Einführung in das mathematische Arbeiten

Michael Dreher, Universität Konstanz, FB Mathematik und Statistik

22.02.2013

Kapitel 0

Vor–Einführung

Dieses Skript ist eine Zusammenstellung der im Tutorium behandelten Aspekte. Von Herzem empfohlen sei die Lektüre des Buches [1] von Kevin Houston.

0.1 Mengen

Definition 0.1. *Eine* Menge *ist eine Sammlung von unterschiedlichen Objekten. Jedes dieser Objekte heißt* Element der Menge*.*

Wichtig ist: kein Element ist zweimal oder öfter in derselben Menge enthalten (deshalb steht dort das Wort „unterschiedlich").
Eine erste Schreibweise: Aufzählung in geschweiften Klammern. Beispiele dafür sind

$$M = \{1, 7, \text{Kirschkuchen}\}, \qquad P = \{3, 4, \text{Kirschkuchen}, \{\text{Hund}, \text{Katze}\}\}.$$

Die Menge M hat genau 3 Elemente, die Menge P hat genau 4 Elemente.

Definition 0.2. *Wenn x ein Element einer Menge X ist, dann schreiben wir $x \in X$. Ansonsten schreiben wir $x \notin X$.*

Definition 0.3. *Die Menge ohne Elemente heißt* leere Menge*, geschrieben als $\emptyset$.*

Stellen Sie sich ein Museum vor, daß alle seine Exponate an befreundete Museen ausgeliehen hat. Dann hat dieses Museum eine leere Sammlung.

Definition 0.4. *Zwei Mengen X und Y heißen* gleich*, wenn jedes Element der einen Menge in der anderen enthalten ist, und umgekehrt. Dann schreiben wir $X = Y$.*

Definition 0.5. *Sei X eine Menge. Eine Menge Y heißt* Teilmenge von X*, wenn jedes Element von Y auch in X enthalten ist. Wir schreiben dies als $Y \subseteq X$.*

Es gibt auch die Schreibweise $Y \subset X$ für diesen Sachverhalt.

Satz 0.6. *Für beliebige Mengen X und Y gilt immer: $X = Y$ genau dann, wenn $X \subseteq Y$ und $Y \subseteq X$.*

Auf einen Beweis verzichten wir.
Man beachte: es ist $x \neq \{x\}$. Denn links steht das Objekt x, und rechts steht diejenige Menge, die genau das Objekt x enthält. Wir kümmern uns übrigens nicht um die Frage, welchen „Typ" das Objekt x hat (es könnte eine reelle Zahl sein oder eine Funktion oder ein Hufeisen oder der Jupiter).
Wenn $x \in X$, dann ist $\{x\} \subseteq X$, aber meistens ist $\{x\} \notin X$. Beachte allerdings, daß $X = \{x, \{x\}\}$ logisch zulässig ist. Diese Menge X hat genau zwei Elemente, und X hat genau vier Teilmengen: nämlich $\emptyset$, $\{x\}$, $\{\{x\}\}$ und $\{x, \{x\}\}$.

Eine zweite Mengenschreibweise ergibt sich durch Rückgriff auf eine zuvor definierte Menge (auch „Obermenge“ genannt), die jetzt eingeschränkt wird.
Man kann die Menge G der geraden Zahlen definieren durch Aufzählung

$$G = \{\dots, -6, -4, -2, 0, 2, 4, 6, 8, \dots\}$$

oder durch Einschränkung einer zuvor definierten Menge

$$G = \{x \in \mathbb{Z} \colon 2 \text{ ist ein Teiler von } x\}.$$

Hierbei steht vor dem : der Bezug auf die Obermenge $\mathbb{Z}$, und hinter dem : steht die Einschränkung der Obermenge.
Offenkundig ist $G \subseteq \mathbb{Z}$. Eine andere Schreibweise tauscht den : durch ein | aus:

$$G = \{x \in \mathbb{Z} \mid 2 \text{ ist ein Teiler von } x\}.$$

Das ist eine Geschmacksfrage. Man kann auch folgendes schreiben:

$$G = \{x \mid x \in \mathbb{Z} \text{ und } 2 \text{ ist ein Teiler von } x\}.$$

Man beachte, daß jetzt die Spezifikation der Obermenge hinter dem Trennstrich | steht.

Definition 0.7. *Seien x und y Objekte. Dann ist (x, y) das* geordnete Paar mit den beiden Komponenten x und y.

Als Beispiel können wir das übliche kartesische Koordinatensystem in der Ebene betrachten, aus dem wir die Erkenntnis ziehen, daß $(1, 3)$ und $(3, 1)$ verschiedene Punkte sind. Deshalb redet man in der Definition von *geordneten* Paaren. Es ist $(x, y) \neq (y, x)$, es sei denn, daß $x = y$ wäre.

Definition 0.8. *Seien X und Y zwei Mengen. Die Menge aller möglichen geordneten Pare (x, y) mit $x \in X$ und $y \in Y$ schreibt man als $X \times Y$. Es ist also*

$$X \times Y = \{(x, y) \mid x \in X \textit{ und } y \in Y\}.$$

Beispiel 0.9. *Wir haben $\mathbb{R}^2 = \{(x_1, x_2) \mid x_1 \in \mathbb{R} \textit{ und } x_2 \in \mathbb{R}\} = \mathbb{R} \times \mathbb{R}$.*

Man beachte, daß normalerweise X keine Teilmenge von $X \times Y$ ist.
Was ist $X \times Y$ wenn $X = \emptyset$?

0.2 Definitionen

Man kann nicht definieren, was eine Definition ist.
Wir versuchen eine Beschreibung, worum es bei einer Definition geht:[1]

- eine Definition bewirkt, daß ein Begriff geboren wird (seine Vorfahren sind meist andere Begriffe, die zuvor definiert wurden),
- man kann keinen Begriff zweimal oder öfter definieren,
- Definitionen dürfen keine Zirkelschlüsse enthalten,
- oft verweisen Definitionen von Begriffen auf einen vorher definierten Oberbegriff, der jetzt eingeschränkt wird,
- bei Definition von Adjektiven (oder allgemeiner: von Eigenschaften) ist der Bezug entscheidend,
- bei Definition von Substantiven gibt es auch oft einen Bezug.

[1] Wenn Sie in der Wikipedia nachschlagen unter dem Begriff *Definition*, dann sehen Sie, welcher Aufwand in der Philosophie getrieben werden muß, um diesen Begriff faßlich zu machen. . .

Beispiel 0.10. *Der* Anstieg *einer Geraden in einem xy–Koordinatenkreuz ist eine reelle Zahl, die angibt, wie sich die y–Koordinate ändert, wenn die x–Koordinate um* 1 *wächst.*
Geben Sie hier an, was

- *der Bezug,*
- *der Oberbegriff,*
- *der einschränkende Halbsatz*

sind.

Definition 0.11. *Ein* Kreis *in einer Ebene ist die Menge all jener Punkte dieser Ebene, die von einem vorgegebenen Punkt einen vorgegebenen Abstand haben.*
Geben Sie hier an, was

- *der Bezug,*
- *der Oberbegriff,*
- *der einschränkende Halbsatz*

sind.

0.3 Mathematische Beweise

Beweise bestehen aus ganzen Sätzen. Diese Sätze enthalten

1. Handlungsbeschreibungen,
2. Aussagen.

Beispiele für Handlungsbeschreibungen sind

- „Wir führen den Beweis mit vollständiger Induktion.“
- „Wir machen eine Fallunterscheidung.“

Beispiele für Aussagen sind

- „$2 + 3 = 7$“
- „$\sqrt{2}$ ist irrational“
- „Der Mond ist eine Butterkremtorte“
- „Jeden Montag im Semester findet das Tutorium statt“

Keine Aussagen sind

- „$c^2 = a^2 + b^2 - 2ab\cos\gamma$“
- „$x^2 + 41x + 12 = 0$“
- „BWL ist interessant“

Definition 0.12. *Eine* Aussage *ist ein grammatischer Satz, der entweder wahr ist oder falsch, aber nicht beides.*

Beschreibung 0.13. *Ein grammatischer Satz, der unbestimmte Objekte (von evtl. sogar unbestimmten Typ) enthält, ist eine* Aussageform *(oder einfach nur Unsinn).*

Wir reden hier von einer *Beschreibung* anstelle einer Definition, weil uns für eine korrekte Definition das philosophische Hintergrundwissen fehlt. Im übrigen ist unser Ziel ja vorwiegend die Mathematik.

Beispiel 0.14. *„$c^2 = a^2 + b^2 - 2ab\cos\gamma$" ist auch dann noch eine Aussageform, wenn wir schon wissen, daß a, b, c, γ unbekannte reelle Zahlen sind.*
Diese Gleichung wird erst dann zu einer Aussage, wenn a, b, c, γ fixierte reelle Zahlen sind.

Aussageformen sind keine Aussagen. Deshalb dürfen sie in Beweisen auch nicht verwendet werden. Man sagt auch, daß sie noch freie Bezeichner enthalten.
Zwei entscheidende Regeln für Ihre Hausaufgaben sind also:

- Formulieren Sie immer in grammatisch vollständigen Sätzen (nach einigem Training werden Sie es dann schaffen, sich kurz auszudrücken).
- Jeder Bezeichner muß eingeführt werden. Zum Einführen eines Bezeichners gehören
  - die Angabe seines Typs (Zahl, Punkt, Funktion, Menge, Abbildung, Matrix, ...),
  - die Angabe, was den Wert dieses Bezeichners festlegt.

  Das Einführen eines Bezeichners kann so geschehen, daß er nicht mehr frei ist.

Später kommen weitere Regeln zur Hausaufgabengestaltung.

0.4 Aussagenverknüpfungen

Beschreibung 0.15. *Seien A und B Aussagen. Dann sind folgende Dinge auch Aussagen:*

- *„nicht (A)",*
- *„A und B",*
- *„A oder B",*

definiert anhand der Wahrheitstabellen *in Tabelle 1.*

A	B	A und B	A oder B
w	w	w	**w**
w	f	f	f
f	w	f	f
f	f	f	f

A	nicht(A)
w	f
f	w

Tabelle 1: Diese Wahrheitstabellen definieren die Verknüpfungen „und", „oder", „nicht".

Wir nennen diese drei Dinge auch Aussagenverknüpfungen.

Warnung: *Die Aussagenverknüpfung „Heute esse ich zu mittag vegetarisch oder Schnitzel" ist mathematisch auch dann wahr, wenn ich heute zu mittag beides esse. Der entsprechende Tabelleneintrag ist fett markiert.*
Im allgemeinen Sprachgebrauch ist das „oder" oft unbewußt exklusiv gemeint (also als „entweder oder"), aber im mathematischen Sprachgebrauch immer inklusiv.

Beschreibung 0.16. *Es seien A, B Aussagen. Dann heißen in „A und B"*

- *A, B: Input,*
- *„A und B": Output.*

Analog für beliebige andere Verknüpfungen. Die Wahrheitstabellen beschreiben Abbildungen vom Input zum Output.

Definition 0.17. *Zwei* Aussagenverknüpfungen *heißen* äquivalent*, wenn sie dieselben Wahrheitstabellen haben.*

Satz 0.18. *Seien A und B Aussagen. Dann sind äquivalent:*

- *„nicht(A oder B)“,*
- *„(nicht A) und (nicht B)“.*

Beweis. Wir bestimmen die Wahrheitstabellen:

A	B	A oder B	nicht(A oder B)
w	w	w	f
w	f	w	f
f	w	w	f
f	f	f	w

A	B	nicht A	nicht B	nicht(A) und nicht(B)
w	w	f	f	f
w	f	f	w	f
f	w	w	f	f
f	f	w	w	w

Die Übereinstimmung beider Tabellen in den ersten beiden Spalten sowie letzten Spalten ergibt die Behauptung. □

Analog beweist man folgende Sätze:

Satz 0.19. *Seien A und B Aussagen. Dann sind äquivalent:*

- *„nicht(A und B)“,*
- *„(nicht A) oder (nicht B)“.*

Satz 0.20. *Seien A und B Aussagen. Dann sind äquivalent:*

- *„A oder B“,*
- *„B oder A“.*

Satz 0.21. *Sei A eine Aussage. Dann sind äquivalent:*

- *„nicht(nicht(A))“,*
- *„A“.*

Beschreibung 0.22. *Seien A, B Aussagen. Dann ist „wenn A, dann B“ (abgekürzt als $A \Longrightarrow B$) auch eine Aussage, und zwar mit folgender Wahrheitstabelle.*

A	B	$A \Longrightarrow B$
w	w	w
w	f	f
f	w	**w**
f	f	**w**

Wir sagen auch „A impliziert B“ bzw. „A ist hinreichend für B“ bzw. „B ist notwendig für A“.

Warnung: *Die Aussage „Wenn der Mond eine Butterkremtorte ist, dann ist er eine kalorienreiche Mahlzeit“, ist* wahr, *denn hier liegt die Gestalt $A \Longrightarrow B$ vor; und A ist falsch, und der Wahrheitswert von B ist irrelevant. Ebenfalls wahr ist: „Wenn der Mond eine Butterkremtorte ist, dann ist 7 eine Quadratzahl“.*

Satz 0.23. *Seien A und B Aussagen. Die Verknüpfungen*

- *„$A \Longrightarrow B$“,*
- *„(nicht A) oder B“*

sind äquivalent.

Beweis. Wir bestimmen die Wahrheitstabelle:

A	B	$A \Longrightarrow B$	nicht (A)	(nicht A) oder B
w	w	w	f	w
w	f	f	f	f
f	w	w	w	w
f	f	w	w	w

Die Übereinstimmung der dritten und fünften Spalte vollendet den Beweis. □

Satz 0.24. *Seien A und B Aussagen. Dann sind äquivalent:*

- *„$A \Longrightarrow B$,*
- *„(nicht B) $\Longrightarrow$ (nicht A)“.*

Wir sagen auch: die Aussage „(nicht B) $\Longrightarrow$ (nicht A)“ ist die *Kontraposition* zur Aussage „$A \Longrightarrow B$“.
Den Beweis könnten wir erneut über einen Vergleich der Wahrheitstabellen führen. Lehrreicher ist es allerdings, wenn wir etwas eleganter vorgehen und unsere bisher bewiesenen Sätze einsetzen. Heuristisch sinnvoll ist es, die kompliziertere der beiden Aussagen als Startpunkt zu nehmen, diese dann solange äquivalent umzuformen, bis die andere Aussage herauskommt.

Beweis. Wir haben folgende Kette von Aussagenverknüpfungs–Äquivalenzen:

- „(nicht B) $\Longrightarrow$ (nicht A)“ ist wegen Satz 0.23 äquivalent zu
- „(nicht(nicht B)) oder (nicht A)“, und dies ist wegen Satz 0.21 äquivalent zu
- „B oder (nicht A)“, und dies ist wegen Satz 0.20 äquivalent zu
- „(nicht A) oder (B)“, und dies ist wegen Satz 0.23 äquivalent zu
- „$A \Longrightarrow B$“.

Für die Aussagenverknüpfungs–Äquivalenzrelation gilt die Eigenschaft der Transitivität (dies beweist man durch Rückgriff auf die Definition der Transitivität als Leser/in bitte selbst !), womit der Beweis vollendet ist. □

Definition 0.25. *Seien A, B Aussagen. Die Verknüpfung „($A \Longrightarrow B$) und ($B \Longrightarrow A$)“ kürzen wir ab als „$A \Longleftrightarrow B$“, und wir sagen dann auch, daß A und B äquivalent zueinander sind.*

Sprechweisen sind:

- „A genau dann, wenn B“,
- „A dann und nur dann, wenn B“.

Satz 0.26. *Diese Relation ist tatsächlich eine Äquivalenzrelation (d.h., sie besitzt die Eigenschaften der Reflexivität, Symmetrie, Transitivität). Und für den Fall, daß A und B ihrerseits Aussagenverknüpfungen sein sollten (d.h. Spezialfälle von Aussagen), ist diese Definition widerspruchsfrei zu Definition 0.17.*

Der Beweis sei den Leserinnen und Lesern anempfohlen.

0.5 Stolpersteine der Logik

Es seien in diesem Abschnitt A und B immer Aussagen. Wir listen einige Situationen auf, wo die mathematische Logik von der Alltagslogik abweicht.

- Sei „A oder B“ wahr. Dann dürfen A und B beide gleichzeitig wahr sein.

- Ein Elter sagt zum Kind: „Wenn Du Dein Zimmer nicht aufräumst, dann bekommst Du kein Eis". Diese Aussage (mathematisch interpretiert) kündigt keinerlei Konsequenzen an, was passieren wird, wenn das Kind das Zimmer aufräumt.

  Das heißt: sei die Verknüpfung „$A \Longrightarrow B$" wahr. Dann wird keine Aussage gemacht, welche Konsequenzen eintreten, wenn A falsch ist.

  Das heißt: die Verknüpfungen „$A \Longrightarrow B$" und „(nicht A) $\Longrightarrow$ (nicht B)" sind in der mathematischen Logik voneinander unabhängig. Die zweite Verknüpfung heißt *Inversion* der ersten Verknüpfung.

  Ein anderes Beispiel: Der Satz „Венн Ду дас лесен каннст, бист Ду кеин думмер Весси" macht keine Aussage darüber, was für diejenigen Menschen gilt, die diesen Satz nicht lesen können.

  Ein weiteres Beispiel: die mittelalterliche Kirchenlogik sagt: „*Wenn* Gott den Geringsten unter euch liebt, um wieviel mehr muß er dann den König lieben !" Die mathematische Logik geht anders: weil der König ja gerade nicht der Geringste ist, macht der erwähnte Satz keinerlei Aussage, ob Gott den König überhaupt liebt.

- Sei „$A \Longrightarrow B$" wahr. Dann ist über den Wahrheitsgehalt von A nichts bekannt. Und es nichts bekannt über den Wahrheitsgehalt von B.

  Ein Beispiel: die mittelalterliche Kirchenlogik sagt: „*Wenn* Gott den Geringsten unter euch liebt, um wieviel mehr muß er dann den König lieben !" Die mathematische Logik geht anders: es ist doch gar nicht gesagt, *daß* Gott den Geringsten liebt. Aus dem Satz kann auch nicht geschlußfolgert werden, ob Gott überhaupt irgendjemanden liebt.

- Sei „$A \Longrightarrow B$" wahr. Dann ist nichts bekannt über den Wahrheitsgehalt von „$B \Longrightarrow A$". Die zweite Verknüpfung heißt *Konversion* bzw. *Umkehrung* der ersten.

- Die Aussage „A **nur** dann, wenn B" ist eine äquivalente sprachliche Umformulierung von „wenn A, dann B".

- Die Primzahldefinition könnte man formulieren als „Eine natürliche Zahl heißt Primzahl, wenn sie mindestens gleich zwei ist, und wenn sie zusätzlich nur durch Eins und sich selbst teilbar ist". Dies ist zwar eine Formulierung der Art „$A \Longleftarrow B$" (also eine Implikation), tatsächlich gemeint ist bei Definitionen aber immer die Äquivalenz (und nicht die Implikation).

  Ein anderes Beispiel: man kann für Dreiecke die Eigenschaft „rechtwinklig" wie folgt definieren: „Wenn ein Dreieck einen rechten Innenwinkel hat, dann heißt dieses Dreieck *rechtwinklig*". So wie dieser Satz formuliert ist, läßt er die (absurde) Möglichkeit offen, daß auch manch ein Dreieck ohne rechten Innenwinkel rechtwinklig genannt werden dürfte. Tatsächlich meint man in der Mathematik aber: „**Genau** dann, wenn ein Dreieck einen rechten Innenwinkel hat, heißt dieses Dreieck *rechtwinklig*". Aus Gründen der sprachlichen Einfachheit läßt man das Wort **genau** immer weg, obwohl es nach den Prinzipien der Logik gesagt werden müßte. Die sprachliche Schönheit hat sich an dieser Stelle gegen die Konsequenz der Logik durchgesetzt.

- In der Mathematik wird anders gezählt: der Satz „Ich habe ein Ohr" ist für praktisch die meisten Menschen mathematisch wahr, denn die mathematische Logik schließt nicht aus, daß ich noch ein weiteres Ohr habe.

0.6 Fortgeschrittene Aspekte der Logik

Wir benötigen noch einige Schlußfolgerungstechniken.

Satz 0.27 (**Abtrennungsregel, Modus Ponens**). *Seien P und Q Aussagen. Angenommen,*

wir wissen: *die Aussage „$P \Longrightarrow Q$" ist wahr*

wir wissen: *die Aussage P ist wahr.*

Dann dürfen wir schlußfolgern, daß die Aussage Q wahr ist.

Bemerkung 0.28. *Diese Schlußfolgerungsregel heißt* Abtrennungsregel*, weil sie uns die Erlaubnis gibt, von der kombinierten Aussage „$P \Longrightarrow Q$" die Aussage Q abzutrennen. Ein anderer gebräuchlicher Name (anstatt Abtrennungsregel) ist* Modus Ponens*, was aber etwas ungenau ist. Genauer müßte die lateinische Bezeichnung*

Modus Ponendo Ponens *lauten, in Abgrenzung zur Bezeichung* Modus Tollendo Ponens*, die sich auf folgende Situation bezieht: wenn wir wissen, daß „P oder Q“ wahr ist, und wir auch noch wissen, daß „nicht(P)“ wahr ist, dann dürfen wir schlußfolgern, daß die Aussage Q wahr ist. Weitere Informationen finden sich in der Wikipedia unter dem Eintrag* Modus Ponens*.*

Geschrieben wird der Modus Ponens in der Logik (betrachtet als Teildisziplin der Mathematik) gelegentlich als

$$P \Longrightarrow Q,\, P \;\vdash\; Q.$$

In die Standardsprache übersetzen wir diese Zeile wie folgt:

„$P \Longrightarrow Q$“ : wir wissen, daß „$P \Longrightarrow Q$“ wahr ist

„ , “: und

„P“ : wir wissen, daß P wahr ist.

„$\vdash$“ : wir dürfen schlußfolgern, daß

„Q“ : es ist Q wahr.

Man sieht schnell, warum es grober Unfug wäre, das Symbol „$\vdash$“ durch „$\Longrightarrow$“ zu ersetzen.

Beweis der Abtrennungsregel. Die Wahrheitstabelle von „$P \Longrightarrow Q$“ ist:

P	Q	$P \Longrightarrow Q$
w	w	w
w	f	f
f	w	w
f	f	w

Die vier Zeilen beschreiben alle überhaupt irgendwie möglichen Situationen. Unser erster Wissensteil schließt die zweite Zeile aus. Unser zweiter Wissensteil schließt die dritte und vierte Zeile aus. Es bleibt bloß die erste Zeile als tatsächlich mögliche Situation übrig. In dieser Zeile ist aber Q als wahr gewählt worden. $\square$

Die Beweisstrategie war: alle überhaupt möglichen Situationen wurden in einer vollständigen Fallunterscheidung erfaßt. Drei dieser Situationen widersprechen dem vorausgesetzten Wissensstand, also muß die vierte Situation vorliegen.
Die Zulässigkeit des Modus Tollendo Ponens beweist man auf ähnliche Weise.

Kapitel 1

Problemlösen I

Im Jahr 1996 erschien ein berühmter Zeitschriftenartikel [3][1] des Mathematik-Didaktikers Heinrich Winter, der darlegte, daß Schülerinnen und Schüler folgende drei Grunderfahrungen im Schulfach Mathematik gemacht haben sollten:

1. Mathematik ist eine Disziplin, die uns dabei hilft, echte Praxisprobleme zu lösen (z.B. beim Aufstellen eines Finanzplans einer Firma, oder bei der Prognose, ob eine noch zu bauende Brücke tragfähig sein wird, oder beim Erstellen eines Fahrplans eines Verkehrsunternehmens, . . .).

2. Mathematik ist ein großes Theoriengebäude, bei dessen Errichtung man von einem festen Fundament von Axiomen startet und auf logisch korrekte Weise eine Aussage nach der anderen beweist.

3. Mathematik ist eine Beschäftigung, bei der wir lernen können, wie man Probleme löst. Hierbei sollen die Schülerinnen und Schüler insbesondere lernen, auch dann noch eine Lösung zu finden, wenn der Lösungsweg noch nicht bekannt ist.

Die erste Grunderfahrung findet sich in der Modellierungsvorlesung des zweiten Semesters des Mathematikstudiums wider, die zweite Grunderfahrung in praktisch jeder Mathematikvorlesung. Die dritte Grunderfahrung sollen Sie in den Übungen der drei Basismodule einüben.
Die theoretischen Hintergründe des Problemlösens sollen in diesem Kapitel erarbeitet werden. Wesentliche Beiträge zur Theorie und Praxis des Problemlösens wurden von George Polya geliefert, dessen Buch *How to solve it* bzw. *Schule des Denkens* [2] auch heute noch lesenswert ist.
Problemlösen besteht aus vier Phasen:

Problem verstehen: wir finden heraus, was wir in der Aufgabe liefern sollen. Wir ermitteln Voraussetzungen und Behauptungen, und wir klären die Bedeutungen aller Fachbegriffe.

Plan erarbeiten: da bei dem jetzt zu lösenden Problem der Lösungsweg nicht bekannt ist (wir haben also gerade nicht die schultypische Situation vor uns, daß wir *einfach eine Rechenschablone runterrechnen*), müssen wir den Lösungsweg suchen. Diese Suche muß irgendwo anfangen. Typischerweise tragen wir unser Wissen zusammen, und wir versuchen uns ähnlicher Aufgaben zu erinnern, die wir früher erfolgreich lösen konnten, und von denen wir hoffentlich einige Ideen wiederverwerten können. Womöglich werden wir bei der zweiten Phase einiges ausprobieren müssen. Sackgassen sind keine Schande, Rumsitzen-und-Nichtsprobieren aber sehr wohl !

Plan durchführen: aus der zweiten Phase entnehmen wir einen Lösungsplan, der zu funktionieren scheint. Diesen führen wir jetzt aus.

Rückblick: wir schauen auf das, was wir geleistet haben, und wir prüfen nach, ob wir wirklich die gestellte Aufgabe gelöst haben (oder ob hier noch etwas fehlt). Wir schauen auch nach, ob unser Lösungsweg vielleicht noch verschönert werden könnte, oder ob er noch einiges mehr leisten könnte. Wir merken uns leistungsstarke Arbeitstechniken.

[1] an die Lehramtsstudierenden und weitere Interssierte: Lektüre sehr empfohlen !

Diese vierschrittige Arbeitsweise ist deshalb allgemeinbildungsrelevant, weil sie in den verschiedensten Berufen vorkommt: ein Anwalt, der die Interessen eines Mandanten zu vertreten hat und deshalb ein Schriftstück aufsetzen muß; oder ein Soziologe, der eine Studie (Meinungsumfrage) zu konzipieren hat; oder ein Automechaniker, der ein Klappergeräusch abstellen soll — stets kann der Arbeitsablauf in diese vier Schritte zerlegt werden.
Ein sehr ähnlicher vierstufiger Ablauf findet sich auch in der mathematischen Modellierung, und das dortige Verfahren hat sogar einen Kreislaufcharakter.
Wir versuchen, das Problemlösen anhand von Beispielen zu verdeutlichen.

1.1 Funktionalgleichungen

Aufgabe 1.1. *Gesucht sind alle Funktionen $f\colon \mathbb{R} \to \mathbb{R}$, die folgende Gleichung erfüllen:*

$$f(xy) = \frac{f(x) + f(y)}{x + y},$$

für jegliche x, $y \in \mathbb{R}$ mit $x + y \neq 0$.

Problem verstehen: eine Funktion $f\colon \mathbb{R} \to \mathbb{R}$ ist *irgendeine* Abbildung, die auf ganz $\mathbb{R}$ definiert ist, und die jeder reellen Zahl eine reelle Zahl zuordnet. Wir dürfen nicht annehmen, daß eine solche Funktion f eine schöne Formel hat. Es ist zulässig, daß eine Lösung f auf die wildeste Weise unstetig ist.

Wir ermitteln irgendwie eine odere mehrere Funktionen f, und dann sind wir verpflichtet, zwei Aussagen zu beweisen:

- die von uns benannten Funktionen sind Lösungen (das bedeutet, daß wir eine Probe machen),
- alle anderen Funktionen von $\mathbb{R}$ nach $\mathbb{R}$ sind keine Lösungen.

Alternativ zum zweiten • können wir auch beweisen: wenn es eine Lösung f der Aufgabe gibt, dann muß sie eine der von uns benannten Funktionen sein (hinter dieser Argumentation steht die Methode der Kontraposition).

In der genannten Gleichung steht rechts eine Division, die nicht durchgeführt werden kann, wenn der Nenner gleich Null wäre, was aber durch die einschränkende Voraussetzung $x + y \neq 0$ vermieden wird.

In der genannten Gleichung bedeutet die linke Seite $f(xy)$ den Funktionswert von f an der Stelle des Produktes $x \cdot y$. Die Gleichung verknüpft also die Werte von f an drei (nicht notwendigerweise verschiedenen) Stellen miteinander.

Da wir bereits einige Kenntnisse in Logik haben, formulieren wir die Aufgabenstellung in der Sprache der Logik:

Sei $f\colon \mathbb{R} \to \mathbb{R}$ eine Funktion. Dann definieren wir

$$f \text{ heißt Lösung} :\Longleftrightarrow$$
$$\Big(\textbf{Wenn } \; x, y \in \mathbb{R} \text{ und } x + y \neq 0, \; \textbf{dann } \; \text{gilt } f(xy) = \frac{f(x) + f(y)}{x + y} \Big) \text{ ist wahr.}$$

Gesucht sind alle solchen f.

Der Doppelpunkt vor dem $\Longleftrightarrow$ besagt, daß die Aussage links davon durch die Aussage rechts vom $\Longleftrightarrow$ definiert wird. In der großen Klammer steht eine Aussage der Form „$P \Longrightarrow Q$".

Wir führen eine abkürzende Schreibweise ein:

$$x, y \in \mathbb{R} \text{ und } x + y \neq 0 \quad \Longrightarrow \quad f(x \cdot y) = \frac{f(x) + f(y)}{x + y}. \tag{1.1}$$

Plan erarbeiten: wir wollen uns an die Sorgfaltsregel halten, daß jeder Bezeichner eingeführt werden muß. Es ist f ein Bezeichner, also sollten wir f einführen.

Wir legen fest: sei f eine Lösung der Funktionalgleichung. Jetzt ist der Bezeichner f eingeführt, und wir sind den allerersten Schritt zum Beweis der alternativen Formulierung des obigen zweiten • gegangen.

Weil gemäß unserer Bezeichnerfestlegung f eine Lösung ist, erhalten wir als Schlußfolgerung: die Aussage (1.1) ist wahr. Dies ist eine Aussage der Form „$P \Longrightarrow Q$“. Es ist Q interessanter als P. Wir kennen den Modus Ponendo Ponens als eine erlaubte Schlußfolgerungsregel.

Es sind $x = 1$ und $y = 1 \in \mathbb{R}$, und $1 + 1 \neq 0$. Nach Modus Ponendo Ponens (MPP) haben wir

$$f(1 \cdot 1) = \frac{f(1) + f(1)}{1 + 1},$$

also $f(1) = f(1)$. Diese Aussage haben wir bewiesen (mittels MPP), und vermutlich ist sie nutzlos. Die Bezeichner x und y, die wir soeben festgelegt hatten, geben wir wieder frei.

Es sind $x = 2$ und $y = 3 \in \mathbb{R}$, und $2 + 3 \neq 0$. Nach MPP haben wir

$$f(2 \cdot 3) = \frac{f(2) + f(3)}{2 + 3},$$

also $5f(6) = f(2)+f(3)$. Diese Aussage ist bewiesen, und ihr Nutzen ist nicht so richtig klar. Die Bezeichner x und y, die wir soeben festgelegt hatten, geben wir wieder frei.

Es sind $x = 1$ und $y = 0 \in \mathbb{R}$, und $1 + 0 \neq 0$. Nach MPP haben wir

$$f(1 \cdot 0) = \frac{f(1) + f(0)}{1 + 0}.$$

Die Bezeichner x und y, die wir soeben festgelegt hatten, geben wir wieder frei. Es folgt $f(0) = f(1)+f(0)$, also

$$f(1) = 0. \tag{1.2}$$

Dies könnte nützlich sein. Unser aktueller Erkenntnisstand ist: die Aussagen (1.1) und (1.2) sind beide wahr, denn die erste ist definitionsgemäß äquivalent zu unserer Bezeichnerfestlegung, daß f eine Lösung sei, und die zweite haben wir bewiesen mittels MPP.

Der MPP war bis hierhin recht nützlich, und wir haben keine anderweitigen Ideen im Moment, also bleiben wir bei unserer bisher einzigen Idee[2], bringen jetzt aber unsere neue Erkenntnis ins Spiel, daß $f(1) = 0$ ist.

Es sind $x = 1$ und $y \in \mathbb{R} \setminus \{-1\}$ reelle Zahlen mit der Eigenschaft $x + y \neq 0$, also folgt mittels MPP für solche x, y, daß

$$f(1 \cdot y) = \frac{f(1) + f(y)}{1 + y},$$

was sich wegen $f(1) = 0$ (nachdem man den Nenner $1 + y$ hochmultipliziert hat) vereinfacht zu

$$y \in \mathbb{R} \setminus \{-1\} \quad \Longrightarrow \quad yf(y) = 0. \tag{1.3}$$

Wir erinnern uns: ein Produkt reeller Zahlen ist genau dann gleich Null, wenn wenigstens einer seiner Faktoren Null ist. Die meisten Elemente der Menge $\mathbb{R} \setminus \{-1\}$ sind ungleich Null, sodaß wir die Idee gewinnen, die zweite Abtrennungsregel (Modus Tollendo Ponens, MTP) zu verwenden. Wir haben eine Kette von Implikationen:

$$\begin{aligned} &y \in \mathbb{R} \setminus \{-1, 0\} \quad \Longrightarrow \quad y \in \mathbb{R} \setminus \{-1\} \quad \Longrightarrow \quad yf(y) = 0 \\ &\quad \Longleftrightarrow \quad y = 0 \text{ oder } f(y) = 0. \end{aligned}$$

Insgesamt ist folgende zusammengesetzte Aussage (Implikation) von uns bewiesen:

$$y \in \mathbb{R} \setminus \{-1, 0\} \quad \Longrightarrow \quad y = 0 \text{ oder } f(y) = 0.$$

Wenn $y \in \mathbb{R} \setminus \{-1, 0\}$, dann ist die Aussage „$y = 0$“ falsch, also liefert uns der Modus Tollendo Ponens:

$$y \in \mathbb{R} \setminus \{-1, 0\} \quad \Longrightarrow \quad f(y) = 0. \tag{1.4}$$

[2] die anschauliche Vorstellung dahinter ist: wir quetschen eine Zitrone solange aus, bis sie keinen Tropfen mehr liefert.

Unser Erkenntnisstand enthält die Aussagen (1.1), (1.2), (1.3), (1.4), die für jede Lösung f gelten, denn wir haben diese vier Aussagen wasserdicht bewiesen.

Wir haben $f(-1)$ und $f(0)$ noch nicht bestimmt. Eine gute Idee ist es, zur Startaussage (1.1) zurückzugehen. Wir können -1 darstellen als Produkt von -7 und $\frac{1}{7}$. Und wir haben $0 = 11 \cdot 0$.

Es sind $x = -7$ und $y = \frac{1}{7} \in \mathbb{R}$ mit der Eigenschaft $x + y \neq 0$. Nach MPP haben wir

$$f\left(-7 \cdot \frac{1}{7}\right) = \frac{f(-7) + f(1/7)}{-7 + \frac{1}{7}},$$

was wegen der bewiesenen Aussage (1.4) die Gleichung $f(-1) = 0$ impliziert.

Es sind $x = 0$ und $y = 11 \in \mathbb{R}$ mit der Eigenschaft $x + y \neq 0$. Nach MPP haben wir

$$f(0 \cdot 11) = \frac{f(0) + f(11)}{0 + 11},$$

was wegen der bewiesenen Aussage (1.4) die Gleichung $f(0) = \frac{1}{11} f(0)$ impliziert. Daraus folgt: $f(0) = 0$.

Wir haben also insgesamt gezeigt: wenn f eine Lösung ist, dann ist

$$f\colon \mathbb{R} \to \mathbb{R}, \qquad f : x \mapsto f(x) = 0.$$

Andere Lösungen kann es also nicht geben.

Der zweite • aus der Phase „Problemverstehen" ist gezeigt, und für den ersten • haben wir den Plan, eine Probe durchzuführen.

Plan durchführen: es gibt eine einzige Lösung, und zwar

$$f\colon \mathbb{R} \to \mathbb{R}, \qquad f : x \mapsto f(x) := 0.$$

Dies ist wirklich eine Lösung der Aufgabe, denn für dieses f gilt tatsächlich:

$$x, y \in \mathbb{R}, \quad x + y \neq 0 \quad \Longrightarrow \quad 0 = \frac{0 + 0}{x + y}.$$

Also ist unser oben erwähntes f tatsächlich eine Lösung.

Jetzt zeigen wir, daß es keine weiteren Lösungen geben kann.

Sei f eine Lösung der Aufgabe. Deshalb gilt (1.1).

Wir setzen $x = 1$ und $y = 0$ in (1.1) ein und erhalten die Aussage $f(1) = 0$.

Wir setzen $x = 1$ und ein beliebiges $y \neq -1$ in (1.1) ein und erhalten wegen $f(1) = 0$ die Aussage $f(y) = \frac{f(y)}{1+y}$, woraus sich $yf(y) = 0$ ergibt für alle $y \neq -1$.

Für $y \neq 0$, $y \neq -1$ ziehen wir daraus die Folgerung $f(y) = 0$.

Einsetzen von $x = -7$ und $y = \frac{1}{7}$ in (1.1) liefert uns dann $f(-1) = 0$.

Einsetzen von $x = 0$ und $y = 11$ in (1.1) bringt uns $f(0) = 0$.

Das bedeutet $f(y) = 0$ für jedes $y \in \mathbb{R}$.

Also kann es keine weiteren Lösungen geben als die von uns angegebene Nullfunktion. Die Aufgabe ist gelöst.[3]

Rückblick: wir haben die beiden • aus Phase 1 tatsächlich abgearbeitet.[4] Das Wählen von $x = 1$ in (1.1) hat den Vorteil, daß wir in dieser Gleichung die Funktion f bloß noch an zwei verschiedenen Stellen auswerten anstelle von vorher drei verschiedenen Stellen. Das eröffnet dann Chancen, einige f–Werte gegeneinander zu verrechnen.

Erinnerungswürdige Arbeitstechniken sind:

[3] eine gut bearbeitete Hausaufgabenlösung sollte ungefähr so aufgeschrieben werden wie die hiesige Phase 3. Im Gegensatz dazu verbleibt der Text von Phase 2 typischerweise auf dem Schmierblatt.

[4] ein subjektiver Rückblick des Autors dieses Skriptes: der Schreibaufwand für Phase 2 war eine Strapaze. In Zukunft wird es sprachlich knapper zugehen.

- das Ausdrücken der Aufgabenstellung mithilfe unseres logischen Formalismus, in Verbindung mit der Benutzung von MPP und MTP
- das *Spezialisieren*: die Aussage (1.1) ist sehr allgemein, denn sie gilt für sehr viele x und y. Wir wählen spezielle x und/oder spezielle y, was uns eine neue Aussage mit kleinerem Gültigkeitsbereich bringt.
- das *Dranbleiben-am-Ball*: unmittelbar nachdem wir die Erkenntnis $f(1) = 0$ gewonnen hatten, haben wir diese Aussage sofort wieder in der Ausgangsgleichung (1.1) eingesetzt und dann zügig die Erkenntnis (1.3) geschlußfolgert.
- das *Zurückgehen-zum-Start:* wenn das Dranbleiben-am-Ball nichts mehr bringt, kann man auch zur Startaussage(1.1) zurückgehen. Auf diesem Wege haben wir $f(-1)$ und $f(0)$ bestimmt.

Wir betrachten eine weitere Aufgabe.

Aufgabe 1.2. *Man ermittle alle Funktionen $f: \mathbb{R} \to \mathbb{R}$, für die*

$$f(xf(x) + f(y)) = y + (f(x))^2$$

gilt für alle $x, y \in \mathbb{R}$.

Problem verstehen: wir dürfen nicht annehmen, daß f eine schöne Funktionsgleichung hat.

Wir ermitteln irgendwie eine odere mehrere Funktionen f, und dann zeigen wir zwei Dinge: diese f sind tatsächlich Lösungen; und alle anderen Funktionen von $\mathbb{R}$ nach $\mathbb{R}$ sind keine Lösungen.

Seltsam bzw. unangenehm ist die linke Seite: der Ausdruck $x \cdot f(x) + f(y)$ wird nochmal in die Funktion f eingesetzt.

Die Umformulierung in die Sprache der Logik ist:

Sei $f: \mathbb{R} \to \mathbb{R}$ eine Funktion. Dann definieren wir

$$f \text{ heißt Lösung} :\Longleftrightarrow \Big(\textbf{Wenn } \ x, y \in \mathbb{R}, \ \textbf{dann } \ \text{gilt } f\Big(x \cdot f(x) + f(y)\Big) = y + (f(x))^2\Big) \text{ ist wahr.}$$

Gesucht sind alle solchen f.

Wir führen eine abkürzende Schreibweise ein:

$$x, y \in \mathbb{R} \quad \Longrightarrow \quad f\Big(x \cdot f(x) + f(y)\Big) = y + (f(x))^2. \tag{1.5}$$

Plan erarbeiten: wir tragen unsere bekannten Lösungsstrategien zusammen:

- wir *erinnern* uns an ähnliche Aufgaben von früher (und kopieren die Ideen von dort),
- das *Spezialisieren.* Das könnte bedeuten, daß wir konkrete Werte für x und y in (1.5) einsetzen.
- das *Dranbleiben-am-Ball.* Das heißt: sobald wir eine Erkenntnis gefunden haben, die nützlich erscheint, verwerten wir diese ohne weiteres Zögern.
- das *Zurückgehen-zum-Start.* Das heißt, daß wir uns gelegentlich nochmal die Aufgabenstellung anschauen; vielleicht stecken dort noch einige Informationen drin.
- das *Literaturstudium.* Wir schauen in Büchern oder in Vorlesungsmitschriften nach und suchen dort Definitionen oder Sätze, die thematisch passen könnten.
- das *Verschönern unangenehmer Terme.* Wir könnten in (1.5) auf solche Weise clever gewählte Werte für x und y einsetzen, daß angenehmere Terme entstehen.

Für uns bringt das *Literaturstudium*, daß wir einige Eigenschaften finden, die eine Funktion von $\mathbb{R}$ nach $\mathbb{R}$ haben könnte: surjektiv, injektiv, bijektiv, stetig, differenzierbar, gerade, ungerade. Wir müßten natürlich beweisen, daß f diese Eigenschaften wirklich hätte (man darf es nicht einfach so behaupten und dann im Lösungsverlauf weitermarschieren).

Sei f eine Lösung (jetzt ist der Bezeichner f eingeführt). Dann wissen wir, daß (1.5) wahr ist. Uns fällt nicht sehr viel ein, also probieren wir unser Glück mit dem Einsetzen einiger Werte in (1.5).

$$
\begin{aligned}
&x = 0, \quad y = 0 \quad \text{in (1.5) einsetzen liefert} && f(0 + f(0)) = 0 + (f(0))^2, \\
&x = 0, \quad y \in \mathbb{R} \quad \text{in (1.5) einsetzen liefert} && f(0 + f(y)) = y + (f(0))^2 \quad \forall y \in \mathbb{R}, \\
&x \in \mathbb{R}, \quad y = 0 \quad \text{in (1.5) einsetzen liefert} && f(xf(x) + f(0)) = 0 + (f(x))^2 \quad \forall x \in \mathbb{R}.
\end{aligned}
$$

Die zweite Aussage sieht vielversprechend aus. Wir schreiben sie nochmal schöner:

$$
y \in \mathbb{R} \quad \Longrightarrow \quad f(f(y)) = y + (f(0))^2. \tag{1.6}
$$

Jetzt benötigen wir einen **scharfen Blick**: aus (1.6) folgt die Surjektivität von f.

Um die Surjektivität zu beweisen (was wir sehr bald leisten werden), *gehen wir zur Definition des Fachbegriffs „Surjektivität" zurück*, und so erkennen wir, daß folgendes Unter-Problem zu lösen ist: sei ein beliebiger Wert $w \in \mathbb{R}$ vorgegeben. Dann sollen wir ein $z \in \mathbb{R}$ finden mit der Eigenschaft $f(z) = w$. Wenn wir dieses Unter-Problem gelöst haben, dann ist die Surjektivität von $f : \mathbb{R} \to \mathbb{R}$ gezeigt.

Jetzt beweisen wir die Surjektivität von f. Sei also ein $w \in \mathbb{R}$ uns vorgegeben. Mit scharfem Blick setzen wir $z := f(w - (f(0))^2)$. Dies ist möglich, weil $f(0)$ eine feste reelle Zahl ist (die wir zwar nicht kennen, aber wir wissen, daß es $f(0)$ gibt). Und tatsächlich ist

$$
f(z) = f(f\Big(w - (f(0))^2\Big)) \overset{(1.6)}{=} \Big(w - (f(0))^2\Big) + (f(0))^2 = w,
$$

wobei wir (1.6) benutzt haben. Die Surjektivität von f ist damit gezeigt.

Wir verwenden nochmal unseren **scharfen Blick**: aus (1.6) folgt die Injektivität von f.

Um die Injektivität zu beweisen (was wir sehr bald leisten werden), *gehen wir zur Definition des Fachbegriffs „Injektivität" zurück*, und so erkennen wir, daß folgendes Unter-Problem zu lösen ist: sei $f(z_1) = f(z_2)$ für zwei Zahlen $z_1, z_2 \in \mathbb{R}$. Dann sollen wir beweisen, daß $z_1 = z_2$. Wenn wir dieses Unter-Problem gelöst haben, dann ist die Injektivität von $f : \mathbb{R} \to \mathbb{R}$ gezeigt.

Jetzt beweisen wir die Injektivität von f. Sei also $f(z_1) = f(z_2)$. Dann ist

$$
z_1 + (f(0))^2 \overset{(1.6)}{=} f(f(z_1)) = f(f(z_2)) \overset{(1.6)}{=} z_2 + (f(0))^2,
$$

also $z_1 = z_2$. Die Injektivität von f ist damit gezeigt.

Also ist $f : \mathbb{R} \to \mathbb{R}$ auch bijektiv.

Wir *gehen zurück zum Start* und schauen uns (1.5) an. Der Term $xf(x) + f(y)$ ist häßlich; wir wollen an seiner Stelle etwas schöneres. Wie können wir ihn *verschönern* ? Wir könnten $x = 0$ setzen, dann wird dieser häßliche Term zum schöneren Term $f(y)$. Wir könnten aber auch x so wählen, daß $f(x) = 0$ wird, und erneut entsteht der schönere Term $f(y)$.

Oder wir könnten y so wählen, daß $f(y) = 0$ wird. Dann wird der häßliche Term $xf(x) + f(y)$ zum dezent schöneren Term $xf(x)$. Eine solche Wahl ist tatsächlich möglich, weil wir f vorhin als bijektiv nachgewiesen hatten.

Diese neue Strategie setzen wir jetzt um. Weil $f : \mathbb{R} \to \mathbb{R}$ bijektiv ist, gibt es genau eine Nullstelle $x_0 \in \mathbb{R}$ von f, d.h. $f(x_0) = 0$ für ein gewisses $x_0 \in \mathbb{R}$. Wir wissen nicht, welchen Wert dieses x_0 hat, aber wir wissen, daß es existiert.

Mit diesem x_0 praktizieren wir die Heuristik des *Spezialisierens*:

$$
\begin{aligned}
&x = x_0, \quad y = 0 \quad \text{in (1.5) einsetzen liefert} && f(0 + f(x_0)) = 0 + (f(x_0))^2, \\
&x = x_0, \quad y \in \mathbb{R} \quad \text{in (1.5) einsetzen liefert} && f(x_0 \cdot 0 + f(y)) = y + (f(x_0))^2 \quad \forall y \in \mathbb{R}, \\
&x \in \mathbb{R}, \quad y = x_0 \quad \text{in (1.5) einsetzen liefert} && f(xf(x) + f(x_0)) = x_0 + (f(x))^2 \quad \forall x \in \mathbb{R}.
\end{aligned}
$$

Diese drei Zeilen stellen wir um zu

$$
f(f(x_0)) = 0,
$$

$$
y \in \mathbb{R} \quad \Longrightarrow \quad f(f(y)) = y, \tag{1.7}
$$

$$
x \in \mathbb{R} \quad \Longrightarrow \quad f(xf(x)) = x_0 + (f(x))^2. \tag{1.8}
$$

Ein *scharfer Blick* auf (1.7) läßt unser *Gedächtnis* aktiv werden: sowas ähnliches hatten wir schonmal gesehen, nämlich in (1.6). Wir verheiraten (1.6) und (1.7):

$$y \in \mathbb{R} \quad \Longrightarrow \quad y + (f(0))^2 \overset{(1.6)}{=} f(f(y)) \overset{(1.7)}{=} y.$$

Es folgt $(f(0))^2 = 0$, also $f(0) = 0$. Also ist 0 eine Nullstelle von f. Wir *bleiben am Ball*, denken also über Nullstellen von f nach, und wir lassen unser *Gedächtnis* rotieren: x_0 war aber auch eine Nullstelle von f ! Und f hatten wir als injektiv nachgewiesen. Also haben wir jetzt $x_0 = 0$ bewiesen. Das ist eine Aussage über x_0.

Wir *bleiben am Ball* und suchen uns Formeln heraus, wo x_0 drin vorkommt. Dabei finden wir die Formel (1.8), die sich jetzt vereinfacht zu

$$x \in \mathbb{R} \quad \Longrightarrow \quad f(x \cdot f(x)) = (f(x))^2.$$

Unsere Schreiberei geht jetzt schon über zwei Seiten, also sollten wir *Ordnung schaffen* und alle bisher bewiesenen Aussagen (bzw. die stärksten Aussagen) übersichtlich zusammentragen:

$$x, y \in \mathbb{R} \quad \Longrightarrow \quad f(xf(x) + f(y)) = y + (f(x))^2, \tag{1.9}$$
$$y \in \mathbb{R} \quad \Longrightarrow \quad f(f(y)) = y, \tag{1.10}$$
$$x \in \mathbb{R} \quad \Longrightarrow \quad f(xf(x)) = (f(x))^2, \tag{1.11}$$
$$f\colon \mathbb{R} \to \mathbb{R} \quad \text{ist bijektiv.} \tag{1.12}$$

Die ersten drei dieser Aussagen kann man umschreiben, indem man den Allquantor $\forall$ benutzt. Man bekommt dann Aussagen im Stile von $\forall\, x, y \in \mathbb{R}\colon f(xf(x) + f(y)) = y + (f(x))^2$ usw.

Mit unseren ordentlich zusammengetragenen Erkenntnissen (1.9)–(1.12) wollen wir jetzt arbeiten.

Könnten wir vielleicht (1.10) und (1.11) verheiraten ? Nimm $y \in \mathbb{R}$ beliebig. Dann könnten wir $x := f(y)$ wählen (wir *spezialisieren* also) und damit in (1.11) hineingehen:

$$y \in \mathbb{R} \quad \Longrightarrow \quad f(f(y) \cdot f(f(y))) = (f(f(y)))^2.$$

Hier benutzen wir jetzt (1.10) auf beiden Seiten des Gleichheitszeichens:

$$y \in \mathbb{R} \quad \Longrightarrow \quad f(f(y) \cdot y) = (y)^2.$$

Der Bezeichner x kommt hier gar nicht vor, also ist der Bezeichner x frei. Also dürfen wir umtaufen (statt y schreiben wir x):

$$x \in \mathbb{R} \quad \Longrightarrow \quad f(xf(x)) = x^2, \tag{1.13}$$

wobei wir auf der linken Seite die Kommutativität der Multiplikation in $\mathbb{R}$ benutzt hatten.

Wir haben ein aufmerksames *Gedächtnis*, das uns daran erinnert, die linke Seite schon mal gesehen zu haben. Und tatsächlich können wir (1.13) mit (1.11) verheiraten:

$$x \in \mathbb{R} \quad \Longrightarrow \quad (f(x))^2 = x^2.$$

Jetzt wollen wir die Wurzel ziehen (aber bitte so, wie es sich in der Mathematik gehört !):

$$x \in \mathbb{R} \quad \Longrightarrow \quad (f(x) = x) \ \text{ oder } \ (f(x) = -x).$$

Rechts vom $\Longrightarrow$ steht eine oder-Verknüpfung zweier Aussagen, die man sich anschaulich als eine „Weiche“ vorstellen kann. Hier ist eine logische Feinheit zu beachten: es kann sein, daß $f(\sqrt{3}) = +\sqrt{3}$, jedoch $f(12) = -12$. Für unterschiedliche x kann diese Weiche also unterschiedlich regeln, so wie es eine Weiche vor einem Bahnhof ja auch tut. Wir haben allerdings die Hoffnung, daß wir jetzt die möglichen Lösungskandidaten weit genug eingegrenzt haben, daß wir durch *zurückgehen auf den Start*, also zur Gleichung (1.9), alle Lösungen f dann schnell finden.

Die Phase des **Plan erarbeitens** nähert sich der Vollendung. Wir schauen nochmal den Gedankengang durch, und wir entdecken, daß wir einen kleinen Umweg gegangen sind: wenn wir irgendwie anders zeigen, daß $f(0) = 0$ eine Nullstelle ist, dann benötigen wir den Nachweis von Surjektivität und Injektivität gar nicht.

Plan ausführen: Sei f eine Lösung. Dann gilt (1.5). Wir wählen in (1.5) die Zahlen $x = 2012$ und $y = -(f(2012))^2$, und es folgt

$$f\Big(2012 \cdot f(2012) + f(-(f(2012))^2)\Big) = -(f(2012))^2 + (f(2012))^2 = 0,$$

also besitzt die Funktion $f\colon \mathbb{R} \to \mathbb{R}$ mindestens eine Nullstelle, nämlich z.B.

$$x_0 := 2012 \cdot f(2012) + f(-(f(2012))^2)$$

ist eine solche.

Wir wählen in (1.5) die Zahl $x = x_0$, und es folgt

$$y \in \mathbb{R} \quad \Longrightarrow \quad f(f(y)) = y.$$

Wir wählen in (1.5) die Zahl $x = 0$, und es folgt

$$y \in \mathbb{R} \quad \Longrightarrow \quad f(f(y)) = y + (f(0))^2.$$

Damit muß $f(0) = 0$ sein.

Jetzt wählen wir in (1.5) die Zahl $y = 0$, und es folgt

$$x \in \mathbb{R} \quad \Longrightarrow \quad f(x \cdot f(x)) = (f(x))^2. \tag{1.14}$$

Die Gleichung rechts vom „$\Longrightarrow$“ gilt für jede reelle Zahl x. Es ist aber $f(x)$ auch eine reelle Zahl, also folgt

$$x \in \mathbb{R} \quad \Longrightarrow \quad f(f(x) \cdot f(f(x))) = (f(f(x)))^2.$$

Hierin benutzen wir nun die oben gezeigte Identität $f(f(x)) = x$:

$$x \in \mathbb{R} \quad \Longrightarrow \quad f(f(x) \cdot x) = x^2.$$

Der Vergleich dieser Erkenntnis mit (1.14) liefert uns dann

$$x \in \mathbb{R} \quad \Longrightarrow \quad (f(x))^2 = x^2,$$

und daraus folgt: wenn $x \in \mathbb{R}$, dann ist $f(x) = +x$ oder $f(x) = -x$.

Offenkundig ist $f(0) = +0 = -0$. Angenommen, es wäre $f(x_1) = +x_1$ und $f(x_2) = -x_2$, wobei $x_1 \neq 0$ und $x_2 \neq 0$. Wir zeigen jetzt durch einen Widerspruchsbeweis, daß dies nicht sein kann, und dieser Widerspruchsbeweis geht so:

Wir setzen $x = x_1$ und $y = x_2$ in (1.5) ein:

$$f(x_1 \cdot x_1 - x_2) = x_2 + x_1^2.$$

Die linke Seite dieser Gleichung hat einen der Werte $+(x_1 \cdot x_1 - x_2)$ oder $-(x_1 \cdot x_1 - x_2)$.

Fall 1: $f(x_1 \cdot x_1 - x_2) = +(x_1 \cdot x_1 - x_2)$**.** Dann folgt $x_1^2 - x_2 = x_2 + x_1^2$, also $x_2 = 0$. Das war aber verboten worden.

Fall 2: $f(x_1 \cdot x_1 - x_2) = -(x_1 \cdot x_1 - x_2)$**.** Dann folgt $-x_1^2 + x_2 = x_2 + x_1^2$, also $x_1 = 0$. Das war aber auch verboten worden.

Diese beiden Fälle sind also beide unmöglich, und einen dritten Fall kann es nicht geben. Mit diesem Widerspruchsbeweis haben wir also gezeigt:

Es kann nicht sein, daß reelle Zahlen $x_1 \neq 0$ und $x_2 \neq 0$ existieren, für die $f(x_1) = +x_1$, aber $f(x_2) = -x_2$.

Das bedeutet: wenn f eine Lösung unserer Aufgabe ist, dann ist f eine der beiden folgenden Funktionen:

$$f\colon \mathbb{R} \to \mathbb{R}, \qquad f\colon x \mapsto f(x) := +x,$$
$$f\colon \mathbb{R} \to \mathbb{R}, \qquad f\colon x \mapsto f(x) := -x.$$

Jetzt zeigen wir, daß diese beiden Lösungskandidaten wirklich Lösungen sind:

Wir geben den Bezeichner f wieder frei. Sei $f\colon \mathbb{R} \to \mathbb{R}$ definiert als die Funktion mit der Funktionsgleichung $f(x) = +x$. Dann ist f eine Lösung der Aufgabe, denn tatsächlich gilt für jegliche x, $y \in \mathbb{R}$, daß

$$x \cdot x + y = y + x^2 \implies x \cdot f(x) + f(y) = y + (f(x))^2 \implies f(xf(x) + f(y)) = y + (f(x))^2.$$

Wir geben den Bezeichner f wieder frei. Sei $f\colon \mathbb{R} \to \mathbb{R}$ definiert als die Funktion mit der Funktionsgleichung $f(x) = -x$. Dann ist f eine Lösung der Aufgabe, denn tatsächlich gilt für jegliche x, $y \in \mathbb{R}$, daß

$$-(x\cdot(-x)+(-y)) = y+(-x)^2 \implies -(x\cdot f(x)+f(y)) = y+(f(x))^2 \implies f(xf(x)+f(y)) = y+(f(x))^2.$$

Rückblick: es gibt genau zwei Lösungen, nämlich die Funktion

$$f\colon \mathbb{R} \to \mathbb{R}, \qquad f\colon x \mapsto f(x) := +x$$

und die Funktion

$$f\colon \mathbb{R} \to \mathbb{R}, \qquad f\colon x \mapsto f(x) := -x.$$

Die Aufgabe 1.2 ist vollständig gelöst.

1.2 Was haben wir gelernt bis hierhin ?

Die hier behandelten Themen sind für ein erfolgreiches Studium in der Mathematik unverzichtbar.
Wir haben einige **Kenntnisse** zumindest ein Stück weit erworben:

- der Begriff der Menge,
- die Verknüpfungen der Logik und ihre Subtilitäten, insbesondere im Vergleich zur Alltagslogik.

Wir haben einiges über **mathematische Syntax** gelernt:

- Beweise bestehen nur aus Aussagen und Handlungsbeschreibungen.
- Aussageformen sind keine Aussagen, weil sie noch freie Objekte enthalten.
- Eine Aussageform wandelt sich in eine Aussage um, wenn man die freien Objekte in ihr festlegt (also Bezeichner einführt). Dies geschieht oft mit den Symbolen $\forall$ und $\exists$.

Wir haben einige mathematische **Arbeitstechniken** kennengelernt:

- das sorgfältige Einführen aller Bezeichner,
- die Schlußfolgerungsregeln der Logik wie z.B. MTP und MPP,
- die Meta-Technik des Problemlösens.

Das Problemlösen ist ein vierstufiger Algorithmus, der in den meisten Berufen vorkommt (wenn auch sehr oft unbewußt). Es geht darum ein Problem zu lösen, das durch eine Rechenschablone eben gerade nicht abgearbeitet werden kann.
Zum erfolgreichen Problemlösen haben wir einige **Heuristiken und Strategien** kennengelernt:

Ähnliche Aufgaben von früher: können wir vielleicht bewährte Ideen von damals jetzt erneut benutzen ?

Literaturstudium: aus den Vorlesungsmitschriften und Büchern entnehmen wir Definitionen und Sätze, die thematisch verwandt mit unserer Aufgabe zu sein scheinen. Wenn wir zeigen wollen, daß eines der Objekte in unserer Aufgabe eine bestimmte Eigenschaft besitzt, dann gehen wir auf die Definition dieser Eigenschaft aus der Literatur zurück. Wenn wir einen Satz aus der Literatur bei der Lösung unserer Aufgabe zitieren wollen, dann beweisen wir, daß die Voraussetzungen dieses Satzes im Falle unserer Aufgabe wirklich erfüllt sind.

Spezialisieren: aus einer allgemeinen Situation machen wir eine spezielle Situation, die hoffentlich übersichtlicher ist. Leider ist die neue Situation weniger allgemein.

Verschönern unangenehmer Objekte: können wir vielleicht noch freie Variablen so mit Werten belegen, daß ein häßlicher Term schöner wird ?

Dranbleiben: wenn wir eine Erkenntnis gewonnen haben, die nützlich erscheint, dann versuchen wir, aus dieser frischen Erkenntnis soviel wie möglich an Schlußfolgerungen herauszuziehen.

Zurück zum Start: gelegentlich schauen wir uns nochmal die Aufgabenstellung an. Haben wir wirklich alles benutzt, was dort steht ?

Ordnung schaffen: hin und wieder schreiben wir alle gewonnenen Erkenntnisse übersichtlich hin. Sonst verliert man den Überblick.

Verheiraten: ein Sachverhalt wird auf zwei Weisen dargestellt, woraus man Erkenntnisse ziehen kann.

Das sind aber noch nicht alle Heuristiken und Strategien; weitere werden folgen.
Wir haben auch einige **allgemeine Fähigkeiten** kennengelernt, die für viele wissenschaftliche Studiengänge (nicht bloß die Mathematik) entscheidend sind:

ein scharfer Blick: öfter, als man glaubt, benötigt man Phantasie und Kreativität.

ein gutes Gedächtnis: es ist einerseits wichtig, Definitionen und Eigenschaften von vielen Fachbegriffen jederzeit sicher präsent zu haben. Ja, auch in der Mathematik muß man viele Dinge **auswendig lernen !**

Andererseits ist es wichtig, sich zu merken oder gut sichtbar zu notieren, wie weit man denn schon mit der Lösung der Aufgabe fortgeschritten ist, und was man alles schon weiß.

Konzentrationsfähigkeit: sehr nützlich ist die Fähigkeit, sich etwa 20 Minuten auf einen Punkt zu konzentrieren. Dazu gehört auch die Kompetenz, Syntaxregeln ganz konsequent einzuhalten und keine eigenen zu erfinden.

Die gute Nachricht ist, daß diese allgemeinen Fähigkeiten mit Motivation erworben und verbessert werden können.

1.3 Vollständige Induktion

Wir betrachten eine Menge $\{A(n) \colon n \in \mathbb{N}_0\}$ von Aussagen. Angenommen, es gelingt uns, folgende beiden • zu zeigen:

- die Aussage $A(0)$ ist wahr,
- für jedes $n \in \mathbb{N}_0$ gilt: wenn die Aussage $A(n)$ wahr ist, dann ist auch die Aussage $A(n+1)$ wahr.

Dann können wir schlußfolgern: weil $A(0)$ wahr ist und weil MPP eine erlaubte Schlußfolgerungsregel ist, ist auch $A(1)$ wahr.
Und wir können weiter schlußfolgern: weil $A(1)$ wahr ist und weil MPP eine erlaubte Schlußfolgerungsregel ist, ist auch $A(2)$ wahr.
Und wir können weiter schlußfolgern: weil $A(2)$ wahr ist und weil MPP eine erlaubte Schlußfolgerungsregel ist, ist auch $A(3)$ wahr.
Und so weiter. Für eine jegliche Zahl $n \in \mathbb{N}_0$ können wir zeigen, daß die Aussage $A(n)$ wahr ist. Dieses Verfahren ist bekannt als *vollständige Induktion*.

Aufgabe 1.3. *Für jedes $n \in \mathbb{N}_+$ und jegliche $x_1, \ldots, x_n \in \mathbb{R}_+$ ist zu zeigen, daß*

$$\sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} \leq \frac{x_1 + \cdots + x_n}{n}.$$

Diese Ungleichung ist bekannt als Ungleichung zwischen dem geometrischen Mittel und dem arithmetischen Mittel. Die Namen erklären sich daraus, daß die Zahlen x_1, $\sqrt{x_1 x_2}$, x_2 eine geometrische Folge bilden; und die Zahlen x_1, $\frac{x_1+x_2}{2}$, x_2 bilden eine arithmetische Folge.

Schritt 1: Problem verstehen. es geht um eine *Beweisaufgabe*, keine *Bestimmungsaufgabe*[5]. Die angegebene Ungleichung soll für alle $n \in \mathbb{N}_+$ bewiesen werden, und für alle positiven reellen $x_1, \dots, x_n$. Es ist im eigentlichen Beweis nicht erlaubt, ein „Lieblings-n" zu wählen. Ganz im Gegenteil: wir sollen die Ungleichung für $n = 1$ beweisen, und wir sollen sie für $n = 2$ beweisen, und wir sollen die Ungleichung für $n = 3$ beweisen, und für alle anderen $n \in \mathbb{N}_+$ auch noch. Das ist viel Arbeit.

Es ist auch nicht erlaubt, „Lieblings-x_j" zu wählen.

Auf dem Schmierzettel (also während Schritt 2 des Problemlöseverfahrens) können wir natürlich uns erstmal konkrete Situationen anschauen, um ein Gefühl für die Aufgabe zu bekommen. Der eigentliche Beweis (im Schritt 3 des Problemlöseverfahrens) muß dann aber allgemein sein.

Was wissen wir vom n ? Nur, daß $n \in \{1, 2, 3, 4, \dots\} = \mathbb{N}_+$ ist.

Was wissen wir von $x_1, \dots, x_n$? Nur, daß jede dieser Zahlen aus der Menge $\mathbb{R}_+ = \{s \in \mathbb{R} \colon s > 0\}$ stammt.

Warum diese Einschränkung auf die positiven Zahlen ? Vermutlich hat das damit zu tun, daß ansonsten das Produkt $x_1 x_2 \dots x_n$ negativ sein könnte, und dann gäbe es die Wurzel daraus nicht.

Von den x_j gibt es genau n Stück, nämlich $x_1, x_2, \dots, x_n$. Wir können also auf unserem Schmierzettel nicht zuerst die $x_1, \dots, x_n$ wählen; und danach wählen wir das n. So eine Wählerei ist logisch nicht möglich.

Schritt 2: Plan ausarbeiten. es könnte sinnvoll sein, mit vollständiger Induktion zu hantieren. Das Kapitel heißt ja nicht ohne Grund so. Vermutlich induzieren wir über die Variable n, denn die x_j sind reell und deshalb für eine Induktion untauglich.

Die Aufgabe hat viele Variablen: nämlich die Variable n, und davon abängig noch die x_j (n Stück davon). Das ist uns zu unübersichtlich, also benutzen wir die Lösungsstrategie des *Spezialisierens*. Das heißt, wir schauen uns erstmal konkrete Werte für n an, womit wir die Hoffnung verbinden, ein Gefühl für die Ungleichung zu entwickeln. Dieses Gefühl wird dann hoffentlich in einen Plan einmünden.

Welches n wollen wir wählen ? Es sind kleine n einfacher zu handhaben als große n, also nehmen wir ein kleines n. Das kleinste in der Aufgabenstellung erlaubte n ist $n = 1$.

Wir versuchen also, die behauptete Ungleichung zu beweisen, wenn $n = 1$ ist. Für jegliches $x_1 \in \mathbb{R}_+$ ist dann zu zeigen, daß

$$\sqrt[1]{x_1} \leq \frac{x_1}{1},$$

und das bedeutet nichts anderes als $x_1 \leq x_1$, was offensichtlich für jedes $x_1 \in \mathbb{R}_+$ stimmt.

Wir haben also die behauptete Ungleichung gezeigt, wenn $n = 1$ ist.

So richtig viel haben wir jetzt noch nicht herausgefunden, und einen Plan sehen wir immer noch nicht, also probieren wir ein weiteres konkretes n. Das einfachste solche n ist $n = 2$.

Wir versuchen also, die behauptete Ungleichung zu beweisen, wenn $n = 2$ ist. Für jegliche $x_1, x_x \in \mathbb{R}_+$ ist dann zu zeigen, daß

$$\sqrt{x_1 x_2} \leq \frac{x_1 + x_2}{2},$$

und diese Ungleichung ist jetzt nicht-banal. Wir wissen nicht, ob diese Ungleichung stimmt (solche Ungleichungen werden wir im Folgenden mit einem Fragezeichen markieren). Die Wurzel ist doof, also wollen wir sie weghaben. Das klappt mittels Quadrieren. Aus der Schule wissen wir, daß wir jetzt aufpassen müssen: ist das Quadrieren wirklich eine äquivalente Umformung ? Eigentlich nicht, aber innerhalb der positiven Zahlen schon. Und wegen $x_1, x_2 > 0$ (wie in der Aufgabenstellung vorausgesetzt) sind beide Seiten der Ungleichung positiv, also ist quadrieren zulässig, und wir haben somit

$$\sqrt{x_1 x_2} \stackrel{?}{\leq} \frac{x_1 + x_2}{2} \quad \Longleftrightarrow \quad x_1 x_2 \stackrel{?}{\leq} \frac{(x_1 + x_2)^2}{4} = \frac{x_1^2 + 2x_1 x_2 + x_2^2}{4},$$

[5] Zur Begriffsklärung: bei einer Bestimmungsaufgabe geht es oft darum, eine Gleichung zu lösen, denn sie lauten oft: „Man bestimme alle $x \in \mathbb{R}$, für die gilt, daß…". Im Gegensatz dazu soll bei einer Beweisaufgabe eine gewisse Aussage bewiesen werden. Bei Bestimmungsaufgaben gibt es eine Probe, aber bei Beweisaufgaben nicht.

und dies ist aber äquivalent zu $0 \overset{?}{\leq} (x_1 - x_2)^2$, was offensichtlich immer stimmt, egal welche positiven Werte man für x_1 und x_2 einsetzt. Also gilt die behauptete Ungleichung für $n = 2$.

Wir probieren noch ein n, nämlich $n = 3$. Für jegliche x_1, x_2, $x_3 \in \mathbb{R}_+$ ist dann zu zeigen, daß

$$\sqrt[3]{x_1x_2x_3} \overset{?}{\leq} \frac{x_1 + x_2 + x_3}{3}.$$

Wieder stört uns die Wurzel, weshalb wir die Ungleichung in die dritte Potenz erheben möchten, was eine äquivalente Umformung ist, denn alle beteiligten Zahlen sind gemäß Aufgabenstellung positiv. Beim Erheben in die dritte Potenz bekommen wir dann die neue Ungleichung

$$x_1x_2x_3 \overset{?}{\leq} \frac{(x_1 + x_2 + x_3)^3}{27} = \frac{x_1^3 + x_2^3 + x_3^3 + 3x_1^2(x_2 + x_3) + 3x_2^2(x_1 + x_3) + 3x_3^2(x_1 + x_2) + 6x_1x_2x_3}{27}.$$

Das sieht kompliziert aus. Selbst wenn wir es hinbekämen, diese Ungleichung irgendwie zu beweisen, spätestens beim Fall von $n = 4$ würden die Rechnungen dann richtig häßlich.

Also haben wir den Fall $n = 3$ jetzt noch nicht in den Griff bekommen.

Wir probieren Induktion, also den Schritt von $n = 2$ auf $n = 3$.

Wir setzen also voraus: wenn x_1 und x_2 positiv sind, dann ist $\sqrt{x_1x_2} \leq \frac{1}{2}(x_1 + x_2)$.

Wir wollen zeigen: wenn y_1, y_2, y_3 positiv sind, dann ist $\sqrt[3]{y_1y_2y_3} \leq \frac{1}{3}(y_1+y_2+y_3)$. Wir dürfen die Zahlen y_1, y_2, y_3 nicht wählen.

Beim Beweis von $\sqrt[3]{y_1y_2y_3} \overset{?}{\leq} \frac{1}{3}(y_1 + y_2 + y_3)$ darf die Ungleichung $\sqrt{x_1x_2} \leq \frac{1}{2}(x_1 + x_2)$ verwendet werden, und zwar mit beliebig wählbaren positiven x_1, x_2.

Die Frage ist also: vorgegeben seien positive y_1, y_2, y_3. Das stellen wir uns so vor, daß wir im Büro sitzen, und jemand reicht uns einen Zettel herein, auf dem diese drei Zahlen stehen. Diese y_1, y_2, y_3 dürfen wir also nicht mehr wählen. Wir sollen für diese drei Zahlen die Behauptung beweisen, aber ohne Zahlenrechnungen zu betreiben (nur die Logik ist erlaubt). Wie wollen/können wir dann positive x_1 und x_2 wählen, sodaß die Ungleichung $\sqrt[3]{y_1y_2y_3} \overset{?}{\leq} \frac{1}{3}(y_1 + y_2 + y_3)$ in Reichweite gelangt ? Wir können natürlich $x_1 := y_1$ und $x_2 := y_2$ setzen, was uns auf $\sqrt[2]{y_1y_2} \leq \frac{1}{2}(y_1 + y_2)$ bringt, aber wie bringen wir hier jetzt noch das y_3 eingebaut ?

Den Induktionsschritt von $n = 2$ auf $n = 3$ haben wir also nicht in den Griff bekommen.

Jetzt machen wir etwas ganz Verrücktes: wir versuchen den Induktionsschritt rückwärts. Wir wollen also von $n = 4$ zurückgehen auf $n = 3$.

Wir setzen also voraus: wenn $x_1, \ldots, x_4$ positiv sind, dann ist $\sqrt[4]{x_1x_2x_3x_4} \leq \frac{1}{4}(x_1 + x_2 + x_3 + x_4)$.

Wir wollen zeigen: wenn y_1, y_2, y_3 positiv sind, dann ist $\sqrt[3]{y_1y_2y_3} \leq \frac{1}{3}(y_1+y_2+y_3)$. Wir dürfen die Zahlen y_1, y_2, y_3 nicht wählen.

Beim Beweis von $\sqrt[3]{y_1y_2y_3} \overset{?}{\leq} \frac{1}{3}(y_1 + y_2 + y_3)$ darf die Ungleichung $\sqrt[4]{x_1x_2x_3x_4} \leq \frac{1}{4}(x_1 + x_2 + x_3 + x_4)$ verwendet werden, und zwar mit beliebig wählbaren positiven $x_1, \ldots, x_4$.

Schön ist, daß es vier Stück von den x_j gibt, aber nur drei Stück von den y_k. Wir probieren also $x_1 := y_1$, $x_2 := y_2$, $x_3 := y_3$, und jetzt haben wir noch Wahlfreiheiten beim x_4. Wir entscheiden uns für $x_4 := \frac{1}{3}(y_1 + y_2 + y_3)$.

Aus der Induktionsvoraussetzung (I.V.) schließen wir

$$\begin{aligned}
& \sqrt[4]{x_1x_2x_3x_4} \leq \frac{x_1 + x_2 + x_3 + x_4}{4} \\
\iff \quad & \sqrt[4]{y_1y_2y_3\frac{1}{3}(y_1 + y_2 + y_3)} \leq \frac{y_1 + y_2 + y_3 + \frac{y_1+y_2+y_3}{3}}{4} \\
\iff \quad & \sqrt[4]{y_1y_2y_3} \cdot \sqrt[4]{\frac{1}{3}(y_1 + y_2 + y_3)} \leq \frac{(y_1 + y_2 + y_3) \cdot (1 + 1/3)}{4} \\
\iff \quad & \sqrt[4]{y_1y_2y_3} \cdot \sqrt[4]{\frac{1}{3}(y_1 + y_2 + y_3)} \leq \frac{y_1 + y_2 + y_3}{3}
\end{aligned}$$

Alle beteiligten Zahlen sind positiv, also ist das Erheben in die vierte Potenz eine äquivalente Umformung, und wir bekommen als äquivalente Aussage also

$$\begin{aligned} y_1y_2y_3 \cdot \frac{1}{3}(y_1+y_2+y_3) &\le \frac{(y_1+y_2+y_3)^4}{3^4} \\ \iff \qquad y_1y_2y_3 &\le \frac{(y_1+y_2+y_3)^3}{3^3}, \end{aligned}$$

wobei wir hier durch $(y_1+y_2+y_3)$ dividiert haben, was eine äquivalente Umformung ist (denn dieser Divisor ist positiv). Wenn wir hier jetzt die dritte Wurzel aus der Ungleichung ziehen (das ist durchführbar und eine äquivalente Umformung, denn alle anwesenden Zahlen sind positiv), dann bekommen wir die zu beweisenden Ungleichung für $n = 3$.

Wir hätten auch $x_4 := \sqrt[3]{y_1y_2y_3}$ setzen können, und dann wären die Rechnungen ähnlich verlaufen.

Wir **schaffen Ordnung**. Folgendes haben wir gezeigt:

- die Behauptung für $n = 1$,
- die Behauptung für $n = 2$,
- einen konkreten Induktionsschritt rückwärts, nämlich von $n = 4$ auf $n = 3$.

Folgendes haben wir *nicht* zeigen können:

- die Behauptung für $n = 3$ auf direktem Wege,
- den Induktionsschritt vorwärts, insbesondere von $n = 2$ auf $n = 3$,
- die Behauptung für $n = 4$.

Wir nutzen unseren **scharfen Blick** und erkennen, daß wir von 2 auf 4 kommen können gemäß $4 = 2 \cdot 2$.

Wir setzen also voraus: wenn x_1 und x_2 positiv sind, dann ist $\sqrt{x_1x_2} \le \frac{1}{2}(x_1+x_2)$.

Wir wollen zeigen: wenn y_1, y_2, y_3, y_4 positiv sind, dann ist $\sqrt[4]{y_1y_2y_3y_4} \le \frac{1}{4}(y_1+y_2+y_3+y_4)$. Wir dürfen die Zahlen y_1, y_2, y_3, y_4 nicht wählen.

Zum Beweis benutzen wir (dreimal !) die Ungleichung zwischen dem arithmetischen und geometrischen Mittel, so wie wir sie für $n = 2$ schon bewiesen haben. Die linke Seite mit der vierten Wurzel erscheint uns komplizierter als die rechte Seite mit der Summe, also fangen wir dort an:

$$\begin{aligned} \sqrt[4]{y_1y_2y_3y_4} &= \sqrt{\sqrt{y_1y_2y_3y_4}} = \sqrt{(\sqrt{y_1y_1}) \cdot (\sqrt{y_3y_4})} \le \frac{1}{2}((\sqrt{y_1y_2}) + (\sqrt{y_3y_4})) \\ &\le \frac{1}{2}\left(\frac{1}{2}(y_1+y_2) + \frac{1}{2}(y_3+y_4)\right) \\ &= \frac{1}{4}(y_1+y_2+y_3+y_4). \end{aligned}$$

Das wollten wir haben.

Insgesamt haben wir also die Hoffnung, daß wir mit ähnlichen Rechnungen auch wie folgt induzieren können:

- von n auf $2n$, falls $n \ge 2$,
- von n auf $n-1$, falls $n \ge 3$.

Dieses Verfahren heißt *Vorwärts–Rückwärts–Induktion von Cauchy.*

Schritt 3: Plan ausführen. [6]

Wir beweisen die Behauptung in vier Teilen.

Teil 1: Die Behauptung gilt im Falle von $n = 1$**.** Denn dann ist die Behauptung äquivalent zu

$$\forall\, x_1 \in \mathbb{R}_+ : \quad \sqrt[1]{x_1} \le \frac{1}{1}x_1,$$

[6] Jetzt wollen wir Wert legen auf Schönheit, und zwar sowohl Schönheit der sprachlichen Darstellung auch Schönheit (im Sinne von Geradlinigkeit) des Gedankengangs.

was offenkundig stimmt.

Teil 2: Die Behauptung gilt im Falle von $n = 2$. Denn seien beliebige positive Zahlen x_1 und x_2 gegeben. Dann haben wir $(\sqrt{x_1} - \sqrt{x_2})^2 \geq 0$, woraus wir

$$x_1 - 2\sqrt{x_1} \cdot \sqrt{x_2} + x_2 \geq 0$$

schließen, und daraus folgt die Behauptung im Falle von $n = 2$ durch elementare Umformungen.

Teil 3: Wenn die Behauptung für eine Zahl $n \in \mathbb{N}_+$ mit $n \geq 2$ gilt, dann gilt die entsprechende Behauptung auch für das Doppelte von n.

Denn seien positive $y_1, \ldots, y_{2n}$ gegeben. Dann gilt (wegen Teil 2 und wegen der Induktionsvoraussetzung), daß

$$\begin{aligned}
\sqrt[2n]{y_1 \cdot \ldots \cdot y_{2n}} &= \sqrt[2]{\sqrt[n]{y_1 \cdot \ldots \cdot y_{2n}}} = \sqrt{\sqrt[n]{y_1 \cdot \ldots \cdot y_n} \cdot \sqrt[n]{y_{n+1} \cdot \ldots \cdot y_{2n}}} \\
&\leq \frac{1}{2}\left(\sqrt[n]{y_1 \cdot \ldots \cdot y_n} + \sqrt[n]{y_{n+1} \cdot \ldots \cdot y_{2n}}\right) \\
&\leq \frac{1}{2}\left(\frac{1}{n}(y_1 + \ldots + y_n) + \frac{1}{n}(y_{n+1} + \ldots + y_{2n})\right) \\
&= \frac{y_1 + \ldots + y_n + y_{n+1} + \ldots + y_{2n}}{2n}.
\end{aligned}$$

Das wollten wir zeigen.

Teil 4: Wenn die Behauptung für eine Zahl $n \in \mathbb{N}_+$ mit $n \geq 3$ gilt, dann gilt die entsprechende Behauptung auch für den Vorgänger von n.

Denn seien positive $y_1, \ldots, y_{n-1}$ gegeben. Dann setzen wir $x_1 := y_1, \ldots, x_{n-1} := y_{n-1}$ sowie $x_n := \sqrt[n-1]{y_1 \ldots y_{n-1}}$, und die Induktionsvoraussetzung lautet

$$\sqrt[n]{x_1 x_2 \cdot \ldots \cdot x_n} \leq \frac{1}{n}(x_1 + x_2 + \ldots + x_n).$$

Diese Ungleichung ist äquivalent zu

$$\sqrt[n]{y_1 y_2 \cdot \ldots \cdot y_{n-1} \cdot \sqrt[n-1]{y_1 y_2 \cdot \ldots \cdot y_{n-1}}} \leq \frac{1}{n}(y_1 + y_2 + \ldots + y_{n-1} + \sqrt[n-1]{y_1 y_2 \cdot \ldots \cdot y_{n-1}}).$$

Gemäß den Potenzrechengesetzen ist die linke Seite gleich

$$(y_1 y_2 \cdot \ldots \cdot y_{n-1})^{(1+\frac{1}{n-1}) \cdot \frac{1}{n}} = (y_1 y_2 \cdot \ldots \cdot y_{n-1})^{\frac{1}{n-1}} = \sqrt[n-1]{y_1 y_2 \cdot \ldots \cdot y_{n-1}},$$

sodaß wir insgesamt bewiesen haben, daß

$$\sqrt[n-1]{y_1 y_2 \cdot \ldots \cdot y_{n-1}} \leq \frac{y_1 + \ldots + y_{n-1}}{n} + \frac{1}{n} \sqrt[n-1]{y_1 y_2 \cdot \ldots \cdot y_{n-1}}.$$

Hieraus folgt die Behauptung durch elementare Umformungen.

Teil 5: Schluß. Nach den Prinzipien der vollständigen Induktion ist die Ungleichung zwischen dem geometrischen und arithmetischen Mittel für sämtliche $n \in \mathbb{N}_+$ bewiesen.

Schritt 4: Rückblick. [7]

Wir haben Fragen identifiziert, die für den Schritt 1 des Problemlösens sinnvoll sind:

- ist es eine *Beweisaufgabe* oder eine *Bestimmungsaufgabe* ?
- brauchen wir eine Probe ?
- welche Variablen dürfen wir wählen, und welche nicht ?
- gibt es vielleicht Abhängigkeiten der Variablen untereinander ?

[7] Bei Beweisaufgaben gibt es keine Probe.

- wenn bei einigen Variablen die Grundmenge eingeschränkt wird (im Beispiel dieser Aufgabe war $x_j \in \mathbb{R}_+$ vorausgesetzt anstatt $x_j \in \mathbb{R}$), dann könnte es dafür einen Grund geben. Welcher Grund wäre das ?

Wir haben folgende Lösungsstrategien gefestigt:

- das *Spezialisieren*. Insbesondere betraf dies die Wahl konkreter kleiner n im Schritt 2, um ein Gefühl für die Ungleichung zu entwickeln.
- das *Ordnung–schaffen*.
- den *scharfen Blick*.

Wir haben folgendes Handwerkszeug gefestigt:

- Potenzrechnen,
- sorgfältiger Umgang mit äquivalenten Umformungen von Ungleichungen,
- sonstige elementare Umformungen.

Folgendes war neu:

- manchmal durchläuft die Induktion die Menge der natürlichen Zahlen in ungewöhnlicher Reihenfolge. Hauptsache ist, daß wir bei jeder natürlichen Zahl vorbeikommen.

Damit ist die Aufgabe vollständig bearbeitet.

Aufgabe 1.4. *Für jedes $n \in \mathbb{N}$ mit $n \geq 2$, und für alle $x_1, \ldots, x_n \in \mathbb{R}_+$, sowie $w_1, \ldots, w_n \in [0,1]$ mit $\sum_{j=1}^n w_j = 1$ ist zu beweisen, daß*

$$x_1^{w_1} \cdot \ldots \cdot x_n^{w_n} \leq w_1 x_1 + \ldots + w_n x_n.$$

Schritt 1: Problem verstehen. Es ist eine Beweisaufgabe. Wir dürfen beim Beweis nichts wählen, weder n noch die x_j oder die w_k. Es ist $[0,1]$ das Intervall von 0 bis 1, wobei die Endpunkte mit dazugehören.

Um ein Gespür für die Ungleichung zu bekommen, verfolgen wir die *Extremierungs–Heuristik*. Wir machen also eine der Variablen so groß wie möglich oder so klein wie möglich und schauen uns dann an, was passiert.

Angenommen, w_2 wäre so groß wie möglich, das heißt $w_2 = 1$. Die Bedingung $\sum_{j=1}^n w_j = 1$ erzwingt dann $w_1 = w_3 = \ldots = w_n = 0$, und die Ungleichung wird zu

$$x_2^1 \overset{?}{\leq} 1x_2,$$

was offenkundig stimmt.

Angenommen, w_2 wäre so klein wie möglich, das heißt $w_2 = 0$. Die anderen w_j seien beliebig, aber die Bedingung $\sum_{j=1}^n w_j = 1$ muß natürlich erfüllt sein. Dann wird die Ungleichung zu

$$x_1^{w_1} \cdot x_3^{w_3} \cdot \ldots \cdot x_n^{w_n} \overset{?}{\leq} w_1 x_1 + w_3 x_3 + \ldots + w_n x_n,$$

und jetzt haben wir links einen Faktor weniger und rechts einen Summanden weniger. Es sieht so aus, als ob wir diese Situation mittels vollständiger Induktion in den Griff bekämen.

Angenommen, x_2 sei so groß wie möglich. Nun hat $\mathbb{R}_+$ kein größtes Element, also nehmen wir an, daß x_2 gigantisch groß sei, insbesondere viel größer als die restlichen x_j. Betrachtet als Funktion von x_2, wächst die linke Seite wie $x_2^{w_2}$, und wir können annehmen, daß $0 < w_2 < 1$ ist, denn die beiden Randfälle sind wenig interessant. Aber die rechte Seite wächst schneller, nämlich wie x_2^1. Dann ist glaubhaft, daß die behauptete Ungleichung gelten wird, falls x_2 riesig groß wird und $w_2 \neq 0$, $w_2 \neq 1$.

Angenommen, x_2 sei so klein wie möglich. Nun hat $\mathbb{R}_+$ kein kleinstes Element, also nehmen wir an, daß x_2 sehr klein und positiv sei, insbesondere viel kleiner als die restlichen x_j. Wir sehen dann, daß die linke Seite (für $x_2 \to 0$) gegen Null strebt, und die rechte Seite strebt gegen den positiven Wert $w_1 x_1 + w_3 x_3 + \ldots + w_n x_n$. Damit ist glaubhaft, daß die behauptete Ungleichung gelten wird, falls x_2 winzig klein wird und $w_2 \neq 0$, $w_2 \neq 1$.

Jetzt vollziehen wir das Gegenteil: anstatt die Parameter „möglichst weit auseinander“ zu positionieren, bringen wir sie möglichst nah zusammen.

Für die w_j hieße das also $w_1 = w_2 = \ldots = w_n = \frac{1}{n}$, und dann wird die Ungleichung zu

$$x_1^{1/n} x_2^{1/n} \cdot \ldots \cdot x_n^{1/n} \overset{?}{\leq} \frac{1}{n} x_1 + \frac{1}{n} x_2 + \ldots + \frac{1}{n} x_n.$$

Diese Ungleichung stimmt, denn das ist genau die vorige Aufgabe.

Wir ziehen die Erkenntnis: **die jetzige Aufgabe 1.4 ist allgemeiner als die vorige Aufgabe 1.3**.

Und wenn wir die x_j alle gleich wählen, also $x_1 = x_2 = \ldots = x_n$, dann wird die Ungleichung zu

$$x_1^{w_1} \cdot x_1^{w_2} \cdot \ldots \cdot x_1^{w_n} \overset{?}{\leq} (w_1 + w_2 + \ldots + w_n) x_1,$$

was ebenfalls stimmt, denn $w_1 + \ldots + w_n = 1$.

Schritt 2: Plan ausarbeiten. Wir versuchen unser Glück mit Induktion über n.

Der Induktionsanfang wäre $n = 2$. Dann haben wir zu zeigen, daß gilt: wenn $w_1 + w_2 = 1$ und $w_1, w_2 \in [0, 1]$, und wenn $x_1, x_2 \in \mathbb{R}_+$, dann ist

$$x_1^{w_1} x_2^{w_2} \overset{?}{\leq} w_1 x_1 + w_2 x_2.$$

Wir dürfen die Zahlen w_1, w_2, x_1, x_2 nicht wählen. Unsere Vorstellung ist: jemand anders gibt uns diese Zahlen und wir sollen jetzt die Ungleichung beweisen, dürfen aber nur Logik verwenden und kein Zahlenrechnen.

Ein *scharfer Blick* bringt uns zur Erkenntnis, daß wir die Variablenzahl reduzieren können durch Division, denn die behauptete Ungleichung ist äquivalent zu

$$1 \overset{?}{\leq} w_1 x_1^{1-w_1} x_2^{-w_2} + w_2 x_1^{-w_1} x_2^{1-w_2} \leq w_1 \left(\frac{x_1}{x_2} \right)^{w_2} + w_2 \left(\frac{x_1}{x_2} \right)^{-w_1}.$$

Wir können jetzt substituieren. Wie nehmen also einen neuen Bezeichner t und setzen ihn fest gemäß $t := \frac{x_1}{x_2}$. Dann ist $t \in \mathbb{R}_+$, und jede positive reelle Zahl kann als Wert von t tatsächlich angenommen werden. Wir wollen also zeigen:

$$\forall\, t \in \mathbb{R}_+ : 1 \overset{?}{\leq} w_1 t^{w_2} + w_2 t^{-w_1}.$$

Der Vorteil ist jetzt, daß wir eine Variable weniger zu verwalten haben.

Um ein Gespür für die Ungleichung zu bekommen, verwenden wir wieder die *Extremierungsheuristik*. Wir wählen also t riesengroß bzw. positiv winzigklein und schauen, was die rechte Seite macht. Offenkundig ist (wenn wir mal annehmen, daß $w_1, w_2 \notin \{0, 1\}$)

$$\lim_{t \to +\infty} t^{w_2} = +\infty, \qquad \lim_{t \to +\infty} t^{-w_1} = 0, \qquad \lim_{t \to +0} t^{w_2} = 0, \qquad \lim_{t \to +0} t^{-w_1} = +\infty,$$

sodaß die rechte Seite der behaupteten Ungleichung also nach $+\infty$ strebt, wenn t riesig oder positiv winzig wird. Das sieht gut aus.

Für welche t ist die rechte Seite minimal ? Offenkundig kann die rechte Seite nicht negativ werden, und sie kann auch nicht Null werden. Wir definieren eine Funktion f,

$$f : \mathbb{R}_+ \to \mathbb{R}_+, \qquad f : t \mapsto f(t) := w_1 t^{w_2} + w_2 t^{-w_1},$$

und diese Funktion ist stetig differenzierbar. Es ist

$$f'(t) = w_1 w_2 t^{-w_1 - 1}(t - 1),$$

also folgt, daß f auf dem Intervall $(0, 1)$ streng monoton fallend ist; und f ist auf dem Intervall $(1, \infty)$ streng monoton steigend. Deshalb hat die Funktion f ein globales Minimum im Punkt 1, und das heißt

$$\forall\, t \in \mathbb{R}_+ : f(t) \geq f(1).$$

Beachte $f(1) = 1$ wegen $w_1 + w_2 = 1$, also gilt tatsächlich $f(t) \geq 1$ für jedes positive t.

Der Induktionsanfang ist bewiesen !

Wir versuchen unser Glück mit dem Induktionsschritt. Sei $n \geq 2$.

Wir setzen voraus: wenn $x_1, \ldots, x_n$ positiv sind und $w_1, \ldots, w_n \in [0, 1]$ mit $w_1 + \ldots + w_n = 1$, dann ist $x_1^{w_1} \ldots x_n^{w_n} \leq w_1 x_1 + \ldots + w_n x_n$.

Wir wollen zeigen: wenn $y_1, \ldots, y_{n+1}$ positiv sind und $u_1, \ldots, u_{n+1} \in [0, 1]$ mit $u_1 + \ldots + u_{n+1} = 1$, dann ist $y_1^{u_1} \ldots y_{n+1}^{u_{n+1}} \leq u_1 y_1 + \ldots + u_{n+1} y_{n+1}$. Wir dürfen die Zahlen u_j und y_k nicht wählen.

Die zu zeigende Ungleichung schreiben wir als $L \overset{?}{\leq} R$. Wir haben

$$R = \Big(u_1 y_1 + \ldots + u_n y_n\Big) + u_{n+1} y_{n+1}.$$

Hier wollen wir die I.V. anwenden. Wenn $u_{n+1} = 0$ ist, dann gibt es den letzten Summanden gar nicht, und die I.V. kann direkt angewandt werden. Und wenn $u_{n+1} = 1$ ist, dann gibt es den Klammerterm nicht. Sei also $0 < u_{n+1} < 1$. Dann ist aber leider $u_1 + \ldots + u_n \neq 1$. Wir reparieren das:

$$R = (u_1 + \ldots + u_n) \cdot \left(\frac{u_1}{u_1 + \ldots + u_n} y_1 + \ldots + \frac{u_n}{u_1 + \ldots + u_n} y_n \right) + u_{n+1} y_{n+1} =: (u_1 + \ldots + u_n) K + u_{n+1} y_{n+1}.$$

Auf den Klammerterm K können wir die I.V. anwenden:

$$K \geq (y_1^{u_1} \cdot \ldots \cdot y_n^{u_n})^{\frac{1}{u_1 + \ldots + u_n}}.$$

Und wir können den Induktionsanfang benutzen:

$$R \geq K^{u_1 + \ldots + u_n} \cdot y_{n+1}^{u_{n+1}}.$$

Der Plan gelingt !

Schritt 3: Plan ausführen. Selbststudium !

Schritt 4: Rückblick. Wir haben neue **Heuristiken** und **Techniken** kennengelernt:

- *Extremierung* (ist eine Variante der *Spezialisierung*),
- Vergleich mit *früheren Aufgaben*,
- Ungleichungen können mittels Differentialrechnung behandelt werden,
- *Generalisierung*.

Wir beobachten: die jetzt behandelte Aufgabe 1.4 ist allgemeiner als die vorherige Aufgabe 1.3, aber trotzdem in gewissem Sinne leichter zu behandeln, denn die seltsame Vorwärts–Rückwärts–Induktion kommt jetzt gar nicht mehr vor.

Die früher behandelte Heuristik der Spezialisierung reduziert eine allgemeine Situation auf eine spezielle Situation.

Im Gegensatz dazu erweitert die jetzt entdeckte Heuristik der Generalisierung eine spezielle Situation zu einer allgemeinen Situation.

Manchmal werden Beweise einfacher, wenn man mehr beweist, als man ursprünglich wollte.

1.4 Vom Verallgemeinern, Interpretieren und Vernetzen

Wir haben folgende Ungleichungen bewiesen für positive x_1, x_2, x_3:

$$\sqrt{x_1 x_2} \leq \frac{x_1 + x_2}{2}, \qquad x_1^{w_1} x_2^{w_2} \leq w_1 x_1 + w_2 x_2 \quad (w_1, w_2 \in [0, 1], \quad w_1 + w_2 = 1)$$

$$\sqrt[3]{x_1 x_2 x_3} \leq \frac{x_1 + x_2 + x_3}{3}, \qquad x_1^{w_1} x_2^{w_2} x_3^{w_3} \leq w_1 x_1 + w_2 x_2 + w_3 x_3 \quad (w_1, w_2, w_3 \in [0, 1], \quad w_1 + w_2 + w_3 = 1).$$

Hierbei sind die Ungleichungen rechts allgemeiner als die Ungleichungen links; und trotzdem sind die allgemeineren Ungleichungen leichter zu beweisen, denn die vollständige Induktion läuft geradlinig durch.

Jetzt wollen wir diese Ungleichungen

interpretieren, in dem wir sie

vernetzen, und zwar mit dem Gebiet der Geometrie, und auf diesem Wege dann

verallgemeinern.

Unser Werkzeug ist die Logarithmusfunktion

$$\ln\colon \mathbb{R}_+ \to \mathbb{R},$$

die bekanntlich streng monoton ist. Dieser Begriff ist definiert wie folgt:

Definition 1.5. [8] *Sei $(a,b) \subset \mathbb{R}$ ein Intervall. Eine Funktion $f\colon (a,b) \to \mathbb{R}$ heißt* streng monoton wachsend, *wenn gilt: Für jegliche z_1, $z_2 \in (a,b)$ mit $z_1 < z_2$ ist $f(z_1) < f(z_2)$.*

Eine Folgerung ist dann: wenn z_1, $z_2 \in (a,b)$ mit $z_1 \le z_2$, dann ist $f(z_1) \le f(z_2)$.
Jetzt nehmen wir für z_1 jeweils die linke Seite der gezeigten vier Ungleichungen und für z_2 die rechte Seite, und es folgt (für jegliche positiven x_1, x_2, x_3), daß

$$\begin{aligned}
\frac{\ln x_1 + \ln x_2}{2} &\le \ln\left(\frac{x_1+x_2}{2}\right), \\
w_1 \ln x_1 + w_2 \ln x_2 &\le \ln(w_1 x_w + w_2 x_2), \quad w_1, w_2 \in [0,1], \quad w_1 + w_2 = 1, \\
\frac{\ln x_1 + \ln x_2 + \ln x_3}{3} &\le \ln\left(\frac{x_1+x_2+x_3}{3}\right), \\
w_1 \ln x_1 + w_2 \ln x_2 + w_3 x_3 &\le \ln(w_1 x_w + w_2 x_2 + w_3 x_3), \quad w_1, w_2, w_3 \in [0,1], \quad w_1 + w_2 + w_3 = 1.
\end{aligned}$$

Jede dieser vier Ungleichungen wollen wir geometrisch interpretieren. Dabei fangen wir aus Gründen der Arbeitsökonomie mit der einfachsten Ungleichung an, und dies ist die erste (weil sie die wenigsten Bezeichner hat). Sie besagt: in einem üblichen Koordinatensystem bilden wir den Mittelpunkt der Verbindungsstrecke der beiden Punkte $(x_1, \ln x_1)^\top$ und $(x_2, \ln x_2)^\top$. Dieser Mittelpunkt liegt unterhalb des Graphen der Logarithmusfunktion. (Eine Skizze dieses Sachverhalts gab es in der Tutoriumssitzung; die damals Anwesenden sind klar im Vorteil !)
Weiterhin verwenden wir die schulbekannte Punkt-Richtungs-Form einer Gerade in einem Koordinatensystem: Die Gerade durch die beiden Punkte $(x_1, \ln x_1)^\top$ und $(x_2, \ln x_2)^\top$ wird beschrieben durch die Menge

$$\left\{ \begin{pmatrix} x_1 \\ \ln x_1 \end{pmatrix} + w \left(\begin{pmatrix} x_2 \\ \ln x_2 \end{pmatrix} - \begin{pmatrix} x_1 \\ \ln x_1 \end{pmatrix} \right) : w \in \mathbb{R} \right\} = \left\{ (1-w) \begin{pmatrix} x_1 \\ \ln x_1 \end{pmatrix} + w \begin{pmatrix} x_2 \\ \ln x_2 \end{pmatrix} : w \in \mathbb{R} \right\},$$

und die Strecke zwischen diesen beiden Punkten (auch *Sehne* genannt) ist dann die Punktmenge

$$\left\{ (1-w) \begin{pmatrix} x_1 \\ \ln x_1 \end{pmatrix} + w \begin{pmatrix} x_2 \\ \ln x_2 \end{pmatrix} : w \in [0,1] \right\},$$

was wir auch schreiben können als

$$\left\{ w_1 \begin{pmatrix} x_1 \\ \ln x_1 \end{pmatrix} + w_2 \begin{pmatrix} x_2 \\ \ln x_2 \end{pmatrix} : \quad w_1, w_2 \in [0,1], \quad w_1 + w_2 = 1 \right\}.$$

Die zweite Ungleichung besitzt dann die geometrische Interpretation, daß der Graph der Logarithmusfunktion oberhalb jeder eigenen Sehne verläuft.
Der Punkt $w_1(x_1, \ln x_1)^\top + w_2(x_2, \ln x_2)^\top$ kann physikalisch wie folgt verstanden werden: Im Punkt $(x_1, \ln x_1)^\top$ positionieren wir eine Punktmasse der Größe w_2, und im Punkt $(x_2, \ln x_2)^\top$ positionieren wir eine Punktmasse der Größe $w_1 = 1 - w_2$. Dann ist $w_1(x_1, \ln x_1)^\top + w_2(x_2, \ln x_2)\top$ gerade der Massenschwerpunkt dieser Konstellation.[9]

[8] Wir denken ein wenig über die Struktur dieser Definition nach. Wir definieren hier kein *Objekt*, wofür wir dann eine Oberkategorie und einen einschränkenden Halbsatz benötigen würden. Sondern wir definieren eine *Eigenschaft* (typographisch markiert durch aufrechte Lettern). Diese Eigenschaft ist grammatisch ein Adjektiv (was bei Eigenschaften meistens so ist). Der einzige Daseinszweck eines Adjektivs besteht darin, ein Substantiv genauer zu beschreiben, und dieses Substantiv wollen wir *Bezug* nennen. Hier ist der Bezug eine Funktion $f\colon (a,b) \to \mathbb{R}$. Beachte, daß für Funktionen $f\colon \mathbb{R}^2 \to \mathbb{R}$ oder $f\colon \mathbb{R} \to \mathbb{R}^4$ kein monotones Wachstum definiert werden kann, weil es weder im $\mathbb{R}^2$ noch im $\mathbb{R}^4$ eine Relation $<$ gibt.

[9] Die Frage, warum hier die Indizes der w_j in vertauschter Weise auftauchen müssen, wurde im Tutorium thematisiert und ist für das Selbststudium exzellent geeignet.

Eines der w_j kann auch negativ sein; und dies können Sie dann selbst anhand der Kräftebilanz an einer Schubkarre deuten.
Für die letzte der vier Ungleichungen betrachten wir das Dreieck in der Ebene mit den drei Eckpunkte $(x_j, \ln x_j)^\top$, $j = 1, 2, 3$. Wir schreiben

$$\begin{aligned} & w_1 \begin{pmatrix} x_1 \\ \ln x_1 \end{pmatrix} + w_2 \begin{pmatrix} x_2 \\ \ln x_2 \end{pmatrix} + w_3 \begin{pmatrix} x_3 \\ \ln x_3 \end{pmatrix} \\ &= \left(w_1 \begin{pmatrix} x_1 \\ \ln x_1 \end{pmatrix} + w_2 \begin{pmatrix} x_2 \\ \ln x_2 \end{pmatrix} \right) + w_3 \begin{pmatrix} x_3 \\ \ln x_3 \end{pmatrix} \\ &= (w_1 + w_2) \left(\frac{w_1}{w_1 + w_2} \begin{pmatrix} x_1 \\ \ln x_1 \end{pmatrix} + \frac{w_2}{w_1 + w_2} \begin{pmatrix} x_2 \\ \ln x_2 \end{pmatrix} \right) + w_3 \begin{pmatrix} x_3 \\ \ln x_3 \end{pmatrix}. \end{aligned}$$

Wir erhalten den Punkt $\frac{w_1}{w_1+w_2}(x_1, \ln x_1)^\top + \frac{w_2}{w_1+w_2}(x_2, \ln x_2)^\top$, indem wir die Verbindungsstrecke der beiden Eckpunkte $(x_1, \ln x_1)^\top$ und $(x_2, \ln x_2)^\top$ im Verhältnis $w_2 : w_1$ teilen. Und die Verbindungsstrecke dieses Zwischenpunktes und der dritten Dreiecksecke wird dann im Verhältnis $w_3 : (w_1 + w_2)$ geteilt.
Man redet davon, daß jeder Punkt im Inneren des genannten Dreiecks mit *baryzentrischen Koordinaten* beschrieben werden kann. Eine andere Bezeichnung ist *Schwerpunktskoordinaten*.

Definition 1.6. [10] *Sei $(a, b) \subset \mathbb{R}$ ein Intervall. Eine Funktion $f : (a, b) \to \mathbb{R}$ heißt* konkav*, wenn gilt: Für jegliche x_1, $x_2 \in (a, b)$ und jegliche $w_1, w_2 \in [0, 1]$ mit $w_1 + w_2 = 1$ ist*

$$f(w_1 x_1 + w_2 x_2) \geq w_1 f(x_1) + w_2 f(x_2).$$

Die geometrische Interpretation ist, daß der Graph der Funktion f oberhalb einer jeden Sehne verläuft. Dabei verlangen wir nicht, daß der Graph „echt oberhalb“ verlaufen möge. Eine Funktion mit der Funktionsgleichung $f(x) = 2x + 7$ ist also auch konkav.
Eine entsprechende Definition des Begriffes *konvex* kann im Selbststudium überlegt werden (es gilt die Merkregel: e^x ist konv**EX**).

Theorem 1.7 (**Ungleichung von Jensen**). *Wenn die Funktion $f : (a, b) \to \mathbb{R}$ konkav ist, dann gilt für jedes $n \in \mathbb{N}_{\geq 2}$ und jegliche $x_1, \ldots, x_n \in (a, b)$ sowie jegliche $w_1, \ldots, w_n \in [0, 1]$ mit $w_1 + \ldots + w_n = 1$, daß*

$$f\left(\sum_{j=1}^{n} w_j x_j\right) \geq \sum_{j=1}^{n} w_j f(x_j).$$

Der Beweis (in voller Länge in der Tutoriumssitzung) benutzt die vollständige Induktion über n, und sein Induktionsschritt hat starke Ähnlichkeiten zum Induktionsschritt im Schritt 2 der Lösung zu Aufgabe 1.4.
Ein entsprechend abgeänderte Variante der Ungleichung von Jensen gilt für konvexe Funktionen.

Korollar 1.8. *Wenn α, β, γ die Innenwinkel eines beliebigen Dreiecks bezeichnen, dann gilt stets*

$$\sin \alpha + \sin \beta + \sin \beta \leq \frac{3}{2}\sqrt{3}.$$

Korollar 1.9 (**Ungleichung zwischen arithmetischem und quadratischen Mittel**). *Für jedes $n \in \mathbb{N}_+$ und jegliche $x_1, \ldots, x_n \in \mathbb{R}_+$ ist*

$$\frac{x_1 + \ldots + x_n}{n} \leq \sqrt{\frac{x_1^2 + \ldots + x_n^2}{n}}.$$

Es ist eine schöne Übungsaufgabe (deren Lernziel darin besteht, sich mit der Technik des *Vernetzens* vertraut zu machen), diese Ungleichung mittels der (aus der Vorlesung zur linearen Algebra bekannten) Cauchy–Schwarz–Ungleichung zu beweisen.

[10] Und auch dies ist eine Definition, welche eine *Eigenschaft* ins Leben ruft. Das Adjektiv ist „konkav“, der Bezug ist die Funktion $f : (a, b) \to \mathbb{R}$. Jetzt allerdings könnte man auch $f : V \to \mathbb{R}$ betrachten, wobei V ein beliebiger Vektorraum über dem Körper $\mathbb{R}$ sein dürfte. Auf diese Verallgemeinerung haben wir fürs Erste verzichtet.

1.5 Die Taxonomie der Lernziele

Die Ziele, die beim Erlernen eines bestimmten Sachverhalts (oder beim Studieren einer Theorie) angestrebt werden sollten, lassen sich in eine Reihenfolge bringen. Diese Reihenfolge lautet (das Fundament ist unten):

7. Kreieren: man ist in der Lage, neue wissenschaftliche Ergebnisse zu erschaffen.

6. Bewerten: man kann wertvolle und belanglose Wissenschaft auseinanderhalten.

5. Synthetisieren: man kann verschiedene Themenfelder miteinander vernetzen.

4. Analysieren: man kann ein größeres Problem in mehrere kleine Probleme zerlegen und dann jedes kleine Problem mit den dafür passenden Methoden bearbeiten.

3. Anwenden: man ist in der Lage, eine verstandene wissenschaftliche Methode auf ein Problem anzuwenden.

2. Verstehen: man hat den wissenschaftlichen Begriff auch verstanden.

1. Kennen: man kann die Definition eines wissenschaftlichen Begriffs aufsagen.

Die jeweils angefügten Erklärungen sollen schlaglichtartig beleuchten, worum es jeweils geht (ein Allgemeindidaktiker würde Alles womöglich etwas anders ausdrücken, aber die hier gewählte Darstellung ist vielleicht verständlicher).

Das im vorigen Abschnitt vorgeführte *Vernetzen* gehört zur Tätigkeit des Synthetisierens. Die Dozierenden an der Universität haben die Erwartung an die Studierenden, ungefähr ab dem vierten oder fünften Semester des Bachelorstudiums aus eigener Kraft heraus Themenfelder vernetzen zu können.

Die Tätigkeit des *Kreierens* wird vorwiegend beim Schreiben einer Doktorarbeit relevant, denn die Inhalte einer Dissertation sollen ja neue Ergebnisse sein. Und vorher sollten Sie sich ein Dissertationsthema suchen, das auch interessant (im Sinne von wertvoll) ist.

Kapitel 2

Quantoren

2.1 Allgemeines

Wir wissen: ein Textfragment der Form $n \geq N_0 \Longrightarrow |a_n - a^*| < \varepsilon$ ist keine Aussage, sondern eine Aussageform, weil noch unbestimmte Bezeichner drinstehen.
Besser wäre ein Textfragment der Form „Sei $(a_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ eine Folge mit Grenzwert a^*. Das bedeutet: wenn $\varepsilon > 0$, dann existiert ein N_0 (das von ε abhängen darf), sodaß gilt: $n \geq N_0 \Longrightarrow |a_n - a^*| < \varepsilon$“.
Die Symbole $\forall$ und $\exists$ können eine Aussageform in eine Aussage überführen, indem damit die noch freien Bezeichner eingeführt werden. Die Symbole $\forall$ und $\exists$ quantifizieren die Anzahl der betreffenden Objekte, deshalb heißen sie *Quantoren*.
Aussagen mit $\forall$ werden geschrieben als

$$\forall\, x \in M\colon \text{Aussageform}(x).$$

Wir lesen den $:$ als „gilt“.
Zu Beginn des Studiums ist es sehr empfehlenswert, die Passage „$\in M$“ immer mit zu erwähnen.
Gelesen wird das $\forall$ als **für jedes** (Singular) bzw. als **für jegliche** (Plural). Die oft gehörte Leseweise „für alle“ ist mathematischer Unfug, denn die Worte *jeder* und *alle* sind keine Synonyme voneinander, wie man sich veranschaulichen kann am Satz

Bei der Siegerehrung der Fußballweltmeisterschaft erhält die siegreiche Mannschaft einen Pokal für alle und eine Medaille für jeden.

Die Aussage „$\forall x \in M\colon A(x)$“ (wobei $A(x)$ eine Aussageform ist) beweist man so:

- setze voraus, daß jemand uns ein $x \in M$ gibt. Wir wissen nicht, welches x wir bekommen werden.
- Beweise die Aussage $A(x)$. Hierbei darf das x von uns nicht gewählt werden (denn es wird ja von einer äußeren Person vorgegeben).

Ein Beispiel einer Allquantoraussage ist das Kommutativgesetz der Addition in $\mathbb{N}_+$:

$$\forall a \in \mathbb{N}_+\colon \forall b \in \mathbb{N}_+\colon a + b = b + a.$$

Man kann es auch schreiben als

$$\forall a, b \in \mathbb{N}_+\colon a + b = b + a.$$

Wir kommen zum Existenzquantor $\exists$. Aussagen damit werden geschrieben als

$$\exists\, x \in M\colon \text{Aussageform}(x).$$

Wir lesen den $:$ als „sodaß“.

Aussagen dieser Form können auf zwei Wegen bewiesen werden:

Weg 1: wir benennen ein konkretes x und erbringen dann den Nachweis, daß für dieses benannte x die Aussage $A(x)$ wahr ist. Im einfachsten Fall hat man lediglich eine Probe zu verifizieren. Wir sind nicht verpflichtet, alle tauglichen x zu benennen; ein taugliches x reicht.

Weg 2: man gibt abstrakte Begründungen an, die erzwingen, daß ein gewünschtes x existieren muß.

Ein Beispiel für den zweiten Weg ist der Beweis für folgenden Satz:

Satz 2.1. *Sei $f\colon [0,1] \to \mathbb{R}$ eine stetige Funktion mit $f(0) < 0$ und $f(1) > 0$. Dann besitzt f eine Nullstelle im Intervall $(0,1)$.*

Klarer wird vielleicht die konsequente Quantorenschreibweise:
Sei $\mathcal{F}([0,1] \to \mathbb{R})$ die Menge aller Funktionen von $[0,1]$ nach $\mathbb{R}$. Dann ist zu beweisen:

$$\forall\, f \in \{g \in \mathcal{F}([0,1] \to \mathbb{R})\colon\; g(0) < 0,\; g(1) > 0,\; g \text{ stetig}\} : \exists\, x_0 \in (0,1)\colon f(x_0) = 0.$$

Wegen des $\forall$ am Zeilenanfang haben wir die Vorstellung, daß jemand anders uns das f vorgibt. Von diesem f ist uns praktisch nichts bekannt. Dann ist der Weg 1 uns aber versperrt, denn wir sehen uns nicht in der Lage, das gewünschte x_0 explizit anzugeben, eben weil wir vom f nur sehr wenig wissen, und weil wir das f nicht wählen können.

Eine häufige Variation der Schreibweise ist diese: das relevante Element, dessen Existenz durch das Symbol $\exists$ verkündet wird, ist ein spezielles Element, ein besonderes Element, also ein ausgezeichnetes Element, was man dadurch symbolisiert, daß es mit einer *Medaille* behängt wird. Damit wird die übliche Schreibweise also zu

$$\exists\, x_0 \in M\colon \text{Aussageform}(x_0).$$

Wir überlegen uns nun Verneinungen zu Quantoraussagen anhand von Beispielen.
Sei $\mathcal{K}$ die Menge aller Katzen. Die Aussage

$$\forall\, K \in \mathcal{K}\colon K \text{ ist grau.}$$

soll verneint werden. Zunächst beobachten wir

$$\begin{aligned} &\forall\, K \in \mathcal{K}\colon K \text{ ist grau} \\ \iff\quad & \mathcal{K} = \{K \in \mathcal{K}\colon K \text{ ist grau}\} \\ \iff\quad & \mathcal{K} \setminus \{K \in \mathcal{K}\colon K \text{ ist grau}\} = \emptyset. \end{aligned}$$

Also ist auch

$$\begin{aligned} &\text{nicht}(\forall\, K \in \mathcal{K}\colon K \text{ ist grau}) \\ \iff\quad & \mathcal{K} \setminus \{K \in \mathcal{K}\colon K \text{ ist grau}\} \neq \emptyset \\ \iff\quad & \exists\, K_0\colon K_0 \in \mathcal{K} \setminus \{K \in \mathcal{K}\colon K \text{ ist grau}\}, \end{aligned}$$

denn eine Menge ist genau dann eine nichtleere Menge, wenn sie mindestens ein Element besitzt. Insgesamt bekommen wir die Aussagenäquivalenz

$$\text{nicht}(\forall\, K \in \mathcal{K}\colon K \text{ ist grau}) \quad\iff\quad \exists\, K_0 \in \mathcal{K}\colon K_0 \text{ ist nicht grau.}$$

Wir kommen zur (richtigen !) Vermutung: wenn M eine Menge ist und $A(x)$ eine Aussagenform, dann gilt

$$\text{nicht}\,(\forall\, x \in M\colon A(x)) \quad\iff\quad \exists\, x \in M\colon (\text{nicht}(A))(x).$$

Jetzt kommen wir zur Verneinung einer Existenzaussage.
Sei $\mathbb{P}$ die Menge aller Primzahlen. Die Aussage

$$\exists\, p_0 \in \mathbb{P}\colon 4|p_0$$

soll verneint werden. Die Verneinung der Aussage $\{p_0 \in \mathbb{P}\colon 4|p_0\} \neq \emptyset$ ist aber $\{p_0 \in \mathbb{P}\colon 4|p_0\} = \emptyset$, was seinerseits äquivalent ist zu

$$\mathbb{P} \setminus \{p_0 \in \mathbb{P}\colon 4|p_0\} = \mathbb{P},$$

und das bedeutet: $\forall\, p \in \mathbb{P}\colon 4 \nmid p$.
Wir kommen zur (richtigen !) Vermutung: wenn M eine Menge ist und $A(x)$ eine Aussagenform, dann gilt

$$\text{nicht}\,(\exists\, x \in M\colon A(x)) \quad \Longleftrightarrow \quad \forall\, x \in M\colon (\text{nicht}(A))(x).$$

Das allgemeine Verneinungsprinzip einer aus $\forall$ und $\exists$ zusammengesetzten Aussage lautet: man gehe von links nach rechts vor, tausche $\forall$ gegen $\exists$ aus (und umgekehrt), und dann verneine man jedesmal den Teil hinter dem : .

2.2 Stetigkeit und Grenzwerte

Wir wollen definieren, was es für eine Funktion $f\colon (a,b) \to \mathbb{R}$ bedeutet, stetig zu sein.
Ein schlechter Definitionsversuch besteht darin zu sagen, daß eine solche Funktion genau dann stetig sei, wenn man den Graphen von links nach rechts ohne absetzen durchzeichnen könne.
Es gibt zwei Gründe, warum dieser Definitionsversuch nicht weit führt: zum einen kann man nicht mathematisch fassen, was denn nun „durchzeichnen“ sein soll; zum anderen ist dieser Zugang nicht tragfähig, wenn man die Stetigkeit von Funktionen $f\colon \Omega \to \mathbb{R}^n$ definieren möchte, wobei Ω eine Teilmenge des $\mathbb{R}^m$ ist, denn weder im $\mathbb{R}^n$ noch im $\mathbb{R}^m$ gibt es links und rechts (es sei denn $n = m = 1$).
Wir gehen beim Erarbeiten der Stetigkeitsdefinition schrittweise vor:
Sei $x_0 \in \mathbb{R}$ irgendeine feste Zahl, und sei $f\colon \mathbb{R} \to \mathbb{R}$ die Funktion mit der Gleichung

$$f(x) = \begin{cases} 1.7 & : x \leq x_0 \\ 2.7 + (x - x_0) & : x > x_0. \end{cases}$$

Diese Funktion vollzieht einen Sprung der Höhe 1 an der Stelle x_0, und wir möchten diese Funktion als unstetig ansehen. Bei dieser Funktion beobachten wir: egal wie nah wir eine Zahl x von rechts an x_0 heranschieben, stets ist $|f(x) - f(x_0)| \geq 1$. Diese Beobachtung formulieren wir wie folgt:

$$\forall\, \delta \in \mathbb{R}_+\colon \exists\, x \in (x_0 - \delta, x_0 + \delta)\colon |f(x) - f(x_0)| \geq 1.$$

Die Verneinung davon ergibt eine Bedingung, die den konkreten Sprung der Höhe 1 verbietet, und diese Verneinung ist dann:

$$\exists\, \delta \in \mathbb{R}_+\colon \forall\, x \in (x_0 - \delta, x_0 + \delta)\colon |f(x) - f(x_0)| < 1.$$

Wenn eine Funktion f an der Stelle x_0 stetig sein will, dann darf sie aber dort auch keinen Sprung der Höhe $\frac{1}{3}$ haben, also muß sie dann folgende Bedingung erfüllen:

$$\exists\, \delta \in \mathbb{R}_+\colon \forall\, x \in (x_0 - \delta, x_0 + \delta)\colon |f(x) - f(x_0)| < \frac{1}{3}.$$

Und alle anderen Sprunghöhen (seien sie auch noch so klein) wollen wir ebenfalls verbieten, womit die folgende Definition motiviert wird:

Definition 2.2 (Stetigkeit in einem Punkt). *Sei $f\colon (a,b) \to \mathbb{R}$ eine Funktion und $x_0 \in (a,b)$. Wir sagen, daß f* stetig im Punkt x_0 *ist, wenn gilt:*

$$\forall\, \varepsilon \in \mathbb{R}_+\colon \exists\, \delta \in \mathbb{R}_+\colon \forall\, x \in (x_0 - \delta, x_0 + \delta) \cap (a,b)\colon |f(x) - f(x_0)| < \varepsilon.$$

Definition 2.3 (Stetigkeit auf einem Intervall). *Sei $f\colon (a,b) \to \mathbb{R}$ eine Funktion. Wir sagen, daß f* auf dem Intervall (a,b) stetig *ist, wenn gilt:*

$$\forall\, x_0 \in (a,b)\colon f \text{ ist stetig in } x_0.$$

Satz 2.4. *Sei $f\colon (a,b) \to \mathbb{R}$ konvex. Dann ist f stetig auf dem Intervall (a,b).*

Die Erläuterungen, wie man auf den Beweis kommt, gab es im Tutorium.

Beweis. Sei $f\colon (a,b) \to \mathbb{R}$ konvex. Zu zeigen ist:

$$\forall\, x_0 \in (a,b)\colon f \text{ ist stetig in } x_0.$$

Sei also $x_0 \in (a,b)$ gegeben[1]. Zu zeigen ist, daß f stetig in x_0 ist, und dies bedeutet, das wir folgendes zu zeigen haben:

$$\forall\, \varepsilon \in \mathbb{R}_+\colon \exists\, \delta \in \mathbb{R}_+\colon \forall\, x \in (x_0-\delta, x_0+\delta) \cap (a,b)\colon |f(x)-f(x_0)| < \varepsilon.$$

Sei jetzt auch noch zusätzlich $\varepsilon \in \mathbb{R}_+$ gegeben. Zu zeigen ist:

$$\exists\, \delta \in \mathbb{R}_+\colon \forall\, x \in (x_0-\delta, x_0+\delta) \cap (a,b)\colon |f(x)-f(x_0)| < \varepsilon.$$

Es reicht, wenn wir ein konkretes $\delta \in \mathbb{R}_+$ angeben.[2] Zu diesem Zwecke drücken wir die vorausgesetzte Konvexität der Funktion f aus entsprechend ihrer Definition:

$$\forall\, x_1, x_2 \in (a,b)\colon \forall w_1, w_2 \in [0,1] \text{ mit } w_1+w_2 = 1\colon f(w_1x_1+w_2x_2) \le w_1 f(x_1) + w_2 f(x_2). \tag{2.1}$$

Wegen $x_0 \in (a,b)$ ist $a < x_0 < b$; insbesondere ist $x_0 \neq a$ und $x_0 \neq b$. Wir definieren

$$x_l := \frac{a+x_0}{2}, \qquad x_r := \frac{x_0+b}{2},$$

und dann gilt die Ungleichungskette $a < x_l < x_0 < x_r < b$.
Wir betrachten ein beliebiges $x \in (x_l, x_0)$.
Wir wenden (2.1) an mit $x_1 := x_l$, $x_2 := x_0$, und w_1, w_2 wählen wir so, daß $w_1x_1 + w_2x_2 \overset{!}{=} x$ wird. Es ergibt sich $w_1 = \frac{x_0-x}{x_0-x_l}$ und $w_2 = \frac{x-x_l}{x_0-x_l}$. Aus (2.1) schlußfolgern wir somit

$$f(x) \le \frac{x_0-x}{x_0-x_l} f(x_l) + \frac{x-x_l}{x_0-x_l} f(x_0). \tag{2.2}$$

Wir wenden (2.1) an mit $x_1 := x$, $x_2 := x_r$, und w_1, w_2 wählen wir so, daß $w_1x_1 + w_2x_2 \overset{!}{=} x_0$ wird. Es ergibt sich $w_1 = \frac{x_r-x_0}{x_r-x}$ und $w_2 = \frac{x_0-x}{x_r-x}$. Aus (2.1) schlußfolgern wir somit

$$f(x_0) \le \frac{x_r-x_0}{x_r-x} f(x) + \frac{x_0-x}{x_r-x} f(x_r). \tag{2.3}$$

Wir formen (2.2) und (2.3) um zu

$$\begin{aligned} f(x) &\le f(x_0) + \frac{x_0-x}{x_0-x_l} \cdot (f(x_l) - f(x_0)), \\ \frac{x_r-x}{x_r-x_0} \cdot f(x_0) &\le f(x) + \frac{x_0-x}{x_r-x_0} f(x_r), \end{aligned}$$

denn die Multiplikation einer Ungleichung mit einem positiven Faktor ist eine äquivalente Umformung. Insgesamt ergibt sich also

$$\begin{aligned} f(x) &\le f(x_0) + \frac{x_0-x}{x_0-x_l} \cdot (f(x_l) - f(x_0)), \\ f(x_0) + \frac{x_0-x}{x_r-x_0} \cdot (f(x_0) - f(x_r)) &\le f(x), \end{aligned}$$

und daraus folgt

$$(x_0-x) \cdot \frac{f(x_0)-f(x_r)}{x_r-x_0} \le f(x) - f(x_0) \le (x_0-x) \cdot \frac{f(x_l)-f(x_0)}{x_0-x_l}.$$

[1] gemeint ist damit: sei x_0 uns von außen vorgegeben. Wir dürfen x_0 nicht mehr wählen

[2] es ist logisch erlaubt, bei der Wahl dieses konkreten δ auch unsere persönlichen Schönheitswünsche zu berücksichtigen.

Die Konsequenz daraus ist dann

$$|f(x) - f(x_0)| \leq |x_0 - x| \cdot \max \left\{ \left| \frac{f(x_0) - f(x_r)}{x_r - x_0} \right|, \left| \frac{f(x_l) - f(x_0)}{x_0 - x_l} \right| \right\},$$

was wir abkürzenderweise[3] schreiben als $|f(x) - f(x_0)| \leq |x_0 - x| \cdot M$.
Unsere Aufgabe ist es sicherzustellen, daß $|f(x) - f(x_0)| < \varepsilon$ ist, wobei $\varepsilon \in \mathbb{R}_+$ uns von außen vorgegeben wurde. Zur Verfügung und freien Wahl haben wir noch ein $\delta \in \mathbb{R}_+$. Wir erhalten als Bedingung an δ, daß $0 < \delta < \frac{\varepsilon}{M}$ erfüllt sein möge.
Falls also $0 < \delta < \frac{\varepsilon}{M}$ und $x_l < x < x_0$ und $|x - x_0| < \delta$, dann folgt

$$|f(x) - f(x_0)| \leq |x - x_0| \cdot M < \delta \cdot M < \frac{\varepsilon}{M} \cdot M = \varepsilon,$$

wie gewünscht.
Es bleibt lediglich noch der Fall $x_0 < x < x_r$ zu betrachten, und dann verlaufen die Rechnungen sehr analog (Übungsaufgabe im Selbststudium).
Wir erkennen, daß das nach unseren Vorlieben wählbare δ noch die beiden Bedingungen $\delta < x_0 - x_l$ und $\delta < x_r - x_0$ erfüllen sollte. Wenn dies gilt, dann erzwingt die Ungleichung $|x - x_0| < \delta$ auch $x \in (x_l, x_r)$, was wir in unseren Betrachtungen vorausgesetzt hatten. □

Bemerkung 2.5. *Die Aussage „wenn $f\colon [a,b] \to \mathbb{R}$ konvex ist, dann ist $f\colon [a,b] \to \mathbb{R}$ auch stetig“ ist übrigens falsch, wie man sich anhand der Funktion $f\colon [0,1] \to \mathbb{R}$ mit*

$$f(x) = \begin{cases} 1 & : 0 \leq x < 1, \\ 13 & : x = 1 \end{cases}$$

überlegt.

Wir kommen nun zu Grenzwerten.

Definition 2.6 (Grenzwert einer Folge in $\mathbb{R}$). *Wir sagen, daß eine Folge $(a_n)_{n \in \mathbb{N}} \subset \mathbb{R}$ den Grenzwert $a^* \in \mathbb{R}$ besitzt, wenn gilt:*

$$\forall\, \varepsilon \in \mathbb{R}_+ \colon \exists\, N_0(\varepsilon) \in \mathbb{N}_+ \colon \forall\, n \geq N_0(\varepsilon) \colon |a_n - a^*| < \varepsilon.$$

Eigentlich hätte man beim zweiten $\forall$ noch $n \in \mathbb{N}$ einfügen müssen (damit unsere syntaktischen Regeln eingehalten werden, die den Verweis auf eine Menge erfordern), aber wir trauen den Lesern zu, diese Offensichtlichkeit mental selbst zu ergänzen.
Gute Verbalisierungen sind:

- a_n kommt mit wachsendem n der Zahl a^* beliebig nah
- a_n strebt gegen a^* für n gegen ∞,
- in jeder noch so kleinen Umgebung von a^* sind alle Folgenglieder bis auf endlich viele.

Für die Wendung *alle bis auf endlich viele* sagen wir auch *fast alle*. Diese Vereinbarung gilt in der gesamten Mathematik.
Eine **falsche Verbalisierung** ist:

- a_n kommt mit wachsendem n der Zahl a^* immer näher.

Denn die Glieder der Folge $(\frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots)$ kommen auch der Zahl -17 immer näher.
Eine **vollkommen falsche Verbalisierung** ist:

- a_n kommt der Zahl a^* immer näher, ohne sie jemals zu erreichen.

[3] es sei also M definiert als der Term $\max\{\dots\}$.

Der schlimmste Fehler steht hinter dem Komma.
Weiterhin ist zu beachten: der Grenzwertbegriff verlangt nicht, daß für jeden Index das Folgenglied a_{n+1} näher am a^* liegt als a_n. Wenn $\lim_{n\to\infty} a_n = a^*$, dann kann für manche Indizes n durchaus $|a_{n+1} - a^*| > |a_n - a^*|$ gelten.
Die Schreibweise $\lim_{n\to\infty} a_n = a^*$ trägt zwei Aussagen in sich:

- der Grenzwert existiert,
- er hat den Wert a^*.

Satz 2.7. *Sei* $f\colon (a,b) \to \mathbb{R}$ *eine Funktion und* $x^* \in (a,b)$.
Dann gilt: f *ist stetig in* x^* *genau dann, wenn gilt: für jede Folge* $(x_n)_{n\in\mathbb{N}} \subset (a,b)$ *mit* $\lim_{n\to\infty} x_n = x^*$ *gilt* $\lim_{n\to\infty} f(x_n) = f(x^*)$.

Beweis. **Zeige „$\Longrightarrow$“:**
Voraussetzung ist: f stetig in x^*, d.h.

$$\forall\, \varepsilon \in \mathbb{R}_+\colon \exists\, \delta \in \mathbb{R}_+\colon \forall\, x \in (x^* - \delta, x^* + \delta) \cap (a,b)\colon |f(x) - f(x^*)| < \varepsilon.$$

Behauptung ist: $\forall\, (x_n)_{n\in\mathbb{N}} \subset (a,b)$ mit $\lim_{n\to\infty} x_n = x^*\colon \quad \lim_{n\to\infty} f(x_n) = f(x^*)$.
Sei eine solche Folge $(x_n)_{n\in\mathbb{N}}$ vorgegeben (extern, von uns nicht wählbar). Das heißt, wir wissen:

$$\forall\, \varepsilon \in \mathbb{R}_+\colon \exists\, N_0(\varepsilon) \in \mathbb{N}\colon \forall\, n \geq N_0(\varepsilon)\colon |x_n - x^*| < \varepsilon.$$

Unser Ziel ist: $\lim_{n\to\infty} f(x_n) = f(x^*)$, das heißt:

$$\forall\, \gamma \in \mathbb{R}_+\colon \exists\, N_1(\gamma) \in \mathbb{N}\colon \forall\, n \geq N_1(\gamma)\colon |f(x_n) - f(x^*)| < \gamma.$$

Sei also $\gamma \in \mathbb{R}_+$ vorgegeben (extern, nicht wählbar). Wir suchen $N_1(\gamma)$.
Lt. Vor. ist f stetig in x^*.
Nimm $\varepsilon := \gamma$. Wegen Stetigkeit von f existiert ein $\delta = \delta(\varepsilon)$ mit

$$|x - x^*| < \delta \quad x \in (a,b) \Longrightarrow |f(x) - f(x^*)| < \gamma.$$

Nimm dieses δ.
Laut Vorgabe[4] ist $(x_n)_{n\in\mathbb{N}}$ konvergent gegen x^*. Das heißt, wir dürfen benutzen:

$$\forall\, \varrho \in \mathbb{R}_+\colon \exists\, N_0(\varrho) \in \mathbb{N}\colon \forall\, n \geq N_0(\varrho)\colon |x_n - x^*| < \varrho.$$

Nimm $\varrho := \delta$. Dann existiert $N_0(\varrho)$ mit der angegebenen Eigenschaft.
Setze $N_1(\gamma) := N_0(\varrho) = N_0(\delta)$.
Dieses $N_1(\gamma)$ erfüllt die gewünschten Bedingungen.
Zeige „$\Longleftarrow$“:
Voraussetzung ist: wenn $(x_n)_{n\in\mathbb{N}} \subset (a,b)$ und $\lim_{n\to\infty} x_n = x^*$ ist, dann gilt $\lim_{n\to\infty} f(x_n) = f(x^*)$.
Behauptung ist: f ist stetig in x^*. Das heißt, wir wollen zeigen:

$$\forall\, \varepsilon \in \mathbb{R}_+\colon \exists\, \delta \in \mathbb{R}_+\colon \forall\, x \in (x^* - \delta, x^* + \delta) \cap (a,b)\colon |f(x) - f(x^*)| < \varepsilon.$$

Sei also $\varepsilon \in \mathbb{R}_+$ (extern) vorgegeben. Wir suchen ein zugehöriges $\delta \in \mathbb{R}_+$.
Wir führen den Beweis der Existenz dieses δ **indirekt**. Sei also die Aussage

$$\exists\, \delta \in \mathbb{R}_+\colon \forall\, x \in (x^* - \delta, x^* + \delta) \cap (a,b)\colon |f(x) - f(x^*)| < \varepsilon \qquad \text{falsch.}$$

Wir wissen dann:

$$\forall\, \delta \in \mathbb{R}_+\colon \exists\, x_\delta \in (x^* - \delta, x^* + \delta) \cap (a,b)\colon |f(x_\delta) - f(x^*)| \geq \varepsilon.$$

Wir nehmen eine Folge $(\delta_1, \delta_2, \delta_3, \dots) \subset \mathbb{R}_+$ mit $\delta_j := \frac{1}{j}$ für jedes j.
Dann erhalten wir eine Folge $(x_{\delta_1}, x_{\delta_2}, \dots)$ mit $x_{\delta_j} \in (x^* - \frac{1}{j}, x^* + \frac{1}{j}) \cap (a,b)$ und $|f(x_{\delta_j}) - f(x^*)| \geq \varepsilon$.
Es ist $\lim_{j\to\infty} x_{\delta_j} = x^*$, denn $|x_{\delta_j} - x^*| \leq \frac{1}{j}$ für jedes j.
Wenn $\lim_{j\to\infty} f(x_{\delta_j})$ existieren sollte, dann ist er ungleich $f(x^*)$ denn $|f(x_{\delta_j}) - f(x^*)| \geq \varepsilon$. Oder der Grenzwert $\lim_{j\to\infty} f(x_{\delta_j})$ existiert gar nicht.
In jedem der beiden Fälle entsteht ein Widerspruch zur Voraussetzung. □

[4] Nicht „Laut Voraussetzung“

Anmerkungen zu Beweisstrategien: Es sei ein Satz zu beweisen, bei dem aus Voraussetzungen $V_1, \ldots, V_N$ eine Behauptung B geschlußfolgert werden soll.

Strategie 1: setze $V_1, \ldots, V_N$ voraus. Beweise B. Das nennt man einen *direkten Beweis.*

Strategie 2: setze $V_1, \ldots, V_N$ und nicht(B) voraus. Beweise irgendeinen Widerspruch. Dies nennt man einen *indirekten Beweis.*

Der indirekte Beweis (den wir im zweiten Teil verwendet hatten) hat den technischen Vorteil, daß man eine Voraussetzung mehr zur Verfügung hat.
Sei $A = A(x)$ eine Aussagenform und X eine Menge.

Wie beweist man $\forall\, x \in X : A(x)$ **?**

- Es sei ein $x \in X$ vorgegeben (wir stellen uns vor, daß jemand anders uns das x gibt), das für uns nicht wählbar ist. Wir beweisen dann für dieses x die Aussage $A(x)$.
- Es sei ein $x \in X$ vorgegeben, das für uns nicht wählbar ist. Wir setzen $(\text{nicht}(A))(x)$ voraus. Wir beweisen irgendeinen Widerspruch.
- Wir setzen voraus: $\exists\, x_0 \in X : (A(x_0)$ ist falsch$)$. Wir beweisen irgendeinen Widerspruch.

Wie beweist man $\exists\, x \in X : A(x)$ **?**

- Wir benennen ein konkretes $x \in X$. Für dieses x beweisen wir dann die Aussage $A(x)$.
- Wir setzen voraus: $\forall\, x \in X : (A(x)$ ist falsch$)$. Wir beweisen irgendeinen Widerspruch.

Wie benutzt man $\forall\, x \in X : A(x)$ **?**

- Wir wählen uns ein Lieblings–x aus X. Für dieses x dürfen wir die Aussage $A(x)$ dann verwenden. Dies ist der Modus Ponendo Ponens.

Wie benutzt man $\exists\, x \in X : A(x)$ **?**

- Es ist uns ein spezielles $x \in X$ vorgegeben, und wir haben die Garantie, daß für dieses x die Aussage $A(x)$ wahr ist. Wir dürfen dieses x nicht wählen. Vom x ist praktisch nichts bekannt außer daß $x \in X$ ist und $A(x)$ wahr ist.

Definition 2.8 (**Grenzwert einer Funktion in einem Punkt**)**.** *Sei* $f : (a, b) \to \mathbb{R}$ *eine Funktion und* $x^* \in [a, b]$*. Wir sagen, daß* $\lim_{x\to x^*} f(x) = y^*$ *genau dann, wenn:*

$$\forall\, (x_n)_{n\in\mathbb{N}} \subset (a, b) \setminus \{x^*\} \text{ mit } \lim_{n\to\infty} x_n = x^* : \lim_{n\to\infty} f(x_n) = y^*.$$

Man beachte, daß x^* auch gleich a oder b sein darf. Dort ist aber f nicht definiert, was absichtlich erlaubt wurde.

Satz 2.9. *Sei* $f : (a, b) \to \mathbb{R}$ *eine Funktion und* $x^* \in [a, b]$*. Dann gilt:* f *ist stetig in* x^* *genau dann, wenn folgende drei Bedingungen erfüllt sind:*

- $f(x^*)$ *existiert,*
- $\lim_{x\to x^*} f(x)$ *existiert,*
- *beide sind gleich.*

Die Bedingung im ersten • bedeutet lediglich, daß x^* im Definitionsbereich von f liegen muß, also $x^* \in (a, b)$.
Der Beweis ist eine kurze Kombination von Satz 2.7 und Definition 2.8.

Bemerkung 2.10. *Wir erhalten somit eine zweite Möglichkeit, die Stetigkeit einer Funktion in einem Punkt* x^* *zu definieren. Nämlich, indem man die drei • fordert. Für die logische Korrektheit der gesamten Theorie ist dabei zu beachten, daß es verboten ist, einen Begriff mehrfach zu definieren. Die Entscheidung für eine der beiden Definitionsmöglichkeiten trifft man typischerweise anhand von Geschmacksvorlieben.*

Satz 2.11. *Sei $f\colon (a,b) \to \mathbb{R}$ eine Funktion und $x^* \in [a,b]$ sowie $y^* \in \mathbb{R}$. Es gilt $y^* = \lim_{x\to x^*} f(x)$ genau dann, wenn:*

$$\forall\, \varepsilon \in \mathbb{R}_+ \colon \exists\, \delta \in \mathbb{R}_+ \colon \forall\, x \in (x^* - \delta, x^* + \delta) \cap (a,b) \setminus \{x^*\} \colon |f(x) - y^*| < \varepsilon.$$

Der Beweis sieht sehr ähnlich aus wie der Beweis zu Satz 2.7, weshalb wir ihn dem Selbststudium überlassen wollen. Die abgesetzte Formelzeile $\forall\, \varepsilon \in \mathbb{R}_+ \ldots < \varepsilon$ kann man auch verwenden, um eine zweite Definition des Ausdrucks $\lim_{x\to x^*} f(x)$ festzulegen. Dann müßte man natürlich Definition 2.8 wieder durchstreichen !

Definition 2.12 (Grenzwert einer Funktion an einer Definitionslücke). *Sei (a,b) ein Intervall und $x^* \in (a,b)$. Sei eine Funktion $f\colon (a,b)\setminus\{x^*\} \to \mathbb{R}$ gegeben. Wir sagen, daß $\lim_{x\to x^*} f(x) = y^*$ genau dann, wenn:*

$$\forall\, (x_n)_{n\in\mathbb{N}} \subset (a,b)\setminus\{x^*\} \text{ mit } \lim_{n\to\infty} x_n = x^* \colon \lim_{n\to\infty} f(x_n) = y^*.$$

Zusammen mit Definition 2.8 ist dies *keine* verbotene Doppeldefinition, denn auf eine gegebene Funktion ist immer nur eine der beiden Definitionen überhaupt anwendbar. Entweder der Definitionsbereich der Funktion f enthält den Punkt x^* (dann kann Definition 2.12 nicht angewandt werden), oder der Definitionsbereich der Funktion f enthält den Punkt x^* nicht, dann ist Definition 2.8 nicht anwendbar.
Hierbei haben wir folgende Vereinbarung verwendet: wenn man bei einer Funktion den Definitionsbereich abändert (indem man z.B. einen Punkt aus dem Definitionsintervall herausstanzt), dann entsteht eine neue Funktion.
Wichtig wird Definition 2.12 beim Definieren der Ableitung:

Definition 2.13 (Ableitung einer Funktion g; Differenzierbarkeit in einem Punkt). *Sei $g\colon (a,b) \to \mathbb{R}$ eine Funktion und $x^* \in (a,b)$.*
Setze $f(x) := \frac{f(x)-f(x^*)}{x-x^*}$ für $x \in (a,b)\setminus\{x^*\}$.
Dann definieren wir $g'(x^*) := \lim_{x\to x^*} f(x)$, falls es diesen Limes gibt, und in diesem Fall sagen wir, daß die Funktion g im Punkt x^* differenzierbar ist. Ansonsten ist $g'(x^*)$ undefiniert.

Es folgt ein Satz zum Üben.

Satz 2.14. *Sei $f\colon (a,b) \to \mathbb{R}$ konvex und auf ganz (a,b) differenzierbar. Dann ist die Ableitung $f'\colon (a,b) \to \mathbb{R}$ eine monoton wachsende Funktion.*

Beweis. Die zu zeigende Monotonie der Funktion f' bedeutet:

$$\forall\, x_l, r_r \in (a,b)\colon \text{ wenn } x_l < x_r \text{ dann } f'(x_l) \le f'(x_r).$$

Seien also x_l, x_r (extern) vorgegeben mit $a < x_l < x_r < b$. Wir wollen zeigen, daß dann gilt $f'(x_l) \le f'(x_r)$. Wegen der vorausgesetzten Differenzierbarkeit der Funktion f ist die Existenz der beiden Ableitungswerte gesichert. Dann liefert uns Definition 2.13 in Verbindung mit Definition 2.12 folgende Erkenntnis:

$$\forall\, (x_{l,n})_{n\in\mathbb{N}} \subset (a,b)\setminus\{x_l\} \text{ mit } \lim_{n\to\infty} x_{l,n} = x_l \colon \lim_{n\to\infty} \frac{f(x_{l,n}) - f(x_l)}{x_{l,n} - x_l} = f'(x_l),$$
$$\forall\, (x_{r,n})_{n\in\mathbb{N}} \subset (a,b)\setminus\{x_r\} \text{ mit } \lim_{n\to\infty} x_{r,n} = x_r \colon \lim_{n\to\infty} \frac{f(x_{r,n}) - f(x_r)}{x_{r,n} - x_r} = f'(x_r).$$

Diese beiden Zeilen wollen wir benutzen, und dafür wählen wir spezielle Folgen $(x_{l,n})_{n\in\mathbb{N}}$ und $(x_{r,n})_{n\in\mathbb{N}}$; nämlich solche, die streng monoton von links bzw. rechts gegen x_l bzw. x_r konvergieren:

$$\forall\, n \in \mathbb{N}\colon a < x_{l,n} < x_l, \qquad \lim_{n\to\infty} x_{l,n} = x_l,$$
$$\forall\, n \in \mathbb{N}\colon x_r < x_{r,n} < b, \qquad \lim_{n\to\infty} x_{r,n} = x_r.$$

Mit diesen gewählten Folgen haben wir nun folgendes Beweisziel:

$$\lim_{n\to\infty} \frac{f(x_{l,n}) - f(x_l)}{x_{l,n} - x_l} \le \lim_{n\to\infty} \frac{f(x_{r,n}) - f(x_r)}{x_{r,n} - x_r} = \lim_{n\to\infty} \frac{f(x_r) - f(x_{r,n})}{x_r - x_{r,n}}.$$

Laut Voraussetzung ist f eine konvexe Funktion. Daraus folgern wir: wenn p, q, r Zahlen sind mit $a < p < q < r < b$, dann liegt der Punkt $(q, f(q))^\top$ unterhalb der Strecke durch die beiden Punkte $(p, f(p))^\top$ und $(r, f(r))^\top$, und daraus folgert man schnell, daß

$$\frac{f(p) - f(q)}{p - q} \leq \frac{f(q) - f(r)}{q - r}.$$

Gemäß unserer Konstruktion ist $a < x_{l,n} < x_l < x_r < x_{r,n} < b$. Mit $(p, q, r) = (x_{l,n}, x_l, x_r)$ folgt

$$\frac{f(x_{l,n}) - f(x_l)}{x_{l,n} - x_l} \leq \frac{f(x_l) - f(x_r)}{x_l - x_r},$$

und die Wahl $(p, q, r) = (x_l, x_r, x_{r,n})$ liefert uns

$$\frac{f(x_l) - f(x_r)}{x_l - x_r} \leq \frac{f(x_r) - f(x_{r,n})}{x_r - x_{r,n}}.$$

In Kombination entsteht dann also

$$\frac{f(x_{l,n}) - f(x_l)}{x_{l,n} - x_l} \leq \frac{f(x_r) - f(x_{r,n})}{x_r - x_{r,n}}.$$

Wir können den Grenzübergang $\lim_{m\to\infty}$ vollziehen (denn die Limites existieren) und erhalten somit die Behauptung. □

Es gilt auch die Umkehrung des Satzes, nämlich: sei $f\colon (a, b) \to \mathbb{R}$ auf ganz (a, b) differenzierbar, und die Ableitung f' sei eine monoton wachsende Funktion. Dann ist f konvex. (Ein Beweis dieses Satzes könnte im Selbststudium indirekt über den Mittelwertsatz der Differentialrechnung geführt werden.)
Weiterhin gilt: sei $f\colon (a, b) \to \mathbb{R}$ eine Funktion, die auf ganz (a, b) zweimal differenzierbar ist. Dann ist f genau dann konvex auf (a, b), wenn überall $f''(x) \geq 0$ gilt.
Als Anwendung dieses Kriteriums schaue man sich nochmal die Ungleichung von Jensen an.

Kapitel 3

Beispiele

Man versteht die Mathematik mit ihren abstrakten Konzepten besser, wenn man viele konkrete Beispiele zur Verfügung hat, um daran dann die Vorstellungskraft anzuhängen. Sehr oft geht es in der Mathematik um zwei Typen von Dingen:

- algebraische Strukturen,
- Abbildungen zwischen diesen Strukturen.

Dafür wollen wir soviel wie möglich aussagekräftige Beispiele zusammentragen.

3.1 Algebraische Strukturen

Sehr oft besteht die Definition einer algebraischen Struktur aus drei Zutaten:

- eine Menge,
- eine oder mehrere Operationen, die auf diese Menge wirken,
- Spielregeln für diese Operationen.

Eine **Halbgruppe** besteht aus einer nichtleeren Menge G und einer zweiargumentigen Operation $\circ\colon G \times G \to G$ darauf. Die Spielregeln sind zwei Stück: es existiert genau ein neutrales Element $e \in G$, das heißt $e \circ x = x \circ e = x$ für jedes $x \in G$. Und $\circ$ ist assoziativ.
Unsere Interpretation ist: eine Halbgruppe hat eine Operation, nämlich $\circ$. Das Standardbeispiel ist $(G, \circ) = (N_0, +)$ mit $e = 0$. Man kann auch $(G, \circ) = (\mathbb{N}_0, \cdot)$ mit $e = 1$ nehmen.

Eine **Gruppe** ist eine Halbgruppe, in der jede Gleichung $a \circ x = b$ und jede Gleichung $y \circ a = b$ lösbar ist (bei gegebenen $a, b \in G$ und gesuchten $x, y \in G$).
Unsere Interpretation ist: eine Halbgruppe hat zwei Operationen, von denen die eine die Umkehrung der anderen ist. Das Standardbeispiel ist $(G, \circ) = (\mathbb{Z}, +)$.

Bemerkung 3.1. *Oft entsteht eine neue algebraische Struktur, indem eine bisherige algebraische Struktur um eine neue Operation ergänzt wird. Dann legt man Wert darauf, daß die neue Operation (so gut es geht) zur bisherigen Struktur kompatibel ist.*

Ab jetzt werden wir bei den weiteren zu benennenden Strukturen aus Gründen der Übersichtlichkeit nicht immer alle Spielregeln aufzählen.

Ein **Ring** ist eine abelsche Gruppe $(G, +)$, die ergänzt wird um eine weitere Operation $\cdot$ mit zwei Argumenten. Diese soll die Spielregel der Assoziativität erfüllen. Als Kompatibilitätsbedingung zwischen $+$ und $\cdot$ haben wir das Distributivgesetz: $a \cdot (b + c) = a \cdot b + a \cdot c$ für jegliche $a, b, c \in G$.
Unsere Interpretation ist: ein Ring hat drei Operation, nämlich $+$, $-$, $\cdot$, von denen $-$ die Umkehrung zu $+$ ist. Das Standardbeispiel lautet $(\mathbb{Z}, +, \cdot)$.

Ein **Körper** ist ein Ring $(G, +, \cdot)$, in dem jede Gleichung $a \cdot x = b$ und jede Gleichung $y \cdot a = b$ lösbar ist (bei gegebenen $a, b \in G$ und gesuchten $x, y \in G$, wobei a nicht das neutrale Element zu $+$ sein darf).
Unsere Interpretation ist: ein Körper hat vier Operationen, von denen zwei die Umkehrungen der beiden anderen sind. Das Standardbeispiel ist $(\mathbb{R}, +, \cdot)$.

Beispiele für Gruppen sind:

- $(\mathbb{Q}_+, \cdot)$, $(\mathbb{Q} \setminus \{0\}, \cdot)$, $(\mathbb{R}_+, \cdot)$, $(\mathbb{R} \setminus \{0\}, \cdot)$
- Sei G die Menge der Abbildungen eines gleichseitigen Dreiecks auf sich, und sei $\circ$ die Nacheinanderausführung. Dann hat G 6 Elemente (nämlich 3 Drehungen um $0°$, $120°$, $240°$ sowie drei Spiegelungen jeweils an den Höhen). Diese Gruppe ist isomorph zur Permutationsgruppe S_3, und sie ist nicht abelsch (was man überprüfen kann, indem man sich zwei Spiegelungen anschaut). Die Nacheinanderausführung zweier Spiegelungen ergibt eine Drehung. Wenn τ und σ zwei Bewegungen sind, dann ist $\tau \circ \sigma$ auch eine Bewegung, und zwar diejenige, bei der σ zuerst ausgeführt wird, anschließend τ. Wir lesen $\circ$ als „nach".
- Sei G die Menge der orientierungserhaltenden Abbildungen eines gleichseitigen Dreiecks auf sich (und $\circ$ sei die Nacheinanderausführung, was ab jetzt bis auf weiteres immer so vereinbart sei). Es besteht G genau aus den drei Drehungen.
- Sei G die Menge aller (Kongruenz–)Abbildungen eines Würfels auf sich. Dann besteht G aus 48 Elementen. Die Verknüpfungstabelle hat 48 Spalten und 48 Zeilen, ist also vergleichsweise unübersichtlich.
- Sei G die Menge aller Drehbewegungen im dreidimensionalen Anschauungsraum, die einen konkreten Punkt festlassen. Diese Gruppe hat unendlich viele Elemente.
- Sei G die Menge aller Verschiebungen in der Ebene. Hierbei verstehen wir den Begriff der Ebene ganz elementargeometrisch — wir führen also keinerlei Koordinatensystem ein.
- Es sei G die Menge der Reste bei Division durch 3, und die Operation sei $\circ = +$. Es ist also $G = \{0, 1, 2\}$, und diese Gruppe ist isomorph zur obengenannten Gruppe der orientierungshaltenden Abbildungen eines gleichseitigen Dreiecks auf sich.

Beispiele für Ringe sind:

- $(3\mathbb{Z}, +, \cdot)$, welcher aus allen durch drei teilbaren Zahlen besteht. Analog $(4\mathbb{Z}, +, \cdot)$, der aus allen durch vier teilbaren Zahlen besteht.
- $(\mathbb{Z}|3\mathbb{Z}, +, \cdot) = (\{0, 1, 2\}, +, \cdot)$, also der Ring der Reste bei Division durch 3. Analog $(\mathbb{Z}|4\mathbb{Z}, +, \cdot) = (\{0, 1, 2, 3\}, +, \cdot)$.
- Die Menge aller Polynome in einer reellen Variablen x mit reellen Koeffizienten. Hierbei definieren wir $+$ und $\cdot$ auf übliche Weise.
- Die Menge $\mathbb{R}^{2\times 2}$ aller Matrizen vom Format 2×2 mit reellen Einträgen. Hierbei ist $+$ die übliche Addition von Matrizen, und $\cdot$ ist die Matrizenmultiplikation. Dieser Ring ist deshalb besonders wertvoll, weil er ein relativ einfach verstehbares Beispiel eines Ringes liefert, in dem die Multiplikation nicht kommutativ ist.

- Beinahe ein Ring ist die Menge aller stetigen Funktionen $[0, \infty) \to \mathbb{R}$. Hierbei definieren wir die Addition zweier Funktionen auf naheliegende Weise, und für die Multiplikation $*$ verwenden wir das sogenannte Faltungsprodukt, das definiert wird über die Formel

$$(u * v)(t) := \int_{s=0}^{t} u(t-s)v(s)\,\mathrm{d}s, \qquad t \in [0, \infty).$$

Die Multiplikation $*$ ist tatsächlich assoziativ und kommutativ. Auch das Distributivgesetz gilt, aber $*$ besitzt kein neutrales Element, weshalb definitionsgemäß kein Ring vorliegt.

Beispiele für Körper sind:

- $(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$, $(\mathbb{C}, +, \cdot)$
- $(\mathbb{Z}|p\mathbb{Z}, +, \cdot)$ genau dann, wenn p eine Primzahl ist. Dieser Körper ist deshalb für das Verständnis wertvoll, weil er nur endlich viele Elemente enthält. Man prüfe selber nach, daß $(\mathbb{Z}|3\mathbb{Z}, +, \cdot)$ ein Körper ist, aber $(\mathbb{Z}|4\mathbb{Z}, +, \cdot)$ nicht (warum nicht ?).
- Die Menge der gebrochenrationalen Funktionen in einer reellen Variablen mit reellen Koeffizienten.
- Der Körper der Quaternionen. Dieser besteht aus allen Termen der Form

$$x + u\mathbf{i} + v\mathbf{j} + w\mathbf{k}, \qquad x, u, v, w \in \mathbb{R},$$

wobei $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ Symbole sind, für deren Multiplikation folgende Regeln vereinbart werden:

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1, \qquad \mathbf{ij} = \mathbf{k}, \quad \mathbf{ji} = -\mathbf{k}, \qquad \mathbf{jk} = \mathbf{i}, \quad \mathbf{kj} = -\mathbf{i}, \qquad \mathbf{ki} = \mathbf{j}, \quad \mathbf{ik} = -\mathbf{j}.$$

Die Addition zweier Quaternionen wird auf naheliegende Weise definiert. Dieser Körper ist schief und deshalb ein für das Verständnis nützliches Beispiel.

Eine algebraische Struktur mit vielen Operationen ist im Allgemeinen wertvoller als eine Struktur mit weniger Operationen, weil jede Operation ein Werkzeug darstellt. Und wenn die Operationen gut zueinander kompatibel sind, dann kann man die Werkzeuge variabler einsetzen.
Wir wollen die Körper $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ mit ihren beiden Operationen $+$ und $\cdot$ (wir verschweigen ab jetzt oft deren Umkehroperationen) um weitere Operationen (oder Relationen) ergänzen.

Betrag: die Betragsfunktion gibt es für $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, und sie ist kompatibel zu $\cdot$ und $+$ wie folgt: $|u \cdot v| = |u| \cdot |v|$ und $|u + v| \le |u| + |v|$. Es wäre schön gewesen, wenn man sogar die Gleichheit $|u + v| = |u| + |v|$ für alle u, v aus dem jeweiligen Körper gehabt hätte, aber dies ließ sich leider nicht verwirklichen.

Konjugation: diese gibt es in $\mathbb{C}$ (und auch in $\mathbb{R}$ und $\mathbb{Q}$, aber dort bewirkt sie überhaupt nichts). Die Konjugation ist kompatibel zu den drei anderen Operationen $+$, $\cdot$ und Betrag im Sinne von $\overline{z + w} = \overline{z} + \overline{w}$, $\overline{z \cdot w} = \overline{z} \cdot \overline{w}$, $|\overline{z}| = z$.

Ordnung: diese existiert in $\mathbb{Q}$ und $\mathbb{R}$, aber nicht in $\mathbb{C}$. Wir haben die Kompatibilität zu $+$ und $\cdot$ gemäß

$$\begin{aligned} a \le b, \quad c \le d \qquad &\Longrightarrow \quad a + c \le b + d, \\ a \le b, \quad 0 \le c \qquad &\Longrightarrow \quad a \cdot c \le b \cdot c. \end{aligned}$$

Man könnte eine Ordnung auch in $\mathbb{C}$ definieren, aber eben nicht auf eine Weise, die kompatibel ist mit $+$ und $\cdot$.

Eine Kompatibilität zwischen $\le$ und der Betragsfunktion ist kaum zu erkennen (abgesehen von $x \le |x|$ für alle $x \in \mathbb{R}$), wird aber auch kaum benötigt.

Ein **Vektorraum** $(V, +, \mathbb{K})$ besteht aus einer abelschen Gruppe $(V, +)$ und einem Körper $(\mathbb{K}, +, \cdot)$. Es gibt zwei relevante Operationen:

- Zahl mal Vektor ergibt Vektor,
- Vektor plus Vektor ergibt Vektor.

(Wir verzichten auf das Mitzählen von Umkehroperationen). Zur Unterscheidung schreiben wir die Additionen als $+_V$ und $+_\mathbb{K}$, und die Zahlenmultiplikation innerhalb von $\mathbb{K}$ sei $\cdot_\mathbb{K}$. Dann gelten die Kompatibilitätsregeln

$$\begin{aligned}
(\lambda +_\mathbb{K} \mu) \cdot \vec{u} &= \lambda \cdot \vec{u} +_V \mu \cdot \vec{u}, \\
\lambda \cdot (\vec{u} +_V \vec{v}) &= \lambda \cdot \vec{u} +_V \lambda \cdot \vec{v}, \\
(\lambda \cdot_\mathbb{K} \mu) \cdot \vec{u} &= \lambda \cdot (\mu \cdot \vec{u}), \\
1 \cdot \vec{u} &= \vec{u},
\end{aligned}$$

wobei 1 das neutrale Element zu $\cdot_\mathbb{K}$ sei. Erst mit der letzten Bedingung kann man beweisen, daß $0_\mathbb{K} \cdot \vec{u} = \vec{0}$ für jedes $\vec{u} \in V$ ist.

Beispiele für Vektorräume sind:

- Alle Verschiebungen in der Ebene (die wir uns völlig ohne Koordinatensystem vorstellen). Jede Verschiebung wird symbolisiert durch einen konkreten Pfeil, der einen Punkt der Ebene mit seinem Bildpunkt verbindet. Man kann dieselbe Verschiebung auch symbolisieren durch jeden äquivalenten Pfeil, wobei zwei Pfeile als äquivalent gelten, wenn sie gleichsinnig parallel und gleichlang sind. Die Nacheinanderausführung symbolisieren wir durch die Addition zweier Pfeile, die wir mit Geodreieck und Bleistift über die Parallelogrammregel verwirklichen können. Die Multiplikation eines Pfeils mit einer reellen Zahl ist die Dehnung dieses Pfeils.
- Der $\mathbb{R}^2$ mit komponentenweise definierten Operationen.
- (Beide ebengenannten Vektorräume sind isomorph zueinander, und eine Isomorphieabbildung kann gebaut werden, nachdem man zwei zueinander nichtparallele Verschiebungen zur Basis proklamiert.)
- Jede Gerade durch den Ursprung in einer Ebene mit Koordinatensystem ist ein Vektorraum der Dimension 1.
- Alle Polynome einer reellen Variablen mit reellen Koeffizienten bilden einen (unendlichdimensionalen) Vektorraum.
- Alle Matrizen des $\mathbb{R}^{2\times 2}$ bilden einen Vektorraum der Dimension $2 \cdot 2 = 4$, wenn man die beiden Operationen komponentenweise definiert.
- Die Menge $C([0, 1]; \mathbb{R})$ der auf dem Intervall $[0, 1]$ stetigen Funktionen mit Werten in $\mathbb{R}$ bilden einen Vektorraum, wenn man die beiden Operationen (Funktion plus Funktion sowie Zahl mal Funktion) punktweise definiert.
- Allgemeiner: wenn X eine nichtleere Menge ist und V ein Vektorraum, dann ist $\mathcal{F}(X; V)$, also die Menge der Abbildungen von X nach V ein Vektorraum. Dieser ist unendlichdimensional, wenn X unendlich viele Elemente hat.
- Alle Folgen $(x_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ bilden einen Vektorraum, wenn man die Operationen (Folge plus Folge sowie Zahl mal Folge) komponentenweise definiert. Dies ist ein Spezialfall des vorherigen •, wenn man $X = \mathbb{N}$ und $V = \mathbb{R}$ setzt.
- Alle beschränkten Folgen in $\mathbb{R}$ bilden einen Vektorraum.
- Alle konvergenten Folgen in $\mathbb{R}$ auch.
- Alle Nullfolgen in $\mathbb{R}$ ebenfalls.

Eine **Algebra** ist eine Struktur, die gleichzeitig Ring und Vektorraum ist. Hierbei teilen sich Ring und Vektorraum die gleiche Addition, und es gibt Kompatibilitätsforderungen.
Beispiele sind:

- Die Menge $\mathbb{R}^{2\times 2}$ der reellen 2×2–Matrizen. Die Vektorraumoperationen sind „Zahl mal Matrix" und „Matrix plus Matrix". Die Ringoperationen sind „Matrix plus Matrix" und „Matrix mal Matrix". Die Kompatibilitätsforderung ist $(\lambda \cdot A) \cdot B = \lambda \cdot (A \cdot B)$ für $\lambda \in \mathbb{R}$ und $A, B \in \mathbb{R}^{2\times 2}$.
- Die Menge der linearen Abbildungen eines Vektorraums in sich. Die Ringmultiplikation ist hierbei die Nacheinanderausführung.
- Die Menge der Polynome einer reellen Variablen mit reellen Koeffizienten.

Schließlich bemerken wir noch: beim Ergänzen eines Vektorraums um eine Norm wurde diese Norm genau so definiert, daß sie kompatibel wird sowohl zur Addition im Vektorraum (über die Dreiecksungleichung) als auch zur Multiplikation von Zahlen mit Vektoren (über die Homogenitätsforderung an die Norm). Wenn man weiterhin den Vektorraum noch um ein Skalarprodukt ergänzt, ist dieses kompatibel zu den beiden Vektorraumoperationen und auch zur Norm.

3.2 Abbildungen

Seien X und Y zwei nichtleere Mengen, und sei $f\colon X \to Y$ eine Abbildung. Wir fragen: welche Eigenschaften vererben sich von X zu Y ? Welche Eigenschaften werden durch f^{-1} von Y zu X vererbt ?
Hierbei definieren wir für Teilmengen $A \subset X$ und $B \subset Y$:

$$\begin{aligned} f(A) &:= \{f(x)\colon x \in A\}, \\ f^{-1}(B) &:= \{x \in X\colon f(x) \in B\}. \end{aligned}$$

Wenn f nur eine Funktion ohne weitere Eigenschaften ist, dann gelten lediglich Vererbungseigenschaften wie

$$f(A_1 \cup A_2) = f(A_1) \cup f(A_2), \quad f(A_1 \cap A_2) \subset f(A_1) \cap f(A_2), \quad f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2)$$

für beliebige A_1, $A_2 \subset X$ und B_1, $B_2 \subset Y$.
Interessanter wird es, wenn X und Y Mengen mit Struktur sind und f passende Eigenschaften hat: Sei $f\colon [a, b] \to \mathbb{R}$ stetig. Dann gilt:

- f bildet in $[a, b]$ konvergente Folgen auf in $\mathbb{R}$ konvergente Folgen ab.
- f bildet in $[a, b]$ kompakte Mengen auf in $\mathbb{R}$ kompakte Mengen ab.
- Das Urbild jeder in $\mathbb{R}$ offenen Menge ist $[a, b]$–offen.

Das wichtigste Beispiel für Abbildungen in der Algebra sind **lineare Abbildungen** $f\colon U \to V$ zwischen zwei Vektorräumen U und V. Definitionsgemäß ist f kompatibel zu den beiden Vektorraumoperationen (Vektor plus Vektor sowie Zahl mal Vektor).
Sei $U = V = \mathbb{R}^2$. Wir denken uns möglichst viele f aus. Interessant sind immer **extremale Beispiele**, also solche, bei denen der Bildraum von f möglichst klein oder möglichst groß ist.

- f bildet U auf $\{0\} \subset V$ ab.
- f sei die identische Abbildung in $\mathbb{R}^2$.
- f ist Drehung um 0 oder Spiegelung an einer Geraden durch 0,
- f ist eine Streckung,
- f ist eine Scherung,
- f ist eine Projektion (schräg oder senkrecht) auf eine Gerade durch 0.

Sei $U = V$ der Vektorraum aller Polynome und f der Ableitungsoperator. Dieser ist surjektiv, aber nicht injektiv. Sei g der Integrationsoperator mit Integrationsstartpunkt Null, also

$$(g\varphi)(x) := \int_{t=0}^{x} \varphi(t)\,\mathrm{d}t,$$

wobei $\varphi = \varphi(x)$ ein beliebiges Polynom sei. Dann ist g eine injektive Abbildung, aber nicht surjektiv.

3.3 Anwendungen

Wir wollen unsere Beispiele verwenden, um algebraische Sachverhalte zu veranschaulichen. Dabei werden wir erkennen, wie entscheidend die Betrachtung von Beispielen für den eigenen Lernerfolg ist.

Wir beginnen mit **Normalteiler**n von Gruppen. Die einfachste uns bekannte nicht-abelsche Gruppe ist S_3, dargestellt durch die Abbildungen eines gleichseitigen Dreiecks auf sich. Wir haben die Rotationen ϱ_0, ϱ_1, ϱ_2 um 0, $2\pi/3$ und $4\pi/3$ im Gegenuhrzeigersinn, und wir haben die Spiegelungen σ_1, σ_2, σ_3. Hierbei sei das Dreieck mit der Spitze nach oben, und σ_1 hält die obere Ecke fest, σ_2 die rechte Ecke, σ_3 die linke Ecke.
Klar ist, daß alle diese Abbildungen eine Gruppe ergeben müssen. Es entsteht folgende Verknüpfungstafel (die oben genannten Abbildungen werden zuerst ausgeführt, die links genannten Abbildungen danach):

$\circ$	ϱ_0	ϱ_1	ϱ_2	σ_1	σ_2	σ_3
ϱ_0	ϱ_0	ϱ_1	ϱ_2	σ_1	σ_2	σ_3
ϱ_1	ϱ_1	ϱ_2	ϱ_0	σ_2	σ_3	σ_1
ϱ_2	ϱ_2	ϱ_0	ϱ_1	σ_3	σ_1	σ_2
σ_1	σ_1	σ_3	σ_2	ϱ_0	ϱ_2	ϱ_1
σ_2	σ_2	σ_1	σ_3	ϱ_1	ϱ_0	ϱ_2
σ_3	σ_3	σ_2	σ_1	ϱ_2	ϱ_1	ϱ_0

Hierbei ergeben sich die blauen Einträge durch besonders einfache geometrische Überlegungen, die roten Einträge durch schwierigere geometrische Überlegungen, und die schwarzen Einträge durch die Regel, daß in jeder Zeile und jeder Spalte jedes Gruppenelement genau einmal vorkommen muß (das ist genau der Inhalt der Bedingung, daß $a \circ x = b$ und $y \circ a = b$ immer eindeutig lösbar sein muß).
Eine Untergruppe N einer Gruppe ist ein Normalteiler genau dann wenn für jedes $g \in G$ gilt, daß $gN = Ng$.
Es ist $(\varrho_0, \varrho_1, \varrho_2)$ ein Normalteiler, denn es ist $\varrho_j \circ (\varrho_0, \varrho_1, \varrho_2) = (\varrho_0, \varrho_1, \varrho_2) = (\varrho_0, \varrho_1, \varrho_2) \circ \varrho_j$ sowie $\sigma_j \circ (\varrho_0, \varrho_1, \varrho_2) = (\sigma_1, \sigma_2, \sigma_3) = (\varrho_0, \varrho_1, \varrho_2) \circ \sigma_j$.
Aber es ist (ϱ_0, σ_1) kein Normalteiler von S_3, wenn wir haben $\varrho_1 \circ (\varrho_0, \sigma_1) = (\varrho_1, \sigma_3)$, jedoch $(\varrho_0, \sigma_1) \circ \varrho_1 = (\varrho_1, \sigma_2)$.
Sei $G = S_3$ und $N = (\varrho_0, \varrho_1, \varrho_2)$. Es ist dann die Menge der Nebenklassen G/N wieder eine Gruppe, und der Satz von Lagrange sagt $|G| = |N| \cdot |G/N|$, also hat G/N genau zwei Elemente.
Es hat N die beiden Nebenklassen $[\varrho_0] = \{\varrho_0, \varrho_1, \varrho_2\}$ und $[\sigma_1] = \{\sigma_1, \sigma_2, \sigma_3\}$. Auf $G/N = \{[\varrho_0], [\sigma_1]\}$ definieren wir die Verknüpfung $\circ$ mittels Repräsentanten. Dies ist auf widerspruchsfreie Weise eben deshalb möglich, weil N ein Normalteiler ist.

Wir kommen zum **Homomorphiesatz für Gruppen.** Sei $G = (\mathbb{Z}, +)$. Wir vergessen, daß $\mathbb{Z}$ in Wirklichkeit noch eine Multiplikation hat. Es ist G eine abelsche Gruppe, also ist jede Untergruppe von G ein Normalteiler. Setze $N = (6\mathbb{Z}, +)$ als die Gruppe der durch 6 teilbaren Zahlen.
Wir nehmen jetzt $\varphi\colon G \to H := \{0, 1, 2\}$, und H versehen wir mit der folgenden Verknüpfungstabelle:

$\circ$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

Wir können uns H vorstellen als die Gruppe der Reste ganzer Zahlen bei Division durch 3. Die Wirkung der Funktion φ definieren wir durch die Forderung, daß $\varphi(n)$ gleich dem Rest von n bei der Division durch 3 sein soll. Dann ist $\ker\varphi = \{\dots, -9, -6, -3, 0, 3, 6, 9, \dots\}$.
Wir haben $N \subset \ker\varphi$.
Es ist $G/N = \{[0], [1], [2], [3], [4], [5]\}$ gleich der Menge der Nebenklassen von N. Beispielsweise haben wir $[1] = \{\dots, -11, -5, 1, 7, 13, \dots\} \subset \mathbb{Z}$.
Wir definieren als Nächstes eine Abbildung $\pi\colon G \to G/N$, die jedem $n \in G$ diejenige Restklasse aus G/N zuordnet, in der n enthalten ist. Zum Beispiel ist $\varphi(-5) = [1]$ und $\varphi(13) = [1]$.
Der Gruppenhomomorphiesatz sagt jetzt: es existiert genau ein Homomorphismus $\overline{\varphi}\colon G/N \to H$ mit $\overline{\varphi} \circ \pi = \varphi$. Dieses $\overline{\varphi}$ können wir fast erraten:

$$\overline{\varphi}([0]) = 0, \quad \overline{\varphi}([1]) = 1, \quad \overline{\varphi}([2]) = 2, \quad \overline{\varphi}([3]) = 0, \quad \overline{\varphi}([4]) = 1, \quad \overline{\varphi}([5]) = 2.$$

Tatsächlich ist $(\overline{\varphi} \circ \pi)(13) = \overline{\varphi}(\pi(13)) = \overline{\varphi}([1]) = 1$ und $\varphi(13) = 1$, wie gewünscht.
Weiterhin ist $\ker\overline{\varphi} = \{[0], [3]\}$ und $\ker(\varphi)/N = \{\dots, -9, -6, -3, -0, 3, 6, 9, \dots\}/\{\dots, -6, -0, 6, \dots\}$, also gilt tatsächlich $\ker\overline{\varphi} = \ker(\varphi)/N$.

Wir kommen zum **Homomorphiesatz für Vektorräume.** Seien V und W die Mengen aller Polynome in einer reellen Variablen x mit reellen Koeffizienten. Sei $f := \frac{\mathrm{d}}{\mathrm{d}x}$ der Ableitungsoperator. Dann ist $f\colon V \to W$ eine lineare Abbildung. Es besteht $\ker f$ aus allen konstanten Polynomen, also $\ker f = \{c\colon c \in \mathbb{R}\}$.
Wir wählen $U := \ker f \subset V$. Wir setzen $\pi\colon V \to V/U$ als diejenige Abbildung, die jedem Polynom $v = v(x) \in U$ diejenige Restklasse aus V/U zuordnet, in der das Polynom v enthalten ist. Wenn z.B. $v = v(x) = 3x^2 + 7$, dann ist $\pi(v) = \{3x^2 + c\colon c \in \mathbb{R}\}$ eine Menge von quadratischen Funktionen, deren Graphen auseinander durch vertikale Verschiebung hervorgehen.
Der Homomorphiesatz für Vektorräume sagt: es existiert genau eine lineare Abbildung $\overline{f}\colon V/U \to W$ mit $\overline{f} \circ \pi = f$.
Zum Beispiel für $v = 3x^2 + 7$ ist $\overline{f}(\pi(v)) = 6x$, als Gleichheit von Polynomen. Es ist $\ker\overline{f}$ gleich derjenigen Restklasse aus V/U, die auf das Nullpolynom aus W abgebildet wird, und diese Restklasse enthält genau die konstanten Polynome.

Wir betrachten als nächstes **Quotientenräume.** Die Merkregel ist: wenn A, B Vektorräume sind mit $A \subset B$, dann ist jedes Element von $B|A$ eine „Kopie“ von A, die in B liegt und parallel zu A ist. Jede solche Kopie von A schreiben wir als $[b]$, wobei $b \in B$. Es ist $[b] = \{b + a\colon a \in A\}$.
Als Beispiel nehmen wir $V = \mathbb{R}^2$ und $U = \{\binom{t}{t}\colon t \in \mathbb{R}\}$. Dann ist

$$V|U = \left\{ \left[\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \right] : v_1, v_2 \in \mathbb{R} \right\} \quad \text{mit} \quad \left[\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \right] = \left\{ \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \begin{pmatrix} t \\ t \end{pmatrix} : t \in \mathbb{R} \right\}.$$

Man zeichne V, U und $[\binom{1}{2}]$, $[\binom{2}{3}]$, $[\binom{2}{1}]$!
Weiterhin wählen wir einen Vektorraumhomomorphismus

$$f\colon V \to \mathbb{R}^1, \qquad \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \mapsto v_1 - v_2.$$

Dann ist

$$\ker f = \{v \in V\colon f(v) = 0_{\mathbb{R}^1}\} = \left\{ \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \in V\colon f\left(\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \right) = 0 \right\} = \left\{ \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \in V\colon v_1 - v_2 = 0 \right\} = U.$$

Wir überlegen uns, daß f surjektiv ist, also ist $\operatorname{img} f = \mathbb{R}^1$. Laut Korollar zum Vektorraumhomomorphiesatz ist $V|\ker f \cong \operatorname{img} f$, was für uns $V|U \cong \mathbb{R}^1$ heißt. Die beiden Vektorräume $V|U$ und $\mathbb{R}^1$ sind also isomorph zueinander, und die Isomorphieabbildungs ist genau $\overline{f}$, definiert durch

$$\overline{f}\left(\left[\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \right] \right) := v_1 - v_2.$$

Ein weiteres Beispiel mit unendlichdimensionalen Vektorräumen ist

$$V = \{(x_n)_{n\in\mathbb{N}} \colon (x_n)_{n\in\mathbb{N}} \text{ ist Cauchyfolge in } \mathbb{Q}\},$$
$$U = \{(x_n)_{n\in\mathbb{N}} \colon (x_n)_{n\in\mathbb{N}} \text{ ist Nullfolge in } \mathbb{Q}\},$$

wobei eine Folge wie üblich „Nullfolge“ heißt, wenn sie den Grenzwert Null hat. Zwei Elemente von V heißen äquivalent, wenn ihre Differenzenfolge eine Nullfolge ist (also in U liegt). Diese Äquivalenzrelation erzeugt eine Zerlegung von V in Äquivalenzklassen. Jede Äquivalenzklasse ist ein Element des Quotientenraumes $V|U$. Wir betrachten einen Vektorraumhomomorphismus

$$f\colon V \to \mathbb{R}^1, \qquad f(v) := \lim_{n\to\infty} v_n \quad \text{wenn} \quad v = (v_n)_{n\in\mathbb{N}}.$$

Es ist $\ker f = U$ und $\operatorname{img} f = \mathbb{R}^1$, denn f ist surjektiv. Das Korollar des Vektorraumhomomorphiesatzes sagt $V|U \cong \mathbb{R}^1$, und tatsächlich wurden in der Analysisvorlesung reelle Zahlen definiert als Äquivalenzklassen von Cauchyfolgen reeller Zahlen.

Wir kommen zum **ersten Isomorphiesatz für Vektorräume.** Wir nehmen

$$V = \mathbb{R}^3, \qquad U = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\}, \qquad W = \left\{ \begin{pmatrix} x_1 \\ 0 \\ x_3 \end{pmatrix} : x_1, x_3 \in \mathbb{R} \right\}.$$

Wir haben also $U \subset V$ und $W \subset V$, und im üblichen $x_1x_2x_3$–Koordinatensystem ist U die x_1x_2–Ebene und W die x_1x_3–Ebene. Dann ist

$$U \cap W = \left\{ \begin{pmatrix} x_1 \\ 0 \\ 0 \end{pmatrix} : x_1 \in \mathbb{R} \right\}$$

also die x_1–Achse, und $U|U \cap W$ besteht gemäß Merkregel aus allen Kopien von $U \cap W$, die in U liegen und parallel zu $U \cap W$ sind. Jedes Element von $U|U \cap W$ ist also eine zur x_1–Achse parallele Gerade, die in der x_1x_2–Ebene liegt. In Formeln ausgedrückt, ist

$$U|U \cap W = \left\{ \left[\begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} \right] : x_1, x_2 \in \mathbb{R} \right\}, \qquad \left[\begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} \right] = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} + \begin{pmatrix} t \\ 0 \\ 0 \end{pmatrix} : t \in \mathbb{R} \right\},$$

also $U|U \cap W \cong \{x_2 \colon x_2 \in \mathbb{R}\}$.
Weiterhin ist

$$U + W = \{u + w \colon u \in U \ \ w \in W\} = \left\{ \begin{pmatrix} u_1 \\ u_2 \\ 0 \end{pmatrix} + \begin{pmatrix} w_1 \\ 0 \\ w_3 \end{pmatrix} : u_1, u_2, w_1, w_3 \in \mathbb{R} \right\} = \mathbb{R}^3 = V.$$

Gemäß Merkregel besteht $U + W|W$ aus allen Kopien von W, die in $U + W = \mathbb{R}^3$ liegen und parallel zu W sind. Wir haben

$$R^3|W = \left\{ \left[\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \right] : x_1, x_2, x_3 \in \mathbb{R} \right\}, \qquad \left[\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \right] = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} t_1 \\ 0 \\ t_3 \end{pmatrix} : t_1, t_3 \in \mathbb{R} \right\},$$

also ist $\mathbb{R}^3|W \cong \{x_2 \colon x_2 \in \mathbb{R}\}$.
Wir beobachten $U|U \cap W \cong (U + W)|W$.

Als weiteres Beispiel für den **ersten Isomorphiesatz für Vektorräume** nehmen wir V als den Vektorraum aller Polynome in $x \in \mathbb{R}$ mit reellen Koeffizienten, U als denjenigen Untervektorraum von V, der aus allen ungeraden Polynomen besteht (also aus allen $u = u(x)$ mit $u(x) = -u(-x)$ für alle $x \in \mathbb{R}$), W als denjenigen

Untervektorraum von V, der aus allen geraden Polynomen besteht (also aus allen $w = w(x)$ mit $w(x) = w(-x)$ für alle $x \in \mathbb{R}$).
Dann enthält $U \cap W$ nur das Nullpolynom, also $U \cap W = \{0\}$, also ist

$$U|U \cap W = U|\{0\} \cong U.$$

Weiterhin kann man jedes Polynom $p \in V$ zerlegen in einen ungeraden und einen geraden Anteil:

$$p(x) = \frac{p(x) - p(-x)}{2} + \frac{p(x) + p(-x)}{2},$$

also ist $U + W = V$. Wegen $U \cap W = \{0\}$ ist die Zerlegung von p in einen geraden und einen ungeraden Anteil eindeutig, also können wir sogar $V = U \oplus W$ schreiben.
Tatsächlich ist dann $(U + W)|W = V|W$ isomorph zur Menge der ungeraden Polynome, denn das Bilden des Quotientenraumes $V|W$ bedeutet ja gerade, daß wir von einem Polynom $v \in V$ seinen geraden Anteil vergessen/wegwerfen/ignorieren, und dann bleibt der ungerade Anteil übrig.

Jetzt wollen wir über die **Begriffsbildung Quotientenräume** nachdenken. Warum heißt $V|U$ so ?
Nur in unserer Anschauung stellen wir uns vor, daß V ein Zähler ist und U ein Nenner. Wir wollen V festhalten und U variieren. Es ist sinnvoll, Extremalbeispiele für U zu betrachten, also U möglichst groß oder klein zu wählen. Wir wissen schon: $V|\{0\} \cong V$, und wir überlegen uns, daß $V|V = \{[0]\}$. Daraus vermuten wir: wenn $U_1 \subset U_2 \subset V$, d.h. wenn U_1 „kleiner“ als U_2 ist, dann sollte $V|U_1$ „größer“ als $V|U_2$ sein, also benähme sich $|$ wie ein Bruchstrich. Die Interpretation (nur in der Anschauung !) ist: wenn man eine Menge in kleinere Stücke zerlegt, dann entstehen mehr Stücke.
Zur Verdeutlichung: sei $V = \mathbb{R}^3$, $U_1 = \{(x_1, 0, 0)^\top : x_1 \in \mathbb{R}\}$ und $U_2 = \{(x_1, x_2, 0)^\top : x_1, x_2 \in \mathbb{R}\}$. Dann ist $U_1 \subset U_2$. Der Quotientenraum $V|U_2$ besteht aus parallelen Kopien der Ebene U_2, also ist $V|U_2$ vorstellbar als ein Stapel Kopierpapier, und zum Durchnumerieren der Blätter in diesem Stapel brauchen wir eine Koordinate, die die Höhe des betreffenden Blattes im Stapel angibt. Als Formel ist dies $V|U_2 \cong \mathbb{R}^1$.
Weiterhin besteht $V|U_1$ aus parallelen Kopien der Gerade U_1, also ist $V|U_1$ vorstellbar als eine Packung Spaghetti, und zum Benennen eines konkreten Spaghettostabs in der Packung benötigt man zwei Koordinaten, denn die Spaghettostäbe liegen nebeneinander und übereinander. Als Formel ist dies $V|U_1 \cong \mathbb{R}^2$, und tatsächlich ist $\mathbb{R}^1$ interpretierbar als eindimensionaler Untervektorraum des $\mathbb{R}^2$, z.B. im Sinne von $\binom{x_1}{0} \in \mathbb{R}^2$ für jedes $x_1 \in \mathbb{R}^1$.

Wir kommen jetzt zum **zweiten Isomorphiesatz für Vektorräume.** Wir nehmen $V = \mathbb{R}^3$ und

$$U = \left\{ \begin{pmatrix} x_1 \\ 0 \\ 0 \end{pmatrix} : x_1 \in \mathbb{R} \right\}, \qquad W = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\}.$$

Dann ist $U \subset W \subset V$.
Nach Merkregel ist $V|U$ die Menge der Geraden im $\mathbb{R}^3$, die parallel zur x_1–Achse sind. Jede solche Gerade wird repräsentiert durch einen Punkt auf dieser Geraden. Es ist egal, welchen Punkt wir als Repräsentanten auf der Geraden nehmen, und wir entscheiden uns für den Durchstoßpunkt der Geraden mit der x_2x_3–Ebene. In diesem Sinne ist $V|U \cong \mathbb{R}^2$, und die Isomorphieabbildung ist

$$V|U \to \mathbb{R}^2, \qquad \left[\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\right] \mapsto \begin{pmatrix} x_2 \\ x_3 \end{pmatrix}. \qquad \text{Beachte } \left[\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\right] = \left[\begin{pmatrix} 7 \\ x_2 \\ x_3 \end{pmatrix}\right] = \left[\begin{pmatrix} 0 \\ x_2 \\ x_3 \end{pmatrix}\right].$$

Weiterhin ist $W|U$ die Menge der zur x_1–Achse parallelen Geraden in der Grundebene W. Jede solche Gerade in der Grundebene wird repräsentiert durch einen Punkt auf dieser Geraden, und es ist egal, welchen Punkt wir als Repräsentanten wählen. Wir entscheiden uns für den Durchstoßpunkt der Geraden mit der x_2–Achse. Es ist dann $W|U \cong \mathbb{R}^1$, und die Isomorphieabbildung ist

$$W|U \to \mathbb{R}^1, \qquad \left[\begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}\right] \mapsto \begin{pmatrix} x_2 \end{pmatrix}. \qquad \text{Beachte } \left[\begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}\right] = \left[\begin{pmatrix} -9 \\ x_2 \\ 0 \end{pmatrix}\right] = \left[\begin{pmatrix} 0 \\ x_2 \\ 0 \end{pmatrix}\right].$$

Schließlich ist $V|W$ die Menge der zur Grundebene W parallelen Ebenen im $\mathbb{R}^3 = V$. Jede solche Ebene wird repräsentiert durch einen Punkt auf der Ebene. Es ist egal, welchen Punkt wir als Repräsentanten auf der Ebene wählen, und wir entscheiden uns für den Schnittpunkt der betreffenden Ebene mit der x_3–Achse. In diesem Sinne ist $V|W \cong \mathbb{R}^1$, und die Isomorphieabbildung ist

$$V|W \to \mathbb{R}^1, \qquad \left[\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\right] \mapsto (x_3). \qquad \text{Beachte} \left[\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\right] = \left[\begin{pmatrix} 12 \\ 47 \\ x_3 \end{pmatrix}\right] = \left[\begin{pmatrix} 0 \\ 0 \\ x_3 \end{pmatrix}\right].$$

Weiterhin ist $W|U$ ein Untervektorraum von $V|U$, denn jedes Element von $W|U$ ist auch ein Element von $V|U$ (es handelt sich hier immer um Geraden, die zur x_1–Achse parallel sind). Das Untervektorraumkriterium zeigt dann, daß tatsächlich $W|U$ ein Untervektoraum von $V|U$ ist.
Der zweite Isomorphiesatz besagt nun

$$(V|U)\Big|(W|U) \cong V|W.$$

Links steht eine Menge, die aus allen parallelen Kopien von $W|U$ besteht, welche in $V|U$ liegen. Das stellen wir uns wie folgt vor: für $W|U$ denken wir an eine Schicht von parallel angeordneten Spaghettostäben. Unendlich viele Kopien dieser Schicht werden übereinander gestapelt, und wir füllen dann den Raum $V|U$ aus.

Genausogut könnten wir aber auch W betrachten (ein horizontales Blatt Papier), und unendlich viele Blätter werden übereinandergestapelt, was den Raum V ausfüllt.

Letztlich ist es äquivalent, ob man eine Schicht von parallel angeordneten Spaghettostäben unendlich oft kopiert und diese Kopien nach oben stapelt, oder ob man ein Blatt unendlich oft kopiert und diese Kopien nach oben stapelt. Das ist eine (stark unwissenschaftlich formulierte) Interpretation des zweiten Isomorphiesatzes.

Unser Lerneffekt: Betrachtungen konkreter Beispiele können hilfreich sein, um sich unter algebraischen Strukturen etwas Bildliches vorzustellen.

Kapitel 4

Problemlösen II (oder: vermischte Aufgaben)

Unser Ziel ist: das Problemlösen soll nochmal eingeübt werden, wobei die Darstellung jetzt abgekürzt wird. Weiterhin wollen wir Bezüge zu behandelten Inhalten der linearen Algebra herstellen.

4.1 Rund ums Skalarprodukt

Im Vektorraum $\mathbb{R}^n$ definieren wir ein Skalarprodukt:

Definition 4.1. *Die Abbildung* $\langle \cdot, \cdot \rangle_n : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, *gegeben durch* $\langle u, v \rangle_n := \sum_{j=1}^n u_j v_j$ *für* $u = (u_1, \dots, u_n)^\top$ *und* $v = (v_1, \dots, v_n)^\top$, *heißt* (Standard)–Skalarprodukt *im* $\mathbb{R}^n$.

Wir beobachten, daß $\langle \cdot, \cdot \rangle_n$ kompatibel ist zu den beiden Operationen im $\mathbb{R}^n$ (Vektor plus Vektor sowie Zahl mal Vektor):

$$\begin{aligned} \langle u + \tilde{u}, v \rangle_n &= \langle u, v \rangle_n + \langle \tilde{u}, v \rangle_n , \qquad \forall u, \tilde{u}, v \in \mathbb{R}^n, \\ \langle \lambda u, v \rangle_n &= \lambda \langle u, v \rangle_n , \qquad \forall \lambda \in \mathbb{R}, \quad \forall u, v \in \mathbb{R}^n, \end{aligned}$$

und zwei analoge Zeilen gelten für den zweiten Faktor.
Wir beobachten auch: wenn $u, v \in \mathbb{R}^n$ (interpretiert als Spaltenvektoren) und wenn wir jeden Vektor als einspaltige Matrix (oder einzeilige Matrix) lesen, dann ist

$$\langle u, v \rangle_n = u^\top v = v^\top u = \langle v, u \rangle_n .$$

Aufgabe 4.2. *Zeigen Sie für beliebige* $A \in \mathbb{R}^{n \times m}$, $x \in \mathbb{R}^m$ *und* $y \in \mathbb{R}^n$, *daß*

$$\langle Ax, y \rangle_n = \left\langle x, A^\top y \right\rangle_m .$$

Das ist nicht sehr schwer: wegen $(PQ)^\top = Q^\top P^\top$ für kompatible Matrizen P und Q ist

$$\langle Ax, y \rangle_n = (Ax)^\top y = (x^\top A^\top) y = x^\top (A^\top y) = \left\langle x, A^\top y \right\rangle_m .$$

Aufgabe 4.3. *Sei* $A \in \mathbb{R}^{n \times m}$ *eine beliebige Matrix. Beweisen Sie, daß* $\mathbb{R}^m = \ker A \oplus \operatorname{img} A^\top$ *sowie* $\mathbb{R}^n = \ker A^\top \oplus \operatorname{img} A$.

Als ersten Schritt zur Lösung tragen wir die Definitionen der relevanten Fachbegriffe zusammen:

Kern einer Matrix: ?

Bild einer Matrix: ?

direkte Summe (geschrieben mit $\oplus$) zweier Untervektorräume: ?

Jetzt haben wir die Definitionen aller vorkommenden (und demnächst genannt werdenden) Fachbegriffe, und wir tragen Eigenschaften dieser Fachbegriffe zusammen (man darf diesen Arbeitsschritt auffassen als Übung im freien Assoziieren, gefolgt von präzisem Hinschreiben):

- $\ker A$ ist ein Untervektorraum von $\mathbb{R}^m$, $\operatorname{img} A$ ist ein Untervektorraum von $\mathbb{R}^n$, $\ker A^\top$ ist ein UVR von $\mathbb{R}^n$, $\operatorname{img} A^\top$ ist ein UVR von $\mathbb{R}^m$
- A erzeugt eine lineare Abbildung von $\mathbb{R}^m$ nach $\mathbb{R}^n$, und $A^\top$ bildet in umgekehrter Richtung ab
- Für lineare Abbildungen gibt es eine Dimensionsformel: $\dim \mathbb{R}^m = \dim\ker A + \dim\operatorname{img} A$ sowie $\dim \mathbb{R}^n = \dim\ker A^\top + \dim\operatorname{img} A^\top$.
- Für UVR gibt es ebenfalls eine Dimensionsformel: $\dim(\ker A + \operatorname{img} A^\top) + \dim((\ker A) \cap (\operatorname{img} A^\top) = \dim\ker A + \dim\operatorname{img} A^\top$ und weiterhin $\dim(\ker A^\top + \operatorname{img} A) + \dim((\ker A^\top) \cap (\operatorname{img} A) = \dim\ker A^\top + \dim\operatorname{img} A$.
- Der Rang einer Matrix ist die Dimension ihres Bildraums (das ist so definiert).
- Die Spalten einer Matrix sind ein Erzeugendensystem für den Bildraum dieser Matrix (das gilt immer; wer es nicht glaubt, mag bitte Beispiele zur Veranschaulichung selber suchen (oder sich überlegen, wer mit wem multipliziert wird bei einer Multiplikation einer Matrix mit einem Spaltenvektor)).
- Also ist der Rang der Matrix gleich der Anzahl ihrer linear unabhängigen Spalten (das ist eine schnelle Folgerung aus den vorigen beiden •).
- Der Rang ist auch gleich der Anzahl der linear unabhängigen Zeilen (das ist ein tiefer Satz mit aufwendigem Beweis).
- Beim Transponieren einer Matrix werden Spalten zu Zeilen und umgekehrt.
- Also ist $\operatorname{rang} A = \operatorname{rang} A^\top$, denn der Zeilenrang ist immer gleich dem Spaltenrang.

Jetzt haben wir schon fünf Gleichungen, also sollten wir unsere Erkenntnisse konsolidieren: sei $r := \operatorname{rang} A = \operatorname{rang} A^\top$. Dann haben wir aus den Dimensionsformeln für Abbildungen, daß

$$\dim\ker A = m - r, \qquad \dim\ker A^\top = n - r.$$

Damit gehen wir in die Dimensionsformeln für UVRe hinein:

$$\dim(\ker A + \operatorname{img} A^\top) + \dim((\ker A) \cap (\operatorname{img} A^\top) = (m - r) + r = m,$$
$$\dim(\ker A^\top + \operatorname{img} A) + \dim((\ker A^\top) \cap (\operatorname{img} A) = (n - r) + r = n,$$

und das sieht schon ziemlich gut aus. Da wir zeigen müssen, daß die UVR-Summen *direkte* Summen sind, kommen wir sowieso nicht daran vorbei, $(\ker A) \cap (\operatorname{img} A^\top) = \{0\}$ und $(\ker A^\top) \cap (\operatorname{img} A) = \{0\}$ zu beweisen. Dies können wir dadurch bewerkstelligen, daß wir zeigen, daß jeder der UVRe auf der linken Gleichungsseite auf dem anderen UVR senkrecht steht.
Seien also $x \in \ker A$ und $y \in \operatorname{img} A^\top$ beliebig gewählt. Dann ist $y = A^\top z$ für ein z, und wir haben

$$\langle x, y\rangle_m = \left\langle x, A^\top z\right\rangle_m = \langle Ax, z\rangle_n = \langle 0, z\rangle_n = 0.$$

Daraus folgt $\ker A \perp \operatorname{img} A^\top$. Wenn nun $q \in (\ker A) \cap (\operatorname{img} A^\top)$ ist, dann folgt $\langle q, q\rangle_m = 0$, aber es ist $\langle q, q\rangle_m = \sum_{j=1}^m q_j^2$, also ist $q = (q_1, \ldots, q_m)^\top = (0, \ldots, 0)^\top$, und somit folgt tatsächlich $(\ker A) \cap (\operatorname{img} A^\top) = \{0\}$. Die andere Relation folgt analog.
Wir tragen die Erkenntnisse zusammen:

$$\dim(\ker A \oplus \operatorname{img} A^\top) + \dim(\{0\}) = m.$$
$$\dim(\ker A^\top \oplus \operatorname{img} A) + \dim(\{0\}) = n.$$

Wenn nun $\ker A \oplus \operatorname{img} A^\top \neq \mathbb{R}^m$ wäre, dann wäre $\ker A \oplus \operatorname{img} A^\top$ ein echter Unterraum des $\mathbb{R}^m$, müßte aber trotzdem die Dimension m haben. Das kann nicht sein. Also ist $\ker A \oplus \operatorname{img} A^\top = \mathbb{R}^m$, wie zu zeigen war. Die zweite Gleichung zeigt man genauso.

Im Aufschrieb bis hierhin gingen Gedankenfindung durch freies Assoziieren und Begründungen durcheinander. Ein anständiger Aufschrieb (so, wie Sie ihn auf den Hausaufgabenzetteln hinschreiben sollen) sähe etwa so aus:

Wir definieren $r := \operatorname{rang} A$. Gemäß Vorlesung sind Zeilenrang und Spaltenrang einer Matrix gleich, also ist auch $r = \operatorname{rang} A^\top$. Aus den Dimensionsformeln für lineare Abbildungen haben wir (wegen $A\colon \mathbb{R}^m \to \mathbb{R}^n$ und $A^\top : \mathbb{R}^n \to \mathbb{R}^m$) dann

$$m = \dim \ker A + r, \qquad n = \dim \ker A^\top + r.$$

Weiterhin ist $(\ker A) \cap (\operatorname{img} A^\top) = \{0\}$, denn sei $q \in (\ker A) \cap (\operatorname{img} A^\top)$, dann existiert ein $z \in \mathbb{R}^n$ mit $q = A^\top z$. Dann ist aber (wegen der vorigen Aufgabe)

$$\sum_{j=1}^m q_j^2 = \langle q, q\rangle_m = \left\langle q, A^\top z\right\rangle_m = \langle Aq, z\rangle_n = \langle 0, z\rangle_n = 0,$$

also $q = (0, \dots, 0)^\top$. Analog zeigt man $(\ker A^\top) \cap (\operatorname{img} A) = \{0\}$.
Weiterhin sind $\ker A$ und $\operatorname{img} A^\top$ UVRe im $\mathbb{R}^m$, also gilt wegen der Dimensionsformel für UVRe im Falle von direkten Summen, daß

$$\dim(\ker A \oplus \operatorname{img} A^\top) = \dim \ker A + \dim \operatorname{img} A^\top = (m - r) + r = m.$$

Wenn nun $\ker A \oplus \operatorname{img} A^\top \neq \mathbb{R}^m$ wäre, dann wäre $\ker A \oplus \operatorname{img} A^\top$ ein echter Unterraum des $\mathbb{R}^m$, müßte aber trotzdem die Dimension m haben. Das kann nicht sein. Also ist $\ker A \oplus \operatorname{img} A^\top = \mathbb{R}^m$, wie zu zeigen war. Die zweite Gleichung zeigt man genauso.
(Ende des Lösungsaufschriebs).

Als Schlußbemerkung stellen wir fest, daß wir auf unserem Schmierzettel mehr gezeigt hatten als in der Aufgabenstellung gefragt war: denn wir hatten sogar bewiesen, daß die beiden Summanden in der direkten Summe aufeinander senkrecht stehen. Das ist bei direkten Summen von UVRen nicht immer der Fall (denn es ist ja gar nicht gesagt, daß der „große Vektorraum“ überhaupt mit einem Skalarprodukt ausgestattet ist).

4.2 Lineare Gleichungen in Vektorräumen

Seien U und V Vektorräume über $\mathbb{R}$, wir wissen nichts über deren Dimensionen. Sei $F\colon U \to V$ eine lineare Abbildung. Wir wollen die Gleichung $F(u) = v$ lösen, wobei $v \in V$ gegeben sei und $u \in U$ gesucht.

Wir beobachten: seien u und $\hat{u}$ Lösungen von $F(\cdot) \stackrel{!}{=} v$. Dann ist $u - \hat{u} \in \ker F$. Denn es ist $F(u) = v$ und $F(\hat{u}) = v$, also wegen der Linearität von F auch $F(u - \hat{u}) = F(u) - F(\hat{u}) = v - v = 0$.

Wir beobachten umgekehrt: sei u Lösung von $F(\cdot) \stackrel{!}{=} v$, und sei $\tilde{u} \in \ker F$. Dann ist auch $\hat{u} := u + \tilde{u}$ eine Lösung der Gleichung $F(\cdot) \stackrel{!}{=} v$.
Wir vergeben Bezeichnungen. Die Aufgabe

finde alle $u \in U$ mit $F(u) = v$

heißt **inhomogenenes Problem.**
Die Aufgabe

finde alle $u \in U$ mit $F(u) = 0$

heißt **homogenenes Problem.** Die Lösungen des homogenen Problems bilden den Kern von F.
Unsere beiden Beobachtungen können wir dann schreiben wie folgt:

- die Differenz zweier Lösungen des inhomogenen Problems löst das homogene Problem,
- man bekommt *jede* Lösung des inhomogenen Problems, indem man irgendeine Lösung des inhomogenen Problems findet und zu dieser dann jede Lösung des homogenen Problems dazuaddiert.

Aufgabe 4.4. *Man bestimme alle Folgen $u = (u_0, u_1, u_2, \dots) \subset \mathbb{R}$ mit $u_{n+1} = 2u_n + 1$ für alle $n \in \mathbb{N}_0$.*

Als Vektorräume nehmen wir $U = V = \{(u_0, u_1, \dots) \colon u_j \in \mathbb{R} \; \forall \, j\}$, was ein unendlichdimensionaler Vektorraum ist, und jedes Element davon ist eine Folge in $\mathbb{R}$. Wenn wir die zu lösenden Gleichungen umschreiben als

$$
\begin{aligned}
u_1 - 2u_0 &= 1, \\
u_2 - 2u_1 &= 1, \\
u_3 - 2u_2 &= 1, \\
\vdots \; &\; \vdots
\end{aligned}
$$

dann ist es naheliegend, eine Abbildung $F \colon U \to V$ zu definieren als

$$F\big((u_0, u_1, u_2, \dots)\big) := (u_1 - 2u_0, u_2 - 2u_1, u_3 - 2u_2, \dots)$$

und dann nach Lösungen von $F(u) = e$ zu fragen, wobei wir $e := (1, 1, 1, 1, \dots) \in V$ definieren. Formal können wir $(F(u))_j := u_{j+1} - 2u_j$ schreiben für eine Definition der Wirkung von F.
Unser Lösungsweg hat zwei Schritte:

- wir bestimmen $\ker F$,
- wir bestimmen irgendeine Lösung u von $F(u) = e$.

Wenn $u = (u_0, u_1, u_2, \dots) \in \ker F$ ist, dann haben wir $0 = (F(u))_j = u_{j+1} - 2u_j$ für jedes j, (die Umkehrung gilt auch), also

$$u_1 = 2u_0, \quad u_2 = 2u_1 = 2^2 u_0, \quad u_3 = 2u_2 = 2^3 u_0, \dots,$$

also per Induktion auch $u_n = 2^n u_0$, und wir erkennen dann

$$\ker F = \{(t, 2^1 t, 2^2 t, 2^3 t, \dots) \colon t \in \mathbb{R}\}.$$

Jetzt fehlt uns noch irgendeine Lösung u von $F(u) = e$. Hier dürfen wir raten (denn es geht ja in diesem Moment nur darum, irgendeine Lösung zu finden). Die rechte Seite e der Gleichung $F(u) = e$ hat einen „störenden Einfluß“, weshalb man e auch als Störterm bezeichnet. Die Erfahrung lehrt, daß man als erstes sein Glück probiert mit einem u, das sich so ähnlich benimmt wie der Störterm. Wir testen also $u = (\alpha, \alpha, \alpha, \alpha, \dots)$ mit einem Parameter $\alpha \in \mathbb{R}$, und dann soll gelten $1 = (F(u))_j = u_{j+1} - 2u_j = \alpha - 2\alpha = -\alpha$, also nehmen wir $\alpha = -1$.
Die Antwort lautet: alle Folgen $(u_0, u_1, u_2, \dots)$ mit $u_{j+1} = 2u_j + 1$ (für alle $j \in \mathbb{N}_0$) werden gegeben durch die Formel

$$u_j = 2^j t - 1, \qquad j \in \mathbb{N}_0,$$

wobei $t \in \mathbb{R}$ ein fester Parameter ist (von j unabhängig).

Aufgabe 4.5. *Gesucht ist eine explizite Formel für das allgemeine Folgenglied u_n der Folge $(u_0, u_1, u_2, \dots) \subset \mathbb{R}$, die gegeben wird durch die Bedingungen*

$$u_0 = u_1 = 1, \qquad u_{n+2} = u_n + u_{n+1} \quad \forall \, n \in \mathbb{N}_0.$$

Diese Zahlenfolge $(u_0, u_1, u_2, \dots) = (1, 1, 2, 3, 5, 8, 13, \dots)$ heißt auch *Folge der Fibonacci–Zahlen.*
Die Rekursionsformel $u_{n+2} = u_n + u_{n+1}$ ist offenkundig bereits homogen. Wir wählen U als den Vektorraum aller Folgen in $\mathbb{R}$, genauso wie in der Lösung zur vorigen Aufgabe.
Wir ignorieren die beiden Anfangsbedingungen $u_0 = 1$ und $u_1 = 1$ für einen Moment, und wir definieren einen UVR von U:

$$U_0 := \{u \in U \colon u_{n+2} = u_n + u_{n+1} \; \forall \, n \in \mathbb{N}_0\}.$$

Wir können U_0 auffassen als Kern einer linearen Abbildung $F \colon U \to U$, wenn wir es möchten.
Wir möchten nicht unbedingt. Stattdessen erscheint es sinnvoller, U_0 so gut wie möglich zu verstehen, denn die von uns gesuchte Folge liegt ja in U_0.

Man versteht einen Vektorraum dann am Besten, wenn man für ihn eine aussagekräftige Basis angeben kann. Wieviele Elemente hat die Basis von U_0 ? Was ist die Dimension von U_0 ? Wir überlegen uns: wenn $u \in U_0$, dann werden alle Folgenglieder von u ab dem u_2 bereits durch u_0 und u_1 festgelegt. Und den Vektor (u_0, u_1) können wir immer erzeugen gemäß $(u_0, u_1) = u_0 \cdot (1, 0) + u_1 \cdot (0, 1)$. Damit gilt: die beiden Folgen

$$(1, 0, 1, 1, 2, 3, 5, \dots), \qquad (0, 1, 1, 2, 3, 5, 8, \dots)$$

bilden ein Erzeugendensystem von U_0. Folglich ist $\dim U_0 \le 2$. Andererseits sind die beiden angegebenen Folgen linear unabhängig. Denn sei

$$\alpha_1 \cdot (1, 0, 1, 1, 2, 3, 5, \dots) + \alpha_2 \cdot (0, 1, 1, 2, 3, 5, 8, \dots) = (0, 0, 0, 0, \dots).$$

Dann ist (wenn wir die ersten Einträge vergleichen) $\alpha_1 \cdot 1 + \alpha_2 \cdot 0 = 0$, andererseits (wenn wir die zweiten Einträge vergleichen) $\alpha_1 \cdot 0 + \alpha_2 \cdot 1 = 0$, also $\alpha_1 = \alpha_2 = 0$, was die lineare Unabhängigkeit der beiden angegebenen Folgen beweist.

Damit ist $\dim U_0 = 2$, und die beiden Folgen bilden eine Basis, allerdings können wir diese beiden Folgen nicht sehr gut verstehen (oder können Sie sofort angeben, was der 47. Eintrag der ersten Folge ist ?).

Wir wollen eine schönere Basis von U_0. Welche schönen Folgen kennen wir ? Wir kennen arithmetische Folgen $(a, a + d, a + 2d, a + 3d, \dots)$, quadratische Folgen wie etwa $(1, 2^2, 3^2, 4^2, \dots)$ oder auch geometrische Folgen wie $(\gamma^0, \gamma^1, \gamma^2, \dots)$.

Wir probieren herum und finden durch Überschlagsrechnungen im Kopf, daß die arithmetischen Folgen und quadratischen Folgen einfach zu langsam wachsen für große n. Deshalb probieren wir, eine Basis von U_0 zu finden mittels geometrischen Folgen. Wir suchen also $\gamma \in \mathbb{R}$ mit

$$(\gamma^0, \gamma^1, \gamma^2, \dots) \in U_0.$$

Dafür brauchen wir $\gamma^{n+2} = \gamma^n + \gamma^{n+1}$ für alle $n \in \mathbb{N}_0$, was nach Kürzen dann $\gamma^2 = 1 + \gamma$ ergibt (die Wahl von $\gamma = 0$ ergibt den Nullvektor in U_0, der aber kein Basisvektor ist. Deshalb ist Kürzen hier zulässig). Lösen dieser quadratischen Gleichung bringt uns dann[1]

$$\gamma_1 = \frac{1 + \sqrt{5}}{2} \approx 1.618034, \qquad \gamma_2 = \frac{1 - \sqrt{5}}{2} \approx -0.618034.$$

Tatsächlich sind die beiden Folgen

$$(\gamma_1^0, \gamma_1^1, \gamma_1^2, \dots), \qquad (\gamma_2^0, \gamma_2^1, \gamma_2^2, \dots)$$

linear unabhängig (Übung im Selbststudium), und nach Konstruktion liegen sie in U_0.

Was haben wir erzielt ? Wir können jedes Element von U_0 schön angeben, weil wir eine schöne Basis von U_0 gefunden haben[2]. Jedes beliebige Element von U_0 ist darstellbar als Linearkombination dieser beiden gut verstehbaren γ–Potenzen–Folgen.

Jetzt suchen wir dasjenige Element u von U_0, das die beiden Anfangsbedingungen $u_0 = 1$ und $u_1 = 1$ erfüllt. Wir suchen also reelle Zahlen α_1 und α_2 mit

$$u_0 = 1 = \alpha_1 \gamma_1^0 + \alpha_2 \gamma_2^0, \qquad u_1 = 1 = \alpha_1 \gamma_1^1 + \alpha_2 \gamma_2^1,$$

also $1 = \alpha_1 + \alpha_2$ und $1 = \alpha_1 \gamma_1 + \alpha_2 \gamma_2$, was zum Gleichungssystem

$$\begin{pmatrix} 1 & 1 \\ \gamma_1 & \gamma_2 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

führt. Nach Cramerscher Regel löst sich dieses System zu

$$\alpha_1 = \frac{\det \begin{pmatrix} 1 & 1 \\ 1 & \gamma_2 \end{pmatrix}}{\det \begin{pmatrix} 1 & 1 \\ \gamma_1 & \gamma_2 \end{pmatrix}} = \frac{\gamma_2 - 1}{\gamma_2 - \gamma_1} = \frac{-\gamma_1}{-\sqrt{5}} = \frac{\gamma_1}{\sqrt{5}}, \qquad \alpha_2 = \frac{\det \begin{pmatrix} 1 & 1 \\ \gamma_1 & 1 \end{pmatrix}}{\det \begin{pmatrix} 1 & 1 \\ \gamma_1 & \gamma_2 \end{pmatrix}} = \frac{1 - \gamma_1}{\gamma_2 - \gamma_1} = \frac{\gamma_2}{-\sqrt{5}}.$$

Die explizite Formel für die Fibonacci–Zahl u_n lautet dann

$$u_n = \alpha_1 \gamma_1^n + \alpha_2 \gamma_2^n = \frac{1}{\sqrt{5}} \left(\left(\frac{1 + \sqrt{5}}{2} \right)^{n+1} - \left(\frac{1 - \sqrt{5}}{2} \right)^{n+1} \right), \qquad n \in \mathbb{N}_0.$$

[1] wir beobachten $\gamma_1 + \gamma_2 = +1$ und $\gamma_1 \gamma_2 = -1$, wie es nach Vietascher Formel auch sein soll

[2] Eine Basis ist sehr oft dazu da, einen Vektorraum handhabbar zu machen. Sehr oft kann man erst dann in einem Vektorraum vernünftig rechnen, wenn man eine schöne Basis gefunden hat.

Anhang A

Anhang: Vom Mathematischen Schreiben

Angenommen,

ein Autor eines mathematikhaltigen Textes schreibt die Formel

$$x = e^y,$$

ohne einen weiteren Kommentar zu hinterlassen. Was könnte er gemeint haben ?

Vielleicht hatte er gemeint:

- Laut Voraussetzung ist $x = e^y$.
- Aus der Aufgabenstellung ergibt sich $x = e^y$ nach kurzer Rechnung.
- Wegen des Noether-Theorems haben wir $x = e^y$.
- Wir definieren $x \in \mathbb{C}$ durch $x := e^y$.
- Wir definieren $y \in \mathbb{R}$ durch $x =: e^y$. Dies ist durchführbar, denn aus dem Erxleben-Theorem folgt $x > 0$.
- Angenommen, es wäre $x = e^y$. Dann würde aber auch blafasel gelten, und daraus würde sich dann die sowieso-Formel ergeben, aber das wäre dann ein Widerspruch zu pongping. Also kann $x = e^y$ nicht sein.
- Wir zeigen jetzt als nächstes, daß $x = e^y$. Denn dies gilt wegen ....
- Wir wollen zeigen, daß $x = e^y$. Bis dahin ist es noch ein weiter Weg, also legen wir uns einige Lemmata bereit: und zwar haben wir zunächst...
- Die Gleichung (7) soll für alle x gelten. Also nehmen wir sie beim Wort, testen sie durch Einsetzen von e^y für die Variable x, und schauen, wohin uns das führt: ...
- Die Zeile $x = e^y$ ist der erste Fall einer Fallunterscheidung, und eine halbe Seite tiefer steht genauso unmotiviert $x \neq e^y$.

Wir wissen es nicht !

Erkenntnis

Der Leser möchte den Gedankengang des Autors, der Autorin nachvollziehen, denn der eigentliche Zweck des Lesens von wissenschaftlichen Texten ist ja, daß man hinterher mehr verstanden haben will als vorher.

Eine Formel ohne weiteren Begleittext ist mehrdeutig. Sie könnte ein Teil der Voraussetzung sein, oder ein Teil der Behauptung, oder eine schon bewiesene Aussage, oder eine Aussage, die im nächsten Moment bewiesen wird, oder eine Aussage, die im nächsten Moment widerlegt wird, oder ein Teil einer Fallunterscheidung, oder eine Definition (wobei nicht komplett klar ist, was hier definiert wird), oder eine Anleitung, einen durch die Formel gegebenen Wert irgendwo einzusetzen, oder was auch immer. Und je nachdem, welche Bedeutung nun zutrifft, ist der Gedankengang höchst unterschiedlich.

Es ist wenig charmant, das Publikum an dieser Stelle alleinzulassen. Es mag sein, daß der Leser es schafft, die gemeinte Bedeutung zu erraten, aber darauf sollte man sich besser nicht verlassen. Wenn der Leser schon kostbare Zeit opfert, sollte man ihn/sie entsprechend höflich behandeln ...

Und im Übrigen kann es gut sein, daß Autor und Leser identisch sind mit einer Phasenverschiebung von vielleicht einem halben Jahr. Dann ist es vorteilhaft, den eigenen Text gut aufgeschrieben zu haben, damit man mit möglichst wenig Zeitverlust die eigene Arbeit wieder nachvollziehen kann.

Empfehlenswerte Formulierungen

Niemand erwartet von Ihnen, daß Sie sich die Finger wundschreiben. Die gute Nachricht ist, daß man auch mit wenig Worten schon viel erreichen kann:

- Setze $\boxed{\cdots} := \boxed{\cdots}$.
- Sei ... .
- Betrachte ... .
- $\boxed{\cdots}$, also $\boxed{\cdots}$. (Der erste Kasten begründet den zweiten.)
- $\boxed{\cdots}$, denn $\boxed{\cdots}$. (Der zweite Kasten begründet den ersten.)
- Wäre $\boxed{\cdots}$, dann $\boxed{\cdots}$. (Das Wort „Wäre" am Satzanfang ist grammatisch ein *Konjunktiv Irrealis* und drückt aus, daß der Inhalt des ersten Kastens unglaubhaft oder unmöglich ist. Auf diese Weise wird also dem Leser ein indirekter Beweis angekündigt, und im zweiten Kasten oder kurz danach muß dann ein Widerspruch auftauchen.)
- Bew. d. Widerspruch. Ann.: ... .
- $\boxed{\cdots}$. Nun ist aber $\boxed{\cdots}$, und somit $\boxed{\cdots}$. (Die Formulierung „Nun ist aber" kündigt an, daß der zweite Kasten gegen den ersten ausgespielt werden wird. Die Vollendung dieses kleinen Bühnenstücks steht im dritten Kasten.)
- Laut Voraussetzung ist ... .
- Aus der Voraussetzung folgt ... . Das ist inhaltlich etwas anderes als die Formulierung im vorigen • !
- Wegen $\boxed{\cdots}$ ist $\boxed{\cdots}$.
- Bekanntlich ... .
- Offensichtlich ... . (Nur verwenden, wenn es wirklich offensichtlich ist. Ansonsten ist diese Formulierung unhöflich !)
- Einfaches Nachrechnen im Kopf zeigt ... .
- Wir zeigen jetzt ... .
- Zunächst ... . (Im Sinne einer Eröffnung)
- Ferner/Weiterhin ... . (Jetzt wird der Gedanke weitergesponnen)
- Schließlich ... . (Und an dieser Stelle wird der Gedankengang vollendet.)
- Setze $x = 0$, dann ... .
- Hinreichend ist ... .
- Notwendig ist ... .
- OBdA ... .
- ... für geeignete $x \in \mathbb{Z}$.

Die Auswahl ist unvollständig und Ergänzungsvorschläge sind willkommen.

Literaturverzeichnis

[1] K. Houston. *Wie man mathematisch denkt.* Springer Spektrum, 2012.

[2] G. Polya. *Schule des Denkens. Vom Lösen mathematischer Probleme.* Francke Verlag, 1980.

[3] H. Winter. Mathematikunterricht und Allgemeinbildung. *Mitteilungen der Gesellschaft für Didaktik der Mathematik*, 61:37–46, 1996.