisierbar ist, läßt sich in jedem Fall entscheiden, indem man wie in Beispiel 24.8(1) die Eigenwerte als Nullstellen des charakteristischen Polynoms bestimmt und dann die Eigenräume nach dem Gaußschen Algorithmus berechnet. Wie aus dem nächsten Satz hervorgeht, genügt es manchmal aber schon, die Eigenwerte selbst samt ihren *Vielfachheiten* zu kennen; damit sind die Vielfachheiten als Nullstellen des charakteristischen Polynoms gemeint.

24.14 Satz V sei ein endlichdimensionaler K-Vektorraum. Für jeden Endomorphismus $f: V \longrightarrow V$ gilt:

(a) Zerfällt χ_f in Linearfaktoren, etwa

$$\chi_f(X) = \prod_{j=1}^{n} (X - \lambda_j) \quad \text{mit } \lambda_j \in K,$$

so ist

$$\operatorname{tr} f = \sum_{j=1}^{n} \lambda_j \quad \text{und} \quad \det f = \prod_{j=1}^{n} \lambda_j.$$

(b) f kann nur dann diagonalisierbar sein, wenn χ_f in $K[X]$ in Linearfaktoren zerfällt.

(c) Wenn χ_f in $K[X]$ in paarweise verschiedene Linearfaktoren zerfällt, dann ist f diagonalisierbar.

Beweis Nach Lemma 24.5 sind $-\operatorname{tr} f$ und $(-1)^n \det f$ die Koeffizienten von X^{n-1} und X^0 in $\chi_f(X)$; daraus folgt (a). — Das charakteristische Polynom einer Diagonalmatrix

$$\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

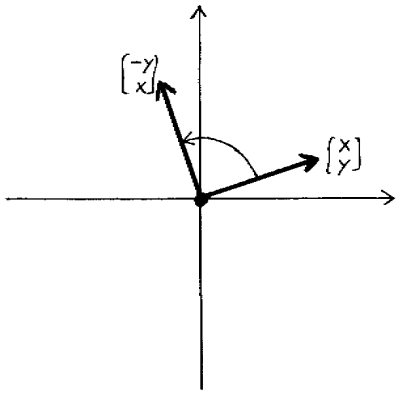
ist $(X-\lambda_1)(X-\lambda_2)\cdots(X-\lambda_n)$, und weil man χ_f bezüglich einer Eigenbasis von V berechnen darf, beweist das (b). — Hat f sogar $n = \dim V$ paarweise verschiedene Eigenwerte, so hat die Summe der Eigenräume mindestens die Dimension n, denn die Summe ist direkt, und jeder einzelne Eigenraum ist mindestens eindimensional. Also ist die Summe zwangsläufig ganz V, und (c) folgt.

Bemerkungen Für die Diagonalisierbarkeit kommt es also wesentlich auf den Körper K an. Insbesondere ist Teil (b) des Lemmas für $K = \mathbb{C}$ gegenstandslos, da in $\mathbb{C}[X]$ ja jedes normierte Polynom in Linearfaktoren zerfällt. Wenn andererseits im Fall $K = \mathbb{R}$ das charakteristische Polynom $\chi_f(X) \in \mathbb{R}[X]$ nicht zerfällt, wird man in der Regel an die Komplexifizierung $f_{\mathbb{C}}$ von f denken: Die hat dasselbe charakteristische Polynom, das aber nun in $\mathbb{C}[X]$ sicher in Linearfaktoren zerfällt und damit kein Hindernis dagegen darstellt, daß wenigstens $f_{\mathbb{C}}$ diagonalisierbar ist. — In der Situation von (c) sind die möglichen Eigenbasen durch f zwar nicht ganz, aber doch weitgehend eindeutig bestimmt: Nur die Reihenfolge der Basisvektoren und für jeden einzelnen ein skalarer Faktor sind noch frei wählbar.

24.15 Beispiele (1) Das charakteristische Polynom der Matrix $a = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in \mathrm{Mat}(2\times 2, \mathbb{R})$

$$\chi_a(X) = \det \begin{pmatrix} X & 1 \\ -1 & X \end{pmatrix} = X^2 + 1 \in \mathbb{R}[X]$$

hat überhaupt keine Nullstellen, a besitzt also keine Eigenvektoren und ist schon gar nicht diagonalisierbar. Das ist auch nicht verwunderlich, denn diese Matrix beschreibt die Drehung der Ebene um einen rechten Winkel (im mathematisch positiven Sinne):



Dagegen zerfällt $\chi_a(X)$ in $\mathbb{C}[X]$ in die Linearfaktoren

$$\chi_a(X) = X^2 + 1 = (X - i)(X + i),$$

nach Teil (c) des Lemmas ist die Komplexifizierung $a_{\mathbb{C}}\colon \mathbb{C}^2 \longrightarrow \mathbb{C}^2$ von a deshalb sehr wohl diagonalisierbar. (Beachten Sie, daß als Matrix $a_{\mathbb{C}} = a$ ist: man muß in diesem Zusammenhang zusätzlich klarmachen, welcher Körper zugrundeliegen soll.) Wenn man den Ehrgeiz hat, die zugehörigen Eigenräume zu bestimmen, muß man die Gleichungen

$$\begin{pmatrix} \pm i & 1 \\ -1 & \pm i \end{pmatrix} x = 0 \quad \text{oder} \quad \begin{pmatrix} 1 & \mp i \end{pmatrix} x = 0$$

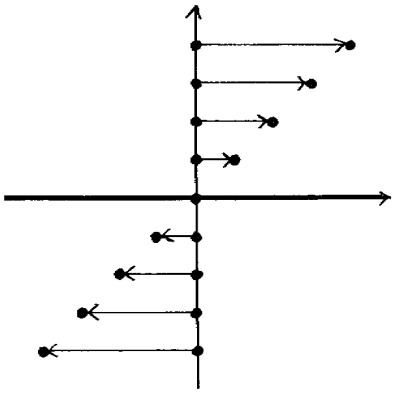
lösen und erhält

$$\mathrm{Lin}\Big(\begin{pmatrix} \pm i \\ 1 \end{pmatrix}\Big) \subset \mathbb{C}^2.$$

Wie vorherzusehen, enthält keiner der beiden Eigenräume einen reellen Eigenvektor.

(2) Wir greifen aus 24.8(2) die Beispielmatrizen $u = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ und $1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ aus $\mathrm{Mat}(2\times 2, \mathbb{R})$ mit dem charakteristischen Polynom $(X-1)^2$ wieder auf. Sie belegen, daß im Fall mehrfacher Eigenwerte die Diagonalisierbarkeit nicht allein anhand des charakteristischen Polynoms entscheidbar ist: die Einheitsmatrix

ist ja selbst schon diagonal, während der einzige Eigenraum von u nur eindimensional ist. Auch das leuchtet anschaulich unmittelbar ein, denn $u\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ beschreibt eine Scherung der Ebene parallel zur ersten Koordinatenachse:



Natürlich macht auch Komplexifizierung diese Matrix nicht diagonalisierbar, denn die Dimension des Eigenraums ändert sich dabei nicht. Dieses Beispiel sei gerade Ihnen zur Einprägung empfohlen, denn manche Physiker neigen dazu, kurzerhand *alle* quadratischen Matrizen zu diagonalisieren oder jedenfalls so zu tun, als ob das möglich sei. Das folgende immer anwendbare Resultat bietet aber einen gewissen Ersatz für die Diagonalisierung.

24.16 Satz V sei ein endlichdimensionaler K-Vektorraum, und f ein Endomorphismus von V, dessen charakteristisches Polynom in Linearfaktoren zerfällt. Dann gibt es eine Basis von V, bezüglich der f durch eine (obere) Dreiecksmatrix

$$\begin{pmatrix} \lambda_1 & * & \dots & \dots & * \\ 0 & \lambda_2 & * & \dots & * \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \lambda_{n-1} & * \\ 0 & \dots & \dots & 0 & \lambda_n \end{pmatrix}$$

beschrieben wird.

In der umgekehrten Richtung gilt natürlich die

24.17 Notiz Das charakteristische Polynom einer oberen Dreiecksmatrix

$$c = \begin{pmatrix} \lambda_1 & * & * \\ & \ddots & * \\ & & \lambda_n \end{pmatrix}$$

(oder auch einer unteren) ist

$$\chi_c = (X - \lambda_1)(X - \lambda_2)\cdots(X - \lambda_n),$$

die Eigenwerte von c sind also die Diagonaleinträge.

Zum Satz selbst: Sowohl für den Beweis als auch für die Anwendungen ist es hilfreich, die Aussage zuerst begrifflicher zu formulieren. Dazu brauchen wir die

24.18 Definition V sei ein n-dimensionaler Vektorraum. Eine Kette

$$\{0\} = V_0 \subset V_1 \subset \cdots \subset V_{k-1} \subset V_k \subset \cdots \subset V_n = V$$

von linearen Teilräumen von V mit $\dim V_k = k$ für alle k nennt man eine Flagge in V. Ist $f\colon V \longrightarrow V$ ein Endomorphismus, so nennt man eine solche Flagge invariant oder stabil unter f, wenn jedes V_k invariant ist:

$$f(V_k) \subset V_k \quad \text{für } k = 1, \dots, n.$$

24.18½ Beispiel Jede Basis $(v_1, \dots, v_n)$ von V definiert die Flagge

$$\{0\} = V_0 \subset \cdots \subset V_k \subset \cdots \subset V_n = V$$

mit $V_k = \mathrm{Lin}(v_1, \dots, v_k)$. Speziell zur Standardbasis von K^n erhält man so die *Standardflagge*

$$\{0\} = K^0 \times \{0\}^n \subset K \times \{0\}^{n-1} \subset \cdots \subset K^k \times \{0\}^{n-k} \subset \cdots \subset K^{n-1} \times \{0\} \subset K^n \times \{0\}^0 = K^n.$$

Sei nun $\underline{v} = (v_1, \dots, v_n)$ eine Basis von V wie im Satz versprochen. Die Dreiecksgestalt der zugehörigen Matrix besagt dann, daß

v_1 ein Eigenvektor von f ist,

v_2 zwar nicht unbedingt ein Eigenvektor ist, aber immerhin in den Unterraum $\mathrm{Lin}(v_1, v_2)$ abgebildet wird,

$f(v_3)$ eine Linearkombination von v_1, v_2, v_3 ist usw.

Das bedeutet aber gerade, daß die durch die Basis $(v_1, \dots, v_k)$ definierte Flagge von V unter f invariant ist. Wir können Satz 24.16 also ebensogut folgendermaßen ausdrücken.

24.16 Satz (begriffliche Fassung) V sei ein endlichdimensionaler K-Vektorraum, und f ein Endomorphismus von V, dessen charakteristisches Polynom in Linearfaktoren zerfällt. Dann gibt es eine Flagge von V, die unter f invariant ist.

Beweis, durch Induktion nach $n = \dim V$. Für $n = 0$ ist die Aussage trivial, sei also $n > 0$. Wir wählen dann einen Eigenwert γ von f. Weil der Endomorphismus $f - \gamma$ nicht surjektiv ist, können wir einen Unterraum $V_{n-1} \subset V$ der Dimension $n-1$ so wählen, daß $\mathrm{Bild}(f - \lambda) \subset V_{n-1}$ ist. Sicher ist V_{n-1} dann stabil unter $f - \gamma$, als linearer Unterraum aber auch unter γ — Multiplikation mit einem Skalar! — und unter $f = (f - \gamma) + \gamma$ selbst. Damit können wir f zu einem Endomorphismus $f' \in \mathrm{End}\, V_{n-1}$ einschränken.

Wenn wir uns eine Basis von V_{n-1} gewählt und mittels eines letzten Vektors zu einer Basis von V ergänzt denken, wird f bezüglich dieser Basis durch eine Matrix der Form

$$c = \left(\begin{array}{ccc|c} & & & * \\ & c' & & \vdots \\ & & & * \\ \hline 0 & \cdots & 0 & \gamma \end{array}\right)$$

dargestellt, in der c' die Matrix von f' ist. Sie erlaubt uns, das charakteristische Polynom von f' zu berechnen: Entwicklung nach der letzten Zeile liefert

$$\chi_f(X) = \chi_c(X) = \det(X - c) = (X - \gamma) \det(X - c') = (X - \gamma)\chi_{c'}(X) = (X - \gamma)\chi_{f'}(X)\,,$$

und weil $\chi_f(X)$ zerfällt und $X - \gamma$ einer der Linearfaktoren ist, muß $\chi_{f'}(X)$ das Produkt der übrigen sein.

Insbesondere zerfällt auch $\chi_{f'}$ in Linearfaktoren. Deshalb finden wir aufgrund der Induktionsannahme eine f'-invariante Flagge $\{0\} = V_0 \subset V_1 \subset \cdots V_{n-1}$ von V_{n-1}, die wir nur mit $V_n := V$ zu einer f-invarianten von V zu ergänzen brauchen.

Bemerkung Über die Einträge oberhalb der Diagonalen macht Satz 24.16 keine Aussage; aber wie eine genauere Analyse zeigt, kann man auch diese durch weitere Verfeinerung der Basiswahl drastisch reduzieren. Die einfachste und im wesentlichen eindeutig bestimmte Gestalt, in die man die Matrix eines Endomorphismus (mit zerfallendem charakteristischem Polynom) bringen kann, ist die sogenannte *jordansche Normalform*. Sie wird in vielen Lehrbüchern der linearen Algebra behandelt — siehe auch den Anhang zu diesem Abschnitt.

24.19 Anwendung Sei V endlichdimensionaler K-Vektorraum, und $f\colon V \longrightarrow V$ ein Endomorphismus mit zerfallendem charakteristischem Polynom. Ist $\lambda \in K$ ein Eigenwert der Vielfachheit $e > 0$ und $E \subset V$ der zugehörige Eigenraum, so ist

$$1 \leq \dim E \leq e.$$

Beweis Aufgrund von Satz 24.16 dürfen wir annehmen, daß $V = K^n$ und f eine obere Dreiecksmatrix

$$f = \begin{pmatrix} \lambda_1 & * & * \\ & \ddots & * \\ & & \lambda_n \end{pmatrix}$$

ist. In der Matrix

$$\lambda - f = \begin{pmatrix} \lambda-\lambda_1 & * & * \\ & \ddots & * \\ & & \lambda-\lambda_n \end{pmatrix}$$

verschwinden dann genau e Diagonaleinträge, also ist $\operatorname{rk}(\lambda - f) \geq n - e$ und deshalb

$$\dim E = \dim \operatorname{Kern}(\lambda - f) \leq e.$$

In 24.19 ist übrigens die Voraussetzung, daß χ_f zerfalle, in Wirklichkeit überflüssig. Im Fall $K = \mathbb{R}$ kann man das sehen, indem man die Komplexifizierung $f_{\mathbb{C}}$ betrachtet.

Übungsaufgaben

24.1 Untersuchen Sie, für welche $\gamma \in \mathbb{C}$ die Matrix

$$c = \begin{pmatrix} i-\gamma & 1 & -i \\ 0 & i & 0 \\ 0 & 2\gamma & i+\gamma \end{pmatrix} \in \operatorname{Mat}(3\times 3, \mathbb{C})$$

diagonalisierbar ist, und bestimmen Sie für jedes solche γ eine Matrix $u \in GL(3, \mathbb{C})$, so daß $u^{-1}cu$ eine Diagonalmatrix ist. Berechnen Sie in jedem Fall alle Eigenräume von c.

24.2 f und g seien miteinander vertauschbare Endomorphismen des Vektorraums V:

$$f \circ g = g \circ f$$

Zeigen Sie, daß dann jeder Eigenraum von g ein unter f invarianter Unterraum von V ist.

24.3 Ist V ein endlichdimensionaler Vektorraum, so gilt

$$\chi_{f\circ g} = \chi_{g\circ f}$$

für je zwei Endomorphismen f, g von V. Beweisen Sie das unter der zusätzlichen (in Wirklichkeit überflüssigen) Annahme, daß einer der beiden Endomorphismen umkehrbar ist. Zeigen Sie, daß man aber $\chi_{f\circ g}$ im allgemeinen nicht aus χ_f und χ_g berechnen kann, selbst dann nicht, wenn f und g miteinander vertauschbar sind.

24.4 Beweisen Sie, daß das Produkt zweier oberer Dreiecksmatrizen wieder eine obere Dreiecksmatrix ist, und zwar

(a) durch direkte Rechnung,

(b) mittels einer Flagge.

24.5 V sei ein endlichdimensionaler K-Vektorraum. Es werden die folgenden Eigenschaften eines linearen Endomorphismus $f\colon V \longrightarrow V$ betrachtet:

(a) es gibt ein $r \in \mathbb{N}$ mit $f^r = f \circ \cdots \circ f = 0$ (solche Endomorphismen nennt man *nilpotent*)

(b) 0 ist der einzig mögliche Eigenwert von f

(c) es gibt eine Basis von V, bezüglich der f durch eine obere Dreicksmatrix c mit lauter Nullen auf der Diagonalen beschrieben wird:

$$c = \begin{pmatrix} 0 & * & * \\ & \ddots & * \\ & & 0 \end{pmatrix}$$

Beweisen Sie: Die Implikationen (a) $\Rightarrow$ (b) sowie (c) $\Rightarrow$ (a) und (c) $\Rightarrow$ (b) gelten immer. Wenn K der Körper der komplexen Zahlen ist, dann gilt auch (b) $\Rightarrow$ (c), also sind dann alle drei Eigenschaften äquivalent.

Zwei natürliche Beispiele nilpotenter Endomorphismen finden Sie in Aufgabe 19.1.

Anhang

Zwei Ergänzungen zu den linearen Endomorphismen möchte ich hier machen. Die erste erklärt, wie man bei der Definition des charakteristischen Polynoms im Fall eines ganz beliebigen Grundkörpers vorgehen kann. Zu einer gegebenen Matrix $a \in \mathrm{Mat}(n\times n, K)$ bestimmt unsere Definition 24.3 das Objekt $\chi_a(X)$ ja als eine Funktion der Unbestimmten X, und zumindest in den Fällen $K = \mathbb{R}$ und $K = \mathbb{C}$ wissen wir als Anwendung des Satzes 3.10, daß diese Funktion die Koeffizienten von $\chi_a(X)$ als algebraischem Term und damit $\chi_a(X)$ selbst im Sinne der Algebra festlegt. Wer etwas Algebra gelernt hat, weiß, daß das auch für andere Körper richtig ist, solange sie unendlich sind — für einen endlichen Körper aber definitiv nicht. In diesem Fall braucht man die folgenden immer gültigen Überlegungen.

Zunächst bemerken wir, daß man für den Begriff einer Matrix natürlich nicht darauf bestehen muß, daß die Einträge in einem Körper liegen. Für die nackte Definition wäre sogar eine bloße Menge gut genug, aber wenn wir statt des Körpers immerhin einen kommutativen Ring R zugrundelegen, bleiben uns auch die elementaren Rechenoperationen mit Matrizen erhalten, so daß wir insbesondere für quadratisches Format den Matrizenring

$$\mathrm{Mat}(n\times n, R) \quad \text{mit einem kommutativen Ring } R$$

erhalten. Gibt es Sinn, von der Determinante einer solchen Matrix über R zu reden? Nun, die Axiome aus dem Abschnitt 23 behalten ihren Sinn; naur liegen die Werte der Determiante jetzt eben in R:

$$\det\colon \mathrm{Mat}(n\times n, R) \longrightarrow R$$

Den Existenzbeweis haben wir damals konstruktiv geführt, nämlich durch Entwicklung nach einer Zeile gemäß 22.5. Wenn Sie diese Konstruktion inspizieren, sehen Sie sofort, daß sie ohne weiteres in R durchführbar ist. Lediglich den Eindeutigkeitsbeweis müssen wir verwerfen, denn er beruhte auf dem gaußschen Algorithmus und zwingt uns damit im allgemeinen zu Divisionen. Es ist aber ganz leicht, einen Eindeutigkeitsbeweis zu führen, der ohne Division auskommt (er wirft aber nicht als Nebenprodukt ein so schönes praktikables Rechenverfahren ab wie unserer). Immerhin ergibt sich damit die Eindeutigkeit der Determinante, und weiter auch die Richtigkeit der wichtigen Sätze 22.8 und 22.10 sowie der Formel 22.19 für Matrizen über R.

Mit diesem erweiterten Werkzeug gewappnet brauchen wir die Definition des charakteristischen Polynoms nur noch neu zu interpretieren. Für jedes $a \in Mat(n\times n, K)$ ist

$$X-a \in \mathrm{Mat}\left(n\times n, K[X]\right)$$

eine Matrix über dem Polynomring $K[X]$, und deren Determinante

$$\chi_a(X) = \det(X-a) \in K[X]$$

eben ein Ringelement, das heißt ein Polynom über K. Das war's schon.

Die zweite Ergänzung ist substanzieller, sie betrifft die sogenannte *jordansche Normalform* von Endomorphismen. Wir setzen durchweg voraus, daß der K-Vektorraum V endlichdimensional und $f \in \mathrm{End}\, V$ ein Endomorphismus mit zerfallendem charakteristischen Polynom ist (letzteres gilt im Fall $K = \mathbb{C}$ automatisch). Satz 24.16 verspricht uns dann eine Basis von V, bezüglich der f durch eine Dreiecksmatrix beschrieben wird. Angesichts der Tatsache, daß Dreiecksmatrizen immer noch ziemlich allgemein sind, mag man daran zweifeln, daß man damit schon die einfachste f beschreibende Matrix vor sich hat! Tatsächlich kann man diese Matrix durch noch geschicktere Basiswahl noch wesentlich vereinfachen, und ich erkläre Ihnen in zwei Schritten wie. Wie wir wissen, stehen in der Diagonalen der Matrix

$$\begin{pmatrix} \lambda_1 & * & \dots & \dots & * \\ 0 & \lambda_2 & * & \dots & * \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \lambda_{n-1} & * \\ 0 & \dots & \dots & 0 & \lambda_n \end{pmatrix} \in \mathrm{Mat}(n\times n, K)$$

ihre Eigenwerte, und noch nach dem Beweis von 24.16 kann man dafür sorgen, daß gleiche Eigenwerte zusammenstehen, die Matrix also die Gestalt

$$\begin{pmatrix} \lambda_1 & * & \dots & \dots & \dots & \dots & \dots & * \\ 0 & \ddots & \ddots & & & & & \vdots \\ \vdots & \ddots & \lambda_1 & \ddots & & & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & & & \vdots \\ \vdots & & & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & & & \ddots & \lambda_r & \ddots & \vdots \\ \vdots & & & & & \ddots & \ddots & * \\ 0 & \dots & \dots & \dots & \dots & \dots & 0 & \lambda_r \end{pmatrix} \in \mathrm{Mat}(n\times n, K)$$

hat, wobei $\lambda_1, \dots, \lambda_r$ jetzt die *verschiedenen* Eigenwerte bezeichnen. Was der Beweis aber nicht zeigt ist, daß alle Einträge, deren Zeilen- und Spaltenindex zu verschiedenen Eigenwerten gehören, zu null machen kann. Damit haben wir eine Matrix der Gestalt

$$\begin{pmatrix} \boxed{d_1} & & \\ & \ddots & \\ & & \boxed{d_r} \end{pmatrix} \in \mathrm{Mat}(n\times n, K),$$

in der jeder einzelne "Kasten" die Form

$$d_j = \begin{pmatrix} \lambda_j & * & \dots & * \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \dots & 0 & \lambda_j \end{pmatrix} \in \mathrm{Mat}(e_j\times e_j, K)$$

hat; hier ist e_j die Vielfachheit des Eigenwertes λ_j als Nullstelle von χ_f. Seien $v_{j1}, \dots, v_{je_j}$ die auf den Kasten d_j entfallenden Basisvektoren und $H_j \subset V$ der von ihnen aufgespannte Unterraum von V. Dann ist $V = H_1 + \cdots + H_r$ und das Zerfallen der Gesamtmatrix in die Kästen besagt gerade, daß jedes H_j ein unter f invarianter Unterraum von V ist. Die H_j übernehmen hier also die Rolle der Eigenräume aus dem diagonalisierbaren Fall.

Natürlich ist die Existenz einer solchen Basis von V eine beweisbedürftige Behauptung. Den Beweis muß ich Ihnen hier vorenthalten, weil man zu ihm zweckmäßigerweise Mittel aus der kommutativen Algebra heranzieht, die wir hier nicht zur Hand haben. Ich zeige Ihnen aber, wie man zu gegebenem f die Unterräume H_j berechnen kann. Eine einfache, aber in dem ganzen Themenkreis sehr nützliche und schon im Beweis von Satz 24.16 zum Zuge gekommene Beobachtung: dieses Problem bleibt unverändert, wenn wir von f ein beliebiges Vielfaches, etwa das λ-fache der identischen Abbildung abziehen: Jede f darstellende Matrix a wird dann eben zu $a-\lambda$, was nur alle Eigenwerte um λ vermindert, aber insbesondere die Räume H_j unverändert läßt. Indem wir das für ein fest gewähltes j mit $\lambda = \lambda_j$ machen, ziehen wir uns also auf den Fall $\lambda_j = 0$ zurück. Die Form von

$$d_j = \begin{pmatrix} 0 & * & \dots & * \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \dots & 0 & 0 \end{pmatrix} \in \mathrm{Mat}(e_j \times e_j, K)$$

zeigt uns, daß dann d_j als Endomorphismus von H_j die durch die Basis $(v_{j1}, \dots, v_{je_j})$ definierte Flagge nicht nur invariant läßt, sondern sie sozusagen komprimiert, nämlich jeden Flaggenraum in seinen Vorgänger abbildet. Dann muß aber die e_j-te Potenz von d_j der Nullhomomorphismus sein! Andererseits ist für jedes $i \neq j$ der Skalar 0 kein Eigenwert von f, also $d_i \in GL(e_i, K)$. Die e_j-te Potenz von f hat — immer in derselben Basis — deshalb die Gestalt

$$\begin{pmatrix} \boxed{d_1} & & & & \\ & \ddots & & & \\ & & \boxed{0} & & \\ & & & \ddots & \\ & & & & \boxed{d_r} \end{pmatrix} \in \mathrm{Mat}(n \times n, K)$$

mit invertierbaren Matrizen $d_i^{e_i} \in GL(e_i, K)$ für $i \neq j$. Insbesondere ist $\mathrm{Kern}\, f^{e_j} = H_j$, und die Verschiebung um λ_j rückgängig machend schließen wir:

$$H_j = \mathrm{Kern}\,(f-\lambda_j)^{e_j} = \mathrm{Kern}\,(f-\lambda_j)^n \quad \text{für jedes } j.$$

Wir fassen das bisher Gesagte wie folgt zitierbar zusammen.

24.20 Definition/Satz Sei V ein n-dimensionaler K-Vektorraum, $f \in \mathrm{End}\, V$ ein Endomorphismus und $\lambda \in K$ ein Skalar. Der Unterrraum

$$H_\lambda = \mathrm{Kern}\,(f-\lambda_j)^n \subset V$$

heißt der *Hauptraum* von f zum Skalar λ. Jeder Hauptraum ist unter f invariant. Ist $e \in \mathbb{N}$ die Vielfachheit von λ als Nullstelle von χ_f, so gilt schon $H_\lambda = \mathrm{Kern}\,(f-\lambda)^e$; insbesondere ist $H_\lambda \neq \{0\}$ genau dann, wenn λ ein Eigenwert von f ist. Wenn χ_f in Linearfaktoren zerfällt, ist V die direkte Summe dieser Haupträume.

Wir wissen jetzt, wie man die Haupträume eines Endomorphismus mit bekannten Eigenwerten bestimmt. Die Suche nach einer möglichst einfachen f beschreibenden Matrix braucht man jetzt nur noch für die Einschränkung von f zu einem Endomorphismus eines jeden Hauptraums durchzuführen — mit anderen Worten für den Fall, daß f einen einzigen (im allgemeinen mehrfachen) Eigenwert λ hat. Auch hier — das ist der zweite Schritt, den ich ohne Beweis mitteile — kommt man wesentlich weiter als Satz 24.16

verspricht, kann nämlich alle Einträge außerhalb der Diagonalen zu null machen mit Ausnahme von Einsen, die unmittelbar rechts der Diagonalen stehen *können* (die möglichen Stellen sind mit Sternen markiert):

$$a = \begin{pmatrix} \lambda & * & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & * \\ 0 & \dots & \dots & 0 & \lambda \end{pmatrix}$$

Jede nicht gesetzte Eins bewirkt eine Zerlegung dieser Matrix in sogennannte *Jordan-Kästchen*

$$J(\lambda, e) = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & 1 \\ 0 & \dots & \dots & 0 & \lambda \end{pmatrix} \in \mathrm{Mat}(e \times e, K)\,,$$

so daß wir alternativ

$$a = \begin{pmatrix} \boxed{J(\lambda, e_1)} & & \\ & \ddots & \\ & & \boxed{J(\lambda, e_s)} \end{pmatrix} \in \mathrm{Mat}(n \times n, K)$$

schreiben können. Kombiniert mit dem vorigen Resultat haben wir also:

24.21 Satz Sei V ein n-dimensionaler K-Vektorraum, $f \in \operatorname{End} V$ ein Endomorphismus mit zerfallendem charakteristischen Polynom. Dann gibt es eine *Jordan-Basis* für f, nämlich eine Basis von V, bezüglich der f durch eine Matrix

$$\begin{pmatrix} \boxed{J(\lambda_1, e_1)} & & \\ & \ddots & \\ & & \boxed{J(\lambda_s, e_s)} \end{pmatrix} \in \mathrm{Mat}(n \times n, K)$$

beschrieben wird, aufgebaut aus einer (unbestimmten) Anzahl s von Jordan-Kästchen

$$J(\lambda, e) = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ & \ddots & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & 0 \\ & & & \ddots & 1 \\ & & & & \lambda \end{pmatrix} \in \mathrm{Mat}(e \times e, K)$$

(mit $\lambda \in K$ und $0 < e \in \mathbb{N}$, und natürlich $\sum_{j=1}^{s} e_j = n$).

Diese sogenannte *jordansche Normalform* für f ist im wesentlichen eindeutig bestimmt: für jedes $\lambda \in K$ und jedes positive $e \in \mathbb{N}$ hängt die Anzahl der auftretenden Kästchen $J(\lambda, e)$ nur von f, nicht von der Wahl der Basis ab.

24.22 Beispiele (1) Es sieht auf den ersten Blick so aus, als ob der diagonalisierbare Fall hier gar keinen Platz fände. Findet er aber doch: es ist genau der Fall, in dem alle Jordan-Kästchen das Formnat 1×1 haben.

(2) Die schon mehrfach betrachteten 2×2-Matrizen $u = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ und $1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ sind selbst in jordanscher Normalform; sie sind die einzig möglichen Normalformen mit dem charakteristischen Polynom $(X-1)^2$.

Konfrontiert mit einem Endomorphismus mit schon in Linearfaktoren zerlegtem charakteristischen Polynom, wie findet man die Normalform? Klar, man bestimmt wie besprochen zu jedem Eigenwert λ den Hauptraum und reduziert dadurch auf den Fall, daß λ der einzige Eigenwert ist. Um zu sehen, welche Jordan-Kästchen auftreten, ziehen wir wieder λ ab und stellen uns ein einziges Kästchen vor:

$$J(0,e) = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ & \ddots & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & 0 \\ & & & \ddots & 1 \\ & & & & 0 \end{pmatrix} \in \mathrm{Mat}(e\times e, K)$$

Dieses läßt nun die Standardflagge

$$\{0\} = K^0\times\{0\}^e \subset K\times\{0\}^{e-1} \subset \cdots \subset K^k\times\{0\}^{e-k} \subset \cdots \subset K^{e-1}\times\{0\} \subset K^e\times\{0\}^0 = K^e$$

nicht nur invariant, sondern komprimiert sie:

$$f\big(K^k\times\{0\}^{e-k}\big) \subset K^{k-1}\times\{0\}^{e-k+1} \quad \text{für jedes } k\,.$$

Diesmal wissen wir noch viel genauer, was passiert: jeder Basisvektor wandert auf seinen Vorgänger, mit Ausnahme des ersten, der auf den Nullvektor geht. Deshalb sind die Flaggenräume gerade die Kerne der Potenzen des Kästchens:

$$K^k\times\{0\}^{e-k} = \mathrm{Kern}\, J(0,e)^k \quad \text{für } k = 0,\dots,e\,.$$

Für den Endomorphismus f folgt daraus aber, daß die Differenz

$$\dim \mathrm{Kern}\,(f-\lambda)^e - \dim \mathrm{Kern}\,(f-\lambda)^{e-1}$$

gerade die Anzahl der Kästchen angibt, deren Größe mindestens e ist. Es ist eine reine Frage der Buchführung, aus dieser Tatsache Art und Anzahl der auftretenden Kästchen zu bestimmen.

Das gilt schließlich auch für die Bestimmung einer Jordan-Basis. Ich begnüge mich mit dem Hinweis, daß sich als letzter Basisvektor für den Endomorphismus $J(0,e)$ jeder eignet, der dessen Anwendung genügend oft überlebt, das heißt jeder Vektor

$$v_e \in \mathrm{Kern}\, J(0,e)^e \setminus \mathrm{Kern}\, J(0,e)^{e-1},$$

und daß die übrigen Basisvektoren dann zwangsläufig

$$v_k := J(0,e)^{e-k}(v_e) \quad \text{für } k = 1,\dots,e-1$$

sind.

Übrigens ist die Zerlegung des Gesamtraumes in Unterräume, die zu den verschiedenen Jordan-Kästchen mit ein und demselben Eigenwert gehören, *nicht* eindeutig bestimmt — anders als die Zerlegung in die Haupträume.

25 Euklidische Vektorräume

Die soweit besprochenen Vektorräume, ganz gleich über welchem Körper, tragen keinerlei metrische Struktur: die Frage nach der Länge eines Vektors oder nach dem Winkel zwischen zwei Vektoren gibt einfach keinen Sinn. Daß es auch ohne diese Begriffe eine reichhaltige und interessante Theorie (und nützliche Praxis) dieser Räume und vor allem der linearen Abbildungen gibt, davon dürften Sie die vorangegangenen Abschnitte überzeugt haben. Aber natürlich sind auch Länge und Winkel interessante, insbesondere in der Physik allgegenwärtige Begriffe, die eine ausführliche Behandlung verdienen. Damit wollen wir jetzt beginnen.

25.1 Definition V sei ein K-Vektorraum. Unter einer symmetrischen Bilinearform auf V versteht man eine Abbildung

$$V \times V \xrightarrow{\beta} K$$

mit den Eigenschaften

• *Bilinearität*: für feste $v, w \in V$ sind die Funktionen $V \ni x \mapsto \beta(x, w) \in K$ und $V \ni y \mapsto \beta(v, y) \in K$ linear, und

• *Symmetrie*: $\beta(v, w) = \beta(w, v)$ für alle $v, w \in V$.

Die Abbildung

$$V \longrightarrow K; \quad v \mapsto \beta(v, v)$$

nennt man die zu β gehörige quadratische Form.

Die Bilinearität verlangt also, daß β in jeder der beiden Variablen linear ist. Ebensowenig wie bei der Determinante folgt daraus, daß β eine lineare Funktion auf dem Produktvektorraum $V \times V$ ist. — Es ist klar, daß es in Gegenwart der Symmetrie genügt, wenn β in *einer* der beiden Variablen linear ist.

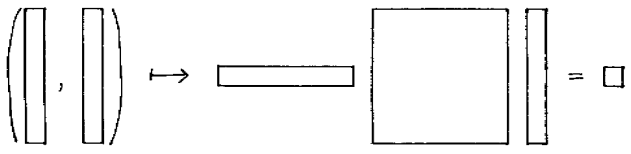
Zur rechnerischen Beschreibung dieser Objekte erklären wir:

25.2 Definition Eine Matrix $s \in \mathrm{Mat}(n \times n, K)$ heißt symmetrisch, wenn $s^{\mathrm{t}} = s$ ist. Die symmetrischen Matrizen bilden den Untervektorraum

$$\mathrm{Sym}(n, K) \subset \mathrm{Mat}(n \times n, K).$$

25.3 Lemma Jede symmetrische Matrix $s \in \mathrm{Sym}(n, K)$ definiert durch

$$K^n \times K^n \ni (x, y) \longmapsto \beta(x, y) := x^{\mathrm{t}} s y \in K$$



eine symmetrische Bilinearform auf K^n, und jede symmetrische Bilinearform auf K^n ist von dieser Art, mit eindeutig bestimmter Matrix $s \in \mathrm{Sym}(n, K)$.

Beweis Die erste Behauptung ist klar; zu zeigen ist nur, daß zu jeder symmetrischen Bilinearform β ein wohlbestimmtes s gehört. Sei also β gegeben. Wenn $\beta(x, y) = x^{\mathrm{t}} s y$ für alle $x, y \in K^n$ gelten soll, muß insbesondere

$$\beta(e_i, e_j) = e_i^{\mathrm{t}} s e_j = s_{ij} \quad \text{für alle } i, j$$

gelten: dadurch ist $s \in \mathrm{Mat}(n\times n, K)$ schon definiert, und dieses s ist auch symmetrisch. Daß die Gleichung $\beta(x,y) = x^{\mathrm{t}}sy$ dann für *alle* $x, y \in K^n$ und nicht nur für die Basisvektoren gilt, ergibt sich mittels der Bilinearität durch einfache Rechnung:

$$\beta(x,y) = \beta\left(\sum_{i=1}^{n} x_i e_i \ , \ \sum_{j=1}^{n} y_j e_j\right) = \sum_{i,j=1}^{n} x_i y_j \beta(e_i, e_j) = \sum_{i,j=1}^{n} x_i s_{ij} y_j = x^{\mathrm{t}} s y$$

Für symmetrische Matrizen haben wir damit eine zweite Interpretation neben der als linearer Abbildung. Es ist nicht überraschend, daß man diese Beschreibung auch auf abstrakte Vektorräume anwenden kann, in denen eine Basis fixiert ist. Wie das genau geht, wird auch hier am übersichtlichsten durch ein kommutatives Diagramm festgehalten. Sei also V ein K-Vektorraum mit Basis $\underline{v} = (v_1, \ldots, v_n)$. Das kommutative Diagramm

$$\begin{array}{ccc} V \times V & \xrightarrow{\beta} & K \\ {\scriptstyle \Phi_{\underline{v}}\times\Phi_{\underline{v}}}\uparrow\ {\scriptstyle \simeq} & & \uparrow{\scriptstyle \mathrm{id}} \\ K^n \times K^n & \xrightarrow{s} & K \end{array}$$

macht aus β eine symmetrische Bilinearform s auf K^n, die wir im Sinne von Lemma 25.3 sofort als Matrix in $\mathrm{Sym}(n, K)$ auffassen dürfen. Natürlich erhält man auf diese Weise auch umgekehrt aus jeder solchen Matrix s eine Form β.

Da das Diagramm anders ist als bei der Interpretation einer Matrix als lineare Abbildung, hat ein Wechsel der Basis auch eine andere Wirkung auf die Matrix. Ist nämlich $\underline{w} = (w_1, \ldots, w_n)$ eine zweite Basis von V, so definiert das zusammengesetzte Diagramm

$$\begin{array}{ccc} K^n \times K^n & \xrightarrow{t} & K \\ {\scriptstyle \simeq}\ {\scriptstyle \Phi_{\underline{w}}\times\Phi_{\underline{w}}}\downarrow & & \downarrow{\scriptstyle \mathrm{id}} \\ V \times V & \xrightarrow{\beta} & K \\ {\scriptstyle u\times u}\uparrow & & \uparrow \\ K^n \times K^n & \xrightarrow{s} & K \end{array}$$

mit $u = \Phi_{\underline{v}}^{-1} \circ \Phi_{\underline{w}} \in GL(n, K)$ die Matrix t. Ihre Wirkung auf Paare von Spalten ist

$$x^{\mathrm{t}} t y = (ux)^{\mathrm{t}} s (uy) = x^{\mathrm{t}} u^{\mathrm{t}} s u y,$$

und deshalb ist

$$t = u^{\mathrm{t}} s u$$

die Matrix, die β bezüglich $\underline{w}$ beschreibt.

Im weiteren Verlauf dieses Abschnitts betrachten wir nur noch *reelle* Vektorräume.

25.4 Definition V sei ein $\mathbb{R}$-Vektorraum. Ein Skalarprodukt auf V ist eine symmetrische Bilinearform β auf V, die positiv definit ist:

$$\beta(v,v) > 0 \quad \text{für alle } v \in V \setminus \{0\}.$$

Bemerkungen Die positive Definitheit ist offensichtlich eine Eigenschaft der zu β gehörigen quadratischen Form. Wegen der Bilinearität gilt ohnehin $\beta(0,0)=0$, also insbesondere $\beta(v,v)\geq 0$ für alle $v\in V$, wenn β ein Skalarprodukt ist. — Im Fall $V=\mathbb{R}^n$ nennt man auch die β entsprechende Matrix $s\in \mathrm{Sym}(n,\mathbb{R})$ positiv definit. Ist V beliebig und in V eine Basis gegeben, so sieht man sofort, daß β genau dann ein Skalarprodukt ist, wenn die zughörige Matrix bezüglich dieser Basis positiv definit ist.

25.5 Definition Einen $\mathbb{R}$-Vektorraum V mit einem Skalarprodukt β auf V — ganz formal geschrieben also das Paar (V,β) — nennt man einen euklidischen Vektorraum. Das Skalarprodukt $(v,w)\mapsto\beta(v,w)$ schreibt man dann gern als

$$V\times V\ni(v,w)\mapsto\langle v,w\rangle\in\mathbb{R},$$

und man nennt die Funktion

$$V\ni v\mapsto\sqrt{\langle v,v\rangle}=:\|v\|\in[0,\infty)$$

die Norm(-abbildung). Die Norm $\|v\|$ des Vektors v nennt man auch den Betrag oder die Länge von v.

Damit haben wir nach langer Zeit wieder ernsthaft mit reellen Zahlen zu tun, und wir beweisen gleich die sogenannte

25.6 Schwarzsche Ungleichung Für je zwei Vektoren v,w eines euklidischen Vektorraums V gilt:

$$|\langle v,w\rangle|\leq\|v\|\cdot\|w\|$$

Gleichheit tritt genau dann ein, wenn v und w linear abhängig sind.

Beweis Für $w=0$ ist $\langle v,w\rangle=0$ und die Aussage trivial.

Für $w\neq 0$ ist wegen der Definitheit $\langle w,w\rangle>0$, und wir setzen trickreich

$$\lambda:=\frac{\langle v,w\rangle}{\langle w,w\rangle}\in\mathbb{R}.$$

Aufgrund der Eigenschaften des Skalarprodukts gilt dann:

$$\begin{aligned}0&\leq\langle v-\lambda w\,,\,v-\lambda w\rangle\\&=\langle v,v\rangle-\langle v,\lambda w\rangle-\langle\lambda w,v\rangle+\langle\lambda w,\lambda w\rangle\\&=\langle v,v\rangle-2\lambda\langle v,w\rangle+\lambda^2\langle w,w\rangle\\&=\|v\|^2-2\frac{\langle v,w\rangle^2}{\|w\|^2}+\frac{\langle v,w\rangle^2}{\|w\|^2}\\&=\|v\|^2-\frac{\langle v,w\rangle^2}{\|w\|^2}\end{aligned}$$

Also

$$\langle v,w\rangle^2\leq\|v\|^2\cdot\|w\|^2$$

oder gleichwertig

$$|\langle v,w\rangle|\leq\|v\|\cdot\|w\|.$$

Wegen der Definitheit des Skalarproduktes gilt in der Abschätzung die Gleichheit nur dann, wenn $v-\lambda w=0$ ist; dann sind v und w sicher linear abhängig. Ist umgekehrt die lineare Abhängigkeit bekannt, so ist entweder $w=0$ und die Gleichheit gilt trivialerweise, oder es ist $v=\mu w$ mit $\mu\in\mathbb{R}$ und die Gleichheit folgt aus

$$|\langle v,w\rangle|=|\langle\mu w,w\rangle|=|\mu|\cdot\|w\|^2=\|\mu w\|\cdot\|w\|=\|v\|\cdot\|w\|.$$

25.7 Folgerung Die Norm $v\mapsto\|v\|$ hat die Eigenschaften

- $\|v\|\geq 0$ immer
- $\|v\|=0\iff v=0$

- $\|\lambda v\| = |\lambda| \cdot \|v$ für alle $\lambda \in \mathbb{R}, v \in V$
- $\|v \pm w\| \le \|v\| + \|w\|$ für alle $v, w \in V$

Beweis Nur die letzte Eigenschaft ist nicht offensichtlich. Man erhält sie aus der Schwarzschen Ungleichung, indem man aus

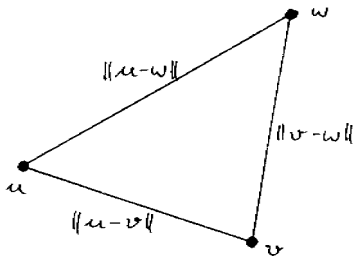
$$\big(\|v\| + \|w\|\big)^2 = \|v\|^2 + 2\|v\| \cdot \|w\| + \|w\|^2 \ge \|v\|^2 + 2\langle v, w\rangle + \|w\|^2 = \langle v+w, v+w\rangle = \|v + w\|^2$$

die Wurzel zieht (und eventuell w durch $-w$ ersetzt).

Für gegebene Vektoren $u, v, w \in V$ kann man die Dreiecksungleichung wegen $u-w = (u-v) + (v-w)$ auch

$$\|u - w\| \le \|u - v\| + \|v - w\|$$

schreiben und so interpretieren:



Die Länge einer jeden Seite des von u, v und w aufgespannten Dreiecks ist höchstens die Summe der beiden anderen Längen. Daher kommt der Name "Dreiecksungleichung"; bei den in 2.15 und 10.3 so bezeichneten Ungleichungen handelt es sich im wesentlichen um Spezialfälle, nämlich die $\mathbb{R}$-Vektorräume $V = \mathbb{R}$ und $V = \mathbb{C} = \mathbb{R}^2$. In welcher Weise das euklidische Vektorräume sind, ist im ersten der folgenden Beispiele beschrieben.

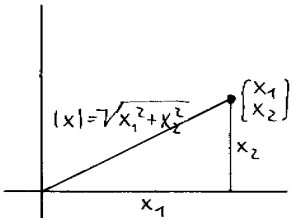
25.8 Beispiele (1) Für jedes $n \in \mathbb{N}$ trägt $\mathbb{R}^n$ als besonders einfaches Skalarprodukt das durch die Einheitsmatrix $1 \in \mathrm{Sym}(n, \mathbb{R})$ bestimmte:

$$\langle x, y\rangle = x^{\mathrm{t}} 1 y = x^{\mathrm{t}} y,$$

in Komponenten ausgeschrieben

$$\left\langle \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \right\rangle = \sum_{j=1}^{n} x_j y_j$$

(natürlich ist das definit: für $x \ne 0$ ist $\|x\| = \sum_j x_j^2 > 0$). Immer dann, wenn man von $\mathbb{R}^n$ als euklidischem Vektorraum redet und nicht ausdrücklich etwas Anderes sagt, ist als Skalarprodukt dieses kanonische oder Standardskalarprodukt gemeint. Der zugehörige Längenbegriff ist der anschauliche nach dem Satz von Pythagoras:



Physiker notieren dieses Produkt meist mit einem mehr oder weniger fetten (oder ganz fehlenden) Punkt, irgendwo zwischen $\mathbf{xy}$ und $\mathbf{x} \bullet \mathbf{y}$. Ich tue das nicht so gern, weil es mit der Matrizenmultiplikation nicht verträglich ist, bleibe lieber bei $x^{\mathrm{t}} y$ oder dem in jedem euklidischen Vektorraum brauchbaren $\langle x, y\rangle$. Den Betrag einer Spalte $x \in \mathbb{R}^n$ schreibt man oft

$$|x| = \sqrt{\sum_{j=1}^{n} x_j^2}$$

mit nur einfachen Strichen, in Übereinstimmung mit der schon in der Analysis verwendeten Notation.

(2) Für ein System von n Massenpunkten in $\mathbb{R}^3$ mit den Massen $m_1, \dots, m_n$ bezeichne

$$\begin{pmatrix} u_j \\ v_j \\ w_j \end{pmatrix} \in \mathbb{R}^3$$

den Geschwindigkeitsvektor des j-ten Massenpunkts. Die Geschwindigkeit des ganzes Systems wird durch den zusammengefaßten Vektor

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} \in \mathbb{R}^{3n}$$

mit

$$u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, \quad v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}, \quad w = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \in \mathbb{R}^n$$

beschrieben. Dann wird durch

$$\mu\left(\begin{pmatrix} u \\ v \\ w \end{pmatrix}, \begin{pmatrix} u' \\ v' \\ w' \end{pmatrix}\right) = \frac{1}{2}\sum_{j=1}^{n} m_j u_j u'_j + \frac{1}{2}\sum_{j=1}^{n} m_j v_j v'_j + \frac{1}{2}\sum_{j=1}^{n} m_j w_j w'_j$$

ein Skalarprodukt μ auf $\mathbb{R}^{3n}$ definiert, dessen zugehörige Matrix eine Diagonalmatrix ist; ihre Diagonaleinträge sind, bis auf den Faktor $\frac{1}{2}$, die Massen m_j. Die durch das Skalarprodukt bestimmte quadratische Form ist die kinetische Energie

$$T\colon \begin{pmatrix} u \\ v \\ w \end{pmatrix} \mapsto \mu\left(\begin{pmatrix} u \\ v \\ w \end{pmatrix}, \begin{pmatrix} u \\ v \\ w \end{pmatrix}\right) = \frac{1}{2}\sum_{j=1}^{n} m_j u_j^2 + \frac{1}{2}\sum_{j=1}^{n} m_j v_j^2 + \frac{1}{2}\sum_{j=1}^{n} m_j w_j^2.$$

Differenziert man sie nach einer der $3n$ Variablen, etwa nach u_j (unter Festhalten der übrigen $3n-1$), so erhält man die entsprechende Impulskomponente $m_j u_j$.

(3) Ganz ähnlich verhält es sich mit der Rotationsenergie eines starren Körpers. Der Bewegungszustand eines starren Körpers mit einem festgehaltenen Punkt ist durch den Vektor der momentanen Winkelgeschwindigkeit charakterisiert, den wir in einem körperfesten kartesischen Koordinatensystem durch eine Spalte $\omega \in \mathbb{R}^3$ beschreiben. Der Trägheitstensor des Körpers ist dann ein Skalarprodukt θ auf $\mathbb{R}^3$, dessen zugehörige quadratische Form bis auf den Faktor $\frac{1}{2}$ die kinetische Energie

$$T\colon \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} \mapsto \frac{1}{2}\theta\left(\begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix}, \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix}\right) = \frac{1}{2}\sum_{j,k=1}^{3} \omega_j \theta_{jk} \omega_k$$

ist. Die Trägheitsmatrix θ ist hier zwar symmetrisch, im allgemeinen aber nicht diagonal. Ihre Definitheit spiegelt die Tatsache wieder, daß die kinetische Energie, ausgenommen im Ruhezustand, stets positiv ist. Differenzieren der Energie nach einer der drei Variablen liefert auch hier die entsprechende (Dreh-)impulskomponente.

(4) Fast selbstverständlich: Jeder Unterraum eines euklidischen Vektorraums ist selbst euklidisch, mit dem eingeschränkten Skalarprodukt natürlich.

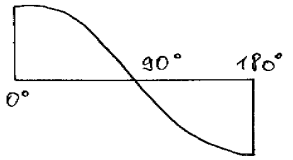
Wir haben schon den durch ein Skalarprodukt gegebenen Längenbegriff erklärt. Die Schwarzsche Ungleichung gestattet es, nun auch Winkel zu definieren.

25.9 Definition V sei ein euklidischer Vektorraum. Für zwei Vektoren $v, w \in V \setminus \{0\}$ definiert man den Winkel t zwischen v und w durch

$$t := \arccos \frac{\langle v, w\rangle}{\|v\| \cdot \|w\|} \in [0, \pi];$$

die Schwarzsche Ungleichung stellt dabei sicher, daß der Bruch in $[-1, 1]$ liegt, wie es für die Anwendung der Arcuscosinusfunktion ja sein muß.

Im Gegensatz zu dem ebenfalls als Winkel interpretierbaren t der komplexen Zahl e^{it} handelt es sich hier um einen "ungerichteten" Winkel, der sich unter Vertauschung von v mit w nicht ändert. Wie aus dem Verlauf der Cosinusfunktion hervorgeht,



nimmt er auch nur Werte zwischen 0 und 180° an. Er ändert sich nicht, wenn man v oder w mit einem positiven Skalar multipliziert. Tatsächlich — und das wird Ihnen aus der Physik schon vertraut sein — spielt der Winkel in der Vektorrechnung nur eine untergeordnete und eher veranschaulichende Rolle. Das liegt daran, daß die gesamte Information über den Winkel ja schon in den drei Skalarprodukten $\langle v, w\rangle$, $\langle v, v\rangle$ und $\langle w, w\rangle$ enthalten ist und sich mit diesen viel besser rechnen läßt als dem Winkel selbst. Für die Anschauung merken wollen wir uns vor allem, daß positives Skalarprodukt einen spitzen, negatives einen stumpfen Winkel zwischen den beiden Vektoren verrät. Besonders wichtig aber ist der dazwischenliegende Fall des rechten Winkels.

25.10 Definition V sei ein euklidischer Vektorraum. Man nennt zwei Vektoren $v, w \in V$ (die auch null sein dürfen) zueinander senkrecht oder orthogonal, wenn

$$\langle v, w\rangle = 0$$

ist. Ein r-tupel $(v_1, \dots, v_r)$ heißt orthonormal oder ein Orthonormalsystem, wenn

$$\langle v_j, v_k\rangle = 1_{jk}\ (= \delta_{jk}) \quad \text{für } j, k = 1, \dots, r$$

gilt.

Die Vektoren v_j eines Orthonormalsystems stehen also paarweise aufeinander senkrecht und sind alle auf die Länge $\|v_j\| = 1$ *normiert*, wie man auch sagt. Als Beispiel eines solchen Orthonormalsystems drängt sich die Standardbasis $(e_1, \dots, e_n)$ von $\mathbb{R}^n$ auf.

25.11 Lemma Jedes Orthonormalsystem ist linear unabhängig.

Beweis Sei $(v_1, \dots, v_r)$ ein solches System, und sei

$$v = \sum_{k=1}^{r} \lambda_k v_k \quad \text{mit } \lambda_k \in \mathbb{R}$$

eine zunächst beliebige Linearkombination. Für jedes $j \in \{1, \dots, r\}$ ist dann

$$\langle v_j, v\rangle = \left\langle v_j, \sum_{k=1}^{r} \lambda_k v_k \right\rangle = \sum_{k=1}^{r} \lambda_k \langle v_j, v_k\rangle = \sum_{k=1}^{r} \lambda_k 1_{jk} = \lambda_j.$$

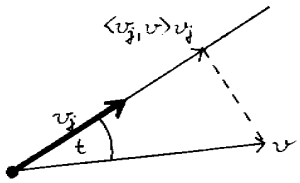
Ist speziell $v = 0$, so folgt $\lambda_j = \langle v_j, 0\rangle = 0$ für alle j, womit wir schon fertig sind.

In einem endlichdimensionalen euklidischen Vektorraum V ist ein orthonormales $(\dim V)$-tupel also stets eine Basis. Aus dem letzten Beweis wollen wir gleich festhalten, wie sich die Koeffizienten eines Vektors bezüglich einer solchen Orthonormalbasis formelmäßig ausdrücken lassen:

25.12 Notiz Ist $(v_1, \dots, v_n)$ eine Orthonormalbasis von V, so ist

$$v = \sum_{j=1}^{n} \langle v_j, v \rangle v_j \quad \text{für jedes } v \in V.$$

Gemäß 25.9 ist der Koeffizient von v_j demnach gleich der Länge von v mal dem Cosinus des Winkels zwischen v_j und v.



Der folgende Satz ist grundlegend; er sichert unter anderem die Existenz von Orthonormalbasen.

25.13 Orthonormalisierungssatz V sei ein endlichdimensionaler euklidischer Vektorraum. Gegeben seien eine Flagge

$$\{0\} = V_0 \subset V_1 \subset \cdots \subset V_{k-1} \subset V_k \subset \cdots \subset V_n = V$$

in V und ein orthonormales r-tupel $(v_1, \dots, v_r)$ mit

$$\mathrm{Lin}(v_1, \dots, v_k) = V_k \text{ für } k = 1, \dots, r.$$

Dieses r-tupel kann man dann zu einer Orthonormalbasis $(v_1, \dots, v_n)$ von V derart ergänzen, daß

$$\mathrm{Lin}(v_1, \dots, v_k) = V_k \text{ für alle } k = 1, \dots, n$$

gilt.

Beweis Es genügt, den Fall $n = r + 1$ zu behandeln, denn dann hilft Induktion (etwa nach $n-r$) weiter (es ist ja klar, daß das Skalarprodukt von V auch jeden Unterraum von V, insbesondere alle Räume V_k selbst zu euklidischen Vektorräumen macht).

Wir fangen mit einem beliebigen Vektor $v \in V_{r+1} \setminus V_r$ an und bilden

$$w := v - \langle v_1, v \rangle v_1 - \langle v_2, v \rangle v_2 - \cdots - \langle v_r, v \rangle v_r;$$

offenbar gilt dann

$$\mathrm{Lin}(v_1, \dots, v_r, w) = \mathrm{Lin}(v_1, \dots, v_r, v) = V_{r+1}.$$

Der Fortschritt, den w gegenüber v darstellt, liegt darin, daß w auf $v_1, \dots, v_r$ senkrecht steht:

$$\langle v_j, w \rangle = \left\langle v_j, v - \sum_{k=1}^{r} \langle v_k, v \rangle v_k \right\rangle = \langle v_j, v \rangle - \langle v_j, v \rangle \langle v_j, v_j \rangle = 0$$

Wegen $w \notin V_r$ ist $w \neq 0$, und der deshalb definierte Vektor

$$v_{r+1} := \frac{1}{\|w\|} w$$

ist zusätzlich noch normiert; damit erfüllt $(v_1, \dots, v_{r+1})$ alle Forderungen.

Bemerkung In der Literatur finden Sie diesen Beweis als "Gram-Schmidtsches Orthonormalisierungsverfahren", aber in der Regel ohne eine zitierfähige Formulierung dessen, was damit eigentlich bewiesen wird. Tatsache ist immerhin, daß das Verfahren als solches wichtig ist und man es sich merken muß.

25.14 Beispiel Die Flagge in V wird in der Praxis meist durch eine Basis angegeben; hier im Standard-$\mathbb{R}^3$ durch

$$V_0 = \mathrm{Lin}() \subset V_1 = \mathrm{Lin}\left(\begin{pmatrix}1\\0\\0\end{pmatrix}\right) \subset V_2 = \mathrm{Lin}\left(\begin{pmatrix}1\\0\\0\end{pmatrix}\ \begin{pmatrix}1\\1\\1\end{pmatrix}\right) \subset V_3 = \mathrm{Lin}\left(\begin{pmatrix}1\\0\\0\end{pmatrix}\ \begin{pmatrix}1\\1\\1\end{pmatrix}\ \begin{pmatrix}1\\1\\0\end{pmatrix}\right).$$

Natürlich ist die Angabe des letzten Basisvektors eigentlich überflüssig, aber wir können damit den Ablauf des Verfahrens ganz festlegen, indem wir als den willkürlich zu wählenden Vektor v immer den nächsten Vektor der gegebenen Basis nehmen.

Der erste Vektor $\begin{pmatrix}1\\0\\0\end{pmatrix}$ ist schon normiert; wir können ihn also gleich als

$$v_1 = \begin{pmatrix}1\\0\\0\end{pmatrix}$$

übernehmen. Der nächste Schritt macht aus

$$v = \begin{pmatrix}1\\1\\1\end{pmatrix}$$

zuerst

$$w = \begin{pmatrix}1\\1\\1\end{pmatrix} - \left\langle \begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}1\\1\\1\end{pmatrix} \right\rangle \begin{pmatrix}1\\0\\0\end{pmatrix} = \begin{pmatrix}1\\1\\1\end{pmatrix} - 1 \cdot \begin{pmatrix}1\\0\\0\end{pmatrix} = \begin{pmatrix}0\\1\\1\end{pmatrix};$$

wegen $|w| = \sqrt{1^2+1^2} = \sqrt{2}$ muß w anschließend noch zu

$$v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix}0\\1\\1\end{pmatrix} = \frac{1}{2}\sqrt{2} \begin{pmatrix}0\\1\\1\end{pmatrix}$$

normiert werden (die bei diesem Verfahren naturgemäß auftretenden Wurzeln sollte man immer sofort durch Erweitern aus dem Nenner entfernen).

Dritter und letzter Schritt:

$$v = \begin{pmatrix}1\\1\\0\end{pmatrix}$$

führt über

$$\begin{aligned} w &= \begin{pmatrix}1\\1\\0\end{pmatrix} - \left\langle \begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}1\\1\\0\end{pmatrix} \right\rangle \begin{pmatrix}1\\0\\0\end{pmatrix} - \left\langle \frac{1}{2}\sqrt{2} \begin{pmatrix}0\\1\\1\end{pmatrix}, \begin{pmatrix}1\\1\\0\end{pmatrix} \right\rangle \frac{1}{2}\sqrt{2} \begin{pmatrix}0\\1\\1\end{pmatrix} \\ &= \begin{pmatrix}1\\1\\0\end{pmatrix} - 1 \cdot \begin{pmatrix}1\\0\\0\end{pmatrix} - \frac{1}{2}\sqrt{2} \cdot \frac{1}{2}\sqrt{2} \begin{pmatrix}0\\1\\1\end{pmatrix} = \frac{1}{2} \begin{pmatrix}0\\1\\-1\end{pmatrix}; \end{aligned}$$

mit $|w| = \frac{1}{2}\sqrt{1^2+1^2} = \frac{1}{2}\sqrt{2}$ zu

$$v_3 = \frac{1}{\frac{1}{2}\sqrt{2}} \cdot \frac{1}{2} \begin{pmatrix}0\\1\\-1\end{pmatrix} = \frac{1}{2}\sqrt{2} \begin{pmatrix}0\\1\\-1\end{pmatrix}.$$

Insgesamt ergibt sich

$$\left(\begin{pmatrix}1\\0\\0\end{pmatrix}, \frac{1}{2}\sqrt{2} \begin{pmatrix}0\\1\\1\end{pmatrix}, \frac{1}{2}\sqrt{2} \begin{pmatrix}0\\1\\-1\end{pmatrix} \right)$$

als die gesuchte Orthonormalbasis von $\mathbb{R}^3$.

Weil jede Basis eines Vektorraums V eine Flagge in V liefert, hat der Orthonormalisierungssatz die unmittelbare

25.15 Folgerung Jeder endlichdimensionale euklidische Vektorraum besitzt eine Orthonormalbasis.

Es liegt nahe, unter den linearen Abbildungen zwischen zwei euklidischen Vektorräumen diejenigen besonders auszuzeichnen, die mit den Skalarprodukten verträglich sind.

25.16 Definition V und W seien euklidische Vektorräume. Eine isometrische Abbildung oder Isometrie von V nach W ist eine lineare Abbildung $f\colon V \longrightarrow W$ mit der Eigenschaft

$$\big\langle f(v), f(w)\big\rangle = \langle v, w\rangle \quad \text{für alle } v, w \in V.$$

25.17 Lemma Isometrien sind stets injektiv. Ist $\underline{v} = (v_1, \dots, v_n)$ eine Orthonormalbasis von V, so ist die lineare Abbildung $f\colon V \longrightarrow W$ genau dann isometrisch, wenn

$$f(\underline{v}) := \big(f(v_1), \dots, f(v_n)\big)$$

ein Orthonormalsystem in W ist.

Beweis Aus $f(v) = 0$ folgt für isometrisches f

$$0 = \big\langle f(v), f(v)\big\rangle = \langle v, v\rangle = \|v\|^2$$

und damit $v = 0$.

Wenn f isometrisch ist, sendet es natürlich Orthonormalsysteme auf Orthonormalsysteme. Zu beweisen bleibt: Wenn $f(\underline{v})$ orthonormal ist, dann ist f isometrisch. Dazu schreiben wir beliebige Vektoren $v, w \in V$ als Linearkombinationen

$$v = \sum_{j=1}^{n} \lambda_j v_j \quad \text{und} \quad w = \sum_{k=1}^{n} \mu_k v_k$$

der Basisvektoren und rechnen: Einerseits ist

$$\langle v, w\rangle = \left\langle \sum_j \lambda_j v_j, \sum_k \mu_k v_k \right\rangle = \sum_{j,k} \lambda_j \mu_k \langle v_j, v_k\rangle = \sum_j \lambda_j \mu_j,$$

andererseits ergibt

$$\begin{aligned}
\big\langle f(v), f(w)\big\rangle &= \left\langle f\left(\sum_j \lambda_j v_j\right), f\left(\sum_k \mu_k v_k\right)\right\rangle \\
&= \left\langle \sum_j \lambda_j f(v_j), \sum_k \mu_k f(v_k)\right\rangle \\
&= \sum_{j,k} \lambda_j \mu_k \big\langle f(v_j), f(v_k)\big\rangle \\
&= \sum_j \lambda_j \mu_j
\end{aligned}$$

dasselbe.

Als bloße Umformulierung der Folgerung 25.15 erhalten wir nun den wichtigen

25.18 Satz Ist V ein n-dimensionaler euklidischer Vektorraum, so gibt es einen isometrischen Isomorphismus zwischen V und dem euklidischen $\mathbb{R}^n$ (mit dem Standardskalarprodukt).

Beweis Es genügt, eine Orthonormalbasis von V zu wählen und die zugehörige Karte $V \xrightarrow{\sim} \mathbb{R}^n$ zu nehmen.

Wie aus Satz 19.2 bekannt ist, kann man den Spaltenraum K^n als Standardmodell für jeden n-dimensionalen K-Vektorraum ansehen; jedes Problem in einem solchen Vektorraum kann man im Prinzip lösen, wenn man das entsprechende Problem in K^n lösen kann. In analogem Sinne sagt Satz 25.18, daß es ein Standardmodell für alle n-dimensionalen euklidischen Vektorräume gibt, nämlich $\mathbb{R}^n$ mit dem Standardskalarprodukt. Jede Fragestellung in einem n-dimensionalen euklidischen Vektorraum ist zu einer Fragestellung in diesem Standard-$\mathbb{R}^n$ gleichwertig (was ihre Behandlung sehr vereinfachen kann, aber nicht muß). Insbesondere ist jedes ganz beliebige Skalarprodukt auf $\mathbb{R}^n$, repräsentiert durch eine positiv definite symmetrische $n \times n$-Matrix, nicht grundsätzlich komplizierter als das Ihnen aus der Physik vertraute Standardprodukt.

Übungsaufgaben

25.1(a) Es seien K ein Körper, der die rationalen Zahlen enthält, V ein K-Vektorraum und β eine symmetrische Bilinearform auf V. Zeigen Sie, daß β durch die zugehörige quadratische Form $q\colon v \mapsto \beta(v,v)$ eindeutig bestimmt ist. (Holen Sie sich eine Idee in $\mathbb{R}$: wie kann man jedes Produkt vw zweier reeller Zahlen v und w allein durch Quadrate ausdrücken?)

(b) V und W seien euklidische Vektorräume. Zeigen Sie, daß jede lineare Abbildung $f\colon V \longrightarrow W$ mit

$$\|f(v)\| = \|v\| \quad \text{für alle } v \in V$$

eine Isometrie ist: "Längentreue impliziert Winkeltreue".

25.2 Im dreidimensionalen Raum werde eine Basis so gewählt, daß der Nullvektor und die drei Basisvektoren zusammen die Ecken eines regulären (platonischen) Tetraeders bilden, dessen Kanten einen Nanometer lang sind. Berechnen Sie das Potential einer im Nullpunkt angebrachten Elementarladung bezüglich dieser Basis.

25.3 In $\mathbb{R}^4$ sei eine Flagge

$$\{0\} = V_0 \subset V_1 = \mathrm{Lin}(u_1) \subset V_2 = \mathrm{Lin}(u_1, u_2) \subset V_3 = \mathrm{Lin}(u_1, u_2, u_3) \subset V_4 = \mathbb{R}^4$$

durch

$$u_1 := \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad u_2 := \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{und} \quad u_3 := \begin{pmatrix} 3 \\ 1 \\ 3 \\ 5 \end{pmatrix}$$

gegeben. Konstruieren Sie eine Orthonormalbasis (v_1, v_2, v_3, v_4) von $\mathbb{R}^4$ mit

$$V_k = \mathrm{Lin}(v_1, \dots, v_k) \quad \text{für } k = 0, 1, 2, 3, 4.$$

25.4 Berechnen Sie eine Orthonormalbasis von

$$V := \left\{ x \in \mathbb{R}^4 \,\middle|\, 3x_1 = x_2 + x_3 + x_4 \right\} \subset \mathbb{R}^4.$$

26 Orthogonale Abbildungen und Komplemente

26.1 Definition V sei ein euklidischer Vektorraum. Eine orthogonale Abbildung von V ist ein linearer Automorphismus $f: V \longrightarrow V$, der zugleich eine Isometrie ist. Es ist offensichtlich, daß die orthogonalen Abbildungen von V unter der Komposition eine Gruppe bilden, man bezeichnet sie mit

$$O(V) \subset GL(V)$$

und nennt sie die orthogonale Gruppe von V. Ist $V = \mathbb{R}^n$ mit dem Standardskalarprodukt, so schreibt man kurz $O(n)$ statt $O(\mathbb{R}^n)$.

Bemerkungen Der Name "orthogonal" ist insofern irreführend, als er suggeriert, daß von der linearen Abbildung f nur erwartet wird, daß sie rechte Winkel erhält. Das wäre aber zu wenig, denn die Multiplikation mit dem Skalar 2 erhält ja sogar alle Winkel, verdoppelt aber die Längen und ist deshalb nicht orthogonal. Eigentlich sollte man besser "ortho*normal*" sagen, aber das hat sich nicht durchgesetzt. Die Terminologie ist in diesem Punkt ziemlich uneinheitlich: "orthogonal", "isometrisch" und auch "unitär" werden alle in mehr oder weniger gleicher Bedeutung verwendet. — Weil Isometrien automatisch injektiv sind, ist bei einem endlichdimensionalen Vektorraum V jede isometrische lineare Abbildung $f: V \longrightarrow V$ umkehrbar und damit orthogonal. In dem praktisch besonders wichtigen Fall des Standard-$\mathbb{R}^n$ redet man natürlich von *orthogonalen Matrizen*, und es gibt eine ganze Reihe von Möglichkeiten, diese zu charakterisieren; sie ergeben sich alle unmittelbar aus den Definitionen und aus Lemma 25.17:

26.2 Notiz Die folgenden Eigenschaften einer quadratischen Matrix $u \in \mathrm{Mat}(n \times n, \mathbb{R})$ sind gleichwertig:

- u ist orthogonal
- die Spalten von u bilden ein Orthonormalsystem in $\mathbb{R}^n$
- $u^t u = 1$ (die Einträge von $u^t u$ sind ja die Skalarprodukte je zweier Spalten von u)
- u ist invertierbar, und $u^{-1} = u^t$
- $u u^t = 1$
- die (transponierten) Zeilen von u bilden ein Orthonormalsystem
- u^t ist orthogonal

Welche Werte kann die Determinante einer orthogonalen Matrix u haben? Nun, wegen $u^t u = 1$ für $u \in O(n)$ gilt

$$(\det u)^2 = \det u^t \det u = \det 1 = 1,$$

also ist $\det u = \pm 1$.

26.3 Definition Die Untergruppe

$$SO(n) := \{u \in O(n) \mid \det u = 1\} = O(n) \cap SL(n, \mathbb{R})$$

nennt man spezielle orthogonale Gruppe.

Selbstverständlich bilden die orthogonalen Matrizen mit Determinante -1 keine Gruppe.

26.4 Beispiele (1) Welche reellen 2×2-Matrizen u gehören zu $SO(2)$? Mit

$$u = \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix} \in \mathrm{Mat}(2 \times 2, \mathbb{R})$$

lauten die Forderungen

(a) $\alpha\gamma + \beta\delta = 0$

(b) $\alpha^2 + \beta^2 = 1,\ \gamma^2 + \delta^2 = 1$

(c) $\alpha\delta - \beta\gamma = 1$

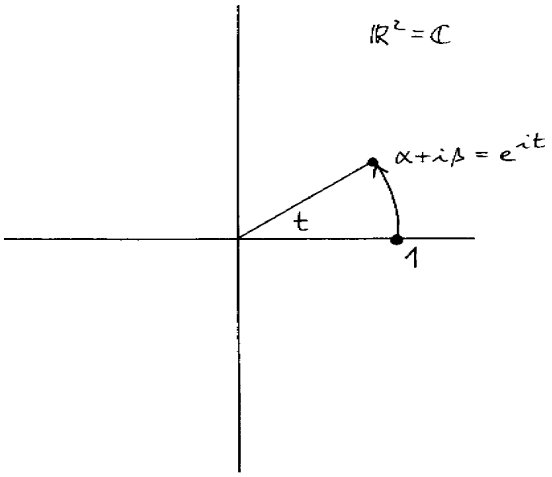
Die erste können wir auch so auffassen: die Spalte $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ und die (kopfgestellte!) Spalte $\begin{pmatrix} -\delta \\ \gamma \end{pmatrix}$ müssen linear abhängig sein. Da nach (b) beide die Länge 1 haben sollen, erlaubt das nur

$$u = \begin{pmatrix} \alpha & \beta \\ \beta & -\alpha \end{pmatrix} \quad \text{oder} \quad u = \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}.$$

Die Forderung (c) sondert die erste Möglichkeit aus, und die verbleibenden Matrizen

$$u = \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix} \in \mathrm{Mat}(2\times 2, \mathbb{R}) \quad \text{mit } \alpha^2 + \beta^2 = 1$$

sind tatsächlich speziell und orthogonal. Sie sind überdies alte Bekannte, denn wenn man die Multiplikation mit der komplexen Zahl $\alpha + i\beta$ als $\mathbb{R}$-lineare Abbildung von $\mathbb{C}$ nach $\mathbb{C}$ ansieht, ergibt sich bezüglich der kanonischen Basis $(1, i)$ nach Lemma 23.2 gerade die angegebene Matrix. Wie wir längst wissen, läßt diese komplexe Zahl vom Betrag 1 sich in der Form $\alpha + i\beta = e^{it}$ mit reellem t schreiben, und bei der Abbildung handelt es sich dann um die Drehung der Ebene um den Winkel t (bei festem Nullpunkt).



(2) Die Elemente von $O(2)\setminus SO(2)$, also die orthogonalen 2×2-Matrizen der Determinante -1, sind die vorhin ausgeschlossenen Matrizen

$$u = \begin{pmatrix} \alpha & \beta \\ \beta & -\alpha \end{pmatrix} \in \mathrm{Mat}(2\times 2, \mathbb{R}) \quad \text{mit } \alpha^2 + \beta^2 = 1$$

Zu ihnen gehört die spezielle Matrix

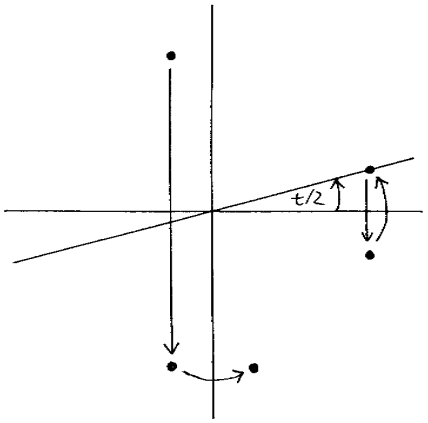
$$s := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \in O(2)\setminus SO(2),$$

die die Ebene an der ersten Koordinatenachse spiegelt. Vermöge

$$\begin{pmatrix} \alpha & \beta \\ \beta & -\alpha \end{pmatrix} = \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

schreibt sich jede Matrix aus $O(2)\setminus SO(2)$ als Komposition einer eindeutig bestimmten Drehung mit der Spiegelung s, und wie eine elementargeometrische Überlegung (oder die Berechnung der Eigenwerte und -räume) zeigt, handelt es sich bei der Komposition selbst um eine Spiegelung, und zwar an der Geraden mit

Steigungswinkel $t/2$, wenn $\alpha + i\beta = e^{it}$ ist: Man verfolge das Schicksal eines Vektors auf dieser Geraden und eines zweiten, dazu senkrechten:



Im zweiten Teil dieses Abschnitts wollen wir uns systematisch mit Mengen zueinander senkrechter Vektoren befassen. Sei A eine zunächst ganz beliebige Teilmenge eines euklidischen Vektorraums V. Wegen der Bilinearität des Skalarproduktes ist die Menge

$$A^\perp := \{v \in V \mid \langle a, v\rangle = 0 \text{ für alle } a \in A\}$$

erstens stets ein linearer Unterraum von V, zweitens nicht von A selbst abhängig, sondern nur von der linearen Hülle von A (der Menge aller aus Vektoren von A zu bildenden Linearkombinationen):

$$A^\perp = \mathrm{Lin}(A)^\perp \subset V.$$

Am interessantesten ist diese neue Bildung, wenn A selbst ein Unterraum von V und V endlichdimensional ist.

26.5 Definition V sei ein endlichdimensionaler euklidischer Vektorraum, $U \subset V$ ein linearer Unterraum. Dann heißt $U^\perp \subset V$ das orthogonale Komplement von U in V.

Dessen wichtigste Eigenschaften:

26.6 Satz und **Definition** V sei ein endlichdimensionaler euklidischer Vektorraum. Dann gilt:

(a) Für jeden linearen Unterraum $U \subset V$ ist $U^\perp$ in der Tat ein Komplement von U in V:

$$U \cap U^\perp = \{0\} \quad \text{und} \quad U + U^\perp = V,$$

insbesondere

$$\dim U + \dim U^\perp = \dim V.$$

(b) Für jeden Unterraum $U \subset V$ gilt

$$U^{\perp\perp} = U.$$

(c) Für je zwei Unterräume $S, T \subset V$ gilt:

$$(S \cap T)^\perp = S^\perp + T^\perp$$
$$(S + T)^\perp = S^\perp \cap T^\perp$$

(d) Zu jedem Unterraum $U \subset V$ existiert ein eindeutig bestimmter Endomorphismus

$$p_U \colon V \longrightarrow V$$

mit

$$p_U(v) = v \text{ für alle } v \in U, \text{ und } p_U|U^\perp = 0.$$

Dieser Endomorphismus heißt die senkrechte oder orthogonale Projektion von V auf U.

Beweis In (a) sei $v \in U \cap U^\perp$: dann ist $\langle v, v\rangle = 0$, also wegen der Definitheit $v = 0$. Damit ist $U \cap U^\perp = \{0\}$ gezeigt. Zum Beweis von $U + U^\perp = V$ wählen wir eine Orthonormalbasis $(u_1, \dots, u_r)$ von U und ergänzen zu einer Orthonormalbasis

$$(u_1, \dots, u_r, u_{r+1}, \dots, u_n)$$

von V, beides nach dem Orthonormalisierungssatz 25.13. Dann ist

$$u_1, \dots, u_r \in U,$$
$$u_{r+1}, \dots, u_n \in U^\perp,$$

und die Behauptung ist klar.

Die Teilaussage $U \subset U^{\perp\perp}$ von (b) folgt sofort aus der Definition. Nach (a) ist nun

$$\dim U^{\perp\perp} = \dim V - \dim U^\perp = \dim U,$$

also in Wirklichkeit $U = U^{\perp\perp}$.

Die zweite der unter (c) angegebenen Gleichungen ist klar: $(S+T)^\perp = S^\perp \cap T^\perp$. Auf sie läßt sich andererseits die erste mittels (b) zurückführen:

$$(S \cap T)^\perp = \left(S^{\perp\perp} \cap T^{\perp\perp}\right)^\perp = \left(S^\perp + T^\perp\right)^{\perp\perp} = S^\perp + T^\perp$$

Zur Konstruktion der orthogonalen Projektion schließlich wählen wir eine Orthonormalbasis $(u_1, \dots, u_n)$ wie im Beweis von (a) und definieren $p_U: V \longrightarrow V$ durch

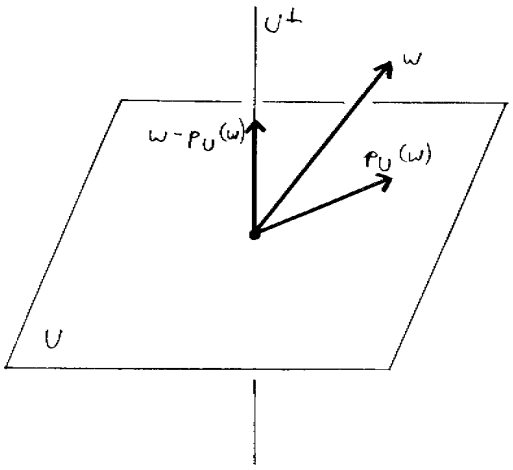
$$p_U(u_i) := \begin{cases} u_i & (i \le r), \\ 0 & (i > r). \end{cases}$$

Das tut's offenbar, und anders können wir es auch nicht machen.

Bemerkung Den Wert eines jeden Vektors $w \in V$ unter p_U — die *orthogonale Projektion* von w in U, wie man auch sagt — kann man direkt durch die Formel

$$p_U(w) = \langle u_1, w\rangle u_1 + \dots + \langle u_r, w\rangle u_r$$

ausdrücken. Diese Tatsache läßt das zum Beweis von 25.13 verwendete Gram-Schmidtsche Orthonormalisierungsverfahren jetzt besser verstehen: dort wird von dem nicht im Unterraum $U = \mathrm{Lin}(u_1, \dots, u_r)$ liegenden Vektor w seine orthogonale Projektion $p_U(w)$ abgezogen, so daß der Rest $w - p_U(w)$ auf U senkrecht steht.



Wenn man sagt, eine Summe $U_1 + \cdots + U_r$ von Unterräumen $U_1, \ldots, U_r$ eines euklidischen Vektorraums V sei orthogonal, meint man natürlich, daß für alle $i \neq j$ jeder Vektor aus U_i auf jedem Vektor von U_j senkrecht steht. Die Schlußfolgerung (a) des Satzes läßt sich damit auch so fassen: V ist orthogonale Summe der beiden Unterräume U und $U^\perp$. Allgemein gilt übrigens:

26.7 Lemma *$U_1, \ldots, U_r$ seien Unterräume eines euklidischen Vektorraums V. Ist die Summe $U_1 + \cdots + U_r$ orthogonal, so ist sie auch direkt.*

Beweis Seien $u_1 \in U_1, \ldots, u_r \in U_r$ Vektoren mit $0 = \sum_{j=1}^{r} u_j$, und sei $i \in \{1, \ldots, r\}$ ein fester Index. Wir bilden das Skalarprodukt mit u_i und erhalten

$$0 = \langle u_i, 0\rangle = \sum_{j=1}^{r} \langle u_i, u_j\rangle = \|u_i\|^2,$$

also $u_i = 0$. Da i beliebig war, folgt die Direktheit der Summe.

Zur Illustration der neuen Konzepte wollen wir noch die Elemente von $SO(3)$, also der orthogonalen 3×3-Matrizen der Determinante 1 untersuchen. Diese Analyse beruht auf dem

26.8 Lemma *Jedes $u \in SO(3)$ besitzt einen Fixvektor $a \in \mathbb{R}^3 \setminus \{0\}$.*

Beweis Das charakteristische Polynom χ_u zerfällt in $\mathbb{C}[X]$ in Linearfaktoren:

$$\chi_u(X) = (X - \lambda)(X - \mu)(X - \nu) \quad \text{mit } \lambda, \mu, \nu \in \mathbb{C},$$

und wir wissen insbesondere

$$\lambda \cdot \mu \cdot \nu = \det u = 1.$$

Nun sind zwei Fälle möglich:

λ, μ und ν seien reell: Ist dann a ein Eigenvektor etwa zu λ, so folgt aus der Orthogonalität von u und aus

$$|\lambda| \cdot \|a\| = \|\lambda a\| = \|ua\| = \|a\|,$$

daß für λ, und ebenso für μ und ν nur die Werte ± 1 in Frage kommen. Wegen $\lambda\mu\nu = 1$ ist mindestens einer der drei Eigenwerte $+1$, und jeder zugehörige Eigenvektor ist ein Fixvektor von u.

Die alternative Möglichkeit ist die, daß nur ein Linearfaktor reell, etwa $\lambda \in \mathbb{R}$ ist; dann sind die beiden anderen zwangläufig zueinander konjugiert: $\nu = \overline{\mu}$. In diesem Fall wird die Gleichung $\lambda\mu\nu = 1$ zu

$$\lambda \cdot |\mu|^2 = 1,$$

woraus man $\lambda > 0$ und weiter wie oben $\lambda = 1$ schließt. Damit ist das Lemma auch für diesen Fall bewiesen.

Mittels des Lemmas und der früheren Analyse von $SO(2)$ können wir jedes gegebene Element $u \in SO(3)$ im Prinzip schon völlig verstehen: Wenn wir gemäß dem Lemma einen Fixvektor $a \in \mathbb{R}^3 \setminus \{0\}$ wählen, bleibt natürlich die ganze Gerade

$$L := \mathrm{Lin}(a) \subset \mathbb{R}^3$$

unter u punktweise fest:

$$ux = x \quad \text{für alle } x \in L.$$

Weil u orthogonal ist, bildet es deshalb auch die zu L senkrechte Ebene $L^\perp = \{a\}^\perp$ von u in sich ab:

$$u(L^\perp) \subset L^\perp,$$

denn aus $y \in L^\perp$, d.h. $\langle a, y\rangle = 0$ folgt

$$\langle a, uy\rangle = \langle ua, uy\rangle = \langle a, y\rangle = 0.$$

Die Einschränkung $u|L^\perp: L^\perp \longrightarrow L^\perp$ ist offenbar selbst ein orthogonaler Endomorphismus und, wie man aus $\det u = 1$ sofort abliest, einer der Determinante $+1$:

$$u|L^\perp \in SO(L^\perp).$$

Insgesamt haben wir damit den Raum so in eine orthogonale Summe $\mathbb{R}^3 = L + L^\perp$ zerlegt, daß u auf dem Summanden L identisch, und auf der Ebene $L^\perp$ als Drehung wirkt; mit anderen Worten ist u selbst eine Drehung des Raumes um die Achse L.

Für konkrete Rechnungen würde man zweckmäßigerweise den Fixvektor a normieren und zu einer Orthonormalbasis von $\mathbb{R}^3$ ergänzen: die Matrix von u bezüglich dieser Basis hätte dann die Gestalt

$$\left(\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & \multicolumn{2}{c}{\multirow{2}{*}{u'}} \\ 0 & & \end{array}\right)$$

mit einer Matrix $u' \in SO(2)$.

Einige ergänzende Fakten machen Sie sich sofort klar: Die Drehachse L ist durch u eindeutig bestimmt, ausgenommen im Fall $u = 1$, wo natürlich jede Gerade die Rolle der Drehachse spielen kann. Ansonsten haben nur sehr spezielle weitere $u \in SO(3)$ drei reelle Eigenwerte, nämlich die Drehungen um $180°$. Und Sie werden sicher Vergnügen daran finden, analog zu Beispiel 26.4(2) die Bedeutung der Elemente von $O(3)\backslash SO(3)$ zu ergründen.

Übungsaufgaben

26.1. V sei ein endlichdimensionaler euklidischer Vektorraum, $U \subset V$ ein linearer Unterraum.

(a) Wenn man eine Basis wie im Beweis von Satz 26.6(a) wählt, wie sieht dann die Matrix von p_U aus?

(b) Begründen Sie die Identität $p_U^2 = p_U$ und zeigen Sie, daß p_U die Eigenschaft

$$\big\langle p_U(v), w\big\rangle = \big\langle v, p_U(w)\big\rangle \text{ für alle } v,\ w \in V$$

hat — *selbstadjungiert* ist, wie wir bald sagen werden.

(c) Zeigen Sie, daß es umgekehrt zu jedem selbstadjungierten Endomorphismus $p: V \longrightarrow V$ mit $p^2 = p$ (genau) einen Unterraum U mit $p = p_U$ gibt. (Offenbar muß man erst mal auf einen Kandidaten für U kommen: wenn man sich dazu vorstellt, das Problem sei schon gelöst, sieht man aber ganz leicht, wie man U aus p_U zurückgewinnt.)

26.2 Jetzt betrachten wir zwei Unterräume S und T eines endlichdimensionalen euklidischen Vektorraums V. Zeigen Sie:

(a) $$S \subset T \iff p_S \circ p_T = p_S \iff p_T \circ p_S = p_S$$

(b) Sind p_S und p_T miteinander vertauschbar, so ist $p_S \circ p_T = p_{S\cap T}$ die orthogonale Projektion auf $S \cap T$.

Tips: Beim Beweis von (a) sind vielleicht (a) und (b) der vorigen Aufgabe nützlich; zum Beweis von (b) können Sie Teil (c) heranziehen.

26.3 Bestimmen Sie die Drehachse und den Drehwinkel der Matrix

$$u = \frac{1}{3}\begin{pmatrix} 2 & 1 & -2 \\ -2 & 2 & -1 \\ 1 & 2 & 2 \end{pmatrix} \in SO(3).$$

26.4 A sei ein nicht-leerer affiner Unterraum des endlichdimensionalen euklidischen Vektorraums V, und $b \in V$ ein Punkt. Beweisen Sie: Es gibt genau einen Punkt $a \in A$, so daß $a - b$ auf A (genauer auf dem zu A parallelen linearen Unterraum) senkrecht steht; unter allen Punkten von A ist a derjenige, der den kleinsten Abstand von b hat:

$$\|a - b\| < \|x - b\| \quad \text{für alle } x \in A \backslash \{a\}.$$

26.5 Seien allgemeiner A und B zwei nicht-leere affine Unterräume des endlichdimensionalen euklidischen Vektorraums V. Beweisen Sie: Es gibt Punkte $a \in A$ und $b \in B$, so daß $a - b$ auf A und auf B senkrecht steht, und genau für diese Punktepaare (a, b) gilt

$$\|a - b\| \leq \|x - y\| \quad \text{für alle } x \in A,\ y \in B.$$

Wie viele solcher Paare gibt es?

27 Dualraum und Skalarprodukt

Der erste Teil dieses Abschnitts hat nicht mit euklidischen Strukturen zu tun, er ist vielmehr ein Nachtrag zu der in den Abschnitten 17 bis 21 behandelten Theorie der Vektorräume und linearen Abbildungen.

27.1 Definition V sei ein K-Vektorraum. Der K-Vektorraum

$$V^\vee := \mathrm{Hom}(V,K) = \{f\colon V \longrightarrow K \mid f \text{ linear}\}$$

heißt der Dualraum von V; seine Elemente nennt man gern Linearformen auf V oder auch, vor allem wenn V ein (unendlichdimensionaler) Raum von Funktionen ist, lineare Funktionale auf V.

Die Vektorraumstruktur des Dualraums ergibt sich natürlich einfach aus der von K als (eindimensionalem) Vektorraum über sich selbst; es werden ja die Werte der Linearformen addiert bzw. mit einem Skalar multipliziert. Ist $\underline{u} = (u_1,\dots,u_n)$ eine Basis von V, so werden gemäß den getroffenen Konventionen die Linearformen auf V durch Zeilen $z \in \mathrm{Mat}(1\times n,K)$ dargestellt. In K als K-Vektorraum verwendet man dabei naheliegenderweise die Standardbasis, so daß sich die $f \in V^\vee$ repräsentierende Zeile aus dem kommutativen Diagramm

$$\begin{array}{ccc} V & \xrightarrow{\ f\ } & K \\ {\scriptstyle \Phi_{\underline{u}}}\uparrow\ {\scriptstyle\simeq} & & \| \\ K^n & \xrightarrow{\ z\ } & K \end{array}$$

zu

$$z = \begin{pmatrix} z_1 & \dots & z_n \end{pmatrix} = \begin{pmatrix} f(u_1) & \dots & f(u_n) \end{pmatrix}$$

ergibt.

Manchmal möchte man $V^\vee$ aber auch als einen eigenständigen Vektorraum ansehen, ohne sich unbedingt daran zu erinnern, daß seine Elemente die Linearformen auf einem anderen Vektorraum sind: zum expliziten Rechnen in $V^\vee$ wird man dann eine Basis von $V^\vee$ wählen und jeden Vektor von $V^\vee$ bezüglich dieser Basis durch die zugehörige Spalte in K^n ausdrücken wollen. Die folgende Definition beschreibt eine naheliegende und besonders einfache Wahl einer solchen Basis.

27.2 Definition $\underline{u} = (u_1,\dots,u_n)$ sei eine Basis des K-Vektorraums V. Das durch

$$\check{u}_i\,(u_j) := \delta_{ij} \in K$$

festgelegte n-tupel $\underline{\check{u}} = (\check{u}_1,\dots,\check{u}_n)$ ist eine Basis von V; man nennt $\underline{\check{u}}$ die zu $\underline{u}$ duale Basis.

Daß es sich bei $\underline{\check{u}}$ tatsächlich um eine Basis handelt, folgt zum Beispiel sofort aus dem im folgenden häufiger und stillschweigend benutzten Satz 19.6: jede Linearform $V \longrightarrow K$ ist durch ihre Werte auf den Basisvektoren $u_1,\dots,u_n$ festgelegt, und diese kann man umgekehrt beliebig vorschreiben und dadurch eine Linearform definieren. — Die zur Standardbasis $(e_1,\dots,e_n)$ von K^n duale Basis ist natürlich die aus den transponierten Vektoren gebildete, also $(e_1^{\mathrm t},\dots,e_n^{\mathrm t})$. Glauben Sie aber deshalb nicht, das sei bei jeder Basis von K^n so. — Das Besondere der (zu einer gegebenen Basis von V) dualen Basis zeigt sich in der

27.3 Notiz Sei $z \in \mathrm{Mat}(1\times n,K)$ die Matrix der Linearform $f\colon V \longrightarrow K$ bezüglich der Basis $\underline{u}$. Dann ist $z^{\mathrm t} \in \mathrm{Mat}(n\times 1,K)$ die Matrix von $f \in V^\vee$ bezüglich der dualen Basis $\underline{\check{u}}$.

Beweis Wie schon bemerkt gilt $z = \begin{pmatrix} f(u_1) & \dots & f(u_n) \end{pmatrix}$; andererseits ist $f = \sum_{j=1}^{n} f(u_j)\,\check{u}_j$ die Darstellung von f in der dualen Basis.

Nicht nur Vektorräume, sondern auch lineare Abbildungen kann man dualisieren:

27.4 Definition und **Lemma** *V und W seien K-Vektorräume; $f\colon V \longrightarrow W$ sei linear. Die durch*

$$W^{\check{}} = \operatorname{Hom}(W,k) \ni \psi \longmapsto \psi \circ f \in \operatorname{Hom}(V,K) = V^{\check{}}$$

definierte lineare Abbildung $f^{\check{}}\colon W^{\check{}} \longrightarrow V^{\check{}}$ heißt die zu f duale lineare Abbildung. Es gelten die Regeln

$$\mathrm{id}^{\check{}} = \mathrm{id} \quad \text{und} \quad (g \circ f)^{\check{}} = f^{\check{}} \circ g^{\check{}}.$$

Ist im endlichdimensionalen Fall $a \in \operatorname{Mat}(p \times n, K)$ die Matrix von f bezüglich Basen $\underline{v}$ von V und $\underline{w}$ von W, so ist $a^{\mathrm{t}} \in \operatorname{Mat}(n \times p, K)$ die Matrix von $f^{\check{}}$ bezüglich der dualen Bases $\underline{\check{w}}$ und $\underline{\check{v}}$.

Beweis Die a definierende Identität

$$f(v_k) = \sum_{j=1}^{p} a_{jk} w_j$$

wird durch Anwenden von $\check{w}_i$ zu

$$f^{\check{}}(\check{w}_i)(v_k) = (\check{w}_i \circ f)(v_k) = \sum_{j=1}^{p} a_{jk}\, \check{w}_i(w_j) = a_{ik}$$

und weiter zu

$$f^{\check{}}(\check{w}_i) = \sum_{j=1}^{n} a_{ij} \check{v}_j.$$

Bemerkungen Damit ist endlich mein Versprechen eingelöst, die begriffliche Bedeutung des Transponierens zu erklären. — Am Anfang tut man sich vielleicht schwer, den Unterschied zwischen einem Vektorraum V und seinem Dualraum $V^{\check{}}$ einzusehen. Macht man nicht zumindest im endlichdimensionalen Fall aus einer Spalte durch Transponieren flugs eine Zeile und damit aus einem Vektor eine Linearform? Gewiß erhält man so einen Isomorphismus zwischen V und $V^{\check{}}$, aber der Haken ist, daß man für diese Zuordnung erst eine Basis von V wählen muß und der entstehende Isomorphismus von der Wahl dieser Basis abhängt. Letztlich liefert diese Idee also bloß die Erkenntnis, daß für endlichdimensionales V überhaupt Isomorphismen zwischen V und $V^{\check{}}$ existieren: das wissen wir aber sowieso, weil die Dimensionen gleich sind. Was man in Abwesenheit weiterer Strukturen dagegen nicht hat, ist ein *kanonischer*, d.h. ein von willkürlichen Wahlen unabhängiger Isomorphismus. Anders sieht es beim Vergleich des Vektorraums V mit seinem Bidualraum $V^{\check{}\check{}} := (V^{\check{}})^{\check{}}$ aus:

27.5 Lemma *Wenn V endliche Dimension hat, dann ist die lineare Abbildung*

$$V \ni v \longmapsto (V^{\check{}} \ni \varphi \mapsto \varphi(v) \in K) \in V^{\check{}\check{}}$$

ein Isomorphismus von V auf seinen Bidualraum.

Beweis Wenn $v \in V$ auf die Nullform $0\colon V^{\check{}} \longrightarrow K$ abgebildet wird, ist $\varphi(v) = 0$ für jedes $\varphi \in V^{\check{}}$, und daraus folgt leicht $v = 0$. Die beschriebene lineare Abbildung ist also injektiv; wegen $\dim V = \dim V^{\check{}\check{}}$ ist sie sogar bijektiv.

Hier darf man, wenn man will, sogar so weit gehen, den Bidualraum eines endlichdimensionalen Vektorraums V mit diesem selbst zu identifizieren: jeder Vektor in V entspricht ja in der im Lemma beschriebenen und von jeder willkürlichen Wahl unabhängigen Weise einem Vektor in $V^{\check{}\check{}}$. Ist $V \xrightarrow{f} W$ ein Homomorphismus,

so wird man dann auch die zu f biduale Abbildung $V^{\check{}\check{}} \xrightarrow{f^{\check{}\check{}}} W^{\check{}\check{}}$ mit f selbst identifizieren wollen. Die Kommutativität (!) des Diagramms

$$\begin{array}{ccc} V & \xrightarrow{\quad f \quad} & W \\ \simeq \downarrow & & \downarrow \simeq \\ V^{\check{}\check{}} & \xrightarrow{\quad f^{\check{}\check{}} \quad} & W^{\check{}\check{}} \end{array}$$

zeigt, daß das auch erlaubt ist.

27.6 Definition V sei ein K-Vektorraum, $A \subset V$ und $\Phi \subset V^{\check{}}$ seien beliebige Teilmengen. Man nennt

$$\operatorname{Ann}^{\check{}} A := \{\varphi \in V^{\check{}} \mid \varphi(v) = 0 \text{ für alle } v \in A\}$$

und

$$\operatorname{Ann} \Phi := \{v \in V \mid \varphi(v) = 0 \text{ für alle } \varphi \in \Phi\} = \bigcap_{\varphi \in \Phi} \operatorname{Kern} \varphi$$

den Annihilator von A bzw. von Φ.

Beide Versionen dieses Begriffs sind dem des orthogonalen Komplementes (in einem euklidischen Vektorraum) sehr ähnlich, nur daß der Annihilator in dem jeweils anderen der Räume V und $V^{\check{}}$ liegt. Ich begnüge mich deshalb hier damit, die wichtigsten und zu Satz 26.6 analogen Eigenschaften des Annihilators aufzuzählen; dabei dürfen Sie jede der folgenden Aussagen noch um die duale ergänzen.

27.7 Lemma V sei ein endlichdimensionaler Vektorraum.

(a) Für jedes $A \subset V$ ist $\operatorname{Ann}^{\check{}} A \subset V^{\check{}}$ ein linearer Unterraum; dieser hängt andererseits nur von der linearen Hülle von A ab:

$$\operatorname{Ann}^{\check{}} A = \operatorname{Ann}^{\check{}} \operatorname{Lin}(A)$$

(b) Für jeden Unterraum $U \subset V$ gilt

$$\dim U + \dim \operatorname{Ann}^{\check{}} U = \dim V$$

sowie

$$\operatorname{Ann} \operatorname{Ann}^{\check{}} U = U.$$

(c) Für je zwei Unterräume $S, T \subset V$ gilt:

$$\operatorname{Ann}^{\check{}} (S \cap T) = \operatorname{Ann}^{\check{}} S + \operatorname{Ann}^{\check{}} T$$
$$\operatorname{Ann}^{\check{}} (S + T) = \operatorname{Ann}^{\check{}} S \cap \operatorname{Ann}^{\check{}} T$$

Aus dem Matrizenkalkül ist uns vertraut, daß das Transponieren einer Matrix ihren Rang nicht ändert, was wir jetzt als die

27.8 Notiz $\operatorname{rk} f = \operatorname{rk} f^{\check{}}$

interpretieren können. Dahinter steckt genauer die folgende Beziehung zwischen den vier Unterräumen Kern und Bild von f und $f^{\check{}}$:

27.9 Lemma $f: V \longrightarrow W$ sei eine lineare Abbildung zwischen endlichdimensionalen K-Vektorräumen. Dann gilt:

$$\operatorname{Bild} f^{\check{}} = \operatorname{Ann}^{\check{}} \operatorname{Kern} f$$
$$\operatorname{Kern} f^{\check{}} = \operatorname{Ann}^{\check{}} \operatorname{Bild} f$$

Beweis Seien $v \in \operatorname{Kern} f$ und $\varphi \in \operatorname{Bild} f^\check{}$, etwa $\varphi = f^\check{}(\psi)$. Dann ist

$$\varphi(v) = \big(f^\check{}(\psi)\big)(v) = (\psi \circ f)(v) = \psi\big(f(v)\big) = 0.$$

Also ist schon mal $\operatorname{Bild} f^\check{} \subset \operatorname{Ann}^\check{} \operatorname{Kern} f$. Wegen

$$\dim \operatorname{Ann}^\check{} \operatorname{Kern} f = \dim V - \dim \operatorname{Kern} f = \operatorname{rk} f = \dim \operatorname{Bild} f$$

folgt daraus die erste behauptete Identität. Die andere ergibt sich analog, oder auch, indem man die schon bewiesene Gleichung für $f^\check{}$ statt f liest und beidseitig den Annihilator bildet.

Wir wollen die neu eingeführten Terminologie jetzt auf etwas ganz Konkretes anwenden. Die Aufgabe, das durch die Matrix $a \in \operatorname{Mat}(p{\times}n, K)$ gegebene homogene lineare Gleichungssystem $ax = 0$ für $x \in K^n$ zu lösen, also den Kern von a zu berechnen, können wir auch so ausdrücken: Die p Zeilen $a_1, \dots, a_p \in \operatorname{Mat}(1{\times}n, K)$ der Matrix a sind Linearformen auf K^n, also Vektoren des Dualraums $(K^n)^\check{}$, und berechnet werden soll

$$\operatorname{Kern} a = \operatorname{Ann}\{a_1, \dots, a_p\} \subset K^n.$$

Der Vorteil dieser Formulierung: Sie läßt sofort erkennen, daß die umgekehrte Frage, nämlich zu einem durch aufspannende Vektoren gegebenen Unterraum $U = \operatorname{Lin}(b_1, \dots, b_p) \subset K^n$ ein Gleichungssystem zu konstruieren, das diesen als Lösungsraum hat, von genau der gleichen Art ist. Zu berechnen ist hier nämlich eine Basis (oder ein Erzeugendensystem) des Annihilators

$$\operatorname{Ann}^\check{} U = \operatorname{Ann}^\check{} \{b_1, \dots, b_p\} \subset (K^n)^\check{} = \operatorname{Mat}(1{\times}n, K)\,.$$

Eine solche Basis besteht aus Linearformen auf K^n, und diese entsprechen den Gleichungen eines homogenen Systems, das gerade $U = \operatorname{Ann} \operatorname{Ann}^\check{} U$ als Lösungsraum hat.

Wenn wir in diesem Problem die Spalten b_j wie üblich zur Matrix

$$b = \begin{pmatrix} b_1 & \dots & b_p \end{pmatrix} \in \operatorname{Mat}(n{\times}p, K)$$

zusammenfassen, wird

$$\operatorname{Ann}^\check{} U = \{z \in \operatorname{Mat}(1{\times}n) \mid zb = 0\}\,,$$

und es wäre konsequent, das Gleichungssystem $zb = 0$ für z durch den Gaußschen Algorithmus in der Spaltenversion lösen. Weil man aber so daran gewöhnt ist, daß in einem Gleichungssystem die Unbekannte hinten steht, rechnet man üblicherweise in der dualen Basis von $(K^n)^\check{}$, schreibt das System also zu $b^{\mathrm{t}} z^{\mathrm{t}} = 0$ um.

27.10 Beispiel Wir suchen Gleichungen, die den Unterraum

$$U = \operatorname{Lin}\left(\begin{pmatrix}1\\2\\3\end{pmatrix}\ \begin{pmatrix}4\\5\\6\end{pmatrix}\right) \subset \mathbb{R}^3$$

beschreiben. In den obigen Bezeichungen ist

$$b = \begin{pmatrix}1 & 4\\2 & 5\\3 & 6\end{pmatrix},$$

und Lösen der Gleichung $b^{\mathrm{t}} z^{\mathrm{t}} = 0$ für z nach dem Gaußschen Algorithmus

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

gibt

$$z^{\mathrm{t}} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \in \mathrm{Mat}(3\times 1, \mathbb{R}).$$

Also spannt die Linearform

$$z = \begin{pmatrix} 1 & -2 & 1 \end{pmatrix} \in \mathrm{Mat}(1\times 3, \mathbb{R})$$

den Raum $\mathrm{Ann}^{\check{}}\, U$ auf, und

$$U = \{x \in \mathbb{R}^3 \,|\, x_1 - 2x_2 + x_3 = 0\}$$

ist eine Darstellung von U durch ein Gleichungssytem.

Aber ist es nicht idiotisch, zu einer vorgegebenen Lösung nach einem passenden Problem zu suchen? Vielleicht, aber das ist nicht die einzige mögliche Interpretation dessen, was wir jetzt gemacht haben. Zur Beschreibung eines Unterraums U etwa von K^n gibt es ja zwei grundsätzlich gleichwertige Möglichkeiten: Einmal die bisher favorisierte durch eine Basis (oder zumindest ein Tupel aufspannender Vektoren) von U, aber eben auch die durch ein endliches System von homogenen Gleichungen. Letztere ist zum Beispiel sicher vorzuziehen, wenn $\dim U = 1000$ und $n = 1001$ ist, weil man dann statt mit 1000 Basisvektoren mit einer einzigen Gleichung auskommt (die zudem noch bis auf einen skalaren Faktor eindeutig bestimmt ist). Außerdem hat die Beschreibung durch Gleichungen immer dann Vorteile, wenn der Durchschnitt zweier Unterräume U und U' berechnet werden soll, denn dazu braucht man offensichtlich nur die Gleichungen für U und U' zusammenzuwerfen.

Jedenfalls wird man in der Praxis gelegentlich Beschreibungen der beiden Arten ineinander umzurechnen. Nun, aus Gleichungen für U eine Basis von U zu konstruieren, ist die klassische Aufgabe, ein lineares Gleichungssystem zu lösen, und die Umkehrung, die also doch auch einen praktischen Sinn hat, habe ich Ihnen gerade vorgeführt.

Entsprechendes gilt für die Beschreibung affiner Unterräume durch einen willkürlichen Punkt und eine Basis des parallelen linearen Teilraums einerseits und durch ein inhomogenes Gleichungssystems andererseits. Die Einzelheiten kann ich wohl Ihnen überlassen.

Nun zurück zu den euklidischen Vektorräumen. In ihnen vereinfacht sich der Umgang mit dem Dualraum dadurch, daß das Skalarprodukt eine kanonische Abbildung in den Dualraum definiert:

27.11 Satz *V sei ein endlichdimensionaler euklidischer Vektorraum. Dann ist die Abbildung*

$$V \xrightarrow{\Sigma} V^{\check{}}; \quad v \longmapsto \big(w \mapsto \langle v, w\rangle\big)$$

ein Isomorphismus von Vektorräumen. Ist das Skalarprodukt auf V bezüglich der Basis $\underline{u}$ durch die Matrix $s \in \mathrm{Sym}(n, \mathbb{R})$ repräsentiert, so ist s auch die Matrix von Σ bezüglich der Basen $\underline{u}$ und $\check{\underline{u}}$.

Beweis Weil das Skalarprodukt in der zweiten Variablen linear ist, definiert $w \mapsto \langle v, w\rangle$ wirklich eine Linearform auf V. Auf der Linearität in der ersten Variablen dagegen beruht es, daß die Abbildung Σ linear ausfällt:

$$\Sigma(v + v')(w) = \langle v + v', w\rangle = \langle v, w\rangle + \langle v', w\rangle = \Sigma(v)(w) + \Sigma(v')(w),$$

d.h.

$$\Sigma(v + v') = \Sigma(v) + \Sigma(v'),$$

und entsprechend für die skalare Multiplikation.

Die Injektivität von Σ folgt sofort aus der Definitheit des Skalarproduktes: $\Sigma(v) = 0$ bedeutet ja $\langle v, w\rangle = 0$ für *alle* $w \in V$, insbesondere für $w = v$; deshalb ist dann $v = 0$. Wir wissen schon, daß $\dim V = \dim V^{\check{}}$ ist, also ist Σ ein Isomorphismus wie behauptet.

Wir bestimmen noch die Matrixdarstellung von Σ. Für beliebige $j, k \in \{1, \dots, n\}$ hat die Linearform $\sum_{j=1}^n s_{jk}\check{u}_j$ auf u_i den Wert

$$\sum_{j=1}^n s_{jk}\check{u}_j(u_i) = s_{ik},$$

die Form $\Sigma(u_k)$ aber auch: $\Sigma(u_k)(u_i) = \langle u_k, u_i\rangle = \langle u_i, u_k\rangle = s_{ik}$. Folglich ist

$$\Sigma(u_k) = \sum_{j=1}^{n} s_{jk}\check{u}_j$$

die Darstellung von $\Sigma(u_k)$ in den angegebenen Basen und damit s die Matrix der Abbildung Σ.

Sie sehen, daß der Isomorphismus Σ von dem zu V gehörigen Skalarprodukt abhängt, was nach den Bemerkungen zur Definition 27.4 ja auch nicht anders zu erwarten war. Sie sehen als Nebenprodukt auch, daß die symmetrische Matrix eines Skalarproduktes immer invertierbar ist.

Wenn V ein endlichdimensionaler euklidischer Vektorraum ist, erlaubt es der Isomorphismus $V \xrightarrow{\Sigma} V^\check{}$, alles im Zusammenhang mit dem Dualraum Gesagte neu zu formulieren, ohne letzteren explizit zu erwähnen. Zuerst und besonders wichtig:

27.12 Definition und **Notiz** V und W seien endlichdimensionale euklidische Vektorräume, $f\colon V \longrightarrow W$ eine lineare Abbildung. Die durch das kommutative Diagramm

$$\begin{array}{ccc} V & \xleftarrow{\ f^*\ } & W \\ {\scriptstyle\Sigma}\downarrow{\scriptstyle\simeq} & & {\scriptstyle\Sigma}\downarrow{\scriptstyle\simeq} \\ V^\check{} & \xleftarrow{\ f^\check{}\ } & W^\check{} \end{array}$$

definierte lineare Abbildung

$$f^*\colon W \longrightarrow V$$

heißt die zu f adjungierte lineare Abbildung. Sie ist durch die Eigenschaft

$$\big\langle f(v), w\big\rangle = \big\langle v, f^*(w)\big\rangle \text{ für alle } v \in V,\ w \in W$$

charakterisiert.

Beweis Wenn man das Schicksal eines Vektors $w \in W$ im Diagramm

$$\begin{array}{ccccc} f^*(w) & \longleftarrow\!\shortmid & w & & w \\ \top\downarrow & & & & \top\downarrow \\ \big(v \mapsto \langle f^*(w), v\rangle\big) & & \big(v \mapsto \langle w, f(v)\rangle\big) & \longleftarrow\!\shortmid & \big(x \mapsto \langle w, x\rangle\big) \end{array}$$

verfolgt, erkennt man, daß die als Charakterisierung angegebene Gleichung gerade dessen Kommutativität zum Inhalt hat.

Sind $\underline{v}$ und $\underline{w}$ Basen von V und W, so werden die Skalarprodukte auf diesen euklidischen Räumen durch symmetrische Matrizen $s \in \mathrm{Sym}(n, \mathbb{R})$ und $t \in \mathrm{Sym}(p, \mathbb{R})$ beschrieben. Sei $a \in \mathrm{Mat}(p\times n, \mathbb{R})$ die Matrix von f bezüglich dieser Basen. Nach 27.4, 27.11 und 27.12 ist klar, daß

$$s^{-1}a^{\mathrm{t}}t \in \mathrm{Mat}(n\times p, \mathbb{R})$$

dann die Matrix von f^* bezüglich $\underline{w}$ und $\underline{v}$ sein muß. Wie immer besonders einfach wird's, wenn $\underline{v}$ und $\underline{w}$ Orthonormalbasen sind: Dann sind s und t Einheitsmatrizen, und die Matrix der Adjungierten ist einfach die Transponierte a^{t}.

Der Annihilator entspricht im euklidischen Fall dem uns schon vertrauten orthogonalen Komplement: für $A \subset V$ gilt nämlich

$$A^\perp = \{v \in V \mid \langle v, w\rangle = 0 \text{ für alle } w \in A\} = \{v \in V \mid \Sigma(v)(w) = 0 \text{ für alle } w \in A\} = \Sigma^{-1}(\check{\text{Ann}}\, A).$$

Es sei Ihnen überlassen, ob Sie die folgenden Eigenschaften der adjungierten Abbildung durch Uminterpretation aus 27.4 bis 27.8 gewinnen oder lieber direkt nachrechnen wollen:

27.13 Regeln für die Adjungierte Es gilt

- $\mathrm{id}^* = \mathrm{id}$ und $(g \circ f)^* = f^* \circ g^*$,
- $f^{**} = f$ sowie
- $\operatorname{rk} f = \operatorname{rk} f^*$.

Außerdem ist

$$\operatorname{Bild} f^* = (\operatorname{Kern} f)^\perp$$
$$\operatorname{Kern} f^* = (\operatorname{Bild} f)^\perp,$$

das heißt, daß

$$V = \operatorname{Kern} f + \operatorname{Bild} f^*$$
$$W = \operatorname{Kern} f^* + \operatorname{Bild} f$$

Zerlegungen von V und W als orthogonale Summen sind.

Zum Schluß dieses Abschnitts wollen wir darüber nachdenken, wie sich die Idee der euklidischen Vektorräume wohl auf komplexe Vektorräume übertragen lassen könnte — weil man reelle Probleme ja oft auf dem Umweg über das Komplexe löst, ist das sicher mehr als nur eine akademische Übung.

Sei also V ein komplexer Vektorraum. Keine Probleme macht der Begriff einer symmetrischen Bilinearform, dessen Definition (25.1) ohnehin für beliebige Vektorräume gilt. Aber die Definitheitsforderung

$$\langle v, v\rangle > 0 \text{ für } 0 \neq v \in V$$

stellt eine echte Schwierigkeit dar, weil $\langle v, v\rangle$ gar keine reelle Zahl zu sein braucht. Das einfach zusätzlich zu verlangen, führt wegen

$$\langle iv, iv\rangle = -\langle v, v\rangle$$

auch nicht weiter. Zum Ziel kommt man aber, wenn man die Forderung der Bilinearität raffiniert abändert, und das geschieht in der folgenden

27.14 Definition V sei ein komplexer Vektorraum. Eine hermitesche Form auf V ist eine Abbildung

$$V \times V \longrightarrow \mathbb{C}; \quad (v, w) \mapsto \langle v, w\rangle$$

mit den folgenden Eigenschaften:

(a) Für jedes feste $v \in V$ ist die Funktion $V \ni y \mapsto \langle v, y\rangle \in \mathbb{C}$ linear, für jedes feste $w \in V$ die Funktion $V \ni x \mapsto \langle x, w\rangle \in \mathbb{C}$ dagegen konjugiert-linear:

$$\langle x + x', w\rangle = \langle x, w\rangle + \langle x', w\rangle$$
$$\langle \lambda x, w\rangle = \overline{\lambda}\langle x, w\rangle$$

für alle $x, x' \in V$ und alle $\lambda \in \mathbb{C}$.

(b) $\langle v, w\rangle = \overline{\langle w, v\rangle}$ für alle $v, w \in V$

Eine solche hermitesche Form nennt man ein (hermitesches) Skalarprodukt auf V, wenn sie außerdem

(c) $\langle v, v\rangle > 0$ für alle $v \in V \setminus \{0\}$

erfüllt. Ein mit einem Skalarprodukt ausgestatteter $\mathbb{C}$-Vektorraum heißt ein unitärer Vektorraum.

Bemerkungen In der Literatur herrscht keine Einigkeit darüber, ob hermitesche Formen nun in der ersten oder der zweiten Variablen konjugiert-linear sein sollen (einfach mangels gewichtiger Gründe, sich für die eine oder andere Variante zu entscheiden). Wichtig ist nur, daß man bei der einmal getroffenen Entscheidung bleibt, was ich natürlich tun werde. Beachten Sie übrigens, daß "konjugiert-linear" nicht weniger ist als "linear", sondern eben anders. Deshalb sollte man in (a) die konjugierte Linearität nicht Semilinearität und die Eigenschaft (a) selbst nicht Sesquilinearität (anderthalbfache Linearität) nennen, wie es manche tun.

Nach (a) ist es nur konsequent, auch die Symmetrieforderung gemäß (b) abzuändern (tatsächlich macht (b) eine der Forderungen in (a) überflüssig).

Erst in (c) zeigt sich der ganze Witz dieser Modifikationen: Für jedes $v \in V$ gilt nach (b)

$$\langle v, v\rangle = \overline{\langle v, v\rangle},$$

also ist $\langle v, v\rangle$ automatisch eine reelle Zahl, und die Definitheitsforderung kann deshalb unverändert aus dem euklidischen Fall übernommen werden.

Wie zu erwarten, ähnelt die Theorie der unitären Vektorräume stark der der euklidischen; was das bisher Besprochene betrifft, begnüge ich mich deshalb mit einigen Hinweisen vor allem zu dem, was doch anders ist. Da ist zunächst die Beschreibung hermitescher Produkte bezüglich einer Basis $\underline{u} = (u_1, \dots, u_n)$ von V. So wie symmetrische Bilinearformen symmetrischen Matrizen entsprechen, entsprechen hermitesche Formen auf V sogenannten hermiteschen Matrizen, nämlich Matrizen $s \in \mathrm{Mat}(n\times n, \mathbb{C})$ mit

$$\overline{s}^{\mathrm{t}} = s.$$

Und zwar gehört zur hermiteschen Matrix s die durch

$$\langle u_j, u_k\rangle = s_{jk},$$

also

$$\big\langle \Phi_{\underline{u}}(x), \Phi_{\underline{u}}(y)\big\rangle = \overline{x}^{\mathrm{t}} s y$$

festgelegte hermitesche Form. Diese ist nicht in jedem Fall ein hermitesches Skalarprodukt auf V, sondern nur dann, wenn die Matrix s auch positiv definit ist, also $\overline{x}^{\mathrm{t}} s x > 0$ für alle komplexen Spalten $x \neq 0$ gilt.

Die Norm eines Vektors v aus einem unitären Raum ist wie im euklidischen Fall durch

$$\|v\| = \sqrt{\langle v, v\rangle} \in [0, \infty)$$

erklärt, und sie hat die in 25.7 aufgezählten Eigenschaften; insbesondere genügt sie der Dreiecksungleichung. Dagegen werden Winkel zwischen zwei Vektoren in einem unitären Raum nicht definiert, auch mangels geometrischen Interesses. Jedoch bleibt als Spezialfall der Begriff der Orthogonalität zweier Vektoren erhalten, und er ist nach wie vor symmetrisch:

$$\langle v, w\rangle = 0 \iff \langle w, v\rangle = 0$$

Deshalb kann man auch in einem unitären Vektorraum von Orthonormalsystemen und -basen reden, und der wichtige Orthnormalisierungssatz 25.13 bleibt uns erhalten, samt dem als Beweis fungierenden Rechenverfahren nach Gram-Schmidt.

Das naheliegende Standardmodell eines n-dimensionalen unitären Vektorraums ist $\mathbb{C}^n$ mit dem durch

$$\langle x, y\rangle = \overline{x}^{\mathrm{t}} y$$

gegebenen hermiteschen Standardprodukt. Der Orthonormalisierungssatz hat als wichtige Konsequenz, daß jeder n-dimensionale unitäre Vektorraum zu diesem Standardraum isometrisch (d.h. unter Erhalt des hermiteschen Produkts) isomorph ist.

Die isometrischen Automorphismen eines unitären Vektorraumes V nennt man nicht wie im euklidischen Fall orthogonale, sondern üblicherweise unitäre Abbildungen von V. Deshalb spricht man auch von der unitären Gruppe

$$U(V) = \big\{ f \in GL(V) \,\big|\, \big\langle f(v), f(w) \big\rangle = \langle v, w \rangle \quad \text{für alle } v, w \in V \big\}$$

von V. Die Bezeichnungen $U(n)$ und

$$SU(n) := U(n) \cap SL(n, \mathbb{C})$$

(spezielle unitäre Gruppe) beziehen sich wieder auf den Fall, daß $V = \mathbb{C}^n$ der Standardraum ist. Beachten Sie, daß der Schritt von $U(n)$ zu der Untergruppe $SU(n)$ "größer" ist als der von $O(n)$ nach $SO(n)$, denn für eine unitäre Matrix $u \in U(n)$ folgt aus

$$1 = \det 1 = \det(\overline{u}^{\mathrm{t}} u) = \det \overline{u}^{\mathrm{t}} \cdot \det u = \overline{\det u} \cdot \det u$$

zwar wie im $SO(n)$-Fall, daß

$$|\det u| = 1$$

ist, aber das läßt für die komplexe Zahl $\det u$ mehr Möglichkeiten als nur ± 1 wie im Reellen.

Von orthogonalen Komplementen, Summen und Projektionen kann man auch in unitären Räumen reden, muß allerdings bei der Projektionsformel

$$p_U(w) = \langle u_1, w \rangle u_1 + \cdots + \langle u_r, w \rangle u_r$$

(für eine Orthonormalbasis $(u_1, \ldots, u_r)$ von U) auf die Reihenfolge in den Klammern achten, sonst könnte p_U ja nicht mehr linear sein! Auf diesen Punkt muß man auch bei der Definition von

$$V \xrightarrow{\Sigma} V^{\check{}}; \quad v \mapsto \big(w \mapsto \langle v, w \rangle\big)$$

in Satz 27.2 achten. Damit die Werte von Σ wirklich lineare und nicht konjugiert-lineare Formen werden, müssen wir hinnehmen, daß Σ selbst ein konjugiert-linearer Isomorphismus von V nach $V^{\check{}}$ wird, insbesondere *kein* Isomorphismus im üblichen Sinne ist:

$$\Sigma(\lambda v) = \overline{\lambda} \Sigma(v) \quad \text{für alle } \lambda \in \mathbb{C},\ v \in V$$

Entsprechend ist die Satzaussage für unitäre Räume also abzuändern.

Schließlich die zu $f\colon V \longrightarrow W$ adjungierte Abbildung f^*: sie ist wie im euklidischen Fall durch die Kommutativität des Diagramms

$$\begin{array}{ccc} V & \xleftarrow{\ f^*\ } & W \\ {\scriptstyle\Sigma}\downarrow{\scriptstyle\simeq} & & {\scriptstyle\Sigma}\downarrow{\scriptstyle\simeq} \\ V^{\check{}} & \xleftarrow{\ f^{\check{}}\ } & W^{\check{}} \end{array}$$

— oder gleichwertig durch die Identität

$$\big\langle f(v), w \big\rangle = \big\langle v, f^*(w) \big\rangle \text{ für alle } v \in V,\ w \in W$$

definiert. Als Komposition einer linearen mit *zwei* konjugiert-linearen Abbildungen ist f^* auch hier linear und nicht konjugiert-linear. Deshalb läßt sich f^* (im Gegensatz zu Σ) in der üblichen Weise durch eine Matrix beschreiben; wenn f bezüglich Orthonormalbasen durch eine Matrix c gegeben ist, so rechnet man schnell nach, daß $\overline{c}^{\mathrm{t}}$ die Matrix zu f^* ist. Überhaupt kann man sich als Faustregel für den Matrizenkalkül in unitären Räumen merken, daß dort, wo im euklidischen Fall transponiert wird, jetzt zusätzlich noch zu konjugieren ist. Eine von den Physikern deshalb sehr geliebte Bezeichnung ist c^* für die zu c transponierte und komplex-konjugierte Matrix. Gegen die ist auch nichts einzuwenden, solange man mit Orthonormalbasen

arbeitet; wenn man aber andere Basen benutzt, muß man darauf achten, daß c^* dann im allgemeinen *nicht* die adjungierte Abbildung beschreibt.

Übungsaufgaben

27.1 $\underline{v} = (v_1, \dots, v_n)$ und $\underline{w} = (w_1, \dots, w_p)$ seien Basen der K-Vektorräume V und W, und $f\colon V \longrightarrow W$ sei eine lineare Abbildung. Zeigen Sie, daß sich die Koeffizienten der f bezüglich $\underline{v}$ und $\underline{w}$ beschreibenden Matrix $a \in \mathrm{Mat}(p\times n, K)$ mittels der zu $\underline{w}$ dualen Basis $\underline{\check{w}} = (\check{w}_1, \dots, \check{w}_p)$

$$a_{ij} = \check{w}_i \circ f(v_j)$$

schreiben lassen.

27.2 $S, T \subset \mathbb{R}^4$ seien die linearen Teilräume

$$S = \mathrm{Lin}\left(\begin{pmatrix}1\\1\\-2\\-1\end{pmatrix}, \begin{pmatrix}2\\1\\-1\\1\end{pmatrix}\right) \quad \text{und} \quad T = \mathrm{Lin}\left(\begin{pmatrix}1\\0\\2\\1\end{pmatrix}, \begin{pmatrix}0\\0\\1\\-1\end{pmatrix}\right).$$

Berechnen Sie eine Basis für $S \cap T$.

27.3 V und W seien endlichdimensionale euklidische Vektorräume. Zeigen Sie, daß

$$\mathrm{Kern}(f^* \circ f) = \mathrm{Kern}\, f \quad \text{und} \quad \mathrm{Bild}(f \circ f^*) = \mathrm{Bild}\, f$$

für jede lineare Abbildung $f\colon V \longrightarrow W$ gilt.

27.4 Das Standardskalarprodukt von $\mathbb{R}^{2n}$ läßt sich natürlich auch als Skalarprodukt auf $(\mathbb{C}^n)_\mathbb{R} = \mathbb{R}^{2n}$ auffassen; wir bezeichnen es hier mit

$$(\mathbb{C}^n)_\mathbb{R} \times (\mathbb{C}^n)_\mathbb{R} \ni (w, z) \longmapsto \langle w, z\rangle_\mathbb{R} \in \mathbb{R}.$$

Andererseits trägt $\mathbb{C}^n$ das hermitesche Standardprodukt

$$\mathbb{C}^n \times \mathbb{C}^n \ni (w, z) \longmapsto \langle w, z\rangle = \overline{w}^{\mathrm{t}} z \in \mathbb{C}.$$

Untersuchen Sie, ob ein Zusammenhang zwischen $\langle w, z\rangle_\mathbb{R}$ und $\langle w, z\rangle$ besteht.

27.5 Die vollständige Lösung dieser Aufgabe ist etwas umfangreicher; die einzelnen Schritte sind aber nicht schwierig, insbesondere dann, wenn Sie den Vorschlägen unten folgen.

(a) Zeigen Sie, daß

$$L := \{c \in \mathrm{Mat}(2\times 2, \mathbb{C})_\mathbb{R} \mid \overline{c}^{\mathrm{t}} = -c \text{ und } \mathrm{tr}\, c = 0\}$$

ein dreidimensionaler (reeller!) Untervektorraum von $\mathrm{Mat}(2\times 2, \mathbb{C})_\mathbb{R}$ ist. Verifizieren Sie, daß die Zuordnung

$$L \times L \ni (c, d) \longmapsto \langle c, d\rangle := -\,\mathrm{tr}\, cd \in \mathbb{R}$$

ein Skalarprodukt auf L definiert, L also zu einem euklidischen Vektorraum macht.

(b) Beweisen Sie, daß die Zuordnung

$$SU(2) \ni u \longmapsto \big(L \ni c \mapsto uc\overline{u}^{\mathrm{t}} = ucu^{-1} \in L\big)$$

einen Gruppenhomomorphismus

$$SU(2) \xrightarrow{h} SO(L)$$

definiert und daß $\operatorname{Kern} h = \{\pm 1\}$ ist.

Anmerkungen zu dieser Aufgabe: Um Teil (a) zu lösen, ist es ganz praktisch, eine Basis von L explizit hinzuschreiben; die Basismatrizen, die Ihnen (wahrscheinlich) als erste einfallen, sind bis auf einen Faktor i die, die bei den Physikern Pauli-Matrizen heißen. Überlegen Sie sich, daß für die Spur ganz allgemein die Regeln

$$\operatorname{tr} c = \operatorname{tr} c^{\mathrm{t}} \quad \text{und} \quad \operatorname{tr} cd = \operatorname{tr} dc$$

gelten. Zur Berechnung des Kerns in (b) sind dann natürlich diejenigen $u \in SU(2)$ zu betrachten, für die $h(u)$ die drei Basismatrizen festläßt.

Wenn Physiker über den Spin reden, wollen sie einem gerne weismachen, eine Drehung um 360° sei etwas Anderes als eine um 0° (also die identische Abbildung), und erst eine Drehung um 720° sei wieder die Identität usw. Das ist natürlich Unsinn, aber es ist etwas Richtiges damit gemeint: Wie Teil (a) der Aufgabe zeigt, darf man sich L als den (klassischen) physikalischen Raum vorstellen — dreidimensional und mit dem gewöhnlichen Skalarprodukt. Wenn man nun in $SU(2)$ zum Beispiel den Weg

$$[0, 2\pi] \ni t \longmapsto u_t := \begin{pmatrix} e^{it} & \\ & e^{-it} \end{pmatrix} \in SU(2)$$

betrachtet, so erweist sich $h(u_t) \in SO(L)$ als eine Drehung von L um den Winkel $2t$ (nachrechnen!); insbesondere ist zwar $h(u_\pi) = \mathrm{id}$, aber eben nicht $u_\pi = \mathrm{id}$, sondern $u_\pi = -\mathrm{id}$. Die Elemente von $SU(2)$ haben neben der durch h vermittelten Wirkung auf den Raum eine (der klassischen Theorie verborgene) Wirkung auf den quantenmechanischen Spin, und $u_\pi = -\mathrm{id} \in SU(2)$ klappt den Spin eines Fermi-Teilchens um.

Man kann zeigen, daß h surjektiv ist (mit den derzeit zur Verfügung stehenden Mitteln wäre das etwas umständlich); zusammen mit dem Resultat $\operatorname{Kern} h = \{\pm 1\}$ folgt dann sofort, daß *alle* Fasern von h genau zwei Elemente haben, nämlich eine Matrix u und die dazu entgegengesetzte $-u$.

28 Normale Endomorphismen

Dieser Abschnitt ist ohne Zweifel der schönste aus dem Bereich der linearen Algebra. Die Resultate, die ich Ihnen hier und im nächsten Abschnitt vorstelle, sind ebenso tiefsinnig und überraschend wie praktisch wichtig; sie werden sich andererseits auf die einfachste Weise daraus ergeben, daß wir früher gewonnene Erkenntnisse auf raffinierte Weise zusammenführen.

28.1 Definition V sei ein euklidischer oder unitärer Vektorraum. Ein Endomorphismus $f: V \longrightarrow V$ heißt normal, wenn f und f^* vertauschbar sind:

$$f^* \circ f = f \circ f^*$$

Mit dieser zunächst schwer zu motivierenden Eigenschaft können wir uns leichter anfreunden, wenn wir feststellen, daß sie einige leicht zu durchschauende Spezialfälle enthält. Ich stelle sie mit einigen schon bekannten in einer Tabelle zusammen:

28.2 Definitionstabelle für Eigenschaften von Endomorphismen eines abstrakten unitären oder euklidischen Vektorraums V sowie von Endomorphismen der Standardräume $\mathbb{C}^n$ und $\mathbb{R}^n$ (alias $n \times n$-Matrizen):

$f \in \mathrm{End}(V)$	$c \in \mathrm{Mat}(n\times n, \mathbb{C})$	$a \in \mathrm{Mat}(n\times n, \mathbb{R})$
selbstadjungiert: $f^* = f$	hermitesch: $\overline{c}^{\mathrm{t}} = c$	symmetrisch: $a^{\mathrm{t}} = a$
anti-selbstadjungiert: $f^* = -f$	schiefhermitesch: $\overline{c}^{\mathrm{t}} = -c$	schiefsymmetrisch: $a^{\mathrm{t}} = -a$
unitär/orthogonal: $f^* = f^{-1}$	unitär: $\overline{c}^{\mathrm{t}} c = 1$	orthogonal: $a^{\mathrm{t}} a = 1$

In jeder der drei Zeilen ist links eine Eigenschaft eines abstrakten Endomorphismus genannt, daneben die gebräuchliche Bezeichnung dafür im Spezialfall des unitären, und schließlich die im Fall des euklidischen Standardraums.

28.3 Notiz Alle in der Definition 28.2 genannten Typen von Endomorphismen sind normal.

28.4 Konkrete Beispiele Die reelle Matrix

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{pmatrix} \in \mathrm{Mat}(3\times 3, \mathbb{R})$$

ist offenbar symmetrisch und deshalb auch hermitesch; dagegen ist

$$c = \begin{pmatrix} 1 & 2i \\ 2i & 0 \end{pmatrix} \in \mathrm{Mat}(2\times 2, \mathbb{C})$$

zwar symmetrisch, aber nicht hermitesch, und tatsächlich auch nicht normal:

$$\begin{aligned} \overline{c}^{\mathrm{t}} &= \begin{pmatrix} 1 & -2i \\ -2i & 0 \end{pmatrix} \\ \overline{c}^{\mathrm{t}} c &= \begin{pmatrix} 1 & -2i \\ -2i & 0 \end{pmatrix} \begin{pmatrix} 1 & 2i \\ 2i & 0 \end{pmatrix} = \begin{pmatrix} 5 & 2i \\ -2i & 4 \end{pmatrix} \\ c\,\overline{c}^{\mathrm{t}} &= \begin{pmatrix} 1 & 2i \\ 2i & 0 \end{pmatrix} \begin{pmatrix} 1 & -2i \\ -2i & 0 \end{pmatrix} = \begin{pmatrix} 5 & -2i \\ 2i & 4 \end{pmatrix} \end{aligned}$$

Andererseits ist

$$\begin{pmatrix} 1 & 2 & -3i \\ 2 & 4 & 5 \\ 3i & 5 & 6 \end{pmatrix} \in \mathrm{Mat}(3\times 3, \mathbb{C})$$

ein Beispiel einer nicht-reellen hermiteschen Matrix. Beachten Sie, daß die Diagonaleinträge einer hermiteschen Matrix stets reell sein müssen.

Beispiele schiefhermitescher Matrizen in $\mathrm{Mat}(3\times 3, \mathbb{C})$ sind

$$\begin{pmatrix} i & 2 & 3i \\ -2 & 0 & -5 \\ 3i & 5 & 0 \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} 0 & 2 & 3 \\ -2 & 0 & -5 \\ -3 & 5 & 0 \end{pmatrix};$$

hier müssen die Diagonaleinträge rein imaginär, im reellen schiefsymmetrischen Fall also null sein.

Bemerkung Es versteht sich von selbst, daß ein Endomorphismus f genau dann normal ist, wenn die zugehörige Matrix c bezüglich irgendeiner (und dann auch jeder) Orthonormalbasis normal ist, d.h. $\overline{c}^{\mathrm{t}} c = c\,\overline{c}^{\mathrm{t}}$ erfüllt: Dafür, daß die Forderung $f^* \circ f = f \circ f^*$ sich in eine Matrizengleichung $c^* c = c\,c^*$ übersetzt, würde zwar schon eine beliebige Basis genügen, aber erst deren Orthonormalität stellt sicher, daß das f^* entsprechende c^* das der Physiker, also $c^* = \overline{c}^{\mathrm{t}}$ ist.

Der Kern dieses Abschnitts ist der sogenannte

28.5 Spektralsatz V sei ein n-dimensionaler unitärer Vektorraum, und

$$f\colon V \longrightarrow V$$

sei ein normaler Endomorphismus. Dann existiert eine Orthonormalbasis $\underline{v}$ von V, so daß f bezüglich $\underline{v}$ durch eine Diagonalmatrix

$$c = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \in \mathrm{Mat}(n\times n, \mathbb{C})$$

beschrieben wird.

Sicher rechtfertigt der Satz das Interesse an dem Begriff "normal"; er besitzt übrigens die folgende ganz leicht einzusehende

28.6 Umkehrung Wird der Endomorphismus f bezüglich einer Orthonormalbasis durch eine Diagonalmatrix beschrieben, so ist f normal.

Beweis der Umkehrung Ist wie oben

$$c = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

diese Matrix, so wird f^* durch

$$\overline{c}^{\mathrm{t}} = \begin{pmatrix} \overline{\lambda_1} & & \\ & \ddots & \\ & & \overline{\lambda_n} \end{pmatrix}$$

beschrieben, und ganz allgemein sind beliebige Diagonalmatrizen miteinander vertauschbar:

$$\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \begin{pmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_n \end{pmatrix} = \begin{pmatrix} \lambda_1\mu_1 & & \\ & \ddots & \\ & & \lambda_n\mu_n \end{pmatrix} = \begin{pmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_n \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

Also ist f normal.

Zum Beweis des Spektralsatzes benötigen wir das

28.7 Lemma $c \in \mathrm{Mat}(n \times n, \mathbb{C})$ sei eine obere Dreiecksmatrix. Ein solches c ist nur dann normal, wenn c schon diagonal ist.

Beweis Speziell für hermitesche oder schiefhermitesche Matrizen wäre das Lemma völlig trivial. Im allgemeinen muß man aber doch ein wenig rechnen. Wir verwenden Induktion nach $n \in \mathbb{N}$; für $n = 0$ ist natürlich nichts zu zeigen. Sei also $n > 0$. Die Voraussetzung über c besagt zunächst

$$c_{jk} = 0 \quad \text{für } j > k.$$

Wir müssen zeigen, daß aus

$$\overline{c}^{\mathrm{t}} c = c\,\overline{c}^{\mathrm{t}}$$

die Diagonalität von c folgt. Dazu rechnen wir von der Matrizengleichung $\overline{c}^{\mathrm{t}} c = c\,\overline{c}^{\mathrm{t}}$ nur die 11-Kompenente (die in der Ecke oben links) aus:

$$\left(\overline{c}^{\mathrm{t}} c\right)_{11} = \sum_{k=1}^{n} \left(\overline{c}^{\mathrm{t}}\right)_{1k} c_{k1} = \sum_{k=1}^{n} \overline{c_{k1}} c_{k1} = \overline{c_{11}} c_{11} = |c_{11}|^2$$

und

$$\left(c\,\overline{c}^{\mathrm{t}}\right)_{11} = \sum_{k=1}^{n} c_{1k} \left(\overline{c}^{\mathrm{t}}\right)_{k1} = \sum_{k=1}^{n} c_{1k} \overline{c_{1k}} = \sum_{k=1}^{n} |c_{1k}|^2;$$

die Normalität liefert also

$$|c_{11}|^2 = \sum_{k=1}^{n} |c_{1k}|^2$$

oder

$$\sum_{k=2}^{n} |c_{1k}|^2 = 0.$$

Das geht nur, wenn

$$c_{12} = c_{13} = \cdots = c_{1n} = 0$$

ist, c also die Form

$$c = \left(\begin{array}{c|ccc} c_{11} & 0 & \dots & 0 \\ \hline 0 & & & \\ \vdots & & c' & \\ 0 & & & \end{array}\right)$$

mit einer $(n-1) \times (n-1)$-Matrix c' hat. Die Normalitätsgleichung $\overline{c}^{\mathrm{t}} c = c\,\overline{c}^{\mathrm{t}}$ reduziert sich jetzt auf

$$\overline{c'}^{\mathrm{t}} c' = c'\,\overline{c'}^{\mathrm{t}}.$$

Nun ist c' wieder eine normale obere Dreiecksmatrix. Nach Induktionsannahme ist c' diagonal, c also auch.

Beweis des Spektralsatzes Den können wir jetzt richtig genießen, besteht er doch nur noch darin, Ergebnisse früherer Anstrengungen wie Mosaiksteine zusammenzusetzen.

Weil das charakteristische Polynom $\chi_f(X) \in \mathbb{C}[X]$ in Linearfaktoren zerfällt, ist Satz 24.18 anwendbar; er liefert uns eine unter f invariante Flagge

$$\{0\} = V_0 \subset V_1 \subset \cdots \subset V_n = V.$$

Auf diese Flagge wenden wir den Orthonormalisierungssatz 25.13 an, und wir erhalten eine Orthonormalbasis $\underline{v} = (v_1, \dots, v_n)$ von V mit $\mathrm{Lin}(v_1, \dots, v_k) = V_k$ für $k = 1, \dots, n$. Die Matrix c von f bezüglich $\underline{v}$ ist daher eine obere Dreiecksmatrix; weil f normal und $\underline{v}$ orthonormal ist, ist c auch normal. Nach dem Lemma ist c also eine Diagonalmatrix, und wir sind schon fertig.

Anmerkungen Der Spektralsatz hängt natürlich eng mit der im Abschnitt 24 besprochenen Eigenwerttheorie zusammen. Er sagt ja insbesondere, daß jeder normale Endomorphismus eines endlichdimensionalen unitären Vektorraums diagonalisierbar ist, verspricht darüber hinaus aber noch, daß die Diagonalisierung sogar durch eine orthonormale Basis erreicht werden kann. Wir erinnern uns daran, wie man einer Diagonalmatrix ihre Eigenräume ansieht: Es empfiehlt sich, durch Vertauschen der Basisvektoren noch dafür zu sorgen, daß die Diagonaleinträge (also die Eigenwerte) in dem Sinne geordnet sind, daß gleiche Einträge beisammen stehen, etwa

$$\underbrace{\lambda_1,\dots,\lambda_1}_{e_1\text{-mal}},\underbrace{\lambda_2,\dots,\lambda_2}_{e_2\text{-mal}},\dots,\underbrace{\lambda_r,\dots,\lambda_r}_{e_r\text{-mal}}.$$

Der zu λ_j gehörige Eigenraum E_j ist dann gerade der Koordinatenunterraum

$$\{0\}\times\mathbb{C}^{e_j}\times\{0\}\subset\mathbb{C}^{e_1+\dots+e_{j-1}}\times\mathbb{C}^{e_j}\times\mathbb{C}^{e_{j+1}+\dots+e_r}=\mathbb{C}^n$$

bzw. dessen Bild unter dem Basisisomorphismus (wenn die Matrix nur dazu dient, einen abstrakten Endomorphismus zu beschreiben).

Alternativ können wir das Verhältnis des Spektralsatzes zu den früheren Überlegungen demnach auch so fassen: Diagonalisierbarkeit von $f\colon V\longrightarrow V$ bedeutet, daß $V=E_1+\dots+E_r$ direkte Summe der Eigenräume von f ist; die Schlußfolgerung des Spektralsatzes verspricht darüber hinaus, daß es sich hier um eine *orthogonale* Summe handelt. Letzteres wollen wir auch gleich formal festhalten:

28.8 Lemma *f sei ein normaler Endomorphismus eines endlichdimensionalen orthogonalen oder unitären Vektorraums. Dann stehen Eigenvektoren zu verschiedenen Eigenwerten von f aufeinander senkrecht.*

Der Spektralsatz macht auch die Stellung der unter 28.3 aufgeführten Spezialfälle innerhalb der Klasse aller normalen Endomorphismen deutlich:

28.9 Satz *$f\colon V\longrightarrow V$ sei ein normaler Endomorphismus eines endlichdimensionalen unitären Vektorraums. Dann gilt:*

$$\begin{aligned} f\text{ selbstadjungiert} &\iff \text{alle Eigenwerte von } f \text{ sind reell}\\ f\text{ antiselbstadjungiert} &\iff \text{alle Eigenwerte von } f \text{ sind rein imaginär}\\ f\text{ unitär} &\iff |\lambda|=1 \text{ für jeden Eigenwert } \lambda \text{ von } f \end{aligned}$$

Beweis Der Spektralsatz erlaubt uns, f bezüglich einer passenden Orthonormalbasis durch eine Diagonalmatrix

$$c=\begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{pmatrix}$$

zu beschreiben, worin die Diagonaleinträge wie immer die Eigenwerte von c, also von f sind. Nun ist f genau dann selbstadjungiert, wenn c hermitesch, d.h. wenn

$$\overline{\lambda_j}=\lambda_j,\quad \text{d.h. } \lambda_j\in\mathbb{R} \text{ für alle } j$$

ist. Entsprechend ist f antiselbstadjungiert, wenn c schiefhermitesch, also wenn

$$\overline{\lambda_j}=-\lambda_j,\quad \text{d.h. } \lambda_j\in i\mathbb{R} \text{ für alle } j$$

gilt. Schließlich ist f unitär genau wenn $\overline{c}^{\mathrm{t}}c=1$, d.h. wenn

$$\overline{\lambda_j}\lambda_j=1,\quad \text{also wenn } |\lambda_j|=1 \text{ für alle } j$$

ist.

Die Methoden, mit denen man die vom Spektralsatz in Aussicht gestellte Diagonalisierung einer normalen Matrix c praktisch berechnen kann, sind uns längst vertraut. Kommt es einem nur darauf an, in *welche* Diagonalmatrizen sich c durch unitären Kartenwechsel (Übergang zu einer anderen Orthonormalbasis)

überführen läßt, braucht man bloß das charakteristische Polynom χ_c in Linearfaktoren zu zerlegen: die gesuchten Diagonaleinträge sind ja genau die Eigenwerte von c unter Berücksichtigung ihrer Vielfachheiten. (Die Reihenfolge, in der sie auf der Diagonalen realisiert werden, ist natürlich ganz willkürlich.) Will man dagegen auch eine Orthonormalbasis berechnen, die die Diagonalisierung leistet, so wird man in einem zweiten Schritt Basen für die Eigenräume von c bestimmen. Wie wir eben bemerkt haben, stehen Basisvektoren aus verschiedenen Eigenräumen ohnehin aufeinander senkrecht, man muß also in einem dritten und letzten Schritt nur die Basen, die man für jeden einzelnen Eigenraum hat, nach dem Verfahren von Gram-Schmidt orthonormalisieren. Das macht um so weniger Arbeit, je kleiner (und deshalb zahlreicher) die Eigenräume sind — sind sogar alle eindimensional, so braucht man die Basisvektoren nur noch auf die Länge 1 zu normieren.

28.10 Beispiel Der durch die lustige Matrix

$$c = \begin{pmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{pmatrix} \in \mathrm{Mat}(4\times 4, \mathbb{C})$$

beschriebene Endomorphismus von $\mathbb{C}^4$ ist selbstadjungiert. Mit einiger Geduld läßt sich sein charakteristisches Polynom

$$\chi_c(X) = X^3(X-4) \in \mathbb{C}[X]$$

ausrechnen. Freilich braucht man sich diese Mühe hier kaum zu machen: man sieht ja, daß $\operatorname{rk} c = 1$, also $\dim \operatorname{Kern} c = 3$ ist, und nach 24.19 oder ganz einfach direkt nach dem Spektralsatz muß folglich $\chi_c(X)$ den Faktor X^3 enthalten. Den zugehörigen Eigenraum, also den Kern, erhalten wir nach dem Standardverfahren fast ohne zu rechnen:

$$\operatorname{Kern} c = \operatorname{Kern} \begin{pmatrix} 1 & -1 & 1 & -1 \end{pmatrix} = \operatorname{Lin}\left(\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \right) \subset \mathbb{C}^4$$

Für den verbleibenden einfachen Eigenwert 4 liefert die Rechnung

$$\begin{pmatrix} 3 & 1 & -1 & 1 \\ 1 & 3 & 1 & -1 \\ -1 & 1 & 3 & 1 \\ 1 & -1 & 1 & 3 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -1 & 1 & 3 \\ -1 & 1 & 3 & 1 \\ 1 & 3 & 1 & -1 \\ 3 & 1 & -1 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -1 & 1 & 3 \\ 0 & 0 & 4 & 4 \\ 0 & 4 & 0 & -4 \\ 0 & 4 & -4 & -8 \end{pmatrix}$$

$$\longrightarrow \begin{pmatrix} 1 & -1 & 1 & 3 \\ & 1 & 0 & -1 \\ & 0 & 4 & 4 \\ & 4 & -4 & -8 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -1 & 1 & 3 \\ & 1 & 0 & -1 \\ & 0 & 1 & 1 \\ & 0 & -4 & -4 \end{pmatrix}$$

$$\longrightarrow \begin{pmatrix} 1 & -1 & 1 & 3 \\ & 1 & 0 & -1 \\ & & 1 & 1 \\ & & & \end{pmatrix}$$

den Eigenraum

$$\operatorname{Lin}\left(\begin{pmatrix} -1 \\ 1 \\ -1 \\ 1 \end{pmatrix} \right).$$

Bezeichnen wir die so erhaltenen Basisvektoren mit w_1, w_2, w_3, w_4, so sehen wir, daß w_1, w_2, w_3 auf w_4 senkrecht stehen, wie es ja auch sein muß (wir hätten das benutzen können, um w_4 ohne Kenntnis von χ_c zu

berechnen). Bei der Konstruktion einer diagonalisierenden Orthonormalbasis brauchen wir w_4 also nur zu

$$v_4 = \frac{1}{2}w_4 = \frac{1}{2}\begin{pmatrix} -1 \\ 1 \\ -1 \\ 1 \end{pmatrix}$$

zu normieren. w_1, w_2, w_3 dagegen müssen wir dem vollständigen Verfahren nach Gram-Schmidt unterwerfen: Nur

$$v_1 = \frac{1}{\sqrt{2}}w_1 = \frac{1}{2}\sqrt{2}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

erhält man allein durch Normieren. Der Vektor w_2 wird zunächst zu

$$w = \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix} - \left(-\frac{1}{2}\sqrt{2}\right)\frac{1}{2}\sqrt{2}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix}$$

und beim Normieren zu

$$v_2 = \frac{1}{6}\sqrt{6}\begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix}.$$

Aus w_3 schließlich ergibt sich

$$w = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} - \left(\frac{1}{2}\sqrt{2}\right)\frac{1}{2}\sqrt{2}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} - \left(-\frac{1}{6}\sqrt{6}\right)\frac{1}{6}\sqrt{6}\begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} - \frac{1}{2}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} + \frac{1}{6}\begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 1 \\ 3 \\ 1 \\ -1 \end{pmatrix}$$

und normiert

$$v_3 = \frac{1}{12}\sqrt{12}\begin{pmatrix} 1 \\ 3 \\ 1 \\ -1 \end{pmatrix}.$$

Wer will, kann sich den Spaß machen, das Endergebnis zu verifizieren, daß nämlich

$$u := \begin{pmatrix} \frac{1}{2}\sqrt{2} & -\frac{1}{6}\sqrt{6} & \frac{1}{12}\sqrt{12} & -\frac{1}{2} \\ 0 & 0 & \frac{1}{4}\sqrt{12} & \frac{1}{2} \\ 0 & \frac{1}{3}\sqrt{6} & \frac{1}{12}\sqrt{12} & -\frac{1}{2} \\ \frac{1}{2}\sqrt{2} & \frac{1}{6}\sqrt{6} & -\frac{1}{12}\sqrt{12} & \frac{1}{2} \end{pmatrix}$$

in der Tat eine unitäre (hier orthogonale) Matrix mit der Eigenschaft

$$u^{-1}c\,u = \overline{u}^{\mathrm{t}}c\,u = \begin{pmatrix} 0 & & & \\ & 0 & & \\ & & 0 & \\ & & & 4 \end{pmatrix}$$

ist.

Daß die Rechnung 28.10 ganz im Reellen abläuft, ist kein Zufall, denn für selbstadjungierte Endomorphismen gilt der Spektralsatz auch reell:

28.11 Satz *V sei ein n-dimensionaler euklidischer Vektorraum, und*

$$f\colon V \longrightarrow V$$

sei ein selbstadjungierter Endomorphismus. Dann existiert eine Orthonormalbasis $\underline{v}$ von V, so daß f bezüglich $\underline{v}$ durch eine Diagonalmatrix

$$a = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \in \mathrm{Mat}(n \times n, \mathbb{R})$$

beschrieben wird.

Beweis Man darf annehmen, daß $V = \mathbb{R}^n$ der Standardraum ist. Auf die (durch dieselbe symmetrische Matrix gegebene) Komplexifizierung von f ist Satz 28.9 über die Eigenwerte normaler Matrizen anwendbar; er garantiert uns, daß $\chi_f(X)$ nicht nur in $\mathbb{C}[X]$, sondern schon in $\mathbb{R}[X]$ in Linearfaktoren zerfällt. Damit läßt sich aber der Beweis des Spektralsatzes 28.5 ganz im Reellen führen, und wir sind fertig.

Für nicht selbstadjungierte Endomorphismen kann es keine rein reelle Version des Spektralsatzes geben, weil sie nach Satz 28.9 ja mindestens einen nicht-reellen Eigenwert haben. Was man da trotzdem erreichen kann, illustriere ich mit dem besonders wichtigen orthogonalen Fall.

28.12 Satz V sei ein n-dimensionaler euklidischer Vektorraum, und

$$f\colon V \longrightarrow V$$

sei ein orthogonaler Endomorphismus. Dann existiert eine Orthonormalbasis von V, bezüglich der f durch eine Matrix der Form

$$\begin{pmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & -1 & & & & & \\ & & & & \ddots & & & & \\ & & & & & -1 & & & \\ & & & & & & \boxed{u_1} & & \\ & & & & & & & \ddots & \\ & & & & & & & & \boxed{u_r} \end{pmatrix} \in O(n)$$

beschrieben wird, in der die "Kästchen" $u_j \in SO(2)$ zweidimensionale Drehmatrizen sind.

Beweis Wir dürfen wieder annnehmen, daß $V = \mathbb{R}^n$, und damit f eine orthogonale Matrix u ist. Nach dem Spektralsatz 28.5 ist $\mathbb{C}^n$ orthogonale Summe der Eigenräume der Komplexifizierung $u_{\mathbb{C}}$ (als Matrix ist natürlich $u_{\mathbb{C}} = u$). Als reelle Eigenwerte von u kommen nur ± 1 in Frage; die zugehörigen Eigenräume $E_\pm$ von $u_{\mathbb{C}}$ sind (als Lösungsraum einer reellen Gleichung) von der Form $(E'_\pm)_{\mathbb{C}}$, wobei $E'_\pm \subset \mathbb{R}^n$ der Eigenraum von u zum Eigenwert ± 1 ist. Insbesondere wird $E_\pm$ von einer *reellen* Orthonormalbasis aufgespannt.

Beim einem Eigenraum E zu einem nicht-reellen Eigenwert λ von u sieht das anders aus: E enthält überhaupt keine reellen Eigenvektoren. Andererseits ist der Raum $\overline{E}$ der zu Vektoren aus E komplex-konjugierten Vektoren gerade der Eigenraum zum Eigenwert $\overline{\lambda}$. Man rechnet nun folgende Tatsachen leicht nach: Ist $(v_1, \dots, v_k)$ eine Orthonormalbasis von E, so ist $(\overline{v_1}, \dots, \overline{v_k})$ eine Orthonormalbasis von $\overline{E}$, und die reellen Vektoren

$$\sqrt{2}\,\mathrm{Re}v_1,\ \sqrt{2}\,\mathrm{Im}v_1,\ \dots,\ \sqrt{2}\,\mathrm{Re}v_k,\ \sqrt{2}\,\mathrm{Im}v_k$$

bilden eine Orthonormalbasis des Summenraums $E + \overline{E}$. Bezüglich dieser (wie angegeben geordneten) Basis ergibt sich die Matrix der Einschränkung $E + \overline{E} \xrightarrow{u_{\mathbb{C}}} E + \overline{E}$ zu

$$\begin{pmatrix} \boxed{c_\lambda} & & \\ & \ddots & \\ & & \boxed{c_\lambda} \end{pmatrix} \in \mathrm{Mat}(2k \times 2k, \mathbb{R})$$

mit dem k-fach wiederholten Kästchen

$$c_\lambda = \begin{pmatrix} \mathrm{Re}\lambda & -\mathrm{Im}\lambda \\ \mathrm{Im}\lambda & \mathrm{Re}\lambda \end{pmatrix} \in SO(2).$$

Wirft man nun alle so konstruierten Basen zusammen, erhält man eine Basis von $\mathbb{R}^n$, die das Gewüschte leistet.

Dieser Satz enthält unsere frühere Diskussion der Elemente von $SO(3)$ und dehnt sie auf beliebige Dimension aus: Da für ein Element $u \in SO(n)$ der Eigenwert -1 eine gerade Vielfachheit haben muß, zeigt der Satz, daß jedes solche u aus $\lfloor \frac{n}{2} \rfloor$ Drehungen in paarweise zueinander senkrechten Ebenen aufgebaut ist, zu denen bei ungeradem n noch eine Fixachse kommt. (Das heißt natürlich nicht, daß man u allein durch die Angabe von $\lfloor \frac{n}{2} \rfloor$ Winkeln festlegen könnte, denn daraus würde man noch nichts über die Lage der Drehebenen erfahren.)

Bemerkung zum Namen "Spektralsatz" In der Physik sind Ihnen sicher schon Frequenzspektren (zum Beispiel das einer Trommel) und Energiespektren (zum Beispiel das eines Atoms) begegnet. Das mathematische Objekt, das beiden zugrundeliegt, ist jeweils ein Differentialoperator ähnlich dem in den Aufgaben 15.1 und 15.2 betrachteten. Diesen Operator kann man als eine selbstadjungierte lineare Abbildung auffassen, allerdings nicht zwischen endlichdimensionalen Räumen, sondern zwischen Funktionenräumen mit Skalarprodukt, genauer sogenannten *Hilbert*-Räumen. Das Spektrum eines solchen Operators ist eine Menge von (in den zitierten Beispielen aus der Physik reellen) Zahlen, die in jedem Fall alle Eigenwerte enthält, darüber hinaus oft noch eine Art verallgemeinerte Eigenwerte, die im Fall endlicher Dimension nicht auftreten. Jedenfalls gibt es auch für diese allgemeinere Situation einen ganz analogen Spektralsatz, und er heißt so, weil er beschreibt, wie man den Operator aus seinem Spektrum (und weiterer Daten) rekonstruieren kann. Im endlichdimensionalen Fall ist das Spektrum dasselbe wie die Menge aller Eigenwerte, und der Spektralsatz reduziert sich auf unseren Satz 28.5, der ja auch letztlich die gegebene lineare Abbildung durch ihre Eigenwerte und -vektoren ausdrückt.

Im allgemeinen wird es nicht möglich sein, zwei diagonalisierbare Endomorphismen eines Vektorraums simultan, d.h. durch ein und dieselbe Basis zu diagonalisieren: Weil je zwei Diagonalmatrizen miteinander vertauschbar sind, kann das nur dann gehen, wenn auch die beiden Endomorphismen vertauschbar sind. Die folgende Verallgemeinerung des Spektralsatzes 28.5 läßt erkennen, daß das auch schon der wesentliche Punkt ist. Sie gilt übrigens entsprechend in der "reinen" linearen Algebra ohne Skalarprodukte, wenn man die Diagonalisierbarkeit der beteiligten Endomorphismen aus anderen Gründen weiß.

28.13 Satz V sei ein n-dimensionaler unitärer Vektorraum, und F eine Menge von normalen Endomorphismen von V, die paarweise miteinander vertauschbar sind:

$$f \circ g = g \circ f \quad \text{für alle } f, g \in F$$

Dann gibt es eine Orthonormalbasis von V, bezüglich der jedes $f \in F$ durch eine Diagonalmatrix beschrieben wird.

Beweis Wir betrachten orthogonale Zerlegungen

$$V = E_1 + \cdots E_r$$

von V in Unterräume, die unter jedem $f \in F$ invariant sind:

$$f(E_j) \subset E_j \quad \text{für jedes } f \in F \text{ und } j = 1, \ldots, r$$

Eine triviale derartige Zerlegung ist $V = E_1$, und wir werden diese durch orthogonale Zerlegung der einzelnen Summanden schrittweise so verfeinern, daß schließlich jedes $f \in F$ auf jedem E_j als Skalar wirkt. Dann brauchen wir bloß noch in jedem E_j eine Orthonormalbasis zu wählen und haben gewonnen.

Der Verfeinerungsschritt läuft so: Wenn jedes $f \in F$ auf jedem E_j schon als Skalar wirkt, ist nichts zu tun. Andernfalls wählen wir willkürlich ein $g \in F$ und $E \in \{E_1, \ldots, E_r\}$, so daß die Einschränkung

$$g' \colon E \xrightarrow{g} E$$

nicht skalar ist. Aber immerhin ist das ein normaler Endomorphismus von E, und nach dem Spektralsatz können wir

$$E = E'_1 + \cdots E'_s$$

orthogonal in die Eigenräume von g' zerlegen. Indem wir die Zerlegung von E anstelle von E in die ursprüngliche einsetzen, haben wir diese jedenfalls verfeinert. Der springende Punkt ist nun, daß auch die Unterräume E'_k unter jedem $f \in F$ invariant sind: Sei $v \in E'_k$ und $\lambda \in \mathbb{C}$ der zugehörige Eigenwert von g'. Dann ist $f(v) \in E$, und es gilt (vergleiche Aufgabe 24.2)

$$g'\big(f(v)\big) = g\big(f(v)\big) = f\big(g(v)\big) = \lambda f(v);$$

also ist auch $f(v)$ ein Eigenvektor von g' zum Eigenwert λ, d.h. $f(v) \in E'_k$. Damit ist der Verfeinerungsschritt abgeschlossen.

Die Verfeinerungskette muß spätestens nach $\dim V$ Schritten damit enden, daß nichts mehr zu tun ist.

Zum Schluß dieses Abschnitts noch zwei Resultate, die die vielfältige Anwendbarkeit des Spektralsatzes illustrieren. Das erste ist ganz einfach, man sollte es in Verbindung mit der in der Aufgabe 24.5 ausführlich diskutierten Tatsache sehen, daß es nicht-triviale nilpotente Matrizen gibt, also Matrizen $a \neq 0$ mit der Eigenschaft $a^r = 0$ für genügend große $r \in \mathbb{N}$.

28.14 Lemma $a \in \mathrm{Sym}(n, \mathbb{R})$ sei eine symmetrische Matrix. Gibt es ein $r \in \mathbb{N}$ mit $a^r = 0$, dann ist $a = 0$.

Beweis Nach dem Spektralsatz 28.11 dürfen wir annehmen, daß a eine Diagonalmatrix ist, und dann ist die Aussage klar.

Eine viel raffiniertere Anwendung:

28.15 Satz V sei ein endlichdimensionaler komplexer Vektorraum, $f\colon V \longrightarrow V$ ein Endomorphismus. Wenn es eine natürliche Zahl r mit $f^r = \mathrm{id}_V$ gibt, dann ist f diagonalisierbar.

Beweis Wir versehen V mit einem zunächst beliebigen hermiteschen Skalarprodukt $\langle\,,\rangle$. Die Formel

$$\langle\!\langle v, w\rangle\!\rangle := \sum_{j=0}^{r-1} \big\langle f^j(v), f^j(w)\big\rangle$$

definiert dann, wie man unmittelbar nachrechnet, ein neues Skalarprodukt auf V. Der Clou ist, daß f bezüglich dieses neuen Skalarprodukts unitär ist:

$$\big\langle\!\big\langle f(v), f(w)\big\rangle\!\big\rangle = \sum_{j=0}^{r-1} \big\langle f^{j+1}(v), f^{j+1}(w)\big\rangle = \sum_{j=1}^{r} \big\langle f^j(v), f^j(w)\big\rangle = \langle\!\langle v, w\rangle\!\rangle$$

Nach dem Spektralsatz ist f also diagonalisierbar (und wir können die unitäre Struktur auf V, die nur als Hilfsmittel gedient hat, wieder vergessen).

Übungsaufgaben

28.1 Die Matrix

$$c = \begin{pmatrix} -1 & & \\ & 1 & i \\ & & 1 \end{pmatrix} \in \mathrm{Mat}(3\times 3, \mathbb{C})$$

werde als Endomorphismus von $\mathbb{C}^3$ aufgefaßt. Dann gibt es keine Basis von $\mathbb{C}^3$, bezüglich der die Matrix dieses Endomorphismus hermitesch ist. Warum?

28.2 Die Matrix $c \in \mathrm{Mat}(n \times n, \mathbb{C})$ sei unitär, zugleich aber auch hermitesch und positiv definit. Beweisen Sie, daß dann nur $c = 1$ sein kann.

28.3 Die in der Vorlesung gegebene Definition des Begriffs "selbstadjungiert" ist nur auf Endomorphismen $f: V \longrightarrow V$ mit endlichdimensionalem V anwendbar, denn sie bezieht sich auf die adjungierte Abbildung f^*. Probieren Sie, ob man Selbstadjungiertheit auch ohne diese Voraussetzung definieren kann, so daß für selbstadjungiertes f (vielleicht nicht gleich der schwierig zu beweisende Spektralsatz, aber immerhin) folgendes gilt:

(a) Alle (komplexen) Eigenwerte von f sind reell, und

(b) Eigenvektoren zu verschiedenen Eigenwerten stehen aufeinander senkrecht.

Die beiden folgenden Aufgabe knüpfen an die Ausführungen gegen Ende von Abschnitt 22 über Permutationen und ihre Darstellung durch Permutationsmatrizen an. Insbesondere wird an die Schreibweise $(j_1\ j_2\ j_3\ \ldots\ j_r)$ für eine zyklische Permutation erinnert.

28.4 Begründen Sie, warum jede Permutationsmatrix orthogonal (als komplexe Matrix aufgefaßt also unitär) ist. Untersuchen Sie, welche Permutationsmatrizen außerdem symmetrisch sind. Gehen Sie dabei von einer Permutation $\sigma \in \mathrm{Sym}_n$ aus, die als Produkt von zyklischen Permutationen gegeben ist, deren einzelne Faktoren nur paarweise verschiedene Elemente bewegen, also

$$\sigma = (j_1\ j_2\ \ldots\ j_{r_1}) \cdot (j_{r_1+1}\ j_{r_1+2}\ \ldots\ j_{r_1+r_2}) \cdots (j_{r_1+\ldots+r_{l-1}+1}\ j_{r_1+\ldots+r_{l-1}+2}\ \ldots\ j_{r_1+\cdots+r_l})$$

mit lauter verschiedenen Ziffern $j_1, \ldots, j_{r_1+\cdots+r_l} \in \{1, \ldots, n\}$. Daß man jede Permutation $\sigma \in \mathrm{Sym}_n$ so schreiben kann, ist übrigens leicht zu sehen, aber nicht unbedingt Teil der Aufgabe.

28.5 Diagonalisieren Sie die Permutationsmatrix der zyklischen Permutation

$$\sigma = (1\ 2\ 3\ \ldots\ n) \in \mathrm{Sym}_n\,.$$

Hinweise: Die Eigenwerte dieser Matrix sind die n-ten Einheitswurzeln (vergleiche Aufgabe 12.4); sie lassen sich mittels der Exponentialfunktion leicht angeben (Skizze!) und sind die Potenzen der speziellen Einheitswurzel

$$\gamma := \exp(2\pi i/n).$$

Wenn Sie anfangen, die Eigenräume nach dem üblichen Verfahren zu berechnen, werden Sie schnell merken, wie der Hase läuft, und den Rest erraten können.

29 Reelle quadratische Formen

Der Titelbegriff ist bei der Erklärung der Skalarprodukte schon beiläufig aufgetaucht (Definition 25.1): Unter einer quadratischen Form auf dem $\mathbb{R}$-Vektorraum V versteht man eine Funktion der Form

$$V \xrightarrow{q} \mathbb{R}; \quad v \mapsto \beta(v,v)$$

mit einer symmetrischen Bilinearform $\beta\colon \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$. Wer die Aufgabe 25.1 studiert hat, weiß, daß die symmetrische Bilinearform β durch q eindeutig bestimmt ist, so daß quadratische Formen und symmetrische Bilinearformen auf V in Wirklichkeit völlig gleichwertige Objekte sind, die man je nach Vorteil oder Interesse gegeneinander austauschen kann.

Ist V mit einer Basis $\underline{v}$ versehen, so entsprechen quadratische Formen q auf V vermöge

$$q\big(\Phi_{\underline{v}}(x)\big) = x^{\mathrm{t}} s\, x$$

den symmetrischen Matrizen $s \in \mathrm{Sym}(n,\mathbb{R})$. Wählen wir eine andere Basis $\underline{w}$, ändert sich die zu q gehörige Matrix in

$$t = u^{\mathrm{t}} s u \in \mathrm{Sym}(n,\mathbb{R}),$$

worin $u = \Phi_{\underline{v}}^{-1} \circ \Phi_{\underline{w}} \in GL(n,\mathbb{R})$ der zu $\underline{v}$ und $\underline{w}$ gehörige Kartenwechsel ist; das hatten wir uns schon damals im Anschluß an Lemma 25.3 überlegt. Nun ist die quadratische Form q natürlich besonders gut zu verstehen, wenn die zugehörige Matrix t eine Diagonalmatrix

$$t = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \in \mathrm{Sym}(n,\mathbb{R})$$

ist; dann ist ja ausgeschrieben

$$q\big(\Phi_{\underline{w}}(x)\big) = \lambda_1 x_1^2 + \cdots + \lambda_n x_n^2$$

(während allgemein auch gemischte Terme $x_j x_k$ mit $j \neq k$ vorkommen). Nachdem wir im vorigen Abschnitt schon so erfolgreich diagonalisiert haben, kann man hoffen, daß das vielleicht auch bei einer quadratischen Form durch geschickte Wahl der Basis $\underline{w}$ immer möglich ist.

Freilich handelt es sich dabei zunächst um eine ganz andere Aufgabe. Interpretieren wir nämlich die symmetrische Matrix s mal versuchsweise als die eines Endomorphismus $f\colon V \longrightarrow V$ bezüglich der Basis $\underline{v}$. Daß f durch die neue Basis $\underline{w}$ diagonalisiert wird, bedeutet dann (in den obigen Bezeichnungen) nicht, daß $u^{\mathrm{t}} s u$, sondern daß $u^{-1} s\, u$ eine Diagonalmatrix ist.

Dieser Unterschied verschwindet aber in dem Moment, wo wir von einem euklidischen Vektorraum V ausgehen und mit Orthonormalbasen arbeiten: die Kartenwechsel zwischen solchen Basen sind ja orthogonale Matrizen $u \in O(n)$, und für die ist u^{t} und u^{-1} dasselbe! Das eröffnet uns die Möglichkeit, den Spektralsatz, der an sich ein Satz über Endomorphismen ist, zur Diagonalisierung einer quadratischen Form zu "mißbrauchen".

29.1 Satz *V sei ein n-dimensionaler euklidischer Vektorraum, q eine quadratische Form auf V. Dann gibt es eine Orthonormalbasis von V, bezüglich der q durch eine Diagonalmatrix*

$$\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \in \mathrm{Sym}(n,\mathbb{R})$$

beschrieben wird.

Beweis Eigentlich habe ich schon alles verraten: Wir wählen zuerst irgendeine Orthonormalbasis $\underline{v}$ von V; bezüglich der wird q durch eine symmetrische Matrix $s \in \mathrm{Sym}(n, \mathbb{R})$ beschrieben. Jetzt interpretieren wir s als die Matrix eines Endomorphismus f von V (bezüglich derselben Basis); f ist dann selbstadjungiert, und nach dem reellen Spektralsatz 28.11 gibt es eine neue Orthonormalbasis $\underline{w}$, die f diagonalisiert. Ausgedrückt mit dem Kartenwechsel $u = \Phi_{\underline{v}}^{-1} \circ \Phi_{\underline{w}} \in O(n)$ heißt das, daß

$$t = u^{-1} s\, u$$

eine Diagonalmatrix ist, und weil wir das ebensogut $t = u^{\mathrm{t}} s\, u$ lesen dürfen, ist damit auch die quadratische Form q diagonalisiert.

Bemerkung Man kann die Matrix von q nicht durch Übergang zu einer anderen orthonormalen Basis weiter vereinfachen, denn die Diagonaleinträge dieser Matrix sind zwangsläufig die mit Vielfachheiten gezählten Eigenwerte der Hilfsabbildung f aus dem Beweis. Diese Koeffizienten sind also jeder auf einem euklidischen Vektorraum definierten quadratischen Form bis auf ihre Reihenfolge eindeutig zugeordnet; man begeht keine große Sünde, wenn man sie mangels eines besseren Namens einfach die *Eigenwerte von* q nennt. Übrigens nennt man den Vorgang, eine quadratische Form im Sinne des Satzes zu diagonalisieren, eine *Hauptachsentransformation*; die rechnerische Durchführung ist mit dem in Beispiel 28.10 erläuterten Verfahren identisch.

29.2 Beispiel Wie Sie wissen und ich in Beispiel 25.8(3) schon einmal erwähnt hatte, wird die Rotationsenergie eines starren Körpers eine quadratische Form

$$\omega \longmapsto T(\omega) = \frac{1}{2}\omega^{\mathrm{t}}\theta\omega$$

auf dem dreidimensionalen Raum der Winkelgeschwindigkeiten gegeben, bezogen auf eine mit dem Körper rotierende Orthonormalbasis. Satz 29.2 erklärt, warum man diese Basis so wählen kann, daß T diagonal wird:

$$T(\omega) = \frac{1}{2}\theta_1\omega_1^2 + \frac{1}{2}\theta_2\omega_2^2 + \frac{1}{2}\theta_3\omega_3^2$$

Die von den Basisvektoren aufgespannten Geraden sind dann die Hauptträgheitsachsen des Körpers (soweit sie definiert sind, nämlich das zugehörige Trägheitsmoment θ_j von den beiden anderen verschieden ist).

Für quadratische Formen, die auf einem gewöhnlichen reellen Vektorraum (ohne euklidische Struktur) leben, erhält man im wesentlichen durch Abschwächung von Satz 29.1 das folgende Ergebnis:

29.3 Satz V sei ein n-dimensionaler reeller Vektorraum, q eine quadratische Form auf V. Dann gibt es natürliche Zahlen $r_\pm$ und eine Basis von V, bezüglich der q durch die Matrix

$$\underbrace{\begin{pmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & -1 & & & & & \\ & & & & \ddots & & & & \\ & & & & & -1 & & & \\ & & & & & & 0 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 0 \end{pmatrix}}_{\in \mathrm{Mat}(n\times n, \mathbb{R})} \qquad \begin{matrix} \overbrace{}^{r_+} & \overbrace{}^{r_-} \end{matrix}$$

beschrieben wird.

Beweis Wir versehen V aushilfsweise mit einem Skalarprodukt, wenden Satz 29.1 an und vergessen das Skalarprodukt wieder. Dann haben wir immerhin eine Basis $\underline{v} = (v_1, \dots, v_n)$, bezüglich der q diagonal ist. Wir ordnen die Basisvektoren so um, daß

$$q(v_j) \begin{cases} > 0 & \text{für } j = 1, \dots, r_+ \\ < 0 & \text{für } j = r_+ + 1, \dots, r_+ + r_- \\ = 0 & \text{für } j > r_+ + r_- \end{cases}$$

ist. Jetzt brauchen wir nur noch v_j für $j \le r_+ + r_-$ durch

$$\begin{array}{ll} \frac{1}{\sqrt{q(v_j)}} v_j & \text{für } 1 \le j \le r_+ \\ \frac{1}{\sqrt{-q(v_j)}} v_j & \text{für } r_+ < j \le r_+ + r_- \end{array}$$

zu ersetzen.

Bemerkung Daß in dieser Situation von den Eigenwerten von q nur noch ihre Vorzeichen bzw. ihr Nullsein übrig bleiben, steht nicht im Widerspruch zu dem, was ich vorhin zur Hauptachsentransformation gesagt habe: Dort war q zusätzlich zu einer schon vorhandenen Struktur gegeben, nämlich dem euklidischen Skalarprodukt, auf das die Basiswahl Rücksicht nehmen muß. Das ist bei Satz 29.3 nicht so; diesen Satz zum Beispiel auf den Trägheitstensor eines starren Körpers anzuwenden, würde zwar die Hauptträgheitsmomente alle zu eins machen, liefe aber darauf hinaus, den Körper dabei in Richtung der Hauptachsen zu strecken oder zu stauchen!

Wie Sie aufgrund der Formulierung von Satz 29.3 vermuten werden, ist aber immerhin die "Anzahl der positiven, negativen und verschwindenden (obwohl selbst gar nicht wohldefinierten) Eigenwerte" jeweils eine der quadratischen Form q zugeordnete Invariante:

29.4 Satz und **Definition** Die in Satz 29.3 eingeführte Zahl r_+ ist die größtmögliche Dimension eines linearen Teilraums $P \subset V$ mit der Eigenschaft, daß die eingeschränkte quadratische Form $q|P$ positiv definit ist; man nennt solche Teilräume kurz positiv definit bezüglich q. Analog ist r_- die größtmögliche Dimension eines negativ definiten (d.h. bezüglich $-q$ positiv definiten) Teilraums von V. Insbesondere ist das Zahlenpaar (r_+, r_-) von der Wahl der diagonalisierenden Basis unabhängig; wir nennen es die Signatur von q.

Beweis Wenn wir eine Basis $\underline{v}$ nach Satz 29.3 wählen, ist

$$q\big(\Phi_{\underline{v}}(x)\big) = \sum_{j=1}^{r_+} x_j^2 \; - \sum_{j=r_+ + 1}^{r_+ + r_-} x_j^2,$$

und man sieht sofort die beiden Unterräume

$$\Phi_{\underline{v}}\big(\mathbb{R}^{r_+} \times \{0\} \times \{0\}\big) \quad \text{und} \quad \Phi_{\underline{v}}\big(\{0\} \times \mathbb{R}^{r_-} \times \{0\}\big)$$

von V, auf denen q positiv bzw. negativ definit ist. Sei nun $P \subset V$ ein beliebiger Teilraum, auf dem q positiv definit ist. Wir müssen noch zeigen, daß $\dim P \le r_+$ ist. Dazu betrachten wir den $(n - r_+)$-dimensionalen Hilfsraum

$$N := \Phi_{\underline{v}}\big(\{0\} \times \mathbb{R}^{r_-} \times \mathbb{R}^{n - r_+ - r_-}\big) \subset V.$$

Auf N ist q zwar nicht unbedingt negativ definit, aber immerhin gilt

$$q\big(\Phi_{\underline{v}}(x)\big) = - \sum_{j=r_+ + 1}^{r_+ + r_-} x_j^2 \le 0 \quad \text{für alle } \Phi_{\underline{v}}(x) \in N.$$

Insbesondere ist also $P \cap N = \{0\}$; das ist nach der Dimensionsformel für Unterräume nur möglich, wenn $\dim P + \dim N \le n$, d.h. wenn $\dim P \le r_+$ ist.

Natürlich gilt für r_- die entsprechende Überlegung.

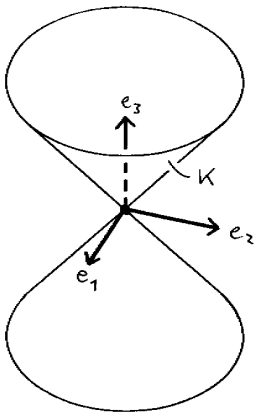
Bemerkungen Der Satz wird oft nach dem Mathematiker Sylvester benannt. — Als Signatur wird häufig nur die Differenz $r_+ - r_-$ bezeichnet, diese auch als *Index*, womit manche aber wieder das r_- meinen: wichtig ist nur, daß jede dieser Zahlen zusammen mit dem Rang $r_+ + r_-$, den man getrost als den Rang von q bezeichnen darf, dieselbe Information wie (r_+, r_-) enthält. — Satz 29.4 ist subtiler als Sie jetzt vielleicht denken; der darin die entscheidende Rolle spielende positiv definite Unterraum P größtmöglicher Dimension ist nämlich keineswegs eindeutig bestimmt (außer wenn q postiv oder negativ definit ist):

29.5 Beispiel Die auf $\mathbb{R}^3$ durch

$$q(x) = x_1^2 + x_2^2 - x_3^2$$

gegebene quadratische Form ist schon diagonal; sie hat offensichtlich die Signatur $(2, 1)$. Die Skizze zeigt den Kegel

$$K = \{x \in \mathbb{R}^3 \mid q(x) = 0\} = \{x \in \mathbb{R}^3 \mid x_1^2 + x_2^2 = x_3^2\} :$$



Während die positiv definiten Ebenen genau diejenigen sind, die K nur im Nullpunkt treffen, sind die negativ definiten Geraden diejenigen, die durch die beiden von K umschlossenen "inneren" Raumgebiete laufen.

Wenn man die Signatur einer quadratischen Form, oder auch eine diagonalisierende Basis im Sinne von Satz 29.3 praktisch berechnen möchte, braucht man nicht die verhältnismäßig aufwendige Hauptachsentransformation durchzuführen. Vielmehr genügt eine leicht modifizierte Form des Orthonormalisierungsverfahrens nach Gram-Schmidt, das man hier am besten gleich an der die Form q beschreibenden symmetrischen Matrix s durchführt. Der Ablauf entspricht fast genau dem des Gaußschen Algorithmus, nur daß jede elementare Zeilenumformung $s \mapsto u^{\mathrm{t}} s$ sofort durch die entsprechende Spaltenumformung ergänzt wird, so daß die symmetrische Matrix $u^{\mathrm{t}} s\, u$, die dieser Doppelschritt hervorbringt, in der Tat q bezüglich einer neuen Basis beschreibt. Statt einer genauen Darstellung des Verfahrens ein kommentiertes Beispiel, in dem alle zu erklärenden Situationen auftreten:

29.6 Beispiel $q\colon \mathbb{R}^3 \longrightarrow \mathbb{R}$ sei durch die Matrix

$$\begin{pmatrix} -4 & 2 & 2 \\ 2 & -1 & 0 \\ 2 & 0 & -1 \end{pmatrix} \in \mathrm{Sym}(3, \mathbb{R})$$

gegeben. Aus der Tatsache, daß alle Diagonalelemente negativ sind, kann man nicht etwa schließen, daß q negativ definit wäre; davor sei hier ausdrücklich gewarnt. (Die nicht definite Form in Beispiel 29.5 hat diese Eigenschaft ja auch, sobald man alle drei Basisvektoren im Inneren des Kegels wählt.)

Nach dem Gaußschen Algorithmus würde man zuerst die erste Zeile durch -4 teilen. Hier müßte man aber anschließend auch die erste Spalte durch -4 teilen: also teilt man von vornherein nicht durch -4, sondern durch 2 (mit -2 ginge es auch, aber nicht besser):

$$\begin{pmatrix} -4 & 2 & 2 \\ 2 & -1 & 0 \\ 2 & 0 & -1 \end{pmatrix} \xrightarrow{(d_{1,\frac{1}{2}})\cdot} \begin{pmatrix} -2 & 1 & 1 \\ 2 & -1 & 0 \\ 2 & 0 & -1 \end{pmatrix} \xrightarrow{\cdot(d_{1,\frac{1}{2}})} \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}$$

Jetzt räumt man ganz normal die erste Spalte auf und — gemäß der Regel — sofort anschließend die erste Zeile:

$$\begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix} \xrightarrow{(u_{31,1}u_{21,1})\cdot} \begin{pmatrix} -1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \xrightarrow{\cdot(u_{12,1}u_{13,1})} \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

Die Null in der Position 22 würde man nach dem Gaußschen Algorithmus jetzt gegen die darunterstehende 1 wegtauschen. Das führt hier nicht weiter, weil die 1 bei der nachfolgenden Spaltenvertauschung ja wieder wegwandern würde. Wir befreien uns aus dieser Sackgasse, indem wir zur zweiten Zeile die dritte addieren (und die Matrix damit vorübergehend wieder komplizierter machen):

$$\begin{pmatrix} -1 & & \\ & 0 & 1 \\ & 1 & 0 \end{pmatrix} \xrightarrow{(u_{23,1})\cdot} \begin{pmatrix} -1 & & \\ & 1 & 1 \\ & 1 & 0 \end{pmatrix} \xrightarrow{\cdot(u_{32,1})} \begin{pmatrix} -1 & & \\ & 2 & 1 \\ & 1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} -1 & & \\ & 2 & 1 \\ & 1 & 0 \end{pmatrix} \xrightarrow{(d_{2,\frac{1}{2}\sqrt{2}})\cdot} \begin{pmatrix} -1 & & \\ & \sqrt{2} & \frac{1}{2}\sqrt{2} \\ & 1 & 0 \end{pmatrix} \xrightarrow{\cdot(d_{2,\frac{1}{2}\sqrt{2}})} \begin{pmatrix} -1 & & \\ & 1 & \frac{1}{2}\sqrt{2} \\ & \frac{1}{2}\sqrt{2} & 0 \end{pmatrix}$$

$$\begin{pmatrix} -1 & & \\ & 1 & \frac{1}{2}\sqrt{2} \\ & \frac{1}{2}\sqrt{2} & 0 \end{pmatrix} \xrightarrow{(u_{32,-\frac{1}{2}\sqrt{2}})\cdot} \begin{pmatrix} -1 & & \\ & 1 & \frac{1}{2}\sqrt{2} \\ & & -\frac{1}{2} \end{pmatrix} \xrightarrow{\cdot(u_{23,-\frac{1}{2}\sqrt{2}})} \begin{pmatrix} -1 & & \\ & 1 & 0 \\ & & -\frac{1}{2} \end{pmatrix}$$

Die Signatur von q ist also $(1,2)$, und wer über die angewendeten Elementarumformungen Buch geführt hat, bekommt auch die diagonalisierende Matrix

$$\begin{aligned} u &= d_{1,\frac{1}{2}} \cdot u_{12,1}u_{13,1} \cdot u_{32,1} \cdot d_{2,\frac{1}{2}\sqrt{2}} \cdot u_{23,-\frac{1}{2}\sqrt{2}} \\ &= \begin{pmatrix} \frac{1}{2} & & \\ & 1 & \\ & & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ & 1 & \\ & & 1 \end{pmatrix} \begin{pmatrix} 1 & & \\ & 1 & \\ & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & & \\ & \frac{1}{2}\sqrt{2} & \\ & & 1 \end{pmatrix} \begin{pmatrix} 1 & & \\ & 1 & -\frac{1}{2}\sqrt{2} \\ & & 1 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{2} & \frac{1}{2}\sqrt{2} & 0 \\ 0 & \frac{1}{2}\sqrt{2} & -\frac{1}{2} \\ 0 & \frac{1}{2}\sqrt{2} & \frac{1}{2} \end{pmatrix} \end{aligned}$$

in $GL(3,\mathbb{R})$.

Zum Schluß möchte ich noch ein wenig über quadratische Formen der Signatur $(1,n)$ auf einem $(n+1)$-dimensionalen Vektorraum plaudern, denn sie zählen (für $n=3$) zu den mathematischen Grundlagen der Relativitätstheorie. Wie Sie wissen, verschmelzen in der Relativitätstheorie Raum und Zeit zu einer vierdimensionalen Raum-Zeit-Welt. Die in dieser Welt verteilten Massen prägen ihr eine Metrik auf, die wegen der ungleichen Massenverteilung von Punkt zu Punkt variiert, in kleinen und von Massenkonzentrationen genügend entfernten Weltgebieten aber als konstant angesehen werden kann (spezielle Relativitätstheorie). In einem solchen Gebiet läßt sich die Welt als ein vierdimensionaler Minkowski-Raum im Sinne der folgenden Definition auffassen — jedenfalls wenn man willkürlich einen Weltpunkt zum Nullpunkt deklariert hat (ein schon zu Beginn des Abschnitts 17 besprochener Schönheitsfehler).

29.7 Definition V sei ein $(n+1)$-dimensionaler reeller Vektorraum. Ein minkowskisches Skalarprodukt auf V ist eine symmetrische Bilinearform $\langle\,,\rangle$ der Signatur $(1,n)$ auf V; ein solches Skalarprodukt macht V zu einem Minkowski-Raum.

Die Minkowski-Räume sind also nach den euklidischen und den unitären Vektorräumen eine dritte Variante von Vektorräumen mit skalarem Produkt als zusätzlicher Struktur. Nach Satz 29.3 kann man in V immer

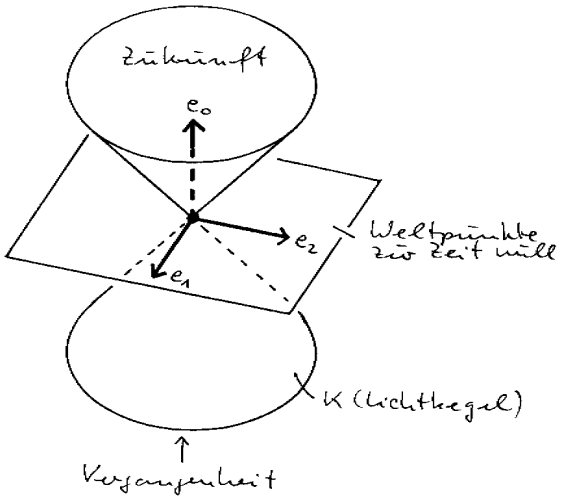
eine Basis wählen, so daß

$$s = \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & \ddots & \\ & & & -1 \end{pmatrix} \in \mathrm{Sym}(n+1, \mathbb{R})$$

die zu $\langle\,,\rangle$ gehörige symmetrische Matrix ist; d.h. man kann sich stets auf den Standardraum $\mathbb{R}^{n+1}$ mit dem Skalarprodukt

$$\langle x, y\rangle = x_0 y_0 - \sum_{j=1}^{n} x_j y_j, \quad \text{insbesondere } \langle x, x\rangle = x_0^2 - \sum_{j=1}^{n} x_j^2$$

zurückziehen (es ist praktisch und üblich, die Spaltenkomponenten hier von 0 an zu zählen). Beachten Sie, daß ich $\langle x, x\rangle$ jetzt nicht mehr als Quadrat einer Norm schreiben kann, denn diese Zahl kann negativ ausfallen. In der physikalischen Interpretation ist $n = 3$ und zunächst $x_0 = ct$, während x_1, x_2, x_3 die gewöhnlichen Raumkoordinaten sind. Für den Fall $n = 2$, den man sich natürlich leichter veranschaulicht, können wir wieder das Bild aus Beispiel 29.5 heranziehen, in dem nur die Bezeichnung der Achsen zu ändern ist.



In einem Minkowski-Raum tritt an die Stelle des Längen- oder Normbegriffs, genauer die Stelle dessen Quadrats, eben die quadratische Form $v \mapsto \langle v, v\rangle$, die Werte beiderlei Vorzeichens annimmt. Damit gehen natürlich die üblichen Eigenschaften einer Norm verloren, insbesondere die Dreiecksungleichung und damit die Tauglichkeit zur Abstandsmessung im gewöhnlichen Sinn. Trotzdem hat die Minkowski-Form in der physikalischen Situation eine handfeste physikalische Bedeutung, die ich in die Skizze schon eingetragen habe: Die Punkte von K, also die, auf denen die Form verschwindet, sind genau diejenigen, die auf einem durch den Nullpunkt gehenden Lichtstrahl liegen; man nennt K deshalb auch den *Lichtkegel* der Minkowski-Form. Der übrigen Weltpunkte gibt es aus physikalischer Sicht zwei, genauer drei Sorten: von 0 aus gesehen *zeitartige* Punkte mit positivem Wert der Form, die innerhalb des Kegels K liegen und gewissermaßen Zukunft und Vergangenheit von 0 bilden, und andererseits von 0 aus gesehen *raumartige* Punkte, auf denen die Minkowski-Form einen negativen Wert hat: sie füllen das ringförmige Gebiet um K. Ereignisse in diesen Punkten können wegen der Endlichkeit der Lichtgeschwindigkeit mit keinem Ereignis in 0 in einem Wirkungszusammenhang stehen. Für einen beliebigen Weltpunkt x ist $\langle x, x\rangle$ ein Maß dafür, um wieviel ein Lichtstrahl "zu schnell" ist, um 0 und x zu verbinden.

Alle Punkte auf der x_0-, also der Zeitachse erscheinen in dem gewählten Koordinatensystem am gleichen Ort, die Punkte der dazu senkrechten Ebene

$$\{e_0\}^\perp = \{x \in \mathbb{R}^{n+1} \mid x_0 = 0\} = \{0\} \times \mathbb{R}^n$$

dagegen als gleichzeitig, aber an verschiedenen Orten. Die putzigen Effekte der Relativitätstheorie beruhen darauf, daß die Begriffe "gleichzeitig" und "an der gleichen Stelle" keinen absoluten Sinn haben, sondern von der willkürlichen Wahl eines Koordinatensystems abhängen. Der Übergang zu einem anderen System

entspricht mathematisch einem bezüglich der Minkowski-Form orthogonalen Kartenwechsel, einer Transformation mit einer Matrix u aus der verallgemeinerten orthogonalen Gruppe

$$O(1,n) = \left\{ u \in \mathrm{Mat}(n \times n, \mathbb{R}) \,\middle|\, u^{\mathrm{t}} s\, u = s \right\}.$$

29.8 Beispiel Wenn wir den Einheitszeitvektor $e_0 \in \mathbb{R}^3$ durch eine Matrix $u \in O(1,2)$ auf den Vektor

$$u_0 = \frac{1}{3} \begin{pmatrix} 5 \\ 0 \\ 4 \end{pmatrix}$$

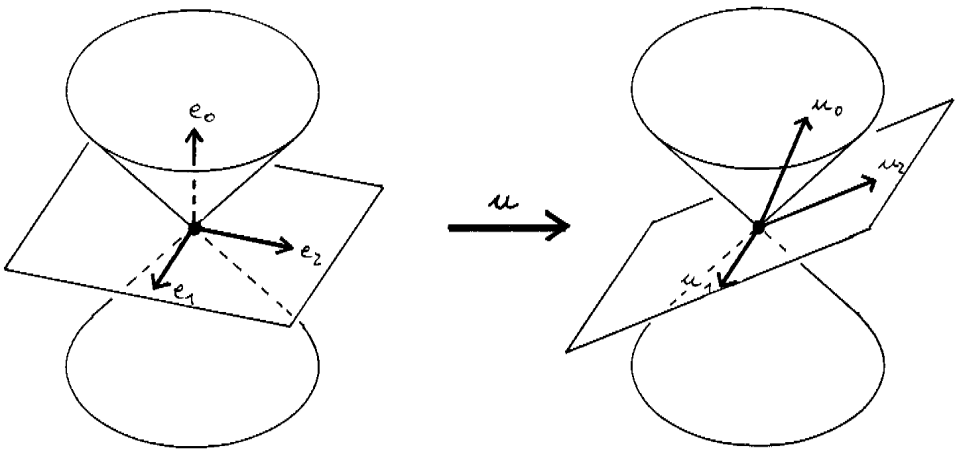
abbilden wollen, der ebenfalls das "Längenquadrat" $\langle u_0, u_0 \rangle = 1$ hat, müssen wir $\{0\} \times \mathbb{R}^2$ auf das orthogonale Komplement

$$\{u_0\}^\perp = \mathrm{Kern} \begin{pmatrix} 5 & 0 & -4 \end{pmatrix} = \mathrm{Lin} \left(\begin{pmatrix} 4 \\ 0 \\ 5 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right)$$

bewegen. Eine offensichtliche Wahl für u ist

$$u = \begin{pmatrix} \frac{5}{3} & 0 & \frac{4}{3} \\ 0 & 1 & 0 \\ \frac{4}{3} & 0 & \frac{5}{3} \end{pmatrix} \in O(1,2).$$

Diese Transformation entspricht physikalisch dem Übergang zu einem Bezugssystem, das sich mit vier Fünfteln der Lichtgeschwindigkeit in Richtung der x_2-Achse bewegt.



Übungsaufgaben

29.1 Berechnen Sie die Signatur der durch

$$q(x) = 2\left(x_1^2 - x_1 x_2 + x_2^2 - x_2 x_3 + x_3^2 - \cdots + x_{n-1}^2 - x_{n-1} x_n + x_n^2\right)$$

gegebenen quadratischen Form q auf $\mathbb{R}^n$ nach dem in Beispiel 29.6 beschriebenen Verfahren.

29.2 Um eine symmetrische Matrix auf Definitheit zu testen, ist vor allem bei größerer Dimension die folgende Regel praktisch:

$$s \in \mathrm{Sym}(n, \mathbb{R}) \text{ ist genau dann positiv definit, wenn } \det \Big(s_{ij} \Big)_{i,j=1}^{k} > 0 \text{ für jedes } k \in \{1, \dots, n\} \text{ ist.}$$

Beweisen Sie die Regel und testen Sie damit einige Matrizen. Tip zum Beweis: Sei V ein n-dimensionaler $\mathbb{R}$-Vektorraum und $U \subset V$ ein $(n-1)$-dimensionaler Unterraum. Welche Signaturen kann eine quadratische Form $q: V \longrightarrow \mathbb{R}$ überhaupt haben, wenn man schon weiß, daß ihre Einschränkung auf U positiv definit ist?

29.3 Wenn man von der Matrix s der Aufgabe 29.2 nur

$$\det \Big(s_{ij} \Big)_{i,j=1}^{k} \neq 0 \quad \text{für jedes } k \in \{1, \dots, n\}$$

voraussetzt, nach welcher Regel kann man dann die Signatur der durch s bestimmen quadratischen Form ablesen?

29.4 V sei ein $(n+1)$-dimensionaler Minkowski-Raum, und $0 \neq v \in V$. Untersuchen Sie, inwieweit die folgenden Aussagen wahr sind:

(a) $\{v\}^{\perp}$ ist ein n-dimensionaler linearer Teilraum von V.

(b) $\{v\}^{\perp}$ ist ein Komplement von $\mathrm{Lin}(v)$ in V.

(c) $\{v\}^{\perp}$ wird durch die Einschränkung des minkowskischen Skalarproduktes selbst zu einem Minkowski-Raum der Dimension n.

29.5 V sei ein $(n+1)$-dimensionaler Minkowski-Raum. Beweisen Sie: Ist $U \subset V$ ein linearer Unterraum mit der Eigenschaft

$$\langle v, v \rangle = 0 \quad \text{für alle } v \in U,$$

so ist $\dim U \leq 1$.

30 Stetige Funktionen in mehreren Variablen

Mit diesem Abschnitt kehren wir in die Analysis zurück. Unsere erste Aufgabe besteht darin, die analytischen Grundbegriffe, die wir im Wintersemester studiert haben, vom Ein- ins Mehrdimensionale zu übertragen.

30.1 Definition Seien n und p natürliche Zahlen, $X \subset \mathbb{R}^n$ und $Y \subset \mathbb{R}^p$ Teilmengen, $f\colon X \longrightarrow Y$ eine Abbildung und $a \in X$ ein Punkt. f heißt bei a stetig, wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ mit

$$|f(x) - f(a)| < \varepsilon \quad \text{für alle } x \in X \text{ mit } |x - a| < \delta$$

gibt.

Die Definition unterscheidet sich von der früheren (7.1) nur durch die Bedeutung der Betragstriche, die jetzt für die euklidische Standardnorm

$$|x| = \sqrt{\sum_{j=1}^{n} x_j^2}$$

auf $\mathbb{R}^n$ stehen. 30.1 birgt auch deswegen keinerlei Überraschungen, weil wir den Fall $n = 2$ und $p \leq 2$ de facto schon als die entsprechende Definition 10.8 für Funktionen einer *komplexen* Variablen kennen: für diesen Zweck ist ja $\mathbb{C} = \mathbb{R}^2$. Ganz genauso verhält es sich nicht nur mit den anderen hier eigentlich zu erklärenden Begriffen (insbesondere Konvergenz von Folgen, Reihen und Funktionen), sondern auch mit deren Eigenschaften: alles zu diesem Thema, was schon den Schritt von $\mathbb{R}$ zu $\mathbb{C}$ überlebt hat, überlebt auch den Schritt zu $\mathbb{R}^n$. Ich begnüge mich deshalb mit einigen zusätzlichen Anmerkungen.

Wie im Eindimensionalen spielt die Zielmenge $Y \subset \mathbb{R}^p$ der Abbildung f für deren Stetigkeitseigenschaften keine Rolle; man erlaubt sich daher wie früher, f und die Komposition $X \xrightarrow{f} Y \hookrightarrow \mathbb{R}^p$ schon mal als dasselbe Objekt anzusehen. — Zu den Regeln für stetige Funktionen und Limites trägt neu noch die Tatsache bei, daß auch die Vektorraumoperationen ebenso wie das Skalarprodukt stetige Funktionen

$$\mathbb{R}^{2n} = \mathbb{R}^n \times \mathbb{R}^n \xrightarrow{+} \mathbb{R}^n, \quad \mathbb{R}^{n+1} = \mathbb{R} \times \mathbb{R}^n \xrightarrow{\cdot} \mathbb{R}^n \quad \text{bzw.} \quad \mathbb{R}^{2n} = \mathbb{R}^n \times \mathbb{R}^n \xrightarrow{\langle,\rangle} \mathbb{R}$$

sind. — Analog zu Lemma 10.7 gilt, daß eine Abbildung $f\colon X \longrightarrow \mathbb{R}^p$ genau dann stetig ist, wenn die einzelnen Komponentenfunktionen

$$f_j\colon X \longrightarrow \mathbb{R} \quad \text{mit} \quad f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_p(x) \end{pmatrix}$$

das sind. Die Stetigkeit einer Funktion *von* mehreren Variablen kann man dagegen nicht auf eine eindimensionale Frage zurückführen:

30.2 Beispiel Die durch

$$f(x,y) := \begin{cases} 0 & \text{für } x = y = 0 \\ \frac{xy}{x^2+y^2} & \text{sonst} \end{cases}$$

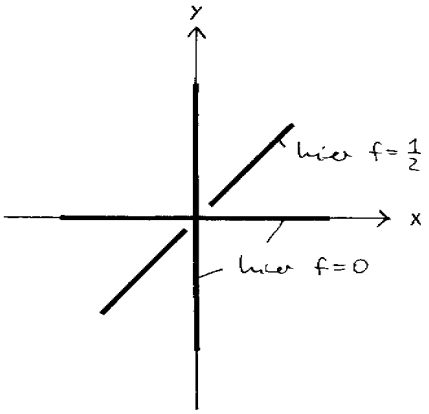
definierte Funktion $f\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ ist im Nullpunkt unstetig, obwohl für beliebige feste $a, b \in \mathbb{R}$ die Funktionen

$$\mathbb{R} \ni x \longmapsto f(x,b) \quad \text{und} \quad \mathbb{R} \ni y \longmapsto f(a,y)$$

überall stetig sind. Das ist nur im Nullpunkt eine interessante Frage, und dort gilt in der Tat etwa

$$\lim_{t\to 0} f(t,t) = \lim_{t\to 0} \frac{t^2}{t^2+t^2} = \frac{1}{2} \neq 0 = f(0,0),$$

während die Funktionen $x \mapsto f(x,0)$ und $y \mapsto f(0,y)$ konstant null und deshalb trivialerweise stetig sind. Ein Blick auf die Skizze



macht das Phänomen verständlich: Die Stetigkeit im Nullpunkt ist eine Forderung an das Verhalten der Funkionswerte bei beliebiger Annäherung an 0, nicht nur der Werte auf dem Achsenkreuz. — Statt $f(x,y)$ hätte ich gemäß den in der linearen Algebra getroffenen Vereinbarungen

$$f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) \quad \text{oder zumindest} \quad f\begin{pmatrix} x \\ y \end{pmatrix}$$

schreiben sollen; diese Mühe macht man sich in der Analysis nur, wenn es sich im Zusammenhang mit dem Matrizenkalkül auch lohnt.

Der eigentliche Unterschied zwischen der eindimensionalen und der (mehrdimensionalen) Vektoranalysis liegt in der Vielfalt der in Betracht zu ziehenden Teilmengen von $\mathbb{R}^n$. Während man da für $n = 1$ nicht viel verpaßt, wenn man Teilmengen, die keine Intervalle sind, kurzerhand für pathologisch erklärt, wäre es im Mehrdimensionalen doch eine störende Einschränkung, würde man als Definitionsbereiche von Funktionen analogerweise etwa nur Quader, also Produkte von (in der Regel beschränkten) Intervallen

$$Q = I_1 \times I_2 \times \cdots \times I_n \subset \mathbb{R}^n$$

zulassen. Wir beschäftigen uns deshalb jetzt genauer mit Eigenschaften von Teilmengen von $\mathbb{R}^n$. Dazu verallgemeinern wir die frühere Definition 10.6:

30.3 Definition Für $a \in \mathbb{R}^n$ und $\varepsilon > 0$ heißt die Menge

$$U_\varepsilon(a) := \left\{x \in \mathbb{R}^n \,\middle|\, |x-a| < \varepsilon\right\}$$

die (n-dimensionale) offene Kugel um a vom Radius ε, und

$$D_\varepsilon(a) := \left\{x \in \mathbb{R}^n \,\middle|\, |x-a| \leq \varepsilon\right\}$$

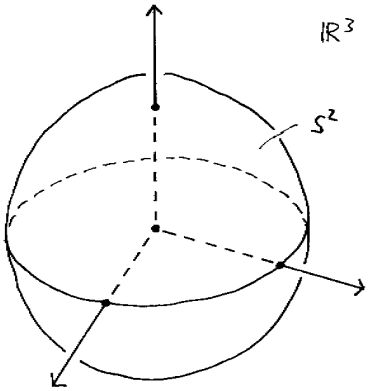
die abgeschlossene Kugel. Die Differenz

$$S_\varepsilon(a) := \left\{x \in \mathbb{R}^n \,\middle|\, |x-a| = \varepsilon\right\}$$

nennt man eine $(n-1)$-dimensionale Sphäre. Im Standardfall $a = 0$ und $\varepsilon = 1$ spricht man von *Einheitskugeln* und *-sphären* und notiert statt dieser Daten die "Dimension":

$$U^n,\ D^n,\ S^{n-1} \subset \mathbb{R}^n$$

Beachten Sie, daß demnach $S^1 \subset \mathbb{R}^2 = \mathbb{C}$ die früher oft mit S bezeichnete Kreislinie ist, daß $U^1 = (-1, 1)$ und $D^1 = [-1, 1]$ Intervalle sind und auch $S^0 = \{\pm 1\}$, $U^0 = D^0 = \{0\}$ und sogar $S^{-1} = \emptyset$ noch definiert sind.



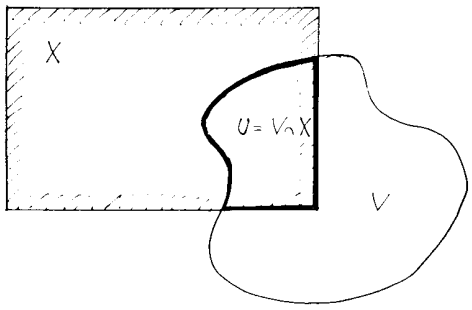
Natürlich sind auch Kugeln nur sehr spezielle Verallgemeinerungen von Intervallen; mit ihrer Hilfe führen wir aber jetzt viel größere Klassen von Mengen ein (auch dazu hatten wir den Anfang schon in der Definition 10.9 gemacht):

30.4 Definition (a) Eine Teilmenge $U \subset \mathbb{R}^n$ heißt offen, wenn es zu jedem $a \in U$ ein $\delta > 0$ mit $U_\delta(a) \subset U$ gibt. Eine Teilmenge $F \subset \mathbb{R}^n$ heißt abgeschlossen, wenn ihr Komplement $\mathbb{R}^n \setminus F$ offen ist.

(b) Wir fixieren eine beliebige Teilmenge $X \subset \mathbb{R}^n$. Eine Teilmenge $U \subset X$ heißt dann (relativ) offen in X, wenn es eine offene Menge $V \subset \mathbb{R}^n$ mit

$$U = V \cap X$$

gibt.



Entsprechend heißt eine Teilmenge $F \subset X$ (relativ) abgeschlossen in X, wenn eine abgeschlossene Menge $G \subset \mathbb{R}^n$ mit

$$F = G \cap X$$

existiert.

Bemerkungen Der Unterschied zwischen den absoluten Begriffen (a) und den relativen (b) wird sprachlich nicht immer klar zum Ausdruck gebracht. Die zu empfehlende Formulierung "... sei U offen *in* X..." ist eindeutig, während man "... sei $U \subset X$ offen..." besser vermeidet, denn dabei könnte man ebensogut an eine in X enthaltene, im absoluten Sinne offene Teilmenge von $\mathbb{R}^n$ denken. Übrigens umfassen die relativen Begriffe die absoluten als Spezialfall, denn "offen/abgeschlossen in $\mathbb{R}^n$" ist dasselbe wie Offen-/Abgeschlossenheit überhaupt. — Man sieht auch sofort, daß die in X abgeschlossenen Teilmengen genau die (in X gebildeten) Komplemente der in X offenen Teilmengen sind. — Aus der Dreiecksungleichung folgt wie in Beispiel 10.10

$$U_{\varepsilon - |x-a|}(x) \subset U_\varepsilon(a) \quad \text{für jedes } x \in U_\varepsilon(a),$$

offene Kugeln sind deshalb tatsächlich offen (zu den abgeschlossenen kommen wir in Kürze).

Die formalen Eigenschaften dieser Begriffe beschreibt das

30.5 Lemma $X \subset \mathbb{R}^n$ sei eine Teilmenge. Dann gilt:

(a) Die Mengen $\emptyset$ und X selbst sind sowohl relativ offen als auch abgeschlossen in X.

(b) Der Durchschnitt endlich vieler und die Vereinigung beliebig vieler in X offener Teilmengen ist wieder offen in X.

(c) Umgekehrt sind endliche Vereinigungen und beliebige Durchschnitte in X abgeschlossener Teilmengen in X abgeschlossen.

Beweis Wir behandeln erst den absoluten Fall $X = \mathbb{R}^n$. Dann ist (a) klar. Zum Beweis von (b) sei $(U_\lambda)_{\lambda\in\Lambda}$ eine Familie offener Teilmengen von $\mathbb{R}^n$. Wir betrachten einen beliebigen Punkt in der Vereinigung

$$a \in \bigcup_{\lambda\in\Lambda} U_\lambda.$$

Etwa sei $a \in U_\alpha$; weil U_α offen ist, finden wir ein $\delta > 0$ mit $U_\delta(a) \subset U_\alpha$. Dann gilt erst recht

$$U_\delta(a) \subset U_\alpha \subset \bigcup_{\lambda\in\Lambda} U_\lambda,$$

und die Offenheit der Vereinigung ist gezeigt.

Sei nun Λ endlich und

$$a \in \bigcap_{\lambda\in\Lambda} U_\lambda$$

ein Punkt des Durchschnitts. Weil alle U_λ offen sind, finden wir zu jedem λ ein $\delta > 0$, so daß

$$U_\delta(a) \subset U_\lambda \quad \text{für alle } \lambda \in \Lambda$$

gilt — da wir diese endlich vielen $\delta > 0$ gleich durch das kleinste unter ihnen ersetzen dürfen, ist es gerechtfertigt, die Abhängigkeit von λ gar nicht erst zu notieren. Aus

$$U_\delta(a) \subset \bigcap_{\lambda\in\Lambda} U_\lambda$$

schließt man jetzt auf die Offenheit des Durchschnitts.

Die absolute Version von (c) ergibt sich sofort durch Komplementbildung, und die relative Version von all dem erhält man, indem man die beteiligten Mengen mit $X \subset \mathbb{R}^n$ schneidet.

30.5½ Folgerung Sei $X \subset \mathbb{R}^n$ eine beliebige Teilmenge. Jede in X enthaltene offene Teilmenge von $\mathbb{R}^n$ ist auch offen in X. Wenn X selbst offen ist, ist umgekehrt jede relativ offene Teilmenge von X auch in $\mathbb{R}^n$ offen. Entsprechendes gilt für Abgeschlossenheit.

Beweis $U \subset X$ impliziert natürlich $U = U \cap X$, daher die erste Beobachtung. Wenn nun neben $V \subset \mathbb{R}^n$ auch X offen ist, dann ist nach dem Lemma auch $V \cap X$ offen.

Die neuen Begriffe ermöglichen es, die Stetigkeit einer Abbildung direkt global zu beschreiben, also ohne sie durch Stetigkeit in jedem Punkt auszudrücken:

30.6 Satz $X \subset \mathbb{R}^n$ und $Y \subset \mathbb{R}^p$ seien beliebige Teilmengen, $f\colon X \longrightarrow Y$ eine Abbildung. Dann sind äquivalent:

(a) f ist stetig;

(b) für jede in Y offene Menge $V \subset Y$ ist $f^{-1}(V)$ offen in X;

(c) für jede in Y abgeschlossene Menge $G \subset Y$ ist $f^{-1}(G)$ abgeschlossen in X.

Beweis Wir setzen (a) voraus: f sei also stetig. Zum Nachweis von (b) sei $V \subset Y$ relativ offen, etwa $V = W \cap Y$ mit offenem $W \subset \mathbb{R}^p$; wir wollen zeigen, daß $f^{-1}(V) = f^{-1}(W)$ offen in X ist. Für jedes $a \in f^{-1}(W)$ ist $f(a) \in W$, und wir finden ein $\varepsilon(a) > 0$ mit $U_{\varepsilon(a)}\big(f(a)\big) \subset W$. Weil f bei a stetig ist, finden wir weiter ein $\delta(a) > 0$ mit

$$f\big(U_{\delta(a)}(a) \cap X\big) \subset U_{\varepsilon(a)}\big(f(a)\big),$$

wie gesagt für jedes $a \in f^{-1}(W)$. Nach Lemma 30.4 ist die Menge

$$U := \bigcup_{a \in f^{-1}(W)} U_{\delta(a)}(a) \subset \mathbb{R}^n$$

offen, andererseits gilt neben $f^{-1}(W) \subset U \cap X$ auch

$$f(U \cap X) = \bigcup_{a \in f^{-1}(W)} f\big(U_{\delta(a)}(a) \cap X\big) \subset \bigcup_{a \in f^{-1}(W)} U_{\varepsilon(a)}\big(f(a)\big) \subset W$$

und damit

$$U \cap X = f^{-1}(W).$$

Also ist $f^{-1}(W) \subset X$ relativ offen.

Zum Beweis der umgekehrten Richtung sei (b) vorausgesetzt; wir beweisen, daß f dann an jeder Stelle $a \in X$ stetig ist. Nun, für gegebenes $\varepsilon > 0$ ist $W := U_{\varepsilon(a)}\big(f(a)\big) \subset \mathbb{R}^p$ eine offene Menge, also $f^{-1}(W) \subset X$ relativ offen, etwa $f^{-1}(W) = U \cap X$ mit offenem U. Natürlich ist dann $a \in U$, und wir wählen $\delta > 0$ so klein, daß $U_\delta(a) \subset U$ gilt. Es folgt

$$f\big(U_\delta(a) \cap X\big) \subset f(U \cap X) \subset W = U_{\varepsilon(a)}\big(f(a)\big),$$

und damit die Stetigkeit von f bei a bewiesen.

Die Äquivalenz zwischen (b) und (c) erweist sich durch Komplementbildung als reine Formalität.

Der Satz ist eine bequeme und unerschöpfliche Quelle offener und abgeschlossener Mengen:

30.7 Beispiele (1) Unter den Intervallen $I \subset \mathbb{R}$ sind die offenen und die abgeschlossenen genau diejenigen, die wir schon früher so bezeichnet hatten; das sieht man sofort ein.

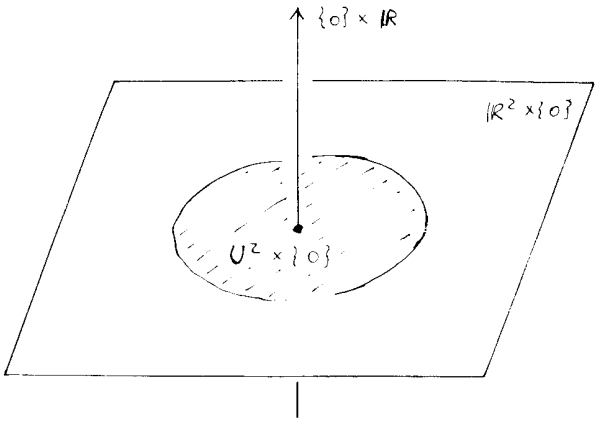
(2) Die abgeschlossenen Kugeln und die Sphären in $\mathbb{R}^n$ sind abgeschlossene Teilmengen, denn wenn man die stetige Funktion $\mathbb{R}^n \ni x \longmapsto |x-a| \in \mathbb{R}$ mit f bezeichnet, dann sind

$$D_\varepsilon(a) = f^{-1}(-\infty, \varepsilon] \quad \text{und} \quad S_\varepsilon(a) = f^{-1}\{\varepsilon\}$$

Urbilder abgeschlossener Mengen unter f.

(3) Jeder affine Unterraum von $\mathbb{R}^n$ ist eine abgeschlossene Teilmenge, weil er sich durch ein (im allgemeinen) inhomogenes lineares Gleichungssystem, das heißt als Faser einer linearen und damit stetigen Abbildung darstellen läßt. Insbesondere ist jede einpunktige und folglich auch jede endliche Menge abgeschlossen.

(4) Man sieht leicht, daß die Kreisscheibe $U^2 \times \{0\} \subset \mathbb{R}^3$ weder offen noch abgeschlossen in $\mathbb{R}^3$ ist.



Aber die Darstellung

$$U^2 \times \{0\} = U^3 \cap \big(\mathbb{R}^2 \times \{0\}\big)$$

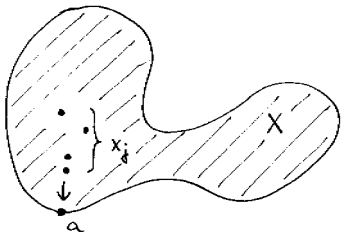
als Durchschnitt einer offenen und einer abgeschlossenen Menge zeigt, daß $U^2 \times \{0\}$ in $\mathbb{R}^2 \times \{0\}$ relativ offen und in U^3 relativ abgeschlossen ist.

Die anschauliche Vorstellung, daß abgeschlossene Mengen alle ihre Randpunkte enthalten, wird durch das folgende Lemma präzisiert:

30.8 Lemma Eine Teilmenge $X \subset \mathbb{R}^n$ ist genau dann abgeschlossen, wenn für jede konvergente Folge $(x_j)_{j=0}^\infty$ in X der Limes

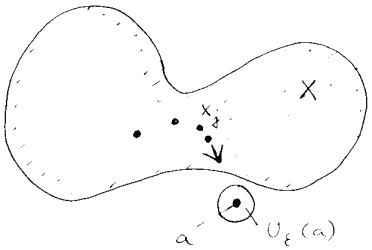
$$\lim_{j\to\infty} x_j \in \mathbb{R}^n$$

zu X gehört. (Eine sprachliche Feinheit: nur wenn das gilt, spricht man auch von einer *in X konvergenten* Folge im Unterschied zu einer konvergenten Folge in X.)



Beweis X sei abgeschlossen, und $(x_j)_{j=0}^\infty$ eine Folge in X, so daß $a = \lim\limits_{j\to\infty} x_j \in \mathbb{R}^n$ existiert. Wir zeigen, daß dann $a \in X$ gilt: Wäre das nicht der Fall, läge a also in der offenen Menge $\mathbb{R}^n \setminus X$, so gäbe es ein $\varepsilon > 0$ mit $U_\varepsilon(a) \subset \mathbb{R}^n \setminus X$, d.h.

$$U_\varepsilon(a) \cap X = \emptyset.$$



Insbesondere enthielte $U_\varepsilon(a)$ kein Glied der Folge (x_j), im Widerspruch zu $a = \lim\limits_{j\to\infty} x_j$.

Jetzt sei X als nicht abgeschlossen vorausgesetzt. Weil dann $\mathbb{R}^n \setminus X$ nicht offen ist, finden wir ein $a \in \mathbb{R}^n \setminus X$ derart, daß jede Kugel $U_\varepsilon(a)$ die Menge X trifft. Mittels der Wahl $\varepsilon := \frac{1}{j}$ erhalten wir eine Folge $(x_j)_{j=1}^\infty$ in X mit

$$|x_j - a| < \frac{1}{j} \quad \text{für } j = 1, 2, \ldots$$

Diese Folge konvergiert offensichtlich gegen $a \in \mathbb{R}^n \setminus X$, und damit ist das Lemma gezeigt.

Einer ganz neuen Definition bedarf der Begriff der Kompaktheit, den ich früher nur ad hoc für Intervalle erklärt hatte.

30.9 Definition Eine Teilmenge $X \subset \mathbb{R}^n$ heißt kompakt, wenn jede Folge $(x_j)_{j=0}^\infty$ in X eine Teilfolge enthält, die in X konvergiert.

Bemerkung Um Kompaktheit zu defineren, gibt es noch andere Möglichkeiten, die für Teilmengen von $\mathbb{R}^n$ äquivalent sind, in manchem allgemeineren Kontext aber nicht. Unsere Version wird dann als Folgenkompaktheit bezeichnet.

Jedenfalls wird die nicht auf Anhieb zu durchschauende Definition Sie auf den ersten Blick überraschen, ist doch nicht mal klar, daß ein kompaktes Intervall in diesem Sinne kompakt ist! Wenn Sie sich aber die damalige Hauptanwendung des Begriffs ansehen, nämlich den Satz 8.4 von der Annahme des Maximums, werden Sie finden, daß in Definition 30.6 genau das formuliert ist, was man bei der Verwendung der Kompaktheit braucht. Der folgende sehr wichtige Satz ist denn auch die mehrdimensionale Verallgemeinerung von 8.4, und der Beweis nur eine Formalisierung des damaligen.

30.10 Satz $X \subset \mathbb{R}^n$ sei kompakt, und $f\colon X \longrightarrow \mathbb{R}^p$ stetig. Dann ist $f(X)$ eine kompakte Teilmenge von $\mathbb{R}^p$.

Beweis Sei $(y_j)_{j=0}^\infty$ eine Folge in $f(X)$, und etwa $y_j = f(x_j)$ mit $x_j \in X$ für alle $j \in \mathbb{N}$. Weil X kompakt ist, enthält die Folge (x_j) eine konvergente Teilfolge $(x_{j_k})_{k=0}^\infty$ mit

$$\lim_{k\to\infty} x_{j_k} = a \in X.$$

Weil f bei a definiert und stetig ist, folgt daraus

$$\lim_{k\to\infty} y_{j_k} = \lim_{k\to\infty} f(x_{j_k}) = f(a) \in f(X).$$

Also ist $f(X)$ kompakt.

Erstaunlich, wie "kompakt" dieser Beweis im Vergleich zu früher ist! So etwas sollte Sie immer mißtrauisch machen. Nun, der Beweis ist deshalb so einfach geworden, weil wir die Definition schon darauf zugeschnitten hatten, und natürlich hat das einen Preis: man sieht einer Menge nicht ohne weiteres an, ob sie kompakt ist. Abhilfe schafft die folgende Charakterisierung, die auch den Anschluß an den alten Kompaktheitsbegriff für Intervalle herstellt.

30.11 Satz Eine Teilmenge $X \subset \mathbb{R}^n$ ist genau dann kompakt, wenn sie abgeschlossen und beschränkt ist.

Beweis X sei kompakt. Jede konvergente Folge $(x_j)_{j=0}^\infty$, etwa mit $\lim x_j = a \in \mathbb{R}^n$, enthält dann eine in X konvergente Teilfolge. Natürlich ist deren Limes ebenfalls a, insbesondere $a \in X$. Nach Lemma 30.5 ist X also abgeschlossen. Wäre X nicht auch beschränkt, so könnte man eine Folge $(x_j)_{j=0}^\infty$ in X mit $|x_j| > j$ für alle $j \in \mathbb{N}$ wählen: diese enthielte sicher keine konvergente Teilfolge, im Widerspruch zur Kompaktheit von X. Das vervollständigt eine Beweisrichtung.

Sei nun umgekehrt X abgeschlossen und beschränkt. Dann ist jede Folge in X ebenfalls beschränkt; nach dem Satz von Bolzano und Weierstraß besitzt sie eine konvergente Teilfolge. Weil X abgeschlossen ist, gehört deren Limes zu X. Also ist X kompakt.

Um den früheren Satz über die Annahme des Maximums als eindimensionalen Spezialfall einzuordnen, brauchen wir jetzt nur noch die folgende Tatsache zu vermerken.

30.11½ Lemma Jede nicht-leere kompakte Teilmenge $X \subset \mathbb{R}$ der Geraden besitzt ein kleinstes und ein größtes Element.

Beweis Weil X nach Satz 30.11 beschränkt ist, können wir $s := \sup X$ bilden. Nach der Eigenschaft 4.12 des Supremums gibt es eine Folge in X, die gegen s konvergiert, und weil X nach Satz 30.11 auch abgeschlossen ist, folgt daraus $s \in X$. Damit ist s das größte Element von X, und entsprechend ist $\inf X \in X$ das kleinste Element.

Bemerkungen Wenn Sie sich den Spaß machen, die Beweise der vorstehenden Sätze mit dem von Satz 8.4 zu vergleichen, werden Sie festzustellen, daß damals wie jetzt insgesamt genau die gleichen Teilschritte vorkommen. Der allgemeine Kompaktheitsbegriff ermöglicht die Aufteilung der Argumente auf mehrere unabhängige Sätze und dadurch eine durchsichtigere Darstellung. — Wir hatten uns schon früher davon überzeugt, daß eine stetige Abbildung keinen Grund hat, offene auf offene, abgeschlossene auf abgeschlossene, beschränkte auf beschränkte Mengen abzubilden, vergleiche Aufgabe 8.1. Es ist bemerkenswert, daß die entsprechende

Behauptung für die Kombination "abgeschlossen und beschränkt", weil äquivalent zu kompakt, eben doch gilt.

Aufgrund von Satz 30.11 ist es leicht, Beispiele kompakter Mengen zu nennen: alle abgeschlossenen Kugeln $D_\varepsilon(a)$, die Sphären $S_\varepsilon(a)$ sowie kompakte Quader, nämlich direkte Produkte

$$[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n] \subset \mathbb{R}^n$$

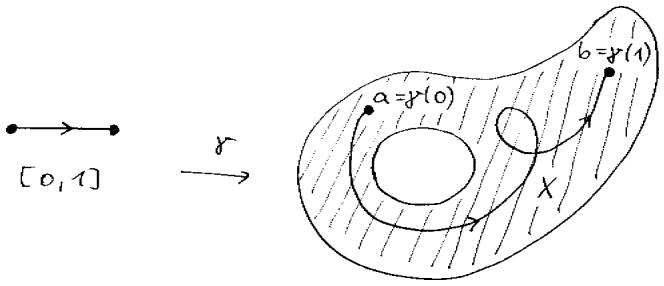
von kompakten Intervallen.

Der Begriff des Zusammenhangs kann wieder direkt aus dem zweidimensionalen Fall (Definition 10.9) übernommen werden:

30.12 Definition Eine Teilmenge $X \subset \mathbb{R}^n$ heißt (weg-)zusammenhängend, wenn es zu je zwei Punkten $a, b \in X$ einen Weg in X von a nach b gibt, nämlich eine stetige Funktion

$$\gamma\colon [0,1] \longrightarrow X$$

mit $\gamma(0) = a$ und $\gamma(1) = b$.

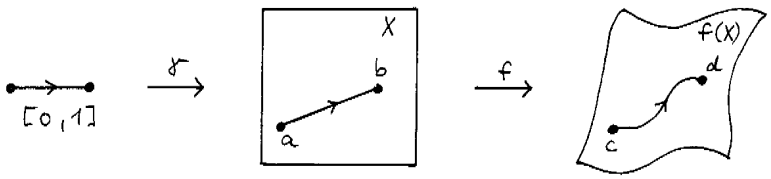


Bemerkung Eigentlich wird Zusammenhang als ein anderer, weniger anschaulicher Begriff erklärt, der bei pathologischen Mengen X vom Wegzusammenhang abweicht. So wichtig ist das aber nicht, und wir wollen für die Zwecke der Vorlesung den Zusammenhang der Einfachheit halber immer als Wegzusammenhang verstehen.

Einfachste Beispiele solcher Mengen sind die Kugeln (egal ob offen oder abgeschlossen), aber auch die Sphären positiver Dimension. Von zusammenhängenden Mengen allgemein handelt das folgende formale Analogon des Zwischenwertsatzes:

30.13 Lemma Sei $X \subset \mathbb{R}^n$ eine zusammenhängende Menge, $f\colon X \longrightarrow \mathbb{R}^p$ eine stetige Abbildung. Dann hängt auch die Bildmenge $f(X)$ zusammen.

Beweis Seien $c, d \in f(X)$, etwa $c = f(a),\ d = f(b)$. Weil X zusammenhängt, gibt es einen Weg $\gamma\colon [0,1] \to X$, der a mit b verbindet. Dann ist $f \circ \gamma\colon [0,1] \longrightarrow Y$ ein Weg in Y von c nach d.



Natürlich kann dieses einfache Lemma nicht auch inhaltlich den Zwischenwertsatz 8.3 enthalten, dessen Beweis ja einige Anstrengung erfordert hatte. Die Funktion des Zwischenwertsatzes besteht in dem neuen Rahmen vielmehr darin, daß er die zusammenhängenden Teilmengen der Geraden charakterisiert:

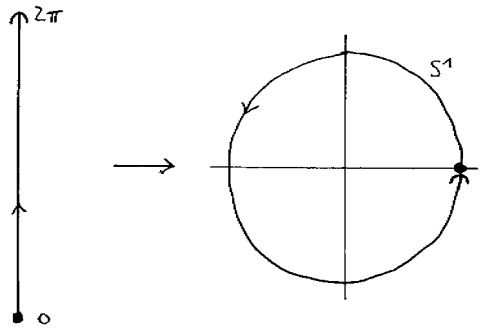
30.14 Satz Die zusammenhängenden Teilmengen von $\mathbb{R}$ sind genau die Intervalle.

Beweis Daß Intervalle zusammenhängen, ist wohl klar. Wir betrachten also nur umgekehrt eine beliebige zusammenhängende Teilmenge $X \subset \mathbb{R}$. Angesichts von Lemma 8.2 genügt es, für X die Zwischenpunkteigenschaft nachzuweisen; wir betrachten also zwei Zahlen $\alpha, \beta \in X$ mit $\alpha < \beta$. Weil X zusammenhängt, gibt es einen Weg, also eine stetige Funktion $\gamma\colon [0,1] \longrightarrow X$ mit $\gamma(0) = \alpha$ und $\gamma(1) = \beta$. Nach dem Zwischenwertsatz 8.3 liegt auch jede Zahl aus dem Intervall (α, β) im Bild von γ und damit in X.

30.15 Beispiel Jede quadratische Form $q\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ ist eine stetige Funktion. Weil die Sphäre $S^{n-1} \subset \mathbb{R}^n$ kompakt ist und — wenn wir noch $n \geq 2$ voraussetzen — auch zusammenhängt, ist die Menge $q(S^{n-1}) \subset \mathbb{R}$ ebenfalls kompakt und zusammenhängend, also ein kompaktes Intervall $[c,d]$: mit anderen Worten nimmt q auf S^{n-1} einen kleinsten Wert c und einen größten Wert d an und außerdem jede dazwischen liegende Zahl. Weil man aus den Werten von $q|S^{n-1}$ alle Werte von q mittels $q(\lambda x) = \lambda^2 q(x)$ erhält, ist q zum Beispiel genau dann positiv definit, wenn $c > 0$ ist. Mittels Hauptachsentransformation (Satz 29.1) sieht man übrigens leicht, daß c der kleinste und d der größte Eigenwert von q ist und beide als die Werte von q auf den zugehörigen Eigenvektoren realisiert werden. Später werden Sie aber auch ein rein analytisches und viel allgemeiner anwendbares Verfahren kennenlernen, um die Extremalstellen einer solchen Funktion zu finden (Satz 37.7 in Verbindung mit Satz 44.6).

Wir haben soweit zwei der drei früheren Sätze über stetige Funktionen auf Intervallen in befriedigender Weise auf das Mehrdimensionale verallgemeinert. Für den dritten, den Satz von der Umkehrfunktion, ist das nicht möglich. Daß man von Monotonie hier nicht reden kann, ist ohnehin klar, aber auch die Aussage über die Stetigkeit der Umkehrabbildung ist im Mehrdimensionalen falsch — wie wir bereits im Anschluß an Satz 12.11 anhand der stetigen Funktion

$$[0, 2\pi) \ni y \mapsto e^{iy} \in S^1 \subset \mathbb{C} = \mathbb{R}^2$$



gesehen haben, deren Umkehrung unstetig ist. Wenn allerdings der Definitionsbereich kompakt ist ...

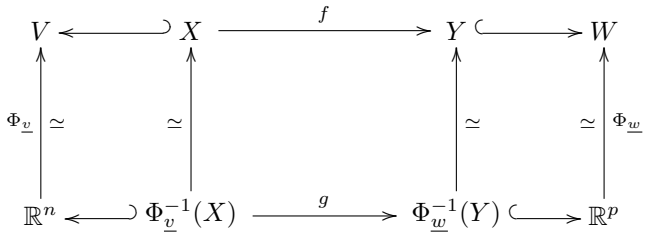
30.16 Satz Sei $X \subset \mathbb{R}^n$ und $Y \subset \mathbb{R}^p$, sowie $f\colon X \longrightarrow Y$ stetig und bijektiv. Wenn X kompakt ist, dann ist auch $f^{-1}\colon Y \longrightarrow X$ stetig.

Beweis Wir planen das Kriterium 30.6 anzuwenden und betrachten eine relativ abgeschlossene Menge $F \subset X$. Dann ist F der Durchschnitt von X mit einer abgeschlossenenen Menge, also selbst abgeschlossen in $\mathbb{R}^n$. Als Teilmenge von X ist F natürlich auch beschränkt und damit nach Satz 30.11 selbst kompakt. Jetzt greift Satz 30.10: auch $f(F)$ ist kompakt, insbesondere abgeschlossen in $\mathbb{R}^n$ und erst recht in Y. Aber diese Menge ist gerade das Urbild von F unter $f^{-1}\colon Y \to X$, und das beweist die Stetigkeit von f^{-1}.

Obwohl die Anwendbarkeit von Satz 30.16 durch die Kompaktheitsvoraussetzung wesentlich einschränkt wird, ist er das wichtigste (nämlich einzige einigermaßen allgemeine) Werkzeug, um die Stetigkeit einer Umkehrabbildung nachzuweisen.

Zum Schluß des Abschnitts sei erwähnt, daß man alle hier besprochenden Begriffe und Resultate ohne jede Schwierigkeit auf Teilmengen beliebiger endlichdimensionaler $\mathbb{R}$-Vektorräume und auf Abbildungen zwischen solchen Mengen übertragen kann. Beispielsweise seien V und W Vektorräume, und $V \supset X \xrightarrow{f} Y \subset W$ eine Abbildung. Um zu erklären, was Stetigkeit von f bedeutet, schreibt man f einfach "in linearen Karten",

d.h. man wählt Basen $\underline{v} = (v_1, \dots, v_n)$ und $\underline{w} = (w_1, \dots, w_p)$, und nennt f stetig, wenn die durch das kommutative Diagramm



definierte Abbildung g stetig ist. Zwar ändert die Wahl anderer Basen g um vor- und nachgeschaltete lineare Kartenwechsel, aber auf die Stetigkeit hat das keinen Einfluß.

Übungsaufgaben

30.1 $X \subset \mathbb{R}^n$ und $Y \subset \mathbb{R}^p$ seien Teilmengen. Verifizieren Sie:

(a) Sind X und Y offen, so ist $X \times Y \subset \mathbb{R}^{n+p}$ offen.

(b) Sind X und Y abgeschlossen, so ist $X \times Y \subset \mathbb{R}^{n+p}$ abgeschlossen.

(c) Sind X und Y kompakt, so ist $X \times Y \subset \mathbb{R}^{n+p}$ kompakt.

Tip: Zur Veranschaulichung können Sie sich erst mal vorstellen, X und Y seien Intervalle in $\mathbb{R}^n = \mathbb{R}^p = \mathbb{R}$ (auch wenn das zum Beweis im allgemeinen Fall formal nichts beiträgt).

30.2 Die Projektion $\mathrm{pr}_1 \colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ werde kurz mit p bezeichnet, also $p(x, y) = x$. Untersuchen Sie, ob die folgenden Aussagen richtig sind:

(a) Ist $X \subset \mathbb{R}^2$ offen, so ist $p(X) \subset \mathbb{R}$ offen.

(b) Ist $X \subset \mathbb{R}^2$ abgeschlossen, so ist $p(X) \subset \mathbb{R}$ abgeschlossen.

(c) Ist $X \subset \mathbb{R}^2$ kompakt, so ist $p(X) \subset \mathbb{R}$ kompakt.

30.3 $(X_\lambda)_{\lambda \in \Lambda}$ sei eine Familie von Teilmengen $X_\lambda \subset \mathbb{R}^n$; es bezeichne

$$X := \bigcup_{\lambda \in \Lambda} X_\lambda$$

ihre Vereinigung. Weiter sei $f \colon X \longrightarrow \mathbb{R}^p$ eine Abbildung. Wenn f stetig ist, ist auch jede Einschränkung $f|X_\lambda$ stetig, das ist klar. Beweisen Sie umgekehrt:

(a) Wenn alle X_λ offen in X und alle $f|X_\lambda$ stetig sind, dann ist f stetig.

(b) Wenn Λ endlich, alle X_λ abgeschlossen in X und alle $f|X_\lambda$ stetig sind, dann ist f stetig.

30.4 Seien $X \subset \mathbb{R}^n$ und $Y \subset \mathbb{R}^p$ Teilmengen, und $f \colon X \longrightarrow Y$ eine Abbildung. Gibt es einen Zusammenhang zwischen den beiden folgenden Eigenschaften von f:

- f ist stetig,
- der Graph $\Gamma_f = \big\{ \big(x, f(x)\big) \,\big|\, x \in X \big\} \subset X \times Y$ ist eine in $X \times Y$ abgeschlossene Teilmenge?

30.5 $(X_j)_{j=0}^{\infty}$ sei eine absteigend geschachtelte Folge nicht-leerer kompakter Teilmengen $X_j \subset \mathbb{R}^n$:

$$X_0 \supset X_1 \supset \cdots \supset X_j \supset X_{j+1} \supset \cdots$$

Beweisen Sie, daß dann

$$\bigcap_{j=0}^{\infty} X_j \neq \emptyset$$

ist. Wenn Sie das Gefühl haben, das sei ja sowieso klar, betrachten Sie vorweg das Beispiel $X_j = [j, \infty)$ mit nur abgeschlossenen, nicht kompakten Mengen, das deswegen in Wirklichkeit ein Unbeispiel ist.

30.6 Die Teilmengen X und Y von $\mathbb{R}^n$ seien nicht-leer und disjunkt: $X \cap Y = \emptyset$. In dieser Aufgabe geht es um die Frage, ob X und Y dann einen positiven Abstand voneinander haben und dieser Abstand angenommen wird, d.h. ob zwei Punkte $a \in X$ und $b \in Y$ existieren, so daß

$$|x - y| \geq |a - b| \quad \text{für alle } x \in X \text{ und alle } y \in Y$$

ist.

(a) Belegen Sie durch ein Beispiel, daß das im allgemeinen falsch ist, selbst dann, wenn man zusätzlich X und Y als abgeschlossen voraussetzt.

(b) Beweisen Sie, daß das aber richtig ist, wenn X und Y kompakt sind. Tip: Untersuchen Sie die Funktion

$$X \times Y \ni (x, y) \overset{f}{\longmapsto} |x - y| \in \mathbb{R}.$$

(c) Es genügt sogar, wenn etwa X kompakt und Y nur abgeschlossen ist: Überlegen Sie sich, daß man dann Y durch eine kompakte Teilmenge $Y' \subset Y$ ersetzen kann, weil Punkte $y \in Y$, die von X weit genug entfernt sind, für das Problem keine Rolle spielen.

30.7 Es sei $X \subset \mathbb{R}^n$ eine beliebige und $Y \subset \mathbb{R}^p$ eine kompakte Teilmenge; $f: X \times Y \longrightarrow \mathbb{R}$ sei eine stetige Funktion. Begründen Sie, warum durch

$$F(x) := \min\left\{f(x, y) \,\middle|\, y \in Y\right\}$$

eine Funktion $F: X \longrightarrow \mathbb{R}$ definiert wird, und beweisen Sie, daß F stetig ist.

Tip: Arbeiten Sie mit dem Folgenkriterium; nehmen Sie die Existenz einer "Verbrecherfolge" an und wählen Sie aus dieser geschickt Teil-Verbrecherfolgen aus ...

Wer Lust dazu hat, kann sich außerdem ein Beispiel einer stetigen Funktion $f: X \times Y \longrightarrow \mathbb{R}$ mit Teilmengen $X \subset \mathbb{R}^n$ und (notwendig nicht-kompaktem) $Y \subset \mathbb{R}^p$ und den folgenden Eigenschaften überlegen:

- Für jedes $x \in X$ existiert das Minimum der Menge $\left\{f(x, y) \,\middle|\, y \in Y\right\}$, aber
- die wie in der vorigen Aufgabe definierte Funktion $F: X \longrightarrow \mathbb{R}$ ist unstetig.

30.8 Beweisen Sie: $\emptyset$ und $\mathbb{R}$ sind die einzigen Teilmengen von $\mathbb{R}$, die zugleich offen und abgeschlossen sind.

30.9 Beweisen Sie allgemeiner: Wenn $X \subset \mathbb{R}^n$ zusammenhängt, dann sind $\emptyset$ und X die einzigen Teilmengen von X, die zugleich relativ offen und abgeschlossen in X sind.

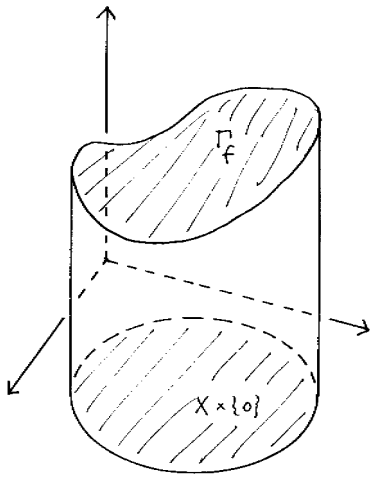
31 Integrieren

Mit der Integralrechnung wenden wir uns einem Thema zu, über das ich auch im Eindimensionalen noch nichts gesagt habe. Der Zweck der Integralrechnung ist leicht erklärt: Es geht darum, Volumina zu berechnen, und dazu natürlich den anschaulichen Begriff des Volumens oder *Maßes* erst mal mathematisch zu präzisieren. Für jedes $n \in \mathbb{N}$ ist damit ein eigener n-dimensionaler Volumen- oder Maßbegriff gemeint, von dem das Volumen der Alltagssprache nur der dreidimensionale Fall ist; für $n = 2$ und $n = 1$ wird das n-dimensionale Volumen dann zu Flächeninhalt und Länge.

Sehen wir uns die Sache etwas näher an: Gegeben seien eine Teilmenge $X \subset \mathbb{R}^n$ und eine überall positive reellwertige Funktion $f\colon X \longrightarrow \mathbb{R}$. Die noch einzuführende Zahl $\int f \in \mathbb{R}$ (Integral von f über X) soll das Volumen der $(n+1)$-dimensionalen "Büchse" mit Boden $X \times \{0\} \subset \mathbb{R}^{n+1}$ und dem Graphen

$$\Gamma_f = \big\{\big(x, f(x)\big) \,\big|\, x \in X\big\}$$

als Deckel sein:



Es erweist sich, daß die Aufgabe, die man da vor sich hat, in zwei ziemlich sauber zu trennende Teile zerfällt: einmal dieses Integral als abstrakte Größe zu konstruieren (was nicht ohne weitere Voraussetzungen über f möglich sein wird), andererseits den Umgang mit dem Integral zu lernen, unter anderem, um die Volumina konkreter, formelmäßig gegebener Büchsen wirklich ausrechnen zu können.

Für die erste Aufgabe haben die Mathematiker erst zu Anfang dieses Jahrhunderts eine wirklich befriedigende Lösung gefunden. Das ist insofern erstaunlich, als ja schon Newton und Leibniz Integrale gekannt haben und mit ihnen effizient zu rechnen wußten. Die Erklärung: Der Fortschritt liegt nicht darin, daß man für früher bekannte Integrale heute andere, in irgendeinem Sinne bessere Werte bekäme, sondern darin, daß man inzwischen von viel mehr Funktionen überhaupt ein Integral bilden kann. Der Integralbegriff hat dadurch ungeheuer an Flexibilität gewonnen und ist zur Grundlage ganz neuer Theorien geworden. Freilich, für eine — zwangsläufig knappe — Darstellung in einer Grundvorlesung eignet sich die Konstruktion des Integrals weniger, nicht weil sie besonders schwierig wäre, sondern weil man die in der Konstruktion steckende Raffinesse eigentlich nur in einem weiter gefaßten Rahmen so recht würdigen kann. Ich werde Ihnen das Integral deshalb als ein schon fertiges Objekt zusammen mit einer Liste seiner grundlegenden Eigenschaften vorstellen. Zu dem zweiten genannten Ziel, nämlich den Umgang mit dem Integral zu lernen, tragen ohnehin nur diese Eigenschaften und nicht die Einzelheiten der Konstruktion bei. Die Situation wird insofern

ähnlich sein wie seinerzeit bei der Determinante, die wir ja schon zu einem Zeitpunkt berechnen konnten, als ihre Existenz noch gar nicht gesichert war. Anders als dort werde ich aber Existenz und Eindeutigkeit des Integrals nicht beweisen, sondern als gegeben ansehen.

Vorweg wollen wir eine Klasse von Mengen vereinbaren, die wegen ihrer Kleinheit bei der Integration keine Rolle spielen und deswegen vernachlässigt werden dürfen. Dabei benutze ich den Begriff des Volumens oder Maßes eines (beschränkten) Quaders

$$Q = I_1 \times I_2 \times \cdots \times I_n \subset \mathbb{R}^n$$

und meine damit für positives n die auf ganz elementare Weise gebildete Zahl

$$\mu(Q) := (b_1 - a_1)(b_2 - a_2)\cdots(b_n - a_n) \in [0, \infty),$$

wenn a_j und b_j Anfangs- und Endpunkt des Intervalls I_j sind:

$$(a_j, b_j) \subset I_j \subset [a_j, b_j] \quad \text{für } j = 1, \ldots, n$$

Der Systematik halber sollte man auch den trivialen Fall $n = 0$ festlegen, was man zweckmäßig durch

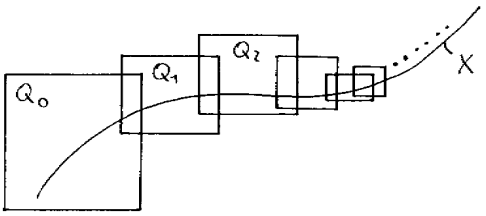
$$\mu(\emptyset) := 0 \quad \text{und} \quad \mu(\mathbb{R}^0) := 1$$

tut; mehr Teilmengen von $\mathbb{R}^0 = \{0\}$ gibt es ja nicht.

31.1 Definition Eine Teilmenge $X \subset \mathbb{R}^n$ heißt eine Nullmenge, wenn es zu jedem $\varepsilon > 0$ eine Folge von Quadern $(Q_j)_{j=0}^{\infty}$ gibt, so daß

$$X \subset \bigcup_{j=0}^{\infty} Q_j \quad \text{und} \quad \sum_{j=0}^{\infty} \mu(Q_j) < \varepsilon$$

ist.



Es ist klar, daß jede Teilmenge einer Nullmenge selbst eine Nullmenge ist. Interessanter ist die folgende Eigenschaft:

31.2 Lemma Die Vereinigung (höchstens) abzählbar vieler Nullmengen ist wieder eine Nullmenge.

Beweis Es genügt, eine Folge $(X_j)_{j=0}^{\infty}$ von Nullmengen $X_j \subset \mathbb{R}^n$ zu betrachten. Sei $\varepsilon > 0$. Für jedes $j \in \mathbb{N}$ wählen wir eine Folge von Quadern $(Q_{jk})_{k=0}^{\infty}$ mit

$$X_j \subset \bigcup_{k=0}^{\infty} Q_{jk} \quad \text{und} \quad \sum_{k=0}^{\infty} \mu(Q_{jk}) < 2^{-j}\varepsilon.$$

Die mit $\mathbb{N} \times \mathbb{N}$ indizierte Familie $(Q_{jk})_{j,k=0}^{\infty}$ ordnen wir auf beliebige Weise zu einer Folge an; aus

$$\bigcup_{j=0}^{\infty} X_j \subset \bigcup_{j,k=0}^{\infty} Q_{jk} \quad \text{und} \quad \sum_{j,k=0}^{\infty} \mu(Q_{jk}) < \sum_{j=0}^{\infty} 2^{-j}\varepsilon = 2\varepsilon$$

folgt dann, daß die Vereinigung $\bigcup_{j=0}^{\infty} X_j$ eine Nullmenge ist. Dabei sind natürlich die Überlegungen aus Abschnitt 6 über Abzählbarkeit und Doppelreihen benutzt.

31.3 Beispiele von Nullmengen (1) Für $n > 0$ ist jede abzählbare Teilmenge X von $\mathbb{R}^n$ eine Nullmenge: man zähle $X = \{x_j \,|\, j \in \mathbb{N}\}$ ab und wähle in der Definition $Q_j = \{x_j\}$, was ja ein Quader mit Volumen null ist. Insbesondere ist also $\mathbb{Q} \subset \mathbb{R}$ eine Nullmenge.

(2) Die *Koordinatenhyperebene*

$$\mathbb{R}^{n-1} \times \{0\} \subset \mathbb{R}^{n-1} \times \mathbb{R} = \mathbb{R}^n$$

ist eine Nullmenge: Sie ist die Vereinigung der Quader

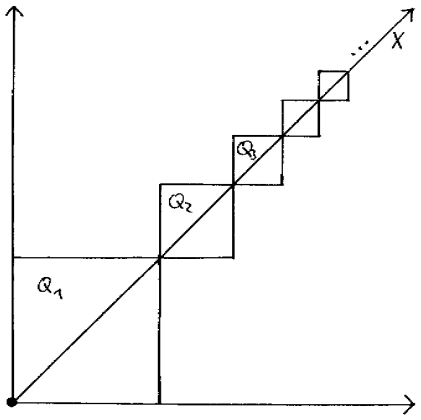
$$Q_j := [-j, j]^{n-1} \times \{0\} \subset \mathbb{R}^{n-1} \times \mathbb{R},$$

die ebenfalls alle das Volumen null haben. Entsprechend für die anderen Koordinatenhyperebenen und die dazu parallelen affinen Teilräume.

(3) Die Anschauung suggeriert, daß allgemeiner jeder echte affine Unterraum von $\mathbb{R}^n$ eine Nullmenge ist. Das direkt mittels der Definition zu beweisen, ist etwas mühsam und lohnt nicht, weil wir das Resultat später umsonst bekommen. Ich will Ihnen hier aber immerhin vorführen, warum der Strahl

$$X = \big\{(t, t) \in \mathbb{R}^2 \,\big|\, t \geq 0\big\}$$

eine Nullmenge in $\mathbb{R}^2$ ist, denn dieses Beispiel illustriert Definition 31.2 besser als die bisherigen. Zu gegebenem $\varepsilon > 0$ reihen wir hier abgeschlossene Quadrate Q_j der Seitenlänge $\sqrt{\varepsilon}/j$ längs des Strahls auf $(j = 1, 2, \ldots)$:



Weil die harmonische Reihe $\displaystyle\sum_{j=1}^{\infty} \frac{\sqrt{\varepsilon}}{j}$ divergiert, wird X von der Vereinigung aller Q_j überdeckt. Die Reihe $\displaystyle\sum \frac{1}{j^2}$ dagegen konvergiert; ist $c\,(= \pi^2/6)$ ihre Summe, so ist

$$\sum_{j=1}^{\infty} \mu(Q_j) = \sum_{j=1}^{\infty} \frac{\varepsilon}{j^2} = c\,\varepsilon,$$

und das beweist die Behauptung.

In der Integralrechnung kann man sich grundsätzlich auf Funktionen beschränken, die auf ganz $\mathbb{R}^n$ (und nicht nur einer Teilmenge X) erklärt sind. Das liegt daran, daß man die zu integrierende Funktion außerhalb von X skrupellos mit Wert null fortsetzen kann, ohne daß sich das Integral daran stört. Zu diesem Zweck vereinbaren wir die

31.4 Notation Sei $X \subset \mathbb{R}^n$ und $f\colon X \longrightarrow \mathbb{R}$ eine Funktion. Dann bezeichnet f_X die auf $\mathbb{R}^n$ mit null fortgesetzte Funktion

$$f_X\colon \mathbb{R}^n \longrightarrow \mathbb{R};\ \ \mathbb{R}^n \ni x \mapsto \begin{cases} f(x) & \text{wenn } x \in X, \\ 0 & \text{sonst.} \end{cases}$$

Ist allgemeiner der Definitionsbereich von f nicht X selbst, sondern eine X umfassenden Menge $Y \supset X$, so schreibt man die Funktion $(f|X)_X$ auch in diesem Fall kurz f_X.

Nun zum Integral selbst:

31.5 Grundeigenschaften des Integrals Sei $n \in \mathbb{N}$. Das n-dimensionale Integral ist ein lineares Funktional, das auf einem mit $\mathcal{L}^1(\mathbb{R}^n)$ bezeichneten linearen Unterraum des $\mathbb{R}$-Vektorraums aller Funktionen $\{f\colon \mathbb{R}^n \longrightarrow \mathbb{R}\}$ definiert ist:

$$\mathcal{L}^1(\mathbb{R}^n) \ni f \longmapsto \int f \in \mathbb{R}$$

Die Funktionen aus $\mathcal{L}^1(\mathbb{R}^n)$ nennt man integrierbar. Eigenschaften des Integrals:

(a) Ist $f, g \in \mathcal{L}^1(\mathbb{R}^n)$, so ist auch $f_{\{x\in\mathbb{R}^n|g(x)>0\}} \in \mathcal{L}^1(\mathbb{R}^n)$.
(b) Ist $f \in \mathcal{L}^1(\mathbb{R}^n)$ und $f(x) \geq 0$ für alle $x \in \mathbb{R}^n$, so ist $\int f \geq 0$.
(c) Ist $X \subset \mathbb{R}^n$ kompakt und $f\colon X \longrightarrow \mathbb{R}$ stetig, so ist $f_X \in \mathcal{L}^1(\mathbb{R}^n)$.
(d) Ist $Q \subset \mathbb{R}^n$ ein kompakter Quader, so ist $\int 1_Q = \mu(Q)$ das Volumen von Q.
(e) Ist $X \subset \mathbb{R}^n$ eine Nullmenge und $f\colon X \longrightarrow \mathbb{R}$ eine beliebige Funktion, so ist $f_X \in \mathcal{L}^1(\mathbb{R}^n)$ und $\int f_X = 0$.

Bemerkungen Diese Eigenschaften bilden noch kein Axiomensystem für den Begriff "Integral", und schon gar kein minimales. Ich habe hier nur ziemlich willkürlich diejenigen Eigenschaften zusammengestellt, die wir als erste brauchen. — Der eigenartig erscheinenden Bezeichnung $\mathcal{L}^1(\mathbb{R}^n)$ für den Raum der integrablen Funktionen liegt die Tatsache zugrunde, daß daneben auch Räume $\mathcal{L}^p(\mathbb{R}^n)$ mit beliebigem $p \in [1, \infty)$ (vor allem $p = 2$) im Leben eine Rolle spielen. Sie haben mit Integrierbarkeit nicht von f selbst, sondern der punktweise gebildeten Potenz f^p zu tun; ihre genaue Definition ist

$$\mathcal{L}^p(\mathbb{R}^n) := \left\{ f\colon \mathbb{R}^n \longrightarrow \mathbb{R} \;\middle|\; |f|^{p-1} f \in \mathcal{L}^1(\mathbb{R}^n) \right\}.$$

— Statt von Integrierbarkeit von f_X spricht man meist von Integrierbarkeit von f *über* X und schreibt

$$\int_X f$$

statt $\int f_X$. Die ausführlichere klassische Schreibweise

$$\int_X f(x)dx = \int_X f(x_1, \ldots, x_n)d(x_1, \ldots, x_n)$$

hat ihre Vorzüge dann, wenn der Integrand f durch einen expliziten Formelausdruck gegeben ist, wie in $\int_X x^2 dx$.

Zuerst wollen wir die Positivitätseigenschaft (b) ein wenig ausarbeiten. Wegen der Linearität des Integrals kann man sie auch so formulieren:

31.6 Notiz und **Schreibweise** Aus $f, g \in \mathcal{L}^1(\mathbb{R}^n)$ und $f \leq g$ folgt $\int f \leq \int g$; dabei ist $f \leq g$ als gängige Abkürzung für

$$f(x) \leq g(x) \quad \text{für alle } x \in \mathbb{R}^n$$

geschrieben.

Beachten Sie, daß man nicht je zwei Funktionen $f, g\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ in diesem Sinne vergleichen kann; im allgemeinen gilt natürlich weder $f \leq g$ noch $f \geq g$. — Noch eine

31.6½ Notation Es ist manchmal nützlich, eine Funktion $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ durch

$$x \mapsto f_+(x) := \begin{cases} f(x) & \text{wenn } f(x) > 0 \\ 0 & \text{sonst} \end{cases} \qquad \text{und} \qquad x \mapsto f_-(x) := \begin{cases} -f(x) & \text{wenn } f(x) < 0 \\ 0 & \text{sonst} \end{cases}$$

in zwei nicht-negative Funktionen $f_\pm \geq 0$ zu zerlegen:

$$f = f_+ - f_-$$
$$|f| = f_+ + f_-$$

Ähnlich wie Summen (gewöhnliche und Reihensummen) genügen Integrale einer

31.7 Dreiecksungleichung Für jedes $f \in \mathcal{L}^1(\mathbb{R}^n)$ ist $|f| \in \mathcal{L}^1(\mathbb{R}^n)$ und

$$\left|\int f\right| \leq \int |f|.$$

Beweis Daß mit f auch $|f|$ integrierbar ist, folgt aus (a): danach sind ja $f_+ = f_{\{x|f(x)>0\}}$ und $f_- = (-f)_+$ integrierbar, also auch $|f| = f_+ + f_-$. Jetzt bleibt nur die Notiz auf die Ungleichung $-|f| \leq f \leq |f|$ anzuwenden:

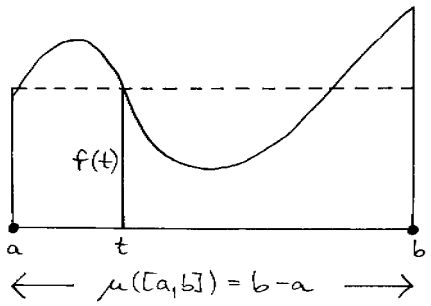
$$-\int |f| \leq \int f \leq \int |f|$$

Auf eine viel speziellere Situation bezieht sich der sogenannte

31.8 Mittelwertsatz der Integralrechnung $X \subset \mathbb{R}^n$ sei eine nicht-leere kompakte zusammenhängende Menge, und $f\colon X \longrightarrow \mathbb{R}$ eine stetige Funktion. Dann gibt es einen Punkt $t \in X$ mit

$$\int_X f = f(t) \cdot \int_X 1.$$

Erklärung und *Beweis* Axiom 31.5(c) stellt sicher, daß f ebenso wie die konstante Funktion 1 über X integrierbar ist. Die nach 31.5(b) nicht-negative Zahl $\mu(X) := \int_X 1$ wird man natürlich als das *Volumen* oder *Maß* von X ansprechen, und der Satz interpretiert das durch $\mu(X)$ geteilte Integral tatsächlich als einen "mittleren" oder durchschnittlichen Wert von f (Skizze für $X = [a, b]$):



Der Beweis beruht zunächst darauf, daß die Bildmenge $f(X) \subset \mathbb{R}$ nach den Sätzen 30.10, 30.13 und 30.14 ein kompaktes Intervall $[b, d]$ ist; nach der Notiz 31.6 ist dann

$$b \cdot \mu(X) = \int_X b \leq \int_X f \leq \int_X d = d \cdot \mu(X).$$

Wir können deshalb

$$\int_X f = c\cdot\mu(X) \quad \text{mit } c \in [b,d]$$

schreiben, und brauchen nur noch ein $t \in X$ so zu nehmen, daß $f(t) = c$ ist.

Im Rest dieses Abschnitts wollen wir uns vor allem anhand des eindimensionalen Falls mit dem Integral vertraut machen. Hier sind natürlich die über ein Intervall $I \subset \mathbb{R}$ gebildeten Integrale $\int_I f$ die wichtigsten. Ist a Anfangs- und b Endpunkt von I, so rechtfertigt die Eigenschaft 31.5(e) die übliche Schreibweise

$$\int_a^b f := \int_I f,$$

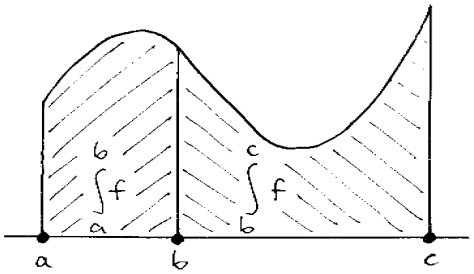
denn die Menge $\{a,b\} \subset \mathbb{R}$ ist als Nullmenge vernachlässigbar. Übrigens muß nur $\infty \le a \le b \le \infty$ sein, es darf sich also auch um unbeschränkte Intervalle handeln. Manchmal ist es praktisch, sogar $a > b$ zuzulassen; für diesen Fall definiert man

$$\int_a^b := -\int_b^a .$$

Im Umgang mit einem so notierten Integral muß man natürlich den gesunden Menschenverstand walten lassen, der einem zum Beispiel sagt, daß man in 31.5 bei den Regeln (b) und (d) das Vorzeichen ändern muß. Auf die Aussage des folgenden ebenso einfachen wie plausiblen Lemmas dagegen ist diese Notation direkt zugeschnitten:

31.9 Lemma Seien $a,b,c \in \mathbb{R}$ drei Punkte. Dann gilt: Eine über zwei der Intervalle (a,b), (a,c) und (b,c) integrierbare Funktion $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ ist über alle drei integrierbar, und es gilt

$$\int_a^c f = \int_a^b f + \int_b^c f.$$



Beweis Wir setzen zunächst $a < b < c$ voraus. Sei f über (a,c) integrierbar, also $f_{(a,c)} \in \mathcal{L}^1(\mathbb{R})$. Nach Eigenschaft (c) des Integrals ist $1_{[a,b]} \in \mathcal{L}^1(\mathbb{R})$, nach (a) also auch

$$f_{(a,b]} = \big(f_{(a,c)}\big)_{[a,b]} \in \mathcal{L}^1(\mathbb{R}),$$

d.h f ist über $(a,b]$ und damit auch über (a,b) integrierbar. Analog folgt die Integrierbarkeit über (b,c) (und tatsächlich über jedes Teilintervall von (a,c)).

Der einzige durch diese Überlegung nicht abgedeckte Fall ist der, daß die Integrierbarkeit von f über (a,b) und (b,c) vorausgesetzt ist: dann folgt die über (a,c) aus

$$f_{(a,c)} = f_{(a,b)} + f_{(b,c)} \in \mathcal{L}^1(\mathbb{R});$$

schließlich ist $L^1(\mathbb{R})$ ein Vektorraum.

Von der Einschränkung $a < b < c$ befreit man sich nun leicht, indem man die paar möglichen Fälle mittels $\int_x^y = -\int_y^x$ durchspielt.

Natürlich habe ich beim Beweis des Lemmas gemogelt: die Funktion $f_{(a,b)} + f_{(b,c)}$ hat ja bei b gar nicht wie $f_{(a,c)}$ den Wert $f(b)$, sondern den Wert null. Aber weil $\{b\} \subset \mathbb{R}$ eine Nullmenge ist, bleibt das ganz egal. Um den Umgang mit derlei Situationen ein für allemal superbequem zu gestalten, vereinbart man die

31.10 Sprechweise Man sagt, eine für alle Punkte von $\mathbb{R}^n$ definierte Aussage gelte fast überall, wenn die Punkte, für die sie nicht gilt, eine Nullmenge bilden.

Die Gleichung aus dem Beweis des Lemmas schreibe ich also lieber

$$f_{(a,c)} = f_{(a,b)} + f_{(b,c)} \quad \text{fast überall,}$$

und damit ist alles wieder in Ordnung.

Die folgende Tatsache ist grundlegend für die praktische Berechnung von Integralen.

31.11 Satz Das Intervall $I \subset \mathbb{R}$ enthalte außer $a \in I$ weitere Punkte, und $f\colon I \longrightarrow \mathbb{R}$ sei eine stetige Funktion. Dann ist die Funktion

$$I \ni x \longmapsto \int_a^x f \in \mathbb{R}$$

eine Stammfunktion von f:

$$\frac{d}{dx}\int_a^x f = f(x) \quad \text{für alle } x \in I$$

Beweis Nach Lemma 31.9 und dem Mittelwertsatz 31.8 ist

$$\int_a^{x+h} f - \int_a^x f = \int_x^{x+h} f = f(t)\cdot h$$

für ein geeignetes t zwischen x und $x+h$, und das bleibt auch bei negativem h richtig. Durchteilen durch h und Limesbildung $h \to 0$, die natürlich $t \to x$ erzwingt, gibt

$$\lim_{h\to 0} \frac{\int_a^{x+h} f - \int_a^x f}{h} = \lim_{t\to x} f(t) = f(x).$$

31.12 Folgerung Sei $a < b$ und $f\colon [a,b] \longrightarrow \mathbb{R}$ stetig. Ist $F\colon [a,b] \longrightarrow \mathbb{C}$ eine Stammfunktion von f, so ist

$$\int_a^b f = F(b) - F(a).$$

Beweis Nach Satz 31.11 ist auch

$$G\colon x \longrightarrow \int_a^x f$$

eine Stammfunktion von f. Die Differenz $F - G$ ist nach der Folgerung 16.3 konstant, also ist

$$F(b) - F(a) = G(b) - G(a) = \int_a^b f - \int_a^a f = \int_a^b f.$$

Die Folgerung wird traditionell "Hauptsatz der Differential- und Integralrechnung" genannt. Na ja. Immerhin ist es damit möglich, eine ganze Menge von Integralen auszurechnen, bei denen man eine Stammfunktion des Integranden einfach schon kennt (wie bei den Potenzen $x \mapsto x^\alpha$ und bei exp, cos, sin etc.) oder nach Regeln der Differentialrechnung ermitteln oder zumindest raten kann. Dabei kann es helfen, die Produkt- und die Kettenregel gleich für Integrale umzuschreiben; so entstehen die bekannten

31.13 Integrationsregeln Sei $[a,b] \subset \mathbb{R}$ ein kompaktes Intervall mit $a < b$.

(a) Sind $f, g\colon [a,b] \longrightarrow \mathbb{R}$ zwei C^1-Funktionen, so ist

$$\int_a^b f'g = (fg)(b) - (fg)(a) - \int_a^b fg'$$

("partielle Integration"; als Bezeichnungen für die Differenz $(fg)(b) - (fg)(a)$ sind auch

$$fg\Big|_a^b \quad \text{und} \quad \big[fg\big]_a^b$$

beliebt).

(b) Ist $\mathbb{R} \supset X \xrightarrow{f} \mathbb{R}$ stetig und $\varphi\colon [a,b] \longrightarrow X$ eine C^1-Funktion, so gilt

$$\int_a^b (f\circ\varphi)\varphi' = \int_{\varphi(a)}^{\varphi(b)} f$$

("Substitutionsregel").

Beweis In (a) ist fg eine Stammfunktion von $f'g + fg'$. Ist in (b) F eine Stammfunktion von f auf dem Intervall $\varphi[a,b]$, so ist $F\circ\varphi$ eine solche von $(f\circ\varphi)\varphi'$.

31.14 Beispiele (1) Immer wieder hat man mit Integralen der Form $\int f(\lambda x)dx$ mit festem $\lambda \neq 0$ zu tun. Um die Substitutionsregel mit $\varphi(x) = \lambda x$, also $\varphi'(x) = \lambda$ anzuwenden, schreibt man

$$\int_a^b f(\lambda x)dx = \frac{1}{\lambda}\int_a^b f(\lambda x)\lambda dx = \frac{1}{\lambda}\int_{\lambda a}^{\lambda b} f(y)dy$$

und führt das Integral so auf ein anderes zurück. Dabei spielt es keine Rolle, ob λ positiv oder negativ ist, also ob φ monoton wächst oder fällt.

(2) Die Substitutionsfunktion φ braucht überhaupt nicht monoton zu sein: Sei $f\colon [-1,1] \longrightarrow \mathbb{R}$ stetig. Zur Berechnung von $\int_0^{2\pi} f(\cos t)\sin t\,dt$ wird man in salopper, aber leicht zu merkender Form $\varphi = \cos t$, also $d\varphi = \varphi'(t)\,dt = -\sin t\,dt$ substituieren und erhält

$$\int_0^{2\pi} f(\cos t)\sin t\,dt = -\int_{\cos 0}^{\cos 2\pi} f(\varphi)d\varphi = 0.$$

Bemerkungen Mit einem Blick auf den "Hauptsatz" 31.12 werden Stammfunktionen einer gegebenen Funktion f häufig mit dem Symbol $\int f$ notiert und als "unbestimmtes Integral" bezeichnet. Das ist eine zunächst mal unzulässige Schreibweise, weil Stammfunktionen auch in dem günstigen Fall, daß f auf einem Intervall oder einem Gebiet erklärt ist, nur bis auf Addition einer Konstanten eindeutig bestimmt sind. Man kann sich aber durch die Vereinbarung helfen, daß in einer Formel mit unbestimmten Integralen wie

$$\int \frac{dx}{x^2+1} = \arctan x$$

das Gleichheitszeichen eine andere als die normale Bedeutung haben soll, nämlich daß die Differenz beider Seiten eine konstante Funktion ist. Nur dann führt die Tatsache, daß auch $x \mapsto -\operatorname{arccot} x$ eine Stammfunktion von $x \mapsto \dfrac{1}{x^2+1}$ ist, nicht zu einem scheinbaren Widerspruch: die Gleichung

$$\arctan x = \int \frac{dx}{x^2+1} = -\operatorname{arccot} x$$

ist im Sinne der Vereinbarung akzeptabel, weil $\arctan + \operatorname{arccot} = \pi/2$ tatsächlich eine konstante Funktion ist.

Obwohl fleißiges Anwenden der Integrationsregeln 31.13 (und vielleicht noch weiterer) zu umfangreichen Formelsammlungen führt, soll man sich der Tatsache bewußt sein, daß die Integrationsregeln anders als die Differentiationsregeln aus den Abschnitten 13 und 14 keine Anleitung enthalten, um zu einer gegebenen "elementaren" (aus den gängigen Grundbausteinen mittels der gängigen Prozesse gebildeten) Funktion eine Stammfunktion zu berechnen: Die Regel der partiellen Integration drückt *nicht* das Integral eines Produkts durch Integrale über die Faktoren aus, und die Substitutionsregel ist *keine* Formel für das Integral einer beliebigen Komposition. Der Grund für das Fehlen systematischer Regeln ist nicht etwa mangelnder Scharfsinn der Mathematiker, sondern die Tatsache, daß die Stammfunktionen vieler Funktionen einer umfangreicheren Klasse angehören als diese selbst. Darauf deutet schon das eben zitierte Beispiel hin: die rationale Funktion $x \mapsto \frac{1}{x^2+1}$ hat als Stammfunktion eine zwar noch elementare, aber schon kompliziertere Funktion, nämlich die Umkehrung einer trigonometrischen. Die einfachsten nicht-elementaren Stammfunktionen sind die sogenannten *elliptischen* (unbestimmten) Integrale

$$\int \frac{dx}{\sqrt{p(x)}} \quad \text{mit einem Polynom } p \text{ vom Grad 3 oder 4;}$$

solche treten bei der Berechnung des Ellipsenumfangs auf (daher der Name), aber auch bei der Berechnung der Periode des Kreispendels.

Systematisch berechnen kann man aber immerhin Stammfunktionen der rationalen Funktionen. Division mit Rest und die in unter 10.12 erklärte Partialbruchzerlegung reduzieren diese Aufgabe darauf, Stammfunktionen der (im allgemeinen komplexen) Funktionen

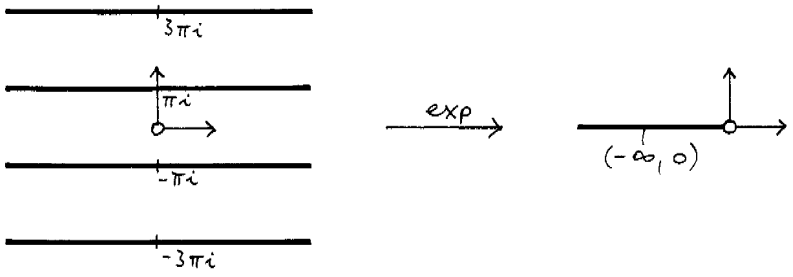
$$z \mapsto z^k \quad (k \in \mathbb{N}) \qquad \text{und} \qquad z \mapsto \frac{1}{(z-c)^k} \quad (0 < k \in \mathbb{N})$$

zu finden. Die kann man in der Form

$$z \mapsto \frac{z^{k+1}}{k+1} \qquad \text{und} \qquad z \mapsto \begin{cases} \log(z-c) & (k=1) \\ -\dfrac{1}{(k-1)(z-c)^{k-1}} & (k>1) \end{cases}$$

sofort hinschreiben, muß sich aber Gedanken über die Bedeutung des komplexen Logarithmus machen:

31.15 Rechnung Wenn c reell ist, die ursprünglich gegebene rationale Funktion $f\colon I \longrightarrow \mathbb{R}$ dort also eine Polstelle hat, ist zu unterscheiden, ob das Intervall I rechts oder links von c liegt. Im ersten Fall ist $x \mapsto \log(x-c)$ durch den gewöhnlichen reellen Logarithmus erklärt. Liegt I dagegen links von c, so liefert jede Wahl von $k \in \mathbb{Z}$ eine mögliche Interpretation



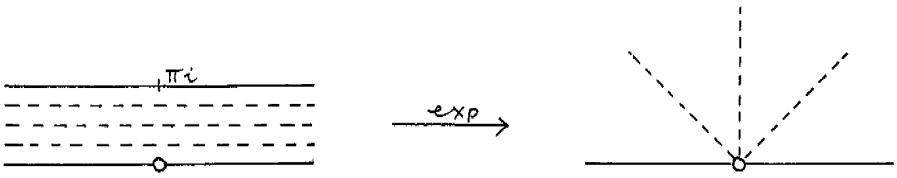
$$\log(x-c) = \log|x-c| + (2k+1)i\pi$$

des Logarithmus als (partielle) Umkehrung der Exponentialfunktion. Zur Verwendung als Stammfunktion von $x \mapsto \frac{1}{x-c}$ können wir den konstanten Imaginärteil aber ignorieren, bei reellem c also durchweg mit

$$I \ni x \mapsto \log|x-c| \in \mathbb{R}$$

rechnen. (Was sich natürlich auch direkt durch Ableiten verifizieren läßt.)

Jetzt sei $c \in \mathbb{C}\backslash\mathbb{R}$ und etwa $\mathrm{Im}c < 0$. Für reelle x liegt dann $x-c \in \mathbb{C}$ in der oberen Halbebene, und es bietet sich an, den Logarithmus hier als die Umkehrung von



$$\mathbb{C} \supset \mathbb{R} \times i\,(0,\pi) \quad \xrightarrow{\exp} \quad \mathbb{R} \times i\,(0,\infty) \subset \mathbb{C}$$

zu lesen. Ist die Stammfunktion oder das damit berechnete Integral bloß ein Zwischenergebnis, soll man es damit genug sein lassen und mit dem so präzisierten komplexen Logarithmus weiterrechnen. Wer aber darauf besteht, als Resultat wirklich eine explizit reelle Zahl zu sehen, muß zur Strafe in den sauren Apfel beißen und die folgende Auswertung durchführen.

Da die Ausgangsfunktion f reell ist, treten die nicht-reellen Terme der Partialbruchzerlegung in komplex-konjugierten Paaren auf; wir müssen also dem Ausdruck

$$\alpha\,\log(x-c) + \overline{\alpha}\,\log(x-\overline{c}) \quad (x \in \mathbb{R})$$

einen Sinn geben, wobei wir uns für die Rechnung auf den Fall $\mathrm{Im}c < 0$ wie oben festlegen dürfen. Wenn wir den betrachteten Logarithmus nach Satz 12.16 in

$$\mathbb{C} \supset \mathbb{R} \times i\,(0,\infty) \ni z = x + iy \longmapsto \log|z| + i\,\mathrm{arccot}\,\frac{x}{y} \in \mathbb{R} \times i\,(0,\pi) \subset \mathbb{C}$$

aufschlüsseln, ergibt sich

$$\begin{aligned}
\alpha\,\log(x-c) + \overline{\alpha}\,\log(x-\overline{c}) &= 2\mathrm{Re}\big(\alpha\log(x-c)\big) \\
&= 2\mathrm{Re}\left(\alpha\,\log|x-c| + i\alpha\,\mathrm{arccot}\,\frac{x-\mathrm{Re}c}{0-\mathrm{Im}c}\right) \\
&= 2\mathrm{Re}\alpha\cdot\log\sqrt{(x-\mathrm{Re}c)^2+(\mathrm{Im}c)^2} - 2\mathrm{Im}\alpha\cdot\mathrm{arccot}\,\frac{x-\mathrm{Re}c}{-\mathrm{Im}c} \\
&= \mathrm{Re}\alpha\cdot\log\big((x-\mathrm{Re}c)^2+(\mathrm{Im}c)^2\big) - 2\mathrm{Im}\alpha\cdot\mathrm{arccot}\,\frac{x-\mathrm{Re}c}{-\mathrm{Im}c}.
\end{aligned}$$

Der Ästhetik halber können wir noch die Rollen von (α, c) und $(\overline{\alpha}, \overline{c})$ vertauschen, also

$$\mathbb{R} \ni x \longmapsto \mathrm{Re}\alpha\cdot\log\big((x-\mathrm{Re}c)^2+(\mathrm{Im}c)^2\big) + 2\mathrm{Im}\alpha\cdot\mathrm{arccot}\,\frac{x-\mathrm{Re}c}{\mathrm{Im}c}$$

als Stammfunktion für

$$\mathbb{R} \ni x \longmapsto \frac{\alpha}{x-c} + \frac{\overline{\alpha}}{x-\overline{c}} \in \mathbb{R} \quad \text{mit } \mathrm{Im}c > 0$$

notieren.

Ganz schön kompliziert. Einfacher wird's in Spezialfällen: Für $\alpha \in \mathbb{R}$ bleibt bloß

$$\alpha\cdot\log\big((x-\mathrm{Re}c)^2+(\mathrm{Im}c)^2\big),$$

für rein imaginäres $\alpha = i\beta$

$$2\beta\cdot\mathrm{arccot}\,\frac{x-\mathrm{Re}c}{\mathrm{Im}c}.$$

Ist ganz konkret $\beta = 1$ und $c = i$, so erhalten wir mit

$$-2\int\frac{dx}{x^2+1} = i\int\frac{dx}{x-i} - i\int\frac{dx}{x+i} = 2\,\mathrm{arccot}\,x$$

wieder die aus Beispiel 15.7(2) bekannte Formel.

Soweit zur Integration der rationalen Funktionen. — Allgemein sollte man übrigens nicht die Möglichkeit übersehen, Integrale von als Potenzreihen gegebenen analytischen Funktionen nach der Notiz 15.10 durch gliedweise Integration der Reihe auszuwerten.

Statt reellwertiger Funktionen $\mathbb{R}^n \supset X \xrightarrow{f} \mathbb{R}$ kann man auch komplexwertige integrieren, einfach indem man das für Real- und Imaginärteil einzeln macht; das Resultat ist dann eine komplexe Zahl. Allgemeiner kann man Integrale nicht nur von $\mathbb{R}^p$-wertigen Funktionen, sondern von Abbildungen

$$\mathbb{R}^n \supset X \xrightarrow{f} V$$

mit Werten in einem beliebigen endlichdimensionalen $\mathbb{R}$-Vektorraum V bilden, indem man mittels einer linearen Karte von V die einzelnen Komponenten von f integriert. Daß der so gebildete Vektor $\int_X f \in V$ nicht von der zugrundegelegten Basis abhängt, folgt sofort aus der Linearität des Integrals: Es gilt

$$\left(\int h\circ f\right)_i = \int (h\circ f)_i = \int \sum_j h_{ij} f_j = \sum_j h_{ij} \int f_j = \sum_j h_{ij} \left(\int f\right)_j = \left(h\left(\int f\right)\right)_i$$

für $\mathbb{R}^p$-wertiges f und jedes lineare $h\colon \mathbb{R}^p \longrightarrow \mathbb{R}^q$, also

$$\int h\circ f = h\left(\int f\right).$$

Typisches Beispiel: Eine Massenverteilung im Raum sei durch die räumliche Dichte $\rho\colon \mathbb{R}^3 \longrightarrow \mathbb{R}$ gegeben. Während deren Integration die Gesamtmasse $\int \rho \in \mathbb{R}$ gibt, ist das vektorwertige Integral (mal in Physiker-Notation)

$$\int \rho(\mathbf{r})\mathbf{r}\, d\mathbf{r} \in \mathbb{R}^3$$

der Schwerpunkt der Verteilung, der logischerweise ja auch — unabhängig von irgendeiner Koordinatenwahl — demselben Raum angehören soll, in dem die Verteilung liegt.

Zum Schluß dieses Abschnitts wollen wir noch etwas über die Bedeutung des Begriffs "Integrierbarkeit" plaudern, unter dem Sie sich aufgrund des bisher Gesagten gewiß noch nichts Konkretes vorstellen können. Woran kann es denn liegen, wenn eine Funktion $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ nicht integrierbar ist? Zum Beispiel daran, daß f zu pathologisch ist. So mag man eigentlich kaum erwarten, die Funktion $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ aus Beispiel 1.8(4) mit

$$f(x) = \begin{cases} 1 & \text{falls } x \in \mathbb{Q} \\ 0 & \text{falls } x \notin \mathbb{Q} \end{cases}$$

integrieren zu können: Der Graph von f sieht aus, als wollte er den Streifen $\mathbb{R}\times[0,1]$ ganz ausfüllen; welchen Flächeninhalt sollte man dann der Menge

$$\big\{(x,y) \in \mathbb{R}\times[0,1] \,\big|\, x = 0 \text{ für irrationales } x\big\}$$

schon zuweisen? Aber beeindruckenderweise ist gerade diese Funktion f doch integrierbar mit $\int f = 0$; das ergibt sich nach 31.5(e) sofort daraus, daß $\mathbb{Q} \subset \mathbb{R}$ eine Nullmenge ist.

Es gibt tatsächlich Funktionen, die allein aufgrund ihrer pathologischen Eigenschaften nicht integrierbar sein können; nur sind sie nicht leicht zu finden, und man muß sich geradezu etwas einfallen lassen, um ihre Existenz zu beweisen. (Sie mögen sich gewundert haben, daß im Axiom 31.5(a) nicht nur die über $\{x \in \mathbb{R}^n | g(x) > 0\}$ zu integrierende Funktion f, sondern auch die "Abschneidefunktion" g als integrierbar vorausgesetzt wird. Der Grund ist, daß ein pathologisches g aus einer braven, integrierbaren Funktion f ein ebenfalls pathologisches $f_{\{x\in\mathbb{R}^n | g(x)>0\}}$ machen kann.)

Wenn man es in der Praxis mit einer konkreten Funktion zu tun hat, wird deren Integrierbarkeit aber kaum je an einer derartigen Pathologie scheitern. Vielmehr geht es bei der Frage der Integrierbarkeit in der Regel

um eine versteckte Endlichkeitseigenschaft. Betrachten wir die konstante Funktion $1\colon \mathbb{R} \longrightarrow \mathbb{R}$ als Beispiel. Gemäß 31.5(d) ist für jedes $j \in \mathbb{N}$

$$\int_0^j 1 = j,$$

und aufgrund der Positivität des Integrals müßte im Falle $1 \in \mathcal{L}^1(\mathbb{R})$

$$\int_{-\infty}^{\infty} 1 \geq j \quad \text{für jedes } j \in \mathbb{N}$$

gelten. Das ist natürlich nicht möglich, und die Nichtintegrierbarkeit dieser Funktion leuchtet auch anschaulich unmittelbar ein: der mit dem Integral zu messende Flächeninhalt ist eben nicht endlich! Auch bei beschränktem Definitionsbereich X kann die Integrierbarkeit von f über X noch daran scheitern, daß f nicht beschränkt ist: So kann zum Beispiel die Funktion

$$f\colon (0,1] \longrightarrow \mathbb{R};\ \ x \mapsto \frac{1}{x^2}$$

nicht über $(0,1]$ integrierbar sein, weil dann

$$\int_0^1 f \geq \int_{1/j}^1 \frac{dx}{x^2} = -\frac{1}{x}\,\Big|_{x=1/j}^1 = j-1 \quad \text{für alle } j$$

gelten müßte.

Auf der anderen Seite schließt Unbeschränktheit einer Funktion nicht von vornherein die Integrierbarkeit aus. Wir brauchen das letzte Beispiel nur ein wenig abzuwandeln und die Funktion

$$f\colon (0,1] \longrightarrow \mathbb{R};\ \ x \mapsto \frac{1}{\sqrt{x}}$$

zu betrachten, die zwar auch unbeschränkt ist, aber bei Annäherung an 0 langsamer wächst. Für $\alpha > 0$ ergibt sich hier

$$\int_\alpha^1 f = \int_\alpha^1 \frac{dx}{\sqrt{x}} = 2\sqrt{x}\,\Big|_{x=\alpha}^1 = 2 - 2\sqrt{\alpha}$$

mit $\lim\limits_{\alpha \searrow 0} \int_\alpha^1 f = 2$, und man ist natürlich geneigt, daraus auf die Integrierbarkeit von f und $\int_0^1 f = 2$ zu schließen. Inwieweit das gerechtfertigt ist, werden wir im nächsten Abschnitt untersuchen.

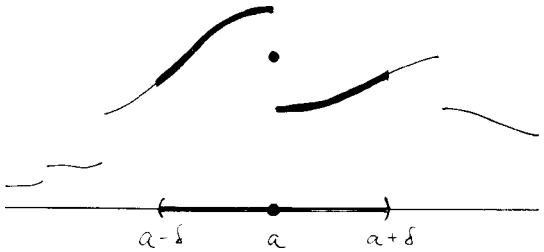
Übungsaufgaben

31.1 Zeigen Sie: Ist $X \subset \mathbb{R}^n$ eine Nullmenge und $Y \subset \mathbb{R}$ eine beliebige Menge, so ist $X \times Y \subset \mathbb{R}^{n+1}$ eine Nullmenge. (Tip: Man braucht dafür nur sehr spezielle Y zu betrachten.)

31.2 Sei $0 < n \in \mathbb{N}$. Konstruieren Sie eine Folge $(W_j)_{j=0}^\infty$ von offenen Würfeln $W_j \subset \mathbb{R}^n$ mit den folgenden Eigenschaften:

- Die Vereinigung $W := \bigcup_{j=0}^\infty W_j$ ist in dem Sinne dicht in $\mathbb{R}^n$, daß W jeden weiteren (nicht-leeren) offenen Würfel in $\mathbb{R}^n$ schneidet.
- Für die Volumina dieser Würfel gilt $\sum_{j=0}^\infty \mu(W_j) < 1$ (was bedeutet, daß die offene dichte Menge W in gewisser Hinsicht zugleich "dünn" ist; es ist klar, daß man die Summe ebensogut kleiner als ein vorgegebenes $\varepsilon > 0$ machen kann).

31.3 Eine auf einem Intervall I definierte Funktion f (mit Werten in $\mathbb{R}^p$) nennt man *stückweise stetig*, wenn es zu jedem $a \in I$ ein $\delta > 0$ gibt, so daß die Einschränkungen von f auf $I \cap (a-\delta, a)$ und $I \cap (a, a+\delta)$ stetig sind und die Grenzwerte $\lim\limits_{x \nearrow a} f(x)$ und $\lim\limits_{x \searrow a} f(x)$, soweit sinnvoll, im eigentlichen Sinne existieren (sie brauchen aber weder miteinander noch mit $f(a)$ übereinzustimmen).



Zeigen Sie, daß jede auf einem kompakten Intervall definierte stückweise stetige Funktion integrierbar ist. — Keine Idee? Lassen Sie sich von Aufgabe 8.6 inspirieren.

31.4 Zeigen Sie, daß die Vorschrift

$$C^0[0,1] \times C^0[0,1] \ni (f,g) \longmapsto \langle f,g \rangle := \int_0^1 fg \in \mathbb{R}$$

ein Skalarprodukt auf dem Funktionenraum $C^0[0,1]$ definiert.

31.5 Berechnen Sie für alle $x, y \in \mathbb{R}$ auf intelligente Weise die unbestimmten Integrale

$$F_{cc}(t) := \int \cos xt \, \cos yt \, dt, \quad F_{cs}(t) := \int \cos xt \, \sin yt \, dt \quad \text{und} \quad F_{ss}(t) := \int \sin xt \, \sin yt \, dt.$$

Kann man es so einrichten, daß F_{cc}, F_{cs} und F_{ss} stetig von $(x, y, t) \in \mathbb{R}^3$ abhängen?

31.6 Sei f eine reelle rationale Funktion. Wie geht man systematisch vor, um eine Stammfunktion von $f \circ \exp$ zu berechnen? Wobei die Exponentialfunktion natürlich auf ein geeignetes Intervall einzuschränken ist. Rechnen Sie das Verfahren für das unbestimmte Integral

$$\int \frac{dx}{e^x - 1}$$

durch.

31.7 Im Grunde genommen nicht anders behandelt man auch Integrale der Form $\int f(e^{ix})dx$. Illustrieren Sie das, indem Sie

$$\int_0^{2\pi} \frac{dx}{5 + 3\cos x}$$

berechnen.

Leitfaden: Dazu ist erst mal der Cosinus durch die Exponentialfunktion auszudrücken, dann das Integral wie in der vorigen Aufgabe auf eines mit rationalem Integranden zurückzuführen. Dessen Partialbruchzerlegung und Integration führt auf einen $\log(\varphi+3)$-haltigen und einen $\log\left(\varphi + \frac{1}{3}\right)$-haltigen Term. Überlegen Sie genau, für welche Werte von φ Sie das brauchen und wie deshalb die komplexen Logarithmen zu interpretieren sind. Warum ist die Antwort für die beiden Terme so wesentlich verschieden? Lesen Sie noch einmal Satz 12.16, aus dem unter anderem hervorgeht, daß man manchmal mehr als eine Formel braucht, um komplexe Logarithmen explizit hinzuschreiben. Ziehen Sie also in Betracht, das Integral vermöge

$$\int_0^{2\pi} = \int_0^{\pi} + \int_{\pi}^{2\pi}$$

in zwei (oder noch mehr) Summanden zu zerlegen.

Wenn Ihnen das alles spanisch vorkommt, haben Sie die komplexe Exponentialfunktion noch nicht verstanden; das wäre ein guter Anlaß, die zweite Hälfte von Abschnitt 12 neu zu lernen.

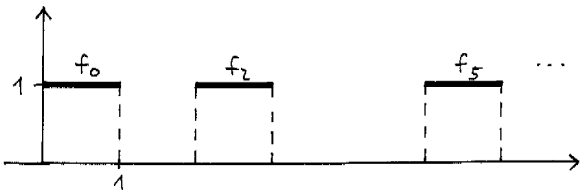
32 Integral und Limes

Wenn man eine Funktion integriert, die der Limes einer Funktionenfolge (f_j) ist, taucht schnell die Frage auf, ob man die Bildung von Limes und Integral miteinander vertauschen darf, also ob man

$$\lim_{j\to\infty}\int f_j = \int \lim_{j\to\infty} f_j$$

schreiben darf. Physiker finden es oft unter ihrer Würde, sich über so was Gedanken zu machen, vor allem dann, wenn sie um jeden Preis ein bestimmtes Resultat erhalten wollen. Die folgenden Beispiele zeigen aber, daß die Frage sehr berechtigt ist.

32.1 Beispiele (1) Für die Folge von Funktionen $f_j\colon\mathbb{R}\longrightarrow\mathbb{R}$ mit $f_j = 1_{[j,\,j+1]}$



gilt offenbar

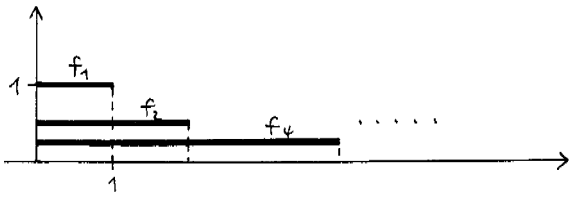
$$\lim_{j\to\infty} f_j = 0, \quad \text{aber} \quad \int f_j = 1 \quad \text{für jedes } j\in\mathbb{N}.$$

Den Grund dafür, daß Integral und Limes hier nicht vertauschbar sind, könnte man darin sehen, daß die Konvergenz der Funktionenfolge gegen die Nullfunktion nur punktweise, aber nicht gleichmäßig ist: es gibt ja offensichtlich kein $j\in\mathbb{N}$ mit $|f_j(x)|<1$ simultan für alle $x\in\mathbb{R}$. Dazu aber das nächste Beispiel:

(2) Wir betrachten die durch

$$f_j = (1/j)_{[0,j]}$$

definierte Funktionenfolge $(f_j)_{j=1}^{\infty}$.



Wegen $|f_j(x)|\le\frac{1}{j}$ für alle $x\in\mathbb{R}$ konvergiert diese Folge sogar gleichmäßig gegen die Nullfunktion, während wie vorhin

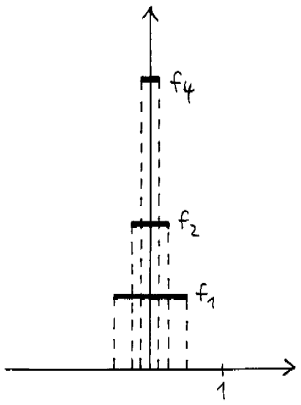
$$\int f_j = \int_0^j \frac{1}{j} = 1 \quad \text{für jedes } j>0$$

ist.

(3) Dieses Beispiel variiert Beispiel (1) in der umgekehrten Richtung. Die Folge $(f_j)_{j=1}^{\infty}$ von Funktionen $f_j\colon\mathbb{R}\longrightarrow\mathbb{R}$ sei durch

$$f_j = j_{[-\frac{1}{2j},\,\frac{1}{2j}]}$$

definiert.



Für jedes $x \neq 0$ ist $f_j(x) = 0$ für $j > \frac{1}{|x|}$; insbesondere konvergiert $f_j(x)$ gegen 0 für $j \to \infty$. Andererseits ist natürlich $\lim f_j(0) = \lim j = \infty$. In diesem Beispiel ist also keine Grenzfunktion $\lim_{j\to\infty} f_j$ erklärt, aber immerhin kann man sagen, daß die Folge (f_j) fast überall gegen die Nullfunktion konvergiert. (Wer damit nicht zufrieden ist, kann alle Werte $f_j(0)$ zu 0 abändern und damit die punktweise Konvergenz *überall* erzwingen.) Vertauschbarkeit von Integral und Limes sollte jedenfalls auch hier $\lim \int f_j = 0$ bedeuten; tatsächlich ist aber

$$\int f_j = \int_{-1/2j}^{1/2j} j = 1 \quad \text{für alle } j > 0.$$

Dieses Beispiel (3) rechtfertigt einen kleinen Exkurs. Wenn Sie einen Physiker nach dem Limes der Folge (f_j) fragen, wird er ohne zu zögern "die Diracsche Delta-Funktion" antworten. In den Physikbüchern findet man denn auch deren Definition; und zwar ist $\delta\colon \mathbb{R} \longrightarrow \mathbb{R}$ durch

$$\delta(x) := 0 \text{ für } x \neq 0, \quad \text{und } \delta(0) \text{ so unendlich, daß trotzdem } \int \delta = 1 \text{ ist}$$

definiert. Natürlich kann es keine solche Funktion geben, denn die erste Forderung zieht ja nach sich, daß das Integral verschwindet. (Auch dadurch, daß man ∞ als Funktionswert zuläßt und den Begriff des Integrals sinnvoll erweitert, läßt sich die Definition nicht retten.) An dieser Stelle gehen die Verfasser der Lehrbücher denn auch getrennte Wege: Manche lassen den Leser mit dieser Definition allein, andere verweisen zu Recht darauf, daß es einen mathematischen Apparat gibt, in dem die Delta-Funktion doch als ein sinnvolles Objekt (aber keine Funktion) vorhanden ist. Am witzigsten sind aber die, die mehr oder weniger elaborate "Konstruktionen" der Delta-Funktion vorführen, die natürlich allesamt Unsinn sind und bestenfalls Beispiele dafür, daß man Integral und Limes eben nicht immer vertauschen kann.

Eine richtige Idee steckt in der Delta-Funktion aber, und ich will zumindest andeuten, wieso. Nehmen wir noch einmal die Funktionen f_j aus Beispiel (3). Statt für die f_j selbst kann man sich auch für eine bestimmte Wirkung interessieren, die sie auf andere Funktionen haben, die man in diesem Zusammenhang *Testfunktionen* nennt. Wir können hier beliebige stetige Funktionen $\varphi\colon \mathbb{R} \longrightarrow \mathbb{R}$ als Testfunktionen nehmen; auf solche wirkt f_j vermöge

$$C^0(\mathbb{R}) \ni \varphi \longmapsto \int f_j\varphi = j \int_{-1/2j}^{1/2j} \varphi \in \mathbb{R};$$

der Testfunktion φ wird ihr Mittelwert über dem Intervall $[-\frac{1}{2j}, \frac{1}{2j}]$ zugeordnet. Gelehrt gesprochen wirkt f_j als ein lineares Funktional auf dem Vektorraum $C^0(\mathbb{R})$. Die sogenannte Delta-Funktion ist nun ihrerseits als ein lineares Funktional auf $C^0(\mathbb{R})$ definiert, und zwar ein ganz einfaches:

$$C^0(\mathbb{R}) \ni \varphi \longmapsto \delta(\varphi) := \varphi(0);$$

d.h. δ wertet die Testfunktionen an der Stelle 0 aus. Die Aussage $\lim\limits_{j\to\infty} f_j = \delta$ erlaubt nun die Interpretation

$$\lim_{j\to\infty} \int f_j\varphi = \delta(\varphi) \quad \text{für jede Testfunktion } \varphi \in C^0(\mathbb{R}),$$

deren einfacher Beweis Ihnen als Übungsaufgabe 32.1 überlassen sei. Lineare Funktionale vom Typ der Delta-Funktion heißen korrekt übrigens *Distributionen*. Sie sind nichts Geheimnisvolles: in der Definition der Delta-Distribution kommt nicht mal der Begriff "unendlich" vor.

Zurück zu unserem eigentlichen Thema. Bisher haben wir nur die Situation betrachtet, daß eine gegebene Funktionenfolge (f_j) schon als konvergent bekannt ist, und haben gefragt, ob man dann den Limes "aus dem Integral" herausziehen darf, insbesondere ob dann die Integralfolge $\left(\int f_j\right)$ konvergiert. Man kann andererseits aber auch hoffen, die Konvergenz einer Funktionenfolge mittels des Integrals zu *erkennnen*. Das liegt daran, daß das Integral des Absolutbetrags einer Funktion zwar nicht alle, aber doch fast alle Eigenschaften einer Norm im Sinne der Folgerung 25.7 hat:

32.2 Definition und **Notiz** Die Funktion

$$\mathcal{L}^1(\mathbb{R}^n) \ni f \longmapsto \|f\| := \int |f| \in \mathbb{R}$$

heißt die (Integral-)Halbnorm des Raumes $\mathcal{L}^1(\mathbb{R})$. Sie hat die Eigenschaften

- $\|f\| \geq 0$ für alle $f \in \mathcal{L}^1(\mathbb{R}^n)$,
- $\|0\| = 0$,
- $\|\lambda f\| = |\lambda| \cdot \|f$ für alle $\lambda \in \mathbb{R}, f \in \mathcal{L}^1(\mathbb{R}^n)$,
- $\|f \pm g\| \leq \|f\| + \|g\|$ für alle $f, g \in \mathcal{L}^1(\mathbb{R}^n)$,

wobei die Dreiecksungleichung nach der Notiz 31.6 aus $|f \pm g| \leq |f| + |g|$ folgt.

Eine wirkliche Norm liegt nicht vor, weil außer der Nullfunktion weitere Funktionen f die Halbnorm $\|f\| = 0$ haben: zumindest diejenigen, die fast überall verschwinden. Nichtsdestotrotz können in völliger Analogie zur Definiton 4.2 erklären:

32.3 Definition Eine Folge $(f_j)_{j=0}^{\infty}$ in $\mathcal{L}^1(\mathbb{R}^n)$ heißt eine Cauchy-Folge (bezüglich der Integralhalbnorm), wenn es zu jedem $\varepsilon > 0$ ein $D \in \mathbb{N}$ gibt mit

$$\|f_{j+l} - f_j\| < \varepsilon \quad \text{für alle } l \in \mathbb{N} \text{ und alle } j > D.$$

Die folgende Eigenschaft des Integrals ist in der vollständigen Darstellung der Integrationstheorie ein Satz. In unserer kurzgefaßten Version mache ich sie zum Inhalt eines weiteren, über die in 31.5 aufgelisteten Integraleigenschaften hinausgehenden Axioms.

32.4 Konvergenzaxiom Für jede Cauchy-Folge $(f_j)_{j=0}^{\infty}$ in $\mathcal{L}^1(\mathbb{R}^n)$ gilt:

- Es gibt eine Teilfolge, die fast überall punktweise konvergiert.
- Je zwei Funktionen $f, g\colon \mathbb{R}^n \longrightarrow \mathbb{R}$, die fast überall Limes einer solchen Teilfolge sind, stimmen fast überall überein und sind integrierbar. Für die Gesamtfolge gilt

$$\lim_{j\to\infty} \|f_j - f\| = 0,$$

insbesondere ist das Integral mit dem Limes vertauschbar:

$$\lim_{j\to\infty} \int f_j = \int f$$

Bemerkungen Die Halbnorm $\|\ \|$ auf $\mathcal{L}^1(\mathbb{R})$ imitiert hier den Absolutbetrag $|\ |$ auf $\mathbb{R}^n$ bei der Beschreibung der Folgenkonvergenz: Die Schlußfolgerung $\lim \|f_j - f\| = 0$ besagt so etwas wie die Konvergenz der Folge

bezüglich der Halbnorm, und das Axiom selbst impliziert eine Art Vollständigkeit von $\mathcal{L}^1(\mathbb{R})$: jede Cauchy-Folge darin konvergiert. — Wie gesagt, hier kein Beweis, aber das "insbesondere" erläutere ich: es ergibt sich nach der Dreiecksungleichung 31.7 aus

$$\left|\int f_j - \int f\right| \le \int |f_j - f| = \|f_j - f\|.$$

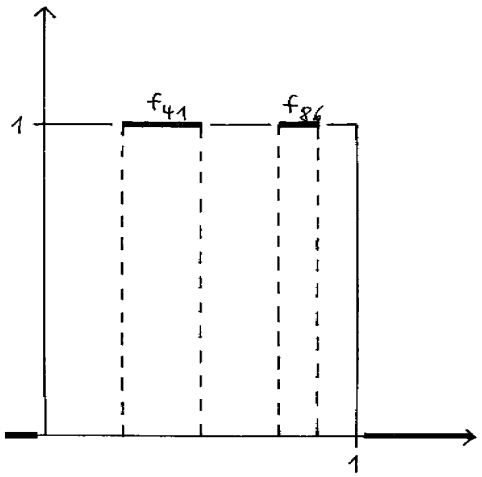
Daß das Axiom über die Konvergenz der Funktionenfolge nur fast überall etwas aussagen kann, ist nicht anders zu erwarten; dagagen mag es überraschen, daß man außerdem zu einer Teilfolge übergehen muß. Das vierte Beispiel in unserer Serie 32.1 erklärt, warum:

(4) Diesmal betrachten wir der Bequemlichkeit halber mittels $m \in \mathbb{N}\backslash\{0\}$ und $k \in \{0, 1, \dots, m-1\}$ doppelt indizierte Funktionen $f_{mk}\colon \mathbb{R} \longrightarrow \mathbb{R}$; sie bilden vermöge der naheliegenden "lexikographischen" Ordnung

$$f_{10};\ f_{20},\ f_{21};\ f_{30},\ f_{31},\ f_{32};\ f_{40},\ \dots$$

trotzdem eine normale Folge. Wir setzen

$$f_{mk} := 1_{\left[\frac{k}{m}, \frac{k+1}{m}\right]}$$



und haben natürlich $f_{mk} \ge 0$, außerdem $\int f_{mk} = \frac{1}{m}$, insbesondere also

$$\lim \|f_{mk}\| = \lim \int f_{mk} = 0.$$

Erst recht liegt damit eine Cauchy-Folge vor. Trotzdem konvergiert die Folge nicht gegen die Nullfunktion, denn zu jedem $x \in [0, 1]$ gibt es unendlich viele Indexpaare (m, k) mit $f_{mk}(x) = 1$.

Ich hatte schon darauf hingewiesen, daß die Integralhalbnorm einer fast überall verschwindenden Funktion natürlich null ist. Wenn man das Integral konstruiert, ergibt sich irgendwann zwischendurch die Erkenntnis, daß diese Aussage auch umgekehrt gilt. Die Umkehrung läßt sich durch ein putziges Argument aber auch formal aus dem Konvergenzaxiom zurückgewinnen:

32.4½ Lemma Für jede Funktion $f \in \mathcal{L}^1(\mathbb{R}^n)$ gilt

$$\|f\| = 0 \quad \Longleftrightarrow \quad f = 0 \text{ fast überall.}$$

Beweis Zu zeigen ist nur, daß eine Funktion f mit $\|f\| = 0$ fast überall verschwindet. Dazu bilden wir die Funktionenfolge, deren Glieder abwechselnd f und die Nullfunktion sind. Das Konvergenzaxiom 32.4

verspricht dann unter anderem, daß die Grenzfunktionen der geraden und der ungeraden Teilfolge fast überall übereinstimmen, d.h. $f = 0$ fast überall ist.

Bemerkung Am liebsten möchte man in der Integrationstheorie zwischen Funktionen, die fast überall gleich sind, gar nicht unterscheiden. Das läßt sich formal dadurch erreichen, daß man alle solchen Funktionen zu einer sogenannten *Äquivalenzklasse* zusammenfaßt (innerhalb einer Klasse unterscheiden sich die Funktionen nur auf Nullmengen) und statt mit den Funktionen mit den Klassen rechnet. Aus dem Vektorraum $\mathcal{L}^1(\mathbb{R}^n)$ wird dadurch ein "gröberer", mit $L^1(\mathbb{R}^n)$ bezeichneter Vektorraum, aus der Integralhalbnorm eine richtige Norm auf $L^1(\mathbb{R}^n)$. Freilich geht beim Übergang von Funktionen zu Klassen etwas verloren, was nach heutiger Auffassung eine Ureigenschaft des Funktionsbegriffs ist: für positives n ist ja jede einpunktige Menge eine Nullmenge, und deshalb kommt einer Äquivalenzklasse integrierbarer Funktionen an keiner einzigen Stelle ein wohlbestimmter Funktionswert zu!

In praktischen Anwendungen benutzt man das Konvergenzaxiom meist in Form der folgenden beiden Sätze, die wir jetzt aus ihm ableiten:

32.5 Satz von der **monotonen Konvergenz** Sei $(f_j)_{j=0}^{\infty}$ eine Folge in $\mathcal{L}^1(\mathbb{R}^n)$, die fast überall monoton wächst:

$$f_0 \le f_1 \le \cdots \le f_j \le f_{j+1} \le \cdots \quad \text{fast überall}$$

Ist die zugehörige Folge der Integrale

$$\left(\int f_j\right)_{j=0}^{\infty}$$

(nach oben) beschränkt, so konvergiert $(f_j)_{j=0}^{\infty}$ fast überall (punktweise) gegen eine Funktion $f \in \mathcal{L}^1(\mathbb{R}^n)$, und es gilt

$$\lim_{j\to\infty} \|f_j - f\| = 0$$

(woraus insbesondere wieder die Vertauschbarkeit von Integral und Limes folgt: $\lim_{j\to\infty} \int f_j = \int f$).

Beweis Zur Formulierung: Die Wörter "nach oben" sind eingeklammert, weil die Folge der Integrale natürlich ebenfalls monoton wächst, und das "punktweise" deswegen, weil nur bei punktweiser Konvergenz der Zusatz "fast überall" einen Sinn hat. — Zum Beweis des Satzes genügt es, die Folge $(f_j)_{j=0}^{\infty}$ als eine Cauchy-Folge in $\mathcal{L}^1(\mathbb{R}^n)$ zu erkennen (bei einer monotonen Folge impliziert die Konvergenz einer Teilfolge schon die der gesamten Folge). Das ist aber ganz einfach: Für beliebige $j, l \in \mathbb{N}$ ist wegen der Monotonie

$$\|f_{j+l} - f_j\| = \int |f_{j+l} - f_j| = \int f_{j+l} - \int f_j = \left|\int f_{j+l} - \int f_j\right|,$$

und die beschränkte monotone Zahlenfolge $\left(\int f_j\right)_{j=0}^{\infty}$ ist konvergent, also ihrerseits eine Cauchy-Folge.

32.6 Satz von der **dominierten Konvergenz** Sei $(f_j)_{j=0}^{\infty}$ eine Folge in $\mathcal{L}^1(\mathbb{R}^n)$, die fast überall gegen eine Funktion $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ konvergiert. Wird diese Folge von einer integrierbaren Funktion g "dominiert", d.h. gibt es ein $g \in \mathcal{L}^1(\mathbb{R}^n)$ mit

$$|f_j| \le g \quad \text{fast überall,}$$

so ist f integrierbar und

$$\lim_{j\to\infty} \|f_j - f\| = 0,$$

also insbesondere $\lim_{j\to\infty} \int f_j = \int f$.

Beweis Indem wir die Werte von f_j und von g auf einer geeigneten Nullmenge zu null machen, dürfen wir annehmen, daß die Voraussetzungen überall (und nicht nur fast überall) erfüllt sind. Wieder kommt es nur darauf an, $(f_j)_{j=0}^{\infty}$ als Cauchy-Folge nachzuweisen. Dazu bilden wir zunächst für jedes $j \in \mathbb{N}$ eine Hilfsfunktion $h_j\colon \mathbb{R}^n \longrightarrow \mathbb{R}$, indem wir

$$h_j(x) := \sup\left\{|f_{j+k}(x) - f_{j+l}(x)| \,\middle|\, k, l \in \mathbb{N}\right\} \quad \text{für jedes } x \in \mathbb{R}^n$$

setzen; wegen der Abschätzung $|f_{j+k} - f_{j+l}| \le 2g$ ist h_j jedenfalls wohldefiniert. Ich zeige darüber hinaus:

Die Funktion h_j ist integrierbar. Tatsächlich ist mit je zwei Funktionen $u, v \in \mathcal{L}^1(\mathbb{R}^n)$ auch die durch

$$\max(u,v)(x) = \max\big\{u(x), v(x)\big\} \quad \text{für jedes } x \in \mathbb{R}^n$$

erklärte Funktion $\max(u,v)$ integrierbar, weil man nämlich listig

$$\max(u,v) = \frac{1}{2}(u+v) + \frac{1}{2}|u-v|$$

schreiben kann. Mittels vollständiger Induktion ergibt sich daraus, daß für jedes $m \ge j$ die Funktion

$$\max_{k,l=0}^{m} |f_{j+k} - f_{j+l}|,$$

in der punktweise das Maximum über alle angegebenen k, l genommen wird, integrierbar ist. Diese Funktionen bilden daher eine Folge in $\mathcal{L}^1(\mathbb{R}^n)$, die offenbar mit m monoton wächst und durch die Funktion $2g \in \mathcal{L}^1(\mathbb{R}^n)$ nach oben beschränkt ist. Ihre Grenzfunktion — die ist aber gerade h_j — ist nach dem vorigen Satz also ebenfalls integrierbar.

Jetzt betrachten wir die Folge $(h_j)_{j=0}^{\infty}$ in $\mathcal{L}^1(\mathbb{R}^n)$. Sie konvergiert gemäß den Voraussetzungen monoton fallend gegen die Nullfunktion. Abermals nach dem Satz von der monotonen Konvergenz folgt daraus

$$\lim_{j\to\infty} \int h_j = \lim_{j\to\infty} \|h_j - 0\| = 0,$$

das heißt insbesondere: Zu jedem $\varepsilon > 0$ gibt es ein $j \in \mathbb{N}$, so daß

$$\|f_{j+k} - f_{j+l}\| = \int |f_{j+k} - f_{j+l}| \le \int \sup_{k,l=0}^{\infty} |f_{j+k} - f_{j+l}| = \int h_j < \varepsilon$$

für alle $k, l \in \mathbb{N}$ gilt. $(f_j)_{j=0}^{\infty}$ ist also eine Cauchy-Folge, und der Beweis damit geführt.

Als erste Anwendung der beiden Konvergenzsätze notieren wir die folgende "Ausschöpfungsmethode":

32.7 Satz Die Menge $X \subset \mathbb{R}^n$ sei Vereinigung einer aufsteigenden Folge $(X_j)_{j=0}^{\infty}$ von Teilmengen:

$$X_0 \subset X_1 \subset \cdots \subset X_j \subset X_{j+1} \subset \cdots \subset \bigcup_{j=0}^{\infty} X_j = X$$

Ist ferner eine über jedes X_j integrierbare Funktion $f\colon X \longrightarrow \mathbb{R}$ gegeben, so gilt: f ist genau dann über X integrierbar, wenn die monoton wachsende Zahlenfolge

$$\left(\int_{X_j} |f|\right)_{j=0}^{\infty}$$

beschränkt ist, und in diesem Fall ist

$$\int_X f = \lim_{j\to\infty} \int_{X_j} f.$$

Beweis Wir setzen $f_j := f_{X_j} \in \mathcal{L}^1(\mathbb{R}^n)$, dann ist $\lim_{j\to\infty} f_j = f_X$. Ist nun f über X integrierbar, so ist nach 31.7 auch $|f|_X = |f_X| \in \mathcal{L}^1(\mathbb{R}^n)$, und wegen $|f_j| \le |f|_X$ gilt die Abschätzung $\int_{X_j} |f| \le \int_X |f|$ für jedes $j \in \mathbb{N}$.

Ist umgekehrt die Beschränktheit der Integralfolge $\left(\int_{X_j} |f|\right)_{j=0}^{\infty}$ bekannt, so genügt die Folge $\big(|f_j|\big)_{j=0}^{\infty}$ dem Satz von der monotonen Konvergenz, sie konvergiert also gegen eine integrierbare Grenzfunktion. Diese

Grenzfunktion ist offensichtlich $|f|_X$, also ist $|f|_X \in \mathcal{L}^1(\mathbb{R}^n)$. Damit eignet sich $|f|_X$ als Dominante für die Folge $(f_j)_{j=0}^\infty$, und aus dem Satz von der dominierten Konvergenz folgt jetzt die Integrierbarkeit von f samt der Vertauschbarkeit von Integral und Limes.

Im Eindimensionalen erlaubt es die Ausschöpfungsmethode, jetzt auch Integrale stetiger Funktionen über nicht-kompakte Intervalle zu analysieren. Es lohnt sich, das noch einmal separat zu formulieren:

32.8 Satz $(a,b) \subset \mathbb{R}$ sei ein offenes (nicht notwendig beschränktes) Intervall, und $f\colon (a,b) \longrightarrow \mathbb{R}$ eine stetige Funktion. Dann gilt: f ist genau dann über (a,b) integrierbar, wenn

$$\lim_{\substack{\alpha \searrow a \\ \beta \nearrow b}} \int_\alpha^\beta |f|$$

im eigentlichen Sinne existiert, und dann ist

$$\int_a^b f = \lim_{\substack{\alpha \searrow a \\ \beta \nearrow b}} \int_\alpha^\beta f.$$

Bemerkungen Weil $\int_\alpha^\beta |f|$ als Funktion von α monoton fällt und als Funktion von β monoton wächst, können Sie den so großzügig notierten Limes hier interpretieren, wie Sie wollen, sei es als kontinuierlichen Limes im Sinne von $(\alpha,\beta) \to (a,b) \in \mathbb{R}^2$, sei es dadurch, daß Sie Folgen (α_j) und (β_j) mit $\lim \alpha_j = a$ und $\lim \beta_j = b$ fixieren und den Grenzwert für $j \to \infty$ anschauen. Dabei geht es wegen der Monotonie in jedem Fall nur um eine Beschränktheitseigenschaft, was die Untersuchung sehr vereinfacht. Achten Sie aber darauf, daß es zum Nachweis der Integrierbarkeit von f über X unerläßlich ist, zunächst Integrale von $|f|$ abzuschätzen, und glauben Sie nicht, in der Praxis werde es auch ohne das schon gutgehen.

32.9 Beispiele (1) Die Funktion $x \mapsto \frac{1}{\sqrt{x}}$ erweist sich jetzt als über $(0,1]$ integrierbar, und es ist

$$\int_0^1 \frac{dx}{\sqrt{x}} = \lim_{\alpha \searrow 0} \int_\alpha^1 \frac{dx}{\sqrt{x}} = \lim_{\alpha \searrow 0} \left(2 - 2\sqrt{\alpha}\right) = 2,$$

wie zum Schluß des vorigen Abschnitt vermutet.

(2) Für die identische Funktion $x \mapsto x$ ergibt sich

$$\int_{-\beta}^\beta x\,dx = \frac{1}{2}x^2\Big|_{x=-\beta}^\beta = \frac{1}{2}\beta^2 - \frac{1}{2}\beta^2 = 0,$$

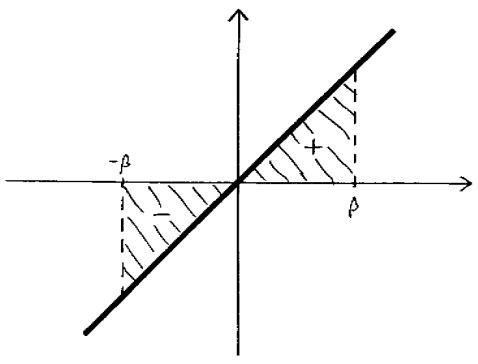
insbesondere

$$\lim_{\beta\to\infty} \int_{-\beta}^\beta x\,dx = 0.$$

Daraus kann man aber keineswegs schließen, daß diese Funktion über $\mathbb{R}$ integrierbar wäre. Vielmehr folgt aus

$$\int_{-\beta}^\beta |x|dx = \int_{-\beta}^0 |x|dx + \int_0^\beta |x|dx = \frac{1}{2}\beta^2 + \frac{1}{2}\beta^2 = \beta^2$$

und $\lim\limits_{\beta\to\infty} \beta^2 = \infty$ nach Satz 32.8 sofort die Nichtintegrierbarkeit. Angesichts des Graphen



leuchtet das sofort ein: Nur wegen der Symmetrie des Integranden *und* des zunächst gewählten Integrationsintervalls $[-\beta,\beta]$ heben sich die positiv und negativ zu zählenden Flächeninhalte gegenseitig weg.

(3) Das Integral

$$\Gamma(x) := \int_0^\infty t^{x-1}e^{-t}dt$$

definiert die sogenannte Gamma-Funktion $\Gamma\colon (0,\infty) \longrightarrow \mathbb{R}$. Weil hier der Integrand keine negativen Werte annimmt, muß man zum Nachweis der Integrierbarkeit nur für die Integrale $\int_\alpha^\beta t^{x-1}e^{-t}dt$ eine von α und β unabhängige obere Schranke finden. Dazu betrachtet man erstens

$$\int_\alpha^1 t^{x-1}e^{-t}dt \le \int_\alpha^1 t^{x-1}dt = \frac{1}{x}t^x\Big|_{t=\alpha}^1 \le \frac{1}{x} \quad \text{für } \alpha \in (0,1),$$

wählt zweitens $D \in (1,\infty)$ so groß, daß

$$t^{x-1} \le e^{t/2} \quad \text{für } t \ge D$$

gilt, folgert daraus für $\beta \in (D,\infty)$

$$\int_D^\beta t^{x-1}e^{-t}dt \le \int_D^\beta e^{-t/2}dt = -2e^{-t/2}\Big|_{t=D}^\beta = -2e^{-\beta/2} + 2e^{-D/2} \le 2$$

und setzt schließlich zusammen:

$$\int_\alpha^\beta \le \int_\alpha^1 + \int_1^D + \int_D^\beta \le \frac{1}{x} + \int_1^D + 2$$

Warum rechnet man übrigens nicht einfach erst das Integral $\int_\alpha^\beta t^{x-1}e^{-t}dt$ aus und sieht nach, was für $\alpha \to 0$ und $\beta \to \infty$ passiert? Nun, das liefe darauf hinaus, eine Stammfunktion des Integranden zu berechnen; der gehört aber zu den nicht elementar integrierbaren Funktionen.

Die Gamma-Funktion ist unter anderem deswegen interessant, weil sie die Fakultäten $k!$ interpoliert:

$$\Gamma(k+1) = k! \quad \text{für jedes } k \in \mathbb{N}$$

(die lästige Verschiebung um 1 ist unglücklicherweise in die etablierte Definition der Gamma-Funktion eingebaut, das mag man heute nicht mehr ändern). Beweis: Aufgabe 32.3.

Bemerkung In früheren Versionen der Integrationstheorie bedurften Integrale von stetigen (und anderen) Funktionen über nicht-kompakte Intervalle einer Sonderbehandlung; sie wurden als *uneigentliche Integrale* bezeichnet, und dieser Ausdruck hat sich bis heute gehalten. Auch hat das Wort "integrierbar" in der älteren (sogenannten Riemannschen) Theorie eine andere, inzwischen obsolete Bedeutung. Wenn Sie mit Literatur arbeiten, müssen Sie unter Umständen beachten, daß unsere Integrierbarkeit in der Sprache der uneigentlichen Integrale der *absoluten* Konvergenz des Integrals entspricht; die dort ebenfalls betrachtete gewöhnliche Konvergenz ist wie bei den Reihen schwierig zu handhaben und auch von untergeordneter Bedeutung.

Das Beispiel der Gamma-Funktion wirft sofort die Frage auf, ob man durch Integration einer Funktion, die stetig von einem "Parameter" abhängt, eine stetige Funktion dieses Parameters erhält. Weil man Stetigkeit nach Satz 7.7 durch die Konvergenz von Folgen ausdrücken kann, ist das letztlich wieder die Frage nach der Vertauschbarkeit von Integral und Limes. Die Konvergenzsätze geben deshalb auch hierzu eine praktisch verwertbare Auskunft. In die folgende Formulierung habe ich auch gleich ein entsprechendes Resultat zur Differenzierbarkeit mit aufgenommen.

32.10 Satz Sei $X \subset \mathbb{R}^n$ eine Teilmenge, und sei $f\colon X \times \mathbb{R}^p \longrightarrow \mathbb{R}$ eine Funktion derart, daß für jedes feste $x \in X$ die Funktion

$$\mathbb{R}^p \ni t \mapsto f(x,t) \in \mathbb{R}$$

integrierbar ist; damit entsteht eine neue Funktion

$$F\colon X \longrightarrow \mathbb{R};\ \ x \mapsto \int f(x,t)dt.$$

Für jedes $t \in \mathbb{R}^p$ bezeichne nun $f_t\colon X \longrightarrow \mathbb{R}$ die durch

$$X \ni x \mapsto f(x,t) \in \mathbb{R}$$

definierte Funktion.

(a) Wenn alle Funktionen f_t an der Stelle $a \in X$ stetig sind und es eine Funktion $g \in \mathcal{L}^1(\mathbb{R}^p)$ mit

$$|f(x,t)| \le g(t) \quad \text{für alle } (x,t) \in X \times \mathbb{R}^p$$

gibt, dann ist F an der Stelle a stetig.

(b) Sei speziell $n = 1$ und $X \subset \mathbb{R}$ ein echtes Intervall. Wenn dann alle f_t stetig differenzierbar sind und es eine Funktion $g \in \mathcal{L}^1(\mathbb{R}^p)$ gibt mit

$$\left|\frac{df}{dx}(x,t)\right| \le g(t) \quad \text{für alle } (x,t) \in X \times \mathbb{R}^p,$$

so sind auch die Ableitungen der f_t integrierbar, F ist stetig differenzierbar und man darf "unter dem Integralzeichen differenzieren":

$$\frac{dF}{dx} = \int \frac{df}{dx}(x,t)dt$$

Beweis Wir testen die Stetigkeit bei a mittels einer gegen a konvergenten Folge $(x_j)_{j=0}^{\infty}$ in X. Wie in (a) vorausgesetzt, konvergiert die durch $f_j(t) := f(x_j,t)$ definierte Funktionenfolge (f_j) punktweise gegen die Funktion $\mathbb{R}^p \ni t \mapsto f(a,t) \in \mathbb{R}$. Wegen $|f_j| \le g$ greift der Satz von der dominierten Konvergenz, und wir schließen

$$\lim_{j\to\infty} F(x_j) = \lim_{j\to\infty} \int f_j = \int f(a,t)dt = F(a).$$

Zum Beweis von (b) prüfen wir die Differenzierbarkeit von F etwa bei $a \in X$. Wir definieren auf $X \times \mathbb{R}^p$ die Hilfsfunktion

$$h\colon (x,t) \mapsto \begin{cases} \dfrac{f(x,t)-f(a,t)}{x-a} & \text{für } x \ne a \\ \dfrac{df}{dx}(a,t) & \text{für } x = a. \end{cases}$$

Nach dem Mittelwertsatz der Differentialrechnung können wir für $x \ne a$

$$h(x,t) = \frac{f(x,t)-f(a,t)}{x-a} = \frac{df}{dx}(\xi,t) \quad \text{mit } \xi \text{ zwischen } a \text{ und } x$$

schreiben, haben also $|h(x,t)| \le g(t)$ für alle $t \in \mathbb{R}^p$. Damit erfüllt h die in (a) an f gestellten Voraussetzungen bis auf die Integrierbarkeit der Funktion $t \mapsto h(a,t)$, die im Beweis aber auch gar nicht benutzt wurde, sondern automatisch folgt. Die Schlußfolgerung von (a) sagt nun

$$\lim_{x\to a} \frac{\int f(x,t)dt - \int f(a,t)dt}{x-a} = \lim_{x\to a} \int \frac{f(x,t)-f(a,t)}{x-a}dt = \lim_{x\to a} \int h(x,t)dt = \int h(a,t)dt = \int \frac{df}{dx}(a,t)dt,$$

also hat F bei a die angegebene Ableitung. Deren Stetigkeit folgt jetzt durch eine weitere Anwendung von (a), diesmal auf die Funktion $(x,t) \mapsto \dfrac{df}{dx}(x,t)$.

Bei der Anwendung dieser Sätze darf man natürlich ausnutzen, daß Stetigkeit und Differenzierbarkeit an einer Stelle a lokale Fragen sind, man deshalb X bei Bedarf mit einer beliebig kleinen Kreisscheibe um a von positivem Radius schneiden darf. Wir greifen etwa das Beispiel der Gamma-Funktion wieder auf: In

$$\Gamma(x) = \int_0^{\infty} t^{x-1}e^{-t}dt$$

ist der Integrand sicher eine C^∞-Funktion von x. Wenn wir Differenzierbarkeitseigenschaften von Γ an der Stelle $a \in (0,\infty)$ testen wollen, dürfen wir x auf ein kleines kompaktes Intervall $[c,d] \subset (0,\infty)$ um a beschränken. Für solches x wird die Funktion $t \mapsto t^{x-1}e^{-t}$ offenbar durch

$$g : t \mapsto \begin{cases} t^{c-1} & \text{für } t \le 1 \\ t^{d-1}e^{-t} & \text{für } t \ge 1 \end{cases}$$

dominiert; und aufgrund der in 32.9(3) etablierten Abschätzungen ist g über $(0,\infty)$ integrierbar. Mit Satz 32.10(a) folgt daraus die Stetigkeit der Gamma-Funktion. Durch wiederholte Anwendung von Teil (b) folgt weiter, daß es sich sogar um eine C^∞-Funktion handelt, denn die höheren Ableitungen

$$\left(\frac{d}{dx}\right)^k t^{x-1}e^{-t} = (\log t)^k t^{x-1}e^{-t}$$

sind ebenso leicht abzuschätzen.

Besonders einfach wird all dies, wenn es bloß um stetiger Funktionen auf einer kompakte Menge geht:

32.11 Satz $X \subset \mathbb{R}^n$ sei der Durchschnitt einer offenen mit einer abgeschlossenen Menge (zum Beispiel X selbst offen oder abgeschlossen), und $T \subset \mathbb{R}^p$ sei kompakt. Ist $f\colon X \times T \longrightarrow \mathbb{R}$ stetig, so ist

$$X \ni x \xrightarrow{F} \int_T f(x,t)\,dt \in \mathbb{R}$$

eine stetige Funktion. Ist die Ableitung nach x

$$(x,t) \mapsto \frac{d}{dx} f(x,t)$$

definiert und eine auf $X \times T$ stetige Funktion, so ist F stetig differenzierbar, und man kann F' durch Differenzieren unter dem Integral berechnen.

Beweis Wie gesagt, ist das eine in X lokale Angelegenheit etwa bei $a \in X$. Ist nun voraussetzungsgemäß $X = F \cap U$ mit abgeschlossenem F und offenem U, so darf man U durch eine ganz in U enthaltene Kugel $D_\delta(a)$, also X durch die kompakte Menge $F \cap D_\delta(a)$ ersetzen. Stetige Funktionen auf der dann kompakten Menge $X \times T$ sind aber beschränkt, und man kann Satz 32.10 mit einer konstanten Funktion g anwenden.

Bemerkung Dieser Satz liegt viel weniger tief als Satz 32.10, ist aber eben auch seltener anwendbar (zum Beispiel schon nicht auf die Gamma-Funktion). Man kann ihn ohne Benutzung des Konvergenzaxioms allein aus den Grundeigenschaften des Integrals herleiten.

Übungsaufgaben

32.1 Zeigen Sie, daß die in Beispiel 32.1(3) betrachtete Funktionenfolge (f_j) mit $f_j = j_{[-\frac{1}{2j},\frac{1}{2j}]}\colon \mathbb{R} \longrightarrow \mathbb{R}$ wie dort behauptet in dem Sinne gegen die Delta-Distribution konvergiert, daß

$$\lim_{j\to\infty} \int f_j \varphi = \delta(\varphi) \quad \text{für jedes } \varphi \in C^0(\mathbb{R})$$

gilt.

32.2 Sei $g\colon \mathbb{R} \longrightarrow \mathbb{R}$ eine Funktion, ferner $(f_j)_{j=0}^\infty$ eine monoton wachsende und $(h_j)_{j=0}^\infty$ eine monoton fallende Folge von integrierbaren Funktionen $f_j, h_j\colon \mathbb{R} \longrightarrow \mathbb{R}$ mit

$$f_j \le g \le h_j \quad \text{für alle } j \in \mathbb{N}$$

(es genügt, wenn diese Voraussetzungen fast überall erfüllt sind).

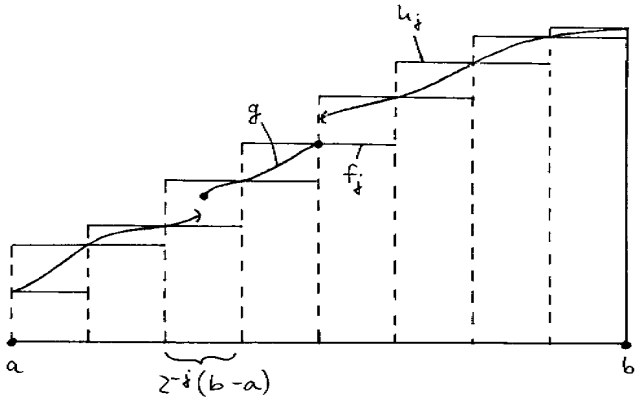
Begründen Sie: Wenn

$$\lim_{j\to\infty}\int_a^b (h_j - f_j) = 0$$

ist, dann ist auch g integrierbar, und es gilt

$$\lim_{j\to\infty} f_j = g = \lim_{j\to\infty} h_j \quad \text{fast überall.}$$

Beweisen Sie als Anwendung, daß jede auf einem kompakten Intervall $[a,b]$ definierte monotone Funktion $g\colon [a,b] \longrightarrow \mathbb{R}$ integrierbar ist:



32.3 Beweisen Sie, daß die Gamma-Funktion aus Beispiel 32.9(3) die Identität

$$\Gamma(x+1) = x\cdot\Gamma(x) \quad \text{für alle } x \in (0,\infty)$$

erfüllt. Zusammen mit dem leicht auszurechnenden Wert $\Gamma(1) = 1$ folgt daraus insbesondere $\Gamma(k+1) = k!$ für jedes $k \in \mathbb{N}$.

32.4 Zeigen Sie, daß

(a) die Funktion $x \mapsto \frac{\cos x}{x^2}$ über das Intervall $[\pi,\infty)$ integrierbar ist,

(b) $\lim\limits_{\beta\to\infty}\int_\pi^\beta \dfrac{\sin x}{x}dx$ im eigentlichen Sinne existiert (benutzen Sie (a)),

(c) die Funktion $x \mapsto \frac{\sin x}{x}$ aber nicht über $[\pi,\infty)$ integrierbar ist.

32.5 Es gibt viele Tricks, um Integrale über ein Intervall auch dann zu berechnen, wenn man keine elementare Stammfumktion des Integranden finden kann. Zwei davon illustriert diese Aufgabe.

(a) Seien $\gamma \in (0,\infty)$ und $\omega \in \mathbb{R}$ beliebig. Für $k \in \mathbb{N}$ bezeichne $p_k(X) \in \mathbb{C}[X]$ das $(k-1)$-te Taylor-Polynom der Funktion $t \mapsto e^{i\omega t}$ an der Stelle 0 (mit $p_0 := 0$). Begründen Sie, warum dann

$$I_k(\omega) := \int_0^\infty e^{-\gamma t}\,\frac{e^{i\omega t} - p_k(t)}{t^k}\,dt \in \mathbb{C}$$

existiert, und berechnen Sie $I_0(\omega)$.

(b) Zeigen Sie, daß I_k eine differenzierbare Funktion (von ω) ist; welche Ableitung hat I_k für $k>0$?

(c) Verwenden Sie (b), um die Integrale

$$\int_0^\infty e^{-\gamma t}\,\frac{\sin t}{t}\,dt = \operatorname{arccot}\gamma$$

und (wenn Sie sich etwas ausgiebiger darin üben wollen)

$$\int_0^\infty e^{-\gamma t}\,\frac{1-\cos t}{t^2}\,dt = \gamma\,\log\gamma - \frac{1}{2}\gamma\,\log(\gamma^2{+}1) + \operatorname{arccot}\gamma$$

explizit auszuwerten.

(d) In den Überlegungen zu (a) bis (c) war die Forderung $\gamma > 0$ wesentlich. Liefern die in (c) hergeleiteten Formeln für $\gamma \searrow 0$ noch eine Information?

32.6 Eine andere Approximation der Delta-Distribution: Zeigen Sie, daß für jedes reelle $x > 0$ und jede beschränkte stetige Funktion $\varphi\colon\mathbb{R}\longrightarrow\mathbb{R}$ das Integral

$$F(x) := \int_{-\infty}^{\infty} \frac{1}{\pi}\frac{x}{x^2+t^2}\varphi(t)dt$$

existiert und daß $\lim\limits_{x\to 0} F(x) = \varphi(0)$ ist.

Anleitung: Beachten Sie erst mal, daß die Limesformel *nicht* aus einer direkten Anwendung der Konvergenzsätze kommen kann. Die (nachzurechnende) Identität $\int_{-\infty}^{\infty}\frac{1}{\pi}\frac{x}{x^2+t^2}dt = 1$ verwandelt die Differenz $F(x)-\varphi(0)$ in ein Integral, dessen Verhalten für $x\to 0$ abzuschätzen ist. Wenn Sie die Stetigkeit von φ bei 0 ausnutzen (mit ε und δ), können Sie das Integral in $\int_{-\infty}^{-\delta}+\int_{-\delta}^{\delta}+\int_{\delta}^{\infty}$ zerlegen und jeden der drei Terme einzeln behandeln.

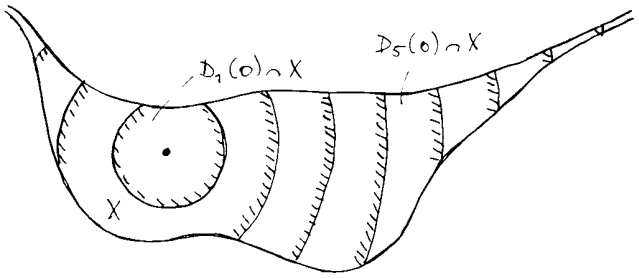
33 Mehrdimensionale Maße und Integrale

Als n-dimensionales Maß oder Volumen der Menge $X \subset \mathbb{R}^n$ hatten wir informell schon die Zahl $\int_X 1$ bezeichnet, vorausgesetzt man kann sie bilden, d.h. die Funktion 1_X ist integrierbar. Wenn das nicht der Fall ist, mag das daran liegen, daß X zu pathologisch ist; es gibt tatsächlich Mengen, die man auf keine vernünftige Art "messen" kann. Es kann aber auch einfach — wie bei $X = \mathbb{R}^n$ — daran liegen, daß X kein endliches Volumen hat. Anders als bei den Integralen, die per definitionem immer endlich sind, sieht man beim Volumen keinen Grund, diesen Fall auszuschließen und solche X für nicht meßbar zu erklären. Die folgende etwas indirekte Definition nimmt darauf Rücksicht.

33.1 Definition Eine Teilmenge $X \subset \mathbb{R}^n$ heißt meßbar, wenn für jedes $r \in \mathbb{N}$ die konstante Funktion 1 über $D_r(0) \cap X$ integrierbar ist. Für jedes solche X nennt man

$$\mu(X) = \mu_n(X) := \lim_{r\to\infty} \int_{D_r(0)\cap X} 1 \in [0, \infty]$$

das (n-dimensionale) Maß von X.



Auch wenn man kaum jemals einer nicht meßbaren Menge explizit begegnen wird, muß man sich mit der Existenz solcher Mengen arrangieren. Das fällt deswegen nicht schwer, weil sehr viele Mengen meßbar sind und weil Mengen, die auf recht großzügige Weise aus gegebenen meßbaren Mengen gebildet werden dürfen, stets wieder meßbar sind. Ich fasse diese Sachverhalte und die wichtigsten Eigenschaften des Maßes zusammen:

33.2 Satz (a) Alle offenen und alle abgeschlossenen Mengen sind meßbar.

(b) Ist $X \subset \mathbb{R}^n$ meßbar, so ist auch das Komplement $\mathbb{R}^n \setminus X$ meßbar. Ist $(X_j)_{j=0}^\infty$ eine Folge meßbarer Mengen, so sind $\bigcap_{j=0}^\infty X_j$ und $\bigcup_{j=0}^\infty X_j$ meßbar; sind diese Mengen paarweise disjunkt, so gilt

$$\mu\left(\bigcup_{j=0}^{\infty} X_j\right) = \sum_{j=0}^{\infty} \mu(X_j).$$

(c) Die Nullmengen sind genau die meßbaren Mengen vom Maß null.

Erläuterung und kurzgefaßter *Beweis* Da das Maß nur nicht-negative Werte annehmen kann, darf man hier in begrenztem Rahmen mit dem Symbol ∞ im Sinne der Limesregeln 9.6 rechnen; zum Beispiel wird man jede Summe, in der mindestens ein Summand unendlich ist, selbst als unendlich lesen. Man muß aber aufpassen, daß man keine Differenzen aus unendlichen Termen bildet. Jedenfalls sind so auch die Formeln gemeint.

Für abgeschlossenes $X \subset \mathbb{R}^n$ folgt die Behauptung (a) sofort daraus, daß dann alle $D_r(0) \cap X$ kompakt, nach 31.5(c) also die Funktionen $1_{D_r(0)\cap X}$ integrierbar sind.

Ist $X \subset \mathbb{R}^n$ meßbar, also für jedes $r \in \mathbb{N}$ die konstante Funktion 1 über die Menge $D_r(0) \cap X$ integrierbar, so ist auch

$$1_{D_r(0)\setminus X} = 1_{D_r(0)} - 1_{D_r(0)\cap X}$$

eine integrierbare Funktion. Wegen $D_r(0) \cap (\mathbb{R}^n \setminus X) = D_r(0) \setminus X$ folgt die Meßbarkeit des Komplements $\mathbb{R}^n \setminus X$, und damit auch die zweite Hälfte von (a).

Die restlichen Behauptungen von (b) beweisen wir zuerst nur für beschränkte Teilmengen von $\mathbb{R}^n$: Der Durchschnitt zweier meßbarer Mengen ist meßbar, weil mit 1_X und 1_Y nach 31.5(a) auch

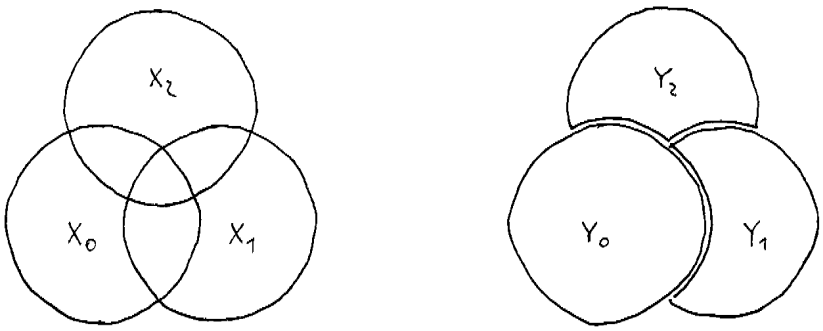
$$1_{X\cap Y} = (1_X)_{\{x\in\mathbb{R}^n | 1_Y > 0\}}$$

integrierbar ist. Nach dem bisher Bewiesenen sind also beliebige endliche Durchschnitte, Vereinigungen und Differenzen von (beschränkten) meßbaren Mengen meßbar; für paarweise disjunkte Mengen addieren sich außerdem die Maße wegen

$$1_{X\cap Y} + 1_{X\cup Y} = 1_X + 1_Y.$$

Eine gegebene Folge meßbarer Mengen $(X_j)_{j=0}^{\infty}$ mit beschränkter Vereinigung wandeln wir vermöge

$$Y_j := X_j \setminus \bigcup_{i=0}^{j-1} X_i \quad \text{für alle } j \in \mathbb{N}$$



in eine solche um, deren Glieder außerdem paarweise disjunkt sind, die aber dieselbe Vereinigung hat:

$$X := \bigcup_{j=0}^{\infty} X_j = \bigcup_{j=0}^{\infty} \left(X_j \setminus \bigcup_{i=0}^{j-1} X_i \right) = \bigcup_{j=0}^{\infty} Y_j.$$

Nach dem Satz von der monotonen (oder dominierten) Konvergenz ist dann

$$1_X = \lim_{k\to\infty} 1_{\bigcup_{j=0}^{k} Y_j} = \lim_{k\to\infty} \sum_{j=0}^{k} 1_{Y_j} = \sum_{j=0}^{\infty} 1_{Y_j}$$

eine integrierbare Funktion, und aufgrund der Vertauschbarkeit von Integral und Limes ergibt sich die in (b) behauptete Summenformel für die Maße. Schließlich befreit man sich von der Beschränktheitsvoraussetzung durch Schneiden mit den Kugeln $D_r(0)$ ($r \in \mathbb{N}$) und Grenzwertbildung: Für jedes r ist der Durchschnitt $D_r(0) \cap X = \bigcup_j (D_r(0) \cap Y_j)$ meßbar, also ist X meßbar; außerdem gilt

$$\mu\big(D_r(0) \cap X\big) = \sum_j \mu\big(D_r(0) \cap X_j\big) \le \sum_j \mu(X_j)$$

und deshalb $\mu(X) \le \sum_j \mu(X_j)$. Andererseits gilt für jedes r und jedes $k \in \mathbb{N}$

$$\mu(X) \ge \mu\Big(\bigcup_{j=0}^{\infty} D_r(0) \cap X\Big) = \sum_{j=0}^{\infty} \mu\big(D_r(0) \cap X_j\big) \ge \sum_{j=0}^{k} \mu\big(D_r(0) \cap X_j\big);$$

das liefert erst mal $\mu(X) \geq \sum_{j=0}^{k} \mu(X_j)$ für jedes k und schließlich die Ungleichung $\mu(X) \geq \sum_{j=0}^{\infty} \mu(X_j)$, die uns noch fehlte.

Die Aussage (c) schließlich ist ein Spezialfall von Lemma 32.4$\frac{1}{2}$: Meßbarkeit von X mit $\mu(X) = 0$ bedeutet dasselbe wie $\|1_X\| = 0$, und das ist nach diesem Lemma dazu gleichwertig, daß X eine Nullmenge ist.

Bemerkungen Aufgrund von Satz 33.2 ist klar, daß man in der Definiton 33.1 statt der kompakten Kugeln $D_r(0)$ ebensogut hätte offene nehmen können, oder auch Würfel oder Quader: es kommt nur darauf an, den ganzen Raum $\mathbb{R}^n$ durch beschränkte meßbare Mengen auszuschöpfen. Natürlich kann auch eine unbeschränkte meßbare Menge X endliches Maß haben; dann ist $\mu(X) = \int_X 1$ und die Limesbildung letztlich überflüssig. — Der eigentliche Fortschritt der modernen Maß- und Integrationstheorie gegenüber der älteren Inhaltstheorie besteht in der in (b) beschriebenen *abzählbaren Additivität* des Maßes, die ja im Gegensatz zur endlichen Additivität nicht ohne weiteres plausibel ist. — Der Meßbarkeitsbegriff erlaubt es, das schwerfällige und wenig plausible Axiom 31.5(a) von der Integrierbarkeit abgeschnittener Funktionen jetzt zu einer Version auszubauen, die theoretisch wie anwendungstechnisch mehr befriedigt:

33.4 Satz (a) Für jedes $g \in \mathcal{L}^1(\mathbb{R}^n)$ ist die Menge

$$g^{-1}(0,\infty) = \{x \in \mathbb{R}^n | g(x) > 0\}$$

meßbar.

(b) Sei $f \in \mathcal{L}^1(\mathbb{R}^n)$ und $X \subset \mathbb{R}^n$ meßbar. Dann ist auch $f_X \in \mathcal{L}^1(\mathbb{R}^n)$.

Beweis Für jedes $r > 0$ ist die Funktion $1_{D_r(0)}$ und damit nach 31.5(a) auch

$$1_{D_r(0)\cap g^{-1}(0,\infty)} = \big(1_{D_r(0)}\big)_{\{x\in\mathbb{R}^n | g(x)>0\}}$$

integrierbar, das beweist (a).

Zu (b): wenn X endliches Maß hat, also 1_X integrierbar ist, dann greift Axiom 31.5(a) direkt:

$$f_X = f_{\{x\in\mathbb{R}^n | 1_X(x)>0\}}$$

Im allgemeinen Fall schöpft man X durch beschränkte meßbare Mengen X_j aus. Dann ist f über jedes der X_j integrierbar und die Folge der Integrale

$$\left(\int_{X_j} |f|\right)_{j=0}^{\infty}$$

durch $\int |f|$ nach oben beschränkt, f also über $X = \bigcup X_j$ integrierbar nach Satz 32.7.

Bemerkung Wenn man von einer gegebenen Funktion f vermutet, daß sie *nicht* integrierbar ist, verhilft einem der vorstehende Satz oft zu einem Beweis: Man wird dann nach einer meßbaren Teilmenge $X \subset \mathbb{R}^n$ suchen, auf der das Verhalten von f besonders leicht zu durchschauen ist, und die Nichtintegrierbarkeit von f_X nachweisen.

Wie man manche Mengen, deren Struktur im einzelnen kaum zu durchschauen ist, durch geschickte Anwendung von Satz 33.2 doch als meßbar erkennen kann, soll das folgende Beispiel illustrieren (das ist der einzige Zweck):

33.4$\frac{1}{2}$ Beispiel Ist $(f_j)_{j=0}^{\infty}$ eine Folge meßbarer Funktionen, so ist die Menge X der Konvergenzpunkte dieser Folge eine meßbare Menge. Denn nach dem Cauchy-Kriterium gehört ein Punkt $x \in \mathbb{R}^n$ zu X, wenn es zu jedem $\varepsilon > 0$ ein $D \in \mathbb{N}$ gibt mit ... Wie wir wissen, genügt es dabei, für ε die Zahlen $1, \frac{1}{2}, \frac{1}{3}, \ldots$ zu nehmen; deshalb ist

$$\begin{aligned} X &= \Big\{x \in \mathbb{R}^n \,\Big|\, \text{zu jedem } j \text{ gibt es ein } D \text{ mit } |f_{k+l}(x) - f_k(x)| < \tfrac{1}{j} \text{ für alle } k > D \text{ und alle } l\Big\} \\ &= \bigcap_{j=1}^{\infty} \bigcup_{D=0}^{\infty} \bigcap_{k=D+1}^{\infty} \bigcap_{l=0}^{\infty} \Big\{x \in \mathbb{R}^n \,\Big|\, \tfrac{1}{j} - |f_{k+l}(x) - f_k(x)| > 0\Big\}. \end{aligned}$$

Nun sind alle Funktionen $|f_{k+l}(x) - f_k(x)|$ integrierbar, und die (nicht integrierbare) konstante Funktion $1/j$ wird durch Abschneiden mit der Kugel $D_r(0)$ integrierbar. Nach Satz 33.4(a) sind alle Durchschnitte $D_r(0) \cap \{x \in \mathbb{R}^n \mid 1/j - |f_{k+l}(x) - f_k(x)| > 0\}$ meßbar, und nach 33.2(b) sind deshalb auch $D_r(0) \cap X$ und schließlich X selbst meßbare Mengen.

Mehrdimensionale Integrale und damit auch mehrdimensionale Maße lassen sich leicht berechnen, weil man sie auf eindimensionale zurückführen kann. Das geschieht mit dem

33.5 Satz von **Fubini** Gegeben seien eine Funktion $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ und eine Zerlegung $n = k + l$; wir schreiben

$$\mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^l \ni (x, y) \overset{f}{\longmapsto} f(x, y) \in \mathbb{R}$$

und definieren für jedes $y \in \mathbb{R}^l$ die Funktion $f_y\colon \mathbb{R}^k \longrightarrow \mathbb{R}$ durch

$$f_y(x) := f(x, y).$$

Wenn f integrierbar ist, so ist für fast alle $y \in \mathbb{R}$ auch f_y integrierbar; die auf einer Nullmenge in $\mathbb{R}^l$ willkürlich ergänzte Funktion

$$\mathbb{R}^l \ni y \longmapsto \int f_y = \int_{\mathbb{R}^k} f(x, y)dx \in \mathbb{R}$$

ist ihrerseits integrierbar, und es gilt die Formel

$$\int f = \int_{\mathbb{R}^l} \left(\int_{\mathbb{R}^k} f(x, y)dx \right) dy.$$

Kommentare Auf den nicht ganz einfachen Beweis soll hier verzichtet werden. — Daß man $f_y \in \mathcal{L}^1(\mathbb{R}^k)$ nur für fast alle und nicht für alle $y \in \mathbb{R}^l$ versprechen kann, ist klar, denn für jede Nullmenge $Y \subset \mathbb{R}^l$ ist $\mathbb{R}^k \times Y$ eine Nullmenge in $\mathbb{R}^n$ (nach Aufgabe 31.1), und man kann f auf dieser Menge beliebig abändern, ohne die Integrierbarkeit zu stören. — In dem doppelten Integral läßt man üblicherweise die Klammern weg, die Regel lautet also, daß die Integrationen von innen nach außen abzuarbeiten sind. In klassischer Notation lautet die Fubini-Formel damit

$$\int_{\mathbb{R}^n} f(x, y)d(x, y) = \int_{\mathbb{R}^l} \int_{\mathbb{R}^k} f(x, y)dx\, dy.$$

Physiker schreiben dx und dy manchmal direkt hinter das zugehörige Integralzeichen und meinen dann mit $\int dy \int dx\, f(x, y)$ oder auch $\int dx\, dy\, f(x, y)$ genau dasselbe. — Selbstverständlich gilt der Satz auch mit vertauschten Rollen von x und y, und man kann ihn wiederholt anwenden, insbesondere um ein mehrdimensionales Integral ganz durch eindimensionale auszudrücken.

Bei der Anwendung des Satzes von Fubini muß man beachten, daß die Integrierbarkeit von f, also die Existenz des n-dimensionalen Integrals, eine Voraussetzung ist. Daß sie auch erfüllt ist, dessen kann man sich a priori nur in besonders einfach gelagerten Fällen sicher sein, vor allem dann (nach 31.5(c) nämlich), wenn der Integrand von der Form f_X mit kompaktem X und stetigem $f\colon X \longrightarrow \mathbb{R}$ ist. Eine häufigere Situtation ist aber die, daß man die Existenz des Doppelintegrals $\int_{\mathbb{R}^l} \int_{\mathbb{R}^k} f(x, y)dx\, dy$ weiß und daraus auf die von $\int_{\mathbb{R}^n} f(x, y)d(x, y)$ schließen möchte oder — vielleicht noch typischer — nur wissen will, daß auch das andere Doppelintegral $\int_{\mathbb{R}^k} \int_{\mathbb{R}^l} f(x, y)dy\, dx$ existiert und denselben Wert hat. Durch zweimalige Anwendung der Fubini-Formel würde das auch sofort folgen, aber solange man die Integrierbarkeit von f nicht weiß, kann man den Satz ja gar nicht anwenden. Das folgende Beispiel zeigt, daß hier ein echtes und nicht ein von übervorsichtigen Mathematikern herbeigeredetes Problem liegt.

33.6 Beispiel Die durch

$$f(x, y) = \frac{y^2 - x^2}{(x^2 + y^2)^2}$$

gegebene Funktion $f\colon (0, 1) \times (0, 1) \longrightarrow \mathbb{R}$ hat bestimmt nichts Exotisches an sich. Nun hat bei festem $y \in (0, 1)$ die Funktion f_y die sogar auf ganz $\mathbb{R}$ erklärte Stammfunktion $x \mapsto \frac{x}{x^2+y^2}$:

$$\frac{d}{dx} \frac{x}{x^2 + y^2} = \frac{(x^2 + y^2) - x(2x)}{(x^2 + y^2)^2} = \frac{y^2 - x^2}{(x^2 + y^2)^2} = f(x, y)$$

Also ist

$$\int_0^1 \int_0^1 f(x,y)dx\,dy = \int_0^1 \left[\frac{x}{x^2+y^2}\right]_{x=0}^1 dy = \int_0^1 \frac{dy}{1+y^2} = \left[\arctan y\right]_{y=0}^1 = \frac{\pi}{4}.$$

Wegen der Symmetrie $f(x,y) = -f(y,x)$ muß für das andere doppelte Integral

$$\int_0^1 \int_0^1 f(x,y)dy\,dx = -\int_0^1 \int_0^1 f(y,x)dy\,dx = -\frac{\pi}{4}$$

gelten. Schlußfolgerung: Zwar sind beide Doppelintegrale sinnvoll, die darin als Integranden auftretenden Funktionen eben integrierbar, aber weil die Integrale verschieden ausfallen, kann die Funktion f selbst nicht integrierbar sein. Ich denke, das Beispiel überzeugt vor allem dadurch, daß überhaupt nichts Auffälliges zu sehen ist, wenn man eines der Doppelintegrale ausrechnet und dann den Fehler begeht, die Reihenfolge der Integrationen zu vertauschen.

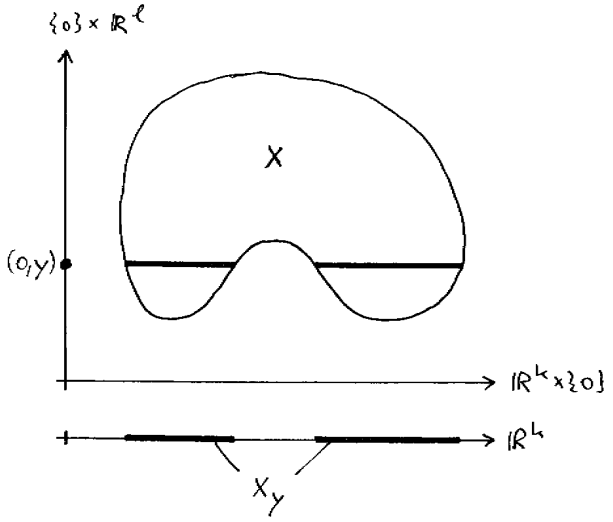
In der schon angesprochenen typischen Situation der Praxis kombiniert man den Satz von Fubini wie folgt mit der Ausschöpfungsmethode 32.7 (Beweis klar):

33.7 Satz Die Menge $X \subset \mathbb{R}^n$ sei Vereinigung einer aufsteigenden Folge $(X_j)_{j=0}^\infty$ von Teilmengen:

$$X_0 \subset X_1 \subset \cdots \subset X_j \subset X_{j+1} \subset \cdots \subset \bigcup_{j=0}^\infty X_j = X$$

Gegeben sei weiter eine Zerlegung $n = k + l$; für jedes $y \in \mathbb{R}^l$ schreiben wir

$$X_y := \left\{x \in \mathbb{R}^k \,\middle|\, (x,y) \in X\right\}.$$



Schließlich sei $f\colon X \longrightarrow \mathbb{R}$ eine Funktion, die über jede der Mengen X_j integrierbar ist. Dann gilt: Genau dann ist f über X integrierbar, wenn die monoton wachsende Zahlenfolge

$$\left(\int_{\mathbb{R}^l} \int_{(X_j)_y} |f(x,y)|dx\,dy\right)_{j=0}^\infty$$

beschränkt ist, und in diesem Fall ist

$$\int_X f = \lim_{j\to\infty} \int_{\mathbb{R}^l} \int_{(X_j)_y} f(x,y)dx\,dy.$$

33.8 Beispiel Die Integrierbarkeit der Funktion

$$(x,y) \mapsto \frac{x}{y^c} \quad (c > 0 \text{ fest})$$

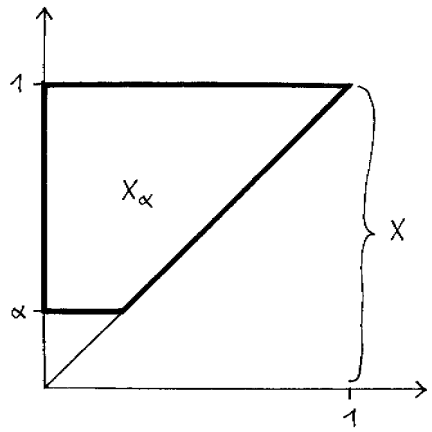
über die Menge

$$X = \left\{(x,y) \in \mathbb{R}^2 \,\middle|\, 0 < x < y < 1\right\}$$

ist zunächst fraglich, nicht aber die über

$$X_\alpha = \left\{(x,y) \in \mathbb{R}^2 \,\middle|\, 0 < x < y < 1,\ y > \alpha\right\},$$

weil X_α sich nur um eine Nullmenge von der kompakten Menge $\left\{(x,y) \in \mathbb{R}^2 \,\middle|\, 0 \le x \le y \le 1,\ y \ge \alpha\right\}$ unterscheidet, auf der der Integrand noch definiert und stetig ist.



Nach Fubini ist daher

$$\int_{X_\alpha} \frac{x}{y^c} d(x,y) = \int_\alpha^1 \int_0^y \frac{x}{y^c} dx\, dy = \int_\alpha^1 \frac{y^2}{2y^c} dy = \left[\frac{1}{2(3-c)} y^{3-c}\right]_{y=\alpha}^1 = \frac{1}{2(3-c)} \left(1 - \alpha^{3-c}\right),$$

jedenfalls für $c \neq 3$. Der Integrand ist überall nicht-negativ, und der Limes für $\alpha \to 0$ existiert genau dann, wenn $c < 3$ ist (auch im Grenzfall $c = 3$ divergiert $\int_\alpha^1 \frac{1}{2y} dy = -\frac{1}{2} \log \alpha$). Genau für $c < 3$ also ist f über X integrierbar, und dann ist

$$\int_X \frac{x}{y^c} d(x,y) = \frac{1}{2(3-c)}$$

der Wert des Integrals.

Es ist klar, daß man mit dem Satz von Fubini auch mehrdimensionale Maße berechnen kann; man braucht ihn ja bloß auf die Funktion 1_X anzuwenden, wenn $X \subset \mathbb{R}^n$ die zu messende Menge ist. Unter Einschluß der Fälle, in denen das Maß unendlich ist, erhält man das bekannte

33.9 Prinzip von **Cavalieri** Sei $X \subset \mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^l$ eine meßbare Menge. Die in 33.7 definierte Menge $X_y \subset \mathbb{R}^k$ ist dann für fast alle $y \in \mathbb{R}^l$ meßbar, und die folgenden Aussagen sind äquivalent:

- $\mu_n(X) < \infty$
- $\mu_k(X_y) < \infty$ für fast alle $y \in \mathbb{R}^l$, und die Funktion $\mathbb{R}^l \ni y \longmapsto \mu_k(X_y) \in \mathbb{R}$ (auf einer μ_l-Nullmenge beliebig ergänzt) ist integrierbar.

Wenn die Aussagen gelten, dann ist

$$\mu_n(X) = \int_{\mathbb{R}^l} \mu_k(X_y) dy.$$

Beweis 33.9 Wenn $\mu(X) < \infty$ vorausgesetzt wird, so liegt offenbar ein Spezialfall des Satzes von Fubini vor. Bezeichnet W_r den kompakten Würfel $W_r = [-r, r]^n$ bzw. $W_r = [-r, r]^k$, so treffen damit beide Aussagen in jedem Fall auf die Mengen $W_r \cap X$ ($r \in \mathbb{N}$) zu. Für jedes $r \in N$ gilt nun

$$W_r \cap X_y = \begin{cases} \emptyset & \text{falls } |y| > r \\ (W_r \cap X)_y & \text{falls } |y| \le r, \end{cases}$$

und insbesondere ist

$$X_y = \bigcup_{r=0}^{\infty} W_r \cap X_y = \bigcup_{r=0}^{\infty} (W_r \cap X)_y$$

für fast alle $y \in \mathbb{R}^l$ eine meßbare Teilmenge von $\mathbb{R}^k$.

Wir setzen nun die zweite Aussage (für X) voraus: die Funktion $\mathbb{R}^l \ni y \mapsto \mu(X_y) \in \mathbb{R}$ sei also fast überall definiert und integrierbar. Wir haben dann

$$\mu(W_r \cap X) = \int_{\mathbb{R}^l} \mu(W_r \cap X)_y \, dy \le \int_{\mathbb{R}^l} \mu(X_y) \, dy \quad \text{für jedes } r \in \mathbb{N}$$

und damit auch

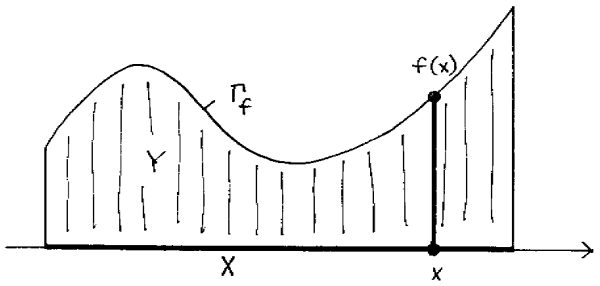
$$\mu(X) = \lim_{r \to \infty} \mu(W_r \cap X) \le \int_{\mathbb{R}^l} \mu(X_y) \, dy;$$

insbesondere ist $\mu(X) < \infty$. Das schließt den Beweis ab.

33.10 Beispiel Sei $X \subset \mathbb{R}^n$ kompakt, und sei $f\colon X \longrightarrow \mathbb{R}$ eine stetige und überall nicht-negative Funktion. Dann hat

$$Y = \left\{ (x, y) \in X \times \mathbb{R} \,\middle|\, 0 \le y \le f(x) \text{ für alle } x \in X \right\} \subset \mathbb{R}^{n+1}$$

das endliche Maß $\mu_{n+1}(Y) = \int_X f$ (so hatten wir ursprünglich ja auch das Integral anschaulich gedeutet).



Tatsächlich ist Y kompakt, und

$$\mu_{n+1}(Y) = \int_X \mu_1([0, f(x)])dx = \int_X f(x)dx$$

nach dem Prinzip von Cavalieri. Es ist klar, wie man die Kompaktheitsvoraussetzung mittels der Ausschöpfungsmethode abmildern kann. (Wenn man sich genauer mit dem Maß befaßt, sieht man, daß f auch nicht stetig zu sein braucht, daß es vielmehr genügt, wenn f über X integrierbar ist.)

Übungsaufgaben

33.1 Sei $X \subset \mathbb{R}^n$ eine meßbare Teilmenge von endlichem Maß, und sei $(f_j)_{j=0}^{\infty}$ eine gleichmäßig konvergente Folge über X integrierbarer Funktionen. Beweisen Sie: Die Grenzfunktion $f := \lim f_j$ ist dann ebenfalls über X integrierbar, und es gilt

$$\lim_{j \to \infty} \int f_j = \int f.$$

33.2 Y und Z seien die beiden Vollzylinder

$$Y = \left\{(x,y,z) \in \mathbb{R}^3 \,\middle|\, x^2 + z^2 \le 1\right\}$$
$$Z = \left\{(x,y,z) \in \mathbb{R}^3 \,\middle|\, x^2 + y^2 \le 1\right\}$$

Berechnen Sie das Volumen von $Y \cap Z$.

33.3 Der starre Körper

$$X = \left\{(x,y,z) \in [0,\infty)^3 \,\middle|\, x + y + z \le 1\right\}$$

habe die konstante Dichte 1. Berechnen Sie das Trägheitsmoment von X bei Rotation um die z-Achse.

33.4 In der folgenden, von den Physikern als Steinerscher Satz bezeichneten Regel wird ein starrer Körper der Masse m betrachtet, der um eine durch seinen Schwerpunkt gehende Achse rotiert. Ersetzt man diese durch eine parallele Achse im Abstand d, so vergrößert sich das Trägheitsmoment um md^2. Beweisen Sie diese Regel.

34 Differenzieren in mehreren Variablen

Wir wollen bei der Differentialrechnung in mehreren Variablen der Einfachheit halber (und bis auf weiteres) nur Abbildungen betrachten, die auf *offenen* Teilmengen von $\mathbb{R}^n$ definiert sind. Zwar bedeutet das selbst im Fall $n = 1$ eine Einschränkung gegenüber dem, was wir im vorigen Semester gemacht haben. Aber einerseits verlieren die jetzt zu erklärenden Konzepte bei allzu beliebigen Definitionsbereichen ihren Sinn, oder zumindest werden schon die einfachsten Sätze darüber falsch, andererseits liegt auf der Hand, welche Modifikationen man für "vernünftige" nicht-offene Definitionsbereiche (zum Beispiel Quader mit positivem Volumen) vornehmen wird.

Die direkte Definition der Ableitung $f'(a)$ als Grenzwert des Diffenzenquotienten $\frac{f(x)-f(a)}{x-a}$ läßt sich nicht auf Funktionen von $n > 1$ Veränderlichen verallgemeinern, weil x und a dann ja Vektoren in $\mathbb{R}^n$ sind, durch die man natürlich nicht teilen kann. Übertragen läßt sich aber das Konzept der Ableitung bei a als der besten linearen Approximation der Abbildung f an dieser Stelle.

34.1 Definition Sei $X \subset \mathbb{R}^n$ offen und $a \in X$. Eine Abbildung $f\colon X \longrightarrow \mathbb{R}^p$ heißt an der Stelle a differenzierbar, wenn es eine lineare Abbildung $l\colon \mathbb{R}^n \longrightarrow \mathbb{R}^p$ gibt, so daß gilt:

- $f(x) = f(a) + l(x-a) + \varphi(x)$ für alle $x \in X$
- $\lim\limits_{x\to a} \dfrac{1}{|x-a|}\varphi(x) = 0$

Erläuterung Die erste Gleichung definiert bloß eine Hilfsfunktion φ, die mißt, um wieviel f von der affinlinearen Abbildung $x \mapsto f(a) + l(x-a)$ abweicht. Die eigentliche Forderung, daß diese Abweichung nämlich für $x \to a$ "schneller als linear" gegen null geht, ist der Inhalt der zweiten Gleichung. Es ist möglich und oft bequem, beide Gleichungen mittels des Landauschen Symbols "o" zu

$$f(x) = f(a) + l(x-a) + o\big(|x-a|\big) \quad \text{für } x \to a$$

zusammenzufassen: Die Definitionen

$$f(x) = o\big(h(x)\big) \quad \Longleftrightarrow \quad \lim \frac{f(x)}{h(x)} = 0$$

$$f(x) = O\big(h(x)\big) \quad \Longleftrightarrow \quad \frac{f(x)}{h(x)} \text{ beschränkt}$$

übertragen sich von $9.6\frac{1}{3}$ ja ohne weiteres auf den Fall einer vektorwertigen Funktion f (während h skalar bleibt).

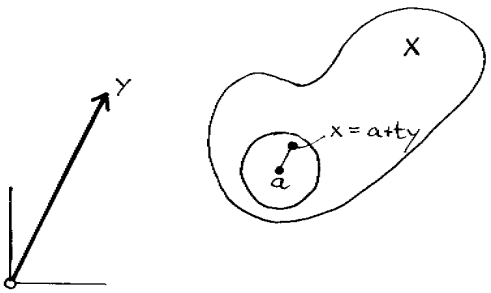
Als nächstes müssen wir klären, was beim Differenzieren eigentlich herauskommt:

34.2 Lemma und **Definition** Die in 34.1 im Falle der Differenzierbarkeit auftretende lineare Abbildung $l\colon \mathbb{R}^n \longrightarrow \mathbb{R}^p$ ist durch f und a eindeutig bestimmt; sie heißt das Differential $Df(a)$ von f an der Stelle a.

Beweis Sei $m\colon \mathbb{R}^n \longrightarrow \mathbb{R}^p$ eine konkurrierende lineare Abbildung; dann gilt

$$l(x-a) - m(x-a) = o\big(|x-a|\big) \quad \text{für } x \to a.$$

Sei nun $y \in \mathbb{R}^n$ fest. Weil X offen ist, ist $x := a + ty \in X$ für alle genügend kleinen reellen $t > 0$,



es ist $|x-a| = t\,|y|$ und wir haben für $t \to 0$

$$l(y) - m(y) = \frac{1}{t}\big(l(ty) - m(ty)\big) = \frac{1}{t}\big(l(x-a) - m(x-a)\big) = \frac{1}{t}o\big(|x-a|\big) = o\left(\frac{1}{t}\,t|y|\right) = o(1),$$

d.h. $l(y) - m(y) = \lim\limits_{t\to 0}\big(l(y) - m(y)\big) = 0$. Weil $y \in \mathbb{R}^n$ beliebig war, heißt das $l = m$.

Wenn man wie üblich die linearen Abbildungen von $\mathbb{R}^n$ nach $\mathbb{R}^p$ mit den reellen $p\times n$-Matrizen identifiziert, wird das Differential von $\mathbb{R}^n \supset X \xrightarrow{f} \mathbb{R}^p$ an der Stelle a zu einer solchen Matrix:

$$Df(a) \in \mathrm{Mat}(p\times n, \mathbb{R})$$

Speziell für $p = n = 1$ ist deren einziger Eintrag $\lambda \in \mathbb{R}$ die Ableitung $f'(a)$ im früheren Sinne, denn dann ist $l(x-a) = \lambda\cdot(x-a)$, und aus der Definition 34.1 wird

$$\frac{f(x) - f(a)}{x - a} = \lambda + o\big(|x - a|\big) \quad \text{für } x \to a,$$

was nur eine Umschreibung des wohlbekannten

$$\lim_{x\to a} \frac{f(x) - f(a)}{x - a} = \lambda = f'(a)$$

ist. Für $n > 1$ ist aber weder der Term "Ableitung" noch die Schreibweise $f'(a)$ gebräuchlich.

Man liest unmittelbar aus den Definitionen ab, daß jede an einer Stelle differenzierbare Abbildung dort auch stetig ist und daß jede affin-lineare Abbildung $\mathbb{R}^n \ni x \mapsto ax + b \in \mathbb{R}^p$ differenzierbar ist und an jeder Stelle das Differential a hat. Darüber hinaus gelten für die Differenzierbarkeit und das Differential aus dem Eindimensionalen entsprechend umzuformulierenden Regeln; unter ihnen sei besonders hervorgehoben die

34.3 Kettenregel Seien $X \subset \mathbb{R}^n$ und $Y \subset \mathbb{R}^p$ offen, $f\colon X \longrightarrow Y$ und $g\colon Y \longrightarrow \mathbb{R}^q$ differenzierbar. Dann ist $g \circ f\colon X \longrightarrow \mathbb{R}^q$ differenzierbar, mit

$$D(g \circ f)(a) = Dg\big(f(a)\big) \circ Df(a).$$

Der Beweis ist nicht schwieriger als der frühere. Interessant ist aber, daß die Aussage der Kettenregel jetzt viel plausibler geworden ist: das Differential der Komposition ist die Komposition der Differentiale. Deswegen habe ich in der Formel auch rechts den Kringel stehen lassen; rechnerisch gesehen handelt es sich um das Matrizenprodukt $Dg\big(f(a)\big)\cdot Df(a) \in \mathrm{Mat}(q\times n, \mathbb{R})$.

Es gibt Situationen, in denen es nicht nur möglich, sondern auch zweckmäßig ist, das Differential einer Abbildung direkt mittels der Definition zu bestimmen:

34.4 Beispiel $f\colon \mathbb{R}^m \times \mathbb{R}^n \longrightarrow \mathbb{R}^p$ sei eine bilineare Abbildung: damit ist natürlich gemeint, daß für feste $a \in \mathbb{R}^m$ und $b \in \mathbb{R}^n$ die Abbildungen $\mathbb{R}^n \ni y \longmapsto f(a, y) \in \mathbb{R}^p$ und $\mathbb{R}^m \ni x \longmapsto f(x, b) \in \mathbb{R}^p$ linear sind. Offenbar ist dann

$$f(x, y) = f\big(a + (x-a), b + (y-b)\big) = f(a, b) + f(x-a, b) + f(a, y-b) + f(x-a, y-b).$$

Man sieht sofort, daß für $(x,y) \to (a,b)$

$$f(x-a,y-b) = O\big(|(x,y)-(a,b)|^2\big) = o\big(|(x,y)-(a,b)|\big)$$

gilt; deshalb ist die lineare Abbildung

$$\mathbb{R}^m \times \mathbb{R}^n \ni (x,y) \longmapsto f(x,b) + f(a,y) \in \mathbb{R}^p$$

das Differential $Df(a,b)$. Beispiele solcher bilinearen Abbildungen sind alle Verknüpfungen, die den Namen "Produkt" verdienen, insbesondere die skalare Multiplikation, die Multiplikation von Matrizen und das Ihnen aus der Physik bekannte Vektorprodukt. Natürlich auch alle symmetrischen Bilinearformen; das Differential der durch $s \in \mathrm{Sym}(n,\mathbb{R})$ bestimmten Form $\mathbb{R}^n \times \mathbb{R}^n \ni (x,y) \longmapsto x^{\mathrm{t}} s\, y \in \mathbb{R}$ an der Stelle (a,b) ist also

$$(x,y) \longmapsto x^{\mathrm{t}} s\, b + a^{\mathrm{t}} s\, y = b^{\mathrm{t}} s\, x + a^{\mathrm{t}} s\, y$$

oder, als Matrix geschrieben

$$\Big(b^{\mathrm{t}} s \;\Big|\; a^{\mathrm{t}} s\Big) \in \mathrm{Mat}(1\times 2n, \mathbb{R}).$$

Daraus ergibt sich auch das Differential der zugehörigen quadratischen Form $x \longmapsto x^{\mathrm{t}} s\, x$ an der Stelle $a \in \mathbb{R}^n$ nach der Kettenregel 34.3, und zwar zu

$$\Big(a^{\mathrm{t}} s \;\Big|\; a^{\mathrm{t}} s\Big) \begin{pmatrix} 1 \\ \hline 1 \end{pmatrix} = 2\, a^{\mathrm{t}} s \in \mathrm{Mat}(1\times n, \mathbb{R}).$$

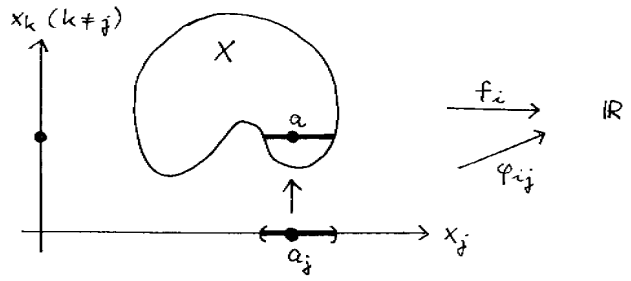
In der Regel stützt man sich bei der praktischen Berechnung von Differentialen aber auf eindimensionale Methoden. Ist wieder $X \subset \mathbb{R}^n$ eine offene Menge, $a \in X$ ein Punkt und $f\colon X \longrightarrow \mathbb{R}^p$ eine Abbildung, so kann man zunächst

$$f = \begin{pmatrix} f_1 \\ \vdots \\ f_p \end{pmatrix}$$

in die p Komponentenfunktionen zerlegen. Für festes $i \in \{1,\dots,p\}$ und festes $j \in \{1,\dots,n\}$ betrachtet man dann die Hilfsfunktion

$$\varphi_{ij}\colon t \mapsto f_i(a_1,\dots,a_{j-1},t,a_{j+1},\dots,a_n) \in \mathbb{R};$$

weil X offen ist, ist φ_{ij} zumindest auf einem offenen Intervall um $a_j \in \mathbb{R}$ definiert.



34.5 Definition Ist φ_{ij} bei a_j differenzierbar, so sagt man, daß die partielle Ableitung von f_i nach der j-ten Variablen an der Stelle a existiert und schreibt sie

$$D_j f_i(a) = \varphi'_{ij}(a_j) \in \mathbb{R}.$$

Existieren diese Ableitungen für alle i (bei festem j), so kann man sie, wenn man will, wieder zur partiellen Ableitung von f

$$D_j f(a) = \begin{pmatrix} \frac{\partial f_1}{\partial x_j}(a) \\ \vdots \\ \frac{\partial f_p}{\partial x_j}(a) \end{pmatrix} \in \mathbb{R}^p$$

zusammenfassen.

Statt der logisch einwandfreien Bezeichnung $D_j f_i(a)$ ist die klassische Schreibweise $\dfrac{\partial f_i}{\partial x_j}(a)$ gebräuchlicher und an sich auch nicht unpraktisch; gelegentlich bringt sie einen aber dadurch in Schwierigkeiten, daß sie an dem festen Namen x_j für die j-te Variable klebt. Die Spalte $D_j f(a)$ schreibt sich damit als

$$D_j f(a) = \frac{\partial f}{\partial x_j}(a) = \begin{pmatrix} \frac{\partial f_1}{\partial x_j}(a) \\ \vdots \\ \frac{\partial f_p}{\partial x_j}(a) \end{pmatrix}.$$

Es liegt nahe, überhaupt alle partiellen Ableitungen von f zu einer $p \times n$-Matrix zusammenzufassen:

34.6 Definition Existieren alle partiellen Ableitungen von f an der Stelle a, so nennt man

$$\frac{df}{dx}(a) = \left(\frac{\partial f_i}{\partial x_j}(a)\right) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \dots & \frac{\partial f_1}{\partial x_n}(a) \\ \vdots & \dots & \vdots \\ \frac{\partial f_p}{\partial x_1}(a) & \dots & \frac{\partial f_p}{\partial x_n}(a) \end{pmatrix} \in \mathrm{Mat}(p \times n, \mathbb{R})$$

die Jacobi-Matrix von f bei a.

Es besteht folgender Zusammenhang mit dem Differential:

34.7 Lemma Ist $f: X \longrightarrow \mathbb{R}^p$ bei a differenzierbar, so existieren alle partiellen Ableitungen von f bei a, und die Jacobi-Matrix stimmt mit dem Differential überein:

$$Df(a) = \left(\frac{\partial f_i}{\partial x_j}\right)$$

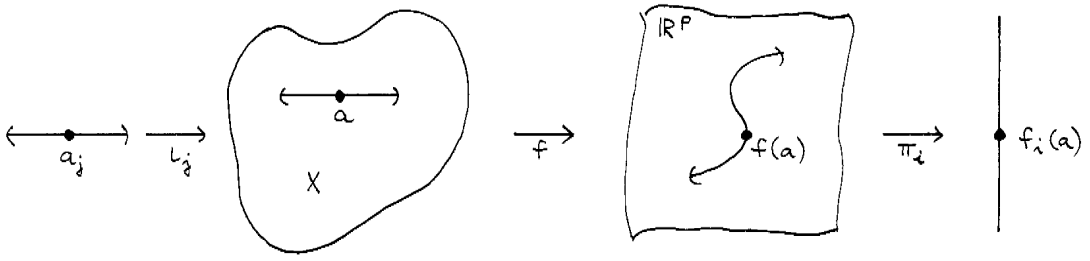
Beweis Wir fixieren i und j; es sei $I \subset \mathbb{R}$ ein offenes a_j enthaltendes Intervall, auf dem die Hilfsfunktion $\varphi_{ij}: t \mapsto f_i(a_1, \dots, a_{j-1}, t, a_{j+1}, \dots, a_n)$ definiert ist. Bezeichnet ι_j die affin-lineare Abbildung

$$\iota_j: \mathbb{R} \longrightarrow \mathbb{R}^n;\ t \mapsto \begin{pmatrix} a_1 \\ \vdots \\ a_{j-1} \\ t \\ a_{j+1} \\ \vdots \\ a_n \end{pmatrix}$$

und π_i die lineare Projektion

$$\pi_i: \mathbb{R}^p \longrightarrow \mathbb{R};\ \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} \mapsto y_i,$$

so können wir φ_{ij} auch gelehrt als $\varphi_{ij} = \pi_i \circ f \circ \iota_j | I$ schreiben.



Aus der Differenzierbarkeit von f bei a folgt nach der Kettenregel die von φ_{ij} bei a_j, mit

$$\frac{\partial f_i}{\partial x_j}(a) = \varphi'_{ij}(a_j) = D\pi_i\big(f(a)\big)\cdot Df(a)\cdot D\iota_j(a_j) = \begin{pmatrix}0 & \dots & 0 & 1 & 0 & \dots & 0\end{pmatrix}\cdot Df(a)\cdot \begin{pmatrix}0\\ \vdots\\ 0\\ 1\\ 0\\ \vdots\\ 0\end{pmatrix} = Df(a)_{ij}.$$

Genau das war die Behauptung.

Aus der Existenz aller partiellen Ableitungen folgt nicht umgekehrt die Differenzierbarkeit. Um das zu verstehen, brauchen wir nur noch mal die Funktion $f\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ aus Beispiel 30.2 anzuschauen:

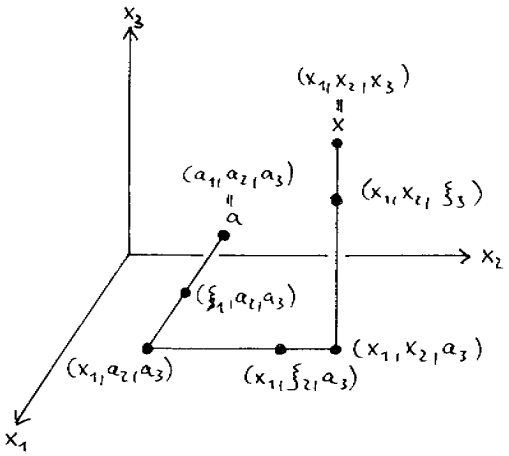
$$f(x,y) = \begin{cases} 0 & \text{für } x = y = 0\\ \frac{xy}{x^2+y^2} & \text{sonst}\end{cases}$$

Die beiden partiellen Ableitungen $\frac{\partial f}{\partial x}$ und $\frac{\partial f}{\partial y}$ existieren an jeder Stelle, auch im Nullpunkt, denn f verschwindet auf dem Achsenkreuz $\big(\mathbb{R}\times\{0\}\big)\cup\big(\{0\}\times\mathbb{R}\big)$ identisch. Andererseits hatten wir schon bemerkt, daß f im Nullpunkt nicht mal stetig, geschweige denn differenzierbar ist. Um so bemerkenswerter ist der folgende

34.8 Satz Sei $X \subset \mathbb{R}^n$ offen, und sei $f\colon X \longrightarrow \mathbb{R}^p$ sei eine Abbildung, deren sämtliche partiellen Ableitungen existieren und stetige Funktionen auf X sind. Dann ist f differenzierbar.

Beweis Direkt aus der Definition folgt, daß f genau dann differenzierbar ist, wenn jede seiner Komponenten $f_i\colon X \longrightarrow \mathbb{R}$ das ist. Wir dürfen uns im Beweis also auf den Fall $p = 1$ beschränken.

Sei $a \in X$ fest, und $U_\delta(a)$ eine ganz in X enthaltene Kugel um a. Wir stellen uns vor, wir wandern von a nach $x \in U_\delta(a)$ auf einem Weg, der aus n achsenparallelen Stücken zusammengesetzt ist:



Dieser Weg verläuft ganz in $U_\delta(a)$ und damit in X, und es ist

$$f(x) - f(a) = \sum_{j=1}^{n} \big(f(x_1,\dots,x_j,a_{j+1},\dots,a_n) - f(x_1,\dots,x_{j-1},a_j,\dots,a_n)\big).$$

Wir wenden auf den j-ten Summanden den Mittelwertsatz der Differentialrechnung 14.1 bezüglich der Variablen x_j an: er liefert eine Zahl ξ_j zwischen a_j und x_j mit

$$f(x_1,\dots,x_j,a_{j+1},\dots,a_n) - f(x_1,\dots,x_{j-1},a_j,\dots,a_n) = \frac{\partial f}{\partial x_j}(x_1,\dots,x_{j-1},\xi_j,a_{j+1},\dots,a_n)\cdot(x_j - a_j).$$

Weil der Grenzübergang $x \to a$ auch $\xi_j \to a_j$ für jedes j nach sich zieht, folgt aus der Stetigkeit der partiellen Ableitungen bei a

$$\frac{\partial f}{\partial x_j}(x_1, \dots, x_{j-1}, \xi_j, a_{j+1}, \dots, a_n) - \frac{\partial f}{\partial x_j}(a_1, \dots, a_n) = o(1)$$

und weiter

$$\begin{aligned} f(x) - f(a) &= \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(x_1, \dots, x_{j-1}, \xi_j, a_{j+1}, \dots, a_n) \cdot (x_j - a_j) \\ &= \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(a_1, \dots, a_n) \cdot (x_j - a_j) + o\big(|x-a|\big). \end{aligned}$$

Damit ist die Differenzierbarkeit von f bei a bewiesen.

Zusammenfassend kann man sagen, daß mit der bloßen Existenz der partiellen Ableitungen von f wenig anzufangen ist, daß dagegen Existenz *und* Stetigkeit dieser Ableitungen zusammen einen Hauch stärker als die Differenzierbarkeit sind. Naheliegenderweise nennt man Abbildungen mit diesen beiden Eigenschaften stetig differenzierbar oder C^1-Abbildungen, wie im Eindimensionalen. Beachten Sie, daß die Stetigkeitsforderung sich auf die einzelne partielle Ableitung $D_j f$ als Funktion *aller* Variablen und nicht nur der j-ten bezieht! Trotzdem ist Satz 34.8 ein robustes Hilfsmittel, um vielen durch Formeln gegebenen Abbildungen ihre Differenzierbarkeit direkt anzusehen und ihr Differential in Gestalt der Jacobi-Matrix zu berechnen.

34.9 Beispiel Die Polar- oder Kugelkoordinatenabbildung

$$\mathbb{R}^3 \ni \begin{pmatrix} r \\ \theta \\ \varphi \end{pmatrix} \overset{\Phi}{\longmapsto} \begin{pmatrix} r\sin\theta\cos\varphi \\ r\sin\theta\sin\varphi \\ r\cos\theta \end{pmatrix} \in \mathbb{R}^3$$

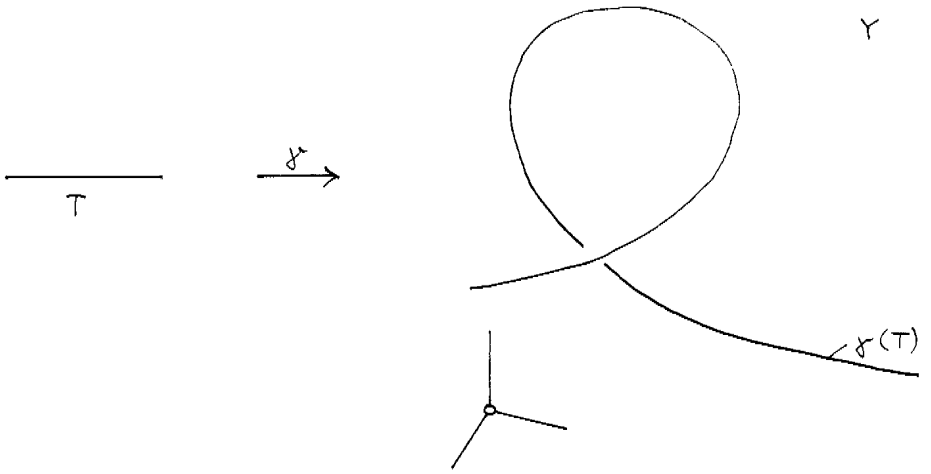
ist stetig differenzierbar, und ihr Differential bei (r, θ, φ) ist

$$D\Phi(r,\theta,\varphi) = \begin{pmatrix} \sin\theta\cos\varphi & r\cos\theta\cos\varphi & -r\sin\theta\sin\varphi \\ \sin\theta\sin\varphi & r\cos\theta\sin\varphi & r\sin\theta\cos\varphi \\ \cos\theta & -r\sin\theta & 0 \end{pmatrix} \in \mathrm{Mat}(3\times 3, \mathbb{R}).$$

Wir wollen noch ein wenig über den Spezialfall $n = 1$ (aber $p \in \mathbb{N}$ beliebig) reden. Tatsächlich unterscheidet der sich kaum von Früherem, ist ja auch klar, weil man es im wesentlichen mit p Funktionen *einer* Veränderlichen zu tun hat. Neu sind erst mal nur einige der Situation angepaßte und besonders suggestive Vokabeln.

34.10 Definition Sei $Y \subset \mathbb{R}^p$. Unter einer (parametrisierten) Kurve in Y versteht man eine stetige Abbildung $\gamma\colon T \longrightarrow Y$ von einem Intervall $T \subset \mathbb{R}$ nach Y.

Erläuterungen Nicht nur Physiker, sondern auch wir Mathematiker stellen uns dabei gern T als ein Zeitintervall vor, und γ als die Bahn eines Massenpunktes, der sich in $Y \subset \mathbb{R}^p$ bewegt.



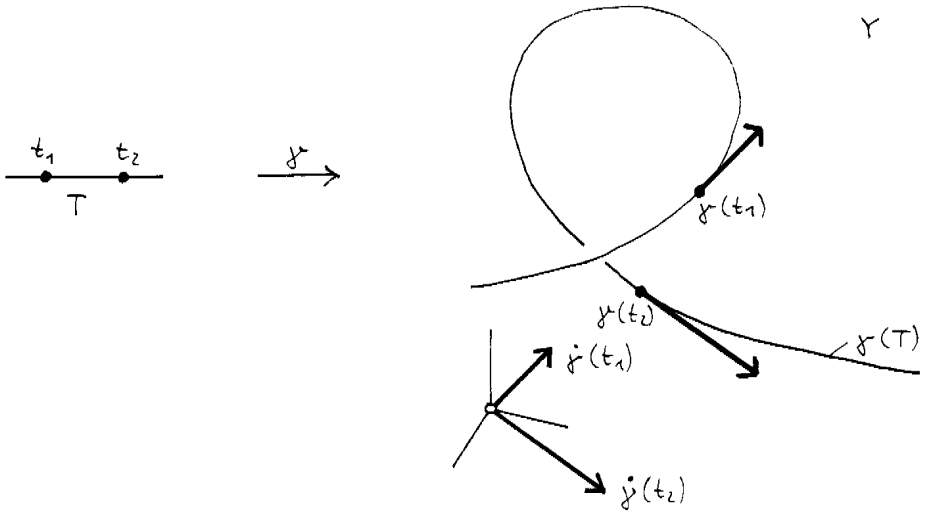
Wenn Ihnen diese Interpretation für $p > 3$ zu weit hergeholt erscheint, denken Sie wie in Beispiel 25.8(2) an n Massenpunkte, deren Bahnen in $\mathbb{R}^3$ man durch Zusammenfassen der Komponenten als eine einzige Bahn in $\mathbb{R}^{3n}$ ansehen kann. — Der Zusatz "parametrisiert" dient dazu, die so bezeichneten Kurven gegen einen anderen Kurvenbegriff abzugrenzen, den wir hier nicht benutzen, der aber in der Mathematik auch wichtig ist und für gewisse Teilmengen von $\mathbb{R}^p$ steht, die eben wie eine Kurve im anschaulichen Sinne aussehen. Etwa würde man erwarten, daß die Bildmenge einer parametrisierten Kurve unter diesen Begriff fällt; das ist aber nicht immer so. Jedenfalls enthält γ in aller Regel mehr Information als die Bildmenge $\gamma(T)$: bei den Planetenbahnen zum Beispiel läßt sich aus dieser ja nur das erste, nicht das zweite oder dritte Keplersche Gesetz ablesen. — Kurven mit dem Definitionsintervall $T = [0, 1]$ hatten wir früher Wege genannt; diese Bezeichnung verwendet man bei Bedarf auch allgemeiner für Kurven, die auf einem beliebigen kompakten Intervall definiert sind.

Wenn das Definitionsintervall einer Kurve $T \xrightarrow{\gamma} Y \subset \mathbb{R}^p$ ein echtes Intervall ist, also nicht nur aus einem einzelnen Punkt besteht, können wir von differenzierbaren Kurven und ihrer Ableitung reden. Die nur für offene T formulierte neue Definition brauchen wir dazu nicht unbedingt, denn wir können alternativ die p Komponentenfunktionen $\gamma_i: T \longrightarrow \mathbb{R}$ einzeln betrachten oder auch gleich vektorwertige Differenzenquotienten

$$\frac{1}{\tau - t}\big(\gamma(\tau) - \gamma(t)\big) \in \mathbb{R}^p$$

bilden. Deshalb geben hier auch höhere Ableitungen und C^k-Kurven für $k = 0, 1, \ldots$ und $k = \infty$ ohne weiteres Sinn. Von den Physikern übernehmen wir für die abgeleiteten Kurven gerne die

34.11 Sprechweise und **Notation** Die Ableitung der differenzierbaren Kurve $\gamma: T \longrightarrow Y$ an der Stelle t nennt man ihren Geschwindigkeitsvektor $\dot\gamma(t) \in \mathbb{R}^p$ zur Zeit t,



die zweite Ableitung (falls definiert) den Beschleunigungsvektor $\ddot\gamma(t) \in \mathbb{R}^p$. Ist γ eine C^1- bzw. C^2-Kurve, so entstehen damit weitere Kurven

$$\dot\gamma: T \longrightarrow \mathbb{R}^p \quad \text{und} \quad \ddot\gamma: T \longrightarrow \mathbb{R}^p.$$

Anmerkungen Der Zusammenhang mit den sonst benutzten Bezeichnungen ist natürlich

$$\dot\gamma(t) = \frac{d\gamma}{dt}(t) = D\gamma(t)\cdot 1 = D\gamma(t);$$

das letzte Gleichheitszeichen steht — wenn man ganz pingelig sein will — dafür, daß man eine auf dem Körper K definierte lineare Abbildung mit ihrem Wert an der Stelle $1 \in K$ identifiziert. — Der Definition nach ist der Geschwindigkeitsvektor $\dot\gamma(t)$ als Punkt in $\mathbb{R}^p$ oder gleichwertig als von 0 ausgehender Pfeil zu zeichnen; anschaulicher ist es aber, sich ihn an der Stelle $\gamma(t)$ an die Kurve "geheftet" vorzustellen, wie ich es auch in der Skizze als zweite Möglichkeit gezeigt habe. Beachten Sie in jedem Fall, daß $\dot\gamma(t)$ keine

Veranlassung hat, wieder zu Y zu gehören und daß deshalb $\dot\gamma$ und $\ddot\gamma$ im allgemeinen nur Kurven in $\mathbb{R}^p$, nicht in Y sind.

Der duale Spezialfall, der einer differenzierbaren Funktion $f\colon X \longrightarrow \mathbb{R}$ mit offenem $X \subset \mathbb{R}^n$, ist in der Physik vielfach durch ortsabhängige "skalare" Größen realisiert. An jeder Stelle $a \in X$ ist das Differential von f eine Linearform

$$Df(a)\colon \mathbb{R}^n \longrightarrow \mathbb{R},$$

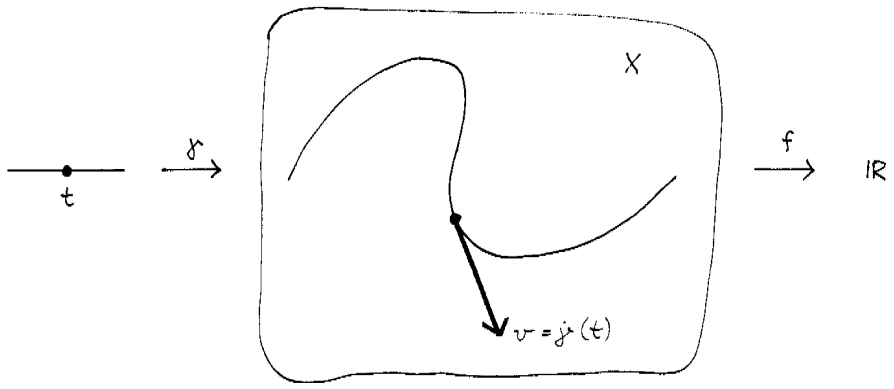
repräsentiert durch eine Zeilenmatrix der Länge n. Durch Einsetzen eines Vektors $v \in \mathbb{R}^n$ erhält man eine Zahl

$$Df(a)v = \begin{pmatrix} \frac{\partial f}{\partial x_1}(a) & \dots & \frac{\partial f}{\partial x_n}(a) \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(a) v_j \in \mathbb{R}.$$

Die Bedeutung dieser Zahl illustriert am schönsten die unmittelbar aus der Kettenregel 34.3 hervorgehende

34.12 Notiz $X \subset \mathbb{R}^n$ sei offen, die Kurve $\gamma\colon T \longrightarrow X$ und die Funktion $f\colon X \longrightarrow \mathbb{R}$ seien differenzierbar. Für jedes $t \in T$ gilt dann

$$(f\circ\gamma)'(t) = Df\big(\gamma(t)\big)\,\dot\gamma(t).$$



Man erhält also $Df(a)v \in \mathbb{R}$, indem man v als Geschwindigkeitsvektor einer durch a laufenden Kurve realisiert und die Veränderung von f längs dieser Kurve durch Ableiten "testet". Man nennt $Df(a)v$ deshalb schon mal die Ableitung von f an der Stelle a *längs* v. Früher hat man das nur für Vektoren v der Länge 1 gemacht und dann von der Richtungsableitung nach v gesprochen, was zwar besonders anschaulich, aber ungeschickt ist, weil dabei der wichtige Aspekt verlorengeht, daß $Df(a)v$ linear von v abhängt. Übrigens erhält man speziell für die achsenparallelen Kurven mit Geschwindigkeit eins wieder die partiellen Ableitungen von f, wie es ja auch sein muß.

Was die soweit zur Differentialrechnung eingeführten Begriffe betrifft, lassen sich die Räume $\mathbb{R}^n$ und $\mathbb{R}^p$ ohne weiteres durch beliebige endlichdimensionale $\mathbb{R}$-Vektorräume V und W zu ersetzen. Das Differential einer Abbildung $V \supset X \xrightarrow{f} W$, die an einer Stelle $a \in X$ differenzierbar ist, ist dann eine lineare Abbildung $Df(a)\colon V \longrightarrow W$; es ist reine Geschmackssache, ob man das aufgrund einer entsprechend allgemeineren Formulierung der Definitionen einsieht oder ob man sich mittels Basen auf den konkreten Fall zurückzieht und dann nachrechnet, daß $Df(a)$ von der Wahl dieser Basen nicht abhängt. Anders ist das bei dem jetzt einzuführenden Begriff des Gradienten: er hängt offenbar nicht nur von der linearen, sondern auch von der euklidischen Struktur von $\mathbb{R}^n$ ab.

34.13 Definition $X \subset \mathbb{R}^n$ sei offen, und $f\colon X \longrightarrow \mathbb{R}$ sei bei $a \in X$ differenzierbar. Der durch die Forderung

$$Df(a)v = \langle \operatorname{grad} f(a), v\rangle \quad \text{für alle } v \in \mathbb{R}^n$$

gemäß Satz 27.11 festgelegte Vektor $\operatorname{grad} f(a) \in \mathbb{R}^n$ heißt der Gradient von f an der Stelle a. Wenn man den durch das Standardskalarprodukt bestimmten Isomorphismus $\mathbb{R}^n \simeq (\mathbb{R}^n)^\vee$ wie in 27.11 mit Σ bezeichnet, ist also $Df(a) = \Sigma\big(\operatorname{grad} f(a)\big)$.

Die in dieser Form auf einen beliebigen euklidischen Vektorraum übertragbare Definition reduziert sich in Matrizensprache natürlich einfach auf

$$\operatorname{grad} f(a) = Df(a)^{\mathrm{t}} \in \mathbb{R}^n.$$

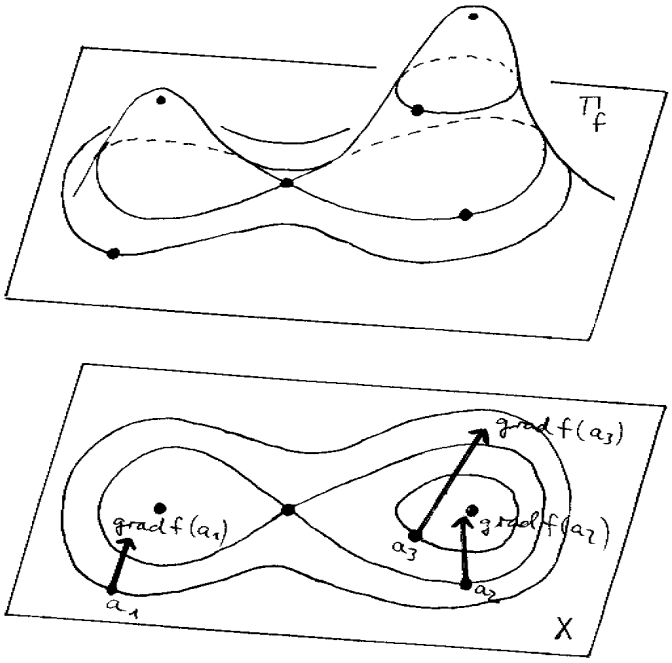
In der physikalischen und vor allem der Ingenieurliteratur wird denn auch häufig $Df(a)$ (für eine skalare Funktion f) selbst als der Gradient angesehen. Das ist aber mit den Konventionen des Matrizenkalküls unverträglich und im allgemeinen ungeschickt. Vorteilhaft erscheint es nur demjenigen, der sich aus Angst vor den Linearformen unbedingt an die vermeintlich vertrauenswürdigeren Vektoren klammern möchte. Wenn man sich übrigens statt in $\mathbb{R}^3$ in der vierdimensionalen Raum-Zeit, also dem Minkowski-Raum $\mathbb{R}^4$ mit dem Standardprodukt bewegt, ist der Gradient logischerweise mittels des Minkowski-Produktes zu bilden; er entsteht dann nicht einfach durch Transponieren von $Df(a)$:

$$\operatorname{grad} f(a) = \begin{pmatrix} \frac{\partial f}{\partial x_0}(a) \\ -\frac{\partial f}{\partial x_1}(a) \\ -\frac{\partial f}{\partial x_2}(a) \\ -\frac{\partial f}{\partial x_3}(a) \end{pmatrix}$$

Immerhin aber hat der Gradient im euklidischen Fall eine interessante anschauliche Bedeutung, die ich Ihnen nicht vorenthalten möchte. Ist $v \in \mathbb{R}^n$ ein Vektor mit $|v| = 1$, so liefert die Schwarzsche Ungleichung 25.6

$$Df(a)v = \langle \operatorname{grad} f(a), v\rangle \le |\operatorname{grad} f(a)|\cdot|v| = |\operatorname{grad} f(a)|,$$

und Gleichheit tritt genau dann ein, wenn $\operatorname{grad} f(a)$ ein nicht-negatives Vielfaches von v ist. Der Gradient gibt deshalb die Stärke, sowie, falls nicht null, auch die Richtung des steilsten Anstiegs der Funktion f an:



Zum Schluß dieses Abschnitts wollen wir noch komplexes Differenzieren in einer mit reellem Differenzieren in zwei Variablen vergleichen.

34.14 Lemma Sei

$$X \ni z = x+iy \overset{f}{\longmapsto} f(z) = u(z)+iv(z) \in \mathbb{C}$$

eine auf der offenen Menge $X \subset \mathbb{C}$ erklärte komplexwertige Funktion. Diese Funktion ist an der Stelle $c \in X$ genau dann komplex differenzierbar, wenn sie als reelle Abbildung dort differenzierbar ist und ihre partiellen Ableitungen die sogenannten Cauchy-Riemann-Gleichungen

$$\frac{\partial u}{\partial x}(c) - \frac{\partial v}{\partial y}(c) = 0$$
$$\frac{\partial u}{\partial y}(c) + \frac{\partial v}{\partial x}(c) = 0$$

erfüllen.

Beweis Die komplexe Differenzierbarkeit bedeutet

$$f(z) = f(c) + l(z-c) + o\big(|z-c|\big) \quad \text{für } z \to c$$

mit einer komplex-linearen Abbildung $l\colon \mathbb{C} \longrightarrow \mathbb{C}$. Die zu solchen l gehörigen reellen 2×2-Matrizen sind nach Lemma 23.2 gerade diejenigen der Form

$$\begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}.$$

Die reelle Differenzierbarkeit bedeutet genau das gleiche, aber mit reell-linearem l, d.h. ohne Forderungen an die Matrix. Diese ist aber in jedem Fall die Jacobi-Matrix

$$Df(c) = \begin{pmatrix} \frac{\partial u}{\partial x}(c) & \frac{\partial u}{\partial y}(c) \\ \frac{\partial v}{\partial x}(c) & \frac{\partial v}{\partial y}(c) \end{pmatrix},$$

und man liest die Behauptung ab.

Übungsaufgaben

34.1 Berechnen Sie die erste Ableitung f' der durch

$$f(x) := \int_{\cos x}^{\sin x} e^{xt^2}\,dt$$

definierten Funktion $f\colon \mathbb{R} \longrightarrow \mathbb{R}$.

34.2 Begründen Sie, warum die Funktion

$$\det\colon \mathrm{Mat}(n\times n, \mathbb{R}) \longrightarrow \mathbb{R}$$

stetig differenzierbar ist, und berechnen Sie ihr Differential an der Stelle $1 \in \mathrm{Mat}(n\times n, \mathbb{R})$. (Das Differential ist definitionsgemäß eine lineare Abbildung; hier wäre es natürlich nicht klug, auch diese Abbildung als Matrix schreiben zu wollen.) Wer Lust dazu hat, kann auch das Differential von det an einer beliebigen Stelle $a \in \mathrm{Mat}(n\times n, \mathbb{R})$ berechnen. Dazu empfiehlt es sich, die Abbildung $\Lambda_a\colon x \mapsto ax$ zu Hilfe zu nehmen und eventuell zuerst nur invertierbare a zuzulassen.

34.3 Durch die Formel

$$\exp x = e^x := \sum_{k=0}^{\infty} \frac{1}{k!} x^k$$

kann man eine Exponentialabbildung

$$\exp\colon \mathrm{Mat}(n\times n, \mathbb{R}) \longrightarrow \mathrm{Mat}(n\times n, \mathbb{R})$$

erklären, die viele Eigenschaften der bekannten skalaren Exponentialfunktion teilt. Überlegen Sie sich zuerst, daß die Formel überhaupt Sinn gibt: Auf $\mathrm{Mat}(n\times n, \mathbb{R}) = \mathbb{R}^{n^2}$ hat man die übliche durch

$$|x| = \sqrt{\sum_{i,j=1}^{n} x_{ij}^2} = \sqrt{\mathrm{tr}\, x^{\mathrm{t}} x}$$

erklärte Norm; wenn man $|xy| \le |x|\,|y|$ für alle $x, y \in \mathrm{Mat}(n\times n, \mathbb{R})$ beweist, erkennt man die absolute Konvergenz der Reihe wie im Eindimensionalen.

Rechnen Sie zur Illustration für jedes $t \in \mathbb{R}$ die Reihe $\exp t \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ explizit aus.

Wenn Sie jetzt anfangen, die Grundeigenschaften der Exponentialabbildung nachzuweisen, werden Sie beim Beweis von $e^{x+y} = e^x e^y$ das Bedürfnis haben, x und y in Produkten miteinander zu vertauschen, was für $n > 1$ im allgemeinen ja nicht möglich ist. Tatsächlich gilt $e^{x+y} = e^x e^y$ auch nicht für beliebige x und y, aber jedenfalls dann, wenn $xy = yx$ ist, was immer noch viele interessante Fälle einschließt. Welche zum Beispiel, und welche weiteren (immer zum eindimensionalen Fall analogen) Eigenschaften der Exponentialabbildung kann man daraus folgern?

34.4 Zeigen Sie, daß die Exponentialreihe

$$x \longmapsto \sum_{k=0}^{\infty} \frac{1}{k!} x^k$$

auf jeder beschränkten Teilmenge von $\mathrm{Mat}(n\times n, \mathbb{R})$ gleichmäßig-absolut konvergiert. Daraus läßt sich nicht nur auf die Stetigkeit der Exponentialabbildung schließen, sondern auch direkt sehen, daß diese Funktion an der Stelle $0 \in \mathrm{Mat}(n\times n, \mathbb{R})$ differenzierbar und $D\exp(0) = \mathrm{id}$ ist: wie nämlich?

34.5 Untersuchen Sie, ob sich der Mittelwertsatz 14.1 statt für Funktionen $f\colon [a,b] \longrightarrow \mathbb{R}$ auch für Wege $\gamma\colon [a,b] \longrightarrow \mathbb{R}^p$ mit $p > 1$ formulieren und beweisen läßt.

34.6 Für die Bahnkurve

$$\gamma(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}$$

eines Massenpunktes in der Ebene gelte

$$3x(t)\dot{x}(t) + 2y(t)\dot{y}(t) = 0 \quad \text{für alle } t \in \mathbb{R}.$$

Zeigen Sie, daß die Bahn des Punktes in einer Ellipse enthalten ist.

34.7 Sei $\gamma\colon T \longrightarrow \mathbb{R}^p$ eine C^1-Kurve, und $t_0 \in T$. Dann heißt die Funktion

$$T \ni t \longmapsto s(t) := \int_{t_0}^{t} |\dot{\gamma}| = \int_{t_0}^{t} |\dot{\gamma}(\tau)|\, d\tau$$

die (vom Zeitpunkt t_0 aus gemessene) Bogenlänge von γ. (Im Zusammenhang mit Kurven wird der Buchstabe s manchmal für diese Bogenlänge reserviert.) Dazu erst mal ein paar Fragen: Wie soll man sich eine C^1-Kurve vorstellen? Kann zum Beispiel

die Bildmenge einer C^1-Kurve sein? Wäre auch

$$\int_{t_0}^{t} \dot{\gamma}(\tau) d\tau$$

eine sinnvolle Größe (wenn auch nicht gleich der Bogenlänge)? Kann die Bogenlänge auch negative Werte annehmen? Was passiert, wenn man ein anderes t_0 wählt?

34.8 Berechnen Sie die Bogenlänge für folgende Kurven:

(a) die parametrisierte Gerade $\gamma\colon \mathbb{R} \longrightarrow \mathbb{R}^p$; $\quad \gamma(t) = at + b$

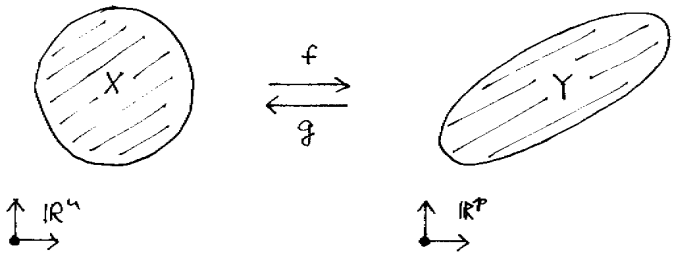
(b) die Parametrisierung der Kreislinie $\gamma\colon \mathbb{R} \longrightarrow \mathbb{R}^2$; $\quad \gamma(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}$

(c) die Zykloide $\gamma\colon \mathbb{R} \longrightarrow \mathbb{R}^2$; $\quad \gamma(t) = \begin{pmatrix} t - \sin t \\ 1 - \cos t \end{pmatrix}$ — Erklären Sie, warum ein Punkt auf dem Umfang eines mit Einheitsgeschwindigkeit rollenden Rades eine solche Zykloide beschreibt. Wie groß ist der Weg, den der Punkt bei einer vollen Umdrehung des Rades zurücklegt?

35 Diffeomorphismen

Diffeomorphismen stehen zu differenzierbaren Abbildungen im gleichen Verhältnis wie Isomorphismen zu Homomorphismen:

35.1 Definition Seien $X \subset \mathbb{R}^n$ und $Y \subset \mathbb{R}^p$ offene Teilmengen. Ein Diffeomorphismus von X nach Y ist eine differenzierbare Abbildung $f\colon X \longrightarrow Y$, zu der eine differenzierbare Abbildung $g\colon Y \longrightarrow X$ mit $g \circ f = \mathrm{id}_X$ und $f \circ g = \mathrm{id}_Y$ existiert. Man sagt auch, f bilde X diffeomorph auf Y ab. Zwei gegebene offene Mengen $X \subset \mathbb{R}^n$ und $Y \subset \mathbb{R}^p$ nennt man diffeomorph, wenn es einen Diffeomorphismus $f\colon X \longrightarrow Y$ gibt.



Bemerkungen Natürlich kann man ebensogut verlangen, daß f bijektiv und neben f auch $g = f^{-1}$ differenzierbar ist. Analog werden C^1- (und später C^k-)Diffeomorphismen erklärt.

In der typischen eindimensionalen Situation sind solche Diffeomorphismen ganz leicht mittels Satz 14.5 zu erkennen: Ist $I \subset \mathbb{R}$ ein offenes Intervall und $f\colon I \longrightarrow \mathbb{R}$ differenzierbar, so ist f genau dann ein Diffeomorphismus von I auf das dann offene Bildintervall $f(I) \subset \mathbb{R}$, wenn $f'(x) \neq 0$ für alle $x \in I$ ist. Die bemerkenswerte Tatsache, daß man hier aus einer lokalen Voraussetzung ($f'(x) \neq 0$ für alle $x \in I$) unter anderem eine globale Schlußfolgerung (f injektiv) ziehen kann, beruht letztlich darauf, daß die Injektivität einer auf einem Intervall definierten stetigen Funktion gleichbedeutend mit ihrer strengen Monotonie ist (Satz 8.5 von der Umkehrfunktion).

Im Mehrdimensionalen gibt es den Begriff der Monotonie nicht, deshalb gibt es auch keinen Satz von der Umkehrabbildung, der 8.5 entspräche, und deshalb liefert auch das gleich zu besprechende Analogon von Satz 14.5 in mehreren Veränderlichen eine bei weitem nicht so umfassende Aussage. Gültig und einfach zu beweisen bleibt immerhin die eine Richtung:

35.2 Lemma Ist $\mathbb{R}^n \supset X \xrightarrow{f} Y \subset \mathbb{R}^p$ ein Diffeomorphismus, so ist das Differential $Df(a)\colon \mathbb{R}^n \longrightarrow \mathbb{R}^p$ für jedes $a \in X$ ein Isomorphismus von Vektorräumen. Insbesondere ist $n = p$ (falls $X \neq \emptyset$), und es gilt

$$Df^{-1}\big(f(a)\big) = \big(Df(a)\big)^{-1}$$

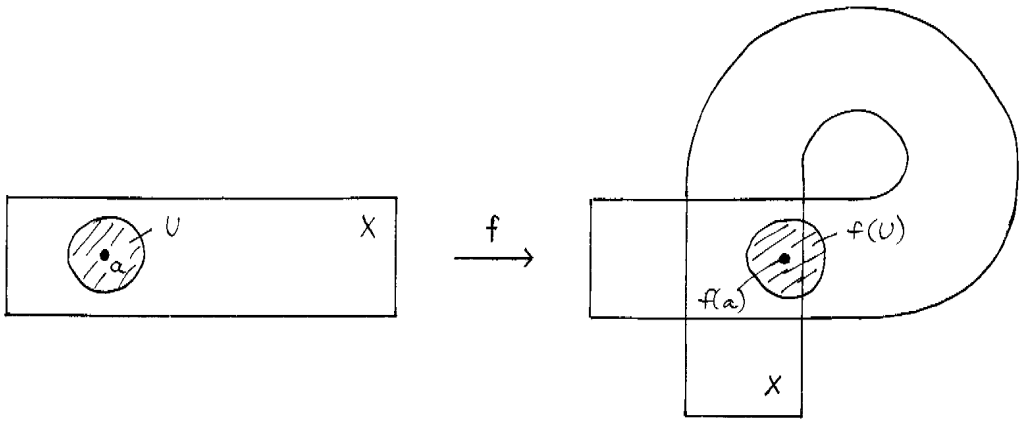
für jedes $a \in X$.

Beweis Die Kettenregel 34.4 liefert mit $g = f^{-1}$ und $b = f(a)$

$$\begin{aligned} Dg(b)Df(a) &= D(g \circ f)(a) = D\,\mathrm{id}_X(a) = 1 \in \mathrm{Mat}(n \times n, \mathbb{R}) \\ Df(a)Dg(b) &= D(f \circ g)(b) = D\,\mathrm{id}_Y(b) \;=\; 1 \in \mathrm{Mat}(p \times p, \mathbb{R}). \end{aligned}$$

Die interessantere Richtung von 14.5 ist aber natürlich die andere, in der man aus den Eigenschaften der Ableitung von f Rückschlüsse auf f selbst ziehen kann. Im Mehrdimensionalen ist das nur noch lokal, nicht global möglich, deshalb ist folgende Definition praktisch:

35.3 Definition Sei $X \subset \mathbb{R}^n$ eine offene Teilmenge, $f\colon X \longrightarrow Y$ eine Abbildung und $a \in X$ ein Punkt. Man sagt, f sei bei a ein lokaler Diffeomorphismus, wenn es eine offene Menge $U \subset X$ mit $a \in U$ gibt, die von f diffeomorph auf eine offene Menge $f(U) \subset \mathbb{R}^n$ abgebildet wird.



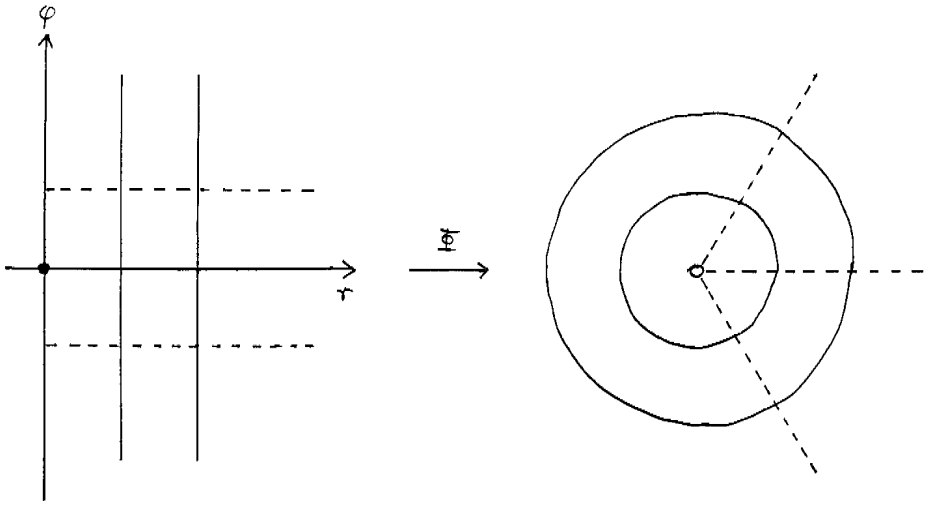
Ist das für jedes $a \in X$ der Fall, läßt man den Zusatz "bei a" weg. Zur Unterscheidung nennt man "richtige" Diffeomorphismen auch *globale* Diffeomorphismen.

35.4 Satz von der **lokalen Umkehrung** Sei $X \subset \mathbb{R}^n$ offen und $a \in X$. Eine C^1-Abbildung $f\colon X \longrightarrow \mathbb{R}^n$ ist genau dann ein lokaler (C^1-)Diffeomorphismus bei a, wenn ihr Differential dort umkehrbar ist:

$$Df(a) \in GL(n, \mathbb{R})$$

Trotz seines nur lokalen Charakters ist das der wichtigste Satz der Differentialrechnung mehrerer Veränderlicher, und wir werden noch zahlreiche Anwendungen kennenlernen. Der nicht einfache Beweis enthält eine Reihe schöner Ideen, würde uns aber zu lange aufhalten.

35.5 Beispiel Wir begnügen uns diesmal mit der ebenen Version der Polarkoordinaten,



betrachten also die stetig differenzierbare Abbildung

$$\mathbb{R}^2 \ni \begin{pmatrix} r \\ \varphi \end{pmatrix} \overset{\Phi}{\longmapsto} \begin{pmatrix} r\cos\varphi \\ r\sin\varphi \end{pmatrix} \in \mathbb{R}^2.$$

Die Determinante ihres Differentials, die man übrigens allgemein Jacobi-Determinante nennt, ist

$$\det D\Phi(r,\varphi) = \det \begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix} = r\big((\cos\varphi)^2 + (\sin\varphi)^2\big) = r.$$

Φ ist also an jeder Stelle (r,φ) mit $r \neq 0$ ein lokaler Diffeomorphismus. Daß das nicht auch an den Stellen mit $r = 0$ gilt, geht ohne Berufung auf den Satz schon daraus hervor, daß $\Phi(0,\varphi) = (0,0)$ für alle $\varphi \in \mathbb{R}$ ist, es um einen Punkt $(0,\varphi)$ also keine offene Menge geben kann, die von Φ injektiv abgebildet wird.

Andererseits schränkt sich Φ wegen der Periodizität $\Phi(r,\varphi+2\pi) = \Phi(r,\varphi)$ nicht zu einem globalen Diffeomorphismus etwa von $(0,\infty)\times\mathbb{R}$ auf die gelochte Ebene $\mathbb{R}^2\backslash\{(0,0)\}$ ein: die Aussage des Satzes ist wirklich nur eine lokale. Man kann Φ auch durch Einschränken auf eine geschickt gewählte offene Teilmenge $X \subset \mathbb{R}^2$ nicht zu einem Diffeomorphismus $X \longrightarrow \mathbb{R}^2\backslash\{(0,0)\}$ machen (Aufgabe 35.1).

Übrigens ist das Beispiel nur eine Illustration, keine echte Anwendung von Satz 35.4, weil wir die lokalen Umkehrungen von Φ auf die bekannte Weise mittels Betrags- und Arcusfunktionen explizit hinschreiben können. Die Stärke des Satzes zeigt sich erst in dem (Normal-)fall, daß die explizite Berechnung lokaler Umkehrungen nicht praktikabel oder unmöglich ist.

Im Gegensatz zur Analysis in einer Veränderlichen kennt man im mehrdimensionalen Fall keine systematischen Methoden, um Abbildungen als (globale) Diffeomorphismen zu erkennen. Die Schlüsselfrage ist dabei die der Injektivität; unter anderem das zeigt der rein formale

35.6 Satz Sei $X \subset \mathbb{R}^n$ offen und $f\colon X \longrightarrow \mathbb{R}^n$ ein lokaler Diffeomorphismus. Dann bildet f jede offene Teilmenge von X auf eine offene Teilmenge von $\mathbb{R}^n$ ab; insbesondere ist die Bildmenge $f(X) \subset \mathbb{R}^n$ offen. Ist f injektiv, so ist $X \xrightarrow{f} f(X)$ ein Diffeomorphismus.

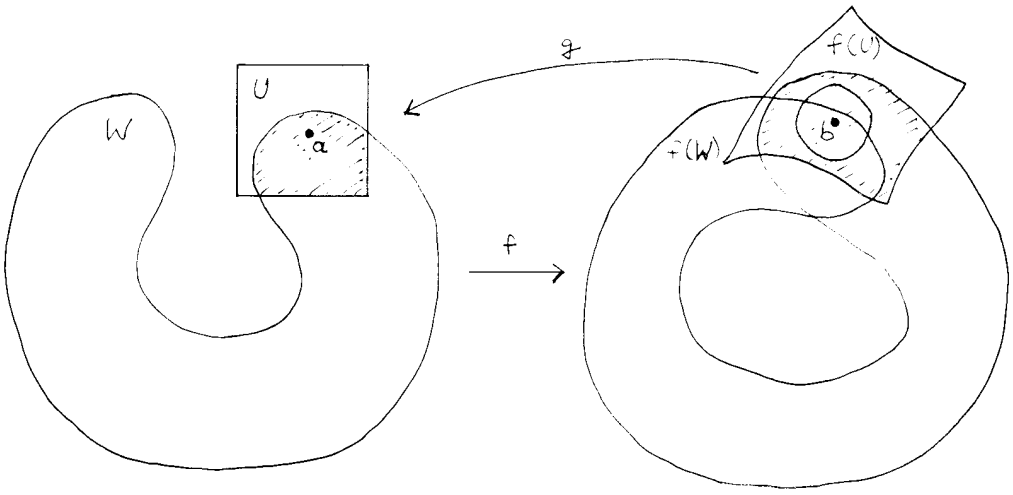
Beweis Sei $W \subset X$ offen und $b \in f(W)$. Wir wählen ein $a \in W$ mit $f(a) = b$. Da f bei a ein lokaler Diffeomorphismus ist, gibt es eine offene Menge $U \subset X$ mit $a \in U$, so daß $f(U)$ offen und die Einschränkung $U \xrightarrow{f} f(U)$ ein Diffeomorphismus ist, etwa mit Umkehrung $g\colon f(U) \longrightarrow U$. Der Durchschnitt $U \cap W$ ist natürlich auch in U offen, und weil $g\colon f(U) \longrightarrow U$ stetig ist, ist nach Satz 30.6 die Bildmenge

$$f(U \cap W) = g^{-1}(U \cap W)$$

relativ offen in $f(U)$. Aber weil $f(U) \subset \mathbb{R}^n$ selbst offen ist, bedeutet das einfach, daß $f(U \cap W)$ als Teilmenge von $\mathbb{R}^n$ offen ist. Diese Menge enthält den Punkt b, also gibt es ein $\varepsilon > 0$ mit

$$U_\varepsilon(b) \subset f(U \cap W) \subset f(W).$$

Das beweist die Offenheit von $f(W)$.



Sei nun f auch injektiv. Wir wissen schon, daß $f(X)$ offen ist, und zu zeigen bleibt, daß $f^{-1}\colon f(X) \longrightarrow X$ differenzierbar ist. Mit den Bezeichnungen a, b, U, g aus dem ersten Teil des Beweises stimmt aber $f^{-1}|f(U)$ zwangsläufig mit der differenzierbaren Abbildung g überein, und daraus folgt's schon, weil Differenzierbarkeit eine lokale Eigenschaft ist.

Klar ist die

35.7 Folgerung Aus einem lokalen oder globalen Diffeomorphismus entsteht durch Einschränken auf eine offene Teilmenge (und, im globalen Fall, entsprechendes Verkleinern der Zielmenge natürlich) wieder ein Diffeomorphismus der gleichen Art.

Der Satz von der lokalen Umkehrung hat außer der direkten eine Fülle von weiteren interessanten Anwendungen. In der, die wir jetzt betrachten, geht es um "implizite Funktionen", nämlich Funktionen f, die nicht direkt durch Angabe ihrer Werte, sondern durch eine Gleichung

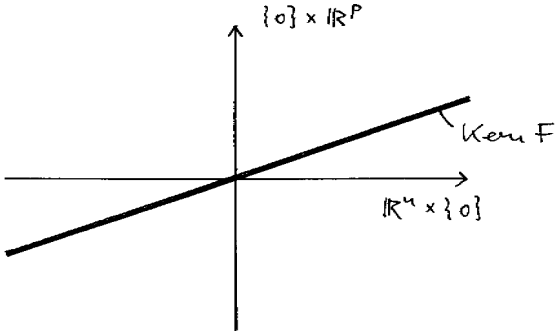
$$F\big(x, f(x)\big) = 0 \quad \text{für alle } x$$

beschrieben werden sollen. Wenn f Werte in $\mathbb{R}^p$, also p Komponenten hat, wird man erwarten, daß dazu ebensoviele Gleichungen nötig sind; man wird also vernünftigerweise von einer Abbildung $F\colon X \times Y \longrightarrow \mathbb{R}^p$ mit offenen Teilmengen $X \subset \mathbb{R}^n$ und $Y \subset \mathbb{R}^p$ ausgehen und als Lösung des Problems auf eine Funktion $f\colon X \longrightarrow Y$ hoffen.

Um besser zu sehen, worauf es dabei ankommt, betrachten wir vorweg den viel einfacheren Fall, daß alles linear ist, also $X = \mathbb{R}^n$, $Y = \mathbb{R}^p$, und das gegebene F ebenso wie das gesuchte f linear. Wenn wir in Matrizensprache dann

$$F = \left(\begin{array}{c|c} F_1 & F_2 \end{array} \right) \in \mathrm{Mat}\,\big(p \times (n+p), \mathbb{R}\big),$$

mit $F_1 \in \mathrm{Mat}(p \times n, \mathbb{R})$ und $F_2 \in \mathrm{Mat}(p \times p, \mathbb{R})$ schreiben, lautet die Frage, ob der lineare Unterraum $\mathrm{Kern}\, F \subset \mathbb{R}^n \times \mathbb{R}^p$ der Graph einer linearen Abbildung $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}^p$ ist.



Nun, bekanntlich ist das genau dann der Fall, wenn $\mathrm{Kern}\, F$ ein Komplement von $\{0\} \times \mathbb{R}^p$ in $\mathbb{R}^n \times \mathbb{R}^p$ ist (vergleiche Aufgabe 18.4). Für die Matrix F bedeutet das, daß die $p \times p$-Teilmatrix F_2 den Rang p hat, also invertierbar ist.

35.8 Satz über "**implizite Funktionen**" $X \subset \mathbb{R}^n$ und $Y \subset \mathbb{R}^p$ seien offen, und

$$X \times Y \ni (x, y) \longmapsto F(x, y) \in \mathbb{R}^p$$

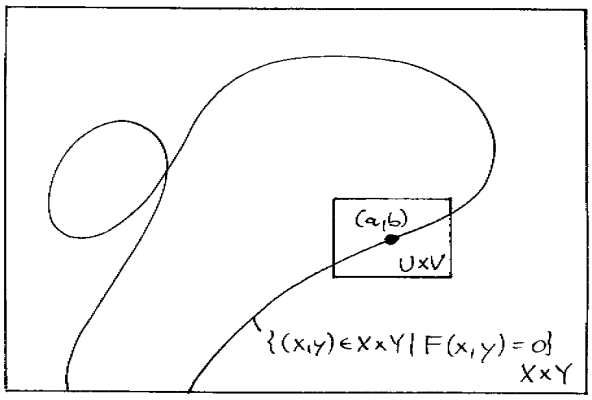
sei eine C^1-Abbildung. Ferner sei $(a, b) \in X \times Y$ eine Nullstelle von F, und die "partielle" Jacobi-Matrix

$$\frac{dF}{dy}(a, b) = \left(\frac{\partial F_i}{\partial y_j}(a, b) \right)_{i,j=1}^{p}$$

sei invertierbar. Dann gibt es offene Teilmengen $U \subset X$ und $V \subset Y$ mit $(a, b) \in U \times V$, so daß

$$\big\{(x, y) \in U \times V \,\big|\, F(x, y) = 0\big\}$$

der Graph einer C^1-Abbildung $f\colon X \longrightarrow Y$ ist.



Das Differential von f an der Stelle a ist

$$Df(a) = -\left(\frac{dF}{dy}(a,b)\right)^{-1} \frac{dF}{dx}(a,b).$$

Anmerkungen Für den, der in der Schlußfolgerung die eingangs erwähnte Formel sehen möchte: Es gibt also eine Funktion $f\colon U \longrightarrow V$, die die Gleichung $F\big(x, f(x)\big) = 0$ nach f auflöst; diese Funktion ist (wenn man passende U und V einmal fixiert hat) eindeutig bestimmt, sie erfüllt $f(a) = b$ und ist stetig differenzierbar. — Der Satz ist ein Musterbeispiel für den grundlegenden Ansatz der Differentialrechnung, nichtlineare Objekte zu untersuchen, indem man durch Differenzieren zu den entsprechenden linearen übergeht: Ist das das vorgelegte Problem, nämlich die Gleichung $F\big(x, f(x)\big) = 0$ nach f aufzulösen, an einer Stelle (a, b) lösbar und ist das entsprechende lineare Problem für das Differential dort lösbar, so ist das Problem auch lokal in der Nähe von (a, b) lösbar.

Beweis des Satzes Ein kleiner Trick verwandelt die Situation in die des Satzes von der lokalen Umkehrung: wir betrachten die Hilfsabbildung

$$H\colon X \times Y \longrightarrow \mathbb{R}^n \times \mathbb{R}^p; \quad (x, y) \mapsto \big(x, F(x, y)\big).$$

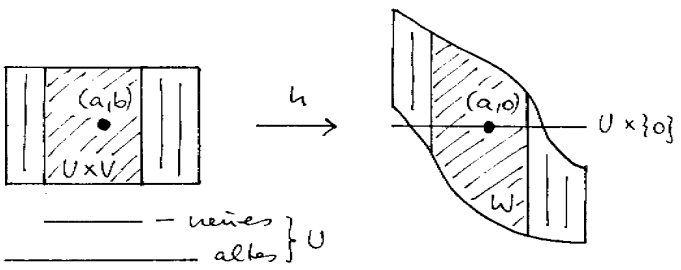
Diese sendet (a, b) auf $(a, 0)$, und ihr Differential bei (a, b) ist

$$DH(a,b) = \left(\begin{array}{c|c} 1 & 0 \\ \hline \frac{dF}{dx}(a,b) & \frac{dF}{dy}(a,b) \end{array}\right).$$

Weil $\dfrac{dF}{dy}(a, b)$ nach Voraussetzung invertierbar ist, ist auch $DH(a, b)$ invertierbar. Nach Satz 35.4 ist H also ein lokaler Diffeomorphismus bei (a, b): Wir finden (unter Verwendung der Folgerung 35.7) offene Mengen $U \subset X$ und $V \subset Y$ mit $(a, b) \in U \times V$ derart, daß die Einschränkung $h := H|(U \times V)$ ein C^1-Diffeomorphismus

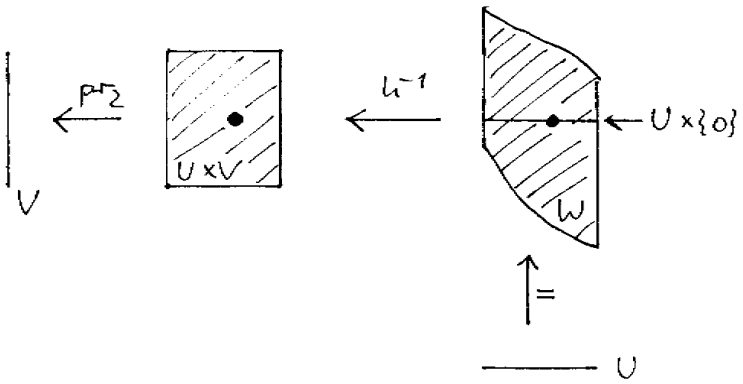
$$U \times V \xrightarrow{\ h\ } W$$

auf die offene Menge $W := H(U \times V) \subset \mathbb{R}^n \times \mathbb{R}^p$ ist. Weil $(a, 0) \in W$ und W offen ist, können wir U so verkleinern, daß $(x, 0) \in W$ für alle $x \in U$ gilt.



Wir definieren jetzt die C^1-Funktion $f\colon U \longrightarrow V$ als die Komposition

$$f\colon U = U \times \{0\} \xrightarrow{\ h^{-1}\ } U \times V \xrightarrow{\ \mathrm{pr}_2\ } V.$$



Ist dann $(x,y) \in U \times V$ beliebig, so bedeutet $F(x,y) = 0$ dasselbe wie $h(x,y) = (x,0)$ oder wie $h^{-1}(x,0) = (x,y)$ oder wie $f(x) = y$. Also ist $\Gamma_f = \big\{(x,y) \in U \times V \,\big|\, F(x,y) = 0\big\}$ wie behauptet.

Bleibt noch die Formel für $Df(a)$ zu beweisen. Dazu differenziert man die Identität $0 = F\big(x, f(x)\big)$ nach der Kettenregel:

$$0 = \begin{pmatrix} \frac{dF}{dx}(a,b) & \frac{dF}{dy}(a,b) \end{pmatrix} \begin{pmatrix} 1 \\ Df(a) \end{pmatrix} = \frac{dF}{dx}(a,b) + \frac{dF}{dy}(a,b) \cdot Df(a)$$

Diese Matrizengleichung braucht man jetzt bloß noch von links mit $\left(\frac{dF}{dy}(a,b)\right)^{-1}$ zu multiplizieren und damit nach $Df(a)$ aufzulösen.

35.9 Beispiel Jede einfache Nullstelle eines reellen Polynoms hängt differenzierbar von dessen Koeffizienten ab. Um das zu sehen, wenden wir den Satz über implizite Funktionen mit

$$F\colon \mathbb{R}^n \times \mathbb{R} \longrightarrow \mathbb{R}; \quad F(x,y) = F(x_0, \dots, x_{n-1}, y) = y^n + \sum_{j=0}^{n-1} x_j y^j$$

an; man nennt F übrigens das *allgemeine (normierte) Polynom* vom Grad n. Ist $b \in \mathbb{R}$ einfache Nullstelle des Polynoms $p(Y) = F(a,Y) = Y^n + \sum_{j=0}^{n-1} a_j Y^j$, so ist

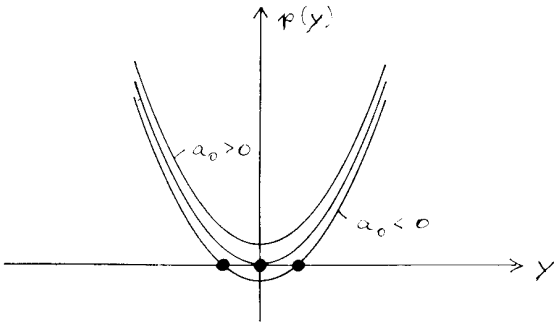
$$p(Y) = (Y - b)\, q(Y) \quad \text{mit } q(b) \neq 0,$$

also $F(a,b) = 0$ und

$$\frac{\partial F}{\partial y}(a,b) = \frac{dp}{dy}(b) = 1 \cdot q(b) + 0 \cdot q'(b) = q(b) \neq 0.$$

Nach Satz 35.8 gibt es also eine offene Kugel $U_\delta(a)$ und ein offenes Intervall J um b derart, daß für jedes $x \in U_\delta(a)$ das Polynom $Y^n + \sum_{j=0}^{n-1} x_j Y^j$ in J eine einzige Nullstelle hat und diese stetig differenzierbar von x abhängt.

Beachten Sie, daß dieses Ergebnis nicht einfach die Differenzierbarkeit einer a priori bekannten Funktion verspricht, sondern auch eine Existenz- und Eindeutigkeitsaussage umfaßt: Wenn man an den Koeffizienten des Polynoms p "wackelt", kann die einfache Nullstelle b weder verschwinden noch in mehrere Nullstellen zerfallen. Für mehrfache Nullstellen ist das falsch: Variiert man das Polynom $p(Y) = Y^2$ zu $Y^2 + x_0$ mit beliebig kleinem $x_0 > 0$, so verschwindet die doppelte Nullstelle 0 von p im Nichts (tatsächlich natürlich in $\mathbb{C}\backslash\mathbb{R}$), für $x_0 < 0$ dagegen zerfällt sie in zwei einfache.



Und selbst wenn eine mehrfache Nullstelle bei einer Variation der Koeffizienten erhalten und eindeutig bleibt, braucht sie nicht differenzierbar von den Koeffizienten abzuhängen, wie das Beispiel $Y^3 + x_0$ mit der eindeutigen Nullstelle $y = -\sqrt[3]{x_0}$ illustriert.

Übungsaufgaben

35.1 Beweisen Sie, daß es keine offene Teilmenge $X \subset \mathbb{R}^2$ gibt, die von der Polarkoordinatenabbildung Φ diffeomorph auf $\mathbb{R}^2 \setminus \{(0,0)\}$ abgebildet wird. (Plausibel ist das ja, aber es geht darum, einen zwingenden Beweis zu finden und zu formulieren.)

35.2 Sei $k \in \mathbb{N}$ positiv. Beweisen Sie, daß jede genügend nahe an der Einheitsmatrix gelegene Matrix $y \in \mathrm{Mat}(n \times n, \mathbb{R})$ eine eindeutig bestimmte k-te Wurzel $x \in \mathrm{Mat}(n \times n, \mathbb{R})$ besitzt, die ihrerseits nahe der Einheitsmatrix liegt (Präzisierung!).

35.3 Sei $X \subset \mathbb{R}^n$ offen, und sei $f\colon X \longrightarrow \mathbb{R}^n$ eine C^1-Abbildung mit einem Fixpunkt $a \in X$ (d.h. $f(a) = a$). Beweisen Sie: Wenn $\det\big(1 - Df(a)\big) \neq 0$ ist, dann gibt es ein $\delta > 0$, so daß f in $U_\delta(a)$ außer a keinen weiteren Fixpunkt hat.

35.4 Zeigen Sie, daß durch die Gleichungen

$$u^4 x + uv^2 y = 2$$
$$u^2 x + v^3 y^2 = 2$$

implizit C^1-Funktionen $(x, y) \mapsto u(x, y)$ und $(x, y) \mapsto v(x, y)$ gegeben sind, die in der Nähe von $(1, 1) \in \mathbb{R}^2$ definiert sind und Werte in der Nähe von $1 \in \mathbb{R}$ annehmen (präzisieren Sie!). Berechnen Sie $Du(1, 1)$ und $Dv(1, 1)$.

35.5 Sei $u \in GL(n, \mathbb{R})$. Berechnen Sie:

- das Differential der Abbildung $f\colon GL(n, \mathbb{R}) \longrightarrow GL(n, \mathbb{R})$ mit $f(x) = x^{-1}$ an der Stelle u,
- das Differential der Abbildung $g_u\colon GL(n, \mathbb{R}) \longrightarrow GL(n, \mathbb{R})$ mit $g_u(x) = uxu^{-1}x^{-1}$ an der Stelle 1,
- das Differential der Abbildung $GL(n, \mathbb{R}) \ni u \longmapsto Dg_u(1) \in \mathrm{End}\big(\mathrm{Mat}(n \times n, \mathbb{R})\big)$ an der Stelle 1.

Kann es übrigens vorkommen, daß $Dg_u(1)$ surjektiv ist?

36 Differenzierbare Karten und Untermannigfaltigkeiten

In der linearen Algebra haben wir immer wieder lineare Karten für einen Vektorraum V benutzt, sei es, um ein abstraktes Objekt konkreten Rechnungen zugänglich zu machen (zum Beispiel eine lineare Abbildung durch eine Matrix auszudrücken), sei es, um ein zunächst schwer zu durchschauendes Objekt durch geschickte Kartenwahl einfach zu machen (zum Beispiel eine quadratische Form durch Hauptachsentransformation zu diagonalisieren).

Die in der Analysis verwendeten Karten werden, für uns jedenfalls, eher unter dem zweiten Gesichtspunkt interessant werden. Von den linearen Karten unterscheiden sie sich vor allem dadurch, daß sie in der Regel nicht global sind, so daß man in einer gegebenen Situation oft nicht eine, sondern mehrere, vielleicht sogar viele Karten benötigt.

36.1 Definition Sei $X \subset \mathbb{R}^n$ offen. Eine (differenzierbare) Karte für X ist ein Diffeomorphismus

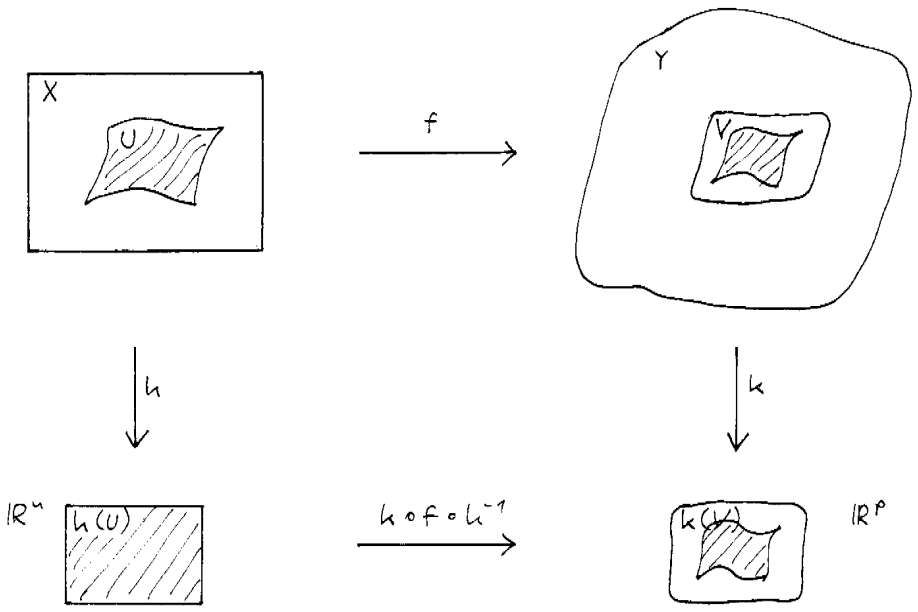
$$U \xrightarrow{h} h(U) \subset \mathbb{R}^n$$

zwischen einer offenen Teilmenge $U \subset X$ und einer offenen Menge $h(U) \subset \mathbb{R}^n$. Ist $a \in X$ ein Punkt und $h(a) = 0$, so spricht man von einer Karte um a. Als wesentliche zu einer Karte gehörige Daten notiert man meist das Paar (U, h).

36.2 Sprechweise Seien $X \subset \mathbb{R}^n$ und $Y \subset \mathbb{R}^p$ offen, $f\colon X \longrightarrow Y$ eine Abbildung. Ist (U, h) eine Karte für X, (V, k) eine Karte für Y und gilt $f(U) \subset V$, so kann man die Komposition

$$k \circ f \circ h^{-1}\colon\ h(U) \xrightarrow{h^{-1}} U \xrightarrow{f} V \xrightarrow{k} k(V)$$

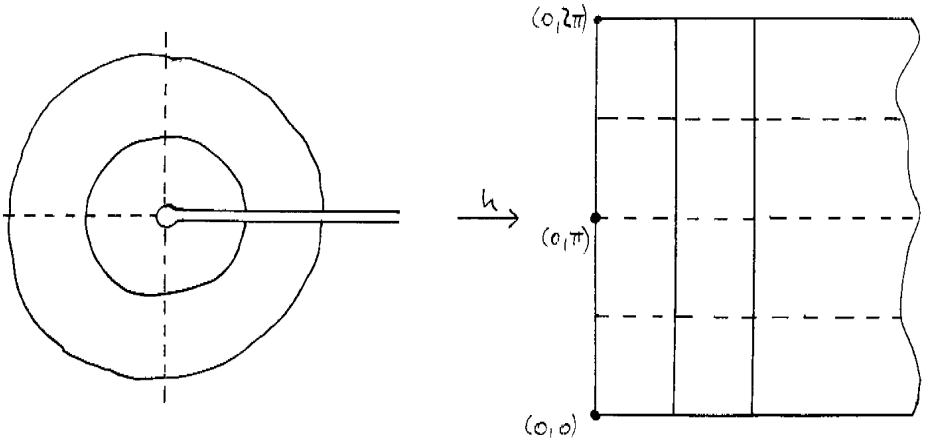
bilden; von dieser Abbildung spricht man als "f (geschrieben) in den Karten h und k".



Ist nur eine der beiden Karten (U, h) und (V, k) genannt, so ist als zweite stillschweigend die identische Karte gemeint.

36.3 Beispiel Sie werden von der Physik daran gewöhnt sein, skalare Funktionen bei Bedarf "in Polarkoordinaten" zu schreiben. Im ebenen Fall mag eine solche Funktion f auf ganz $\mathbb{R}^2$ erklärt sein. Wählt man eine offene Menge $U \subset \mathbb{R}^2$, auf der man die Polarkoordinatenabbildung Φ umkehren kann, so erhält man dann in Φ^{-1} eine Karte h für $\mathbb{R}^2$; zum Beispiel ist

$$\begin{array}{rcl} h\colon \mathbb{R}^2 \setminus \big([0,\infty) \times \{0\}\big) & \longrightarrow & (0,\infty) \times (0,2\pi) \\ (r\cos\varphi, r\sin\varphi) & \mapsto & (r,\varphi) \end{array}$$



eine mögliche Wahl. "f in Polarkoordinaten" ist in mathematischer Ausdrucksweise dann "f in der Karte h", also die Funktion

$$f \circ h^{-1}\colon (0,\infty) \times (0,2\pi) \longrightarrow \mathbb{R}.$$

Wenn Physiker sich erlauben, statt $f \circ h^{-1}$ einfach $f(r,\varphi)$ — im Gegensatz zu $f(x,y)$ — zu schreiben, geht das deswegen im großen und ganzen gut, weil in physikalischen Formeln den Buchstaben, hier eben x, y, r und φ eine feste physikalische oder geometrische Bedeutung zukommt (was freilich ist mit $f(1,2)$ gemeint?). Als korrekte mathematische Notation kann das aber nicht gelten; schließlich sind f und $f \circ h^{-1}$ ganz verschiedene Funktionen.

Nicht nur im Beispiel der Polarkoordinaten steht das Wort "Karte" für das, was in der Physik meist Koordinaten oder Satz/System von Koordinaten heißt. Um zu betonen, daß es sich nicht notwendig um die besonders einfachen linearen Karten handelt, haben die Physiker noch den lustigen, aber treffenden Ausdruck "*krummlinige* Koordinaten". Beachten Sie übrigens, daß Polarkoordinaten — gleich welcher Dimension — gerade dort, wo man am ehesten meinen könnte, nämlich im Nullpunkt, auf keine Weise zu einer differenzierbaren Karte gemacht werden können.

Das Interesse an Karten steht und fällt damit, inwieweit sich eine gegebene Situation durch Wahl einer geeigneten Karte vereinfachen läßt. Bei Polarkoordinaten ist die Vereinfachung häufig offensichtlich, wenn die zu behandelnden Objekte rotationssymmetrisch sind. Es gibt aber auch interessante Situationen, in denen man die guten Karten nicht von vornherein hat, sondern erst konstruieren muß. Die vielleicht wichtigste Variante davon erkläre ich jetzt.

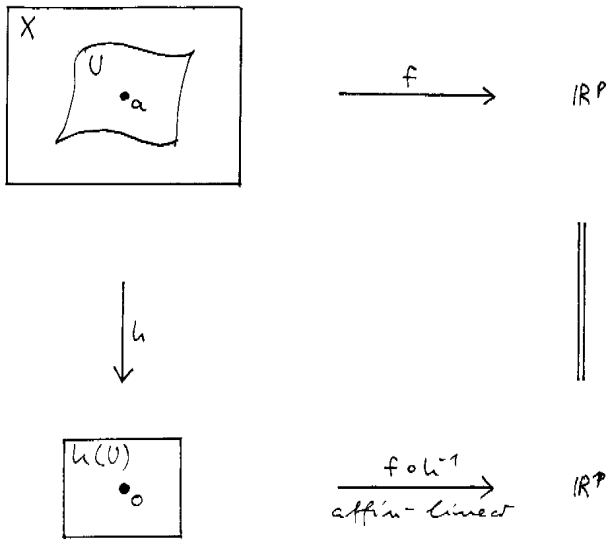
36.4 Definition Sei $X \subset \mathbb{R}^n$ offen und $f\colon X \longrightarrow \mathbb{R}^p$ differenzierbar. Unter dem Rang von f an einer Stelle $a \in X$ versteht man den Rang des Differentials $Df(a)$. Man nennt a einen regulären Punkt von f, wenn f bei a den Rang p hat; andernfalls heißt a ein kritischer Punkt von f.

Bemerkungen Ob a regulärer oder kritischer Punkt ist, hängt zwar nicht vom genauen Zielbereich von f ab, aber man darf zum Beispiel nicht $f\colon X \longrightarrow \mathbb{R}^p$ mit $f\colon X \longrightarrow \mathbb{R}^p \hookrightarrow \mathbb{R}^{p+1}$ identifizieren. — Nur für $n \geq p$ kann es reguläre Punkte geben; in einem solchem Punkt hat f den größtmöglichen Rang.

36.5 Satz vom **regulären Punkt** Sei $X \subset \mathbb{R}^n$ offen und $f\colon X \longrightarrow \mathbb{R}^p$ eine C^1-Abbildung. Um jeden regulären Punkt a von f gibt es eine C^1-Karte (U,h), in der f die Gestalt

$$f \circ h^{-1}\colon \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \longmapsto f(a) + \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$$

annimmt.



Lokal um einen regulären Punkt wird f also bezüglich einer geeignet gewählten Karte durch eine affin-lineare Abbildung beschrieben — ein durchaus eindrucksvolles Versprechen, wenn man sich vor Augen hält, wie kompliziert differenzierbare Abbildungen im Vergleich zu linearen sein können und im allgemeinen auch sind. Der

Beweis des Satzes ist dem des Satzes über implizite Funktionen ganz ähnlich. Nach Voraussetzung hat die Jacobi-Matrix $Df(a) \in \mathrm{Mat}(p\times n, \mathbb{R})$ den Rang p, sie enthält also eine invertierbare $p\times p$-Teilmatrix. Der Bequemlichkeit halber numerieren wir die Komponenten von $\mathbb{R}^n$ so um, daß eine solche Teilmatrix ganz links steht. Wenn wir dann $\mathbb{R}^n = \mathbb{R}^p \times \mathbb{R}^{n-p}$ und $a = (a', a'')$ zerlegen und f entsprechend in der Form

$$X \ni (x,y) \longmapsto f(x,y) \in \mathbb{R}^p$$

schreiben, ist die partielle Jacobi-Matrix $\dfrac{df}{dx}(a)$ invertierbar; die Abbildung

$$X \ni (x,y) \overset{H}{\longmapsto} \big(f(x,y)-f(a',a''),\ y-a''\big) \in \mathbb{R}^n$$

mit der Jacobi-Matrix

$$DH(a) = \left(\begin{array}{c|c} \dfrac{df}{dx}(a) & \dfrac{df}{dy}(a) \\ \hline 0 & 1 \end{array}\right)$$

ist deshalb nach Satz 35.3 ein lokaler Diffeomorphismus an der Stelle a. Wir finden also eine offene Menge $U \subset X$ um a, so daß die Einschränkung $h := H|U$ eine C^1-Karte

$$X \supset U \overset{h}{\longrightarrow} h(U) \subset \mathbb{R}^n$$

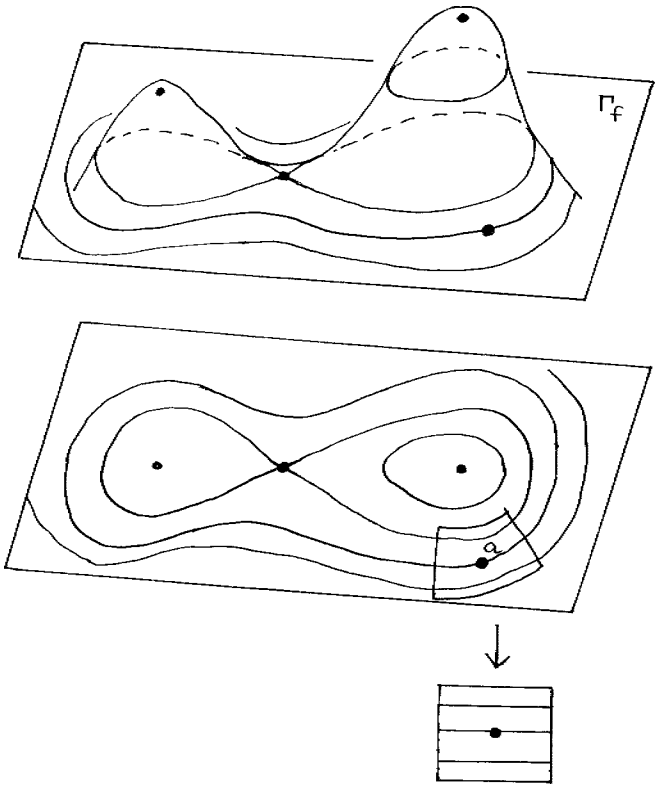
ist. Wegen $h(a) = (0,0)$ ist das eine Karte um den Punkt a, und wegen

$$h(U) \ni \big(f(x,y)-f(a), y-a''\big) \overset{h^{-1}}{\longmapsto} (x,y) \overset{f}{\longmapsto} f(x,y) \in \mathbb{R}^p$$

hat $f\circ h^{-1}$ in dieser Karte die gewüschte Form $f\circ h^{-1}\colon (x,y) \longmapsto f(a)+x$.

Anschaulich gesprochen laufen nach dem Satz vom regulären Punkt zum Beispiel die "Höhenlinien" einer C^1-Funktion $f\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ (mathematisch gesprochen die Fasern $f^{-1}\{b\} \subset \mathbb{R}^2$) in der Nähe eines regulären Punktes ordentlich nebeneinander. Bei den beiden Typen von kritischen Punkten, die die Skizze zeigt, ist

das dagegen nicht der Fall: In den beiden lokalen Maxima entarten die Höhenlinien zu einem Punkt, und die Höhenlinie durch den Sattelpunkt besteht aus zwei sich kreuzenden Zweigen.



In diesem speziellen Fall sagt der Satz vom regulären Punkt also unter anderem etwas über die Lösungsmenge einer differenzierbaren Gleichung, darum handelt es sich bei einer Höhenlinie ja. Was weiß man überhaupt über solche Lösungsmengen, also die Fasern einer (sagen wir C^1-)differenzierbaren Abbildung $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}^p$? Nun, im allgemeinen nichts, außer daß es sich nach Satz 30.6(c) um abgeschlossene Mengen handeln muß. Ganz abgesehen davon, daß selbst abgeschlossene Mengen noch recht pathologisch sein können, hat man vor allem über die Größe der Lösungsmengen ohne weitere Voraussetzungen keinerlei Kontrolle.

36.6 Beispiele (1) Daß eine große Zahl von Gleichungen nicht unbedingt zu einer kleinen Lösungsmenge führt, ist uns aus der linearen Algebra schon vertraut: Ist a eine $p\times n$-Matrix, so hat das Gleichungssystem $ax = 0$ mit p Gleichungen und n Variablen einen Lösungsraum der Dimension $n - \operatorname{rk} a$; der Lösungsraum fällt also immer dann größer als "erwartet" (nämlich $(n-p)$-dimensional) aus, wenn der Rang von a kleiner als p ist. Freilich brauchte uns das in der linearen Algebra nicht sonderlich zu stören, weil wir dort effiziente Techniken haben, um den Rang und überhaupt den Lösungsraum zu berechnen. Diese Methoden kann man im nichtlinearen Fall natürlich nicht mehr anwenden.

(2) Bei nichtlinearen Gleichungen ist auch das umgekehrte Phänomen möglich, daß die Lösungsmenge einer kleinen Anzahl von Gleichungen unerwartet klein ausfällt: Ganz gleich, wie groß $p \in \mathbb{N}$ ist, läßt die Lösungsmenge des von p Funktionen $f_i\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ gebildeten Gleichungssystems

$$X := \big\{x \in \mathbb{R}^n \,\big|\, f_1(x) = \cdots = f_p(x) = 0\big\}$$

ja auch die Beschreibung

$$X = \big\{x \in \mathbb{R}^n \,\big|\, f_1^2(x) + \cdots + f_p^2(x) = 0\big\}$$

durch eine einzige Gleichung zu.

Für differenzierbare Gleichungen darf man eben keine Theorie erwarten, die auch nur im entferntesten so vollständig und befriedigend wäre wie die der linearen Gleichungen. Aber in wichtigen Spezialfällen weiß man doch eine Menge über die Struktur der Lösungsmengen.

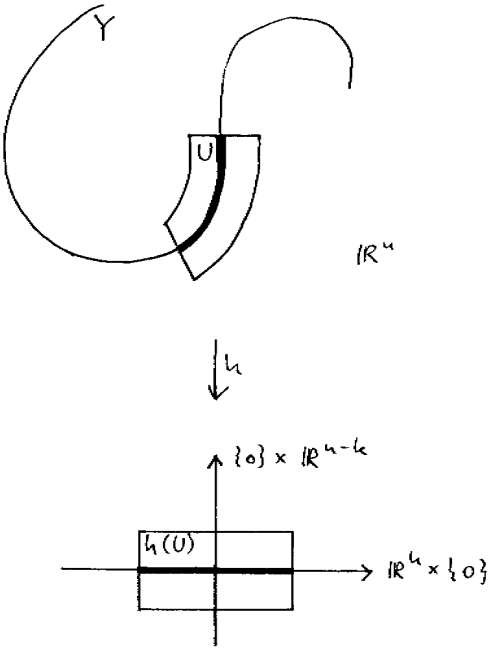
36.7 Definition Eine Teilmenge $Y \subset \mathbb{R}^n$ heißt eine k-dimensionale Untermannigfaltigkeit von $\mathbb{R}^n$, wenn es zu jedem Punkt $a \in Y$ eine Karte

$$U \xrightarrow{h} h(U) \subset \mathbb{R}^n = \mathbb{R}^k \times R^{n-k}$$

mit $a \in U$ gibt, derart daß

$$h(U \cap Y) = h(U) \cap \big(\mathbb{R}^k \times \{0\}\big)$$

ist. Man nennt h dann eine Untermannigfaltigkeitskarte (für Y in $\mathbb{R}^n$).



Bemerkungen Natürlich kann man durch Verschieben jederzeit $h(a) = 0$ erreichen und damit h zu einer Karte *um* a machen. Aufgabe von h ist es jedenfalls, die Untermannigfaltigkeit X lokal zu $\mathbb{R}^k \times \{0\}$ "flachzudrücken". Die Tatsache, daß die Menge $h(U) \cap \big(\mathbb{R}^k \times \{0\}\big)$ relativ offen in $\mathbb{R}^k \times \{0\}$ ist, läßt sich auch so ausdrücken, daß $\big\{x \in \mathbb{R}^k \,|\, (x,0) \in h(U)\big\}$ eine offene Teilmenge von $\mathbb{R}^k$ ist. Demnach sehen k-dimensionale Untermannigfaltigkeiten lokal wie offene (im allgemeinen aber gebogene) Stücke von $\mathbb{R}^k$ aus. Speziell sind die n-dimensionalen Untermannigfaltigkeiten von $\mathbb{R}^n$ einfach die offenen Teilmengen von $\mathbb{R}^n$. — Jede relativ offene Teilmenge $Z \subset Y$ ist wieder eine Untermannigfaltigkeit derselben Dimension: ist $Z = V \cap Y$ mit offenem $V \subset \mathbb{R}^n$, so braucht man bloß (U,h) durch $\big(U \cap V, h|(U \cap V)\big)$ zu ersetzen. — Wenn es um jeden Punkt von Y eine C^1-Karte für Y gibt, spricht man von einer C^1-Untermannigfaltigkeit. — Es gibt auch einen Begriff "*k-dimensionale Mannigfaltigkeit*", zu dem die Untermannigfaltigkeiten im gleichen Verhältnis stehen wie Untervektorräume zu Vektorräumen, und den man konsequenterweise zuerst behandeln sollte. Tue ich aber nicht, wegen der damit verbundenen höheren Abstraktionsstufe. Wenn ich im folgenden schon mal von Mannigfaltigkeiten rede, sind Untermannigfaltigkeiten gemeint. Zweidimensionale Mannigfaltigkeiten nennt man übrigens auch Flächen, eindimensionale Kurven.

36.7½ Beispiel Die Sphäre $S^2 \subset \mathbb{R}^3$ ist eine zweidimensionale Untermannigfaltigkeit. Dazu betrachten wir die offene Menge $U = U^1 \times (0,\infty) \subset \mathbb{R}^2 \times \mathbb{R} = \mathbb{R}^3$ und die durch

$$U \ni (x,y,z) \longmapsto h(x,y,z) := \big(x,\, y,\, z - \sqrt{1-x^2-y^2}\big)$$

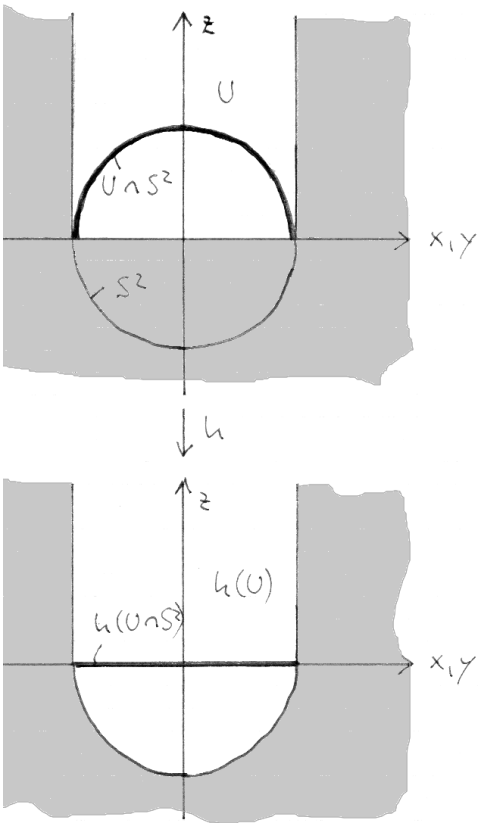
definierte differenzierbare Abbildung. Ihre Wertemenge ist

$$h(U) = \big\{(x,y,z) \in U^1 \times \mathbb{R} \,\big|\, z > -\sqrt{1-x^2-y^2}\big\},$$

und die durch $(x,y,z) \longmapsto h(x,y,z) := \big(x,\, y,\, z+\sqrt{1-x^2-y^2}\big)$ explizit gegebene Umkehrung zeigt, daß $h\colon U \to h(U)$ ein Diffeomorphismus ist. Ersichtlich gilt

$$U \cap S^2 = \big\{(x,y,z) \in U^1 \times (0,\infty) \,\big|\, z = \sqrt{1-x^2-y^2}\big\}$$

und damit $h(U \cap S^2) = U^1 \times \{0\} = h(U) \cap \big(\mathbb{R}^2 \times \{0\}\big)$: die nördliche Hemisphäre wird durch h senkrecht in die Äquatorebene projiziert und plattgedrückt.



Natürlich hat man eine entsprechende Karte für die südliche Hemisphäre und vier weitere, bei denen die besondere Rolle der Koordinate z von einer der beiden andere gespielt wird. Zusammen decken die Definitionsbereiche dieser sechs Karten die ganze Sphäre ab.

Die Definition 36.7 ist nicht von der Sorte, die man im konkreten Fall gern verifiziert. Das braucht man oft auch nicht, denn viele interessante Untermannigfaltigkeiten liefert der sogenannte Satz vom regulären Wert.

36.8 Definition Sei $X \subset \mathbb{R}^n$ offen, und $f\colon \mathbb{X} \longrightarrow \mathbb{R}^p$ differenzierbar. Ein Punkt $b \in \mathbb{R}^p$ heißt ein regulärer Wert von f, wenn jedes $a \in f^{-1}\{b\}$ ein regulärer Punkt von f ist. Die übrigen $b \in \mathbb{R}^p$, die also Wert mindestens eines kritischen Punktes von f sind, heißen kritische Werte von f.

Bemerkung Daß diejenigen $b \in \mathbb{R}^p$, die überhaupt keine Werte von f sind, danach als reguläre Werte gelten, klingt zwar paradox, ist aber wirklich so gemeint und auch zweckmäßig.

36.9 Satz vom **regulären Wert** Ist $X \subset \mathbb{R}^n$ offen, $f\colon X \longrightarrow \mathbb{R}^p$ eine C^1-Abbildung und $b \in \mathbb{R}^p$ ein regulärer Wert von f, so ist die Faser $f^{-1}\{b\} \subset X$ eine Untermannigfaltigkeit der Dimension $n-p$.

Beweis Wir dürfen von f die Konstante b abziehen und so $b = 0$ annehmen. Sei nun $a \in f^{-1}\{0\}$, dann ist a regulärer Punkt von f. Nach Satz 36.5 gibt es also eine Karte (U,h) um a, so daß $f \circ h^{-1}$ durch

$$\mathbb{R}^{n-p} \times \mathbb{R}^p \supset h(U) \ni (x,y) \overset{f\circ h^{-1}}{\longmapsto} y \in \mathbb{R}^p$$

beschrieben wird. Dann ist aber

$$U \cap f^{-1}\{0\} = \big\{u \in U \,\big|\, f(u) = 0\big\} = \big\{u \in U \,\big|\, (f \circ h^{-1})\big(h(u)\big) = 0\big\}$$

und deshalb

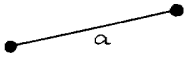
$$h\big(U \cap f^{-1}\{0\}\big) = h(U) \cap \big(\mathbb{R}^{n-p} \times \{0\}\big).$$

Also ist h eine Karte für $f^{-1}\{0\} \subset \mathbb{R}^n$.

36.10 Beispiele (1) Physiker reden gern von der "Zahl der Freiheitsgrade" eines Systems. Zum Beispiel hat ein aus n Massenpunkten bestehendes mechanisches System $3n$ Freiheitsgrade, weil der zugehörige sogenannte *Konfigurationsraum*, dessen Punkte den möglichen Lagen der Massenpunkte entsprechen, in offensichtlicher und schon mehrfach besprochener Weise $\mathbb{R}^{3n}$ ist — na ja, nicht ganz, denn weil sich nicht zwei Massenpunkte an ein und derselben Stelle aufhalten können, ist der Konfigurationsraum nur die offene Menge

$$X = \big\{(x_1, \dots, x_n) \in (\mathbb{R}^3)^n \,\big|\, x_i \neq x_j \text{ für } i \neq j\big\}.$$

Jedenfalls handelt es sich um eine $3n$-dimensionale Mannigfaltigkeit. Oft sind solche Systeme von Massenpunkten zusätzlich *Zwangsbedingungen* unterworfen, im günstigsten Fall differenzierbaren auf X erklärten Gleichungen. In einem einfachen Fall wäre $n = 2$, und der Abstand zwischen den beiden Massenpunkte durch eine (gedachte) masselose Stange der Länge $a > 0$ fixiert.



Statt $X \subset \mathbb{R}^6$ wäre der Konfigurationsraum dann nur

$$Y = \big\{(x_1, x_2) \in \mathbb{R}^3 \times \mathbb{R}^3 \,\big|\, |x_1 - x_2| = a\big\},$$

und das ist die Faser $f^{-1}\{a^2\}$ der Funktion

$$f\colon \mathbb{R}^3 \times \mathbb{R}^3 \longrightarrow \mathbb{R};\ (x_1, x_2) \mapsto |x_1 - x_2|^2.$$

Daß die von den Physikern vorgetragene Rechnung

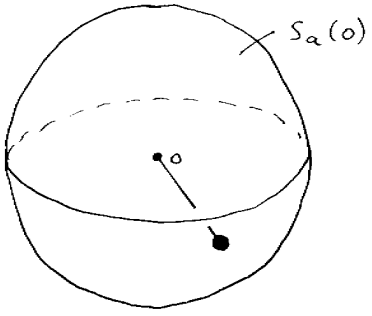
"sechs Freiheitsgrade minus eine Zwangsbedingung, bleiben fünf Freiheitsgrade"

hier aufgeht, liegt am Satz vom regulären Wert: Das gemäß Beispiel 34.4 flugs bestimmte Differential von f an der Stelle (x_1, x_2)

$$Df(x_1, x_2) = 2\,\big(\,x_1^{\mathrm{t}} - x_2^{\mathrm{t}} \quad x_2^{\mathrm{t}} - x_1^{\mathrm{t}}\,\big) \in \mathrm{Mat}(1 \times 6, \mathbb{R})$$

verschwindet genau dann, wenn $x_1 = x_2$, d.h. wenn $f(x_1, x_2) = 0$ ist. Deshalb ist $a^2 > 0$ ein regulärer Wert von f, und nach dem Satz $Y \subset X$ ist eine 5-dimensionale Untermannigfaltigkeit.

(2) Legt man in demselben Beispiel zusätzlich die Koordinaten des ersten Massenpunktes zu 0 fest, hat man ein Kugelpendel vor sich. Natürlich kann man den ersten Massenpunkt dann auch ganz weglassen, und der Konfigurationsraum wird die zweidimensionale Sphäre $S_a(0) \subset \mathbb{R}^3$.



Der Satz vom regulären Wert zeigt in noch einfacherer Weise als vorhin (wir werden das gleich in Beispiel (4) noch einmal sehen), daß es sich um eine zweidimensionale Untermannigfaltigkeit handelt, also — in Physikersprache — zwei Freiheitsgrade verbleiben.

(3) Wieviele Freiheitsgrade hat ein System von drei Massenpunkten, deren gegenseitige Abstände a_{12}, a_{13} und a_{23} festliegen? Die Physiker würden $9-3=6$ antworten und hätten damit meistens auch Recht; wann genau, das wird in Aufgabe 36.2 geklärt.

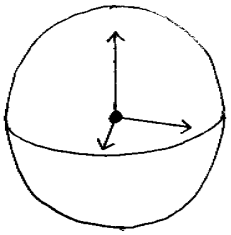
(4) Die entsprechende naive Rechnung für n Massenpunkte mit paarweise vorgeschriebenen Abständen kann nicht für alle $n \in \mathbb{N}$ richtig sein. Von den $3n$ Freiheitsgraden wären ja $n(n-1)/2$ Zwangsbedingungen abzuziehen, was für große n einen negativen Saldo läßt. Der Grund liegt darin, daß diese Zwangsbedingungen ab $n=5$ nicht mehr unabhängig sein können und deshalb alle Werte der durch sie definierten Abbildung $\mathbb{R}^{3n} \longrightarrow \mathbb{R}^{n(n-1)/2}$ kritisch sind — soweit sie überhaupt Werte sind, der Konfigurationsraum also nicht wegen Unverträglichkeit der Zwangsbedingungen leer ausfällt. Tatsächlich ist man schon bei $n=3$ (wenn es sich nicht eine spezielle Lage handelt) bei einem starren Körper angelangt, der sechs Freiheitsgrade hat, wie wir gleich noch genauer begründen werden.

(5) Sei $q\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ eine quadratische Form, durch eine Matrix $s \in \mathrm{Sym}(n,\mathbb{R})$ gegeben: $q(x) = x^{\mathrm{t}} s\, x$ für alle $x \in \mathbb{R}^n$. Im Beispiel 34.4 haben wir das Differential von q an der Stelle $x \in \mathbb{R}^n$ schon zu

$$Dq(x) = 2x^{\mathrm{t}} s \in \mathrm{Mat}(1\times n, \mathbb{R})$$

berechnet. Ist $x \in \mathbb{R}^n$ ein kritischer Punkt von q, so ist also $x^{\mathrm{t}} s = 0$ und damit erst recht $q(x) = x^{\mathrm{t}} s\, x = 0$, deshalb ist $0 \in \mathbb{R}$ der einzige kritische Wert von q. Damit ist für jedes von null verschiedene $b \in \mathbb{R}$ die Faser $q^{-1}\{b\} \subset \mathbb{R}^n$ eine $(n-1)$-dimensionale Untermannigfaltigkeit. Solche Untermannigfaltigkeiten nennt man Quadriken; hier einige populäre Beispiele in $\mathbb{R}^3$:

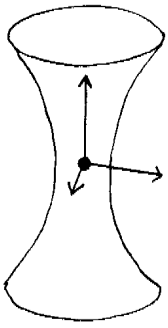
- $q(x,y,z) = x^2 + y^2 + z^2$:



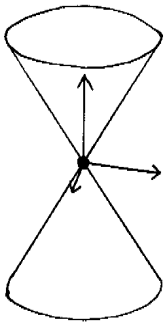
$q^{-1}\{1\} = S^2$

$q^{-1}\{-1\} = \emptyset$ (die leere Menge gilt als Mannigfaltigkeit jeder Dimension)

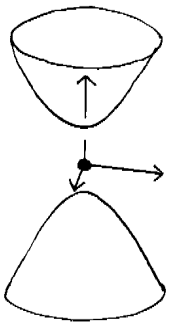
- $q(x,y,z) = x^2 + y^2 - z^2$:



$q^{-1}\{1\}$
einschaliges Hyperboloid



$q^{-1}\{0\}$
Kegel (keine Mannigfaltigkeit)



$q^{-1}\{-1\}$
zweischaliges Hyperboloid

(6) Die allgemeine lineare Gruppe $GL(n,\mathbb{R}) = \{x \in \mathrm{Mat}(n\times n,\mathbb{R}) \mid \det x \neq 0\}$ ist als offene Teilmenge von $\mathrm{Mat}(n\times n,\mathbb{R}) = \mathbb{R}^{n^2}$ eine Mannigfaltigkeit der Dimension n^2. Aber auch viele interessante Untergruppen von $GL(n,\mathbb{R})$ erweisen sich als Mannigfaltigkeiten; ich führe das für die orthogonale Gruppe $O(n) = \{x \in GL(n,\mathbb{R}) \mid x^{\mathrm{t}}x = 1\}$ als Beispiel vor.

Für jede Matrix $x \in \mathrm{Mat}(n\times n,\mathbb{R})$ ist $x^{\mathrm{t}}x$ eine symmetrische Matrix; wir haben deshalb eine Abbildung

$$\begin{array}{ccc} \mathrm{Mat}(n\times n,\mathbb{R}) & \xrightarrow{f} & \mathrm{Sym}(n,\mathbb{R}) \\ x & \mapsto & x^{\mathrm{t}}x \end{array}$$

in den Raum $\mathrm{Sym}(n,\mathbb{R}) = \mathbb{R}^{n(n+1)/2}$, und $O(n) = f^{-1}\{1\}$ ist die Faser dieser Abbildung über der Einheitsmatrix. Wieder nach Beispiel 34.4, angewendet auf die Matrizenmultiplikation in $\mathrm{Mat}(n\times n,\mathbb{R})$ als bilineare Abbildung, ist für beliebiges $a \in \mathrm{Mat}(n\times n,\mathbb{R})$

$$Df(a)(\xi) = \xi^{\mathrm{t}}a + a^{\mathrm{t}}\xi \quad \text{für alle } \xi \in \mathrm{Mat}(n\times n,\mathbb{R}),$$

und speziell für $a \in O(n)$ ist $Df(a)\colon \mathrm{Mat}(n\times n,\mathbb{R}) \longrightarrow \mathrm{Sym}(n,\mathbb{R})$ surjektiv: Sei $\sigma \in \mathrm{Sym}(n,\mathbb{R})$ beliebig, dann ist

$$Df(a)\left(\frac{1}{2}a\sigma\right) = \frac{1}{2}\sigma^{\mathrm{t}}a^{\mathrm{t}}a + \frac{1}{2}a^{\mathrm{t}}a\sigma = \sigma.$$

Die Einheitsmatrix ist also ein regulärer Wert von f, und nach Satz 36.9 die Gruppe $O(n)$ eine Untermannigfaltigkeit von $\mathrm{Mat}(n\times n,\mathbb{R})$ der Dimension $n^2 - n(n+1)/2 = n(n-1)/2$.

Untergruppen der allgemeinen linearen Gruppen, die zugleich Untermannigfaltigkeiten sind, zählen zu den sogenannten Lie-Gruppen. Weitere Beispiele sind $SL(n,\mathbb{R})$, $SL(n,\mathbb{C})$, $U(n)$ und viele andere. Aus der Tatsache, daß $SO(3) = \{x \in GL(n,\mathbb{R}) \mid x^{\mathrm{t}}x = 1,\ \det x > 0\}$ als relativ offene Untergruppe von $O(3)$ ebenso wie diese eine dreidimensionale Lie-Gruppe ist, ergeben sich die insgesamt sechs Freiheitsgrade eines starren Körpers (drei der Rotation zusätzlich zu denen der Translation).

Übungsaufgaben

36.1 $T \subset \mathbb{R}$ sei ein offenes Intervall mit $0 \in T$, und $\gamma\colon T \longrightarrow \mathbb{R}^p$ sei eine C^1-Kurve mit $\dot\gamma(0) \neq 0$. Beweisen Sie: Es gibt ein $\delta > 0$, so daß $\gamma(-\delta,\delta) \subset \mathbb{R}^p$ eine eindimensionale Untermannigfaltigkeit ist.

36.2 Wieviele Freiheitsgrade hat ein System von drei Massenpunkten im Raum, deren gegenseitige Abstände a_{12}, a_{13} und a_{23} festliegen? (Die Antwort hängt von den Werten dieser Abstände ab; bei der Rechnung ergibt sich ein spezieller Fall, der aber auch leicht zu raten ist.)

36.3 Konstruieren Sie für die Sphäre $S^n \subset \mathbb{R}^{n+1}$ explizit eine Menge von Untermannigfaltigkeitskarten (U_λ, h_λ), derart daß jeder Punkt von S^n in mindestens einem U_λ liegt. (Man käme dafür mit nur zwei Karten aus, aber es gibt eine besonders bequem hinzuschreibende Lösung mit $2n+2$ Karten.)

36.4 Konstruieren Sie für die Sphäre $S^n \subset \mathbb{R}^{n+1}$ zwei Untermannigfaltigkeitskarten $(U_\pm, h_\pm)$, so daß $S^n \subset U_+ \cup U_-$ gilt. Warum kann man nicht mit einer einzigen Karte auskommen?

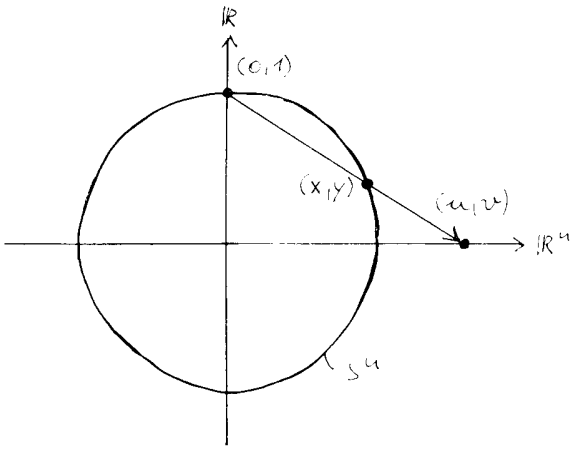
Tip: Die skizzierte sogenannte *stereographische Projektion* der Sphäre wird durch die Formel

$$(x,y) \longmapsto \left(\frac{1}{1-y}\,x,\ 1 - \frac{|x|^2 + (1-y)^2}{2(1-y)}\right)$$

bewerkstelligt, und

$$(u,v) \longmapsto \left(\frac{2(1-v)}{1+|u|^2}\, u,\ 1 - \frac{2(1-v)}{1+|u|^2} \right)$$

ist die Umkehrformel.



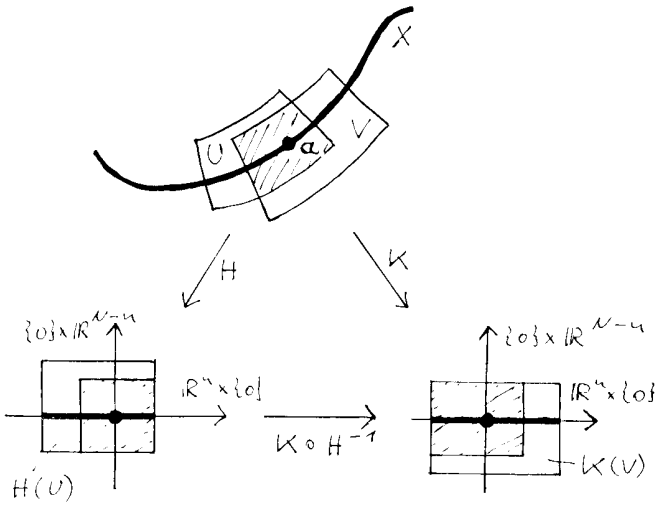
37 Tangentialvektoren

37.1 Lemma und **Definition** Sei $X \subset \mathbb{R}^N$ eine n-dimensionale Untermannigfaltigkeit und $a \in X$. Der durch Wahl einer Untermannigfaltigkeitskarte (U, H) für X um a definierte Vektorraum

$$T_aX := DH(a)^{-1}\big(\mathbb{R}^n \times \{0\}\big) \subset \mathbb{R}^N$$

hängt von dieser Wahl in Wirklichkeit nicht ab; er heißt der Tangentialraum von (oder an) X bei a. Die Elemente des Tangentialraums nennt man Tangentialvektoren.

Beweis Sei (V, K) eine weitere Untermannigfaltigkeitskarte für X um a. Der damit entstehende *Kartenwechsel*



$$K \circ H^{-1} \colon H(U \cap V) \longrightarrow K(U \cap V)$$

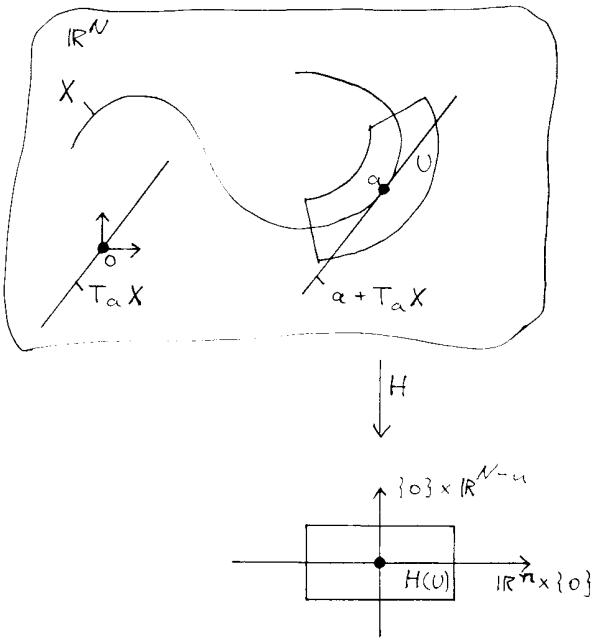
ist ein Diffeomorphismus, der Punkte der Form $(x, 0) \in \mathbb{R}^n \times \{0\}$ in solche derselben Art überführt; das gilt deshalb auch für sein Differential im Nullpunkt $D(K \circ H^{-1})(0) = DK(a) \circ DH(a)^{-1} \colon \mathbb{R}^N \longrightarrow \mathbb{R}^N$. Es ist also

$$DK(a) \circ DH(a)^{-1}\big(\mathbb{R}^n \times \{0\}\big) = \mathbb{R}^n \times \{0\}$$

oder gleichwertig

$$DH(a)^{-1}\big(\mathbb{R}^n \times \{0\}\big) = DK(a)^{-1}\big(\mathbb{R}^n \times \{0\}\big).$$

Ganz ähnlich wie man sich den Geschwindigkeitsvektor einer Kurve gern an die entsprechende Stelle geheftet denkt, stellt man den Tangentialraum T_aX der Anschaulichkeit halber meist um den Punkt $a \in \mathbb{R}^N$ verschoben dar, denkt also eher an den *affinen* Unterraum $a + T_aX$. Für die theoretischen Überlegungen und auch zum Rechnen ist das aber nicht praktisch.



Der Begriff des Tangentialraums erlaubt es, Abbildungen jetzt nicht nur zwischen offenen Teilmengen von $\mathbb{R}^N$, sondern auch zwischen Untermannigfaltigkeiten zu differenzieren. Für die Definition (und auch für konkrete Rechnung) zieht man wieder willkürliche Karten zu Hilfe und zeigt dann, daß deren genaue Wahl unerheblich ist. Sei also $X \subset \mathbb{R}^N$ eine n-dimensionale Untermannigfaltigkeit und (U, H) eine Untermannigfaltigkeitskarte für X. Wenn wir

$$H = \begin{pmatrix} H_1 \\ \vdots \\ H_N \end{pmatrix}$$

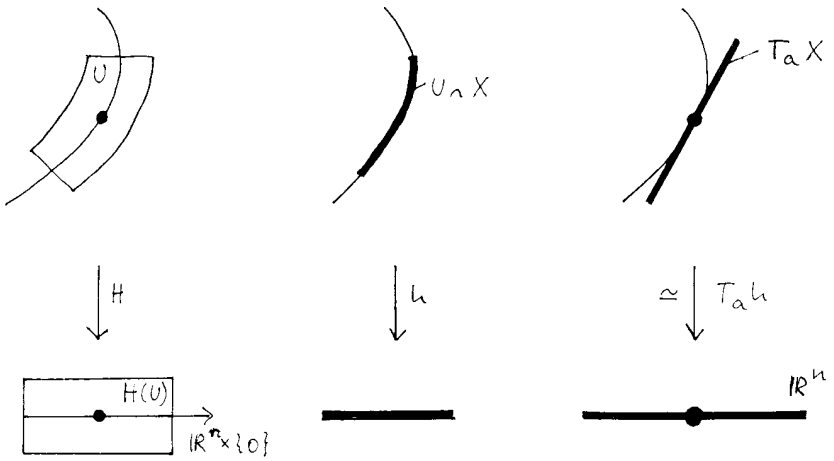
jetzt weniger zur Beschreibung von X selbst als vielmehr zur Beschreibung auf X lebender Objekte verwenden wollen, sind die letzten $N-n$ Komponenten von geringem Interesse; sie sind auf $U \cap X$ ja identisch null. Wir werden deshalb künftig statt mit H viel mehr mit der durch die ersten n Komponenten von H gegebenen Abbildung

$$h\colon U \cap X \longrightarrow h(U \cap X) \subset \mathbb{R}^n; \quad x \mapsto \begin{pmatrix} H_1(x) \\ \vdots \\ H_n(x) \end{pmatrix}$$

zu tun haben, die wir kurz als die zur Untermannigfaltigkeitskarte H gehörige *Karte von* X bezeichnen wollen. h ist bijektiv und ebenso wie h^{-1} stetig, und die Bildmenge $h(U \cap X) \subset \mathbb{R}^n$ ist offen. Eine Bezeichnung aus der folgenden Definition vorwegnehmend schreiben wir den aus dem Differential $DH(a)\colon \mathbb{R}^N \longrightarrow \mathbb{R}^N$ durch Einschränken gebildeten linearen Isomorphismus als

$$T_a h\colon T_a X \longrightarrow \mathbb{R}^n \times \{0\} = \mathbb{R}^n;$$

er wird gleich das Differential von h an der Stelle a heißen. Beachten Sie aber, daß im Augenblick nicht mal die Frage einen Sinn hat, ob h überhaupt differenzierbar ist, weil der Definitionsbereich von h keine offene Teilmenge von $\mathbb{R}^N$ ist. Lediglich in dem Sonderfall $n = N$, wo $h = H$ eine Karte für X im früheren Sinne ist, ist in der Tat $T_a h = DH(a)$.



37.2 Lemma und **Definition** $X \subset \mathbb{R}^N$ sei eine n-dimensionale, und $Y \subset \mathbb{R}^P$ eine p-dimensionale Untermannigfaltigkeit. Eine Abbildung $f\colon X \longrightarrow Y$ heißt an der Stelle $a \in X$ differenzierbar, wenn für irgendeine (und dann für jede) Wahl einer Untermannigfaltigkeitskarte (U, H) für X um a die Abbildung

$$f \circ h^{-1}\colon h(U \cap X) \longrightarrow \mathbb{R}^P$$

an der Stelle 0 differenzierbar ist, wobei $h\colon U \cap X \longrightarrow h(U \cap X)$ die zu H gehörige Karte von X ist. Unter dem Differential von f bei a versteht man dann

$$T_a f := D(f \circ h^{-1})(0) \circ T_a h;$$

dies ist eine von der Wahl der Karte H unabhängige lineare Abbildung $T_a f\colon T_a X \longrightarrow T_{f(a)} Y$.

Beweis Es sei $f \circ h^{-1}$ bei 0 differenzierbar. Übergang zu einer anderen Untermannigfaltigkeitskarte K mit zugehöriger Karte k macht aus $f \circ h^{-1}$, nötigenfalls nach Verkleinerung der offenen Menge U um a, die ebenfalls bei 0 differenzierbare Komposition $f \circ k^{-1} = (f \circ h^{-1}) \circ (h \circ k^{-1})$. Dabei bleibt das Differential nach der Kettenregel dasselbe:

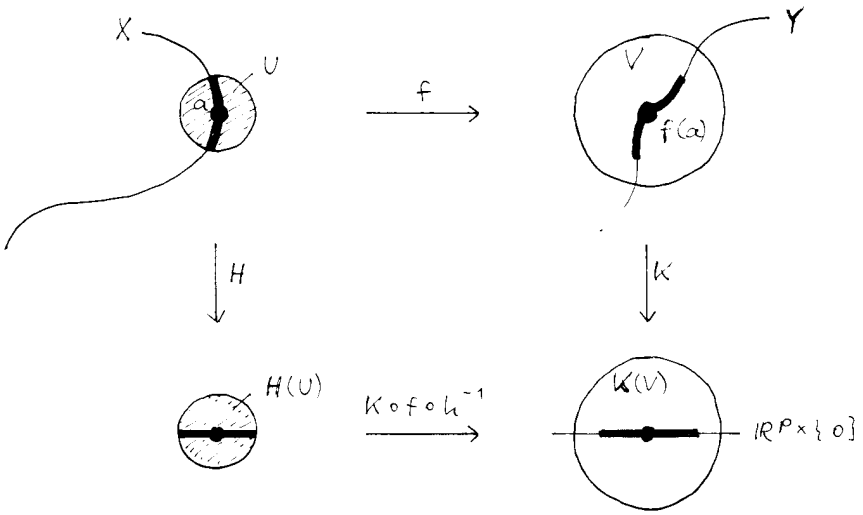
$$\begin{aligned} D(f \circ k^{-1})(0) \circ T_a k &= D\big((f \circ h^{-1}) \circ (h \circ k^{-1})\big)(0) \circ T_a k \\ &= D(f \circ h^{-1})(0) \circ D(h \circ k^{-1})(0) \circ T_a k \\ &= D(f \circ h^{-1})(0) \circ T_a h; \end{aligned}$$

die letzte Umformung macht davon Gebrauch, daß das Diagramm

$$\begin{array}{ccccc} \mathbb{R}^N & \xrightarrow{DK(a)} & \mathbb{R}^N & \xrightarrow{D(H \circ K^{-1})(0)} & \mathbb{R}^N \\ \uparrow & & \uparrow & & \uparrow \\ & & \mathbb{R}^n \times \{0\} & & \mathbb{R}^n \times \{0\} \\ & & \| & & \| \\ T_a X & \xrightarrow{T_a k} & \mathbb{R}^n & \xrightarrow{D(h \circ k^{-1})(0)} & \mathbb{R}^n \end{array}$$

gemäß den Definitionen kommutativ ist. Damit ist etabliert, daß der Begriff der Differenzierbarkeit und das Differential nicht von der Wahl von (U, H) abhängen.

Zu beweisen bleibt noch, daß $T_a f$ eine Abbildung nach $T_{f(a)} Y \subset \mathbb{R}^P$ ist. Wir wählen dazu zuerst eine Untermannigfaltigkeitskarte (V, K) um $f(a) \in Y$, dürfen dabei $V = U_\varepsilon\big(f(a)\big)$ voraussetzen. Weil f bei a stetig ist, können wir die Karte (U, H) so einschränken, daß $U \cap X$ unter f in V abgebildet wird.



Die jetzt definierte Komposition $K\circ f\circ h^{-1}$ nimmt ihre Werte in $\mathbb{R}^p\times\{0\}\subset\mathbb{R}^p\times\mathbb{R}^{P-p}$, und deshalb hat das Differential dieser Abbildung

$$D(K\circ f\circ h^{-1})(0)=DK\big(f(a)\big)\circ D(f\circ h^{-1})(0)$$

dieselbe Eigenschaft; also ist

$$\operatorname{Bild} T_af=\operatorname{Bild} D(f\circ h^{-1})(0)\subset DK\big(f(a)\big)^{-1}\big(\mathbb{R}^p\times\{0\}\big)=T_{f(a)}Y.$$

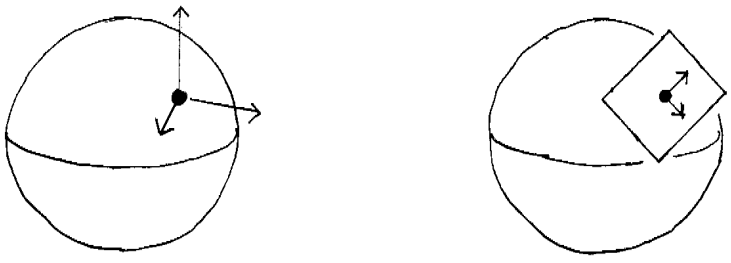
Bemerkungen Im Fall $n=N$ erhält man offensichtlich den alten Differenzierbarkeitsbegriff, und es wird $T_af=Df(a)$ das Differential im bisherigen Sinne. — Man kann die Definition auch so verstehen: Die einfachste auf X (lokal) definierte Abbildung, für die bisher kein Differential erklärt ist, ist die eingeschränkte Karte $h\colon U\cap X\longrightarrow\mathbb{R}^n$. Deren Differential T_ah wird nun wie oben als Einschränkung von $DH(a)$ definiert (wodurch h als Abbildung $U\cap X\xrightarrow{h}h(U\cap X)$ zu einem Diffeomorphismus wird), und der allgemeine Fall ergibt sich dann automatisch aus der Forderung, daß auch für die neuen Differentiale die Kettenregel gelten soll. Man überzeugt sich sofort davon, daß letzteres auch wirklich der Fall ist, daß also stets gilt:

37.3 Notiz $T_a(g\circ f)=T_{f(a)}\circ T_af$

Als Differential der Inklusion $X\hookrightarrow\mathbb{R}^N$ an der Stelle $a\in X$ kommt wie nicht anders zu erwarten die Inklusion $T_aX\hookrightarrow T_a\mathbb{R}^N=\mathbb{R}^N$ des Tangentialraums heraus, und damit ergibt sich die weitere

37.3½ Notiz Ist $f\colon X\longrightarrow\mathbb{R}^P$ die Einschränkung einer auf einer offenen Teilmenge von $\mathbb{R}^N$ definierten differenzierbaren Funktion F, so ist $T_af=(T_aF)|T_aX=DF(a)|T_aX$.

Beachten Sie, daß man nicht mehr von partiellen Ableitungen einer auf einer Untermannigfaltigkeit $X\subset\mathbb{R}^n$ definierten Funktion f reden kann; bewegt man sich von $a\in X$ aus auf einer achsenparallelen Geraden, so wird man im allgemeinen X und damit den Definitionsbereich von f sofort verlassen. Man kommt der Sache schon näher, wenn man versucht, Richtungsableitungen zu bilden, indem man sich längs in T_aX enthaltenen Geraden bewegt; aber auch das kann noch daran scheitern, daß der Tangentialraum bei a (in der anschaulichen Version $a+T_aX$) mit X zu wenige Punkte gemeinsam hat, wie das Beispiel der Sphären zeigt.

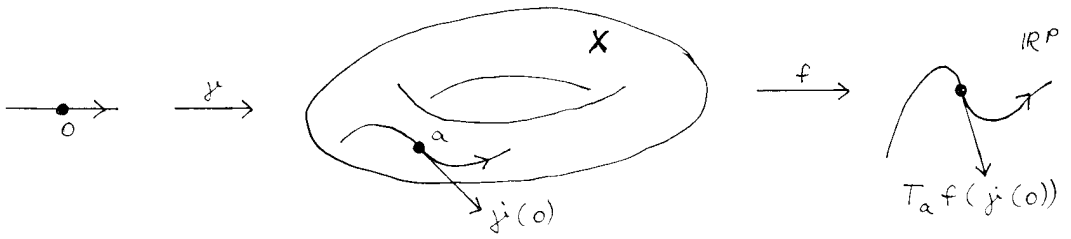


Es funktioniert aber der schon in der Notiz 34.12 propagierte Ansatz mittels Ableitungen längs differenzierbarer Kurven durch a; er liefert die folgende sehr anschauliche und zur Definition 37.2 alternative Beschreibung des Differentials.

37.4 Lemma Sei $X \subset \mathbb{R}^n$ eine Untermannigfaltigkeit und $a \in X$. Dann gilt:

(a) $T_aX \subset \mathbb{R}^n$ besteht aus allen Geschwindigkeitsvektoren $\dot\gamma(0)$ von differenzierbaren Kurven $\gamma\colon I \longrightarrow X$ mit $0 \in I$ und $\gamma(0) = a$.

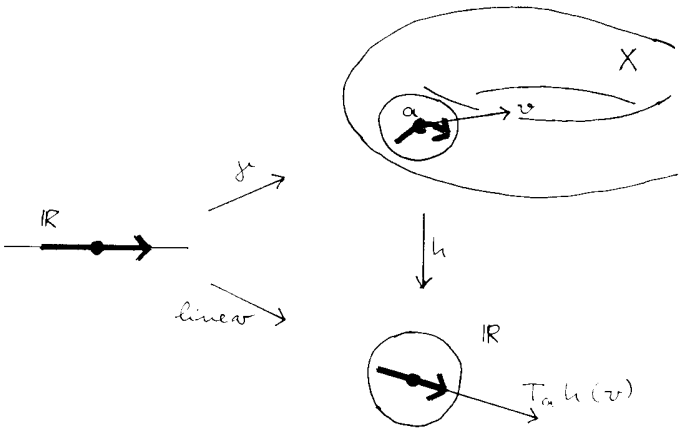
(b) Ist $f\colon X \longrightarrow \mathbb{R}^p$ bei a differenzierbar, so wirkt das Differential vermöge $T_af\big(\dot\gamma(0)\big) = (f\circ\gamma)^{\cdot}(0)$.



Beweis Die Vektoren $\dot\gamma(0) = T_0\gamma(1)$ gehören definitionsgemäß zu T_aX. Ist umgekehrt $v \in T_aX$ beliebig und (U, h) eine Karte für X um a, so definiert

$$t \mapsto h^{-1}\big(t\cdot Dh(a)v\big)$$

in einem kleinen offenen Intervall um 0 eine Kurve durch a mit Geschwindigkeitsvektor v.



Das beweist (a). Andererseits ist (b) nur eine andere Schreibweise der Kettenregel $T_af \circ T_0\gamma = T_0(f\circ\gamma)$.

Ist $Y \subset \mathbb{R}$ als Faser einer C^1-Abbildung über einem regulären Wert gegeben, so lassen sich die Tangentialräume von Y besonders einfach beschreiben:

37.5 Satz Sei $X \subset \mathbb{R}^n$ offen, $f\colon X \longrightarrow \mathbb{R}^p$ eine C^1-Abbildung und $b \in \mathbb{R}^p$ ein regulärer Wert von f. Für jedes $a \in Y := f^{-1}\{b\}$ ist dann $T_aY = \operatorname{Kern} Df(a)$.

Beweis Da $f|Y$ eine konstante Funktion ist, ist $Df(a)|T_aY = T_a(f|Y) = 0$, also $T_aY \subset \operatorname{Kern} Df(a)$. Weil beide Vektorräume dieselbe Dimension $n-p$ haben, folgt $T_aY = \operatorname{Kern} Df(a)$.

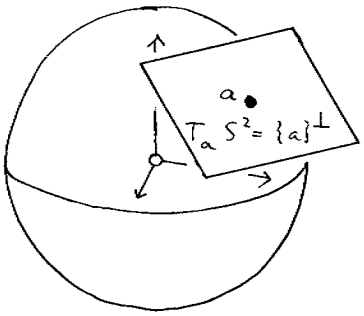
37.6 Beispiele (1) Die $(n-1)$-dimensionale Einheitssphäre ist die Faser $S^{n-1} = f^{-1}\{1\}$ der Funktion

$$\mathbb{R}^n \ni x \xrightarrow{f} |x|^2 \in \mathbb{R}.$$

Wegen $Df(a) = 2a^{\mathrm{t}} \in \mathrm{Mat}(1\times n, \mathbb{R})$ ist für $a \in S^{n-1}$ der Tangentialraum

$$T_a S^{n-1} = \mathrm{Kern}\, Df(a) = \{v \in \mathbb{R}^n \mid a^{\mathrm{t}} v = 0\} = \{a\}^{\perp} \subset \mathbb{R}^n$$

das orthogonale Komplement von a.



(2) Die orthogonale Gruppe $O(n)$ ist, wie wir in Beispiel 36.10(6) gesehen haben, die Faser der Abbildung

$$\mathrm{Mat}(n\times n, \mathbb{R}) \ni x \overset{f}{\longmapsto} x^{\mathrm{t}} x \in \mathrm{Sym}(n, \mathbb{R})$$

über dem regulären Wert 1. An der Stelle $a \in O(n)$ hat f das Differential

$$Df(a)\colon \mathrm{Mat}(n\times n, \mathbb{R}) \longrightarrow \mathrm{Sym}(n, \mathbb{R});\ \xi \mapsto \xi^{\mathrm{t}} a + a^{\mathrm{t}} \xi;$$

nach Lemma 37.5 ist der Tangentialraum dort also

$$T_a O(n) = \mathrm{Kern}\, Df(a) = \left\{\xi \in \mathrm{Mat}(n\times n, \mathbb{R}) \,\middle|\, \xi^{\mathrm{t}} a + a^{\mathrm{t}} \xi = 0\right\}.$$

Speziell für $a = 1$ wird

$$T_1 O(n) = \left\{\xi \in \mathrm{Mat}(n\times n, \mathbb{R}) \,\middle|\, \xi^{\mathrm{t}} + \xi = 0\right\}$$

der $n(n-1)/2$-dimensionale Vektorraum der schiefsymmetrischen Matrizen. Diese Matrizen erscheinen hier also als Tangentialvektoren an die Gruppe $O(n)$; die Physiker sprechen deshalb (vor allem für $n = 3$) von *infinitesimalen* Drehungen. Zum Beispiel definieren die Drehungen um die z-Achse eine Kurve

$$\mathbb{R} \ni t \overset{\gamma}{\longmapsto} \begin{pmatrix} \cos t & -\sin t & 0 \\ \sin t & \cos t & 0 \\ 0 & 0 & 1 \end{pmatrix} \in O(3)$$

in $O(3)$ mit Geschwindigkeitsvektor

$$\dot\gamma(0) = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \in T_1 O(3).$$

(3) Bei anderen Lie-Untergruppen von $GL(n, \mathbb{R})$ verhält sich das ganz analog; zum Beispiel ist die spezielle lineare Gruppe $SL(n, \mathbb{R})$ geradezu als die Faser $\det^{-1}\{1\} \subset \mathrm{Mat}(n\times n, \mathbb{R})$ definiert, und sie erweist sich als (n^2-1)-dimensionale Mannigfaltigkeit mit dem Tangentialraum

$$T_1 SL(n, \mathbb{R}) = \left\{x \in \mathrm{Mat}(n\times n, \mathbb{R}) \,\middle|\, \mathrm{tr}\, x = 0\right\}:$$

Aufgabe 34.2 liefert alles, was man dazu wissen muß.

Allgemein besteht eine sehr wirkungsvolle Standardtechnik im Umgang mit Lie-Gruppen darin, Rechnungen in der Gruppe auf solche in ihrem Tangentialraum an der Stelle 1, der sogenannten *Lie-Algebra* zurückzuführen, in Physikersprache also, statt mit wirklichen Gruppenelementen mit infinitesimalen zu rechnen.

Es ist klar, daß man auch bei auf Mannigfaltigkeiten definierten differenzierbaren Abbildungen von deren Rang (an einer Stelle) und damit von kritischen und regulären Punkten und Werten sprechen kann. Auch wenn all diese Begriffe im Prinzip über die Karten zugänglich sind, mag zum Beispiel die Suche nach kritischen Punkten ein praktisches Problem darstellen, denn es ist nicht leicht, mit Karten konkret zu rechnen (denken Sie daran, daß es ja schon in einfachen Fällen unmöglich sein kann, die Umkehrung einer Karte formelmäßig zu beschreiben). Für Funktionen, die auf einer nach dem Satz vom regulären Wert beschriebenen Untermannigfaltigkeit definiert sind, gibt es aber eine Methode, um die kritischen Werte ohne Verwendung von Karten zu finden.

37.7 Satz Sei $X \subset \mathbb{R}^n$ offen, $F\colon X \longrightarrow \mathbb{R}^p$ eine C^1-Abbildung und $b \in \mathbb{R}^p$ ein regulärer Wert von F, insbesondere also $Y := F^{-1}\{b\} \subset X$ eine $(n-p)$-dimensionale Untermannigfaltigkeit. Weiter sei $f\colon X \longrightarrow \mathbb{R}$ eine differenzierbare Funktion, und $a \in Y$ ein Punkt. Dann sind äquivalent:

- a ist kritischer Punkt von $f|Y$.
- Es gibt Zahlen $\lambda_1, \dots, \lambda_p \in \mathbb{R}$, so daß a kritischer Punkt der Funktion $f - \lambda_1 F_1 - \dots - \lambda_p F_p$ ist.

Bemerkungen Die Aufgabe, die kritischen Punkte von $f|Y$ zu finden, ist unter einer weniger klaren klassischen Formulierung bekannt, nämlich: "bestimme die kritischen Punkte a von f unter der Nebenbedingung $F(a) = b$". — Die Koeffizienten λ_i sind eindeutig bestimmt und heißen *Lagrange-Multiplikatoren*.

Beweis Sei (U, h) eine Karte um a, wie sie der Satz vom regulären Punkt liefert; zerlegt man die Punkte von $\mathbb{R}^n$ in $(x, y) \in \mathbb{R}^p \times \mathbb{R}^{n-p}$, so ist dann $(F \circ h^{-1})(x, y) = x$. Weil h ein Diffeomorphismus ist, gilt für jede differenzierbare Funktion $g\colon U \longrightarrow \mathbb{R}$

$$a \text{ kritischer Punkt von } g \iff 0 \text{ kritischer Punkt von } g \circ h^{-1}.$$

Speziell ist a genau dann ein kritischer Punkt der Funktion $f - \sum_{i=1}^{p} \lambda_i F_i$, wenn 0 ein kritischer Punkt von $f \circ h^{-1} - \sum_{i=1}^{p} \lambda_i F_i \circ h^{-1}$ ist, also wenn

$$D_j \left(f \circ h^{-1} - \sum_{i=1}^{p} \lambda_i F_i \circ h^{-1} \right)(0) = 0 \quad \text{für } j = 1, \dots, n$$

oder explizit

$$D_j(f \circ h^{-1})(0) = \begin{cases} \lambda_j & \text{für } j = 1, \dots, p \\ 0 & \text{für } j = p{+}1, \dots, n \end{cases}$$

gilt. Natürlich *existieren* Koeffizienten λ_i mit dieser Eigenschaft genau dann, wenn

$$D_j(f \circ h^{-1})(0) = 0 \quad \text{für } j = p{+}1, \dots, n$$

ist. Weil h (bis auf die vertauschten Rollen von x und y) eine Untermannigfaltigkeitskarte für $Y \subset X$ ist, also $U \cap Y$ genau auf $h(U) \cap \big(\{0\} \times \mathbb{R}^{n-p}\big)$ sendet, bedeutet diese letzte Bedingung gerade, daß a ein kritischer Punkt von $f|Y$ ist.

37.8 Beispiel Wir greifen Beispiel 30.15 wieder auf: Die durch die symmetrische Matrix $s \in \mathrm{Sym}(n, \mathbb{R})$ gegebene quadratische Form $q\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ werde auf die Sphäre $S^{n-1} \subset \mathbb{R}^n = \big\{x \in \mathbb{R}^n \,\big|\, |x|^2 = 1\big\}$ eingeschränkt. Um die kritischen Punkte von $q|S^{n-1}$ zu bestimmen, sind zunächst die kritischen Punkte der Hilfsfunktion $\mathbb{R}^n \ni x \longmapsto q(x) - \lambda|x|^2 \in \mathbb{R}$ mit dem noch unbekannten Lagrange-Multiplikator λ zu ermitteln: das sind die Lösungen der Gleichung

$$2x^{\mathrm{t}} s - \lambda \cdot 2x^{\mathrm{t}} = 0$$

oder der äquivalenten Gleichung

$$sx = \lambda x.$$

Die kritischen Punkte von $q|S^{n-1}$ sind deshalb genau die auf die Länge 1 normierten Eigenvektoren der Matrix s und damit der quadratischen Form q; die zugehörigen kritischen Werte sind die Eigenwerte selbst.

Insbesondere ist die Zahl der kritischen Punkte genau dann endlich (und zwar $2n$), wenn die n Eigenwerte paarweise verschieden sind.

Wir wollen jetzt kurz darüber reden, was im mehrdimensionalen Rahmen zwei- und mehrmalige Differenzierbarkeit einer Abbildung bedeutet. Das ist zwar nicht so offensichtlich wie bei auf Intervallen definierten Funktionen $f\colon I \longrightarrow \mathbb{R}$, wo die Ableitung f' eine Funktion derselben Art ist, aber im Prinzip doch analog. Sei zunächst $X \subset \mathbb{R}^n$ eine offene Menge und $f\colon X \longrightarrow \mathbb{R}^p$ differenzierbar. Wir bilden aus den Differentialen $Df(x)$ für alle $x \in X$ provisorisch die Abbildungen

$$X \times \mathbb{R}^n \longrightarrow \mathbb{R}^p;\ (x,v) \mapsto Df(x)\cdot v$$

und

$$X \longrightarrow \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^p);\ x \mapsto Df(x);$$

offenbar enthalten beide genau dieselbe Information. Die Frage, ob sie ihrerseits differenzierbar sind, läuft rechnerisch in beiden Fällen auf die Frage hinaus, ob die pn partiellen Ableitungen $\dfrac{\partial f_i}{\partial x_j}$ differenzierbare Funktionen auf X sind. Ist das der Fall, nennt man f zweimal differenzierbar. Von größerer praktischer Bedeutung ist aber die zweimalige *stetige* Differenzierbarkeit, und man verfügt allgemein die induktive

37.9 Definition Sei $X \subset \mathbb{R}^n$ offen. Eine stetige Abbildung $f\colon X \longrightarrow \mathbb{R}^p$ nennt man auch eine $\underline{C^0\text{-Abbildung}}$; man nennt sie eine $\underline{C^k\text{-Abbildung}}$ für $k > 0$, wenn sie differenzierbar ist und die Abbildung $X \times \mathbb{R}^n \ni (x,v) \mapsto Df(x)(v) \in \mathbb{R}^p$ eine C^{k-1}-Abbildung ist. Ist f eine C^k-Abbildung für jedes $k \in \mathbb{N}$, so spricht man von einer $\underline{C^\infty\text{-Abbildung}}$.

Aus Satz 34.8 ergibt sich sogleich die praktisch wichtige

37.10 Notiz $f\colon X \longrightarrow \mathbb{R}^p$ ist genau dann eine C^k-Abbildung, wenn für jedes $r \in \{0,1,\dots,k\}$ alle partiellen Ableitungen r-ter Ordnung

$$D_{j_1}D_{j_2}\cdots D_{j_r}f_i = \frac{\partial^r f_i}{\partial x_{j_1}\partial x_{j_2}\cdots\partial x_{j_r}} \quad \text{mit } i \in \{1,\dots,p\} \text{ und } j_1,\dots,j_r \in \{1,\dots,n\}$$

existieren und stetige Funktionen auf X sind.

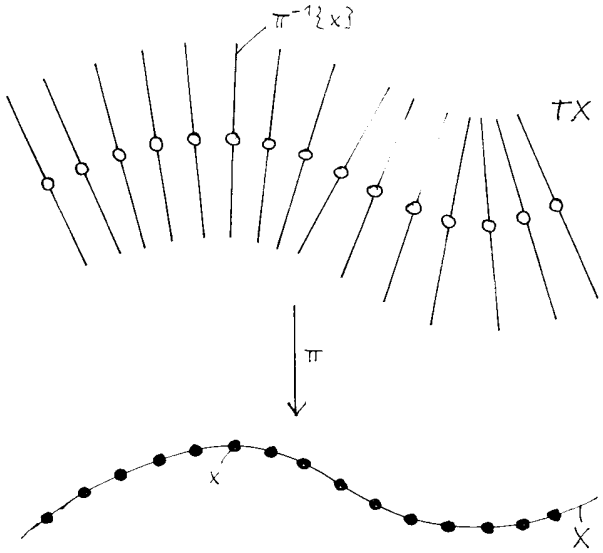
Nicht beantwortet wird damit die Frage, welches Objekt man denn nun etwa als die zweite Ableitung einer C^2-Abbildung $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ ansehen soll. Darauf werde ich im Abschnitt 43 zurückkommen; vorerst wollen wir uns noch mit den ersten Ableitungen befassen. Festhalten können wir soweit aber, daß alles in den beiden vorigen Abschnitten über C^1-Abbildungen Gesagte sinngemäß auch für C^k mit $k \geq 1$ gilt; man kann von C^k-Diffeomorphismen, -Karten und damit auch von C^k-Mannigfaltigkeiten reden, und es ist auch nicht schwer einzusehen, daß der Satz von der lokalen Umkehrung und seine Folgerungen C^k-Ergebnisse liefern, wenn man C^k-Daten einfüttert. Tatsächlich sind bei all dem die Unterschiede zwischen C^1 und höherer Differenzierbarkeitsordnung nicht besonders spannend. Für die meisten Zwecke kommt man denn auch mit C^1 oder C^2 aus; wir wollen es uns aber bequem machen und treffen die großzügige *Vereinbarung*, daß mit "differenzierbar" ab jetzt "C^∞-differenzierbar" gemeint sein soll, wenn nicht ausdrücklich etwas Anderes gesagt wird.

Wir wollen jetzt die an den einzelnen Stellen gebildeten Differentiale T_xf einer Abbildung zu einem globalen Differential Tf zusammenfassen. Wenn f auf einer offenen Teilmenge $X \subset \mathbb{R}^n$ definiert ist, wäre an sich jede der beiden genannten Abbildungen $X \times \mathbb{R}^n \longrightarrow \mathbb{R}^p$ und $X \longrightarrow \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^p)$ ein geeigneter Kandidat dafür. Keiner der beiden Ansätze läßt sich aber auf den allgemeineren Fall übertragen, daß der Definitionsbereich von f eine Untermannigfaltigkeit $X \subset \mathbb{R}^N$ ist: die Differentiale T_xf sind dann ja auf lauter verschiedenen Vektorräumen, nämlich den Tangentialräumen T_xX mit $x \in X$ erklärt. Abhilfe schafft ein neues Gebilde, das man sich als eine disjunkte Vereinigung all dieser Tangentialräume vorstellen kann.

37.11 Definition $X \subset \mathbb{R}^N$ sei eine n-dimensionale C^k-Untermannigfaltigkeit ($k \geq 1$). Dann heißt

$$TX := \big\{(x,v) \in X \times \mathbb{R}^N \,\big|\, v \in T_xX\big\}$$

zusammen mit der Projektion $TX \ni (x,v) \overset{\pi}{\longmapsto} x \in X$ das Tangentialbündel von X. Die Projektion hat als Fasern $\pi^{-1}\{x\} = \{x\} \times T_xX$ also im wesentlichen die Tangentialräume T_xX.



(Wenn die Tangentialräume T_xX hier scheinbar nicht mehr tangential zu X sind, liegt das daran, daß unter der Inklusion $TX \subset X \times \mathbb{R}^N \subset \mathbb{R}^N \times \mathbb{R}^N$ die Punkte von X in den ersten, die Tangentialvektoren dagegen in den zweiten Faktor $\mathbb{R}^N$ fallen.)

Ist $Y \subset \mathbb{R}^p$ eine weitere Mannigfaltigkeit und $f\colon X \longrightarrow Y$ eine C^k-Abbildung ($k>0$), so heißt

$$Tf\colon TX \longrightarrow TY;\ (x,v) \mapsto \big(f(x), T_xf(v)\big)$$

das Differential von f.

Bemerkungen Speziell für eine offene Teilmenge $X \subset \mathbb{R}^n$ sind alle Tangentialräume von X gleich, nämlich ganz $\mathbb{R}^n$, und das Tangentialbündel ist das Produkt $TX = X \times \mathbb{R}^n$. — Ist $X \subset \mathbb{R}^N$ eine n-dimensionale Untermannigfaltigkeit, so ist $TX \subset X \times \mathbb{R}^N \subset \mathbb{R}^N \times \mathbb{R}^N$ eine Untermannigfaltigkeit der Dimension $2n$, denn ist (U,H) eine Untermannigfaltigkeitskarte für X, so ist

$$TH\colon U \times \mathbb{R}^N \ni (x,v) \longmapsto \big(H(x), DH(x)\cdot v\big) \in H(U) \times \mathbb{R}^N$$

eine (ebenfalls beliebig oft differenzierbare) Untermannigfaltigkeitskarte für $TX \subset \mathbb{R}^N \times \mathbb{R}^N$, bis auf eine unbedeutende Koordinatenvertauschung: es ist

$$\begin{aligned} TH\big((U \times \mathbb{R}^N) \cap TX\big) &= TH\big(T(U \cap X)\big) \\ &= H(U \cap X) \times \big(\mathbb{R}^n \times \{0\}\big) \\ &= \Big(H(U) \cap \big(\mathbb{R}^n \times \{0\}\big)\Big) \times \big(\mathbb{R}^n \times \{0\}\big) \\ &= TH(U \times \mathbb{R}^N) \cap \big(\mathbb{R}^n \times \{0\} \times \mathbb{R}^n \times \{0\}\big). \end{aligned}$$

Diese Karte hat noch eine bemerkenswerte zusätzliche Eigenschaft: Sie bildet die im Bündel TX enthaltene Faser $T_xX = \{x\} \times T_xX$ linear isomorph auf die Faser $T_{H(x)}\big(\mathbb{R}^n \times \{0\}\big)$ des Tangentialbündels von $\mathbb{R}^n \times \{0\}$ ab. — Das Differential Tf enthält definitionsgemäß eine Kopie von f selbst; sind etwa im Eindimensionalen $X \subset \mathbb{R}$ und $Y \subset \mathbb{R}$ offene Intervalle, so ist Tf die Abbildung

$$TX = X \times \mathbb{R} \ni (x,v) \longmapsto \big(f(x), f'(x)v\big) \in Y \times \mathbb{R} = TY.$$

Tf ist nur noch eine C^{k-1}-Abbildung; der Hauptvorteil unserer C^∞-Vereinbarung ist, daß Tf damit ebenso differenzierbar bleibt wie f selbst. Tf ist außerdem "fasernweise linear": Für jedes $x \in X$ bildet Tf die Faser

$\{x\} \times T_xX$ in $\{f(x)\} \times T_{f(x)}Y$ ab, und zwar vermöge der linearen Abbildung $T_xf: T_xX \longrightarrow T_{f(x)}Y$. — Die Kettenregel nimmt jetzt die elegante Form

$$T(g \circ f) = Tg \circ Tf$$

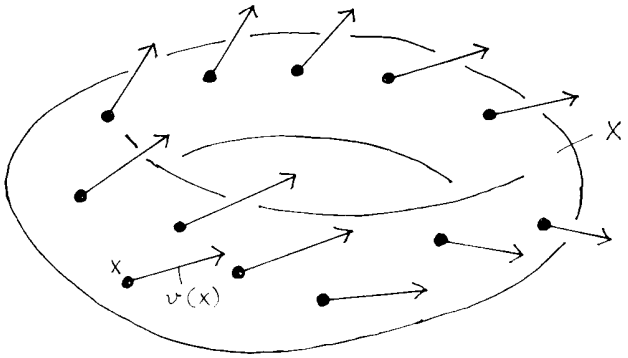
an, freilich ohne daß damit etwas inhaltlich Neues verbunden wäre.

Das Tangentialbündel ist auch der richtige Rahmen, um über Vektorfelder zu reden, die in der Physik ja allgegenwärtig sind.

37.12 Definition Sei X eine Mannigfaltigkeit. Unter einem Vektorfeld auf X verstehen die Mathematiker eine Abbildung $v: X \longrightarrow TX$ derart, daß

$$X \xrightarrow{v} TX \xrightarrow{\pi} X$$

die identische Abbildung ist, v also jedem Punkt $x \in X$ einen Tangentialvektor an der Stelle x zuweist.



Manchmal spricht man von einem *tangentialen* Vektorfeld, um zu betonen, daß die Vektoren $v(x) \in T_xX$ eben tangential zu X sind. Warum Physiker die so erklärten Vektorfeldern zusätzlich noch als kontravariant bezeichnen, werden wir bald noch besprechen.

Wir wollen genau studieren, wie man ein solches Vektorfeld lokal in einer Karte beschreiben kann. An dieser Stelle sei daran erinnert, daß mit einer Mannigfaltigkeit X genauer eine Untermannigfaltigkeit $X \subset \mathbb{R}^N$ (für irgendein $N \in \mathbb{N}$) gemeint ist. Bezüglich der Karten vereinbaren wir nun die

37.12½ Sprechweise Unter einer Karte für eine Mannigfaltigkeit X wollen wir künftig einfach einen Diffeomorphismus

$$U \xrightarrow{h} h(U) \subset \mathbb{R}^n$$

zwischen einer in X offenen Menge $U \subset X$ und einer offenen Menge $h(U) \subset \mathbb{R}^n$ verstehen.

Wie zu Anfang dieses Abschnitts besprochen erhält man solche Karten durch Einschränken von Untermannigfaltigkeitskarten für $X \subset \mathbb{R}^N$; man kann sogar zeigen, daß jede Karte so entsteht. Im übrigen ist die neue Sprechweise völlig analog zur Definition 36.1, in der X eine offene Teilmenge von $\mathbb{R}^n$ ist. Genau wie dort notiert man als Karte meist (U, h), und mit einer Karte um $a \in X$ meint man wieder eine mit $h(a) = 0$.

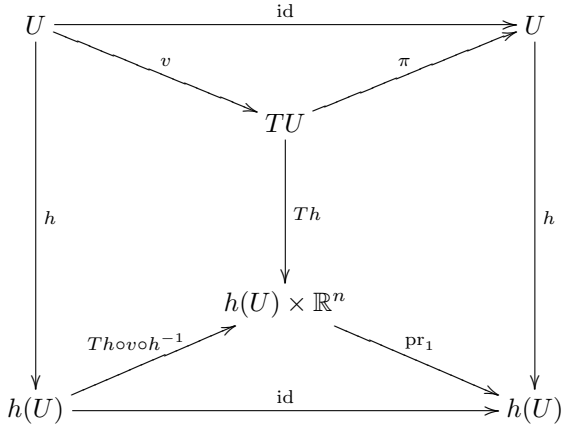
Bemerkung Wenn Sie sich mit dem Begriff der Mannigfaltigkeit (bisher) gar nicht anfreunden können, können Sie vieles von dem, was folgt, trotzdem sinnvoll studieren, indem Sie statt "n-dimensionale Mannigfaltigkeit" durchweg "offene Teilmenge von $\mathbb{R}^n$" lesen. Die Karten werden dann zu den Karten im Sinne der Definition 36.1, jeder Tangentialraum ist $\mathbb{R}^n$, und das Differential ist das klassische unter 34.2 eingeführte, das Sie auf Wunsch als Jacobi-Matrix auffassen können. Bei diesem begrifflich einfacheren Standpunkt entgeht Ihnen natürlich manches, was im Leben wichtig ist; der Konfigurationsraum des Kugelpendels ist nun einmal eine Sphäre und keine offene Teilmenge der Ebene. Unter rechnerischen Aspekten laufen beide Standpunkte aber auf dasselbe hinaus: Die Möglichkeit, Differentialrechnung auf Mannigfaltigkeiten zu treiben, spiegelt letztlich die Tatsache wieder, daß die in offenen Teilmengen von $\mathbb{R}^n$ betrachteten Objekte (wie Vektorfelder

und die bald zu besprechenden Differentialformen) in beliebigen Karten (im Sinne der alten Definition 36.1) beschrieben werden können und dabei in bestimmter Weise mit Koordinatenwechseln verträglich sind.

Sei nun also v ein Vektorfeld auf der n-dimensionalen Mannigfaltigkeit X, und sei (U,h) eine Karte für X. Das Differential Th ist ein fasernweise linearer Diffeomorphismus

$$TU = \big\{(x,v) \in U \times T_xX \,\big|\, v \in T_xX\big\} \xrightarrow{Th} T\big(h(U)\big) = h(U) \times \mathbb{R}^n,$$

und es entsteht das folgende kommutative Diagramm:



Man wird demnach das Vektorfeld $Th \circ v \circ h^{-1}$ auf $h(U)$ als das "in der Karte h geschriebene" Vektorfeld v ansehen. Aufgrund der Kommutativität des unteren Dreiecks hat es die Gestalt

$$\mathbb{R}^n \supset h(U) \ni z \longmapsto \big(z, g(z)\big) \in h(U) \times \mathbb{R}^n,$$

worin die zweite Komponente

$$h(U) \ni z \longmapsto g(z) = \begin{pmatrix} g_1(z) \\ \vdots \\ g_n(z) \end{pmatrix} \in \mathbb{R}^n$$

keinen Einschränkungen mehr unterliegt. In dieser Form kennen Sie die Vektorfelder aus der Physik: als $\mathbb{R}^n$-wertige Funktionen von n Variablen. Die Mathematiker ziehen es aber in der Regel vor, den Bezug auf die Karte h sichtbar zu machen, und schreiben das g entsprechende Vektorfeld v (genauer $v|U$) traditonell

$$v = \sum_{j=1}^{n} (g_j \circ h)\, \frac{\partial}{\partial h_j};$$

in dem konkreten Beispiel $n = 3$ und $g(z) = \begin{pmatrix} z_1^3 \\ z_2 z_3 \\ \cos z_2 \end{pmatrix}$ also

$$v = h_1^3\, \frac{\partial}{\partial h_1} + h_2 h_3\, \frac{\partial}{\partial h_2} + \cos h_2\, \frac{\partial}{\partial h_3}.$$

Die auf den ersten Blick völlig absurd erscheinende Idee, das der konstanten Funktion $g(z) = e_j \in \mathbb{R}^n$ entsprechende Vektorfeld $v\colon U \longrightarrow TU$ mit $\dfrac{\partial}{\partial h_j}$ zu bezeichnen, findet ihre Rechtfertigung, wenn man mit dem Differential einer Funktion $f\colon U \longrightarrow \mathbb{R}$ komponiert: Die die interessante Information tragende zweite Komponente der Komposition $Tf \circ v\colon U \xrightarrow{v} TU \xrightarrow{Tf} T\mathbb{R} = \mathbb{R} \times \mathbb{R}$ ergibt sich an der Stelle $x \in U$ zu

$$D(f \circ h^{-1})\big(h(x)\big)\, g\big(h(x)\big) = D(f \circ h^{-1})\big(h(x)\big)\, e_j = D_j((f \circ h^{-1})\big(h(x)\big) = \frac{\partial (f \circ h^{-1})}{\partial h_j}\,\big(h(x)\big);$$

in diesem Sinne wirkt $\frac{\partial}{\partial h_j}$ auf die Funktion f tatsächlich durch partielles Differenzieren nach der Koordinaten h_j. Übrigens muß ich zugeben, daß auch die Mathematiker diese partielle Ableitung schon mal salopp als $\frac{\partial f}{\partial h_j}$ schreiben, obwohl die partielle Ableitung nicht von f, sondern eben von $f \circ h^{-1}$ gebildet wird. Eine gewisse Rechtfertigung liegt darin, daß die Verwendung der Karte $U \xrightarrow{h} h(U) \subset \mathbb{R}^n$ den Koordinaten des Ziel-$\mathbb{R}^n$ automatisch die Namen $h_1, \dots, h_n$ zuweist und damit $f \circ h^{-1}$ die Bedeutung von "f als Funktion der h_j" bekommt. Leider suggeriert die Notation außerdem, für je zwei (differenzierbare) Funktionen f und h sei eine partielle Ableitung $\frac{\partial f}{\partial h}$ definiert, was keineswegs der Fall ist: die partielle Ableitung $\frac{\partial f}{\partial h_j}$ nimmt auf die gesamte Karte h Bezug, nicht nur auf ihre j-te Komponente.

37.13 Beispiel Wir verwenden in $\mathbb{R}^2$ einmal die identische Karte $x = (x_1, x_2)$ und einmal die Karte $h = (h_1, h_2)$ mit

$$h_1 = x_1 \quad \text{und} \quad h_2 = x_1 + x_2.$$

Weil das Differential von h hier die konstante Matrix

$$Dh = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \in \mathbb{R}^2$$

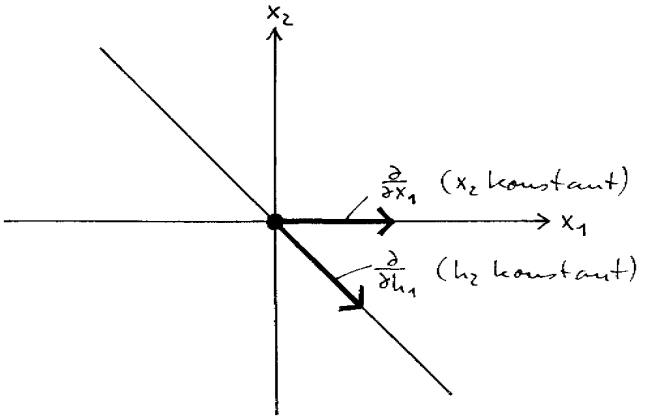
ist, entnimmt man den beiden kommutativen Diagrammen (für $j = 1, 2$)

$$\begin{array}{ccc} \mathbb{R}^2 & \xrightarrow{\frac{\partial}{\partial x_j}} & T\mathbb{R}^2 \\ \downarrow{\scriptstyle h} & & \downarrow{\scriptstyle Th} \\ \mathbb{R}^2 & \xrightarrow{Th \circ \frac{\partial}{\partial x_j} \circ h^{-1}} & T\mathbb{R}^2 \end{array}$$

leicht $\frac{\partial}{\partial x_1} = \frac{\partial}{\partial h_1} + \frac{\partial}{\partial h_2}$ und $\frac{\partial}{\partial x_2} = \frac{\partial}{\partial h_2}$. Das Vektorfeld

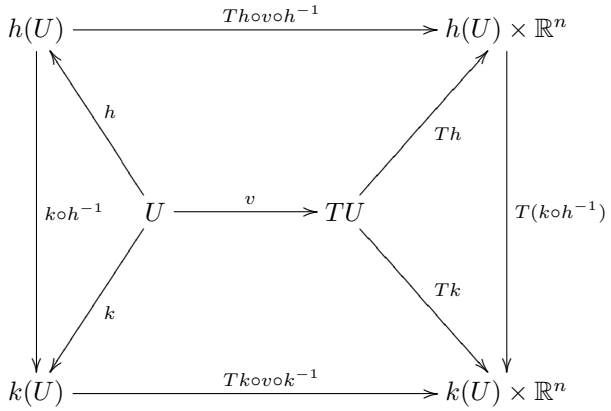
$$\frac{\partial}{\partial h_1} = \frac{\partial}{\partial x_1} - \frac{\partial}{\partial x_2}$$

ist also trotz $h_1 = x_1$ nicht dasselbe wie $\frac{\partial}{\partial x_1}$! Das leuchtet auch anschaulich unmittelbar ein, denn wenn man f partiell nach x_1 oder h_1 differenziert, hält man ja die jeweils andere Variable x_2 bzw. h_2 fest: die beiden Funktionen einer Variablen, die man dabei differenziert, sind also ganz verschieden.



Jeder, der in seinem Leben mit Vektorfeldern zu tun hat (selbst wenn es nur Felder auf offenen Teilmengen von $\mathbb{R}^n$ sind), muß wissen, wie die Kartendarstellung eines Vektorfelds auf Kartenwechsel reagiert. Allgemein

hat man ein Vektorfeld $v\colon X \longrightarrow TX$ auf der n-dimensionalen Mannigfaltigkeit X zu betrachten, und man möchte die Darstellungen von v bezüglich zweier Karten (U,h) und (V,k) miteinander vergleichen. Dabei darf man U und V durch ihren Durchschnitt ersetzen, also annehmen, daß $U = V$ ist. Nun, die aus der linearen Algebra vertraute Methode besteht darin, die maßgeblichen kommutativen Diagramme zusammenzusetzen:



Ist also v bezüglich der Karte k durch

$$k(U) \ni z \longmapsto g(z) \in \mathbb{R}^n$$

gegeben, so liest sich dasselbe Vektorfeld in der Karte h so:

$$h(U) \ni y \longmapsto D(k\circ h^{-1})(y)^{-1}\cdot(g\circ k\circ h^{-1})(y) \in \mathbb{R}^n$$

37.14 Beispiel Das in kartesischen Koordinaten $k = (x,y)$ durch die $\mathbb{R}^2$-wertige Funktion g gegebene Vektorfeld auf $\mathbb{R}^2$ soll (auf einer geeigneten offenen Menge) in Polarkoordinaten $h = (r,\varphi)$ ausgedrückt werden. Der Kartenwechsel ist hier die als Beispiel 35.5 eingeführte Polarkoordinatenabbildung

$$k\circ h^{-1} = \Phi\colon \begin{pmatrix} r \\ \varphi \end{pmatrix} \longmapsto \begin{pmatrix} r\cos\varphi \\ r\sin\varphi \end{pmatrix}$$

mit $D\Phi(r,\varphi) = \begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix}$, also $D\Phi(r,\varphi)^{-1} = \frac{1}{r}\begin{pmatrix} r\cos\varphi & r\sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}$. Folglich ist

$$\begin{pmatrix} r \\ \varphi \end{pmatrix} \longmapsto \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\frac{1}{r}\sin\varphi & \frac{1}{r}\cos\varphi \end{pmatrix} g(r\cos\varphi, r\sin\varphi)$$

die Polarkoordinatendarstellung des Vektorfelds.

Ich persönlich mag Diagramme und deshalb auch diese Methode. Eine mehr rechnerische Alternative benutzt die $\frac{\partial}{\partial h_j}$-Symbolik und funktioniert ebenso zuverlässig; sie ist allerdings nicht mit dem Matrizenkalkül verträglich. Das Vektorfeld v wird hier in der lässigen Form

$$v = \sum_{j=1}^{n} g_j \frac{\partial}{\partial h_j}$$

geschrieben; dabei ist unterstellt, daß mit g_j eigentlich die Funktion $g_j\circ h$ auf $h(U)$ gemeint ist. Es geht darum, darin $\frac{\partial}{\partial h_j}$ durch die $\frac{\partial}{\partial k_i}$ $(i = 1,\dots,n)$ auszudrücken. Dazu stellt man sich hilfsweise eine Funktion f vor, auf die die Vektorfelder wirken, rechnet nach der Kettenregel (in weiterhin lässiger Schreibweise)

- $$\frac{\partial}{\partial h_j} f = \frac{\partial}{\partial h_j}\big((f\circ k^{-1})\circ k\big) = \sum_{i=1}^{n}\left(\frac{\partial}{\partial k_i} f\right)\frac{\partial k_i}{\partial h_j}$$

und läßt das f dann wieder weg:

$$\frac{\partial}{\partial h_j} = \sum_{i=1}^{n} \frac{\partial k_i}{\partial h_j} \frac{\partial}{\partial k_i}$$

Man braucht das nur oben einzusetzen und erhält

$$v = \sum_{i,j=1}^{n} g_j \frac{\partial k_i}{\partial h_j} \frac{\partial}{\partial k_i}.$$

Speziell in unserem Beispiel ergibt sich

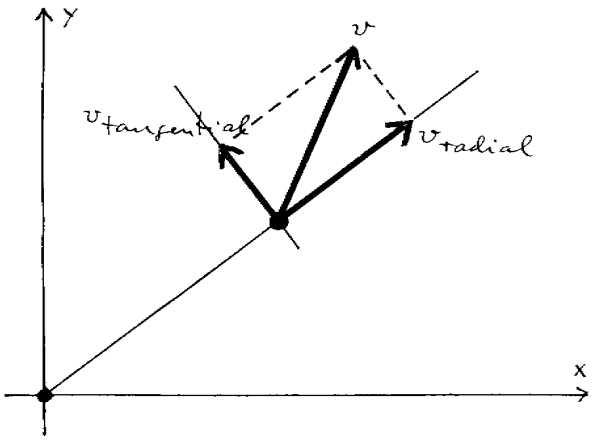
$$\begin{aligned} \frac{\partial}{\partial r} &= \frac{\partial x}{\partial r}\frac{\partial}{\partial x} + \frac{\partial y}{\partial r}\frac{\partial}{\partial y} = \cos\varphi \frac{\partial}{\partial x} + \sin\varphi \frac{\partial}{\partial y} \\ \frac{\partial}{\partial \varphi} &= \frac{\partial x}{\partial \varphi}\frac{\partial}{\partial x} + \frac{\partial y}{\partial \varphi}\frac{\partial}{\partial y} = -r\sin\varphi \frac{\partial}{\partial x} + r\cos\varphi \frac{\partial}{\partial y}, \end{aligned}$$

woraus man durch Invertieren der Matrix dieselbe Umrechnungsformel wie oben erhält (natürlich!):

$$\begin{aligned} \frac{\partial}{\partial x} &= \cos\varphi \frac{\partial}{\partial r} - \tfrac{1}{r}\sin\varphi \frac{\partial}{\partial \varphi} \\ \frac{\partial}{\partial y} &= \sin\varphi \frac{\partial}{\partial r} + \tfrac{1}{r}\cos\varphi \frac{\partial}{\partial \varphi} \end{aligned}$$

Bemerkung Gelegentlich wird ein Vektorfeld v auf $X \subset \mathbb{R}^2 \setminus \{0\}$ in einen radialen und einen tangentialen Anteil zerlegt. Das bedeutet fast dasselbe, wie v in Polarkoordinaten zu schreiben, aber nicht ganz. Denn gemäß den Formeln, die $\frac{\partial}{\partial r}$ und $\frac{\partial}{\partial \varphi}$ durch ihre kartesichen Komponenten auszudrücken, hat der Vektor $\frac{\partial}{\partial r}(r,\varphi)$ die euklidische Länge 1 und $\frac{\partial}{\partial \varphi}(r,\varphi)$ die Länge r, deshalb ist

$$v = v_{\text{radial}} \frac{\partial}{\partial r} + v_{\text{tangential}} \frac{1}{r} \frac{\partial}{\partial \varphi}.$$



Sie werden jetzt verstehen, warum die Definition "ein Vektorfeld ist eine $\mathbb{R}^3$-wertige Funktion von drei Variablen", die man Ihnen anderweitig vorsetzt, ganz unzulänglich ist. Zum Beispiel ist das Geschwindigkeitsfeld eines im Raum strömenden Mediums (zu einem festen Zeitpunkt) zweifellos ein solches Vektorfeld, aber warum erhalte ich nicht auch ein Vektorfeld, wenn ich jedem Raumpunkt den Vektor zuordne, dessen Komponenten die dort herrschenden Werte von Druck, Dichte und Temperatur sind? Weil das drei skalare Felder sind, heißt es. Aber was ist dann der Unterschied zwischen drei Skalaren und einem Vektor? Bei der korrekten Definition 37.12 stellt sich diese Frage gar nicht; ein Vektorfeld auf der n-dimensionalen Mannigfaltigkeit X ist eine Abbildung von X in TX, und n Skalare geben eine Abbildung $X \longrightarrow \mathbb{R}^n$.

Wenn man sich nun als Physiker aber mit der Auffassung eines Vektorfeldes als einer Abbildung $g\colon X \longrightarrow \mathbb{R}^n$ (mit offenem $X \subset \mathbb{R}^n$) einfach wohler fühlt, weil sie so konkret erscheint? Dann soll man die Definition um

die Forderung ergänzen, daß g sich unter einem Kartenwechsel $k \circ h^{-1}$ wie besprochen transformiert, nämlich in

$$y \longmapsto D(k \circ h^{-1})(y)^{-1} \cdot (g \circ k \circ h^{-1})(y)$$

übergeht. Das erklärt den Begriff "Vektorfeld" auf eine Weise, die ästhetisch weniger befriedigt, aber den wesentlichen Punkt implizit erfaßt und praxisnah ist. Und die Verwechslung mit einem Satz von Skalaren ausschließt, denn ein für n Skalare stehendes g transformiert sich unter demselben Kartenwechsel natürlich in

$$y \longmapsto (g \circ k \circ h^{-1})(y).$$

Übungsaufgaben

37.1 Beweisen Sie, daß die Menge

$$Y := \left\{(w,z) \in \mathbb{C}^2 \,\middle|\, w^3 + z^3 = 1\right\}$$

eine 2-dimensionale Untermannigfaltigkeit von $\mathbb{C}^2 = \mathbb{R}^4$ ist, und bestimmen Sie eine Basis des Tangentialraums $T_{(1,0)}Y$.

37.2 Sei

$$X := \left\{(x,y,z) \in \mathbb{R}^3 \,\middle|\, x^2 + y^2 = z^2,\ z \geq 0\right\}$$

die obere Hälfte des Kegels. Beweisen Sie, daß X keine Untermannigfaltigkeit von $\mathbb{R}^3$ ist.

Tip: Für jede ganz in X verlaufende differenzierbare Kurve $\gamma\colon I \longrightarrow \mathbb{R}^n$ mit $\gamma(0) = 0$ verschwindet der Geschwindigkeitsvektor zur Zeit 0.

37.3 Zeigen Sie, daß

$$Y := \left\{(x,y,z) \in \mathbb{R}^3 \,\middle|\, x^2 + y^2 = 1,\ x^2 + 2yz = 0\right\}$$

eine Untermannigfaltigkeit von $\mathbb{R}^3$ ist, und berechnen Sie alle kritischen Punkte der durch $f(x,y,z) = z$ definierten Funktion $f\colon Y \longrightarrow \mathbb{R}$.

37.4 Übertragen Sie den Begriff "lokaler Diffeomorphismus" aus Definition 35.3 und den Satz 35.4 von der lokalen Umkehrung auf den Fall einer Abbildung zwischen Mannigfaltigkeiten.

37.5 Zeigen Sie, daß unter den Voraussetzungen der Aufgabe 36.1 die vom Zeitpunkt 0 aus gemessene Bogenlänge eine Karte für die Mannigfaltigkeit $\gamma(-\delta,\delta)$ definiert, wenn man $\delta > 0$ genügend klein wählt.

37.6 Ist $G \subset GL(n,\mathbb{R})$ eine Lie-Gruppe, so bildet die Exponentialabbildung $\exp\colon \mathrm{Mat}(n\times n,\mathbb{R}) \longrightarrow GL(n,\mathbb{R})$ aus Aufgabe 34.3 die in Beispiel 37.6 erklärte Lie-Algebra T_1G in G selbst ab. Beweisen Sie das in zwei beispielhaften Fällen:

- $G = O(n)$: Ist $x \in \mathrm{Mat}(n\times n,\mathbb{R})$ schiefsymmetrisch, so ist e^x orthogonal.
- $G = SL(n,\mathbb{R})$: Hat $x \in \mathrm{Mat}(n\times n,\mathbb{R})$ die Spur 0, so ist $\det e^x = 1$.

Tip zum zweiten Fall: Betrachten Sie den Weg $\mathbb{R} \ni t \longmapsto e^{tx} \in \mathrm{Mat}(n\times n,\mathbb{R})$.

37.7 Wie lautet die im Beispiel 37.14 mit dem Punkt markierte Zeile in pedantischer Schreibweise?

37.8 In Texten über Thermodynamik stößt man auf Behauptungen wie das Lemma in Abschnitt 1.5 aus K. Huang: *Statistical Mechanics*. Schauen Sie nach, ob man das verstehen kann (der Autor scheint es ja zu erwarten). Jedenfalls besteht die Aufgabe darin, herauszufinden, was gemeint ist, das mathematisch korrekt zu formulieren und zu beweisen. Mehr als auf raffinierte Rechentricks kommt es darauf an, die relevante Geometrie zu verstehen, die übrigens in den einzelnen Behauptungen (a), (b) und (c) jeweils etwas verschieden ist.

37.9 Verifizieren Sie, daß die Zuordnung

$$v\colon (x,y) \longmapsto \begin{pmatrix} -y \\ x \end{pmatrix}$$

ein Vektorfeld auf der Kreislinie $S^1 \subset \mathbb{R}^2$ definiert. Wie schreibt v sich in Polarkoordinaten? Versuchen Sie, mittels v ein stetiges Vektorfeld o auf der Sphäre $S^2 \subset \mathbb{R}^3$ zu konstruieren, das außer an den beiden Polen überall nach Osten weist. Schreiben Sie o auch in Kugelkoordinaten.

37.10 Zeigen Sie, daß die Formel

$$n\colon (x,y,z) \longmapsto \begin{pmatrix} -xz \\ -yz \\ 1-z^2 \end{pmatrix}$$

ein stetiges Vektorfeld n auf S^2 definiert, das überall (mit Ausnahme der Pole) nach Norden zeigt. Schreiben Sie auch n in Kugelkoordinaten. Wie hängt die Darstellung eines beliebigen Vektorfelds v auf S^2 in Kugelkoordinaten mit der Zerlegung von v in eine laterale und eine longitudinale Komponente (parallel zu den Längen- bzw. Breitenkreisen) zusammen?

37.11 Konstruieren Sie einen Diffeomorphismus $f\colon Z \longrightarrow H$ zwischen dem Zylinder $Z := S^1 \times \mathbb{R} \subset \mathbb{R}^3$ und dem Hyperboloid $H := \left\{(x,y,z) \in \mathbb{R}^3 \,\middle|\, x^2+y^2-z^2 = 1\right\}$. Welches Vektorfeld auf H entspricht unter f dem Vektorfeld v auf Z mit

$$v(x,y,z) = -y\,\frac{\partial}{\partial x} + x\,\frac{\partial}{\partial y} + \frac{\partial}{\partial z}\ ?$$

(Die Antwort mag verschieden ausfallen, je nachdem, welches f Sie gewählt haben.)

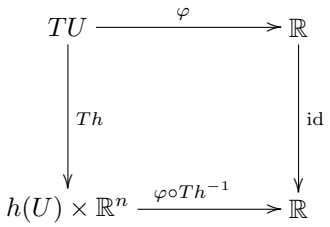
38 Pfaffsche Formen

38.1 Definition Sei X eine n-dimensionale Mannigfaltigkeit. Unter einer Pfaffschen Form auf X versteht man eine Abbildung $\varphi: TX \longrightarrow \mathbb{R}$, derart daß φ auf jedem Tangentialraum linear ist, genauer nämlich für jedes $x \in X$ die Funktion

$$\varphi_x: T_xX \longrightarrow \mathbb{R}; \quad v \mapsto \varphi(x, v)$$

eine Linearform auf T_xX ist. Physiker kennen keine Pfaffschen Formen, nennen sie vielmehr kovariante Vektorfelder.

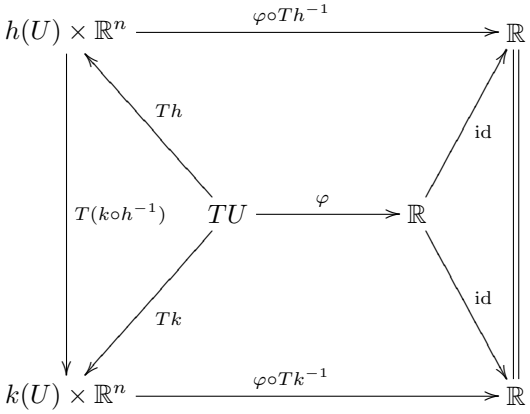
Alternativ und vielleicht einen Hauch anschaulicher: Die Pfaffsche Form φ ordnet jedem $x \in X$ eine Linearform auf dem Tangentialraum T_xX zu, nämlich φ_x. Natürlich kann man jede Pfaffsche Form φ lokal bezüglich Karten beschreiben. Das zu einer Karte (U, h) für X gehörige kommutative Diagramm



fällt sogar einfacher aus als bei (kontravarianten) Vektorfeldern, und $\varphi \circ Th^{-1}$ entspricht der Abbildung

$$h(U) \ni z \mapsto g(z) = \begin{pmatrix} g_1(z) & \dots & g_n(z) \end{pmatrix} \in \mathrm{Mat}(1 \times n, \mathbb{R})$$

mit $g_j(z) = \varphi\big(h^{-1}(z), Dh^{-1}(z) \cdot e_j\big)$ für $j = 1, \dots, n$. Bei groben Hinsehen sieht eine Pfaffsche Form also genau so aus wie ein Vektorfeld, und das erklärt schon mal, warum Physiker beide als Vektorfelder bezeichnen. Daß andererseits Vektoren und Linearformen aus mathematischer Sicht etwas ganz Verschiedenes sind, haben wir in Abschnitt 27 ausführlich besprochen. An diesem Unterschied kommt man auch als Physiker letztlich nicht vorbei, weil sich Pfaffsche Formen bei Kartenwechsel anders verhalten als Vektorfelder. Nehmen wir zu (U, h) wieder eine zweite Karte (U, k) hinzu. Es entsteht das folgende kommutative Diagramm:



Die bezüglich der Karte k durch

$$k(U) \ni z \longmapsto g(z) \in \mathrm{Mat}(1 \times n, \mathbb{R})$$

gegebene Pfaffsche Form schreibt sich in der Karte h also

$$h(U) \ni y \longmapsto (g \circ k \circ h^{-1})(y) \cdot D(k \circ h^{-1})(y) \in \mathrm{Mat}(1 \times n, \mathbb{R}).$$

38.2 Beispiel Die in kartesischen Koordinaten $k = (x, y)$ durch die Zeile $(x, y) \mapsto g(x, y) \in \mathrm{Mat}(1 \times 2, \mathbb{R})$ gegebene Pfaffsche Form auf $\mathbb{R}^2$ hat die Polarkoordinatendarstellung

$$\begin{pmatrix} r \\ \varphi \end{pmatrix} \longmapsto g(r\cos\varphi, r\sin\varphi) \cdot \begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix} \in \mathrm{Mat}(1 \times 2, \mathbb{R}).$$

Die Transformationsformel unterscheidet sich von der für die Vektorfelder insbesondere dadurch, daß hier nur der Kartenwechsel $k \circ h^{-1}$ und sein Differential eingehen, während bei jener der Kartenwechsel direkt und das *Inverse* seines Differentials eingehen (oder umgekehrt, wenn man die Rollen von h und k vertauscht). Wahrscheinlich ist das der Grund für die Attribute "ko-" bzw. "kontravariant".

Natürliche Beispiele von Pfaffschen Formen erhält man aus differenzierbaren Funktionen:

38.3 Definition Sei X eine Mannigfaltigkeit und $f\colon X \longrightarrow \mathbb{R}$ differenzierbar. Dann ist die zweite Komponente der Abbildung

$$Tf\colon TX \longrightarrow T\mathbb{R} = \mathbb{R} \times \mathbb{R}, \text{ also } (x, v) \longmapsto T_x f(v)$$

offenbar eine Pfaffsche Form auf X; sie wird mit df bezeichnet und ebenso wie Tf das Differential von f genannt. Die so entstehenden Pfaffschen Formen df heißen exakt, und man nennt f dann eine Stammfunktion von df.

Bemerkung Die damit eingeführte Mehrfachbedeutung des Wortes ist einigermaßen vertretbar, weil df immerhin den interessantesten Teil der Information aus Tf enthält; außerdem stehen ja zur Unterscheidung die Symbole zur Verfügung.

Schreibt man die Funktion $f\colon X \longrightarrow \mathbb{R}$ in einer Karte (U, h) für X, so wird df in dieser Karte zu

$$h(U) \ni z \mapsto D(f \circ h^{-1})(z) = \begin{pmatrix} \frac{\partial f}{\partial h_1}(z) & \dots & \frac{\partial f}{\partial h_n}(z) \end{pmatrix} \in \mathrm{Mat}(1 \times n, \mathbb{R}).$$

Weil speziell auch die Komponenten $h_1, \dots, h_n$ von h differenzierbare Funktionen auf $U \subset X$ sind, kann man von deren Differentialen $dh_1, \dots, dh_n$ reden. In der Karte h entsprechen sie den konstanten Abbildungen, deren Werte die Standardzeilen $e_1^{\mathrm{t}}, \dots, e_n^{\mathrm{t}}$ sind. Auf der Hand liegt deshalb die

38.3½ Notiz Ist (U, h) eine Karte für die n-dimensionale Mannigfaltigkeit X, so sind für jedes $x \in X$ die n-tupel

$$\left(\frac{\partial}{\partial h_1}(x), \dots, \frac{\partial}{\partial h_n}(x)\right) \quad \text{und} \quad \big((dh_1)_x, \cdots, (dh_n)_x\big)$$

zueinander duale Basen von $T_x X$ und $(T_x X)\check{}$.

Eine beliebige, etwa bezüglich h durch

$$z \longmapsto g(z) = \begin{pmatrix} g_1(z) & \dots & g_n(z) \end{pmatrix}$$

gegebene Pfaffsche Form φ auf U kann man also auch in der Form

$$\varphi = (g_1 \circ h)\, dh_1 + \cdots + (g_n \circ h)\, dh_n$$

notieren. Diese beliebte und praktische Schreibweise eröffnet auch eine zweite Möglichkeit, die Umrechnung auf eine andere Karte k zu bewerkstelligen; denn nach der Kettenregel ist

$$dk_i = d\big((k_i \circ h^{-1}) \circ h\big) = \big(D(k_i \circ h^{-1}) \circ h\big) \cdot \begin{pmatrix} dh_1 \\ \vdots \\ dh_n \end{pmatrix} = \sum_{j=1}^{n} \frac{\partial k_i}{\partial h_j}\, dh_j.$$

Der im Beispiel 38.2 betrachtete Wechsel zwischen kartesischen und Polarkoordinaten ergibt sich daraus in der alternativen Form

$$\begin{aligned} dx &= \frac{\partial x}{\partial r}\,dr + \frac{\partial x}{\partial \varphi}\,d\varphi &&= \cos\varphi\,dr - r\sin\varphi\,d\varphi \\ dy &= \frac{\partial y}{\partial r}\,dr + \frac{\partial y}{\partial \varphi}\,d\varphi &&= \sin\varphi\,dr + r\cos\varphi\,d\varphi \end{aligned}$$

oder

$$\begin{aligned} dr &= \frac{x}{\sqrt{x^2+y^2}}\,dx + \frac{y}{\sqrt{x^2+y^2}}\,dy \\ d\varphi &= -\frac{y}{x^2+y^2}\,dx + \frac{x}{x^2+y^2}\,dy. \end{aligned}$$

38.3$\frac{3}{4}$ Beispiele (1) Die Funktion $f\colon\mathbb{R}^3\to\mathbb{R}$ sei durch $f(x,y,z)=x+y^2+e^z$ gegeben. Ihr Differential ist die Pfaffsche Form

$$df = \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy + \frac{\partial f}{\partial z}\,dz = dx + 2y\,dy + e^z\,dz.$$

Was passiert, wenn man f und df auf die Untermannigfaltigkeit $S^2\subset\mathbb{R}^3$ einschränkt? Nun, die Formeln $f(x,y,z)=x+y^2+e^z$ und $df=dx+2y\,dy+e^z\,dz$ bleiben sinnvoll und gültig, denn die drei kartesischen Koordinaten definieren definieren auch drei Funktionen auf der Sphäre. Daß sie dort nicht mehr die Komponentenfunktionen einer Karte sind, tut dem keinen Abbruch.

Lokal ist es aber auf Wunsch möglich, eine der drei Funktionen durch die anderen auszudrücken, etwa

$$z=\sqrt{1-x^2-y^2}\quad\text{auf der durch } z>0 \text{ definierten Nordhemisphäre.}$$

Wir können das gleich in f substituieren und $f(x,y)=x+y^2+e^{\sqrt{1-x^2-y^2}}$ schreiben; Differenzieren liefert dann

$$df = \Big(1 - x\frac{e^{\sqrt{1-x^2-y^2}}}{\sqrt{1-x^2-y^2}}\Big)dx\Big(2y - y\frac{e^{\sqrt{1-x^2-y^2}}}{\sqrt{1-x^2-y^2}}\Big)\,dy.$$

Ebensogut können wir aber aus der Gleichung $z=\sqrt{1-x^2-y^2}$ die auf S^2 gültige Darstellung

$$dz = \frac{-x\,dx - y\,dy}{e^{\sqrt{1-x^2-y^2}}}$$

ablesen und das in df substituieren, natürlich mit demselben Ergebnis.

(2) Wir können auf einer geeigneten offenen Teilmeng von $\mathbb{R}^3$ auch in Polarkoordinaten arbeiten, haben in der Physiker-Notation also

$$f(r,\theta,\varphi) = r\sin\theta\cos\varphi + r^2(\sin\theta)^2(\sin\varphi)^2 + e^{r\cos\theta}$$

und damit

$$\begin{aligned} df = {} & \big(\sin\theta\cos\varphi + 2r(\sin\theta)^2(\sin\varphi)^2 + e^{r\cos\theta}\cos\theta\big)\cdot dr \\ & + \big(r\cos\theta\cos\varphi + 2r^2\sin\theta\cos\theta(\sin\varphi)^2 - re^{r\cos\theta}\sin\theta\big)\cdot d\theta \\ & + \big(-r\sin\theta\sin\varphi + 2r^2(\sin\theta)^2\sin\varphi\cos\varphi\big)\cdot d\varphi\,. \end{aligned}$$

Auch auf der Sphäre S^2 ist das eine gültige Identität, aber weil dort $r=1$ konstant und deshalb $dr=0$ ist, wird der dr-Term entbehrlich.

In physikalisch relevanten Situationen tragen die vorkommenden Tangentialräume in aller Regel eine metrische Struktur (im einfachsten Fall das Standardskalarprodukt von $\mathbb{R}^n$): das heißt, sie sind euklidische Vektorräume. Weil dadurch gemäß Satz 27.11 ein Isomorphismus des Tangentialraums mit seinem Dualraum gegeben ist, werden Vektorfelder und Pfaffsche Formen zu im Prinzip gegeneinander austauschbaren Objekten. In der Physik spricht man denn auch gern durchweg von Vektorfeldern, die man (bei gegebener Karte) nach Wahl durch seine ko- oder kontravarianten Komponenten ausdrücken kann. Es ist trotzdem nichts damit

gewonnen (und manches verloren), wenn man Vektorfelder und Pfaffsche Formen generell miteinander identifiziert. Denn bei einem Kartenwechsel sind ja auch die Skalarprodukte auf den einzelnen Tangentialräumen zu transformieren, so daß der Isomorphismus zwischen diesen und ihren Dualräumen nicht mehr durch bloßes Transponieren beschrieben wird: die Formeln werden dadurch nur komplizierter.

Bei vielen populären Vektorfeldern der Physik erscheint es natürlicher, sie als kovariante Felder, also Pfaffsche Formen zu schreiben. Das trifft insbesondere auf Felder zu, die ein *Potential* besitzen. Etwa tritt dann beim elektrischen Feld E der Elektrostatik an die Stelle der Gleichung $E = -\operatorname{grad}\varphi$ die mathematisch einfachere Gleichung $E = -d\varphi$, die keinen Bezug auf die euklidische Struktur nimmt und sich deshalb bei Kartenwechsel (zum Beispiel auf Polarkoordinaten) auch korrekt transformiert.

Stichwort Potential: Wie Sie wissen, ist die Frage von großem Interesse, ob ein Vektorfeld ein Potential besitzt, in mathematischer Sprache also, ob eine gegebene Pfaffsche Form exakt ist und damit eine Stammfunktion hat (das Minuszeichen, das in der Physik Tradition hat, spielt dabei natürlich keine Rolle). Wir werden sie bald systematisch in einem allgemeineren Rahmen behandeln, wollen sie hier aber schon ein bißchen beschnuppern. Sei $X \subset \mathbb{R}^n$ offen, und sei eine differenzierbare Pfaffsche Form in kartesischen Koordinaten

$$\varphi = g_1\,dx_1 + \cdots + g_n\,dx_n$$

gegeben; $g_1, \ldots, g_n$ sind also differenzierbare Funktionen auf X. Exakt ist φ genau dann, wenn es eine differenzierbare Funktion $f\colon X \longrightarrow \mathbb{R}$ mit $df = \varphi$, also mit

$$\frac{\partial f}{\partial x_j} = g_j \quad \text{für } j = 1, \ldots, n$$

gibt. Wenn das der Fall ist, folgt

$$\frac{\partial g_j}{\partial x_i} = \frac{\partial^2 f}{\partial x_i x_j} \quad \text{und} \quad \frac{\partial g_i}{\partial x_j} = \frac{\partial^2 f}{\partial x_j x_i}$$

für alle i, j. Nun erweist sich die Reihenfolge der Differentiation in den höheren partiellen Ableitungen als unerheblich:

38.4 Satz Sei $X \subset \mathbb{R}^n$ offen und $f\colon X \longrightarrow \mathbb{R}^p$ eine C^k-Abbildung. Für je k Indizes $j_1, \ldots, j_k \in \{1, \ldots, n\}$ und jede Permutation $\sigma \in \mathrm{Sym}_k$ gilt dann

$$D_{j_1} D_{j_2} \cdots D_{j_k} f = D_{j_{\sigma 1}} D_{j_{\sigma 2}} \cdots D_{j_{\sigma k}} f.$$

Beweis Es genügt offenbar, als wesentlichen Fall $n=2,\ p=1,\ k=2$, also eine C^2-Funktion $\mathbb{R}^2 \supset X \xrightarrow{f} \mathbb{R}$ zu betrachten und

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$$

zu beweisen. Dazu bilden wir an der Stelle $(a, b) \in X$ den "doppelten Differenzenquotienten"

$$F(x, y) = \frac{f(x, y) - f(x, b) - f(a, y) + f(a, b)}{(x-a)\,(y-b)},$$

der für alle (x, y) aus einer kleinen Kreisscheibe um (a, b) definiert ist, für die $x \neq a$ und $y \neq b$ gilt. Wir halten ein solches (x, y) zunächst fest. Der Mittelwertsatz 14.1, angewendet auf die Funktion $u \longmapsto f(u, y) - f(u, b)$, macht aus dem doppelten den gewöhnlichen Differenzenquotienten

$$F(x, y) = \frac{D_1 f(s, y) - D_1 f(s, b)}{y - b}$$

für ein geeignetes s zwischen a und x. Jetzt wenden wir den Mittelwertsatz auf die Funktion $v \longmapsto D_1 f(s, v)$ an und erhalten

$$F(x, y) = D_2 D_1 f(s, t)$$

mit einem passenden t zwischen b und y. Weil nun $(x,y) \to (a,b)$ auch $(s,t) \to (a,b)$ erzwingt, folgt aus der Stetigkeit der zweiten Ableitungen

$$\lim_{(x,y)\to(a,b)} F(x,y) = D_2 D_1 f(a,b).$$

Da F in den beiden Variablen symmetrisch gebaut ist, gilt ebenso

$$\lim_{(x,y)\to(a,b)} F(x,y) = D_1 D_2 f(a,b),$$

und der Satz ist bewiesen.

Aufgrund unserer obigen Überlegungen ergibt sich sofort die

38.5 Folgerung Die differenzierbare Pfaffsche Form $\varphi = g_1\,dx_1 + \cdots + g_n\,dx_n$ kann nur dann exakt sein, wenn

$$\frac{\partial g_j}{\partial x_i} = \frac{\partial g_i}{\partial x_j} \quad \text{für alle } i \neq j$$

(als Identität von Funktionen) gilt.

38.6 Beispiel Die auf $\mathbb{R}^2\backslash\{0\}$ erklärte Pfaffsche Form

$$r\,d\varphi = -\frac{y}{\sqrt{x^2+y^2}}\,dx + \frac{x}{\sqrt{x^2+y^2}}\,dy$$

ordnet jedem Vektor seinen tangentialen Koeffizienten zu. Sie kann nicht exakt sein, denn:

$$\frac{\partial}{\partial x}\frac{x}{\sqrt{x^2+y^2}} = \frac{\sqrt{x^2+y^2} - \frac{x\cdot 2x}{2\sqrt{x^2+y^2}}}{x^2+y^2} = \frac{y^2}{(x^2+y^2)^{3/2}} \neq -\frac{x^2}{(x^2+y^2)^{3/2}} = \frac{\partial}{\partial y}\left(-\frac{y}{\sqrt{x^2+y^2}}\right)$$

Mußte man sich wirklich die Mühe machen, das auszurechnen? Mußte man nicht, denn weil die Konzepte "Pfaffsche Form" und "Exaktheit" mit der Wahl einer speziellen Karte gar nichts zu tun haben, darf man die Rechnung auch in Polarkoordinaten durchführen, wo sie trivial ist:

$$\frac{\partial}{\partial r}\,r = 1 \neq 0 = \frac{\partial}{\partial \varphi}\,0$$

Die Kombination von Pfaffschen Formen und Kurven führt zu einer neuen Art von eindimensionalen Integralen.

38.7 Definition Sei φ eine stetige Pfaffsche Form auf der Mannigfaltigkeit X, und sei $\gamma\colon [a,b] \longrightarrow X$ eine (C^1-)differenzierbare Kurve. Dann heißt die Zahl

$$\int_\gamma \varphi := \int_a^b \varphi\big(\gamma(t), \dot\gamma(t)\big)\,dt$$

das Kurvenintegral von φ längs γ.

Bemerkungen Weil $\dot\gamma(t)$ ein Tangentialvektor von X an der Stelle $\gamma(t)$ ist, gibt die Funktion unter dem Integralzeichen Sinn; sie kommt dadurch zustande, daß die Geschwindigkeitsvektoren von γ mittels φ (linear) bewertet werden. Ein typischer in der Physik auftretender Fall ist der, daß die Form φ ein (kovariantes) Kraftfeld, und γ die Bahn eines Teilchens ist; das Integral ist dann die vom Feld am Teilchen geleistete Arbeit. — Natürlich kann man als φ auch nichtstetige Pfaffsche Formen sowie statt $[a,b]$ nichtkompakte Intervalle in Betracht ziehen, muß sich dann aber um die Existenz des Integrals im Sinne der Integrationstheorie Gedanken machen. — Ist die Kurve γ *geschlossen*, d.h. $\gamma(a) = \gamma(b)$, so schreiben vor allem Physiker statt dem gewöhnlichen Integralzeichen gern das lustige (aber eigentlich überflüssige) $\oint$.

Wenn man als Kurve γ die identische Funktion $[a,b] \to [a,b]$ wählt und die zu integrierende Pfaffsche Form $\varphi(t) = f(t)\,dt$ schreibt, dann reduziert sich das Kurvenintegral wegen $dt \cdot \dot{\varphi} = 1$ auf das "alte" Integral einer Funktion über ein Intervall

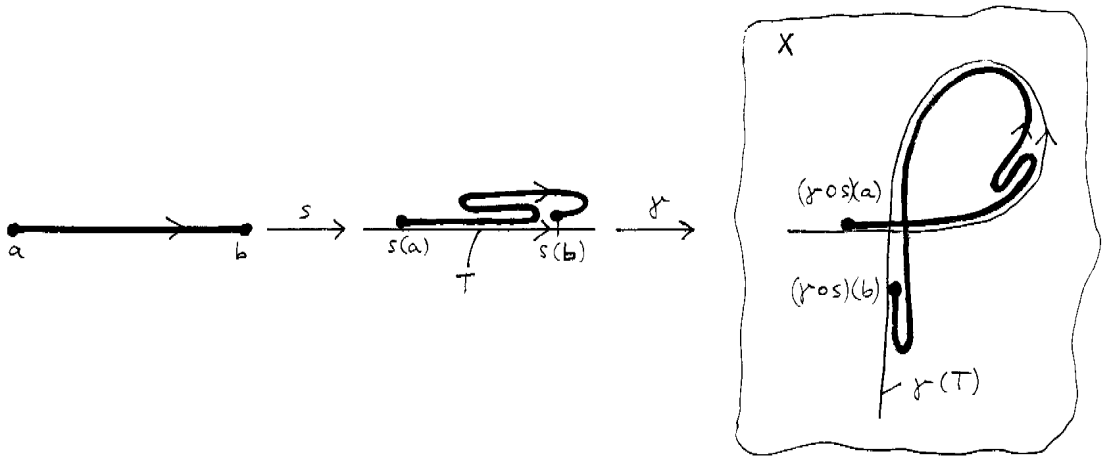
$$\int_\gamma \varphi = \int_a^b f(t)\,dt$$

und gibt diesem damit eine neue Interpretation, in der auch dt allein und nicht nur der gesamte Integralausdruck eine Bedeutung hat.

Inwieweit hängt das Integral von der genauen Wahl des Intergrationsweges ab? Nun, die wohl mildeste Manipulation eines Weges $\gamma\colon [c,d] \longrightarrow X$ besteht darin, ihn "umzuparametrisieren", indem man einen monoton wachsenden Diffeomorphismus $s\colon [a,b] \longrightarrow [c,d]$ vorschaltet. Dieser Vorgang ändert das Kurvenintegral nie; schon unter etwas allgemeineren Voraussetzungen gilt nämlich:

38.8 Lemma Sei $T \subset \mathbb{R}$ ein Intervall, X eine Mannigfaltigkeit, $\gamma\colon T \longrightarrow X \subset \mathbb{R}^n$ eine C^1-Kurve und $s\colon [a,b] \longrightarrow T$ eine C^1-Funktion mit $s(a) \le s(b)$. Für jede stetige Pfaffsche Form φ auf X gilt dann

$$\int_{\gamma\circ s} \varphi = \int_{\gamma|[s(a),s(b)]} \varphi.$$



Beweis Das folgt aus der Substitutionsregel 31.13(b):

$$\begin{aligned}
\int_{\gamma\circ s} \varphi &= \int_a^b \varphi\big((\gamma\circ s)(t), (\gamma\circ s)^{\cdot}(t)\big)\,dt \\
&= \int_a^b \varphi\big((\gamma\circ s)(t), (\dot\gamma\circ s)(t)\cdot s'(t)\big)\,dt \\
&= \int_a^b \varphi\big((\gamma\circ s)(t), (\dot\gamma\circ s)(t)\big)\cdot s'(t)\,dt \\
&= \int_{s(a)}^{s(b)} \varphi\big((\gamma(s), \dot\gamma(s)\big)\,ds \\
&= \int_{\gamma|[s(a),s(b)]} \varphi.
\end{aligned}$$

Das Lemma erlaubt es, bei der Angabe des Weges γ schon mal etwas großzügiger zu sein. Oft wird nur die Bildmenge von γ hingeschrieben und dabei eine mehr oder weniger naheliegende Parametrisierung unterstellt. Dann muß allerdings auch deren Richtung erkennbar sein, denn wie man etwa aus dem vorstehenden Beweis

sofort abliest, ändert Umparametrisieren mit einem monoton fallenden Diffeomorphismus das Vorzeichen des Integrals.

Auch Kurvenintegrale werden Sie in physikalischen Darstellungen am häufigsten in koordinatengebundener Notation antreffen, die den Unterschied zwischen Linearformen und Vektoren verwischt. Dann würde etwa $\int_\gamma \vec{v}\cdot d\vec{s}$ dasselbe wie unser $\int_\gamma \varphi$ bedeuten, wobei das Vektorfeld $\vec{v}$ und die Pfaffsche Form φ sich unter Transponieren entsprechen, deshalb das Skalarprodukt $\cdot$ an die Stelle der Matrizenmultiplikation zwischen $\varphi_{\gamma(t)}$ und $\dot\gamma(t)$ treten muß. Wie an früherer Stelle erwähnt, wird der Buchstabe s bei Kurven bevorzugt für die Bogenlänge verwendet, und das von den Physikern "Linienelement" genannte $d\vec{s}$ symbolisiert die vektorielle Version von $ds = |\dot\gamma(t)|dt$ (siehe die Aufgabe 34.7).

In speziellen Fällen geht die Wegunabhängigkeit des Kurvenintegrals viel weiter:

38.9 Lemma Sei $f\colon X \longrightarrow \mathbb{R}$ differenzierbar, und $\gamma\colon [a,b] \longrightarrow X$ ein differenzierbarer Weg. Dann ist

$$\int_\gamma df = f\big(\gamma(b)\big) - f\big(\gamma(a)\big).$$

Beweis Man braucht bloß

$$\int_\gamma df = \int_a^b df\big(\gamma(t),\dot\gamma(t)\big)dt = \int_a^b T_{\gamma(t)}f\big(\dot\gamma(t)\big)dt = \int_a^b (f\circ\gamma)'(t)dt = \Big[(f\circ\gamma)(t)\Big]_{t=a}^b = f\big(\gamma(b)\big) - f\big(\gamma(a)\big)$$

nach dem "Hauptsatz" auszuwerten (Satz 31.12 aus der Analysis einer einzelnen Variablen).

38.10 Folgerung Ist φ eine exakte Pfaffsche Form, so hängt $\int_\gamma \varphi$ nur vom Anfangs- und vom Endpunkt von γ ab, insbesondere ist bei geschlossenem Integrationsweg $\oint \varphi = 0$.

Bemerkung Es ist nicht schwer zu sehen, daß auch die Umkehrung gilt: Hängt das Kurvenintegral $\int_\gamma \varphi$ für alle differenzierbaren Wege in X nur von deren Anfangs- und Endpunkt ab, so ist φ exakt.

38.11 Beispiel Die Pfaffsche Form

$$\psi := -\frac{y}{x^2+y^2}\,dx + \frac{x}{x^2+y^2}\,dy$$

auf $\mathbb{R}^2\setminus\{0\}$ ist nicht exakt. Integration von ψ über den geschlossenen Weg

$$[0,2\pi] \ni t \longmapsto \begin{pmatrix}\cos t\\ \sin t\end{pmatrix} \in \mathbb{R}^2\setminus\{0\}$$

liefert nämlich

$$\oint \psi = \int_0^{2\pi} \begin{pmatrix}-\sin t & \cos t\end{pmatrix}\begin{pmatrix}-\sin t\\ \cos t\end{pmatrix} dt = \int_0^{2\pi} dt = 2\pi.$$

Durch Anwenden der Folgerung 38.5 hätten wir die Nichtexaktheit von ψ nicht entdecken können, denn wenn wir diesmal gleich in Polarkoordinaten rechnen, ist $\psi = \frac{1}{r}(r\,d\varphi) = d\varphi$ und ... Nanu? Geht daraus denn nicht hervor, daß ψ doch exakt ist, im Widerspruch zu dem, was wir gerade gesehen haben? Nein, es ist schon alles in Ordnung, denn damit die Koordinatenfunktion φ überhaupt definiert ist, muß man die Polarkoordinatenabbildung Φ genügend eingeschränkt haben, und es ist nicht möglich, das so zu machen, daß das Bild der Einschränkung ganz $\mathbb{R}^2\setminus\{0\}$ bleibt. Es ist also zwar $d\varphi$, nicht aber φ selbst global definiert. Die Nichtexaktheit von $\psi = d\varphi$ ist damit von ganz anderer Art ist als die der Pfaffschen Form $r\,d\varphi$ aus Beispiel 38.6: Während letztere auch bei Einschränkung auf beliebige (nicht-leere) offene Teilmengen von $\mathbb{R}^2\setminus\{0\}$ nicht-exakt bleibt, ist ψ immerhin in dem Sinne *lokal* exakt, daß es um jeden Punkt von $\mathbb{R}^2\setminus\{0\}$ eine offene Teilmenge gibt, auf der ψ exakt ist. Die Nichtexaktheit von ψ muß ihren Grund also in der Geometrie des Definitionsbereiches $\mathbb{R}^2\setminus\{0\}$ haben. Wir kommen auf diesen Punkt später zurück. Wenn Sie übrigens Aufgabe 35.1 noch einmal studieren, werden Sie sehen, daß das Phänomen dort implizit vorweggenommen ist.

Bemerkung Nachdem wir das verstanden haben, können wir als einen harmlosen Mangel der Notation abtun, daß im vorstehenden Beispiel die nicht-exakte Form $d\varphi$ den Anschein einer exakten Form hat. Ganz anders verhält sich das bei dem üblichen Umgang mit dem d in der Thermodynamik. Physiker lieben es ja, mit df etc. eine "infinitesimale" Änderung von f zu bezeichnen, d.h. eine, die zwar nicht null ist, aber so klein, daß f im Bereich dieser Änderung näherungsweise als affin-linear angesehen werden kann. Daß man das Differential einer Funktion f ebenso bezeichnet, ist kein Zufall, vielmehr ist es geradezu der Zweck der Pfaffschen Formen, die Argumentation mit den infinitesimalen Größen zu präzisieren, indem man diese durch "richtige" Größen ersetzt: möglich gemacht wird das dadurch, daß Pfaffsche Formen teilweise lineare, nämlich in einer der beiden Variablen lineare Objekte sind. In der Thermodynamik wird das d nun üblicherweise wahllos durcheinander in beiden Bedeutungen verwendet: einerseits werden alle Pfaffschen Formen, ob exakt oder nicht, als dQ, dW, dU etc. geschrieben, andererseits spielt gerade die Frage eine physikalisch wichtige Rolle, ob diese Formen exakt sind, ob es also ein Q mit $dQ = dQ$ gibt (sic). Es bleibt einem nur übrig, jedes d zunächst anhand des Zusammenhangs daraufhin abklopfen, ob es im mathematischen Sinne ernst gemeint ist.

Die Pfaffschen Formen sind Spezialfälle einer sehr viel größeren Klasse von Objekten, den sogenannten Differentialformen. Zu ihrem Verständnis müssen wir noch etwas aus der linearen Algebra lernen; dem dient der folgende Abschnitt.

Übungsaufgaben

38.1 Das "Physiker-Vektorfeld" g habe in kartesischen Koordinaten x, y, z die konstanten Komponenten $(1, 2, 3)$, und bezüglich der Koordinaten

$$h_1 = x,\ h_2 = x+y,\ h_3 = x+y+z$$

die Komponenten $(1, 3, 6)$. Entscheiden Sie, ob es sich bei g um ein (kontravariantes) Vektorfeld oder eine Pfaffsche Form handelt; prüfen Sie auch, ob eventuell beides (oder keines) zutreffen kann.

38.2 Begründen Sie, warum die Van-der-Waals-Zustandsfläche

$$X := \left\{ (P,V,T) \in (0,\infty)^3 \;\middle|\; \left(V-\frac{1}{3}\right)\left(P+\frac{3}{V^2}\right) = \frac{8}{3}T \right\}$$

eine 2-dimensionale Untermannigfaltigkeit von $\mathbb{R}^3$ ist, und warum die Projektionen auf V und T eine bei dem Punkt $(1, 1, 1) \in X$ definierte Karte (V, T) für X bilden. Berechnen Sie die Darstellung der Pfaffschen Form dP in dieser Karte.

38.3 Ein Teilchen bewege sich auf der Ellipse

$$E := \left\{ (x,y,z) \in \mathbb{R}^3 \;\middle|\; \frac{x^2}{16} + \frac{y^2}{9} = 1,\ z = 0 \right\} \subset \mathbb{R}^3$$

in dem kovarianten Kraftfeld

$$\varphi = (3x-4y+2z)\,dx + (4x+2y-3z^2)\,dy + (2xz-4y^2+z^3)\,dz.$$

Welche Arbeit leistet das Feld bei einem einmaligen Umlauf des Teilchens im mathematisch positiven Sinn? Besitzt das Feld ein Potential?

39 Alternierende Multilinearformen

In diesem Abschnitt bezeichnet V einen n-dimensionalen reellen Vektorraum. Für jedes $k \in \mathbb{N}$ ist mit V^k das kartesische Produkt

$$V^k = V \times \cdots \times V$$

mit k gleichen Faktoren gemeint.

39.1 Definition Eine Multilinearform vom Grad k auf V ist eine Funktion

$$\varphi\colon V^k \longrightarrow \mathbb{R},$$

die multilinear, d.h. in jeder der k Variablen linear ist.

Das brauche ich Ihnen nicht mehr groß zu erklären, denn Sie kennen als Beispiele schon die Skalarprodukte ($k = 2$) und die Determinante, die man offenbar als Multilinearform

$$\det\colon (\mathbb{R}^n)^n \longrightarrow \mathbb{R}$$

vom Grad n auffassen kann. Die Definition ist übrigens auch für $k = 0$ wörtlich zu nehmen: mangels Variablen wird von der Funktion $\varphi\colon \{0\} \longrightarrow \mathbb{R}$ eben nichts verlangt.

Es ist klar, daß die Multilinearformen auf V von festem Grad k unter punktweiser Addition und skalarer Multiplikation einen Vektorraum bilden, den wir mit $\mathrm{Mult}^k V$ bezeichnen wollen.

39.2 Definition Eine Form $\varphi \in \mathrm{Mult}^k V$ heißt alternierend, wenn

$$\varphi(v_1, \ldots, v_k) = 0$$

immer dann gilt, wenn zwei der Vektoren v_i übereinstimmen.

Auch diesen Begriff kennen Sie schon; er ist ja eine der definierenden Eigenschaften der Determinante. Natürlich bilden die alternierenden Multilinearformen vom Grad k einen Untervektorraum

$$\mathrm{Alt}^k V \subset \mathrm{Mult}^k V.$$

Einige Eigenschaften dieser Formen:

39.3 Satz Für jede Form $\varphi \in \mathrm{Alt}^k V$ gilt:

(a) $\varphi(v_1, \ldots, v_k) = 0$ immer dann, wenn das k-tupel $(v_1, \ldots, v_k)$ linear abhängig ist; insbesondere ist $\mathrm{Alt}^k V = \{0\}$ für $k > n$.

(b) Ist $f\colon \mathrm{Lin}(v_1, \ldots, v_k) \longrightarrow \mathrm{Lin}(v_1, \ldots, v_k)$ linear, so gilt

$$\varphi\big(f(v_1), \ldots, f(v_k)\big) = \det f \cdot \varphi(v_1, \ldots, v_k).$$

(c) Für jede Permutation $\sigma \in \mathrm{Sym}_k$ und beliebige $v_1, \ldots, v_k \in V$ gilt

$$\varphi(v_{\sigma 1}, \ldots, v_{\sigma k}) = (-1)^\sigma \varphi(v_1, \ldots, v_k).$$

Beweis Zum Beweis von (a) schreibt man einen der Vektoren, etwa v_j, als Linearkombination der anderen und verwendet die Multilinearität:

$$\varphi(v_1,\dots,v_k)=\varphi\Big(v_1,\dots,\sum_{i\neq j}\lambda_i v_i,\dots,v_k\Big)=\sum_{i\neq j}\varphi(v_1,\dots,v_i,\dots,v_k);$$

weil v_i jetzt in jedem Summanden an der i-ten und der j-ten Stelle vorkommt, verschwindet alles.

Nun zu (b): Wenn das k-tupel $(v_1,\dots,v_k)$ linear abhängig ist, so ist $\big(f(v_1),\dots,f(v_k)\big)$ erst recht linear abhängig und die behauptete Identität wahr, weil nach (a) beide Seiten verschwinden. Wir dürfen ab jetzt also $(v_1,\dots,v_k)$ als linear unabhängig voraussetzen. Die Endomorphismen f von $\mathrm{Lin}(v_1,\dots,v_k)$ entsprechen dann in der bekannten Weise, bezüglich der Basis $(v_1,\dots,v_k)$ nämlich, den Matrizen in $\mathrm{Mat}(k\times k,\mathbb{R})$. Wenn wir den zur Matrix a gehörigen Endomorphismus hier mit f_a bezeichnen, ist die Funktion

$$\mathrm{Mat}(k\times k,\mathbb{R})\ni a\longmapsto\varphi\big(f_a(v_1),\dots,f_a(v_k)\big)\in\mathbb{R}$$

eine alternierende Multilinearform in den Spalten von a: gerade das, was wir bei der Einführung der Determinante ad hoc als eine Prädeterminante bezeichnet hatten. Nach dem damaligen Eindeutigkeitssatz (Satz 22.1) muß es sich um ein Vielfaches der Determinantenfunktion selbst handeln; es gibt also eine von a unabhängige Zahl $\lambda\in\mathbb{R}$, so daß

$$\varphi\big(f_a(v_1),\dots,f_a(v_k)\big)=\lambda\det a\quad\text{für alle } a\in\mathrm{Mat}(k\times k,\mathbb{R})$$

ist. Die Wahl $a=1$, also $f_a=\mathrm{id}$ ergibt $\lambda=\varphi(v_1,\dots,v_k)$, und damit folgt

$$\varphi\big(f_a(v_1),\dots,f_a(v_k)\big)=\det a\cdot\varphi(v_1,\dots,v_k)=\det f_a\cdot\varphi(v_1,\dots,v_k)$$

wie behauptet.

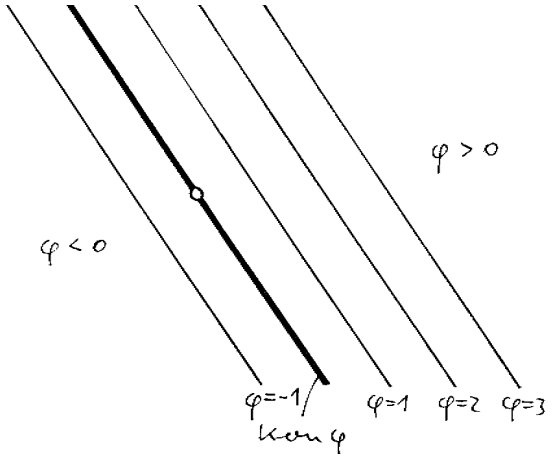
Um (c) zu beweisen, wendet man (b) mit der durch $f(v_i)=v_{\sigma i}$ gemäß Satz 19.6 definierten linearen Abbildung $f\colon\mathrm{Lin}(v_1,\dots,v_k)\longrightarrow\mathrm{Lin}(v_1,\dots,v_k)$ an: Die Matrix von f ist dann die zu σ gehörige Permutationsmatrix, also ergibt sich

$$\varphi(v_{\sigma 1},\dots,v_{\sigma k})=\det f\cdot\varphi(v_1,\dots,v_k)=(-1)^\sigma\varphi(v_1,\dots,v_k)$$

wie behauptet.

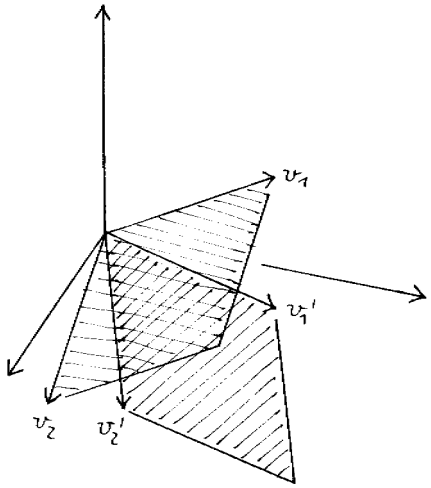
Bemerkung Man sieht sofort, daß umgekehrt jede Multilinearform, die (c) erfüllt, alternierend sein muß: beim Vertauschen zweier gleicher Vektoren ändert sich der Wert einer solchen Form einerseits um das Vorzeichen, andererseits überhaupt nicht. Auch wenn die Eigenschaft (c) komplizierter als die ursprüngliche Definition ist, wird sie sich als besonders günstige Charakterisierung der alternierenden Formen erweisen.

Wir können uns jetzt auch ein anschauliches Bild von den alternierenden Multilinearformen machen. Zunächst ist eine Form vom Grad eins einfach eine Linearform auf V, also eine Funktion, die die Vektoren von V in linearer Weise bewertet.

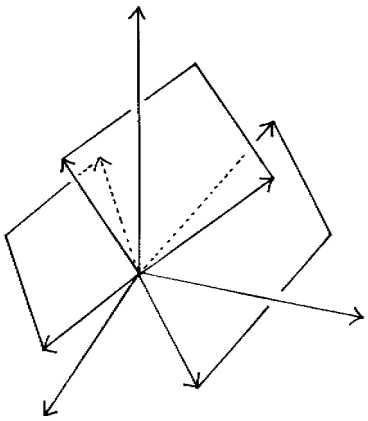


Gehen wir jetzt gleich zum Grad n: Wenn man in V eine Basis fixiert, wird die Determinante ein Beispiel einer solchen Form $V^n \longrightarrow \mathbb{R}$, und nach 39.3(b) ist dann jedes $\varphi \in \operatorname{Alt}^n V$ zu dieser speziellen Form proportional. Aufgrund unserer früheren Interpretation der Determinante als des orientierten Volumens darf man sich φ also als eine Funktion vorstellen, die Volumina in dem Sinne mißt, daß sie das orientierte Volumen des von n Vektoren aufgespannten Parallelepipeds in linearer Weise bewertet. Die verschiedenen alternierenden Formen von Grad n unterscheiden sich bei dieser Bewertung dann bloß im Maßstab, nämlich um einen festen (möglicherweise negativen) skalaren Faktor: $\operatorname{Alt}^n V$ ist eindimensional.

Daß damit das Wesentliche schon durchschaut ist, sehen Sie vor allem an der Aussage 39.3(b). Sie zeigt ja, daß der Wert einer alternierenden k-Form auf einem k-tupel $(v_1, \dots, v_k)$ sich nicht ändert, wenn man die Vektoren v_i durch andere ersetzt, die ein in demselben k-dimensionalen Unterraum gelegenes Parallelepiped gleicher Orientierung und gleichen Volumens aufspannen.



Alternierende k-Formen bewerten also die Volumina von k-dimensionalen in V gelegenen Parallelepipeden. Der naiven Anschauung nicht so direkt zugänglich — und gerade deswegen eine Erkenntnis — ist, daß diese Bewertung in einer linearen Weise möglich ist; die Linearität sagt ja auch etwas über die Bewertung von Volumina, die in ganz verschiedenen k-dimensionalen Unterräumen von V enthalten sind.



Für das weitere ist es günstig, die für die alternierenden Formen charakteristische Eigenschaft 39.3(c) noch etwas glatter schreiben zu können.

39.4 Notation Ist $\varphi \in \operatorname{Mult}^k V$ eine beliebige Multilinearform und $\sigma \in \operatorname{Sym}_k$ eine Permutation, so bezeichnen wir die mit σ transformierte Form mit $\sigma\varphi \in \operatorname{Mult}^k V$:

$$(\sigma\varphi)(v_1, \dots, v_k) := (-1)^\sigma \varphi(v_{\sigma 1}, \dots, v_{\sigma k})$$

Für alternierende Formen ist dann gerade $\sigma\varphi = \varphi$ für alle σ. Ist $\tau \in \mathrm{Sym}_k$ eine weitere Permutation, so gilt

$$\begin{aligned}\big(\tau(\sigma\varphi)\big)(v_1,\dots,v_k) &= (-1)^\tau(\sigma\varphi)(v_{\tau 1},\dots,v_{\tau k})\\ &= (-1)^\tau(-1)^\sigma\varphi(v_{\tau\sigma 1},\dots,v_{\tau\sigma k})\\ &= (-1)^{\tau\sigma}\varphi(v_{\tau\sigma 1},\dots,v_{\tau\sigma k})\\ &= \big(\varphi(\tau\sigma)\big)(v_1,\dots,v_k),\end{aligned}$$

d.h.

$$\tau(\sigma\varphi) = (\tau\sigma)\varphi,$$

so daß wir auf die Klammern verzichten können.

Multilinearformen kann man auf die naheliegende Weise miteinander multiplizieren: Sind $\varphi \in \mathrm{Mult}^k V$ und $\psi \in \mathrm{Mult}^l V$ Formen der Grade k und l, so hat die durch

$$(\varphi\psi)(v_1,\dots,v_k,w_1,\dots,w_l) := \varphi(v_1,\dots,v_k)\cdot\psi(w_1,\dots,w_l)$$

definierte Form $\varphi\psi \in \mathrm{Mult}^{k+l} V$ den Grad $k+l$. Allerdings hat das Produkt zweier alternierender Formen keinen Grund, wieder alternierend zu sein, und da man das gern hätte, hilft man der Natur etwas nach:

39.5 Lemma und **Definition** Für beliebige $k, l \in \mathbb{N}$ wird durch

$$\varphi\wedge\psi := \frac{1}{k!\,l!}\sum_{\sigma\in\mathrm{Sym}_{k+l}} \sigma(\varphi\psi)$$

eine Abbildung

$$\mathrm{Alt}^k V \times \mathrm{Alt}^l V \xrightarrow{\wedge} \mathrm{Alt}^{k+l} V$$

erklärt; man nennt sie das Dachprodukt oder äußere Produkt.

Beweis Wir müssen nur verifizieren, daß die Multilinearform $\varphi\wedge\psi$ wieder alternierend ist; dazu weisen wir die Eigenschaft 39.3(c) nach. Sei also $\tau \in \mathrm{Sym}_{k+l}$. Dann ist

$$\tau(\varphi\wedge\psi) = \frac{1}{k!\,l!}\sum_{\sigma\in\mathrm{Sym}_{k+l}} ,$$

und weil die Abbildung $\mathrm{Sym}_{k+l} \ni \sigma \longmapsto \tau\sigma \in \mathrm{Sym}_{k+l}$ eine Bijektion ist, können wir weiter

$$\tau(\varphi\wedge\psi) = \frac{1}{k!\,l!}\sum_{\tau\sigma\in\mathrm{Sym}_{k+l}} \tau\sigma(\varphi\psi) = \frac{1}{k!\,l!}\sum_{\rho\in\mathrm{Sym}_{k+l}} \rho(\varphi\psi) = \varphi\wedge\psi$$

schreiben.

Bemerkung Der an sich unwesentliche Faktor $\frac{1}{k!\,l!}$ in der Definition erweist sich aus folgendem Grund als praktisch. Die Untergruppe $\mathrm{Sym}_k \times \mathrm{Sym}_l \subset \mathrm{Sym}_{k+l}$ derjenigen Permutationen, die die ersten k und die letzten l Ziffern jeweils nur untereinander vertauschen, hat auf die alternierenden Formen φ und ψ, und damit auch auf $\varphi\psi$ keine Wirkung. In der Summe taucht jeder Summand deshalb de facto $(k!\,l!)$-mal auf, und der Vorfaktor beseitigt diese Vielfachheit wieder.

Das Dachprodukt hat die folgenden sympathischen Eigenschaften:

39.6 Lemma Das Dachprodukt ist

- bilinear,
- assoziativ und,
- wie man sagt, graduiert kommutativ: $\psi\wedge\varphi = (-1)^{kl}\varphi\wedge\psi$ für $\varphi \in \mathrm{Alt}^k V$ und $\psi \in \mathrm{Alt}^l V$

Beweis Die Bilinearität ist klar. Zum Beweis der Assoziativität rechnet man die Produkte von $\varphi \in \mathrm{Alt}^k V$, $\psi \in \mathrm{Alt}^l V$ und $\chi \in \mathrm{Alt}^m V$ aus:

$$(\varphi\wedge\psi)\wedge\chi = \frac{1}{(k+l)!\,m!}\sum_\tau \tau\big((\varphi\wedge\psi)\chi\big) = \frac{1}{k!\,l!}\frac{1}{(k+l)!\,m!}\sum_{\sigma,\tau}\tau\big(\sigma(\varphi\psi))\chi\big);$$

in den Summen läuft σ über Sym_{k+l} und τ über Sym_{k+l+m}. Wenn wir Sym_{k+l} als diejenige Untergruppe von Sym_{k+l+m} ansehen, die die letzten m Ziffern festläßt und damit χ nicht verändert, können wir das auch

$$(\varphi\wedge\psi)\wedge\chi = \frac{1}{k!\,l!}\frac{1}{(k+l)!\,m!}\sum_{\sigma,\tau}\tau\sigma(\varphi\psi\chi)$$

schreiben. Nun wird ein gegebenes Element $\rho \in \mathrm{Sym}_{k+l+m}$ auf genau $(k+l)!$ verschiedene Weisen als Produkt $\tau\sigma$ realisiert, nämlich mit beliebigem $\sigma \in \mathrm{Sym}_{k+l}$ und dadurch festgelegtem $\tau = \rho\sigma^{-1} \in \mathrm{Sym}_{k+l+m}$. Nach Zusammenfassen der je $(k+l)!$ gleichen Summanden bleibt

$$(\varphi\wedge\psi)\wedge\chi = \frac{1}{k!\,l!\,m!}\sum_\rho \rho(\varphi\psi\chi).$$

Es liegt auf der Hand, daß die Auswertung von $\varphi\wedge(\psi\wedge\chi)$ dasselbe liefert.

Zum Beweis des (graduierten) Kommutativgesetzes brauchen wir die spezielle Permutation $\rho \in \mathrm{Sym}_{k+l}$, die die ersten l Ziffern unter Beibehaltung ihrer Reihenfolge an den hinteren k vorbeischiebt. Für beliebige Vektoren $v_i, w_j \in V$ gilt

$$\begin{aligned}\big(\rho(\psi\varphi)\big)(v_1,\dots,v_k,w_1,\dots,w_l) &= (-1)^\rho(\psi\varphi)(w_1,\dots,w_l,v_1,\dots,v_k)\\ &= (-1)^\rho\psi(w_1,\dots,w_l)\cdot\varphi(v_1,\dots,v_k)\\ &= (-1)^\rho\varphi(v_1,\dots,v_k)\cdot\psi(w_1,\dots,w_l)\\ &= (-1)^\rho(\varphi\psi)(v_1,\dots,v_k,w_1,\dots,w_l),\end{aligned}$$

also ist

$$\rho(\psi\varphi) = (-1)^\rho\varphi\psi = (-1)^{kl}\varphi\psi.$$

Weil $\mathrm{Sym}_{k+l} \ni \sigma \longmapsto \sigma\rho \in \mathrm{Sym}_{k+l}$ bijektiv ist, folgt wie behauptet

$$\psi\wedge\varphi = \frac{1}{k!\,l!}\sum_\sigma\sigma(\psi\varphi) = \frac{1}{k!\,l!}\sum_\sigma\sigma\rho(\psi\varphi) = (-1)^{kl}\frac{1}{k!\,l!}\sum_\sigma\sigma(\varphi\psi) = (-1)^{kl}\varphi\wedge\psi.$$

Die im Assoziativitätsbeweis hergeleitete Formel für das Dachprodukt dreier alternierender Formen verallgemeinert sich leicht auf Produkte mit beliebig vielen Faktoren, und man erhält als

39.7 Korollar Für beliebige Formen $\varphi_i \in \mathrm{Alt}^{k_i} V$ $(i = 1,\dots,r)$ gilt

$$\varphi_1\wedge\varphi_2\wedge\cdots\wedge\varphi_r = \frac{1}{k_1!\,k_2!\cdots k_r!}\sum_\sigma\sigma(\varphi_1\varphi_2\cdots\varphi_r),$$

worin die Summe über alle $\sigma \in \mathrm{Sym}_{k_1+k_2+\cdots+k_r}$ zu bilden ist.

Speziell für ein Dachprodukt von lauter Linearformen ergibt sich die

39.8 Formel $(\varphi_1\wedge\cdots\wedge\varphi_k)(v_1,\dots,v_k) = \det\Big(\varphi_i(v_j)\Big)_{i,j=1}^k$ für beliebige Linearformen $\varphi_1,\dots,\varphi_k$ auf V.

Beweis Die Darstellung aus dem Korollar

$$\begin{aligned}(\varphi_1\wedge\varphi_2\wedge\cdots\wedge\varphi_k)(v_1,v_2,\dots,v_k) &= \sum_{\sigma\in\mathrm{Sym}_k}\big(\sigma(\varphi_1\varphi_2\cdots\varphi_k)\big)(v_1,v_2,\dots,v_k)\\ &= \sum_{\sigma\in\mathrm{Sym}_k}(-1)^\sigma(\varphi_1\varphi_2\cdots\varphi_k)(v_{\sigma1},v_{\sigma2},\dots,v_{\sigma k})\\ &= \sum_{\sigma\in\mathrm{Sym}_k}(-1)^\sigma\varphi_1(v_{\sigma1})\varphi_2(v_{\sigma1})\cdots\varphi_k(v_{\sigma k})\end{aligned}$$

führt nach der Formel 22.19 auf die angegebene Determinante.

Die konkrete Beschreibung alternierender Formen auf V kann von einer Basis von V ausgehen. Im Hinblick auf die im nächsten Abschnitt geplante Anwendung in der Vektoranalysis wollen wir stattdessen primär eine Basis des Dualraums $V^\check{} = \mathrm{Hom}(V,\mathbb{R}) = \mathrm{Alt}^1 V$ zugrundelegen, was natürlich völlig gleichwertig ist, siehe die Definition der dualen Basis 17.2. Außerdem werden wir diese Basis von $V^\check{}$ durchweg mit $(dh_1,\dots,dh_n)$ bezeichnen. Im Rahmen der multilinearen Algebra mag dh_j einfach als exzentrische Namenswahl für eine Linearform gelten; die analytische Bedeutung des Symbols, die Sie ja schon kennen, spielt in diesem Abschnitt noch keine Rolle.

39.9 Satz Sei $(dh_1,\dots,dh_n)$ eine Basis von $V^\check{}$ und $k\in\mathbb{N}$. Dann bilden die Dachprodukte

$$dh_{i_1}\wedge dh_{i_2}\wedge\cdots\wedge dh_{i_k}$$

mit $1\le i_1<i_2<\cdots<i_k\le n$ eine Basis des Vektorraums $\mathrm{Alt}^k V$. Insbesondere ist $\dim\mathrm{Alt}^k V=\binom{n}{k}$.

Beweis Wir verwenden als Hilfsmittel die zu $(dh_1,\dots,dh_n)$ duale Basis $(v_1,\dots,v_n)$ von V, die sich gemäß der Definition 17.2 durch

$$dh_i(v_j)=\delta_{ij}\quad\text{für } i,j=1,\dots,n$$

bestimmt. Es ist klar, daß eine Multilinearform $\varphi\in\mathrm{Mult}^k V$ durch ihre Werte auf allen k-Tupeln von Basisvektoren festgelegt ist. Ist φ alternierend, so genügen aufgrund der Eigenschaft 39.3(c) sogar schon die Werte auf den k-Tupeln $(v_{j_1},v_{j_2},\dots,v_{j_k})$ mit $1\le j_1<j_2<\cdots<j_k\le n$. Da es genau $\binom{n}{k}$ solche k-tupel gibt — nämlich ebensoviele wie k-elementige Teilmengen von $\{1,\dots,n\}$ — ist $\dim\mathrm{Alt}^k V\le\binom{n}{k}$. Es genügt jetzt, die angegebenen Basisformen als linear unabhängig zu erkennen. Dazu braucht man nur nach der Formel 39.8

$$(dh_{i_1}\wedge\cdots\wedge dh_{i_k})(v_{j_1},\dots,v_{j_k})=\det\Big(dh_{i_r}(v_{j_s})\Big)_{r,s=1}^{k}=\begin{cases}1 & \text{wenn } (i_1,\dots,i_k)=(j_1,\dots,j_k)\\ 0 & \text{sonst}\end{cases}$$

auszurechnen.

39.10 Beispiel Kartesische Koordinaten auf $\mathbb{R}^3$ liefern die Basen

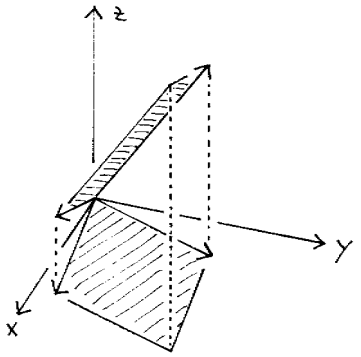
$$\begin{aligned}1 &\quad\text{für } \mathrm{Alt}^0(\mathbb{R}^3)=\mathbb{R},\\ (dx,dy,dz) &\quad\text{für } \mathrm{Alt}^1(\mathbb{R}^3)=(\mathbb{R}^3)^\check{},\\ (dy\wedge dz,\ dz\wedge dx,\ dx\wedge dy) &\quad\text{für } \mathrm{Alt}^2(\mathbb{R}^3)\quad\text{und}\\ dx\wedge dy\wedge dz &\quad\text{für } \mathrm{Alt}^3(\mathbb{R}^3).\end{aligned}$$

Im Grad zwei bin ich dabei der allgemein üblichen und von der Formulierung 39.9 abweichenden Konvention gefolgt, die die Basisformen vom Grad $n-1$ nach dem weggelassenen Faktor ordnet und die übrigen Faktoren in zyklische Reihenfolge bringt. (Für $1<k<n-1$ enthält jede Anordnung ein größeres Maß an Willkür: Etwa hat man in einem $\mathbb{R}^4$ mit Koordinatennamen t,x,y,z im Grad zwei die Basisformen

$$dt\wedge dx,\ dt\wedge dy,\ dt\wedge dz,\ dx\wedge dy,\ dx\wedge dz\text{ und }dy\wedge dz$$

zu betrachten.)

Was diese Formen machen, ist schnell verstanden: Etwa bewertet die Form $dx\wedge dy$ jedes in der xy-Ebene gelegene Vektorpaar mit dem orientierten Flächeninhalt des aufgespannten Parallelogramms, Paare in den beiden anderen Koordinatenebenen dagegen mit null. Zwei beliebige Vektoren im Raum werden so bewertet wie ihre orthogonale Projektion in die xy-Ebene.



Bezüglich der angegebenen Basen wird insbesondere das äußere Produkt

$$\mathrm{Alt}^1(\mathbb{R}^3) \times \mathrm{Alt}^1(\mathbb{R}^3) \xrightarrow{\wedge} \mathrm{Alt}^2(\mathbb{R}^3)$$

zu einer bilinearen Abbildung $\mathbb{R}^3 \times \mathbb{R}^3 \longrightarrow \mathbb{R}^3$, die man aus

$$\begin{aligned}
&(\lambda_1 dx + \mu_1 dy + \nu_1 dz) \wedge (\lambda_2 dx + \mu_2 dy + \nu_2 dz) \\
&= \lambda_1\mu_2\, dx\wedge dy + \lambda_1\nu_2\, dx\wedge dz + \mu_1\lambda_2\, dy\wedge dx + \mu_1\nu_2\, dy\wedge dz + \nu_1\lambda_2\, dz\wedge dx + \nu_1\mu_2\, dz\wedge dy \\
&= (\mu_1\nu_2 - \nu_1\mu_2)\, dy\wedge dz + (\nu_1\lambda_2 - \lambda_1\nu_2)\, dz\wedge dx + (\lambda_1\mu_2 - \mu_1\lambda_2)\, dx\wedge dy
\end{aligned}$$

explizit ablesen kann und die Ihnen als "Kreuzprodukt" wohlbekannt ist.

Bemerkung Es gibt noch andere Arten, das Vektorprodukt invariant (koordinatenfrei) zu interpretieren; daran wollen wir uns hier nicht aufhalten. Die Tatsache, daß das Produkt zweier Vektoren des Raumes "von Natur aus" in einem anderen, nur eben auch dreidimensionalen Raum liegt, ist zwar in der Koordinatenschreibweise nicht erkennbar, spiegelt sich aber in der Gesamtheit der physikalischen Formeln dadurch wieder, daß gewisse Ausdrücke nie vorkommen, obwohl sie rein rechnerisch möglich wären. So werden Sie kaum jemals auf die Summe aus einem Impuls und einem Drehimpuls, einem Geschwindigkeitsvektor und einer Winkelgeschwindigkeit, oder einem elektrischen und einem magnetischen Feldstärkenvektor etc. stoßen: das wäre nicht nur wegen der unterschiedlichen Einheiten sinnlos (die man durch geeignete skalare Faktoren wohl noch zurechtbiegen könnte), sondern vor allem deswegen, weil man dann Vektoren aus zwei ganz verschiedenen Vektorräumen addieren müßte.

Jetzt sei W ein zweiter, sagen wir p-dimensionaler reeller Vektorraum, und $f\colon V \longrightarrow W$ eine lineare Abbildung. Die Definition 27.4

$$f^{\check{}}(\psi) = \psi \circ f$$

der zu f dualen linearen Abbildung $f^{\check{}}\colon W^{\check{}} \longrightarrow V^{\check{}}$ verallgemeinert sich ohne weiteres auf Multilinearformen: die Vorschrift

$$\big((\mathrm{Mult}^k f)(\psi)\big)(v_1, \dots, v_k) := \psi\big(f(v_1), \dots, f(v_k)\big)$$

definiert für jedes $k \in \mathbb{N}$ eine lineare Abbildung

$$\mathrm{Mult}^k f\colon \mathrm{Mult}^k W \longrightarrow \mathrm{Mult}^k V$$

(auch hier die Richtung beachten!). Da dabei alternierende Formen wieder zu alternierenden werden, entsteht durch Einschränken eine weitere lineare Abbildung

$$\mathrm{Alt}^k f\colon \mathrm{Alt}^k W \longrightarrow \mathrm{Alt}^k V.$$

Wenn keine Verwechslungsgefahr besteht, bezeichnet man all diese Abbildungen kurz mit f^*. Sie genügen offensichtlich der Kettenregel

$$\mathrm{id}^* = \mathrm{id}, \quad (g\circ f)^* = f^* \circ g^*,$$

und die alternierende Version ist auch mit dem äußeren Produkt verträglich:

39.11 Lemma Für beliebige alternierende Formen φ, ψ gilt $f^*(\varphi\wedge\psi) = f^*(\varphi)\wedge f^*(\psi)$.

Beweis Für $\varphi \in \mathrm{Alt}^k V$ und $\psi \in \mathrm{Alt}^l V$ ist

$$f^*(\varphi\wedge\psi) = f^*\Big(\frac{1}{k!\,l!}\sum_\sigma \sigma(\varphi\psi)\Big) = \frac{1}{k!\,l!}\sum_\sigma f^*\big(\sigma(\varphi\psi)\big) = \frac{1}{k!\,l!}\sum_\sigma \sigma\big(f^*(\varphi)f^*(\psi)\big) = f^*(\varphi)\wedge f^*(\psi).$$

Natürlich läßt sich f^* auch in Karten berechnen. Seien dazu wieder $(dh_1, \dots, dh_n)$ und $\underline{v} = (v_1, \dots, v_n)$ zueinander duale Basen von $V^\vee$ und V, sowie $(dk_1, \dots, dk_p)$ und $\underline{w} = (w_1, \dots, w_p)$ ebensolche Basen von $W^\vee$ und W. Wird $f\colon V \longrightarrow W$ bezüglich $\underline{v}$ und $\underline{w}$ durch die Matrix $a \in \mathrm{Mat}(p\times n, \mathbb{R})$ beschrieben, so geht es darum, die Wirkung von f^* auf die Basisformen $dk_{i_1}\wedge\dots\wedge dk_{i_m}$ mittels der Koeffizienten von a und der Basisformen $dh_{j_1}\wedge\dots\wedge dh_{j_m}$ auszudrücken. Für die Linearformen ($m=1$) ergibt sich einfach

$$f^*(dk_i)(v_j) = (dk_i\circ f)(v_j) = a_{ij}$$

(vergleiche Aufgabe 27.1), was man auch in der Form

$$f^*(dk_i) = \sum_{j=1}^n a_{ij}\,dh_j$$

schreiben kann, wenn man das vorzieht. Mittels der Formeln aus 39.8 und 39.11 schließt man nun leicht auf den allgemeinen Fall:

$$\begin{aligned} f^*(dk_{i_1}\wedge\dots\wedge dk_{i_m})(v_{j_1},\dots,v_{j_m}) &= \big(f^*(dk_{i_1})\wedge\dots\wedge f^*(dk_{i_m})\big)(v_{j_1},\dots,v_{j_m}) \\ &= \det\Big(\big(f^*(dk_{i_r})\big)(v_{j_s})\Big)_{r,s=1}^m \\ &= \det\Big(a_{i_r,j_s}\Big)_{r,s=1}^m \end{aligned}$$

oder

$$f^*(dk_{i_1}\wedge\dots\wedge dk_{i_m}) = \sum \det\Big(a_{i_r j_s}\Big)_{r,s=1}^m dh_{j_1}\wedge\dots\wedge dh_{j_m},$$

worin die Summe über alle m-tupel $(j_1,\dots,j_m)$ mit $1\le j_1<j_2<\dots<j_m\le n$ zu nehmen ist.

Bemerkung Für $V=W$ und $m=n$ ist $f^*\colon \mathrm{Alt}^n V \longrightarrow \mathrm{Alt}^n V$ also die Multiplikation mit $\det a = \det f$, wie man auch schon aus Aussage (b) von Satz 39.3 ablesen kann. Wenn man auch im Fall $m=n-1$ zweimal dieselbe Basis in $V=W$ verwendet, die n Basisformen in $\mathrm{Alt}^{n-1}V$ nach dem fehlenden Faktor dh_{i_r} ordnet und noch mit einem geeigneten Vorzeichen versieht, wird die Matrix von f^* die unter der Nummer 22.12 definierte Adjunkte $\tilde a$ von a.

Übungsaufgaben

39.1 Sei V ein n-dimensionaler reeller Vektorraum. Es ist klar, was mit einer *symmetrischen* Multilinearform vom Grad k auf V gemeint ist, und daß diese Formen einen Untervektorraum $\mathrm{Sym}^k V \subset \mathrm{Mult}^k V$ bilden. Beweisen Sie, daß $\mathrm{Sym}^2 V$ und $\mathrm{Alt}^2 V$ zueinander komplementäre Teilräume von $\mathrm{Mult}^2 V$ sind. Ist allgemeiner $\mathrm{Mult}^k V$ die direkte Summe von $\mathrm{Sym}^k V$ und $\mathrm{Alt}^k V$?

39.2 Natürlich ist die alternierende Form $dt\wedge dx + dy\wedge dz \in \mathrm{Alt}^2\mathbb{R}^4$ von jeder der sechs Basisformen $dt\wedge dx$, $dt\wedge dy$ ... verschieden. Aber vielleicht findet man Linearformen $\varphi, \psi \in \mathrm{Alt}^1\mathbb{R}^4$, so daß

$$dt\wedge dx + dy\wedge dz = \varphi\wedge\psi$$

ist?

39.3 V sei ein endlichdimensionaler reeller Vektorraum und $\omega \in \mathrm{Alt}^1 V$ eine Linearform, $\omega \neq 0$. Beweisen Sie, daß dann für jede Form $\varphi \in \mathrm{Alt}^k V$ gilt:

$$\omega \wedge \varphi = 0 \iff \text{es gibt ein } \psi \in \mathrm{Alt}^{k-1} V \text{ mit } \varphi = \omega \wedge \psi$$

(Wer pingelig ist, muß sich dafür die Definition um $\mathrm{Alt}^{-1} V := \{0\}$ ergänzt denken.) Tip: Man darf natürlich in einer bequem gewählten Basis rechnen.

39.5 Sei $L \subset \mathbb{R}^3$ eine Gerade (durch 0), etwa $L = \mathrm{Lin}(v)$. Die Tatsache, daß $\mathbb{R}^3$ und L euklidische Vektorräume sind, erlaubt es, ihre Dualräume $(\mathbb{R}^3)^\vee = \mathrm{Alt}^1 \mathbb{R}^3$ und $L^\vee = \mathrm{Alt}^1 L$ als $\mathbb{R}^3$ bzw. L selbst aufzufassen. Wenn man das tut, welche Bedeutung bekommt dann die Einschränkungsabbildung

$$\mathrm{Alt}^1 \mathbb{R}^3 \longrightarrow \mathrm{Alt}^1 L; \quad \varphi \mapsto \varphi | L \ ?$$

39.6 Die Ebene $H \subset \mathbb{R}$ sei als orthogonales Komplement $H = \{n\}^\perp$ eines Vektors $n \in \mathbb{R}^3$ mit $|n| = 1$ gegeben, dann bildet die einzelne Form

$$H \times H \ni (v, w) \longmapsto \det \begin{pmatrix} v & w & n \end{pmatrix} \in \mathbb{R}$$

eine Basis von $\mathrm{Alt}^2 H$. Wie sieht die Einschränkungsabbildung

$$\mathrm{Alt}^2 \mathbb{R}^3 \longrightarrow \mathrm{Alt}^2 H; \quad \varphi \mapsto \varphi | H$$

bezüglich dieser Basis und der als Beispiel 39.10 beschriebenen Basis von $\mathrm{Alt}^2 \mathbb{R}^3$ aus?

40 Differentialformen

Wir kehren zur Vektoranalysis zurück. Die Differentialformen, die wir jetzt definieren, verallgemeinern die Pfaffschen Formen in folgendem Sinne: Während diese jedem Punkt einer Mannigfaltigkeit eine Linearform auf dem Tangentialraum zuordnen, wollen wir jetzt in jedem Punkt eine Multilinearform erklären. Dazu brauchen wir zuerst:

40.1 Definition $X \subset \mathbb{R}^N$ sei eine n-dimensionale Untermannigfaltigkeit, k eine natürliche Zahl. Dann heißt

$$T^kX := \big\{(x, v_1, \dots, v_k) \in X \times \mathbb{R}^N \times \dots \times \mathbb{R}^N \,\big|\, v_j \in T_xX \text{ für } j = 1, \dots, k\big\}$$

zusammen mit der Projektion $T^kX \ni (x, v_1, \dots, v_k) \overset{\pi}{\longmapsto} x \in X$ das k-Tangentialbündel von X.

Bemerkungen und *Ergänzungen* Die Projektion des k-Tangentialbündels hat als Faser über $x \in X$ im wesentlichen das direkte Produkt von k Kopien des Tangentialraums T_xX:

$$\pi^{-1}\{x\} = \{x\} \times T_xX \times \dots \times T_xX$$

Für $k = 0$ ist natürlich $T^0X = X$, und für $k = 1$ erhält man das gewöhnliche Tangentialbündel $T^1X = TX$. — Wenn $X \subset \mathbb{R}^n$ eine offene Teilmenge ist, ist einfach $T^kX = X \times (\mathbb{R}^n)^k$. Auch im allgemeinen Fall ist nicht schwer zu sehen, daß T^kX eine Mannigfaltigkeit der Dimension $(k+1)n$ ist und daß jede Karte (U, h) von Y einen Diffeomorphismus

$$T^kU \xrightarrow{\quad T^kh \quad} T^k\big(h(U)\big) = h(U) \times (\mathbb{R}^n)^k,$$

also eine Karte für T^kX liefert; sie sendet die Faser $\{x\} \times (T_xX)^k$ Komponente für Komponente mittels Th auf $\{h(x)\} \times (\mathbb{R}^n)^k$.

40.2 Definition Sei X eine Mannigfaltigkeit und $k \in \mathbb{N}$. Eine Differentialform vom Grad k auf X ist eine Funktion $\varphi\colon T^kX \longrightarrow \mathbb{R}$, die sich auf jeder Faser zu einer alternierenden Multilinearform $\varphi_x \in \operatorname{Alt}^k T_xX$ einschränkt:

$$\varphi_x(v_1, \dots, v_k) := \varphi(x, v_1, \dots, v_k) \quad \text{für jedes } x \in X$$

Üblicherweise spricht man kurz von einer k-Form auf X. Am interessantesten sind *differenzierbare* k-Formen, womit wir C^∞-Formen meinen; diese vereinfachende Vereinbarung soll weiterhin in Kraft bleiben. Die differenzierbaren k-Formen auf X bilden in offensichtlicher Weise einen Vektorraum, den man mit $\mathcal{A}^kX$ bezeichnet.

Bemerkung 1-Formen sind also dasselbe wie Pfaffsche Formen, während eine 0-Form $T^0X = X \longrightarrow \mathbb{R}$ einfach eine reellwertige Funktion auf X ist.

Die Bildungen der linearen Algebra aus dem vorigen Abschnitt lassen sich mühelos auf Differentialformen übertragen.

40.3 Definition (a) Sei φ eine k-Form und ψ eine l-Form auf der Mannigfaltigkeit X. Das Dachprodukt oder äußere Produkt von φ und ψ ist die durch

$$(\varphi \wedge \psi)_x = \varphi_x \wedge \psi_x$$

definierte $(k+l)$-Form $\varphi \wedge \psi$. Insbesondere ist so eine Multiplikation differenzierbarer Formen

$$\mathcal{A}^kX \times \mathcal{A}^lX \overset{\wedge}{\longrightarrow} \mathcal{A}^{k+l}X$$

erklärt.

(b) Sei ψ eine m-Form auf Y und $f\colon X \longrightarrow Y$ eine differenzierbare Abbildung. Dann bestimmt die Formel

$$(f^*\psi)_x = (T_x f)^*(\psi_x)$$

eine m-Form $f^*\psi$ auf X; man nennt diesen Vorgang Zurückziehen der Form ψ mittels f. Insbesondere ist so eine lineare Abbildung

$$f^*\colon \mathcal{A}^m Y \longrightarrow \mathcal{A}^m X$$

erklärt. Ist speziell $f\colon X \hookrightarrow Y$ die Inklusion einer Untermannigfaltigkeit, so ist $f^*\psi = \psi|T^k X$; man schreibt dafür kurz $\psi|X$ und spricht von der auf X eingeschränkten Form.

Bemerkungen Für $k = 0$ reduziert sich das äußere Produkt auf die skalare Multiplikation einer Funktion $\varphi\colon X \longrightarrow \mathbb{R}$ mit der l-Form ψ. — Wegen der Kettenregel 37.3 gilt $\mathrm{id}^*\,\psi = \psi$ und $(g\circ f)^*\psi = g^* f^*\psi$, und wegen Lemma 39.11 auch $f^*(\varphi\wedge\psi) = f^*\varphi \wedge f^*\psi$. — Im Fall einer Pfaffschen Form ψ ist $f^*\psi = \psi\circ Tf$, und für eine 0-Form ψ ist $f^*\psi$ nur eine gelehrte Schreibweise für $\psi\circ f$.

Seien in der Situation von (b) auf X und Y Karten (U,h) und (V,k) mit $f(U) \subset V$ gegeben. Nach Satz 39.9 schreibt sich ψ auf U als

$$\psi = \sum g_{i_1 i_2 \dots i_m}\, dk_{i_1} \wedge dk_{i_2} \wedge \cdots \wedge dk_{i_m}$$

mit eindeutig bestimmten Funktionen $g_{i_1 i_2 \dots i_m}\colon V \longrightarrow \mathbb{R}$; natürlich ist dabei über alle $(i_1, \dots, i_m)$ mit $i_1 < \cdots < i_m$ zu summieren. (Die dk_i haben jetzt wieder ihre analytische Bedeutung als Differentiale der Koordinatenfunktionen k_i.) Wir wollen in analoger Weise die zurückgezogene Form $f^*\psi$ auf U durch die dh_j ausdrücken; dazu genügt es im Prinzip, in

$$f^*\psi = \sum (g_{i_1 i_2 \dots i_m} \circ f)\; f^* dk_{i_1} \wedge f^* dk_{i_2} \wedge \cdots \wedge f^* dk_{i_m}$$

alle f^*dk_i gemäß

$$f^* dk_i = dk_i \circ Tf = d(k_i \circ f) = \sum_{j=1}^{n} \frac{\partial(k_i \circ f)}{\partial h_j}\, dh_j$$

zu substituieren und die entstehenden Dachprodukte tapfer auszurechnen. Tatsächlich haben wir das im Anschluß an Lemma 39.11 schon durchgeführt, müssen in die dort erhaltene Formel bloß noch die Jacobi-Matrix von f einsetzen und erhalten:

$$f^*(dk_{i_1} \wedge dk_{i_2} \wedge \cdots \wedge dk_{i_m}) = \sum \det \left(\frac{\partial(k_{i_r} \circ f)}{\partial h_{j_s}} \right)_{r,s=1}^{m} dh_{j_1} \wedge \cdots \wedge dh_{j_m}$$

40.4 Beispiel Die differenzierbare Abbildung

$$\mathbb{R}^2 \ni \begin{pmatrix} u \\ v \end{pmatrix} \longmapsto \begin{pmatrix} u^2 + v^3 \\ uv^2 \\ v^2 \end{pmatrix} \in \mathbb{R}^3$$

in den $\mathbb{R}^3$ mit Standardkoordinaten x, y, z hat die Jacobi-Matrix

$$Df(u,v) = \begin{pmatrix} 2u & 3v^2 \\ v^2 & 2uv \\ 0 & 2v \end{pmatrix}$$

und zieht deshalb die 1-Form $\cos x\, dy + \sin y\, dz$ zu

$$\begin{aligned} f^*(\cos x\, dy + \sin y\, dz) &= \big(\cos(u^2+v^3)\big)(v^2\, du + 2uv\, dv) + \big(\sin(uv^2)\big)(2v\, dv) \\ &= v^2 \cos(u^2+v^3)\, du + \big(2uv\cos(u^2+v^3) + 2v\sin(uv^2)\big) dv, \end{aligned}$$

die 2-Form $e^z dx\wedge dy + dx\wedge dz$ zu

$$\begin{aligned} f^*(e^z dx\wedge dy + dx\wedge dz) &= e^{v^2} \det \begin{pmatrix} 2u & 3v^2 \\ v^2 & 2uv \end{pmatrix} du\wedge dv + \det \begin{pmatrix} 2u & 3v^2 \\ 0 & 2v \end{pmatrix} du\wedge dv \\ &= \big((4u^2v - 3v^4)e^{v^2} + 4uv\big) du\wedge dv \end{aligned}$$

und jede 3-Form von $\mathbb{R}^3$ natürlich zu null zurück, weil es auf $\mathbb{R}^2$ keine nicht-trivialen 3-Formen gibt.

Die Formeln für das Zurückziehen einer Differentialform unter f enthalten zugleich die Umrechnungsvorschrift bei Kartenwechsel, denn man braucht bloß für f die identische Abbildung einzusetzen:

$$dk_{i_1}\wedge dk_{i_2}\wedge\cdots\wedge dk_{i_m} = \sum \det \left(\frac{\partial k_{i_r}}{\partial h_{j_s}} \right)_{r,s=1}^{m} dh_{j_1}\wedge\cdots\wedge dh_{j_m}$$

Zum Beispiel gilt für ebene Polarkoordinaten

$$dx\wedge dy = \det \begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix} dr\wedge d\varphi = r\, dr\wedge d\varphi.$$

40.5 Tabelle Wir wollen auf $\mathbb{R}^3$ erklärte Vektorfelder und Differentialformen der verschiedenen Grade einander in mathematischer und gängiger physikalischer Koordinatennotation gegenüberstellen. Außerdem interessieren wir uns dafür, wie sich die physikalische Darstellung transformiert, wenn man von den Standardkoordinaten x, y, z zu beliebigen affin-linearen Koordinaten x', y', z' übergeht; denkt man sich die Koordinatentripel als Spalten $\vec{r}$ und $\vec{r}\,'$ geschrieben, so ist der Zusammenhang zwischen ihnen also durch

$$\vec{r} = a\vec{r}\,' + b$$

mit Konstanten $a \in GL(3,\mathbb{R})$ und $b \in \mathbb{R}^3$ gegeben.

Mathematiker	Physiker	Transformation
Vektorfeld $\mathbb{R}^3 \xrightarrow{v} T\mathbb{R}^3$ $v = g_1\frac{\partial}{\partial x} + g_2\frac{\partial}{\partial y} + g_3\frac{\partial}{\partial z}$	(kontravariantes polares) Vektorfeld $g_i = g_i(\vec{r}) \quad (i = 1,2,3)$	$g_i' =$ $\sum_j (a^{-1})_{ij}\, g_j(a\vec{r}\,'+b)$
Funktion $\mathbb{R}^3 \xrightarrow{g} \mathbb{R}$ g	Skalar(-enfeld) $g = g(\vec{r})$	$g' =$ $g(a\vec{r}\,'+b)$
1-Form $T\mathbb{R}^3 \xrightarrow{\varphi} \mathbb{R}$ $\varphi = g_1\, dx + g_2\, dy + g_3\, dz$	(kovariantes polares) Vektorfeld $g_i = g_i(\vec{r}) \quad (i = 1,2,3)$	$g_i' =$ $\sum_j (a^{\mathrm{t}})_{ij}\, g_j(a\vec{r}\,'+b)$
2-Form $T^2\mathbb{R}^3 \xrightarrow{\varphi} \mathbb{R}$ $\varphi = g_1\, dy\wedge dz + g_2\, dz\wedge dx + g_3\, dy\wedge dz$	(axiales, Pseudo-) Vektorfeld $g_i = g_i(\vec{r}) \quad (i = 1,2,3)$	$g_i' =$ $\sum_j \tilde{a}_{ij}\, g_j(a\vec{r}\,'+b)$
3-Form $T^2\mathbb{R}^3 \xrightarrow{\varphi} \mathbb{R}$ $\varphi = g\, dx\wedge dy\wedge dz$	Pseudoskalar $g = g(\vec{r}) \quad (i = 1,2,3)$	$g' =$ $\det a \cdot g(a\vec{r}\,'+b)$

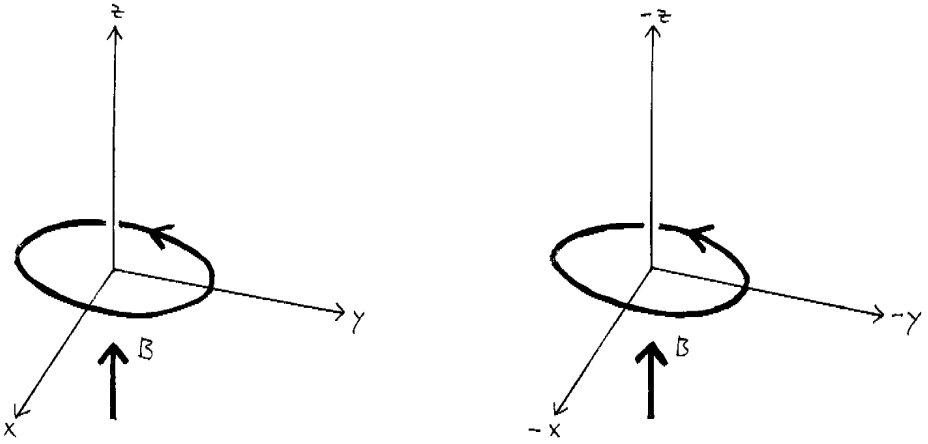
Von den Transformationsformeln ist nur diejenige für die 2-Formen (mit der Adjunkten $\tilde{a}$, siehe 22.12) nicht offensichtlich. Man erhält sie aber zuverlässig aus

$$\begin{aligned} dy\wedge dz &= (a_{21}\, dx' + a_{22}\, dy' + a_{23}\, dz') \wedge (a_{31}\, dx' + a_{32}\, dy' + a_{33}\, dz') \\ &= (a_{22}a_{33} - a_{23}a_{32})\, dy'\wedge dz' + (a_{23}a_{31} - a_{21}a_{33})\, dz'\wedge dx' + (a_{21}a_{32} - a_{22}a_{31})\, dx'\wedge dy' \\ &= \det \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} dy'\wedge dz' - \det \begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{pmatrix} dz'\wedge dx' + \det \begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} dx'\wedge dy' \end{aligned}$$

und den entsprechenden Darstellungen von $dz\wedge dx$ und $dx\wedge dy$.

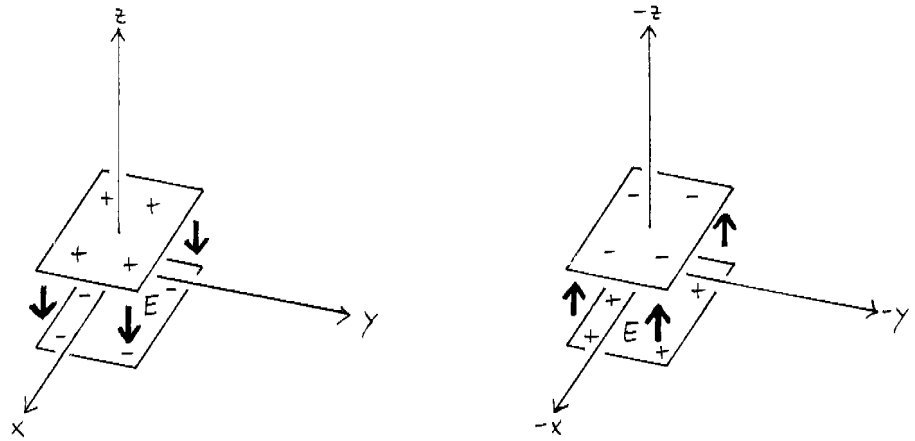
Erwartungsgemäß erkennt man die fünf verschiedenen Typen von Objekten auch in koordinatengebundener Schreibweise an ihrem Transformationsverhalten. Wenn man nun aber statt beliebiger affin-linearer Karten nur noch orthogonale, also solche mit $a \in O(n)$ zuläßt, werden kontra- und kovariante Vektorfelder ununterscheidbar, denn dann ist ja $a^{-1} = a^{\mathrm{t}}$. Dagegen verraten sich die *axial* genannten oder *Pseudovektoren* noch dadurch, daß sie unter Spiegelungen der Koordinaten im Vergleich zu den "richtigen", *polaren* Vektorfeldern zusätzlich das Vorzeichen wechseln: nach Satz 22.13 gilt für die Adjunkte $\tilde{a} = \frac{1}{\det a} \cdot a^{-1}$, hier also $\tilde{a} = -a^{-1}$. Entsprechend unterscheiden sich *Pseudoskalare* von gewöhnlichen Skalaren durch den Vorzeichenwechsel bei Raumspiegelungen. In der Physik scheinen Pseudoskalare nicht sehr geläufig zu sein; jedenfalls weiß ich nur ein konkretes Beispiel, die sogenannte Helizität von Elementarteilchen. Natürlich sind weder Pseudovektoren noch -skalare als solche erkennbar, wenn man noch einen (kleinen) Schritt weiter geht und neben den ohnehin harmlosen Koordinatenverschiebungen nur Drehungen erlaubt, für die ja $a \in SO(n)$, also $\det a = 1$ ist.

Die elektrische Feldstärke E und die magnetische Induktion B bilden ein schönes Beispielpaar polarer und axialer Vektorfelder. Obwohl es letztlich Willkür ist, welchen der beiden man als axialen und welchen als polaren Vektor ansehen will, bietet sich in der dreidimensionalen Formulierung der Elektrodynamik eher B als axialer Vektor an. Zum Beispiel erzeugt ein positiv orientierter Ringstrom in der xy-Ebene ein (innerhalb des Ringes) nach oben weisendes Magnetfeld B. Wenn man nun alle Raumkoordinaten spiegelt, bleibt die Orientierung des Ringstroms und deshalb auch B bezüglich der neuen Koordinaten unverändert; B weist also auch in diesen nach oben:



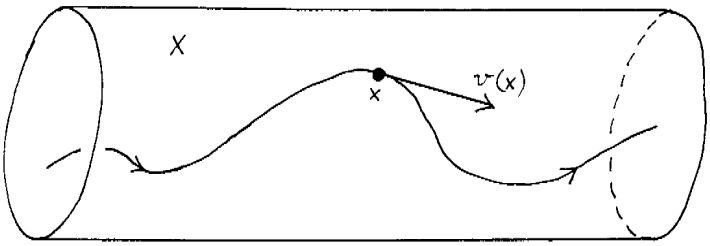
Das Paradoxon, daß B aus der Sicht der ursprünglichen Koordinaten nun scheinbar nach unten weist, wird durch die Bemerkung im Anschluß an 39.10 aufgelöst, nach der die Werte von B (mathematisch gesehen alternierende Multillinearformen vom Grad zwei) überhaupt nicht dem skizzierten Konfigurationsraum, sondern einem anderen dreidimensionalen Raum angehören.

Anders verhält sich das elektrische Feld E zwischen zwei Kondensatorplatten als polarer Vektor: es erscheint in den neuen Koordinaten umgeklappt.

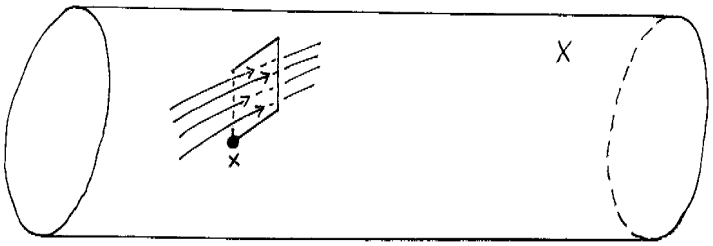


Bei manchen Vektorfeldern der Physik ist es gar nicht nötig, die günstigste Interpretation aus dem Transformationsverhalten zu rekonstruieren, weil diese sich schon von der physikalischen Bedeutung her aufdrängt.

40.6 Beispiele (1) Wir denken an ein in einem Raumgebiet $X \subset \mathbb{R}^3$ strömendes Medium. Daß dessen Geschwindigkeitsfeld v kontravariant, also ein Vektorfeld im mathematischen Sinne ist, geht fast zwingend daraus hervor, daß v an jeder Stelle $x \in X$ der Geschwindigkeitsvektor der Bahn ist, die ein Partikel dieses Mediums durchläuft. Es wäre demnach ganz unnatürlich, $v(x)$ als etwas Anderes anzusehen als einen Tangentialvektor $v(x) \in T_x X$.

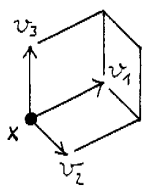


(2) Mit einem strömenden Medium sind auch Stromdichten j als Beispiele weiterer Vektorfelder im physikalischen Sinne assoziiert, zum Beispiel Massen- oder elektrische Ladungsstromdichte. Hier die wohl einfachste und auch physikalischste Art, den Begriff "Stromdichte" zu erklären: Die Stromdichte an einer Stelle, j_x, gibt zu jeder kleinen — sagen wir ruhig mal infinitesimalen — bei x in den Raum gestellten Meßfläche an, wieviel Masse oder Ladung oder ... in der Zeiteinheit durch diese Fläche strömt, natürlich mit einem Vorzeichen versehen, das Auskunft darüber gibt, in welcher Richtung die Fläche dabei durchsetzt wird.



So werden Stromdichten ja oft experimentell bestimmt! j_x bewertet also in den Raum gestellte infinitesimale Flächen; diese Bewertung ist sicher zumindest in dem Sinne linear, daß sie proportional zur Größe der Meßfläche erfolgt und bei Orientierungswechsel der Fläche das Vorzeichen wechselt. Deshalb genügen schon Meßflächen einer bestimmten Form, etwa Parallelogramme. Tatsächlich erweist sich die Bewertung als überhaupt linear, d.h. j_x als eine alternierende Bilinearform auf $T_x X$ und damit $j: T^2 X \longrightarrow \mathbb{R}$ als eine 2-Form. Beachten Sie besonders, daß diese Überlegung an keinerlei Koordinaten gebunden ist! Stromdichten sind also natürlicherweise 2-Formen.

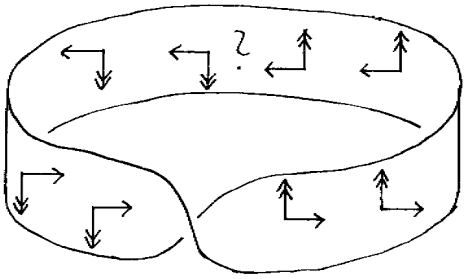
(3) In der gleichen Situation wird man die Dichte ρ des Mediums naheliegenderweise als eine 3-Form auf X ansehen: die Dichte an der Stelle x, wir schreiben ρ_x, bewertet jedes infinitesimale räumliche Parallelepiped an der Stelle x mit der darin enthaltenen Masse bzw. Ladung.



Damit die Beschreibung der beiden Beispiele vollständig wird, ist allerdings eine Orientierungsvereinbarung nötig. Dazu ein ganz knapp gefaßter

40.7 Bericht über **Orientierungen** Ein n-dimemsionaler reeller Vektorraum V wird dadurch orientiert, daß man gewisse Basen von V als *positiv orientiert* auszeichnet, und zwar so, daß zwei Basen von V, deren

Kartenwechsel positive Determinante hat, entweder beide positiv oder beide negativ orientiert sind. Natürlich gibt es für $n > 0$ genau zwei Orientierungen von V, und im Fall $n = 0$ erzwingt man das durch eine Sondervereinbarung. Einen linearen Isomorphismus zwischen zwei n-dimensionalen orientierten Vektorräumen nennt man *orientierungstreu (-erhaltend)* oder *orientierungsumkehrend* je nach dem Vorzeichen der bezüglich positiv orientierter Basen gebildeten Determinante. — Eine n-dimensionale Mannigfaltigkeit zu orientieren, heißt alle Tangentialräume von X zu orientieren, aber in einer "stetigen" Weise, nämlich so, daß es um jedes $x \in X$ eine Karte (U, h) derart gibt, daß die Isomorphismen $T_y h: T_y X \longrightarrow \mathbb{R}^n$ für alle $y \in X$ dasselbe Verhalten — orientierungstreu oder -umkehrend — aufweisen. Es gibt Mannigfaltigkeiten, die keine solche Orientierung erlauben, eben *nicht orientierbar* sind; die bekannteste ist das Möbiusband.

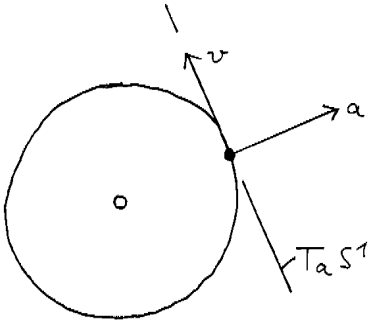


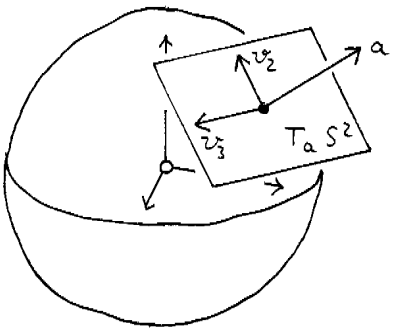
Versucht man hier, eine an einer Stelle vorgegebene Orientierung durch Verschieben längs des Bandes auszubreiten, so entsteht nach einem vollem Umlauf unweigerlich eine Orientierungskollision.

Man sieht andererseits ohne große Schwierigkeiten, daß jede nicht-leere zusammenhängende Mannigfaltigkeit X, die überhaupt orientierbar ist, genau zwei Orientierungen besitzt; wenn man sich für eine davon entschieden hat, bezeichnet man die entgegengesetzt orientierte Mannigfaltigkeit symbolisch mit $-X$.

Natürlich soll der Vektorraum $\mathbb{R}^n$ immer stillschweigend die kanonische Orientierung tragen, in der die Standardbasis positiv orientiert ist. Damit ist auch jede offene Teilmenge $X \subset \mathbb{R}^n$ als Mannigfaltigkeit kanonisch orientiert, denn alle ihre Tangentialräume sind ja $\mathbb{R}^n$. Viele weitere orientierte Mannigfaltigkeiten erhält man nach dem Satz vom regulären Wert. Dort entsteht eine $(n–p)$-dimensionale Untermannigfaltigkeit von $\mathbb{R}^n$ als die Faser $Y = f^{-1}\{b\} \subset X$ der Abbildung $f: X \longrightarrow \mathbb{R}^p$ über dem regulären Wert $b \in \mathbb{R}^p$. Die Tangentialräume $T_a Y$ orientiert man nun durch die folgende Regel: Ist $\underline{v} = (v_{p+1}, \ldots, v_n)$ eine Basis von $T_a Y$, so ergänzt man sie durch vorangestellte Vektoren $v_1, \ldots, v_p \in \mathbb{R}^n$ mit $(Df_i)(a)v_j = \delta_{ij}$ zu einer Basis $(v_1, \ldots, v_n)$ von $\mathbb{R}^n$ und weist $\underline{v}$ dasselbe Orientierungsvorzeichen zu wie dieser. Daß man ergänzende Vektoren $v_1, \ldots, v_p$ mit den geforderten Eigenschaften findet, sieht man sofort etwa mittels des Satzes vom regulären Punkt, den man zum Beweis dessen vom regulären Wert ja sowieso braucht. Es ist auch nicht schwer nachzuprüfen, daß die Regel wirklich eine Orientierung von Y liefert, die wohlgemerkt von der Darstellung von Y als Faser $f^{-1}\{b\}$ abhängt.

Für die Sphären in der Standarddarstellung $S^{n-1} = f^{-1}\{1\}$ mit $\mathbb{R}^n \ni x \overset{f}{\longmapsto} |x| \in \mathbb{R}$ sagt die Regel, daß eine Basis $(v_2, \ldots, v_n)$ von $T_a S^{n-1} = \{a\}^\perp$ genau dann positiv orientiert ist, wenn sie durch Voranstellen eines nach außen weisenden Vektors zu einer positiv orientierten Basis von $\mathbb{R}^n$ wird, d.h. wenn $(a, v_2, \ldots, v_n)$ positiv orientiert ist.





Nun nochmal zu den Beispielen 40.6. Nicht nur für 3-Formen auf $\mathbb{R}^3$, sondern allgemein für n-Formen auf der orientierten n-dimensionalen Mannigfaltigkeit X ist jetzt klar: als positiv orientiert gelten genau die von einem positiv orientierten n-tupel von Tangentialvektoren aufgespannten Parallelepipede. Eine als n-Form aufgefaßte Dichte bewertet also die darin vorhandenen positiven Ladungen positiv. Und für die Interpretation einer Stromdichte als $(n-1)$-Form vereinbaren wir, daß das $(n-1)$-tupel $(v_2, \dots, v_n)$ von Tangentialvektoren genau dann positiv zu bewerten ist, wenn $(u, v_1, \dots, v_{n-1})$ positiv orientiert ist, wobei $u \in T_xX$ in die Stromrichtung (der positiven Ladungen) zeigt.

Bemerkung Da Dichten in der Physik ganz geläufige Objekte sind, scheint Beispiel 40.6(3) zunächst die anläßlich der Tabelle 40.5 geäußerte Vermutung zu widerlegen, nach der Pseudoskalare wenig gebräuchlich sind. Das liegt aber nur daran, daß Physiker anstelle der 3-Formen traditionell einen speziellen mathematischen Begriff "Dichte" vorziehen, bei dem die Jacobi-Determinante nur mit ihrem Absolutbetrag in die Transformationformel eingeht, so daß solche Dichten sich unter Raumspiegelungen (aber nicht beliebigen Kartenwechseln!) doch wie echte Skalare verhalten. Einen Vorteil bringt das aber nur auf nicht-orientierbaren Mannigfaltigkeiten, ansonsten ist die Auffassung physikalischer Dichten als 3-Formen vorzuziehen.

Übungsaufgaben

40.1 Drücken Sie eine der drei Basisformen $dy \wedge dz$, $dz \wedge dx$ und $dx \wedge dy$ auf $\mathbb{R}^3$ in Kugelkoordinaten aus (oder auch alle drei).

40.2 ψ sei die Einschränkung der 2-Form $dy \wedge dz$ auf die Sphäre $S^2 \subset \mathbb{R}^3$. Berechnen Sie ψ auf der oberen Hemisphäre $\{(x, y, z) \in S^2 \mid z > 0\}$ erstens in der aus den Koordinatenfunktionen x und y bestehenden Karte, und zweitens in der Karte (r, φ), die aus den ersten beiden Komponenten der Zylinderkoordinaten r, φ, z gebildet ist.

40.3 Durch die Vorschrift

$$\varphi_a(w_1, w_2) := \det \begin{pmatrix} a & w_1 & w_2 \end{pmatrix}$$

wird eine Differentialform $\varphi \in \mathcal{A}^2\mathbb{R}^3$ erklärt.

(a) Wie lautet die Standarddarstellung von φ in kartesischen Koordinaten?

(b) Welche anschauliche Bedeutung hat die Einschränkung $\psi := \varphi|S^2$ auf die 2-Sphäre $S^2 \subset \mathbb{R}^3$?

(c) Berechnen Sie $f^*\psi$, wenn $f: S^2 \longrightarrow S^2$; $x \mapsto -x$ die sogenannte *Antipodenabbildung* ist.

40.4 Was wird mit einem orientierungstreuen lokalen Diffeomorphismus zwischen zwei orientierten Mannigfaltigkeiten gemeint sein? Begründen Sie: Sind X und Y zwei solche Mannnigfaltigkeiten und ist $X \neq \emptyset$ zusammenhängend, so ist jeder lokale Diffeomorphismus $f: X \longrightarrow Y$ entweder orientierungstreu oder orientierungsumkehrend. Welches Orientierungsverhalten hat die Antipodenabbildung $f: S^n \longrightarrow S^n$, die natürlich auch hier durch $x \mapsto -x$ erklärt ist?

41 Das Cartansche Differential

Vom rein mathematischen Standpunkt gesehen haben wir im vorigen Abschnitt wenig geleistet: Die Algebra der alternierenden Multilinearformen überträgt sich eben reibungslos von den Vektorräumen auf die Tangentialbündel von Mannigfaltigkeiten. Dieser Algebra und Analysis in oberflächlicher Weise verbindende Vorgang hätte übrigens ebensogut mit beliebigen Multilinearformen und vielen weiteren Objekten der linearen Algebra funktioniert. Freilich erweckt all das den Eindruck einer reinen Spielerei. Daß dies eine Fehleinschätzung ist, erkennt man im Fall der alternierenden Formen an der Möglichkeit, für solche Formen eine Ableitung zu definierenen und damit eine wirkliche Verzahnung der Algebra mit der Analysis zu schaffen. Diese jetzt zu erklärende Ableitung ist exklusiv für die Differentialformen im Sinne der Definition 40.2 erklärt, also solche, die auf den Fasern alternierend sind; es gibt keine analoge Konstruktion für beliebige fasernweise multilineare Formen.

41.1 Satz und **Definition** Sei X eine Mannigfaltigkeit. Es gibt genau eine Folge von linearen Abbildungen

$$\mathcal{A}^k X \xrightarrow{d} \mathcal{A}^{k+1} X \quad (k \in \mathbb{N})$$

mit den folgenden Eigenschaften:

(a) $d\colon \mathcal{A}^0 X \longrightarrow \mathcal{A}^1 X$ hat die alte Bedeutung, weist also der Funktion f ihr Differential df im Sinne der Definieiton 38.3 zu;

(b) es gilt die Produktregel

$$d(\varphi\wedge\psi) = d\varphi\wedge\psi + (-1)^k\varphi\wedge d\psi \quad \text{für alle } \varphi \in \mathcal{A}^k X,\ \psi \in \mathcal{A}^l X;$$

(c) für jedes $k \in \mathbb{N}$ ist die Komposition

$$\mathcal{A}^k X \xrightarrow{d} \mathcal{A}^{k+1} X \xrightarrow{d} \mathcal{A}^{k+2} X$$

die Nullabbildung.

Man nennt d die äußere Ableitung oder das (nach dem französchen Mathematiker Élie Cartan Cartansche) Differential (im jeweiligen Grad k).

Bemerkung Das Vorzeichen in der Produktregel (b) merkt man sich am besten, indem man sich d als einen Operator vom Grad eins vorstellt; tatsächlich hebt er ja den Grad einer Form um eins. Im zweiten Summanden der Produktregel ist d formal mit ψ vertauscht, was ein Vorzeichen im Einklang mit dem Kommutativgesetz aus Lemma 39.6 nach sich zieht.

Beweis des Satzes Ich will den Beweis nur unter der zusätzlichen Annahme durchführen, daß X mit einer einzigen Karte (X, h) beschrieben werden kann (was äquivalent zu dem Fall ist, daß X selbst eine offene Teilmenge von $\mathbb{R}^n$ ist). Jede Form $\varphi \in \mathcal{A}^k X$ schreibt sich in dieser Karte als

$$\varphi = \sum g_{i_1\dots i_k}\, dh_{i_1}\wedge\cdots\wedge dh_{i_k}$$

mit eindeutig bestimmten Funktionen $g_{i_1\dots i_k} \in \mathcal{A}^0 X$, und natürlich genügt es, die äußere Ableitung für einen einzelnen Summanden, etwa $g\, dh_{i_1}\wedge\cdots\wedge dh_{i_k}$ zu erklären. Aber ein Blick auf (b) und (c) zeigt, daß wir angesichts der nach (a) schon verfügten Bedeutung der Pfaffschen Form dg gar keine andere Wahl haben, als

$$d(g\, dh_{i_1}\wedge\cdots\wedge dh_{i_k}) = d(g\wedge dh_{i_1}\wedge\cdots\wedge dh_{i_k}) := dg\wedge dh_{i_1}\wedge\cdots\wedge dh_{i_k}$$

zu definieren, denn der nach der Produktregel (b) zu erwartende zweite Term

$$g\wedge d(dh_{i_1}\wedge\cdots\wedge dh_{i_k})$$

muß nach (b) und (c) verschwinden. Mit dieser Definition ist (a) offenbar erfüllt, und es bleiben noch (b) und (c) verifizieren. Bei der Produktregel (b) kann man sich natürlich ebenfalls auf Formen $\varphi = f\,dh_{i_1}\wedge\cdots\wedge dh_{i_k}$ und $\psi = g\,dh_{j_1}\wedge\cdots\wedge dh_{j_l}$ beschränken, und sie folgt letztlich daraus, daß sie für 0-Formen, also Funktionen gilt:

$$\begin{aligned}
d(\varphi\wedge\psi) &= d(f\,dh_{i_1}\wedge\cdots\wedge dh_{i_k}\wedge g\,dh_{j_1}\wedge\cdots\wedge dh_{j_l})\\
&= d(fg\,dh_{i_1}\wedge\cdots\wedge dh_{i_k}\wedge dh_{j_1}\wedge\cdots\wedge dh_{j_l})\\
&= (df\cdot g + f\cdot dg)\wedge dh_{i_1}\wedge\cdots\wedge dh_{i_k}\wedge dh_{j_1}\wedge\cdots\wedge dh_{j_l}\\
&= (df\wedge dh_{i_1}\wedge\cdots\wedge dh_{i_k})\wedge(g\,dh_{j_1}\wedge\cdots\wedge dh_{j_l}) + (-1)^k(f\,dh_{i_1}\wedge\cdots\wedge dh_{i_k})\wedge(dg\wedge dh_{j_1}\wedge\cdots\wedge dh_{j_l})\\
&= d\varphi\wedge\psi + (-1)^k\varphi\wedge d\psi
\end{aligned}$$

Die Eigenschaft (c) brauchen wir jetzt sogar nur noch für $k = 0$ zu beweisen, denn dann folgt

$$dd(\varphi) = dd(g\,dh_{i_1}\wedge\cdots\wedge dh_{i_k}) = d(dg\wedge dh_{i_1}\wedge\cdots\wedge dh_{i_k}) = 0$$

nach der schon bewiesenen Produktregel auch allgemein. Sei also $f\in\mathcal{A}^0X$ eine Funktion: es ist

$$\begin{aligned}
df &= \sum_{j=1}^{n}\frac{\partial f}{\partial h_j}\,dh_j\\
ddf &= \sum_{i,j=1}^{n}\frac{\partial^2 f}{\partial h_i\partial h_j}\,dh_i\wedge dh_j = 0
\end{aligned}$$

wegen der Vertauschbarkeit der partiellen Ableitungen (Satz 38.4) und $dh_i\wedge dh_j = -dh_j\wedge dh_i$.

Damit ist der Satz unter der genannten Zusatzannahme bewiesen. Im allgemeinen Fall wäre das Problem damit nur lokal gelöst, und es bliebe die Übertragung auf die ganze Mannigfaltigkeit X mittels mehrerer Karten zu leisten. Ein etwas subtiler Punkt ist dabei der Nachweis, daß die zu definierenden Cartanschen Differentiale zwangläufig *lokale* Operatoren sein müssen in dem Sinne, daß der Wert $(d\varphi)_x$ an einer Stelle $x\in X$ schon durch die Werte φ_y für alle y in der Nähe von x, nämlich alle y aus einer beliebig klein vorgebbaren x enthaltenden offenen Teilmenge von X festgelegt ist. Zum Vergleich: $\wedge$ und f^* sind nicht nur lokale, sondern sogar *punktweise* Operationen, denn um das Dachprodukt $(\varphi\wedge\psi)_x$ und bei gegebenem f die zurückgezogene Form $(f^*\psi)_x$ zu kennen, braucht man die Ausgangsdaten nur an einer einzigen Stelle, nämlich nur φ_x und ψ_x bzw. $\psi_{f(x)}$. Die Lokalität von d geht aus den Forderungen (a), (b) und (c) zwar nicht unmittelbar hervor, ist aber eine beweisbare Folgerung. Der Rest des Beweises stützt sich dann vor allem darauf, daß die zunächst lokal erklärte Cartansche Ableitung auch mit dem Zurückziehen von Formen in einfachster Weise verträglich ist. Das gilt dann natürlich auch global und verdient eine eigene Formulierung.

41.2 Lemma Ist $X\xrightarrow{f}Y$ eine differenzierbare Abbildung, so ist für jedes $k\in\mathbb{N}$ das Diagramm

$$\begin{array}{ccc}
\mathcal{A}^kY & \xrightarrow{\;d\;} & \mathcal{A}^{k+1}Y\\
\downarrow{\scriptstyle f^*} & & \downarrow{\scriptstyle f^*}\\
\mathcal{A}^kX & \xrightarrow{\;d\;} & \mathcal{A}^{k+1}X
\end{array}$$

kommutativ, es gilt also $f^*d\psi = d(f^*\psi)$ für jede Differentialform ψ auf Y.

Beweis Für eine 0-Form $g\in\mathcal{A}^0Y$ ergibt sich aufgrund der Kettenregel für Differentiale sowohl f^*dg als auch $d(f^*g)$ als die Komposition

$$\underbrace{T\,X\xrightarrow{Tf}TY\xrightarrow{Tg}T\,\mathbb{R}}_{T(g\circ f)} = \mathbb{R}\times\mathbb{R}\xrightarrow{\mathrm{pr}_2}\mathbb{R}.$$

Die Behauptung gilt also für 0-Formen, dann aber auch für jede exakte 1-Form $dg \in \mathcal{A}^1Y$, denn es ist

$$f^*d(dg) = 0$$

ebenso wie

$$d(f^*dg) = d\big(d(f^*g)\big) = 0.$$

Für Differentialformen höheren Grades rechnen wir lokal in Karten, jede beliebige Form schreibt sich dann als Summe von Produkten der schon behandelten Typen. Es genügt deshalb, folgendes zu beweisen: Gilt die behauptete Vertauschbarkeit von d und f^* bei Anwendung auf zwei Formen $\varphi \in \mathcal{A}^kY$ und $\psi \in \mathcal{A}^lY$, so gilt sich auch bei Anwendung auf $\varphi\wedge\psi$. Das aber rechnet man direkt nach:

$$\begin{aligned} f^*d(\varphi\wedge\psi) &= f^*\big(d\varphi\wedge\psi + (-1)^k\varphi\wedge d\psi\big) \\ &= f^*d\varphi \wedge f^*\psi + (-1)^k f^*\varphi \wedge f^*d\psi \\ &= d(f^*\varphi) \wedge f^*\psi + (-1)^k f^*\varphi \wedge d(f^*\psi) \\ &= d(f^*\varphi\wedge f^*\psi) \\ &= d\big(f^*(\varphi\wedge\psi)\big) \end{aligned}$$

Damit ist auch Lemma 41.2 bewiesen.

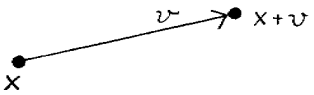
Die Cartansche Ableitung ist das zentrale Objekt der Vektoranalysis schlechthin, und man sollte deshalb auch eine gute anschauliche Vorstellung davon haben, was das d eigentlich macht. Aus der lokalen Formel

$$d(g\,dh_{i_1}\wedge\cdots\wedge dh_{i_k}) = dg\wedge dh_{i_1}\wedge\cdots\wedge dh_{i_k}$$

wissen wir jedenfalls, daß d ein Differentialoperator erster Ordnung ist; um $d\varphi$ an einer einzelnen Stelle zu verstehen, dürfen und werden wir deshalb annehmen, daß alle Koeffizienten von φ affin-linear sind. Das hat den Vorteil, daß die partiellen Ableitungen dann die Differenzenquotienten selbst und nicht erst deren Limites sind. (Die bekannte Argumentation mit infinitesimalen Größen läuft auf genau dasselbe hinaus.) Von affin-linearen Koeffizienten zu reden gibt auf einer ganz beliebigen Mannigfaltigkeit natürlich keinen Sinn. Um die Koordinatenunabhängigkeit des Folgenden zu betonen, wollen wir aber nicht gleich in $\mathbb{R}^n$ rechnen, sondern von Formen ausgehen, die auf einem abstrakten (aber natürlich endlichdimensionalen reellen) Vektorraum V definiert sind.

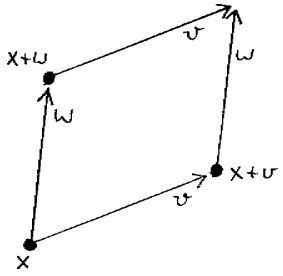
Die Wirkung von d auf eine 0-Form, also eine Funktion $f\colon V \longrightarrow \mathbb{R}$, ist uns längst vertraut: Jedem Tangentialvektor $v \in T_xV = V$ wird die Richtungsableitung von f längs v zugeordnet; weil f affin-linear ist, ist das hier einfach der Zuwachs von f längs v:

$$df(x) = f(x+v) - f(x)$$



Sei als nächstes $\varphi\colon TV \longrightarrow \mathbb{R}$ eine 1-Form. $d\varphi$ muß dann jedem Vektorpaar $(v,w) \in T_xV \times T_xV$ auf bilineare und alternierende Weise einen Wert zuweisen. Wollen wir einfach mal raten: Als ersten Versuch könnte man an die Änderung der Linearform φ längs v oder w denken, wobei der jeweils andere Vektor als Variable stehen bleibt; das hieße also

$$d\varphi_x(v,w) = \varphi_{x+v}(w) - \varphi_x(w) \quad \text{oder} \quad d\varphi_x(v,w) = \varphi_{x+w}(v) - \varphi_x(v).$$



In beiden Fällen wäre $d\varphi_x$ bilinear, aber nicht alternierend. Jedoch ist die Differenz zwangsläufig auch alternierend, und damit

$$d\varphi_x(v,w) = \varphi_{x+v}(w) - \varphi_x(w) - \varphi_{x+w}(v) + \varphi_x(v)$$

ein Kandidat für $d\varphi_x(v,w)$. Ob nun auch der richtige, das zeigt sich, wenn wir zwei linear unabhängige Vektoren v und w einsetzen. Dazu dürfen wir uns v und w als Teil einer Basis von V vorstellen; wenn wir die entsprechenden Vektoren der dualen Basis dh und dk nennen, schreibt sich φ in der Form

$$\varphi = f\,dh + g\,dk + \cdots$$

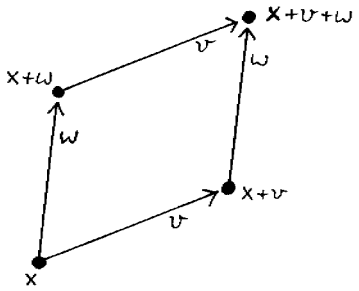
(die weiteren Terme brauchen wir nicht zu bezeichnen, weil sie auf v und w ohenehin keinen Beitrag liefern. Eine kleine Rechnung bestätigt dann unseren Ansatz:

$$\begin{aligned} d\varphi(v,w) &= (df\wedge dv + dg\wedge dw)(v,w) \\ &= df(v)\,dh(w) - df(w)\,dh(v) + dg(v)\,dk(w) - dg(w)\,dk(v) \\ &= dg(v) - df(w) \\ &= \big(\varphi_{x+v}(w) - \varphi_x(w)\big) - \big(\varphi_{x+w}(v) - \varphi_x(v)\big) \end{aligned}$$

Die Bedeutung des Ausdrucks für $d\varphi_x(v,w)$ ist vielleicht besser zu durchschauen, wenn man ihn in der Form

$$\big(\varphi_x(v) + \varphi_{x+v}(w)\big) - \big(\varphi_x(w) + \varphi_{x+w}(v)\big)$$

schreibt. Jeder der beiden eingeklammerten Terme steht für den Zuwachs von φ längs eines bestimmten Weges von x nach $x+v+w$: Er setzt sich im ersten Fall aus dem Zuwachs von φ_x längs v und dem von φ_{x+v} längs w zusammen, im zweiten aus den Zuwächsen von φ_x längs w und φ_{x+w} längs v.



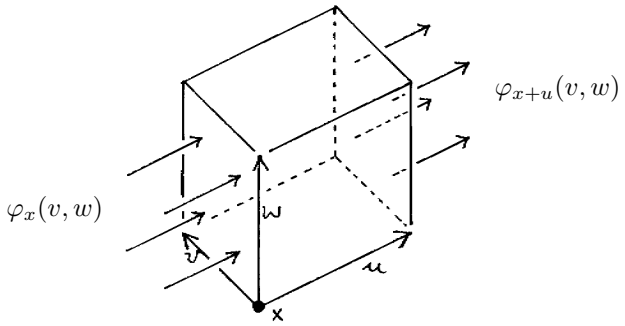
$d\varphi(v,w)$ mißt nun den Unterschied der beiden so berechneten Zuwächse von x nach $x+v+w$, und allgemein beschreibt $d\varphi$ daher die Wegabhängigkeit der nach diesem Prinzip längs gegenüberliegenden Parallelogrammseiten bestimmten Variation von φ, das, was Physiker und Ingenieure (infinitesimale) *Wirbel* des kovarianten Vektorfelds φ nennen.

Der Anschauung am besten zugänglich ist das Differential $d\varphi$ einer 2-Form auf einem dreidimensionalen Raum V (und allgemein einer $(n–1)$-Form auf einer n-dimensionalen Mannigfaltigkeit). Die bei den 1-Formen schon durchgeführte Überlegung zeigt, daß $d\varphi$ das Vektortripel (u,v,w) mit

$$d\varphi_x(u,v,w) = \big(\varphi_{x+u}(v,w) - \varphi_x(v,w)\big) + \big(\varphi_{x+v}(w,u) - \varphi_x(w,u)\big) + \big(\varphi_{x+w}(u,v) - \varphi_x(u,v)\big)$$

und allgemein das Vektor-n-tupel $(v_1,\dots,v_n)$ mit der Summe der analog gebildeten n Differenzen bewertet.

Interpretieren wir nun φ gemäß Beispiel 40.6(2) als Stromdichte, wobei (u,v,w) positiv orientiert sei! $\varphi_{x+u}(v,w)$ ist dann die Substanzmenge, die in der Zeiteinheit durch das Parallelogramm mit den Ecken $x+u$, $x+u+v$, $x+u+w$ und $x+u+v+w$ fließt, während $\varphi_x(v,w)$ dieselbe Bedeutung für das Parallelogramm mit den Ecken x, $x+v$, $x+w$ und $x+v+w$ hat.



Die Differenz mißt demnach, wieviel Substanz durch diese beiden gegenüberliegenden Seitenflächen zusammen aus dem von u, v und w aufgespannten Parallelepiped P *heraus*fließt, und gemeinsam mit den übrigen Termen ergibt sich die gesamte aus P in der Zeiteinheit ausfließende Substanzmenge. Damit wird klar, daß $d\varphi(v,w)$ eine Strömungsbilanz ist:

> $d\varphi$ bewertet das von u, v und w in T_xV aufgspannte orientierte (wenn man will, infinitesimale) Parallelepiped P mit der unter der Stromdichte φ pro Zeiteinheit aus P ausfließenden Substanzmenge.

Und diese Erklärung gilt praktisch wörtlich auch dann, wenn φ eine $n-1$-Form auf einer n-dimensionalen Mannigfaltigkeit ist: die $2n$ in $d\varphi_x(v_1,\dots,v_n)$ enthaltenen Terme messen dann die durch die $2n$ Seitenflächen aus P fließende Substanz.

Soweit zur anschaulichen Bedeutung der Cartanschen Ableitung. Nun werden Ihnen als angehenden Physikern die Differentialformen von kleinem Grad auf offenen Mengen $X \subset \mathbb{R}^3$ häufig nicht als solche, sondern in koordinatengebundener Schreibweise als die verschiedenen Typen von Vektorfeldern entgegentreten. Logischerweise wird dazu auch das Cartansche Differential nicht explizit als d serviert, sondern in einer der folgenden

41.3 Tarnpackungen für die **Cartansche Ableitung** *Grad* 0: Das Differential einer Funktion $f \in \mathcal{A}^0X$ ist die 1-Form

$$df = \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy + \frac{\partial f}{\partial z}\,dz,$$

wird also, wenn man Pfaffsche Formen über das euklidische Skalarprodukt mit Vektorfeldern identifiziert, zum *Gradientenfeld* $\operatorname{grad} f$ auf X, alternativ auch ∇f geschrieben. Der Operator d selbst wird so zu

$$\operatorname{grad}\colon f \longmapsto \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{pmatrix}.$$

Grad 1: In Koordinaten geschrieben, wird $\mathcal{A}^1X \xrightarrow{d} \mathcal{A}^2X$ zu einem Operator, der aus Vektorfeldern wieder Vektorfelder macht. Wegen

$$\begin{aligned} d(u\,dx + v\,dy + w\,dz) &= du\wedge dx + dv\wedge dy + dw\wedge dz \\ &= \left(\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z}\right) dy\wedge dz + \left(\frac{\partial u}{\partial z} - \frac{\partial w}{\partial x}\right) dz\wedge dx + \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) dx\wedge dy \end{aligned}$$

wird d in brutaler Koordinatenschreibweise zu der *Rotation* genannten Abbildung

$$\operatorname{rot}\colon \begin{pmatrix} u \\ v \\ w \end{pmatrix} \longmapsto \begin{pmatrix} \frac{\partial w}{\partial y} - \frac{\partial v}{\partial z} \\ \frac{\partial u}{\partial z} - \frac{\partial w}{\partial x} \\ \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \end{pmatrix}.$$

Weil man die Rotation auch als formales Vektorprodukt mit dem Gradientenoperator ansehen kann, ist auch die Schreibweise $\operatorname{rot} f = \nabla \times f$ gebräuchlich.

Grad 2: Das Differential $\mathcal{A}^2X \xrightarrow{d} \mathcal{A}^3X$ schließlich, mit

$$\begin{aligned} d(u\,dy\wedge dz + v\,dz\wedge dx + w\,dx\wedge dy) &= du\wedge dy\wedge dz + dv\wedge dz\wedge dx + dw\wedge dx\wedge dy \\ &= \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right) dx\wedge dy\wedge dz \end{aligned}$$

erscheint in Koordinaten als sogenannte *Divergenz*

$$\mathrm{div}\colon \begin{pmatrix} u \\ v \\ w \end{pmatrix} \longmapsto \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z},$$

deren Werte skalar sind. Auch hier hat man eine alternative Schreibweise, nämlich als formales Skalarprodukt $\mathrm{div}\, f = \nabla \bullet f$.

Wenn man den Vektorraum der C^∞-Vektorfelder auf X mit $\mathrm{Vect}X$ bezeichnet, hat man zusammengefaßt also für offene $X \subset \mathbb{R}^3$ das kommutative Diagramm

$$\begin{array}{ccccccc} \mathcal{A}^0X & \xrightarrow{d} & \mathcal{A}^1X & \xrightarrow{d} & \mathcal{A}^2X & \xrightarrow{d} & \mathcal{A}^3X \\ \downarrow\simeq & & \downarrow\simeq & & \downarrow\simeq & & \downarrow\simeq \\ C^\infty(X) & \xrightarrow{\mathrm{grad}} & \mathrm{Vect}X & \xrightarrow{\mathrm{rot}} & \mathrm{Vect}X & \xrightarrow{\mathrm{div}} & C^\infty(X) \end{array}$$

mit den im einzelnen schon angegebenen koordinatenabhängigen vertikalen Isomorphismen. (Bei genauerem Hinsehen merkt man, daß sie immerhin nur von der euklidischen Struktur und der Orientierung von $\mathbb{R}^3$ abhängen; man achte dazu auf die in der Tabelle 40.5 aufgeführten Transformationsformeln und die daran anschließende Diskussion.)

Natürlich kann man die dreidimensionale Vektoranalysis statt mit d auch mit diesen veralteten Bezeichnungen formulieren. Man muß dann aber eine ziemliche Schwerfälligkeit in Kauf nehmen, und eine solche Darstellung der an sich gar nicht schwierigen Gesetzmäßigkeiten der Vektoranalysis verkümmert leicht zu einer reinen Formelsammlung; davon können Sie sich durch einen Blick in eines der vielen einschlägigen Lehrbücher schnell überzeugen. Leider hängen die Physiker (ganz zu schweigen von den Ingenieuren) sehr an dem alten Kram; wenn er eines Tages von der Bildfläche verschwinden sollte, wird man ihm aber nicht nachtrauern müssen.

Gradient und Divergenz lassen sich immerhin auf beliebige Dimension übertragen; dagegen ist die Rotation eine besonders unglückliche und auf das Dreidimensionale beschränkte Bildung.

41.4 Beispiel In den klassischen Maxwellschen Gleichungen

$$\begin{aligned} \mathrm{rot}\vec{B} - \dot{\vec{E}} &= \vec{\jmath} & \mathrm{div}\vec{E} &= \rho \\ \mathrm{rot}\vec{E} + \dot{\vec{B}} &= 0 & \mathrm{div}\vec{B} &= 0 \end{aligned}$$

spielt die Rotation gewiß eine prominente Rolle. Nun ist die Elektrodynamik geradezu Paradigma einer relativistisch invarianten Theorie, und man sollte diese Gleichungen deswegen auch in einer invarianten Gestalt, also zumindest im vierdimensionalen Minkowski-Raum formulieren. Da ist die Rotation aber schon am Ende, während es in der modernen Sprache der Vektoranalysis ganz einfach ist: $\vec{E}$ und $\vec{B}$ werden zu einem elektromagnetischen Feldstärke*tensor* F verschmolzen, der mathematisch gesehen eine 2-Form auf $\mathbb{R}^4$ ist, während ρ und $\vec{\jmath}$ zur Viererstromdichte (Physikersprache) $J \in \mathcal{A}^3\mathbb{R}^4$ zusammengefaßt werden. Eine Auftrennung in die ein- und dreidimensionalen Anteile gibt wie immer erst Sinn, wenn ein bestimmtes Bezugssystem, zumindest eine Zerspaltung von $\mathbb{R}^4$ in Zeit und Raum gewählt ist. Sind t, x, y, z zugehörige Koordinaten, so ist die Zerlegung durch

$$F = -E_x\,dt\wedge dx - E_y\,dt\wedge dy - E_z\,dt\wedge dz + B_x\,dy\wedge dz + B_y\,dz\wedge dx + B_z\,dx\wedge dy$$

und

$$J = \rho\, dx\wedge dy\wedge dz - j_x\, dt\wedge dy\wedge dz - j_y\, dt\wedge dz\wedge dx - j_z\, dt\wedge dx\wedge dy$$

beschrieben. Die Maxwell-Gleichungen lauten dann schlicht und einfach

$$dF = 0 \quad \text{und} \quad d*F = J;$$

der in der zweiten Gleichung auftretende *Sternoperator* $\mathcal{A}^k X \xrightarrow{*} \mathcal{A}^{n-k}X$ wird in gleichem Sinne wie das Dachprodukt punktweise mittels linearer Algebra definiert, hängt aber anders als jenes von einer metrischen (euklidischen oder hier minkowskischen) Struktur ab, die die zugrundegelegte Mannigfaltigkeit X mitbringen muß.

Übrigens ist die aus $d*F = J$ wegen $dd = 0$ folgende *Kontinuitätsgleichung*

$$dJ = 0$$

der infinitesimale Erhaltungssatz für die elektrische Ladung: In Zeit- und Raumanteil zerlegt lautet sie nämlich $\dot\rho + d\,\vec\jmath = 0$, worin die Stromdichte jetzt als 2-Form $\vec\jmath \in \mathcal{A}^2\mathbb{R}^3$ und die Ladungsdichte als eine 3-Form $\rho \in \mathcal{A}^3\mathbb{R}^3$ aufgefaßt ist. Man erkennt, wie die aus einem infinitesimalen Volumen abfließende Ladung $d\vec\jmath$ durch die entgegengesetzte Änderung der Ladungsdichte ρ kompensiert wird.

Bemerkung Wenn man den Feldstärketensor F zu einem festen Zeitpunkt t auf den Raum-$\mathbb{R}^3$ (genauer also auf $\{t\}\times\mathbb{R}^3$) einschränkt, ist die verbleibende 2-Form

$$B_x\, dy\wedge dz + B_y\, dz\wedge dx + B_z\, dx\wedge dy \in \mathcal{A}^2\mathbb{R}^3$$

gerade die magnetische Induktion $B \in \mathcal{A}^2\mathbb{R}^3$ (axiales Vektorfeld!). Andererseits erhält man durch partielles Auswerten von F auf dem konstanten in Zeitrichtung weisenden Vektorfeld $\dfrac{\partial}{\partial t}$ die Pfaffsche Form

$$T_{(x,y,z)}\mathbb{R}^3 \ni v \longmapsto F_{(t,x,y,z)}\left(\frac{\partial}{\partial t}, v\right) \in \mathbb{R}$$

auf $\mathbb{R}^3$, die sich sofort zu

$$-E_x\, dx - E_y\, dy - E_z\, dz \in \mathcal{A}^1\mathbb{R}^3,$$

ergibt, bis aufs Vorzeichen also die elektrischen Feldstärke in kovarianter Schreibweise ist.

Wir führen die schon begonnene Verallgemeinerung der Definition 38.3 auf Differentialformen höherer Grade jetzt zu Ende und vereinbaren die folgende

41.5 Sprechweise Eine Differentialform $\psi \in \mathcal{A}^k X$ heißt geschlossen, wenn $d\psi = 0$ ist; für $k > 0$ heißt sie exakt, wenn es eine Form $\varphi \in \mathcal{A}^{k-1}X$ mit $d\varphi = \psi$ gibt.

Trivial ist die

41.6 Notiz Jede exakte Form ist geschlossen. (Insbesondere — in physikalischer Sprache: Jedes Vektorfeld mit skalarem Potential ist wirbelfrei, und jedes Feld, das ein Vektorpotential besitzt, ist quellenfrei.)

Der in der Folgerung 38.5 behandelte Fall einer Pfaffschen Form auf einer offenen Menge $X \subset \mathbb{R}^n$ ordnet sich hier als Spezialfall ein; die dort an die Form $\varphi = \varphi_1\, dx_1 + \cdots + \varphi_n\, dx_n$ gestellte Bedingung

$$\frac{\partial\varphi_j}{\partial x_i} = \frac{\partial\varphi_i}{\partial x_j} \quad \text{für alle } i \neq j$$

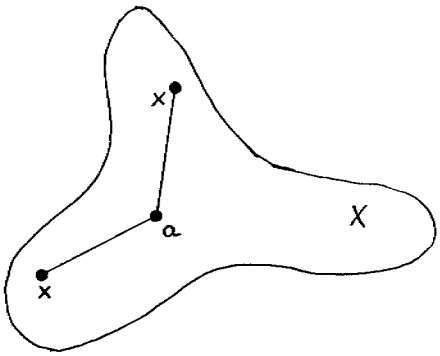
bedeutet ja gerade, daß

$$d\varphi = \sum_{i<j}\left(\frac{\partial\varphi_j}{\partial x_i} - \frac{\partial\varphi_i}{\partial x_j}\right) dx_i\wedge dx_j$$

die Nullform ist. Die Frage, ob umgekehrt eine geschlossene Form auch exakt ist, hatten wir damals ebenfalls schon angeschnitten. In der inzwischen entwickelten Sprache ist die Pfaffsche Form aus Beispiel 38.11, die auf

$\mathbb{R}^2\backslash\{0\}$ definiert ist und in Polarkoordinaten $d\varphi$ lautet, eine geschlossene, aber nicht exakte 1-Form. Der folgende Satz stützt die bei der Diskussion dieses Beispiels geäußerte Vorstellung, daß der Unterschied zwischen geschlossenen und exakten Differentialformen auf einer Mannigfaltigkeit X deren Geometrie wiederspiegelt. Zumindest stellt der Satz nämlich sicher, daß gewisse besonders einfach gebaute Mannigfaltigkeiten diesen Unterschied überhaupt nicht zeigen. Obwohl es sich vom Kaliber her um einen echten Satz handelt, spricht man traditionsgemäß vom

41.7 Poincaré-Lemma *$X \subset \mathbb{R}^n$ sei eine offene Menge, die bezüglich eines ihrer Punkte a sternförmig ist: für jedes $x \in X$ ist die ganze Verbindungsstrecke $\big\{(1-t)a + tx \,\big|\, t \in [0,1]\big\}$ in X enthalten.*



Dann ist für $k > 0$ jede geschlossene Form aus $\mathcal{A}^k X$ auch exakt.

Beweis Nach einer Verschiebung um a dürfen wir annehmen, daß X bezüglich $0 \in X$ sternförmig ist; so läßt sich das folgende einfacher hinschreiben. Wir erklären lineare Abbildungen

$$H\colon \mathcal{A}^k X \longrightarrow \mathcal{A}^{k-1} X \quad (k > 0)$$

durch die Formel

$$(H\psi)_x(v_2, \dots, v_k) = \int_0^1 t^{k-1} \psi_{tx}(x, v_2, \dots, v_k)\, dt \quad \text{für } \psi \in \mathcal{A}^k X.$$

Weil X sternförmig ist, ist der Integrand definiert und eine C^∞-Funktion, und nach Satz 32.11 ist deshalb das (über ein kompaktes Intervall gebildete) Integral eine C^∞-Funktion der verbleibenden Variablen. Außerdem ist $(H\psi)_x$ natürlich multilinear und alternierend in $v_2, \dots, v_n \in \mathbb{R}^n$.

Wir werden beweisen, daß $d \circ H + H \circ d$ die identische Abbildung von $\mathcal{A}^k X$ ist. Wozu? Weil daraus das Poincaré-Lemma sofort folgt; für jede geschlossene Form ψ ist dann nämlich

$$d(H\psi) = d(H\psi) + H(d\psi) = \psi.$$

Also los! Wir bemerken erst mal, daß H und d nicht nur linear, sondern auch mit Koodinatenvertauschungen verträglich sind: $h^* \circ H = H \circ h^*$ rechnet man sogar für *jedes lineare* $h\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$ ohne Mühe nach, und nach Lemma 41.2 gilt $h^* \circ d = d \circ h^*$ ohnehin für jedes differenzierbare h. Deshalb brauchen wir die gewünschte Identität $dH\psi + Hd\psi = \psi$ nur für die speziellen $\psi \in \mathcal{A}^k X$ der Form $\psi = g\, dx_1 \wedge \cdots \wedge dx_k$ zu beweisen. Wir schreiben dazu

$$\begin{aligned}
&(dx_1 \wedge dx_2 \wedge \cdots \wedge dx_k)(v_1, v_2, \dots, v_k) \\
&\quad = \det\Big((dx_i)(v_j)\Big)_{i,j=1}^{k} \\
&\quad = \sum_{r=1}^{k} (-1)^{r+1} (dx_r)(v_1) \cdot \det\Big((dx_i)(v_j)\Big)_{\substack{i \neq r \\ j \neq 1}} \\
&\quad = \sum_{r=1}^{k} (-1)^{r+1} (dx_r)(v_1) \cdot (dx_1 \wedge \cdots \wedge dx_{r-1} \wedge dx_{r+1} \wedge \cdots \wedge dx_k)(v_2, \dots, v_k)
\end{aligned}$$

(Entwicklung der Determinante aus Formel 39.8 nach der ersten Spalte). Damit wird

$$\begin{aligned}
&(H\psi)_x(v_2,\dots,v_k)\\
&\quad=\int_0^1 t^{k-1}g(tx)\cdot(dx_1\wedge dx_2\wedge\cdots\wedge dx_k)(x,v_2,\dots,v_k)\,dt\\
&\quad=\int_0^1 t^{k-1}g(tx)\cdot\sum_{r=1}^{k}(-1)^{r+1}(dx_r)(x)\cdot(dx_1\wedge\cdots\wedge dx_{r-1}\wedge dx_{r+1}\wedge\cdots\wedge dx_k)(v_2,\dots,v_k)\,dt\\
&\quad=\int_0^1 t^{k-1}g(tx)\,dt\cdot\sum_{r=1}^{k}(-1)^{r+1}x_r\cdot(dx_1\wedge\cdots\wedge dx_{r-1}\wedge dx_{r+1}\wedge\cdots\wedge dx_k)(v_2,\dots,v_k),
\end{aligned}$$

und wir erhalten eine Formel für $H\psi$, die ohne die Hilfsvektoren $v_2,\dots,v_k$ auskommt:

- $$H(g\,dx_1\wedge\cdots\wedge dx_k)_x=\int_0^1 t^{k-1}g(tx)\,dt\cdot\sum_{r=1}^{k}(-1)^{r+1}x_r\,dx_1\wedge\cdots\wedge dx_{r-1}\wedge dx_{r+1}\wedge\cdots\wedge dx_k.$$

Unter Verwendung von Satz 32.11 (Differenzieren unter dem Integralzeichen) berechnen wir nun

$$\begin{aligned}
d\left(\int_0^1 t^{k-1}g(tx)dt\cdot x_r\right)&=\sum_{j=1}^{n}\frac{\partial}{\partial x_j}\left(\int_0^1 t^{k-1}g(tx)dt\cdot x_r\right)dx_j\\
&=\sum_{j=1}^{n}\left(\int_0^1 t^{k}\frac{\partial g}{\partial x_j}(tx)dt\cdot x_r+\int_0^1 t^{k-1}g(tx)dt\cdot\delta_{jr}\right)dx_j\\
&=\sum_{j=1}^{n}\int_0^1 t^{k}\frac{\partial g}{\partial x_j}(tx)dt\cdot x_r\,dx_j+\int_0^1 t^{k-1}g(tx)dt\cdot dx_r
\end{aligned}$$

und damit

$$\begin{aligned}
(dH\psi)_x&=\sum_{r=1}^{k}(-1)^{r+1}d\left(\int_0^1 t^{k-1}g(tx)dt\cdot x_r\right)\wedge dx_1\wedge\cdots\wedge dx_{r-1}\wedge dx_{r+1}\wedge\cdots\wedge dx_k\\
&=\sum_{j=1}^{n}\sum_{r=1}^{k}(-1)^{r+1}\int_0^1 t^{k}\frac{\partial g}{\partial x_j}(tx)dt\cdot x_r\,dx_j\wedge dx_1\wedge\cdots\wedge dx_{r-1}\wedge dx_{r+1}\wedge\cdots\wedge dx_k\\
&\qquad+k\int_0^1 t^{k-1}g(tx)dt\cdot dx_1\wedge\cdots\wedge dx_k.
\end{aligned}$$

Andererseits ist

$$(d\psi)_x=\sum_{j=1}^{n}\frac{\partial g}{\partial x_j}(x)\,dx_j\wedge dx_1\wedge\cdots\wedge dx_k,$$

und die mit dem Punkt markierte Formel wird für $d\psi\in\mathcal{A}^k\mathbb{R}^n$ anstelle $\psi\in\mathcal{A}^{k-1}\mathbb{R}^n$ zu

$$(Hd\psi)_x=\sum_{j=1}^{n}\int_0^1 t^{k}\frac{\partial g}{\partial x_j}(tx)dt\cdot\Big(x_j\,dx_1\wedge\cdots\wedge dx_k+\sum_{r=1}^{k}(-1)^{r}x_r\,dx_j\wedge dx_1\wedge\cdots\wedge dx_{r-1}\wedge dx_{r+1}\wedge\cdots\wedge dx_k\Big).$$

Beim Aufaddieren fällt die Doppelsumme freundlicherweise heraus, und wir kommen mit

$$\begin{aligned}
(dH\psi+Hd\psi)_x&=\left(\int_0^1 kt^{k-1}g(tx)\,dt+\int_0^1 t^{k}\frac{d}{dt}g(tx)\,dt\right)dx_1\wedge\cdots\wedge dx_k\\
&=\left[t^k\cdot g(tx)\right]_{t=0}^{t=1}\cdot dx_1\wedge\cdots\wedge dx_k\\
&=g(x)\cdot dx_1\wedge\cdots\wedge dx_k\\
&=\psi_x
\end{aligned}$$

glücklich ins Ziel.

Bemerkungen Ich gebe zu, daß der Beweis nicht besonders erhellend ist; er ist dafür konkret und gibt eine Anleitung, wie man zu gegebener geschlossener Form ψ eine Stammform φ (also eine mit $d\varphi = \psi$) berechnen kann. — H trifft unter diesen Formen φ nur eine spezielle, stark von der konkreten Situation abhängige Wahl: zu φ kann man ja immer eine beliebige geschlossene (zum Beispiel exakte) Form addieren, ohne $d\varphi$ zu ändern. — H ist ein schönes Beispiel eines nicht-lokalen Operators: In $(H\psi)_x$ fließen die Werte von ψ_y für alle y aus der Verbindungsstrecke von 0 nach y ein.

Physiker gehen mit dem Poincaré-Lemma häufig sehr nachlässig um, indem sie die geometrische Voraussetzung über X einfach ignorieren: Jedes wirbelfreie Vektorfeld besitzt dann angeblich ein Potential, und jedes quellenfreie Feld ein Vektorpotential — dabei kommt es in Wirklichkeit doch noch darauf an, wo diese Vektorfelder definiert sind. Pedanterie der Mathematiker? Denken Sie an eine komplizierte elektrische Maschine mit ihren vielen Eisenkernen und Drahtwindungen, die ebenso komplizierte Löcher in das Feldgebiet fressen; dort spielt das durchaus eine praktische Rolle. Anders sieht es aus, wenn man nur an der lokalen Existenz von Potentialen interessiert ist:

41.8 Folgerung Ist X eine beliebige Mannigfaltigkeit, so ist für $k > 0$ jede geschlossene k-Form ψ auf X lokal exakt.

Erklärung und *Beweis* Gemeint ist, daß es um jeden Punkt $x \in X$ eine in X offene Menge $U \subset X$ und eine Form $\varphi \in \mathcal{A}^{k-1}U$ mit $d\varphi = \psi|U$ gibt. Zum Beweis wählt man eine Karte (U, h) um x und schrumpft dann U so, daß $h(U) \subset \mathbb{R}^n$ eine offene Kugel um 0 wird. Dann ist das Poincaré-Lemma auf $h(U)$, also auch auf U selbst anwendbar, und die Behauptung folgt.

Es sei nochmal betont, daß die lokalen Stammformen, die die Folgerung 41.8 zu einer geschlossenen Form liefert, sich im allgemeinen nicht zu einer globalen Stammform zusammenfügen lassen (Beispiel 38.11 belegt das).

Bemerkung Wie schon erwähnt, sehen Mathematiker das Ganze gern unter einem anderen Blickwinkel: Die Abweichung zwischen exakten und geschlossenen Differentialformen auf einer Mannigfaltigkeit X ist ein Maß für die geometrische Kompliziertheit von X; das erlaubt es, Mannigfaltigkeiten mit wesentlich verschiedener Geometrie voneinander zu unterscheiden, ja die Geometrie von Mannigfaltigkeiten in gewisser Weise geradezu zu berechnen. Das Poincaré-Lemma ist ein erster Schritt in diese Richtung: Wenn man auf X eine geschlossene, aber nicht exakte Differentialform findet, dann kann X jedenfalls nicht zu einer sternförmigen offenen Teilmenge von $\mathbb{R}^n$ diffeomorph sein.

Übungsaufgaben

41.1 $\varphi \in \mathcal{A}^k X$ und $\psi \in \mathcal{A}^l X$ seien Differentialformen auf einer Mannigfaltigkeit X. Beweisen Sie:

- Sind φ und ψ geschlossen, so ist auch $\varphi \wedge \psi$ geschlossen.
- Wenn φ geschlossen und ψ exakt ist, dann ist $\varphi \wedge \psi$ exakt.

41.2 In einer Aufgabe aus K. Meyberg, P. Vachenauer: *Höhere Mathematik 1* zum Thema soll der Leser sich wundern, daß für die Divergenz eines Spatproduktes keine der (von Rechenregeln mit Skalar- und Kreuzprodukt suggerierten) Formeln $\nabla \bullet (\mathbf{f} \times \mathbf{g}) = \mathbf{g} \bullet (\nabla \times \mathbf{f})$ oder $\nabla \bullet (\mathbf{f} \times \mathbf{g}) = -\mathbf{f} \bullet (\nabla \times \mathbf{g})$ gilt, sondern eine andere.

Helfen Sie dem Leser, indem Sie die relevante Formel aus dem Differentialformenkalkül in die altmodische Sprache übersetzen.

41.3 Im Nullpunkt von $\mathbb{R}^3$ befinde sich eine Punktladung q; das von ihr in $X := \mathbb{R}^3\backslash\{0\}$ erzeugte elektrische Potential schreibt sich in Kugelkoordinaten bekanntlich ganz einfach:

$$U = \frac{q}{r}$$

Berechnen Sie das zugehörige Feld E auf geschickte Art, indem Sie E nämlich als 1-Form $E \in \mathcal{A}^1X$ auffassen, die Koordinatenunabhängigkeit der Cartanschen Ableitung benutzen und erst zum Schluß auf kartesische Koordinaten umrechnen. Vergleichen Sie das Ergebnis mit dem, was Sie in der Physik gelernt haben.

41.4 Zwei Punktladungen entgegengesetzter Größe $\pm q/2\lambda$ an den Stellen $(0, 0, \pm\lambda) \in \mathbb{R}^3$ erzeugen natürlich ein komplizierteres Potential U_λ. Für $\lambda \to 0$ konvergiert U_λ aber gegen eine wieder recht einfache Funktion $U \in \mathcal{A}^0X$ (elektrisches Dipolpotential). Welches? Recht bequem ist es, dabei Zylinderkoordinaten (r, φ, z) zu verwenden und sich der Taylor-Reihen, insbesondere der binomischen Reihen 16.5(3) zu erinnern, konkret nämlich U_λ als Potenzreihe in λ zu schreiben. Analog zur vorigen Aufgabe können Sie dann das zugehörige Dipolfeld leicht aus- und nach Wunsch auf Kugel- oder kartesische Koordinaten umrechnen.

41.5 Überzeugen Sie sich davon, daß die Gleichung $dF = 0$ mit dem in Beispiel 41.4 definierten Feldstärketensor F bei Zerspaltung in Zeit und Raum tatsächlich die Gruppe der homogenen Maxwellschen Gleichungen in der klassischen Form liefert. Wie wird in diesem Fall der in der Vorlesung erwähnte Sternoperator wirken, damit $d*F = J$ die Bedeutung der restlichen Maxwell-Gleichungen bekommt?

41.6 Zeigen Sie, daß

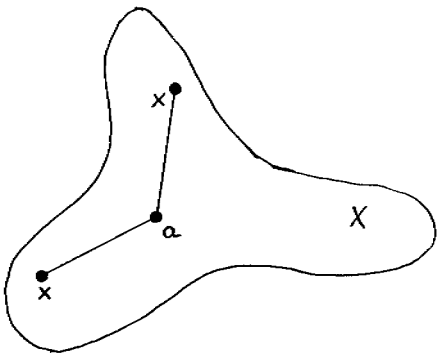
$$\psi = 2xz\,dy\wedge dz - dz\wedge dx - (e^x + z^2)\,dx\wedge dy$$

eine geschlossene 2-Form auf $\mathbb{R}^3$ ist. Verwenden Sie den zum Beweis des Poincaré-Lemmas konstruierten Operator $H\colon \mathcal{A}^{"}X \longrightarrow \mathcal{A}^1X$ dazu, eine Stammform von ψ konkret auszurechnen.

Anhang

Wenn man den Begriff der Differentialform nicht über den der Pfaffschen Form hinaus kennt, kann man vom Poincaré-Lemma 41.7 nur den Anfang formulieren und beweisen:

Poincaré-Lemma $X \subset \mathbb{R}^n$ sei eine offene Menge, die bezüglich eines ihrer Punkte a sternförmig ist: für jedes $x \in X$ ist die ganze Verbindungsstrecke $\big\{(1-t)a + tx \,\big|\, t \in [0,1]\big\}$ in X enthalten.



(Bekanntere Spezialfälle sternförmiger Mengen sind konvexe Mengen X, nämlich solche, die bezüglich eines jeden ihrer Punkte sternförmig sind, das heißt mit je zwei Punkten $a, x \in X$ die Verbindungsstrecke von a nach x umfassen. Dazu zählen offenbar Kugeln und Quader aller Art.)

Sei nun

$$\psi = \sum_{j=1}^{n} g_j \, dx_j$$

eine C^1-Form auf X mit

$$\frac{\partial g_i}{\partial x_j} = \frac{\partial g_j}{\partial x_i} \quad \text{für alle } i \neq j.$$

Dann ist ψ exakt.

Beweis Nach einer Verschiebung um a dürfen wir annehmen, daß X bezüglich $0 \in X$ sternförmig ist; so läßt sich eine Stammfunktion von ψ einfacher hinschreiben, nämlich als

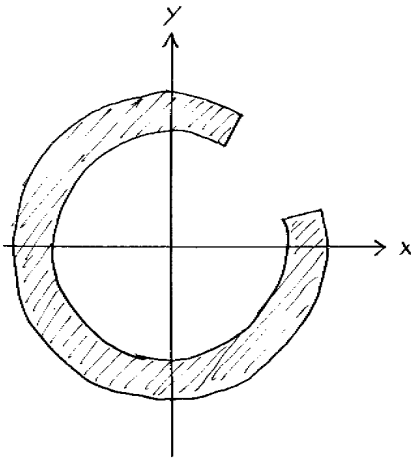
$$f(x) := \sum_{i=1}^{n} \int_0^1 g_i(tx) \, dt \cdot x_i \quad \text{für alle } x \in X.$$

Satz 32.10 über Integrale mit Parametern schließt problemlos auch mehrdimensionale Differenzierbarkeit ein — wir konnten das damals bloß noch nicht formulieren. Hier garantiert uns schon die vereinfachte Version 32.11, daß f eine C^1-Funktion ist, deren Differential wir durch Differenzieren nach x unter dem Integralzeichen berechnen können. Das tun wir auch und rechnen tapfer:

$$\begin{aligned}
df(x) &= \sum_{i,j=1}^{n} \int_0^1 \frac{\partial g_i}{\partial x_j}(tx) \, t \, dt \cdot x_i \, dx_j + \int_0^1 g_j(tx) \, dt \cdot dx_j \\
&= \sum_j \int_0^1 \left(\sum_i \frac{\partial g_j}{\partial x_i}(tx) \, t x_i + g_j(tx) \right) dt \cdot dx_j \\
&= \sum_j \int_0^1 \frac{d}{dt} \Big(t \cdot g_j(tx) \Big) dt \cdot dx_j \\
&= \sum_j \Big[t \cdot g_j(tx) \Big]_{t=0}^{1} \cdot dx_j \\
&= \sum_j g_j(x) \, dx_j = \psi(x)
\end{aligned}$$

42 Differentialformen und Integral

Wir wenden uns jetzt noch einmal der Integralrechnung zu, und zwar speziell der mehrdimensionalen. Unsere bisher einzige Möglichkeit, ein mehrdimensionales Integral auszuwerten, besteht darin, es nach dem Satz von Fubini in ein mehrfaches Integral zu verwandeln und dieses durch sukzessives Integrieren in einer Variablen zu berechnen. Dieses Verfahren wird nicht immer Freude bereiten:



Ist eine Funktion etwa über das schraffierte Gebiet zu integrieren und dort in Polarkoordinaten gegeben, so würde man schon einiges dafür geben, ganz in Polarkoordinaten rechnen zu dürfen! Was uns dazu fehlt, ist eine mehrdimensionale Substitutionsregel. Hier ist sie:

42.1 Satz (Integraltransformationsformel) Sei $U \subset \mathbb{R}^n$ eine offene Menge und

$$U \xrightarrow{\Phi} \Phi(U)$$

ein C^1-Diffeomorphismus auf die offene Teilmenge $\Phi(U) \subset \mathbb{R}^n$. Ist dann $f\colon \Phi(U) \longrightarrow \mathbb{R}$ eine über $\Phi(U)$ integrierbare Funktion, so ist die Funktion

$$(f\circ\Phi)\cdot|\det D\Phi|\colon U \longrightarrow \mathbb{R}$$

über U integrierbar, und

$$\int_{\Phi(U)} f = \int_U (f\circ\Phi)\cdot|\det D\Phi|.$$

Bemerkungen Aus der Existenz von $\int_U (f\circ\Phi)\cdot|\det D\Phi|$ kann man umgekehrt auf die von $\int_{\Phi(U)} f$ schließen, denn wenn man $g\colon U \longrightarrow \mathbb{R}$ durch

$$g := (f\circ\Phi)\cdot|\det D\Phi|$$

definiert, dann ist

$$(g\circ\Phi^{-1})\cdot|\det D\Phi^{-1}| = (g\circ\Phi^{-1})\cdot\frac{1}{|\det D\Phi|\circ\Phi^{-1}} = f,$$

und man liest Satz 42.1 in umgekehrter Richtung. — Häufig möchte man Funktionen f integrieren, die nur auf einer (nicht notwendig offenen) Teilmenge $X \subset U$ definiert sind. Die Transformationsformel ist dann eben auf die künstlich auf X (sogar ganz $\mathbb{R}^n$) erweiterte Funktion f_X anzuwenden:

$$\int_{\Phi(X)} f = \int_X (F\circ\Phi)\cdot|\det D\Phi|$$

Am leichtesten verstehen läßt sich die Transformationsformel an folgendem Speziallfall:

42.2 Folgerung (Maßtransformationsformel) Sei $U \subset \mathbb{R}^n$ eine offene Menge und

$$U \xrightarrow{\Phi} \Phi(U)$$

ein C^1-Diffeomorphismus auf eine offene Teilmenge $\Phi(U) \subset \mathbb{R}^n$. Außerdem sei $X \subset U$ derart, daß $\Phi(X)$ meßbar ist. Unter diesen Voraussetzungen ist das Maß $\mu\big(\Phi(X)\big)$ genau dann endlich, wenn die Funktion $|\det D\Phi| \colon U \longrightarrow \mathbb{R}$ über U integrierbar ist, und in diesem Fall ist

$$\mu\big(\Phi(X)\big) = \int_X |\det D\Phi(x)|.$$

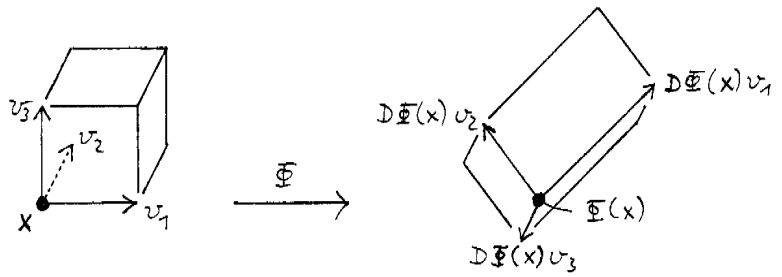
Beweis Daß $\Phi(X)$ meßbar von endlichem Maß ist, bedeutet, daß die Funktion $1_{\Phi(X)}$ integrierbar ist, und nach Satz 42.1 folgt daraus die Integrierbarkeit von $|\det D\Phi|$ über X ebenso wie die Formel. Meßbare Mengen $\Phi(X)$ von unendlichem Maß schöpft man wie immer durch ihre Schnitte mit großen Würfeln oder Kugeln aus; aus der Monotonie des Integrals folgt dann, daß $|\det D\Phi|$ in diesem Fall nicht integrierbar ist.

Bemerkung Die etwas umständlich zu lesenden Voraussetzungen ergeben sich so durch Spezialisierung aus 42.1. Tatsächlich ist bei gegebenem Φ die Menge $\Phi(X)$ genau dann meßbar, wenn $X \subset U$ selbst meßbar ist. Beachten Sie aber: Auch wenn $\Phi(X)$ endliches Maß hat, kann $\mu(X) = \infty$ sein (und umgekehrt); man betrachte etwa

$$U = (0, \infty) = \Phi(U); \quad \Phi(x) = \frac{1}{x}$$

und $X = (1, \infty)$, also $\Phi(X) = (0, 1)$.

Die Maßtransformationsformel besagt, daß das Maß bei der Abbildung von U nach $\Phi(U)$ eine "Verzerrung" erleidet, und zwar um den (ortsabhängigen) überall positiven Faktor $|\det D\Phi|$. Das deckt sich gut mit unserer anschaulichen Vorstellung vom Determinantenbetrag als einem Volumen: Der Diffeomorphismus Φ sendet jeden an einer Stelle $x \in U$ sitzenden infinitesimalen Würfel auf ein infinitesimales Parallelepiped, und die Zahl $|\det D\Phi(x)|$ gibt an, um welchen Faktor sich das Volumen dabei verändert.



$|\det(v_1, v_2, v_3)|$ $\qquad$ $|\det D\Phi(x)| \cdot |\det(v_1, v_2, v_3)|$

Bei der wirklichen (nicht infinitesimalen) Volumenmessung durch ein Integral geht dieser Faktor deshalb als Integrand ein. Tatsächlich bildet die Transformationsformel auch die Brücke zwischen der "naiven" Maßtheorie der Parallelepipede, auf die sich unser Verständnis der Determinante gründet, und der analytischen Maßtheorie, die man braucht, wenn kompliziertere Mengen zu messen sind. — Weil Satz 42.1 nicht leicht zu beweisen, vielmehr von ähnlichem Kaliber wie der Satz von Fubini ist, begnüge ich mit anstelle eines Beweises mit diesen erläuternden Worten und einigen Hinweisen zum Umgang mit der Formel.

Zu den praktisch wichtigen Diffeomorphismen Φ zählen die Abbildungen, die kartesische Koordinaten in Polar-, Kugel-, Zylinder- oder anderen einer konkreten Situation angepaßten Koordinaten ausdrücken. Die etwa bei ebenen Polarkoordinaten sonst so lästige Tatsache, daß als Definitionsintervall für die Winkelvariable das kompakte Intervall $[0, 2\pi]$ zu groß, das offene $(0, 2\pi)$ aber eigentlich zu klein ist, wird hier belanglos. Denn man darf den Integrationsbereich beider beteiligter Integrale schadlos um Nullmengen vergrößern oder verkleinern, insbesondere $\Phi \colon (r, \varphi) \mapsto (r\cos\varphi, r\sin\varphi)$ einerseits zu einem Diffeomorphismus

$$(0, \infty) \times (0, 2\pi) \longrightarrow \mathbb{R}^2 \setminus \big([0, \infty) \times \{0\}\big)$$

einschränken (um Satz 42.1 anwenden zu können), andererseits bei Bedarf davon Gebrauch machen, daß Φ auch auf $[0,\infty)\times[0,2\pi]$ (ja sogar ganz $\mathbb{R}^2$) noch als differenzierbare Abbildung erklärt ist.

Wie schon beim Satz von Fubini ist die typische Situation die, daß die Existenz des interessierenden Integrals zunächst fraglich ist und mit der Auswertung erst etabliert werden muß. Auch hier wird man dann auf die Ausschöpfungsmethode von Satz 32.7 zurückgreifen, meist also damit anfangen, den Betrag eines stetigen Integranden über kompakte Teilbereiche zu integrieren.

42.3 Beispiele (1) "Die in Kugelkoordinaten durch

$$f(r,\theta,\varphi)=\frac{e^{-r}}{r^2\sin\theta}$$

gegebene Funktion f soll über den Raum integriert werden." Wenn Sie das in einem Physikbuch finden, ist natürlich nicht an $\int f(r,\theta,\varphi)d(r,\theta,\varphi)$ gedacht. Vielmehr ist der Integrand $f=f(x,y,z)$ eigentlich eine Funktion der kartesischen Koordinaten, die aber längs der z-Achse nicht definiert ist; mathematisch präzise, gewiß auch schwerfälliger, müßte man sich also etwa so ausdrücken: Mittels der Kugelkoordinatenabbildung $\Phi\colon(r,\theta,\varphi)\mapsto(x,y,z)$ ist durch $f\circ\Phi\colon(r,\theta,\varphi)\longmapsto\frac{e^{-r}}{r^2\sin\theta}$ eine Funktion f auf $X:=\mathbb{R}^3\setminus\big(\{0\}\times\{0\}\times\mathbb{R}\big)$ definiert, und gefragt ist nach $\int_X f$ — wobei es ebenso akzeptabel wäre, gleich $\int_{\mathbb{R}^3} f$ zu schreiben.

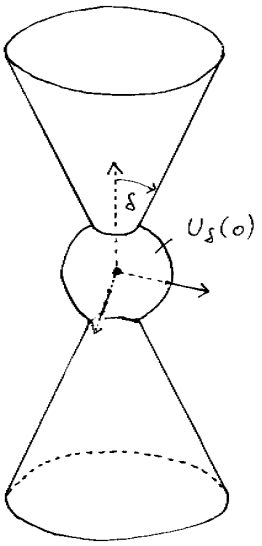
Es liegt auf der Hand, auch die Auswertung des Integrals in Kugelkoordinaten anzupeilen. Nun ist f zwar auf X stetig, aber X nicht kompakt; man integriert f deshalb zuerst nur über die kompakten Teilmengen $X_\delta\subset X$, die man exakt

$$X_\delta:=\Phi\left(\left[\delta,\tfrac{1}{\delta}\right]\times[\delta,\pi-\delta]\times[0,2\pi]\right),$$

aber prägnanter und wohl unmißverständlich

$$X_\delta:=\left\{(x,y,z)\in X\;\middle|\;\delta\le r\le\tfrac{1}{\delta},\ \delta\le\theta\le\pi-\delta\right\}$$

schreibt; aus der kompakten Kugel $D_{1/\delta}$ wird also die skizzierte offene Menge entfernt:



X_δ ist tatsächlich kompakt: Das folgt nach Satz 30.10 daraus, daß Φ sich als stetige auf $\left[\delta,\frac{1}{\delta}\right]\times[\delta,\pi-\delta]\times[0,2\pi]$ definierte Abbildung auffassen läßt: als solche ist Φ zwar nicht injektiv und schon gar kein Diffeomorphismus, aber darauf kommt es hier auch nicht an. Jedenfalls ist so die Integrierbarkeit der stetigen Funktion f über X_δ sichergestellt. Erst für den nächsten Schritt, nämlich die Anwendung der Transformationsformel 42.1, denken wir uns Φ auf die kleinere und offene Menge $\left(\delta,\frac{1}{\delta}\right)\times(\delta,\pi-\delta)\times(0,2\pi)$ eingeschränkt; dadurch wird Φ zu einem Diffeomorphismus, und an der Integrierbarkeit ändert das Weglassen der Ränder ebensowenig wie an dem Wert der Integrale.

Die Jacobi-Determinante von Φ ergibt sich nach Beispiel 34.9 zu $\det D\Phi(r,\varphi,\theta) = r^2\sin\theta$, also ist nach der Transformationsformel und dem Satz von Fubini

$$\begin{aligned}\int_{X_\delta} f(x,y,z)d(x,y,z) &= \int_{[\delta,\frac{1}{\delta}]\times[\delta,\pi-\delta]\times[0,2\pi]} \frac{e^{-r}}{r^2\sin\theta}\,|r^2\sin\theta|\,d(r,\theta,\varphi)\\ &= \int_\delta^{1/\delta}\int_\delta^{\pi-\delta}\int_0^{2\pi} e^{-r}\,dr\,d\theta\,d\varphi\\ &= \int_\delta^{1/\delta} e^{-r}dr\cdot\int_\delta^{\pi-\delta} d\theta\cdot\int_0^{2\pi} d\varphi\\ &= \big[-e^{-r}\big]_{r=\delta}^{r=1/\delta}\cdot(\pi-2\delta)\cdot 2\pi\\ &= \big(e^{-\delta}-e^{-1/\delta}\big)\cdot(\pi-2\delta)\cdot 2\pi.\end{aligned}$$

Wegen $\bigcup_{\delta>0} X_\delta = X$ und weil f überall positiv ist, folgen nach Satz 33.7 Existenz und Wert von

$$\int_X f = \lim_{\delta\to 0}\int_{X_\delta} f = (1-0)\cdot\pi\cdot 2\pi = 2\pi^2.$$

(2) Ändert man unter Beibehaltung der übrigen Vorgaben f zu

$$f(r,\theta,\varphi) = \frac{e^{-r}\cos\theta}{r^2(\sin\theta)^2}$$

ab, so wird

$$\begin{aligned}\int_{X_\delta} f &= \int_\delta^{1/\delta}\int_\delta^{\pi-\delta}\int_0^{2\pi} \frac{e^{-r}\cos\theta}{\sin\theta}\,dr\,d\theta\,d\varphi\\ &= \int_\delta^{1/\delta} e^{-r}dr\cdot\int_\delta^{\pi-\delta}\cot\theta\,d\theta\cdot\int_0^{2\pi} d\varphi\\ &= 0\end{aligned}$$

wegen der Symmetrie des Cotangens. Aber über die Integrierbarkeit von f sagt das gar nichts, denn man muß ja zuerst $|f|$, d.h. f_+ und f_- einzeln anschauen. Das bedeutet hier, das θ-Integral getrennt über $[\delta,\pi/2]$ und über $[\pi/2,\pi-\delta]$ auswerten. Man erhält $\int_\delta^{\pi/2} = \big[\log\sin\theta\big]_{\theta=\delta}^{\theta=\pi/2}$ und folglich zu $\int_{X_\delta} f$ den Beitrag

$$\big(e^{-\delta}-e^{-1/\delta}\big)\cdot(-\log\sin\delta)\cdot 2\pi$$

mit Limes ∞: die Funktion f ist deshalb nicht integrierbar.

Die Integraltransformationsformel ist auch aus theoretischen Gründen sehr wichtig. Der Determinantenfaktor in der Formel erinnert ja an das Transformationsverhalten einer n-Form auf $U\subset\mathbb{R}^n$ unter differenzierbarem Kartenwechsel. Deutet das vielleicht darauf hin, daß man statt Funktionen auch n-Formen integrieren kann? Das wäre aus der Sicht physikalischer Anwendungen sogar besonders natürlich, denn n-Formen stehen in der Regel für Dichten, und das typische Raumintegral der Physik drückt eine globale Größe wie die Ladung durch Integration einer Dichte (wie der Ladungsdichte) über einen Raumbereich aus. Mathematisch gesehen bringt dieser Standpunkt auf den ersten Blick kaum einen Unterschied:

42.4 Definition Sei $U\subset\mathbb{R}^n$ offen und φ eine n-Form auf U, also

$$\varphi = f\,dx_1\wedge dx_2\wedge\cdots\wedge dx_n$$

mit einer eindeutig bestimmten Funktion $f\colon U\longrightarrow\mathbb{R}$. Für eine Teilmenge $X\subset U$ wird das Integral $\int_X\varphi$ dann als

$$\int_X\varphi := \int_X f$$

definiert, falls dieses existiert.

Was nun, wenn φ statt in der natürlichen in einer anderen Karte (U,h) vorliegt, von der wir nur voraussetzen wollen, daß sie orientierungstreu ist? Wir haben dann also

$$\varphi = (g\circ h)\,dh_1\wedge dh_2\wedge\cdots\wedge dh_n$$

mit einer neuen Koeffizientenfunktion $g\colon h(U)\longrightarrow\mathbb{R}$ und müssen zur Berechnung von $\int_X\varphi$ das f der obigen Darstellung ausrechnen. Wie, das haben wir uns im Anschluß an die Definition 40.3 überlegt: Weil es sich um Formen vom höchstmöglichen Grad handelt, ist einfach

$$dh_1\wedge dh_2\wedge\cdots\wedge dh_n = \det Dh\cdot dx_1\wedge dx_2\wedge\cdots\wedge dx_n$$

und deshalb

$$f = (g\circ h)\cdot\det Dh.$$

Wegen der Orientierungstreue von h ist die Determinante durchweg positiv, die Transformationsformel 42.1 sagt uns also:

$$\int_X\varphi = \int_X f = \int_X(g\circ h)\cdot\det Dh = \int_{h(X)} g$$

Wir brauchen in Wirklichkeit also gar nichts umzurechnen, dürfen vielmehr die Definition 42.4 von vornherein in einer beliebigen orientierungstreuen Karte für U lesen (oder einer orientierungsumkehrenden, und dann das Vorzeichen ändern)! Wer mit der in den vorangehenden Abschnitten entwickelten Denkweise einigermaßen vertraut geworden ist, wird sofort erkennen, welche Möglichkeiten sich damit auftun, und daß es auch aus mathematischer Sicht natürlicher ist, n-Formen und nicht Funktionen zu integrieren.

Sei nämlich X eine orientierte n-dimensionale Mannigfaltigkeit, und φ eine n-Form auf X. Wenn es eine orientierungtreue Karte (U,h) für X gibt, so daß φ außerhalb von U identisch verschwindet, dann können wir $\int_X\varphi$ nach dem eben beschriebenen Prozeß, in einer Formel also durch

$$\int_X\varphi := \int_{h(U)}(h^{-1})^*\varphi$$

definieren. Von der konkreten Wahl der Karte hängt das so definierte Integral nicht ab. Freilich werden die meisten interessanten Differentialformen uns nicht den Gefallen tun, nur innerhalb des Definitionsbereichs einer einzigen Karte zu leben. Hier hilft aber eine wichtige, *Zerlegung der Eins* genannte Methode weiter, für deren genaue Beschreibung ich Sie auf spezielle Bücher oder Vorlesungen über Vektoranalysis verweisen muß. Grob gesagt erlaubt sie es, beliebige lineare Objekte auf einer Mannigfaltigkeit, zum Beispiel eben Differentialformen, in eine Summe von solchen Objekten zu zerlegen, die außerhalb von vorgeschrieben kleinen Teilmengen von X identisch verschwinden. Das geschieht in einer solchen Weise, daß eventuelle Stetigkeits- und Differenzierbarkeitseigenschaften auf die einzelnen Summanden vererbt werden, und so, daß die Summe entweder endlich oder doch zumindest *lokal endlich* ist, d.h. beim Aufaddieren lokal immer nur endlich viele Summanden wirksam sind. Vorgeschrieben klein? Hat man zum Beispiel die gesamte Mannigfaltigkeit X mit einer Kollektion $(U_\lambda,h_\lambda)_\lambda$ von Karten beschrieben $(\bigcup_\lambda U_\lambda = X)$, so darf man unter "vorgeschrieben klein" dasselbe verstehen wie "in mindestens einem der U_λ ganz enthalten". Mit dieser Technik kann man den Integrierbarkeits- und Integralbegriff ohne weiteres auf beliebige Differentialformen auf X ausdehnen; und es ergibt sich insbesondere als a-priori-Aussage, daß bei kompaktem X jede stetige n-Form auf X integrierbar ist.

Für offene $X\subset\mathbb{R}^n$ bringt dieser erweiterte Integralbegriff zunächst auf formaler Ebene die neue Schreibweise

$$\int_X f(x)\,dx_1\wedge dx_2\wedge\cdots\wedge dx_n$$

für $\int_X f$. Eine inhaltliche Neuerung ist, daß das Integral nun auf eine Orientierung der Grundmannigfaltigkeit Bezug nimmt und wir mit

$$\int_{-X} f(x)\,dx_1\wedge\cdots\wedge dx_n = -\int_X f(x)\,dx_1\wedge\cdots\wedge dx_n$$

auch über negativ orientierte offene $X \subset \mathbb{R}^n$ integrieren können: das hatten wir im Eindimensionalen mit der Schreibweise $\int_a^b = -\int_b^a$ ja schon getan und praktisch gefunden. Achten Sie darauf, daß zum Beispiel

$$\int_{\mathbb{R}^n} f(x)\, dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n = -\int_{\mathbb{R}^n} dx_2 \wedge dx_1 \wedge dx_3 \wedge \cdots \wedge dx_n$$

ist, während bei dem alten, unorientierten Integral die Reihenfolge der Koordinaten ganz egal war. Beim orientierten Integral entfällt mit der Jacobi-Determinante auch der Absolutbetrag in der Transformationsformel (er war in der eindimensionalen Substitutionsregel ja auch nicht vorhanden), und die Formel lautet stattdessen

$$\int_Y \psi = \int_X \Phi^*\psi \quad \text{für jeden orientierungstreuen Diffeomorphismus } X \xrightarrow{\Phi} Y.$$

Letztlich an der neuen Auffassung vom Integral mag es auch liegen, daß man die Transformationsformel oft anwendet, ohne sich dessen bewußt zu sein. Denn etwa wenn man das Integral $\int_{\mathbb{R}}^2 f(x,y)\, dx \wedge dy$ in Polarkoordianten auswertet und dazu neben $x = r\cos\varphi,\ y = r\sin\varphi$ auch $dx \wedge dy = r\, dr \wedge d\varphi$ substituiert, ist das ja mehr als die im Beispiel 31.14(2) eingeführte symbolische Schreibweise, nämlich die sinnvolle und korrekte Berechnung der mittels der Polarkoordinatenabbildung zurückgezogenen 2-Form $f\, dx \wedge dy$, und es erübrigt sich, sich ausdrücklich auf die Transformationsformel zu berufen.

Wie alle bisher eingeführten Integralbegriffe geht der des Kurvenintegrals aus Definition 38.7 in dem jetzt betrachteten als Spezialfall auf. Ist nämlich $\varphi \in \mathcal{A}^1 X$ und $\gamma\colon [a,b] \longrightarrow X$ eine differenzierbare Kurve, so ist $(\gamma^*\varphi)_t(v) = \varphi\big(\gamma(t), \dot\gamma(t)\big)\cdot v$ für jeden Tangentialvektor $v \in T_t[a,b] = \mathbb{R}$, also

$$(\gamma^*\varphi)_t = \varphi\big(\gamma(t), \dot\gamma(t)\big)\, dt$$

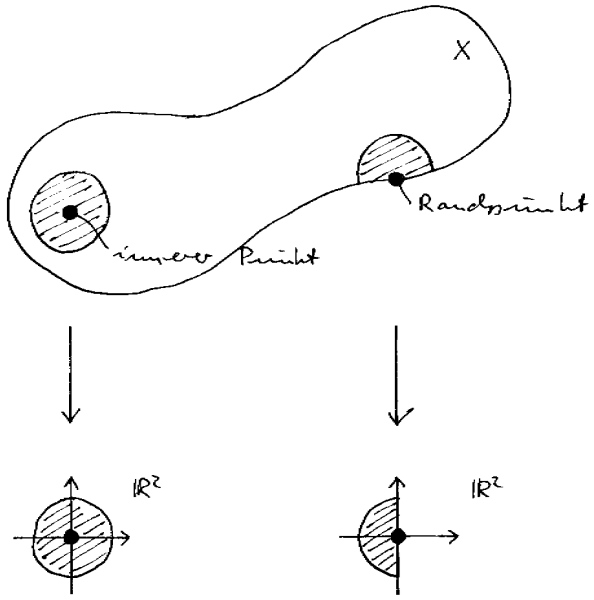
und damit

$$\int_{[a,b]} \gamma^*\varphi = \int_a^b \varphi\big(\gamma(t), \dot\gamma(t)\big)\, dt = \int_\gamma \varphi$$

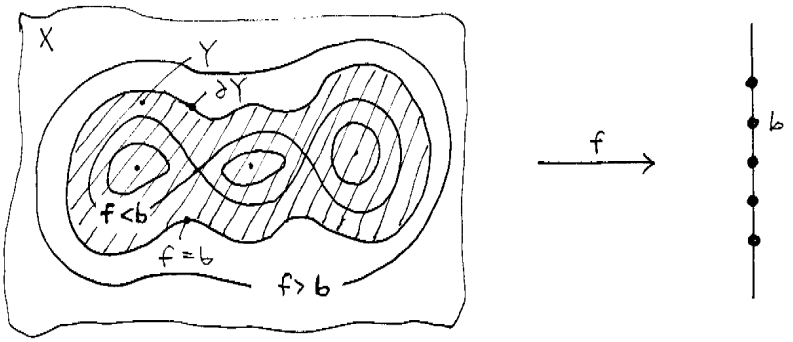
das Kurvenintegral von φ längs γ. Und ganz nebenbei hat in dem alten Integralsymbol $\int f(t)dt$ das dt im nachhinein eine inhaltliche Bedeutung bekommen: dt ist die Standardbasisform auf dem Integrationsintervall, und integriert wird nicht die Funktion f, sondern die 1-Form $f(t)dt$. (Das ist auch der historische Gang der Dinge: $\int \ldots dt$ war zunächst als Ganzes ein Symbol, das an $\sum \ldots \Delta t$ erinnern sollte, und erst die Entdeckung der Pfaffschen Formen hat dem dt seine selbständige Bedeutung gegeben.)

Es bleibt uns keine Zeit, die soweit besprochene Vektoranalysis in der gleichen Ausführlichkeit zum Abschluß zu bringen. Über eines der wichtigsten noch ausstehenden Unterthemen will ich aber wenigstens kurz erzählen.

42.5 Bericht über den **Satz von Stokes** In dem Satz geht es um Differentialformen auf einer sogenannten *berandeten* Mannigfaltigkeit. Bei diesem Begriff handelt es sich um eine Verallgemeinerung des gewöhnlichen Mannigfaltigkeitsbegriffs; eine n-dimensionale berandete Mannigfaltigkeit darf außer den "inneren" Punkten (wo sie aussieht wie $\mathbb{R}^n$) auch Randpunkte haben, in denen sie wie ein abgeschlossener Halbraum in $\mathbb{R}^n$ aussieht.



Statt einer genauen Definition erläutere ich, wie manche berandete Mannigfaltigkeiten als Anwendung des Satzes vom regulären Wert entstehen. Wir gehen von einer offenen Menge $X \subset \mathbb{R}^n$ und einem regulären Wert $b \in \mathbb{R}$ der differenzierbaren Funktion $f\colon X \longrightarrow \mathbb{R}$ aus. Dann ist das Urbild $Y := f^{-1}(-\infty, b]$ eine n-dimensionale Untermannigfaltigkeit von X mit dem Rand $\partial Y = f^{-1}\{b\}$.



Der Rand von Y ist also (nach dem Satz 36.9 vom regulären Wert) selbst eine (randlose) Untermannigfaltigkeit von um eins kleinerer Dimension, und das ist auch allgemein so. Konkretes Beispiel: Die kompakte Kugel $D^n \subset \mathbb{R}^n$ ist eine n-dimensionale berandete Mannigfaltigkeit mit dem Rand $\partial D^n = S^{n-1}$, denn mit $\mathbb{R}^n \ni x \longmapsto f(x) = |x|^2 \in \mathbb{R}$ ist $D^n = f^{-1}(-\infty, 1]$ und $S^{n-1} = f^{-1}\{1\}$.

Diese Beispielklasse illustriert auch, wie sich eine eventuelle Orientierung einer berandeten Mannigfaltigkeit auf ihren Rand vererbt: Alle Tangentialräume an $Y = f^{-1}(-\infty, b]$ sind mit $\mathbb{R}^n$ identisch und deshalb kanonisch orientiert, während wir den Rand $\partial Y = f^{-1}\{b\}$ in der Orientierumgsdiskussion 40.7 durch eine Übereinkunft orientiert hatten: Eine Basis von $T_a(\partial Y) = T_a f^{-1}\{b\}$ gilt als positiv orientiert, wenn durch Voranstellen eines nach außen weisenden Tangentialvektors eine positiv orientierte Basis von $T_a Y = \mathbb{R}^n$ entsteht. Gemäß dieser sinngemäß auch auf andere berandete Mannigfaltigkeiten zu übertragenden Konvention trägt also zum Beispiel S^{n-1} als Rand von D^n die Standardorientierung, als Rand des Außengebietes $\mathbb{R}^n \backslash U^n = f^{-1}[1, \infty)$ aber die entgegengesetzte:

$$\partial(\mathbb{R}^n \backslash U^n) = -S^{n-1} = -\partial D^n$$

Nun zum Satz von Stokes, hier in einer gängigen, schon ziemlich allgemeinen Fassung:

Sei X eine *kompakte* orientierte berandete Mannigfaltigkeit der Dimension n. Dann gilt

$$\int_{\partial X} \varphi = \int_X d\varphi$$

für jede Differentialform $\varphi \in \mathcal{A}^{n-1}X$.

Mit $\int_{\partial X} \varphi$ ist hier natürlich das Integral von $\varphi|\partial X$ über ∂X gemeint, was ja Sinn gibt, weil auch ∂X orientiert ist und φ den richtigen Grad $n-1$ hat.

In dieser Form noch nie gesehen? Betrachten wir mal den Spezialfall, daß $X \subset \mathbb{R}^3$ dreidimensional ist (etwa $X = D^3$), und begeben wir uns in die Niederungen des Koordinatenrechnens. Die 2-Form φ schreibt sich dann

$$\varphi = g_1\, dy{\wedge}dz + g_2\, dz{\wedge}dx + g_3\, dx{\wedge}dy.$$

Auf der linken Seite der Stokesschen Formel wird φ auf die Fläche ∂X eingeschränkt. Dieser Vorgang bedeutet in Vektornotation, den Vektor $\vec{g}$ senkrecht auf die nach außen weisende Flächennormale zu projizieren (siehe Aufgabe 39.6); Physiker würden die 2-Form $\varphi|\partial X$ also als Skalarprodukt $\vec{g}{\cdot}d\vec{f}$ mit dem "Flächenelement" $d\vec{f}$ schreiben. Das d auf der anderen Seite der Stokesschen Formel wird in Koordinatenschreibweise zur Divergenz, so daß sich die Formel als der *Gaußsche Integralsatz*

$$\int_{\partial X} \vec{g}{\cdot}d\vec{f} = \int_X \operatorname{div} \vec{g}$$

präsentiert.

Nehmen wir jetzt eine 1-Form φ, bleiben aber dabei, daß $\varphi = g_1\, dx + g_2\, dy + g_3\, dz$ auf ganz $\mathbb{R}^3$ definiert ist. Um den Satz von Stokes auf einem zweidimensionalen $X \subset \mathbb{R}^3$ (einer "Fläche") anzuwenden, ist φ auf deren Rand (eine "Kurve") einzuschränken. Wie Sie aus Aufgabe 39.5 wissen, bedeutet das in Vektorschreibweise, den Vektor $\vec{g}$ senkrecht auf seine Tangentialkomponente längs ∂X zu projizieren, so daß ein Physiker diese Einschränkung als Skalarprodukt $\vec{g} \cdot d\vec{s}$ mit dem "Linienelement" $d\vec{s}$ schreiben würde. Auf der anderen Seite der Stokesschen Formel haben wir nach Lemma 41.2 (Vertauschbarkeit von d mit dem Zurückziehen) $d(\varphi|X) = (d\varphi)|X$, also $\int_X d(\varphi|X) = \int_X d\varphi$. Weil nun $d\colon \mathcal{A}^1\mathbb{R}^3 \longrightarrow \mathcal{A}^2\mathbb{R}^3$ der Rotation, und das Einschränken von $d\varphi$ auf X wie besprochen dem Projizieren auf die Flächennormale entspricht, ergibt sich genau das, was die Physiker Ihnen als Satz von Stokes verkaufen:

$$\int_{\partial X} \vec{w}{\cdot}d\vec{s} = \int_X \operatorname{rot} \vec{w}{\cdot}\vec{f}$$

Na ja, welche die schönere, durchsichtigere, einfacher anzuwendende und vielseitigere Stokessche Formel ist, ist wohl keine Frage.

Schließlich läßt der Satz sich auf eine 0-Form, also eine Funktion f auf einem kompakten Intervall $[a,b]$ anwenden. Natürlich ist $\partial[a,b] = \{a,b\}$ als Menge, aber eine konsequente Interpretation der Orientierungskonventionen zeigt, daß man linke Randpunkte als negativ orientiert zählen soll: $\partial[a,b] = -\{a\} \cap \{b\}$. Die linke Seite der Stokesschen Gleichung wird dann

$$\int_{\partial[a,b]} f = \int_{-\{a\}} f + \int_{\{b\}} f = f(b) - f(a),$$

die rechte

$$\int_{[a,b]} df = \int_a^b f(t)dt.$$

So erweist sich der Satz von Stokes als mehrdimensionale Fassung des sogenannten Hauptsatzes, und er würde diesen Namen schon mit größerem Recht tragen dürfen als jener.

Übungsaufgaben

42.1 Berechnen Sie den Flächeninhalt der in Polarkoordinaten durch

$$r^2 \leq \cos 2\varphi$$

gegebenen Fläche (Skizze!).

42.2 Berechnen Sie den Trägheitstensor der Achtelkugel

$$X = \big\{(x,y,z) \in D^3 \,\big|\, x \geq 0,\, y \geq 0,\, z \geq 0\big\}$$

mit der konstanten Dichte 1, bezogen auf einen Punkt Ihrer Wahl (man wird in jedem Fall zuerst die Spitze bei 0 nehmen, weil das am bequemsten ist; das Resultat kann man dann etwa zum schwerpunktbezogenenen Trägheitstensor umrechnen, vergleiche den Steinerschen Satz aus Aufgabe 33.4).

42.3 Das Integral $\int_{-\infty}^{\infty} e^{-t^2/2} dt$ läßt sich mit der Standardmethode nicht auswerten, weil der Integrand keine elementare Stammfunktion besitzt.

(a) Es geht aber ganz einfach, wenn man das Integral

$$\int_{\mathbb{R}^2} e^{-(x^2+y^2)/2} dx \wedge dy$$

einmal in kartesischen und einmal in Polarkoordinaten bestimmt.

(b) Ist allgemeiner $q\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ eine positiv definite quadratische Form, so existiert das Integral

$$\int_{\mathbb{R}^n} e^{-q(x)/2} dx,$$

und es läßt sich leicht berechnen, wenn man vorweg q mittels der Hauptachsentransformation nach Satz 29.1 diagonalisiert.

42.4 Man hat oft mit Funktionen $f\colon \mathbb{R}^3\backslash\{0\} \longrightarrow \mathbb{R}$ zu tun, die aus zwei stetigen Funktionen $R\colon (0,\infty) \longrightarrow \mathbb{R}$ und $\Psi\colon S^2 \longrightarrow \mathbb{R}$ gemäß der Formel (in Kugelkoordinaten!)

$$f(r,\theta,\varphi) = R(r)\cdot\Psi(\theta,\varphi)$$

gebildet sind. Welche Bedingungen muß R erfüllen,

(a) damit f sich im Nullpunkt stetig ergänzen läßt,

(b) damit f über D^3 (oder eine andere Kugel um den Nullpunkt) integrierbar ist,

(c) damit f über $\mathbb{R}^3\backslash U^3$ (oder das Komplement irgendeiner Kugel um 0) integrierbar ist?

Eine einfache Antwort auf diese Fragen, die zwar nicht erschöpfend ist, aber für praktische Zwecke oft ausreicht, läßt sich mittels der Landauschen Symbole formulieren.

42.5 Diese Aufgabe illustriert ein wenig die Idee einer Zerlegung der Eins. Konstruieren Sie eine Familie $(f_n)_{n\in\mathbb{Z}}$ von C^∞-Funktionen $f_n\colon \mathbb{R} \longrightarrow [0,1]$ mit den folgenden Eigenschaften:

- $f_n|(-\infty, n-1] = 0$ und $f_n|[n+1,\infty) = 0$ für jedes $n \in \mathbb{Z}$
- $\displaystyle\sum_{n=-\infty}^{\infty} f_n = 1$

Tip: Aus zwei Exemplaren der Funktion f aus Aufgabe 14.8 bildet man leicht eine Funktion, die auf einem vorgeschriebenen offenen Intervall positiv ist und sonst identisch verschwindet.

42.6 Sei $X \subset \mathbb{R}^n$ offen und $f\colon X \longrightarrow \mathbb{R}$ eine differenzierbare Funktion ohne kritische Punkte. Sei außerdem $\varphi \in \mathcal{A}^{n-1}X$ eine Differentialform, für die $df \wedge \varphi \in \mathcal{A}^n X$ integrierbar ist. Zeigen Sie, daß dann

$$\int_X df \wedge \varphi = \int_{-\infty}^{\infty} \left(\int_{f^{-1}(t)} \varphi \right) dt$$

gilt. Benutzen Sie dabei ruhig die in der Vorlesung unter dem Stichwort "Zerlegung der Eins" erwähnte Tatsache, daß man nur solche φ zu betrachten braucht, die außerhalb einer vorgeschrieben kleinen offenen Menge U identisch verschwinden, und rechnen Sie in einer Karte (U, h) nach dem Satz vom regulären Punkt.

42.7 Die Differentialform $\psi \in \mathcal{A}^2 S^2$ sei die Einschränkung von $z\,dx \wedge dy \in \mathcal{A}^2\mathbb{R}^3$. Bestimmen Sie das Integral $\int_{S^2} \psi$, indem Sie in den beiden auf der nördlichen bzw. und südlichen Hemisphäre definierten Karten $h_\pm = (x, y)$ rechnen (den Äquator darf man für die Berechnung des Integrals ja ignorieren). Warum kann man von vornherein sicher sein, daß das Ergebnis eine positive Zahl ist?

43 Höhere Ableitungen in mehreren Variablen

Wir haben uns bisher nicht ernsthaft mit den höheren Ableitungen einer mehrdimensionalen Abbildung beschäftigt, und viel Zeit bleibt uns auch nicht mehr dafür. Daß diese Ableitungen in meiner Darstellung einen vegleichsweise kleinen Raum einnehmen, liegt nicht daran, daß sie nicht wichtig oder nützlich wären, sondern einfach daran, daß im Mehrdimensionalen schon die Theorie des ersten Differentials so reichhaltig ist, wie Sie gesehen haben.

Betrachten wir ein auf einer offenen Menge $X \subset \mathbb{R}^n$ definiertes genügend differenzierbares $f\colon X \longrightarrow \mathbb{R}^p$. Nach unserem bisherigen Verständnis ist das Differential von f eine Abbildung

$$Df\colon X \longrightarrow \mathrm{Mat}(p\times n, \mathbb{R}) = \mathbb{R}^{pn},$$

deshalb das zweite Differential eine Abbildung $X \longrightarrow \mathrm{Mat}(pn\times n, \mathbb{R}) = \mathbb{R}^{pn^2}$ und allgemein das k-te Differential eine Abbildung

$$Df\colon X \longrightarrow \mathrm{Mat}(pn^{k-1}\times n, \mathbb{R}) = \mathbb{R}^{pn^k}.$$

Rechnerisch wird dieses Differential durch die Gesamtheit der partiellen Ableitungen k-ter Ordnung

$$D_{j_1}D_{j_2}\cdots D_{j_k}f_i = \frac{\partial^k f_i}{\partial x_{j_1}\partial x_{j_2}\cdots\partial x_{j_k}}$$

ausgedrückt, die ihrerseits Funktionen auf X sind. (Aufgrund der Vertauschbarkeit der Differentiationsreihenfolge nach Satz 38.4 stimmen viele dieser partiellen Ableitungen überein.) Jedenfalls sind die höheren Ableitungen — anders als im Eindimensionalen — von ihrer Art her komplizierter als die Ausgangsabbildung. Klar ist auch, daß das nicht an $p>1$, sondern an $n>1$ liegt, und überhaupt kann man sich wie so oft dadurch auf den Fall $p=1$ zurückziehen, daß man die p Komponenten $f_1,\dots,f_p$ von f als eigenständige Funktionen betrachtet. Das wollen wir im folgenden der Übersichtlichkeit halber auch tun.

Um die höheren Ableitungen in den Griff zu bekommen, braucht man ein System, nach dem man die vielen einzelnen Ableitungen zu einem einzigen Objekt zusammenfassen kann, mit dem man gut rechnen kann. Im Eindimensionalen war ein solches Objekt das Taylor-Polynom: Das k-te Taylorpolynom der entsprechend differenzierbaren Funktion f an der Stelle a enthält ja vermöge

$$T_a^k f(x) = \sum_{j=0}^{k} \frac{f^{(j)}(a)}{j!}\,(x-a)^j$$

genau die gleiche Information wie die Gesamtheit der Ableitungen von f bis zur Ordnung k an dieser Stelle. Im Mehrdimensionalen ist das ganz genau so; der Unterschied ist höchstens, daß das Rechnen mit einzelnen Ableitungen in einer Variablen nur oft ungeschickt, in mehreren Variablen dagegen durchweg hoffnungslos ist.

Um über Taylor-Polynome in mehreren Variablen reden zu können, muß man natürlich wissen, was ein Polynom in n Variablen ist. Wie im bekannten Fall einer Variablen ist ein solches Polynom eine Linearkombination von Monomen, nur daß zur Kennzeichnung eines solchen Monoms eine einzelne Zahl $j \in \mathbb{N}$ (der Grad) jetzt nicht mehr ausreicht, sondern ein ganzes n-tupel $j = (j_1,\dots,j_n) \in \mathbb{N}^n$ erforderlich ist (man spricht in diesem Zusammenhang oft von "Multiindizes").

43.1 Definition Sei $n \in \mathbb{N}$. Unter einem Monom in n Variablen versteht man einen Ausdruck der Form

$$X^j := X_1^{j_1}X_2^{j_2}\cdots X_n^{j_n}$$

in den n Veränderlichen $X_1, \dots, X_n$ mit $j = (j_1, \dots, j_n) \in \mathbb{N}^n$. (Wie im eindimensionalen Fall verwendet man vorzugsweise Großbuchstaben als Platzhalter für die noch einzusetzenden Variablen.) Die Zahl

$$|j| := j_1 + j_2 + \cdots + j_n \in \mathbb{N}$$

nennt man die Ordnung von j und den Grad von X^j. Ist K ein Körper, so ist ein Polynom vom Grad d über K eine Linearkombination

$$f(X_1, \dots, X_n) = \sum_{|j| \le d} a_j X^j,$$

worin die Summe wie angedeutet über alle $j = (j_1, \dots, j_n) \in \mathbb{N}^n$ mit $|j| \le d$ zu nehmen ist und mindestens ein a_j mit $|j| = d$ von null verschieden ist.

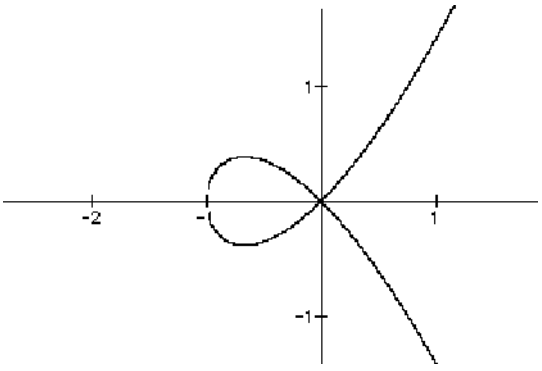
Die Polynome in n Variablen über K bilden in der üblichen Weise nicht nur einen K-Vektorraum, sondern vor allem einen Ring, den man mit $K[X_1, \dots, X_n]$ bezeichnet oder, wenn die Zahl der Variablen aus dem Zusammenhang hervorgeht, auch kurz mit $K[X]$. (Hat man nur wenige Variablen, kann man natürlich durch Schreibweisen wie $K[X, Y, Z]$ Indizes einsparen.)

43.2 Beispiele (1) $X^2+X^3-Y^2 \in \mathbb{R}[X, Y]$ und $X^2Y-3Z^2+X^3Y^2+YZ^4 \in \mathbb{R}[X, Y, Z]$ sind reelle Polynome in zwei bzw. drei Variablen; ihre Grade sind drei und fünf.

(2) Ein Polynom vom Grad zwei in drei Variablen lautet ausgeschrieben und nach konstantem, linearem und quadratischem Anteil geordnet

$$\begin{aligned} f(X, Y, Z) = {} & a_{000} \\ & + a_{100}X + a_{010}Y + a_{001}Z \\ & + a_{200}X^2 + a_{020}Y^2 + a_{002}Z^2 + a_{011}YZ + a_{101}XZ + a_{110}XY. \end{aligned}$$

Bemerkungen Man kann bei Polynomen in mehr als einer Variablen nicht mehr von einem Leitterm reden, höchstens noch von der sogenannten *Leitform*, die alle Terme des höchsten Grades umfaßt. Auch der Begriff "normiertes Polynom" gibt keinen Sinn mehr. — Selbst über dem Körper $\mathbb{C}$ zerfallen Polynome in mehreren Variablen im allgemeinen nicht in Linearfaktoren. Wie man leicht beweist, ist das schon für die einfachen Beispiele in (1) nicht der Fall. Das hat übrigens eine einfache geometrische Bedeutung: etwa müßte die Nullfaser $\{(x, y) \mid x^2+x^3-y^2 = 0\}$ sonst ja eine Vereinigung von Geraden sein.



Die Multiindizes identifizieren nun nicht nur die Monome, sondern auch die partiellen Ableitungen. Ist $f: X \longrightarrow \mathbb{R}$ wie oben eine C^k-Funktion, so sind an (möglicherweise) verschiedenen partiellen Ableitungen nach Satz 38.4 nur

$$D_j f := (D_1)^{j_1}(D_2)^{j_2} \cdots (D_n)^{j_n} f = \frac{\partial^{|j|} f}{(\partial x_1)^{j_1}(\partial x_2)^{j_2} \cdots (\partial x_n)^{j_n}}$$

zu betrachten, für die Multiindizes $j \in \mathbb{N}^n$ mit $|j| \le k$. Unser erstes Ziel soll es sein, die Ableitungen eines Polynoms an der Stelle 0 berechnen. Dazu vereinbaren wir für die Multiindizes noch weitere

43.3 Notationen Für $j = (j_1, j_2, \dots, j_n) \in \mathbb{N}^n$ kürzt man

$$j! := j_1! j_2! \cdots j_n!$$

ab; ist $k = (k_1, k_2, \dots, k_n) \in \mathbb{N}^n$ ein weiterer Multiindex, so kann man vermöge

$$j \pm k := (j_1 \pm k_1, j_2 \pm k_2, \dots, j_n \pm k_n) \quad \text{und} \quad j \le k :\iff j_1 \le k_1,\ j_2 \le k_2, \dots, j_n \le k_n$$

auch rechnen und vergleichen.

43.4 Lemma Das Polynom

$$f(X_1, \dots, X_n) = \sum_{|j| \le d} a_j X^j$$

hat an der Stelle $0 \in \mathbb{R}^n$ die Ableitungen

$$D_k f(0) = k! \cdot a_k \quad \text{für alle } k \in \mathbb{N}^n,$$

wobei für $|k| > d$ natürlich $a_k = 0$ zu lesen ist.

Beweis Ebenso einfach wie im Eindimensionalen: k_1-maliges Differenzieren nach x_1 gibt

$$(D_1)^{k_1} f(X) = \sum_{j_1 \ge k_1} a_j j_1 (j_1 - 1) \cdots (j_1 - k_1 + 1)\, X_1^{j_1 - k_1} X_2^{j_2} \cdots X_n^{j_n} = \sum_{j_1 \ge k_1} a_j \frac{j_1!}{(j_1 - k_1)!}\, X_1^{j_1 - k_1} X_2^{j_2} \cdots X_n^{j_n},$$

nach Anwendung aller $|j|$ Differentialoperatoren hat man

$$D_k f(X) = \sum_{j \ge k} a_j \frac{j!}{(j-k)!}\, X^{j-k},$$

und Auswerten an der Stelle $X = 0$ läßt davon nur den Summanden zu $j = k$:

$$D_k f(0) = a_k \frac{k!}{0!}\, X^0 = k! \cdot a_k$$

Man kann sich auf dieses Lemma berufen, wenn man zwischen dem Polynom $f(X)$ als Rechenausdruck und der durch den Ausdruck definierten Funktion $\mathbb{R}^n \ni x \mapsto f(x) \in \mathbb{R}$ nicht unterscheidet, denn ersichtlich lassen sich aus dieser Funktion alle Koeffizienten von $f(X)$ vermöge der Ableitungen rekonstruieren. Das Lemma zeigt aber auch, wie man die mehrdimensionalen Taylor-Polynome definieren muß.

43.5 Definition Sei $X \subset \mathbb{R}^n$ offen und $f\colon X \longrightarrow \mathbb{R}$ eine C^k-Funktion. Das k-te Taylor-Polynom von f an der Stelle $a \in X$ ist das Polynom

$$T_a^k f(X) := \sum_{|j| \le k} \frac{D_j f(a)}{j!}\, (X - a)^j \in \mathbb{R}[X_1, \dots, X_n].$$

Warum ist das die "richtige" Definition? Weil sich aus Lemma 43.4 ergibt:

43.6 Notiz $T_a^k f(X)$ ist das einzige Polynom in $\mathbb{R}[X]$, dessen Grad höchstens k ist und dessen partielle Ableitungen an der Stelle a mit denen von f selbst bis zur Ordnung k einschließlich übereinstimmen. Es gilt also $T_a^k(T_a^k f) = T_a^k f$ und allgemeiner

$$T_a^k(T_a^l f) = T_a^k f \quad \text{für jedes } k \le l.$$

Beweis Für $a = 0$ ist das die Aussage des Lemmas, und auf diesen Fall kann man sich hier und an vielen anderen Stellen leicht zurückziehen: definiert man die Hilfsfunktion $\tilde{f}$ durch $\tilde{f}(x) = f(x+a)$, so gilt offenbar

$$T_a^k f(X) = T_0^k \tilde{f}(X-a).$$

Bemerkung Das Taylor-Polynom einer C^k-Abbildung $f\colon X \longrightarrow \mathbb{R}^p$ erklärt man durch komponentenweise Anwendung der obigen Definition; in

$$T_a^k f(X) = \sum_{|j|\le k} \frac{D_j f(a)}{j!}\,(X-a)^j$$

sind die Koeffizienten dann Vektoren (Spalten aus $\mathbb{R}^p$). Das einzige Problem dabei ist ein sprachliches: Der aus der Algebra stammende Begriff "Polynom" ist üblicherweise für skalare Funktionen reserviert; sonst hilft man sich, indem man von polynomialen Abbildungen redet. Weil man davor nicht gut noch "Taylor-" setzen kann, muß man hinnehmen, daß das Taylor-Polynom in diesem Fall gar kein Polynom, sondern eben eine solche polynomiale Abbildung ist.

Ich habe schon erwähnt, daß das Rechnen mit höheren Ableitungen gerade in mehreren Veränderlichen praktisch nur als Rechnen mit den Taylor-Polynomen Sinn gibt. Freilich will das dem Anfänger oft nicht einleuchten, weil angeblich $\dfrac{\partial^5 f}{\partial x^2 \partial y\,\partial z^2}(a)$ ebenso einfach wie $T_a^5 f(a)$ schwierig zu verstehen sei. Wahr ist daran bloß, daß in der Regel elementarste Kenntnisse der Differentialrechnung genügen, um die partiellen Ableitungen (aber damit doch auch das Taylor-Polynom) auszurechnen; nur was soll man mit einer einzelnen partiellen Ableitung denn schon anfangen? Ich will Ihnen aber helfen, die Taylor-Hemmschwelle zu überwinden, indem ich zunächst die Taylor-Polynome kleiner Ordnung mit Ihnen explizit durchgehe. Dazu sei $f\colon X \longrightarrow \mathbb{R}$ jetzt wieder skalar.

Ordnung 0: Es gibt nur einen Multiindex vom Grad 0, nämlich $j = (0,\dots,0) \in \mathbb{N}^n$, also ist

$$T_a^0 f(X) = f(a)$$

das konstante Polynom.

Ordnung 1: Der Multiindizes vom Grad 1 gibt es n Stück, nämlich

$$j = (0,\dots,0,1,0,\dots,0) \in \mathbb{N}^n$$

mit der 1 an einer beliebigen Stelle. Damit ist

$$\begin{aligned} T_a^1 f(X) &= f(a) + \sum_{|j|=1} \frac{D_j f(a)}{j!}\,(X-a)^j \\ &= f(a) + \sum_{r=1}^{n} \frac{\partial f}{\partial x_r}(a)\cdot(X_r - a_r) \\ &= f(a) + Df(a)\cdot(X-a). \end{aligned}$$

Daß wir hier mit der zuletzt verwendeten eine besonders glatte Schreibweise zur Verfügung haben würden, war zu erwarten, denn mit dem ersten Taylor-Polynom in Gestalt des Differentials haben wir uns in den letzten Abschnitten ja schon ausgiebig befaßt. Übrigens ist hier, wo wir nur skalare Funktionen betrachten, a genau dann ein kritischer Punkt von f, wenn $Df(a) = 0$ ist, also wenn $T_a^1 f(X) = f(a)$ nur aus dem konstanten Term besteht.

Ordnung 2: Es gibt zwei Sorten von Multiindizes vom Grad 2: einmal die n Stück vom Typ

$$j = (0,\dots,0,2,0,\dots,0) \in \mathbb{N}^n \quad \text{mit } j! = 2,$$

und dann die $\binom{n}{2} = \frac{1}{2}n(n-1)$ vom Typ

$$j = (0,\dots,0,1,0,\dots,0,1,0,\dots,0) \in \mathbb{N}^n \quad \text{mit } j! = 1.$$

Dementsprechend kann man das zweite Taylor-Polynom aufschlüsseln:

$$\begin{aligned} T_a^2 f(X) &= f(a) + Df(a)\cdot(X-a) + \sum_{|j|=2} \frac{D_j f(a)}{j!}\,(X-a)^j \\ &= f(a) + Df(a)\cdot(X-a) + \frac{1}{2}\sum_{r=1}^{n} \frac{\partial^2 f}{\partial x_r^2}(a)\cdot(X_r-a_r)^2 + \sum_{r<s}^{n} \frac{\partial^2 f}{\partial x_r \partial x_s}(a)\cdot(X_r-a_r)(X_s-a_s) \end{aligned}$$

Das, was zum ersten Taylor-Polynom hinzukommt, ist ersichtlich eine quadratische Form in der Vektorvariablen $X-a$. Wenn man die partiellen Ableitungen zweiter Ordnung zu der quadratischen Matrix

$$Hf(a) := \Big(D_r D_s f(a)\Big)_{r,s=1}^{n} = \Big(\frac{\partial^2 f}{\partial x_r \partial x_s}\,(a)\Big)_{r,s=1}^{n} \in \mathrm{Sym}(n,\mathbb{R})$$

zusammenfaßt, kann man

$$T_a^2 f(X) = f(a) + Df(a)\cdot(X-a) + \frac{1}{2}(X-a)^{\mathrm{t}} Hf(a)(X-a)$$

schreiben.

43.7 Definition Ist $f\colon X \longrightarrow \mathbb{R}$ eine C^2-Funktion und $a \in X$, so nennt man die Matrix

$$Hf(a) = \Big(D_r D_s f(a)\Big)_{r,s=1}^{n} \in \mathrm{Sym}(n,\mathbb{R})$$

die Hesse-Matrix von f an der Stelle a, und die zugehörige quadratische oder Bilinearform auf $\mathbb{R}^n$ die Hesse-Form.

Ordnung 3: Hier will ich allgemein nur noch die möglichen $j \in \mathbb{N}^n$ mit $|j| = 3$ auflisten:

$$\begin{aligned} j = (0,\ldots,0,3,0,\ldots,0) &\quad \text{mit } j! = 6, \\ j = (\ldots,2,\ldots,1,\ldots) &\quad \text{mit } j! = 2, \\ j = (\ldots,1,\ldots,2,\ldots) &\quad \text{mit } j! = 2, \\ j = (\ldots,1,\ldots,1,\ldots,1,\ldots) &\quad \text{mit } j! = 1. \end{aligned}$$

Ist speziell $n = 3$ und $a = 0$ und schreiben wir X, Y, Z statt X_1, X_2, X_3, so sind

$$X^3,\ Y^3,\ Z^3\ X^2Y,\ X^2Z,\ Y^2Z,\ XY^2,\ XZ^2,\ YZ^2,\ XYZ$$

die möglichen Monome vom Grad drei, und

$$\frac{1}{6}\frac{\partial^3 f}{\partial x^3}(0),\ldots,\frac{1}{2}\frac{\partial^3 f}{\partial x^2 \partial y}(0),\ldots,\frac{\partial^3 f}{\partial x \partial y \partial z}(0)$$

die zugehörigen Koeffizienten im Taylor-Polynom.

Die meisten Taylor-Polynome in einer Variablen hatten wir durch Abschneiden der Taylor-*Reihe* einer analytischen Funktion erhalten. Natürlich gibt es auch eine Theorie der Potenzreihen und analytischen Funktionen von mehreren Variablen; weil deren systematisches Studium an dieser Stelle aber weniger lohnt, sage ich lieber etwas mehr zur Berechnung des Taylor-Polynoms einer C^k-Abbildung. Das wichtigste Werkzeug dazu ist die schon unter den Regeln 16.10 aufgeführte

43.8 Kettenregel Die Abbildungen $\mathbb{R}^n \supset X \xrightarrow{f} Y \subset \mathbb{R}^p$ und $Y \xrightarrow{g} \mathbb{R}^q$ seien C^k-differenzierbar. Dann gilt

$$T_a^k(g\circ f) = T_a^k\big(T_{f(a)}^k g \circ T_a^k f\big)$$

für jedes $a \in X$.

Neben der ohnehin offensichtlichen Linearität des Taylor-Polynoms ist noch die Produktregel merkenswert; sie ist — anders als im Eindimensionalen — in der Kettenregel als Spezialfall enthalten:

43.9 Produktregel $X \xrightarrow{f} \mathbb{R}^p$ und $X \xrightarrow{g} \mathbb{R}^q$ seien C^k-Abbildungen, und $\mathbb{R}^p \times \mathbb{R}^q \ni (y,z) \longmapsto y*z \in \mathbb{R}^r$ sei ein bilineares Produkt. Für jedes $a \in X$ gilt dann

$$T_a^k(f*g) = T_a^k\left(T_a^k f * T_a^k g\right).$$

Beweis $f*g\colon X \longrightarrow \mathbb{R}^r$ ist die Komposition

$$x \longmapsto \begin{pmatrix} f(x) \\ g(x) \end{pmatrix} = \begin{pmatrix} y \\ z \end{pmatrix} \longmapsto y*z$$

Nun ist die Multiplikation $*$ als bilineare Abbildung polynomial vom Grad zwei; für $k \geq 2$ stimmt sie nach der Notiz 43.6 also mit ihrem k-ten Taylor-Polynom an einer beliebigen Stelle überein. Nach der Kettenregel ist deshalb

$$T_a^k(f*g) = T_a^k\left(\left(T^k_{f(a),g(a)}*\right) \circ \begin{pmatrix} T_a^k f \\ T_a^k g \end{pmatrix}\right) = T_a^k\left(T_a^k f * T_a^k g\right)$$

wie behauptet, und für $k=0$ und $k=1$ folgt die Regel daraus durch weiteres Abschneiden.

Im übrigen soll man beim Rechnen mit Taylor-Polynomen den gesunden Menschenverstand benutzen. Was das heißen soll, illustriert das folgende

43.10 Beispiel Wir wollen die Taylor-Polynome kleiner Ordnung der Funktion

$$f\colon \mathbb{R}^2 \longrightarrow \mathbb{R};\ (x,y) \mapsto (\sin x)^2 \cdot \sin(x+y) \cdot \sqrt[10]{1 + \frac{\arctan x \cdot e^y}{1+y^4}}$$

im Nullpunkt bestimmen. Wenn wir dazu

$$f(x,y) = \underbrace{(\sin x)^2}_{u(x,y)} \cdot \underbrace{\sin(x+y)}_{v(x,y)} \cdot \underbrace{\sqrt[10]{1 + \frac{\arctan x \cdot e^y}{1+y^4}}}_{w(x,y)}$$

schreiben, sind die ersten nicht-verschwindenden Taylor-Polynome der drei Faktoren

$$\begin{aligned} T_0^2 u(x,y) &= x^2 \\ T_0^1 v(x,y) &= x+y \\ T_0^0 w(x,y) &= 1; \end{aligned}$$

daraus ergibt sich schon $T_0^3 f(x,y) = x^2\cdot(x+y)\cdot 1 = x^3+x^2y$, ohne daß man die vollständigen dritten Taylor-Polynome von u, v und w ansehen müßte. Wenn wir nun etwa an $T_0^5 f$ interessiert sind, genügt es, bei jedem Faktor um zwei Schritte weiterzugehen. $T_0^4 u$ ergibt sich aus

$$(\sin x)^2 = \left(x - \frac{1}{6}x^3 + \cdots\right)^2 = x^2 - \frac{1}{3}x^4 + \cdots,$$

$T_0^3 v(x+y) = x+y - \frac{1}{6}(x+y)^3$ ist klar, und bei $T_0^2 w$ gehen wir von den Potenzreihenentwicklungen

$$\sqrt[10]{1+z} = (1+z)^{1/10} = \binom{1/10}{0} + \binom{1/10}{1}z + \binom{1/10}{2}z^2 + \cdots = 1 + \frac{1}{10}z - \frac{9}{200}z^2 + \cdots$$

und

$$\frac{\arctan x \cdot e^y}{1+y^4} = (x - \frac{1}{3}x^3 + \cdots)\cdot(1 + y + \frac{1}{2}y^2 + \cdots)(1 - y^4 + \cdots) = x + xy + \cdots$$

aus. Nach Ketten- und Produktregel erhalten wir schließlich

$$\begin{aligned} f(x,y) &= \left(x^2 - \frac{1}{3}x^4 + \cdots\right)\left(x+y - \frac{1}{6}(x+y)^3 + \cdots\right)\left(1 + \frac{1}{10}(x+xy) - \frac{9}{200}(x+xy)^2 + \cdots\right) \\ &= \left(x^2(x+y) - \frac{1}{6}x^2(x+y)^3 - \frac{1}{3}x^4(x+y) + \cdots\right)\left(1 + \frac{1}{10}x(1+y) - \frac{9}{200}x^2(1+y)^2 + \cdots\right) \\ &= x^2(x+y) - \frac{1}{6}x^2(x+y)^3 - \frac{1}{3}x^4(x+y) + \frac{1}{10}x^3(x+y)(1+y) - \frac{9}{200}x^4(x+y) + \cdots, \end{aligned}$$

womit

$$T_0^5 f(x,y) = x^2(x+y) + \frac{1}{10}x^3(x+y) - \frac{1}{6}x^2(x+y)^3 - \frac{1}{3}x^4(x+y) + \frac{1}{10}x^3(x+y)y - \frac{9}{200}x^4(x+y)$$

berechnet ist.

Die Kettenregel regelt auch das Verhalten der Taylor-Polynome unter einem differenzierbaren Kartenwechsel. Das Ergebnis ist im allgemeinen natürlich ziemlich kompliziert, kann aber auch ganz leicht überschaubar werden, wenn genügend viele Ableitungen niederer Ordnung verschwinden. Ein einfaches und wichtiges Beispiel dafür:

43.11 Lemma Sei $X \subset \mathbb{R}^n$ offen und a ein kritischer Punkt der C^2-Funktion $f\colon X \longrightarrow \mathbb{R}$. Für jede differenzierbare Karte (U,h) um a gilt dann

$$Hf(a) = Dh(a)^{\mathrm{t}} \cdot H(f\circ h^{-1})(0) \cdot Dh(a).$$

Beweis Weil $0 \in \mathbb{R}^n$ ein kritischer Punkt von $f\circ h^{-1}$ ist, besteht das Taylor-Polynom nur aus dem konstanten und dem quadratischen Term:

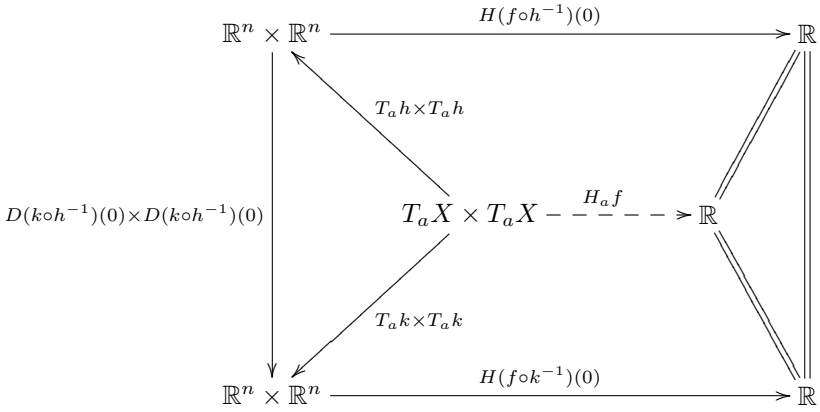
$$T_0^2(f\circ h^{-1})(Y) = (f\circ h^{-1})(0) + Y^{\mathrm{t}} \cdot H(f\circ h^{-1})(0) \cdot Y$$

Nach der Kettenregel folgt

$$\begin{aligned} T_a^2 f(X) &= T_a^2\big((f\circ h^{-1})\circ h\big)(X) \\ &= T_a^2\big(T_0^2(f\circ h^{-1})\circ T_a^2 h\big)(X) \\ &= T_a^2\left((f\circ h^{-1})(0) + \frac{1}{2}T_a^2h(X)^{\mathrm{t}} \cdot H(f\circ h^{-1})(0) \cdot T_a^2h(X)\right) \\ &= f(a) + \frac{1}{2}\big(Dh(a)\cdot(X-a)\big)^{\mathrm{t}} \cdot H(f\circ h^{-1})(0) \cdot \big(Dh(a)\cdot(X-a)\big); \end{aligned}$$

wegen $h(a) = 0$ kommen im letzten Schritt nur die ersten Ableitungen von h zum Zuge.

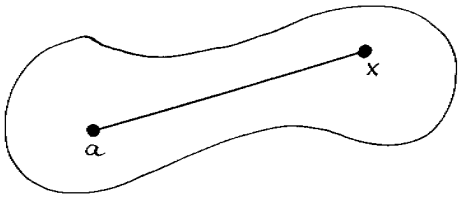
Die Formel des Lemmas ist genau die, die man braucht, um dem Begriff der Hesse-Form an einer kritischen Stelle einer auf einer Mannigfaltigkeit erklärten Funktion Sinn zu geben. Sei also X jetzt eine n-dimensionale Mannigfaltigkeit und $a \in X$ ein kritischer Punkt von $f\colon X \longrightarrow \mathbb{R}$. Sind dann (U,h) und (U,k) zwei Karten um a, so erlaubt das aufgrund von Lemma 43.11 kommutative Diagramm



es, die Hesse-Form von f bei a als den gestrichelten Pfeil $H_a f\colon T_a X \times T_a X \longrightarrow \mathbb{R}$ zu definieren. Die Hesse-Form $H_a f$ ist also eine von jeder Koordinatenwahl unabhängige symmetrische Bilinearform auf dem Tangentialraum in einem kritischen Punkt. In einem regulären Punkt wäre das falsch, und deswegen ist in solchen Punkten die im Fall $X = \mathbb{R}^n$ ja immer noch definierte Hesse-Form oder -Matrix von geringerem Interesse.

An Beziehungen zwischen einer C^k-Funktion f und ihren Taylor-Polynomen an der Stelle a kennen wir bisher nur die Notiz 43.6: die Taylor-Polynome imitieren f bestmöglich bezüglich der Ableitungen bei a. Die Abweichung zwischen Funktion und Taylor-Polynom läßt sich aber auch konkreter beschreiben, etwa mittels Landauscher Symbole, wie wir es im Eindimensionalen in Satz 16.11 getan hatten. Weil wir inzwischen die Integralrechnung zur Verfügung haben, können wir gleich noch weiter gehen:

43.12 Satz von **Taylor** Sei $X \subset \mathbb{R}^n$ offen, $f\colon X \longrightarrow \mathbb{R}^p$ eine C^{k+1}-Abbildung und $a \in X$. Für jeden Punkt $x \in X$, für den die Verbindungsstrecke von a nach x ganz in X liegt,



gilt dann:

$$f(x) - T_a^k f(x) = (k+1) \sum_{|j|=k+1} \int_0^1 (1-t)^k \frac{D_j f\big((1-t)a + tx\big)}{j!}\, dt \cdot (x-a)^j$$

Beweis Die Formel wird sich bei näherem Hinsehen als Aussage über die Funktion

$$\varphi\colon [0,1] \longrightarrow \mathbb{R};\ t \mapsto f\big((1-t)a + tx\big)$$

erweisen. Wir beweisen deshalb zuerst die eindimensionale Version

$$\varphi(1) - T_0^k \varphi(1) = \frac{1}{k!} \int_0^1 (1-t)^k \varphi^{(k+1)}(t)\, dt \quad \text{für jede } C^{k+1}\text{-Funktion } \varphi\colon [0,1] \longrightarrow \mathbb{R},$$

und zwar durch Induktion nach $k \in \mathbb{N}$. Für $k=0$ stimmt's: $\varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)\, dt$. Und für $k>0$ ist die Differenz der Integrale

$$\begin{aligned}
&\frac{1}{(k-1)!} \int_0^1 (1-t)^{k-1} \varphi^{(k)}(t)\, dt - \frac{1}{k!} \int_0^1 (1-t)^k \varphi^{(k+1)}(t)\, dt \\
&= \frac{1}{k!} \int_0^1 \Big(k(1-t)^{k-1} \varphi^{(k)}(t) - (1-t)^k \varphi^{(k+1)}(t) \Big) dt \\
&= \frac{1}{k!} \Big[-(1-t)^k \varphi^{(k)}(t) \Big]_{t=0}^{t=1} \\
&= \frac{1}{k!} \varphi^{(k)}(0)
\end{aligned}$$

gerade gleich dem zusätzlichen Term in $T_0^{k+1}\varphi(1)$.

Die allgemeine Taylor-Formel ergibt sich jetzt, wenn man φ speziell wie oben wählt. Die affin-lineare Abbildung $t \mapsto (1-t)a + tx$ stimmt für jedes $k \geq 1$ mit ihrem k-ten Taylor-Polynom (an beliebiger Stelle) überein. Nach der Kettenregel folgt erstens

$$T_0^k \varphi(1) = T_a^k f(x),$$

und wenn wir t vorübergehend festhalten und hilfsweise eine Polynom-Variable T einführen, außerdem

$$\begin{aligned}\frac{1}{(k+1)!}\,\varphi^{(k+1)}(t)\cdot(T-t)^{k+1} &= T_t^{k+1}\varphi(T) - T_t^k\varphi(T)\\ &= T^{k+1}_{(1-t)a+tx} f\big((1-T)a+Tx\big) - T^k_{(1-t)a+tx} f\big((1-T)a+Tx\big)\\ &= \sum_{|j|=k+1} \frac{D_j f\big((1-t)a+tx\big)}{j!}\,\Big(\big((1-T)a+Tx\big) - \big((1-t)a+tx\big)\Big)^j\\ &= \sum_{|j|=k+1} \frac{D_j f\big((1-t)a+tx\big)}{j!}\,(x-a)^j\cdot(T-t)^{k+1}.\end{aligned}$$

Aus dieser Identität zweier Polynome in T lesen wir ab, daß die Koeffizienten von $(T-t)^{k+1}$ übereinstimmen, notieren also

$$\frac{1}{(k+1)!}\,\varphi^{(k+1)}(t) = \sum_{|j|=k+1} \frac{D_j f\big((1-t)a+tx\big)}{j!}\,(x-a)^j.$$

Damit sind wir in der Lage, alle φ-haltigen Terme der speziellen Taylor-Formel durch f-haltige ersetzen, und erhalten die allgemeine Version:

$$\begin{aligned}f(x) - T_a^k f(x) &= \varphi(1) - T_0^k\varphi(1)\\ &= \frac{1}{k!}\int_0^1 (1-t)^k\varphi^{(k+1)}(t)\,dt\\ &= (k+1)\int_0^1 (1-t)^k \sum_{|j|=k+1} \frac{D_j f\big((1-t)a+tx\big)}{j!}\,(x-a)^j\,dt\end{aligned}$$

Es gibt noch einige andere Varianten des Satzes von Taylor; alle drücken die traditionell *Restglied* genannte Differenz $f(x) - T_a^k f(x)$ durch die Ableitungen der nächsthöheren Ordnung aus. Wenigstens erwähnt sei die Version von Lagrange, die leicht zu merken und deshalb besonders beliebt, allerdings relativ grob ist und nur für skalare Funktionen gilt:

43.13 Restglied nach **Lagrange** Sei $X \subset \mathbb{R}^n$ offen, $f\colon X \longrightarrow \mathbb{R}$ eine C^{k+1}-Funktion und $a \in X$. Zu jedem Punkt $x \in X$, für den die Verbindungsstrecke von a nach x ganz in X liegt, gibt es dann ein $t \in [0,1]$ mit

$$f(x) - T_a^k f(x) = \sum_{|j|=k+1} \frac{D_j f\big((1-t)a+tx\big)}{j!}\,(x-a)^j.$$

Bemerkung Für $k=0$ reduziert sich 43.13 auf den Mittelwertsatz der Differentialrechnung 14.1 (mit geringfügig strengeren Voraussetzungen).

Die Integralversion der Taylor-Formel hat die folgende hübsche Anwendung:

43.14 Lemma $X \subset \mathbb{R}^n$ sei offen und sternförmig bezüglich des Punktes $a \in X$. Weiter sei $1 \le k \le l$, und für die C^l-Funktion $f\colon X \longrightarrow \mathbb{R}$ gelte $T_a^{k-1} f = 0$. Dann gibt es Funktionen $f_j \in C^{l-k}(X)$, wobei j alle Multiindizes mit $|j| = k$ durchläuft, so daß

$$f(x) = \sum_{|j|=k} (x-a)^j\cdot f_j(x) \quad \text{für alle } x \in X$$

ist. Speziell gibt es zu jeder Funktion $f \in C^l(X)$ Funktionen $f_1, \dots, f_n \in C^{l-1}(X)$ mit

$$f(x) - f(a) = \sum_{r=1}^n (x_r - a_r)\cdot f_r(x) \quad \text{für alle } x \in X.$$

Beweis Von der Taylor-Formel 43.12 (mit $k-1$ statt k) bleibt nur

$$f(x) = \sum_{|j|=k} \underbrace{k\int_0^1 (1-t)^{k-1} \frac{D_j f\big((1-t)a+tx\big)}{j!}\, dt}_{f_j(x)} \cdot (x-a)^j,$$

und Satz 32.11 garantiert, daß es sich bei den f_j um C^{l-k}-Funktionen handelt. Der Spezialfall ergibt sich, indem man $k = 0$ wählt und f durch $f - f(a)$ ersetzt: Das nullte Taylor-Polynom dieser Funktion bei a verschwindet, und die Monome vom Grad eins in $x-a$ sind gerade $x_r - a_r$ für $r = 1, \ldots, n$.

43.14½ Folgerung Unter denselben Voraussetzungen über X ist die Menge

$$\big\{f \in C^\infty(X) \,\big|\, T_a^{k-1} f = 0\big\}$$

das von allen Funktionen $x \mapsto (x-a)^j$ mit $|j| = k$ erzeugte Ideal.

Dieses Resultat ist auch im Eindimensionalen neu für uns: Wenn für eine C^∞-Funktion f die Ableitungen $f(a)$, $f'(a)$, ..., $f^{(k-1)}(a)$ alle verschwinden, kann man aus $f(x)$ den Faktor $(x-a)^k$ herausziehen, und es bleibt immer noch eine C^∞-Funktion. Zwar ist das für analytische Funktionen trivial, aber für beliebige C^∞-Funktionen nicht.

43.15 Beispiel Die C^∞-Funktion $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ aus den Aufgaben 14.8 und 14.9 mit

$$f(x) = \begin{cases} 0 & (x \le 0) \\ e^{-\frac{1}{x}} & (x > 0) \end{cases}$$

hat im Nullpunkt die triviale Taylor-Reihe. Aus $f(x)$ kann man hier also so viele Faktoren x herausziehen, wie man will! Eine analytische Funktion mit dieser Eigenschaft müßte dagegen identisch verschwinden.

Übungsaufgaben

43.1 $f(X,Y)$, $g(X,Y) \in K[X,Y]$ seien zwei Polynome der Grade $d, e \in \mathbb{N}$ (mit dieser Voraussetzung soll wie früher auch ausgeschlossen sein, daß f oder g das Nullpolynom ist). Beweisen Sie, daß $f \cdot g \in K[X,Y]$ dann ein Polynom vom Grad $d+e$ ist. (Das ist auch für Polynome in $n > 2$ Variablen richtig und ebenso zu beweisen, der Beweis aber etwas komplizierter zu formulieren.) Begründen Sie, warum das Polynom $X^2 + X^3 - Y^2 \in \mathbb{R}[X,Y]$ nicht Produkt zweier Polynome von echt kleinerem Grad sein kann.

43.2 Sei X eine Mannigfaltigkeit. Im Abschnitt 37 ist anläßlich der Notation $\frac{\partial}{\partial h_j}$ erklärt, wie ein Vektorfeld $v \in \mathrm{Vect}X$ auf Funktionen $f \in C^\infty(X)$ wirkt:

$$v(f) := df \circ v\colon X \xrightarrow{v} TX \xrightarrow{df} \mathbb{R}$$

ist eine neue C^∞-Funktion auf X. Beweisen Sie: Ist $w \in \mathrm{Vect}X$ ein weiteres Vektorfeld und $a \in X$ ein kritischer Punkt von f, so gilt für die Hesse-Form von f dort

$$H_f\big(v(a), w(a)\big) = v\big(w(f)\big)(a) = w\big(v(f)\big)(a).$$

43.3 Berechnen Sie das sechste Taylor-Polynom der Funktion

$$f\colon \mathbb{R}^3 \longrightarrow \mathbb{R}; \quad f(x,y,z) := (x + z^3) \cdot \sin \frac{y}{1+z^2}$$

im Nullpunkt.

43.4 Beweisen Sie die folgende Aussage, die man sich als eine parametrisierte Version von Lemma 43.14 vorstellen kann: Die C^∞-Funktion $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ verschwinde auf $\{0\} \times \mathbb{R}^{n-1}$ identisch. Dann gibt es eine C^∞-Funktion $g\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ mit

$$f(x) = x_1 \cdot g(x) \quad \text{für alle } x \in \mathbb{R}^n.$$

44 Morse-Punkte

Ich gebe zu, daß ich Ihnen bisher keine echte Anwendung der höheren Ableitungen genannt habe; jetzt kommt aber eine. Erinnern wir uns noch mal an den Satz vom regulären Punkt: Jede differenzierbare Funktion $f\colon X \longrightarrow \mathbb{R}$ schreibt sich in einer geeigneten Karte um einen solchen Punkt a als die lineare Projektion auf eine der Koordinaten, jedenfalls wenn man den Wert $f(a)$ als Konstante abzieht. Man kann das auch so auslegen, daß alle differenzierbaren Funktionen lokal, in der Nähe ihrer regulären Punkte, gleich aussehen: diese Deutung ist besonders dann sinnvoll, wenn X eine Mannigfaltigkeit ist, auf der ja von vornherein keine bestimmte Karte gegeben sind. Kann man nun in ähnlicher Weise auch etwas über die Struktur der kritischen Punkte von f aussagen? Ob man das kann und wie f in der Nähe eines solchen Punktes aussieht, darüber entscheiden zunächst mal die zweiten Ableitungen, also die Hesse-Form von f. Wie, das werde ich genau beschreiben. Der Einfachheit halber seien jetzt wieder alle Daten beliebig oft differenzierbar.

44.1 Definition X sei eine n-dimensionale Mannigfaltigkeit, $f\colon X \longrightarrow \mathbb{R}$ eine differenzierbare Funktion und $a \in X$ ein kritischer Punkt von f. Dann heißt a ein Morse-Punkt von f, wenn die Hesse-Form $H_a f$ den größtmöglichen Rang, also den Rang n hat. Wenn alle kritischen Punkte von f solche Morse-Punkte sind, nennt man f eine Morse-Funktion.

Bemerkungen Man kann zeigen, daß "fast jede" differenzierbare (skalare) Funktion auf einer gegebenen Mannigfaltigkeit eine Morse-Funktion ist, in einem ähnlichen Sinne, wie wir diesen Ausdruck in der Maßtheorie verwendet haben. — Der amerikanische Mathematiker Marston Morse hat zwar nicht den Bekanntheitsgrad seines Namensvetters Samuel, aber während dessen Telegraphenalphabet schon wieder in Vergessenheit gerät, dürfte es sich bei den Morse-Funktionen und der Theorie, die er darauf aufgebaut hat, um eine bleibende Schöpfung handeln.

44.2 Satz (**Morse-Lemma**) $a \in X$ sei ein Morse-Punkt der Funktion $f\colon X \longrightarrow \mathbb{R}$. Dann gibt es eine Karte (U, h) um a, in der f die Gestalt

$$f \circ h^{-1}\colon h(U) \longrightarrow \mathbb{R};\ x \mapsto f(a) + q(x)$$

mit einer quadratischen Fom $q\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ hat.

Bemerkung Während f in der Nähe eines regulären Punktes in einer geeigneten Karte linear ist, ist f in der Nähe eines Morse-Punktes also in einer geeigneten Karte quadratisch.

Beweis Es handelt sich um eine besonders pfiffige Anwendung des Satzes von der lokalen Umkehrung. Wir treffen erst einige Vorbereitungen: Mittels einer beliebigen Karte um den Punkt a reduzieren wir auf den Fall, daß $a = 0 \in \mathbb{R}^n$ und $X \subset \mathbb{R}^n$ eine offene Kugel ist. Außerdem dürfen wir von f den Wert $f(0)$ als Konstante abziehen (und am Schluß wieder addieren), also $f(0) = 0$ annehmen.

Damit bleibt von der Taylor-Formel aus Satz 43.12 für $k=1$ nur das Restglied:

$$f(x) = 2\sum_{|j|=2} \int_0^1 (1-t)\frac{D_j f(tx)}{j!}\,dt \cdot x^j = \frac{1}{2}\sum_{r,s=1}^{n} H_{rs}(x)\,x_r x_s = \frac{1}{2}x^{\mathrm{t}} H(x)\,x \quad \text{für alle } x \in X,$$

wenn man die differenzierbare Abbildung $H\colon X \longrightarrow \mathrm{Sym}(n, \mathbb{R})$ durch

$$H_{rs}(x) := 2\int_0^1 (1-t) D_r D_s f(tx)\,dt = 2\int_0^1 (1-t)\frac{\partial^2 f}{\partial x_r \partial x_s}(tx)\,dt$$

erklärt. Dann ist $H(0) = Hf(0)$ die Hesse-Matrix, und wir haben f als eine "verlogen-quadratische" Form geschrieben, in der die symmetrische Koeffizientenmatrix H selbst von x abhängt; so war die Behauptung des

Morse-Lemmas freilich nicht gemeint! Vielmehr geht es darum, durch Wechsel zu einer geschickt konstruierten Karte $X \supset U \xrightarrow{h} \mathbb{R}^n$ die Abhängigkeit der Matrix H von x zu beseitigen. Passenderweise setzen wir denn auch h als "verlogen-lineare" Abbildung $h(x) = \varphi(x)\cdot x$ mit einer differenzierbaren Abbildung

$$\varphi\colon U \longrightarrow \mathrm{Mat}(n\times n, \mathbb{R})$$

an. Worauf es jetzt ankommt, das formulieren wir in einem

Hilfssatz Es gibt eine den Nullpunkt enthaltende offene Menge $U \subset X$ und eine differenzierbare Abbildung $\varphi\colon U \longrightarrow \mathrm{Mat}(n\times n, \mathbb{R})$ mit $\varphi(0) = 1$ und

$$H(x) = \varphi(x)^{\mathrm{t}} H(0)\, \varphi(x) \quad \text{für alle } x \in U.$$

Aus diesem Hilfssatz folgt das Morse-Lemma schnell: Wegen

$$Dh(0) = D\varphi(0)\cdot 0 + \varphi(0)\cdot 1 = \varphi(0) = 1$$

ist die durch φ bestimmte Abbildung $h\colon U \longrightarrow \mathbb{R}^n$ ein lokaler Diffeomorphismus bei $0 \in X$; durch Schrumpfen von U wird h dann zu einem globalen Diffeomorphismus $X \supset U \xrightarrow{h} h(U) \subset \mathbb{R}^n$, also einer Karte für X. Außerdem gilt

$$f(x) = \frac{1}{2} x^{\mathrm{t}} H(x)\, x = \frac{1}{2} x^{\mathrm{t}} \varphi(x)^{\mathrm{t}} H(0) \varphi(x)\, x = \frac{1}{2} h(x)^{\mathrm{t}} H(0)\, h(x) \quad \text{für alle } x \in U,$$

oder gleichwertig:

$$(f \circ h^{-1})(y) = \frac{1}{2} y^{\mathrm{t}} H(0)\, y \quad \text{für alle } y \in h(U)$$

Die Karte (U, h) leistet damit das Gewünschte.

Bleibt der Hilfssatz zu beweisen. Wir haben dort nicht ganz die Situation des Satzes über implizite Funktionen vor uns, denn die gesuchte Abbildung $\varphi\colon U \longrightarrow \mathrm{Mat}(n\times n, \mathbb{R})$ hat n^2 Komponenten, für die wir aber wegen $\varphi(x)^{\mathrm{t}} H(0)\, \varphi(x) \in \mathrm{Sym}(n, \mathbb{R})$ nur $n(n+1)/2$ Gleichungen haben. Tatsächlich beeinträchtigt das nicht die Lösbarkeit des Problems, lediglich deren Eindeutigkeit, die ja auch nicht behauptet wird.

Das Differential der Abbildung

$$\begin{array}{ccc} \mathrm{Mat}(n\times n, \mathbb{R}) & \xrightarrow{\Phi} & \mathrm{Sym}(n, \mathbb{R}) \\ y & \longmapsto & y^{\mathrm{t}} H(0)\, y \end{array}$$

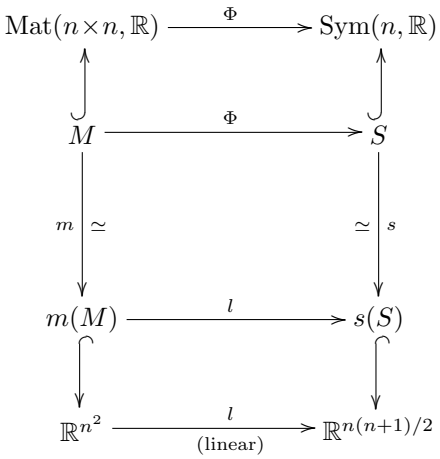
an der Stelle 1 ist

$$D\Phi(1)\colon \eta \longmapsto 1^{\mathrm{t}} H(0)\, \eta + \eta^{\mathrm{t}} H(0)\, 1 = H(0)\, \eta + \eta^{\mathrm{t}} H(0).$$

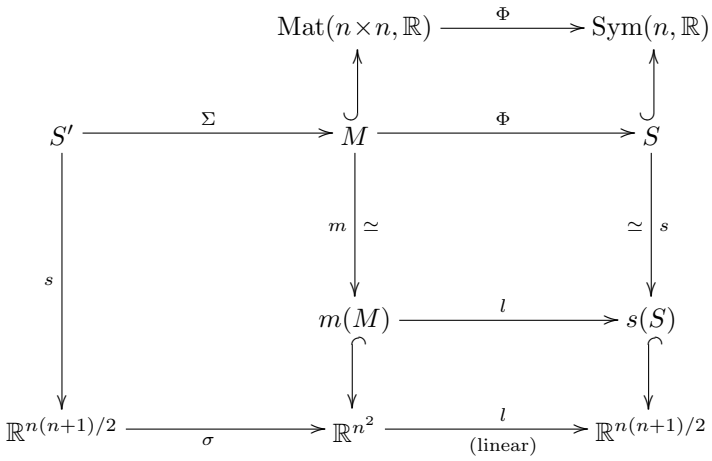
Nun ist $H(0)$, die Hesse-Matrix von f an der Stelle 0, nach Voraussetzung invertierbar, und daraus folgt durch direkte Rechnung, daß $D\Phi(1)$ surjektiv ist:

$$D\Phi(1)\left(\frac{1}{2} H(0)^{-1} \zeta\right) = \frac{1}{2}\zeta + \frac{1}{2}\zeta^{\mathrm{t}} = \zeta \quad \text{für jedes } \zeta \in \mathrm{Sym}(n, \mathbb{R}).$$

Mit anderen Worten ist 1 ein regulärer Punkt von Φ, und nach dem einschlägigen Satz ist Φ in geeigneten Karten (M, m) um $1 \in \mathrm{Mat}(n\times n, \mathbb{R})$ und (S, s) um $\Phi(1) = H(0) \in \mathrm{Sym}(n, \mathbb{R})$ linear:



Weil l surjektiv ist, gibt es eine lineare Abbildung $\sigma\colon \mathbb{R}^{n(n+1)/2} \longrightarrow \mathbb{R}^{n^2}$ mit $l \circ \sigma = \mathrm{id}$, und auf der offenen Menge $S' := S \cap s^{-1}\sigma^{-1}m(M)$ um $1 \in \mathrm{Sym}(n,\mathbb{R})$ definiert die Formel $\Sigma = m^{-1} \circ \sigma \circ s$ eine differenzierbare Abbildung Σ, die das Diagramm



kommutativ macht. Es gilt $\Sigma\big(H(0)\big) = 1$ und die Komposition $\Phi \circ \Sigma = \mathrm{id}$ ist die Inklusion $S' \subset S$. Wir schrumpfen jetzt U so weit, daß $H(U) \subset S'$ ist, was wegen der Stetigkeit von H ja möglich ist. Die Abbildung

$$\varphi\colon U \xrightarrow{H} S' \xrightarrow{\Sigma} M \subset \mathrm{Mat}(n\times n,\mathbb{R})$$

genügt dann allen Anforderungen: Es ist $\varphi(0) = \Sigma\big(H(0)\big) = 1$, und für alle $x \in U$ gilt

$$\varphi(x)^{\mathrm{t}} H(0)\varphi(x) = \Phi\big(\varphi(x)\big) = (\Phi\circ\Sigma)\big(H(x)\big) = H(x).$$

Jetzt ist der Hilfssatz, und damit auch das Morse-Lemma vollständig bewiesen.

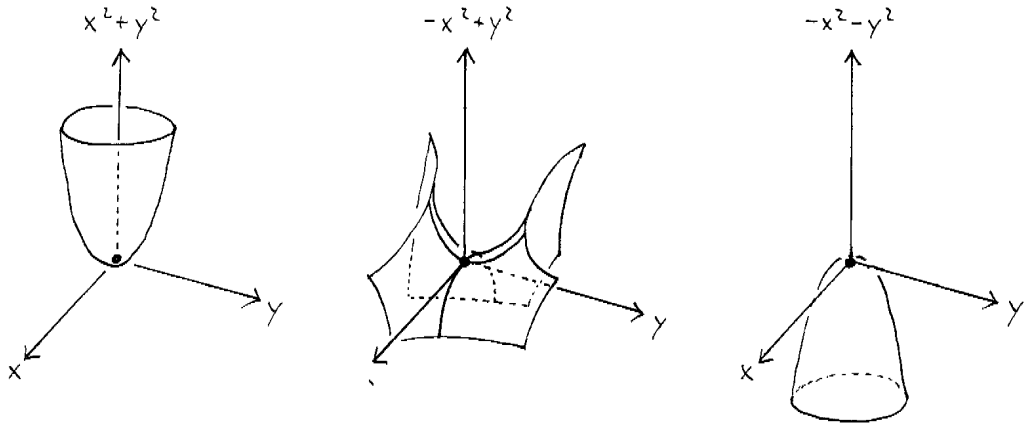
Das Morse-Lemma verspricht unter den genannten Voraussetzungen, daß man die Funktion f lokal um den Morse-Punkt a durch geeignete Kartenwahl in die Gestalt $x \longmapsto f(a) + q(x)$ mit einer quadratischen Form q bringen kann. Natürlich ist dann $2q$ die in dieser Karte geschriebene Hesse-Form von f; insbesondere hat q den vollen Rang n. Aufgrund unserer Kenntnisse über quadratische Formen wissen wir nun, daß q durch einen weiteren linearen Kartenwechsel diagonalisiert werden kann, nach Satz 29.3 sogar so, daß alle Diagonalkoeffizienten ± 1 oder 0 sind — aber 0 kann ja nicht vorkommen, weil $\mathrm{rk}\, q = n$ ist. Wir können das Morse-Lemma deshalb auch in der folgenden verschärften Fassung formulieren:

44.3 Folgerung und **Definition** Unter den Voraussetzungen des Morse-Lemmas 44.2 kann man die Karte h so einrichten, daß

$$(f\circ h^{-1})(x) = f(a) - \sum_{j=1}^{\lambda} x_j^2 + \sum_{j=\lambda+1}^{n} x_j^2 \quad \text{für alle } x \in h(U)$$

gilt. Die durch die Signatur $(n-\lambda,\lambda)$ von $H_a f$ offenbar eindeutig bestimmte Zahl $\lambda \in \{0,1,\ldots,n\}$ nennt man den Morse-Index von a bezüglich f.

Die Morse-Punkte sind schon deshalb interessanter als die regulären Punkte einer Funktion, weil auf einer n-dimensionalen Mannigfaltigkeit nicht nur einer, sondern $n+1$ verschiedene Typen von Morse-Punkten möglich sind. Für $n = 2$ kann man die zugehörigen Graphen leicht zeichnen:



Offenbar ist der Nullpunkt der einzige kritische Punkt der Funktion $\mathbb{R}^n \ni x \longmapsto -\sum_{j=1}^{\lambda} x_j^2 + \sum_{j=\lambda+1}^{n} x_j^2$, denn ihr Differential ist $(-x_1 \quad \cdots \quad -x_\lambda \quad x_{\lambda+1} \quad \ldots \quad x_n)$. Für die Morse-Punkte folgt daraus die

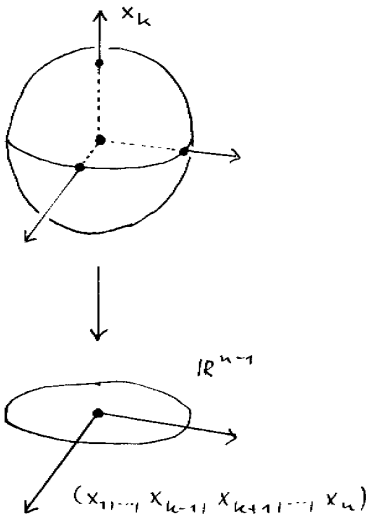
44.4 Notiz Jeder Morse-Punkt a einer Funktion $f\colon X \longrightarrow \mathbb{R}$ ist ein *isolierter* kritischer Punkt von f: es gibt eine a enthaltende offene Teilmenge U von X, so daß a der einzige kritische Punkt von f in U ist.

Ob ein kritischer Punkt einer gegebenen Funktion f ein Morse-Punkt ist und welchen Morse-Index dieser Punkt dann hat, läßt sich im Prinzip leicht feststellen: es kommt dafür ja nur auf den Rang bzw. die Signatur der (in irgendeiner Karte berechneten) Hesse-Matrix an. Ich will zur Illustration das Beispiel 37.8 fortführen, in dem wir die kritischen Punkte mittels der Methode der Lagrange-Multiplikatoren ermittelt hatten.

44.5 Beispiel f sei die Einschränkung der quadratischen Form $q\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ auf die Sphäre $S^{n-1} \subset \mathbb{R}^n$. Wir hatten in 37.8 festgestellt, daß die kritischen Punkte von f gerade die zu S^{n-1} gehörenden Eigenvektoren von q (d.h. der Matrix von q) sind, also die Eigenvektoren der Länge 1. Ein solcher kritischer Punkt von f ist nur dann isoliert, wenn der zugehörige Eigenwert einfach ist; nur dann kann es sich also um einen Morse-Punkt von f handeln. Um zu klären, ob das der Fall ist, empfiehlt es sich, q als diagonal anzunehmen, was nach dem Satz über die Hauptachsentransformation ja keine Einschränkung ist. Dann ist also

$$q(x) = \sum_{j=1}^{n} \lambda_j x_j^2$$

mit den Eigenwerten $\lambda_1, \ldots, \lambda_n$. Sei nun λ_k ein einfacher Eigenwert; wir wollen sehen, ob die beiden zugehörigen kritischen Punkte $\pm e_k \in S^{n-1}$ Morse-Punkte von f sind. Dazu schreiben wir f in der schon häufiger verwendeten Hemisphärenkarte $(x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_n)$,



müssen also $x_k = \pm\sqrt{1 - \sum_{j\neq k} x_j^2}$ in $q(x)$ substituieren und erhalten

$$(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n) \longmapsto \lambda_k \left(1 - \sum_{j\neq k} x_j^2\right) + \sum_{j\neq k} \lambda_j x_j^2 = \lambda_k + \sum_{j\neq k} (\lambda_j - \lambda_k) x_j^2.$$

Daraus liest man sofort alles ab: $\pm e_k$ sind tatsächlich Morse-Punkte von f, und ihr Morse-Index ist die Anzahl der Eigenwerte λ_j, die kleiner als λ_k sind. Sind insbesondere alle Eigenwerte paarweise verschieden, so kommt bei den kritischen Punkten von f jeder der möglichen Morse-Indizes $0, 1, \dots, n-1$ genau zweimal vor.

Vorsicht Die Hesse-Form von $f = q|S^{n-1}$ an der Stelle a ist nicht dasselbe wie die Einschränkung von $H_a q$ auf den Tangentialraum $T_a S^{n-1}$. Man muß in einer derartigen Situation wirklich eine Karte h um a heranziehen, darf diese allerdings nach der Kettenregel durch ihr zweites Taylor-Polynom bei a ersetzen.

Der Satz vom regulären Punkt und das Morse-Lemma geben einem ganz nebenbei eine zwar nicht vollständige, aber praktisch oft ausreichende Auskunft darüber, an welchen Stellen eine differenzierbare Funktion ein lokales Extremum annimmt. Denn diese Frage kann man in Karten prüfen, und weil eine nicht-konstante lineare Funktion sicher kein lokales Extremum besitzt und unter den quadratische Formen von vollem Rang nur die definiten im Nullpunkt extremal werden, ergibt sich der

44.6 Satz Sei X eine n-dimensionale Mannigfaltigkeit und $f\colon X \longrightarrow \mathbb{R}$ differenzierbar. Dann gilt:

- Wenn f bei $a \in X$ ein lokales Extremum hat, dann ist a ein kritischer Punkt von f.
- Ist a ein Morse-Punkt von f, so hat f bei a ein lokales Minimum/Maximum genau dann, wenn a den Morse-Index 0 bzw. n hat.

Der Satz ist auch in der Praxis leicht anzuwenden: Man berechnet die Nullstellen von Df — gegebenenfalls mit dem Verfahren 37.7 der Lagrange-Multiplikatoren — und prüft etwa mittels der Regel der Aufgabe 29.2, ob die Hesse-Form dort definit ist. Bevor man dieses Verfahren ablaufen läßt, sollte man sich aber vergewissern, daß der Definitionsbereich der Funktion f tatsächlich eine (randlose) Mannigfaltigkeit ist. Das zu vergessen, ist ein häufiger, aber gravierender und plumper Fehler: Es würde ja auch niemand behaupten, die Funktion $[0,1] \ni x \mapsto x \in \mathbb{R}$ nehme kein Minimum an, nur weil ihre Ableitung nirgends verschwindet! So wie man in diesem Beispiel die Endpunkte eben separat untersuchen muß, ist der Definitionsbereich im allgemeinen Fall nötigenfalls als eine Vereinigung von Mannigfaltigkeiten zu schreiben, die durchaus verschiedene Dimensionen haben dürfen.

44.7 Beispiel Es seien die globalen Extrema der Funktion

$$f\colon D^2 \longrightarrow \mathbb{R};\ f(x,y) = x^2 + x + 2y^2$$

gesucht. Man bestimmt dann zuerst getrennt die lokalen Extrema von $f|U^2$ und $f|S^1$. Die Gleichung

$$0 = Df(x,y) = \begin{pmatrix} 2x+1 & 4y \end{pmatrix}$$

gibt $(-\frac{1}{2}, 0)$ als einzigen kritischen Punkt; die Hesse-Form dort ist $(x,y) \mapsto x^2 + 2y^2$, wir haben also ein ein lokales Minimum mit Wert $-\frac{1}{4}$ vor uns. Die Einschränkung auf S^1 kann man mit einem Lagrange-Multiplikator λ behandeln; aus der Gleichung

$$\begin{pmatrix} 2x+1 & 4y \end{pmatrix} = Df(x,y) = \lambda \begin{pmatrix} 2x & 2y \end{pmatrix}$$

ergibt sich $\lambda = 2$ oder $y = 0$ und in jedem Fall $2(\lambda-1)x = 1$; also hat $f|S^1$ die vier kritischen Punkte

$$(x,y) = \left(\frac{1}{2}, \pm\frac{1}{2}\sqrt{3}\right) \quad \text{und} \quad (x,y) = (\pm 1, 0)$$

mit den Werten $\frac{9}{4}$, $\frac{9}{4}$, 2 und 0. Damit nimmt f das globale Minimum $-\frac{1}{4}$ bei $(-\frac{1}{2}, 0)$ und das globale Maximum $\frac{9}{4}$ zweimal, bei $\left(\frac{1}{2}, \pm\frac{1}{2}\sqrt{3}\right)$ an; die beiden Stellen $(\pm 1, 0)$ kommen nicht zum Zuge (ohne daß deswegen ihre Bestimmung überflüssig gewesen wäre).

Das unerläßliche Zerlegen des Definitionsbereiches in Mannigfaltigkeiten kann auch schon mal mehr Mühe machen: Ist zum Beispiel der Definitionsbereich der zu minimierenden Funktion f ein kompakter Würfel, so hat man f jeweils separat auf Würfelinnerem, -flächen, -kanten und -ecken zu untersuchen.

Der Satz 44.6 über die lokalen Extrema liegt viel weniger tief als das Morse-Lemma, auf das ich mich als Beweis berufen habe. Insbesondere folgt die Tatsache, daß eine C^2-Funktion in einem kritischen Punkt mit positiv definiter Hesse-Form ein lokales Minimum hat, ziemlich leicht aus dem Satz von Taylor. Darüber hinaus liefert dieser Zugang auch noch eine Teilinformation für kritische Punkte, die keine Morse-Punkte sind:

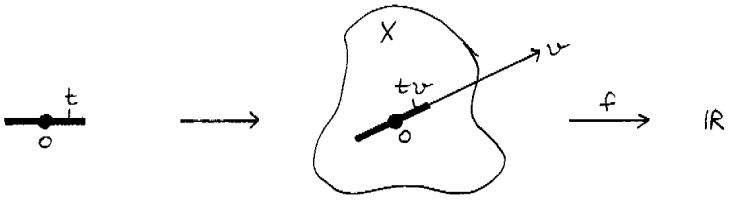
44.8 Lemma Sei X eine Mannigfaltigkeit, $f\colon X \longrightarrow \mathbb{R}$ eine C^2-Funktion und $a \in X$ ein kritischer Punkt von f. Wenn die Hesse-Form $H_a f$ mindestens einen negativen Eigenwert besitzt, dann hat f bei a sicher kein lokales Minimum.

Beweis Wir dürfen annehmen, daß $X \subset \mathbb{R}^n$ offen und sternförmig bezüglich $a = 0$, und daß $f(a) = 0$ ist. Mit den Bezeichnungen aus dem Beweis des Morse-Lemmas können wir f dann als "verlogen-quadratische" Form

$$f(x) = \frac{1}{2} x^{\mathrm{t}} H(x)\, x \quad \text{für alle } x \in X$$

schreiben, in der $H\colon X \longrightarrow \mathrm{Sym}(n, \mathbb{R})$ diesmal immerhin noch stetig ist. Nach Voraussetzung gibt es nun einen Vektor $v \in \mathbb{R}^n$ mit $v^{\mathrm{t}} H(0)\, v < 0$. Die auf einem offenen Intervall um $0 \in \mathbb{R}$ erklärte Funktion

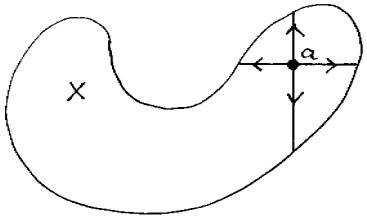
$$t \longmapsto f(tv) = \frac{1}{2}(tv)^{\mathrm{t}} H(tv)\,(tv) = \frac{1}{2}\big(v^{\mathrm{t}} H(tv)\, v\big) \cdot t^2$$



hat dann für genügend kleine $|t| \neq 0$ negative Werte, deshalb kann f bei 0 kein lokales Minimum haben.

Grundsätzlich liegen die Verhältnisse aber wie im Eindimensionalen: Nur im "gewöhnlichen" Fall kann man aus der Kenntnis der ersten und zweiten Ableitungen die lokale Geometrie der Funktion (insbesondere das eventuelle Extremalverhalten) ablesen; es bleiben aber immer "entartete" Fälle (zum Beispiel kritische Punkte mit verschwindender Hesse-Form), in denen das nicht möglich ist und die weitere Untersuchung beliebig kompliziert werden kann.

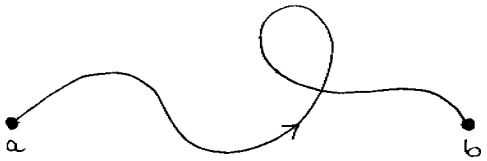
Auch die Tatsache, daß ein lokales Extremum nur in einem kritischen, nicht aber einem regulären Punkt vorliegen kann, ist viel einfacher zu begründen als mit dem Satz vom regulären Punkt. Ist nämlich $X \subset \mathbb{R}^n$ offen und hat $f\colon X \longrightarrow \mathbb{R}$ bei $a \in X$ ein lokales Extremum, so hat dort ja auch die Einschränkung von f auf jede achsenparallele Gerade durch a ein lokales Extremum; deshalb müssen alle partiellen Ableitungen von f an der Stelle a verschwinden, solange sie nur existieren.



Bei der Suche nach den lokalen Extrema einer gegebenen Funktion kommt der erste Schritt, die Suche nach den kritischen Punkten als Kandidaten für die zu bestimmenden Stellen, also ganz mit den Methoden der eindimensionalen Analysis aus, und erst die subtilere Frage, ob bei einem solchen Kandidaten wirklich ein

lokales Extremum vorliegt, ist ein wesentlich mehrdimensionales Problem. Nachdem wir die mehrdimensionale Analysis ohnehin studiert haben, wäre diese Bemerkung an sich belanglos. Sie wird aber doch interessant, wenn man sich für Extrema von Funktionen interessiert, die auf nicht endlichdimensionalen Räumen definiert sind: dann sind ja auch unsere mehrdimensionalen Methoden nicht mehr anwendbar. Derartige Fragestellungen sind in der Physik als Variationsprinzipien geläufig, und ich will das zum Abschluß der Vorlesung mit einem bekannten Beispiel illustrieren.

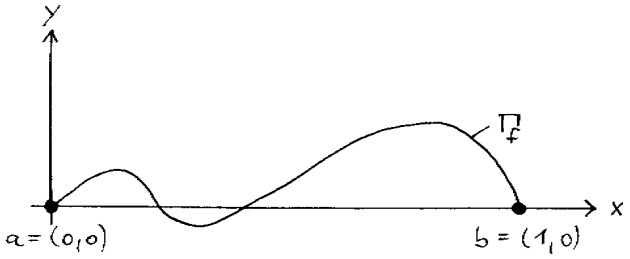
44.9 Beispiel Das Fermatsche Prinzip der geometrischen Optik besagt, daß die Bahn eines Lichtstrahls zwischen zwei gegebenen Punkten a und b unter allen denkbaren Bahnen die schnellste, also die mit der kürzesten Gesamtreisezeit ist.



Mittels dieses Prinzips den Lichtstrahl zu bestimmen, bedeutet die Funktion T zu betrachten, die jedem Weg von a nach b die Reisezeit zuordnet, und herauszufinden, wo T das absolute Minimum annimmt. Es ist klar, daß der Definitionsbereich X von T kein endlichdimensionaler Raum sein kann, so daß wir uns nicht auf die in diesem Abschnitt besprochenen Sätze berufen können. Wenn es aber nur darum geht, einen Kandidaten für den rechten Weg zu finden, kann man sich helfen. Wir beschränken uns der Einfachheit halber auf das ebene Problem, setzen $a = (0,0) \in \mathbb{R}^2$ und $b = (1,0) \in \mathbb{R}^2$, und ziehen nur Wege in Betracht, die sich als Graph einer Funktion aus

$$X = \big\{ f \in C^2[0,1] \,\big|\, f(0) = f(1) = 0 \big\}$$

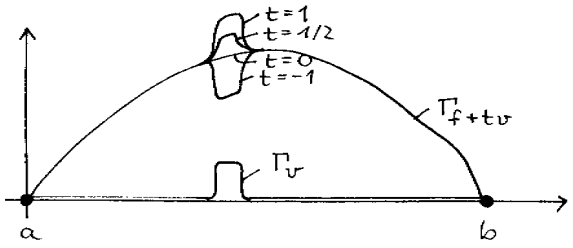
schreiben lassen.



Wenn wir uns die Ebene mit einem isotropen Medium von ortsabhängigem Brechungsindex $n = n(x,y)$ ausgefüllt denken, dann ist

$$T(f) = \int_0^1 n\big(x, f(x)\big) \sqrt{1 + f'(x)^2}\, dx.$$

Für jede Wahl von $v \in X$ beschreibt nun $\mathbb{R} \ni t \mapsto f + tv \in X$ eine Art Deformation von f; man denke dabei insbesondere an solche v, die nur auf einem kurzen in $(0,1)$ enthaltenen Teilintervall von null verschiedene Werte annehmen.



Wir setzen jetzt voraus, daß T bei bei f den kleinsten Wert annimmt, also daß Γ_f die Bahn des wirklichen Lichtstrahls ist. Die zu v gehörige Hilfsfunktion

$$\mathbb{R} \ni t \longmapsto T(f+tv) \in \mathbb{R}$$

muß dann an der Stelle $t = 0$ ihren kleinsten Wert annehmen, also ist

$$0 = \frac{d}{dt}T(f)\bigg|_{t=0} = \frac{d}{dt}\int_0^1 n\big(x, f(x)+tv(x)\big)\sqrt{1+\big(f'(x)+tv'(x)\big)^2}\,dx\bigg|_{t=0}$$

oder, weil man unter dem Integral differenzieren darf,

$$0 = \int_0^1 \left(\frac{\partial n}{\partial y}\big(x, f(x)\big)\,v(x)\sqrt{1+f'(x)^2} + n\big(x, f(x)\big)\frac{f'(x)\,v'(x)}{\sqrt{1+f'(x)^2}}\right)dx.$$

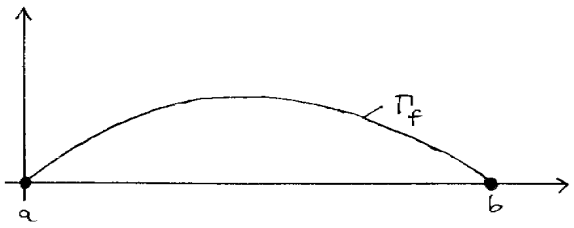
In dieser Form nützt das nicht viel, denn wir können die Tatsache, daß v frei wählbar war, nicht so recht ins Spiel bringen. Wenn man aber den zweiten Summanden mit partieller Integration behandelt und dabei $v(0) = v(1) = 0$ berücksichtigt, erhält man stattdessen

$$0 = \int_0^1 \left(\frac{\partial n}{\partial y}\big(x, f(x)\big)\sqrt{1+f'(x)^2} - \frac{d}{dx}\frac{n\big(x, f(x)\big)f'(x)}{\sqrt{1+f'(x)^2}}\right)v(x)\,dx.$$

Wie ein einfaches Stetigkeitsargument zeigt, kann das nur dann für jede beliebige Wahl von $v \in X$ gelten, wenn der Inhalt der großen Klammer identisch verschwindet. Wenn man den ausrechnet und etwas aufräumt, erhält man in

$$n\big(x, f(x)\big)f''(x) = \frac{\partial n}{\partial y}\big(x, f(x)\big)(1+f'(x)^2) - \frac{\partial n}{\partial x}\big(x, f(x)\big)f'(x)(1+f'(x)^2)$$

eine Differentialgleichung für den tatsächlichen Lichtstrahl Γ_f. Sie ist im allgemeinen nicht explizit lösbar (hängt ja auch von der Funktion n ab); immerhin reduziert sie sich für konstantes $n \neq 0$ auf $f''(x) = 0$ mit der geradlinigen Verbindung von a nach b als Lösung, was ja Vertrauen weckt. Als weiterer explizit lösbarer Fall erweist sich $n(x, y) = \frac{1}{\nu+y}$ mit $\nu > 0$; das Licht folgt dann einem Kreisbogen mit Mittelpunkt $\big(\frac{1}{2}, -\nu\big)$:



$$+^{\left(\frac{1}{2}, -\nu\right)}$$

$$\left(x - \frac{1}{2}\right)^2 + \big(f(x) + \nu\big)^2 = \nu^2 + \frac{1}{4}$$

Übungsaufgaben

44.1 Verifizieren Sie, daß die Funktion $f\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ mit

$$f(x,y) = x^3 + y^3 - 3xy$$

eine Morse-Funktion ist, und bestimmen Sie Lage und Morse-Index der kritischen Punkte (insbesondere der lokalen Extrema).

44.2 Sei $X \subset \mathbb{R}^n$ offen und $a \in X$ ein kritischer Punkt der C^2-Funktion $f\colon X \longrightarrow \mathbb{R}$. Beweisen Sie direkt mittels des Satzes von Taylor: Ist $Hf(a)$ positiv definit, so hat f an der Stelle a ein strenges lokales Minimum.

44.3 Verifizieren Sie am Beispiel 44.5, daß die dort ausgesprochene Warnung berechtigt ist; man kann $H_f(\pm e_k)$ im allgemeinen tatsächlich nicht als Einschränkung der Hesse-Form auf den Tangentialraum berechnen. Andererseits scheint das Lemma 43.11 diese Vorgehensweise gerade zu rechtfertigen. Warum nur scheinbar?

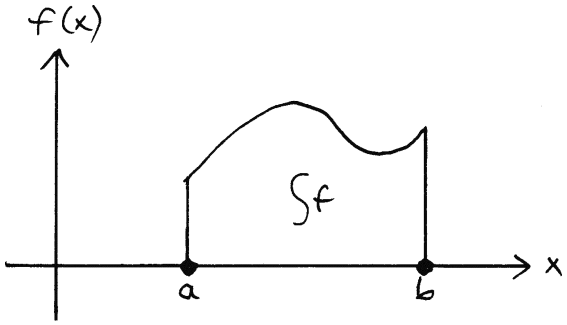
44.4 Die Funktion $f\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ sei durch

$$f(x,y) = x^4 + 3x^2y + 2y^2$$

definiert. Zeigen Sie, daß die Einschränkung von f auf jede durch den Nullpunkt laufende Gerade dort ein lokales Minimum hat. Hat auch f selbst ein lokales Minimum im Nullpunkt?

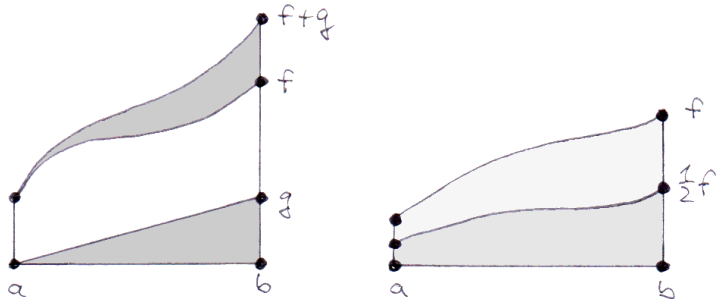
15 1/2 Einstieg in die Integralrechnung

Die Kernidee des Integrierens ist die Flächenberechnung. Im einfachsten Fall geht es darum, für eine auf einem kompakten Intervall definierte stetige Funktion $f\colon [a,b] \to \mathbb{R}$ mit nur positiven Werten den unter ihrem Graphen liegenden Flächeninhalt — eine $\int f$ geschriebene reelle Zahl — zu erklären und zu berechnen.

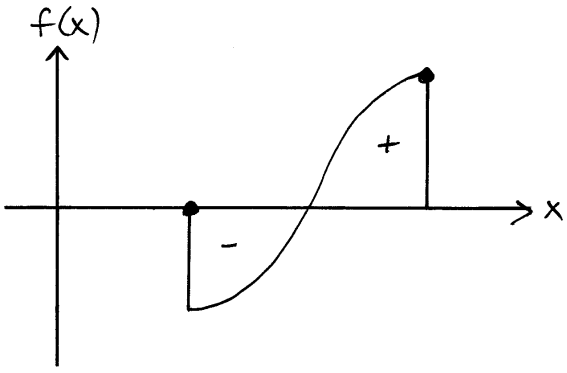


Welche Gesetzmäßigkeiten legt die Anschauung dabei nahe? Nun, erst mal eine Linearität genannte Eigenschaft

$$\int (f+g) = \int f + \int g$$
$$\int (\lambda f) = \lambda \cdot \int f\,,$$



jedenfalls solange der konstante Faktor $\lambda \in \mathbb{R}$ nicht negativ ist. Ein erster mathematischer Kniff besteht darin, auch Funktionen mit nicht unbedingt positiven Werten zuzulassen, dabei die Flächenanteile unter der Abszissenachse negativ zu zählen:



Dann dürfen wir die Linearitätsformel sogar für beliebige $\lambda \in \mathbb{R}$ ins Auge fassen; wir wissen ja, daß mit f und g auch $f+g$ und λf stetige Funktionen sind.

Damit erhält eine zweite — wie alles soweit natürlich nur auf die Anschauung gegründete — Regel Sinn. Der bequemen Verständigung halber vereinbaren wir:

$15\frac{1}{2}$.1 Schreibweise Sei X eine beliebige Menge. Für je zwei reellwertige Funktionen $f, g\colon X \to \mathbb{R}$ schreiben wir $f \le g$ als Kurzform für

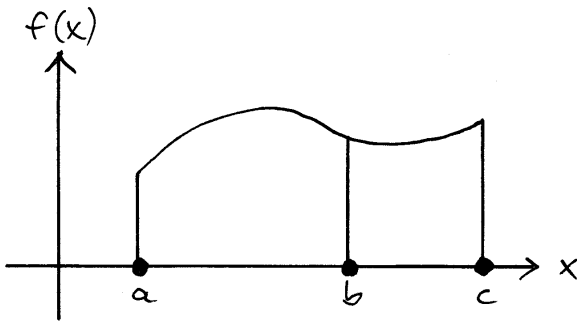
$$f(x) \le g(x) \quad \text{für alle } x \in X\,.$$

Eine Funktion f mit $f \ge 0$ nennen wir konsequenterweise *nicht-negativ*.

Beachten Sie, daß das zwar eine Art Ordnung auf der Menge der Funktionen erklärt, diese aber nicht je zwei Funktionen auf X miteinander vergleichbar macht: anders als bei den reellen Zahlen würde man hier nur von einer *partiellen Ordnung* sprechen.

Die Regel, die sich nun aufdrängt: Wenn $f\colon [a,b] \to \mathbb{R}$ eine Funktion mit $f \ge 0$ ist, dann ist auch $\int f \ge 0$.

In jedem Fall erwarten wir auch, daß der Flächeninhalt stückweise berechnet werden kann: für $a \le b \le c$ und stetiges $f\colon [a,c] \to \mathbb{R}$ wird die Regel $\int f = \int f|[a,b] + \int f|[b,c]$

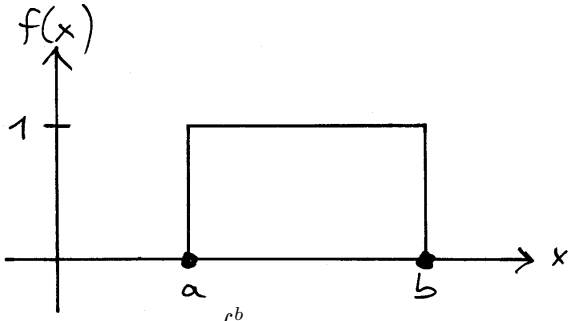


gelten, prägnanter geschrieben

$$\int_a^c f = \int_a^b f + \int_b^c f\,,$$

wenn wir vereinbaren, bei Bedarf auch die Intervallgrenzen in die Integralnotation aufzunehmen.

Schließlich brauchen wir eine Regel, die sozusagen den Maßstab festlegt, nämlich den Flächeninhalt eines Rechtecks der Höhe 1:



$$\int_a^b 1 = b - a.$$

All das sind erst mal nur heuristische Überlegungen; es ist gar nicht klar, ob so ein Fächen- oder Integralbegriff existiert und ob er dann eindeutig ist. Wie schon bei früheren Gelegenheiten ist es deshalb zweckmäßig, die Regeln zu Axiomen zu machen und damit die Existenz- und Eindeutigkeitsfrage zu verselbständigen.

15½.2 Definition Für jedes Paar von reellen Zahlen $a \le b$ bezeichne $C^0[a,b]$ die Menge aller stetigen Funktionen $f\colon [a,b] \to \mathbb{R}$. Eine Familie von Funktionen

$$\left(\int_a^b : C^0[a,b] \longrightarrow \mathbb{R}\right)_{a\le b}$$

mit den Eigenschaften

- Linearität: $\int_a^b (f+g) = \int_a^b f + \int_a^b g$ und $\int_a^b (\lambda f) = \lambda \cdot \int_a^b f$ für konstantes $\lambda \in \mathbb{R}$,
- Positivität: $f \ge 0 \quad\Longrightarrow\quad \int_a^b f \ge 0$,
- Unterteilbarkeit: $\int_a^c f = \int_a^b f + \int_b^c f$ falls $a \le b \le c$, und
- Normiertheit: $\int_a^b 1 = b-a$

heißt ein *Integral* (für stetige Funktionen auf kompakten Intervallen).

Anmerkungen Eine "Familie", wie sie hier erstmalig auftaucht, ist letztlich eine Zuordnung, hier eine, die jedem Paar $a \le b$ eine Funktion $C^0[a,b] \to \mathbb{R}$ zuweist. — Wie schon in unserer heuristischen Vorüberlegung kann man die *Integrationsgrenzen* a und b weglassen, wenn sie aus dem Zusammenhang klar sind; in umgekehrter Richtung kann man vor allem dann, wenn die zu integrierende Funktion — der *Integrand* — f durch eine explizite Formel gegeben ist, das Bedürfnis haben, ausführlicher

$$\int_a^b f = \int_a^b f(x)\,dx\,,$$

zum Beispiel $\int_1^2 x^3\,dx$ zu schreiben, eine ganz klassische Notation, in der das x analog zum i in $\sum_{i=1}^n x_i$ bloß ein austauschbarer Platzhalter ist. Übrigens verwendet man das Wort Integral nicht nur für die gesamte Familie, sondern auch für jede einzelne der Funktionen $\int_a^b : C^0[a,b] \longrightarrow \mathbb{R}$ — das Integral *von a bis b* — und deren Werte in $\mathbb{R}$: das Integral *über* eine Funktion (von a bis b).

Im ersten Teil dieses Abschnitts unterstellen wir beim Aufbau der Integralrechnung stillschweigend eine beliebige Wahl eines Integrals. Unsere Ergebnisse werden also für jedes Integral gelten, sollte es mehrere geben — und gegenstandslos (damit auch wahr) sein, falls gar kein Integral existiert. Im zweiten Teil beweisen wir dann, daß es in Wirklichkeit genau ein Integral gibt. Das rechtfertigt dann auch, daß wir für das Integral das spezielle Symbol $\int$ verwenden.

15½.3 Notiz Seien $f, g \in C^0[a,b]$. Dann gilt

$$f \le g \quad\Longrightarrow\quad \int f \le \int g$$

und

$$\left|\int f\right| \le \int |f|.$$

Beweis Aus $f \le g$ folgt $g-f \ge 0$, also

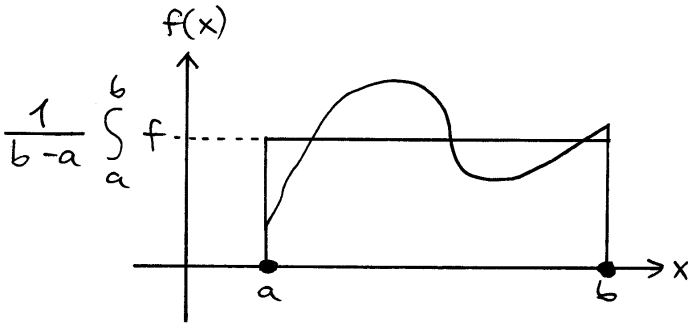
$$\int g - \int f = \int (g-f) \ge 0.$$

Wegen $-|f| \le f \le |f|$ haben wir insbesondere

$$-\int |f| \le \int f \le \int |f|.$$

$15\frac{1}{2}$.4 Mittelwertsatz (der Integralrechnung) Sei $a \le b$ und $f \in C^0[a,b]$. Dann gibt es eine Zahl $t \in [a,b]$ mit

$$\int_a^b f = f(t) \cdot (b-a).$$



Erläuterung Wie die Skizze zeigt, kann man (für $a < b$) die Zahl $\frac{1}{b-a}\int_a^b f$ als den "mittleren" Wert von f interpretieren. Versprochen wird also, daß dieser Mittelwert als tatsächlicher Wert von f vorkommt.

Beweis Nach dem Satz von der Annahme des Maximums ist

$$f([a,b]) = [c,d]$$

ein kompaktes Intervall, und wenn wir $c, d \in C^0[a,b]$ als konstante Funktionen lesen, gilt $c \le f \le d$. Nach den Integralaxiomen und $15\frac{1}{2}$.3 folgt

$$c \cdot (b-a) = c\int_a^b 1 = \int_a^b c \le \int_a^b f \le \int_a^b d = d \cdot \int_a^b 1 = d \cdot (b-a),$$

also gibt es ein $y \in [c,d]$ mit

$$\int_a^b f = y \cdot (b-a)$$

(im Fall $a < b$ ist $y = \int f/(b-a)$ eindeutig bestimmt). Wegen $f([a,b]) = [c,d]$ schließlich gibt es wie behauptet ein $t \in [a,b]$ mit $f(t) = y$.

Es ist sehr praktisch, die Bedeutung des Symbols $\int_a^b$ auf die Fälle auszudehnen, in denen $a > b$ ist, nämlich als ein Integral mit "Orientierungsvorzeichen":

$15\frac{1}{2}$.5 Definition und **Notiz** Sei $a \ge b$ und $f \in C^0[b,a]$. Dann wird

$$\int_a^b f := -\int_b^a f$$

definiert — widerspruchsfrei, weil im Fall $a = b$ ohnehin $\int_a^a f = 0$ ist (konstante Funktion auf dem Intervall $[a,a]$).

Während das Positivitätsaxiom unter der Voraussetzung $a \ge b$ natürlich zu

$$f \ge 0 \quad \Longrightarrow \quad \int_a^b f \le 0$$

anzuändern ist, gelten die übrigen drei Integralaxiome auch für das orientierte Integral wörtlich. Insbesondere darf man die Unterteilungsformel

$$\int_a^c = \int_a^b + \int_b^c$$

ohne Rücksicht auf die Lage der Punkte $a, b, c \in \mathbb{R}$ anwenden (zum Beweis braucht man nur die sechs möglichen Fälle durchzugehen).

Die folgende Differential- und Integralrechnung verbindende Tatsache ist grundlegend unter anderem für die Berechnung von Integralen — sie würde geradezu den Namen "Hauptsatz der Differential- und Integralrechnung" verdienen.

$15\frac{1}{2}$.6 Satz Sei $I \subset \mathbb{R}$ ein echtes Intervall, $a \in I$ und $f\colon I \to \mathbb{R}$ stetig. Dann ist die Funktion

$$I \ni x \longmapsto \int_a^x f \in \mathbb{R}$$

eine Stammfunktion von f:

$$\frac{d}{dx}\int_a^x f = f(x) \quad \text{für alle } x \in I.$$

Beweis Wir beweisen die Differenzierbarkeit an der Stelle $x \in I$. Sei also $\varepsilon > 0$. Weil f stetig ist, finden wir ein $\delta > 0$, so daß $|f(y)-f(x)| < \varepsilon$ für alle $y \in I$ mit $|y-x| < \delta$ gilt. Für ein beliebiges solches $y \neq x$ liefern die Notiz $15\frac{1}{2}$.5 und der Mittelwertsatz $15\frac{1}{2}$.4 ein t zwischen x und y mit

$$\int_a^y f - \int_a^x f = \int_x^y f = f(t)\cdot(y-x)$$

— auch wenn y links von x liegt, bleibt das richtig. Für den mit y gebildeten Differenzenquotienten folgt

$$\left|\frac{\int_a^y f - \int_a^x f}{y-x} - f(x)\right| = \big|f(t)-f(x)\big| < \varepsilon\,,$$

letzteres wegen $|t-x| \le |y-x| < \delta$. Damit sind wir schon fertig.

Traditionell bezeichnet man als Hauptsatz der Differential- und Integralrechnung aber nur die

$15\frac{1}{2}$.7 Folgerung Sei $I \subset \mathbb{R}$ ein echtes Intervall und $f\colon I \to \mathbb{R}$ stetig. Ist $F\colon I \to \mathbb{C}$ eine Stammfunktion von f, so gilt für alle $a, b \in I$

$$\int_a^b f = F(b) - F(a).$$

Beweis Nach $15\frac{1}{2}$.6 ist auch

$$G\colon x \longmapsto \int_a^x f$$

eine Stammfunktion von f. Die Differenz $F-G$ ist nach der Notiz 15.9 konstant, also

$$F(b) - F(a) = G(b) - G(a) = \int_a^b f - \int_a^a f = \int_a^b f.$$

Anmerkungen (1) Für die auftretende Differenz sind auch Schreibweisen wie

$$F(b) - F(a) = F(x)\big|_{x=a}^b = \big[F(x)\big]_{x=a}^b$$

gebräuchlich; sie lohnen, sobald $F(x)$ ein längerer Ausdruck ist.

(2) Der Sachverhalt $F' = f$ wird ebenso traditionell wie unlogisch auch

$$F(x) = \int f(x)\,dx$$

(ohne Grenzen am Integralzeichen) geschrieben, und F als "unbestimmtes" Integral von f bezeichnet. Die Tatsache, daß es sich um eine Aussage der Differential- und nicht der Integralrechnung handelt, wird dabei unter den Teppich gekehrt. Außerdem ist, wie wir wissen, F durch f nur bis auf eine additive Konstante festgelegt, und die Logik wird deshalb nur durch die etwas holprige Vereinbarung gerettet, daß ein Gleichheitszeichen zwischen solchen unbestimmten Integralen in Wirklichkeit nur Gleichheit bis auf Addition einer Konstanten anzeigt (wer aus der Algebra mit dem Begriff der Äquivalenzrelation oder Kongruenz vertraut ist, erkennt sofort, daß es sich um Kongruenz modulo der Untergruppe der konstanten Funktionen handelt). Natürlich muß man aufpassen, wenn man aus solchen "Gleichungen" Integralzeichen wegkürzt: zum Beispiel liefern gleich zu besprechende Standardmethoden alternativ

$$\int (x+1)\,dx = \frac{1}{2}(x+1)^2 \quad \text{oder} \quad \int (x+1)\,dx = \frac{1}{2}x^2 + x;$$

wenn man daraus durch scheinbar harmlosen Vergleich den Schluß

$$\frac{1}{2}(x+1)^2 = \frac{1}{2}\,x^2 + x$$

und damit $\frac{1}{2} = 0$ zieht, ist das eben auch nur bis auf Addition einer konstanten Funktion wahr, bloß wird man durch kein Integralzeichen mehr daran erinnert.

(3) Es mag Sie erstaunen, wenn ich hier und im folgenden ausdrücklich auch Stammfunktionen mit nichtreellen Werte in Betracht ziehe, obwohl die Integralrechnung nur von reellwertigen Funktionen auf Intervallen handelt. Wie die Sätze sofort zeigen, muß jede Stammfunktion einer solchen Funktion konstanten Imaginärteil haben, und der fällt bei der Differenzbildung sowieso wieder heraus. Trotzdem ist der komplexe Standpunkt bei den *unbestimmten* Integralen — also den Stammfunktionen — nicht nur legitim, sondern auch nützlich, weil uns die komplexe Differentialrechnung zusätzliche Methoden zur Bestimmung von Stammfunktionen zur Verfügung stellt.

Am einfachsten ist die Anwendung von $15\frac{1}{2}.7$ natürlich dann, wenn man eine Stammfunktion des Integranden "zufällig" kennt:

$15\frac{1}{2}.8$ Beispiele unbestimmter Integrale:

$$\int e^x\,dx = e^x\,,$$
$$\int \frac{dx}{\sqrt{1-x^2}} = \arcsin x\,, \quad \int \frac{dx}{1+x^2} = \arctan x\,,$$
$$\int x^b\,dx = \frac{1}{b+1}\,x^{b+1} \quad \text{für alle } b \in \mathbb{R}\setminus\{-1\};$$

die Lücke füllt

$$\int \frac{dx}{x} = \int \frac{1}{x}\,dx = \log|x|$$

(beachten Sie, daß hier in einem Integrationsintervall das Vorzeichen von x nicht wechseln kann, der Betrag also durchweg x oder durchweg $-x$ ist).

Um zu einer gegebenen Funktion eine Stammfunktion zu bestimmen, bietet es sich an, die bekannten Regeln für das Differenzieren entsprechend umzuschreiben. Völlig problemlos gelingt das mit der Linearität:

$$\int (\lambda f + \mu g) = \lambda \cdot \int f + \mu \cdot \int g\,.$$

Umschreiben der Ketten- und der Produktregel der Differentialrechnung führt dagegen nicht etwa auf eine Ketten- oder Produktregel der Integralrechnung, sondern zu spezielleren und weniger flexibel anwendbaren Regeln.

15½.9 Integrationsregeln (a) *Substitutionsregel*: für stetiges f und stetig differenzierbares φ derart, daß $f \circ \varphi$ definiert ist, gilt

$$\int (f \circ \varphi) \cdot \varphi' = \left(\int f \right) \circ \varphi.$$

(b) *Regel der partiellen Integration*: für stetig differenzierbare f und g gilt

$$\int f'g = f \cdot g - \int fg'.$$

Beweis (a) Sei X der Definitionsbereich von φ und F eine Stammfunktion von f auf einem $\varphi(X)$ umfassenden Intervall oder Gebiet. Dann gilt

$$(F \circ \varphi)' = (F' \circ \varphi) \cdot \varphi',$$

das heißt, daß $F \circ \varphi$ eine Stammfunktion von $(F' \circ \varphi) \cdot \varphi'$ ist, und genau das war behauptet. — In (b) ist nach der Produktregel $f \cdot g$ eine Stammfunktion von $f'g + fg'$.

15½.10 Beispiele (1) Eine Standardanwendung der Substitutionregel betrifft die Integrale der Form $\int f(\lambda x)\,dx$ mit konstantem $\lambda \neq 0$. Um die Regel mit $\varphi(x) = \lambda x$, also $\varphi'(x) = \lambda$ anzuwenden, schreibt man

$$\int f(\lambda x)\,dx = \frac{1}{\lambda} \int f(\lambda x)\lambda\,dx = \frac{1}{\lambda} \int f(y)\,dy \Big|_{y=\varphi(x)}.$$

Wie immer, wenn man die Substitutionregel zur Berechnung eines (bestimmten) Integrals anwendet, darf man nicht vergessen, die Substitution auch auf die Integrationsgrenzen anzuwenden. Also:

$$\int_a^b f(\lambda x)\,dx = \frac{1}{\lambda} \int_{\varphi(a)}^{\varphi(b)} f(y)\,dy = \frac{1}{\lambda} \int_{\lambda a}^{\lambda b} f(y)\,dy$$

Ganz egal für die Korrektheit der Substitution ist dagegen, ob ein reelles λ positiv oder negativ ist, ob φ also monton wächst oder fällt.

(2) Mit der Wahl $\lambda = i$ wird zum Beispiel

$$\int e^{ix}\,dx = \frac{1}{i} \int e^y\,dy \Big|_{y=ix} = -i\,e^{ix},$$

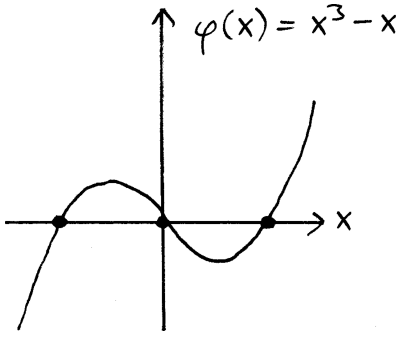
und als Real- und Imaginärteil dieser Gleichung lesen wir nochmal

$$\int \cos x\,dx = \sin x \quad \text{und} \quad \int \sin x\,dx = -\cos x$$

ab.

(3) Die substituierende Funktion φ braucht keineswegs injektiv zu sein: auch

$$\varphi(x) = x^3 - x$$



erfüllt die Voraussetzungen, liefert mit $\varphi'(x) = 3x^2 - 1$ also

$$\int f(x^3-x)\cdot(3x^2-1)\,dx = \int f(y)\,dy\,\Big|_{y=x^3-x}$$

und so zum Beispiel $\int_{-1}^{1} f(x^3-x)\cdot(3x^2-1)\,dx = \int_0^0 f(y)\,dy = 0$ für jede stetige Funktion f.

(4) Angesichts der Notation $\varphi'(x) = \frac{d\varphi}{dx}$ kann man symbolisch $d\varphi = \varphi'(x)\cdot dx$ schreiben, obwohl $d\varphi$ und dx allein gar keine Bedeutung haben. Wenn man in einem unbestimmten Integral

$$\int f\left(\varphi(x)\right)\varphi'(x)\,dx$$

dann φ substituiert und obige Identität einsetzt, erhält man

$$\int f\left(\varphi(x)\right)\varphi'(x)\,dx = \int f(\varphi)\,d\varphi$$

und macht es damit automatisch richtig! Diese einfache und empfehlenswerte Merkregel erinnert leider nicht an die dann noch vorzunehmende Substitution der Integrationsgrenzen.

Obwohl fleißiges Anwenden der Integrationsregeln $15\frac{1}{2}.9$ (und vielleicht noch weiterer) zu umfangreichen Formelsammlungen führt, soll man sich der Tatsache bewußt sein, daß die Integrationsregeln anders als die Differentiationsregeln aus den Abschnitten 13 und 14 keine Anleitung enthalten, um zu einer gegebenen "elementaren" (aus den gängigen Grundbausteinen mittels der gängigen Prozesse gebildeten) Funktion eine Stammfunktion zu berechnen: Die Regel der partiellen Integration drückt *nicht* das Integral eines Produkts durch Integrale über die Faktoren aus, und die Substitutionsregel ist *keine* Formel für das Integral einer beliebigen Komposition. Der Grund für das Fehlen systematischer Regeln ist nicht etwa mangelnder Scharfsinn der Mathematiker, sondern die Tatsache, daß die Stammfunktionen vieler Funktionen einer umfangreicheren Klasse angehören als diese selbst. Darauf deutet schon die Tatsache hin, daß die rationale Funktion $x \mapsto 1/x$ als Stammfunktion den Logarithmus hat, der zwar noch als elementar gilt, das aber unbestreitbar weniger ist als die rationale Funktion selbst. Die einfachsten nicht-elementaren Stammfunktionen sind die sogenannten *elliptischen* (unbestimmten) Integrale

$$\int \frac{dx}{\sqrt{p(x)}} \quad \text{mit einem Polynom } p \text{ vom Grad 3 oder 4;}$$

solche treten bei der Berechnung des Ellipsenumfangs auf (daher der Name), aber auch bei der Berechnung der Periode des Pendels.

Systematisch berechnen kann man aber immerhin die Stammfunktionen einer jeden rationalen Funktion, jedenfalls wenn man die Nullstellen und damit die Linearfaktorzerlegung ihres Nenners kennt. Division mit Rest und die in unter 10.12 erklärte Partialbruchzerlegung reduzieren diese Aufgabe darauf, Stammfunktionen der (im allgemeinen nicht-reellen) Funktionen

$$z \mapsto z^k \quad (k \in \mathbb{N}) \qquad \text{und} \qquad z \mapsto \frac{1}{(z-c)^k} \quad (0 < k \in \mathbb{N})$$

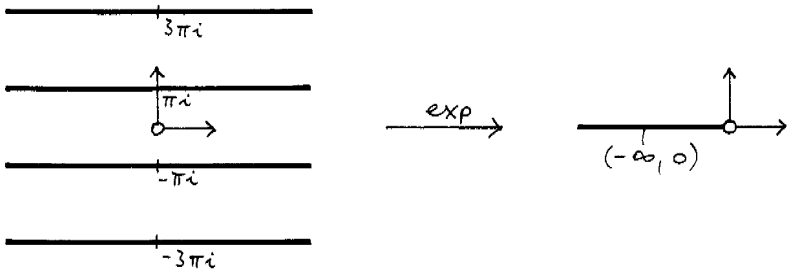
zu finden. Die kann man in der Form

$$z \mapsto \frac{z^{k+1}}{k+1} \qquad \text{und} \qquad z \mapsto \begin{cases} \log(z-c) & (k=1) \\ -\dfrac{1}{(k-1)(z-c)^{k-1}} & (k>1) \end{cases}$$

sofort hinschreiben, muß sich aber Gedanken über die Bedeutung des komplexen Logarithmus machen:

$15\frac{1}{2}$.11 Rechnung Wenn c reell ist, die ursprünglich gegebene rationale Funktion $f\colon I \longrightarrow \mathbb{R}$ dort also eine Polstelle hat, ist zu unterscheiden, ob das Intervall I rechts oder links von c liegt. Im ersten Fall ist

$x \mapsto \log(x-c)$ durch den gewöhnlichen reellen Logarithmus erklärt. Liegt I dagegen links von c, so liefert jede Wahl von $k \in \mathbb{Z}$ eine mögliche Interpretation



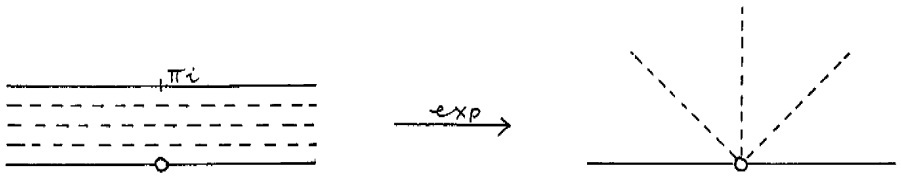
$$\log(x-c) = \log|x-c| + (2k+1)i\pi$$

des Logarithmus als (partielle) Umkehrung der Exponentialfunktion. Zur Verwendung als Stammfunktion von $x \mapsto 1/(x-c)$ können wir den konstanten Imaginärteil aber ignorieren, bei reellem c also durchweg mit

$$I \ni x \mapsto \log|x-c| \in \mathbb{R}$$

rechnen. (Was sich natürlich auch direkt durch Ableiten verifizieren läßt.)

Jetzt sei $c \in \mathbb{C}\backslash\mathbb{R}$ und etwa $\mathrm{Im}c < 0$. Für reelle x liegt dann $x-c \in \mathbb{C}$ in der oberen Halbebene, und es bietet sich an, den Logarithmus hier als die Umkehrung von



$$\mathbb{C} \supset \mathbb{R} \times i\,(0,\pi) \quad \xrightarrow{\exp} \quad \mathbb{R} \times i\,(0,\infty) \subset \mathbb{C}$$

zu lesen. Ist die Stammfunktion oder das damit berechnete Integral bloß ein Zwischenergebnis, soll man es damit genug sein lassen und mit dem so präzisierten komplexen Logarithmus weiterrechnen. Wer aber darauf besteht, als Resultat wirklich eine explizit reelle Zahl zu sehen, muß zur Strafe in den sauren Apfel beißen und die folgende Auswertung durchführen.

Da die Ausgangsfunktion f reell ist, treten die nicht-reellen Terme der Partialbruchzerlegung in komplex-konjugierten Paaren auf; wir müssen also dem Ausdruck

$$\alpha\,\log(x-c) + \overline{\alpha}\,\log(x-\overline{c}) \quad (x \in \mathbb{R})$$

einen Sinn geben, wobei wir uns für die Rechnung auf den Fall $\mathrm{Im}c < 0$ wie oben festlegen dürfen. Wenn wir den betrachteten Logarithmus nach Satz 12.16 in

$$\mathbb{C} \supset \mathbb{R} \times i\,(0,\infty) \ni z = x + iy \longmapsto \log|z| + i\,\mathrm{arccot}\,\frac{x}{y} \in \mathbb{R} \times i\,(0,\pi) \subset \mathbb{C}$$

aufschlüsseln, ergibt sich

$$\begin{aligned}
\alpha\,\log(x-c) + \overline{\alpha}\,\log(x-\overline{c}) &= 2\mathrm{Re}\big(\alpha\log(x-c)\big) \\
&= 2\mathrm{Re}\left(\alpha\,\log|x-c| + i\alpha\,\mathrm{arccot}\,\frac{x-\mathrm{Re}c}{0-\mathrm{Im}c}\right) \\
&= 2\mathrm{Re}\alpha\cdot\log\sqrt{(x-\mathrm{Re}c)^2+(\mathrm{Im}c)^2} - 2\mathrm{Im}\alpha\cdot\mathrm{arccot}\,\frac{x-\mathrm{Re}c}{-\mathrm{Im}c} \\
&= \mathrm{Re}\alpha\cdot\log\big((x-\mathrm{Re}c)^2+(\mathrm{Im}c)^2\big) - 2\mathrm{Im}\alpha\cdot\mathrm{arccot}\,\frac{x-\mathrm{Re}c}{-\mathrm{Im}c}.
\end{aligned}$$

Der Ästhetik halber können wir noch die Rollen von (α, c) und $(\overline{\alpha}, \overline{c})$ vertauschen, also

$$\mathbb{R} \ni x \longmapsto \mathrm{Re}\alpha \cdot \log\big((x-\mathrm{Re}c)^2 + (\mathrm{Im}c)^2\big) + 2\mathrm{Im}\alpha \cdot \operatorname{arccot} \frac{x - \mathrm{Re}c}{\mathrm{Im}c}$$

als Stammfunktion für

$$\mathbb{R} \ni x \longmapsto \frac{\alpha}{x-c} + \frac{\overline{\alpha}}{x-\overline{c}} \in \mathbb{R} \quad \text{mit } \mathrm{Im}c > 0$$

notieren.

Ganz schön kompliziert. Einfacher wird's in Spezialfällen: Für $\alpha \in \mathbb{R}$ bleibt bloß

$$\alpha \cdot \log\big((x-\mathrm{Re}c)^2 + (\mathrm{Im}c)^2\big),$$

für rein imaginäres $\alpha = i\beta$

$$2\beta \cdot \operatorname{arccot} \frac{x - \mathrm{Re}c}{\mathrm{Im}c}.$$

Ist ganz konkret $\beta = 1$ und $c = i$, so erhalten wir mit

$$-2\int \frac{dx}{x^2+1} = i\int \frac{dx}{x-i} - i\int \frac{dx}{x+i} = 2\operatorname{arccot} x$$

im wesentlichen die aus $15\frac{1}{2}.8$ bekannte Formel $\displaystyle\int \frac{dz}{1+x^2} = \arctan x$.

Soweit zur Integration der rationalen Funktionen. — Allgemein sollte man übrigens nicht die Möglichkeit übersehen, Integrale von als Potenzreihen gegebenen analytischen Funktionen nach der Notiz 15.10 durch gliedweise Integration der Reihe auszuwerten.

Jetzt kehren wir zu den Grundlagen zurück und beweisen, daß es genau ein Integral im Sinne der Definition $15\frac{1}{2}.2$ gibt. Die Eindeutigkeit macht inzwischen keine Mühe mehr. Zu ihrem Beweis dürfen wir natürlich die Existenz voraussetzen. Daß immer $\int_a^a = 0$ gilt, folgt direkt aus dem Normiertheitsaxiom. Ist nun $a < b$ und $f \in C^0[a,b]$ eine stetige Funktion, so existieren nach Satz $15\frac{1}{2}.6$ Stammfunktionen von f, und die damit anwendbare Folgerung $15\frac{1}{2}.7$ berechnet das Integral über f aus einer beliebig gewählten Stammfunktion.

Unsere eigentliche Aufgabe ist es, die Existenz des Integrals zu beweisen, indem wir eines explizit konstruieren. Wir werden uns dabei wesentlich auf eine Eigenschaft der zu integrierenen Funktionen stützen, die eine gewisse Analogie zum Begriff der gleichmäßigen Konvergenz einer Funktionenfolge hat.

$15\frac{1}{2}.12$ Definition Sei $X \subset \mathbb{C}$ und $f\colon X \to \mathbb{C}$ eine Funktion. Man nennt sie *gleichmäßig stetig*, wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ mit

$$\big|f(x) - f(y)\big| < \varepsilon \quad \text{für alle } x, y \in X \text{ mit } |x-y| < \delta$$

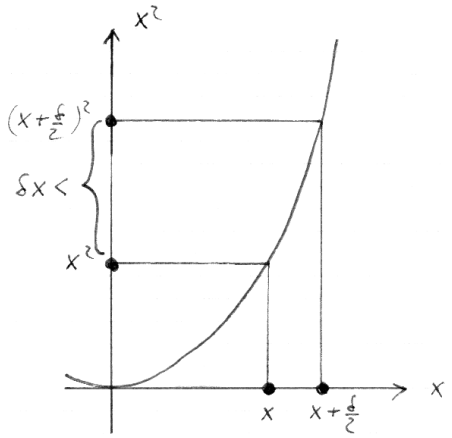
gibt.

Im Unterschied zur gewöhnlichen Stetigkeit, das heißt der Stetigkeit an jeder Stelle von X, wird also ein δ gefordert, das die bekannte Abschätzung simultan überall in X nach sich zieht.

$15\frac{1}{2}.13$ Beispiel Die bekanntermaßen stetige Funktion $\mathbb{R} \ni x \mapsto x^2 \in \mathbb{R}$ ist nicht gleichmäßig stetig. Denn schon zur Wahl $\varepsilon = 1$ gäbe es sonst ein $\delta > 0$ mit (insbesondere)

$$\left(x + \frac{\delta}{2}\right)^2 - x^2 < 1 \quad \text{für alle } x \geq 0,$$

woraus sofort $\delta x < 1$ für alle $x \geq 0$ folgen würde.



Der folgende Satz aber ist eine wichtige Quelle für gleichmäßig stetige Funktionen.

15½.14 Satz Sei $K \subset \mathbb{R}$ ein kompaktes Intervall und $f\colon K \to \mathbb{C}$ eine stetige Funktion. Dann ist f automatisch gleichmäßig stetig.

Beweis Wie nehmen das Gegenteil an, finden also ein $\varepsilon > 0$, so daß es zu jedem $\delta > 0$ ein Paar $x, y \in K$ mit $|x-y| < \delta$, aber $|f(x)-f(y)| \geq \varepsilon$ gibt. Speziell für $\delta := 1/n$ wählen wir solche Verbrecher $x = x_n$ und $y = y_n$ aus und haben damit zwei Folgen $(x_n)_{n=1}^\infty$ und $(y_n)_{n=1}^\infty$ in K konstruiert. Nun ist K nach Voraussetzung kompakt; wenn wir die erste Folge durch eine geeignete Teilfolge ersetzen, verlieren wir nichts und gewinnen, daß die Folge konvergiert und ihr Grenzwert zu K gehört:

$$a := \lim_{n\to\infty} x_n \in K\,.$$

Wir ersetzen auch $(y_n)_{n=1}^\infty$ durch die entsprechende Teilfolge: sie konvergiert wegen $|x_n - y_n| < 1/n$ ebenfalls gegen a.

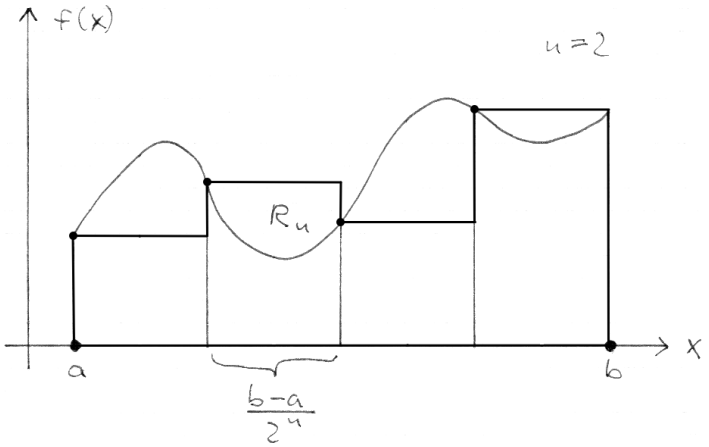
Nun ist f an der Stelle a stetig. Wir finden also ein $\delta > 0$ mit $|f(x) - f(a)| < \varepsilon/2$ für alle $x \in X$ mit $|x - a| < \delta$. Zu diesem δ wählen ein $n \in \mathbb{N}$ mit $n > 1/\delta$, also $1/n < \delta$ und schließen

$$\big|f(x_n) - f(a)\big| < \varepsilon/2 \quad \text{und} \quad \big|f(y_n) - f(a)\big| < \varepsilon/2$$

und damit $|f(x_n) - f(y_n)| < \varepsilon$ im Widerspruch zur Konstruktion unserer Folgen.

15½.15 Konstruktion eines **Integrals** Seien $a \leq b$ und $f \in C^0[a,b]$. Für jedes $n \in \mathbb{N}$ definieren wir die n-te *riemannsche Summe*

$$R_n = \sum_{j=0}^{2^n-1} f\big(a + \frac{j}{2^n}(b-a)\big) \cdot \frac{b-a}{2^n} \in \mathbb{R}\,.$$



Die Folge $(R_n)_{n=0}^{\infty}$ ist eine Cauchy-Folge, und wir definieren das Integral von f als den Limes

$$\int_a^b f := \lim_{n\to\infty} R_n \in \mathbb{R}\,.$$

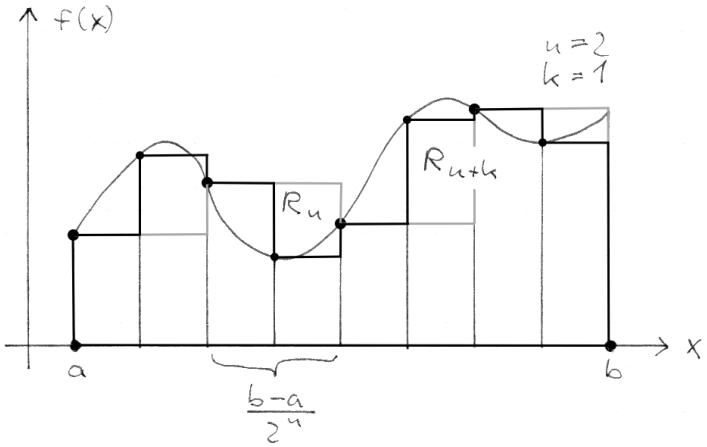
Anmerkung Die R_n sind im Vergleich zur üblichen Terminologie nur sehr spezielle riemannschen Summen, für unsere Zwecke aber ausreichend. Die zugrundliegende Idee ist, den zu bestimmenden Flächeninhalt durch das Ausmessen von schmalen senkrechten Streifen zu approximieren. Die klassische Integralnotation $\int f(x)\,dx$ erinnert daran, indem sie $\lim\sum$ zum Integralzeichen $\int$ und die Streifenbreite, hier $(b-a)/2^n$, zu dx werden läßt, einer traditonell "infinitesimal" genannten Größe: das heißt kleiner als jede positive Zahl, aber selbst noch positiv — wir wissen gut, daß so eine Vorstellung zumindest in dieser Form logisch nicht haltbar ist.

Beweis der Cauchy-Eigenschaft Sei $\varepsilon > 0$. Weil f nach Satz $15\frac{1}{2}.14$ gleichmäßig stetig ist, finden wir ein $\delta > 0$ mit $|f(x) - f(y)| < \varepsilon$ für alle $x, y \in [a, b]$ mit $|x-y| < \delta$; wir wählen dann $D \in \mathbb{N}$ genügend groß, so daß $2^{-D}(b-a) < \delta$ wird.

Seien jetzt $n, k \in \mathbb{N}$ mit $n > D$. Wir schreiben die riemannschen Summen R_{n+k} und R_n als

$$R_{n+k} = \sum_{i=0}^{2^n-1}\sum_{j=0}^{2^k-1} f\big(a + \frac{i2^k+j}{2^{n+k}}(b-a)\big)\cdot\frac{b-a}{2^{n+k}}$$

$$R_n = \sum_{i=0}^{2^n-1}\sum_{j=0}^{2^k-1} f\big(a + \frac{i\,2^k}{2^{n+k}}\,(b-a)\big)\cdot\frac{b-a}{2^{n+k}}$$



und können die Differenz durch

$$\begin{aligned}\big|R_{n+k} - R_n\big| &\le \sum_{i=0}^{2^n-1}\sum_{j=0}^{2^k-1}\Big|f\big(a + \frac{i2^k+j}{2^{n+k}}(b-a)\big) - f\big(a + \frac{i\,2^k}{2^{n+k}}\,(b-a)\big)\Big|\cdot\frac{b-a}{2^{n+k}}\\ &\le \sum_{i=0}^{2^n-1}\sum_{j=0}^{2^k-1}\varepsilon\cdot\frac{b-a}{2^{n+k}}\\ &= \varepsilon\cdot(b-a)\end{aligned}$$

abschätzen.

Bleibt zu zeigen, daß das so definierte Integral die vier Integralaxiome erfüllt. Für die Linearität, Positivität und Normiertheit ist das evident, so daß wir uns nur um die Unterteilbarkeit kümmern müssen. Dazu beweisen wir den folgenden Hilfssatz, der in der Tat nur innerhalb des Beweises von Interesse ist, da er für das fertige Integral von Satz $15\frac{1}{2}.6$ überholt wird.

$15\frac{1}{2}$.16 Hilfssatz Seien $a \le b$ und $f \in C^0[a,b]$. Die nach $15\frac{1}{2}$.15 gebildete Funktion $[a,b] \ni \beta \mapsto \int_a^\beta f \in \mathbb{R}$ ist stetig.

Beweis Wir zeigen die Stetigkeit an der Stelle $\beta \in [a,b]$. Sei $\varepsilon > 0$. Wieder aufgrund der gleichmäßigen Stetigkeit von f finden wir ein $\delta > 0$, so daß $|f(x) - f(y)| < \varepsilon$ für alle $x, y \in [a,b]$ mit $|x-y| < \delta$ gilt, und wir dürfen gleich $\max\{|f(x)| \,|\, x \in [a,b]\}\,\delta \le \varepsilon$ annehmen.

Sei nun $\xi \in [a,b]$ ein Punkt mit $|\xi - \beta| < \delta$. Für jedes $n \in \mathbb{N}$, jedes $j \in \{0, \dots, 2^n - 1\}$ und jedes $\xi \in [a,b]$ gilt dann

$$\Big|\big(a + \frac{j}{2^n}(\xi - a)\big) - \big(a + \frac{j}{2^n}(\beta - a)\big)\Big| < \delta\,.$$

Für die riemannschen Summen — deren Abhängigkeit von der Intervallgrenze ξ wir jetzt mit notieren müssen — gilt deshalb

$$\begin{aligned}
\big|R_n(\xi) - R_n(\beta)\big| &\le \sum_{j=0}^{2^n-1} \Big|f\big(a + \frac{j}{2^n}(\xi - a)\big) - f\big(a + \frac{j}{2^n}(\beta - a)\big)\Big| \cdot \frac{\xi - a}{2^n} \\
&\qquad + \sum_{j=0}^{2^n-1} \Big|f\big(a + \frac{j}{2^n}(\beta - a)\big)\Big| \cdot \Big|\frac{\xi - a}{2^n} - \frac{\beta - a}{2^n}\Big| \\
&\le \sum_{j=0}^{2^n-1} \varepsilon \cdot \frac{b-a}{2^n} + \sum_{j=0}^{2^n-1} \max\{|f(x)| \,|\, x \in [a,b]\} \cdot \frac{\delta}{2^n} \\
&\le (b - a + 1)\,\varepsilon\,.
\end{aligned}$$

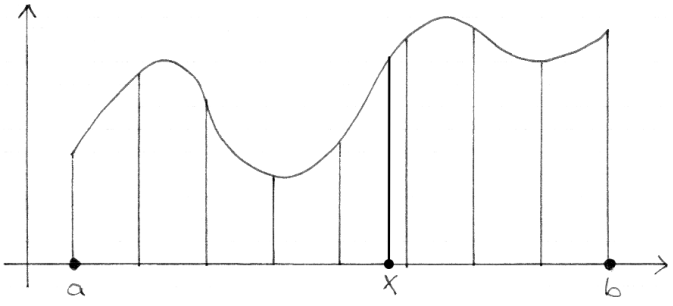
Für die Grenzwerte folgt dann auch

$$\Big|\int_a^\xi f - \int_a^\beta f\Big| \le (b - a + 1)\,\varepsilon\,,$$

und das beweist die behauptete Stetigkeit.

Natürlich gilt das Lemma entsprechend, wenn man die rechte Integrationsgrenze festhält und die linke variiert. Zum Beweis des Unterteilbarkeitsaxioms betrachten wir nun für einen festen Integranden $f \in C^0[a,c]$ die Funktion

$$[a,c] \ni x \longmapsto F(x) := \int_a^x f + \int_x^c f \in \mathbb{R};$$



sie ist nach dem Lemma stetig. Wenn nun $x \in [a,c]$ die Form $a + \frac{j}{2^n}(c-a)$ hat, dann addieren sich die beiden n-ten ebenso wie alle weiteren riemannschen Summen zur n-ten riemannschen Summe über das Gesamtintervall $[a,c]$, und wir schließen $F(x) = \int_a^c f$ für all diese x.

Daraus folgt, daß F überhaupt den konstanten Wert $\int_a^c f$ hat: Sei $b \in [a,c)$, für genügend großes $n \in \mathbb{N}$ ist dann $b + 1/n \le c$, und das Intervall $(b, b+1/n)$ enthält nach der naheliegenden Varianten von Satz 2.13 eine Zahl x_n der Form $a + \frac{j}{2^n}(c-a)$. Die so entstehende Folge (x_n) konvergiert gegen b, und weil F bei b stetig ist, folgt

$$F(b) = \lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} \int_a^c f = \int_a^c f\,.$$

Das vervollständigt den Beweis des Unterteilbarkeitsaxioms und damit auch unsere Konstruktion des Integrals.

30 1/3 Messen

Wir steuern jetzt auf die Integralrechnung zu. Deren Zweck ist leicht erklärt: Es geht darum, Volumina zu berechnen, und natürlich muß man dazu den anschaulichen Begriff des Volumens oder *Maßes* erst mal mathematisch präzisieren. Genauer ist für jedes $n \in \mathbb{N}$ ein eigener n-dimensionaler Volumen- oder Maßbegriff gemeint, von dem das Volumen der Alltagssprache nur der dreidimensionale Fall ist; für $n = 2$ und $n = 1$ wird das n-dimensionale Volumen dann zu Flächeninhalt und Länge.

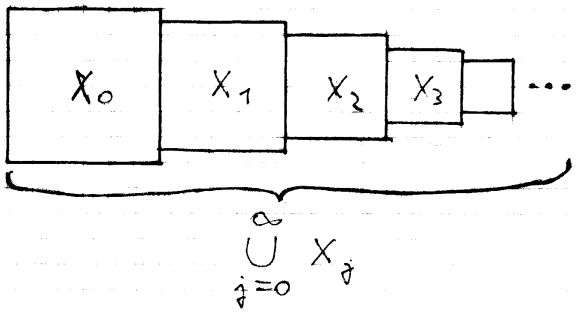
Auch den Begriff des Maßes wollen wir axiomatisch angehen und uns erst mal überlegen, was wir von ihm erwarten. Seine Aufgabe wird es sein, Teilmengen von $\mathbb{R}^n$ zu *messen*, also jedenfalls jeder solchen Teilmenge X eine Zahl $\mu(X)$ zuzuweisen — und zwar eine nicht-negative, denn ein Orientierungsvorzeichen wie bei der Determinante wollen wir hier nicht ins Auge fassen. Als primitive Eigenschaften erwarten wir gewiß $\mu(\emptyset) = 0$ und

$$\mu(X \cup Y) = \mu(X) + \mu(Y) \quad \text{für disjunkte } X, Y \subset \mathbb{R}^n,$$

und vielleicht sogar eine unendliche Additivität

$$\mu\Big(\bigcup_{\lambda\in\Lambda} X_\lambda\Big) = \sum_{\lambda\in\Lambda} \mu(X_\lambda)$$

für den Fall, daß die Mengen der Familie $(X_\lambda)_{\lambda\in\Lambda}$ paarweise disjunkt sind.



Auf mehr geometrischer Ebene sollte das Maß eines Quaders

$$Q = I_1 \times I_2 \times \cdots \times I_n \subset \mathbb{R}^n$$

das Produkt der Kantenlängen sein, und allgemein das Maß einer Menge sich nicht ändern, wenn man sie in $\mathbb{R}^n$ bewegt, etwa verschiebt oder dreht.

Ganz so einfach ist es freilich nicht. Was kann denn das n-dimensionale Maß von $\mathbb{R}^n$ selbst sein? Es muß größer als jede reelle Zahl sein; man muß also auch ∞ als Wert der Maßfunktion zulassen. Andererseits ist der vorgetragene Wunsch nach einer Additivität des Maßes für beliebige Familien disjunkter Mengen ebenso unklar wie illusorisch, denn für $n > 0$ ist das Maß jeder einpunktigen Teilmenge von $\mathbb{R}$ (ein Quader mit Kantenlängen null!) null, andererseits etwa

$$[0,1]^n = \bigcup_{x\in[0,1]^n} \{x\}$$

eine Darstellung des Einheitswürfels als disjunkte Vereinigung solcher Mengen. Eine Interpretation von

$$\sum_{x\in[0,1]^n} \mu(\{x\}) = \sum_{x\in[0,1]^n} 0,$$

die aus (wenn auch vielen) Summanden, die alle null sind, den Wert $\mu([0,1]^n) = 1$ macht, ist schwer vorstellbar.

Geben wir uns also besser doch mit der soliden endlichen Additivität zufrieden? Nun, eine verhältnismäßig tiefliegende und junge Erkenntnis ist, daß es überaus hilfreich ist, die Additivität des Maßes zwar nicht für beliebige, aber doch für abzählbare Familien disjunkter Teilmengen von $\mathbb{R}^n$ sicherzustellen. Diese sogenannte σ-Additivität werden wir deshalb in den Forderungskatalog für das Maß aufnehmen.

Damit sind aber noch nicht alle Hindernisse beseitigt. Vielmehr zeigt eine subtilere Überlegung, daß jedes Maß mit den bisher besprochenen Eigenschaften an gewissen Teilmengen von $\mathbb{R}^n$ scheitern muß, die einfach zu kraus oder diffus sind, um sie messen zu können. Das zwingt uns, den Begriff der *meßbaren* Menge einzuführen und zu akzeptieren, daß $\mu(X)$ eben nur für meßbare $X \subset \mathbb{R}^n$ erklärt ist.

Damit ist der Rahmen für diesen Abschnitt im Groben abgesteckt. Als erstes sprechen wir über Systeme von Mengen, die als meßbar in Frage kommen. Wir tun das gleich in dem ganz allgemeinen axiomatischen Rahmen der sogenannten Maßtheorie; das erhöht nicht nur die Übersicht, sondern macht unsere Überlegungen auch auf andere mathematische Teilgebieten übertragbar, speziell die moderne Wahrscheinlichkeitstheorie oder Stochastik, die eine Spielart der Maßtheorie ist. Den größeren Teil des Abschnitts nimmt dann aber die Konstruktion unseres konkreten Maßes auf $\mathbb{R}^n$ ein.

30$\frac{1}{3}$.1 Definition Sei M eine Menge. Eine Menge M von Teilmengen von M heißt eine σ-Algebra in M, wenn sie folgenden Axiomen genügt.

(a) $\emptyset \in \mathsf{M}$;

(b) aus $X \in \mathsf{M}$ folgt $M \setminus X \in \mathsf{M}$;

(c) ist $(X_j)_{j=0}^{\infty}$ eine Folge in M, so ist $\bigcup_{j=0}^{\infty} X_j \in \mathsf{M}$.

Wenn eine bestimmte σ-Algebra in M fixiert ist, nennt man ihre Elemente die meßbaren Teilmengen von M.

Eine σ-Algebra ist also ein System von Teilmengen, das zumindest die leere Menge enthält sowie unter Komplementbildung und abzählbarer Vereinigung abgeschlossen ist. Das gilt dann übrigens aufgrund der Identität

$$\bigcap_{j=0}^{\infty} X_j = M \setminus \bigcup_{j=0}^{\infty} (M \setminus X_j)$$

automatisch auch für abzählbare Durchschnitte.

Der algebraisch klingende Name kommt daher, daß es in einer früheren, kurz Algebra genannten Version nur um die Vereinigung *zweier* Mengen ging, ein Vorgang, den man ja als Verknüpfung auf der Menge M lesen kann. Das vorgesetzte Sigma ist ein nicht nur in der Maßtheorie gängiges Kürzel für Rezepte mit abzählbar vielen Zutaten.

30$\frac{1}{3}$.2 Definition Sei M eine Menge und M eine σ-Algebra in M. Ein Maß auf M ist eine Funktion

$$\mu\colon \mathsf{M} \longrightarrow [0, \infty]$$

mit den Eigenschaften

- $\mu(\emptyset) = 0$

und

- $\mu(\bigcup_{j=0}^{\infty} X_\lambda) = \sum_{j=0}^{\infty} \mu(X_j)$ für jede Folge $(X_j)_{j=0}^{\infty}$ von paarweise disjunkten Mengen in M.

Erklärung Was ist mit der Reihensumme gemeint, wenn die Reihe divergiert? Es handelt sich a priori um eine Reihe mit nicht-negativen Gliedern, und wie wir aus dem Abschnitt 5 wissen, bedeutet die Konvergenz für solche Reihen dasselbe wie die Beschränktheit der Partialsummenfolge. Das legt es nahe, die Summe

im Fall der Divergenz als ∞ zu erklären, und das auch in dem Fall zu tun, daß mindestens ein Summand gar keine reelle Zahl, sondern selbst schon unendlich ist. Um in der Maßtheorie immer wieder auftretende Sonderfälle in den Formulierungen mit erfassen zu können, ist es darüber hinaus zweckmäßig, auch Formeln wie

$$\lambda + \infty = \infty \text{ für jedes } \lambda \in (-\infty, \infty] \quad \text{und} \quad \lambda\infty = \infty \text{ für jedes } \lambda \in (0, \infty]$$

als richtig zu akzeptieren — der sinnvolle Umgang mit ihnen regelt sich nach den Limesregeln 9.6 und ähnlichen. Desgleichen ist es zweckmäßig, auch nicht nach oben beschränkten Mengen einschließlich Teilmengen $X \subset [-\infty, \infty]$ ein Supremum zuzuordnen, nämlich $\sup X = \infty$, was ja dann offenbar die kleinste obere Schranke von X, bloß eben keine reelle Zahl mehr ist. Mit diesen Vereinbarungen ist zum Beispiel für beliebige $\mu_j \in [0, \infty]$

$$\sum_{j=0}^{\infty} \mu_j = \sup_{k \in \mathbb{N}} \sum_{j=0}^{k} \mu_j$$

eine immer korrekte und nützliche Formel.

Schließlich vereinbaren wir:

$30\frac{1}{3}$.3 Definition Ein Tripel (M, M, μ) aus einer Menge M, einer σ-Algebra M in M und einem Maß μ auf M heißt ein Maßraum. Eine meßbare Menge $X \in \mathsf{M}$ mit $\mu(X) = 0$ nennt man eine Nullmenge.

$30\frac{1}{3}$.4 Beispiele (1) Natürlich ist für jede Menge M die Potenzmenge (die Menge aller Teilmengen) $\mathbf{P}M$ eine σ-Algebra in M. Wenn M endlich ist, können wir die beiden Daten durch das Zählmaß

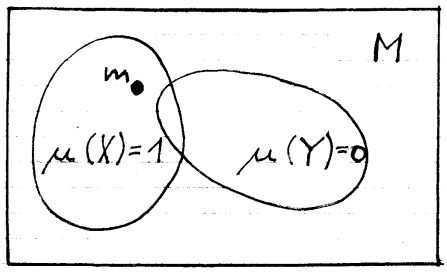
$$\mu(X) := |X| \in \mathbb{N}$$

zu einem Maßraum $(M, \mathbf{P}M, \mu)$ ergänzen.

(2) Auch ohne M als endlich vorauszusetzen, sind alle Maßraumaxiome flott verifiziert, wenn man für unendliche Teilmengen $X \subset M$ eben $\mu(X) = \infty$ setzt. Interessant ist dieser Maßraum vor allem für abzählbar unendliche M, also etwa $(\mathbb{N}, \mathbf{P}\mathbb{N}, |?|)$.

(3) Die Maßaxiome sind auch auf ganz plumpe Weise zu erfüllen, nämlich durch die Festsetzung $\mu(X) = 0$ für jedes $X \in \mathsf{M}$. Nicht ganz so triviale und schon nützliche Maße sind die sogenannten *Einheitsmassen*: man zeichne in einer Menge M ein beliebiges Element $m \in M$ aus und definiere das Maß $\mu\colon \mathbf{P}M \to [0, 1]$ durch

$$\mu(X) = \begin{cases} 1 & \text{falls } m \in X, \\ 0 & \text{sonst.} \end{cases}$$



Freilich ist keines dieser Beispiele von dem Typ, den wir uns als "das" Maß auf $\mathbb{R}^n$ vorgestellt haben. Das liegt an der Schwierigkeit, dieses Maß zu konstruieren — eine Aufgabe, die den ganzen Rest des Abschnitts ausmachen wird. Wir fixieren als unser Ziel:

$30\frac{1}{3}.4\frac{1}{2}$ Satz und **Definition** Sei $n \in \mathbb{N}$. Es gibt einen Maßraum $(\mathbb{R}^n, \mathsf{M}, \mu)$ mit der Eigenschaft, daß jeder Quader $Q \subset \mathbb{R}^n$ meßbar ist und $\mu(Q)$ das Produkt der Kantenlängen ist. Wir nennen es das (n-dimensionale) Lebesgue-Maß.

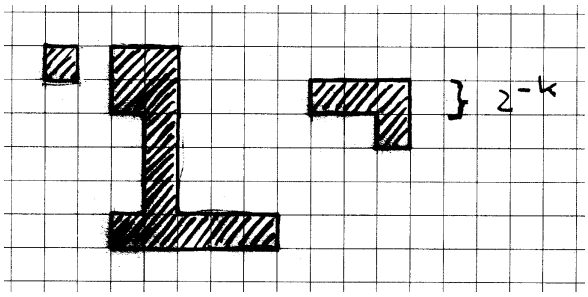
Bemerkungen Den trivialen Fall $n = 0$ habe ich wie so oft der Systematik halber mit aufgenommen; der einzige Quader $Q = \{0\} \subset \mathbb{R}^0$ hat als Maß $\mu(Q) = 1$, das leere Produkt. — Die genannten Eigenschaften bestimmen den Maßraum nicht wirklich eindeutig, aber die im folgenden konstruierte Version ist es, die nach ihrem Erfinder das Lebesgue-Maß genannt wird.

Der relativ langwierige, aber nicht tiefsinnige Beweis des Satzes besteht darin, $\mu(X)$ zuerst für ganz einfache und dann für zunehmend kompliziertere X zu definieren. Zuerst wollen wir ein paar Begriffe ad hoc, nur für die Zwecke der Konstruktion prägen; die sehr treffenden Namen habe ich übrigens von Klaus Jänich gelernt.

$30\frac{1}{3}$.5 Definition Sei $k \in \mathbb{N}$. Ein k-Elementarwürfel ist ein abgeschlossener Würfel in $\mathbb{R}^n$ der Form

$$\left\{x \in \mathbb{R}^n \,\middle|\, 2^{-k}(z_j - 1) \le x_j \le 2^{-k} z_j \text{ für } j = 1, \dots, n\right\}$$

mit $z_j \in \mathbb{Z}$ für alle j. Jede Vereinigung von k-Elementarwürfeln wollen wir ein k-Aggregat nennen:



Enthält ein k-Aggregat A genau r der k-Elementarwürfel, so erklären wir das Volumen $v(A)$ von A als

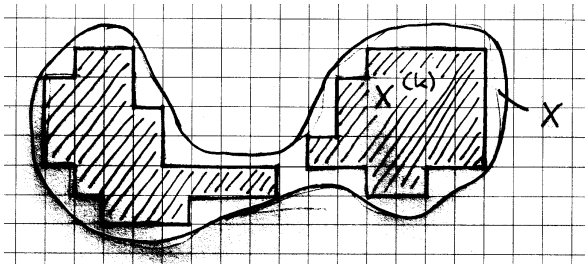
$$v(A) = r \cdot 2^{-nk} \in [0, \infty].$$

Diese Definition ist zulässig, denn zwar ist ein k-Aggregat A ebenso ein $(k+1)$-Aggregat, aber als solches enthält es genau $2^n r$ der $(k+1)$-Elementarwürfel, und die definierende Formel liefert denselben Wert

$$v(A) = 2^n r \cdot 2^{-n(k+1)} = r \cdot 2^{-nk}.$$

Im nächsten Schritt definieren wir mittels von Aggregaten das Maß offener Teilmengen von $\mathbb{R}^n$. Wir abstrahieren dabei die Idee, eine ebene Menge X dadurch auszumessen, daß wir durchsichtiges Millimeterpapier darüberlegen und die ganz in X enthaltenen Kästchen zählen.

$30\frac{1}{3}$.6 Definition Sei $X \subset \mathbb{R}^n$ offen. Für jedes $k \in \mathbb{N}$ sei dann $X^{(k)}$ die Vereinigung aller in X enthaltenen k-Elementarwürfel, also das größte in X enthaltene k-Aggregat.



Damit definieren wir das (Lebesgue-)Maß von X als

$$\mu(X) = \lim_{k \to \infty} v(X^{(k)});$$

der Limes dieser wegen $X^{(k)} \subset X^{(k+1)}$ monoton wachsenden Folge existiert in $[0, \infty]$ stets.

Von den Eigenschaften des Maßes offener Mengen notieren wir neben der Trivialität

$$\mu(X) \le \mu(Y) \quad \text{für } X \subset Y$$

erst mal die folgende.

$30\frac{1}{3}$.7 Lemma Für $r > 0$ sei

$$W_r := \big\{x \in \mathbb{R}^n \,\big|\, |x_j| < r \text{ für } j = 1, \dots, n\big\}$$

der offene Würfel. Für jedes offene $X \subset \mathbb{R}^n$ gilt dann

$$\mu(X) = \lim_{r\to\infty} \mu(X \cap W_r) = \lim_{\substack{r\in\mathbb{N}\\ r\to\infty}} \mu(X \cap W_r).$$

Beweis Aufgrund der wachsenden Monotonie macht es nichts, ob man den Limes über positive reelle oder nur über ganzzahlige r bildet. Klar ist auch

$$v(X^{(k)}) = \lim_{r\to\infty} v\big((X \cap W_r)^{(k)}\big) \quad \text{für jedes } k \in \mathbb{N},$$

so daß sich die Behauptung in der Form

$$\lim_{k\to\infty} \underbrace{\lim_{r\to\infty} v\big((X \cap W_r)^{(k)}\big)}_{v(X^{(k)})} = \lim_{r\to\infty} \underbrace{\lim_{k\to\infty} v\big((X \cap W_r)^{(k)}\big)}_{\mu(X\cap W_r)}$$

lesen läßt. Da sich die beiden Seiten nur dadurch unterscheiden, in welcher Reihenfolge die beiden Limites gebildet werden, ist die Versuchung groß, den Beweis damit für beendet zu erklären. Das wäre aber voreilig, denn im allgemeinen ist nicht vorherzusagen, welche Wirkung eine Vertauschung zweier Grenzwertbildungen hat. Daß es hier — und an vielen Stellen im folgenden — doch klappt, liegt, daran, daß $v\big((X \cap W_r)^{(k)}\big)$ sowohl von r als auch von k im wachsenden Sinne monoton abhängt und deshalb beide Seiten mit

$$\sup_{k,r\in\mathbb{N}} v\big((X \cap W_r)^{(k)}\big)$$

übereinstimmen. Aus unserer alten Regel 3.8(d) folgt nämlich, daß die Grenzwerte nicht größer sind als das Supremum, und da umgekehrt jeder Limes eine obere Schranke für die beteiligten Folgenglieder ist, sind sie auch nicht kleiner (übrigens beschreibt Satz 6.8 über Doppelreihen im wesentlichen denselben Sachverhalt in anderer Sprache).

Auf Lemma $30\frac{1}{3}$.7 werden wir uns häufig berufen, um die zu messenden Teilmengen von $\mathbb{R}^n$ ohne Einschränkung der Allgemeinheit als beschränkt vorauszusetzen und damit die Endlichkeit ihrer Maße zu erzwingen. Das tun wir auch gleich mit dem folgenden Satz, in dem man eine Form der endlichen Additivität erkennt.

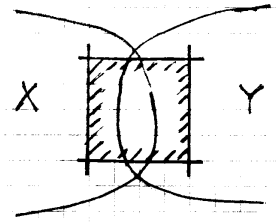
$30\frac{1}{3}$.8 Satz Sind $X \subset \mathbb{R}^n$ und $Y \subset \mathbb{R}^n$ offene Teilmengen, so ist

$$\mu(X) + \mu(Y) = \mu(X \cap Y) + \mu(X \cup Y).$$

Beweis Wie gesagt dürfen wir X und Y als beschränkt voraussetzen. Sei $k \in \mathbb{N}$. Während nach Definition $X^{(k)} \cap Y^{(k)} = (X \cap Y)^{(k)}$ ist, gilt für die Vereinigung im allgemeinen nur

$$X^{(k)} \cup Y^{(k)} \subset (X \cup Y)^{(k)},$$

weil ein in $X \cup Y$ enthaltener Elementarwürfel weder ganz in X noch ganz in Y zu liegen braucht.



Ich behaupte, daß aber

$$X^{(k)} \cup Y^{(k)} \subset (X \cup Y)^{(k)} \subset X^{(l)} \cup Y^{(l)} \quad \text{für alle genügend großen } l \in \mathbb{N}$$

gilt.

Zum Beweis nehmen wir das Gegenteil an und finden eine Folge $(x_l)_{l=0}^{\infty}$ mit

$$x_l \in K := (X \cup Y)^{(k)} \quad \text{und} \quad x_l \notin X^{(l)} \cup Y^{(l)}$$

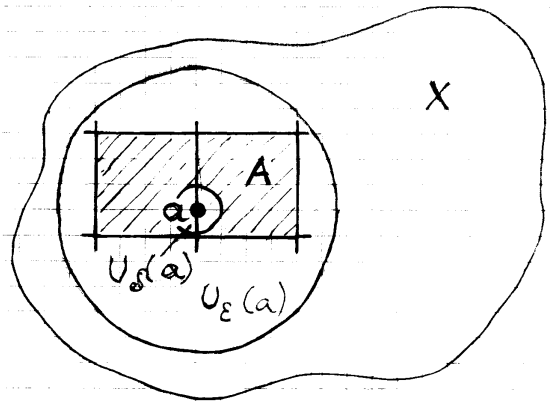
für alle l. Weil K als Vereinigung endlich vieler Elementarwürfel kompakt ist, konvergiert eine geeignete Teilfolge in K, und wegen $X^{(l)} \cup Y^{(l)} \subset X^{(l+1)} \cup Y^{(l+1)}$ dürfen wir annehmen, daß schon die Ausgangsfolge konvergiert und etwa

$$\lim_{l\to\infty} x_l = a \in K$$

liefert.

Nun ist sicher $a \in X \cup Y$, ohne Einschränkung etwa $a \in X$. Weil X offen ist und deshalb eine Kugel $U_\varepsilon(a)$ umfaßt, finden wir für genügend großes $m \in \mathbb{N}$ ein m-Aggregat A und ein $\delta > 0$ mit

$$U_\delta(a) \subset A \subset X :$$



die Skizze illustriert, wie das geht, sobald $\sqrt{n} \cdot 2^{-m} < \varepsilon$ ist. Wegen $A \subset X^{(m)}$ folgt $x_l \notin U_\delta(a)$ für alle $l \geq m$, im Widerspruch zu $\lim x_l = a$.

Das beweist die Zwischenbehauptung, und wir sehen insbesondere

$$\lim_{k\to\infty} v(X^{(k)} \cup Y^{(k)}) = \lim_{k\to\infty} v\big((X \cup Y)^{(k)}\big).$$

Der Rest ist einfach: Abzählen der Elementarwürfel gibt die Identität

$$v(X^{(k)}) + v(Y^{(k)}) = v\big((X \cap Y)^{(k)}\big) + v(X^{(k)} \cup Y^{(k)}),$$

und durch Übergang zum Limes für $k \to \infty$ folgt die Behauptung.

Wir kommen zu einer Vorstufe der σ-Additivität.

$30\frac{1}{3}$.9 Lemma Für jede Folge $(X_j)_{j=0}^{\infty}$ offener Mengen $X_j \subset \mathbb{R}^n$ gilt $\mu(\bigcup_{j=0}^{\infty} X_j) \le \sum_{j=0}^{\infty} \mu(X_j)$.

Beweis Wir dürfen nach Lemma $30\frac{1}{3}$.7 annehmen, daß die Vereinigung $X := \bigcup_{j=0}^{\infty} X_j$ beschränkt ist. Für jedes $k \in \mathbb{N}$ ist dann das Aggregat $X^{(k)}$ kompakt und deshalb schon in der Vereinigung endlich vieler X_j enthalten, etwa

$$X^{(k)} \subset \bigcup_{j=0}^{l} X_l.$$

Denn sonst fänden wir eine Folge $(x_l)_{l=0}^{\infty}$ mit

$$x_l \in X^{(k)} \quad \text{und} \quad x_l \notin \bigcup_{j=0}^{l} X_l$$

für alle l, diese hätte nach Auswahl einer geeigneten Teilfolge einen Grenzwert $a \in X^{(k)}$, und wir kämen zu einem Widerspruch genau wie im vorigen Beweis.

Nun liefert wiederholtes Anwenden von Lemma $30\frac{1}{3}$.8 die Abschätzung

$$\mu(\bigcup_{j=0}^{l} X_j) \le \sum_{j=0}^{l} \mu(X_j).$$

Daher gilt für jedes $k \in \mathbb{N}$

$$v(X^{(k)}) \le \mu(\bigcup_{j=0}^{l} X_j) \le \sum_{j=0}^{l} \mu(X_j) \le \sum_{j=0}^{\infty} \mu(X_j),$$

und wir schließen

$$\mu(X) = \lim_{k\to\infty} v(X^{(k)}) \le \sum_{j=0}^{\infty} \mu(X_j).$$

$30\frac{1}{3}$.10 Beispiele (1) $Q \subset \mathbb{R}^n$ sei ein offener Quader mit den Kantenlängen $q_1, \dots, q_n$. Für jedes $k \in \mathbb{N}$ ist $Q^{(k)}$ ein kompakter Quader; seine j-te Kantenlänge $q_j^{(k)}$ ist mindestens $q_j - 2 \cdot 2^{-k}$, aber kleiner als q_j. Damit ist

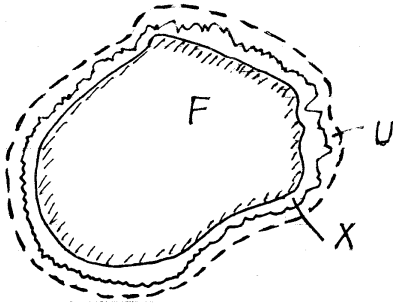
$$\mu(Q) = \lim_{k\to\infty} v(Q^{(k)}) = \lim_{k\to\infty} \prod_{j=1}^{n} q_j^{(k)} = \prod_{j=1}^{n} q_j$$

das Produkt der Kantenlängen, wie wir es uns gewünscht hatten.

Eine fundamentaler Trick der Maßtheorie ist nun, daß man beliebige Teilmengen von $\mathbb{R}^n$ nicht direkt durch Ausschöpfen mit Aggregaten zu messen versucht, sondern durch Einschließen zwischen eine offene und eine abgeschlossene Menge.

$30\frac{1}{3}$.11 Definition Sei $X \subset \mathbb{R}^n$ eine Teilmenge. Ein Meßvorgang für X ist ein Paar (U, F) aus einer offenen Menge $U \subset \mathbb{R}^n$ und einer abgeschlossenen Menge $F \subset \mathbb{R}^n$ mit

$$F \subset X \subset U.$$



Das Maß $\mu(U\setminus F)$ der (offenen!) Menge $U\setminus F$ heißt der Meßfehler des Vorgangs.

X heißt meßbar, wenn es zu jedem $\varepsilon > 0$ einen Meßvorgang (U, F) für X gibt, für dessen Meßfehler

$$\mu(U\setminus F) < \varepsilon$$

gilt. In diesem Fall heißt

$$\mu(X) := \inf\big\{\mu(U)\,\big|\,U \subset \mathbb{R}^n \text{ offen mit } X \subset U\big\} \in [0,\infty]$$

das (n-dimensionale Lebesgue-)Maß von X.

Anmerkung Sollte $X \subset \mathbb{R}^n$ selbst offen sein, provoziert die Bezeichnung $\mu(X)$ natürlich Verwechslungen mit dem "alten" Maß von X als offener Menge. Wir nehmen das ausnahmsweise in Kauf, da sich die Unsicherheit ohnehin schnell erledigen wird. — Ist (U, F) ein Meßvorgang für die meßbare Menge X, so ist definitionsgemäß $\mu(X) \le \mu(U)$, andererseits nach Lemma $30\frac{1}{3}.9$

$$\mu(U) = \mu((U\setminus F)\cup V) \le \mu(U\setminus F) + \mu(V) \quad \text{für jedes } X \text{ umfassende offene } V \subset U$$

und damit $\mu(U) \le \mu(U\setminus F) + \mu(X)$. Wir erhalten die Abschätzungen

$$\mu(U) - \mu(U\setminus F) \le \mu(X) \le \mu(U)$$

für $\mu(X)$ und rechtfertigen so die Bezeichnung "Meßfehler".

Ich stelle gleich ein paar einfach einzusehende Eigenschaften des eben definierten Maßes zusammen:

$30\frac{1}{3}.12$ Lemma (a) Beschränkte offene Mengen $X \subset \mathbb{R}^n$ sind meßbar, und $\mu(X)$ hat dann die alte Bedeutung.

(b) Ist $X \subset \mathbb{R}^n$ meßbar, so auch $\mathbb{R}^n\setminus X$.

(c) Sind X und Y meßbar, so sind $X\cap Y$ und $X\cup Y$ meßbar, mit $\mu(X\cup Y) \le \mu(X) + \mu(Y)$.

(d) Sind X und Y überdies disjunkt, so ist $\mu(X\cup Y) = \mu(X) + \mu(Y)$.

Beweis (a) Für jedes $k \in \mathbb{N}$ ist das Paar $(X, X^{(k)})$ ein Meßvorgang für X, dessen Meßfehler wir abschätzen können: für jedes $l \ge k$ gilt wegen

$$\big(X\setminus X^{(k)}\big)^{(l)} \subset X^{(l)} \setminus X^{(k)}$$

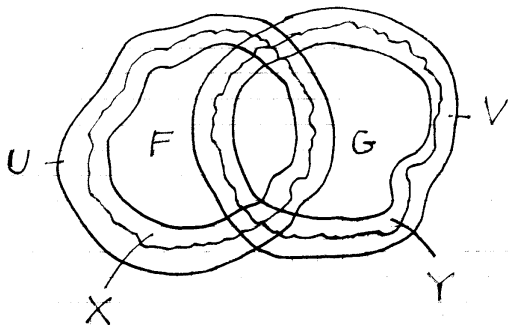
die Ungleichung $v\big(\big(X\setminus X^{(k)}\big)^{(l)}\big) \le v(X^{(l)}) - v(X^{(k)})$, und für $l\to\infty$ folgt

$$\mu(X\setminus X^{(k)}) \le \mu(X) - v(X^{(k)}).$$

Da $k \in \mathbb{N}$ beliebig war, ergeben sich Meßbarkeit und Maß von X wie behauptet.

(b) Ist (U, F) ein Meßvorgang für X, so ist $(\mathbb{R}^n\setminus F, \mathbb{R}^n\setminus U)$ ein Meßvorgang für $\mathbb{R}^n\setminus X$ mit demselben Meßfehler.

(c) (U, F) und (V, G) seien Meßvorgänge für X beziehungsweise Y; dann ist $(U \cup V, F \cup G)$ einer für $X \cup Y$, und die Meßfehler können sich höchstens addieren:



$$(U \cup V) \setminus (F \cup G) \subset (U \setminus F) \cup (V \setminus G).$$

Also ist $X \cup Y$ meßbar, und nach (b) folgt auch die Meßbarkeit von $X \cap Y$. Aus der Ungleichung

$$\mu(X \cup Y) \le \mu(U \cup V) \le \mu(U) + \mu(V)$$

sehen wir weiter $\mu(X \cup Y) \le \mu(X) + \mu(Y)$.

(d) Sind X und Y außerdem disjunkt, so liefern die in (c) gewählten Meßvorgänge

$$\begin{aligned}\mu(X) + \mu(Y) &\le \mu\big((U \cup V) \setminus G\big) + \mu\big((U \cup V) \setminus F\big) \\ &= \mu(U \cup V) + \mu\big((U \cup V) \setminus (F \cup G)\big) \\ &\le \mu(X \cup Y) + 2 \cdot \mu\big((U \cup V) \setminus (F \cup G)\big),\end{aligned}$$

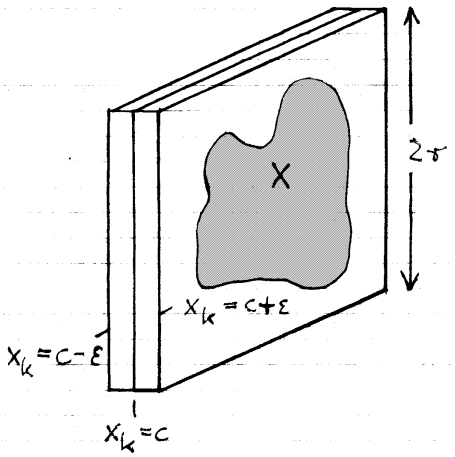
und wir schließen $\mu(X) + \mu(Y) \le \mu(X \cup Y)$ und damit die Gleichheit beider Seiten.

Weitere Beispiele:

(2) Sei $X \subset \mathbb{R}^n$ beschränkt und ganz in einer affinen Koordinatenhyperebene

$$\big\{x = (x_1, \dots, x_n) \in \mathbb{R}^n \,\big|\, x_k = c\big\}$$

($k \in \{1, \dots, n\}$ und $c \in \mathbb{R}$ fest) enthalten (ganz allgemein bezeichnet man als *Hyperebenen* in einem n-dimensionalen Vektorraum die Unterräume der Dimension $n-1$). Die Menge X ist dann meßbar, und zwar eine Nullmenge: $\mu(X) = 0$. Denn ist etwa $X \subset W_r$, so ist



$$U := \big\{x \in W_r \,\big|\, |x_k - c| < \varepsilon\big\}$$

ein offener Quader mit $\mu(U) = (2r)^{n-1} \cdot 2\varepsilon$ und $(U, \emptyset)$ ein Meßvorgang für X.

Weil nach (c) auch jede Vereinigung endlich vieler solcher Mengen eine Nullmenge ist, sehen wir jetzt, daß nicht nur die offenen, sondern alle Quader meßbar sind und daß ihr Maß das Produkt der Kantenlängen ist, unabhängig davon, welche Seiten(hyper)flächen mit zum Quader gehören.

(3) Die Menge $\mathbb{Q} \subset \mathbb{R}$ ist eine Nullmenge. Wir wählen zum Beweis eine Abzählung $\mathbb{Q} = \{x_k \,|\, k \in \mathbb{N}\}$ und definieren Intervalle

$$U_k = (x_k - 2^{-k}\varepsilon,\ x_k + 2^{-k}\varepsilon).$$

Dann ist

$$U := \bigcup_{k=0}^{\infty} U_k$$

offen, also $(U, \emptyset)$ ein Meßvorgang für $\mathbb{Q}$. Für den Meßfehler garantiert Lemma $30\frac{1}{3}.9$

$$\mu(U) \le \sum_{k=0}^{\infty} \mu(U_k) = 2 \cdot \sum_{k=0}^{\infty} 2^{-k}\varepsilon = 4\varepsilon,$$

woraus die Behauptung sofort folgt.

Dieses Ergebnis ist wirklich erstaunlich: Einerseits muß jede $\mathbb{Q}$ umfassende Menge U gewiß mindestens so dicht in $\mathbb{R}$ liegen wie $\mathbb{Q}$ selbst. Andererseits sind offene Mengen $U \subset \mathbb{R}$ definitionsgemäß von einer gewissen "Dicke": um jedes $a \in U$ liegt ein noch ganz in U enthaltenes Intervall $(a-\delta, a+\delta)$. Man sollte also meinen, daß die einzige $\mathbb{Q}$ umfassende offenen Menge $\mathbb{R}$ selbst ist. Weit gefehlt: wie wir gerade gesehen haben, gibt es solche Mengen sogar mit beliebig klein vorgegebenem positiven Maß!

Jetzt verstehen wir schon besser, warum es nicht geschickt gewesen wäre, beliebige Teilmengen von $\mathbb{R}$ durch Ausschöpfen mit Aggregaten messen zu wollen. Weil zwischen je zwei rationalen Zahlen eine irrationale liegt, enthält die Menge $\mathbb{Q}$ kein einziges Elementarintervall, so daß diese Methode zum korrekten Maß null führen würde:

$$\lim_{k\to\infty} v(\mathbb{Q}^{(k)}) = \lim_{k\to\infty} v(\emptyset) = 0.$$

Aber das trifft aus demselben Grund auch auf die komplementäre Menge $\mathbb{R}\backslash\mathbb{Q}$ der irrationalen Zahlen zu:

$$\lim_{k\to\infty} v\big((\mathbb{R}\backslash\mathbb{Q})^{(k)}\big) = \lim_{k\to\infty} v(\emptyset) = 0$$

Wenn also auch nur die endliche Additivität des Maßes gelten soll, muß $\mathbb{R} = \mathbb{Q} \cup (\mathbb{R}\backslash\mathbb{Q})$ dann selbst eine Nullmenge sein, was absurd ist. In Wirklichkeit liegen aus der Sicht der Maßtheorie Welten zwischen der Mengen $\mathbb{Q}$ und $\mathbb{R}\backslash\mathbb{Q}$. Was wir bisher über μ schon gelernt haben, reicht ja zum Beispiel aus, um für jedes offene Intervall $(a, b) \subset \mathbb{R}$ das Maß

$$\mu\big((a,b)\backslash\mathbb{Q}\big) = \mu\big((a,b)\big) - \mu\big((a,b) \cap \mathbb{Q}\big) = \mu\big((a,b)\big) - 0 = b-a$$

zu bestimmen: die rationalen Punkte des Intervalls tragen zum Maß einfach nicht bei.

Die Subtilität der lebesgueschen Maßkonstruktion liegt darin, die zu messende Menge von innen mit abgeschlossenen Mengen auszuschöpfen (Aggregate oder offene Mengen wären dafür zu "dick"), aber in offene einzuschließen (wozu wieder die abgeschlossenen nicht so geeignet wären). Die offenen (per Komplementbildung auch die abgeschlossenen) selbst darf man ruhig nach der "naiven" Methode messen, wie wir es ja auch gemacht haben.

Jetzt haben wir schon viele der in Satz $30\frac{1}{3}.4\frac{1}{2}$ versprochenen Eigenschaften für das Lebesgue-Maß bewiesen, und wir wenden uns den Axiomen zu, die mit (unendlichen) Folgen von meßbaren Mengen zu tun haben:

(e) Ist $X_j \subset \mathbb{R}^n$ für jedes $j \in \mathbb{N}$ meßbar, so ist $\bigcup_{j=0}^{\infty} X_j$ meßbar, und

(f) wenn die X_j außerdem paarweise disjunkt sind, gilt $\mu(\bigcup_{j=0}^{\infty} X_j) = \sum_{j=0}^{\infty} \mu(X_j)$.

Die entsprechenden Aussagen für endlich viele meßbare Mengen folgen sofort aus Lemma $30\frac{1}{3}.12$ (c,d) durch vollständige Induktion. Und auch zum Beweis von (e) dürfen wir wegen

$$X := \bigcup_{j=0}^{\infty} X_j = X_0 \cup (X_1 \setminus X_0) \cup \big(X_2 \setminus (X_0 \cup X_1)\big) \cup \cdots$$

die X_j als paarweise disjunkt voraussetzen. Außerdem betrachten wir zuerst nur den Fall, daß X beschränkt, etwa im Würfel $W_r \subset \mathbb{R}^n$ enthalten ist.

Sei $\varepsilon > 0$. Für jedes j wählen wir einen Meßvorgang (U_j, F_j) für X_j mit $U_j \subset W_r$ und $\mu(U_j \setminus F_j) < 2^{-j}\varepsilon$. Dann ist für jedes $k \in \mathbb{N}$ das Paar

$$(\bigcup_{j=0}^{\infty} U_j, \bigcup_{j=0}^{k} F_j)$$

ein Meßvorgang für X, dessen Meßfehler wir unter Verwendung von (d) und Lemma $30\frac{1}{3}.9$ durch

$$\begin{aligned}
\mu(\bigcup_{j=0}^{\infty} U_j \setminus \bigcup_{j=0}^{k} F_j) &\le \mu(\bigcup_{j=0}^{k}(U_j \setminus F_j) \cup \bigcup_{j=k+1}^{\infty} U_j) \\
&\le \sum_{j=0}^{k} \mu(U_j \setminus F_j) + \sum_{j=k+1}^{\infty} \mu(U_j) \\
&< \qquad 2\varepsilon \qquad + \sum_{j=k+1}^{\infty} \mu(U_j)
\end{aligned}$$

abschätzen. Wegen $X \subset W_r$ und der endlichen Additivität (d) von μ sind die Partialsummen $\sum_{j=0}^{k} \mu(X_j)$ durch $\mu(W_r) = (2r)^n$ nach oben beschränkt, also konvergieren die Reihe $\sum_{j=0}^{\infty} \mu(X_j)$ und wegen

$$\mu(U_j) < \mu(X_j) + 2^{-j}\varepsilon \quad \text{für alle } j \in \mathbb{N}$$

auch die Reihe $\sum_{j=0}^{\infty} \mu(U_j)$ nach dem Majorantenkriterium:

$$\sum_{j=0}^{\infty} \mu(U_j) < \infty$$

(bisher stand das Symbol $\sum_{?}^{\infty}$ ja nur für einen Wert in $[0, \infty]$). Damit wird für genügend große k der Reihenrest $\sum_{j=k+1}^{\infty} \mu(U_j)$ kleiner als ε und damit der gesamte Meßfehler von $(\bigcup_{j=0}^{\infty} U_j, \bigcup_{j=0}^{k} F_j)$ kleiner als 3ε. Das beweist die Meßbarkeit von X.

Die Formel (f) folgt nun leicht aus der endlichen Additivität (d): einerseits ist für jedes k

$$\sum_{j=0}^{k} \mu(X_j) = \mu(\bigcup_{j=0}^{k} X_j) \le \mu(X)$$

und damit $\sum_{j=0}^{\infty} \mu(X_j) \le \mu(X)$, andererseits

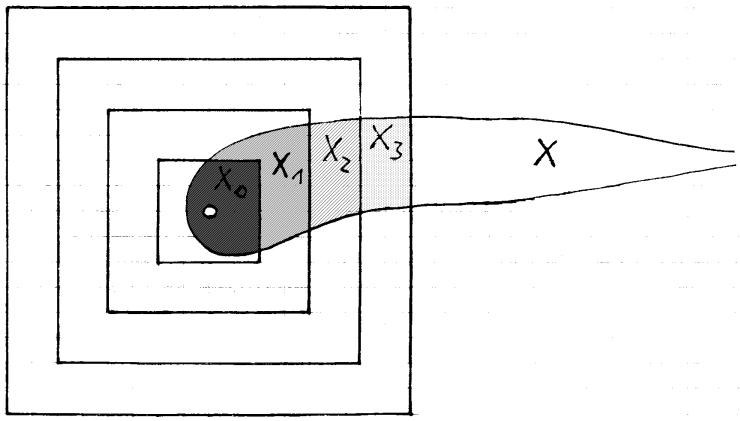
$$\mu(X) \le \mu(\bigcup_{j=0}^{\infty} U_j) \le \sum_{j=0}^{\infty} \mu(U_j) \le \sum_{j=0}^{\infty} \big(\mu(X_j) + 2^{-j}\varepsilon\big) = \sum_{j=0}^{\infty} \mu(X_j) + 2\varepsilon,$$

also $\mu(X) \le \sum_{j=0}^{\infty} \mu(X_j) + 2\varepsilon$ für jedes $\varepsilon > 0$ und folglich auch $\mu(X) \le \sum_{j=0}^{\infty} \mu(X_j)$.

Um uns schließlich von der Voraussetzung zu befreien, die zu messende Menge sei beschränkt, zeigen wir noch: Ist $X \subset \mathbb{R}^n$ eine beliebige Teilmenge mit der Eigenschaft, daß für jedes $j \in \mathbb{N}$ der Durchschnit $X \cap W_j$ meßbar ist, so ist X selbst meßbar und es gilt

$$\mu(X) = \lim_{j \to \infty} \mu(X \cap W_j) = \sum_{j=0}^{\infty} \mu(X \cap W_{j+1} \setminus W_j).$$

Dazu setzen wir $X_j := X \cap W_{j+1} \setminus W_j$



und wählen Meßvorgänge (U_j, F_j) für X_j wie oben. Diesmal garantiert aber Lemma 30.8, daß die unendliche Vereinigung $\bigcup_{j=0}^{\infty} F_j$ eine abgeschlossene Teilmenge von $\mathbb{R}^n$ ist, denn die Glieder einer in $\mathbb{R}^n$ konvergenten Folge in dieser Menge können nur endlich viele der Mengen $W_{j+1} \setminus W_j$ treffen. Deshalb ist

$$(\bigcup_{j=0}^{\infty} U_j, \bigcup_{j=0}^{\infty} F_j)$$

ein Meßvorgang für X, woraus alles Weitere sofort folgt.

Damit ist die Konstruktion des Lebesgue-Maßes abgeschlossen und Satz $30\frac{1}{3}.4\frac{1}{2}$ vollständig bewiesen. Wir halten als Nebenergebnisse noch fest:

$30\frac{1}{3}$.13 Korollar Beliebige (und nicht nur beschränkte) offene oder abgeschlossene Mengen X sind meßbar, ebenso alle Aggregate. In allen Fällen hat $\mu(X)$ die alte Bedeutung beziehungsweise die von $v(X)$. — Ist $Y \subset \mathbb{R}^n$ eine Nullmenge und $X \subset Y$, so ist auch X eine Nullmenge.

Die letzterwähnte Tatsache ist zwar aufgrund der Definition völlig klar, aber trotzdem bemerkenswert, weil sie nicht formal aus den Axiomen folgt und auch nicht folgen kann, weil sie nicht in jedem Maßraum gültig ist.

Wenn man bedenkt, daß die Menge M der meßbaren Teilmengen von $\mathbb{R}^n$ eine σ-Algebra ist und deshalb auch alle Mengen umfassen muß, die aus abzählbar vielen der im Korollar genannten durch Mengenoperationen entstehen, fragt man sich, wie denn eine nicht meßbare Menge aussieht. Nun ist es nicht all zu schwer, solche Mengen zu konstruieren, aber die Konstruktionen sind wenig explizit und geben keine konkrete Vorstellung: man kann wohl mit Recht sagen, daß eine Menge unvorstellbar diffus sein muß, um nicht meßbar zu sein!

Zum Schluß beweisen wir die am Anfang erwähnte, aber nicht zum Axiom erklärte Translationsinvarianz des Lebesgue-Maßes (die schwieriger zu beweisende Invarianz unter Drehungen stellen wir noch zurück).

$30\frac{1}{3}$.14 Satz Sei $X \subset \mathbb{R}^n$ meßbar, und $a \in \mathbb{R}$. Dann ist auch die Menge

$$a + X = \{a+x \,|\, x \in X\} \subset \mathbb{R}^n$$

meßbar, und $\mu(a+X) = \mu(X)$.

Beweis Für offene X ist

$$\mu(X) = \sup\Big\{\sum_{j=1}^{r} \mu(Q_j) \,\Big|\, Q_1, \ldots, Q_r \subset X \text{ sind paarweise disjunkte offene Quader}\Big\}$$

eine alternative Beschreibung des Maßes, denn die rechte Seite ist gewiß nicht zu groß, aber auch nicht zu klein, weil man als Q_j die ihrer Oberflächen beraubten Elementarwürfel eines in X enthaltenen Aggregates nehmen kann. Damit haben wir eine a priori translationsinvariante Beschreibung des Maßes von offenen Mengen. Sie charakterisiert über die Meßvorgänge dann aber auch Meßbarkeit und Maß beliebiger Mengen translationsinvariant.

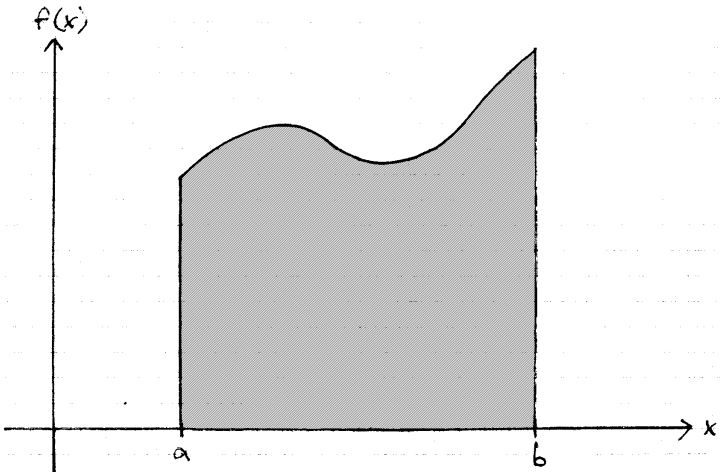
30 2/3 Vom Maß zum Integral

Natürlich ist der inzwischen übliche Integralbegriff nicht von heute auf morgen entstanden — das ändert nichts an der Notwendigkeit, ihn in der Grundvorlesung von jetzt auf gleich zu erklären. Immerhin sind die Grundzüge schnell beschrieben.

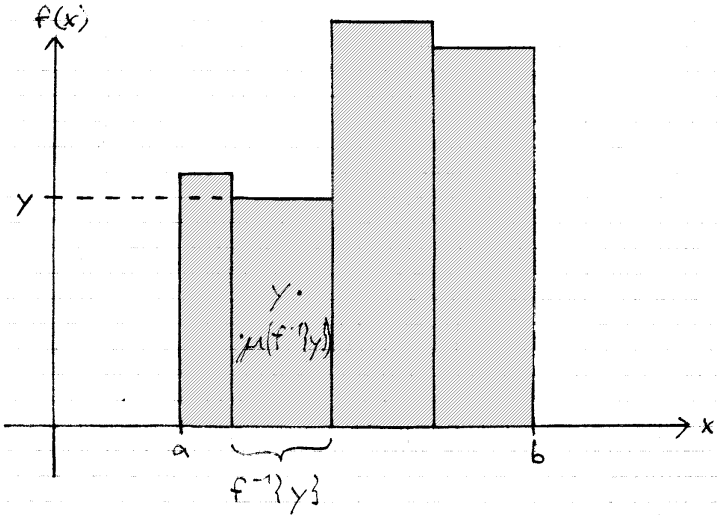
Wie schon erwähnt geht es darum, Volumina zu berechnen, genauer gesagt erst mal solche einer speziellen Art. Im einfachsten Fall betrachten wir eine überall positive reellwertige Funktion $f\colon [a,b] \to \mathbb{R}$ auf einem kompakten Intervall. Das (noch zu definierende) Integral von f soll der Flächeninhalt der zwischen $[a,b]\times\{0\}$ und dem Graphen

$$\Gamma_f = \big\{\big(x, f(x)\big) \,\big|\, x \in [a,b]\big\} \subset \mathbb{R}^2$$

liegenden Teil der Ebene berechnen:



Sicher ist nicht auf den ersten Blick klar, wie das allgemein gehen soll, aber in dem sehr speziellen Fall, daß f eine "Treppenfunktion" ist, brauchen wir nur für jedes Treppenniveau y nachzuschauen, was unter der zugehörigen Stufe liegt,



und die Fächeninhalte der zugehörigen "Rechtecke" aufzuaddieren, in gelehrter Sprache also

$$\sum_{\text{alle Niveaus } y} y \cdot \mu(f^{-1}\{y\})$$

zu berechnen. Eigentlich alle Ansätze zur Integrationstheorie gehen von diesem einfachen Sachverhalt aus und übertragen ihn auf allgemeinere Funktionen f, indem sie diese durch Treppenfunktionen approximieren, das heißt f als Grenzwert einer Folge von Treppenfunktionen darstellen.

Wie weit man damit kommt, hängt von Feinheiten ab, vor allem davon, welche Mengen man als die Treppenstufen, also die Fasern $f^{-1}\{y\}$ zuläßt und welche Art von Konvergenz man bei der Approximation ins Auge faßt. Bei der heute überholten Version von Riemann aus der Mitte des 19. Jahrhundert sind die Treppenstufen im wesentlichen selbst Intervalle (obige "Rechtecke" also wirkliche Rechtecke), und der Konvergenzbegriff ist der der gleichmäßigen Konvergenz. Viel zweckmäßiger und aus heutiger Sicht naheliegend dagegen ist der Ansatz von Lebesgue, der von den Stufen nichts weiter vorauszusetzt als daß sie meßbare Mengen von endlichem Maß sind, und sich für die Approximation mit einem schwächeren Konvergenzbegriff begnügt. Man kann dabei schon vorhandene Kenntnisse über das Maß einbringen und wird mit einer insgesamt einfacheren und befriedigenderen Integrationstheorie belohnt.

Ziel dieses Abschnittes ist es, auf der Basis des Maßbegriffs den des Integrals zu konstruieren. Da ersterer ganz allgemein ist, besteht kein Grund, sich bei den Integranden auf Funktionen von nur einer Variablen zu beschränken — im Gegenteil ist die Konstruktion so formal, daß wir von einen ganz beliebigen Maßraum (M, M, μ) als Geschäftsgrundlage des Abschnitts ausgehen werden, ohne daß das zusätzliche Mühe kostet. Sollte Sie daran aber der Grad der Abstraktheit stören, denken Sie ruhig an den im vorigen Abschnitt beschriebenen lebesgueschen Maßraum mit $M = \mathbb{R}^n$; nur den werden wir später verwenden.

$30\frac{2}{3}$.1 Definition Eine Funktion $f\colon M \to \mathbb{R}$ heißt eine Treppenfunktion, wenn sie nur endlich viele Werte annimmt und für jedes $y \in \mathbb{R}$ mit $y \neq 0$ die "Stufe"

$$f^{-1}\{y\} \subset M$$

meßbar ist und endliches Maß $\mu(f^{-1}\{y\}) < \infty$ hat. Das Integral dieser Treppenfunktion f ist die Zahl

$$\int f := \sum_{\substack{y \in \mathbb{R} \\ y \neq 0}} y \cdot \mu(f^{-1}\{y\}) \in \mathbb{R}.$$

Bemerkungen Die Funktion $f\colon \mathbb{R} \to \mathbb{R}$ aus Beispiel 1.8 (4) mit $f(x) = 1$ für rationale und $f(x) = 0$ für irrationale x ist eine Treppenfunktion! — Wer das verwendete Maß mit in das Integralsymbol aufnehmen will, schreibt $\int f d\mu$. — Beim Umgang mit der das Integral definierenden Formel ist die Zerlegung von M in die Stufen von f häufig zu starr, und man zieht dann die folgende Variante vor:

Ist $M = \bigcup_{j=1}^{r} X_j$ eine endliche Zerlegung der Menge M in meßbare Teilmengen X_j derart, daß $f|X_j$ den konstanten Wert y_j hat und im Fall $y_j \neq 0$ das Maß $\mu(X_j)$ endlich ist, dann gilt

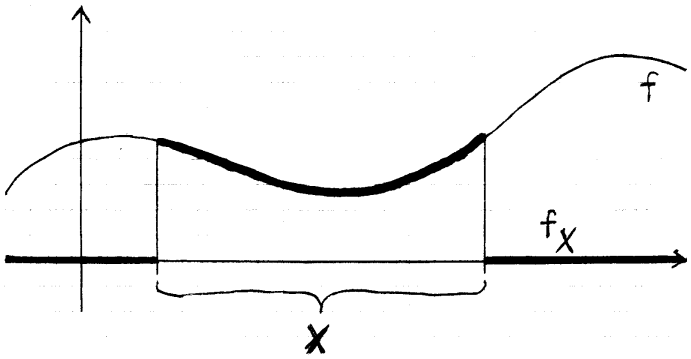
$$\int f = \sum_{\substack{j=1 \\ y_j \neq 0}}^{r} y_j \cdot \mu(X_j).$$

Denn die gröbste Zerlegung dieser Art ist gerade die in die Stufen, und alle weiteren entstehen aus ihr durch Verfeinerung, wobei sich der Wert der Summe aufgrund der endlichen Additivität des Maßes nicht ändert.

$30\frac{2}{3}$.2 Notationen Ist $X \subset M$ eine Teilmenge und $f\colon X \to \mathbb{R}$ eine auf X oder auch einer X umfassenden Teilmenge von M definierte Funktion, so sei

$$f_X\colon M \longrightarrow \mathbb{R}$$

als die Fortsetzung durch die Null



$$f_X(x) := \begin{cases} f(x) & \text{für } x \in X \\ 0 & \text{sonst} \end{cases}$$

erklärt. Ist X meßbar und f eine Treppenfunktion auf M, so ist auch f_X eine Treppenfunktion, und wir schreiben

$$\int_X f := \int f_X \quad \text{oder ausführlicher} \quad \int_X f d\mu := \int f_X d\mu.$$

Die Treppenfunktionen $f\colon M \to \mathbb{R}$ bilden unter punktweisen Rechenoperationen einen reellen Vektorraum, den wir mit T oder $\mathsf{T}(\mu)$ bezeichnen, und das Integral definiert eine Linearform

$$\begin{aligned} \mathsf{T}(\mu) &\longrightarrow \mathbb{R} \\ f &\mapsto \int f d\mu. \end{aligned}$$

Schließlich dürfen wir

$$\|f\| := \int |f| d\mu \quad \text{für jedes } f \in \mathsf{T}(\mu)$$

setzen, denn mit f ist auch $|f|$ eine Treppenfunktion.

$30\frac{2}{3}$.3 Lemma Die Funktion $\mathsf{T} \ni f \longmapsto \|f\| \in \mathbb{R}$ ist eine Halbnorm, das heißt:

- $\|f\| \geq 0$ für alle $f \in \mathsf{T}$,
- $\|\lambda f\| = |\lambda| \|f\|$ für alle $\lambda \in \mathbb{R},\ f \in \mathsf{T}$,
- $\|f + g\| \leq \|f\| + \|g\|$ für alle $f, g \in \mathsf{T}$.

Beweis Ganz einfach, man zerlegt M in meßbare Teilmengen, auf denen f *und* g konstante Werte haben.

Übrigens handelt es sich in der Regel nicht um eine Norm, weil $\|f\| = 0$ auch für nicht-triviale Treppenfunktionen f möglich ist. Der Begriff der Halbnorm reicht aber aus, um den der Cauchy-Folge aus dem Abschnitt 4 wörtlich auf Folgen in T zu übertragen:

$30\frac{2}{3}$.4 Definition Eine Folge $(f_k)_{k=0}^{\infty}$ in T heißt eine Cauchy-Folge, wenn es zu jedem $\varepsilon > 0$ ein $D \in \mathbb{N}$ gibt mit

$$\|f_j - f_k\| < \varepsilon \quad \text{für alle } k > D \text{ und alle } j \geq k.$$

Beachten Sie, daß hier explizit $\|f_j - f_k\| = \int |f_j - f_k|$ ist. Unser Vorhaben ist, das bisher ja nur auf ganz naive Weise für Treppenfunktionen erklärte Integral allgemeiner für jede Grenzfunktion einer Cauchy-Folge dadurch zu definieren, daß wir die Vertauschbarkeit von Limes und Integral fordern:

$$\int \lim_{k\to\infty} f_k := \lim_{k\to\infty} \int f_k.$$

Das ist allerdings erst mal nur eine Idee, da wir über das Konvergenzverhalten dieser Cauchy-Folgen gar nichts wissen. Wie wir später sehen werden, ist es möglich, daß eine Cauchy-Folge an keiner einzigen Stelle konvergiert — was zwar kein unmittelbares Hindernis für die Realisierung ist, aber doch eher nahelegt, daß es sich um eine Schnapsidee handelt: es scheint ja nicht mal sicher zu sein, *wovon* $\int f$ das Integral sein soll. Wunderbarerweise geht es aber doch; der Schlüssel ist der

30$\frac{2}{3}$.5 Satz Sei $(f_k)_{k=0}^{\infty}$ eine Cauchy-Folge in T. Dann gibt es eine Teilfolge, die fast überall punktweise konvergiert. Zu jedem $\varepsilon > 0$ gibt es eine Menge $Z \in \mathsf{M}$ mit $\mu(Z) < \varepsilon$, derart daß die Konvergenz der Teilfolge auf $M \setminus Z$ gleichmäßig ist.

Zum ersten Mal benutze ich dabei die praktische

30$\frac{2}{3}$.6 Sprechweise Man sagt, eine (von einem Punkt in M abhängige) Aussage gilt fast überall (in M), wenn sie in jedem Punkt außerhalb einer Nullmenge $N \in \mathsf{M}$ gilt.

Beweis des Satzes Klar ist, daß das nicht ganz einfach sein kann, denn zunächst ist kein naheliegender Kandidat für die Teilfolge oder die Menge Z in Sicht. Listig wählen wir nun aber erstere so aus, daß sich für die Teilfolge, durch die wir die Ausgangsfolge gleich ersetzen, die Cauchy-Eigenschaft zu

$$\|f_j - f_k\| < 2^{-2k} \quad \text{für alle } j \geq k \in \mathbb{N}$$

konkretisiert; es ist klar, daß sich das durch vollständige Induktion machen läßt.

Die Mengen

$$Y_k := \left\{x \in M \,\middle|\, |f_{k+1} - f_k| \geq 2^{-k}\right\} \subset M$$

sind (als Vereinigungen von Treppenstufen) meßbare Teilmengen von M, und es gilt

$$2^{-k} \cdot \mu(Y_k) \leq \int |f_{k+1} - f_k| = \|f_{k+1} - f_k\| < 2^{-2k},$$

also $\mu(Y_k) < 2^{-k}$. Setzen wir nun

$$Z_k = \bigcup_{j=k}^{\infty} Y_j,$$

so ist $\mu(Z_k) < 2 \cdot 2^{-k} = 2^{-k+1}$, und $x \notin Z_k$ bedeutet

$$|f_{j+1}(x) - f_j(x)| < 2^{-j} \quad \text{für jedes } j \geq k.$$

Nach der Dreiecksungleichung impliziert das (geometrische Reihe!) allgemeiner

$$|f_i(x) - f_j(x)| < 2^{-j+1} \quad \text{für alle } i \geq j \geq k,$$

und damit erweist sich die Folge der auf $M \setminus Z_k$ eingeschränkten Funktionen $\big(f_j|(M \setminus Z_k)\big)_{j=0}^{\infty}$ als eine gleichmäßige Cauchy-Folge, mithin als gleichmäßig konvergent — zwar habe ich das gleichmäßige Cauchy-Kriterium 11.5 im damaligen Kontext für auf einer Menge von komplexen Zahlen definierte Funktionen formuliert, aber der Beweis nimmt auf diese Voraussetzung überhaupt keinen Bezug.

Da all diese Aussagen für beliebiges $k \in \mathbb{N}$ gelten, ist der zweite Teil der Satzbehauptung damit bewiesen. Der erste folgt nun daraus, daß die Folge $(f_j)_{j=0}^{\infty}$ außerhalb der Menge

$$Z := \bigcap_{k=0}^{\infty} Z_k \subset M$$

jedenfalls punktweise konvergiert, und daß Z wegen $\mu(Z) \leq \mu(Z_k) < 2^{-k+1}$ eine Nullmenge ist.

Viel einfacher haben wir es mit der Folge der Integrale einer Cauchy-Folge in T.

30$\frac{2}{3}$.7 Notiz Ist $(f_k)_{k=0}^{\infty}$ eine Cauchy-Folge in T, so ist die Zahlenfolge $\left(\int f_k\right)_{k=0}^{\infty}$ konvergent.

Beweis Natürlich gilt $\left|\int f\right| \le \int |f|$ für jede Treppenfunktion f. Wegen

$$\left|\int f_i - \int f_j\right| \le \int |f_i - f_j| = \|f_i - f_j\|$$

ist die Folge der Integrale also eine Cauchy-Folge in $\mathbb{R}$, das heißt eine konvergente Folge.

Wir fühlen uns durch die beiden letzten Resultate ermutigt, das Integral zu definieren:

30$\frac{2}{3}$.8 Definition Eine Funktion $f\colon M \to \mathbb{R}$ heißt integrierbar, wenn es eine Cauchy-Folge $(f_k)_{k=0}^{\infty}$ in T gibt, die fast überall punktweise gegen f konvergiert. Die Menge aller integrierbaren Funktionen wird mit L^1 oder, wenn man das zugrundegelegte Maß mit notieren will, $\mathsf{L}^1(\mu)$ bezeichnet. Ist f integrierbar, so heißt die reelle Zahl

$$\int f := \lim_{k\to\infty} \int f_k$$

das Integral von f (über M bezüglich des Maßes μ).

Daß der Grenzwert überhaupt existiert, folgt aus der Notiz, und das verbleibende Eindeutigkeitsproblem wird durch den folgenden Satz gelöst, zu dessen Beweis wir uns noch einmal anstrengen müssen.

30$\frac{2}{3}$.9 Satz Seien $(f_k)_{k=0}^{\infty}$ und $(g_k)_{k=0}^{\infty}$ Cauchy-Folgen in T, die beide fast überall gegen dieselbe Funktion konvergieren. Dann gilt

$$\lim_{k\to\infty} \int |f_k - g_k| = 0,$$

erst recht also

$$\lim_{k\to\infty} \int f_k = \lim_{k\to\infty} \int g_k.$$

Beweis "Erst recht" wegen $\left|\int f_k - \int g_k\right| \le \int |f_k - g_k|$.

Zum Beweis des Satzes bilden wir die Differenzfolge $h_k = f_k - g_k$. Sie ist eine Cauchy-Folge, die fast überall gegen die Nullfunktion konvergiert, und zu beweisen ist $\lim_{k\to\infty} \|h_k\| = 0$. Ich behaupte, daß es genügt, das für eine Teilfolge zu beweisen. Ist das geschehen, finden wir nämlich zu jedem $\varepsilon > 0$ ein $D \in \mathbb{N}$ mit $\|h_k - h_l\| < \varepsilon$ für alle $k, l > D$. Wir verfügen dann $l > D$ als einen der in der Teilfolge auftretenden Indizes und außerdem so groß, daß $\|h_l\| < \varepsilon$ ist: es folgt

$$\|h_k\| \le \|h_k - h_l\| + \|h_l\| < 2\varepsilon \quad \text{für alle } k > D.$$

Wir entscheiden uns für eine Teilfolge nach Satz 30$\frac{2}{3}$.5, dürfen also annehmen, daß es zu jedem $\varepsilon > 0$ eine Menge $Z \in \mathsf{M}$ mit $\mu(Z) < \varepsilon$ derart gibt, daß die Folge $(h_j)_{j=0}^{\infty}$ auf $M\backslash Z$ gleichmäßig gegen die Nullfunktion konvergiert.

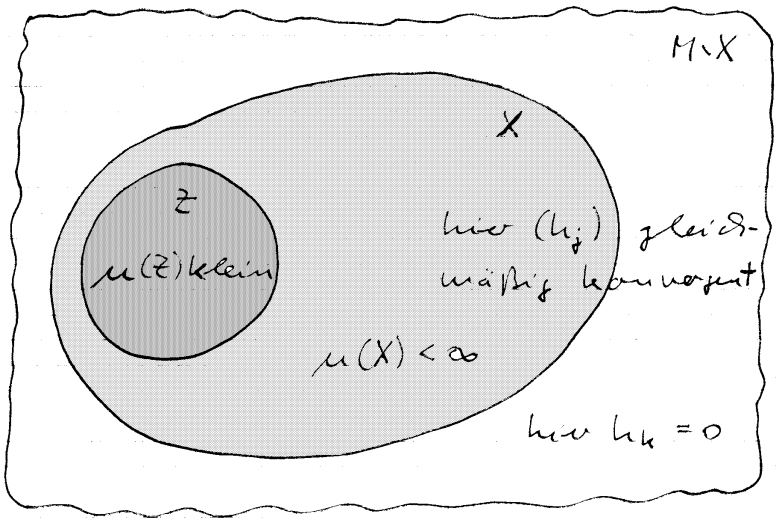
Sei nun $\varepsilon > 0$ gegeben. Wir fixieren ein genügend großes $k \in \mathbb{N}$, so daß

$$\|h_j - h_k\| < \varepsilon \quad \text{für alle } j \ge k$$

gilt. Außerdem wählen wir $Z \in \mathsf{M}$ so, daß

$$\max\left\{|h_k(x)| \,\middle|\, x \in M\right\} \cdot \mu(Z) < \varepsilon$$

ist und die Folge $(h_j)_{j=0}^{\infty}$ auf $M\backslash Z$ gleichmäßig konvergiert: das Maximum existiert einfach deswegen, weil h_k eine Treppenfunktion ist und daher nur endlich viele Werte annimmt. Aus demselben Grund finden wir noch eine Z umfassende Menge $X \in \mathsf{M}$ von endlichem Maß mit $h_k|(M\backslash X) = 0$.



Jetzt zerlegen wir

$$\|h_j\| = \int_M |h_j| = \int_Z + \int_{X\setminus Z} + \int_{M\setminus X}$$

und schätzen einzeln ab:

$$\int_Z \;\le\; \int_Z |h_j - h_k| + \int_Z |h_k| \;\le\; \|h_j - h_k\| + \max\big\{|h_k(x)| \,\big|\, x \in M\big\} \cdot \mu(Z) \;<\; \varepsilon + \varepsilon$$

für $j \ge k$,

$$\int_{X\setminus Z} \;\le\; \max\big\{|h_j(x)| \,\big|\, x \in M\big\} \cdot \mu(X) \;<\; \varepsilon$$

für alle genügend großen j (gleichmäßige Konvergenz gegen die Nullfunktion!), und

$$\int_{M\setminus X} \;\le\; \int_{M\setminus X} |h_j - h_k| + \int_{M\setminus X} |h_k| \;\le\; \|h_j - h_k\| \;<\; \varepsilon$$

für $j \ge k$ — insgesamt also $\|h_j\| < 4\varepsilon$ für alle genügend großen j.

Zusammenfassung Wir haben das Integral als eine Funktion

$$\int : \mathrm{L}^1 \longrightarrow \mathbb{R}$$

so definiert: für Treppenfunktionen f durch die elementare Formel, und allgemein dadurch, daß wir die zu integrierende Funktion f fast überall als Grenzfunktion einer Cauchy-Folge $(f_k)_{k=0}^{\infty}$ von Treppenfunktionen auffassen und $\int f = \lim \int f_k$ setzen. Die Definition von L^1 garantiert, daß es überhaupt solche Folgen gibt, und Satz $30\frac{2}{3}.9$ stellt sicher, daß das Resultat nicht von der speziellen Wahl dieser Folge abhängt. Übrigens kommt der Name L^1 daher, daß allgemeiner auch L^p für reelle $p > 0$ eine Rolle spielt, wobei es da um die Integrierbarkeit p-ten Potenzen geht.

Wie schon bei den Treppenfunktionen — die ja Spezialfälle integrierbarer Funktionen sind — vereinbaren wir: Ist $X \subset M$ eine Teilmenge, so nennen wir eine Funktion $f: X \to \mathbb{R}$ über X integrierbar, wenn $f_X: M \to \mathbb{R}$ integrierbar ist, und schreiben dann $\int_X f = \int f_X$. Bei Bedarf schreiben wir $\mathrm{L}^1 X$ für die Menge der über X integrierbaren Funktionen.

Ohne die einfachen Beweise notieren wir die ersten

$30\frac{2}{3}.10$ Regeln für Integrale:

(a) L^1 ist ein reeller Vektorraum und $\int\colon \mathsf{L}^1 \to \mathbb{R}$ ist linear.

(b) Ist $f \in \mathsf{L}^1$ und $f \geq 0$ im Sinne von

$$f(x) \geq 0 \quad \text{für jedes } x \in M,$$

so ist $\int f \geq 0$.

(c) Ist $X \in \mathsf{M}$ eine meßbare Teilmenge endlichen Maßes, so ist $1_X \in \mathsf{L}^1$ und

$$\int 1_X = \int_X 1 = \mu(X).$$

(d) Sind X und Y disjunkte meßbare Teilmengen von M, so ist eine Funktion $f\colon X \cup Y \to \mathbb{R}$ genau dann über $X \cup Y$ integrierbar, wenn sie über X und über Y integrierbar ist, und dann gilt

$$\int_{X\cup Y} f = \int_X f + \int_Y f.$$

(e) Aus $f \in \mathsf{L}^1$ folgt $|f| \in \mathsf{L}^1$ und

$$\left|\int f\right| \leq \int |f|.$$

(f) Die Funktion

$$\mathsf{L}^1 \ni f \longmapsto \|f\| := \int |f|$$

ist eine Halbnorm auf L^1.

(g) Eine Funktion $f \in \mathsf{L}^1$ auf einer Nullmenge abzuändern hat weder auf die Integrierbarkeit noch auf den Wert des Integrals einen Einfluß, ändert insbesondere die Halbnorm $\|f\|$ nicht.

Die letztgenannte Aussage erlaubt auch eine Umkehrung, deren Beweis aber keineswegs auf der Hand liegt:

$30\frac{2}{3}$.11 Satz Eine Funktion $f\colon M \to \mathbb{R}$ ist genau dann integrierbar mit $\|f\| = 0$, wenn f fast überall verschwindet.

Beweis Regel (g) gibt die einfache Richtung. Zu betrachten bleibt eine beliebige Funktion $f \in \mathsf{L}^1$ mit $\|f\| = 0$. Wir wählen eine Cauchy-Folge $(f_k)_{k=0}^\infty$ in T, die fast überall gegen f konvergiert, und dazu eine Nullmenge $N \in \mathsf{M}$ so, daß

$$\lim_{k\to\infty} f_k(x) = f(x) \quad \text{für alle } x \in M\backslash N$$

gilt. Wenn wir auf N alle f_k und auch f selbst zu null abändern, was im Hinblick auf die zu beweisende Behauptung ja erlaubt ist, erreichen wir, daß die Folge $(f_k)_{k=0}^\infty$ sogar überall gegen f konvergiert.

Nun ist für jedes positive $k \in \mathbb{N}$ die Menge $Y_k := \left\{x \in M \,\middle|\, |f(x)| \geq \frac{1}{k}\right\}$ meßbar; dazu schreiben wir sie listig in der Form

$$\begin{aligned} Y_k &= \bigcap_{\varepsilon>0} \bigcup_{D=0}^{\infty} \bigcap_{j=D+1}^{\infty} \left\{x \in M \,\middle|\, |f_j(x)| > \frac{1}{k} - \varepsilon\right\} \\ &= \bigcap_{l=1}^{\infty} \bigcup_{D=0}^{\infty} \bigcap_{j=D+1}^{\infty} \left\{x \in M \,\middle|\, |f_j(x)| > \frac{1}{k} - \frac{1}{l}\right\}, \end{aligned}$$

beachten, daß jedes f_j eine Treppenfunktion ist, und erinnern uns an die Eigenschaften einer σ-Algebra. Aus der Abschätzung

$$\frac{1}{k} \cdot \mu(Y_k) \leq \int |f| = \|f\| = 0$$

folgt aber $\mu(Y_k) = 0$, und damit ist auch

$$\big\{x \in M \,\big|\, f(x) \neq 0\big\} = \bigcup_{k=1}^{\infty} Y_k$$

eine Nullmenge.

Was Sie an der soweit entwickelten Integrationstheorie stören mag, ist, daß wir noch gar keine Vorstellung davon haben, welche Funktionen nun integrierbar sind, abgesehen davon, daß jedenfalls die Treppenfunktionen es sind. Weil das für den uns vor allem interessierenden Fall des Lebesgue-Integrals auch nicht so leicht zu sagen ist, trösten wir uns erst mal mit einem

30$\frac{2}{3}$.12 Beispiel Wir legen den Maßraum $(\mathbb{N}, \mathbf{P}\mathbb{N}, |?|)$ aus 30$\frac{1}{3}$.4 zugrunde. Meßbar sind also alle Teilmengen von $\mathbb{N}$, und das Maß zählt einfach die Punkte. Insbesondere ist $\emptyset$ die einzige Nullmenge. Eine Funktion $f\colon \mathbb{N} \to \mathbb{R}$ ist das, was man üblicherweise eine Zahlenfolge nennt, und um eine Treppenfunktion handelt es sich genau dann, wenn fast alle Glieder null sind.

Sei nun $f\colon \mathbb{N} \to \mathbb{R}$ integrierbar. Nach Regel 30$\frac{2}{3}$.10(e) ist auch $|f|$ integrierbar. Weiter ist für jedes $n \in \mathbb{N}$ die Funktion $|f|_{\{0,\dots,n\}}$ eine Treppenfunktion (insbesondere integrierbar) mit $|f|_{\{0,\dots,n\}} \leq |f|$, was nach den Regeln (a,b)

$$\int_{\{0,\dots,n\}} |f| = \int |f|_{\{0,\dots,n\}} \leq \int |f|$$

nach sich zieht. Aber das links stehende Integral ist die Summe $\sum_{x=0}^{n} |f(x)|$, und wir sehen, daß die Partialsummenfolge der Reihe $\displaystyle\sum_{x=0}^{\infty} |f(x)|$ nach oben beschränkt, diese Reihe also konvergent ist.

Sei nun umgekehrt $f\colon \mathbb{N} \to \mathbb{R}$ eine Funktion mit der Eigenschaft, daß die Reihe $\displaystyle\sum_{x=0}^{\infty} f(x)$ absolut konvergiert. Ihre Partialsummenfolge hat die Cauchy-Eigenschaft, und wegen

$$\|f_{\{0,\dots,k+l\}} - f_{\{0,\dots,k\}}\| = \int |f_{\{0,\dots,k+l\}} - f_{\{0,\dots,k\}}| = \sum_{x=k+1}^{k+l} |f(x)| \quad \text{für alle } k, l \in \mathbb{N}$$

ist die Folge $(f_{\{0,\dots,n\}})_{n=0}^{\infty}$ eine Cauchy-Folge in T. Da sie offensichtlich punktweise gegen f konvergiert, schließen wir auf $f \in \mathsf{L}^1$ mit

$$\int f = \lim_{n\to\infty} \int f_{\{0,\dots,n\}} = \lim_{n\to\infty} \sum_{x=0}^{n} f(x) = \sum_{x=0}^{\infty} f(x).$$

Damit ist unser Beispiel eine Neuauflage der Theorie der unendlichen Reihen und speziell der absoluten Konvergenz. Nach unserem langen Weg durch die abstrakte Wüste der Maßtheorie ist diese Erkenntnis gewiß Balsam für die Seele. Sie läßt uns aber auch erwarten, daß das Integral im allgemeinen wesentliche Züge der Reihentheorie aufweist. Speziell im Fall des Lebesgue-Integrals, mit dem wir im folgenden arbeiten werden, trifft diese Erwartung weitgehend zu: das Integral über eine auf $\mathbb{R}^n$ definierte Funktion verhält sich wie eine kontinuierliche Version der Summe einer absolut konvergenten Reihe.

35 1/2 Zum Satz von der lokalen Umkehrung

Diesen Satz zu beweisen ist das einzige Ziel des Abschnitts:

35.4 Satz von der **lokalen Umkehrung** Sei $X \subset \mathbb{R}^n$ offen und $a \in X$. Eine C^1-Abbildung $f\colon X \to \mathbb{R}^n$ ist genau dann ein lokaler (C^1-)Diffeomorphismus bei a, wenn ihr Differential dort umkehrbar ist:

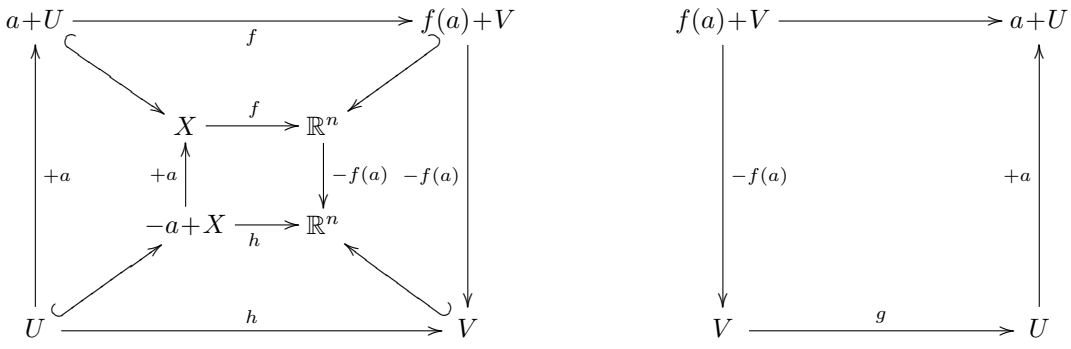
$$Df(a) \in GL(n, \mathbb{R})$$

Die einfache Richtung ist mit 35.2 erledigt, und wir müssen uns also eine C^1-Abbildung $f\colon X \to \mathbb{R}^n$ mit bei a invertierbarem Differential $Df(a) \in GL(n, \mathbb{R})$ vornehmen: Ziel ist es, offene Mengen $U \subset X$ um a und $V \subset \mathbb{R}^n$ um $f(a)$ so zu konstruieren, daß f die Menge U bijektiv auf V abbildet und die dadurch definierte Umkehrung $g\colon V \to U$ wieder stetig differenzierbar ist. Um dieses anspruchsvolle Projekt übersichtlich zu machen, gliedere ich es in vier kleine Abschnitte — lassen Sie sich von dieser Methode ruhig inspirieren, wenn Sie später mal eine eigene Arbeit mündlich oder schriftlich präsentieren wollen:

- Beweisvorbereitungen,
- Beweisidee,
- Beweisdurchführung und
- Aufräumen.

• *Die Beweisvorbereitungen* Längere Beweise beginnt man zweckmäßig damit, daß man die Ausgangssituation auf gewisse Standardfälle reduziert. Mit einem wirklichen Beweisansatz hat das oft noch nichts zu tun, aber in jedem Fall wird die spätere Beweisführung übersichtlicher.

Hier bietet sich zunächst die Normierung von a und $f(a)$ zu $0 \in \mathbb{R}^n$ an. Wenn der Satz für diesen speziellen Fall gilt, können wir ihn nämlich auf die auf der Menge $-a+X$ definierte Komposition $h\colon x \mapsto f(w+a) - f(a)$ anwenden. Der Satz liefert dann offene Mengen U und V um den Nullpunkt derart, daß h sich zu einem Diffeomorphismus $U \simeq V$ einschränkt. Diese Mengen definieren das linke, und die Umkehrung $V \xrightarrow{g} U$ dann auch das rechte der beiden kommutativen Diagramme

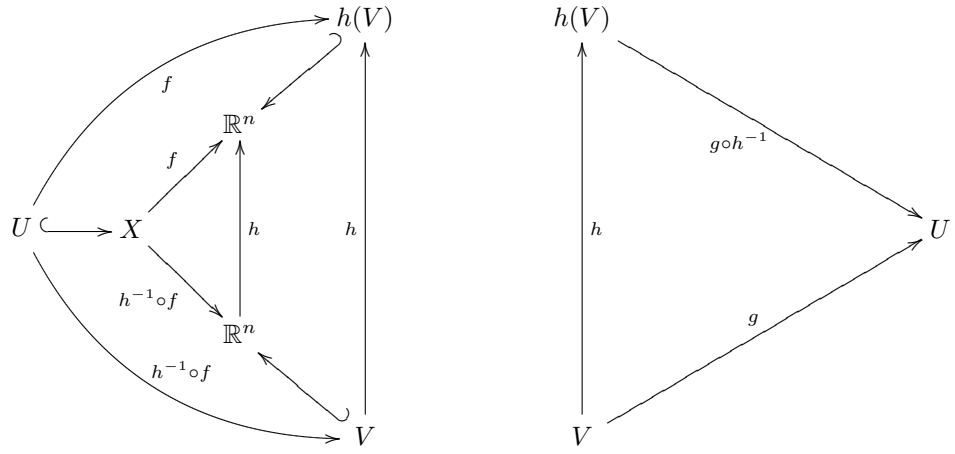


in denen alle vertikalen Pfeile Verschiebungen und damit Diffeomorphismen sind. Der obere unmarkierte Pfeil des rechten Diagaramms ist die gesuchte lokale Umkehrung von f.

Nur wenig raffinierter ist die weitere Reduktion auf den Fall, daß $h := Df(0) \in GL(n, \mathbb{R})$ die identische lineare Abbildung ist. In jedem Fall hat die Komposition $h^{-1} \circ f\colon X \to \mathbb{R}^n$ nach der Kettenregel diese Eigenschaft:

$$D(h^{-1} \circ f)(0) = Dh^{-1}(0) \circ Df(0) = h^{-1} \circ Df(0) = \mathrm{id}\,.$$

Indem wir den Satz für diesen Fall anwenden, erhalten wir eine lokale Umkehrung $V \xrightarrow{g} U$ für $h^{-1} \circ f$, und wie die Diagramme



zeigen, löst dann $g \circ h^{-1}$ das Problem.

Ab jetzt gelten also die vereinfachten Voraussetzungen $a = f(a) = 0$ und $Df(0) = 1$. Wir werden durchweg

$$f(x) = x + \varphi(x) \quad \text{für alle } x \in X$$

schreiben und wissen: die Abbildung $\varphi\colon X \to \mathbb{R}^n$ ist an der Stelle 0 in dem Sinne klein, daß neben $\varphi(0) = 0$ auch $D\varphi(0) = 0$ ist. Weil f stetig differenzierbar, die Abbildung $D\varphi\colon X \to \mathrm{Mat}(n\times n, \mathbb{R})$ also stetig ist, bleibt $D\varphi$ auch in der Nähe von 0 klein, genauer: zu jedem $\varepsilon > 0$ finden wir ein $\delta > 0$ mit

$$|D\varphi(x)| < \varepsilon \quad \text{für alle } x \in X \text{ mit } |x| \le \delta$$

(wobei gemäß der Auffassung $\mathrm{Mat}(n\times n, \mathbb{R}) = \mathbb{R}^{n^2}$ die links stehende Norm einer Matrix h explizit durch

$$|h|^2 = \sum_{i,j} h_{ij}^2 \in [0, \infty)$$

gegeben ist). Wir denken uns ein solches $\delta > 0$ gleich so klein, daß die abgeschlossene Kugel $D_\delta(0)$ ganz in X enthalten ist.

Wir werden die Kleinheit von $D\varphi$ in der Form zweier konkreter Abschätzungen anwenden.

Abschätzung 1 $|D\varphi(x)\cdot v| \le \sqrt{n}\,\varepsilon\cdot|v|$ für alle $x \in D_\delta(0)$ und alle $v \in \mathbb{R}^n$, und

Abschätzung 2 $|\varphi(u) - \varphi(v)| \le \sqrt{n}\,\varepsilon\cdot|u - v|$ für alle $u, v \in D_\delta(0)$.

Beweise Für jedes feste i zwischen 1 und n ist $D\varphi_i(x)$ eine Zeile, und wir können das Matrixprodukt $D\varphi_i(x)\cdot v$ ebensogut als das Skalarprodukt $\langle \mathrm{grad}\,\varphi_i(x), v\rangle$ schreiben. Die Schwarzsche Ungleichung 25.6 stiftet nun die Abschätzung

$$|D\varphi_i(x)\cdot v| = |\langle \mathrm{grad}\,\varphi_i(x), v\rangle| \le |D\varphi_i(x)|\cdot|v| \le \varepsilon\cdot|v|,$$

und durch Aufaddieren über alle i ergibt sich die Abschätzung 1.

Zum Beweis der zweiten Abschätzung seien $u, v \in D_\delta(0)$ gegeben. Wir kehren zu festem i zurück und wenden den Mittelwertsatz auf die Funktion

$$\begin{array}{ccc} [0,1] & \longrightarrow & \mathbb{R} \\ t & \mapsto & \varphi_i\big(tu + (1-t)v\big) \end{array}$$

an — weil die Kugel $D_\delta(0)$ konvex ist, ist φ längs der Verbindungstrecke zwischen u und v ja definiert. Der Mittelwertsatz garantiert

$$\varphi_i(u) - \varphi_i(v) = \frac{d}{dt}\varphi_i(tu + (1-t)v)\Big|_{t=\tau} = D\varphi_i\big(\tau u + (1-\tau)v\big)\cdot(u - v)$$

für eine geeigenete Zahl $\tau \in (0,1)$, insbesondere

$$|\varphi_i(u) - \varphi_i(v)| \le \varepsilon\cdot|u - v|.$$

Die Abschätzung 2 folgt daraus durch Aufaddieren.

Mit einem Blick auf die beiden Abschätzungen verfügen wir das bisher ja noch unbestimmte $\varepsilon > 0$ zu

$$\varepsilon = \frac{1}{2\sqrt{n}}$$

und fixieren ein dazu wie oben passendes δ. Im folgenden können wir uns also auf die Ungleichungen

$$|D\varphi(x)\cdot v| \le \frac{1}{2}\cdot|v| \quad \text{für alle } x \in D_\delta(0) \text{ und alle } v \in \mathbb{R}^n$$

und

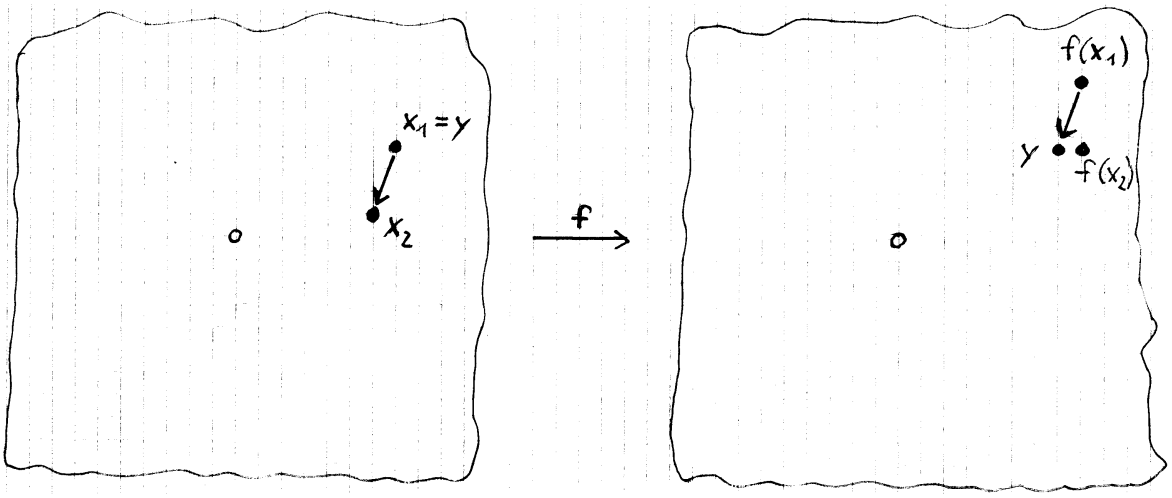
$$|\varphi(u) - \varphi(v)| \le \frac{1}{2}\cdot|u - v| \quad \text{für alle } u, v \in D_\delta(0)$$

als gültige Tatsachen stützen. Dabei hat die zweite eine ganz einfache anschauliche Bedeutung: φ bildet nicht nur $D_\delta(0)$ in sich ab (wie die Wahl $v = 0$ zeigt), sondern verkürzt dabei noch alle Abstände mindestens um den Faktor $\frac{1}{2}$. Solche Abbildungen nennt man *kontrahierend* (mit Faktor $\frac{1}{2}$).

• *Die Beweisidee* Im Kern besteht unsere Aufgabe darin, zu jedem dem Nullpunkt genügend nahe gelegenen $y \in \mathbb{R}^n$ ein Urbild aufzuspüren, also ein $x \in \mathbb{R}^n$ mit $f(x) = y$. Weil nun f immerhin *ungefähr* die identische Abbildung ist, ist als erster Versuch

$$x_1 := y$$

nicht abwegig: zwar werden wir damit y nicht genau treffen, aber doch ungefähr, dürfen erwarten, daß der Fehler $y - f(x_1)$ klein ausfällt. Er zeigt uns auch gleich an, wie wir x_1 korrigieren sollten, um die Näherung zu verbessern, nämlich zu



$$x_2 := x_1 + y - f(x_1).$$

Es ist also $x_2 - x_1 = y - f(x_1)$, und der neue Fehler

$$y - f(x_2) = \big(x_2 - x_1\big) - \big(f(x_2) - f(x_1)\big)$$

ist gerade die Diskrepanz zwischen den Differenzen $x_2 - x_1$ und $f(x_2) - f(x_1)$. Und jetzt kommt das Entscheidende: Weil f ungefähr identisch wirkt, wird diese Diskrepanz nicht nur absolut gesehen klein sein, sondern sogar klein gegen $x_2 - x_1 = y - f(x_1)$, also gegen den ersten Fehler! Es bestehen daher gute Aussichten, daß wir durch Wiederholen des Korrekturvorgangs zu einer Folge $(x_1, x_2, \ldots)$ gelangen, die nicht planlos umherirrt, sondern gegen ein wahres Urbild von y konvergiert.

• *Die Beweisdurchführung* Gesagt, getan. Was wir jetzt beweisen, ist:

Zu jedem $y \in U_{\delta/2}(0)$ gibt es genau ein $x \in U_\delta(0)$ mit $f(x) = y$.

Dazu sei $y \in U_{\delta/2}(0)$ gegeben. Wir bemerken zuerst, daß die durch die Vorschrift $F(u) := y - \varphi(u)$ definierte Abbildung $F\colon D_\delta(0) \to D_\delta(0)$ ebenso kontrahierend ist wie φ:

$$|F(u) - F(v)| = |-\varphi(u) + \varphi(v)| \le \frac{1}{2}|u - v|$$

und speziell sogar Werte in der *offenen* Kugel $U_\delta(0)$ annimmt:

$$|F(u)| \le |F(u) - F(0)| + |F(0)| \le \frac{1}{2}|u - 0| + |y| < \frac{1}{2}\delta + \frac{\delta}{2} = \delta.$$

Die Bedeutung der Abbildung F für unsere Fragestellung ergibt sich direkt aus den logischen Äquivalenzen

$$f(u) = y \quad\Longleftrightarrow\quad u + \varphi(u) = y \quad\Longleftrightarrow\quad F(u) = u$$

für $u \in D_\delta(0)$; uns interessieren also die in $U_{\delta/2}(0)$ gelegenenen *Fixpunkte* von F. Die durch

$$x_k := F^k(0) = \underbrace{(F \circ \cdots \circ F)}_{k-\text{mal}}(0) \quad \text{für jedes } k \in \mathbb{N}$$

definierte Folge in $D_\delta(0)$ ist gerade die in der Beweisidee beschriebene, denn wir haben

$$x_1 = F(0) = y \quad \text{und} \quad x_{k+1} = F(x_k) = y - \varphi(x_k) = x_k + y - f(x_k).$$

Weil für beliebige $k, l \in \mathbb{N}$ offenbar

$$|x_{k+l} - x_k| = |F^k(x_l) - F^k(0)| \le \frac{1}{2^k}\,|x_l| \le \frac{1}{2^k}\,\delta$$

gilt, handelt es sich um eine Cauchy-Folge und damit wie erhofft um eine konvergente Folge, und wegen der Abgeschlossenheit der Kugel $D_\delta(0)$ enthält diese den Grenzwert

$$x := \lim_{k\to\infty} x_k \in D_\delta(0).$$

Weil F stetig ist, muß x ein Fixpunkt von F sein:

$$F(x) = F(\lim_{k\to\infty} x_k) = \lim_{k\to\infty} F(x_k) = \lim_{k\to\infty} x_{k+1} = x,$$

und damit ist x eine Lösung unseres Problems, denn als Wert von F liegt x automatisch in der offenen Kugel $U_\delta(0)$.

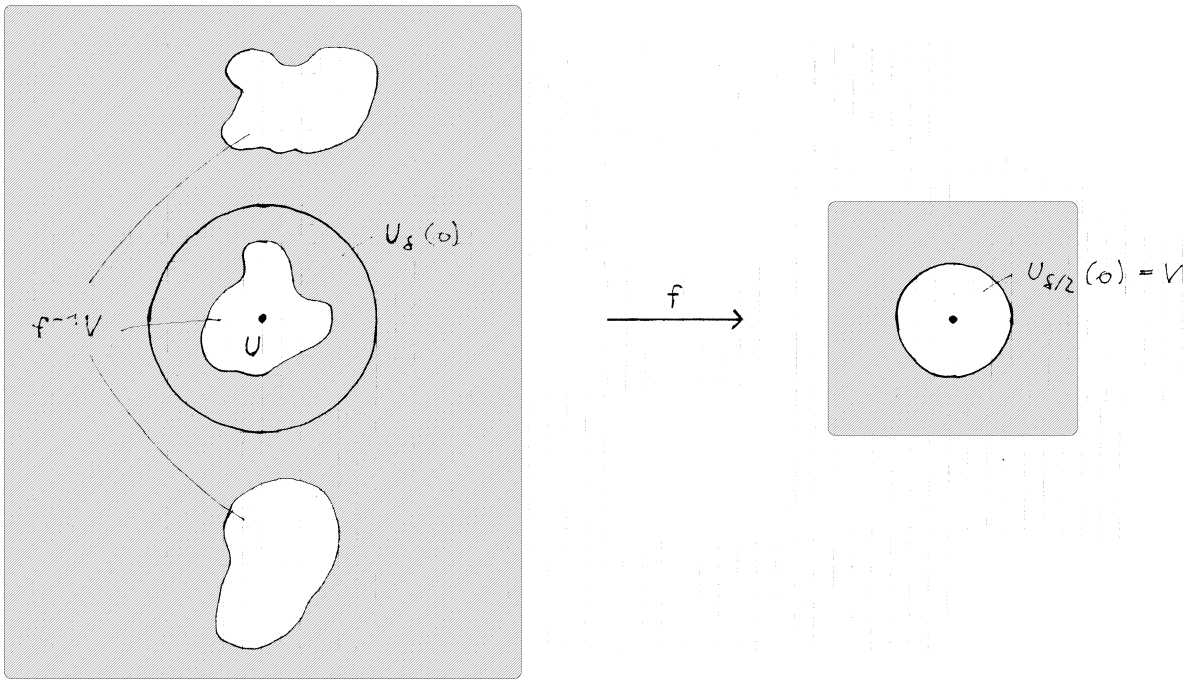
Andererseits ist x der einzige Fixpunkt von F, denn ist $x' \in D_\delta(0)$ ein weiterer Fixpunkt, so liefert die Kontraktionseigenschaft sofort

$$|x - x'| = |F(x) - F(x')| \le \frac{1}{2}\,|x - x'|$$

und damit $x = x'$.

Damit ist der Beweis unserer Behauptung geführt: für gegebenes $y \in U_{\delta/2}(0)$ hat die Gleichung $f(x) = y$ in $U_\delta(0)$ eine eindeutige Lösung x. Hier endet der Hauptteil des Beweises.

• *Das Aufräumen* Wir können jetzt die offenen Mengen U und V definieren, nämlich als



$$V = U_{\delta/2}(0) \subset \mathbb{R}^n$$

und

$$U = U_\delta(0) \cap f^{-1}V \subset X \subset \mathbb{R}^n.$$

Da f stetig ist, ist nicht nur V, sondern auch U offen. Eben haben wir gesehen, daß f als Abbildung von U nach V bijektiv ist; wir definieren $g\colon V \to U$ als ihre Umkehrung.

Es bleibt deren stetige Differenzierbarkeit zu beweisen. Beginnen wir mit der Stetigkeit: Weil φ kontrahierend ist, gilt für alle $x, x' \in D_\delta(0)$

$$|f(x) - f(x')| = |x + \varphi(x) - x' - \varphi(x')| \geq |x - x'| - |\varphi(x) - \varphi(x')| \geq \frac{1}{2}\,|x - x'|.$$

Speziell für $x, x' \in U$ können wir diese Abschätzung als

$$|y - y'| \geq \frac{1}{2}\,|g(y) - g(y')| \quad \text{für alle } y, y' \in V$$

lesen, woraus schon folgt, daß g stetig ist.

Sei nun $b \in V$ beliebig und $a = g(b)$. Nach der Abschätzung 1 gilt für jedes $v \in \mathbb{R}^n$

$$|Df(a)\cdot v| = |v + D\varphi(a)\cdot v| \geq |v| - |D\varphi(a)\cdot v| \geq \frac{1}{2}\cdot|v|,$$

so daß insbesondere $Df(a)$ injektiv ist: $Df(a) \in GL(n,\mathbb{R})$. Weil wir schon wissen, daß g bei b stetig ist, gilt $g(y) \to g(b) = a$ für $y \to b$, deshalb können wir die Differenzierbarkeit von f bei a, also die Aussage

$$f(x) - f(a) = Df(a)(x-a) + o(|x-a|) \quad \text{für } x \to a$$

zu

$$y - b = Df(a)\big(g(y) - g(b)\big) + o(|g(y) - g(b)|) \quad \text{für } y \to b$$

umschreiben. Wenn wir den Fehlerterm rechts $\psi(y)$ taufen, haben wir also

$$g(y) - g(b) = Df(a)^{-1}(y-b) - Df(a)^{-1}\psi(y) \quad \text{mit } \lim_{y\to b} \frac{1}{|g(y)-g(b)|}\,\psi(y) = 0.$$

Aus der Schreibweise

$$\frac{1}{|y-b|}\,Df(a)^{-1}\psi(y) = \underbrace{\frac{|g(y)-g(b)|}{|y-b|}}_{\le 2} \cdot Df(a)^{-1}\,\underbrace{\frac{1}{|g(y)-g(b)|}\,\psi(y)}_{\to 0}$$

lesen wir für den neuen Fehlerterm

$$-Df(a)^{-1}\psi(y) = o(|y-b|) \quad \text{für } y \to b$$

ab und haben damit gezeigt, daß g an der Stelle b differenzierbar ist und das Differential $Dg(b) = Df\big(g(b)\big)^{-1}$ hat — wie es nach Lemma 35.2 ja auch sein muß. Hier lesen wir aus der Formel für das Differential nur die letzte noch fehlende Aussage ab, daß nämlich $Dg(b)$ eine stetige Funktion von b ist. Dazu schreiben wir sie als Komposition

$$Dg\colon V \xrightarrow{g} U \xrightarrow{Df} GL(n,\mathbb{R}) \xrightarrow{h\mapsto h^{-1}} GL(n,\mathbb{R})$$

und berufen uns auf die Stetigkeit von $g\colon V \to U$ und $Df\colon U \to GL(n,\mathbb{R})$ sowie auf den zweiten Teil des folgenden auch allgemein interessanten Lemmas.

> $GL(n,\mathbb{R}) \subset \mathrm{Mat}(n\times n,\mathbb{R})$ ist eine offene Teilmenge, und das Invertieren $GL(n,\mathbb{R}) \ni h \mapsto h^{-1} \in GL(n,\mathbb{R})$ ist eine C^1-Abbildung.
>
> Aus der Charakterisierung $GL(n,\mathbb{R}) = \{h \in \mathrm{Mat}(n\times n,\mathbb{R}) \mid \det h \neq 0\}$ folgt nämlich die Offenheit, und die Adjunktenformel $h^{-1} = \frac{1}{\det h}\cdot\tilde{h}$ aus Satz 22.13 stellt die Inverse durch einen Rechenausdruck in den Koeffizienten von h dar.

Der Satz von der lokalen Umkehrung ist damit vollständig bewiesen.

43 Höhere Ableitungen in mehreren Variablen (Ergänzungen)

Statt vom k-ten Taylor-Polynom spricht man in der heutigen Zeit auch vom k-Jet einer C^k-Funktion oder einer C^k-Abbildung. Das Rechnen mit solchen Jets beruht neben der offensichtlichen Linearität

$$T_a^k(\lambda f + \mu g) = \lambda T_a^k f + \mu T_a^k g$$

vor allem auf der Produkt- und der Kettenregel, beide hier mit vollständigen Beweisen:

43.8 Kettenregel Die Abbildungen $\mathbb{R}^n \supset X \xrightarrow{f} Y \subset \mathbb{R}^p$ und $Y \xrightarrow{g} \mathbb{R}^q$ seien C^k-differenzierbar. Dann gilt

$$T_a^k(g \circ f) = T_a^k\left(T_{f(a)}^k g \circ T_a^k f\right)$$

für jedes $a \in X$.

43.9 Produktregel $X \xrightarrow{f} \mathbb{R}^p$ und $X \xrightarrow{g} \mathbb{R}^q$ seien C^k-Abbildungen, und $\mathbb{R}^p \times \mathbb{R}^q \ni (y,z) \longmapsto y * z \in \mathbb{R}^r$ sei ein bilineares Produkt. Für jedes $a \in X$ gilt dann

$$T_a^k(f*g) = T_a^k\left(T_a^k f * T_a^k g\right).$$

Zunächst ein erläuterndes

43.9$\frac{1}{2}$ Beispiel Die früher, wenn auch nur im Eindimensionalen besprochenen Potenzreihen sind die bequemste Quelle für Taylor-Polynome. Denn nach Satz 15.6 erhält man die erste Ableitung einer Potenzreihe

$$f(x) = \sum_{j=0}^{\infty} a_j (x-a)^j$$

im Konvergenzkreis durch gliedweises Differenzieren, woraus wie in Lemma 43.4 folgt, daß für jedes $k \in \mathbb{N}$ das k-te Taylor-Polynom

$$T_a^k f(x) = \sum_{j=0}^{k} a_j (x-a)^j$$

einfach durch Abschneiden der Reihe zu bilden ist. Also etwa

$$T_0^4 \exp(x) = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{24}x^4 \quad \text{und} \quad T_0^4 \sin(x) = x - \frac{1}{6}x^3$$

Die Produktregel verspricht uns für die Funktion $h\colon x \mapsto e^x \cdot (x - \sin x)$ also

$$\begin{aligned} T_0^4 h(x) &= T_0^4\left(T_0^4 \exp(x) \cdot T_0^4(x - \sin x)\right) \\ &= T_0^4\left(\left(1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{24}x^4\right) \cdot \left(\frac{1}{6}x^3\right)\right) \\ &= \frac{1}{6}x^3 + \frac{1}{6}x^4. \end{aligned}$$

Tatsächlich kann man daraus, daß der 2-Jet von $x - \sin x$ verschwindet, schon im voraus ablesen, daß man vom Exponentialfaktor gar nicht den vollen 4-Jet braucht, sondern schon mit dem 1-Jet auskommt.

Beweis der Produktregel Besonders gut induktiv umgehen kann man mit der Aussage $T_a^k f = 0$. Sie verspricht von f, daß alle partiellen Ableitungen bis zur Ordnung k bei a verschwinden. Für positives k sagt sie gleichwertig, daß jede erste partielle Ableitung $\frac{\partial f}{\partial x_i}$ bei a verschwindenden $(k-1)$-Jet hat.

Für $k = 0$ ist die Produktregel $T_a^k(f*g) = T_a^k\big(T_a^k f * T_a^k g\big)$ gewiß wahr, denn sie verspricht dann bloß $(f*g)(a) = f(a) * g(a)$.

Den Beweis für $k > 0$ führen wir erst unter der zusätzlichen Annahme $T_a^k f = 0$, und zwar durch Induktion nach k. Für jede Wahl von $i \in \{1, \dots, n\}$ gilt nach der altbekannten Produktregel — der für die erste Ableitung

$$\frac{\partial(f*g)}{\partial x_i} = \frac{\partial f}{\partial x_i} * g + f * \frac{\partial g}{\partial x_i}.$$

Wegen $T_a^{k-1}\frac{\partial f}{\partial x_i} = 0$ und sowieso $T_a^{k-1} f = 0$ folgt nach der Induktionsannahme $T_a^{k-1}\frac{\partial(f*g)}{\partial x_i} = 0$. Damit ist $T_a^k(f*g) = 0$, also die Produktregel für diesen speziellen Fall bestätigt.

Im allgemeinen Fall schreiben wir

$$f = T_a^k f + \varphi \text{ und } g = T_a^k g + \psi \quad \text{mit } T_a^k\varphi = 0 \text{ und } T_a^k\psi = 0$$

und lesen aus

$$f*g = (T_a^k f + \varphi)*(T_a^k g + \psi) = T_a^k f * T_a^k g + T_a^k f * \psi + \varphi * T_a^k g + \varphi * \psi$$

alles ab.

Die Produktregel überträgt sich in naheliegender Weise auf den Fall von mehr Faktoren.

Beweis der Kettenregel Durch Verschiebungen in $\mathbb{R}^n$ und $\mathbb{R}^p$ ziehen wir uns wie folgt auf den Fall $a = 0$ und $f(a) = 0$ zurück. Wir setzen $\tilde f(x) = f(x+a) - f(a)$ und $\tilde g(y) = f\big(y+f(a)\big)$. Dann gilt $\tilde f(0) = 0$ und $(\tilde g \circ \tilde f)(x) = (g \circ f)(x+a)$, und die Kettenregel für $\tilde f$ und $\tilde g$ impliziert die für f und g:

$$\begin{aligned} T_a^k(g \circ f)(X) &= T_0^k(\tilde g \circ \tilde f)(X-a) \\ &= T_0^k\big(T_0^k\tilde g \circ T_0^k\tilde f\big)(X-a) \\ &= T_a^k\big(T_0^k\tilde g \circ (T_a^k f - f(a))\big)(X) \\ &= T_a^k\big(T_{f(a)}^k g \circ T_a^k f\big)(X)\,. \end{aligned}$$

Damit ist die gewünschte Reduktion erreicht.

Für $k = 0$ ist die Formel wieder klar. Für $k > 0$ machen wir zunächst die Annahme $T_0^k g = 0$ und schließen aus der bekannten Ableitungsregel

$$\frac{\partial(g \circ f)}{\partial x_i} = (Dg \circ f) \cdot \frac{\partial f}{\partial x_i}$$

unter Benutzung der Produktregel induktiv $T_0^{k-1}\dfrac{\partial(g \circ f)}{\partial x_i} = 0$ für jedes i und damit $T_0^k(g \circ f) = 0$.

Im allgemeinen Fall schreiben wir $g = T_0^k g + \psi$ mit $T_0^k\psi = 0$, haben also

$$g \circ f = T_0^k g \circ f + \psi \circ f.$$

Nun ist das Polynom $T_0^k g(Y)$ eine Linearkombination von Monomen Y^j mit $|j| \le k$. Die Komposition $Y^j \circ f$ auszuführen bedeutet nun einfach, gewisse Komponenten von f miteinander zu multiplizieren, und nach der schon bewiesenen Produktregel folgt die erste Gleichheit in

$$T_0^k(Y^j \circ f) = T_0^k(Y^j \circ T_0^k f) = T_0^k(T_0^k Y^j \circ T_0^k f);$$

die zweite gilt einfach deswegen, weil das Polynom Y^j sich beim Anwenden von T_0^k nicht ändert. Wir fassen alle Monome wieder zusammen und erhalten (T_0^k ist linear)

$$T_0^k(T_0^k g \circ f) = T_0^k(T_0^k g \circ T_0^k f).$$

Wegen $T_0^k\psi = 0$ folgt nun

$$T_0^k(g\circ f) = T_0^k(T_0^k g\circ f) + T_0^k(\psi\circ f) = T_0^k(T_0^k g\circ T_0^k f)$$

wie behauptet.

Die bisherigen Überlegungen sagen noch wenig darüber, was der k-Jet einer Abbildung mit der Abbildung selbst zu tun hat. Als ersten Schritt in dieser Richtung übertragen wir Satz 16.11 aus dem Eindimensionalen.

43.11$\frac{1}{2}$ Satz Sei $X\subset\mathbb{R}^n$ offen, $f\colon X\to\mathbb{R}^p$ eine C^k-Funktion für ein $k\in\mathbb{N}$, und $a\in X$ ein Punkt. Das Polynom $T_a^k f$ approximiert f bei a im Sinne von

$$f(x) = T_a^k f(x) + o(|x-a|^k) \quad \text{für } x\to a.$$

Unter allen polynomialen Abbildungen $\mathbb{R}^n\to\mathbb{R}^p$ vom Grad höchstens k ist $T_a^k f$ die einzige mit dieser Eigenschaft.

Beweis Zur Vereinfachung dürfen wir die Normierungen $p=1$ und $a=0$ machen und f durch die Differenz $f-T_0^k f$ ersetzen: zu beweisen ist dann einerseits

$$f(x) = o(|x|^k) \quad \text{für } x\to 0$$

unter der Voraussetzung $T_0^k f = 0$ und andererseits, daß unter allen Polynomen f vom Grad höchstens k mit $f(x)=o(|x|^k)$ für $x\to 0$ das Nullpolynom das einzige ist.

Die erste Behauptung beweisen wir induktiv nach $k\in\mathbb{N}$. Für $k=0$ ist sie wahr, denn

$$f(x) = f(0) + o(1) \quad \text{für } x\to 0$$

stellt nur die Stetigkeit von f bei 0 fest. Sei also $k>0$. Das Differential $Df\colon X\to \mathrm{Mat}(1\times n,\mathbb{R})$ erfüllt $T_0^{k-1}(Df)=0$, und nach der Induktionsannahme folgt

$$Df(x) = o(|x|^{k-1}) \quad \text{für } x\to 0.$$

Zu vorgegebenem $\varepsilon>0$ finden wir konkret also ein $\delta>0$ mit $U_\delta(0)\subset X$ und

$$|Df(\xi)| \le |\xi|^{k-1}\cdot\varepsilon \quad \text{für alle } \xi\in D_\delta(0).$$

Sei $x\in U_\delta(0)$ beliebig. Wir wenden den Mittelwertsatz der Differentialrechnung auf die Hilfsfunktion

$$[0,1]\ni t\overset{\varphi}{\longmapsto} tx\in U_\delta(0)$$

an und finden ein $t\in(0,1)$ mit

$$f(x) = \varphi(1)-\varphi(0) = \varphi'(t)\cdot(1-0) = Df(tx)\cdot x.$$

Das letzte Produkt genügt als euklidisches Skalarprodukt $\langle \mathrm{grad} f(tx), x\rangle$ gelesen der schwarzschen Ungleichung, das liefert die Abschätzung

$$|f(x)| \le |Df(tx)|\cdot|x| \le |tx|^{k-1}\varepsilon\cdot|x| \le |x|^k\cdot\varepsilon,$$

aus der die Behauptung sofort folgt.

Zur Eindeutigkeitfrage nehmen wir an, $f\in\mathbb{R}[X]$ sei vom Nullpolynom verschieden und erfülle $f(x)=o(|x|^k)$ für $x\to 0$. Wir werden zeigen, daß dann $\deg f > k$ sein muß. Zum Beweis wählen wir ein $x\in\mathbb{R}^n$, das keine Nullstelle von f ist, und bilden das Polynom in einer Variablen

$$\varphi(T) := f(Tx)\in\mathbb{R}[T].$$

Auch φ ist nicht das Nullpolynom, etwa sei $\varphi(T) = \sum_{j=c}^{d} a_j T^j$ mit $a_c \neq 0 \neq a_d$. Wir wissen

$$t^c \cdot \sum_{j=c}^{d} a_j t^{j-c} = \varphi(t) = f(tx) = o(|tx|^k) = o(t^k) \quad \text{für } t \to 0,$$

und weil die Summe gegen $a_c \neq 0$ konvergiert, impliziert das $t^c = o(t^k)$ für $t \to 0$. Das ist offenbar gleichbedeutend mit $c > k$, damit schließen wir weiter

$$\deg f \geq \deg \varphi = d \geq c > k$$

und sind fertig.

Jetzt zur Herleitung des Lagrange-Restgliedes für den Satz von Taylor:

43.13 Restglied nach **Lagrange** Sei $X \subset \mathbb{R}^n$ offen, $f\colon X \longrightarrow \mathbb{R}$ eine C^{k+1}-Funktion und $a \in X$. Zu jedem Punkt $x \in X$, für den die Verbindungsstrecke von a nach x ganz in X liegt, gibt es dann ein $\tau \in [0,1]$ mit

$$f(x) - T_a^k f(x) = \sum_{|j|=k+1} \frac{D_j f\big((1-\tau)a + \tau x\big)}{j!} (x-a)^j.$$

Beweis Der Mittelwertsatz der Integralrechnung 31.8 verspricht für jede stetige Funktion $h\colon [0,1] \to \mathbb{R}$ ein $\tau \in [0,1]$ mit

$$\int_0^1 h(t)\, dt = h(\tau).$$

Angewendet auf die Funktion

$$[0,1] \ni t \longmapsto h(t) = \sum_{|j|=1} D_j f\big((1-t)a + tx\big)\,(x-a)^j$$

regelt das den Fall $k = 0$.

Auch für größere k läßt sich das Lagrange-Restglied als ein Mittelwert verstehen, aber einer, bei dem das Integral von h mit der Funktion

$$\varphi\colon [0,1] \ni t \longmapsto \varphi(t) = (k+1)(1-t)^k$$

so "gewichtet" wird, daß die Werte $h(t)$ für kleine $t \in [0,1|$ stärker zählen. Möglich wird diese Interpretation durch die Eigenschaft

$$\int_0^1 \varphi(t)\, dt = \int_0^1 (k+1)(1-t)^k\, dt = \big[-(1-t)^{k+1}\big]_{t=0}^{1} = 1,$$

die dafür sorgt, daß der gewichtete Mittelwert einer konstanten Funktion deren konstanter Wert wird. Die Zuordnung $h \mapsto \int_0^1 h\varphi$ hat damit die in Definition $15\frac{1}{2}.2$ genannten Eigenschaften Linearität, Positivität und Normiertheit, und die erlauben es, den Mittelwertsatz 31.8 (alias $15\frac{1}{2}.4$) auf eine Version mit Gewichtsfunktion zu verallgemeinern: diese verspricht ein $\tau \in [0,1]$ mit

$$\int_0^1 h(t)\varphi(t)\, dt = h(\tau)\,.$$

Sind nämlich c und d der kleinste und der größte Wert von h, so folgt aus der Funktionenungleichung $c \leq h \leq d$

$$c = \int_0^1 c\,\varphi \leq \int_0^1 h\,\varphi \leq d\,\varphi = d\,,$$

und nach dem Zwischenwertsatz muß $\int h\varphi$ ein Wert von h sein.

Die spezielle Wahl

$$h(t) = (k+1) \sum_{|j|=k+1} (1-t)^k \frac{D_j f\big((1-t)a + tx\big)}{j!} (x-a)^j$$

ergibt das Restglied von Lagrange für beliebiges $k \in \mathbb{N}$.

43.13$\frac{1}{2}$ Beispiele (1) Für $k = 0$ verspricht die Taylor-Formel mit dem Lagrange-Restglied ein $\tau \in [0,1]$ mit

$$f(x) - f(a) = \sum_{j=1}^{n} D_j f\big((1-\tau)a + \tau x\big)\,(x_j - a_j) = Df\big((1-\tau)a + \tau x\big)\,(x-a),$$

das ist im großen und ganzen der Mittelwertsatz der Differentialrechnung 14.1.

(2) Die Wahl $k=9$, $f=\cos$ und $a=0$ liefert ein $\tau \in [0,1]$ mit

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \frac{\cos \tau x}{10!}\, x^{10}$$

und insbesondere die Abschätzung

$$\left|\cos x - \left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!}\right)\right| \le \frac{1}{10!}\,|x|^{10}$$

für jedes $x \in \mathbb{R}$ — was zum Beispiel zeigt, daß $\cos x$ für $|x| \le \frac{\pi}{4}$ durch das achte Taylor-Polynom mit einem Fehler von höchstens

$$\frac{(\pi/4)^{10}}{10!} < \frac{1}{10!} < 3 \cdot 10^{-7}$$

approximiert wird.

44 E Lokale Extrema

Die folgende Darstellung ist als Alternative zum Abschnitt 44 des Skriptes gedacht; sie verzichtet auf das aus theoretischer Sicht sehr interessante Morse-Lemma und befaßt sich von vornherein nur mit einem Teilaspekt, nämlich der Untersuchung von Funktionen auf lokale Extrema — das aber detaillierter, insbesondere was die rechnerische Durchführung betrifft.

Die Fragestellung ist Ihnen aus dem Eindimensionalen vertraut: Gegeben eine Funktion $f: X \to \mathbb{R}^n$ auf irgendeiner Menge X, interessiert man sich für die (eventuell vorhandenen) Stellen, an denen f den kleinsten oder größten Wert, also das globale Minimum beziehungsweise Maximum annimmt. Wenn X eine Teilmenge von $\mathbb{R}^n$ ist, können wir allgemeiner nach den Stellen fragen, an denen f immerhin ein *lokales* Extremum hat. Beachten Sie, daß die Fragestellung von vornherein nur für Funktionen f (mit skalaren Werten) Sinn hat, andererseits nichts mit Stetigkeits- oder Differenzierbarkeitseigenschaften von f zu tun hat.

Der erste Beitrag der eindimensionalen Analysis ist die als Satz 14.8(a) verbuchte Tatsache, daß jede Stelle, an der eine in einem offenen Intervall differenzierbare Funktion ein lokales Extremum annimmt, ein kritischer Punkt von f sein muß. Das überträgt sich direkt auf die mehrdimensionale Situation, und das sogar ohne daß man dazu die mehrdimensionale Differentialrechnung überhaupt braucht. Das liegt an einem ganz elementaren Sachverhalt, den ich noch vorher formulieren möchte:

44E.1 Notiz Sei $\varphi: T \to X$ eine Abbildung, und sei $f: X \to \mathbb{R}$ eine Funktion. Dann gilt für jedes $t \in T$:

- Hat f bei $\varphi(t)$ ein globales Extremum, so hat $f \circ \varphi$ bei t ein globales Extremum (derselben Art natürlich);
- ist $\varphi: T \to X$ eine bei t stetige Abbildung zwischen Teilmengen von $\mathbb{R}^m$ beziehungsweise $\mathbb{R}^n$ und hat f bei $\varphi(t)$ ein lokales Extremum, so hat $f \circ \varphi$ bei t ein lokales Extremum.

44E.2 Lemma Sei X eine offene Teilmenge von $\mathbb{R}^n$ oder allgemeiner eine n-dimensionale Mannigfaltigkeit. Hat die differenzierbare Funktion $f: X \to \mathbb{R}$ an der Stelle a ein lokales Extremum, so ist a eine kritische Stelle von f:

$$Df(a) = 0 \quad \text{beziehungsweise} \quad T_a f = 0$$

Beweis Sei $X \subset \mathbb{R}^n$ offen. Für jeden Index $j \in \{1, \dots, n\}$ ist die Funktion $\varphi: t \mapsto f(a + te_j)$ in einem offenen Intervall um $0 \in \mathbb{R}$ definiert und differenzierbar. Sie hat nach 44E.1 bei 0 ein lokales Extremum, so daß nach 14.8(a) ihre Ableitung dort verschwindet. Das heißt aber

$$\frac{\partial f}{\partial x_j}(a) = \varphi'(0) = 0,$$

und damit folgt überhaupt $Df(a) = 0$. Man überträgt das auf den Fall einer n-dimensionalen Mannigfaltigkeit X, indem man eine Karte $X \supset U \xrightarrow{h} h(U) \subset \mathbb{R}^n$ um a und den zugehörigen Isomorphismus $T_a X \xrightarrow{\simeq} \mathbb{R}^n$ benutzt.

Ganz anders steht es um hinreichende Bedingungen dafür, daß f an einer kritischen Stelle a tatsächlich ein lokales Extremum hat: hier kommt man nicht ohne die Konzepte der mehrdimensionalen Analysis aus.

44E.3 Satz Sei X eine offene Teilmenge von $\mathbb{R}^n$ oder allgemeiner eine n-dimensionale Mannigfaltigkeit in $\mathbb{R}^N$. Es sei $a \in X$ ein kritischer Punkt der C^2-Funktion $f: X \to \mathbb{R}$. Dann gilt:

- Wenn es einen Vektor $v \in T_a X$ mit $H_a f(v) < 0$ gibt, dann hat f an der Stelle a kein lokales Minimum.
- Wenn $H_a f$ positiv definit ist, dann hat f an der Stelle a ein strenges lokales Minimum.

Natürlich kann man für in $\mathbb{R}^n$ offenes X auch $v \in \mathbb{R}^n$ und $Hf(a)$ statt $H_a f$ lesen.

Beweis Es ist klar, daß wir uns durch Wahl einer Untermannigfaltigkeitskarte um a auf den erstgenannten Fall zurückziehen können und $X \subset \mathbb{R}^n$ sogar als offene Kugel um a voraussetzen dürfen. Durch die einfachen Normierungen von a zu $0 \in \mathbb{R}^n$ und von $f(a)$ zu $0 \in \mathbb{R}$ machen wir uns die Lage noch übersichtlicher.

Wir wenden die Taylor-Formel aus Satz 43.12 mit $k=1$ an. Weil 0 ein kritischer Punkt von f ist, bleibt auf der rechten Seite nur das Restglied:

$$f(x) = 2\sum_{|j|=2} \int_0^1 (1-t)\frac{D_j f(tx)}{j!}\,dt \cdot x^j = \frac{1}{2}\sum_{r,s=1}^{n} H_{rs}(x)\,x_r x_s = \frac{1}{2}x^{\mathrm{t}} H(x)\,x \quad \text{für alle } x \in X,$$

worin wir die stetige Abbildung $H\colon X \longrightarrow \mathrm{Sym}(n,\mathbb{R})$ durch

$$H_{rs}(x) := 2\int_0^1 (1-t) D_r D_s f(tx)\,dt = 2\int_0^1 (1-t)\frac{\partial^2 f}{\partial x_r \partial x_s}(tx)\,dt$$

erklärt haben. Wegen $2\int_0^1 (1-t)dt = 1$ ist $H(0) = Hf(0)$ die Hesse-Matrix, und wir schreiben

$$f(x) = \frac{1}{2}x^{\mathrm{t}} \cdot Hf(0) \cdot x + \frac{1}{2}x^{\mathrm{t}}\big(H(x) - H(0)\big)\,x$$

mit $\lim\limits_{x\to 0}\big(H(x)-H(0)\big) = 0$. Für jedes gegebene $\varepsilon > 0$ können wir die Kugel X um a also so verkleinern, daß $|H(x)-H(0)| < \varepsilon/\sqrt{n}$ und damit nach der bekannten Abschätzung

$$\big|x^{\mathrm{t}}\big(H(x)-H(0)\big)\,x\big| < \varepsilon \cdot |x|^2 \quad \text{für alle } x \in X$$

wird (bei genauerem Hinschauen wie in Aufgabe 34.3 sieht man, daß der Faktor $1/\sqrt{n}$ unnötig ist).

Unter der Voraussetzung der ersten Satzaussage wählen wir einen Vektor $v \in \mathbb{R}^n$ mit $v^{\mathrm{t}} Hf(0)\,v < 0$ und $|v| = 1$; wir setzen $\varepsilon = -v^{\mathrm{t}} Hf(0)\,v$ und haben für genügend kleine $t > 0$

$$f(tv) = \frac{1}{2}v^{\mathrm{t}} Hf(0)\,v \cdot t^2 + \frac{1}{2}v^{\mathrm{t}}\big(H(x)-H(0)\big)\,v \cdot t^2 < -\frac{1}{2}\varepsilon \cdot t^2 + \frac{1}{2}\varepsilon \cdot t^2 = 0.$$

Also hat f bei 0 kein lokales Minimum.

Ist $Hf(0)$ dagegen positiv definit, so setzen wir $\varepsilon = \min\big\{v^{\mathrm{t}} \cdot Hf(0) \cdot v \,\big|\, v \in S^{n-1}\big\} > 0$, verkleinern X entsprechend und haben

$$f(x) = \frac{1}{2}x^{\mathrm{t}} \cdot Hf(0) \cdot x + \frac{1}{2}x^{\mathrm{t}}\big(H(x)-H(0)\big)\,x > \frac{1}{2}\varepsilon \cdot |x|^2 - \frac{1}{2}\varepsilon \cdot |x|^2 = 0$$

für alle $x \in X \setminus \{0\}$. Damit hat f im Nullpunkt ein strenges lokales Minimum.

44E.4 Beispiele (1) Die einfachsten und zugleich lehrreichsten Beispiele sind die quadratischen Formen $q\colon \mathbb{R}^n \to \mathbb{R}$ als Funktion: selbstverständlich ist (für $n > 0$) der Nullpunkt ein kritischer Punkt und $Hq = 2q$ im wesentlichen die Funktion selbst. Wie wir aus Abschnitt 29 wissen, dürfen wir q als diagonal annehmen:

$$q(x) = \lambda_1 x_1^2 + \cdots + \lambda_n x_n^2 \quad \text{für alle } x \in \mathbb{R}^n.$$

Auch ohne Satz 44E.3 ist evident, daß q im Nullpunkt sicher kein lokales Minimum hat, wenn einer der Eigenwerte λ_j negativ ist, und daß q im Falle der positiven Definitheit — also genau wenn alle Eigenwerte positiv sind — im Nullpunkt das echte sogar globale Minimum 0 annimmt. Was Satz 44E.3 eigentlich sagt: eine im Nullpunkt kritische C^2-Funktion verhält sich in den Richtungen, in denen die Hesse-Form nicht null ist, ähnlich wie diese.

Die quadratischen Formen als Beispiele zeigen auch, daß es weitere Arten von kritischen Punkten gibt, bei denen *kein* lokales Extremum vorliegt. Das ist ja ersichtlich der Fall, sobald ein Eigenwert positiv, ein anderer negativ ist (man nennt die Form dann *indefinit*), und Satz 44E.3 überträgt diesen Sachverhalt auf

kritische Punkte beliebiger C^2-Funktionen. Die Anschauung im kleinstmöglichen Fall $n = 2$ legt nahe, von *Sattelpunkten* der Funktion zu sprechen. Man kann die Vorstellung präzisieren, daß solche Punkte stabil in dem Sinne sind, daß sie nicht durch eine beliebig kleine Störung der Funktion beseitigt werden können. Im Eindimensionalen gibt es dieses Phänomen nicht: kritische Punkte, bei denen kein Extremum vorliegt, kommen nur als gewissermaßen entartete Fälle zwischen den Minima und den Maxima vor.

(2) Es liegt auf der Hand, wie man Satz 44E.3 explizit anwendet. Etwa für die Funktion

$$\mathbb{R}^2 \ni (x,y) \longmapsto f(x,y) = x\,y\,(x+y-1) \in \mathbb{R}$$

sucht man die Nullstellen des Differentials

$$Df(x,y) = \begin{pmatrix} y(x+y-1)+xy & x(x+y-1)+xy \end{pmatrix}$$

und findet die vier kritischen Punkte

$$(0,0),\ (1,0),\ (0,1) \text{ und } (\tfrac{1}{3},\tfrac{1}{3}).$$

Deren Hesse-Matrizen ergeben sich aus den zweiten partiellen Ableitungen

$$\frac{\partial^2 f}{\partial x^2}(x,y) = 2y, \quad \frac{\partial^2 f}{\partial x \partial y}(x,y) = 2x + 2y - 1 \quad \text{und} \quad \frac{\partial^2 f}{\partial y^2}(x,y) = 2x$$

zu

$$Hf(0,0) = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, \quad Hf(1,0) = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \quad Hf(0,1) = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{und} \quad Hf(\tfrac{1}{3},\tfrac{1}{3}) = \frac{1}{3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

Man sieht sofort, daß die ersten drei Matrizen indefinit, die letzte dagegen positiv definit ist: nur bei $(\frac{1}{3},\frac{1}{3})$ liegt also ein lokales Extremum, und zwar ein Minimum.

Bei größeren Dimensionen prüft man die Definitheitseigenschaften der Hesse-Form mit Standardmethoden der linearen Algebra, etwa nach Aufgabe 29.2.

Spannender ist die Suche nach den lokalen Extrema einer auf einer Untermannigfaltigkeit definierten Funktion. Vorweg eine ganz plumpe Methode: wenn man das Glück hat, eine Untermannigfaltigkeit $X \subset \mathbb{R}^N$ gut "parametrisieren", das heißt als Bildmenge einer einfachen stetigen Abbildung $\varphi\colon T \to \mathbb{R}^N$ schreiben zu können, dann mag einem mit der Notiz 44E.1 schon geholfen sein. Musterbeispiel dafür ist die Kreislinie $S^1 \subset \mathbb{R}^2$ mit der Parametrisierung

$$\mathbb{R} \ni t \mapsto \varphi(t) = (\cos t, \sin t) \in \mathbb{R}^2.$$

Ist $\varphi(t) \in S^1$ lokale oder globale Extremstelle der Funktion $f\colon S^1 \to \mathbb{R}$, so muß t eine ebensolche der Komposition $f\circ\varphi\colon \mathbb{R} \to \mathbb{R}$ sein. Im konkreten Beispiel der Kreislinie liegen die Verhältnisse besonders günstig, weil die Parametrisierung kein zusätzlichen kritischen Punkte und damit keine zusätzlichen Kandidaten für Extremstellen erzeugt — im allgemeinen hängt das natürlich von den Qualitäten der Parametrisierung φ ab. Jedenfalls hilft die Überlegung oft dabei, die Suche auf eine überschaubare Menge von Kandidaten einzugrenzen.

Wenn man keine passende Parametrisierung zur Hand hat, kann man häufig auf die als Satz 37.7 formulierte Methode der Lagrange-Multiplikatoren zurückgreifen. Dort ist $X \subset \mathbb{R}^{n+p}$ die Faser einer auf einer offenen Teilmenge von $\mathbb{R}^{n+p}$ definierten C^2-Abbildung F über einem regulären Wert. Auch die (skalare) C^2-Funktion f ist auf der X enthaltenden offenen Teilmenge von $\mathbb{R}^{n+p}$ gegeben — gesucht sind aber die lokalen Extrema der Einschränkung $f|X$. Gemacht wird's dann so:

44E.5 Verfahren Nach Lemma 44E.2 können lokale Extrema nur an den kritischen Stellen von $f|X$ vorliegen, zuerst bestimmt man also diese nach der Methode von Satz 37.7 als die kritischen Punkte der Hilfsfunktion $f - \sum_{j=1}^{p} \lambda_j F_j$. Wir hoffen, daß wir das entstehende Gleichungssystem lösen können und so eine überschaubare (zum Beispiel endliche) Menge von Kandidaten erhalten.

Sei $a \in X$ so ein Kandidat, das heißt ein kritischer Punkt von $f|X$. Wir ersetzen jetzt die ursprüngliche Funktion f durch $f - \sum_{j=1}^{p} \lambda_j F_j$ mit den zu a gehörigen (eindeutig bestimmten) Werten von $\lambda_1, \dots, \lambda_p$ — auf X bedeutet das nur, eine Konstante zu addieren — und machen damit a zu einem kritischen Punkt von f selbst (und nicht nur der Einschränkung auf X). Die Berechnung der jetzt benötigten Hesse-Form $H_a(f|X)$ leistet dann das folgende

44E.5$\frac{1}{3}$ Lemma Sei $W \subset \mathbb{R}^N$ offen, $X \subset W$ eine Untermannigfaltigkeit, $f\colon W \to \mathbb{R}$ eine C^2-Funktion mit kritischem Punkt $a \in X$. Dann ist die Einschränkung der Hesse-Form von f

$$T_aX \times T_aX \ni (v, w) \longmapsto H_a f(v, w) \in \mathbb{R}$$

die Hesseform $H_a(f|X)$ der Einschränkung von f.

Beweis Sei $H\colon U \to H(U) \subset \mathbb{R}^N = \mathbb{R}^n \times \mathbb{R}^p$ eine Untermannigfaltigkeitskarte um a, also

$$H(U \cap X) = H(U) \cap \big(\mathbb{R}^n \times \{0\}\big).$$

Der einfacheren Schreibweise halber ersetzen wir W durch U und X durch $U \cap X$, und wie üblich bezeichnen wir die aus H durch Einschränken entstehende Karte für X selbst mit $h\colon X \to h(X) \subset \mathbb{R}^n = \mathbb{R}^n \times \{0\}$. Wir rechnen

$$\begin{aligned} H_a(f|X) &= H_0\big(f|X \circ h^{-1}\big) \circ (T_a h \times T_a h) \\ &= H_0\big(f \circ H^{-1}|H(X)\big) \circ (T_a h \times T_a h) \\ &= H_0(f \circ H^{-1})|(\mathbb{R}^n \times \mathbb{R}^n) \circ (T_a h \times T_a h) \\ &= H_a f \circ \big(DH^{-1}(0) \times DH^{-1}(0)\big) \circ \big(DH(a) \times DH(a)\big)|(T_aX \times T_aX) \\ &= H_a f|(T_aX \times T_aX); \end{aligned}$$

die erste Gleichheit ergibt sich aus der Definition von $H_a(f|X)$ und die zweite daraus, daß H eine Untermannigfaltigkeitskarte ist; die dritte und die letzte sind klar, und die vierte folgt nach Lemma 43, weil a ein kritischer Punkt von f ist.

44E.5 Fortsetzung Nach Satz 37.5 ist $T_aX = \operatorname{Kern} DF(a)$, und das Standardverfahren 20.10 der linearen Algebra berechnet aus $DF(0)$ eine Parametrisierung von $\operatorname{Kern} DF(a)$, das heißt eine injektive lineare Abbildung $b\colon \mathbb{R}^n \to \mathbb{R}^{n+p}$ mit $\operatorname{Bild} b = T_aX$. Damit stellt die Matrix

$$b^{\mathrm{t}} Hf(a)\, b \in \operatorname{Sym}(n, \mathbb{R})$$

die Hesse-Form $T_a(f|X)$ in derjenigen linearen Karte $l\colon T_aX \to \mathbb{R}^n$ dar, mit der $l \circ b = \mathrm{id}$ gilt, und aus ihr lesen wir die relevanten Informationen wie immer ab.

44E.6 Beispiele (1) Wir greifen das Beispiel 44.5 der quadratischen Form

$$S^{n-1} \ni x \mapsto q(x) = \sum_{j=1}^{n} \lambda_j x_j^2 \quad \text{mit paarweise verschiedenen } \lambda_j$$

noch einmal auf. Wir wissen schon, daß $\pm e_k$ für jedes k eine kritische Stelle von $q|S^{n-1}$ ist, haben jetzt aber den Ehrgeiz, die Hesse-Form von $q|S^{n-1}$ dort ohne Verwendung einer Karte zu berechnen. Dazu ersetzen wir q durch die Funktion $q - \lambda|?|^2$ mit dem für $\pm e_k$ zuständigen Lagrange-Multiplikator $\lambda = \lambda_k$, also

$$\mathbb{R}^n \ni x \longmapsto \sum_{j=1}^{n} \lambda_j x_j^2 - \lambda_k \sum_{j=1}^{n} x_j^2 = \sum_{j=1}^{n} (\lambda_j - \lambda_k) x_j^2 \in \mathbb{R}\,.$$

Die diagonale Hesse-Matrix mit Einträgen $2(\lambda_j - \lambda_k)$ ist evident, Einschränken auf den Tangentialraum $T_{\pm e_k} S^{n-1} = \big\{x \in \mathbb{R}^n \,\big|\, x_k = 0\big\}$ löscht gerade den Nullterm, und wir sehen erneut, daß die beiden Stellen Morse-Punkte sind und ihr Index die Anzahl der Eigenwerte λ_j unterhalb von λ_k ist.

(2) Wir wählen als Mannigfaltigkeit

$$X = \{(x,y) \in \mathbb{R}^2 \mid y^2 = 2x - x^3\}$$

und wollen nach lokalen Extrema der auf X eingeschränkten Funktion $f\colon (x,y) \mapsto (x+1)y$ suchen. Es bietet sich an, $F\colon \mathbb{R}^2 \to \mathbb{R}$ durch $F(x,y) = y^2 + x^3 - 2x$ zu erklären: dann ist $X = F^{-1}\{0\}$, und aus dem Differential

$$DF(x,y) = \begin{pmatrix} 3x^2-2 & 2y \end{pmatrix}$$

liest man sofort ab, daß 0 in der Tat ein regulärer Wert von F und damit $X \subset \mathbb{R}^2$ eine eindimensionale Untermannigfaltigkeit oder "glatte ebene Kurve" ist, wie man gern sagt. Die Funktion f ist formelmäßig auf ganz $\mathbb{R}^2$ gegeben und hat das Differential $Df(x,y) = \begin{pmatrix} y & x+1 \end{pmatrix}$. Wir setzen also mit einem Lagrange-Multiplikator λ

$$\begin{pmatrix} y & x+1 \end{pmatrix} = \lambda \cdot \begin{pmatrix} 3x^2-2 & 2y \end{pmatrix}$$

an. Die lineare Abhängigkeit der beiden 1×2-Matrizen bedeutet

$$2y^2 = (x+1)(3x^2-2)$$

und führt zusammen mit der Gleichung $F(x,y) = 0$ auf die kubische Gleichung

$$5x^3 + 3x^2 - 6x - 2 = 0$$

mit einer erkennbaren ersten Lösung $x = 1$; sie führt zu den beiden kritischen Punkten $(1, \pm 1)$ von $f|X$, mit Werten ± 1 für λ. Die beiden anderen Nullstellen des kubischen Polynoms liegen im Intervall $(-\sqrt{2}, 0)$, und für sie wird die Gleichung $F(x,y) = 0$ für kein reelles y erfüllt. Damit bleiben die kritischen Stellen $(1, \pm 1)$ die einzigen.

Zur Berechnung der Hesse-Form ist f durch $f - \lambda F$ zu ersetzen, wir haben ab jetzt also

$$f(x,y) = (x+1)y \pm (-y^2 - x^3 + 2x)$$

mit der Hesse-Matrix

$$Hf(1,\pm 1) = \begin{pmatrix} \mp 6x & 1 \\ 1 & \mp 2 \end{pmatrix}\Big|_{x=1} = \begin{pmatrix} \mp 6 & 1 \\ 1 & \mp 2 \end{pmatrix}.$$

Einzuschränken ist auf den Tangentialraum

$$T_{(1,\pm 1)}X = \operatorname{Kern} DF(1,\pm 1) = \operatorname{Kern} \begin{pmatrix} 1 & \pm 2 \end{pmatrix} = \operatorname{Bild} \begin{pmatrix} \mp 2 \\ 1 \end{pmatrix};$$

die Hesse-"Matrix" der Einschränkung ergibt sich zu

$$\begin{pmatrix} \mp 2 & 1 \end{pmatrix} \begin{pmatrix} \mp 6 & 1 \\ 1 & \mp 2 \end{pmatrix} \begin{pmatrix} \mp 2 \\ 1 \end{pmatrix} = \pm 17.$$

Die Einschränkung $f|X$ hat deshalb an der Stelle $(1,1)$ ein strenges lokales Maximum und bei $(1,-1)$ ein strenges lokales Minimum (Morse-Extrema im Sinne von Abschnitt 44).

Übungsaufgabe

44E.1 Sei $X \subset \mathbb{R}^n$ offen, $a \in X$ und $f\colon X \to \mathbb{R}^n$ ein lokaler C^l-Diffeomorphismus bei a. Zeigen Sie, wie sich der k-Jet $T^k_{f(a)} f^{-1}$ aus $T^k_a f$ für alle $k = 1, \dots, l$ sukzessiv — das heißt unter Verwendung von $T^{k-1}_{f(a)} f^{-1}$ — berechnen läßt, indem man eine der Jet-Identitäten

$$T^k_a\big(T^k_{f(a)} f^{-1} \circ T^k_a f\big) = \mathrm{id} \quad \text{oder} \quad T^k_{f(a)}\big(T^k_a f \circ T^k_{f(a)} f^{-1}\big) = \mathrm{id}$$

auflöst.