Lineare Algebra

Dr. Stefan Kühnlein

Institut für Algebra und Geometrie, Karlsruher Institut für Technologie

Vorwort

Die Lineare Algebra ist eine mathematische Disziplin, die aus geometrischen Fragestellungen und aus dem Determinanten- und Matrizenkalkül entstand. Die eigentlichen geometrischen Objekte, die dabei studiert wurden, sind dabei ein wenig in den Hintergrund getreten; von daher ist auch die Analytische Geometrie in diesem Buch nicht sehr ausführlich behandelt worden. Ich hoffe, dass trotzdem gelegentlich die Geometrie immer wieder aufleuchtet, denn sie bildet einen Kontrapunkt zur sehr abstrakt vorgehenden algebraischen Sichtweise und ist oft ein guter Ideengeber. Die algebraische Sichtweise ist aber diejenige, die für die Bedeutung der Linearen Algebra in vielen Bereichen der Mathematik, Informatik und Physik entscheidend ist.

Dabei ist es wesentlich, gemeinsame Strukturen zu erkennen, die vielen verschiedenen Phänomenen zugrunde liegen. Diese Strukturen sind es, die in der Algebra thematisiert werden. Unsere Hauptstrukturen in der Linearen Algebra sind Gruppen, Körper und Vektorräume. Ein Vektor ist einfach ein Element eines Vektorraumes, und ohne den Begriff Vektorraum ist der Begriff Vektor sinnlos. Wichtig ist nicht ein einzelner Vektor, sondern die Gesamtheit und das Zusammenspiel aller Elemente eines Vektorraums, das sich in der Form des Additionsgesetzes und der skalaren Multiplikation ausdrückt.

Im Buch habe ich versucht, den Vorlesungsstil beizubehalten. Es war mir ein Anliegen, Begriffsbildungen durch Beispiele (ein sehr dehnbarer Begriff übrigens) zu motivieren. Oft wird sich auch ein Argumentationstyp an mehreren Stellen des Buches finden, damit sich ein gewisser Gewöhnungseffekt einstellen kann. Einige Beispiele, die erst mühsam sind, hätten an späterer Stelle weniger Arbeit erfordert. Das soll auch zeigen, dass es sich lohnt, den umfangreichen Begriffsapparat der Linearen Algebra aufzubauen, auch wenn gerade er für viele Neulinge ein großes Hindernis darstellt. Ich habe auch nicht immer zwischen Definitionen und Bemerkungen unterschieden, oft bot es sich an, unmittelbar in einer Definition noch ein Argument einzuarbeiten, das zur Klärung hilfreich ist. Andernorts habe ich Begriffe auch innerhalb eines Hilfssatzes definiert.

Die Überschriften, die ich den meisten Definitionen, Bemerkungen und Sätzen gegeben habe, sollen bei der Orientierung hilfreich sein, aber natürlich nicht den Rest der jeweiligen Nummer ersetzen. Im Zweifelsfall gilt immer der Volltext.

Da es nicht möglich ist, eine mathematische Disziplin zu erlernen, ohne viel zu üben, sind im Buch immer wieder Aufgaben eingestreut, die manchmal eine Routine einstudieren sollen, manchmal auch mehr Kreativität und Sitzfleisch erfordern.

Ich habe versucht, die Notation konsequent durchzuhalten. Viele Querverweise sollen das Nachschlagen erleichtern, erzwingen aber auch die teilweise sperrige oder pedantische Nummerierung.

Ich habe versucht, an machen Stellen über den Horizont der Linearen Algebra hinauszublicken und ihre Bedeutung für andere Disziplinen auch außerhalb der Mathematik zu beleuchten. Diese „Blicke über den Tellerrand“ sind nicht essentiell für den Fortgang des Buches, mögen aber als Anregungen und Motivationsquelle dienen.

Schließlich möchte ich es nicht Versäumen, meinen Hörern der Vorlesung in den vergangenen zwei Jahrzehnten zu danken, die mich durch ihre Aufnahme eines Skriptums darin bestärkt haben, dass es sich lohnt, Energie und Zeit in diese Arbeit zu stecken. Einige Kollegen – insbesondere Prof. Dr. Frank Herrlich und Dr. Hendrik Kasten – sollte ch erwähnen, die die unvermeidlichen Fehler entdeckt haben. Ich danke allen für ihre konstruktiven Hinweise. Ein paar Fehler sind sicher noch übrig geblieben; diese bitte ich mir anzukreiden und mitzuteilen.

An vielen Stellen des Buches sind Ideen mit eingeflossen, die sich vor einigen Jahren bei der Zusammenarbeit mit Prof. Herrlich anlässlich der damaligen parallel gehaltenen Vorlesungen ergaben. Auch dafür will ich meinen Dank nicht verhehlen. Schließlich habe ich auch die Vorlesung von Prof. Dr. Wulf-Dieter Geyer, die ich selbst als Student hören durfte, bei manchen Argumenten verwendet.

Karlsruhe, Stefan Kühnlein

Inhaltsverzeichnis

Kapitel 1

Allgemeine Grundlagen

In diesem Kapitel sollen einige Tatsachen und vor allem Ausdrucksweisen sowie Notationen der Logik und der Mengenlehre vermittelt werden. Dabei werden wir den Mengenbegriff nicht problematisieren, also im Bereich der so genannten naiven Mengenlehre verbleiben. Für die Zwecke der Linearen Algebra reicht dies vollkommen aus, manche Leser werden später noch sehen, dass dies nicht alles ist, was die Mengenlehre zu bieten hat.

1.1 Logisches

Die Logik beschäftigt sich mit Aussagen. Das sind Sätze, die entweder wahr oder falsch sind. Fragesätze wie zum Beispiel „Meinst Du, dass es morgen regnet?“ sind keine Aussagen. Auch Befehle wie „Komm sofort her!“ sind keine Aussagen. Beide Beispielsätze haben keinen „Wahrheitswert“.

Im Gegensatz dazu ist ein seltsam anmutender Satz wie „Wenn 2 ungerade ist, dann ist 1 gleich 0.“ eine Aussage. Noch dazu ist diese Aussage wahr, denn die Bedingung, an die der zweite Satzteil geknüpft ist, wird niemals eintreten.

Schlichtere Aussagen sind zum Beispiel die folgenden: „Alle Quadrate sind rund.“ „Draußen regnet es.“ „Ich habe heute Geburtstag.“ Die zwei letzteren Aussagen beziehen sich (direkt oder indirekt) auf einen Zeitpunkt. In der Mathematik werden wir es immer mit Aussagen zu tun haben, deren Wahrheitswert für alle Zeiten ungeändert bleibt (zumindest idealer Weise).

Natürlich ist man nicht unbedingt an jeder einzelnen Aussage für sich interessiert, sondern eher an Zusammenhängen zwischen verschiedenen Aussagen. Die Logik hat einige Möglichkeiten, aus vorhandenen Aussagen neue zu machen, formalisiert. Stellen Sie sich also vor, Sie hätten zwei Aussagen A und B aus einer großen Kiste mit Aussagen herausgezogen und wollten aus diesen neue Aussagen basteln. Dazu gibt es einige einfache Möglichkeiten, die natürlich auch im Alltag

verwendet werden, für unsere Zwecke aber formal etwas enger zu fassen sind.

Bemerkung 1.1.1 (Verknüpfung von Aussagen)

(a) Die **Konjunktion** $A \wedge B$:

Diese Aussage ist wahr, wenn A und B beide wahr sind, ansonsten ist sie falsch. Oft werden wir auf die symbolische Notation mit dem $\wedge$ verzichten und stattdessen so etwas wie „A und B“, „sowohl A als auch B“ oder „A sowie B“ schreiben.

(b) Die **Negation** $\neg A$:

Diese Aussage ist wahr, wenn A falsch ist und falsch, wenn A wahr ist. Die Negation von „mein Fahrrad ist schwarz“ ist nicht „mein Fahrrad ist weiß“, sondern – entgegen allem Schwarz-Weiß-Denken – „mein Fahrrad ist nicht schwarz“. Die Aussage $A \wedge (\neg A)$, die durch Konjunktion der beiden Aussagen A und $\neg A$ gebildet wird, ist immer falsch. Wahr ist:

$$\boxed{\neg[A \wedge (\neg A)].}$$

(c) Die **Disjunktion** $A \vee B$:

Diese Aussage ist wahr, wenn A wahr ist oder B wahr ist oder auch beide wahr sind. Sie ist falsch, wenn sowohl A als auch B falsch sind. Wir sagen oft auch „A oder B“. Im allgemeinen Sprachgebrauch meint man damit oft das ausschließende oder, also das „Entweder - Oder“. In der Mathematik wird das „oder“ immer im nicht ausschließenden Sinn verwendet. Es ist also $A \vee B$ dieselbe Aussage wie $\neg((\neg A) \wedge (\neg B))$. Demnach ist zum Beispiel die Aussage $A \vee (\neg A)$ für jede Aussage A wahr: wenn A wahr ist, ist sie wahr, und wenn A falsch ist, ist ja $\neg A$ wahr und damit auch einer der beiden Partner in $A \vee (\neg A)$ wahr.

An solchen Beispielen sieht man schon, dass es oft sinnvoll ist, in längeren Aussagengefügen die Zutaten durch Klammern zusammenzufassen, sodass die Struktur überhaupt erkennbar ist. So ist zunächst nicht klar, was die Aussage $A \wedge B \vee C$ bedeutet. Dafür gibt es ja die zwei Möglichkeiten

$$(A \wedge B) \vee C \quad \text{bzw.} \quad A \wedge (B \vee C).$$

Wenn A falsch und C wahr ist, dann ist die linke Aussage wahr, aber die rechte falsch.

Die Klammern geben dabei an, in welcher Reihenfolge die Aussagen verknüpft werden.

Bitte lassen Sie sich durch ein langes Klammerngewusel nicht abschrecken, sondern nehmen Sie es als Grundgerüst zur Auflösung einer längeren Aussage!

(d) Die **Implikation** $A \Rightarrow B$:

„aus A folgt B“, „wenn A wahr ist, so auch B“. Die Implikation ist wahr, wenn entweder A falsch ist oder sowohl A als auch B wahr sind. Dies ist eine Formalisierung der Tatsache, dass die Voraussetzung A die Folgerung B nach sich zieht. Also ist $A \Rightarrow B$ dieselbe Aussage wie $\neg A \vee (A \wedge B)$, oder auch dieselbe wie $(\neg A) \vee B$. Als Beispiel sei noch einmal der Satz „Wenn 2 ungerade ist, dann ist 1 gleich 0“ benutzt. Auch der Satz „Wenn 2 ungerade ist, dann ist 1 gleich 1“ ist eine wahre Aussage, nicht aber der Satz „Wenn 1 gleich 1 ist, dann ist 2 ungerade.“ Hier ist ja die Voraussetzung wahr, aber die Folgerung falsch.

Bemerkung 1.1.2 (Widerspruchsprinzip)

Eine der wichtigsten Tatsachen der klassischen Logik ist das **Widerspruchsprinzip**. Es sagt, dass die Aussage $A \Rightarrow B$ dasselbe bedeutet wie die Aussage $(\neg B) \Rightarrow (\neg A)$. Dies gilt, denn $A \Rightarrow B$ bedeutet $\neg A \vee B$ und $\neg B \Rightarrow \neg A$ bedeutet $\neg(\neg B) \vee \neg A$, wobei wir ja $\neg(\neg B)$ durch B ersetzen dürfen.

Alternativ können wir das durch Angabe von **Wahrheitstafeln** der logischen Verknüpfungen erkennen. Dabei werden in einer Tabelle die möglichen Wahrheitswerteverteilungen der Aussagen A, B und der betrachteten Verknüpfung aufgestellt. Beispiele:

A	w	w	f	f
B	w	f	w	f
$\neg A$	f	f	w	w
$\neg B$	f	w	f	w
$A \Rightarrow B$	w	f	w	w
$\neg B \Rightarrow \neg A$	w	f	w	w

Die letzte Zeile ergibt sich hierbei aus den Wahrheitswerten von $\neg A$ und $\neg B$. Sie stimmt mit der vorletzten überein. Dies ist die Grundlage dafür, dass in der Mathematik Beweise immer wieder durch Widerspruch geführt werden: wenn die Wahrheit der Implikation $A \Rightarrow B$ zu zeigen ist, dann nimmt man an, B sei falsch, und kann daraus mit etwas Glück folgern, dass dann auch A falsch sein muss. Wenn aber A wie angenommen richtig ist, muss demnach die Annahme, B sei falsch, einen Widerspruch darstellen, also muss B auch wahr sein. Beispiele hierfür werden wir noch häufig zu sehen bekommen.

Hilfreich ist diese Beweistechnik dann, wenn die Annahme der Falschheit von B eine Denkrichtung vorgibt, die unter der Annahme der Wahrheit von A vielleicht nicht so naheliegend ist.

Viele Aussagen sind von der Form „alle X haben die Eigenschaft Y“. In unserer

Sprache könnte man das schreiben als

$$(m \text{ ist X}) \Rightarrow (m \text{ hat Y}).$$

Dies lässt sich dann per Widerspruch beweisen, indem man zeigt, dass jedes Objekt m, das die Eigenschaft Y nicht hat auch kein X ist. Dazu bedarf es natürlich scharfer Definitionen für Y und X, und genau dies ist eine der großen Stärken der Mathematik. Ein wichtiges sprachliches Mittel hierfür ist die Sprache der Mengenlehre, die wir im nächsten Abschnitt einführen.

Bemerkung 1.1.3 (Äquivalenz von Aussagen)

Die **Äquivalenz** zweier Aussagen A, B bedeutet, dass sie den selben Wahrheitswert haben. Wir schreiben dann $A \Leftrightarrow B$. Dies ist eine Kurzschreibweise für

$$(A \Rightarrow B) \wedge (B \Rightarrow A).$$

Zum Beispiel sind für eine natürliche Zahl n die Aussagen „n ist gerade“ und „$n+1$ ist ungerade“ äquivalent, was Sie alle wissen und was auch leicht bewiesen werden kann.

Wir halten in dieser Notation noch einmal die Grundregel des Widerspruchsbeweises fest:

$$\boxed{(A \Rightarrow B) \iff (\neg B \Rightarrow \neg A).}$$

Beispiel 1.1.4 (Ein Distributivgesetz)

Wir wollen einsehen, dass für drei Aussagen A, B, C stets gilt:

$$[A \wedge (B \vee C)] \Leftrightarrow [(A \wedge B) \vee (A \wedge C)].$$

Das heißt: Wir müssen zeigen, dass die beiden Aussagen in eckigen Klammern für alle Möglichkeiten von A, B, C den gleichen Wahrheitswert haben. Da es dabei nur auf die Wahrheitswerte von A, B, C ankommt, bedienen wir uns der Wahrheitstafeln. Es gibt 8 mögliche Verteilungen der Wahrheitswerte für A, B, C :

A	w	w	w	w	f	f	f	f
B	w	w	f	f	w	w	f	f
C	w	f	w	f	w	f	w	f
$B \vee C$	w	w	w	f	w	w	w	f
$A \wedge (B \vee C)$	w	w	w	f	f	f	f	f
$A \wedge B$	w	w	f	f	f	f	f	f
$A \wedge C$	w	f	w	f	f	f	f	f
$(A \wedge B) \vee (A \wedge C)$	w	w	w	f	f	f	f	f

Das zeigt die gewünschte Übereinstimmung der Wahrheitswerte der Aussagen in allen Situationen.

1.2 Mengen

Wir stellen uns auf den naiven Standpunkt: Eine **Menge** M ist eine Ansammlung von Objekten (was auch immer das ist; Dinge, Aussagen, andere Mengen, Abbildungen), sodass von jedem Objekt x prinzipiell entschieden werden kann, ob es zu M gehört oder nicht.

Statt „x gehört zu M“ schreibt man meistens kurz $x \in M$. Statt „x gehört nicht zu M“ schreibt man entweder (selten) $\neg(x \in M)$ oder meistens $x \notin M$.

Ein wichtiges Beispiel einer Menge ist die leere Menge $\emptyset$[1]. Das ist die Menge, bei der für alle x gilt, dass sie nicht dazu gehören.

$$\boxed{\text{Es gilt für alle } x : x \notin \emptyset.}$$

Man könnte zum Beispiel $\emptyset$ als die Menge aller viereckigen Kreise definieren.

Viele Leuten erscheint diese Menge vielleicht überflüssig zu sein. Aber sie ist insofern notwendig, als sie uns immer wieder dazu verhilft, Fallunterscheidungen zu vermeiden, die ohne sie notwendig wären.

Die folgenden Mengen werden wir als bekannt voraussetzen:

$\mathbb{N} = \{1, 2, 3, 4 \dots\}$, die Menge der natürlichen Zahlen.

$\mathbb{N}_0 = \{0, 1, 2, 3, 4 \dots\}$.

$\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$, die Menge der ganzen Zahlen.

$\mathbb{Q}$, die Menge der rationalen Zahlen, und $\mathbb{R}$, die Menge der reellen Zahlen.

„Kleine“ Mengen können durch die Angabe aller zugehörigen Elemente angegeben werden. Zum Beispiel schreibt man die Menge M, deren Elemente die Zahlen 2, 3, 5 und 7 sind, als $M := \{2, 3, 5, 7\}$.

Dabei bedeutet der **Doppelpunkt** beim Gleichheitszeichen, dass die auf der Seite des Doppelpunktes befindliche Größe durch die Größe auf der anderen Seite definiert wird.

Man könnte hier M auch definieren als die Menge aller Primzahlen, die nicht größer als 10 sind. (Eine Primzahl ist eine natürliche Zahl ≥ 2, die sich nicht als Produkt von kleineren natürlichen Zahlen schreiben lässt.)

Eine inhaltliche Charakterisierung der Elemente einer Menge wird umso wichtiger, je komplexer die Menge ist. Man schreibt zum Beispiel in unserem Fall:

$$M := \{n \mid n \text{ ist Primzahl und } 1 \leq n \leq 10\}.$$

In diesem Sinne werden Mengen meistens dadurch angegeben, dass man charakteristische Eigenschaften ihrer Elemente nennt. Zwischen den Zeilen haben wir

[1] Das sollten Sie nicht mit der Null verwechseln!

das eben gesehen bei der Definition der Primzahlen. Um dies aber jetzt noch eleganter aufzuschreiben, brauchen wir ein Symbol, den **Allquantor** $\forall$. Er wird verwendet, um zu sagen, dass für alle Objekte x mit einer bestimmten Eigenschaft eine Aussage $A(x)$ gilt. Statt zum Beispiel zu sagen: „Für jede natürliche Zahl n ist auch $n+1$ eine natürliche Zahl“ schreibt man kurz

$$\forall n \in \mathbb{N} : n+1 \in \mathbb{N}.$$

Mit dem Allquantor kann man die Menge aller Primzahlen definieren durch

$$\mathbb{P} := \{n \in \mathbb{N} \mid 2 \leq n \wedge \forall a, b \in \mathbb{N} : [(a < n) \wedge (b < n)] \Rightarrow a \cdot b \neq n\}.$$

Wir werden allerdings versuchen, den Inhalt mathematischer Formeln nicht durch eine Überfrachtung mit Notation unkenntlich zu machen. Trotzdem sei an dieser Stelle auch noch auf den **Existenzquantor** $\exists$ hingewiesen, den man verwendet, um zu sagen, dass es mindestens ein Objekt x mit einer speziellen Eigenschaft gibt. Also: statt „es gibt (mindestens) eine Primzahl, die bei Division durch 4 den Rest 1 lässt und bei Division durch 7 den Rest 6“ könnte man zum Beispiel schreiben:

$$\exists p \in \mathbb{P} : [\exists m, n \in \mathbb{N} : p = 4m + 1 \text{ und } p = 7n + 6].$$

Dabei ist gleichzeitig miterklärt, was es heißt, Rest 1 (oder 6) nach Division durch 4 (oder 7) zu lassen. Die Richtigkeit einer solchen Aussage hat man zum Beispiel gezeigt, indem man eine solche Primzahl (etwa $13 = 4 \cdot 3 + 1 = 7 \cdot 1 + 6$) angibt.

Manchmal aber liegen die Dinge so verzwickt, dass man zwar abstrakt zeigen kann, dass es ein x mit der und der Eigenschaft gibt, aber trotzdem kein einziges Beispiel dafür angeben kann. Dann spricht man von einem **reinen Existenzbeweis**. Man kann zum Beispiel zeigen, dass es für beliebige natürliche Zahlen $a < b$ mit größtem gemeinsamen Teiler 1 eine Primzahl gibt, die bei Division durch b den Rest a lässt[2], kann solch eine Primzahl p jedoch nicht explizit für alle möglichen Wahlen von a und b *konstruieren*. Für jede feste Wahl von a und b findet sich trotzdem oft sehr schnell eine solche Primzahl (was natürlich noch kein Beweis ist – dieser sieht ganz anders aus und wird mit analytischen Methoden geführt).

Definition 1.2.1 (Teilmenge, Mengengleichheit)

Eine Menge N heißt **Teilmenge** der Menge M, falls alle ihre Elemente auch in M liegen:

$$\forall x : (x \in N \Rightarrow x \in M).$$

Dann schreibt man $N \subseteq M$ oder $M \supseteq N$.

[2]Diese Aussage heißt *Dirichlets Primzahlsatz*

Zwei Mengen sind **gleich**, wenn sie sich gegenseitig als Teilmengen enthalten. In Zeichen:

$$M = N :\Longleftrightarrow (M \subseteq N \wedge N \subseteq M).$$

Bevor wir das an einem Beispiel illustrieren, führen wir noch ein paar Operationen mit Mengen ein, mit denen sich neue Mengen bilden lassen.

Definition 1.2.2 (Durchschnitt, Vereinigung, Produkt, Tupel)

Für zwei Mengen M und N treffen wir die folgenden Definitionen.

(a) Der **Durchschnitt** $M \cap N$ ist definiert als

$$M \cap N := \{x \mid x \in M \text{ und } x \in N\}.$$

(b) Die **Vereinigung** $M \cup N$ ist definiert als

$$M \cup N := \{x \mid x \in M \text{ oder } x \in N\}.$$

(c) Die **Differenzmenge** $M \setminus N$ ist definiert als

$$M \setminus N := \{x \mid x \in M \text{ und } x \notin N\}.$$

(d) Das **kartesische Produkt** $M \times N$ ist definiert als die Menge aller geordneten Paare mit einem ersten Eintrag aus M und einem zweiten aus N. In Zeichen:

$$M \times N := \{(m, n) | m \in M \text{ und } n \in N\}.$$

(e) Für $k \in \mathbb{N}$ setzen wir

$$M^k := \{(m_1,\ m_2,\ \ldots,\ m_k) \mid \forall i :\ m_i \in M\}$$

und nennen die Elemente von M^k auch k-**Tupel** in M. Später werden wir die Elemente von M^k oft als „Spalten“ schreiben, also als

$$(m_1\ m_2\ \ldots\ m_k)^\top := \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_k \end{pmatrix}.$$

Dabei steht das Symbol $^\top$ für die **Transposition**, siehe 4.1.13.

Beispiel 1.2.3 (Konkreter)

Wir betrachten die beiden Teilmengen der rationalen Zahlen

$$A = \{\frac{m}{2} \mid m \in \mathbb{N}\}, \quad \text{und} \quad B = \{\frac{n}{3} \mid n \in \mathbb{N}\}.$$

Behauptung: $A \cap B = \mathbb{N}$.

Um das zu belegen müssen wir beide Inklusionen zeigen.

$A \cap B \supseteq \mathbb{N}$ ist klar, da für $k \in \mathbb{N}$ auch $m = 2k \in \mathbb{N}$ und $n = 3k \in \mathbb{N}$, und mit diesen Wahlen ist

$$k = \frac{m}{2} \in A \quad \text{und} \quad k = \frac{n}{3} \in B,$$

also $k \in A \cap B$.

$A \cap B \subseteq \mathbb{N}$ ist etwas subtiler und kann so gezeigt werden: Sei $x \in A \cap B$. Dann lässt sich x schreiben als

$$x = \frac{m}{2} = \frac{n}{3},$$

wobei m, n geeignete natürliche Zahlen sind. Es folgt $3m = 2n$. Da demnach $3m$ gerade sein muss, ist auch m gerade, und es folgt $x = \frac{m}{2} \in \mathbb{N}$. ◯

Aufgabe 1.2.4 (Noch mehr Brüche)

Zeigen Sie, dass die beiden folgenden Mengen gleich sind:

$$M = \{\frac{a}{2} + \frac{b}{3} \mid a, b \in \mathbb{Z}\} \text{ und } N = \{\frac{c}{6} \mid c \in \mathbb{Z}\}.$$

Gilt auch Gleichheit der beiden Mengen

$$\tilde{M} = \{\frac{a}{2} + \frac{b}{3} \mid a, b \in \mathbb{N}\} \text{ und } \tilde{N} = \{\frac{c}{6} \mid c \in \mathbb{N}\}?$$

Definition/Bemerkung 1.2.5 (Die Potenzmenge)

Für eine Menge M heißt die Menge, deren Elemente die Teilmengen von M sind, die **Potenzmenge** von M:

$$\mathcal{P}(M) := \{x \mid x \subseteq M\}.$$

Zum Beispiel gilt

$$\mathcal{P}(\{0,1,2\}) = \big\{\emptyset, \{0\}, \{1\}, \{2\}, \{0,1\}, \{0,2\}, \{1,2\}, \{0,1,2\}\big\}.$$

Wichtig ist hierbei, dass $M \in \mathcal{P}(M)$ immer gilt, aber $M \subseteq \mathcal{P}(M)$ in aller Regel falsch ist. Das hängt eng damit zusammen, dass man zwischen $\in$ und $\subseteq$ streng unterscheiden muss.

$$\boxed{\emptyset \subseteq \emptyset, \text{ aber } \emptyset \notin \emptyset.}$$

Beispiel 1.2.6 (Graphen)

Ein **Graph** – hier ist nicht der Funktionsgrapgh gemeint – wird in diesem Buch immer verstanden als ein einfacher schleifenfreier und ungerichteter Graph. Damit meint man eine Menge E von sogenannten **Ecken**, von denen einige durch **Kanten** verbunden sind. Dabei soll eine Kante immer zwei verschiedene Ecken verbinden („schleifenfrei“) und zwei verschiedene Ecken sollen durch höchstens eine Kante verbunden sein („einfach“).

Ein Lehrbuch zur Graphentheorie ist zum Beispiel das Buch *Graphentheorie* von Reinhard Diestel, Springer 2017.

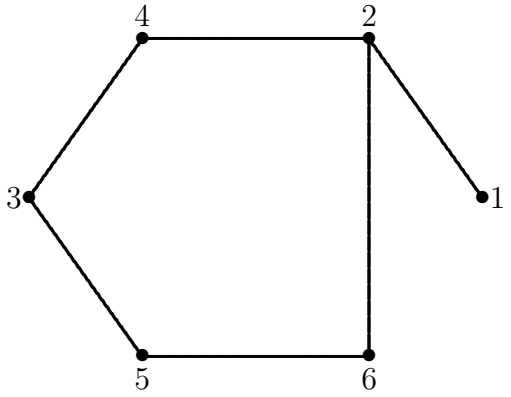
Graphen treten in vielen Bereichen auf, zum Beispiel bei Streckenplänen von Verkehrsgesellschaften oder anderen Netzwerken, sie können dann je nach Kontext auch nicht einfach oder nicht ungerichtet sein, aber wir bleiben der Einfachheit halber bei unseren Prämissen.

Formal kann ein solcher Graph definiert werden als eine Menge E zusammen mit einer Teilmenge von $K \subset \mathcal{P}(E)$, die aus (einigen) zweielementigen Teilmengen besteht. Man schreibt dann auch, dass $\Gamma = (E, K)$ gilt.

Der Graph heißt endlich, wenn die Menge der Ecken endlich ist. Dann kann man die Ecken durchnummerieren. Ein Beispiel für einen Graphen mit 6 Ecken ist gegeben durch

$$E = \{1, 2, 3, 4, 5, 6\}, \quad K = \{\{1, 2\}, \{2, 4\}, \{2, 6\}, \{3, 4\}, \{3, 5\}, \{5, 6\}\}.$$

Er lässt sich bildlich darstellen wie folgt:



Ein Graph wie dieser, der sich durch eine Punktmenge mit überschneidungsfreien Verbindungslinien, die übrigens nicht gerade sein müssen, in der Ebene darstellen lässt, heißt **planarer Graph**. Nicht jeder Graph hat diese Eigenschaft.

Definition 1.2.7 (Durchschnitt relaoded)

Nun sei eine nichtleere Menge I gegeben, und für jedes $i \in I$ eine Menge M_i.

Dann setzt man

$$\bigcap_{i\in I} M_i := \{x \mid \forall i \in I : x \in M_i\} \text{ und } \bigcup_{i\in I} M_i := \{x \mid \exists i \in I : x \in M_i\}.$$

Dies sind die naheliegenden Verallgemeinerungen der Durchschnitte und Vereinigungen zweier Mengen.

Manchmal liegt der Spezialfall vor, dass (die sogenannte **Indexmenge**) I selber schon eine Teilmenge von $\mathcal{P}(M)$ ist für eine Menge M. Dann hat man die Mengen

$$\bigcap_{i\in I} i, \quad \bigcup_{i\in I} i$$

als Durchschnitt und Vereinigung, auch wenn dies etwas gewöhnungsbedürftig aussieht.

Als letzte Notation führen wir noch eine Schreibweise für die Anzahl der Elemente einer Menge ein.

Definition 1.2.8 (Mächtigkeit)

Die Anzahl der Elemente einer Menge M nennt man die **Mächtigkeit** oder auch die **Kardinalität** von M. Wir schreiben dafür $|M|$ oder auch $\#M$.

Diese Anzahl kann endlich oder auch unendlich sein, und tatsächlich muss man verschiedene Typen der Unendlichkeit unterscheiden, aber das wird für die Lineare Algebra keine Rolle spielen, und wir bleiben auch bei diesem Begriff im Naiven und verweisen die Neugierigen auf Literatur, wie etwa *Einführung in die Mengenlehre* von Ebbinghaus, 2003.

Bemerkung 1.2.9 (Prinzip der vollständigen Induktion)

Eine Teilmenge $S \subseteq \mathbb{N}$, die nicht leer ist, enthält mindestens ein Element n. Da es nur endlich viele natürliche Zahlen gibt, die kleiner sind als n, gibt es auch in S nur endlich viele solcher Zahlen, also enthält S ein kleinstes Element s_0. Wenn nun für jedes $n \in S$ gilt, dass auch $n+1$ in S liegt, dann ist

$$\{s_0, s_0+1, s_0+2, \dots\} \subseteq S,$$

und da es sonst keine natürlichen Zahlen gibt, die größer sind als s_0, gilt hier sogar

$$\{s_0, s_0+1, s_0+2, \dots\} = S.$$

Das begründet das Prinzip der **vollständigen Induktion.** Um zu zeigen, dass eine von $n \in \mathbb{N}$ abhängige Aussage $A(n)$ für alle natürlichen Zahlen n ab einer gegebenen natürlichen Zahl N gilt, setzt man

$$S := \{n \in \mathbb{N} \mid n \geq N \text{ und } A(n) \text{ wahr}\}.$$

Nun hat man also noch zu zeigen, dass

$$N \in S \text{ und } \forall n : (n \in S) \Rightarrow (n+1 \in S).$$

Beispiel 1.2.10 (Zweielementige Teilmengen)

Wir behaupten, dass für jede natürliche Zahl n die Menge $M = \{1, 2, \ldots, n\} = \{x \in \mathbb{N} \mid x \leq n\}$ genau $n(n-1)/2$ zweielementige Teilmengen hat.

Dies kann man mit vollständiger Induktion beweisen. Für $n = 1$ gibt es offensichtlich keine zweielementige Teilmengen von $M = \{1\}$ – das ist der Induktionsanfang. Wenn die Aussage für ein n stimmt, so ist jede zweielementige Teilmenge von $\{1, 2, \ldots, n+1\}$ entweder schon in $\{1, 2, \ldots, n\}$ enthalten (davon gibt es also nun $n(n-1)/2$ viele) oder sie ist von der Gestalt $\{x, n+1\}$, $x \leq n$ – davon gibt es n Stück. Insgesamt hat $\{1, 2, \ldots, n+1\}$ also

$$n(n-1)/2 + n = n(n+1)/2$$

zweielementige Teilmengen, was genau die Behauptung ist.

Insbesondere heißt das zusammen mit der folgenden Aufgabe, dass es $2^{n(n-1)/2}$ verschiedene Strukturen eines einfachen ungerichteten und schleifenfreien Graphen (siehe 1.2.6) mit der Eckenmenge $\{1,\ 2, \ldots, n\}$ gibt.

Aufgabe 1.2.11 (Alle Teilmengen)

Zeigen Sie mit vollständiger Induktion, dass für jede natürliche Zahl n gilt: Die Menge $\{1, 2, \ldots, n\}$ hat genau 2^n Teilmengen.

Beispiel 1.2.12 (Die Existenz von Primfaktoren)

Wir behaupten, dass jede natürliche Zahl sich als Produkt von Primzahlen schreiben lässt, d.h.

$$\forall n \in \mathbb{N} : \exists k \in \mathbb{N}_0, p_1, \ldots, p_k \in \mathbb{P} : n = p_1 \cdot \ldots \cdot p_k.$$

Der Fall $k = 0$ bedeutet hierbei, dass das Produkt keinen Faktor hat, und man setzt dann das Produkt gleich 1.

Um die Behauptung per Induktion beweisen zu können, formulieren wir die Aussage scheinbar allgemeiner in Abhängigkeit von einer natürlichen Zahl N :

$$A(N) := \left[\forall n \in \mathbb{N} : \left[n \leq N \Rightarrow \exists k \in \mathbb{N}_0, p_1, \ldots, p_k \in \mathbb{P} : n = p_1 \cdot \ldots \cdot p_k\right]\right].$$

Das heißt: welche Zahl N auch immer wir vorgeben, jede natürliche Zahl n, die höchstens N ist, ist Produkt von Primzahlen.

$A(1)$ ist also korrekt, da ja 1 das (wenn auch leere) Produkt von Primzahlen ist und keine andere natürliche Zahl n unterhalb von $N = 1$ existiert.

Wir nehmen nun an, $A(N)$ sei für irgendeine Zahl N korrekt, und wollen daraus $A(N+1)$ folgern. Da wir aber wegen $A(N)$ schon von allen $n \leq N$ wissen, dass sie Produkte von Primzahlen sind, langt es, dies noch für $n = N+1$ zu zeigen.

Dazu gibt es zwei Fälle zu unterscheiden:

Fall 1: $N+1$ ist eine Primzahl: Dann setzen wir $k = 1,\ p_1 = N+1$ und sehen: $N+1 = p_1$ ist ein Produkt (mit nur einem Faktor).

Fall 2: $N+1$ ist keine Primzahl. Da $N+1 > N \geq 1$ gilt, ist daher $N+1$ in zwei Faktoren zerlegbar, die kleiner sind:

$$\exists m, n \in \mathbb{N} : m, n < N+1,\ N+1 = m \cdot n.$$

Da m, n also höchstens N sind, greift die Voraussetzung $A(N)$ und wir können m und n als Produkte von Primzahlen schreiben, und damit auch $m \cdot n = N+1$.

1.3 Abbildungen

Wir machen erst eine Art Absichtserklärung und sagen, was wir uns unter einer Abbildung f zwischen zwei Mengen M und N vorstellen: Es soll eine „Vorschrift“ sein, die jedem $m \in M$ ein $n \in N$ zuordnet.

Da wir nicht zu eng fassen wollen, wie so eine „Vorschrift“ aussehen soll, gehen wir das anders an und definieren etwas weniger eingänglich, aber dafür sehr weitreichend:

Definition 1.3.1 (Abbildung)

Eine **Abbildung** f zwischen zwei Mengen M und N ist eine Teilmenge $f \subseteq M \times N$, sodass für alle $m \in M$ genau ein $n \in N$ existiert, sodass $(m, n) \in f$. Für dieses n schreiben wir kurz $n = f(m)$.

M heißt der **Definitionsbereich** von f, N heißt der **Wertebereich**. Die Menge aller Abbildungen von M nach N bezeichnen wir mit $\mathrm{Abb}(M, N)$. Auch die Notation N^M ist hierfür gebräuchlich.

Eine Abbildung im Sinne unserer obigen Definition ist also „eigentlich“ der aus der Schule bekannte Funktionsgraph. Es ist in der Tat schwierig, die Absichtserklärung anders zu präzisieren. Wenn einmal die präzise Definition gemacht ist, schreibt man dafür dann doch wieder

$$f : M \longrightarrow N,\ m \mapsto f(m)$$

oder auch etwas verkürzend

$$M \ni m \mapsto f(m) \in N.$$

Wie gesagt: $f(m)$ ist <u>das</u> Element von N, für das $(m, f(m)) \in f$ gilt. Die Zugheörigkeit von $(m, f(m))$ zur Menge f ist also die Abbildungsvorschrift.

Beispiel 1.3.2 (Woher kommt die andere Notation?)

Wir wollen verstehen, wieso die Menge aller Abbildungen von M nach N auch als N^M geschrieben wird.

Dazu seien $N = \{1, \ldots, n\}, M = \{1, \ldots, m\}$ gegeben. Eine Abbildung von M nach N ist dann eine Liste von Paaren – eine Wertetabelle! –

$$(1, f(1)),\ (2, f(2)), \ldots, (m, f(m))$$

und die einzige Bedingung dabei ist, dass die Werte $f(1), \ldots, f(m)$ alle in N liegen. Da diese ansonsten vollkommen unabhängig gewählt werden können, entsteht eine solche Abbildung durch die Wahl von jeweils einem aus n Kandidaten für jedes $x \in M$, und dafür gibt es insgesamt n^m Möglichkeiten.

Das motiviert die andere Notation, die zuweilen etwas suggestiver ist als das sperrige $\mathrm{Abb}(M, N)$.

Definition 1.3.3 (Gleichheit von Abbildungen, Identität)

a) Zwei Abbildungen $f, g : M \longrightarrow N$ sind **gleich**, wenn für alle $m \in M$ die Gleichheit $f(m) = g(m)$ gilt. Das bedeutet gerade, dass die entsprechenden Teilmengen von $M \times N$ gleich sind.

b) Die Abbildung $\mathrm{Id}_M : M \longrightarrow M$, die durch

$$\forall m \in M : \mathrm{Id}_M(m) := m$$

definiert ist, heißt die **Identität auf** M**.**

Zum Beispiel sind die zwei reellwertigen Abbildungen f und g, die auf $\{0, 1, -1\}$ durch

$$f(x) := 0 \ \text{ und } \ g(x) := x^3 - x$$

definiert sind, gleich, auch wenn sie zunächst verschieden angegeben werden. Es gibt ja nur drei erlaubte Argumente, und für diese sieht man leicht, dass f und g dieselben Werte annehmen.

Wenn zwei Abbildungen $f : M \longrightarrow N$ und $g : N \longrightarrow O$ vorliegen, so kann man diese Abbildungen zusammensetzen (man sagt auch verknüpfen, komponieren oder hintereinander ausführen) und damit eine neue Abbildung $g \circ f$ (sprich: „g nach f“) von M nach O definieren. Erst einmal machen wir das formal als Graph, wie es die Definition einer Abbildung verlangt:

Definition 1.3.4 (Komposition von Abbildungen)

In der eben beschriebenen Situation ist die **Komposition** $g \circ f$ definiert durch

$$g \circ f := \{(m, o) \in M \times O | \exists n \in N : (m, n) \in f \text{ und } (n, o) \in g\}.$$

Da n hier das eindeutig festliegende $n = f(m)$ ist und $o = g(n)$ auch durch n, und damit durch m eindeutig festgelegt wird, ist klar, dass diese Menge wieder die Eigenschaften aus der Definition einer Abbildung besitzt. Es gilt

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

Wenn $f : M \longrightarrow N$, $g : N \longrightarrow O$ und $h : O \longrightarrow P$ Abbildungen sind, so kann man auf zwei Arten die Hintereinanderausführung bilden:

$$(h \circ g) \circ f \quad \text{oder } h \circ (g \circ f).$$

Es ist offensichtlich, dass beide Möglichkeiten zum selben Ergebnis führen:

$$\begin{array}{lll} \forall m \in M : & & ((h \circ g) \circ f)(m) \\ & = & (h \circ g)(f(m)) = h(g(f(m))) = h((g \circ f)(m)) \\ & = & (h \circ (g \circ f))(m). \end{array}$$

Dabei wird in jedem Schritt nur die Definition von $\circ$ benutzt.

Als Konsequez ergibt sich die Assoziativität der Komposition, die uns immer wieder als wichtiges Hilfsmittel begegnen wird:

Fazit 1.3.5 (Assoziativität der Komposition von Abbildungen)

Wenn $f : M \longrightarrow N$, $g : N \longrightarrow O$ und $h : O \longrightarrow P$ Abbildungen sind, so gilt

$$\boxed{(h \circ g) \circ f = h \circ (g \circ f)}$$

Bemerkung 1.3.6 (Urbild und Bild)

Nun wenden wir uns wieder einer einzelnen Abbildung $f : M \longrightarrow N$ zu. Zu dieser Abbildung gibt es eine Abbildung zwischen den Potenzmengen

$$f^{-1} : \mathcal{P}(N) \longrightarrow \mathcal{P}(M), \text{ wobei } f^{-1}(B) := \{m \in M | f(m) \in B\}.$$

Man nennt $f^{-1}(B)$ das **Urbild der Teilmenge** $B \subseteq N$ **unter** f.

Wenn zum Beispiel f die Abbildung ist, die jeder Studentin ihren Geburtstag zuordnet (eine kalenderwertige Abbildung auf der Menge der Studentinnen sozusagen), dann ist $f^{-1}(\{29.\text{ Februar } 1996\})$ eben die Menge aller Studentinnen, die an diesem Tag Geburtstag hatten. Und $f^{-1}(\{29.\text{ Februar } 1997\})$ ist die leere Menge.

Oft wird man statt $f^{-1}(\{a\})$ die kürzere Notation $f^{-1}(a)$ benutzen, auch wenn dies formal nicht ganz korrekt ist. Wir versuchen, dies hier zu vermeiden.

Für eine Abbildung $f : M \longrightarrow N$ und eine Teilmenge $A \subseteq M$ bezeichnen wir mit $f(A) := \{f(a) \mid a \in A\} \subseteq N$ die Menge aller Funktionswerte von f auf der Menge A. Diese Menge heißt auch das **Bild von A unter f.**

Bemerkung 1.3.7 (Die Umkehrabbildung)

Oft ist eine Abbildung $f : M \to N$ gegeben, und es gilt zu entscheiden, ob es für ein gegebenes $n \in N$ ein $m \in M$ gibt, sodass $f(m) = n$.

Das heißt: Wir wollen eine Gleichung lösen!

Es gibt eine Situation, in der dies besonders einfach ist, nämlich dann, wenn die Menge $\tilde{f} = \{(f(m), m) \mid m \in M\}$ eine Abbildung von N nach M ist. In diesem Fall ist für gegebenes $n \in N$ das Element $m \in M$ genau dann eine Lösung von $f(m) = n$, wenn $m = \tilde{f}(n)$.

Sie kennen dieses Phänomen von der e-Funktion

$$f : \mathbb{R} \to \mathbb{R}_{>0},\ f(x) = e^x.$$

Dann ist $\{(f(x), x) \mid x \in \mathbb{R}\} \subset \mathbb{R}_{>0} \times \mathbb{R}$ der Graph der natürlichen Logarithmusfunktion.

Die Bedingung, dass $\tilde{f}$ Graph einer Abbildung ist, ist einfach, dass jedes $n \in N$ sich auf genau eine Art als $f(m),\ m \in M$, schreiben lässt.

Wenn dies erfüllt ist, dann gilt

$$f \circ \tilde{f} = \mathrm{Id}_N \ \text{ und } \ \tilde{f} \circ f = \mathrm{Id}_M.$$

Wir nennen $\tilde{f}$ dann die **Umkehrabbildung** von f und schreiben in Zukunft dafür $f^{-1} : N \to M$.

Vorsicht: Ob mit dem Symbol f^{-1} die Umkehrfunktion oder die Urbildabbildung aus 1.3.6 gemient ist, entscheidet sich daran, ob Elemente von N oder Teilmengen von N als Argumente eingesetzt werden.

Die vorhin genannte Bedingung an f wird gerne in zwei unabhängige Bedingungen zerlegt, die in der folgenden Definition einen Namen erhalten.

Definition 1.3.8 (injektiv, surjektiv, bijektiv)

a) Eine Abbildung $f : M \longrightarrow N$ heißt **injektiv**, wenn für alle $m_1, m_2 \in M$ gilt :

$$[f(m_1) = f(m_2)] \Rightarrow [m_1 = m_2].$$

Das bedeutet, dass man m eindeutig daran erkennen kann, was $f(m)$ ist. Noch anders gesagt ist f injektiv, wenn für alle $n \in N$ gilt:

$$|f^{-1}(\{n\})| \leq 1.$$

Also: es gibt höchstens ein m mit $f(m) = n$.

b) f heißt **surjektiv**, wenn $f(M) = N$ gilt, also wenn für jedes $n \in N$ gilt:

$$|f^{-1}(\{n\})| \geq 1.$$

Also: es gibt mindestens ein m mit $f(m) = n$. Oder auch:

$$\forall n \in N : f^{-1}(\{n\}) \neq \emptyset.$$

c) Die Abbildung f heißt **bijektiv**, wenn sie sowohl injektiv als auch surjektiv ist. Also: für jedes $n \in N$ gibt es genau ein $m \in M$ mit $f(m) = n$. Bijektive Abbildungen werden auch **Bijektionen** genannt.

Beispiel 1.3.9 Die Menge M hat dann und nur dann genau n Elemente, wenn es eine Bijektion zwischen M und der Menge $\{1, 2, 3, \ldots, n\} \subseteq \mathbb{N}$ gibt. Solch eine Bijektion tut ja nichts anderes, als die Elemente aus M durchzunummerieren.

Beispiel 1.3.10 (Isomorphismen von Graphen)

Es ist sinnvoll, zwei Graphen $\Gamma = (E, K), \tilde{\Gamma} = (\tilde{E}, \tilde{K})$ als im wesentlichen gleich zu betrachten, wenn sie durch Umbenennung der Ecken auseinander hervorgehen.

Präziser meint man damit, dass es eine bijektive Abbildung

$$f : E \to \tilde{E}$$

gibt, sodass für alle $x, y \in E$ gilt:

$$\{x, y\} \in K \Leftrightarrow (f(x), f(y)) \in \tilde{K}.$$

Solch eine Abbildung nennen wir einen **Isomorphismus** zwischen Γ und $\tilde{\Gamma}$. Wenn es solch einen Isomorphismus gibt, heißen die Graphen **isomorph**.

Da f bijektiv ist, haben zwei isomorphe Graphen immer gleich viele Ecken. Auch die Anzahl der Kanten ist gleich. Dass die nicht genügt, sieht man am folgenden Beispel:

$$\Gamma = (\{1, 2, 3, 4\},\ \{\{1, 2\}, \{3, 4\}\}) \text{ und } \tilde{\Gamma} = ((\{1, 2, 3, 4\},\ \{\{1, 2\}, \{2, 3\}\}).$$

Im ersten Graphen gibt es nämlich keine Ecke, die zu zwei verschiedenen Kanten gehört, und das muss ein Isomorphismus respektieren.

Veranschaulichen Sie sich die beiden Graphen durch ein Bild!

Aufgabe 1.3.11 (Graphen mit drei Ecken)

Geben Sie eine Menge von Graphen mit drei Ecke an, die paarweise nicht isomorph sind und folgende Eigenschaft haben: Jeder Graph mit drei Ecken ist zu einem Ihrer Graphen isomorph.

Sie können dazu zum Beispiel alle Graphen mit Eckenmenge $\{1, 2, 3\}$ auflisten und daraus eine geeignete Teilmenge auswählen.

Aufgabe 1.3.12 (Automorphismen von Graphen)

Es sei $\Gamma = (E, K)$ en Graph. Ein Isomorphismus $f : \Gamma \to \Gamma$ heißt auch ein **Automorphismus** von Γ.

Zeigen Sie, dass für zwei Automorphismen f, g von Γ auch die Komposition $g \circ f$ ein Automorphismus von Γ ist.

Um noch eine etwas andere Sichtweise auf die obigen Eigenschaften von Abbildungen zu bekommen, geben wir ein anderes Kriterium für Injektivität und Surjektivität an.

Satz 1.3.13 (Injektivität und Surjektivität)

Es sei $f : M \longrightarrow N$ eine Abbildung zwischen den Mengen M und N, und M sei nicht leer. Dann gelten die folgenden Aussagen:

a) (f *ist injektiv*) $\iff$ ($\exists g : N \longrightarrow M$ *mit* $g \circ f = \mathrm{Id}_M$).
b) (f *ist surjektiv*) $\iff$ ($\exists h : N \longrightarrow M$ *mit* $f \circ h = \mathrm{Id}_N$).
c) (f *ist bijektiv*) $\iff$ (*es gibt* g *und* h *wie in* a) *und* b)).

In diesem letzten Fall gilt außerdem $g = h = f^{-1}$.

Beweis. [3]

a) „$\Longrightarrow$“ Wir nehmen zunächst an, f sei injektiv. Zu zeigen ist die Existenz einer Abbildung g mit den gewünschten Eigenschaften. Um diese zu konstruieren wählen wir zunächst ein $m_0 \in M$, was geht, da M nicht leer ist. Dann setzen wir für $n \in N$

$$g(n) := \begin{cases} m & \text{falls } n = f(m) \in f(M), \\ m_0 & \text{falls } n \notin f(M). \end{cases}$$

Diese Abbildung ist sinnvoll definiert: für alle $n \in f(M)$ gibt es genau ein $m \in M$ mit $f(m) = n$, denn f ist injektiv. Nun rechnet man nach

$$\forall m \in M : (g \circ f)(m) = g(f(m)) = m,$$

[3] Um eine Äquivalenz zweier Aussagen zu zeigen, zeigt man oft, dass die eine die andere impliziert und umgekehrt. Dies wird – wie hier im Beweis – oft dadurch kenntlich gemacht, dass man den Äquivalenzpfeil in zwei Implikationspfeile zerlegt und eben einmal „$\Longrightarrow$“ zeigt, das ist die Implikation von links nach rechts, und dann auch noch „$\Longleftarrow$“, die andere Implikation.

also $g \circ f = \mathrm{Id}_M$ nach Definition der identischen Abbildung.

„$\Longleftarrow$“ Nun gibt es nach Voraussetzung ein g wie im Satz, und wir müssen daraus folgern, dass f injektiv ist. Wenn aber $m_1, m_2 \in M$ Elemente mit $f(m_1) = f(m_2)$ sind, dann folgt

$$\begin{aligned} m_1 &= \mathrm{Id}_M(m_1) = (g \circ f)(m_1) = g(f(m_1)) \\ &\overset{f(m_1)=f(m_2)}{=} g(f(m_2)) = (g \circ f)(m_2) = \mathrm{Id}_M(m_2) = m_2. \end{aligned}$$

Also ist f injektiv.

b) „$\Longrightarrow$“ Hier ist f zunächst als surjektiv vorausgesetzt. Wir wählen für jedes $n \in N$ ein $m \in M$ mit $f(m) = n$ und nennen dieses gewählte m geschickter Weise $h(n)$.[4] Damit ist eine Abbildung $h : N \longrightarrow M$ ausgewählt, und es gilt für alle $n \in N$:

$$(f \circ h)(n) = f(h(n)) = n$$

nach Wahl von $h(n)$: $f \circ h = \mathrm{Id}_N$.

„$\Longleftarrow$“ Nun nehmen wir an, wir hätten eine Abbildung h von N nach M mit $f \circ h = \mathrm{Id}_N$. Dann gilt wieder für jedes $n \in N$:

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

also $h(n) \in f^{-1}(\{n\})$ und damit ist f surjektiv, da n beliebig war.

c) Nach Definition ist f genau dann bijektiv, wenn es sowohl in- als auch surjektiv ist. Nach den Teilen a) und b) (die ja schon bewiesen sind!) ist das äquivalent zur Existenz von g und h. Nur $g = h$ ist noch zu zeigen. Wir benutzen nun das Assoziativitätsgesetz (Fazit 1.3.5) für die Hintereinanderausführung von Abbildungen und sehen, dass gilt:

$$g = g \circ \mathrm{Id}_N = g \circ (f \circ h) = (g \circ f) \circ h = \mathrm{Id}_M \circ h = h.$$

$\bigcirc$

Beispiel 1.3.14 (Ein bekanntes Beispiel dazu)

Wir betrachten die Funktion

$$f : \mathbb{R} \to \mathbb{R}_{\geq 0}, \quad f(x) = x^2.$$

Aus der Schule ist bekannt, dass jede nicht negative reelle Zahl y sich als $y = x^2$ schreiben lässt, mit anderen Worten: f ist surjektiv.

[4] Hier benutzen wir das so genannte Auswahlaxiom. Das soll hier nicht problematisiert werden.

Nun wollen wir eine Abbildung $h : \mathbb{R}_{\geq 0} \to \mathbb{R}$ wie in 1.3.13(b) angeben. Dazu müssen wir für jede nicht negative Zahl y ein $x \in \mathbb{R}$ wählen, sodass $x^2 = y$ gilt. Bekanntlich ist x hier nur bis aufs Vorzeichen festgelegt, und wir könnten und müssen formal für jedes y frei entscheiden, welchen der beiden möglichen x-Werte wir auswählen. In der gegenwärtigen Situation ist das dadurch konsistent möglich, dass wir uns für die nicht negative Lösung entscheiden. Dann haben wir die Funktion $y \mapsto \sqrt{y}$ erfunden, die für jedes $y \geq 0$ tatsächlich $y = f(h(y)) = (\sqrt{y})^2$ liefert.

In anderen Situationen wird es nicht so leicht konsistent möglich sein, ein Element aus dem Urbild auszuwählen. Durch die Konstruktion in 1.4.11 wird das ein Stück weit behoben.

Bemerkung 1.3.15 (noch einmal Tupel)

Die Menge aller k-Tupel in der Menge M aus 1.2.2 kann man sich auch denken als die Menge aller Abbildungen von $\{1, \ldots, k\}$ nach M. Die Vorschrift, die dem Tupel $(m_1, \ldots, m_k)$ die Abbildung $[i \mapsto m_i] \in \mathrm{Abb}(\{1, \ldots, k\}, M)$ zuordnet, ist eine Bijektion zwischen M^k und $\mathrm{Abb}(\{1, \ldots, k\}, M)$. Dies gibt uns die Möglichkeit, auch M^0 noch einen Sinn zu geben:

$$M^0 = \mathrm{Abb}(\emptyset, M).$$

Diese Menge hat genau ein Element, denn es gibt genau eine Teilmenge von $\emptyset \times M = \emptyset$, und diese Teilmenge erfüllt die Bedingung aus der Definition von Abbildungen (1.3.1).

Definition 1.3.16 (Einschränkung einer Abbildung)

Es seien $f : M \longrightarrow N$ eine Abbildung und T eine Teilmenge von M. Dann heißt die Abbildung

$$f|_T : T \longrightarrow N, \quad t \mapsto f(t)$$

die **Einschränkung** oder auch **Restriktion von f nach T.** Man merkt sich hier im Symbol der Abbildung den künstlich verkleinerten Definitionsbereich.

Wenn wir zum Beispiel in 1.3.14 den Definitionsbereich von f auf die nicht negativen reellen Zahlen einschränken, dann ist die neue Abbildung

$$f|_{\mathbb{R}_{\geq 0}} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}, \quad x \mapsto x^2,$$

bijektiv und besitzt damit eine Umkehrfunktion, womit wir wieder bei der üblichen Quadratwurzel landen.

Wenn $M = N$ gilt und $f(T) \subseteq T$, dann bezeichnet man mit $f|_T$ oft auch die Abbildung von T nach T, die durch f gegeben ist; siehe z.B. Definition 8.2.1.

1.4 Relationen

Definition 1.4.1 (Relationen)

Es sei M eine beliebige Menge. Eine (zweistellige) **Relation** auf M ist eine Teilmenge $R \subseteq M \times M$.

Statt $(x, y) \in R$ schreibt man zumeist kürzer xRy.

Beispiel 1.4.2 (Zwei bekannte Relationen)

a) Für jede Menge M ist die Relation $R := \{(m, m) | m \in M\}$ die Gleichheitsrelation auf M:
$$\forall x, y \in M : \; xRy \iff x = y.$$

b) Für das Intervall $M = [0, 1] \subseteq \mathbb{R}$ sei $S := \{(x, y) | x \leq y\}$. Das ist die Kleinergleich–Relation.

Nun interessiert man sich in aller Regel nicht für alle Relationen, sondern nur für solche, die günstige Eigenschaften haben. Für uns von besonderem Interesse sind die folgenden Eigenschaften.

Definition 1.4.3 (Eigenschaften von Relationen)

Es sei $R \subseteq M \times M$ eine Relation. Dann heißt R

- **reflexiv**, wenn für alle $x \in M$ gilt: $(x, x) \in R$.
- **symmetrisch**, wenn für alle $x, y \in M$ gilt:
$$xRy \iff yRx.$$
- **antisymmetrisch**, wenn für alle $x, y \in M$ gilt:
$$[xRy \text{ und } yRx] \Longrightarrow x = y.$$
- **transitiv**, wenn für alle $x, y, z \in M$ gilt:
$$[xRy \text{ und } yRz] \Longrightarrow xRz.$$

Es ist sicher instruktiv, für jeden dieser Begriffe Relationen zu haben, die ihn erfüllen und auch solche, die dies nicht tun.

Die Gleichheitsrelation (Bsp. 1.4.2) ist reflexiv, symmetrisch und transitiv. Die Relation $\leq$ in Beispiel 1.4.2 ist antisymmetrisch. Sie ist auch nicht symmetrisch, aber reflexiv. (Die $<$-Relation wäre nicht reflexiv.)

Beide aber sind transitiv. Eine Relation, die nicht transitiv ist, ist zum Beispiel die folgende Relation U („Ungleichheit“) auf der Menge $\{0,1\}$:

$$U := \{(0,1),(1,0)\}.$$

Es gilt ja $0U1$ und $1U0$, und Transitivität würde verlangen, dass aus diesen beiden auch $0U0$ folgt, was aber nicht stimmt.

Definition 1.4.4 (Äquivalenzrelation)

Eine Relation R auf der Menge M heißt eine **Äquivalenzrelation**, wenn sie reflexiv, symmetrisch und transitiv ist.

Beispiel 1.4.5 (Kongruenz)

a) Die Gleichheit ist eine Äquivalenzrelation auf jeder Menge M.

b) Die $\leq$-Relation auf $[0,1]$ aus obigem Beispiel ist keine Äquivalenzrelation, denn sie ist nicht symmetrisch (z.B. gilt $0 \leq 1$ aber nicht $1 \leq 0$).

c) Nun sei $M = \mathbb{Z}$ und n eine natürliche Zahl.

Dann definieren wir die Relation $\equiv_n$ durch

$$x \equiv_n y \iff (x-y) \text{ ist teilbar durch } n.$$

Dabei heißt „teilbar durch n“, dass für eine geeignete ganze Zahl k die Gleichheit $x - y = k \cdot n$ gilt.

Jetzt weisen wir nach, dass dies eine Äquivalenzrelation ist. Der Nachweis besteht aus drei Schritten.

Reflexivität: Für alle $x \in \mathbb{Z}$ gilt

$$x - x = 0 = 0 \cdot n.$$

Also ist $x \equiv_n x$.

Symmetrie: Für alle $x, y \in \mathbb{Z}$ gilt:

$$\begin{aligned} x \equiv_n y &\implies \exists k \in \mathbb{Z} : x - y = k \cdot n \\ &\implies \exists k \in \mathbb{Z} : y - x = (-k) \cdot n \\ &\implies y \equiv_n x. \end{aligned}$$

Transitivität: Für alle $x, y, z \in \mathbb{Z}$ gilt:

$$\begin{aligned} (x \equiv_n y \wedge y \equiv_n z) &\implies \exists k, l \in \mathbb{Z} : x - y = k \cdot n \wedge y - z = l \cdot n \\ &\implies x - z = (x-y) + (y-z) = (k+l) \cdot n \\ &\implies x \equiv_n z. \end{aligned}$$

Oft schreibt man übrigens statt $x \equiv_n y$ lieber $x \equiv y \pmod{n}$ und sagt dafür: x und y sind **kongruent modulo** n**.** Diese Relation und ihre weitläufige Verwandtschaft ist in vielen Teilen der Mathematik, speziell auch in der (linearen) Algebra, von großer Bedeutung. Außerdem braucht man sie in vielen Anwendungen, zum Beispiel in der Kryptographie.

Äquivalenzrelationen gehen Hand in Hand mit einer Zerlegung der Menge M in disjunkte Teilmengen. Das heißt, dass man M als Vereinigung von Teilmengen schreibt, von denen jeweils zwei verschiedene die leere Menge als Schnitt haben. Um diesen Zusammenhang zu präzisieren machen wir die folgende Definition. Dabei benutzen wir die übliche Notation $\sim$ für eine beliebige Äquivalenzrelation (anstelle des Buchstaben R aus Definition 1.4.1).

Aufgabe 1.4.6 (Äquivalenz?)

a) Es sei $\mathcal{D}$ die Menge aller Dreiecke in der Tafelebene. Zwei Dreiecke heißen **kongruent**, wenn bei geeigneter Nummerierung ihre Seitenlängen übereinstimmen.

Zeigen Sie, dass das eine Äquivalenzrelation auf $\mathcal{D}$ liefert.

b) Wir sagen, dass zwei Punkte P, Q der Tafelebene benachbart sind, wenn ihr Abstand nicht größer als 1 ist.

Zeigen Sie, dass diese Relation symmetrisch und reflexiv ist, aber nicht transitiv.

Definition 1.4.7 (Äquivalenzklassen)

Es sei $\sim$ eine Äquivalenzrelation auf der Menge M. Dann heißt für $x \in M$ die Teilmenge

$$[x]_\sim := \{y \in M \mid x \sim y\} \subseteq M$$

die **Äquivalenzklasse von** x (bezüglich $\sim$).

Wir erhalten den folgenden Satz.

Satz 1.4.8 (Zerlegung in Äquivalenzklassen)

Es sei M eine Menge.

a) Für jede Äquivalenzrelation $\sim$ auf M sind die Äquivalenzklassen bezüglich $\sim$ nicht leer und es gilt

$$M = \bigcup_{x \in M} [x]_\sim.$$

Außerdem gilt für alle $x, y \in M$

$$[x]_\sim \cap [y]_\sim = \emptyset \text{ oder } [x]_\sim = [y]_\sim.$$

b) Ist umgekehrt $\mathcal{S} \subseteq \mathcal{P}(M)$ ein System von Teilmengen von M, sodass $\emptyset \notin \mathcal{S}$ gilt sowie

$$M = \bigcup_{A \in \mathcal{S}} A \quad \text{und} \quad \forall A, B \in \mathcal{S} : [A \cap B = \emptyset \text{ oder } A = B],$$

dann gibt es eine Äquivalenzrelation $\sim$ auf M, für die $\mathcal{S}$ die Menge aller Äquivalenzklassen ist, das heißt:

$$\mathcal{S} = \{[x]_\sim \mid x \in M\}.$$

Beweis.

a) Da für jedes $x \in M$ wegen der Reflexivität von $\sim$ insbesondere auch $x \in [x]_\sim$ gilt, ist $[x]_\sim \neq \emptyset$, und es folgt außerdem $M = \bigcup_{x \in M} [x]_\sim$. Wenn $[x]_\sim \cap [y]_\sim \neq \emptyset$ gilt, dann gibt es ein Element z in $[x]_\sim \cap [y]_\sim$. Für dieses gilt $x \sim z$ und $y \sim z$, und wegen Symmetrie und Transitivität folgt $x \sim y$. Wiederum Symmetrie und Transitivität implizieren dann, dass für alle $m \in M$ gilt: $x \sim m \iff y \sim m$. Also sind die Äquivalenzklassen $[x]_\sim$ und $[y]_\sim$ gleich.

b) Wir definieren die Relation $\sim$ durch

$$x \sim y :\iff \exists A \in \mathcal{S} : x \in A \text{ und } y \in A.$$

Dies ist eine Äquivalenzrelation, denn sie ist

- reflexiv, weil M die Vereinigung aller $A \in \mathcal{S}$ ist, also für jedes $x \in M$ ein $A \in \mathcal{S}$ existiert mit $x \in A$.
- symmetrisch: klar.
- transitiv, weil aus $x, y \in A$ und $y, z \in B$ für $A, B \in \mathcal{S}$ folgt, dass $A \cap B$ nicht leer ist (denn y liegt im Schnitt), also $A = B$ und damit auch $x, z \in A$.

Es ist klar, dass für $x \in M$ die Äquivalenzklasse von x bezüglich $\sim$ gerade diejenige Menge $A \in \mathcal{S}$ ist, für die $x \in A$. Also gilt

$$\mathcal{S} \supseteq \{[x]_\sim \mid x \in M\}.$$

Umgekehrt ist jedes $A \in \mathcal{S}$ nicht leer, enthält also ein $x \in M$, und damit ist $A = [x]_\sim$. Das zeigt

$$\mathcal{S} \subseteq \{[x]_\sim \mid x \in M\}.$$

Also sind diese zwei Mengensysteme gleich. ◯

Beispiel 1.4.9 (noch einmal die Kongruenz)

Zur Illustration betrachten wir den Fall der Äquivalenzrelation $\equiv_n$ (Kongruenz modulo n) aus dem letzten Beispiel 1.4.5. Die Äquivalenzklasse von $x \in \mathbb{Z}$ ist die Menge aller $y \in \mathbb{Z}$, für die $x - y$ durch n teilbar ist, also die Menge aller y, die bei Division durch n denselben Rest lassen wie x. Offensichtlich ist das

$$\{x + kn | k \in \mathbb{Z}\} = [x]_{\equiv_n}.$$

Wenn n nicht gerade 0 ist, so gibt es genau n Äquivalenzklassen, nämlich

$$[0]_{\equiv_n}, [1]_{\equiv_n}, \ldots, [n-1]_{\equiv_n}.$$

Der Rest von x bei Division durch n ist ja eine nicht negative Zahl kleiner als n, und zwei verschiedene solcher Zahlen sind nicht kongruent modulo n.

Betrachten wir noch den **Spezialfall** $n = 2$, so bekommen wir zwei Äquivalenzklassen: die Menge der geraden Zahlen und die der ungeraden Zahlen.

Die Klassenbildung aus dem letzten Satz verallgemeinert also einen Typus der Unterscheidung, der uns allen vertraut ist.

Aufgabe 1.4.10 (Zusammenhangskomponenten in Graphen)

Es sei $\Gamma = (E, K)$ ein Graph. Zwei Ecken $x, y \in E$ heißen **benachbart**, wenn $\{x, y\}$ ein Kante ist. Benachbart zu sein ist eine Relation auf E.

Ein **Kantenzug** der Länge $l \in \mathbb{N}_0$ von $x \in E$ nach $y \in E$ ist eine Sequenz von Ecken $x_0 = x, x_1, \ldots, x_{l-1}, x_l = y$, sodass für jedes $i \in \{0, \ldots, l-1\}$ die Ecken x_i und x_{i+1} benachbart sind.

Wir sagen, x sei mit y **verbindbar**, wenn es einen Kantenzug (irgendeiner Länge) von x nach y gibt. Auch verbindbar zu sein ist eine Relation auf E.

a) Zeigen Sie, dass benachbart zu sein zwar symmetrisch ist, aber niemals reflexiv, wenn $E \neq \emptyset$ und niemals transitiv wenn $K \neq \emptyset$.

b) Zeigen Sie, dass verbindbar zu sein eine Äquivalenzrelation auf E ist.

c) Finden Sie für die beiden Graphen Γ und $\tilde{\Gamma}$ aus 1.3.10 die Äquivalenzklassen bezüglich Verbindbarkeit und sehen Sie ein, dass es sinnvoll ist, sie die **Zusammenhangskomponenten** von Γ zu nennen.

Wir wollen noch eine wichtige Konstruktion für Äquivalenzrelationen angeben:

Hilfssatz 1.4.11 (Abbildungen und Äquivalenzrelationen)

Es sei $f : M \longrightarrow N$ *eine Abbildung. Dann wird durch*

$$x \sim y : \Longleftrightarrow \ f(x) = f(y)$$

eine Äquivalenzrelation auf M *definiert, und die Äquivalenzklasse von* x *ist* $f^{-1}(f(x))$.

Den einfachen *Beweis* können Sie selbst als Übungsaufgabe durchführen. Die Eigenschaften einer Äquivalenzrelation werden sehr leicht auf die entsprechenden Eigenschaften der Gleichheitsrelation zurückgeführt.

Bemerkung 1.4.12 (Quotientenbildung)

Jede Äquivalenzrelation auf M lässt sich aus der Konstruktion des letzten Hilfssatzes gewinnen, wenn man nur N und f richtig wählt. Konkreter sei $\sim$ irgendeine Äquivalenzrelation auf M. Wähle $N := \mathcal{P}(M)$ die Potenzmenge von M und setze

$$f(x) := [x]_\sim.$$

Dann rechnet man leicht nach, dass man aus f durch die Konstruktion des letzten Hilfssatzes die alte Relation $\sim$ zurückgewinnt.

Dies ist das Prinzip, das der so genannten Quotientenbildung zu Grunde liegt, auf das wir noch des öfteren zu sprechen kommen werden. Sie werden hoffentlich im Laufe der Zeit feststellen, dass dieses Prinzip eines der fundamentalsten Prinzipien der Mathematik ist. Genauer nennt man das Bild von f

$$f(M) = \{[x]_\sim | x \in M\} =: M/_\sim$$

die **Quotientenmenge von** M **nach der Äquivalenzrelation** $\sim$. Diese Quotientenmenge wird oft benutzt, um Abbildungen $g : M \longrightarrow P$, die die Bedingung

$$\forall x, y \in M : x \sim y \Longrightarrow g(x) = g(y)$$

erfüllen, zu schreiben als

$$g = \widetilde{g} \circ f,$$

wobei die Abbildung $\widetilde{g} : M/_\sim \longrightarrow P$ definiert ist durch

$$\widetilde{g}([x]_\sim) := g(x).$$

Das ist sinnvoll definiert, da g ja auf jeder Äquivalenzklasse $[x]_\sim$ konstant ist.

Oft ist es sogar so, dass man nicht an M interessiert ist, sondern M nur braucht um die eigentlich viel interessantere Menge $M/_\sim$ hinzuschreiben, die man anders nicht in den Griff bekommt. Dieses Credo wird auch noch verschiedentlich eine Rolle spielen.

Kapitel 2

Gruppen

In diesem Kapitel werden algebraische Objekte eingeführt, die uns durch den Rest der Vorlesung begleiten werden. Eine der Grundfragen der Linearen Algebra ist die Frage nach der Lösbarkeit linearer Gleichungssysteme. Es hat sich dabei herausgestellt, dass man sich nicht von vorneherein auf reelle Zahlen als Koeffizienten beschränken sollte, sondern dass man sinnvoller Weise die Theorie in größerer Allgemeinheit entwickelt, sodass sie besser auf andere Situationen übertragbar ist. Man abstrahiert also, und benutzt nur solche Eigenschaften der reellen Zahlen, die man zum Beispiel für die Durchführung des Gauß-Algorithmus braucht. Das wird später bei der Einführung des Körperbegriffs in Kapitel 3 geschehen. Vorbereitend aber kommen wir zu einer der zentralsten Begriffsbildungen der Mathematik, zum Gruppenbegriff, der noch stärker abstrahiert und insbesondere eine Analogie zwischen (Mengen von) Abbildungen und (Mengen von) Zahlen zum Ausdruck bringt.

2.1 Gruppen – Definition und Beispiele

Definition 2.1.1 (Verknüpfung, Assoziativität, Kommutativität)

Es sei M eine Menge.

a) Eine **Verknüpfung auf** M ist eine Abbildung $* : M \times M \longrightarrow M$, also eine Vorschrift, die je zwei Elementen aus M ein Element aus M zuordnet. Dabei kommt es auf die Reihenfolge der Argumente an. Statt (formal korrekt) $*(m_1, m_2)$ schreibt man kürzer $m_1 * m_2$.

b) Eine Verknüpfung $*$ auf M heißt **assoziativ**, wenn gilt:

$$\forall m_1, m_2, m_3 \in M : (m_1 * m_2) * m_3 = m_1 * (m_2 * m_3).$$

In diesem Fall ist die kürzere Schreibweise $m_1 * m_2 * m_3$ legitim und unmissverständlich.

c) Eine Verknüpfung $*$ auf M heißt **kommutativ**, wenn gilt

$$\forall m_1, m_2 \in M : m_1 * m_2 = m_2 * m_1.$$

Beispiel 2.1.2 (Ein paar Verknüpfungen)

a) Auf den Mengen $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ sind Addition und Multiplikation jeweils assoziative und kommutative Verknüpfungen. Die Differenz ist auf $\mathbb{N}$ gar nicht als Verknüpfung definiert (da etwa $1 - 2 \notin \mathbb{N}$). Auf den anderen drei Mengen ist sie eine Verknüpfung, die weder assoziativ noch kommutativ ist.

b) Für eine beliebige Menge D ist auf $M := \mathrm{Abb}(D, D)$ die Hintereinanderausführung $\circ$ eine assoziative Verknüpfung (siehe Fazit 1.3.5). Sie ist aber nicht kommutativ, wenn D mindestens 2 Elemente hat. Sind nämlich $d, e \in D$ zwei verschiedene Elemente, so gibt es die folgenden Abbildungen $f, g \in M$:

$$\forall x \in D : f(x) := d,\ g(x) := e.$$

Dann sieht man sofort

$$\begin{aligned} \forall x \in D : \quad & (f \circ g)(x) = f(g(x)) = f(e) = d \\ & (g \circ f)(x) = g(f(x)) = g(d) = e. \end{aligned}$$

Da D insbesondere nicht leer ist, gibt es also wirklich mindestens ein x mit $(f \circ g)(x) \neq (g \circ f)(x)$, und damit gilt $f \circ g \neq g \circ f$.

c) Auf jeder Menge M ist die Vorschrift

$$\forall x, y \in M : \ x * y := x$$

eine assoziative Verknüpfung.

d) Für eine beliebige Menge D sind auf der Potenzmenge $M := \mathcal{P}(D)$ die Verknüpfungen $\cap$ und $\cup$ beide assoziativ und kommutativ. Die Mengendifferenz $\setminus$ ist im Allgemeinen weder assoziativ noch kommutativ. Beispiele:

$$\begin{aligned} \text{keine Assoziativität} : \quad & (\{1,2,3\} \setminus \{1,2\}) \setminus \{1\} = \{3\}, \text{ aber} \\ & \{1,2,3\} \setminus (\{1,2\} \setminus \{1\}) = \{1,3\}; \\ \text{keine Kommutativität} : \quad & \forall A \neq \emptyset : A \setminus \emptyset = A,\ \emptyset \setminus A = \emptyset \neq A. \end{aligned}$$

e) Auf der Menge $M := \{a, b\}$ mit den zwei (verschiedenen) Elementen a und b wird durch

$$a * a = b, \quad b * b = a * b = b * a = a$$

eine kommutative Verknüpfung festgelegt, die nicht assoziativ ist:

$$b * (a * a) = b * b = a, \quad \text{aber} \quad (b * a) * a = a * a = b.$$

Bemerkung 2.1.3 Analogie zu Graphen

Eine Verknüpfung auf M ist – als Abbildung von $M \times M$ nach M – vorstellbar als Menge M zusammen mit einer Teilmenge $*$ von $M^3 = M \times M \times M$, wobei jetzt die entscheidende Bedingung ist, dass für jedes Paar $(m_1, m_2) \in M \times M$ genau ein $m_3 \in M$ existiert mit $(m_1, m_2, m_3) \in *$.

Eine gewisse Analogie zur Welt der Graphen ist nicht zu verkennen, wobei hier natürlich die Reihenfolge in (m_1, m_2, m_3) eine wichtige Rolle spielt, während bei Graphen die Ecken nicht gerichtet waren.

Nun kommt der eingangs versprochene Gruppenbegriff, der allen sehr ans Herz gelegt sei. Man abstrahiert hierbei von den Objekten (wie z.B. Zahlen) und definiert eine Klasse von Verknüpfungen, die formal sehr nahe an dem sind, was man von Zahlen her kennt.

Definition 2.1.4 (Gruppe)

Es sei M eine Menge und $*$ eine Verknüpfung auf M. Dann heißt das Paar $(M, *)$ eine **Gruppe**, wenn die folgenden Bedingungen a) – c) erfüllt sind:

a) Die Verknüpfung $*$ ist assoziativ.

b) Es gibt (mindestens) ein Element $e \in M$, sodass für alle $x \in M$ gilt:

$$x * e = e * x = x.$$

Bemerkung: Wenn $\tilde{e}$ ein weiteres Element mit dieser Eigenschaft ist, so folgt $e = e * \tilde{e} = \tilde{e}$, indem man erst $x = \tilde{e}$ und dann $x = e$ in der obigen Gleichung setzt. Also ist e eindeutig charakterisiert. Man nennt es das **neutrale Element** von $(M, *)$ und schreibt dafür oft e_M (statt des präziseren $e_{(M,*)}$).

c) Für jedes $x \in M$ gibt es (mindestens) ein $y \in M$, sodass

$$x * y = y * x = e$$

gilt. Dabei ist e das neutrale Element von $(M, *)$.

Bemerkung: Wenn $\tilde{y}$ ein weiteres Element in M mit der Eigenschaft

$$x * \tilde{y} = \tilde{y} * x = e$$

ist, dann folgt unter Ausnutzung der Assoziativität:

$$y = y * e = y * (x * \tilde{y}) = (y * x) * \tilde{y} = e * \tilde{y} = \tilde{y}.$$

Also ist y eindeutig durch die charakterisierende Gleichung festgelegt. Man nennt es das zu x **inverse Element** in $(M, *)$.

Speziell ist zum Beispiel e zu sich selbst invers.

Beispiel 2.1.5 (Zahlen, symmetrische Gruppe)

a) Die ganzen Zahlen $\mathbb{Z}$ mit der Addition bilden eine Gruppe. Die Assoziativität ist aus der Schule bekannt, das neutrale Element ist 0, und das inverse Element zur ganzen Zahl x ist $-x$.

Das zeigt auch schon, dass die $(\mathbb{N}, +)$ mangels neutralem Element keine Gruppe ist, und dass $(\mathbb{N}_0, +)$ keine Gruppe ist, da zum Beispiel das inverse Element zur 1 nicht darin liegt.

Wie $\mathbb{Z}$ so bilden auch die rationalen Zahlen $\mathbb{Q}$ und die reellen Zahlen $\mathbb{R}$ mit der Addition als Verknüpfung eine Gruppe.

Bezüglich der (wie üblich definierten) Multiplikation muss man mehr aufpassen. Wir finden aber zum Beispiel die Gruppen

$$(\{\pm 1\}, \cdot), \quad (\mathbb{Q} \setminus \{0\}, \cdot), \quad (\mathbb{R} \setminus \{0\}, \cdot).$$

b) Für eine beliebige Menge D ist $\mathrm{Abb}(D, D) =: M$ mit der Verknüpfung $\circ$ zwar assoziativ, aber keine Gruppe, sobald D mindestens zwei Elemente hat. Das hat nicht direkt mit der dann fehlenden Kommutativität zu tun, auch wenn die Begründung wieder die Funktion f von oben benutzt. Zunächst einmal halten wir fest, dass die identische Abbildung Id_D das neutrale Element von $(\mathrm{Abb}(D, D), \circ)$ ist. Für ein beliebiges Element $d \in D$ sei die Abbildung f definiert durch $f(x) := d$. Wenn D mindestens zwei Elemente enthält, ist dieses f dann weder injektiv noch surjektiv, also gibt es nicht einmal eine der einseitigen Umkehrabbildungen (siehe 1.3.13).

c) Für eine beliebige Menge D sei $M := \{f \in \mathrm{Abb}(D, D) \mid f \text{ bijektiv}\}$ die Menge aller Bijektionen von D nach D. Die Verknüpfung $\circ$ ist darauf assoziativ (da man eine Teilmenge von $\mathrm{Abb}(D, D)$ hat). Wieder ist Id_D das neutrale Element, und jetzt gibt es zu jedem Element auch ein inverses Element (die Umkehrabbildung nämlich). Also ist $(M, \circ)$ eine Gruppe.

Sie heißt die symmetrische Gruppe von D, meistens wird sie als $\mathrm{Sym}(D)$, Sym_D oder S_D notiert. Sie (oder ihre „Untergruppen“ – siehe später –) werden benutzt um Symmetrieeigenschaften von Mengen zu charakterisieren.

d) Eine Menge M mit genau einem Element m wird durch die einzig mögliche Verknüpfung darauf – $m * m = m$ – zu einer Gruppe; diese Gruppe heißt eine **triviale Gruppe**. Sie kennen zwei Beispiele hierfür: $(\{0\}, +)$ und $(\{1\}, \cdot)$.

Für jede Gruppe $(G, *)$ ist $(\{e_G\}, *)$ eine (oft sagt man auch die) triviale Gruppe.

e) Nun habe die Menge M genau zwei Elemente e und m. Wenn wir festlegen, dass e neutrales Element sein soll, so gibt es nur eine Möglichkeit der Gruppenstruktur auf M:

$$e * e = e,\ e * m = m * e = m,\ m * m = e.$$

Die ersten drei Gleichungen werden von den Eigenschaften des neutralen Elements erzwungen, die letzte von der Existenz eines zu m inversen Elements. Die Assoziativität ist offensichtlich erfüllt.

f) Es sei $n \geq 1$ eine natürliche Zahl. Auf $\mathbb{Z}$ haben wir seit Beispiel 1.4.5 die Äquivalenzrelation

$$x \equiv y \pmod{n} \iff n \text{ teilt } (x - y).$$

Es sei $\mathbb{Z}/n\mathbb{Z}$ die Menge aller Äquivalenzklassen dieser Äquivalenzrelation:

$$\mathbb{Z}/n\mathbb{Z} := \{[k] \mid k \in \mathbb{Z}\} = \{[0], [1], [2], \ldots, [n-1]\}.$$

Auf dieser Menge definieren wir eine Verknüpfung $+_n$ mittels

$$[k] +_n [l] := [k + l].$$

Damit dies wirklich eine Verknüpfung ist, darf die Zuordnung nur von den jeweiligen Äquivalenzklassen, nicht aber von der konkreten Wahl von k oder l abhängen. Man muss also Folgendes überprüfen: wenn für $\tilde{k}, \tilde{l} \in \mathbb{Z}$ die Voraussetzung erfüllt ist, dass $[k] = [\tilde{k}]$ und $[l] = [\tilde{l}]$, dann gilt $[k+l] = [\tilde{k} + \tilde{l}]$. Dies verifiziert man wie folgt: $[k] = [\tilde{k}]$ bedeutet, dass eine ganze Zahl a mit $\tilde{k} = k + an$ existiert; genauso gibt es eine ganze Zahl b mit $\tilde{l} = l + bn$. Dann ist aber

$$\tilde{k} + \tilde{l} = k + an + l + bn = k + l + (a + b) \cdot n,$$

also sind $\tilde{k} + \tilde{l}$ und $k + l$ kongruent modulo n, und ihre Äquivalenzklassen stimmen überein.

Nun ist wegen

$$([k] +_n [l]) +_n [m] = [k + l + m] = [k] +_n ([l] +_n [m])$$

die Verknüpfung assoziativ, $[0]$ ist ein neutrales Element, und für $[k] \in \mathbb{Z}/n\mathbb{Z}$ gilt

$$[k] +_n [-k] = [0].$$

Also ist $(\mathbb{Z}/n\mathbb{Z}, +_n)$ eine Gruppe.

In den Beispielen a) und d) – f) ist hier die Verknüpfung kommutativ (Definition 2.1.1).

Aufgabe 2.1.6 Automorphismen von Graphen

Es sei $\Gamma = (E, K)$ ein Graph. Erinnern Sie sich an den Begriff eines Automorphismus von Γ, 1.3.12.

Zeigen Sie, dass die Menge aller Automorphismen von Γ bezüglich der Komposition eine Gruppe ist.

Definition 2.1.7 (abelsche Gruppe)

Eine Gruppe $(G, *)$ heißt **kommutativ**, wenn $*$ eine kommutative Verknüpfung ist. Oft sagt man dann auch, die Gruppe sei **abelsch**. (Diese Bezeichnung erinnert an den norwegischen Mathematiker Niels Henrik Abel.)

Schreibweisen: Oft benutzt man als Zeichen für die Verknüpfung einen Malpunkt (und lässt dann meistens auch diesen noch weg) und schreibt x^{-1} für das Inverse zu x.

Die „additive Schreibweise“ mit $+$ als Symbol für die Verknüpfung (und $-x$ als zu x inversem Element), benutzt man höchstens für kommutative Gruppen.

Wenn klar ist, welche Verknüpfung man auf G betrachtet, so sagt man meistens, dass G eine Gruppe ist, ohne explizit die Verknüpfung mit zu erwähnen.

2.2 Untergruppen

Definition 2.2.1 (Untergruppe)

Es sei $(G, *)$ eine Gruppe und H eine Teilmenge von G. Auf H sei eine Verknüpfung $\circ$ gegeben. Dann heißt $(H, \circ)$ eine **Untergruppe** von $(G, *)$, wenn $(H, \circ)$ eine Gruppe ist und außerdem

$$\forall h_1, h_2 \in H : h_1 \circ h_2 = h_1 * h_2$$

gilt.

Das bedeutet, etwas weniger formal formuliert, dass die Einschränkung von $*$ auf $H \times H$ aus H eine Gruppe macht, also: $\forall h_1, h_2 \in H : h_1 * h_2 \in H$, $e_G \in H$ und $\forall h \in H : h^{-1} \in H$ (das Inverse, das es in G gibt, liegt sogar schon in H).

Beispiel 2.2.2 (Untergruppen)

$(\mathbb{Z}, +)$ ist eine Untergruppe von $(\mathbb{Q}, +)$, $(\{\pm 1\}, \cdot)$ ist eine Untergruppe von $(\mathbb{Q} \setminus \{0\}, \cdot)$.

Da die Verknüpfung auf einer Untergruppe von $(G, *)$ durch die auf G festgelegt ist, spricht man immer davon, dass eine Teilmenge H von G eine Untergruppe von G ist, wenn der zweite Absatz der Definition erfüllt ist.

Hilfssatz 2.2.3 (Untergruppenkriterium)

Es seien $(G, *)$ *eine Gruppe und* $H \subseteq G$ *eine Teilmenge von* G.

H *ist genau dann eine Untergruppe von* G, *wenn gilt:*

$$H \neq \emptyset \quad \text{und} \quad \forall h_1, h_2 \in H : h_1 * h_2^{-1} \in H. \tag{$\dagger$}$$

Beweis. Wenn H eine Untergruppe ist, dann ist $e_H = e_G \in H$, also H nicht leer. Außerdem liegt für alle $h_2 \in H$ das Inverse h_2^{-1} in H, und wenn auch h_1 in H liegt, so auch $h_1 * h_2^{-1}$ als Produkt von zwei Elementen der Untergruppe.

Wenn umgekehrt die zwei Bedingungen aus ($\dagger$) erfüllt sind, so gibt es ein $h \in H$, und damit liegt auch $e_G = h * h^{-1}$ in H (wobei in der Bedingung $h_1 = h_2 = h$ gesetzt wird). Damit liegt mit jedem Element h aus H auch $h^{-1} = e_G * h^{-1} \in H$, und schließlich ist H auch unter $*$ abgeschlossen, da für alle $h_1, h_2 \in H$ gilt:

$$h_2^{-1} \in H \quad \text{und daher wegen } (\dagger) \quad h_1 * h_2 = h_1 * (h_2^{-1})^{-1} \in H.$$

Also ist H eine Untergruppe. (Die Assoziativität von $(H, *)$ versteht sich von selbst.) ◯

Beispiel 2.2.4 (Untergruppen der ganzen Zahlen)

Wenn wir von $\mathbb{Z}$ als Gruppe sprechen, meinen wir immer die Addition als Verknüpfung. In diesem Beispiel wollen wir alle Untergruppen von $\mathbb{Z}$ kennenlernen.

Die triviale Untergruppe (2.1.5 e)) ist $\{0\}$. Es ist außerdem klar, dass für jede ganze Zahl n die Teilmenge

$$n\mathbb{Z} := \{nk \mid k \in \mathbb{Z}\}$$

eine Untergruppe ist, denn diese Menge ist nicht leer und mit nk und nl ist auch $nk - nl = n(k - l)$ in $n\mathbb{Z}$ enthalten. Für $n = 0$ erhalten wir wieder die triviale Untergruppe.

Wir zeigen nun umgekehrt, dass jede Untergruppe von $\mathbb{Z}$ eine der eben genannten ist. Es sei also $H \subseteq \mathbb{Z}$ eine Untergruppe, und H sei nicht die triviale Untergruppe (sonst wählen wir $n = 0$ und sind fertig). Dann gibt es in H ein von 0 verschiedenes Element x. Mit diesem liegt auch $-x$ in H, und es gibt demnach ein positives x in H. Die Menge $H \cap \mathbb{N}$ ist also nicht leer, und enthält damit auch ein kleinstes Element, welches wir n nennen. Die Behauptung ist nun, dass $H = n\mathbb{Z}$. Die Inklusion $\supseteq$ ist klar. Wenn umgekehrt $h \in H$ beliebig gewählt ist, so gilt für die größte Zahl k mit der Eigenschaft $kn \leq h$ die Ungleichung

$$0 \leq h - nk < n.$$

Da mit h und nk auch $h - nk$ in H liegt, muss nach Wahl von n die Differenz $h - nk$ gleich Null sein, also $h \in n\mathbb{Z}$.

Fazit 2.2.5

Die Untergruppen von $\mathbb{Z}$ sind genau die Mengen $n\mathbb{Z}$, wobei n die nichtnegativen ganzen Zahlen durchläuft.

Hilfssatz 2.2.6 (Durchschnitt von Untergruppen)

Es sei G eine Gruppe, I eine nichtleere Menge, und für jedes $i \in I$ sei eine Untergruppe U_i von G gegeben. Dann ist auch $\cap_{i\in I}U_i$ eine Untergruppe von G.

Beweis. In jedem U_i liegt das neutrale Element e_G, also gilt $e_G \in \cap_{i\in I}U_i$. Außerdem gilt für $h_1, h_2 \in \cap_{i\in I}U_i$:

$$\forall i \in I : h_1 * h_2^{-1} \in U_i,$$

und damit auch $h_1 * h_2^{-1} \in \cap_{i\in I}U_i$. Nach dem Untergruppenkriterium (Hilfssatz 2.2.3) ist damit alles gezeigt. ◯

Definition 2.2.7 (Gruppenerzeugnis, zyklische Gruppe)

a) Für eine Teilmenge M der Gruppe G sei I die Menge aller Untergruppen von G, die M enthalten. Dazu gehört zum Beispiel G selbst. Dann ist aber nach dem Vorhergehenden auch

$$\langle M\rangle := \bigcap_{i\in I} i$$

eine Gruppe, sie heißt das **(Gruppen-)Erzeugnis von** M oder die **von** M **erzeugte Untergruppe von** G. Es ist offensichtlich die kleinste Untergruppe von G, die M enthält.

b) Eine Gruppe G heißt **zyklisch**, wenn es ein Element $a \in G$ gibt, sodass $G = \langle\{a\}\rangle$. Hierfür schreibt man kürzer auch $G = \langle a\rangle$.

Beispiel 2.2.8 (zyklische Gruppen)

a) Für jede natürliche Zahl n ist die Gruppe $\mathbb{Z}/n\mathbb{Z}$ von $[1]$ erzeugt.

b) Für beliebiges $g \in G$ und für $n \in \mathbb{N}_0$ setzen wir $g^0 := e_G$ und für $n > 0$ schreiben wir

$$g^n := g * g * \cdots * g \text{ (n Faktoren)}, \quad g^{-n} := (g^{-1})^n.$$

Dann ist

$$\langle g\rangle = \{g^k \mid k \in \mathbb{Z}\}$$

die von g erzeugte zyklische Gruppe.

Definition 2.2.9 (Ordnung)

Die Kardinalität einer Gruppe nennt man auch ihre **Ordnung**. Die **Ordnung eines Elementes** $g \in G$ ist definiert als die Ordnung der von g erzeugten Untergruppe.

Bemerkung 2.2.10 Wenn $g \in G$ endliche Ordnung hat, dann ist diese gleich der kleinsten natürlichen Zahl k, für die $g^k = e_G$ gilt.

Satz 2.2.11 („von Lagrange“)

Es sei G eine endliche Gruppe und H eine Untergruppe von G. Dann ist die Ordnung von H ein Teiler der Ordnung von G.

Beweis. Wir definieren auf G die Relation $\sim$ durch

$$g_1 \sim g_2 :\iff g_1 g_2^{-1} \in H.$$

Dann ist $\sim$ eine Äquivalenzrelation, wie man leicht nachrechnet (ähnlich wie bei $\equiv \pmod n$ in 1.4.5). Die Äquivalenzklasse eines Elements g ist

$$[g] = Hg := \{hg \mid h \in H\}.$$

Nun ist G aber die disjunkte Vereinigung der Äquivalenzklassen (Satz 1.4.7), und wir sind fertig, wenn wir gezeigt haben, dass jede Äquivalenzklasse genauso viele Elemente hat wie H. Dies zeigen wir durch die Angabe einer Bijektion von $H = [e_G]$ nach $[g]$:

$$F : H \longrightarrow Hg,\ h \mapsto hg.$$

Diese Abbildung ist surjektiv, wie man der vorletzten Gleichung entnimmt. Sie ist injektiv, denn

$$\forall h_1, h_2 \in H : F(h_1) = F(h_2) \Rightarrow h_1 g = h_2 g \Rightarrow h_1 g g^{-1} = h_2 g g^{-1} \Rightarrow h_1 = h_2.$$

◯

Bemerkung 2.2.12 (Primzahlordnung)

Speziell ist in jeder endlichen Gruppe die Ordnung jedes Elements ein Teiler der Gruppenordnung. Die Ordnung von $[6]$ in $\mathbb{Z}/39\mathbb{Z}$ ist zum Beispiel 13. Das sieht man an der Primfaktorzerlegung: Ein Vielfaches $l \cdot 6 = 2 \cdot l \cdot 3$ ist genau dann durch $39 = 3 \cdot 13$ teilbar, wenn 13 ein Teiler von $2l$ ist, also von l.

Eine Gruppe G, deren Ordnung eine Primzahl p ist, ist immer zyklisch. Es gilt hier sogar:

$$\forall g \in G : [G = \langle g \rangle \iff g \neq e_G].$$

Insbesondere gibt es für alle natürlichen Zahlen $k < p$ eine natürliche Zahl l, sodass $\sum_{i=1}^{l} [k] = [1]$. Das heißt nichts anderes, als dass p ein Teiler von $kl - 1$ ist. Darauf kommen wir später zurück.

Aufgabe 2.2.13 Euklid

Bestimmen Sie für $k = 17$ und $p = 31$ eine Zahl l, sodass $kl-1$ durch 31 teilbar ist.

Zeigen Sie, dass es für $k = 51$ und $p = 119$ kein l gibt, sodass $kl - 1$ durch 119 teilbar ist.

2.3 Homomorphismen von Gruppen

In diesem Abschnitt geht es darum, für zwei gegebene Gruppen solche Abbildungen zu studieren, die mit den Gruppenstrukturen „verträglich“ sind. Dies wird präzisiert durch den Begriff des Gruppenhomomorphismus. Wenn wir uns Verknüpfungen wie in 2.1.3 „visualisieren“, dann legt 1.3.10 nahe, was ein Isomorphismus von Gruppen sein sollte. Wir fangen erst ohne die Bedingung der Bijektivität an.

Definition 2.3.1 (Gruppenhomomorphismus)

Es seien $(G, *)$ und $(H, \bullet)$ zwei Gruppen. Ein **(Gruppen-)Homomorphismus** von G nach H ist eine Abbildung $f : G \longrightarrow H$, für die gilt:

$$\forall x, y \in G : f(x * y) = f(x) \bullet f(y).$$

Die oben nahegelegte Analogie wäre hier zu schreiben als

$$\forall x, y, z \in G : \ (x, y, z) \in * \Rightarrow (f(x), f(y), f(z)) \in \bullet.$$

Die Menge aller Homomorphismen von G nach H nennen wir $\mathrm{Hom}(G, H)$. Dies ist streng genommen eine Abkürzung für die umständlichere aber aussagekräftigere Notation $\mathrm{Hom}_{\mathrm{Gruppen}}((G, *), (H, \bullet))$.

Beispiel 2.3.2 (Gruppenhomomorphismen)

a) Für beliebige Gruppen G und H ist die Abbildung

$$f : G \longrightarrow H, \ \ \forall x \in G : f(x) := e_H,$$

ein Gruppenhomomorphismus, der so genannte **triviale** Homomorphismus.

b) Für $G = \mathbb{Z}$ und beliebiges h in beliebigem H ist die Abbildung (mit Notation aus 2.2.8b))

$$f : \mathbb{Z} \longrightarrow H, \ \ \forall x \in \mathbb{Z} : f(x) := h^x,$$

ein Homomorphismus von $\mathbb{Z}$ nach H:

$$f(x + y) = h^{x+y} = h^x \bullet h^y = f(x) \bullet f(y).$$

c) Für $G = (\mathbb{R}, +)$ und $H = (\mathbb{R}_{>0}, \cdot)$ ist die e-Funktion

$$\forall x \in \mathbb{R} : x \mapsto e^x$$

ein Gruppenhomomorphismus: $e^{x+y} = e^x \cdot e^y$.

Wir wollen einige grundsätzliche Eigenschaften von Gruppenhomomorphismen kennenlernen.

Hilfssatz 2.3.3 (Eigenschaften von Homomorphismen)
Es sei $f : G \longrightarrow H$ ein Homomorphismus von Gruppen. Dann gelten die folgenden Aussagen:

a) $f(e_G) = e_H$.

b) $\forall g \in G : \ f(g^{-1}) = f(g)^{-1}$. *Dabei ist links das Inverse in G, rechts das in H gemeint.*

c) $f^{-1}(\{e_H\})$ *(das Urbild des neutralen Elements von H) ist eine Untergruppe von G.*

d) $f(G)$ *ist eine Untergruppe von H.*

e) f *ist genau dann injektiv, wenn* $f^{-1}(\{e_H\}) = \{e_G\}$.

Beweis. a) Es gilt

$$f(e_G) = f(e_G * e_G) = f(e_G) \bullet f(e_G).$$

Diese Gleichung wird nun mit dem zu $f(e_G)$ inversen Element multipliziert, und es folgt

$$e_H = f(e_G).$$

b) Es gilt

$$f(g) \bullet f(g^{-1}) = f(g * g^{-1}) = f(e_G) = e_H.$$

Genauso gilt auch $f(g^{-1}) \bullet f(g) = e_H$.
Nach Definition des inversen Elements heißt das $f(g^{-1}) = f(g)^{-1}$.

c) Wegen Teil a) gilt $e_G \in f^{-1}(\{e_H\})$, also ist $f^{-1}(\{e_H\})$ nicht leer. Nach dem Untergruppenkriterium 2.2.3 ist noch zu zeigen, dass für alle $x, y \in f^{-1}(\{e_H\})$ auch das Produkt $x * y^{-1}$ in $f^{-1}(\{e_H\})$ liegt. Dies folgt aber wegen b) aus

$$f(x * y^{-1}) = f(x) \bullet f(y^{-1}) = e_H \bullet e_H^{-1} = e_H.$$

d) Wegen a) ist $e_H = f(e_G) \in f(G)$. Nach dem Untergruppenkriterium genügt es also zu zeigen, dass für alle $a, b \in f(G)$ auch $a \bullet b^{-1} \in f(G)$ gilt. Wegen $a, b \in f(G)$ gibt es aber definitionsgemäß $x, y \in G$ mit

$$a = f(x), \quad b = f(y).$$

Nach b) folgt dann

$$a \bullet b^{-1} = f(x) \bullet f(y)^{-1} = f(x) \bullet f(y^{-1}) = f(x * y^{-1}) \in f(G).$$

e) Wenn f injektiv ist, dann liegt in $f^{-1}(\{e_H\})$ nicht mehr als ein Element, aber e_G liegt nach a) darin, also folgt

$$f^{-1}(\{e_H\}) = \{e_G\}.$$

Gilt umgekehrt diese Mengengleichheit, dann folgt für $x, y \in G$ mit $f(x) = f(y)$:

$$e_H = f(y) \bullet f(y)^{-1} = f(x) \bullet f(y^{-1}) = f(x * y^{-1})$$

und damit $x * y^{-1} \in f^{-1}(\{e_H\}) = \{e_G\}$. Das heißt aber $x = y$, und f muss injektiv sein. ◯

Definition 2.3.4 (Kern)

Ist $f : G \longrightarrow H$ ein Homomorphismus zwischen zwei Gruppen, so heißt die Untergruppe $f^{-1}(\{e_H\}) \subseteq G$ der **Kern** von f. Wir haben also gerade gezeigt:

$$\boxed{f \in \mathrm{Hom}(G, H) \text{ ist genau dann injektiv, wenn } \mathrm{Kern}\,(f) = \{e_G\}.}$$

Wegen dieses Sachverhaltes und wegen des damit eng verknüpften Homomorphiesatzes führt man den Kern überhaupt ein.

Beispiel 2.3.5 (Kerne)

a) Der Kern des trivialen Homomorphismus (siehe 2.3.2a)) von G nach H ist G, sein Bild ist $\{e_H\}$.

b) Im Beispiel 2.3.2b) ist das Bild des Homomorphismus

$$\mathbb{Z} \longrightarrow H, \quad x \mapsto h^x,$$

die von h erzeugte Gruppe $\langle h \rangle$, und der Kern ist entweder $\{0\}$, nämlich wenn h nicht endliche Ordnung hat, oder er ist die Untergruppe von $\mathbb{Z}$, die von der Ordnung von h erzeugt wird.

c) Die Exponentialabbildung $\mathbb{R} \ni x \mapsto e^x \in \mathbb{R}_{>0}$ ist surjektiv, ihr Kern besteht nur aus der 0, also ist sie auch injektiv. Sie ist ein bijektiver Homomorphismus von $(\mathbb{R}, +)$ nach $(\mathbb{R}_{>0}, \cdot)$.

Definition 2.3.6 (Endo-, Auto-, Isomorphismus)

a) Für eine Gruppe G heißt ein Homomorphismus von G nach G auch ein **Endomorphismus.** Die Menge aller Endomorphismen wird mit $\mathrm{End}(G)$ notiert.

b) Ein bijektiver Homomorphismus zwischen zwei Gruppen G und H heißt ein **Isomorphismus** zwischen G und H.

c) Einen bijektiven Endomorphismus der Gruppe G nennt man **Automorphismus** von G. Die Menge aller Automorphismen wird mit $\mathrm{Aut}(G)$ notiert.

Schreibweise: Wenn es (mindestens) einen Isomorphismus zwischen G und H gibt, so nennt man sie **isomorph**, und schreibt dafür $G \cong H$. Isomorph zu sein ist eine Äquivalenzrelation auf jeder Menge von Gruppen.

Beispiel 2.3.7 Wir haben gerade gesehen, dass die Exponentialabbildung ein Isomorphismus ist. Ein zweites Beispiel gewinnen wir wie folgt.

Es sei $G = \{1, -1\}$ mit Multiplikation und $H = \mathbb{Z}/2\mathbb{Z}$. Dann ist die Abbildung

$$f : G \longrightarrow H, \quad f(1) = [0], \ f(-1) = [1],$$

ein Gruppenisomorphismus.

Hilfssatz 2.3.8 (Invertieren eines Isomorphismus)

Es sei $f : G \longrightarrow H$ ein Isomorphismus von Gruppen. Dann ist auch die Umkehrabbildung $f^{-1} : H \longrightarrow G$ ein Gruppenisomorphismus.

Beweis. Es seien a, b in H beliebig. Dann gilt

$$\begin{aligned} f^{-1}(a \bullet b) &= f^{-1}\Big(f\big(f^{-1}(a)\big) \bullet f\big(f^{-1}(b)\big)\Big) \\ &= f^{-1}\Big(f\big(f^{-1}(a) * f^{-1}(b)\big)\Big) = f^{-1}(a) * f^{-1}(b). \end{aligned}$$

○

Aufgabe: Lassen Sie sich diesen Beweis auf der Zunge zergehen.

Bemerkung 2.3.9 Es sei K der Kern des Homomorphismus $f : G \longrightarrow H$. Dann gilt für alle $g \in G$ und alle $k \in K$, dass auch $g * k * g^{-1} \in K$:

$$f(g * k * g^{-1}) = f(g) \bullet f(k) \bullet f(g)^{-1} = f(g) \bullet e_H \bullet f(g)^{-1} = e_H.$$

Dieser Eigenschaft von K gibt man einen Namen und nennt K einen **Normalteiler**. Dieser Begriff ist in der Gruppentheorie sehr wichtig, für die Zwecke der Linearen Algebra spielt er nicht die entscheidende Rolle.

Bemerkung 2.3.10 (Verknüpfung von Homomorphismen)

a) Es seien $(G, *), (H, \bullet), (I, \#)$ drei Gruppen und

$$\Phi : G \longrightarrow H, \quad \Psi : H \longrightarrow I$$

zwei Gruppenhomomorphismen. Dann ist auch

$$\Psi \circ \Phi : G \longrightarrow I, \ g \mapsto \Psi(\Phi(g)),$$

ein Gruppenhomomorphismus:

$$\begin{aligned} \forall g_1, g_2 \in G : (\Psi \circ \Phi)(g_1 * g_2) \ &= \Psi(\Phi(g_1 * g_2)) = \Psi(\Phi(g_1) \bullet \Phi(g_2)) = \\ &= \Psi(\Phi(g_1)) \# \Psi(\Phi(g_2)). \end{aligned}$$

b) Insbesondere ist für zwei Automorphismen von G auch die Komposition ein Automorphismus. Da die Identität ein Automorphismus ist und auch die Inverse eines Automorphismus wieder ein solcher ist, bilden die Automorphismen von G eine Untergruppe der symmetrischen Gruppe (siehe 2.1.5c)) der Menge G. $(\mathrm{Aut}(G), \circ)$ heißt die **Automorphismengruppe von** G.

Ein Blick auf 2.1.6 zeigt, dass auch andere Objekte Automorphismengruppen haben können. Auch uns wird dieser Begriff immer wieder begegnen.

2.4 Die symmetrische Gruppe

In diesem Abschnitt wollen wir das Beispiel der symmetrischen Gruppe einer endlichen Menge D (siehe 2.1.5) noch etwas eingehender studieren. Die Ergebnisse werden wir später bei der Einführung der Determinanten noch einmal brauchen.

Da für unsere Zwecke die Namen der Elemente der Menge D, deren symmetrische Gruppe betrachtet wird, keine Rollen spielen, wählen wir für die Menge D mit n Elementen die Menge

$$D := \{1, 2, \ldots, n\} = \{i \in \mathbb{N} \mid 1 \leq i \leq n\}.$$

Statt $\mathrm{Sym}\,(D)$ schreiben wir kürzer S_n.

Die Elemente von S_n sind die Permutationen von D, also die bijektiven Abbildungen von D nach D. Eine solche Bijektion geben wir zunächst durch ihre Wertetabelle an. Zum Beispiel finden wir für $n = 3$ die folgenden 6 Permutationen:

	1	2	3
Id	1	2	3
ζ_1	2	3	1
ζ_2	3	1	2
τ_1	1	3	2
τ_2	3	2	1
τ_3	2	1	3

Um allgemein eine Permutation σ von D festzulegen, hat man für die Wahl von $\sigma(1)$ zunächst n Möglichkeiten. Wenn $\sigma(1)$ gewählt ist, hat man für $\sigma(2)$ nur noch $n-1$ Möglichkeiten, denn σ muss ja injektiv sein. Für $\sigma(3)$ gibt es dann nur noch $n-2$ Möglichkeiten und so fort, und insgesamt finden sich

$$\#S_n = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1 =: n!$$

Permutationen von D – die symmetrische Gruppe wird schnell sehr groß, wenn n groß wird. Deswegen will man ihre Elemente übersichtlicher hinschreiben.

Definition 2.4.1 (Zykel, Transpositionen)

a) Es seien $x_1, x_2, \ldots, x_k \in D$ k paarweise verschiedene Elemente. Dann wird durch

$$\zeta(x) := \begin{cases} x_{i+1} & \text{wenn } x = x_i,\ i < k, \\ x_1 & \text{wenn } x = x_k, \\ x & \text{wenn } x \notin \{x_1, \ldots, x_k\}. \end{cases}$$

eine Abbildung ζ von D nach D definiert, die offensichtlich bijektiv und damit eine Permutation ist. Die so definierte Abbildung heißt ein k-**Zykel**. Wir schreiben dafür anstelle der sperrigen und unübersichtlichen Wertetabelle von ζ:

$$\zeta =: (x_1\ x_2\ \ldots\ x_k).$$

Zum Beispiel haben wir eben in der Tabelle für S_3 die Dreizykel $\zeta_1 = (1\ 2\ 3)$ und $\zeta_2 = (1\ 3\ 2)$ gesehen.

b) Einen 2-Zykel nennt man auch eine **Transposition.**

Im Beispiel der S_3 gibt es die Transpositionen $\tau_1 = (2\ 3)$, $\tau_2 = (1\ 3)$ und $\tau_3 = (1\ 2)$.

c) Die Ordnung eines k-Zykels (siehe 2.2.9) ist k. Speziell gilt für Transpositionen τ stets:

$$\tau^{-1} = \tau.$$

Jeden k-Zykel kann man auch als Produkt von $k-1$ Transpositionen schreiben:

$$(x_1\ x_2\ \ldots\ x_k) = (x_1\ x_2) \circ (x_2\ x_3) \circ \cdots \circ (x_{k-1}\ x_k).$$

Um dies zu verifizieren müssen wir uns überlegen, was das Produkt π auf der rechten Seite mit einem Element $x \in D$ macht. Wenn x nicht zu $\{x_1, \ldots, x_k\}$ gehört, dann lässt jede Transposition rechts x fest, also gilt $\pi(x) = x$. Für $x = x_i$ überlegt man sich, dass x_i für $i < k$ rechts im i-ten Faktor vorkommt, der es mit x_{i+1} vertauscht. Keine weiter rechts stehende (also vorher auszuführende) Transposition hat x_i schon bewegt, und kein weiter links stehender Faktor bewegt x_{i+1}. Also macht die rechte Seite mit x_i dasselbe wie die linke Seite. Die letzte

Zahl x_k wird von rechts nach links „durchgeschoben“ und landet am Ende bei x_1.

d) Wir fassen stets S_{n-1} als Untergruppe von S_n auf, nämlich als die Menge all der Permutationen von D, die n auf sich selbst abbilden.

Hilfssatz 2.4.2 (S_n wird von Transpositionen erzeugt)

Es sei $\sigma \in S_n$.

Dann gibt es ein $k \in \mathbb{N}_0$ und Transpositionen $\tau_1, \ldots, \tau_k \in S_n$, sodass

$$\sigma = \tau_1 \circ \tau_2 \circ \cdots \circ \tau_k.$$

Beweis. Wir machen vollständige Induktion nach n. Für $n = 0$ oder 1 folgt die Behauptung mit $k = 0$: ein leeres Produkt ist das neutrale Element. Aber auch die Fälle $n = 2$ oder $n = 3$ sind klar.

Sei also $n \geq 2$ und die Behauptung für kleinere Werte von n bewiesen. Wähle ein $\sigma \in S_n$. Wenn $\sigma(n) = n$ gilt, dann liegt σ schon in S_{n-1} und die Induktionsvoraussetzung greift direkt. Wenn $a = \sigma(n) \neq n$ gilt, so benutze die Transposition $\tau = (a\ n)$:

$$(\tau \circ \sigma)(n) = \tau(a) = n, \quad \text{folglich} \quad \tau \circ \sigma \in S_{n-1}.$$

Also ist $\tau \circ \sigma$ nach Induktionsvoraussetzung ein Produkt von Transpositionen und damit auch σ selbst, da $\tau^{-1} = \tau$ eine Transposition ist. ◯

Jetzt haben wir aber unsere Permutationen schon fast wieder zu klein gehackt... es gibt auch den folgenden Zwischenweg.

Bemerkung 2.4.3 (Zerlegung in disjunkte Zykel)

Jede Permutation $\sigma \in S_n$ lässt sich wie folgt auf noch überschaubarere Weise als Produkt von Zykeln schreiben. Wir werden das in Beispiel 2.5.5 a) noch einmal anders verstehen.

Wir betrachten $\{1, \sigma(1), \sigma(\sigma(1)), \ldots, \sigma^n(1)\}$. Diese Menge habe $k \leq n$ Elemente, und es sei für $1 \leq i \leq k$ die Zahl x_i definiert durch $x_i := \sigma^{i-1}(1)$. Dann sind diese x_i k paarweise verschiedene Elemente von D und σ macht auf diesen Elementen dasselbe wie der k-Zykel $(x_1\ x_2\ \ldots\ x_k)$.

Wenn $k < n$ gilt, so sei $y_1 = \min(D \setminus \{x_1, \ldots, x_k\})$. Wir analysieren die Menge $\{y_1, \sigma(y_1), \ldots, \sigma^n(y_1)\}$. Es sei l die Anzahl ihrer Elemente. Diese schreiben wir ähnlich wie eben auf als

$$\forall i \in \{1, \ldots, l\} : y_i := \sigma^{i-1}(y_1).$$

Dann macht σ auf $\{y_1, \ldots, y_l\}$ dasselbe wie der Zykel $(y_1\ y_2\ \ldots\ y_l)$. Die zwei Mengen $\{x_1, \ldots, x_k\}$ und $\{y_1, \ldots, y_l\}$ sind disjunkt (d.h. ihr Durchschnitt ist

leer). So kann man sukzessive weitermachen und dabei ist am Ende (wenn man alle Elemente aus D in jeweils einem Zykel untergebracht hat) σ das Produkt der so erhaltenen Zykel.

Diese Überlegungen werden in Abschnitt 2.5 präzisiert.

Beispiel 2.4.4 (Zykelzerlegung)

Es sei $\sigma \in S_7$ definiert durch seine Wertetabelle

x	1	2	3	4	5	6	7
$\sigma(x)$	2	5	4	7	1	3	6

Dann ist $\sigma(1) = 2, \sigma(2) = 5, \sigma(5) = 1,$ also steckt 1 in dem 3-Zykel (1 2 5). Die kleinste Zahl, die hier noch nicht verarbeitet ist, ist 3. Wir finden $\sigma(3) = 4, \sigma(4) = 7, \sigma(7) = 6, \sigma(6) = 3,$ und damit gilt

$$\sigma = (1\ 2\ 5) \circ (3\ 4\ 7\ 6).$$

Definition 2.4.5 (Signum)

Das **Signum** einer Permutation $\sigma \in S_n$ ist definiert durch

$$\operatorname{sign}(\sigma) := \prod_{1 \le i < j \le n} \frac{\sigma(j) - \sigma(i)}{j - i}.$$

Da oben und unten bis auf das Vorzeichen und die Reihenfolge dieselben Faktoren stehen, ist der Wert des Signums stets 1 oder -1.

Tatsächlich kommt es bei der Definition des Signums nicht darauf an, dass man wirklich $i < j$ verwendet; wichtig ist nur, dass man für jedes Paar $1 \le i \ne j \le n$ entscheidet, in welcher Reihenfolge die Differenz gebildet wird, und dies in Zähler und Nenner gleichermaßen handhabt. Das heißt insbesondere, dass für jede Permutation τ sich das Signum von σ auch durch

$$\operatorname{sign}(\sigma) = \prod_{1 \le i < j \le n} \frac{\sigma(\tau(j)) - \sigma(\tau(i))}{\tau(j) - \tau(i)}$$

berechnen lässt. Daraus folgt aber

$$\begin{aligned}
\operatorname{sign}(\sigma \circ \tau) &= \textstyle\prod_{1 \le i < j \le n} \frac{\sigma(\tau(j)) - \sigma(\tau(i))}{j - i} \\
&= \textstyle\prod_{1 \le i < j \le n} \left[\frac{\sigma(\tau(j)) - \sigma(\tau(i))}{j - i} \cdot \frac{\tau(j) - \tau(i)}{\tau(j) - \tau(i)} \right] \\
&= \textstyle\prod_{1 \le i < j \le n} \frac{\sigma(\tau(j)) - \sigma(\tau(i))}{\tau(j) - \tau(i)} \cdot \frac{\tau(j) - \tau(i)}{j - i} \\
&= \operatorname{sign}(\sigma) \cdot \operatorname{sign}(\tau).
\end{aligned}$$

Fazit 2.4.6

Das Signum ist ein Gruppenhomomorphismus von S_n nach $(\{\pm 1\}, \cdot)$.

Folgerung 2.4.7 (Berechnung des Signums)

Nun wollen wir die ursprüngliche Definition des Signums schnell vergessen und sehen, wie sich das Signum einer Permutation bequemer berechnen lässt. Dazu beschränken wir uns auf den Fall $n \geq 2$, denn S_0 und S_1 sind ohnehin die triviale Gruppe. Für $n \geq 2$ berechnen wir zunächst das Signum für die Transposition $\tau := (1\ 2)$. Das ist die letzte Gelegenheit, bei der wir die Definition des Signums verwenden. Für $1 \leq i < j \leq n$ gilt:

$$\tau(j) > \tau(i) \iff (i,j) \neq (1,2),$$

wie man leicht nachprüft. Damit gibt es in der Formel aus der Definition von $\operatorname{sign}(\tau)$ genau einen Faktor, der negativ ist, und es gilt

$$\operatorname{sign}((1\ 2)) = -1.$$

Wenn $a \neq b$ beliebige Elemente aus D sind, dann gibt es eine (im allgemeinen sogar sehr viele) Permutation $\pi \in S_n$ mit

$$\pi(a) = 1, \quad \pi(b) = 2.$$

Dann gilt aber für $x \in D$:

$$[\pi^{-1} \circ (1\ 2) \circ \pi](x) = \begin{cases} a & \text{falls } x = b, \\ b & \text{falls } x = a, \\ x & \text{falls } x \notin \{a, b\}. \end{cases}$$

Das aber heißt, dass

$$\pi^{-1} \circ (1\ 2) \circ \pi = (a\ b),$$

und daraus folgt mit 2.3.1 und 2.3.3

$$\operatorname{sign}((a\ b)) = (\operatorname{sign}(\pi))^{-1} \cdot \operatorname{sign}(1\ 2) \cdot \operatorname{sign}(\pi) = \operatorname{sign}((1\ 2)) = -1.$$

Jede Transposition hat Signum -1.

Hat man nun eine beliebige Permutation σ gegeben, so schreibe man sie mit Hilfssatz 2.4.2 als Produkt von Transpositionen. Falls man hierzu l Faktoren benötigt, dann ist das Signum von σ gleich

$$\operatorname{sign}(\sigma) = (-1)^l.$$

Zum Beispiel hat ein k-Zykel Signum $(-1)^{k-1}$.

Aufgabe 2.4.8 (Bitte konkreter!)

Wir betrachten in S_{18} die Permutation σ, die durch folgende Wertetabelle gegeben ist:

$x =$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
$\sigma(x) =$	2	7	3	15	14	4	1	6	17	16	9	10	5	8	11	12	13	18

Bestimmen Sie die Zykelzerlegung und das Signum von σ.

2.5 Gruppenoperationen

Dieser Abschnitt ist in der Linearen Algebra von optionaler Natur.

Die Gruppentheorie dient dem Zweck, verschiedene Beispiele von Gruppen, die man ohnehin kennt und benutzt, unter einem einheitlichen Gesichtspunkt zu betrachten, indem eben die Gruppenaxiome als gemeinsames Wesensmerkmal der Beispiele herausdestilliert werden.

Wir haben bisher zwei Typen von Gruppen kennengelernt: Gruppen von Zahlen mit Addition oder Multiplikation als Verknüpfung und damit Verwandte (die Gruppen $\mathbb{Z}/n\mathbb{Z}$) stellen den einen Typ dar, die symmetrischen Gruppen den anderen. Der zweite Typ von Gruppen ist also dazu da, etwas mit Elementen einer Menge anzufangen. Dieser Aspekt soll hier etwas vertieft werden.

Definition 2.5.1 (Gruppenoperation)

Es seien $(G, *)$ eine Gruppe und M eine Menge. Dann ist eine **Operation von G auf M** definiert als eine Abbildung

$$\bullet : G \times M \longrightarrow M,$$

sodass die folgenden Bedingungen erfüllt sind:

a) $\forall m \in M : e_G \bullet m = m$,

b) $\forall m \in M, g_1, g_2 \in G : g_1 \bullet (g_2 \bullet m) = (g_1 * g_2) \bullet m$.

Wenn $G \subseteq S_M$ eine Untergruppe der symmetrischen Gruppe von M ist, dann wird solch eine Abbildung $\bullet$ zum Beispiel durch

$$g \bullet m := g(m)$$

gegeben. Dies ist die Urmutter aller Operationen, wie wir gleich sehen werden.

Hilfssatz 2.5.2 (Operationen und symmetrische Gruppe)

Es seien G eine Gruppe und M eine Menge.

a) Für jeden Homomorphismus $\Phi : G \longrightarrow S_M$ wird durch

$$g \bullet m := \Phi(g)(m)$$

eine Operation von G auf M festgelegt.

b) Für jede Operation $\bullet$ von G auf M gibt es einen Homomorphismus Φ, sodass $\bullet$ wie in Teil a) konstruiert werden kann.

Beweis. a) Dass hierbei aus einem Homomorphismus eine Operation gewonnen wird, ist klar.

b) Sei umgekehrt eine beliebige Operation $\bullet$ gegeben. Wir zeigen, wie man aus ihr den passenden Homomorphismus konstruiert. Für jedes $g \in G$ sei $\Phi_g : M \longrightarrow M$ die Abbildung, die durch

$$\forall m \in M : \ \Phi_g(m) := g \bullet m$$

gegeben wird. Die Abbildung Φ_g ist eine Bijektion, da die Abbildung $\Phi_{g^{-1}}$ zu ihr invers ist:

$$\begin{aligned} \forall m \in M : \quad & (\Phi_g \circ \Phi_{g^{-1}})(m) = g \bullet (g^{-1} \bullet m) \\ & = (g * g^{-1}) \bullet (m) = e_G \bullet m \\ & = m \\ & = (g^{-1} * g) \bullet (m) \\ & = (\Phi_{g^{-1}} \circ \Phi_g)(m). \end{aligned}$$

Also ist $g \mapsto \Phi_g$ eine Abbildung von G nach S_M, und diese ist ein Gruppenhomomorphismus wegen der zweiten Bedingung an die Operation. ◯

Beispiel 2.5.3 (für Operationen)

a) Im Fall $G = M$ wird eine wichtige Operation durch die Gruppenverknüpfung selbst festgelegt: $\bullet = *$. Man sieht leicht, dass der zugehörige Homomorphismus Φ von G in die symmetrische Gruppe S_G injektiv ist, denn

$$\Phi(g)(e_G) = g * e_G = g,$$

also kann man g aus $\Phi(g)$ ablesen.

Das Bild von Φ ist also eine zu G isomorphe Untergruppe von S_G, und damit ist jede Gruppe isomorph zu einer Untergruppe einer symmetrischen Gruppe. Diese Aussage nennt man oft den **Satz von Cayley**, den wir hiermit bewiesen haben.

b) Eine Untergruppe $G \subseteq S_M$ operiert auf der Potenzmenge von M durch

$$\forall \sigma \in G, \ A \subseteq M : \sigma \bullet A := \sigma(A).$$

Auf ähnlichem Wege „induziert“ jede Gruppenoperation einer Gruppe auf einer Menge M eine Operation derselben Gruppe auf der Potenzmenge von M.

Hilfssatz 2.5.4 *Es sei G eine Gruppe, die auf der Menge M operiert. Dann wird auf M durch die Vorschrift*

$$m_1 \sim m_2 : \Longleftrightarrow \exists g \in G : m_1 = g \bullet m_2$$

eine Äquivalenzrelation definiert.

Beweis. Die Relation ist reflexiv, da

$$\forall m \in M : m = e_G \bullet m, \text{ also } m \sim m.$$

Sie ist symmetrisch, da für alle $m_1, m_2 \in M$ gilt:

$$\begin{aligned} m_1 \sim m_2 &\Rightarrow \exists g \in G : g \bullet m_1 = m_2 \\ &\Rightarrow \exists g \in G : g^{-1} \bullet (g \bullet m_1) = g^{-1} \bullet m_2 \\ &\Rightarrow \exists g \in G : m_1 = g^{-1} \bullet m_2 \Rightarrow m_2 \sim m_1. \end{aligned}$$

Sie ist transitiv, da aus $m_1 \sim m_2$ und $m_2 \sim m_3$ die Existenz von $g_1, g_2 \in G$ mit

$$m_1 = g_1 \bullet m_2, \quad m_2 = g_2 \bullet m_3, \text{ also } m_1 = (g_1 * g_2) \bullet m_3$$

folgt und damit $m_1 \sim m_3$. ◯

Wie für jede Äquivalenzrelation, so gibt es auch hier wieder eine disjunkte Zerlegung von M in Äquivalenzklassen (siehe 1.4.8). Die Äquivalenzklasse von m wird hier zumeist als Gm oder (präziser) $G \bullet m$ notiert und die **Bahn von** m (unter der gegebenen Operation von G) genannt.

Es ist $G \bullet m = \{g \bullet m \mid g \in G\}$.

Beispiel 2.5.5 (Binomialkoeffizienten, Satz von Lagrange)

a) Für die natürliche Zahl n sei $M := \{1, \ldots, n\}$ und $G := S_n$. Wir haben am Anfang von Abschnitt 2.4 gesehen, dass G genau $n!$ Elemente enthält.

Es sei $\sigma \in G$ gegeben und $H := \langle \sigma \rangle = \{\sigma^k \mid k \in \mathbb{Z}\}$. Dann operiert auch H auf M, und die Zerlegung von M in Bahnen unter H entspricht der Zykelzerlegung von σ aus 2.4.3. Insbesondere ist hierdurch geklärt, dass diese Zykel wirklich disjunkt sind.

b) Wenn G eine Gruppe und H eine Untergruppe von G ist, dann operiert H auf G durch die Gruppenmutiplikation

$$\bullet : H \times G \longrightarrow G, \quad h \bullet g := hg.$$

Die Bahnen dieser Operation sind gerade die Mengen $Hg := \{hg \mid h \in H\}$, die im Beweis von Satz 2.2.11 eine entscheidende Rolle spielten. Sie heißen die (Links-)Nebenklassen von G nach H.

c) Nun seien zwei natürliche Zahlen $k \leq n$ gegeben. Wie viele Teilmengen der Mächtigkeit k enthält $M = \{1, \ldots, n\}$?

Dazu sei $B := \{A \subseteq M \mid \#A = k\}$. Zu B gehört

$$A_0 := \{1, \ldots, k\}.$$

Wir erinnern uns daran, dass G auf $\mathcal{P}(M)$ operiert. Die Kardinalität einer Teilmenge von M bleibt dabei unverändert. Also liegt die Bahn von A_0 in B. Es ist umgekehrt klar, dass es für jedes $A \in B$ mindestens eine Permutation $\sigma \in G$ gibt, die $\sigma(A_0) = A$ erfüllt. Also ist die Bahn von A_0 gleich B.

Nun ist es so, dass für zwei Permutationen $\sigma, \tau \in G$ mit $\sigma(A_0) = \tau(A_0)$ die Beziehung

$$(\sigma^{-1} \circ \tau)(A_0) = A_0$$

gilt. Man rechnet leicht nach, dass

$$F := \{\rho \in G \mid \rho(A_0) = A_0\}$$

eine Untergruppe von G ist. Also gilt

$$\sigma(A_0) = \tau(A_0) \iff \sigma^{-1} \circ \tau \in F.$$

Das liefert eine analoge Äquivalenzrelation wie in Satz 2.2.11 (Lagrange). Die Elemente von B entsprechen damit bijektiv den Äquivalenzklassen $\sigma F \subseteq G$.

Die Untergruppe F enthält genau $k! \cdot (n-k)!$ Elemente, denn die k Elemente von A_0 und die $n-k$ Elemente von $M \setminus A_0$ werden unabhängig voneinander jeweils unter sich permutiert. Daher gibt es nach Lagrange

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$$

Elemente in B.

Dies ist ein Beispiel für die sogenannte Bahnbilanzformel, welche ein allgemeines Faktum aus der Theorie der Gruppenoperationen ist.

Kapitel 3

Ringe und Körper

Die Gruppentheorie hat unter Anderem das Ziel, formale Gemeinsamkeiten von arithmetischen Aspekten (Gruppen von Zahlen) und von Symmetrieaspekten (Permutationsgruppen) einheitlich zu beschreiben. Es geht dabei um die Betrachtung jeweils einer Verknüpfung. In der Arithmetik ist es nun meistens so, dass man sich darüber hinaus für das Wechselspiel von Addition und Multiplikation interessiert. Dies führt zum abstrakten Ringbegriff.

3.1 Ringe und Ringhomomorphismen

Definition 3.1.1 (Ring, Teilring)

a) Eine Menge R mit zwei Verknüpfungen $+$ (Addition genannt) und $\cdot$ (Multiplikation genannt) heißt ein **Ring**, wenn die folgenden Bedingungen erfüllt sind:

- $(R, +)$ ist eine kommutative Gruppe. (Ihr neutrales Element schreiben wir dann als 0_R bzw. in aller Regel einfach 0.)
- Die Verknüpfung $\cdot$ ist assoziativ.
- Für die Multiplikation existiert ein neutrales Element $1_R \in R$ (oder einfacher 1), das **Einselement**:

$$\forall x \in R : 1_R \cdot x = x \cdot 1_R = x.$$

- Es gelten die Distributivgesetze, das heißt für alle $x, y, z \in R$ gelten

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

b) Ein Ring $(R, +, \cdot)$ mit kommutativer Multiplikation wird ein **kommutativer Ring** genannt.

c) Es sei R ein Ring. Eine Teilmenge $T \subseteq R$ heißt ein **Teilring** von R, wenn $1_R \in T$, wenn zusätzlich für alle $t_1, t_2 \in T$ gilt, dass auch $t_1 + t_2$ und $t_1 \cdot t_2$ in T liegen, und wenn schließlich T mit diesen von R ererbten Verknüpfungen ein Ring ist. Hierfür muss man nur noch für alle $t \in T$ fordern, dass auch $-t \in T$. Assoziativ- und Distributivgesetze in T gelten dann, weil sie in R gelten und nur Allquantoren eine Rolle spielen.

d) Wir machen uns die Konvention „Punkt vor Strich“ zu Eigen, das heißt, ein Ausdruck der Gestalt $a \cdot b + c$ mit a, b, c in einem Ring R ist immer zu lesen als $(a \cdot b) + c$. Außerdem werden wir oft den Malpunkt unterdrücken: $ab := a \cdot b$.

Beispiel 3.1.2 (ein paar Ringe)

a) $\mathbb{Z}, \mathbb{Q}$ und $\mathbb{R}$ mit der üblichen Addition und Multiplikation sind Ringe, und sogar kommutative Ringe.

b) Die Menge der stetigen reellwertigen Funktionen auf dem Intervall $[0, 1]$ (oder sonst einer Teilmenge von $\mathbb{R}$) ist ein Ring, wobei Addition und Multiplikation punktweise erklärt werden:

$$(f + g)(x) = f(x) + g(x), \quad (f \cdot g)(x) = f(x) \cdot g(x).$$

Summen und Produkte stetiger Funktionen sind wieder stetig, die konstante Nullfunktion ist neutral für die Addition, und die konstante Einsfunktion ist neutral für die Multiplikation.

Genauso ist die Menge der reellen Cauchy-Folgen ein Ring, wie man in der Analysis lernt (oft ohne das so zu nennen).

c) Für eine natürliche Zahl n wird die Menge $\mathbb{Z}/n\mathbb{Z} = \{[0], \ldots, [n-1]\}$ ein Ring, wenn wir zusätzlich zur Addition aus Beispiel 2.1.5 f) vereinbaren, dass

$$[k] \cdot [l] := [k \cdot l].$$

(Im Gegensatz zu 2.1.5 schreiben wir nicht $\cdot_n$ als Malpunkt. Auch bei der Addition lassen wir aus Bequemlichkeit ab jetzt das $_n$ im Index weg.)

Wie in 2.1.5 f) muss man sich überlegen, dass dies wohldefiniert ist, also nicht von k und l sondern nur von der jeweiligen Äquivalenzklasse abhängt:

$$\begin{aligned} \forall a, b \in \mathbb{Z} : [(k + an)(l + bn)] &= [kl + kbn + anl + abn^2] \\ &= [kl + (kb + al + abn) \cdot n] = [kl]. \end{aligned}$$

Damit haben wir eine Multiplikation, und die Assoziativ- und Distributivgesetze erhalten wir wieder durch die entsprechenden Sätze für die ganzen Zahlen, z.B.

$$([k] + [l]) \cdot [m] = [k + l] \cdot [m] = [(k + l) \cdot m] = [km + lm] = [k] \cdot [m] + [l] \cdot [m].$$

Die so konstruierten Ringe $\mathbb{Z}/n\mathbb{Z}$ (sie heißen die **Restklassenringe von** $\mathbb{Z}$) sind äußerst wichtig, auch für viele Anwendungen in der Informatik.

d) Wenn R ein beliebiger Ring ist und M eine nichtleere Menge, so können wir die Menge $\mathrm{Abb}(M, R)$ zu einem Ring machen, indem wir für $f, g \in \mathrm{Abb}(M, R)$ die folgenden Verknüpfungen verwenden:

$$\forall m \in M : \ (f+g)(m) := f(m) +_R g(m), \quad (f \cdot g)(m) := f(m) \cdot_R g(m).$$

Beispiel 3.1.3 (Die Neunerprobe)

Sie kennen vermutlich die Regel, dass eine natürliche Zahl genau dann durch 9 teilbar ist, wenn dies für ihre Quersumme gilt. Wichtig ist dabei, dass wir die Quersumme im 10er-System nehmen, das heißt wir schreiben $a \in \mathbb{N}$ als

$$a = c_0 + c_1 \cdot 10 + c_2 \cdot 100 + \cdots + c_d \cdot 10^d.$$

Dann ist die Quersumme gerade $c_0 + c_1 + \cdots + c_d$. Nun gilt im Restklassenring $\mathbb{Z}/9\mathbb{Z}$ offensichtlich, dass $[10] = [1]$, und es folgt

$$[a] = [c_0] + [c_1] \cdot [1] + [c_2] \cdot [1]^2 + \cdots + [c_d] \cdot [1]^d = [c_0 + c_1 + \cdots + c_d],$$

oder anders gesagt: Die Restklasse von a ist die Restklase der Quersumme von a. Insbesondere ist die Restklasse von a genau dann 0, wenn die der Quersumme 0 ist. Aber die Restklasse ist genau dann 0, wenn a durch 9 teilbar ist.

Aufgabe 3.1.4 (Die Elferprobe)

Für die natürliche Zahl

$$a = c_0 + c_1 \cdot 10 + c_2 \cdot 100 + \cdots + c_d \cdot 10^d$$

definieren wir die alternierende Quersumme als

$$c_0 - c_1 + c_2 - c_3 \pm \cdots + (-1)^d c_d.$$

Zeigen Sie, dass a genau dann durch 11 teilbar ist, wenn die alternierende Quersumme von a durch 11 teilbar st.

Bemerkung 3.1.5 All die bisherigen Beispiele sind kommutative Ringe. Wir werden später (4.1.14) auch noch nicht kommutative Ringe kennenlernen. Eine echte Einschränkung, die wir uns hier sinnvoller Weise auferlegen, ist die dritte Forderung aus der Definition der Ringe, die Existenz eines Einselements. Ohne Forderung nach der Existenz eines Einselements wäre es viel schwieriger, interessante Aussagen über Ringe zu erhalten, und die Ringe in der linearen Algebra brauchen ohnehin ein Einselement. In manchen Quellen verzichtet man zunächst auf die Existenz einer Eins und nennt „unsere“ Ringe dann **Ringe mit Eins**.

Hilfssatz 3.1.6 (Zwei Rechenregeln)

Es sei R ein Ring. Dann gilt für alle $x, y \in R$:

$$0_R \cdot x = x \cdot 0_R = 0_R, \quad (-x) \cdot y = -(x \cdot y) = x \cdot (-y).$$

Beweis. Für die erste Gleichung überlegt man sich

$$0_R \cdot x = (0_R + 0_R) \cdot x = (0_R \cdot x) + (0_R \cdot x),$$

und einen dieser Summanden $0_R \cdot x$ darf man nun kürzen, da $(R, +)$ eine Gruppe ist. Dann bleibt aber gerade $0_R = 0_R \cdot x$ übrig. Analog geht das für $x \cdot 0_R$.

Die zweite Gleichung sagt, dass das Element $(-x) \cdot y$ zum Element $(x \cdot y)$ additiv invers ist. Das folgt aber aus

$$(-x) \cdot y + (x \cdot y) = (-x + x) \cdot y = 0_R \cdot y = 0_R.$$

Die letzte Gleichung beweist man analog. ◯

Aufgabe 3.1.7 Binomische Formeln

Es sei R ein kommutativer Ring. Für $x \in R$ setzen wir $x^2 = x \cdot x$ und $2x = x + x$. Zeigen Sie für beliebige $a, b \in R$ die Gleichungen

$$(a + b)^2 = a^2 + 2ab + b^2 \text{ und } (a + b) \cdot (a - b) = a^2 - b^2.$$

Wie lässt sich $(a + b)^3$ berechnen?

Wie für Gruppen, so interessiert man sich auch bei Ringen für Abbildungen, die die Verknüpfungen erhalten. Man nennt sie auch hier Homomorphismen.

Definition 3.1.8 (Ringhomomorphismus)

Ein **Homomorphismus zwischen zwei Ringen** $(R, +_R, \cdot_R)$ und $(S, +_S, \cdot_S)$ (kurz **Ringhomomorphismus**) ist eine Abbildung $\Phi : R \longrightarrow S$, sodass die folgenden Regeln gelten:

- $\forall x, y \in R : \Phi(x +_R y) = \Phi(x) +_S \Phi(y)$,
- $\forall x, y \in R : \Phi(x \cdot_R y) = \Phi(x) \cdot_S \Phi(y)$,
- $\Phi(1_R) = 1_S$.

Das Urbild $\Phi^{-1}(0_S)$ heißt der **Kern von** Φ. Da Φ ja insbesondere ein Homomorphismus von $(R,+_R)$ nach $(S,+_S)$ ist, ist der Kern von Φ also der Kern dieses Gruppenhomomorphismus, das ist eine Untergruppe von $(R,+_R)$.

Die Menge aller Homomorphismen von R nach S heißt $\mathrm{Hom}(R,S)$, etwas vorsichtiger müsste man sagen $\mathrm{Hom}_{\mathrm{Ring}}(R,S)$, wenn man betonen will, dass man nicht nur die additive Gruppe von R und S betrachtet. Einen Homomorphismus von R nach R nennt man wieder einen Endomorphismus von R, einen bijektiven Homomorphismus von R nach S nennt man einen Isomorphismus zwischen R und S, und einen Isomorphismus von R mit sich selbst (einen bijektiven Endomorphismus also) nennt man einen Automorphismus von R.

Beispiel 3.1.9 (ein paar Homomorphismen)

a) Ein Homomorphismus Φ von $\mathbb{Z}$ in einen beliebigen Ring S muss die Zahl 1 nach 1_S schicken, und es folgen

$$\Phi(2)=\Phi(1+1)=1_S+1_S=:2\cdot 1_S,\ \Phi(3)=\Phi(1+1+1)=1_S+1_S+1_S,\ \ldots$$

sowie

$$\Phi(0)=0_S,\ \Phi(-1)=-1_S,\ldots$$

Also sind alle Funktionswerte von Φ durch die Definition des Begriffs Ringhomomorphismus festgelegt. Das heißt, dass es höchstens einen Ringhomomorphismus von $\mathbb{Z}$ nach S geben kann. Das Distributivgesetz in S erzwingt, dass die eben beschriebene Abbildung von $\mathbb{Z}$ nach S tatsächlich ein Ringhomomorphismus ist.

Es gibt also (egal was für ein Ring S ist) genau einen Homomorphismus von $\mathbb{Z}$ nach S. Der Kern dieses Homomorphismus ist eine Teilmenge von $\mathbb{Z}$ von der Gestalt $c\cdot\mathbb{Z}$, $c\in\mathbb{N}_0$ geeignet. Diese Zahl c heißt die **Charakteristik** $\mathrm{char}(S)$ von S. Wenn $c>0$ gilt, so ist c die Ordnung (siehe 2.2.9) von 1_S in der Gruppe $(S,+)$. Für $c=0$ hat 1_S unendliche Ordnung.

Für alle $n\in\mathbb{N}_0$ etwa ist $n=\mathrm{char}(\mathbb{Z}/n\mathbb{Z})$.

b) Wenn R der Ring der reellen Cauchy-Folgen ist, dann ist die Abbildung

$$\Phi: R\longrightarrow\mathbb{R},\quad (a_i)\mapsto\lim_{i\to\infty}a_i$$

ein Ringhomomorphismus. Der Kern ist die Menge aller Nullfolgen.

Ähnlich funktioniert die „Auswertungsabbildung“ vom Ring der stetigen Funktionen auf $[0,1]$ nach $\mathbb{R}$, die f nach $f(\frac{\pi}{8})$ schickt. Der Kern ist die Menge aller stetigen Funktionen, die bei $\frac{\pi}{8}$ den Wert 0 annehmen. ($\frac{\pi}{8}$ ist hier natürlich eine willkürlich gewählte Zahl im betrachteten Intervall.)

c) Für jeden Ring R ist die Identität auf R ein Automorphismus von R.

d) Die Komposition von Homomorphismen ist wieder ein Homomorphismus. Die Inversen von Automorphismen sind wieder Automorphismen (was hiermit zum selbständigen Nachrechnen überlassen wird – siehe 2.3.8)

Wie bei den Gruppen (siehe 2.3.10) ist die Menge aller Automorphismen von R eine Untergruppe $\mathrm{Aut}(R) \subseteq \mathrm{Sym}(R)$.

Definition 3.1.10 (Einheiten, Einheitengruppe)

Es sei R ein Ring. Ein Element $x \in R$ heißt **invertierbar in** R oder auch eine **Einheit in** R, wenn ein Element $y \in R$ existiert, sodass $x \cdot y = y \cdot x = 1_R$.

Wie bei den Gruppen (siehe 2.1.4 c)) ist dieses y eindeutig bestimmt, und man schreibt dafür x^{-1}.

Die **Einheitengruppe** $R^\times$ ist die Menge aller Einheiten von R.

Sie wird durch die Multiplikation in R zu einer Gruppe. Zunächst ist klar, dass 1_R eine Einheit ist: $1_R \cdot 1_R = 1_R$. Dann gilt für alle $x_1, x_2 \in R^\times$:

$$(x_1x_2)\cdot(x_2^{-1}x_1^{-1}) = x_1\cdot(x_2x_2^{-1})\cdot x_1^{-1} = x_1\cdot 1_R\cdot x_1^{-1} = 1_R = \cdots = (x_2^{-1}x_1^{-1})\cdot(x_1x_2).$$

Also ist $x_1x_2 \in R^\times$, und die Multiplikation liefert eine Verknüpfung auf $R^\times$, die noch dazu assoziativ ist, da dies von der Multiplikation im Ring R ja gefordert wird. Schließlich ist mit x auch x^{-1} ein Element von $R^\times$, und damit sind alle Gruppenaxiome erfüllt.

Notation: Sind R ein kommutativer Ring, $r \in R^\times$ eine Einheit und $s \in R$ beliebig, so schreiben wir auch $\frac{s}{r}$ anstelle von $sr^{-1} = r^{-1}s$. Speziell ist $r^{-1} = \frac{1}{r}$.

Beispiel 3.1.11 (Einheiten in $\mathbb{Z}/n\mathbb{Z}$)

a) Die Einheitengruppe des Ringes $\mathbb{Z}$ der ganzen Zahlen besteht nur aus 1 und -1. Hingegen gilt

$$\mathbb{Q}^\times = \mathbb{Q} \smallsetminus \{0\}, \quad \mathbb{R}^\times = \mathbb{R} \smallsetminus \{0\}.$$

b) Für die natürliche Zahl n untersuchen wir nun die Einheitengruppe des Ringes $\mathbb{Z}/n\mathbb{Z}$. Für $a \in \mathbb{Z}$ ist das Element $[a]$ genau dann invertierbar, wenn es eine ganze Zahl b gibt, sodass $ab-1$ durch n teilbar ist. Das gilt genau dann, wenn es ganze Zahlen b, k gibt, sodass $ab - nk = 1$. Man sieht daran, dass der größte gemeinsame Teiler von a und n gleich 1 ist, wenn $[a] \in (\mathbb{Z}/n\mathbb{Z})^\times$.

[NB: Es gilt sogar: wenn $d = \mathrm{ggT}(a,n) \neq 1$, dann ist $a \cdot (n/d) \in n \cdot \mathbb{Z}$, also $[a]\cdot[n/d] = [0]$, obwohl $[a]$ und $[n/d]$ selbst nicht $[0]$ sind.]

Nehmen wir umgekehrt an, der größte gemeinsame Teiler von a und n sei eins. Die Ordnung von $[a]$ in der additiven Gruppe von $\mathbb{Z}/n\mathbb{Z}$ ist das minimale $k \in \mathbb{N}$, sodass n ein Teiler von ak ist. Wegen der Teilerfremdheit von a und n muss für dieses minimale k hier $k = n$ gelten, also wird $\mathbb{Z}/n\mathbb{Z}$ additiv von $[a]$ erzeugt,

und daher liegt auch $[1]$ in der von $[a]$ erzeugten Gruppe. Es gibt also ein $l \in \mathbb{N}$ mit $[l][a] = l[a] = [1]$.

Da $\mathbb{Z}/n\mathbb{Z}$ kommutativ ist, gilt auch $[a][l] = 1$, und a ist eine Einheit. Insgesamt haben wir gezeigt:

$$(\mathbb{Z}/n\mathbb{Z})^\times = \{[a] \mid \mathrm{ggT}(a, n) = 1\}.$$

Speziell sehen wir daran, dass im Falle, dass n eine Primzahl ist, jedes von $[0]$ verschiedene Element von $\mathbb{Z}/n\mathbb{Z}$ invertierbar ist: es gibt $n - 1$ Einheiten.

Wer nach 2.2.12 zurückblättert, kann dies dort auch schon sehen.

Satz 3.1.12 (Kleiner Satz von Fermat)

Es sei p eine Primzahl.

Dann gilt für jede ganze Zahl $a \in \mathbb{Z}$:

$$p \text{ teilt } a^p - a.$$

Beweis. Wenn a selbst ein Vielfaches von p ist, ist die Aussage klar. Wenn a kein Vielfaches von p ist, dann sind a und p teilerfremd, und nach dem eben gesehenen ist $[a] \in (\mathbb{Z}/p\mathbb{Z})^\times$ invertierbar. Da diese Einheitengruppe $p - 1$ Elemente hat (auch das haben wir eben festgehalten, wobei eben die Primzahl noch n hieß), sagt der Satz von Lagrange (2.2.11), dass die Ordnung e von $[a] \in (\mathbb{Z}/p\mathbb{Z})^\times$ ein Teiler von $p - 1$ ist: $\exists f \in \mathbb{Z} : \; p - 1 = e \cdot f$. Damit gilt

$$[a]^{p-1} = [a]^{e \cdot f} = ([a]^e)^f = [1]^f = [1].$$

Dann gilt aber auch

$$[a^p - a] = [a] \cdot ([a]^{p-1} - [1]) = [0],$$

und das heißt nach Definition unseres Ringes gerade, dass p ein Teiler von $a^p - a$ ist. ◯

Für $p = 2$ und $p = 3$ ist der Satz auch elementar leicht einzusehen, da

$$a^2 - a = a \cdot (a - 1) \quad \text{und} \quad a^3 - a = a \cdot (a - 1) \cdot (a + 1).$$

Dabei ist immer einer der Faktoren links durch 2 teilbar, und einer der Faktoren rechts durch 3 teilbar, weshalb $a^2 - a$ gerade und $a^3 - a$ durch 3 teilbar ist.

Bemerkung 3.1.13 (RSA-Kryptographie)

Der kleine Satz von Fermat ist einer der Dreh- und Angelpunkte für viele Anwendungen der Mathematik. In der Kryptographie wird er oft benutzt. Er liefert zum Beispiel ein Kriterium, mit dem man oft ausschließen kann, dass eine gegebene

Zahl eine Primzahl ist: 6 ist keine Primzahl, denn $5^6 - 5 = 15620$ ist nicht durch 6 teilbar. Das wird für größere Zahlen noch interessanter...

Die eigentlich relevante Information im Satz von Fermat ist genau betrachtet, dass für jede natürliche Zahl k die gegebene Primzahl p ein Teiler von $a^{k(p-1)+1} - a$ ist. Dies ist der Kerngedanke der RSA-Kryptographie. Hier wählt man sich zwei Primzahlen $p \neq q$ und setzt $N = pq$. Eine ganze Zahl ist genau dann durch N teilbar, wenn sie sowohl durch p als auch durch q teilbar ist. Nun sei $e \in \mathbb{N}$ zu $(p-1)(q-1)$ teilerfremd. Dann haben wir eben gelernt, dass es ein $f \in \mathbb{N}$ gibt sodass $(p-1)(q-1)$ ein Teiler von $ef - 1$ ist. Das heißt aber wegen Fermat: p und q teilen beide $a^{ef} - a = a^{k(p-1)(q-1)+1} - a$.

Benutzt wird das zur Verschlüsselung auf folgende Art: Ich wähle p, q, e wie oben, verrate davon aber nur $N = pq$ und e. Um mir verschlüsselte Information schicken zu lassen bitte ich nun alle, die das tun wollen, mir ihre Botschaft $a \in \mathbb{Z}/N\mathbb{Z}$ als a^e zu schicken. Da ich nun aber p und q kenne, kann ich f wie oben berechnen und damit a rekonsturieren. Insbesondere ist das Potenzieren mit e bijektiv, aber die Inverse ist für großes N ohne Zusatzinformation (hier die Pirmfaktorzerlegung) nicht gut berechenbar.

Die Sicherheit des RSA-Verfahrens steht und fällt also damit, dass es schwierig oder zumindest zeitaufwändig ist, große Zahlen in Primfaktoren zu zerlegen.

Falls irgendwann einmal hinreichend starke Quantencomputer verfügbar sein sollten, dann ist diese Prämisse nicht mehr gegeben. Daher gibt es bereits viele neue Ansätze, kryptographische Verfahren zu entwickeln, die gegen Angriffe mit Quantencomputern sicher sind.

Aufgabe 3.1.14 $N = 323,\ e = 5$

Wir setzen $p = 17,\ q = 19$, also $pq = 323$. Weiter sei $e = 5$. Finden Sie eine ganze Zahl f sodass $5f$ durch $16 \cdot 18$ teilbar ist. Verifizieren Sie, dass 323 ein Teiler von $2^{ef+1} - 2$ ist.

Nun wollen wir eine interessante Konsequenz aus der Existenz des Einselements in unseren Ringen ziehen.

Hilfssatz 3.1.15 (Ringhomomorphismen und Einheiten)

Es seien R und S zwei Ringe und $\Phi : R \longrightarrow S$ ein Ringhomomorphismus. Dann ist die Einschränkung Ψ von Φ auf die Einheitengruppe $R^\times$ ein Gruppenhomomorphismus

$$\Psi : R^\times \longrightarrow S^\times.$$

Beweis. Von Φ wird gefordert, dass $\Phi(1_R) = 1_S$. Daraus folgt für jede Einheit $x \in R^\times$ und ihr Inverses y:

$$\Phi(x) \cdot \Phi(y) = \Phi(xy) = \Phi(1_R) = 1_S = \Phi(yx) = \Phi(y) \cdot \Phi(x).$$

Also ist $\Phi(y)$ zu $\Phi(x)$ in S multiplikativ invers, und $\Phi(x)$ ist eine Einheit in S. Damit ist Ψ tatsächlich eine Abbildung von $R^\times$ nach $S^\times$. Dass diese dann ein Gruppenhomomorphismus ist, ist wegen der Multiplikativität von Φ klar. ◯

3.2 Körper

Definition 3.2.1 (Körper)

Ein **Körper** ist ein kommutativer Ring K, in dem $0_K \neq 1_K$ gilt und jedes von Null verschiedene Element invertierbar ist (siehe 3.1.10): $K^\times = K \setminus \{0\}$.

Beispiel 3.2.2 $\mathbb{Q}$ und $\mathbb{R}$ sind Körper, und auch der Ring $\mathbb{Z}/p\mathbb{Z}$ für eine Primzahl p (wegen Beispiel 3.1.11b)). In diesem letzten Fall schreiben wir $\mathbb{F}_p := \mathbb{Z}/p\mathbb{Z}$. Das F steht dabei für „Feld“. Manchmal liest man auch $GF(p) = \mathbb{Z}/p\mathbb{Z}$, das ist ein „Galois-Feld“ (zu Ehren des Franzosen Evariste Galois so bezeichnet).

Keine Körper sind zum Beispiel die Ringe $\mathbb{Z}$ oder $\mathbb{Z}/6\mathbb{Z}$. Auch die Polynomringe, die wir im nächsten Abschnitt kennenlernen werden, sind keine Körper.

Hilfssatz 3.2.3 *Es seien K ein Körper und R ein Ring, in dem $1_R \neq 0_R$ gilt. Dann ist jeder Ringhomomorphismus von K nach R injektiv.*

Beweis. Die Voraussetzung an R erzwingt, dass $0_R \notin R^\times$ gilt. Wegen Hilfssatz 3.1.15 haben wir aber $\Phi(K^\times) \subseteq R^\times$. Für den Körper K gilt $K^\times = K \setminus \{0\}$, also ist 0_K das einzige Element, das im Kern von Φ liegen kann. Da es dies auch tut, haben wir

$$\mathrm{Kern}(\Phi) = \{0_K\}.$$

Da Φ insbesondere ein Homomorphismus zwischen den additiven Gruppen von K und R ist, sagt 2.3.4, dass Φ injektiv sein muss. ◯

Beispiel 3.2.4 (für die Konstruktion eines größeren Körpers)

Nun sei K ein Körper und $d \in K$ ein Element, sodass für alle $a \in K$ die Ungleichung $d \neq a^2$ gilt. Als Beispiele denke man etwa an $K = \mathbb{R}$, $d = -1$, oder an $K = \mathbb{Q}$, $d = 2$, oder an $K = \mathbb{F}_5$, $d = [3]$.

Wir wollen einen Körper L konstruieren, der K als Teilring (3.1.1 c)) enthält und in dem es dann doch ein Element w gibt, für das $d = w^2$ gilt.

Nehmen wir zunächst einmal an, wir hätten so einen Körper L. Dann betrachten wir

$$K(w) := \{a + bw \mid a, b \in K\} \subseteq L.$$

Offensichtlich sind 0_K und 1_K in $K(w)$ enthalten. Für zwei Elemente $a+bw, \tilde{a}+\tilde{b}w \in K(w)$ gelten dann

$$\begin{aligned}(a+bw)+(\tilde{a}+\tilde{b}w) &= (a+\tilde{a})+(b+\tilde{b})w,\\ (a+bw)\cdot(\tilde{a}+\tilde{b}w) &= a\tilde{a}+a\tilde{b}w+\tilde{a}bw+b\tilde{b}w^2 = (a\tilde{a}+b\tilde{b}d)+(a\tilde{b}+\tilde{a}b)w,\end{aligned}$$

und diese Elemente liegen wieder in $K(w)$. Also ist $K(w)$ unter der Bildung von Summen und Produkten abgeschlossen. Da mit $a+bw$ auch $-(a+bw)$ zu $K(w)$ gehört, sieht man damit sogar, dass $K(w)$ ein Teilring von L ist.

Wenn a und b nicht beide 0 sind, dann ist auch $a+bw \neq 0$, da sonst $w = -ab^{-1} \in K$, was ausgeschlossen war. Dann gilt aber

$$(a+bw)\cdot(a-bw) = a^2-b^2d \in K^\times,$$

denn aus $a^2-b^2d=0$ folgte, dass $d=(a/b)^2$ in K als Quadrat geschrieben werden könnte (was wir ja explizit ausgeschlossen hatten).

Damit ist aber $a+bw$ in $K(w)$ invertierbar:

$$(a+bw)^{-1} = (a^2-b^2d)^{-1}\cdot(a-bw).$$

All dies hatten wir unter der Annahme gemacht, dass der größere Körper L existiert, in dem sich w findet. Nun drehen wir den Spieß um und konstruieren mit den Rechenregeln aus $K(w)$ einen passenden Körper L.

Ansatz: Wir setzen $L := K\times K = \{(a,b) \mid a,b\in K\}$ und definieren auf L die Addition und Multiplikation so, wie das von den Überlegungen zu $K(w)$ nahegelegt wird:

$$\begin{aligned}(a,b)+(\tilde{a},\tilde{b}) &:= (a+\tilde{a}, b+\tilde{b}),\\ (a,b)\cdot(\tilde{a},\tilde{b}) &:= (a\tilde{a}+b\tilde{b}d, a\tilde{b}+b\tilde{a}).\end{aligned}$$

Es ist dabei klar, dass $(L,+)$ eine abelsche Gruppe ist. Die Multiplikation ist assoziativ, denn für alle $a,b,e,f,k,l\in K$ gilt

$$\begin{aligned}[(a,b)\cdot(e,f)]\cdot(k,l) &= (ae+bfd, af+be)\cdot(k,l)\\ &= (aek+bfdk+afld+beld, ael+bfdl+afk+bek)\\ &= (a,b)\cdot(ek+fld, fk+el)\\ &= (a,b)\cdot[(e,f)\cdot(k,l)].\end{aligned}$$

Die Multiplikation auf L ist kommutativ, und es gilt das Distributivgesetz: für alle $a,b,e,f,k,l\in K$ haben wir ja

$$\begin{aligned}[(a,b)+(e,f)]\cdot(k,l) &= (a+e,b+f)\cdot(k,l)\\ &= ((a+e)k+(b+f)ld, (a+e)l+(b+f)k)\\ &= (ak+bld, al+bk)+(ek+fld, el+fk)\\ &= [(a,b)\cdot(k,l)]+[(e,f)\cdot(k,l)].\end{aligned}$$

Schließlich gibt es ein neutrales Element bezüglich der Multiplikation, nämlich $(1,0)$, und jedes Element $(a,b) \neq (0,0)$ ist bezüglich der Multiplikation invertierbar, die Inverse ist $(\frac{a}{a^2-db^2}, \frac{-b}{a^2-db^2})$.

Wir erhalten insgesamt das folgende Ergebnis.

Satz 3.2.5 *Die Menge L mit der eben angegebenen Addition und Multiplikation ist ein Körper, der K als Teilring enthält.*

Beweis. Nur die Aussage über den Teilring muss noch begründet werden. Die Abbildung

$$K \longrightarrow L, \quad x \mapsto (x,0)$$

ist ein (injektiver) Ringhomomorphismus und man identifiziert K mit seinem Bild in L. ◯

Aufgabe 3.2.6 (Mehr endliche Körper)

a) Im Körper $K = \mathbb{F}_7$, 3.2.1, schreiben wir die Restklasse von $x \in \mathbb{Z}$ als x_K. In K gibt es die Quadrate

$$0_K = 0_K^2,\ 1_K = 1_K^2 = (-1_K)^2,\ 2_K = 3_K^2 = (-3_K)^2,\ 4_K = (2_K^2 = (-2_K)^2.$$

Wir konstruieren wie eben einen Körper, der K enthält und Element w mit der Eigenschaft $w^2 = -1_K$.

Wie viele Elemente enthält dieser neue Körper?

b) Gibt es einen Körper mit 25 Elementen?

Bemerkung 3.2.7 (Komplexe Zahlen)

Ein besonders wichtiger Spezialfall ist der Fall, dass es sich bei K um den Körper der reellen Zahlen handelt, und $d = -1$. Dann heißt der konstruierte größere Körper der Körper $\mathbb{C}$ der **komplexen Zahlen**, und statt w (wie oben) schreibt man üblicherweise i für ein Element aus $\mathbb{C}$, dessen Quadrat -1 ergibt. Wir fassen das zusammen:

Es ist $\mathbb{C} = \mathbb{R}(\mathrm{i}) = \{a + b\mathrm{i} \mid a, b \in \mathbb{R}\}$, wobei
$(a + b\mathrm{i}) + (c + d\mathrm{i}) = a + c + (b + d)\mathrm{i}$ und
$(a + b\mathrm{i}) \cdot (c + d\mathrm{i}) = ac - bd + (ad + bc)\mathrm{i}$.

Die reelle Zahl a heißt der **Realteil**, b der **Imaginärteil** von $a + b\mathrm{i}$. Zwei komplexe Zahlen sind genau dann gleich, wenn sie denselben Real- und denselben Imaginärteil haben. Das benutzen wir jetzt, um zu zeigen, dass in $\mathbb{C}$ selbst jedes Element ein Quadrat ist.

Es sei $a + b\mathrm{i} \in \mathbb{C}$ gegeben. Gesucht ist ein $x + y\mathrm{i} \in \mathbb{C}$, sodass

$$a + b\mathrm{i} = (x + y\mathrm{i})^2 = x^2 - y^2 + 2xy\mathrm{i}.$$

Hierbei müssen die Real- und Imaginärteile übereinstimmen, also

$$a = x^2 - y^2, \quad b = 2xy.$$

Im Fall $b = 0$ ist die Lösung einfach: Im Fall $a \geq 0$ setze man dann $x = \sqrt{a}$, $y = 0$, im Fall $a < 0$ setze man $x = 0, y = \sqrt{-a}$ und rechne in beiden Fällen das Gewünschte nach.

Nur der Fall $b \neq 0$ ist also wirklich interessant. Hier versuchen wir unser Glück mit dem Ansatz $y = b/(2x)$ und müssen dann

$$x^2 - \frac{b^2}{4x^2} = a$$

lösen. Multiplikation mit x^2 liefert

$$x^4 - ax^2 - b^2/4 = 0,$$

und die Lösungsformel für quadratische Gleichungen sagt uns, dass

$$x^2 = \frac{a \pm \sqrt{a^2 + b^2}}{2}.$$

Da $\frac{a+\sqrt{a^2+b^2}}{2}$ positiv ist, gibt es eine reelle Zahl $x > 0$, die diese Gleichung löst. Man rechnet nach, dass $x + \frac{b}{2x}\mathrm{i}$ dann eine Quadratwurzel von $a + b \cdot \mathrm{i}$ ist.

Für die Zwecke der Multiplikation hat es sich als hilfreich erwiesen, eine komplexe Zahl $z = a + b\mathrm{i}$ zu schreiben als

$$z = r \cdot (\cos(\alpha) + \sin(\alpha) \cdot \mathrm{i}), \quad r := \sqrt{a^2 + b^2}.$$

Dabei heißt die reelle Zahl $r \geq 0$ der **Betrag** von z und wir schreiben in Zukunft oft $r = |z|$. Der Winkel α muss (und kann – das verspricht uns die Analysis) passend gewählt werden, sodass die Gleichung stimmt. Die Zahlen r und α heißen die Polarkoordinaten von z.

Ist $u = c + d\mathrm{i} = s \cdot (\cos(\beta) + \sin(\beta)\mathrm{i})$ eine weitere komplexe Zahl, so folgt

$$\begin{aligned} z \cdot u &= r \cdot s \cdot (\cos(\alpha) + \sin(\alpha) \cdot \mathrm{i}) \cdot (\cos(\beta) + \sin(\beta)\mathrm{i}) = \\ &= rs(\cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta) + (\cos(\alpha)\sin(\beta) + \sin(\alpha)\cos(\beta))\mathrm{i}) = \\ &= rs(\cos(\alpha + \beta) + \sin(\alpha + \beta)\mathrm{i}). \end{aligned}$$

Dabei leihen wir uns aus der Analysis das Additionstheorem für Sinus und Cosinus. Wir sehen daran für festes komplexes $u \neq 0$: Die Abbildung

$$\mu : \mathbb{C} \longrightarrow \mathbb{C}, \quad \mu(z) := u \cdot z$$

ist geometrisch gesehen eine Drehstreckung, denn der Betrag wird mit s multipliziert und der Winkel ändert sich um die Addition von β.

Jetzt finden wir auch noch viel schneller eine komplexe Quadratwurzel von z:

$$z = r \cdot (\cos(\alpha) + \sin(\alpha) \cdot \mathrm{i}) = [\sqrt{r} \cdot (\cos(\alpha/2) + \sin(\alpha/2) \cdot \mathrm{i})]^2.$$

Wir werden später noch Verwendung für die folgende Abbildung

$$^- : \mathbb{C} \to \mathbb{C}, \quad \overline{x + y\mathrm{i}} := x - y\mathrm{i},$$

haben. Diese heißt die **komplexe Konjugation** und ist ein Ringautomorphismus von $\mathbb{C}$, was sich leicht nachrechnen lässt.

Aufgabe 3.2.8 Einheitswurzeln

Sei $n \in \mathbb{N}$ gegeben. Weisen Sie nach, dass die Zahlen

$$z_k := \cos(2\pi k/n) + \sin(2\pi k/n)\mathrm{i}, \ 0 \leq k \leq n-1,$$

die einzigen komplexen Zahlen sind, die die Gleichung

$$z^n = 1$$

lösen.

3.3 Polynomringe

In diesem Abschnitt sei R ein kommutativer Ring. Das intuitive Verständnis dessen, was ein Polynom sein soll, nämlich ein Ausdruck der Gestalt

$$\sum_{i=0}^{d} a_i X^i,$$

legt nahe, wie man Polynome (unter Wahrung der Assoziativ- und Distributivgesetze) addiert und multipliziert. Dies wollen wir jetzt formalisieren, indem wir uns nur die Koeffizienten $(a_0, \ldots, a_d)$ merken. Da wir für verschiedene Polynome verschiedene Obergrenzen d werden verwenden müssen, benutzen wir unendliche Folgen, bei denen die Folgenglieder ab einer gewissen vom individuellen Polynom abhängigen Grenze alle 0 werden.

Definition 3.3.1 (Polynome, Polynomring)

Ein **Polynom** über R ist eine Folge $(a_i)_{i\in\mathbb{N}_0}$ mit Einträgen aus R, sodass eine natürliche Zahl N existiert, für die die Bedingung

$$\forall i \geq N : a_i = 0$$

erfüllt ist. Diese letzte Bedingung nennen wir die „Abbruchsbedingung“.

Mit $R[X]$ bezeichnen wir die Menge aller Polynome (mit Koeffizienten in R). $R[X]$ heißt der **Polynomring** über R in der Veränderlichen X. Auf $R[X]$ gibt es zwei Verknüpfungen. Dazu seien (a_i) und (b_i) zwei Polynome in $R[X]$. Dann setzen wir

$$\begin{aligned}(a_i)_{i\in\mathbb{N}_0} + (b_i)_{i\in\mathbb{N}_0} &:= (a_i + b_i)_{i\in\mathbb{N}_0}\\ (a_i)_{i\in\mathbb{N}_0} \cdot (b_i)_{i\in\mathbb{N}_0} &:= (c_k)_{k\in\mathbb{N}_0},\\ \text{wobei}\quad c_k &:= \textstyle\sum_{i=0}^{k} a_i b_{k-i}.\end{aligned}$$

Diese Definitionen sind motiviert durch die naheliegenden Formeln, wie die Ausdrücke $\sum a_i X^i$ und $\sum b_j X^j$ zu addieren beziehungsweise zu multiplizieren sind. Wohlgemerkt wissen wir noch nicht, was X ist, und deswegen gehen wir erst diesen sehr formalen Weg.

Bemerkung 3.3.2 Bei der Summe ist wirklich klar, dass es sich wieder um ein Polynom handelt. Beim Produkt muss man sich überlegen, dass die Abbruchsbedingung erfüllt ist. Wenn zwei natürliche Zahlen M, N gefunden sind, sodass für $i \geq N$ die Bedingung $a_i = 0$ und für $j \geq M$ die Bedingung $b_j = 0$ gilt, dann folgt für $k \geq M + N$:

$$c_k = \sum_{i=0}^{k} a_i b_{k-i} = \sum_{i=0}^{N-1} a_i b_{k-i} = 0,$$

denn für $i \geq N$ ist ja $a_i = 0$ und daher der Summand unerheblich, und für $i < N$ ist wegen $k - i \geq M + N - i \geq M$ der Faktor b_{k-i} Null.

Hilfssatz 3.3.3 (der Polynomring ist ein Ring)

Es sei R ein kommutativer Ring. Dann gelten:

a) Die Menge $R[X]$ wird mit den eben definierten Verknüpfungen zu einem kommutativen Ring. Das Einselement ist die Folge $(1, 0, 0, \dots)$, bei der alle Einträge ab dem ersten gleich 0 sind.

b) Der Ring R kann als Teilring von $R[X]$ aufgefasst werden, indem man $r \in R$ mit der Folge $(r, 0, 0, \dots)$ “identifiziert”.

Beweis. a) Dass $(R[X], +)$ eine abelsche Gruppe ist, ist evident. Stellvertretend für die restlichen Ringaxiome rechnen wir die Assoziativität der Multiplikation

nach. Es seien $f = (a_i),\ g = (b_i),\ h = (c_i)$ Polynome mit Koeffizienten in R. Dann ist

$$\begin{aligned} f \cdot (g \cdot h) &= f \cdot (\textstyle\sum_{i \le k} b_i \cdot c_{k-i})_k \\ &= (\textstyle\sum_{j \le l} a_j \cdot (\sum_{i \le l-j} b_i \cdot c_{l-j-i}))_l \\ &= (\textstyle\sum_{i \le l} (\sum_{j \le i} a_j \cdot b_{i-j}) \cdot c_{l-i})_l \\ &= (\textstyle\sum_{j \le i} a_j \cdot b_{i-j})_i \cdot h \\ &= (f \cdot g) \cdot h. \end{aligned}$$

Die übrigen Axiome lassen sich ähnlich verifizieren.

b) Die Abbildung $R \ni r \mapsto (r, 0, 0, \dots) \in R[X]$ ist ein injektiver Ringhomomorphismus, liefert also einen Ringisomorphismus von R mit einem Teilring von $R[X]$. ◯

Insbesondere können wir jetzt Polynome mit Elementen aus R multiplizieren, indem wir die Elemente aus R auch als Polynome auffassen: sogenannte „konstante Polynome“.

Jetzt wollen wir sehen, dass diese sehr formale Konstruktion eines Rings tatsächlich etwas liefert, was unserem naiven Ausgangsverständnis der Polynome entspricht.

Dazu wählen wir ein spezielles Polynom und nennen es X, nämlich:

$$\begin{aligned} X &:= (0, 1, 0, 0, 0, 0 \dots) \\ &= (a_i)_{i \in \mathbb{N}_0}, \quad \text{wobei } a_1 = 1 \text{ und } \forall i \neq 1 :\ a_i = 0 \end{aligned}$$

Diese Definition liegt nahe, wenn wir uns daran erinnern, was wir eingangs als Motivation für die formale Konstruktion gesagt hatten.

Dann gilt für ein beliebiges Polynom (b_i):

$$\begin{aligned} X \cdot (b_i) &= (c_k)_{k \in \mathbb{N}_0}, \quad c_0 = 0,\ \forall k \ge 1 : c_k = b_{k-1}, \\ &= (0, b_0, b_1, b_2, \dots). \end{aligned}$$

Jetzt setzen wir $X^0 := 1$ und rekursiv $X^{i+1} := X \cdot X^i$, dann ist für die natürliche Zahl i das Polynom X^i die Folge, die als i-ten Eintrag 1 enthält und ansonsten nur Nullen. Damit ist

$$(b_i)_{i \in \mathbb{N}_0} = \sum_{i=0}^{\infty} b_i \cdot X^i,$$

wobei die Summe in Wirklichkeit endlich ist, da ja fast alle b_i Null sind. Wir können jetzt endlich den Polynomring so schreiben, wie wir das von vorneherein tun wollten:

$$\boxed{\begin{array}{c} R[X] = \{\textstyle\sum_{i=0}^{d} r_i X^i \mid d \in \mathbb{N},\ r_0, \dots, r_d \in R\}. \\ R \subseteq R[X] \text{ via } R \ni r \mapsto rX^0 \in R[X]. \end{array}}$$

Definition 3.3.4 (Grad eines Polynoms, Leitkoeffizient)
Der **Grad des Polynoms** $f = \sum_{i=0}^{d} r_i X^i \in R[X]$ ist definiert als

$$\mathrm{Grad}(f) := \begin{cases} \max(\{i \in \mathbb{N}_0 \mid r_i \neq 0\}), & f \neq 0, \\ -\infty, & f = 0. \end{cases}$$

Dabei ist $-\infty$ ein Symbol für ein Element außerhalb der natürlichen Zahlen. Wir vereinbaren jetzt, dass für $a \in \mathbb{N}_0 \cup \{-\infty\}$ gelten soll:

$$\max(a, -\infty) = a, \quad a + (-\infty) = (-\infty) + a = -\infty.$$

Diese Konvention brauchen wir für den nächsten Hilfssatz.

Für $f \neq 0$ heißt der Koeffizient $r_{\mathrm{Grad}(f)}$ der **Leitkoeffizient** von f.

Hilfssatz 3.3.5 (Regeln für das Rechnen mit dem Grad) *Es seien $f, g \in R[X]$ Polynome. Dann gelten die folgenden Regeln für die Grade:*

- $\mathrm{Grad}(f+g) \leq \max(\mathrm{Grad}(f), \mathrm{Grad}(g))$.
- $\mathrm{Grad}(f \cdot g) \leq \mathrm{Grad}(f) + \mathrm{Grad}(g)$.
- $\mathrm{Grad}(f \cdot g) = \mathrm{Grad}(f) + \mathrm{Grad}(g)$, *falls R die folgende Eigenschaft hat:*

$$\forall a, b \in R \setminus \{0\} : a \cdot b \neq 0.$$

Beweis. Es sei $m := \max(\mathrm{Grad}(f), \mathrm{Grad}(g))$. Dann lassen sich f und g schreiben als

$$f = \sum_{i=0}^{m} r_i X^i, \quad g = \sum_{i=0}^{m} s_i X^i,$$

und damit ist

$$f + g = \sum_{i=0}^{m} (r_i + s_i) X^i,$$

und man braucht keinen Summationsindex größer als m. Das zeigt die erste Ungleichung.

Nun seien $d = \mathrm{Grad}(f)$, $e = \mathrm{Grad}(g)$. Weiter schreiben wir

$$f = \sum_{i=0}^{d} r_i X^i, \quad g = \sum_{i=0}^{e} s_i X^i,$$

wobei r_d und s_e beide nicht Null sind. Dann ist

$$f \cdot g = \sum_{k=0}^{d+e} \left(\sum_{i=0}^{k} r_i s_{k-i} \right) X^k,$$

und das zeigt, dass $\mathrm{Grad}(f \cdot g) \leq d + e$.

Der Koeffizient, der in fg vor X^{d+e} steht, ist $r_d s_e$. Unter der zusätzlich gemachten Voraussetzung an den Ring ist dieses Produkt nicht 0, da r_d und s_e nicht Null sind. ◯

Definition 3.3.6 (Nullteilerfreiheit)

Einen Ring $R \neq \{0_R\}$ mit der Eigenschaft $\forall a, b \in R \setminus \{0_R\} : a \cdot b \neq 0$ nennt man einen **nullteilerfreien Ring.**

Wir haben zum Beispiel in 3.1.11 b) schon gesehen, dass für $n \in \mathbb{N}$ der Ring $\mathbb{Z}/n\mathbb{Z}$ genau dann nullteilerfrei ist, wenn n eine Primzahl ist.

Körper sind immer nullteilerfrei.

Für einen nullteilerfreien Ring R gilt $(R[X])^\times = R^\times$. Denn eine Einheit im Polynomring muss ja Grad 0 haben.

Sicher werden viele Leser und Leserinnen denken, dass Polynome eigentlich Abbildungen seien. Diesen „grundfalschen“ Gedanken wollen wir jetzt wenigstens zur Hälfte legitimieren. Eigentlich sollte man sich ein Polynom als einen Platzhalter für eine ganze Reihe von Abbildungen vorstellen, die „Form“ einer Abbildung, die man auf ganz verschiedenen Bereichen definieren kann. Das wird jetzt präzisiert, wenn auch nicht in der allgemeinst möglichen aller Situationen.

Definition/Bemerkung 3.3.7 (Potenzen, Zentrum, Einsetzabbildung)

Es sei A ein Ring.

a) Für $a \in A$ setzen wir $a^0 := 1$ und rekursiv für $i \in \mathbb{N}_0 :\quad a^{i+1} := a \cdot a^i$.
Diese a^i heißen die **Potenzen** von a.

b) Die Menge
$$Z(A) := \{a \in A \mid \forall x \in A : a \cdot x = x \cdot a\}$$
heißt das **Zentrum** von A. Das Zentrum ist ein kommutativer Teilring von A. Ist A kommutativ, so gilt $Z(A) = A$.

c) Es sei R ein Teilring von $Z(A)$. Dann ist für $f = \sum_{i=0}^{d} r_i X^i \in R[X]$ die Abbildung
$$\tilde{f} : A \longrightarrow A, \quad \tilde{f}(a) := \sum_{i=0}^{d} r_i a^i,$$
definiert. Die Zuordnung $f \mapsto \tilde{f}$ ist ein Ringhomomorphismus von $R[X]$ in den Ring der Abbildungen von A nach A (vgl. 3.1.2 d)). Für $\tilde{f} \cdot \tilde{g} = \widetilde{f \cdot g}$ brauchen wir gerade die Voraussetzung, dass R im Zentrum von A liegt.

Ist A ein endlicher Ring, so ist auch $\mathrm{Abb}(A,A)$ endlich, während der Polynomring $R[X]$ für $R \neq \{0\}$ unendlich ist: ein Polynom f lässt sich also im Allgemeinen nicht aus der ihm zugeordneten Abbildung $\tilde{f}$ rekonstruieren.

Notation: Wir werden oft anstelle von $\tilde{f}(a)$ auch $f(a)$ schreiben. Insbesondere gilt in diesem Sinne im Fall $A = R[X]$, $a = X$ die zunächst bizarr anmutende Gleichheit $f(X) = f$.

d) Es sei R ein Teilring von $Z(A)$. Für jedes $a \in A$ ist dann

$$E_a : R[X] \longrightarrow A, \quad f \mapsto E_a(f) := f(a),$$

die **Einsetzabbildung bei** a. (Man nennt diese auch die Auswertungsabbildung.) E_a ist ein Ringhomomorphismus, wie man leicht nachrechnet. Es gilt $E_a(1) = 1$, da $a^0 = 1$ gesetzt wurde. Weiter gilt für Polynome $f = \sum_{i=0}^m r_i X^i$, $g = \sum_{i=0}^m s_i X^i$:

$$\begin{aligned} E_a(f+g) &= \textstyle\sum_{i=0}^m (r_i + s_i)a^i = \sum_{i=0}^m r_i a^i + \sum_{i=0}^m s_i a^i \\ &= E_a(f) + E_a(g). \\ E_a(f \cdot g) &= \textstyle\sum_{k=0}^{2m} \sum_{i=0}^k (r_i \cdot s_{k-i})a^k = \sum_{i=0}^m r_i a^i \cdot \sum_{i=0}^m s_i a^i \\ &= E_a(f) \cdot E_a(g). \end{aligned}$$

Dabei benutzt man wieder für die Multiplikativität, dass R im Zentrum von A liegt, denn man braucht $a^i s_{k-i} = s_{k-i} a^i$ beim Umsortieren.

Das Bild von E_a wird meistens mit $R[a]$ bezeichnet. Es ist

$$R[a] = \{\sum_{i=0}^d r_i a^i \mid d \in \mathbb{N},\ r_0, \ldots, r_d \in R\}.$$

Dies ist ein kommutativer Teilring von A, und zwar der kleinste Teilring, der R und a enthält.

Beispiel 3.3.8 a) Es sei $R = \mathbb{Z}$ und $A = \mathbb{Q}$, sowie $a = \frac{1}{2}$. Dann ist

$$\mathbb{Z}[\frac{1}{2}] = \{\sum_{i=0}^d r_i 2^{-i} \mid d \in \mathbb{N},\ r_0, \ldots, r_d \in R\} = \{\frac{a}{2^d} \mid a \in \mathbb{Z}, d \in \mathbb{N}\},$$

wie man leicht nachrechnet; dazu muss man alle Summanden in $\sum_{i=0}^d r_i 2^{-i}$ auf einen Hauptnenner bringen, der offensichtlich gleich 2^d gewählt werden kann.

b) Es sei $R = \mathbb{R}$ und A der Ring der stetigen reellwertigen Funktionen auf $\mathbb{R}$. Weiter sei a die Exponentialabbildung. Dann ist $\mathbb{R}[a]$ die Menge aller Funktionen der Gestalt

$$x \mapsto \sum_{k=0}^d r_k e^{kx}.$$

Der Ring $\mathbb{R}[a]$ ist in diesem Fall zum Polynomring isomorph, auch wenn die zugehörigen Abbildungen von $\mathbb{R}$ nach $\mathbb{R}$ alles andere als Polynome sind!

Definition 3.3.9 (Teilbarkeit im Polynomring)

Es seien R ein kommutativer Ring und $f, g \in R[X]$ Polynome über R. Dann heißt g ein **Teiler** von f, wenn es ein Polynom $h \in R[X]$ gibt, sodass

$$f = g \cdot h.$$

Beispiel 3.3.10 Lineare Teiler

Für $r \in R$ ist das Polynom $g := X - r$ genau dann ein Teiler von $f \in R[X]$, wenn $f(r) = 0$.

Denn: Wenn für ein Polynom h die Gleichheit $f = (X - r) \cdot h$ gilt, dann gilt auch

$$f(r) = (r - r) \cdot h(r) = 0,$$

denn E_r ist ein Ringhomomorphismus.

Umgekehrt sei $f(r) = 0$. Wir setzen $\hat{f}(X) := f(X + r)$, dann ist $\hat{f}(0) = 0$. Wir schreiben $\hat{f}$ als

$$\hat{f}(X) = \sum_{i=0}^{d} a_i X^i,$$

und sehen wegen $f(r) = \hat{f}(0) = a_0$, dass $a_0 = 0$ und damit

$$\hat{f}(X) = X \cdot \sum_{i=1}^{d} a_i X^{i-1} = X \cdot h(X).$$

Dabei ist $h(X) := \sum_{i=1}^{d} a_i X^{i-1} \in R[X]$.

Es folgt

$$f = f(X) = \hat{f}(X - r) = (X - r) \cdot h(X - r),$$

denn die Auswertungsabbildung E_{X-r} ist ein Ringhomomorphismus. Und natürlich ist $h(X - r)$ ein Polynom: Es ist das Bild von h bei der Einsetzabbildung $E_{X-r} : R[X] \to R[X]$.

Aufgabe 3.3.11 (Die Anzahl der Nullstellen)

Sei R ein nullteilerfreier kommutativer Ring und $f \in R[X]$ ein Polynom vom Grad $d \geq 0$. Zeigen Sie mit vollständiger Induktion nach d, dass f höchstens d Nullstellen in R haben kann.

Bemerkung 3.3.12 Gerade mit der Anzahl der Nullstellen eines Polynoms muss man bei nicht gegebener Nullteilerfreiheit aufpassen. Zum Beispiel hat $X^2 - 1$ in $\mathbb{Z}/8\mathbb{Z}$ sogar 4 Nullstellen, nämlich $[1], [3], [5]$ und $[7]$. Das ist nichts anderes als die Aussage, dass das Quadrat einer ungeraden ganzen Zahl bei Division durch 8 immer Rest 1 lässt. Ist nämlich $k = 2l + 1$ mit $l \in \mathbb{Z}$, so gilt

$$k^2 = 4(l^2 + l) + 1,$$

aber $l^2 + l$ ist immer gerade, denn l^2 und l haben dieselbe Parität.

Im Zusammenhang mit linearen Gleichungssystemen werden wir in Kürze auch nichtkommutative Ringe kennenlernen, für die wir die Einsetzabbildungen oft benutzen werden.

Kapitel 4

Lineare Gleichungssysteme und Matrizen

Nun nähern wir uns einer der zentralen Fragestellungen der linearen Algebra, nämlich der Frage nach der Lösbarkeit linearer Gleichungssysteme und nach der „Struktur“ ihrer Lösungsmengen. Eine bequeme Notation hierfür werden wir in Form der Matrizenschreibweise zur Verfügung stellen, wobei der Mechanismus der Ringtheorie benutzt wird. Wir formulieren die grundlegende Frage erst in größerer Allgemeinheit als wir sie später lösen werden. Dazu sei R in diesem Kapitel stets ein kommutativer Ring.

4.1 Lineare Gleichungssysteme – Grundlegendes

Definition 4.1.1 (Lineares Gleichungssystem)

Es sei R ein kommutativer Ring. Ein **Lineares Gleichungssystem über** R (kurz oft auch LGS genannt) mit p Gleichungen und q Unbekannten ist ein System

$$\begin{array}{ccccccc} a_{11}x_1 & + a_{12}x_2 & + \cdots + & a_{1q}x_q & = b_1 \\ a_{21}x_1 & + a_{22}x_2 & + \cdots + & a_{2q}x_q & = b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{p1}x_1 & + a_{p2}x_2 & + \cdots + & a_{pq}x_q & = b_p \end{array} \qquad (*)$$

wobei die **Koeffizienten** $a_{ij}, 1 \leq i \leq p,\ 1 \leq j \leq q$ und auch die $b_i,\ 1 \leq i \leq p$, in R liegen und Lösungstupel $(x_j)_{1 \leq j \leq q} \in R^q$ gesucht sind.

Die Menge aller Lösungen des Systems $(*)$ bezeichnen wir mit $\mathcal{L}(*)$.

Das System, das aus $(*)$ durch die Ersetzung der rechten Seite durch 0 entsteht, heißt das zu $(*)$ gehörende **homogene Gleichungssystem**.

Statt $(*)$ schreiben wir kürzer

$$\sum_{j=1}^{q} a_{ij}x_j = b_i, \quad 1 \leq i \leq p.$$

Nun halten wir fest, dass R^q (die Menge aller q-Tupel in R; siehe 1.2.2 e)) mit komponentenweiser Addition eine Gruppe bildet. Das Nullelement ist das Tupel, dessen Einträge allesamt 0 sind. Das (additiv) inverse Element zu $(r_1\ r_2\ \ldots\ r_q)^\top$ ist $(-r_1\ -r_2\ \ldots\ -r_q)^\top$.

Hilfssatz 4.1.2 (Das LGS aus der Sichtweise der Gruppentheorie)
Mit der eben eingeführten Notation ist die Abbildung

$$\Phi : R^q \longrightarrow R^p,\ \Phi((x_j)_{1\leq j\leq q}) = (\sum_{j=1}^{q} a_{ij}x_j)_{1\leq i\leq p}$$

ein Gruppenhomomorphismus. Statt $\mathcal{L}(*)$ schreiben wir dann auch $\mathcal{L}(\Phi, b)$.

Der Lösungsraum des zugehörigen homogenen Gleichungssystems ist dann gerade der Kern von Φ. Wenn $\mathcal{L}(\Phi, b)$ nicht leer ist, so gilt für jede beliebige „spezielle Lösung“ $x^{(s)}$ von $(*)$ die Aussage

$$\mathcal{L}(\Phi, b) = \{x^{(h)} + x^{(s)} \mid x^{(h)} \in \mathrm{Kern}(\Phi)\},$$

Beweis. Dass Φ ein Gruppenhomomorphismus ist, folgt aus dem Distributivgesetz im Ring R.

Die Aussage über den Kern ist einfach die Definition desselben (siehe 2.3.4).

Es bleibt noch die behauptete Mengengleichheit zu zeigen; dazu sei $x^{(s)} \in \mathcal{L}(\Phi, b)$ fest gewählt.

„$\subseteq$“: Sei $x \in \mathcal{L}(\Phi, b)$. Dann gilt $x = (x - x^{(s)}) + x^{(s)}$ und

$$\Phi(x - x^{(s)}) = \Phi(x) - \Phi(x^{(s)}) = b - b = 0,$$

also $x^{(h)} := x - x^{(s)} \in \mathrm{Kern}(\Phi)$.

„$\supseteq$“: Sei $x^{(h)} \in \mathrm{Kern}(\Phi)$ beliebig und $x := x^{(s)} + x^{(h)}$. Dann gilt

$$\Phi(x) = \Phi(x^{(s)} + x^{(h)}) = \Phi(x^{(s)}) + \Phi(x^{(h)}) = b + 0 = b,$$

also $x \in \mathcal{L}(\Phi, b)$. ◯

Beispiel 4.1.3 Wir lösen nun ein Lineares Gleichungssystem über den reellen Zahlen. Das Gleichungssystem sei gegeben durch

$$\begin{array}{llll} 1\cdot x & +2\cdot y & +3\cdot z & =10 \\ 4\cdot x & +5\cdot y & +6\cdot z & =20 \end{array}$$

Wie in der Schule ziehen wir das Vierfache der ersten Zeile von der zweiten ab und teilen dann die zweite Zeile durch -3. Dann erhalten wir

$$\begin{array}{llll} 1\cdot x & +2\cdot y & +3\cdot z & =10 \\ 0\cdot x & +1\cdot y & +2\cdot z & =20/3 \end{array}$$

Schließlich ziehen wir das Doppelte der zweiten Zeile wieder von der ersten ab und erhalten

$$\begin{array}{llll} 1\cdot x & +0\cdot y & -1\cdot z & =-10/3 \\ 0\cdot x & +1\cdot y & +2\cdot z & =20/3 \end{array}$$

sodass sich die Lösungsmenge ergibt zu

$$\mathcal{L} = \{ \begin{pmatrix} z-10/3 \\ -2\cdot z+20/3 \\ z \end{pmatrix} \mid z\in\mathbb{R}\} = \begin{pmatrix} -10/3 \\ 20/3 \\ 0 \end{pmatrix} + \mathbb{R}\cdot \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}.$$

Beim Lösen eines Linearen Gleichungssystems muss man also in geeigneter Weise die Koeffizienten manipulieren. Um dies effizient handhaben zu können, fasst man die Koeffizienten zusammen zu einem Objekt.

Definition 4.1.4 (Matrizen)

Es seien R ein kommutativer Ring und p, q natürliche Zahlen. Eine **$p\times q$-Matrix mit Einträgen in** R ist eine Abbildung

$$A : \{1,2,\ldots,p\}\times\{1,2,\ldots,q\} \longrightarrow R.$$

Dabei heißt p die Anzahl der Zeilen und q die Anzahl der Spalten von A. Wir schreiben meistens $a_{ij} := A(i,j)$, und notieren die Matrix A suggestiv als

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1q} \\ a_{21} & a_{22} & \ldots & a_{2q} \\ \vdots & \ldots & \ldots & \vdots \\ a_{p1} & a_{p2} & \ldots & a_{pq} \end{pmatrix}.$$

Die Menge aller $p\times q$-Matrizen mit Einträgen in R notieren wir als $R^{p\times q}$. Speziell ist $R^p := R^{p\times 1}$ die Menge aller Matrizen mit nur einer Spalte der Länge p.

Definition 4.1.5 (Φ_A, Produkt zweier Matrizen)

Die Abbildung Φ aus Hilfssatz 4.1.2 und die Matrix A legen sich gegenseitig fest. Wenn A gegeben ist, so schreiben wir in Zukunft oft Φ_A für die zugehörige Abbildung von R^q nach R^p. Statt $\mathcal{L}(\Phi_A, b)$ schreiben wir meistens $\mathcal{L}(A, b)$

Den engen Zusammenhang zwischen A und Φ_A benutzt man nun.

Wenn nämlich neben dem linearen Gleichungssystem $(*)$ noch ein weiteres mit q Zeilen gegeben ist:

$$\sum_{k=1}^{r} c_{jk} y_k = d_j, \quad 1 \leq j \leq q,$$

dann gehört zur Matrix $C = (c_{jk}) \in R^{q \times r}$ ein Homomorphismus

$$\Phi_C : R^r \to R^q, \quad \Phi_C((y_k)_{1 \leq k \leq r}) := (\sum_{k=1}^{r} c_{jk} y_k)_{1 \leq j \leq q}.$$

Dann ist aber $\Phi_A \circ \Phi_C$ der Homomorphismus von $R^r \mapsto R^p$, der durch

$$\begin{aligned} \Phi_A \circ \Phi_C((y_k)_{1 \leq k \leq r}) &= \Phi_A(\Phi_C((y_k)_{1 \leq k \leq r})) \\ &= (\textstyle\sum_{j=1}^{q} a_{ij} \sum_{k=1}^{r} c_{jk} y_k)_{1 \leq i \leq p} \\ &=: (\textstyle\sum_{k=1}^{r} f_{ik} y_k)_{1 \leq i \leq p}. \end{aligned}$$

gegeben wird. Auf diese Art haben wir eine neue Matrix $F \in R^{p \times r}$ definiert mit den Einträgen

$$f_{ik} := \sum_{j=1}^{q} a_{ij} c_{jk}, \ 1 \leq i \leq p, \ 1 \leq k \leq r.$$

Wir nennen diese Matrix F das **Produkt** $A \cdot C$ der Matrizen $A \in R^{p \times q}$ und $C \in R^{q \times r}$. Das Matrizenprodukt ist so gemacht, dass gilt:

$$\boxed{\Phi_A \circ \Phi_C = \Phi_{A \cdot C}.}$$

Bemerkung 4.1.6 (Wichtige Merkregel)

Der Eintrag von $A \cdot C$ an der Stelle (i, k) ergibt sich durch Multiplikation der i-ten Zeile von A mit der k-ten Spalte von C. Dabei wird eine Zeile der Länge q (also eine $1 \times q$-Matrix) mit einer Spalte der Länge q (also mit einer $q \times 1$-Matrix) multipliziert, indem man die j-ten Einträge beider Faktoren multipliziert und diese Produkte aufsummiert:

$$(a_1 \ \ldots \ a_q) \cdot \begin{pmatrix} c_1 \\ \vdots \\ c_q \end{pmatrix} = a_1 c_1 + a_2 c_2 + \ldots + a_q c_q.$$

Beispiel 4.1.7 a) Das Tupel $x = \begin{pmatrix} x_1 \\ \vdots \\ x_q \end{pmatrix} \in R^{q\times 1}$ ist selbst eine $q \times 1$-Matrix, und man kann die linke Seite von $(*)$ als $A \cdot x$ schreiben. Das ist die Matrizenschreibweise für das Lineare Gleichungssystem:

$$A \cdot x = b.$$

b) Vielleicht ist doch auch ein Zahlenbeispiel hilfreich. Wir nehmen Matrizen mit Einträgen in den ganzen Zahlen:

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} = \begin{pmatrix} 22 & 28 \\ 49 & 64 \end{pmatrix}.$$

Dabei ist zum Beispiel $28 = 1 \cdot 2 + 2 \cdot 4 + 3 \cdot 6$.

Aufgabe 4.1.8 (Konkrete Produkte)

Gegeben seien die Matrizen mit ganzzahligen Einträgen

$$A = \begin{pmatrix} 1 & -2 & 3 \\ 1 & 2 & -3 \end{pmatrix}, \quad B = \begin{pmatrix} -1 & 2 \\ 3 & -4 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & 1 & -1 \\ 0 & 3 & -3 \\ 0 & -7 & 7 \end{pmatrix}.$$

Entscheiden Sie, welches der Produkte AB, AC, BA, BC, CA, CB definiert ist, und berechnen Sie diese Matrix gegebenenfalls.

Aufgabe 4.1.9 Die Adjazenzmatrix eines Graphen

Es sei $n \in \mathbb{N}$ gegeben und $E = \{1, \ldots, n\}$ die Eckenmenge eines Graphen Γ mit Kantenmenge K. Die **Adjazenzmatrix** von Γ ist gegeben als

$$A = (a_{i,j})_{1\le i,j\le n}, \quad \text{wobei} \quad a_{i,j} = \begin{cases} 1, & \text{falls } \{i,j\} \in K, \\ 0, & \text{sonst.} \end{cases}$$

Das heißt, dass $a_{i,j}$ genau dann 1 ist, wenn die Ecken i und j durch eine Kante verbunden sind, es also einen Kantenzug der Länge 1 gibt, siehe 1.4.10.

a) Stellen Sie die Adjazenzmatrizen der beiden Graphen aus 1.3.10 auf.

b) Überlegen Sie sich zunächst für die beiden Graphen aus Teil a) und dann möglichst allgemein, wieso die Einträge des Quadrates $A^2 = A \cdot A$ der Adjazenzmatrix eines Graphen angeben, wieviele Kantenzüge der Länge 2 zwischen 2 Ecken es gibt.

c) Kann man ähnlich auch die Anzahl der Kantenzüge von Länge l zwischen zwei Ecken als Eintrag von $A^l = A \cdot A \cdot \ldots \cdot A$ (l Faktoren) berechnen? Hier ist natürlich eine Begründung gefragt.

Streng genommen müssten wir hier schon wissen, dass die Matrizenmultiplikation assoziativ ist. Das sehen wir als nächstes:

Bemerkung 4.1.10 (Assoziativität der Matrizenmultiplikation)

Da die Komposition von Abbildungen assoziativ ist (Fazit 1.3.5) und die Zuordnung

$$A \mapsto \Phi_A, \quad R^{p\times q} \longrightarrow \mathrm{Abb}(R^q, R^p),$$

injektiv ist, überträgt sich die Assoziativität der Komposition von Abbildungen auf das Multiplizieren von Matrizen:

$$\forall A \in R^{p\times q}, B \in R^{q\times r}, C \in R^{r\times s} : (A \cdot B) \cdot C = A \cdot (B \cdot C).$$

Das wird uns noch oft hilfreich sein und lässt sich auch explizit nachrechnen; für $1 \leq i \leq p$ und $1 \leq l \leq s$ gilt nämlich:

$$\begin{aligned}((A \cdot B) \cdot C)(i,l) &= \textstyle\sum_{k=1}^{r} (A \cdot B)(i,k) \cdot C(k,l) = \\ &= \textstyle\sum_{k=1}^{r} \sum_{j=1}^{q} A(i,j) \cdot B(j,k) \cdot C(k,l) = \\ &= \textstyle\sum_{j=1}^{q} \sum_{k=1}^{r} A(i,j) \cdot B(j,k) \cdot C(k,l) = \\ &= \textstyle\sum_{j=1}^{q} A(i,j) \cdot (B \cdot C)(j,l) = \\ &= (A \cdot (B \cdot C))(i,l).\end{aligned}$$

Dabei ist es tatsächlich hilfreich, die Matrizen als Abbildungen aufzufassen, damit man nicht für alle Produkte neue Buchstaben verwenden muss.

Definition 4.1.11 (Summe zweier Matrizen)

Für zwei Matrizen $A, B \in R^{p\times q}$ derselben Größe definieren wir eine neue Matrix $S \in R^{p\times q}$ durch

$$\forall 1 \leq i \leq p,\ 1 \leq j \leq q :\ S(i,j) := A(i,j) + B(i,j)$$

S heißt die **Summe** $A + B$ **von** A **und** B.

Es gilt für alle $x \in R^q : A \cdot x + B \cdot x = (A + B) \cdot x$, das heißt

$$\Phi_{A+B} = \Phi_A + \Phi_B,$$

und daraus ergibt sich sofort das Distributivgesetz

$$\forall C \in R^{r\times p}, D \in R^{q\times s} : C \cdot (A+B) = C \cdot A + C \cdot B, \quad (A+B) \cdot D = A \cdot D + B \cdot D,$$

denn Φ_C und Φ_D sind Gruppenhomomorphismen.

Die Addition von Matrizen ist assoziativ und kommutativ, die Matrix, die alle Einträge gleich 0 hat (die **Nullmatrix** genannt und auch als 0 geschrieben) ist neutrales Element für die Addition, und die Matrix B mit Einträgen $B(i,j) := -A(i,j)$ ist additiv zu A invers:

Fazit 4.1.12

$\boxed{(R^{p\times q},+) \text{ ist eine kommutative Gruppe.}}$

Schließlich sei $I_p \in R^{p\times p}$ die Matrix mit den Einträgen

$$I_p(i,j) := \begin{cases} 1 & \text{falls } i = j, \\ 0 & \text{falls } i \neq j; \end{cases}$$

Man nennt I_p die $p \times p$-**Einheitsmatrix**. Dann rechnet man nach, dass für alle $A \in R^{p\times q}$ die Gleichungen

$$I_p \cdot A = A = A \cdot I_q$$

gelten.

Definition 4.1.13 (Multiplikation mit Skalaren, Transponieren)

a) Für eine Matrix $A = (a_{ij})_{i,j} \in R^{p\times q}$ und $r \in R$ definieren wir

$$r \cdot A := A \cdot r := (r \cdot a_{ij})_{i,j}.$$

In Worten: die Einträge der Matrix A werden mit r multipliziert. Dabei ist, da R kommutativ ist, egal, ob diese Multiplikation von links oder von rechts durchgeführt wird. Offensichtlich gelten Regeln wie

$$\forall A \in R^{n\times p}, B \in R^{p\times q}, r \in R : A \cdot r \cdot B = r \cdot A \cdot B = A \cdot B \cdot r.$$

Wieso werden hier keine Klammern benötigt?

b) Für eine Matrix $A = (a_{ij})_{i,j} \in R^{p\times q}$ definieren wir die **transponierte Matrix** $A^\top \in R^{q\times p}$ durch

$$A^\top(j,i) := A(i,j).$$

Für die transponierte Matrix gibt es in der Literatur mehrere konkurrierende Notationen, z.B. A^t oder ${}^tA$.

Für das Produkt $A \cdot B$ zweier Matrizen gilt (bitte nachrechnen!):

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

Am Ende diese Abschnittes halten wir fest:

Fazit 4.1.14

$(R^{p\times p}, +, \cdot)$ ist ein Ring.
Das Einselement ist I_p. Das Nullelement ist die Nullmatrix.
Ist $R \neq \{0\}$ und $p \geq 2$, so ist $R^{p\times p}$ nicht kommutativ.

Wir nennen $R^{p\times p}$ mit dieser Struktur den **Matrizenring** der $p \times p$-Matrizen.

Der Ring R kann (für jedes $p \geq 1$) als Teilring von $R^{p\times p}$ aufgefasst werden, indem man R mithilfe des injektiven Ringhomomorphismus

$$\iota : R \longrightarrow R^{p\times p}, \iota(r) := r \cdot I_p$$

in den Matrizenring „einbettet“ (das Bild von ι ist zu R isomorph). Wie man sofort nachrechnen kann, gilt für alle Matrizen $A \in R^{p\times p}$ und alle $r \in R$:

$$\iota(r) \cdot A = r \cdot A = A \cdot \iota(r),$$

wobei in der Mitte die skalare Multiplikation wie in 4.1.13 verwendet wird.

Daher erhält man wie in 3.3.7 d) für jede Matrix $A \in R^{p\times p}$ einen Ringhomomorphismus

$$E_A : R[X] \longrightarrow R^{p\times p},\ \sum_{i=0}^{d} r_i X^i \mapsto \sum_{i=0}^{d} r_i A^i,$$

wobei wie immer (siehe 3.3.7a)) A^0 das Einselement in dem Ring ist, in dem man A betrachtet, hier also $A^0 = I_p$.

NB: Lassen Sie sich nicht dadurch verwirren, dass der Buchstabe A jetzt nicht mehr einen Ring bezeichnet wie in 3.3.7, sondern das, was dort a war. Eine Matrix heißt nun einmal typischer Weise A und nicht a. Der Ring, für den dort im allgemeinen Fall A stand (was kurz für Algebra steht), ist hier $R^{p\times p}$. Buchstaben sind Namen, die ihre Bedeutung ändern können. Es sollte allerdings im jeweiligen Kontext die Bedeutung immer geklärt sein.

Aufgabe 4.1.15 (2×2-Matrizen in Polynomen)

Es sei R ein beliebiger kommutativer Ring. Weiter betrachten wir das Polynom $f = X^2 + sX + t \in R[X]$.

Rechnen Sie nach, dass $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in R^{2\times 2}$ sicher dann die Gleichung $f(A) = 0$ erfüllt, wenn $a + d = -s$ und $ad - bc = t$ gelten.

Belegen Sie damit, dass es im Fall $R = \mathbb{R}$ unendlich viele Matrizen $A \in \mathbb{R}^{2\times 2}$ gibt, für die $A^2 = I_2$ gilt.

4.2 Invertierbare Matrizen

Definition 4.2.1 (Invertierbare Matrizen)

Es sei R ein kommutativer Ring. Die Einheitengruppe des Rings $R^{p\times p}$ bezeichnet man mit $\mathrm{GL}_p(R)$, was das englische „general linear group“ abkürzt:

$$\mathrm{GL}_p(R) = \{A \in R^{p\times p} \mid \exists B \in R^{p\times p} : AB = BA = I_p\}.$$

Die Matrizen in $\mathrm{GL}_p(R)$ heißen **invertierbare** oder auch **reguläre** Matrizen.

Beispiel 4.2.2 a) Für jedes $\alpha \in R$ ist die Matrix

$$A := \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix}$$

invertierbar:

$$A^{-1} = \begin{pmatrix} 1 & -\alpha \\ 0 & 1 \end{pmatrix}.$$

Wird diese Matrix inks an eine Matrix mit zwei Zeilen multipliziert, so ergibt sich

$$\begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} a & b & \dots & c \\ d & e & \dots & f \end{pmatrix} = \begin{pmatrix} a+\alpha d & b+\alpha e & \dots & c+\alpha f \\ d & e & \dots & f \end{pmatrix}.$$

Die Multiplikation mit dieser Matrix addiert also das α-fache der zweiten Zeile zur ersten und lässt die zweite Zeile unverändert: diese Art von Zeilenmanipulation brauchen wir für Beispiel 4.1.3!

b) Genauso ist die Matrix

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

invertierbar, sie ist sogar zu sich selbst invers (nachrechnen!). Beim Multiplizieren mit einer Matrix mit zwei Zeilen ergibt sich

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} a & b & \dots & c \\ d & e & \dots & f \end{pmatrix} = \begin{pmatrix} d & e & \dots & f \\ a & b & \dots & c \end{pmatrix}.$$

Die zwei Zeilen werden vertauscht, was wir auch zum systematischen Lösen von Linearen Gleichungssystemen verwenden werden.

c) Schließlich ist

$$D := \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}$$

als Matrix invertierbar, wenn α und β dies in R sind. Die Inverse ist dann

$$D^{-1} = \begin{pmatrix} \alpha^{-1} & 0 \\ 0 & \beta^{-1} \end{pmatrix}.$$

Wenn D von links an eine Matrix mit zwei Zeilen multipliziert wird, so ergibt sich

$$\begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} \cdot \begin{pmatrix} a & b & \dots & c \\ d & e & \dots & f \end{pmatrix} = \begin{pmatrix} \alpha a & \alpha b & \dots & \alpha c \\ \beta d & \beta e & \dots & \beta f \end{pmatrix}.$$

Die zwei Zeilen der rechten Matrix werden einfach mit α bzw. β multipliziert.

Was wir jetzt für Matrizen mit zwei Zeilen schon können, wollen wir für beliebige Matrizen auch erreichen. Dazu definieren wir einige spezielle Matrizen und fangen an mit den Elementarmatrizen.

Definition 4.2.3 (Elementarmatrizen)

Es seien R ein kommutativer Ring und p, q natürliche Zahlen. Dann ist für $1 \leq i \leq p,\ 1 \leq j \leq q$ die **Elementarmatrix** $E_{i,j} \in R^{p \times q}$ definiert durch ihre Einträge $E_{i,j}(k,l),\ 1 \leq k \leq p,\ 1 \leq l \leq q,$ die man auf folgende Art festlegt:

$$E_{i,j}(k,l) := \begin{cases} 1 & \text{falls } i = k \text{ und } j = l, \\ 0 & \text{sonst.} \end{cases}$$

Der Eintrag, der sowohl in der i-ten Zeile als auch in der j-ten Spalte steht, ist 1, alle anderen Einträge sind 0. Die Größe von $E_{i,j}$ wird in der Notation nicht mit festgehalten. Sie wird sich oft aus dem Kontext ergeben.

Beispiel : $R^{3\times 4} \ni E_{2,3} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$

Es ist klar, dass sich jede Matrix $A = (a_{ij}) \in R^{p \times q}$ schreiben lässt als

$$A = \sum_{i,j} a_{ij} \cdot E_{i,j}.$$

Eine Auswertung der Multiplikationsregeln führt auf folgende Merkregel:

Fazit 4.2.4 Für $E_{i,j} \in R^{p\times q}, E_{k,l} \in R^{q\times s},$ gilt:

$$\boxed{E_{i,j} \cdot E_{k,l} = \begin{cases} E_{i,l} & \text{falls } j = k \\ 0 & \text{sonst.} \end{cases}}$$

Denn: Das Produkt kann höchstens in der i-ten Zeile einen von Null verschiedenen Eintrag haben, und höchstens in der l-ten Spalte. Dort ist der Eintrag genau dann nicht 0 (sondern 1), wenn die 1 aus der i-te Zeile von $E_{i,j}$ beim Multiplizieren auf die 1 aus der l-ten Spalte von $E_{k,l}$ trifft, also wenn $j = k$.

Ein Spezialfall ist der einer einzelnen Spalte. Wir schreiben dann kürzer

$$e_j := E_{j,1} \in R^q = R^{q\times 1}.$$

$$\boxed{\text{Für } A \in R^{p\times q} \text{ und } e_j \in R^q \text{ gilt } A \cdot e_j = \sum_{i=1}^{p} a_{ij} e_i \in R^p.}$$

Das ist die j-te Spalte von A. Diese Regel wird uns noch oft begegnen.

Nun verallgemeinern wir die drei Matrizentypen, die wir in Beispiel 4.2.2 schon im 2×2–Fall kennengelernt haben.

Definition/Bemerkung 4.2.5 (Additionsmatrizen)

Für $1 \leq i \neq j \leq p$ und $\alpha \in R$ definieren wir die Matrix $A_{i,j}(\alpha) \in R^{p\times p}$ durch

$$A_{i,j}(\alpha) := I_p + \alpha E_{i,j}.$$

Diese Matrix heißt eine **Additionsmatrix.** Sie hat als Einträge Einsen auf der Diagonalen, α an der Stelle (i, j) und Null überall sonst. Es gilt:

$$A_{i,j}(\alpha)\cdot A_{i,j}(-\alpha) = (I_p+\alpha E_{i,j})\cdot(I_p-\alpha E_{i,j}) = I_p+\alpha E_{i,j}-\alpha E_{i,j}-\alpha^2 E_{i,j}\cdot E_{i,j} = I_p,$$

da $i \neq j$. Genauso gilt

$$A_{i,j}(-\alpha) \cdot A_{i,j}(\alpha) = I_p.$$

Es folgt, dass Additionsmatrizen invertierbar sind.

$$A_{i,j}(\alpha) \in \mathrm{GL}_p(R).$$

Für $M = \sum_{i,j} m_{ij} \cdot E_{i,j} \in R^{p\times q}$ gilt nun

$$\begin{aligned} A_{i,j}(\alpha) \cdot M &= (I_p + \alpha E_{i,j}) \cdot M \\ &= M + \alpha E_{i,j} \cdot M \\ &= M + \alpha E_{i,j} \cdot \textstyle\sum_{k,l} m_{kl} \cdot E_{k,l} \\ &= M + \textstyle\sum_l \alpha m_{jl} E_{i,l}. \end{aligned}$$

Das ist die Matrix, die aus M entsteht, indem man zur i-ten Zeile das α-fache der j-ten Zeile addiert – daher der Name Additionsmatrix!

Definition/Bemerkung 4.2.6 (Vertauschungsmatrizen)

Für $1 \leq i, j \leq p$ sei die **Vertauschungsmatrix** $V_{i,j} \in R^{p\times p}$ definiert durch

$$V_{i,j} := I_p - E_{i,i} - E_{j,j} + E_{i,j} + E_{j,i}.$$

Bildlich gesprochen ersetzt man also die Einsen, die in der Einheitsmatrix an den Stellen (i,i) und (j,j) stehen, durch Einsen an den Stellen (i,j) und (j,i). Ähnlich wie bei den Additionsmatrizen rechnet man nach, was die Multiplikation von links mit $V_{i,j}$ mit einer Matrix $M \in R^{p\times q}$ tut:

$$V_{i,j} \cdot M = M + \sum_l (m_{jl} - m_{il}) E_{i,l} + \sum_l (m_{il} - m_{jl}) E_{j,l}.$$

Das ist die Matrix, die aus M durch Vertauschen der i-ten und j-ten Zeile entsteht – woher kommt also der Name Vertauschungsmatrix? Speziell gilt

$$V_{i,j} \cdot V_{i,j} = I_p,$$

also insbesondere

$$V_{i,j} \in \mathrm{GL}_p(R).$$

Definition/Bemerkung 4.2.7 (Diagonalmatrizen)

Für Elemente $\alpha_1, \ldots, \alpha_p \in R$ definieren wir die **Diagonalmatrix** $\mathrm{diag}\,(\alpha_1, \ldots, \alpha_p)$ durch

$$\mathrm{diag}(\alpha_1, \ldots, \alpha_p) := \sum_{i=1}^{p} \alpha_i E_{i,i} \in R^{p\times p}.$$

(Woher kommt der Name?) Dann haben wir für $M \in R^{p\times q}$:

$$\mathrm{diag}(\alpha_1, \ldots, \alpha_p) M = \sum_{i,j} \alpha_i m_{ij} E_{i,j}.$$

Das ist die Matrix, die aus M entsteht, indem für alle i die i-te Zeile mit α_i multipliziert wird.

Wenn insbesondere die α_i allesamt Einheiten in R sind, dann gilt:

$$\mathrm{diag}(\alpha_1, \ldots, \alpha_p) \cdot \mathrm{diag}(\alpha_1^{-1}, \ldots, \alpha_p^{-1}) = \mathrm{diag}(\alpha_1^{-1}, \ldots, \alpha_p^{-1}) \cdot \mathrm{diag}(\alpha_1, \ldots, \alpha_p) = I_p,$$

also gilt in diesem Fall

$$\mathrm{diag}(\alpha_1, \ldots, \alpha_p) \in \mathrm{GL}_p(R).$$

Manchmal ist es hilfreich, die Einträge einer Matrix zu kleineren Untermatrizen zusammenzufassen. Diese kompaktere Schreibweise benutzen wir im folgenden Hilfssatz, den wir im übernächsten Abschnitt verwenden wollen.

Hilfssatz 4.2.8 *Für die natürliche Zahl $k \leq p$ sei eine Matrix $A \in R^{p\times p}$ durch folgende Blockgestalt gegeben:*

$$A = \begin{pmatrix} I_k & B \\ 0 & D \end{pmatrix},$$

wobei I_k *die* $k \times k$*-Einheitsmatrix,* 0 *die Nullmatrix der Größe* $(p-k) \times k$*,* $B \in R^{k\times(p-k)}$*, und* $D \in R^{(p-k)\times(p-k)}$*. Dann ist* A *genau dann invertierbar, wenn* D *invertierbar ist. In diesem Fall gilt*

$$A^{-1} = \begin{pmatrix} I_k & -BD^{-1} \\ 0 & D^{-1} \end{pmatrix}.$$

Beweis. Wenn D invertierbar ist, dann rechnet sich leicht nach, dass die angegebene Matrix zu A invers ist. Es ist also nur noch zu zeigen, dass aus der Invertierbarkeit von A auch die von D folgt.

Dazu sei also A invertierbar mit inverser Matrix $M \in R^{p\times p}$. Dann ist

$$M \cdot A = I_p,$$

aber für $1 \leq i \leq k$ ist die i-te Spalte dieses Produkts wegen 4.2.4 gleich

$$M \cdot e_i = i-\text{te Spalte von } M.$$

Also hat auch M eine „Blockgestalt“ wie A:

$$M = \begin{pmatrix} I_k & E \\ 0 & F \end{pmatrix}.$$

Die Produkte $A \cdot M$ und $M \cdot A$ zeigen dann, dass $D \cdot F = F \cdot D = I_{p-k}$. ◯

Zu guter Letzt wollen wir für spätere Zwecke hier noch das Beispiel der 2×2-Matrizen diskutieren.

Beispiel 4.2.9 (invertierbare 2×2-Matrizen)

Es seien R ein kommutativer Ring und $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in R^{2\times 2}$. Dann ist A genau dann regulär, wenn $\det(A) := ad - bc$ eine Einheit in R ist. Denn:

- Wenn A regulär ist, dann gibt es eine zu ihr inverse Matrix $E = \left(\begin{smallmatrix} e & f \\ g & h \end{smallmatrix}\right)$, und es gilt
$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = A \cdot E = \begin{pmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{pmatrix}.$$
Vergleicht man hier die Einträge der Einheitsmatrix mit denen des Produkts ganz rechts, dann folgt $af + bh = 0$ und
$$\begin{aligned} 1 &= \det(I_2) = 1 \cdot 1 \\ &= (ae+bg)\cdot(cf+dh) = acef + adeh + bcfg + bdgh \\ &= -bceh + adeh + bcfg - adfg = (ad-bc)\cdot(eh-fg). \end{aligned}$$
Also ist $1 = \det(A) \cdot \det(E)$, und $\det(A)$ ist in R invertierbar.

- Wenn $\det(A)$ in R invertierbar ist, dann gilt für

$$E := (\det(A))^{-1} \cdot \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

dass

$$AE = EA = I_2,$$

also ist A invertierbar.

Die Größe $\det(A) \in R$ heißt die **Determinante** von A. Wir werden in Kapitel 7 auf diesen Begriff zurückkommen.

Aufgabe 4.2.10 (Permutationsmatrizen)

Es sei $p \in \mathbb{N}$ und $\sigma \in S_p$ eine Permutation der Menge $\{1, \ldots, p\}$. Wir bilden damit eine Matrix

$$A_\sigma = \sum_{i=1}^{p} E_{\sigma(i),i} \in \mathbb{Z}^{p \times p}.$$

Dabei sind alle Elementarmatrizen in $\mathbb{Z}^{p \times p}$.

In jeder Zeile und jeder Spalte von A_σ steht genau eine 1, alle anderen Einträge sind 0. Solche Matrizen heißen **Permutationsmatrizen**.

Rechnen Sie nach, dass für zwei Permutationen $\sigma, \tau \in S_p$ stets

$$A_\sigma \cdot A_\tau = A_{\sigma \circ \tau}$$

gilt. Dabei wird links das Matrizenprodukt verwendet und rechts die Komposition der Permutationen.

Rechtfertigen Sie mithilfe dieser Identität die Tatsache, dass A_σ invertierbar ist. Genauer: Welche Permutationsmatrix ist zu A_σ invers?

4.3 Die Gauß-Normalform

Wir kehren zurück zum Gleichungssystem $Ax = b$ aus 4.1.7.

Unser Ziel ist es nach wie vor, durch Manipulation der Zeilen der Matrix $A \in R^{p \times q}$ eine „möglichst einfache“ neue Matrix herauszubekommen, die aber bezüglich des linearen Gleichungssystems dieselbe Information enthält. Was wir unter dieser möglichst einfachen Matrix verstehen, das wird nun präzisiert.

Definition 4.3.1 (Treppenform, Gauß-Normalform, Rang)

Eine Matrix $T = (t_{ij}) \in R^{p \times q}$ hat **Treppenform** oder auch **Gauß-Normalform**, wenn es eine Zahl $r \in \mathbb{N}_0$ und natürliche Zahlen $1 \leq s_1 < s_2 < s_3 < \cdots < s_r \leq q$ gibt, sodass die folgenden Bedingungen erfüllt sind:

- Für alle i mit $1 \leq i \leq r$ gilt: $t_{i,s_i} = 1$ und $\forall k \neq i : t_{k,s_i} = 0$ und $\forall k < s_i : t_{i,k} = 0$.
- Für alle $i \geq r+1$ und alle $j \in \{1, \ldots, q\}$ gilt $t_{i,j} = 0$.

Wenn T Treppenform hat, so heißt die Zahl r der **Rang** von T, und $s_1, \ldots, s_r$ heißen die **Stufenindizes** von T.

In Worten sagen die Bedingungen: Für $1 \leq i \leq r$ ist die s_i-te Spalte von T gleich e_i (siehe 4.2.4), links von der Eins an der Stelle (i, s_i) stehen nur Nullen, und ab der $(r+1)$-ten Zeile sind alle Zeilen Null.

Über die übrigen Einträge werden keine Vorschriften erlassen.

Beispiel 4.3.2 Was bedeuten die Bedingungen an die Treppenform für $p = 4,\ q = 6, r = 3, s_1 = 2, s_2 = 3, s_3 = 5$? Eine Treppenform mit diesem Rang und diesen Stufenindizes muss die folgende Form haben:

$$\begin{pmatrix} 0 & 1 & 0 & * & 0 & * \\ 0 & 0 & 1 & * & 0 & * \\ 0 & 0 & 0 & 0 & 1 & * \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

wobei anstelle der Sternchen beliebige Einträge aus dem Ring R stehen können.

Der folgende Hilfssatz zeigt uns, dass ein lineares Gleichungssystem mit einer Koeffizientenmatrix in Treppenform tatsächlich leicht zu handhaben ist. Man kann genau sagen, wann $Tx = b$ lösbar ist, und wenn dem so ist, findet man eine spezielle Lösung (siehe 4.1.2) und den Lösungsraum des homogenen Gleichungssystems.

Hilfssatz 4.3.3 (Lösen eines LGS mit einer Matrix in Treppenform)

Es seien eine Matrix $T \in R^{p \times q}$ in Treppenform (vom Rang r und mit Stufenindizes $s_1, \ldots, s_r$) und eine Spalte $b = (b_1 \ \ldots \ b_p)^\top \in R^p$ gegeben. Dann gelten:

a) Das Lineare Gleichungssystem $T \cdot x = b$ ist genau dann lösbar, wenn die Einträge $b_{r+1}, \ldots, b_p$ von b alle 0 sind. In diesem Fall ist zum Beispiel

$$x^{(s)} := \sum_{i=1}^{r} b_i e_{s_i} \in \mathcal{L}(T, b).$$

eine spezielle Lösung.

b) Für $j \in J := \{1, \ldots, q\} \setminus \{s_1, \ldots, s_r\}$ ist $F^{(j)} := e_j - \sum_{i=1}^{r} t_{ij} e_{s_i}$ eine Lösung des homogenen Gleichungssystems $T \cdot x = 0$. Die $F^{(j)}$ nennen wir die Fundamentallösungen *des homogenen Gleichungssystems.*

c) Die Lösungsmenge $\mathcal{L}(T,0)$ des zu T gehörigen homogenen Gleichungssystems ist gegeben durch

$$\mathcal{L}(T,0) = \{\sum_{j\in J} x_j F^{(j)} \mid x_j \in R\}.$$

Jedes $v \in \mathcal{L}(T,0)$ hat genau eine Darstellung in der Form $v = \sum_{j\in J} x_j F^{(j)}$ mit Koeffizienten $x_j \in R$.

Beweis: a) Wenn das LGS lösbar ist, müssen natürlich die Einträge $b_{r+1} = \cdots = b_p = 0$ Null sein, da in den Zeilen $r+1, \ldots, p$ der Matrix T nur Nullen als Koeffizienten auftauchen. Wenn umgekehrt diese Koeffizienten alle Null sind, so gilt für den vorgeschlagenen Lösungsvektor wegen 4.2.4 und 4.3.1

$$T \cdot x^{(s)} = \sum_{i=1}^{r} b_i \cdot T \cdot e_{s_i} = \sum_{i=1}^{r} b_i \cdot e_i = b.$$

b) Hier gilt:

$$T \cdot F^{(j)} = T \cdot e_j - \sum_{i=1}^{r} t_{ij} \cdot T \cdot e_{s_i} = \sum_{k=1}^{r} t_{kj} e_k - \sum_{i=1}^{r} t_{ij} e_i = 0.$$

c) Die Inklusion $\supseteq$ ist klar, da $\mathcal{L}(T,0)$ als Gruppe unter der Addition abgeschlossen ist und für $r \in R$ und $j \in J$ auch $T \cdot (rF^{(j)}) = r \cdot (TF^{(j)}) = 0$, also $rF^{(j)} \in \mathcal{L}(T,0)$ gilt.

Sei umgekehrt $x \in \mathcal{L}(T,0)$. Dann ist auch

$$x - \sum_{j\in J} x_j F^{(j)} \in \mathcal{L}(T,0),$$

aber diese Spalte hat für $j \in J$ als j-ten Eintrag eine 0. Einsetzen in das homogene LGS zeigt, dass diese Differenz auch an den Stellen $s_1, \ldots, s_r$ eine 0 stehen hat, also insgesamt die Nullspalte ist. Das zeigt $x = \sum_{j\in J} x_j F^{(j)}$, und damit die Inklusion $\subseteq$.

Die Eindeutigkeit der Darstellung ist wieder klar, es gibt ja jeweils nur ein $F^{(j)}$, das an einer gegebenen Stelle $j_0 \in J$ einen von Null verschiedenen Eintrag hat, und dieser ist sogar 1. ◯

Bemerkung 4.3.4 (Der (-1)-Trick)[1]

Die Fundamentallösungen aus dem letzten Hilfssatz lassen sich mit folgendem Verfahren aus der Gauß-Normalform T ablesen.

[1]Der Name ist historisch gewachsen. Hier ist das eher ein $(1-)$-Trick.

Für jedes $1 \leq i \leq r$ sei die i-te Zeile von T die s_i-te Zeile einer neuen $q \times q$-Matrix S, deren übrige Zeilen 0 sind.

Dann sind die von Null verschiedenen Spalten der Matrix

$$I_q - S$$

genau die Fundamentallösungen von $Tx = 0$.

Genauer ist $F^{(j)}$ die j-te Spalte in $I_q - S$.

Als Beispiel betrachten wir noch einmal den Fall $p = 4,\ q = 6, r = 3, s_1 = 2, s_2 = 3, s_3 = 5$.

$$T = \begin{pmatrix} 0 & 1 & 0 & a & 0 & c \\ 0 & 0 & 1 & b & 0 & d \\ 0 & 0 & 0 & 0 & 1 & e \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

„Einpflanzen“ der von Null verschiedenen Zeilen an der richtigen Stelle führt zu

$$S = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & a & 0 & c \\ 0 & 0 & 1 & b & 0 & d \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & e \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Schließlich bilden wir

$$I_6 - S = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -a & 0 & -c \\ 0 & 0 & 0 & -b & 0 & -d \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -e \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

In der ersten, vierten und sechsten Spalte stehen jetzt die drei Fundamentallösungen

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -a \\ -b \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -c \\ -d \\ 0 \\ -e \\ 1 \end{pmatrix}.$$

Das sollten Sie jetzt mal selber ausprobieren:

Aufgabe 4.3.5 (Ein Beispiel)

Gegeben sei die Matrix

$$A = \begin{pmatrix} 1 & a & b & 0 & c & 0 & 0 \\ 0 & 0 & 0 & 1 & d & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \in \mathbb{Z}^{5\times 7}$$

und $t = (t_1, t_2, t_3, t_4, 0)^\top \in \mathbb{Z}^4$. Bestimmen Sie den Rang von A, die Stufenindizes, die Fundamentallösungen und schließlich den Lösungsraum $\mathcal{L}(A, t)$ in $\mathbb{Z}^7$.

4.4 Das Gauß-Verfahren

In diesem Abschnitt wollen wir endlich den Sack zubinden und das Verfahren angeben, mit dem sich zu einer gegebenen Matrix eine passende Gauß-Normalform findet. Dass wir dabei keine Information über unser LGS verlieren, zeigt der folgende Hilfssatz. Dazu erinnern wir uns daran, dass die Zeilenmanipulationen durch Multiplikation von links mit invertierbaren Matrizen vollzogen werden können.

Hilfssatz 4.4.1 (weshalb das Gaußverfahren erlaubt ist)

Wenn ein lineares Gleichungssystem $A\cdot x = b$ mit $A \in R^{p\times q}$ und $b \in R^p$ gegeben ist, dann stimmt für jede invertierbare Matrix $C \in \mathrm{GL}_p(R)$ die Gleichung

$$\mathcal{L}(A, b) = \mathcal{L}(CA, Cb).$$

Beweis. Aus $A\cdot x = b$ folgt durch Multiplikation (von links) mit C: $CA\cdot x = Cb$, also $x \in \mathcal{L}(CA, Cb)$, also $\mathcal{L}(A, b) \subseteq \mathcal{L}(CA, Cb)$.

Die umgekehrte Inklusion folgt durch Multiplikation mit C^{-1}. ◯

Wir wollen nun eine invertierbare Matrix C finden, sodass $C \cdot A$ Gauß-Normalform hat. Das wird im Allgemeinen nicht möglich sein, zum Beispiel lässt sich die Matrix $\left(\begin{smallmatrix} 2 & 0 \\ 0 & 1 \end{smallmatrix}\right) \in \mathbb{Z}^{2\times 2}$ nicht auf Gauß-Normalform bringen, weil man in $\mathbb{Z}$ im Allgemeinen nicht durch 2 dividieren kann.

Aber alles geht gut, wenn wir voraussetzen, dass der Grundring R ein Körper ist – was wir jetzt tun, und wir schreiben dann auch K statt R. Nun dürfen wir durch jedes von 0_K verschiedene Element teilen, das heißt: mit seinem inversen multiplizieren.

Satz 4.4.2 (Gauß-Normalform; Gauß-Algorithmus)

Es seien K ein Körper und $A \in K^{p\times q}$ eine Matrix. Dann gibt es eine invertierbare Matrix $C \in \mathrm{GL}_p(K)$, sodass $C \cdot A$ Gauß-Normalform hat.

Diese Gauß-Normalform (nicht aber C!) ist eindeutig durch A bestimmt.

Beweis. Existenz: Wir beweisen die Existenz von C per Induktion nach der Anzahl p der Zeilen von A.

Für $p = 0$ ist nichts zu zeigen, das ist aber auch der Induktionsanfang für Puristen.

Wer dem skeptisch gegenübersteht, wird sich über den Fall $p = 1$ freuen. In diesem Fall ist entweder A die Nullmatrix und damit schon in Treppenform, oder A ist nicht 0. Dann gibt es $s_1 := \min\{j \mid a_{1j} \neq 0\}$. Das Produkt von A mit der (invertierbaren) 1×1-Matrix (a_{1,s_1}^{-1}) ist eine Treppenform. Damit ist der Induktionsanfang gemacht – und schon hier brauchen wir, dass K ein Körper ist.

Für den Induktionsschritt nehmen wir an, es sei $p \geq 2$ und die Existenz von C für alle Matrizen mit $p-1$ Zeilen schon gezeigt. Ist $A = 0$, so sind wir wieder fertig. Ansonsten wählen wir

$$s_1 := \min\{j \mid \exists i : a_{ij} \neq 0\}.$$

Weiter sei

$$i_0 := \min\{i \mid a_{is_1} \neq 0\}.$$

Dann ist

$$V_{1,i_0}\Big(\prod_{i\neq i_0} A_{i,i_0}(-a_{i,s_1})\Big)DA =: \left(\begin{array}{cc|c} 0 & 1 & z \\ \hline \vdots & 0 & \\ \vdots & \vdots & \widetilde{A} \\ 0 & 0 & \end{array}\right)$$

eine Matrix, deren erste von 0 verschiedene Spalte die Spalte e_1 an der s_1-ten Stelle ist, wobei $\widetilde{A} \in K^{(p-1)\times(q-s_1)}$ und $z = (z_{s_1+1}, \ldots, z_q) \in K^{1\times(q-s_1)}$. Dabei ist D diejenige Diagonalmatrix, die Einsen auf der Diagonale stehen hat, außer dem Eintrag a_{i_0,s_1}^{-1} an der i_0-ten Stelle.

Die Multiplikation mit D normiert den Eintrag von A an der Stelle (i_0, s_1) auf 1. Anschließend sorgen die Additionsmatrizen $A_{i,i_0}(-a_{i,s_1})$ dafür, dass in der s_1-ten Spalte alles außerhalb der i_0-ten Zeile 0 wird. Schließlich vertauscht V_{1,i_0} die erste Zeile mit der i_0-ten.

Nun hat aber $\widetilde{A}$ nur noch $p-1$ Zeilen, sodass die Induktionsvoraussetzung greift und eine Matrix $\widetilde{C} \in \mathrm{GL}_{p-1}(K)$ existiert, für die $\widetilde{C}\widetilde{A} =: \widetilde{T}$ Treppenform hat. Diese hat also einen Rang, den wir $r-1$ nennen und Stufenindizes, die wir mit $\tilde{s}_2, \ldots, \tilde{s}_r$ bezeichnen.

Nach 4.2.8 ist die $p \times p$-Matrix $\left(\begin{smallmatrix} 1 & 0 \\ 0 & \widetilde{C} \end{smallmatrix}\right)$ invertierbar. Es gilt außerdem

$$\begin{pmatrix} 1 & 0 \\ 0 & \widetilde{C} \end{pmatrix} \cdot \left(\begin{array}{cc|c} 0 & 1 & z \\ \hline \vdots & 0 & \\ \vdots & \vdots & \widetilde{A} \\ 0 & 0 & \end{array}\right) = \left(\begin{array}{cc|c} 0 & 1 & z \\ \hline \vdots & 0 & \\ \vdots & \vdots & \widetilde{T} \\ 0 & 0 & \end{array}\right),$$

und dies ist schon fast eine Treppenform, wenn man von der ersten Zeile absieht. Diese wird nun noch durch Multiplikation mit den Matrizen $A_{1,i}(-z_{\tilde{s}_i+s_1})$ (wobei $2 \leq i \leq r$) an den Stellen, die in der großen Matrix den Stufenindizes von $\widetilde{A}$ entsprechen, zu 0 gemacht. Damit erhalten wir: Die Matrix

$$C := \prod_{i=2}^{r} A_{1,i}(-z_{\tilde{s}_i+s_1}) \cdot \begin{pmatrix} 1 & 0 \\ 0 & \widetilde{C} \end{pmatrix} \cdot V_{1,i_0}(\prod_{i \neq i_0} A_{i,i_0}(-a_{i,s_1}))D$$

ist invertierbar und $T := C \cdot A$ hat Treppenform. Der Rang von T ist r, und die Stufenindizes sind $s_1, \ldots, s_r$, wenn man für $i \geq 2$ die Größen s_i durch

$$s_i := \tilde{s}_i + s_1$$

definiert.

Damit ist die Existenz von C und die der Treppenform gezeigt.

Eindeutigkeit: Wenn C und $\widetilde{C}$ zwei invertierbare Matrizen sind, für die $C \cdot A$ und $\widetilde{C} \cdot A$ Treppenform haben, dann gilt

$$C \cdot A = C \cdot \widetilde{C}^{-1} \cdot \widetilde{C} \cdot A = (C \cdot \widetilde{C}^{-1}) \cdot (\widetilde{C} \cdot A),$$

wobei auch $C \cdot \widetilde{C}^{-1}$ invertierbar ist. Zu zeigen ist also für die Eindeutigkeit der Treppenform das Folgende:

Wenn für Treppenformen T und $\widetilde{T}$ eine invertierbare Matrix D existiert, sodass $D \cdot T = \widetilde{T}$, dann gilt $T = \widetilde{T}$.

Diese Aussage wollen wir nun beweisen.

Wieder machen wir vollständige Induktion nach p.

Wieder ist der Induktionsanfang $p = 0$ einfach, aber für die meisten Leser unbefriedigend.

Also schauen wir auch den Fall $p = 1$ noch an: Es seien T und $\widetilde{T}$ $1 \times q$-Treppenformen, also Zeilen, deren erster von Null verschiedener Eintrag 1 ist. Weiter sei $d \cdot T = \widetilde{T}$ für ein von Null verschiedenes $d \in K$. Dann ist $T = 0 \Longleftrightarrow \widetilde{T} = 0$, und wenn $T \neq 0$ gilt, haben T und $\widetilde{T}$ an derselben Stelle den ersten von 0 verschiedenen Eintrag, was $d = 1$ und damit auch $T = \widetilde{T}$ impliziert.

Nun kommt der Induktionsschritt. Es sei $p \geq 2$ und die Behauptung der Eindeutigkeit für alle Treppenformen mit höchstens $p-1$ Zeilen gezeigt. Es seien $T, \widetilde{T} \in K^{p\times q}$ zwei Treppenformen, für die ein invertierbares $D \in K^{p\times p}$ existiert mit $D \cdot T = \widetilde{T}$. Die Matrix T habe den Rang r und die Stufenindizes $s_1, \ldots, s_r$, $\widetilde{T}$ habe den Rang $\tilde{r}$ und die Stufenindizes $\tilde{s}_1, \ldots, \tilde{s}_{\tilde{r}}$.

Dann ist klar, dass die erste von 0 verschiedene Spalte von $D \cdot T$ die s_1-te Spalte ist: Alle vorherigen Spalten von T sind 0, die s_1-te ist e_1, und $D \cdot e_1 \neq 0$, da D invertierbar ist. Andererseits ist die erste von Null verschiedene Spalte von $D \cdot T$ die $\tilde{s}_1$-te Spalte, denn $D \cdot T = \widetilde{T}$. Also ist $s_1 = \tilde{s}_1$, und außerdem ist

$$e_1 = \widetilde{T} \cdot e_{s_1} = D \cdot T \cdot e_{s_1} = D \cdot e_1 = \text{ erste Spalte von } D.$$

Damit ist aber D eine Matrix der Gestalt

$$D = \begin{pmatrix} 1 & z \\ 0 & \hat{D} \end{pmatrix}, \quad \hat{D} \in K^{(p-1)\times(p-1)},$$

und nach dem Hilfssatz 4.2.8 ist $\hat{D}$ invertierbar.

Nun führt aber $\hat{D}$ die Treppenmatrix $\hat{T}$, die aus T durch Entfernen der ersten Zeile entsteht, in die entsprechende Treppenmatrix $\hat{\widetilde{T}}$ über, womit diese nach Induktionsvoraussetzung gleich sind.

Das zeigt aber analog zum Argument für s_1, dass auch die anderen Stufenindizes übereinstimmen, und damit auch die Ränge:

$$r = \tilde{r}, s_1 = \tilde{s}_1, \ldots, s_r = \tilde{s}_r.$$

Dann folgt sofort aus 4.2.4 und 4.3.1, dass für $1 \leq i \leq r$ gilt:

$$D \cdot e_i = D \cdot T \cdot e_{s_i} = \widetilde{T} \cdot e_{s_i} = e_i.$$

Also ist D von der Gestalt

$$D = \begin{pmatrix} I_r & B \\ 0 & D_{p-r} \end{pmatrix}, \; B \in K^{r\times(p-r)}, D_{p-r} \in K^{(p-r)\times(p-r)}.$$

Daran sieht man dann aber $\widetilde{T} = D \cdot T = T$, da T ja von Null verschiedene Einträge nur in den ersten r Zeilen hat. ◯

Definition 4.4.3 (Rang einer Matrix)

Der **Rang** $\operatorname{Rang}(A)$ einer Matrix $A \in K^{p\times q}$ wird definiert als der Rang der eindeutig bestimmten Treppenform T von Gestalt $T = C \cdot A$, $C \in \mathrm{GL}_p(K)$ (siehe Definition 4.3.1).

Mit Satz 4.4.2 ergibt sich eine Lösungsstrategie für ein gegebenes Lineares Gleichungssystem wie folgt.

Fazit 4.4.4 (Lösungsstrategie zur Lösung eines LGS)

Konstruiere für $A \in K^{p\times q}$ ein $C \in \mathrm{GL}_p(K)$, sodass CA Treppenform hat. Berechne dann die spezielle Lösung $x^{(s)}$ von $CAx = Cb$ und die Fundamentallösungen wie in 4.3.3.

Insbesondere ist $Ax = b$ genau dann lösbar, wenn $\mathrm{Rang}\,(A) = \mathrm{Rang}(A|b)$ gilt, wobei $(A|b)$ die $p \times (q+1)$ Matrix ist, die aus A entsteht, indem b als letzte Spalte angehängt wird. Die Matrix $(A|b)$ heißt **erweiterte Matrix**.

$$\boxed{\mathcal{L}(A,b) \neq \emptyset \iff \mathrm{Rang}(A) = \mathrm{Rang}(A|b).}$$

Wenn eine Lösung existiert, ist sie genau dann eindeutig, wenn $\mathrm{Rang}\,(A) = q$ gilt. Denn genau dann wird das homogene Gleichungssystem nur von der Null gelöst (es gibt also keine Fundamentallösungen des homogenen Gleichungssystems), siehe 4.3.3. Anders gesagt:

$$\boxed{\Phi_A \text{ ist injektiv} \iff \mathrm{Rang}(A) = q.}$$

Für gegebenes A gilt: $Ax = b$ ist für jedes b lösbar genau dann, wenn A den Rang $r = p$ hat. Denn genau dann kann der Rang der erweiterten Matrix durch keine Wahl von b größer als $\mathrm{Rang}\,(A)$ werden. Anders gesagt:

$$\boxed{\Phi_A \text{ ist surjektiv} \iff \mathrm{Rang}(A) = p.}$$

Aufgabe 4.4.5 (Die Hessesche Normalform)

Es seien K ein Körper und $v, w \in K^3$ zwei Spalten, von denen keine ein Vielfaches der anderen ist. Wir schreiben $U = \{av + bw \mid a, b \in K\}$ für die von v, w „aufgespannte“ Ebene (die Terminologie wird später klarer). Weiter sei $A = \binom{v^\top}{w^\top} \in K^{2\times 3}$.

Zeigen Sie:

a) A hat Rang 2 und es gibt genau eine Fundamentallösung $n \in K^3$ des homogenen Gleichungssystems $Ax = 0$.

b) Es gibt genau zwei Fundamentallösungen des Gleichungssystems $n^\top x = 0$, und
$$\mathcal{L}(n^\top, 0) = U.$$

c) Für festes $z \in K^3$ und $z + U = \{z + u \mid u \in U\}$ gilt:
$$z + U = \{x \in K^3 \mid n^\top x = n^\top z\}.$$

Tipp: Schauen Sie für die letzte Aussage noch einmal in die Nummer 4.1.2.

Was wir hier lernen, ist die Sicht der linearen Algebra für die **Hessesche Normalform** für affine Ebenen im dreidimensionalen Raum. Das hat auch die Benennung n der Lösung motiviert: geometrisch ist das ein Normalenvektor, also ein Vektor, der auf U senkrecht steht.

Wir können das später mit mehr Theorie noch einmal eleganter ausdrücken und vor allem deutlich verallgemeinern.

Beispiel 4.4.6 (praktische Durchführung des Gauß-Verfahrens)

Wir möchten die Matrix A genau so auf Treppenform bekommen, wie es einige in der Schule gelernt haben. Dabei war wahrscheinlich gar nicht von invertierbaren Matrizen die Rede, sondern die Zeilenumformungen wurden eben „nur“ als Zeilenumformungen eingeführt.

Um sich zu merken, welche Matrix man dabei insgesamt an A von links dranmultipliziert, führt man oft zur Buchhaltung die größere Matrix $\widetilde{A} := (A|I_p)$ mit, die aus A durch Anhängen der $p \times p$-Einheitsmatrix entsteht. Dann gilt $C \cdot (A|I_p) = (C \cdot A|C)$. Am Ende aller Zeilenumformungen, die man nun mit $\widetilde{A}$ durchführt, bis A Treppenform hat, steht also rechts vom „Trennstrich“ die Matrix C, die man dabei benützt hat. Diese ist wie gesagt im Allgemeinen nicht eindeutig, und man wird oft beim Rechnen von Hand einen anderen Weg gehen als den, der sich im Beweis von 4.4.2 als allgemein möglich herausstellte.

Nun also ein Zahlenbeispiel. Dazu sei

$$A := \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 5 \\ 1 & 2 & 5 & 6 \end{pmatrix} \in \mathbb{R}^{3\times 4}.$$

Abziehen der ersten Zeile von der zweiten und dritten führt zu

$$\left(\begin{array}{cccc|ccc} 1 & 2 & 3 & 4 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 2 & 2 & -1 & 0 & 1 \end{array}\right),$$

wobei wir schon die entsprechende Additionsmatrix neben dem senkrechten Strich vermerkt haben. Unterhalb der ersten Zeile steht nun der erste von Null verschiedene Eintrag in der dritten Spalte, aber nicht in der zweiten Zeile, sondern in der dritten. Durch Vertauschung dieser zwei Zeilen erhalten wir

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

Der erste Eintrag in der zweiten Zeile soll für die Treppenform ja 1 und nicht 2 sein, also sollten wir durch 2 teilen:

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

Daran sehen wir schon, dass die Treppenform Rang 3 haben wird, und dass die Stufenindizes 1, 3 und 4 sind. Aber die dritte und vierte Spalte sind noch nicht e_2 und e_3. Dazu ziehen wir nun geeignete Vielfache der dritten Zeile von den ersten beiden ab, nämlich das Vierfache von der ersten und das Einfache von der zweiten:

$$\left(\begin{array}{cccc|ccc} 1 & 2 & 3 & 0 & 5 & -4 & 0 \\ 0 & 0 & 1 & 0 & 1/2 & -1 & 1/2 \\ 0 & 0 & 0 & 1 & -1 & 1 & 0 \end{array}\right).$$

Schließlich ziehen wir das Dreifache der zweiten Zeile von der ersten ab und erhalten links vom Strich eine Treppenform:

$$\left(\begin{array}{cccc|ccc} 1 & 2 & 0 & 0 & 7/2 & -1 & -3/2 \\ 0 & 0 & 1 & 0 & 1/2 & -1 & 1/2 \\ 0 & 0 & 0 & 1 & -1 & 1 & 0 \end{array}\right).$$

Rechts vom Strich steht nun die Matrix

$$C := \begin{pmatrix} 7/2 & -1 & -3/2 \\ 1/2 & -1 & 1/2 \\ -1 & 1 & 0 \end{pmatrix}.$$

Diese ist invertierbar, da sie als Produkt von Additions-, Vertauschungs- und (invertierbaren) Diagonalmatrizen entstanden ist. Es gilt

$$C \cdot A = \begin{pmatrix} 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Da der Rang von A gleich $3 = 4 - 1$ ist, liefert 4.3.3 genau eine Fundamentallösung, nämlich

$$F^{(2)} := \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

Ist nun zum Beispiel

$$A \cdot x = e_1$$

zu lösen, so ersetzen wir dies durch

$$CA \cdot x = Ce_1 = \begin{pmatrix} 7/2 \\ 1/2 \\ -1 \end{pmatrix},$$

und lesen nach 4.3.3 die spezielle Lösung ab:

$$x^{(s)} = \begin{pmatrix} 7/2 \\ 0 \\ 1/2 \\ -1 \end{pmatrix}.$$

Dann gilt

$$\mathcal{L}(A, e_1) = \{x^{(s)} + tF^{(2)} \mid t \in \mathbb{R}\}.$$

Folgerung 4.4.7 (Regularität und Rang)

Es seien K ein Körper und $A \in K^{p\times p}$. Dann sind äquivalent:

i) A *ist regulär.*

ii) $\mathrm{Rang}\,(A) = p$.

iii) $\exists S \in K^{p\times p} : AS = I_p$.

Beweis.

i)$\Rightarrow$ ii) Wenn A invertierbar ist, dann ist $A^{-1} \cdot A = I_p$ die Treppenform, die zu A gehört, und diese hat Rang p. Nach Definition hat also auch A Rang p..

ii)$\Rightarrow$iii) Wenn A Rang p hat, dann wähle ein $C \in \mathrm{GL}_p(K)$, für das CA Treppenform hat. Die einzige $p \times p$-Treppenform vom Rang p ist die Einheitsmatrix I_p. Also ist $A = C^{-1}$ und damit auch $AC = I_p$.

iii)$\Rightarrow$i) Aus $AS = I_p$ folgt, dass $A(Sb) = b$ für alle $b \in K^p$. Daher ist $Ax = b$ für alle b lösbar, und A hat nach dem oben gelernten Rang p. Daher gibt es ein invertierbares C mit $CA = I_p$, und nach Multiplikation mit C^{-1} von links folgt $A = C^{-1} \in \mathrm{GL}_p(K)$. ◯

Insbesondere liefert damit das Gauß-Verfahren eine Möglichkeit, für eine reguläre Matrix A die Inverse A^{-1} zu berechnen:

Konstruktion 4.4.8 Invertieren einer Matrix

Es sei $A \in K^{p\times p}$ eine Matrix. Dann ist A genau dann invertierbar, wenn der Rang von A gleich p ist. Wenn wir das Gaußverfahren für die Matrix

$$(A|I_p) \in K^{p\times 2p}$$

durch führen, bis links vom Strich die einheitsmatrix steht, dann steht rechts vom Strich die Inverse A^{-1}.

Aufgabe 4.4.9 (Ein Beispiel)

Seien K ein Körper und $a, b, c, d \in K$. Zeigen Sie, dass die Matrix

$$A = \begin{pmatrix} 0 & 0 & 0 & -a \\ 1 & 0 & 0 & -b \\ 0 & 1 & 0 & -c \\ 0 & 0 & 1 & -d \end{pmatrix} \in K^{4\times 4}$$

genau dann invertierbar ist, wenn $a \neq 0$, und bestimmen Sie in diesem Fall die inverse Matrix A^{-1}.

Rechnen Sie nach, dass

$$A^4 + dA^3 + cA^2 + bA + aA^0 = 0.$$

Dabei ist A^0 die Einheitsmatrix.

Diese Matrix A heißt die **Begleitmatrix zum Polynom** $X^4 + dX^3 + cX^2 + bX + a$.

Kapitel 5

Vektorräume

Endlich kommen wir zu den zentralen Objekten der Linearen Algebra. Wir kennen schon viele Beispiele von Vektorräumen, ohne aber das Wesentliche, was allen gemeinsam ist, herausdestilliert zu haben. Das wird nun passieren.

5.1 Grundlegende Definitionen

Definition 5.1.1 (Vektorraum)

Es sei K ein Körper. Ein **Vektorraum über dem Körper** K (kurz auch: K-Vektorraum) ist eine kommutative Gruppe $(V, +)$, für die zusätzlich eine Abbildung

$$\cdot : K \times V \longrightarrow V, \quad (a, v) \mapsto a \cdot v$$

definiert ist (die **skalare Multiplikation**), sodass die folgenden Bedingungen erfüllt sind:

- Für alle $v \in V$ gilt $1_K \cdot v = v$.
- Für alle $a, b \in K$ und alle $v \in V$ gilt $a \cdot (b \cdot v) = (a \cdot b) \cdot v$.
- Für alle $a, b \in K$ und alle $u, v \in V$ gilt

$$\begin{aligned} a \cdot (u + v) &= a \cdot u + a \cdot v, \\ (a + b) \cdot v &= a \cdot v + b \cdot v. \end{aligned}$$

Hierbei verwenden wir wieder die Regel „Punkt vor Strich“.

Bemerkung 5.1.2 Während explizit gefordert ist, dass $1_K \cdot v = v$ für alle $v \in V$, wird über $0_K \cdot v$ nichts ausgemacht. Das ist aber auch gar nicht nötig, denn mit $n := 0_K \cdot v$ gilt:

$$n = 0_K \cdot v = (0_K + 0_K) \cdot v = 0_K \cdot v + 0_K \cdot v = n + n.$$

Wir haben also $n = n + n$, und wenn wir hier links und rechts $-n$ addieren – das gehört zur additiven Grupenstruktur von V –, so folgt

$$n = 0_V.$$

Wir haben also bewiesen, dass in einem K-Vektorraum V gilt:

$$\forall v \in V : 0_K \cdot v = 0_V.$$

Im Weiteren werden wir häufig sowohl für 0_K als auch für 0_V und für alle möglichen weiteren Nullen nur 0 schreiben und hoffen, dass der Kontext klärt, welche Null jeweils gemeint ist.

Beispiel 5.1.3 (mein erster Vektorraum)

Es sei K ein Körper.

a) K selbst ist bezüglich der Addition eine abelsche Gruppe. Wenn man die Körpermultiplikation als skalare Multiplikation benutzt, kann K selbst als K-Vektorraum aufgefasst werden. Dies sollten wir auch immer wieder tun.

b) Für eine natürliche Zahl d ist $V := K^d$ ein K-Vektorraum, mit der skalaren Multiplikation

$$\cdot : K \times V \longrightarrow V, \quad (a, \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}) \mapsto a \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} := \begin{pmatrix} a \cdot x_1 \\ \vdots \\ a \cdot x_d \end{pmatrix}$$

aus 4.1.13. Die Axiome für die skalare Multiplikation folgen unmittelbar aus dem Assoziativ- und Distributivgesetz von K (siehe 3.1.1, K ist ja auch ein Ring), da Komponente für Komponente die skalare Multiplikation aus a) wirkt.

Spezialfälle sind die $\mathbb{R}$-Vektorräume $\mathbb{R}^2$ („Tafelebene“) und $\mathbb{R}^3$ („Anschauungsraum“).

c) Noch allgemeiner (siehe 1.3.15 warum das allgemeiner ist!) sei M eine beliebige Menge und $V := \mathrm{Abb}(M, K)$. Dann ist V durch „punktweise“ Addition eine kommutative Gruppe:

$$\forall f, g \in V, \forall m \in M : (f + g)(m) := f(m) + g(m).$$

Nun wird die skalare Multiplikation wie folgt eingeführt:

$$\forall a \in K,\ f \in V : \forall m \in M : (a \cdot f)(m) := a \cdot (f(m)).$$

Dadurch ist offensichtlich $a \cdot f$ wieder eine Abbildung von M nach K, also ein Element von V. Die geforderten Eigenschaften der skalaren Multiplikation folgen wieder aus den Ringeigenschaften von K.

d) Schließlich sei W ein beliebiger K-Vektorraum und M eine beliebige Menge. Dann ist $V := \mathrm{Abb}(M, W)$ ein K-Vektorraum, wobei die Addition und skalare Multiplikation wieder „punktweise“, also für jedes $m \in M$ einzeln, festgelegt wird (genauso wie in c)).

e) Ein etwas exotischer anmutendes Beispiel für einen Vektorraum ist $\mathbb{R}$ als $\mathbb{Q}$-Vektorraum, wobei die Addition die übliche ist, und die skalare Multiplikation die Einschränkung der üblichen Multiplikation von $\mathbb{R} \times \mathbb{R}$ nach $\mathbb{Q} \times \mathbb{R}$ ist.

f) Allgemeiner ist jeder Ring A, der einen Körper K als Teilring enthält, ein K-Vektorraum durch Einschränkung der Multiplikation von $A \times A$ nach $K \times A$. Dies gilt zum Beispiel für den Polynomring $K[X]$ und für den Matrizenring $K^{n\times n}$.

Um noch mehr Beispiele von Vektorräumen konstruieren zu können (und weil es inhaltlich ohnehin wichtig ist), führen wir als nächstes den Begriff des Untervektorraums ein.

Definition 5.1.4 (Untervektorraum)

Es seien K ein Körper und V ein K-Vektorraum. Ein (K-)**Untervektorraum** von V ist eine Teilmenge $U \subseteq V$, die bezüglich der Addition eine Untergruppe von V ist und für die gilt:

$$\forall a \in K, u \in U : a \cdot u \in U.$$

Dann ist also U selber auch ein Vektorraum. Um U von beliebigen Teilmengen von V in der Notation zu unterscheiden, schreiben wir oft $U \leq V$.

Hilfssatz 5.1.5 (Untervektorraumkriterium)

Es seien K ein Körper, V ein K-Vektorraum, und $U \subseteq V$. Dann sind die folgenden zwei Aussagen äquivalent:

i) U ist ein Untervektorraum von V.

ii) U ist nicht leer, $\forall u_1, u_2 \in U : u_1 + u_2 \in U$ und $\forall a \in K, u \in U : a \cdot u \in U$.

Beweis. Wir zeigen, dass jede der Aussagen die jeweils andere impliziert.

$i) \Rightarrow ii)$: Zunächst nehmen wir also an, U sei ein Untervektorraum. Dann ist U eine Untergruppe von V, also nicht leer. Außerdem ist U eine Untergruppe, also unter Addition abgeschlossen. Die dritte Bedingung aus $ii)$ wird in der Definition von Untervektorraum explizit auch gefordert. Demnach gilt $ii)$.

$ii) \Rightarrow i)$: Nun ist U nicht leer. Mit $u \in U$ ist auch $(-1) \cdot u \in U$, aber das ist additiv invers zu u, denn mit 5.1.2 folgt

$$u + (-1) \cdot u = (1 + (-1)) \cdot u = 0 \cdot u = 0.$$

Also ist mit u auch $-u$ in U. Dann sind aber mit u_1, u_2 auch $u_1, -u_2$ in U, und damit wegen der Annahme $ii)$ auch $u_1 + (-u_2) = u_1 - u_2$.

Nach dem Untergruppenkriterium 2.2.3 ist also U eine Untergruppe. Wegen des Rests von $ii)$ ist es dann sogar ein Untervektorraum. ◯

Beispiel 5.1.6 (Lineare Codes)

Eine Anwendung von Untervektorräumen in der Informatik findet sich bei **Linearen Codes**.

Grundsätzlich ist in der Codierungstheorie ein endliches Alphabet A gegeben, und Ziel ist es, Wörter der Länge d zu übermitteln. Um Übertragungsfehler behandeln zu können, hat sich folgende Idee etabliert: Wir schauen nicht alle Elemente von A^d an, sondern definieren eine Teilmenge von A^d als erlaubte Wörter. Wenn sich dabei je zwei verschiedene erlaubte Wörter an mindestens drei Stellen unterscheiden, dann kann eine fehlerhafte Übertragung, bei der nur eine Stelle falsch übermittelt wurde, korrigiert werden, denn das ursprüngliche Wort ist das einzige erlaubte Wort, das sich von dem übermittelten nur an einer Stelle unterscheidet.

Wenn sich je zwei verschiedene erlaubte Wörter an mindestens 5 Stellen unterscheiden, können wir sogar zwei fehlerhafte Stellen korrigieren – und so weiter.

Wenn nun A ein endlicher Körper ist, wie etwa $\mathbb{F}_p$ für eine Primzahl p, und wenn U ein Untervektorraum von A^d ist, dann gibt es aufgrund der algebraischen Zusatzstrukturen praktikable Verfahren zur Fehlerbehebung. Dazu brauchen wir vor allem gute Möglichkeiten, Untervektorräume zu beschreiben. Wir kommen darauf in 5.5.14 zurück.

Bemerkung 5.1.7 (Durchschnitt von Untervektorräumen)

Es seien V ein K-Vektorraum, I eine nichtleere Menge, und für jedes $i \in I$ sei U_i ein Untervektorraum von V. Dann ist

$$\bigcap_{i \in I} U_i$$

ein Untervektorraum von V. Der Nachweis geht ähnlich wie in 2.2.6 für Untergruppen.

Beispiel 5.1.8 (noch mehr Vektorräume)

a) Für jede Matrix $A \in K^{p\times q}$ ist die Lösungsmenge $\mathcal{L}(A,0)$ des zugehörigen homogenen linearen Gleichungssystems ein Vektorraum. Denn es ist eine Teilmenge des Vektorraums K^q, für die die Bedingungen aus Teil *ii)* des Untervektorraumkriteriums 5.1.5 erfüllt sind.

Die Menge aller $b \in K^p$, für die $\mathcal{L}(A,b)$ nicht leer ist, ist ein Untervektorraum von K^p.

b) Die Menge aller stetigen Abbildungen von einem (festen) Intervall I nach $\mathbb{R}$ ist ein $\mathbb{R}$-Untervektorraum von $\mathrm{Abb}(I,\mathbb{R})$.

c) Nun sei $V = \mathbb{R}^2$. Was sind die Untervektorräume dieses $\mathbb{R}$-Vektorraums?

Wie immer, so gibt es auch hier die Untervektorräume $\{0\}$ und V. Welche sonst noch? Es sei $U \subseteq V$ ein Untervektorraum, der keiner der beiden genannten ist.

Dann liegt in U sicher ein von 0 verschiedener Vektor $u = \binom{a}{b}$, und mit ihm auch alle Vielfachen, also

$$\mathbb{R}\cdot u := \{r\cdot u \mid r \in \mathbb{R}\} \subseteq U.$$

Es ist klar, dass $\mathbb{R}\cdot u$ selbst schon ein Untervektorraum ist. Nun ist die Behauptung, dass $U = \mathbb{R}\cdot u$. Dazu müssen wir noch $U \subseteq \mathbb{R}\cdot u$ zeigen, und nehmen dazu an, wir hätten einen Vektor $v = \binom{c}{d} \in U \setminus \mathbb{R}\cdot u$.

Im Fall $a \neq 0$ folgt hier $d \neq bc/a$, da sonst ja

$$v = \begin{pmatrix} c \\ d \end{pmatrix} = \frac{c}{a}\begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}\cdot u.$$

Analog folgt für $b \neq 0$, dass $c \neq ad/b$.

Da aber a und b nicht beide Null sein können (wegen $u \neq 0$) folgt insgesamt auf jeden Fall

$$ad - bc \neq 0.$$

Die Matrix $A = \left(\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}\right)$ ist dann regulär, denn (siehe 4.2.9) $A^{-1} := \frac{1}{ad-bc}\cdot \left(\begin{smallmatrix} d & -c \\ -b & a \end{smallmatrix}\right)$ ist dazu invers. Für ein $w \in V$ sei $A^{-1}w = \binom{s}{t}$. Dann gilt

$$s\cdot u + t\cdot v = A\cdot \begin{pmatrix} s \\ t \end{pmatrix} = A\cdot A^{-1}\cdot w = w.$$

Also ist w in jedem Untervektorraum von $\mathbb{R}^2$, der u und v enthält.

Das heißt aber: sobald im Untervektorraum $U \subseteq V = \mathbb{R}^2$ noch ein Vektor außerhalb $\mathbb{R}\cdot u$ liegt, muss $U = V$ gelten, was ausgeschlossen war.

Im Vektorraum $\mathbb{R}^2$ gibt es genau die folgenden Untervektorräume:

$$\{0\}, \mathbb{R}^2, \text{ und } \mathbb{R}\cdot u, \text{ wobei } u \in \mathbb{R}^2 \setminus \{0\}.$$

Insbesondere hat $\mathbb{R}^2$ unendlich viele Untervektorräume. Die letztgenannten in der Liste heißen „Geraden“, auf Lateinisch also „linea“, was die Hälfte des Namens des Buches erklärt. Die andere Hälfte kommt aus dem Arabischen und bedeutet so etwas wie ergänzen oder heilen.

Die Diskussion des letzten Beispiels legt es nahe, für eine endliche Teilmenge $\{v_1, \ldots, v_q\}$ des Vektorraums V die folgende Menge anzusehen:

$$\langle\{v_1, \ldots, v_q\}\rangle := \{\sum_{i=1}^{q} \alpha_i v_i \mid \alpha_i \in K\}.$$

Es ist klar, dass dies ein Untervektorraum von V ist, und zwar der kleinste Untervektorraum von V, der $\{v_1, \ldots, v_q\}$ enthält. Allgemeiner definiert man das Erzeugnis wie folgt:

Definition/Bemerkung 5.1.9 (Erzeugnis, Linearkombination, Träger)

Es sei $M \subseteq V$ Teilmenge eines K-Vektorraums V. Dann heißt für eine Abbildung $\alpha : M \longrightarrow K$, die für alle bis auf endlich viele Elemente $m \in M$ den Wert 0 annimmt, die (endliche!) Summe

$$\sum_{m \in M} \alpha(m) \cdot m \in V$$

eine **Linearkombination** von M.

Die Menge $\langle M\rangle$ aller Linearkombinationen von M heißt das **(Vektorraum-)Erzeugnis** oder auch die **lineare Hülle** von M.

M heißt ein **Erzeugendensystem** von $\langle M\rangle$.

Die Bedingung, dass α für alle bis auf endlich viele Elemente $m \in M$ den Wert 0 annimmt, formulieren wir in Zukunft kürzer so: α hat endlichen Träger. Dabei ist der **Träger** von $\alpha \in \mathrm{Abb}(M, K)$ die Menge aller $m \in M$ mit $\alpha(m) \neq 0$.

Sehr oft wird bei uns die Menge K endlich sein. Wenn wir dann mit

$$M = \{m_1, \ldots, m_n\}$$

alle Elemente erfasst haben, dann ist eine Linearkombination von M einfach eine Summe der Gestalt

$$c_1 m_1 + c_2 m_2 + \ldots + c_n m_n, \quad \text{wobei alle } \; c_i \in K.$$

In diesem Sinn ist eine Linearkombination einer beliebigen Menge einfach eine Linearkombination einer (geeigneten) endlichen Teilmenge.

Bemerkung 5.1.10 (Vorsicht ist geboten)

a) **Vorsicht**: Vergleiche die Notation $\langle M \rangle$ mit der aus Definition 2.2.7. Aus dem Kontext wird hoffentlich immer klar sein, wann das Gruppenerzeugnis und wann das Vektorraumerzeugnis gemeint ist. Ein Ausweg wäre es, für das Vektorraum-Erzeugnis von M die Notation $\langle M \rangle_{K-\mathrm{VR}}$ zu benutzen.

b) $\langle M \rangle$ ist der kleinste Untervektorraum von V, der M enthält. Es gilt

$$\langle M \rangle = \bigcap_{M \subseteq U \leq V} U.$$

Das ist der Durchschnitt aller Untervektorräume von V, die M als Teilmenge enthalten.

c) Für eine Menge M und einen Körper K ist

$$\mathrm{Abb}(M,K)_0 := \{f \in \mathrm{Abb}(M,K) \mid f \text{ hat endlichen Träger}\}$$

ein Untervektorraum von $\mathrm{Abb}\,(M,K)$. Er wird erzeugt von der Menge aller Funktionen mit einelementigem Träger. Genauer bezeichnen wir für $m \in M$ mit δ_m die Abbildung

$$\delta_m : M \to K, \quad \delta_m(x) = \begin{cases} 1, & \text{falls } x = m, \\ 0, & \text{sonst.} \end{cases}$$

Dann lässt sich $f \in \mathrm{Abb}(M,K)_0$ schreiben als

$$f = \sum_{m \in M} f(m) \cdot \delta_m.$$

Diese Summe ist endlich, da ja nur endlich viele Funktionswerte nicht 0 sind. Wir haben also eine Linearkombination der δ_m vorliegen.

Es gilt

$$\mathrm{Abb}(M,K)_0 = \mathrm{Abb}(M,K) \iff \#M < \infty.$$

d) Wenn M nur aus einem Element m besteht, so schreiben wir statt $\langle \{m\} \rangle$ suggestiver $K \cdot m$ (wie in Beispiel 5.1.8c)). Das ist die Menge aller Vielfachen von m.

Beispiel 5.1.11 (Die Erzeugendensysteme von K^p)

Da der K-Vektorraum K^p das endliche Erzeugendensystem $\{e_1, \ldots, e_p\}$ hat, enthält jedes Erzeugendensystem von K^p auch ein endliches Erzeugendensystem, denn wir brauchen nur endlich viele seiner Elemente, um $e_1, \ldots, e_p$ als Linearkombination zu schreiben, und diese langen dann auch für den gesamten Vektorraum.

Wie sieht ein endliches Erzeugendensystem von K^p aus? Oder besser gefragt: Wie teste ich, ob eine endliche Menge von Vektoren in K^p diesen Raum erzeugt?

Es seien $v_1, \ldots, v_q \in K^p$ und $A \in K^{p\times q}$ die Matrix mit den Spalten $v_1, \ldots, v_q$. Dann gilt

$$K^p = \langle\{v_1, \ldots, v_q\}\rangle \iff \text{Rang}(A) = p.$$

Denn die Bedingung auf der rechten Seite ist nach 4.4.4 gleichbedeutend damit, dass für jedes $u \in K^p$ das LGS

$$A \cdot \lambda = u$$

eine Lösung λ hat, u sich also als $u = \sum_{i=1}^{q} \lambda_i v_i$ schreiben lässt.

Aufgabe 5.1.12 (Zahlenbeispiel)

Es sei $K = \mathbb{F}_5$ der Körper mit 5 Elementen, 3.2.1. In $V = \mathbb{F}_5^4$ seien die Vektoren

$$\begin{pmatrix} 1_K \\ 2_K \\ 0_K \\ 4_K \end{pmatrix}, \begin{pmatrix} 3_K \\ 1_K \\ 0_K \\ a \end{pmatrix}, \begin{pmatrix} 0_K \\ 2_K \\ 1_K \\ b \end{pmatrix}, \begin{pmatrix} 0_K \\ 1_K \\ 3_K \\ 0_K \end{pmatrix}$$

gegeben, wobei $a, b \in \mathbb{F}_5$ Parameter sind und x_K für die Restklasse von x modulo 5 steht.

Ermitteln Sie, für welche Wahlen von $a, b \in K$ diese Vektoren den Vektorraum K^4 erzeugen.

Definition 5.1.13 (Summe von Untervektorräumen)

Es sei V ein K-Vektorraum, I eine Indexmenge und für jedes $i \in I$ sei ein Untervektorraum U_i von V gegeben. Dann ist

$$\sum_{i\in I} U_i := \langle \bigcup_{i\in I} U_i \rangle$$

die **Summe** der U_i, $i \in I$. Das ist der kleinste Untervektorraum, der alle U_i enthält.

Speziell gilt für $I = \{1, \ldots, n\}$:

$$\sum_{i=1}^{n} U_i = \{u_1 + u_2 + \cdots + u_n \mid \forall i : u_i \in U_i\}.$$

Vorsicht: In den wenigsten Fällen gilt $\sum U_i = \cup U_i$.

Aufgabe 5.1.14 (Summe und Vereinigung)

Es seien K ein Körper, V ein K-Vektorraum und $U_1, U_2 \leq V$ Untervektorräume.

Weisen Sie nach, dass

$$U_1 \cup U_2 = U_1 + U_2 \Leftrightarrow [U_1 \subseteq U_2 \text{ oder } U_2 \subseteq U_1].$$

5.2 Homomorphismen

Wie bei Gruppen, so wollen wir auch zwischen Vektorräumen Abbildungen betrachten, die die Vektorraumstrukturen „respektieren“. Präziser sieht das so aus:

Definition 5.2.1 (Vektorraumhomomorphismus)

Es seien K ein Körper und V, W zwei K-Vektorräume. Ein **K-Vektorraum-Homomorphismus** (meistens einfach **Homomorphismus** oder auch (K-)**lineare Abbildung** genannt) von V nach W ist eine Abbildung $\Phi : V \longrightarrow W$, für die gilt:

$$\begin{aligned} \forall u, v \in V: &\quad \Phi(u+v) = \Phi(u) + \Phi(v), \\ \forall a \in K, v \in V: &\quad \Phi(av) = a\Phi(v). \end{aligned}$$

Insbesondere ist ein Vektorraumhomomorphismus also auch ein Gruppenhomomorphismus von $(V, +)$ nach $(W, +)$.

Einen bijektiven Homomorphismus nennt man einen **Isomorphismus**, einen Homomorphismus von V nach V einen **Endomorphismus** von V, und einen bijektiven Endomorphismus nennt man einen **Automorphismus** von V.

Notationen: $\mathrm{Hom}(V, W)$ ist die Menge aller Homomorphismen von V nach W (wieder wäre $\mathrm{Hom}_{K-\mathrm{VR}}(V, W)$ präziser, wenngleich mühsamer), $\mathrm{End}(V)$ ist die Menge aller Endomorphismen von V, und $\mathrm{Aut}(V)$ die Menge aller Automorphismen. Wenn zwei Vektorräume V und W isomorph sind (also wenn es mindestens einen Isomorphismus zwischen ihnen gibt), dann schreiben wirdafür $V \cong W$. Isomorphie ist eine Äquivalenzrelation (auf jeder Menge von K-Vektorräumen). $\mathrm{Aut}(V)$ ist eine Gruppe bezüglich der Komposition von Abbildungen.

Beispiel 5.2.2 (für Homomorphismen)

a) Für eine Matrix $A \in K^{p\times q}$ ist (siehe 4.1.5)

$$\Phi_A : K^q \longrightarrow K^p, \quad v \mapsto \Phi_A(v) := A \cdot v,$$

eine K-lineare Abbildung. Das folgt sofort aus den Rechenregeln für die Matrizenmultiplikation und -addition (siehe den Beweis von 5.2.5).

b) Es sei V ein K-Vektorraum. Für zwei Untervektorräume U und W von V ist

$$U \times W := \{(u, w) \mid u \in U,\ w \in W\}$$

ein Vektorraum mit komponentenweiser Addition und Skalarmultiplikation. Dann ist die Abbildung

$$\alpha : U \times W \longrightarrow V, \quad \forall (u, w) \in U \times W : \alpha((u, w)) := u - w,$$

ein Homomorphismus von Vektorräumen.

Der Kern von α (siehe 2.3.4 und 5.2.3) ist

$$\begin{aligned}\mathrm{Kern}(\alpha) &= \{(u,w) \in U \times W \mid \alpha((u,w)) = 0\} \\ &= \{(u,w) \in U \times W \mid u - w = 0\} \\ &= \{(u,w) \in U \times W \mid u = w\} \\ &= \{(v,v) \mid v \in U \cap W\},\end{aligned}$$

und dies ist via $(v,v) \mapsto v$ isomorph zu $U \cap W$.

Das Bild von α ist

$$\mathrm{Bild}(\alpha) = \{\alpha(u,w) \mid u \in U, w \in W\} = \{u - w \mid u \in U, w \in W\} = U + W.$$

c) Die Abbildung $^\top : K^{p\times q} \longrightarrow K^{q\times p}$, die eine Matrix auf ihre transponierte abbildet (siehe 4.1.13), ist K–linear und bijektiv; sie ist zu sich selbst invers.

d) Wenn V der Vektorraum aller unendlich oft differenzierbaren, reellwertigen Funktionen auf $\mathbb{R}$ ist, dann ist die Ableitung

$$D : V \to V,\ f \mapsto Df = f',$$

ein Endomorphismus.

Bemerkung 5.2.3 (Kern eines Homomorphismus)

Für einen Homomorphismus $\Phi : V \longrightarrow W$ ist der Kern wie in 2.3.4 definiert als

$$\mathrm{Kern}(\Phi) := \{v \in V \mid \Phi(v) = 0\} = \Phi^{-1}(\{0\}).$$

Dies ist der Kern von Φ aufgefasst als Homomorphismus zwischen den additiven Gruppen $(V,+)$ und $(W,+)$, und daher gilt wieder

$$\boxed{\Phi \text{ ist injektiv genau dann, wenn Kern}\,(\Phi) = \{0\}.}$$

Der Kern ist sogar ein Untervektorraum von V.

Für eine Matrix $A \in K^{p\times q}$ ist der Kern von Φ_A (siehe 4.1.5)genau der Lösungsraum $\mathcal{L}(A,0)$ des homogenen Gleichungssystems $Ax = 0$.

Bemerkung 5.2.4 (Der Vektorraum aller Homomorphismen)

$\mathrm{Hom}\,(V,W)$ ist ein Untervektorraum von $\mathrm{Abb}\,(V,W)$ (siehe 5.1.3 für die Vektorraumstruktur hiervon).

Denn: $\mathrm{Hom}\,(V,W)$ ist nicht leer, da die Nullabbildung (die alles auf $0 \in W$ abbildet) offensichtlich dazu gehört. Weiter gilt für alle $\Phi, \Psi \in \mathrm{Hom}(V,W)$:

$$\begin{aligned}\forall u, v \in V:\ & (\Phi+\Psi)(u+v) = \Phi(u+v) + \Psi(u+v) = \\ & = \Phi(u) + \Phi(v) + \Psi(u) + \Psi(v) = (\Phi+\Psi)(u) + (\Phi+\Psi)(v), \\ \forall a \in K, v \in V:\ & (\Phi+\Psi)(av) = \cdots = a(\Phi+\Psi)(v).\end{aligned}$$

Also liegt auch $\Phi + \Psi$ in $\mathrm{Hom}\,(V,W)$. Ähnlich rechnet man nach, dass für alle $a \in K$ und $\Phi \in \mathrm{Hom}(V,W)$ auch $a\Phi \in \mathrm{Hom}(V,W)$. Dann liefert aber das Untervektorraumkriterium 5.1.5 die gewünschte Aussage.

Wie sieht $\mathrm{Hom}\,(V,W)$ für die Vektorräume aus, die wir am besten kennen, also für die Räume K^p? Noch einmal erinnern wir an Φ_A aus Beispiel 5.2.2a). Es gilt:

Hilfssatz 5.2.5 ($\mathbf{Hom}\,(K^q,K^p) \cong K^{p\times q}$)

Es seien p,q natürliche Zahlen und K ein Körper. Dann ist die Abbildung

$$K^{p\times q} \ni A \mapsto \Phi_A \in \mathrm{Hom}(K^q,K^p)$$

ein Isomorphismus von K-Vektorräumen.

Beweis. Zunächst müssen wir klären, dass für jede Matrix A die Abbildung Φ_A tatsächlich ein Homomorphismus ist. Das ist aber nach den Rechenregeln für die Matrizenmultiplikation (siehe 4.1.11 und 4.1.13) klar:

$$\begin{aligned} \forall v_1,v_2 \in K^q: \quad & \Phi_A(v_1+v_2) = A\cdot(v_1+v_2) = A\cdot v_1 + A\cdot v_2 \\ & = \Phi_A(v_1)+\Phi_A(v_2), \\ \forall a\in K, v\in K^q: \quad & \Phi_A(av) = A\cdot(av) = a\cdot Av = a\cdot\Phi_A(v). \end{aligned}$$

Außerdem ist klar, dass $A \mapsto \Phi_A$ eine injektive Abbildung ist, denn für $1 \le i \le q$ ist $\Phi_A(e_i)$ die i-te Spalte von A (siehe 4.2.4). Also stehen verschiedene Matrizen für verschiedene Homomorphismen.

Dies legt aber auch nahe, wie die Umkehrabbildung auszusehen hat. Dazu sei $\Psi \in \mathrm{Hom}(K^q,K^p)$ beliebig und für $1 \le i \le q$ sei $v_i := \Psi(e_i) \in K^p$. Weiter sei A die Matrix

$$A := (v_1\ v_2\ \ldots\ v_q) \in K^{p\times q}.$$

Jedes $x \in K^q$ lässt sich schreiben als

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_q \end{pmatrix} = \sum_{j=1}^{q} x_j e_j,$$

und das führt zu

$$\Psi(x) = \Psi(\sum_{j=1}^{q} x_j e_j) = \sum_{j=1}^{q} \Psi(x_j e_j) = \sum_{j=1}^{q} x_j \Psi(e_j) = \sum_{j=1}^{q} x_j v_j = \sum_{j=1}^{q} x_j A e_j = A\cdot x.$$

Also folgt $\Psi = \Phi_A$.

Wir müssen nun nur noch zeigen, dass die Zuordnung $A \mapsto \Phi_A$ ein K-lineare Abbildung ist.

Dazu seien $A, B \in K^{p\times q}$. Aus 4.1.11 kennen wir die Gleichheit $\Phi_{A+B} = \Phi_A + \Phi_B$, die wir brauchen.

Weiter seien $a \in K$ und $A \in K^{p\times q}$. Dann gilt für alle $x \in K^q$:

$$\Phi_{aA}(x) = (aA) \cdot x = a \cdot (Ax) = a\Phi_A(x),$$

und daher auch $\Phi_{aA} = a\Phi_A$.

Damit ist alles gezeigt. ◯

5.3 Basen

Ein abstrakter Vektorraum ist häufig für konkrete Rechnungen nicht ohne Weiteres zugänglich. Um diese zu bewerkstelligen brauchen wir meistens einen Standardvektorraum, der zu ihm isomorph ist und – wenn es hart auf hart kommt – einen Isomorphismus.

Eigentlich ist genau das der Grund, den Begriff der Basis einzuführen. Später werden wir lernen, wie „geeignete“ Basen für spezielle Untersuchungen gewählt werden können, um sich von einer einmal getroffenen Wahl wieder zu befreien, wenn diese sich als ungeschickt erweisen sollte.

Definition 5.3.1 (Basis)

Es seien K ein Körper und V ein K-Vektorraum. Eine Teilmenge $B \subseteq V$ heißt eine **Basis** von V, falls sich jeder Vektor $v \in V$ auf genau eine Art als Linearkombination von B schreiben lässt:

$$\forall v \in V : \ \exists_1 \lambda \in \mathrm{Abb}(B, K)_0 : v = \sum_{b\in B} \lambda(b) \cdot b.$$

Beispiel 5.3.2 (Bezug zum LGS)

Es sei $A \in K^{p\times q}$ eine Matrix. Wir haben in Abschnitt 4.4 gelernt, dass der Lösungsraum $\mathcal{L}(A, 0)$ berechnet werden kann als $\mathcal{L}(T, 0)$, wobei T die zu A gehörende Gauß-Normalform ist, und 4.3.3 b) stellt in unserer jetzigen Sprache sicher, dass die Fundamentallösungen eine Basis des Vektorraumes der Lösungen des homogenenen LGS $Ax = 0$ sind. Ihre Anzahl ist $q - \mathrm{Rang}(A)$.

Beispiel 5.3.3 (Basen von K^p)

Im vorletzten Abschnitt hatten wir alle Untervektorräume von $\mathbb{R}^2$ aufgelistet und dabei mit überlegt, dass für jedes $u \neq 0$ in $\mathbb{R}^2$ und jedes $v \notin \mathbb{R} \cdot u$ gilt:

$$\mathbb{R}^2 = \langle u, v\rangle.$$

Da die Matrix $(u \mid v) \in \mathbb{R}^{2\times 2}$ Rang zwei hat, lässt sich jedes $x \in \mathbb{R}^2$ auf eindeutig bestimmte Art als Linearkombination von u und v schreiben. Also ist $\{u, v\}$ eine Basis von $\mathbb{R}^2$.

Allgemeiner besitzt eine Basis B von K^p höchstens p Elemente, denn sonst enthält sie $p+1$ Elemente $v_1, \ldots, v_{p+1}$, und nach 4.4.4 gibt es eine von 0 verschiedene Lösung des homogenen LGS

$$x_1 v_1 + x_2 v_2 + \cdots + x_{p+1} v_{p+1} = 0.$$

Also lässt sich der Nullvektor nicht eindeutig darstellen.

Eine Teilmenge des K^p aus q Elementen $v_1, \ldots, v_q$ ist genau dann eine Basis von K^p, wenn für die Matrix A mit Spalten $v_1, \ldots, v_q$ die Abbildung Φ_A bijektiv ist. Denn genau dann lässt sich jedes $w \in K^p$ auf genau eine Art als

$$w = \lambda_1 v_1 + \cdots + \lambda_q v_q = \Phi_A(\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_q \end{pmatrix})$$

schreiben. Wegen 4.4.4 ist die Bijektivität von Φ_A dazu äquivalent, dass $p = \mathrm{Rang}(A) = q$, also dazu, dass $q = p$ und dass A invertierbar ist.

Wir können also festhalten:

Fazit 5.3.4 Jede Basis von K^p hat genau p Elemente.

Vorbehaltlich der Frage, ob jeder K-Vektorraum eine Basis hat, gilt der folgende Sachverhalt.

Hilfssatz 5.3.5 ($V \cong \mathrm{Abb}(B, K)_0$)

Es sei V ein K-Vektorraum mit einer Basis B. Dann gibt es einen Isomorphismus zwischen $\mathrm{Abb}(B, K)_0$ und V.

Beweis. Die Abbildung

$$\Lambda : \mathrm{Abb}(B, K)_0 \longrightarrow V,\ \lambda \mapsto \sum_{b\in B} \lambda(b) \cdot b,$$

ist linear. Da sich jedes $v \in V$ als Linearkombination von B schreiben lässt, ist sie surjektiv. Da diese Möglichkeit eindeutig ist, ist Λ injektiv. Also ist Λ ein Isomorphismus. ◯

Bemerkung 5.3.6 (Koordinatenvektor)

Die Umkehrabbildung zu Λ aus dem eben vorgeführten Beweis nennen wir

$$D_B : V \longrightarrow \mathrm{Abb}(B, K)_0.$$

Für ein $v \in V$ heißt die Abbildung $D_B(v)$ der **Koordinatenvektor** von v bezüglich der Basis B, der Isomorphismus D_B heißt die **Koordinatenabbildung** bezüglich der Basis B.

Konkreter gehen wir auf den Fall ein, dass B endlich viele Elemente hat. Es sei $B = \{b_1, \ldots, b_q\}$ mit q Elementen. Wir identifizieren die Abbildungen von B nach K mit den q-Tupeln in K. Dabei kommt es auf die Reihenfolge der Basisvektoren entscheidend an! Wir sprechen dann von einer **geordneten Basis**, dem Tupel $(b_1, \ldots, b_q)$.

$$D_B(v) = D_B(\lambda_1 b_1 + \ldots + \lambda_q b_q) \hat{=} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_q \end{pmatrix} \in K^q.$$

Warnhinweis: Es gibt einen einzigen Fall, in dem jeder Vektor gleich seinem Koordinatenvektor ist, und das ist im Fall des Standardraumes $V = K^q$ mit der geordneten Basis $(e_1, \ldots, e_q)$.

Aufgabe 5.3.7 (Geraden)

Es sei V ein K-Vektorraum und $b \in V, b \neq 0$. Zeigen Sie, dass $\{b\}$ eine Basis von $Kb = \langle\{b\}\rangle$ ist. Was ist der Koordinatenvektor von λb, $\lambda \in K$ bezüglich dieser Basis? Gibt es andere Basen? Wie ändert sich der Koordinatenvektor von λb beim Übergang zu einer anderen Basis?

Definition 5.3.8 (Lineare Unabhängigkeit)

Eine Teilmenge $M \subset V$ des K-Vektorraums V heißt **linear unabhängig**, wenn die einzige Möglichkeit, den Nullvektor als Linearkombination von M zu schreiben, die triviale ist:

$$\forall \lambda \in \mathrm{Abb}(M, K)_0 : \Big[\sum_{m \in M} \lambda(m) \cdot m = 0 \iff \lambda = 0 \Big]$$

Ansonsten heißt M **linear abhängig.**

Beispiel 5.3.9 (Konkretisierung im K^p)

Wie nun schon fast üblich seien $v_1, \ldots, v_q \in K^p$ und $A \in K^{p \times q}$ die Matrix mit den Spalten $v_1, \ldots, v_q$. Dieses Mal seien die Vektoren paarweise verschieden. Dann gilt

$$\{v_1, \ldots, v_q\} \text{ ist linear unabhängig } \iff \mathrm{Rang}(A) = q.$$

Denn die rechte Seite ist nach 4.4.4 damit gleichbedeutend, dass $\mathcal{L}(A, 0) = \{0\}$, und das ist genau die lineare Unabhängigkeit der Menge $\{v_1, \ldots, v_q\}$.

Aufgabe 5.3.10 (Zwei Erhaltungseigenschaften)

Es seien K ein Körper und V, W zwei K-Vektorräume. Weiter sei $\Phi : V \to W$ ein Homomorphismus und $S = \{s_1, \ldots, s_q\} \subset V$ eine Teilmenge.

Begründen Sie:

a) Wenn S linear unabhängig ist und Φ injektiv, dann ist $\Phi(S)$ linear unabhängig.

b) Wenn S ein Erzeugendensystem von V ist und Φ surjektiv, dann ist $\Phi(S)$ ein Erzeugendensystem von W.

Übrigens: Das stimmt auch, wenn S unendlich ist. Versuchen Sie ruhig, das auch in diesem Fall aufzuschreiben.

Aufgabe 5.3.11 (Logarithmen)

Wir fassen $\mathbb{R}$ als Vektorraum über $\mathbb{Q}$ auf. Für d paarweise verschiedene Primzahlen $p_1, \ldots, p_d$ sei

$$x_i = \ln(p_i),\ 1 \leq i \leq d.$$

Weisen Sie nach, dass die Zahlen $x_1, \ldots, x_d$ über $\mathbb{Q}$ linear unabhängig sind.

Tipp: Nehmen Sie an, die Zahlen seien linear abhängig, multiplizieren Sie eine nichttriviale darstellung der 0 als Linearkombination mit einem gemeinsamen Nenner der Vorfaktoren durch und wenden Sie die Exponentialfunktion an. Welche Eigenschaft der natürlichen Zahlen können Sie jetzt benutzen?

Satz 5.3.12 (Charakterisierende Eigenschaften einer Basis)

Es sei V ein Vektorraum über dem Körper K. Dann sind für $B \subseteq V$ die folgenden Aussagen äquivalent:

i) B ist eine Basis.

ii) B ist maximal unter den linear unabhängigen Teilmengen von V, d.h. jede echt größere Teilmenge von V ist linear abhängig.

iii) B ist minimal unter den Erzeugendensystemen von V, d.h. jede echte Teilmenge von B ist kein Erzeugendensystem von V.

iv) B ist ein linear unabhängiges Erzeugendensystem.

Beweis.

i) $\Rightarrow$ ii)

Wenn B eine Basis ist, so ist B auch linear unabhängig, denn der Nullvektor lässt sich eindeutig als Linearkombination von B darstellen. In jeder Teilmenge M von V, die B enthält aber ungleich B ist, gibt es einen Vektor $v \in M \setminus B$. Dieser Vektor lässt sich als Linearkombination von B schreiben:

$$v = \sum_{b \in B} \lambda(b) \cdot b.$$

Setzt man $\lambda(v) = -1$ und $\lambda(m) = 0$ für $m \in M \setminus (B \cup \{v\})$, so folgt

$$0 = \sum_{m \in M} \lambda(m) \cdot m,$$

also ist M nicht linear unabhängig, sobald es wirklich größer als B ist.

ii) $\Rightarrow$ iii)

Nun sei B also maximal unter den linear unabhängigen Teilmengen von V. Eine echte Teilmenge S von B kann dann nicht V erzeugen, denn sonst wäre $b \in B \setminus S$ eine Linearkombination von S, und man erhielte ähnlich wie für M gerade eben, dass B nicht linear unabhängig ist.

Andererseits ist für $v \in V \setminus B$ die Menge $B \cup \{v\}$ linear abhängig, also gibt es eine Abbildung $\lambda \in \mathrm{Abb}(B, K)_0$ und ein $\mu \in K$, sodass

$$\mu \cdot v + \sum_{b \in B} \lambda(b) \cdot b = 0, \text{ wobei } (\mu, \lambda) \neq (0, 0) \in K \times \mathrm{Abb}(B, K)_0.$$

Wäre $\mu = 0$, so wäre $\lambda \neq 0$, und B wäre linear abhängig. Also ist $\mu \neq 0$, und wir können schreiben:

$$v = -\sum_{b \in B} \mu^{-1} \lambda(b) \cdot b \in \langle B \rangle,$$

das heißt: $V = \langle B \rangle$. Also ist B ein Erzeugendensystem, aber keine kleinere Menge erzeugt V.

iii) $\Rightarrow$ iv)

Nun sei B minimal unter den Erzeugendensystemen, das heißt: keine kleinere Menge erzeugt V.

Annahme: B ist nicht linear unabhängig.

Dann gibt es ein $0 \neq \lambda \in \mathrm{Abb}(B, K)_0$ mit

$$\sum_{b \in B} \lambda(b) \cdot b = 0.$$

Aus den beteiligten b's wählen wir ein b_0 mit $\lambda(b_0) \neq 0$, und dieses ist dann eine Linearkombination der übrigen:

$$b_0 = -\sum_{b_0 \neq b \in B} \lambda(b_0)^{-1} \lambda(b) b.$$

Damit sieht man, dass $B \setminus \{b_0\}$ V erzeugt, denn der kleinste Untervektorraum, der $B \setminus \{b_0\}$ enthält, enthält auch b_0 und damit $\langle B \rangle = V$. Da B aber ein minimales Erzeugersystem ist, folgt ein Widerspruch, und B ist – entgegen der Annahme – linear unabhängig.

iv) $\Rightarrow$ i)

Nun sei B ein linear unabhängiges Erzeugendensystem. Dann lässt sich jedes $v \in V$ als Linearkombination von B schreiben. Für $\lambda, \mu \in \mathrm{Abb}(B, K)_0$ gelte nun

$$v = \sum_{b \in B} \lambda(b) \cdot b = \sum_{b \in B} \mu(b) \cdot b.$$

Dann erhalten wir

$$0 = v - v = \sum_{b \in B} (\lambda(b) - \mu(b)) \cdot b,$$

also $\lambda = \mu$, da B linear unabhängig ist. Also ist die Darstellung von v als Linearkombination eindeutig und B ist Basis von V. ◯

Wenn Sie sich hier daran stören sollten, dass B so beliebig ist, und wenn Ihnen zum Beispiel $\mathrm{Abb}(B, K)_0$ nicht gefällt, dann versuchen Sie, den Beweis noch einmal im Fall $B = \{b_1, \ldots, b_d\}$ aufzuschreiben, wobei natürlich $b_i \neq b_j$ für $i \neq j$ vorausgesetzt sein sollte.

Folgerung 5.3.13 (Existenz einer Basis)

Der K-Vektorraum V besitze ein endliches Erzeugendensystem. Dann gelten:

a) V hat eine Basis.

b) Jedes endliche Erzeugendensystem von V enthält eine Basis von V.

c) Jede linear unabhängige Teilmenge von V lässt sich durch Hinzunahme endlich vieler Vektoren zu einer Basis ergänzen.

d) Je zwei Basen von V besitzen gleich viele Elemente.

Vor dem Beweis bemerken wir, dass Aussage c) oft als Basisergänzungssatz bezeichnet.

Beweis.

a) folgt aus b). Aussage b) gilt, da jedes endliche Erzeugendensystem auch ein minimales Erzeugendensystem enthalten muss. Entfernt man nämlich einen Vektor aus dem Erzeugendensystem, der sich als Linearkombination der anderen schreiben lässt, so bleibt ein kleineres Erzeugendensystem übrig. Dieses Vorgehen wiederholt man, bis ein linear unabhängiges Erzeugendensystem entstanden ist.

c) Da V eine endliche Basis B besitzt, ist die Koordinatenabbildung D_B ein Isomorphismus von V mit dem Vektorraum K^p (für $p = |B|$). Eine linear unabhängige Teilmenge M von V wird hierbei wegen 5.3.10 auf eine linear unabhängige Teilmenge von K^p abgebildet, hat also wegen 5.3.9 höchstens p Elemente. Wenn diese noch nicht V erzeugen, wählen wir ein Element $v \in V \setminus \langle M \rangle$ aus. Dann ist $M \cup \{v\}$ linear unabhängig und hat damit höchstens p Elemente. So fahren wir fort und haben nach spätestens $p - |M|$ Schritten eine linear unabhängige Teilmenge von V erreicht, die maximal unter den linear unabhängigen ist, da jede größere Teilmenge mindestens $p + 1$ Elemente enthält.

d) Da nach 5.3.4 jede Basis von K^p genau p Elemente enthält, enthält auch jede Basis von V genau p Elemente, wenn es mindestens eine Basis mit p Elementen gibt. Also enthalten alle Basen gleich viele Elemente. ◯

Definition 5.3.14 (Dimension)

Es sei V ein K-Vektorraum, der ein endliches Erzeugendensystem enthält. Dann wird die Mächtigkeit einer Basis B von V die **Dimension** von V genannt.

$$\boxed{\dim_K(V) := |B|.}$$

Nach der letzten Folgerung ist diese Größe von der Wahl einer Basis nicht abhängig, sondern nur davon, dass es überhaupt eine endliche Basis gibt. V heißt dann auch **endlichdimensional**. Wenn V kein endliches Erzeugendensystem besitzt, nennt man V **unendlichdimensional**.

Aufgabe 5.3.15 Zur Abwechslung mal Polynome

Es sei $V = \{f \in \mathbb{R}[X] \mid \mathrm{Grad}(f) < 5\}$. Dies ist ein reeller Vektorraum der Dimension 5, als Basis können wir nach Definition des Grades die Monome $1, X, X^2, X^3, X^4$ benutzen – und tun dies auch.

Bestimmen Sie eine Basis und die Dimension des Untervektorraumes

$$W = \{f \in V \mid f(1) = f(2) = 0\}.$$

Aufgabe 5.3.16 (Kombinatorische Ableitung)

Es sei $E = \{1, \ldots, n\}$ die Menge der Ecken eines Graphen $\Gamma = (E, K)$. Jede Kante k von Γ ist von der Form $\{\min(k), \max(k)\}$. Es sei $V = \mathrm{Abb}(E, \mathbb{R})$ die Menge aller Abbildungen von E nach $\mathbb{R}$ – also die n-Tupel reeller Zahlen. Für $f \in V$ sei $Df : K \to \mathbb{R}$ die Abbildung

$$Df(k) = f(\max(k)) - f(\min(k)).$$

Damit ist $D : V \to W := \mathrm{Abb}(K, \mathbb{R})$ definiert.

Weisen Sie nach, dass D linear ist und dass $f \in V$ genau dann im Kern von D liegt, wenn f auf jeder Zusammenhangskomponente von Γ konstant ist (vgl. 1.4.10).

Bestimmen Sie die Dimension des Kernes von D.

Bemerkung 5.3.17 (Unendliche Dimension)

Jeder Vektorraum besitzt eine Basis. Dass dies auch für Vektorräume gilt, die nicht endlich erzeugbar sind, folgt aus dem Auswahlaxiom. Wir werden darauf in §5.6 zurückkommen, *aber dies ist für den Rest des Buches nicht essentiell.*

Hilfssatz 5.3.18 (Monotonie der Dimension)

Es sei V ein endlichdimensionaler Vektorraum über dem Körper K. Dann ist jeder Untervektorraum U von V endlichdimensional, und es gilt

$$\dim_K(U) \leq \dim_K(V).$$

Gleichheit der Dimensionen gilt genau dann, wenn $U = V$.

Beweis. Es sei $M \subseteq U$ linear unabhängig. Dann hat M nach 5.3.13c) höchstens $\dim_K(V)$ Elemente, also gibt es eine maximale linear unabhängige Teilmenge von U und damit eine endliche Basis. Diese lässt sich zu einer Basis von V ergänzen und hat damit höchstens $\dim_K(V)$ Elemente. Das zeigt die Behauptung. ◯

Beispiel 5.3.19 (noch einmal der $\mathbb{R}^2$)

Wir haben in 5.1.8c) gesehen – damals etwas mühsam – wie die Untervektorräume von $\mathbb{R}^2$ aussehen. Das fügt sich jetzt nahtlos in unser Bild ein. Es gibt Untervektorräume von Dimension 0 (nämlich $\{0\}$), Dimension 1 (nämlich die $\mathbb{R}\cdot u,\ u \neq 0$) und Dimension 2 (nämlich $\mathbb{R}^2$ selbst), und keine anderen.

5.4 Summen von Untervektorräumen

Definition 5.4.1 (direkte Summe von Untervektorräumen)

Es seien V ein K-Vektorraum und $U_1, \ldots, U_n$ Untervektorräume von V. Dann ist die Summe von $U_1, \ldots, U_n$ in 5.1.13 definiert.

Diese Summe heißt eine **direkte Summe**, wenn gilt:

$$\forall u_i \in U_i : [u_1 + u_2 + \cdots + u_n = 0 \Longrightarrow u_1 = 0, u_2 = 0, \ldots, u_n = 0].$$

Das heißt, dass es nur eine Möglichkeit gibt, den Nullvektor als Summe von Vektoren $u_i \in U_i$ zu schreiben. Insbesondere gilt im Falle der Direktheit der Summe für $1 \leq i \neq j \leq n$ die Gleichheit $U_i \cap U_j = \{0\}$.

Falls die Summe direkt ist, so schreiben wir auch $\bigoplus_{i=1}^n U_i$ statt $\sum_{i=1}^n U_i$.

Beispiel 5.4.2 (Basen und Direktheit von Summen)

a) Es sei $B = \{b_1, \ldots, b_d\}$ eine Basis des d-dimensionalen Vektorraums V. Dann gilt

$$V = \bigoplus_{i=1}^{d} K \cdot b_i = \bigoplus_{i=1}^{d} \langle b_i \rangle.$$

Klar, das ist gerade die Definition 5.3.1 der Basis.

b) Wenn die Summe der U_i direkt ist und in jedem U_i eine Basis B_i gewählt wurde, dann sind diese Basen paarweise disjunkt (haben also leeren Durchschnitt) und ihre Vereinigung ist eine Basis von $\bigoplus_{i=1}^{n} U_i$. Das bedeutet im Fall endlicher Dimension

$$\dim_K(\bigoplus_{i=1}^{n} U_i) = \sum_{i=1}^{n} \dim_K(U_i).$$

c) Wenn umgekehrt die Gleichung

$$\dim_K(\sum_{i=1}^{n} U_i) = \sum_{i=1}^{n} \dim_K(U_i).$$

gilt und Basen der U_i gewählt sind, so ist ihre Vereinigung ein Erzeugendensystem von $\sum_{i=1}^{n} U_i$, und da die Kardinalität dieser Vereinigung sicher nicht größer ist als die Summe der Dimensionen, muss die Vereinigung der Basen disjunkt sein und eine Basis des Summenraumes liefern.

Im Endlichdimensionalen gilt also

$\sum_{i=1}^{n} U_i$ ist genau dann eine direkte Summe, wenn gilt:

$$\dim_K(\sum_{i=1}^{n} U_i) = \sum_{i=1}^{n} \dim_K(U_i).$$

Hilfssatz 5.4.3 (Dimensionsformel)

Es seien U und W endlichdimensionale Untervektorräume des K-Vektorraumes V. Dann gilt

$$\dim(U + W) = \dim(U) + \dim(W) - \dim(U \cap W).$$

Beweis. Es sei B_1 eine Basis von $U \cap W$. Diese lässt sich nach 5.3.13 c) zu einer Basis B von U ergänzen sowie zu einer Basis C von W. Es gilt dann, dass $B \cap C = B_1$, denn die Elemente von $B \cap C$ liegen ja alle in $U \cap W$, also in dem von B_1 erzeugten Vektorraum. Außerdem ist $B \cup C$ linear unabhängig. Denn aus

$$\sum_{v \in B \cup C} \lambda_v v = 0, \quad \lambda_v \in K,$$

folgt

$$\sum_{v \in B} \lambda_v v = - \sum_{v \in C \setminus B_1} \lambda_v v.$$

Die rechte Seite liegt in W, die linke Seite in U, also sind linke und rechte Seite im Durchschnitt $U \cap W$. Das geht für die rechte Seite nur, wenn alle λ_v, $v \in C \setminus B_1$, verschwinden (also Null sind). Damit sind aber überhaupt alle λ_v's Null, also $B \cup C$ linear unabhängig. Also ist $B \cup C$ eine Basis von $U + W$, und es gilt

$$\begin{aligned} \dim(U + W) &= |B \cup C| = |B| + |C| - |B \cap C| \\ &= \dim(U) + \dim(W) - \dim(U \cap W). \end{aligned}$$

○

Definition 5.4.4 (komplementärer Untervektorraum)

Es sei U ein Untervektorraum des Vektorraums V. Dann heißt W ein zu U **komplementärer Untervektorraum in** V oder auch **Vektorraumkomplement zu** U, wenn

$$V = U \oplus W.$$

Das bedeutet konkret, dass $V = U + W$ und $U \cap W = \{0\}$.

Wenn V endlichdimensional ist, dann gibt es zu jedem Untervektorraum mindestens einen komplementären Untervektorraum. Um solch einen zu finden wähle man eine Basis B_U von U und ergänze sie zu einer Basis B von V. Dann ist der von $B \setminus B_U$ erzeugte Untervektorraum von V ein zu U komplementärer Untervektorraum. Im Allgemeinen wird also ein Untervektorraum sehr viele veschiedene Vektorraumkomplemente haben.

Abschnitt 5.6 zeigt, dass dies auch ohne Endlichkeit der Dimension stimmt.

Für jedes Vektorraumkomplement W zu U in V gilt $U \cap W = \{0\}$ und wegen 5.4.2 haben wir die Formel:

$$\dim(U) + \dim(W) = \dim(V).$$

Aufgabe 5.4.5 (Geraden und Ebenen)

Im reellen Vektorraum $V = \mathbb{R}^3$ sei $U = \mathbb{R} \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$.

Begründen Sie, wieso die beiden Vektoren

$$w_1 = \begin{pmatrix} a \\ b \\ c \end{pmatrix},\ w_2 = \begin{pmatrix} d \\ e \\ f \end{pmatrix}$$

genau dann einen zu U komplementären Vektorraum aufspannen, wenn $bf - ce \neq 0$.

5.5 Faktorräume

Hinführung: was sind und sollen Faktorräume?

Gegeben seien ein Körper K und ein Homomorphismus $\Phi : V \longrightarrow W$ zwischen zwei K-Vektorräumen. Aus Hilfssatz 1.4.11 wissen wir dann, dass durch

$$v_1 \sim v_2 \iff \Phi(v_1) = \Phi(v_2)$$

eine Äquivalenzrelation auf V definiert wird, und dass die Äquivalenzklasse von v genau die Menge $\Phi^{-1}(\Phi(v))$ ist. Das lässt sich hier konkretisieren: es gilt ja

$$\Phi(v_1) = \Phi(v_2) \iff v_1 - v_2 \in \mathrm{Kern}(\Phi),$$

also ist die Äquivalenzklasse von v genau die Menge

$$[v] = v + \mathrm{Kern}(\Phi) = \{v + k \mid k \in \mathrm{Kern}(\Phi)\}.$$

Bitte vergleichen Sie das mit Hilfssatz 4.1.2; wir sollten uns für diese Äquivalenzklassen interessieren, sie sind Lösungsmengen von (abstrakten) Linearen Gleichungssystemen der Gestalt

$$\Phi(x) = \Phi(v),$$

bei denen sozusagen die eine Lösung v schon gefunden ist.

Wenn Φ surjektiv ist, vermittelt die Abbildung Φ eine Bijektion $\widetilde{\Phi}$ zwischen der Menge $V/\sim$ aller Äquivalenzklassen und der Menge W. Psychologisch fällt es manchen vielleicht leichter zu sehen, dass – immer noch im Fall der Surjektivität – die Urbildabbildung Φ^{-1} jeder einelementigen Teilmenge $\{w\}$ von W die Äquivalenzklasse $\{v \in V \mid \Phi(v) = w\}$ zuordnet, und dass dies eine Bijektion zwischen W und der Menge der Äquivalenzklassen liefert.

Nun ist aber W ein Vektorraum, und das legt nahe, dass sich die Vektorraumstruktur von W nach $V/\sim$ „zurückziehen“lässt. Das heißt, dass wir für $v_1, v_2 \in V, \alpha \in K$ folgende Addition und skalare Multiplikation auf $V/\sim$ festlegen:

$$\begin{aligned} [v_1] + [v_2] &:= \widetilde{\Phi}^{-1}(\widetilde{\Phi}([v_1]) + \widetilde{\Phi}([v_2])) = \Phi^{-1}(\Phi(v_1) + \Phi(v_2)) = [v_1 + v_2], \\ \alpha \cdot [v_1] &:= \widetilde{\Phi}^{-1}(\alpha\widetilde{\Phi}([v_1])) = \Phi^{-1}(\Phi(\alpha v_1)) = [\alpha \cdot v_1]. \end{aligned}$$

Mit diesen Verknüpfungen wird $V/\sim$ ein K-Vektorraum, was wir gleich noch vorrechnen werden, nachdem wir uns von Φ befreit haben.

Die durch Φ gegebene Äquivalenzrelation können wir bereits berechnen, wenn wir den Kern von Φ kennen. Dieser ist ein Untervektorraum von V.

Definition 5.5.1 (die Menge V/U)

Nun seien K ein Körper, V ein K-Vektorraum und $U \leq V$ ein beliebiger Untervektorraum. Dann definieren wir auf V die Äquivalenzrelation

$$v_1 \sim v_2 : \Longleftrightarrow\ v_1 - v_2 \in U.$$

Dass dies eine Äquivalenzrelation ist, haben wir im Prinzip schon in Beispiel 1.4.5c) gesehen, bei der Kongruenz modulo n auf den ganzen Zahlen. Dieselbe Rechnung funktioniert auch hier:

Reflexivität: Für alle $v \in V$ gilt

$$v - v = 0 \in U.$$

Also ist $v \sim v$.

Symmetrie: Für alle $v, w \in V$ gilt:

$$\begin{aligned} v \sim w &\Longrightarrow v - w \in U \\ &\Longrightarrow w - v \in U \\ &\Longrightarrow w \sim v. \end{aligned}$$

Transitivität: Für alle $u, v, w \in V$ gilt:

$$\begin{aligned} (u \sim v \wedge v \sim w) &\Longrightarrow u - v \in U \wedge v - w \in U \\ &\Longrightarrow u - w = (u - v) + (v - w) \in U \\ &\Longrightarrow u \sim w. \end{aligned}$$

Wir sehen, dass die drei Eigenschaften der Untergruppe U (siehe 2.2.1) für die drei Eigenschaften der Relation verantwortlich sind. Die Existenz der 0 liefert die Reflexivität, die Möglichkeit der Inversenbildung innerhalb U die Symmetrie, ud die Abgeschlossenheit von U unter der Addition die Transitivität.

Eine Äquivalenzklasse dieser Relation nennt man auch eine **Nebenklasse von V modulo U.** Die Menge aller Äquivalenzklassen nennen wir V **modulo** U und schreiben dafür V/U.

Beispiel 5.5.2 (Kulinarisches)

a) Es seien $K = \mathbb{R}$, $V = \mathbb{R}^3$, und $U = \mathbb{R} \cdot v$ für einen Vektor $v \neq 0$. Dann ist U also die Gerade, die 0 und v verbindet. Die Nebenklassen von V modulo U sind genau die zu U parallelen Geraden in V. Das erinnert ein bisschen an eine Tüte mit Spaghetti, die alle brav nebeneinander liegen: das „Spaghetti-Modell"des Faktorraums.

Wenn U ein zweidimensionaler Untervektorraum von V ist, dann sind die Nebenklassen von V modulo U genau die zu U parallelen Ebenen in V. Das ist das „Lasagne-Modell".

b) Das Beispiel für leidenschaftliche Matrizenrechner ist das übliche: wir nehmen eine Matrix $A \in K^{p\times q}$ und $U := \mathcal{L}(A, 0)$. Dann sind die Nebenklassen von K^q modulo U genau die nichtleeren Lösungsräume, also

$$V/U = \{\mathcal{L}(A, b) \mid b \in K^p, \exists v \in K^q : b = Av\}.$$

Insbesondere ist das v, das hier auftaucht, ein Repräsentant der Klasse $\mathcal{L}(A, b)$.

Beispiel 5.5.3 Stammfunktionen

Wir erinnern uns an Beispiel 5.2.2d), wo V der Vektorraum der unendlich oft differenzierbaren Funktionen auf $\mathbb{R}$ war und durch die Ableitung ein Endomorphismus D von V gegeben wurde.

Dieser ist surjektiv, da jede unendlich oft differenzierbare Funktion stetig ist und daher eine Stammfunktion besitzt, die aber nur bis auf eine additive Konstante bestimmt ist: Der Kern U von D is der Raum der konstanten Funktionen. Das unbestimmte Integral, das ja dem Versuch geschuldet ist, die Differentiation zu invertieren, ist nur bis auf eine additive Konstante festgelegt, ist also eine Element in V/U.

Definition 5.5.4 (Vektorraumstruktur auf V/U)

Es seien V ein K-Vektorraum und U ein Untervektorraum.

Wir wollen zwei Nebenklassen in V/U addieren. Dazu wählen wir $v_1, v_2 \in V$ und setzen

$$\begin{aligned}[v_1] + [v_2] &:= \{\tilde{v}_1 + \tilde{v}_2 \mid \tilde{v}_i \in [v_i]\} \\ &= \{(v_1 + u_1) + (v_2 + u_2) \mid u_1, u_2 \in U\} \\ &= \{(v_1 + v_2) + u \mid u \in U\} \\ &= [v_1 + v_2].\end{aligned}$$

Hier haben wir gleich mitgerechnet, dass diese Addition nicht von v_1 und v_2 abhängt, sondern nur von den durch sie repräsentierten Nebenklassen. Wir erhalten eine wohldefinierte Addition auf V/U.

Ganz ähnlich definieren wir die skalare Multiplikation auf $K \times (V/U)$, indem wir für $\alpha \in K^\times,\ v \in V$ schreiben:

$$\alpha \cdot [v] := \{\alpha \cdot \tilde{v} \mid \tilde{v} \in [v]\} = \{\alpha \cdot (v+u) \mid u \in U\} = \{(\alpha \cdot v) + \hat{u} \mid \hat{u} \in U\} = [\alpha \cdot v].$$

Für $\alpha = 0$ setzen wir

$$0 \cdot [v] := U = [0].$$

Wieder ist klar, dass diese Definitionen nur von der Äquivalenzklasse von v abhängen (deren Elemente ja genau die Elemente $v+u,\ u \in U$ sind!) und dass somit eine wohldefinierte Abbildung von $K \times (V/U)$ nach V/U vorliegt.

Nun ergeben sich die geforderten Bedingungen für Addition und skslare Multiplikation auf V/U leicht aus den Regeln in V.

Fazit 5.5.5

Mit den Verknüpfungen

$$[v_1] + [v_2] := [v_1 + v_2]$$
$$\alpha[v] := [\alpha v]$$

wird aus V/U ein K-Vektorraum.
Dieser heißt der **Faktorraum von V modulo U**.

Beispiel 5.5.6 (Konstruktion der reellen Zahlen)

Um einzusehen, dass solch eine Konstruktion die Mühe lohnt, soll hier ein vollkommen unerwartetes Beispiel erwähnt werden.

Es sei $K = \mathbb{Q}$ und V der Vektorraum der rationalen Cauchy-Folgen. Dieser Vektorraum lässt sich ohne Rückgriff auf reelle Zahlen definieren! Weiter sei U der Untervektorraum von V, der aus den Nullfolgen besteht.

Die Analysis sagt uns, dass jede Cauchy-Folge gegen eine reelle Zahl konvergiert, und zwei Cauchyfolgen liefern denselben Limes genau dann, wenn ihre Differenz eine Nullfolge ist. Wir haben also eine surjektive (und $\mathbb{Q}$-lineare) Abbildung von V nach $\mathbb{R}$, die einer Folge ihren Grenzwert zuordnet. Der Kern dieser Abbildung ist U.

Allerdings sagt uns die Analysis in aller Regel nicht, was eine reelle Zahl ist, oder wieso es einen Körper mit diesen Eigenschaften gibt!

Wir können aber jetzt den Spieß umdrehen und die Menge der reellen Zahlen als V/U definieren! Darauf gibt es dann eine Addition (wie stets in einem Faktorraum), eine Multiplikation (wie in 3.1.2 c)), eine archimedische Anordnung, und V/U ist vollständig (was zu zeigen bliebe). Es ergibt sich also ein mathematisch konsistentes Modell für die reellen Zahlen, indem man durch solch eine Quotientenbildung die Eigenschaften erzwingt, die man haben will. Wir brauchen dafür

nur die rationalen Zahlen und den Begriff der Cauchy-Folge, der sich ohne reelle Zahlen fassen lässt.

Dieses Verfahren findet sich sehr häufig, in aller Regel dann, wenn der Vektorraum V, den man gerade hat, zu viel an Information enthält und einiges davon für den eigentlichen Zweck irrelevant. Bei den reellen Zahlen „vergessen“ wir, auf welche Art wir sie jeweils durch eine individuelle rationale Cauchy-Folge approximieren.

Moral: Oft ist es so, dass sich mathematische Objekte mit „erhofften“ Eigenschaften auf diese Art aus bereits bestehenden Objekten herleiten lassen. Das allgemeine Vorgehen der Faktorbildung (oder Quotientenbildung) trägt oft dazu bei, dass man nicht erst langwierig jedesmal neu Strukturen legitimieren muss. Wenn man es richtig anfängt, dann „ererben“ sie sich von den Ausgangsobjekten. So etwas werden wir zum Beispiel in 10.3.3 noch einmal sehen.

Bemerkung 5.5.7 (kanonische Projektion)

Gleich kommt eine der Hauptanwendungen der Bildung des Faktorraums. Diese besteht darin, dass man einen gegebenen Homomorphismus von V nach W ohne große Willkür zerlegt als Produkt von drei Abbildungen, deren erste surjektiv, die zweite ein Isomorphismus und die dritte injektiv ist. Diese Schritte kann man dann einzeln analysieren.

Zuerst halten wir fest, dass für jeden Untervektorraum U von V die Abbildung

$$\pi_{V/U} : V \longrightarrow V/U,\ v \mapsto [v] = v + U,$$

ein surjektiver Vektorraumhomomorphismus ist. Die Linearität folgt sofort aus der Definition der Vektorraumstruktur von V/U. Die Surjektivität ist klar, weil es in V/U gar keine Elemente außer den Nebenklassen $[v]$ mit $v \in V$ gibt, und die werden alle von $\pi_{V/U}$ getroffen. Der Kern von $\pi_{V/U}$ ist gerade U.

$\pi_{V/U}$ heißt die **kanonische Projektion** von V auf V/U.

Satz 5.5.8 (Homomorphiesatz)

Es seien K ein Körper, V, W Vektorräume über K, und $\Phi \in \mathrm{Hom}(V, W)$. Es sei $U \leq \mathrm{Kern}(\Phi)$ ein Untervektorraum.

a) Es gibt genau eine lineare Abbildung

$$\widetilde{\Phi} : V/U \longrightarrow \Phi(V) \leq W,$$

sodass gilt:

$$\forall v \in V : \Phi(v) = \widetilde{\Phi}([v]).$$

b) Wenn sogar $U = \mathrm{Kern}(\Phi)$ gilt, dann ist die Abbildung $\widetilde{\Phi}$ ein Isomorphismus zwischen V/U und $\Phi(V)$.

Beweis.

a) Zunächst ist klar, wie $\widetilde{\Phi}$ zu definieren ist, wenn es überhaupt geht. Die Formel steht ja schon da:

$$\forall v \in V : \widetilde{\Phi}([v]) := \Phi(v).$$

Wir müssen nachprüfen, ob diese Vorschrift wohldefiniert ist, das heißt, ob tatsächlich für jede Klasse $[v] \in V/U$ nur ein Wert herauskommt, oder ob dieser vom gewählten Repräsentanten v abhängt.

Wir nehmen also an, es gelte $[v_1] = [v_2]$. Dann gilt $v_1 - v_2 \in U \subseteq \mathrm{Kern}(\Phi)$. Damit folgt aber

$$\Phi(v_1) - \Phi(v_2) = \Phi(v_1 - v_2) = 0,$$

also

$$\Phi(v_1) = \Phi(v_2).$$

Das zeigt, dass die Vorschrift für $\widetilde{\Phi}$ tatsächlich nur von der Äquivalenzklasse abhängt, $\widetilde{\Phi}$ also wohldefiniert ist.

Nun rechnen wir nach, dass $\widetilde{\Phi}$ linear ist:

$$\begin{aligned}
\forall v_1, v_2 \in V : \quad \widetilde{\Phi}([v_1] + [v_2]) &= \widetilde{\Phi}([v_1 + v_2]) = \Phi(v_1) + \Phi(v_2) \\
&= \widetilde{\Phi}([v_1]) + \widetilde{\Phi}([v_2]), \\
\forall \alpha \in K, v \in V : \quad \widetilde{\Phi}(\alpha[v]) &= \widetilde{\Phi}([\alpha v]) = \Phi(\alpha v) = \alpha\Phi(v) \\
&= \alpha\widetilde{\Phi}([v]).
\end{aligned}$$

Diese Regeln waren nachzuprüfen. Also ist $\widetilde{\Phi}$ linear.

b) Nun gelte sogar $U = \mathrm{Kern}(\Phi)$. Dann ist noch zu prüfen, dass $\widetilde{\Phi}$ ein Isomorphismus von V/U nach $\Phi(V)$ ist.

Surjektivität: Klar, jedes $w \in \Phi(V)$ lässt sich schreiben als $w = \Phi(v)$ für (mindestens) ein geeignetes $v \in V$, also $w = \widetilde{\Phi}([v])$.

Injektivität: Aus $\widetilde{\Phi}([v_1]) = \widetilde{\Phi}([v_2])$ folgt $\Phi(v_1) = \Phi(v_2)$, also $\Phi(v_1 - v_2) = 0$, also $v_1 - v_2 \in \mathrm{Kern}(\Phi) = U$, also $[v_1] = [v_2]$ in V/U. ◯

Bemerkung 5.5.9 (ein kommutatives Diagramm)

Es ist recht elegant, die Aussage des Homomorphiesatzes auf folgende Art durch ein Diagramm zu veranschaulichen. Dazu sei ι die Einbettung von $\Phi(V)$ nach W, d.h.

$$\iota : \Phi(V) \longrightarrow W, \quad \iota(w) := w.$$

Das ist nur ein Kunstgriff, um Definitions- und Bildbereiche der beteiligten Abbildungen abzustimmen. Dann gilt nach dem letzten Satz

$$\Phi = \iota \circ \widetilde{\Phi} \circ \pi_{V/U}.$$

Dies malen wir so auf:

$$\begin{array}{rcl} V & \xrightarrow{\Phi} & W \\ \pi_{V/U} \downarrow & & \uparrow \iota \\ V/U & \xrightarrow{\widetilde{\Phi}} & \Phi(V) \end{array}$$

Wir sagen, dass dieses Diagramm kommutiert, weil alle möglichen Kombinationen von Pfeilen, die von einem ersten Eck zu einem (nicht notwendig anderen) zweiten Eck gehen, dieselbe Abbildung liefern. In diesem Fall ist ja nur die Gleichung

$$\Phi = \iota \circ \widetilde{\Phi} \circ \pi_{V/U}$$

zu prüfen, die wir gerade bewiesen haben.

Bemerkung 5.5.10 (Basis des Faktorraums)

Wenn $U \subseteq V$ Vektorräume über K sind, und wir eine Basis B von V haben, die eine Basis B_U von U enthält, dann ist

$$C := \{b + U \mid b \in B \setminus B_U\}$$

eine Basis von V/U. Es gilt ja $V = U \oplus \langle B \setminus B_U \rangle$ und daher

$$V/U = \pi_{V/U}(V) = \pi_{V/U}(U \oplus \langle B \setminus B_U \rangle) = \langle \pi_{V/U}(B \setminus B_U) \rangle = \langle C \rangle,$$

da U unter $\pi_{V/U}$ auf die Null abgebildet wird. Außerdem ist C linear unabhängig, denn aus

$$\sum_{c \in C} \lambda(c) \cdot c = 0, \quad \lambda \in \mathrm{Abb}(C, K)_0,$$

folgt

$$\sum_{b \in B \setminus B_U} \lambda([b])b \in U = \langle B_U \rangle,$$

was wegen der linearen Unabhängigkeit von B zwangsläufig $\lambda = 0$ nach sich zieht.

Also ist C ein linear unabhängiges Erzeugendensystem und damit eine Basis von V/U. Insbesondere gilt im endlichdimensionalen Fall

$$\boxed{\dim(V/U) = \dim(V) - \dim(U).}$$

Denn nach 5.3.13 kann eine Basis von U stets zu einer Basis von V ergänzt werden.

Bemerkung 5.5.11 (zwei Dimensionsformeln)

a) In Beispiel 5.2.2b) hatten wir zwei Untervektorräume U und W des K-Vektorraums V betrachtet und dazu den Homomorphismus

$$\alpha : U \times W \longrightarrow V, \quad \forall (u,w) \in U \times W : \alpha(u,w) := u - w$$

studiert.

Der Homomorphiesatz sagt, dass

$$\text{Bild}(\alpha) \cong (U \times W)/\text{Kern}(\alpha).$$

Nun gilt aber nach 5.2.2b), dass $\text{Bild}(\alpha) = U + W$ und $\text{Kern}(\alpha) \cong U \cap W$. Wenn V endlichdimensional ist, so folgt daraus

$$\dim(U + W) = \dim(U) + \dim(W) - \dim(U \cap W).$$

Das ist ein neuer Beweis von Dimensionsformel 5.4.3.

b) Teil a) ist ein Spezialfall folgender Situation. Es seien V, W zwei K-Vektorräume, V endlichdimensional, $\Phi : V \longrightarrow W$ ein Homomorphismus. Dann sind auch $\text{Bild}\,(\Phi)$ und $\text{Kern}\,(\Phi)$ endlichdimensional, und es gilt

$$\text{Bild}(\Phi) \cong V/\text{Kern}(\Phi),$$

insbesondere also

$$\dim(\text{Bild}(\Phi)) = \dim(V) - \dim(\text{Kern}(\Phi)).$$

Definition 5.5.12 (Rang eines Homomorphismus)

Es seien V und W zwei endlichdimensionale Vektorräume über einem Körper K und Φ ein Homomorphismus von V nach W. Dann heißt die Dimension des Bildes von Φ der **Rang von** Φ.

Es gilt die Dimensionsformel

$$\boxed{\text{Rang}(\Phi) + \dim(\text{Kern}(\Phi)) = \dim(V).}$$

Die Dimension des Kernes wird oft auch der **Defekt** von Φ genannt, und dann heißt die letzte Formel auch die Rang-Defekt-Formel.

Ist $A \in K^{p \times q}$ eine Matrix, so gilt für die zugehörige lineare Abbildung $\Phi_A : K^q \longrightarrow K^p$, dass $\text{Rang}(\Phi_A) = \text{Rang}(A)$.

Aufgabe 5.5.13 (Hesse Reloaded)

Wir nehmen die Aufgabe 4.4.5 wieder auf und versuchen uns an einer Verallgemeinerung.

Denken Sie zunächst noch einmal über die alte Aufgabe nach. Dann geht es so weiter:

Sei K ein Körper und $A \in K^{p\times q}$ eine Matrix von Rang r. Sei U der von den Spalten von $A^\top$ erzeugte Untervektorraum von K^q.

Zeigen Sie:

a) Es gibt genau $q-r$ Fundamentallösungen des Linearen Gleichungssystems $Ax = 0$.

b) Sei B eine Basis von $\mathcal{L}(A, 0)$ und $N \in K^{q\times(q-r)}$ die Matrix mit den Vektoren aus B als Spalten. Dann gilt

$$U = \mathcal{L}(N^\top, 0).$$

c) Für $z \in K^q$ gilt

$$z + U = \{x \in K^q \mid N^\top x = N^\top z\}.$$

Bemerkung 5.5.14 (Lineare Codes – konkreter)

Wir nehmen die Fragestellung von 5.1.6 wieder auf und sehen uns ein Beispiel dazu an, das zwar in der Praxis nicht verwendet wird, dafür aber die Funktionsweise gut beleuchtet.

Es sei $K = \mathbb{F}_2 = \{0,\ 1\}$ der Körper mit zwei Elementen und $V = \mathbb{F}_2^5$ der fünfdimensionale Standardvektorraum über $\mathbb{F}_2$. Wenn Sie i Rest dieser bemerkung hie und da ein Minuszeichen vermissen, so liegt das daran, dass in $\mathbb{F}_2$ die Gleichheit $-1 = 1$ gilt.

Wir nehmen in V den Untervektorraum U, der von den beiden Vektoren

$$u_1 = \begin{pmatrix}1\\1\\1\\0\\0\end{pmatrix} \quad \text{und} \quad u_2 = \begin{pmatrix}0\\0\\1\\1\\1\end{pmatrix}$$

erzeugt wird. Es gilt

$$U = \{0, u_1, u_2, u_1 + u_2\}.$$

Jeder von Null verschiedene Vektor in U hat mindestens drei Einträge ungleich 0, wir sollten also bei der Datenübertragung von Codewörtern aus U ein falsch übertragenes Bit korrigieren können.

Um das praktisch umzusetzen, beschreiben wir U so wie eben gelernt. Die Matrix N hat als Spalten die Fundamentallösungen des Linearen Gleichungssystems

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix} x = 0,$$

die sich nach dem Gaußverfahren berechnen als

$$F^{(2)} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \; F^{(4)} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \; F^{(5)} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$

Nun gilt mit der Einsicht aus 5.5.13

$$U = \{u \in V \mid N^\top u = 0\},$$

wobei

$$N^\top = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

Diese Matrix wird im Kontext der Codierungstheorie auch gerne die **Kontrollmatrix** genannt, da sie eben kontrolliert, ob ein Element aus V zu U gehört.

Wenn nun ein Element $u \in U$ als Signal versandt wird und fehlerhaft als ein $v \in V$ ankommt, das sich von u nur an einer Stelle unterscheidet, dann gibt es ein $i \in \{1, \ldots, 5\}$ mit $v = u + e_i$, und es gilt

$$N^\top v = N^\top (u + e_i) = N^\top e_i = i\text{−te Spalte von } N^\top.$$

Da die Spalten von $N^\top$ paarweise verschieden sind – das kommt letztlich daher, dass es keinen Vektor in U gibt, der nur an zwei Stellen etwas von Null verschiedenes stehen hat –, können wir demnach dem Vektor $N^\top v$ ansehen, an welcher Stelle er sich von u unterscheidet und damit den Fehler beheben.

Das mag bei diesem Beispiel banal aussehen, ist aber der Clou, mit dem Lineare Codes, die etwas komplizierter sind, arbeiten.

Aufgabe 5.5.15 (Zusammenspiel)

Es sei V ein endlichdimensionaler Vektorraum und $\Phi : V \to V$ ein Endomorphismus.

Zeigen Sie, dass Φ genau dann surjektiv ist, wenn Φ injektiv ist.

Tipp: In Wirklichkeit gilt diese Aussage auch für einen Homomorphismus Φ zwischen zwei eindlichdimensionalen Vektorräumen gleicher Dimension.

5.6 Existenz von Basen

Wir haben bisher nur im Fall endlich erzeugbarer Vektorräume gesehen, dass sie eine Basis besitzen. Dass dies eine sachlich unnötige Einschränkung war (wenn auch didaktisch gerechtfertigt) soll hier erläutert werden.

Wir wollen dabei nicht wirklich abstrakt über geordnete Mengen sprechen, sondern uns konkret an der Situation orientieren, die uns interessiert. Allgemeineres zum Lemma von Zorn, das wir hier als Hilsmittel aus der Mengenlehre benötigen, findet sich zum Beispiel in Lang's *Algebra*, Springer 2002.

In diesem Abschnitt sei – wie sonst ja eigentlich auch – V ein Vektorraum über dem Körper K.

Satz 5.6.1 (Stets existiert eine Basis)

Es seien K ein Körper und V ein K-Vektorraum. Dann existiert eine Basis von V.

Beweis. Wir bezeichnen mit $\mathcal{M}$ die Menge aller linear unabhängigen Teilmengen von V. Diese Menge ist nicht leer, da $\emptyset \in \mathcal{M}$. Eine nichtleere Teilmenge $\mathcal{S} \subseteq \mathcal{M}$ heißt **total geordnet**, wenn gilt

$$\forall A, B \in \mathcal{S}: \ A \subseteq B \ \text{ oder } B \subseteq A.$$

Insbesondere ist die Vereinigung zweier Elemente einer total geordneten Menge $\mathcal{S}$ selbst wieder eine dieser beiden Mengen, und induktiv folgt, dass die Vereinigung endlich vielen Mengen aus $\mathcal{S}$ ebenfalls zu $\mathcal{S}$ gehört.

Daraus folgt: Für eine total geordnete Teilmenge $\mathcal{S}$ von $\mathcal{M}$ haben wir

$$\bigcup_{A \in \mathcal{S}} A \in \mathcal{M}.$$

Denn wäre diese Vereinigung linear abhängig, so enthielte sie eine endliche linear abhängige Teilmenge $\{v_1, \ldots, v_n\}$. Jedes v_i liegt in einem $A_i \in \mathcal{S}$, und da nur endlich viele A_i benötigt werden, liegen alle v_i auch in

$$\bigcup_{i=1}^{n} A_i = A_{i_0} \in \mathcal{S} \ (i_0 \text{ geeignet}).$$

Aber A_{i_0} ist linear unabhängig, was einen Widerspruch zur linearen Abhängigkeit von $\{v_1, \ldots, v_n\}$ herstellt.

Nun nehmen wir an, es gebe in V keine Basis. Dann sagt uns Satz 5.3.12, dass es in $\mathcal{M}$ kein maximales Element gibt. Das heißt, für alle $A \in \mathcal{M}$ gibt es ein $v_A \in V \setminus A$, sodass auch

$$f(A) := A \cup \{v_A\} \in \mathcal{M}$$

linear unabhängig ist. Bei der „simultanen“ Auswahl der v_A benutzen wir das so genannte Auswahlaxiom, das ebenfalls in Ebbinghaus Buch Einführung in die Mengenlehre diskutiert wird. Wir hätten damit eine Abbildung

$$f : \mathcal{M} \longrightarrow \mathcal{M}$$

konstruiert, für die gilt:

$$\forall A \in \mathcal{M} : A \subseteq f(A), \quad A \neq f(A).$$

Der folgende Satz führt dies zum Widerspruch und beendet damit den Beweis dieses Satzes. ◯

Satz 5.6.2 („Fixpunktsatz von Bourbaki“) *Für die Menge $\mathcal{M}$ aus dem letzten Satz sei $F : \mathcal{M} \longrightarrow \mathcal{M}$ eine Abbildung mit*

$$\forall A \in \mathcal{M} : A \subseteq F(A).$$

Dann gibt es ein $A \in \mathcal{M}$ mit $F(A) = A$.

Beweis. Wie nennen eine Teilmenge $\mathcal{S}$ von $\mathcal{M}$ **zulässig**, wenn die folgenden drei Bedingungen gelten: $\emptyset \in \mathcal{S}$, $F(\mathcal{S}) \subseteq \mathcal{S}$ und für jede total geordnete Teilmenge $\mathcal{T} \subseteq \mathcal{S}$ liegt auch die Vereinigung $\bigcup_{T \in \mathcal{T}} T$ in $\mathcal{S}$.

Zum Beispiel ist $\mathcal{M}$ selbst zulässig. Nun sei $\mathcal{S}_0$ der Durchschnitt aller zulässigen Teilmengen von $\mathcal{M}$. Da in jeder zulässigen Teilmenge auch die leere Menge liegt, enthält der Durchschnitt zumindest die leere Menge (als Element!), und außerdem gelten auch die beiden anderen Bedingungen der Zulässigkeit (nachrechnen!). Das heißt: $\mathcal{S}_0$ ist selbst zulässig und damit die kleinste aller zulässigen Teilmengen von $\mathcal{M}$.

Wenn wir nun zeigen können, dass $\mathcal{S}_0$ total geordnet ist, dann folgt daraus für $T_0 := \cup_{T \in \mathcal{S}_0} T$ zum Einen, dass $T_0 \in \mathcal{S}_0$ das größte Element von $\mathcal{S}_0$ ist. Andererseits gilt aber wegen der Zulässigkeit $F(T_0) \in \mathcal{S}_0$. Wir bekommen insgesamt

$$T_0 \subseteq F(T_0) \subseteq T_0$$

und damit die gewünschte Gleichheit.

Noch zu zeigen ist also die folgende Behauptung

Behauptung: $\mathcal{S}_0$ ist total geordnet.

Dies gilt, denn: Für den Beweis nennen wir $A \in \mathcal{S}_0$ ein **extremales Element**, wenn für alle $B \in \mathcal{S}_0$ mit $B \subset A$, $B \neq A$ gilt, dass $F(B) \subseteq A$. Es gibt extremale Elemente, zum Beispiel $A = \emptyset$.

Für ein extremales A setzen wir

$$\mathcal{S}_A := \{B \in \mathcal{S}_0 \mid B \subseteq A \vee F(A) \subseteq B\}.$$

Dann ist für jedes extremale A die Menge $\mathcal{S}_A$ zulässig:

- Die leere Menge liegt in $\mathcal{S}_A$, da sie eine Teilmenge von A ist.
- Ist $B \in \mathcal{S}_A$, so folgt aus $B \subset A$ schon $F(B) \subseteq A$, aus $B = A$ folgt $A \subseteq F(A) = F(B)$, und aus $B \not\subseteq A$ folgt $A \subseteq F(A) \subseteq B \subseteq F(B)$. Also gilt für jede dieser drei Möglichkeiten $F(B) \in \mathcal{S}_A$, folglich $F(\mathcal{S}_A) \subseteq \mathcal{S}_A$.
- Wenn $\mathcal{T}$ eine total geordnete Teilmenge von $\mathcal{S}_A$ ist, dann gilt: wenn alle $T \in \mathcal{T}$ in A enthalten sind, dann auch $\cup_{T\in\mathcal{T}} T$, und wenn mindestens ein T nicht in A enthalten ist, dann ist $F(A) \subseteq T \subseteq \cup_{T\in\mathcal{T}} T$. Also gilt

$$\cup_{T\in\mathcal{T}} T \in \mathcal{S}_A.$$

Da aber $\mathcal{S}_0$ die kleinste zulässige Teilmenge von $\mathcal{M}$ ist, muss also für alle extremalen A gelten:

$$\mathcal{S}_A = \mathcal{S}_0.$$

Nun müssen wir noch zeigen, dass jedes $A \in \mathcal{S}_0$ extremal ist. Dann folgt nämlich für $B \in \mathcal{S}_0$:

$$B \in \mathcal{S}_A \text{ also } [B \subseteq A] \vee [A \subseteq F(A) \subseteq B.]$$

Das sagt aber, dass $\mathcal{S}_0$ total geordnet ist.

Um zu beweisen, dass jedes $A \in \mathcal{S}_0$ extremal ist, betrachten wir

$$\mathcal{E} := \{A \in \mathcal{S}_0 \mid A \text{ ist extremal}\}.$$

Nun weisen wir nach, dass $\mathcal{E}$ zulässig ist, und damit gleich $\mathcal{S}_0$.

$$\emptyset \in \mathcal{E}: \quad \text{klar.}$$

Abgeschlossenheit von $\mathcal{E}$ unter F:

$$\forall A \in \mathcal{E}: \ \forall B \in \mathcal{S}_0 = \mathcal{S}_A, \ B \subset F(A): F(B) \subseteq F(A).$$

Diese letzte Inklusion gilt, da $F(A) \not\subseteq B$ und damit $B \subseteq A$. Nun greift die Extremalität von A.

Nun sei noch $\mathcal{T} \subseteq \mathcal{E}$ total geordnet, und

$$T_\infty := \bigcup_{T\in\mathcal{T}} T.$$

Zu zeigen ist $T_\infty \in \mathcal{E}$. Sei dazu $B \in \mathcal{S}_0$, $B \subset T_\infty$. Wenn für jedes $C \in \mathcal{T}$ die Inklusion $F(C) \subseteq B$ gelten würde, dann wäre T_∞ als Vereinigung von extremalen Teilmengen selbst eine Teilmenge von B – Widerspruch. Also gibt es ein extremales $C \in \mathcal{T}$ mit $F(C) \not\subseteq B$, und da $\mathcal{S}_0 = \mathcal{S}_C$ gilt, folgt daraus zwangsweise $B \subseteq C$. Aber $B \neq C$ impliziert dann $F(B) \subseteq C \subseteq T_\infty$, und aus $B = C$ folgt $F(B) = F(C) \in \mathcal{E}$, denn $\mathcal{E}$ ist unter F abgeschlossen. Also ist auch in diesem Fall $F(B) \subseteq T_\infty$. Damit folgt insgesamt, dass T_∞ extremal ist.

Das beendet den Beweis. ◯

Bemerkung 5.6.3 (zum Lemma von Zorn)

Das Lemma von Zorn verallgemeinert den Beweis von Satz 5.6.1 und sagt etwas allgemeiner das Folgende: wenn auf einer Menge M eine Ordnungsrelation gegeben ist, und wenn jede nichtleere total geordnete Teilmenge T von M eine obere Schranke besitzt, dann besitzt M ein maximales Element.

Dabei ist eine Ordnungsrelation auf M eine reflexive und transitive Relation $\leq$, für die noch dazu gilt:

$$\forall x, y \in M : [x \leq y \wedge y \leq x] \Rightarrow x = y.$$

Eine Teilmenge T von M heißt total geordnet, wenn

$$\forall x, y \in T : x \leq y \vee y \leq x.$$

Eine obere Schranke von T ist eine Element $z \in M$, sodass

$$\forall x \in T : x \leq z.$$

Eine kleinste obere Schranke von T ist eine obere Schranke z_0, sodass für alle oberen Schranken z von T gilt:

$$z_0 \leq z.$$

Ein **maximales Element** m von M ist eines, für das gilt:

$$\forall x \in M : x \geq m \Rightarrow x = m.$$

In unserem Beispiel ist eben M die Menge der linear unabhängigen Teilmengen von V, und $\leq$ die Inklusionsrelation $\subseteq$. Im Prinzip haben wir am Beispiel den allgemeinen Beweis des Lemmas von Zorn vorgeführt, wobei im allgemeinen Fall erst einmal ein Ersatz für die leere Menge gefunden werden muss. Dazu verkleinert man für ein beliebiges $a \in M$ die Menge M küntlich zu

$$\{x \in M \mid x \geq a\},$$

wo a das kleinste Element ist. Ein maximales Element in M_a ist dann auch maximal in M.

Als obere Schranke einer total geordneten Menge T konnten wir im Beweis von 5.6.2 immer die Vereinigung der Elemente von T verwenden. Das ist eine kleinste obere Schranke, was das Leben etwas einfacher macht. Im Allgemeinen braucht man noch einen Extratrick, um die Schwierigkeit zu umschiffen, dass es vielleicht keine kleinste obere Schranke gibt, auch wenn obere Schranken für total geordnete Teilmengen existieren.

Ansonsten geht der Beweis so durch wie vorgeführt: wenn es kein maximales Element gäbe, so gäbe es eine Abbildung $F : M \longrightarrow M$ mit $F(x) > x$ für alle $x \in M$. Aber jedes F mit $x \leq F(x)$ hat einen „Fixpunkt“, d.h. es gibt ein x_0 mit $F(x_0) = x_0$.

Satz 5.6.4 (je zwei Basen haben gleiche Kardinalität)

Es sei V ein Vektorraum über K, und B und C seien zwei Basen von V. Dann gibt es eine Bijektion zwischen B und C.

Beweis. Wir betrachten die Menge aller Tripel

$$\begin{aligned}(B_1, C_1, f_1): \quad & B_1 \subseteq B \text{ und } C_1 \subseteq C,\ f_1 : B_1 \longrightarrow C_1 \text{ bijektiv},\\ & (B \setminus B_1) \cap C_1 = \emptyset,\\ & (B \setminus B_1) \cup C_1 \text{ linear unabhängig}.\end{aligned}$$

Wir ordnen die Menge M dieser Tripel durch die folgende Ordnungsrelation:

$$(B_1, C_1, f_1) \leq (B_2, C_2, f_2) :\Longleftrightarrow\ B_1 \subseteq B_2,\ C_1 \subseteq C_2 \text{ und } f_1 = f_2|_{B_1}.$$

Es ist klar, dass dies eine Ordnungsrelation ist. Das kleinste Element von M ist $(\emptyset, \emptyset, \mathrm{Id}_\emptyset)$. Nun sei mithilfe einer Indexmenge I eine total geordnete Teilmenge

$$T := \{(B_i, C_i, f_i) \mid i \in I\}$$

von M gegeben. Zu zeigen ist, dass diese Menge T eine kleinste obere Schranke hat. Diese wird gegeben durch

$$(\widetilde{B}, \widetilde{C}, \widetilde{f}),\ \text{wobei } \widetilde{B} := \bigcup_{i\in I} B_i, \widetilde{C} := \bigcup_{i\in I} C_i \text{ und } \widetilde{f}(b) := f_i(b)\ (b \in B_i).$$

Es ist klar, dass dieses Tripel wieder in M liegt. Wäre zum Beispiel $(B \setminus \widetilde{B}) \cup \widetilde{C}$ linear abhängig, dann gäbe es eine endliche linear abhängige Teilmenge S davon, aber die läge schon in einem einzigen $(B \setminus B_{i_0}) \cup C_{i_0}$ für ein geeignetes $i_0 \in I$. Widerspruch. Die anderen Kriterien sieht man auf ähnliche Weise ein.

Das Lemma von Zorn sagt dann, dass die Menge M ein größtes Element $(\hat{B}, \hat{C}, \hat{f})$ besitzt. Wenn wir nun noch zeigen können, dass

$$B = \hat{B},\ C = \hat{C}$$

gilt, dann sind wir fertig.

Wäre aber zum Beispiel $B \neq \hat{B}$, so fände sich auch $C \neq \hat{C}$, denn

$$(B \setminus \hat{B}) \cup \hat{C}$$

wäre linear unabhängig und würde $\hat{C}$ echt enthalten, also wäre $\hat{C}$ noch nicht maximal linear unabhängig, also keine Basis, also ungleich C. Dann lässt sich durch Wahlen von geeigneten $b_1 \in (B \setminus \hat{B})$ und $c_1 \in (C \setminus \hat{C})$ das Element $(\hat{B}, \hat{C}, \hat{f})$ in naheliegender Weise vergrößern, wäre also nicht ein größtes Element.

Also folgt $B = \hat{B}$, und $C = \hat{C}$ geht analog. Insgesamt ist damit

$$\hat{f} : B \longrightarrow C$$

eine Bijektion zwischen den gegebenen Basen von V. ◯

Definition 5.6.5 (Dimension)

Auch im allgemeinen Fall heißt die Mächtigkeit einer Basis von V die Dimension von V. Da wir hier nicht wirklich in die Grundlagen der Mengenlehre eintreten wollen, sei nur am Rande erwähnt, dass es zu jeder Menge M eine echt größere Menge N gibt in dem Sinne, dass es keine surjektive Abbildung von M nach N geben kann. Insbesondere gibt es auch nicht zueinander isomorphe unendlichdimensionale Vektorräume.

Beispiel 5.6.6 (Nun finde mal eine Basis!)

Fast die einzigen unendlichdimensionalen Vektorräume, für die eine Basis bekannt ist, sind die, die durch Vorgabe einer Basis konstruiert werden. Dazu gehören die Vektorräume $\mathrm{Abb}(M,K)_0$. Solch ein Vektorraum hat zum Beispiel die Basis, die aus allen Funktionen besteht, deren Träger genau ein Element enthält, und die dort den Wert 1 annehmen. Im Prinzip gehört in diese Klasse von Vektorräumen auch der Polynomring über K, der als Basis zum Beispiel die Menge $\{1, X, X^2, X^3, \dots\}$ besitzt. Interessanter ist der Körper der rationalen Funktionen über K (siehe Abschnitt 8.3) für den sich mithilfe der sogenannten Partialbruchzerlegung eine K-Basis angeben lässt.

Hingegen ist man weit davon entfernt, eine Basis von $\mathbb{R}$ als $\mathbb{Q}$-Vektorraum zu kennen.

Kapitel 6

Basen und lineare Abbildungen

Hier wollen wir erklären, wie Basen benutzt werden können, um lineare Abbildungen zu beschreiben. Das führt uns wieder zum Matrizenkalkül zurück.

6.1 Lineare Fortsetzung

Hilfssatz 6.1.1 (Rekonstruktion einer Linearen Abbildung)

Es seien K ein Körper, V, W zwei K-Vektorräume, und B eine Basis von V. Weiter sei $\Phi : V \longrightarrow W$ ein Homomorphismus von K-Vektorräumen. Dann ist Φ eindeutig durch die Einschränkung $\Phi|_B : B \longrightarrow W$ festgelegt.

Beweis. Schreibe $v \in V$ als $v = \sum_{b \in B} \lambda(b) \cdot b$. Das geht mit einem geeigneten (sogar eindeutigen, aber das ist hier irrelevant) $\lambda \in \mathrm{Abb}(B, K)_0$. Dann gilt

$$\Phi(v) = \Phi(\sum_{b \in B} \lambda(b) \cdot b) = \sum_{b \in B} \lambda(b) \cdot \Phi(b) = \sum_{b \in B} \lambda(b) \cdot \Phi|_B(b).$$

Das zeigt die Behauptung. ◯

Dieser Hilfssatz sagt, dass die Zuordnung $\Phi \mapsto \Phi|_B$ eine injektive Abbildung von $\mathrm{Hom}(V, W)$ nach $\mathrm{Abb}(B, W)$ ist. Dass sie linear ist, ist offensichtlich.

Nun drehen wir den Spieß um und konstruieren durch beliebige Vorgabe von Werten auf der Basis eine lineare Abbildung von V nach W.

Satz 6.1.2 (Lineare Fortsetzung)

Es seien K ein Körper, V, W zwei K-Vektorräume, und B eine Basis von V. Weiter sei $f : B \longrightarrow W$ eine Abbildung. Dann gibt es genau eine lineare Abbildung $\Phi : V \longrightarrow W$ mit $\Phi|_B = f$.

Diese heißt die **lineare Fortsetzung** *von f.*

Beweis. Es sei also $f : B \longrightarrow W$ gegeben. Wenn es überhaupt eine Möglichkeit gibt, dazu eine passende lineare Abbildung auf V zu finden, dann muss sie durch die Formel im Beweis von 6.1.1 definiert sein:

$$\Phi : V \longrightarrow W, \quad \Phi(v) := \sum_{b \in B} \lambda(b) \cdot f(b),$$

wobei $v \in V$ als Linearkombination $v = \sum_{b \in B} \lambda(b) \cdot b$ (mit $\lambda \in \mathrm{Abb}(B, K)_0$) geschrieben wird. Da λ aufgrund der Basiseigenschaft von B eindeutig bestimmt ist, ist diese Abbildung Φ wohldefiniert. Zu zeigen ist noch die Linearität von Φ. Das passiert jetzt:

Seien $u, v \in V$. Dann lassen sich u, v als Linearkombinationen von B schreiben:

$$\exists \lambda, \mu \in \mathrm{Abb}(B, K)_0 : u = \sum_{b \in B} \mu(b) \cdot b, \quad v = \sum_{b \in B} \lambda(b) \cdot b.$$

Dann gilt aber

$$u + v = \sum_{b \in B} (\mu(b) + \lambda(b)) \cdot b = \sum_{b \in B} (\mu + \lambda)(b) \cdot b.$$

Dabei ist $\mu + \lambda \in \mathrm{Abb}(B, K)_0$ und es folgt

$$\begin{aligned} \Phi(u + v) &= \Phi(\textstyle\sum_{b \in B}((\mu + \lambda)(b)) \cdot b) = \\ &= \textstyle\sum_{b \in B}(\mu(b) + \lambda(b)) \cdot f(b) = \\ &= \textstyle\sum_{b \in B} \mu(b) f(b) + \sum_{b \in B} \lambda(b) f(b) = \\ &= \Phi(u) + \Phi(v). \end{aligned}$$

Genauso gilt für $\alpha \in K$

$$\alpha \cdot v = \sum_{b \in B} (\alpha\lambda)(b) \cdot b,$$

also wegen $\alpha\lambda \in \mathrm{Abb}(B, K)_0$ auch

$$\Phi(\alpha v) = \Phi(\sum_{b \in B} (\alpha\lambda)(b) \cdot b = \sum_{b \in B} \alpha\lambda(b) \cdot f(b) = \alpha\Phi(v).$$

Damit sind die zwei Bedingungen der Linearität nachgewiesen. ◯

Folgerung 6.1.3

Die Abbildung $\mathrm{Hom}(V, W) \longrightarrow \mathrm{Abb}(B, W), \ \Phi \mapsto \Phi\|_B,$ ist ein Isomorphismus von Vektorräumen

Bemerkung 6.1.4 (Matrizen, Dimension)

a) Vielleicht ist es nicht ganz abwegig, dies im Spezialfall $V = K^q,\ W = K^p$ noch einmal explizit zu machen. Wir wählen in V die Standardbasis

$$B := \{e_1, \ldots, e_q\}.$$

Eine Abbildung $f : B \longrightarrow K^p$ entspricht der Angabe von q Vektoren

$$f(e_i) := f_i \in K^p,\ 1 \leq i \leq q.$$

Dann ist die lineare Fortsetzung von f die lineare Abbildung Φ_A von K^q nach K^p, die durch Multiplikation mit der Matrix

$$A := (f_1\ f_2\ \ldots\ f_q) \in K^{p \times q}$$

gegeben wird: Erstens ist dies eine lineare Abbildung, und zweitens gilt für $1 \leq i \leq q$:

$$\Phi_A(e_i) = A \cdot e_i = f_i = f(e_i).$$

Wir haben also eigentlich nur Hilfssatz 5.2.5 neu ausgeleuchtet.

b) Eine Anwendung des Prinzips der linearen Fortsetzung ist die Feststellung, dass zwei endlichdimensionale Vektorräume genau dann isomorph sind, wenn sie dieselbe Dimension haben. Denn ein Isomorphismus bildet eine Basis des einen auf eine Basis des anderen ab, erhält also die Dimension. Wenn umgekehrt Basen gleicher Kardinalität existieren, dann setzt sich eine Bijektion zwischen diesen zu einem Isomorphismus der erzeugten Vektorräume fort. (Warum?)

c) Seien nun V und W wieder beliebige KVektorräume. Da jede linear unabhängige Teilmenge S von V sich zu einer Basis von V ergänzen lässt, lässt sich auch jede Abbildung von S nach W zu einer linearen Abbildung von V nach W fortsetzen. Dies funktioniert allerdings nicht eindeutig, wenn S keine Basis von V ist.

Aufgabe 6.1.5 (mit Parameter)

Es sei $K = \mathbb{F}_3$ und $V = \{f \in K[X] \mid \text{grad}(f) \leq 3\}$. Dies ist ein K-Vektorraum der Dimension 4. Für ein festes $a \in K$ sei

$$S = \{1 - X^2, X - X^3, 1 + X + aX^2 + aX^3\} \subset V.$$

Entscheiden Sie, für welche Werte $a \in K$ sich die Abbildung

$$\phi : S \to V,\ \phi(1 - X^2) = 1,\ \phi(X - X^3) = X, \phi(1 + X + aX^2 + aX^3) = 1 + aX$$

zu einem Endomorphismus von V fortsetzen lässt.

6.2 Der Dualraum

Definition 6.2.1 (Linearformen, Dualraum)

Es seien K ein Körper und V ein Vektorraum über K. Eine **Linearform** auf V ist eine K-lineare Abbildung von V nach K.

Der Raum $\mathrm{Hom}(V, K)$ aller Linearformen heißt auch der **Dualraum** von V und wird oft mit V^* notiert.

Der Dualraum ist also ein Spezialfall der Vektorräume $\mathrm{Hom}(V, W)$, allerdings ein besonders wichtiger.

Bemerkung 6.2.2 (duale Basis)

Wenn V endlichdimensional und in V eine Basis $B := \{b_1, \ldots, b_d\}$ gewählt ist, dann wird nach 6.1.2 eine Linearform auf eindeutige Art durch eine Abbildung von B nach K vorgegeben:

$$V^* \cong \mathrm{Abb}(B, K).$$

Speziell sehen wir, dass auch V^* Dimension d hat. Aus dieser Einsicht ergibt sich eine Basis von V^*, indem für $1 \leq i \leq d$ die Linearform $b_i^* \in V^*$ definiert wird als Lineare Fortsetzung der Vorschrift

$$b_i^*(b_j) := \begin{cases} 1, & \text{falls } i = j, \\ 0, & \text{falls } i \neq j. \end{cases}$$

Dies ist so gemacht, dass für $v = \sum_{i=1}^{d} c_i b_i,\ c_i \in K,$ die Gleichheit

$$c_j = b_j^*(v)$$

gilt. Für die zu B gehörige Koordinatenabbildung $D_B : V \to K^d$ gilt also wegen

$$\boxed{v = \sum_{i=1}^{d} b_i^*(v) \cdot b_i,}$$

die Formel $D_B(v) = (b_i^*(v))_{1 \leq i \leq d}$.

Die Menge $\{b_i^* \mid 1 \leq i \leq d\}$ besteht aus d Elementen im d-dimensionalen Vektorraum V^*. Wenn wir zeigen, dass sie V^* erzeugt, muss es sich um eine Basis handeln.

Um dies zu zeigen, wählen wir eine beliebige Linearform $\lambda \in V^*$. Dann gilt:

$$\lambda = \sum_{i=1}^{d} \lambda(b_i) \cdot b_i^*.$$

Um das zu verifizieren, rechnen wir für ein beliebiges $v \in V$ nach:

$$\begin{aligned}\lambda(v) &= \lambda(\textstyle\sum_{i=1}^{d} b_i^*(v) b_i) = \\ &= \textstyle\sum_{i=1}^{d} b_i^*(v)\lambda(b_i) = \\ &= \left(\textstyle\sum_{i=1}^{d} \lambda(b_i) \cdot b_i^*\right)(v).\end{aligned}$$

Das zeigt die gewünschte Gleichheit.

Die Basis $\{b_1^*, \ldots, b_d^*\}$ von V^* heißt die zu $\{b_1, \ldots, b_d\}$ **duale Basis**.

Da die duale Basis genauso viele Elemente hat wie die ursprüngliche Basis liefert die Vorschrift $b_i \mapsto b_i^*$ nach Linearer Fortsetzung (siehe 6.1.2) einen Isomorphismus zwischen V und V^*. Dieser hängt allerdings von der Wahl einer Basis in V ab und hat daher etwas willkürliches an sich. Trotzdem halten wir fest:

$$\boxed{\dim(V) < \infty \Longrightarrow \dim(V) = \dim(V^*) \Rightarrow V \cong V^*.}$$

Vorsicht: Der Vektor b_i^* hängt nicht nur von b_i ab, sondern von der gesamten Basis $B = \{b_1, \ldots, b_d\}$. Dies in der Notation zu dokumentieren wäre allerdings nicht sehr leserfreundlich.

Beispiel 6.2.3 (K^d)

Als konkretes Beispiel wählen wir wieder einmal den K^d. Eine Lineare Abbildung von K^d nach K wird gegeben durch die Multiplikation mit einer $(1 \times d)$-Matrix, denn wir haben ja in Hilfssatz 5.2.5

$$\mathrm{Hom}(K^d, K) \cong K^{1\times d}$$

gesehen. Wenn nun $\{b_1, \ldots, b_d\} =: B$ eine Basis von K^d ist, dann stellt sich die Frage, welche Zeilen $z_1, \ldots, z_d \in K^{1\times d}$ der dazu dualen Basis entsprechen. Die Bedingung dafür ist nach Definition der dualen Basis:

$$z_i \cdot b_j = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

Das bedeutet aber gerade, dass die Matrix Z mit den Zeilen $z_1, \ldots, z_d$ zu der Matrix A mit den Spalten $b_1, \ldots, b_d$ invers ist. Daraus folgt für beliebiges $v \in K^d$:

$$v = A \cdot (Z \cdot v), \quad D_B(v) = Z \cdot v.$$

Aufgabe 6.2.4 (Taylorpolynome)

Für eine natürliche Zahl d sei $V_d = \{f \in \mathbb{R}[X] \mid \mathrm{Grad}(f) < d\}$. In diesem $\mathbb{R}$-Vektorraum der Dimension d haben wir die naheliegende Basis $B = \{1, X, \ldots, X^{d-1}\}$. Wir schreiben $b_i = X^i,\ 0 \leq i \leq d-1$.

Weisen Sie nach, dass für festes $k \in \mathbb{N}_0$ und festes $x \in \mathbb{R}$ die Abbildung

$$D_x^k : V_d \to \mathbb{R}, \quad f \mapsto f^{(k)}(x),$$

wobei $f^{(k)}$ die k-fache Ableitung von f bezeichnet, eine Linearform auf V_d ist. Weisen Sie zusätzlich nach, dass die duale Basis zu B gegeben ist durch

$$b_i^*(f) = f^{(i)}(0)/i!, \ 0 \leq i \leq d-1.$$

Begründen Sie damit die Identität

$$\forall f \in V_d : f = \sum_{i=0}^{d-1} \frac{f^{(i)}(x)}{i!} X^i.$$

Bemerkung 6.2.5 (unendliche Weiten)

Wenn V unendlichdimensional ist, dann gilt immer noch $V^* \cong \mathrm{Abb}(B, K)$. Aber nun hat V^* größere Dimension als V in dem Sinn, dass es keinen surjektiven Homomorphismus von V nach V^* gibt. Es ist eine delikate Aufgabe, dieses Defizit in speziellen Situationen wenn möglich zu entschärfen. Manchmal kann man einen interessanten Teilraum von V^* sinnvoll definieren, der zu V isomorph ist. Am Besten geht das in der Theorie der Hilberträume, die (reelle oder komplexe) Vektorräume mit einem Skalarprodukt sind und die noch dazu bezüglich dieses Skalarproduktes „vollständig“ sind. Dies ist ein beliebtes Thema der Funktionalanalysis.

Definition/Bemerkung 6.2.6 (duale Abbildung)

Es seien V, W Vektorräume über dem Körper K und $\Phi : V \longrightarrow W$ eine lineare Abbildung. Dann ist für jede Linearform $\kappa \in W^*$ die Abbildung $\kappa \circ \Phi : V \longrightarrow K$ eine Linearform auf V. Wir erhalten also eine Abbildung

$$\Phi^* : W^* \longrightarrow V^*, \ \Phi^*(\kappa) := \kappa \circ \Phi.$$

Diese Abbildung Φ^* ist linear, denn für alle $\kappa, \lambda \in W^*$, für alle $\alpha \in K$ und für alle $v \in V$ gilt:

$$\begin{aligned}
(\Phi^*(\kappa + \lambda))(v) &= (\kappa + \lambda)(\Phi(v)) \\
&= \kappa(\Phi(v)) + \lambda(\Phi(v)) = (\Phi^*(\kappa))(v) + (\Phi^*(\lambda))(v), \\
(\Phi^*(\alpha \cdot \kappa))(v) &= (\alpha \cdot \kappa)(\Phi(v)) \\
&= \alpha \cdot \kappa(\Phi(v)) = \alpha(\Phi^*(\kappa))(v).
\end{aligned}$$

Dabei wird natürlich die Vektorraumstruktur von W^* und V^* benutzt, die in 5.2.4 festgelegt wurde.

Diese Abbildung Φ^* heißt die zu Φ **duale Abbildung**.

Wenn $\Phi \in \mathrm{Hom}(V,W)$ surjektiv ist, dann ist Φ^* injektiv. Denn wenn $\kappa \in W^*$ im Kern von Φ^* liegt, dann haben wir für alle $v \in V$:

$$\kappa(\Phi(v)) = \Phi^*(\kappa)(v) = 0(v) = 0,$$

und da Φ surjektiv ist, lässt sich jedes $w \in W$ als $\Phi(v)$ schreiben, also gilt für alle $w \in W$:

$$\kappa(w) = 0.$$

Damit ist $\kappa = 0$ und der Kern von Φ^* besteht nur aus dem Nullelement. Bemerkung 5.2.3 sagt, dass dies die Injektivität von Φ^* liefert.

Wenn Φ injektiv ist, dann ist Φ^* surjektiv. Denn für eine beliebige Linearform $\lambda \in V^*$ lässt sich ein Urbild unter Φ^* wie folgt konstruieren: Wir wählen ein Vektorraumkomplement U zu $\Phi(V) \subseteq W$, was aufgrund des Satzes von der Basisergänzung funktioniert. Dann lässt sich (wegen der Injektivität von Φ) jedes $w \in W$ auf genau eine Art schreiben als

$$w = \Phi(v) + u, \quad v \in V,\ u \in U.$$

Wenn man auf W die Linearform $\kappa(w) := \lambda(v)$ definiert, dann gilt

$$\lambda = \Phi^*(\kappa).$$

Aufgabe 6.2.7 (Isomorphismen und duale Basen)

a) Es seien V, W endlichdimensionale Vektorräume über dem Körper K und $\Phi : V \to W$ ein Isomorphismus. Weiter sei $B = \{b_1, \ldots, b_d\}$ eine Basis von V und $c_i = \Phi(b_i),\ 1 \le i \le d$. Zur Basis $C = \{c_1, \ldots, c_d\}$ von W gehört die duale Basis $\{c_1^*, \ldots, c_d^*\}$ von W^*.

Zeigen Sie, dass die zu B duale Basis $\{b_1^*, \ldots, b_d^*\}$ durch

$$b_i^* = \Phi^*(c_i^*)$$

gegeben ist.

Weiter gilt $(\Phi^*)^{-1} = (\Phi^{-1})^*$.

b) Nun sei $V = W = V_d$ in der Situation von Aufgabe 6.2.4 und für ein festes $a \in \mathbb{R}$ sei

$$\Phi : V_d \to V_d,\ f = f(X) \mapsto f(X - a).$$

Für $X^i = b_i,\ 0 \le i \le d-1$, sei $c_i = \Phi(X^i)$.

Zeigen Sie: Es gilt

$$c_i^*(f) = ((\Phi^*)^{-1}(b_i))(f) = b_i^*(\Phi^{-1}(f)) = f^{(i)}(a)/(i!),$$

und es folgt

$$f(X) = f(a + (X - a)) = \sum_{i=0}^{d-1} b_i^*(\Phi(f)) \cdot (X - a)^i = \sum_{i=0}^{d-1} \frac{f^{(i)}(a)}{i!} \cdot (X - a)^i.$$

NB: Lassen Sie sich nicht davon verwirren, dass der Laufindex in Teil b) bei 0 losgeht. Das passt besser zur „Natur“ der Situation.

Bemerkung 6.2.8 (Der Bidualraum)

Da V^* ein Vektorraum ist, hat auch er einen Dualraum, $(V^*)^* =: V^{**}$. Dieser heißt der **Bidualraum** von V. Zu jeder Abbildung $\Phi : V \longrightarrow W$ gibt es eine „biduale Abbildung“

$$(\Phi^*)^* : (V^*)^* \longrightarrow (W^*)^*.$$

Für jedes $v \in V$ ist durch

$$\Lambda(v) : V^* \longrightarrow K, \quad \kappa \mapsto \kappa(v),$$

eine Linearform auf V^*, also ein Element von V^{**} definiert. Diese Abbildung Λ ist linear (nachrechnen!) und injektiv. Das beweisen wir unter der Annahme, V sei endlichdimensional. Dann gibt es nämlich nach 5.4.4 zu $w \in V \setminus \{0\}$ einen Untervektorraum U mit $V = \langle w \rangle \oplus U$, und jedes $v \in V$ lässt sich schreiben als $v = \alpha \cdot w + u$ mit einem $u \in U$. Da $w \neq 0$ vorausgesetzt ist, ist α eindeutig bestimmt. Daher ist die Abbildung

$$\kappa : V \longrightarrow K, \quad v = \alpha \cdot w + u \mapsto \alpha,$$

eine Linearform auf V. Wegen $\Lambda(w)(\kappa) = \kappa(w) = 1$ folgt

$$\Lambda(w) \neq 0.$$

Für endlichdimensionale Vektorräume stimmen die Dimensionen von V und V^{**} beide mit der Dimension von V^* überein. Also ist die injektive Abbildung Λ in diesem Fall wegen der Dimensionsformel 5.5.11b) und der Monotonie der Dimension 5.3.18 auch surjektiv und damit ein Isomorphismus. Da Λ definiert werden kann, ohne eine Basis zu wählen, spricht man von einem „natürlichen Isomorphismus“.

$$\boxed{\dim V < \infty \Rightarrow V \cong V^{**}}$$

6.3 Die Abbildungsmatrix

Nun seien wieder V, W beliebige endlichdimensionale K-Vektorräume. Wir wählen in V eine Basis $B := \{b_1, \ldots, b_q\}$ und in W eine Basis $C := \{c_1, \ldots, c_p\}$.

Außerdem sei ein Homomorphismus Φ von V nach W gegeben. Diesen wollen wir mithilfe der Basen beschreiben.

Etwas präziser wollen wir eine Methode angeben, wie man für $v \in V$ die Koeffizienten von $\Phi(v)$ bezüglich C ausrechnen kann, wenn die Koeffizienten von v bezüglich B bekannt sind.

Dazu schreiben wir erst einmal die Vektoren $\Phi(b_j),\ 1 \leq j \leq q,$ als Linearkombinationen von $c_1, \ldots, c_p$:

$$\Phi(b_j) = \sum_{i=1}^{p} a_{ij} c_i.$$

Diese Koeffizienten fassen wir zur $p \times q$-Matrix $A \in K^{p \times q}$ zusammen. Dann gilt für $v = \sum_{j=1}^{q} \alpha_j b_j$:

$$\Phi(v) = \sum_{j=1}^{q} \alpha_j \Phi(b_j) = \sum_{j=1}^{q} \sum_{i=1}^{p} \alpha_j a_{ij} c_i = \sum_{i=1}^{p} (\sum_{j=1}^{q} a_{ij} \alpha_j) c_i = \sum_{i=1}^{p} \beta_i c_i,$$

wobei

$$\beta = A \cdot \alpha.$$

Definition 6.3.1 (Abbildungsmatrix)

Die eben eingeführte Matrix A heißt die **Abbildungsmatrix** von Φ bezüglich der Basen B und C. Oft werden wir hierfür $D_{CB}(\Phi)$ schreiben.

Der Buchstabe D in dieser Notation kommt vom Wort „darstellen“ her: Die Matrix stellt die Abbildung bezüglich gegebener Basen dar. Im Index $_{CB}$ merken wir uns die Basen, bezüglich derer dargestellt wird.

Wenn wir uns nun noch an die Koordinaten-Abbildungen

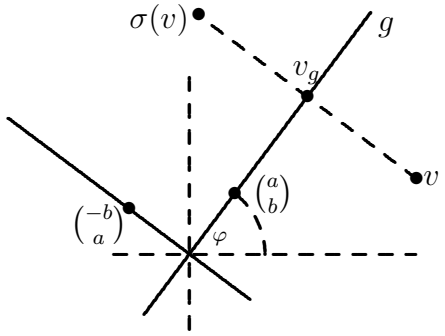
$$D_B : V \longrightarrow K^q, \quad D_C : W \longrightarrow K^p$$

erinnern, die jedem Vektor die Koordinaten bezüglich der betrachteten (geordneten) Basis zuordnen, so erhalten wir die folgende Merkregel:

$$\boxed{\textbf{Merkregel}\ D_C(\Phi(v)) = D_{CB}(\Phi) \cdot D_B(v).}$$

Beispiel 6.3.2 (Spiegelung an einer Geraden)

Als Beispiel hierfür wollen wir uns in der reellen Ebene $\mathbb{R}^2$ die Spiegelung σ an der Geraden g, die von $\binom{a}{b} \in V := \mathbb{R}^2$ mit $a^2 + b^2 = 1$ erzeugt wird, ansehen. Also:



Jedes $v \in \mathbb{R}^2$ lässt sich schreiben als $v = v_g + (v - v_g)$, wobei $v_g \in g$ liegt und $(v - v_g)$ auf g „senkrecht“ steht. Konkreter: $\binom{a}{b}$ und $\binom{-b}{a}$ bilden eine Basis von V, und es gilt:

$$v = \lambda_1 \begin{pmatrix} a \\ b \end{pmatrix} + \lambda_2 \begin{pmatrix} -b \\ a \end{pmatrix}$$

für eindeutig bestimmte $\lambda_1, \lambda_2 \in \mathbb{R}$.

Dann ist $\sigma(v) = \lambda_1 \binom{a}{b} - \lambda_2 \binom{-b}{a}$. Dies ist eine lineare Abbildung von V nach V, und wir wollen die Abbildungsmatrix $D_{SS}(\sigma)$ bezüglich der Standardbasis $S := \{e_1, e_2\}$ finden. Dazu müssen wir die Bilder von e_1 und e_2 unter σ berechnen, und erinnern uns deshalb an die Formeln

$$e_1 = a \begin{pmatrix} a \\ b \end{pmatrix} - b \begin{pmatrix} -b \\ a \end{pmatrix}, \quad e_2 = b \begin{pmatrix} a \\ b \end{pmatrix} + a \begin{pmatrix} -b \\ a \end{pmatrix}.$$

Wir sehen dann

$$\begin{aligned} \sigma(e_1) &= a\tbinom{a}{b} + b\tbinom{-b}{a} = \tbinom{a^2-b^2}{2ab} &&= (a^2 - b^2) \cdot e_1 &&+ 2ab \cdot e_2, \\ \sigma(e_2) &= b\tbinom{a}{b} - a\tbinom{-b}{a} = \tbinom{2ab}{b^2-a^2} &&= 2ab \cdot e_1 &&+ (b^2 - a^2) \cdot e_2. \end{aligned}$$

In die Abbildungsmatrix von σ bezüglich S gehören nun spaltenweise die Koeffizienten von $\sigma(e_1)$ und $\sigma(e_2)$ bezüglich e_1 und e_2, also gilt

$$D_{SS}(\sigma) = \begin{pmatrix} a^2 - b^2 & 2ab \\ 2ab & b^2 - a^2 \end{pmatrix}.$$

Wählen wir hierbei einen Winkel $\varphi \in \mathbb{R}$ so, dass $\binom{a}{b} = \binom{\cos(\varphi)}{\sin(\varphi)}$ gilt, dann folgt mit den Additionstheoremen von Sinus und Cosinus:

$$D_{SS}(\sigma) = \begin{pmatrix} \cos(2\varphi) & \sin(2\varphi) \\ \sin(2\varphi) & -\cos(2\varphi) \end{pmatrix}.$$

Aber eigentlich haben wir hierbei die ganze Zeit in einer anderen Basis gerechnet, die viel besser an das Problem angepasst ist: Bezüglich der Basis $B = \{\binom{a}{b}, \binom{-b}{a}\}$ gilt ja

$$D_{BB}(\sigma) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Moral:

Diese zweite Matrix ist viel einfacher als die erste, die wir nur deshalb hingeschrieben haben, weil wir gerne bezüglich der Standardbasis denken. Daraus lernen wir, dass es oft nützlich ist, den Standpunkt zu wechseln und die Bereitschaft aufzubringen, ein Problem aus einem anderen Blickwinkel als einem vorgegebenen zu betrachten.

Aufgabe 6.3.3 (Multiplikation in den komplexen Zahlen)

$\mathbb{C}$ ist ein zweidimensionaler reeller Vektorraum mit geordneter Basis $(1, \mathrm{i})$. Wenn $w = u + v\mathrm{i}$ eine feste komplexe Zahl ist, dann betrachten wir die Abbildung

$$\mu_w : \mathbb{C} \to \mathbb{C},\ z \mapsto w \cdot z.$$

Zeigen Sie, dass μ_w $\mathbb{R}$-linear ist und bestimmen Sie die Abbildungsmatrix $D(\mu_w)$ von μ_w bezüglich obiger Basis.

Zusatz: Rechnen Sie nach, dass die Abbildung $w \to D(w)$ ein injektiver Ringhomomorphismus von $\mathbb{C}$ nach $\mathbb{R}^{2\times 2}$ ist.

Beispiel 6.3.4 (duale Abbildung)

Es seien V und W endlichdimensionale K-Vektorräume und $\Phi : V \longrightarrow W$ ein Homomorphismus zwischen denselben. Dazu haben wir in 6.2.6 die duale Abbildung $\Phi^* : W^* \longrightarrow V^*$ kennen gelernt.

Die Abbildung Φ habe bezüglich zweier Basen B und C von V und W die Abbildungsmatrix

$$A := D_{CB}(\Phi).$$

Wie sieht die Abbildungsmatrix von Φ^* bezüglich der dualen Basen C^* und B^* von W^* und V^* aus? Nach 6.2.2 gilt für die Abbildungsmatrix A zunächst

$$A = (c_i^*(\Phi(b_j)))_{i,j}.$$

Die Definition von Φ^* sagt, dass

$$(\Phi^*(c_i^*))(b_j) = c_i^*(\Phi(b_j))$$

ist, also nach 6.2.2

$$\Phi^*(c_i^*) = \sum_j c_i^*(\Phi(b_j))b_j^*,$$

und da in $D_{B^*C^*}(\Phi^*)$ gerade die Koeffizienten von $\Phi^*(c_i^*)$ bezüglich B^* in der i-ten Spalte stehen, gilt

$$D_{B^*C^*}(\Phi^*) = D_{CB}(\Phi)^\top.$$

Bezüglich der dualen Basen wird die duale Abbildung durch die transponierte Matrix beschrieben.

Bemerkung 6.3.5 (Der Rang ist der Rang ist der Rang)

Es seien V und W endlichdimensionale K-Vektorräume.

Der Rang eines Homomorphismus $\Phi : V \longrightarrow W$ ist definiert als die Dimension des Bildraums:

$$\mathrm{Rang}(\Phi) = \dim(\mathrm{Bild}(\Phi)).$$

Dies ist gleich der Dimension des Vektorraums, der von den Spalten einer Abbildungsmatrix A von Φ (bezüglich irgendwelcher Basen) erzeugt wird. Diese Größe heißt der Spaltenrang von A. Andererseits sagt uns die Dimensionsformel 5.5.11 b), dass

$$\mathrm{Rang}(\Phi) = \dim(V) - \dim(\mathrm{Kern}(\Phi)).$$

Da aber offensichtlich

$$\dim(\mathrm{Kern}(\Phi)) = \dim \mathcal{L}(A, 0)$$

gilt, ist dies (nach 4.4.4 und Hilfssatz 4.3.3 b) und c)) auch der Rang von A wie er in 4.4.3 definiert wurde – die Fundamentallösungen sind eine Basis von $\mathcal{L}(A, 0)$.

6.4 Basiswechsel für Homomorphismen

Hinleitung zur Basiswechselformel

In diesem Abschnitt soll systematisch untersucht werden, wie sich aus der Abbildungsmatrix $A := D_{CB}(\Phi)$ eines Homomorphismus von V nach W bezüglich gegebener Basen $B = \{b_1, \ldots, b_q\}$ von V und $C := \{c_1, \ldots, c_p\}$ von W die Abbildungsmatrix von Φ bezüglich „neuer“ Basen $\widetilde{B}, \widetilde{C}$ berechnen lässt.

Dazu schreiben wir $\widetilde{b}_j \in \widetilde{B}$ als

$$\widetilde{b}_j = \sum_{i=1}^{q} s_{ij} b_i,$$

fassen also die Koeffizienten von $\widetilde{B}$ bezüglich B in einer Matrix $S = (s_{ij})_{1 \le i,j \le q} \in \mathrm{GL}_q(K)$ zusammen. Diese Matrix ist nichts anderes als

$$S = D_{B\widetilde{B}}(\mathrm{Id}_V).$$

Genauso schreiben wir ein $c_k \in C$ bezüglich $\widetilde{C}$ als

$$c_k = \sum_{l=1}^{p} t_{lk} \widetilde{c}_l,$$

also

$$T = (t_{lk})_{1\leq l,k\leq p} = D_{\widetilde{C}C}(\mathrm{Id}_W).$$

Dann ergibt sich für die Koeffizienten von $\Phi(\widetilde{b}_j)$ bezüglich $\widetilde{C}$ das Folgende:

$$\Phi(\widetilde{b}_j) = \sum_{i=1}^{q} s_{ij}\Phi(b_i) = \sum_{i=1}^{q} s_{ij} \sum_{k=1}^{p} a_{ki}c_k = \sum_{i=1}^{q} s_{ij} \sum_{k=1}^{p} a_{ki} \sum_{l=1}^{p} t_{lk}\widetilde{c}_l.$$

Daran lesen wir ab, dass die Abbildungsmatrix von Φ bezüglich $\widetilde{C}$ und $\widetilde{B}$ gegeben ist durch

Fazit 6.4.1

$$\boxed{\widetilde{A} := D_{\widetilde{C}\widetilde{B}}(\Phi) = TAS.}$$

Das führt uns zu der folgenden Definition.

Definition 6.4.2 (Äquivalenz von Matrizen)

Es seien $A, B \in K^{p\times q}$ zwei Matrizen. A und B heißen dann **äquivalent**, wenn es invertierbare Matrizen $S \in \mathrm{GL}_q(K)$ und $T \in \mathrm{GL}_p(K)$ gibt, sodass

$$B = TAS.$$

Nach dem Vorangehenden sind zwei Matrizen genau dann äquivalent, wenn sie dieselbe lineare Abbildung von K^q nach K^p bezüglich zweier Basenpaare beschreiben.

Aufgabe 6.4.3 (Eine Äquivalenzrelation)

Seien K ein Körper und $p, q \in \mathbb{N}$. Weisen Sie nach , dass die Äquivalenz von Matrizen in $K^{p\times q}$ tatsächlich eine Äquivalenzrelation ist.

Hilfssatz 6.4.4 (Äquivalenzklassen auf $K^{p\times q}$)

Es seien p, q natürliche Zahlen, K ein Körper und $m := \min(p, q)$. Dann gibt es in $K^{p\times q}$ genau $m + 1$ Äquivalenzklassen von Matrizen. Zwei Matrizen sind genau dann äquivalent, wenn sie denselben Rang haben.

Genauer ist die Menge der Blockmatrizen

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}, \quad 0 \leq r \leq m,$$

(mit Nullmatrizen von passender Größe) ein Vertretersystem der Äquivalenzklassen.

Beweis. Zwei äquivalente Matrizen $A, \widetilde{A}$ haben sicher denselben Rang, denn A und $\widetilde{A}$ beschreiben dieselbe Abbildung $\Phi : K^q \longrightarrow K^p$ bezüglich verschiedener Basen, und der Rang von A ist die Dimension des Bildvektorraumes von Φ, was auch für den Rang von $\widetilde{A}$ gilt.

Zu zeigen ist also noch, dass jede Matrix A vom Rang r zur Matrix $\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$ äquivalent ist. Dazu sei $\{v_{r+1}, \ldots, v_q\}$ eine Basis von $\mathcal{L}(A, 0)$, die durch Hinzunahme von weiteren Vektoren $\{v_1, \ldots, v_r\}$ zu einer Basis B von K^q ergänzt werden kann. Konkret könnte man für $\{v_{r+1}, \ldots, v_q\}$ die Fundamentallösungen in $\mathcal{L}(A, 0)$ aus 4.3.3 nehmen (neu durchnummerieren!) und ergänzen durch die Vektoren $v_i := e_{s_i}$, wobei $s_1, \ldots, s_r$ die Stufenindizes sind.

Die Vektoren $w_1 := Av_1, \ldots, w_r := Av_r$ erzeugen dann das Bild $A \cdot K^q$. Da der Bildraum AK^q die Dimension r hat, muss $\{Av_1, \ldots, Av_r\}$ eine Basis von AK^q sein. Diese lässt sich zu einer Basis $C = \{w_1, \ldots, w_p\}$ von K^p ergänzen. Es gilt dann:

$$A \cdot v_i = \begin{cases} w_i, & 1 \leq i \leq r, \\ 0, & i > r. \end{cases}$$

Das bedeutet aber

$$D_{CB}(\Phi_A) = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix},$$

sodass tatsächlich A in der Äquivalenzklasse dieser Matrix liegt. ◯

Genau betrachtet liefert uns dieser Beweis auch noch einmal einen Beweis der Dimensionsformel aus 5.5.12, der etwas greifbarer ist und keine Faktorräume benutzt.

Aufgabe 6.4.5 (Mit Zahlen)
Es seien $V = \mathbb{Q}^4$, $W = \mathbb{Q}^5$ und $\Phi : V \to W$ die Multiplikation mit der Matrix

$$A = \begin{pmatrix} 1 & 2 & 0 & -3 \\ 0 & 1 & -1 & 1 \\ 1 & 0 & -2 & 1 \\ 0 & 0 & 1 & -1 \\ 1 & -1 & 3 & -3 \end{pmatrix}.$$

Finden Sie Bases von V und W, sodass Φ bezüglich dieser Basen durch eine Matrix wie im letzten Hifssatz dargestellt wird.

Um die Basiswechselformel besser zu verstehen, leiten wir sie jetzt noch einmal her. Dabei verwenden wir den folgenden Hilfssatz, der noch einmal zeigt, wie passend die Notation für die Abbildungsmatrizen ist.

Hilfssatz 6.4.6 *Es seien U, V, W endlichdimensionale Vektorräume über dem Körper K mit Basen B, C, D. Weiter seien $\Phi : U \longrightarrow V$ und $\Psi : V \longrightarrow W$ Homomorphismen. Dann gilt*

$$D_{DB}(\Psi \circ \Phi) = D_{DC}(\Psi) \cdot D_{CB}(\Phi).$$

Beweis. Wir verwenden die Merkregel am Ende von 6.3.1. Diese sagt für alle Vektoren $u \in U$:

$$\begin{aligned} D_{DB}(\Psi \circ \Phi) \cdot D_B(u) &= D_D(\Psi(\Phi(u))) \\ &= D_{DC}(\Psi) \cdot D_C(\Phi(u)) \\ &= D_{DC}(\Psi) \cdot D_{CB}(\Phi) \cdot D_B(u). \end{aligned}$$

Da hierbei $D_B(u)$ jeder Vektor in $K^{\dim(U)}$ sein kann, folgt Gleichheit der Matrizen. ◯

Folgerung 6.4.7 (Basiswechselformel)

Es seien V, W zwei K-Vektorräume und Φ ein Homomorphismus von V nach W. Weiter seien B, $\widetilde{B}$ zwei Basen von V und C, $\widetilde{C}$ zwei Basen von W. Dann gilt

$$D_{\widetilde{C}\widetilde{B}}(\Phi) = D_{\widetilde{C}C}(\mathrm{Id}_W) \cdot D_{CB}(\Phi) \cdot D_{B\widetilde{B}}(\mathrm{Id}_V).$$

◯

Nach dem eben Gelernten gibt es nur sehr wenige verschiedene Typen von Homomorphismen zwischen zwei verschiedenen endlichdimensionalen Vektorräumen. Dies ändert sich grundlegend bei der Betrachtung von Endomorphismen. Bei diesen kann man ja Vektoren und ihre Bilder vergleichen und sollte dann nur noch Abbildungsmatrizen der Form $D_{BB}(\Phi)$ zulassen. Dies führt zu einer Theorie, die viel facettenreicher ist. Wir kommen in Kapitel 8 darauf zurück. Das nächste Kapitel bereitet das aber vor.

Kapitel 7

Determinanten

Die Determinante einer quadratischen Matrix (mit Einträgen in einem Körper) wird uns Aufschluss darüber geben, ob diese Matrix invertierbar ist oder nicht. Wie im Spezialfall der 2×2-Matrix in 4.2.9 ist also die Determinante einer quadratischen Matrix insbesondere ein Maß dafür, ob deren Spalten (oder Zeilen) linear abhängig sind oder nicht.

7.1 Die Determinantenform

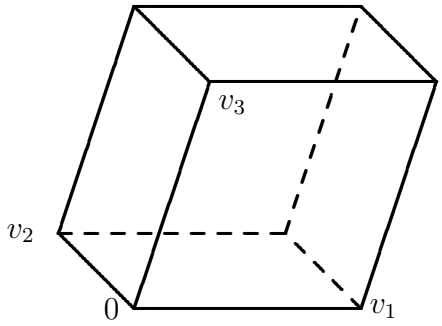
Zunächst wollen wir am Beispiel des dreidimensionalen Anschauungsraumes motivieren, was nachher zu den Axiomen der Determinantenform wird.

Bemerkung 7.1.1 (Parallelepipede und ihr Volumen)

Drei Vektoren v_1, v_2, v_3 im $\mathbb{R}^3$ spannen ein sogenanntes Parallelepiped auf, das ist die Menge

$$\mathcal{P}(v_1, v_2, v_3) := \{a_1v_1 + a_2v_2 + a_3v_3 \mid 0 \leq a_i \leq 1\}.$$

Es ist so etwas wie ein verbogener Quader:



Das Volumen $V(v_1, v_2, v_3)$ dieses Parallelepipeds berechnet sich als Grundfläche mal Höhe (oder Seitenfläche mal Breite, oder Frontfläche mal Tiefe, das ist alles gleich gut).

Die drei Vektoren sind genau dann linear unabhängig, wenn sie nicht in einer Ebene liegen, also wenn das Volumen nicht Null ist. Welche Eigenschaften hat die „Volumenfunktion“ V?

Das Volumen des Würfels mit Kantenlänge 1 ist 1:

$$V(e_1, e_2, e_3) = 1. \tag{1}$$

Wie ändert sich das Volumen, wenn man die Vektoren ändert?

Wenn man v_1 verdoppelt, dann verdoppelt sich das Volumen, genauso auch für v_2 und v_3. Allgemeiner gilt für positive $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$:

$$V(\lambda_1 v_1, \lambda_2 v_2, \lambda_3 v_3) = \lambda_1 \lambda_2 \lambda_3 V(v_1, v_2, v_3). \tag{2}$$

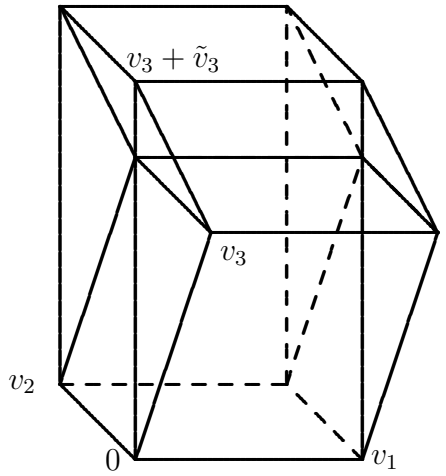
Wenn man v_2 parallel zu v_1 verschiebt, dann ändern sich weder Grundfläche noch Höhe, also bleibt auch das Volumen gleich:

$$V(v_1, v_2 + \lambda v_1, v_3) = V(v_1, v_2, v_3). \tag{3*}$$

Das folgt auch aus Regel (3), denn $V(v_1, \lambda v_1, v_3) = 0$, da das Parallelepiped ganz in einer Ebene liegt.

Wenn man zu v_3 einen anderen Vektor $\tilde{v}_3$ addiert, dann addieren sich die Volumina der zugehörigen Parallelepipede, denn die Höhen addieren sich. Genauso auch mit v_2 und v_1.

$$V(v_1, v_2, v_3 + \tilde{v}_3) = V(v_1, v_2, v_3) + V(v_1, v_2, \tilde{v}_3). \tag{3}$$



Aber jetzt sind wir auf ein suggestives Bild hereingefallen.

Denn wenn $\tilde{v}_3 = -v_3$ ist, dann hat das neue Parallelepiped Höhe Null, also Volumen Null, und das kann nicht die Summe der Volumina sein, denn diese sind positiv.

Man rettet sich, indem man eine Funktion einführt, die die beschriebenen und eigentlich ja auch gewünschten Eigenschaften hat, und zunächst von Fragen der Positivität absieht – diese spielen für beliebige Körper ohnehin keine sinnvolle Rolle. Insbesondere fordern wir die Gültigkeit von Gleichung (2) für alle Werte der λ_i.

Die so beschriebene Funktion heißt die Determinantenform. Sie lässt sich für jeden Körper definieren. Diese Determinantenform hat noch eine Eigenschaft, die sich aus der Gleichung (3) herleiten lässt, wenn man die Bemerkung nach Gleichung (3^*) noch berücksichtigt:

$$\begin{aligned} 0 \quad &= V(v_1+v_2, v_1+v_2, v_3) = V(v_1, v_1+v_2, v_3) + V(v_2, v_1+v_2, v_3) = \\ &= V(v_1, v_1, v_3) + V(v_1, v_2, v_3) + V(v_2, v_1, v_3) + V(v_2, v_2, v_3) = \\ &= V(v_1, v_2, v_3) + V(v_2, v_1, v_3). \end{aligned}$$

Das heißt:

$$V(v_1, v_2, v_3) = -V(v_2, v_1, v_3). \tag{4}$$

Beim Vertauschen zweier Vektoren ändert sich das Vorzeichen des „Volumens“.

Wir nehmen die eben gesehenen Eigenschaften als definierende Eigenschaften einer Abbildung von $(K^n)^n$ nach K für einen beliebigen Körper K.

Definition 7.1.2 (Determinantenform)

Es seien K ein Körper und $n \in \mathbb{N}$. Eine Abbildung

$$D : (K^n)^n \longrightarrow K$$

heißt eine **Determinantenform** auf K^n, wenn die folgenden Bedingungen erfüllt sind (dabei seien die Vektoren $w, v_1, \ldots, v_n \in K^n$ und $\alpha \in K$ beliebig):

D1 Für die Standardbasisvektoren gilt $D(e_1, e_2, \ldots, e_n) = 1$

D2 Für $1 \leq i \leq n$ gilt

$$D(v_1, \ldots, v_{i-1}, v_i+w, v_{i+1}, \ldots, v_n) = D(v_1, \ldots, v_n) + D(v_1, \ldots, v_{i-1}, w, v_{i+1}, \ldots, v_n)$$

D3 Für $1 \leq i \leq n$ gilt

$$D(v_1, v_2, \ldots, \alpha \cdot v_i, \ldots, v_n) = \alpha \cdot D(v_1, v_2, \ldots, v_n).$$

D4 Wenn für zwei Indizes $1 \leq i < j \leq n$ die Spalten v_i und v_j übereinstimmen, dann ist

$$D(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) = 0.$$

Aufgabe 7.1.3 (2×2-Matrizen)

Es sei K ein Körper. Aus 4.2.9 kennen wir bereits den Begriff der Determinante einer Matrix $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in K^{2\times 2}$, nämlich

$$\det(A) = ad - bc.$$

Rechnen Sie nach, dass die Abbildung

$$D : K^2 \times K^2 \to K, \quad D(\begin{pmatrix} a \\ c \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix}) = \det(\begin{pmatrix} a & b \\ c & d \end{pmatrix})$$

die Eigenschaften D1 bis D4 einer Determinantenform hat.

Folgerung 7.1.4 (Eigenschaften der Determinantenform)

Es sei D eine Determinantenform auf K^n.

a) Aus D2 und D3 folgt: Wenn (für irgendein i zwischen 1 und n) die Vektoren $v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n$ festgehalten werden, dann ist die Abbildung

$$v_i \mapsto D(v_1, \ldots, v_i, \ldots, v_n)$$

eine Linearform auf K^n. Man sagt für diese Eigenschaft, dass D eine n-fache **Multilinearform** ist.

b) Für $i \neq j$ und $\alpha \in K$ gilt (Addition eines Vielfachen der j-ten Spalte zur i-ten Spalte):

$$D(v_1, \ldots, v_i + \alpha v_j, \ldots, v_j, \ldots, v_n) = D(v_1, \ldots, v_n).$$

Denn die linke Seite ist wegen der Multilinearität

$$D(v_1, \ldots, v_n) + \alpha D(v_1, \ldots, v_j, \ldots, v_j, \ldots, v_n),$$

und der zweite Summand ist wegen D4 0.

c) Wenn für ein Paar (i, j) von Indizes mit $i < j$ die Vektoren in der i-ten und j-ten Spalte vertauscht werden, so ändert sich D dabei um den Faktor (-1):

$$D(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) = -D(v_1, \ldots, v_{i-1}, v_j, v_{i+1}, \ldots, v_{j-1}, v_i, v_{j+1}, \ldots, v_n).$$

Denn wir finden mit D2 und D4 aus 7.1.2:

$$\begin{aligned}
0 \overset{\text{D4}}{=}\ & D(v_1,\ldots,v_{i-1},v_i+v_j,v_{i+1},\ldots,v_{j-1},v_i+v_j,v_{j+1}\ldots,v_n)\\
\overset{\text{D2}}{=}\ & D(v_1,\ldots,v_{i-1},v_i,v_{i+1},\ldots,v_{j-1},v_i+v_j,v_{j+1}\ldots,v_n)\\
& +D(v_1,\ldots,v_{i-1},v_j,v_{i+1},\ldots,v_{j-1},v_i+v_j,v_{j+1}\ldots,v_n)\\
\overset{\text{D2}}{=}\ & D(v_1,\ldots,v_{i-1},v_i,v_{i+1},\ldots,v_{j-1},v_i,v_{j+1}\ldots,v_n)\\
& +D(v_1,\ldots,v_{i-1},v_i,v_{i+1},\ldots,v_{j-1},v_j,v_{j+1}\ldots,v_n)\\
& +D(v_1,\ldots,v_{i-1},v_j,v_{i+1},\ldots,v_{j-1},v_i,v_{j+1}\ldots,v_n)\\
& +D(v_1,\ldots,v_{i-1},v_j,v_{i+1},\ldots,v_{j-1},v_j,v_{j+1}\ldots,v_n)\\
\overset{\text{D4}}{=}\ & D(v_1,\ldots,v_{i-1},v_i,v_{i+1},\ldots,v_{j-1},v_j,v_{j+1}\ldots,v_n)\\
& +D(v_1,\ldots,v_{i-1},v_j,v_{i+1},\ldots,v_{j-1},v_i,v_{j+1}\ldots,v_n).
\end{aligned}$$

d) Für beliebige $v_1,\ldots,v_n \in K^n,\ \alpha_1,\ldots,\alpha_n \in K$ gilt die Gleichung

$$D(\alpha_1\cdot v_1,\ldots,\alpha_n\cdot v_n) = \alpha_1\cdot\alpha_2\cdot\ldots\cdot\alpha_n\cdot D(v_1,v_2,\ldots,v_n).$$

e) Um gleich den Gauß-Algorithmus ins Spiel zu bringen halten wir fest, dass damit für die Additions-, Vertauschungs- und Diagonalmatrizen aus Abschnitt 4.2 die folgenden Rechenregeln gezeigt sind. Wir fassen dazu $v_1,\ldots,v_n$ zu einer Matrix $M \in K^{n\times n}$ zusammen und betrachten D als Abbildung von $K^{n\times n}$ nach K. Dann formulieren sich die Eigenschaften D2, D3 und D4 bequemer so:

Bemerkung 7.1.5

Für eine Determinantenform D gilt für beliebiges $M \in K^{n\times n}$

$$\begin{array}{rll}
D(M\cdot A_{ij}(\alpha)) &= D(M) & \text{für } 1\le i,j\le n, i\ne j, \alpha\in K,\\
D(M\cdot V_{ij}) &= -D(M) & \text{für } 1\le i,j\le n, i\ne j,\\
D(M\cdot \operatorname{diag}(\alpha_1,\ldots,\alpha_n)) &= (\prod_{i=1}^n \alpha_i)\cdot D(M) & \text{für } \alpha_i\in K.
\end{array}$$

Diese Merkregel benutzen wir jetzt, um D auszurechnen, immer unter der Annahme, dass es eine Determinantenform überhaupt gibt, die wir erst im nächsten Abschnitt rechtfertigen werden.

Zunächst gilt wegen 7.1.2 a):

$$D(I_n) = 1.$$

Mit $M = I_n$ folgt dann aus der Merkregel insbesondere

$$D(A_{ij}(\alpha)) = 1,\quad D(V_{ij}) = -1,\quad D(\operatorname{diag}(\alpha_1,\ldots,\alpha_n)) = \prod_{i=1}^n \alpha_i.$$

Speziell ist $D(X) \neq 0$, wenn X eine Additions-, Vertauschungs- oder invertierbare Diagonalmatrix ist (dann sind die α_i alle ungleich 0). Diese Matrizen werden wir im Rest dieses Abschnitts immer als **spezielle Matrizen** bezeichnen.

Die Merkregel in Bemerkung 7.1.5 wird dabei zu

$$D(MX) = D(M) \cdot D(X),$$

wenn X eine spezielle Matrix ist.

Bemerkung 7.1.6 (Auftritt Gauß)

Wir benutzen den Gauß-Algorithmus: Nach dem Beweis von 4.4.2 gibt es spezielle Matrizen $X_1, \ldots, X_d$, sodass die Matrix $(M \cdot X_1 \cdot \ldots \cdot X_d)^\top$ Gauß-Normalform besitzt. Speziell gilt wegen 4.4.7:

$$M \cdot X_1 \cdot \ldots \cdot X_d = \begin{cases} I_n & , \text{ falls } \operatorname{Rang}(M) = n, \\ S & , \text{ falls } \operatorname{Rang}(M) < n, \end{cases}$$

wobei S eine $n \times n$-Matrix ist, deren letzte Spalte die Nullspalte ist. Es gilt daher

$$D(S) = D(S \cdot \operatorname{diag}(1, \ldots, 1, 0)) = 0.$$

Das führt zu

$$D(M) \cdot \prod_{i=1}^{d} D(X_i) = \begin{cases} 1 & , \text{ falls } \operatorname{Rang}(M) = n, \\ 0 & , \text{ falls } \operatorname{Rang}(M) < n. \end{cases}$$

Oder anders gesagt, weil die $D(X_i)$ alle von Null verschieden sind:

$$D(M) = \begin{cases} \prod_{i=1}^{d} D(X_i)^{-1} & , \text{ falls } \operatorname{Rang}(M) = n, \\ 0 & , \text{ falls } \operatorname{Rang}(M) < n. \end{cases}$$

Wir erinnern daran (siehe 4.4.7), dass der Rang von $M \in K^{n \times n}$ genau dann n ist, wenn M invertierbar ist, also in $\mathrm{GL}_n(K)$ liegt.

Folgerung 7.1.7 (wichtige Eigenschaften der Determinante)

Es seien K ein Körper und $n \in \mathbb{N}_0$.

Wenn es auf K^n eine Determinantenform D gibt, dann gelten die Regeln

a) $\forall M \in K^{n \times n} : \; D(M) \neq 0 \iff M \in \mathrm{GL}_n(K)$.

b) $\forall M, N \in K^{n \times n} : \; D(M \cdot N) = D(M) \cdot D(N)$.

c) Es gibt genau eine Determinantenform.

d) $\forall M \in K^{n\times n}:\ D(M) = D(M^\top)$.

Beweis. Nur die Regeln b) und d) sind noch zu zeigen.

b) Der Rang von $M \cdot N$ ist genau dann kleiner als n, wenn M oder N Rang kleiner als n hat. Also gilt für diesen Fall die Multiplikativität. Wenn M und N beide Rang n haben, so schreibt man sie sich als Produkte der speziellen Matrizen hin, und dieselben Faktoren kann man verwenden, um $M \cdot N$ als Produkt spezieller Matrizen zu schreiben. Also folgt die Behauptung aus 7.1.6.

d) Diese Regel sagt, dass sich beim Transponieren der Wert von D nicht ändert. Wir unterscheiden zwei Fälle.

Wenn M nicht regulär ist, dann ist auch $M^\top$ nicht regulär, und für beide ist nach a)

$$D(M) = 0 = D(M^\top).$$

Wenn M regulär ist, dann ist M ein Produkt

$$M = X_1 \cdot X_2 \cdot \ldots \cdot X_d,$$

wobei die X_ν spezielle Matrizen sind. Für diese gilt offensichtlich

$$D(X_\nu) = D(X_\nu^\top),$$

Dann folgt aber wegen b)

$$\begin{aligned} D(M^\top) &= D(X_d^\top \cdot X_{d-1}^\top \cdot \ldots \cdot X_1^\top) = \textstyle\prod_{i=1}^d D(X_i^\top) \\ &= \textstyle\prod_{i=1}^d D(X_i) = D(X_1 \cdot X_2 \cdot \ldots \cdot X_d) = D(M). \end{aligned}$$

○

Beispiel 7.1.8 (für die Berechnung einer Determinante)

Immer noch unter der Voraussetzung, dass es eine Determinantenform D auf K^n gibt, wollen wir einige Determinanten ausrechnen.

a) **Obere Dreiecksmatrizen** sind Matrizen der Gestalt

$$A = \begin{pmatrix} d_1 & * & \ldots & * \\ 0 & d_2 & * & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \ldots & 0 & d_n \end{pmatrix} \in K^{n\times n}.$$

Also: unterhalb der Diagonalen stehen nur Nullen als Einträge. Auf der Diagonalen stehen die Werte $d_1, \ldots, d_n$. Es gilt:

$$D(A) = d_1 \cdot d_2 \cdot \ldots \cdot d_n.$$

Um dies einzusehen unterscheiden wir zwei Fälle.

Wenn es ein i gibt mit $d_i = 0$, dann sind offensichtlich die ersten i Spalten von A linear abhängig (enthalten in $\langle e_1, \dots e_{i-1} \rangle$), also ist A nicht invertierbar und es ist

$$D(A) = 0 = d_1 \cdot d_2 \cdot \ldots \cdot d_i \cdot \ldots \cdot d_n.$$

Wenn keines der d_i Null ist, dann ist wegen Definition 7.1.2 c)

$$D(A) = d_1 \cdot d_2 \cdot \ldots \cdot d_n \cdot D(\tilde{A}),$$

wobei

$$\tilde{A} = \begin{pmatrix} 1 & * & \dots & * \\ 0 & 1 & * & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \dots & 0 & 1 \end{pmatrix}$$

eine obere Dreiecksmatrix ist, deren Diagonaleinträge sämtlich gleich 1 sind. Wir müssen zeigen, dass $D(\tilde{A}) = 1$ gilt. Dies ist wahr, da A sich als Produkt von Additionsmatrizen schreiben lässt: andere braucht man offensichtlich nicht um den Gauß-Algorithmus für $\tilde{A}$ durchzuführen. Als Produkt von Additionsmatrizen hat $\tilde{A}$ dann wegen 7.1.6 und Merkregel 7.1.5 die Determinante $D(\tilde{A}) = 1$.

b) Im folgenden Zahlenbeispiel benutzen wir zunächst Additionsmatrizen, nachher Diagonalmatrizen und am Ende Beispiel 7.1.8 a).

$$D\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & -1 & -2 \\ 1 & 4 & 1 & 4 \\ 1 & 8 & -1 & -8 \end{pmatrix} = D\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -2 & -3 \\ 0 & 3 & 0 & 3 \\ 0 & 7 & -2 & -9 \end{pmatrix} = D\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -2 & -3 \\ 0 & 0 & 6 & 12 \\ 0 & 0 & 12 & 12 \end{pmatrix}$$

$$= 6 \cdot 12 \cdot D\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -2 & -3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 \end{pmatrix} = 72 \cdot D\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -2 & -3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & -1 \end{pmatrix} = -72.$$

Aufgabe 7.1.9 (Ein Zahlenbeispiel)

Glauben Sie, dass es eine Determinantenform D auf $\mathbb{Q}^4$ gibt, und berechnen Sie

$$D\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \\ 3 & 4 & 1 & 2 \\ 4 & 1 & 2 & 3 \end{pmatrix}.$$

7.2 Die Leibnizformel

In diesem Abschnitt wird sichergestellt, dass es immer eine Determinantenform gibt. Anschließend dürfen wir die Aussagen über Determinanten aus dem letzten Abschnitt ohne Einschränkung verwenden.

Gleich brauchen wir das Signum einer Permutation, siehe 2.4.5 und die dort hergeleiteten Eigenschaften des Signums.

Definition 7.2.1 (Leibniz-Formel)

Es sei n eine natürliche Zahl und $A = (a_{ij}) \in K^{n\times n}$. Dann wird die **Determinante** $\det(A)$ von A definiert durch die **Leibniz-Formel**, nämlich

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

Die Abbildung det ist eine Abbildung von $K^{n\times n}$ nach K.

Satz 7.2.2 (Existenz der Determinantenform)

Die Abbildung

$$\det : K^{n\times n} \longrightarrow K$$

ist eine Determinantenform auf K^n.

Beweis. Wir müssen die vier Punkte aus der Definition abhaken.

a) Wir bezeichnen die Einträge der Einheitsmatrix mit dem „Kronecker-Delta“:

$$I_n = (\delta_{ij}), \quad \text{wobei} \quad \delta_{ij} := \begin{cases} 1 & \text{falls } i = j, \\ 0 & \text{falls } i \neq j. \end{cases}$$

Wenn $\sigma \in S_n$ eine Permutation ungleich der Identität ist, dann gibt es ein $i \in \{1, \ldots, n\}$ mit $i \neq \sigma(i)$, also $\delta_{i\sigma(i)} = 0$. Für so ein σ ist dann aber

$$\prod_{i=1}^{n} \delta_{i,\sigma(i)} = 0.$$

Also trägt zur Summe in der Leibniz-Formel für $\det(I_n)$ nur die Identität bei:

$$\det(I_n) = \prod_{i=1}^{n} \delta_{ii} = 1.$$

b) Wenn zur i_0-ten Spalte von A eine Spalte $(b_1 \ \dots \ b_n)^\top$ addiert wird und dadurch eine neue Matrix $\tilde{A}$ entsteht, dann sagt die Leibniz-Formel:

$$\begin{aligned}
\det(\tilde{A}) &= \sum_{\sigma\in S_n} \operatorname{sign}(\sigma) \prod_{i=1}^{n} \tilde{a}_{i,\sigma(i)} \\
&= \sum_{\sigma\in S_n} \operatorname{sign}(\sigma)[\prod_{\substack{i=1\\ \sigma(i)\neq i_0}}^{n} a_{i,\sigma(i)}] \cdot (a_{\sigma^{-1}(i_0),i_0} + b_{\sigma^{-1}(i_0)}) \\
&= \sum_{\sigma\in S_n} \operatorname{sign}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)} + \sum_{\sigma\in S_n} \operatorname{sign}(\sigma)[\prod_{\substack{i=1\\ \sigma(i)\neq i_0}}^{n} a_{i,\sigma(i)}] \cdot b_{\sigma^{-1}(i_0)} \\
&= \det(A) + \det(\hat{A}),
\end{aligned}$$

wobei $\hat{A}$ die Matrix ist, die aus A durch Ersetzen der i_0-ten Spalte durch b entsteht. Dann ist aber genau die Bedingung b) aus der Definition der Determinantenform erfüllt.

c) Wenn die i-te Spalte von A mit $\alpha \in K$ multipliziert wird, dann ändert sich die Leibnizformel dahingehend, dass in jedem Produkt der Faktor $a_{\sigma^{-1}(i),i}$ durch $\alpha \cdot a_{\sigma^{-1}(i),i}$ ersetzt wird. Der Faktor α lässt sich ausklammern, was in der Definition der Determinantenform die Bedingung c) impliziert.

d) Wenn die k-te und die l-te Spalte in $A \in K^{n\times n}$ gleich sind (mit $k \neq l$), dann bezeichnen wir wie in 2.4.1 b) mit $(k\ l) \in S_n$ die Transposition, die k und l vertauscht. In der Leibnizformel stimmen dann die Summanden zu $\sigma \in S_n$ und $(k\ l) \circ \sigma$ bis aufs Vorzeichen überein, denn in den Produkten

$$\prod_{i=1}^{n} a_{i,\sigma(i)} \quad \text{und} \quad \prod_{i=1}^{n} a_{i,(k\ l)\circ\sigma(i)}$$

stimmen die Faktoren für die i überein, für die $\sigma(i) \notin \{k,l\}$, und die für die zwei Werte von i mit $i \in \{k,l\}$ werden durch die Transposition $(k\ l)$ gerade vertauscht. Da aber die k-te und l-te Spalte übereinstimmen, ändert sich das Produkt insgesamt nicht. Allerdings ist $\operatorname{sign}(\sigma) = -\operatorname{sign}((k\ l) \circ \sigma)$, womit auch die Bedingung d) aus der Definition verifiziert ist. ◯

Bemerkung 7.2.3 (Endlich ohne Einschränkung)

Wir haben im ersten Abschnitt dieses Kapitels einige Eigenschaften der Determinantenform herggeleitet und dabei immer voraussetzen müssne, dass eine Deterinantenform existiert. Das wird nun durch den letzten Satz sichergestellt, und wir wissen damit, dass det alle Eigenschaften der ominösen Determinantenform D hat. Dies machen wir uns nun zu eigen und argumentieren gerne mit diesen Regeln, anstatt die Leibniz-Formel anzuwenden.

Aufgabe 7.2.4 (Cramers Regel)

Es seien $A \in K^{n\times n}$ gegeben sowie $b, x \in K^n$ sodass $Ax = b$. Weiter sei für $1 \leq i \leq n$ die Matrix A_i diejenige Matrix, die aus A entsteht, indem die i-te Spalte durch b ersetzt wird.

Nutzen Sie die Eigenschaften aus Definition 7.1.2 sowie die Tatsache, dass det eine Determinantenform ist, um zu zeigen, dass

$$\det(A_i) = x_i \cdot \det(A).$$

Ersetzen Sie dazu das b in A_i durch $\sum_{j=1}^{n} x_j A e_j$ und verwenden Sie D2, D3 und D4.

Folgern Sie im Fall, dass A regulär ist, dass

$$x_i = \frac{\det(A_i)}{\det(A)},$$

da wegen 7.1.7 a) reguläre Matrizen Determinante ungleich 0 haben.

Diese Gesetzmäßigkeit heißt Cramersche Regel.

Bemerkung 7.2.5 (Theorie und Praxis)

Die Leibniz-Formel sollten Sie bei der Berechnung von Determinanten möglichst vermeiden. Nur für kleine Werte von n (nämlich für $n = 0, 1, 2, 3$ und vielleicht noch $n = 4$) liefert sie eine sinnvolle Rechenmethode. Die Determinante einer 0×0-Matrix ist 1, denn die symmetrische Gruppe S_0 enthält ein Element, und für dieses hat man das leere Produkt auszuwerten, das 1 ist. Die Determinante der 1×1-Matrix (a) ist a. Weiter finden wir

$$\det(\begin{pmatrix} a & b \\ c & d \end{pmatrix}) = ad - bc,$$

und die Determinante einer 3×3-Matrix wird oft über die „Regel von Sarrus“ ausgerechnet, die hier aber nicht wiedergegeben wird, um den falschen Eindruck zu vermeiden, dass sie in irgendeiner anderen Form zu verallgemeinern sei als eben durch die Leibniz-Formel.

Wir halten aber fest:

Fazit 7.2.6

$$\begin{array}{|c|}\hline \forall A \in K^{n\times n} : A \text{ regulär} \Leftrightarrow \det(A) \neq 0. \\ \forall A, B \in K^{n\times n} : \det(AB) = \det(A) \cdot \det(B). \\ \hline \end{array}$$

Für die praktische Berechnung von Determinanten erinnern wir an Abschnitt 7.1 und legen eher das Gauß-Verfahren nahe, lernen nun jedoch noch ein Verfahren kennen, das allerdings auch eher in theoretischen Situationen und zur rekursiven Berechnung der Determinanten rekursiv definierter Familien von größer werdenden Matrizen Verwendung findet.

7.3 Die Laplace-Entwicklung

Hilfssatz 7.3.1 (Blockmatrizen)

Es seien $a, c \in \mathbb{N}_0$ *und* $A \in K^{a\times a}$, $C \in K^{c\times c}$. *Dann gilt für jede* $a \times c$*-Matrix* B*:*

$$\det \begin{pmatrix} A & B \\ 0 & C \end{pmatrix} = \det(A) \cdot \det(C).$$

Beweis. Wenn C Determinante 0 hat, dann sind die letzten c Zeilen von $\left(\begin{smallmatrix} A & B \\ 0 & C \end{smallmatrix}\right)$ linear abhängig, also

$$\det \begin{pmatrix} A & B \\ 0 & C \end{pmatrix} = 0 = \det(A) \cdot \det(C).$$

Wenn C Determinante ungleich 0 hat, dann gilt

$$\det \begin{pmatrix} A & B \\ 0 & C \end{pmatrix} = \det \begin{pmatrix} I_a & -BC^{-1} \\ 0 & I_c \end{pmatrix} \cdot \det \begin{pmatrix} A & B \\ 0 & C \end{pmatrix} = \det \begin{pmatrix} A & 0 \\ 0 & C \end{pmatrix},$$

denn der erste Faktor in der Mitte ist die Determinante einer Dreiecksmatrix mit Einsen auf der Diagonale, also nach 7.1.8 a) gleich 1.

Wenn $A = I_a$ die Einheitsmatrix ist, dann braucht man zur Berechnung von $\det(\left(\begin{smallmatrix} I_a & 0 \\ 0 & C \end{smallmatrix}\right))$ nur solche Spaltenumformungen, die man auch zur Berechnung von $\det(C)$ braucht. Also gilt

$$\det \begin{pmatrix} I_a & 0 \\ 0 & C \end{pmatrix} = \det(C) \text{ und analog } \det \begin{pmatrix} A & 0 \\ 0 & I_c \end{pmatrix} = \det(A).$$

Die Multiplikativität der Determinante (siehe 7.1.7 b)) liefert dann

$$\det \begin{pmatrix} A & 0 \\ 0 & C \end{pmatrix} = \det(\begin{pmatrix} I_a & 0 \\ 0 & C \end{pmatrix} \cdot \begin{pmatrix} A & 0 \\ 0 & I_c \end{pmatrix}) = \det(C) \cdot \det(A).$$

○

Beispiel 7.3.2 (Vorbereitung des Entwicklungssatzes)

Nun sei $A \in K^{n\times n}$ eine beliebige quadratische Matrix, $n \geq 1$. Für $1 \leq j \leq n$ sei A_{1j} die Matrix, die aus A durch Streichen der ersten Zeile und der j-ten Spalte entsteht, und s_j sei die Spalte, die aus der j-ten Spalte von A durch Streichen von a_{1j} entsteht. Wir schreiben uns nun die Spalten einzeln auf, zerlegen dann die erste Zeile als Summe von n Zeilen, was (wegen 7.1.2 für die transponierte

Matrix und wegen 7.1.7 d)) eine Summe für die Determinante ergibt, und dann vertauschen wir im i-ten Summanden zyklisch die ersten i Spalten:

$$\begin{aligned}
\det(A) &= \det\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1,n-1} & a_{1n} \\ s_1 & s_2 & \dots & s_{n-1} & s_n \end{pmatrix} \\
&\overset{7.1.2b)}{=} \det\begin{pmatrix} a_{11} & 0 & \dots & 0 \\ s_1 & s_2 & \dots & s_n \end{pmatrix} + \det\begin{pmatrix} 0 & a_{12} & 0 & \dots \\ s_1 & s_2 & s_3 & \dots \end{pmatrix} + \dots + \\
&\quad + \det\begin{pmatrix} 0 & \dots & 0 & a_{1n} \\ s_1 & s_2 & \dots & s_n \end{pmatrix} \\
&\overset{7.1.4c)}{=} \det\begin{pmatrix} a_{11} & 0 \\ s_1 & A_{11} \end{pmatrix} - \det\begin{pmatrix} a_{12} & 0 \\ s_2 & A_{12} \end{pmatrix} + \dots + \\
&\quad + (-1)^{n+1} \det\begin{pmatrix} a_{1n} & 0 \\ s_n & A_{1n} \end{pmatrix} \\
&\overset{7.3.1}{=} \textstyle\sum_{j=1}^{n} (-1)^{j+1} a_{1j} \cdot \det(A_{1j}).
\end{aligned}$$

Man sagt dazu, die Determinante von A wurde durch Entwicklung nach der ersten Zeile berechnet.

Was wir hier für die erste Zeile gemacht haben, geht genauso für jede andere Zeile und analog (zum Beweis kann man zum Beispiel transponieren) für jede Spalte. Das zeigt den folgenden Satz.

Satz 7.3.3 (Laplace-Entwicklung)

Es sei A eine $n \times n$-Matrix über dem Körper K. Für $1 \leq i, j \leq n$ sei A_{ij} die Matrix, die aus A durch Streichen der i-ten Zeile und der j-ten Spalte entsteht. Dann gilt für festes k zwischen 1 und n:

$$\det(A) = \sum_{j=1}^{n} (-1)^{j+k} a_{kj} \cdot \det(A_{kj}).$$

Diese Formel heißt Entwicklung der Determinante nach der k-ten Zeile.

Analog geht die Entwicklung nach der k-ten Spalte:

$$\det(A) = \sum_{i=1}^{n} (-1)^{k+i} a_{ik} \cdot \det(A_{ik}).$$

○

Beispiel 7.3.4 (Ein charakteristisches Polynom)

Es seien $t, a_0, \ldots, a_{n-1} \in K$. In Vorbereitung auf das nächste Kapitel, insbesondere 8.2.7 und 8.4.3, untersuchen wir die folgende Matrix:

$$A := \begin{pmatrix} t & 0 & 0 & \ldots & 0 & a_0 \\ -1 & t & 0 & \ldots & 0 & a_1 \\ 0 & -1 & t & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ & \ldots & & -1 & t & a_{n-2} \\ 0 & 0 & \ldots & 0 & -1 & t + a_{n-1} \end{pmatrix} \in K^{n \times n}.$$

Es gilt:

$$\det(A) = t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0 = t^n + \sum_{i=0}^{n-1} a_i t^i.$$

Dies sehen wir mit vollständiger Induktion nach n, wobei wir, um es konkret vor Augen zu haben, die Fälle $n \leq 2$ alle diskutieren:

Für $n = 0$ ist A die „leere“ Matrix mit Determinante $\det(A) = 1 = t^0 + \sum_{i=0}^{-1} a_i t^i$. Für $n = 1$ ist die Matrix A gerade die Matrix $(t + a_0)$ und diese hat Determinante $t + a_0$. Für $n = 2$ gilt

$$\det \begin{pmatrix} t & a_0 \\ -1 & t + a_1 \end{pmatrix} = t^2 + a_1 t + a_0.$$

Für beliebiges $n \geq 3$ gilt nun nach Laplace-Entwicklung nach der ersten Spalte

$$\det(A) = t \cdot \det(A_{11}) + 1 \cdot \det(A_{21}).$$

Hierbei kennen wir die Determinante von A_{11} nach Induktionsvoraussetzung,

$$\det(A_{11}) = t^{n-1} + \sum_{i=0}^{n-2} a_{i+1} t^i,$$

und Laplace-Entwicklung von A_{21} nach der ersten Zeile sagt uns:

$$\det(A_{21}) = (-1)^n \cdot (a_0) \cdot \det((A_{21})_{1,n-1}) = a_0,$$

da die $(n-2) \times (n-2)$-Matrix $(A_{21})_{1,n-1}$ eine obere Dreiecksmatrix ist, deren Diagonaleinträge alle gleich (-1) sind (in A wurden die ersten zwei Zeilen und die erste und letzte Spalte gestrichen); wir können also 7.1.8 zur Berechnung verwenden. Insgesamt folgt die Behauptung.

Aufgabe 7.3.5 (Die Vandermondematrix)

Um ein Polynom vom Grad $< d$ zu finden, das an d gegebenen Stellen Wunschwerte annimmt, kann man entweder die Interpolationsformel von Legendre benutzen

oder die Wünsche als Lineares Gleichungssystem an die Koeffizienten des Polynoms formulieren. Der zweite Ansatz führt auf folgende Situation:

Es sei K ein Körper und $a_1, \ldots, a_d \in K$ gegeben. Dann ist die zugehörige **Vandermondematrix** definiert als

$$V = (a_j^{i-1})_{1 \leq i,j \leq d} \in K^{d \times d}.$$

Machen Sie sich für $d = 3$ klar, wie diese Matrix aussieht.

Zegen Sie anschließend die allgemeine Formel

$$\det(V) = \prod_{1 \leq i < j \leq d} (a_j - a_i).$$

Das geht zum Beispiel mit vollständiger Induktion.

Definition/Bemerkung 7.3.6 (Adjunkte Matrix)

Für $A \in K^{n \times n}$ heißt die $(n \times n)$-Matrix $A^{\#} = (\alpha_{ij})_{1 \leq i,j \leq n}$ mit

$$\alpha_{ij} := (-1)^{i+j} \det(A_{ji})$$

die zu A **adjunkte Matrix**.

Es gilt $A \cdot A^{\#} = \det(A) \cdot I_n$. Denn: die Einträge auf der Diagonalen sind nach der Laplace-Entwicklungsformel gerade die Determinante von A, und der Eintrag an der Stelle (k, l) mit $l \neq k$ ist

$$\sum_{j=1}^{n} a_{kj} \cdot (-1)^{j+l} \det(A_{lj}) = \det(B) = 0,$$

wobei B die Matrix ist, die aus A entsteht, indem die l-te Zeile durch die k-te Zeile ersetzt wird (wieder wird Laplace benutzt!), die also zwei gleiche Zeilen und damit Determinante 0 hat.

Insbesondere sehen wir für $\det(A) \neq 0$ die Gleichung (vgl. 4.2.9)

$$\boxed{A^{-1} = (\det(A))^{-1} \cdot A^{\#}.}$$

Kapitel 8

Endomorphismen

Ein Endomorphismus eines Vektorraumes V ist eine lineare Abbildung $\Phi : V \longrightarrow V$. Wenn V endlichdimensional ist, so kann man wieder Φ durch Abbildungsmatrizen A beschreiben, wobei es jetzt sinnvoll ist, mit nur einer Basis B von V zu arbeiten:

$$A := D_{BB}(\Phi).$$

8.1 Basiswechsel

Bemerkung 8.1.1 (Basiswechsel bei Endomorphismen)

Wir erinnern uns an unser Vorgehen in Abschnitt 6.4 und übertragen die dortigen Ergebnisse in die sezielle Situaion von Endomorphismen.

Es seien V endlichdimensional, $\Phi \in \mathrm{End}(V)$ ein Endomorphismus von V und $B = \{b_1, \ldots, b_d\}$ eine Basis von V. Dann lässt sich $\Phi(b_j)$ schreiben als

$$\Phi(b_j) = \sum_{i=1}^{d} a_{ij} b_i,$$

und die Matrix $A = (a_{ij})_{1 \le i,j \le d} \in K^{d \times d}$ ist die Abbildungsmatrix von Φ bezüglich B:

$$D_{BB}(\Phi) = A.$$

Wenn wir nun von B zu einer neuen Basis $\widetilde{B}$ übergehen, dann gibt es dazu sogenannte Basiswechselmatrizen $S = D_{B\widetilde{B}}(\mathrm{Id}_V)$ und $T = D_{\widetilde{B}B}(\mathrm{Id}_V)$.

NB: Die Identität ist der einzige Endomorphismus, bei dem es hilfreich ist, zwei verschiedene Basen für die Abbildungsmatrix zu verwenden!

Es gilt dann aber wegen $D_{BB}(\mathrm{Id}_V) = D_{\widetilde{B}\widetilde{B}}(\mathrm{Id}_V) = I_d$:

$$TS = TD_{BB}(\mathrm{Id}_V)S = D_{\widetilde{B}\widetilde{B}}(\mathrm{Id}_V) = I_d = D_{BB}(\mathrm{Id}_V) = ST$$

oder kurz

$$T = S^{-1}.$$

Damit ergibt sich die Abbildungsmatrix $\widetilde{A}$ von Φ bezüglich $\widetilde{B}$ durch

$$\widetilde{A} = S^{-1}AS.$$

Wie in Abschnitt 6.4 führt uns diese Einsicht zu einer Definition.

Definition 8.1.2 (Ähnlichkeit)

Es sei d eine natürliche Zahl. Zwei Matrizen $A, \widetilde{A} \in K^{d\times d}$ heißen **ähnlich**, wenn es (mindestens) eine invertierbare Matrix $S \in \mathrm{GL}_d(K)$ gibt mit

$$\widetilde{A} = S^{-1}AS.$$

Manchmal sieht man auch $\widetilde{A} = TAT^{-1}$ für ein reguläres T als definierendes Kriterium, was dasselbe liefert, wenn man $T = S^{-1}$ setzt.

Bemerkung 8.1.3 (Ähnlichkeitsinvarianten, Spur)

Anders als im Fall der Äquivalenz von Matrizen gibt es für die Ähnlichkeit, die auch eine Äquivalenzrelation ist, im Allgemeinen unendlich viele Klassen. Es gibt genau dann nur endlich viele Klassen in $K^{d\times d}$ unter der Ähnlichkeitsrelation, wenn $d = 0$ gilt oder K endlich ist. In beiden Fällen ist ja $K^{d\times d}$ endlich und es kann nur endlich viele Klassen geben. Wenn aber $d \geq 1$ gilt, dann sind die Matrizen $\alpha \cdot I_d$ (mit $\alpha \in K$) paarweise verschieden. Da für jedes invertierbare $S \in \mathrm{GL}_d(K)$ die Gleichung

$$S^{-1} \cdot (\alpha I_d) \cdot S = \alpha I_d$$

gilt, repräsentieren diese Matrizen im Falle eines unendlichen Körpers unendlich viele paarweise verschiedene Äquivalenzklassen (die noch dazu jeweils aus nur einem Element bestehen).

Die nächsten Kapitel der Linearen Algebra werden sich eingehender mit der Frage beschäftigen, wie man (unter einer Nebenbedingung) eine leicht zu verstehende Matrix findet, die zu einer gegebenen Matrix ähnlich ist.

Um die Frage der Ähnlichkeit zweier Matrizen entscheiden zu können, ist es immer gut, wenn man einige Größen hat, die sich beim Übergang zu einer ähnlichen Matrix nicht ändern. Solche Größen nennt man **Ähnlichkeitsinvarianten**.

Wenn eine solche Ähnlichkeitsinvariante für zwei Matrizen verschiedene Werte annimmt, dann sind diese Matrizen nicht ähnlich. Die Übereinstimmung einiger

Ähnlichkeitsinvarianten wird hingegen meistens nicht ausreichen, um die Ähnlichkeit zu begründen. Es könnte ja noch eine weitere Ähnlichkeitsinvariante geben, wo die Werte verschieden sind...

Auf jeden Fall aber sind Ähnlichkeitsinvarianten eine hilfreiche Sache.

Zum Beispiel ist der Rang einer Matrix eine Ähnlichkeitsinvariante:

$$\forall A \in K^{d\times d}, S \in \mathrm{GL}_d(K): \ \mathrm{Rang}(A) = \mathrm{Rang}(S^{-1}AS)$$

Eine andere Ähnlichkeitsinvariante einer Matrix $A \in K^{d\times d}$ ist die Summe der Diagonalelemente; diese heißt die **Spur** von A:

$$\mathrm{Spur}(A) := \sum_{i=1}^{d} a_{ii}.$$

Deren Ähnlichkeitsinvarianz wird in Abschnitt 8.4 mit nachgewiesen; wir könnten sie aber auch hier schon direkt nachrechnen.

Aufgabe 8.1.4 (Die Ähnlichkeitsinvarianz der Spur)

Es seien $A = (a_{i,j}),\ B = (b_{i,j}) \in K^{n\times n}$ gegeben. Rechnen Sie nach, dass

$$\mathrm{Spur}(AB) = \sum_{i,j=1}^{n} a_{i,j}b_{j,i}$$

gilt, folgern Sie

$$\mathrm{Spur}(AB) = \mathrm{Spur}(BA)$$

und schließlich, falls B invertierbar ist,

$$\mathrm{Spur}(A) = \mathrm{Spur}(BAB^{-1}).$$

Definition/Bemerkung 8.1.5 (Determinante eines Endomorphismus)

Es seien V ein endlichdimensionaler Vektorraum über K und Φ ein Endomorphismus von V. Weiter sei B eine Basis von V und $D_{BB}(\Phi)$ die Abbildungsmatrix von Φ bezüglich B. Dann definieren wir die Determinante von Φ durch

$$\det(\Phi) := \det(D_{BB}(\Phi)).$$

Wir müssen dabei nachweisen, dass dies nicht von der Wahl von B abhängt. Wenn C eine weitere Basis von V ist, dann gibt es nach 8.1.1 eine invertierbare Matrix S, sodass

$$D_{CC}(\Phi) = S^{-1} \cdot D_{BB}(\Phi) \cdot S.$$

Daraus aber folgt zu unserer Beruhigung aufgrund der Multiplikativität der Determinante (siehe Folgerung 7.1.7 b))

$$\begin{aligned}\det(D_{CC}(\Phi)) &= \det(S^{-1} \cdot D_{BB}(\Phi) \cdot S) = \det(S)^{-1} \cdot \det(D_{BB}(\Phi)) \cdot \det(S) \\ &= \det(D_{BB}(\Phi)).\end{aligned}$$

Mit anderen Worten: die Determinante ist eine Ähnlichkeitsinvariante (siehe 8.1.3).

Aufgabe 8.1.6 (Ähnlich oder nicht?)

Entscheiden Sie bei den folgenden Matrizen $A, B, C \in \mathbb{Q}^4$, ob sie ähnlich sind oder nicht.

$$A = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}, B = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 1 & 2 \end{pmatrix}, C = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

8.2 Invariante Unterräume

Definition 8.2.1 (invarianter Untervektorraum)

Es seien V ein K-Vektorraum und Φ ein Endomorphismus von V. Ein Untervektorraum U von V heißt ein unter Φ **invarianter Untervektorraum**, wenn $\Phi(U) \subseteq U$ gilt. Für einen Φ-invarianten Untervektorraum U bezeichnen wir mit $\Phi|_U$ den Endomorphismus von U, der durch $u \mapsto \Phi(u)$ gegeben wird.

Im Gegensatz zum allgemeinen Gebrauch des Symbols $|_U$ machen wir also auch den Wertebereich kleiner, nicht nur den Definitionsbereich, siehe 1.3.16.

Im Fall $\Phi = \mathrm{Id}_V$ ist jeder Untervektorraum von V Φ-invariant.

Bemerkung 8.2.2 (Divide et impera!)

Allgemeiner heißt für eine Menge M und eine Abbildung $\Phi : M \longrightarrow M$ eine Teilmenge $U \subseteq M$ eine Φ-invariante Teilmenge, wenn $\Phi(U) \subseteq U$ gilt. Man kann solche Teilmengen oft benutzen, um die Abbildung Φ in kleinere Portionen zu zerlegen, die sich einzeln vielleicht besser verstehen lassen.

Aufgabe 8.2.3 (für invariante Untervektorräume)

Es sei $\Phi \in \mathrm{End}(V)$.

Zeigen Sie: Die Untervektorräume $\{0\}$, V, $\mathrm{Kern}(\Phi)$, $\mathrm{Bild}(\Phi)$ von V sind alle Φ-invariant.

Aufgabe 8.2.4 (Kommutierende Endomorphismen)

Seien V ein K-Vektorraum und $\Phi, \Psi : V \to V$ zwei Endomorphismen, für die $\Phi \circ \Psi = \Psi \circ \Phi$ gilt. Wir sprechen dann von **kommutierenden Endomorphismen**.

Zeigen Sie in diesem Fall, dass Kern(Φ) auch unter Ψ invariant ist, dass aber nicht jeder Φ-invariante Unterraum auch unter Ψ invariant sein muss.

Bemerkung 8.2.5 (Blockgestalt, Faktorraum)

Wenn V ein endlichdimensionaler K-Vektorraum ist und $U \leq V$ unter dem Endomorphismus Φ von V invariant bleibt, dann wählt man eine Basis $\widetilde{B} := \{b_1, \ldots, b_e\}$ von U und ergänzt sie zu einer Basis $B = \{b_1, \ldots, b_e, c_1, \ldots c_f\}$ von V mit $e+f = \dim(V)$. Bezüglich der Basis B hat dann Φ eine Abbildungsmatrix der folgenden **Blockgestalt**:

$$D_{BB}(\Phi) = \begin{pmatrix} D_1 & M \\ 0 & D_2 \end{pmatrix}, \quad D_1 \in K^{e\times e},\ M \in K^{e\times f}, 0 \in K^{f\times e}, D_2 \in K^{f\times f},$$

wobei 0 die Nullmatrix bezeichnet; dies gilt, da für $b_i \in \widetilde{B} \subseteq U$ der Vektor $\Phi(b_i)$ in U liegt, was die lineare Hülle von $\widetilde{B}$ ist.

Dabei ist $D_1 = D_{\widetilde{B}\widetilde{B}}(\Phi|_U)$ die Abbildungsmatrix des Endomorphismus $\Phi|_U$ von U.

Was ist D_2?

Wenn $\{c_1, \ldots, c_f\}$ einen Φ-invarianten Unterraum W aufspannt, dann ist $M = 0$, und D_2 beschreibt den Endomorphismus $\Phi|_W$ von W bezüglich der Basis $\{c_1, \ldots, c_f\}$. Leider gibt es nicht immer ein Φ-invariantes Komplement zu U, also hilft diese Erklärung nicht immer. Allgemein gilt das Folgende:

Durch Φ wird ein Endomorphismus des Faktorraumes V/U definiert. Denn für die Abbildung

$$\pi_{V/U} : V \longrightarrow V/U$$

gilt ja (siehe 5.5.7) Kern$(\pi_{V/U}) = U$, und da U ein Φ-invarianter Unterraum von V ist, ist U auch im Kern von $\pi_{V/U} \circ \Phi$. Dann gibt es aber nach dem Homomorphiesatz 5.5.8 eine lineare Abbildung

$$\widetilde{\Phi} : V/U \longrightarrow V/U,$$

sodass

$$\forall [v] \in V/U : \ \widetilde{\Phi}([v]) = [\Phi(v)].$$

Diese Abbildung nennt man die durch Φ auf V/U **induzierte Abbildung**.

Die Bilder von $c_1, \ldots, c_f$ in V/U biden dort eine Basis C (siehe 5.5.10), und es gilt für die obige Matrix D_2

$$D_{CC}(\widetilde{\Phi}) = D_2.$$

Beispiel 8.2.6 (invariante Unterräume)

a) Es sei $\Phi = \Phi_A : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$, wobei

$$A := \begin{pmatrix} 1 & 2 & -1 \\ 2 & 4 & -2 \\ 3 & 6 & -3 \end{pmatrix}.$$

Die Matrix A hat Rang 1, also einen zweidimensionalen Kern. Eine Basis von $U := \mathrm{Kern}(\Phi)$ besteht zum Beispiel aus

$$b_1 := \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \text{ und } b_2 := \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}.$$

Diese lässt sich durch $b_3 := e_3$ zu einer Basis B von $\mathbb{R}^3$ ergänzen, und bezüglich B beschreibt sich Φ durch die Abbildungsmatrix $D_{BB}(\Phi) = S^{-1}AS$, wobei

$$S = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 2 & 1 \end{pmatrix} \text{ und } S^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -2 & 1 \end{pmatrix}.$$

Das liefert

$$D_{BB}(\Phi) = \left(\begin{array}{cc|c} 0 & 0 & -1 \\ 0 & 0 & -2 \\ \hline 0 & 0 & 2 \end{array}\right).$$

Die Striche hier trennen optisch die verschiedenen Blöcke im Sinne von 8.2.5. Die 2×2-Matrix mit Nullen links oben beschreibt die Wirkung von Φ auf dem Kern von Φ, wo alles zu 0 gemacht wird. Die zwei Nullen darunter kommen von der Invarianz des Kerns her. Die letzte Spalte sagt, dass auf dem eindimensionalen Vektorraum V/U als Abbildung die Multiplikation mit 2 induziert wird.

Kann man das noch verbessern? Finden wir ein invariantes Komplement zu U? Das müsste ein eindimensionaler Untervektorraum sein, auf dem Φ die Multiplikation mit 2 ist, denn die auf V/U induzierte Abbildung kennen wir ja schon. Wir müssen also lösen:

$$A \cdot v = 2 \cdot v, \quad v \in \mathbb{R}^3, \ v \neq 0.$$

Also muss v im Bildraum von Φ sein, der eindimensional ist. Tatsächlich gilt die zu lösende Gleichung für

$$v = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Wenn wir nun diesen Vektor zum dritten Basisvektor erwählen: $\tilde{B} := \{b_1, b_2, v\}$, dann gilt:

$$D_{\tilde{B}\tilde{B}}(\Phi) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

b) Nun sei K ein beliebiger Körper, $V = K^2$, und der Endomorphismus Φ von V gegeben durch

$$\Phi(\begin{pmatrix} x \\ y \end{pmatrix}) := \begin{pmatrix} y \\ 0 \end{pmatrix}.$$

Der Rang von Φ ist 1, das Bild ist $K \cdot \binom{1}{0}$. Dies ist auch der Kern von Φ.

Wenn nun $U \subseteq V$ ein invarianter Unterraum $\neq \{0\}, V$ ist, dann ist seine Dimension 1. Gilt $\Phi(U) = \{0\}$, dann liegt U im Kern von Φ und stimmt aus Dimensionsgründen mit ihm überein. Ansonsten ist $\Phi(U) \subseteq U$ eindimensional und damit gleich dem Bild von Φ.

Daher ist Bild(Φ) der einzige nichttriviale Φ-invariante Unterraum und besitzt damit insbesondere keinen Φ-invarianten Komplementärraum. Hier ist also die „einfachste“ Matrix, die Φ beschreibt, die Abbildungsmatrix bezüglich der Standardbasis S:

$$D_{SS}(\Phi) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Definition 8.2.7 (zyklischer Untervektorraum)

Es seien V ein endlichdimensionaler Vektorraum und Φ ein Endomorphismus von V. Ein Φ-invarianter Untervektorraum U von V heißt ein **zyklischer Untervektorraum**, wenn es einen Vektor $u \in U$ gibt, sodass mit $d := \dim_K(U)$ gilt, dass

$$B := \{u, \Phi(u), \ldots, \Phi^{d-1}(u)\}$$

eine Basis von U ist. Das heißt: Es gibt einen Vektor $u \in U$, sodass U der kleinste u enthaltende Φ-invariante Untervektorraum von V ist.

Es gibt dann eindeutig bestimmte $a_0, \ldots, a_{d-1} \in K$, sodass

$$\Phi(\Phi^{d-1}(u)) = \Phi^d(u) = -\sum_{i=0}^{d-1} a_i \Phi^i(u).$$

Die Abbildungsmatrix von $\Phi|_U$ bezüglich der Basis B ist daher

$$D_{BB}(\Phi|_U) = \begin{pmatrix} 0 & 0 & \ldots & 0 & 0 & -a_0 \\ 1 & 0 & 0 & \ldots & 0 & -a_1 \\ 0 & 1 & 0 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & \ldots & 0 & 1 & 0 & -a_{d-2} \\ 0 & 0 & \ldots & 0 & 1 & -a_{d-1} \end{pmatrix},$$

Matrizen der Gestalt $tI_d - D_{BB}(\Phi|_U)$ haben wir schon gesehen, und zwar in 7.3.4. Das dort eingeführte Polynom (wir fassen t bald als Variable auf) werden wir in 8.4.3 allgemeiner zu einem wichtigen Objekt machen.

Die obige Matrix heißt die **Begleitmatrix** zum Polynom

$$a_0 + a_1X + a_2X^2 + \cdots + a_{d-1}X^{d-1} + X^d.$$

Auch in Aufgabe 4.4.9 sind uns schon Begleitmatrizen über den Weg gelaufen.

8.3 Eigenräume

Wir suchen jetzt nach den kleinstmöglichen Φ-invarianten Untervektorräumen, die vom Nullvektorraum verschieden sind. Diese sind idealer Weise eindimensional.

Definition 8.3.1 (Eigenvektoren und -werte)

Es sei Φ ein Endomorphismus des K-Vektorraumes V.

a) Ein Vektor $v \in V$ heißt ein **Eigenvektor** von Φ, wenn $\langle v\rangle = K \cdot v$ ein eindimensionaler Φ-invarianter Unterraum ist.

Konkreter heißt das:

$$v \neq 0 \text{ und } \exists \lambda \in K : \Phi(v) = \lambda \cdot v.$$

b) Das Element $\lambda \in K$ heißt ein **Eigenwert** von Φ, wenn es einen Eigenvektor v von Φ gibt, sodass $\Phi(v) = \lambda \cdot v$ gilt.

c) Die Menge aller Eigenwerte eines Endomorphismus Φ nennen wir sein **Spektrum** und schreiben dafür $\mathrm{Spec}(\Phi)$.

Definition/Bemerkung 8.3.2 („Säkulargleichung“, Eigenraum)

a) Die Gleichung $\Phi(v) = \alpha \cdot v$ heißt die **Eigenwertgleichung** oder auch **Säkulargleichung** (lat. saeculum = Jahrhundert). Dieser zweite Name entstammt der Astronomie. Beim Studium der Bahnbewegung von Planeten waren die Astronomen auf diese Gleichung gestoßen (für gewisse Endomorphismen Φ), deren Lösungen Periodizitätsaussagen für die Bahnbewegungen lieferten, also Konstellationen, die sich *cum grano salis* alle paar hundert Jahre wieder einstellen.

b) Die Eigenwertgleichung formen wir jetzt um. Es gilt ja

$$\begin{aligned}\Phi(v) = \alpha \cdot v \quad &\iff \Phi(v) - \alpha \cdot v = 0 \iff (\Phi - \alpha \cdot \mathrm{Id}_V)(v) = 0\\ &\iff v \in \mathrm{Kern}(\Phi - \alpha \cdot \mathrm{Id}_V).\end{aligned}$$

Insbesondere ist die Menge aller $v \in V$, die die Säkulargleichung für festes $\alpha \in K$ lösen, ein Φ-invarianter Untervektorraum von V. Er heißt der **Eigenraum** von Φ zu $\alpha \in K$, notiert als $\mathrm{Eig}(\Phi, \alpha)$:

$$\boxed{\mathrm{Eig}(\Phi, \alpha) := \mathrm{Kern}(\Phi - \alpha \cdot \mathrm{Id}_V)}$$

$\alpha \in K$ ist genau dann ein Eigenwert von Φ, wenn $\mathrm{Eig}(\Phi, \alpha) \neq \{0\}$.

Um für $\alpha \in K$ den Vektorraum $\mathrm{Eig}(\Phi, \alpha)$ auszurechnen, kann man wieder einmal den Gauß-Algorithmus benutzen, nachdem eine Abbildungsmatrix für Φ gewählt wurde. Wenn V die Dimension $n < \infty$ hat und Φ bezüglich einer Basis B durch $A := D_{BB}(\Phi)$ dargestellt wird, dann liegt ein Vektor $v \in V$ genau dann in $\mathrm{Eig}(\Phi, \alpha)$, wenn sein Koordinatenvektor $D_B(v)$ in $\mathcal{L}(A - \alpha I_n, 0)$ liegt. Außerdem hat Φ genau dann den Eigenwert α, wenn $\mathrm{Rang}(A - \alpha I_n) < n$. Die Dimension des Eigenraumes ist dann $n - \mathrm{Rang}(A - \alpha I_n)$.

Diese Zusammenhänge werden wir uns im nächsten Kapitel zunutze machen, um ein Verfahren zu finden, mit dem grundsätzlich alle Eigenwerte eines Endomorphismus eines endlichdimensionalen Vektorraumes bestimmt werden können.

c) Statt von Eigenwerten, -vektoren und -räumen von Endomorphismen spricht man auch von Eigenwerten, -vektoren und -räumen von (quadratischen) Matrizen. Dabei kann man zur Definition dieser Größen für die Matrix A die entsprechenden Größen des Endomorphismus Φ_A benutzen.

Beispiel 8.3.3 (für Eigenwerte und -vektoren)

a) Es sei $A = \mathrm{diag}(\alpha_1, \ldots, \alpha_n) \in K^{n \times n}$ eine Diagonalmatrix, $\Phi = \Phi_A$ der Endomorphismus von $V = K^n$, der durch Multiplikation mit A gegeben ist. Dann ist für $1 \leq i \leq n$

$$\Phi(e_i) = A \cdot e_i = \alpha_i \cdot e_i,$$

also ist der i-te Standardbasisvektor ein Eigenvektor von Φ zum Eigenwert α_i. Damit sind die Elemente $\alpha_1, \ldots, \alpha_n$ sicher Eigenwerte von Φ. Wenn $\lambda \notin \{\alpha_1, \ldots, \alpha_n\}$ ein weiteres Element aus K ist, dann gilt

$$A - \lambda I_n = \mathrm{diag}(\alpha_1 - \lambda, \alpha_2 - \lambda, \ldots, \alpha_n - \lambda),$$

und dies ist eine Diagonalmatrix, deren Diagonaleinträge alle von 0 verschieden sind. Also sind diese Einträge Einheiten in K, und damit ist $A - \lambda I_n$ invertierbar (siehe 4.2.7). Somit ist λ kein Eigenwert von Φ.

Wir halten fest: Die Eigenwerte eines Endomorphismus, der bezüglich einer geeigneten Basis durch eine Diagonalmatrix gegeben wird, sind genau die Diagonaleinträge dieser Diagonalmatrix.

$$\boxed{\mathrm{Spec}(\mathrm{diag}(\alpha_1, \ldots, \alpha_n)) = \{\alpha_1, \ldots, \alpha_n\}.}$$

b) Konkreter betrachten wir wieder einmal die Spiegelung σ an einer Geraden $\mathbb{R}\binom{a}{b}$ in der reellen Ebene, wobei $a^2+b^2=1$ gelte (siehe 6.3.2). Bezüglich der Basis $\{\binom{a}{b},\binom{-b}{a}\}$ hatten wir schon einmal die Abbildungsmatrix $\left(\begin{smallmatrix}1&0\\0&-1\end{smallmatrix}\right)$ ausgerechnet, die nach Beispiel a) zeigt, dass σ genau die Eigenwerte 1 und -1 hat. Die Spiegelachse $\mathbb{R}\binom{a}{b}$ ist der Eigenraum zum Eigenwert 1, die dazu „senkrechte" Gerade $\mathbb{R}\binom{-b}{a}$ ist der Eigenraum zum Eigenwert -1.

c) Nun wollen wir für einen Körper K untersuchen, wann die Matrix $A=\left(\begin{smallmatrix}0&-1\\1&0\end{smallmatrix}\right)$ Eigenwerte hat.

• *geometrisch:* wenn $K=\mathbb{R}$ gilt, dann beschreibt A die Drehung in der Ebene um 90° um den Nullpunkt, die offensichtlich keinen eindimensionalen invarianten Unterraum besitzt. Also gibt es für $K=\mathbb{R}$ keinen Eigenwert.

• Nun sei K beliebig. Dann ist $\lambda\in K$ genau dann ein Eigenwert von A, wenn der Rang von $A-\lambda I_2$ kleiner ist als 2, also wenn $A-\lambda I_2$ nicht invertierbar ist, also wenn (siehe 4.2.9)

$$\det(A-\lambda I_2)=\det(\begin{pmatrix}-\lambda & -1\\ 1 & -\lambda\end{pmatrix})=\lambda^2+1=0.$$

Die Nullstellen des Polynoms X^2+1 in K sind also die Eigenwerte von A. Folglich gibt es genau dann in K einen Eigenwert, wenn es ein Element $\lambda\in K$ gibt mit $\lambda^2=-1$. Dies ist zum Beispiel für $K=\mathbb{C}$ der Fall, oder auch für $K=\mathbb{F}_5$ (wähle $\lambda=[2]$), nicht aber für $K=\mathbb{F}_3$.

Satz 3.2.5 zeigt uns: selbst wenn A in K noch keinen Eigenwert besitzt, gibt es immer einen Körper L, der K enthält, und in dem ein Eigenwert von A liegt.

Beispiel 8.3.4 (Die Fibonacci-Zahlen)

Welche Eigenwerte hat die Matrix

$$M=\begin{pmatrix}0&1\\1&1\end{pmatrix}?$$

Ein Eigenvektor ist eine von Null verschiedene Spalte $v=\binom{x}{y}$ für die Mv ein Vielfaches von v ist. Konkret heißt das, dass ein γ existiert mit

$$y=\gamma x \text{ und } x+y=\gamma y.$$

Da x und y nicht beide 0 sind, muss $x\neq 0$ gelten, und wir dürfen $x=1$ setzen. Es folgt $y=\gamma$ und damit $1+\gamma=\gamma^2$. Das ist die Gleichung des **goldenen Schnittes**, und wir erhalten zwei Eigenwerte, nämlich $\gamma=\frac{1+\sqrt{5}}{2}$ (das ist der goldene Schnitt) und $\tilde{\gamma}=\frac{1-\sqrt{5}}{2}=\frac{-1}{\gamma}$.

Wir setzen $F_0=0, F_1=1, F_2=1$ und können dann

$$M=\begin{pmatrix}F_0&F_1\\F_1&F_2\end{pmatrix}$$

schreiben. Rekursiv ergibt sich für die Potenzen von M

$$M^n = \begin{pmatrix} F_{n-1} & F_n \\ F_n & F_{n+1} \end{pmatrix},$$

wobei $F_{n+1} = F_n + F_{n-1}$ gilt. Dies ist die Rekursionsvorschrift für die Folge der **Fibonacci-Zahlen**: $0, 1, 1, 2, 3, 5, 8, 13, 21, \ldots$

Wenn wir die beiden Eigenvektoren $(1\ \gamma)^\top$ und $(1\ \tilde{\gamma})^\top$ in die Matrix S setzen, so folgt

$$MS = S \cdot \begin{pmatrix} \gamma & 0 \\ 0 & \tilde{\gamma} \end{pmatrix}.$$

Also ist $S^-1MS = \text{diag}\gamma\ \tilde{\gamma}$. Dies gibt eine Alterative zur Berechnung der Potenzen von M :

$$M^n = S \cdot \begin{pmatrix} \gamma^n & 0 \\ 0 & \tilde{\gamma}^n \end{pmatrix} S^{-1}.$$

Damit lassen sich die Fibonacci-Zahlen konkret berechnen als

$$F_n = \frac{1}{\sqrt{5}}(\gamma^n - \tilde{\gamma}^n).$$

Im Umkehrschuss ergibt sich daraus, weil $|\gamma| > |\tilde{\gamma}|$ gilt, dass

$$\gamma = \lim_{n\to\infty} \frac{F_{n+1}}{F_n}.$$

In allen Beispielen, die wir eben gesehen haben, ist aus unseren Untersuchungen klar, dass die Anzahl der Eigenwerte eines Endomorphismus Φ von V höchstens so groß ist wie die Dimension von V. Dass dies allgemein gilt, folgt aus dem nächsten Hilfssatz. Vorher sehen wir uns mal wieder einen Graphen an.

Aufgabe 8.3.5 (k-reguläre Graphen)

Es sei $\Gamma = (E, K)$ ein Graph und $k \in \mathbb{N}$. Γ heißt **k-regulär**, wenn jede Ecke genau k benachbarte Ecken hat, also zu genau k Kanten gehört. Es sei $V = \text{Abb}(E, \mathbb{R})$.

Γ heißt **bipartit**, wenn E sich in zwei Teilmengen E_1 und E_2 zerlegen lässt, sodass keine Kante innerhalb E_1 oder E_2 verläuft, sondern eine Ecke aus E_1 mit einer aus E_2 verbindet.

Wir definieren den **Adjazenzoperator** $A : V \to V$ durch

$$\forall x \in E : (Af)(x) = \sum_{y\in E:\{x,y\}\in K} f(y),$$

das heißt für gegebenes $f \in V$ ist Af die Funktion, die bei x alle Funktionswerte von f bei zu x benachbarten Ecken aufsummiert.

Wir setzen jetzt voraus, dass Γ k-regulär ist. Zeigen Sie:

a) k ist der größte Eigenwert von A und die Dimension des Eigenraumes zu k ist die Anzahl der Zusammenhangskomponenten von Γ.

b) Γ ist genau dann bipartit, wenn auch $-k$ ein Eigenwert von A ist.

c) Γ ist genau dann bipartit, wenn für jeden Eigenwert λ von A auch $-\lambda$ ein Eigenwert von A ist.

Hilfssatz 8.3.6 (die Summe von Eigenräumen ist eine direkte Summe)
Es seien V ein K-Vektorraum, $\Phi \in \mathrm{End}(V)$, und $\lambda_1, \ldots, \lambda_n \in K$ paarweise verschieden. Dann gilt

$$\sum_{i=1}^{n} \mathrm{Eig}(\Phi, \lambda_i) = \bigoplus_{i=1}^{n} \mathrm{Eig}(\Phi, \lambda_i).$$

Beweis. Wir machen vollständige Induktion nach n. Für ($n = 0$ und) $n = 1$ ist die Behauptung klar. Sei also $n \geq 2$ und die Behauptung für $n-1$ bewiesen. Zu zeigen ist nun, dass es nur eine Möglichkeit gibt, 0 als Summe

$$0 = u_1 + u_2 + \cdots + u_n \qquad (*)$$

mit $u_i \in \mathrm{Eig}(\Phi, \lambda_i)$ zu schreiben, nämlich mit $u_1 = 0, u_2 = 0, \ldots, u_n = 0$. Wir nehmen nun $(*)$ als gegeben an. Dann folgt durch Anwendung von Φ :

$$0 = \Phi(0) = \Phi(\sum_{i=1}^{n} u_i) = \sum_{i=1}^{n} \Phi(u_i) = \sum_{i=1}^{n} \lambda_i \cdot u_i.$$

Andererseits macht die Multiplikation mit λ_n aus $(*)$ die Gleichung

$$0 = \sum_{i=1}^{n} \lambda_n \cdot u_i.$$

Die Differenz dieser zwei neuen Darstellungen der 0 liefert

$$0 = \sum_{i=1}^{n} (\lambda_i - \lambda_n) u_i = \sum_{i=1}^{n-1} (\lambda_i - \lambda_n) u_i.$$

Nun ist aber für $1 \leq i \leq n-1$ der Vektor $(\lambda_i - \lambda_n)u_i$ ein Vielfaches von u_i und damit in $\mathrm{Eig}\,(\Phi, \lambda_i)$ enthalten. Das erzwingt aber nach Induktionsvoraussetzung

$$\forall 1 \leq i \leq n-1 : (\lambda_i - \lambda_n)u_i = 0.$$

Da $\lambda_i - \lambda_n \neq 0$ vorausgesetzt ist, folgt $u_i = 0$ für $1 \leq i \leq n-1$.
In $(*)$ eingesetzt muss dann auch noch $u_n = 0$ gelten. ◯

Folgerung 8.3.7 (Spektrum und Dimension)

Es sei Φ ein Endomorphismus des endlichdimensionalen K-Vektorraumes V. Dann ist die Anzahl der Eigenwerte von Φ höchstens so groß wie die Dimension von V.

Beweis Es seien $\lambda_1, \ldots, \lambda_n \in \mathrm{Spec}(\Phi)$ paarweise verschieden. Dann ist für $1 \leq i \leq n$ die Dimension $\dim_K(\mathrm{Eig}(\Phi, \lambda_i))$ mindestens 1. Wegen 8.3.6 und wegen des Fazits in 5.4.2 gilt

$$\dim_K V \geq \dim_K(\bigoplus_{i=1}^{n} \mathrm{Eig}(\Phi, \lambda_i)) = \sum_{i=1}^{n} \dim_K(\mathrm{Eig}(\Phi, \lambda_i)) \geq n.$$

○

Definition 8.3.8 (Diagonalisierbarkeit)

Ein Endomorphismus Φ des K-Vektorraumes V heißt **diagonalisierbar**, wenn V eine Basis aus Eigenvektoren zu Φ besitzt.

Im endlichdimensionalen Fall wird Φ bezüglich solch einer Basis aus Eigenvektoren durch eine Abbildungsmatrix in Diagonalgestalt beschrieben, was den Namen erklärt und ebenfalls eine Charakterisierung der Diagonalisierbarkeit ist: Eine (beliebige) Abbildungsmatrix von Φ ist ähnlich zu einer Diagonalmatrix.

Eine weitere Möglichkeit, die Diagonalisierbarkeit zu charakterisieren, ist

$$V = \bigoplus_{\lambda \in \mathrm{Spec}(\Phi)} \mathrm{Eig}(\Phi, \lambda).$$

Dies wiederum ist wegen 5.4.2 und 8.3.6 äquivalent zu

$$\dim V = \sum_{\lambda \in \mathrm{Spec}(\Phi)} \dim \mathrm{Eig}(\Phi, \lambda).$$

Aufgabe 8.3.9 (Endomorphismen von Rang 1)

Es sei V ein endlichdimensionaler Vektorraum und Φ ein Endomorphismus von V vom Rang 1.

Welche Dimension hat der Eigenraum $\mathrm{Eig}(\Phi, 0)$?

Zeigen Sie, dass Φ genau dann diagonalisierbar ist, wenn es einen von 0 verschiedenen Eigenwert gibt, und dass dies genau dann der Fall ist, wenn die Spur von Φ nicht 0 ist.

Der Rest dieses Kapitels wird sich der Frage widmen, wie man Diagonalisierbarkeit nachweisen kann.

8.4 Das charakteristische Polynom

Wir wollen jetzt unter Ausnutzung der Eigenschaften der Determinante ein Verfahren sehen, mit dessen Hilfe die Eigenwerte eines Endomorphismus beziehungsweise einer quadratischen Matrix grundsätzlich bestimmt werden können. Dazu müssen wir uns wieder ein wenig Spielraum verschaffen. Zur Motivation:

Ein Element $\lambda \in K$ ist nach 8.3.2 b) genau dann ein Eigenwert von Φ, wenn $\mathrm{Kern}(\Phi - \lambda \mathrm{Id}_V) \neq \{0\}$ ist. Das wiederum ist wegen 5.5.11 b) dazu äquivalent, dass $\Phi - \lambda \mathrm{Id}_V$ nicht invertierbar ist. 7.1.7 a) sagt, dass dies genau dann der Fall ist, wenn

$$\det(\Phi - \lambda \mathrm{Id}_V) = 0.$$

Die Leibniz-Formel zeigt, dass dies eine polynomiale Bedingung an λ ist: die Determinante einer Matrix mit Einträgen im Polynomring ist wieder ein Polynom.

Bemerkung 8.4.1 (Ringe statt Körpern)

Wenn man an quadratischen Matrizen über einem kommutativen Ring R statt über einem Körper interessiert ist, dann kann mit der Leibniz-Formel noch immer eine Determinante definiert werden, die die Eigenschaften einer Determinantenabbildung hat. Allerdings können wir nicht mehr auf den Gauß-Algorithmus zurückgreifen, um die Eindeutigkeit zu zeigen. Die Laplace-Entwicklung und der Determinanten-Multiplikationssatz bleiben allerdings richtig, und es gilt

$$A \in \mathrm{GL}_n(R) \iff \det(R) \in R^\times.$$

Auch die Formel für die Adjunkte, $A \cdot A^\# = \det(A) \cdot I_n$, bleibt gültig.

Dies stimmt insbesondere für Teilringe von Körpern, für die man die Ergebnisse für Körper benutzen darf. Da wir Determinanten von Matrizen mit Einträgen im Polynomring brauchen, überlegen wir uns also nun, dass für jeden Körper K der Polynomring $K[X]$ in einem Körper enthalten ist.

Bemerkung 8.4.2 (Der Körper der rationalen Funktionen)

Es sei K ein Körper und $R = K[X]$ der Polynomring in einer Variablen X über K. Wir betrachten zunächst

$$S := R \times (R \setminus \{0\}) = \{(f, g) \mid f, g \in R,\ g \neq 0\}.$$

Auf S definieren wir die Äquivalenzrelation

$$(f, g) \sim (\tilde{f}, \tilde{g}) : \iff \tilde{g} \cdot f = g \cdot \tilde{f}.$$

Beim Nachrechnen dafür, dass das eine Äquivalenzrelation ist, braucht man die Nullteilerfreiheit von $K[X]$. Wir führen das hier nicht vor.

Nun sei $K(X) := S/\sim$ die Menge der Äquivalenzklassen in S, siehe 1.4.7, 1.4.8. Die Äquivalenzklasse von $(f,g) \in S$ notieren wir suggestiv als $\frac{f}{g}$. Das motiviert auch die Äquivalenzrelation, die wir nun so schreiben können:

$$\frac{f}{g} = \frac{\tilde{f}}{\tilde{g}} : \Longleftrightarrow \ \tilde{g} \cdot f = g \cdot \tilde{f}.$$

Addition und Multiplikation auf $K(X)$ definieren wir nun durch die vertrauten Formeln

$$\frac{f}{g} + \frac{k}{l} := \frac{fl + gk}{gl}, \quad \frac{f}{g} \cdot \frac{k}{l} := \frac{fk}{gl}.$$

Man muss (tun Sie das!) nachrechnen, dass diese Vorschriften wohldefiniert sind und aus $K(X)$ einen Körper machen, der $K[X]$ mithilfe des injektiven Ringhomomorphismus

$$K[X] \ni f \mapsto \frac{f}{1} \in K(X)$$

als Teilring enthält. $K(X)$ heißt der Körper der rationalen Funktionen (über K).

Wenn $R \neq \{0\}$ ein beliebiger kommutativer Ring ohne Nullteiler ist, dann lässt sich wortgleich zum eben Gesehenen ein Körper konstruieren, der R enthält. Dieser heißt der **Quotientenkörper** von R.

Das, was wir vorhin für quadratische Matrizen gemacht haben, wollen wir benutzen, um wieder die Frage nach den Eigenwerten von Endomorphismen aufzugreifen. Wir hatten ja in 8.3.2 gesehen, dass $\lambda \in K$ genau dann ein Eigenwert eines Endomorphismus Φ eines endlichdimensionalen Vektorraumes V ist, wenn $\Phi - \lambda \mathrm{Id}_V$ nicht regulär ist.

Definition 8.4.3 (Charakteristisches Polynom)

Es seien V ein n-dimensionaler K-Vektorraum mit Basis B und Φ ein Endomorphismus von V.

Das Polynom

$$\mathrm{CP}_\Phi(X) := \det(X \cdot I_n - D_{BB}(\Phi))$$

heißt das **charakteristische Polynom** von Φ.

Wir werden gleich sehen, dass es nicht von der gewählten Basis B abhängt.

Zunächst aber folgt aus der Bemerkung am Ende von 8.1.5 das wichtige

Fazit 8.4.4 (Nullstellen und Eigenwerte)

Die Eigenwerte von Φ sind genau die Nullstellen von $\mathrm{CP}_\Phi(X)$.

Das charakteristische Polynom hat Grad n und ist normiert (d.h. der Leitkoeffizient ist 1). Sein konstanter Term ist gleich $(-1)^n \cdot \det(\Phi)$, und der Koeffizient bei X^{n-1} ist $-\mathrm{Spur}(D_{BB}(\Phi))$, wobei die Spur wie in 8.1.3 definiert ist. Insbesondere sehen wir daran, dass die Spur eine Ähnlichkeitsinvariante ist, denn dies gilt auch für das charakteristische Polynom.

Bemerkung 8.4.5 (Ähnlichkeitsinvarianz des charakteristischen Polynoms)

Wie das charakteristische Polynom eines Endomorphismus wird auch das charakteristische Polynom einer Matrix $A \in K^{n\times n}$ definiert durch

$$\mathrm{CP}_A(X) := \det(XI_n - A).$$

Dies ist immer ein normiertes Polynom vom Grad n. Wie in 8.1.5 rechnet man nach, dass für eine zu A ähnliche Matrix $\widetilde{A} \in K^{n\times n}$ stets gilt:

$$\mathrm{CP}_A = \mathrm{CP}_{\widetilde{A}}.$$

Das charakteristische Polynom (einer quadratischen Matrix) ist also eine Ähnlichkeitsinvariante und das impliziert die in 8.4.3 versprochene Unabhängigkeit des charakteristischen Polynoms (eines Endomorphismus) vom Ergebnis einer Basiswahl.

Vorsicht: Es gibt Matrizen mit demselben charakteristischen Polynom, die nicht zueinander ähnlich sind, zum Beispiel die 2×2-Matrizen

$$A := \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \widetilde{A} := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Sie sind nicht ähnlich, da für beliebiges $S \in \mathrm{GL}_2(K)$ gilt:

$$S^{-1} \cdot A \cdot S = 0 = A \neq \widetilde{A}.$$

Aber beide haben das charakteristische Polynom X^2.

Aufgabe 8.4.6 (mit Zahlen)

Berechnen Sie das charakteristische Polynom der folgenden Matrix

$$\begin{pmatrix} 2 & 2 & 1 \\ 1 & 3 & 1 \\ 1 & 2 & 2 \end{pmatrix}$$

und finden Sie seine Nullstellen.

Tipp: Machen Sie Zeilenumformungen über dem Körper der rationalen Funktionen! Wenn sich beim Berechnen Linearfaktoren des charakteristischen Polynoms ergeben, sollten Sie das nicht wieder ausmultiplizieren. Wir suchen ja Nullstellen!

Definition 8.4.7 (algebraische und geometrische Vielfachheit)

Es seien Φ ein Endomorphismus eines endlichdimensionalen K-Vektorraums und $\lambda \in K$. Dann heißt

$$\mu_g(\Phi, \lambda) := \dim(\mathrm{Eig}(\Phi, \lambda))$$

die **geometrische Vielfachheit** von λ (für Φ). Die Zahl

$$\mu_a(\Phi, \lambda) := \max\{e \in \mathbb{N}_0 \mid 0 \leq e \leq \dim(V) \text{ und } (X-\lambda)^e \text{ teilt } \mathrm{CP}_\Phi(X)\}$$

heißt die **algebraische Vielfachheit** von λ (für Φ). Das ist die Nullstellenordnung der polynomialen Abbildung, die durch CP_Φ gegeben wird, im Punkt λ.

Es ist klar, dass

$$\mu_g(\Phi, \lambda) \geq 1 \iff \lambda \in \mathrm{Spec}(\Phi) \iff \mu_a(\Phi, \lambda) \geq 1.$$

Denn die erste Äquivalenz definiert geradezu die Eigenwerte und die zweite Äquivalenz nutzt aus, dass $\mathrm{CP}_\Phi(X)$ genau dann durch $(X-\lambda)$ teilbar ist, wenn $\mathrm{CP}_\Phi(\lambda) = 0$ (siehe 3.3.9).

Nun sei $\lambda \in \mathrm{Spec}(\Phi)$. Wir wählen eine Basis $\{b_1, \ldots, b_d\}$ von $\mathrm{Eig}(\Phi, \lambda)$ und ergänzen sie zu einer Basis $B := \{b_1, \ldots, b_e\}$ von V. 8.2.5 sagt uns

$$D_{BB}(\Phi) = \begin{pmatrix} \lambda \cdot I_d & C \\ 0 & D \end{pmatrix},$$

mit $C \in K^{d \times (e-d)}$ und $D \in K^{(e-d)\times(e-d)}$. Daraus aber ergibt sich

$$\begin{aligned} \mathrm{CP}_\Phi(X) &= \det \begin{pmatrix} (X-\lambda) \cdot I_d & -C \\ 0 & XI_{e-d} - D \end{pmatrix} \\ &= \det((X-\lambda) \cdot I_d) \cdot \det(XI_{e-d} - D) \\ &= (X-\lambda)^d \cdot \mathrm{CP}_D(X). \end{aligned}$$

Wir verwenden dabei 7.3.1.

Nach Konstruktion ist $d = \mu_g(\Phi, \lambda)$, und es folgt:

Fazit 8.4.8

$$\boxed{\text{Für } \lambda \in \mathrm{Spec}(\Phi) \text{ gilt } 1 \leq \mu_g(\Phi, \lambda) \leq \mu_a(\Phi, \lambda).}$$

Satz 8.4.9 (Diagonalisierbarkeit)

Es sei V ein endlichdimensionaler K-Vektorraum und Φ ein Endomorphismus von V. Dann ist Φ genau dann diagonalisierbar, wenn sich sein charakteristisches Polynom als Produkt von Linearfaktoren schreiben lässt und wenn außerdem gilt

$$\forall \lambda \in \mathrm{Spec}(\Phi) : \mu_g(\Phi, \lambda) = \mu_a(\Phi, \lambda).$$

Beweis.

Für diagonalisierbares Φ folgt das Kriterium durch Berechnen des charakteristischen Polynoms – wir dürfen ja eine Abbildungsmatrix in Diagonalgestalt benutzen.

Wenn umgekehrt das charakteristische Polynom ein Produkt von Linearfaktoren ist, dann gilt in Anlehnung an 8.3.8 und den daran anschließenden Satz

$$\mathrm{CP}_\Phi(X) = \prod_{\lambda \in \mathrm{Spec}(\Phi)} (X - \lambda)^{\mu_a(\Phi,\lambda)},$$

was $\sum_{\lambda \in \mathrm{Spec}(\Phi)} \mu_a(\Phi, \lambda) = \dim(V)$ erzwingt. Die Gleichheit der Multiplizitäten sagt dann auch

$$\sum_{\lambda \in \mathrm{Spec}(\Phi)} \mu_g(\Phi, \lambda) = \dim(V),$$

was wir schon im Zuge der Definition 8.3.8 als Kriterium der Diagonalisierbarkeit erkannt hatten. ◯

Aufgabe 8.4.10 (Entscheidung)

Entscheiden Sie, ob die Matrix aus Aufgabe 8.4.6 diagonalisierbar ist.

Satz 8.4.11 (Cayley-Hamilton)

Es seien V ein endlichdimensionaler K-Vektorraum und Φ ein Endomorphismus von V. Dann gilt

$$\mathrm{CP}_\Phi(\Phi) = 0.$$

Wir sagen dann auch: $\mathrm{CP}_\Phi(X) \in I(\Phi)$ *ist ein annullierendes Polynom für* Φ.

Beweis. Es seien $f(X) := \mathrm{CP}_\Phi(X)$ und $v \in V$. Wir zeigen, dass dann gilt

$$[f(\Phi)](v) = 0,$$

das heißt: $f(\Phi)$ annulliert jeden einzelnen Vektor in V und ist damit die Nullabbildung. Dazu betrachten wir den kleinsten Φ-invarianten Untervektorraum U von V, der v enthält. Er wird erzeugt von

$$v, \Phi(v), \Phi^2(v), \ldots, \Phi^n(v), \ldots,$$

und weil V und damit auch U endlichdimensional ist, gibt es ein minimales k, sodass

$$\{v, \Phi(v), \Phi^2(v), \ldots, \Phi^k(v)\}$$

linear abhängig ist:

$$\Phi^k(v) = -\sum_{i=0}^{k-1} a_i \Phi^i(v),$$

wobei wir das Vorzeichen für unsere Zwecke passend gewählt haben. Dann ist aber

$$\Phi^{k+1}(v) = \Phi(\Phi^k(v)) = \Phi(\sum_{i=0}^{k-1} a_i \Phi^i(v)) = \sum_{i=0}^{k-1} a_i \Phi^{i+1}(v) \ \ldots,$$

und daher ist

$$B_U := \{v, \Phi(v), \Phi^2(v), \ldots, \Phi^{k-1}(v)\}$$

eine Basis von U. Diese ergänzen wir zu einer Basis B von V, und 8.2.5 zeigt uns dann, dass Φ bezüglich dieser Basis durch eine Blockmatrix $\left(\begin{smallmatrix} A & * \\ 0 & C \end{smallmatrix}\right)$ beschrieben wird, wobei A eine $k \times k$-Matrix ist, $*$ eine passende Matrix, und $C \in K^{l \times l}$ eine Matrix, die Φ auf V/U beschreibt. Außerdem ist 0 die Nullmatrix passender Größe. Unsere Wahl von B_U zeigt uns mit dem Beispiel 8.2.7, dass

$$A = D_{B_U B_U}(\Phi|_U) = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & -a_0 \\ 1 & 0 & 0 & \ldots & 0 & -a_1 \\ 0 & 1 & 0 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ & \ldots & & 1 & 0 & -a_{n-2} \\ 0 & 0 & \ldots & 0 & 1 & -a_{n-1} \end{pmatrix} \in K^{k \times k}.$$

Dann gilt aber mit Hilfssatz 7.3.1

$$f(X) = \det(X \cdot I_n - D_{BB}(\Phi)) = \det(X \cdot I_k - A) \cdot \det(X \cdot I_l - C),$$

also ist (siehe Beispiel 7.3.4)

$$g(X) := \det(X \cdot I_k - A) = X^k + \sum_{i=0}^{k-1} a_i X^i$$

ein Teiler von f. Andererseits zeigt die Wahl der Basis B_U, dass

$$g(\Phi)(v) = 0,$$

also muss erst Recht auch

$$f(\Phi)(v) = 0$$

gelten, denn g teilt f. ◯

8.5 Polynome und Eigenwerte

Hilfssatz 8.5.1 ($f(\mathrm{Spec}(\Phi)) \subseteq \mathrm{Spec}(f(\Phi))$)

Es seien V ein K-Vektorraum, Φ ein Endomorphismus von V, und $f \in K[X]$ ein Polynom. Wenn $\lambda \in K$ ein Eigenwert von Φ ist, dann ist $f(\lambda)$ ein Eigenwert von $f(\Phi)$.

Beweis. Wir schreiben $f(X) = \sum_{i=0}^{d} a_i X^i$ mit $a_i \in K$. Außerdem wählen wir uns einen Eigenvektor $v \in \mathrm{Eig}(\Phi, \lambda)$. Induktiv sieht man, dass für alle $i \in \mathbb{N}_0$ die Identität

$$\Phi^i(v) = \lambda^i \cdot v$$

gilt, wobei wir uns sicherheitshalber an $\Phi^0 := \mathrm{Id}_V$ und $\lambda^0 = 1$ erinnern (siehe 3.3.7).

Dann haben wir aber auch

$$\begin{aligned}(f(\Phi))(v) &= (\textstyle\sum_{i=0}^{d} a_i \Phi^i)(v) = \sum_{i=0}^{d} a_i(\Phi^i(v)) \\ &= \textstyle\sum_{i=0}^{d} a_i(\lambda^i \cdot (v)) = (\sum_{i=0}^{d} a_i \lambda^i) \cdot (v) \\ &= f(\lambda) \cdot v.\end{aligned}$$

Also ist $f(\lambda)$ ein Eigenwert von $f(\Phi)$. Es gilt sogar

$$\mathrm{Eig}(\Phi, \lambda) \subseteq \mathrm{Eig}(f(\Phi), f(\lambda)).$$

○

Folgerung 8.5.2 *Wenn in der Situation von Hilfssatz 8.5.1 $f(\Phi) = 0$ gilt, dann ist das Spektrum von Φ in der Menge aller Nullstellen von f (in K) enthalten.*

Definition 8.5.3 (annullierende Polynome)

Es sei Φ ein Endomorphismus eines K-Vektorraumes.

a) Ein Polynom $f \in K[X]$ mit $f(\Phi) = 0$ heißt ein **annullierendes Polynom** von Φ.

b) Wir bezeichnen mit $I(\Phi) \subseteq K[X]$ die Menge aller annullierenden Polynome von Φ. Diese Menge heißt das **Verschwindungsideal** von Φ.

Gleich wird sich herausstellen, dass es in dieser Situation immer ein annullierendes Polynom gibt, das alle annullierenden Polynome teilt. Diese Tatsache spiegelt eine Eigenschaft des Polynomringes wieder: jedes Ideal in $K[X]$ ist ein Hauptideal. Etwas mehr dazu findet sich in Abschnitt 9.1.

Satz 8.5.4 (Existenz des Minimalpolynoms)

Es sei V ein endlichdimensionaler K-Vektorraum und Φ ein Endomorphismus von V. Dann gilt:

a) $I(\Phi) \neq \{0\}$, d.h. es gibt ein Polynom $f \in K[X]$, $f \neq 0$, sodass $f(\Phi) = 0$.

b) Es gibt ein Polynom $M \in I(\Phi)$ von kleinstmöglichem Grad ≥ 0 mit Leitkoeffizient 1.

c) Sei M wie in Teil b) gewählt. Dann existiert für alle Polynome $f \in I(\Phi)$ ein Polynom $g \in K[X]$, sodass

$$f = M \cdot g.$$

Beweis. a) Wir bezeichnen mit n die Dimension von V. Da $\mathrm{End}(V)$ zum Vektorraum der $n \times n$-Matrizen isomorph ist und damit Dimension n^2 hat, sind die $n^2 + 1$ Potenzen

$$\Phi^0, \Phi^1, \Phi^2, \ldots, \Phi^{n^2}$$

linear abhängig: es gibt also Koeffizienten $a_0, a_1, \ldots, a_{n^2}$, sodass nicht alle a_i Null sind, und trotzdem gilt

$$\sum_{i=0}^{n^2} a_i \Phi^i = 0.$$

Das aber heißt, dass Φ eine Nullstelle des von Null verschiedenen Polynoms $\sum_{i=0}^{n^2} a_i X^i$ ist.

b) Die Menge der Grade aller von Null verschiedenen Polynome in $I(\Phi)$ ist eine nichtleere Teilmenge von $\mathbb{N}_0$ und enthält daher ein minimales Element d. Sei $F \in I(\Phi)$ ein Element vom Grad d. Dann ist der Leitkoeffizient α von F von Null verschieden, also in K invertierbar, und auch das „normierte“ Polynom $\alpha^{-1}F =: M$ ist in $I(\Phi)$.

c) Wir nehmen an, es gebe ein Element von $I(\Phi)$, das sich nicht als Vielfaches von M schreiben lässt. Dann gibt es auch in $I(\Phi) \setminus K[X] \cdot M$ ein Element F von minimalem Grad. Dieses Element habe Grad $e \geq \mathrm{Grad}(M) = d$, da d der kleinstmögliche Grad von Elementen in $I(\Phi) \setminus \{0\}$ überhaupt ist. Es sei α der Leitkoeffizient von F. Dann ist

$$F - \alpha X^{e-d} \cdot M \in I(\Phi)$$

ein Polynom von kleinerem Grad als F, also muss es sich als $g \cdot M$ schreiben lassen. Das impliziert aber $F = \alpha X^{e-d} \cdot M + g \cdot M = (\alpha X^{e-d} + g) \cdot M$, im Widerspruch zur Annahme. ◯

Definition/Bemerkung 8.5.5 (Minimalpolynom)

Das Polynom M aus dem vorherigen Satz heißt das **Minimalpolynom** von Φ. Wir schreiben dafür $\mathrm{MP}_\Phi(X)$. Das Minimalpolynom ist eindeutig bestimmt. (Wieso?)

Analog kann man Verschwindungsideale und Minimalpolynome für quadratische Matrizen definieren. Zwei ähnliche quadratische Matrizen haben dasselbe Verschwindungsideal und dasselbe Minimalpolynom, das Minimalpolynom ist also eine Ähnlichkeitsinvariante (siehe 8.1.3).

Denn: Seien $A, B \in K^{n\times n}$ ähnlich. Dann gibt es nach Definition 8.1.2 ein $S \in \mathrm{GL}_n(K)$, sodass $B = S^{-1}AS$. Daher gilt für $i \in \mathbb{N}_0$:

$$B^i = B \cdot \ldots \cdot B = (S^{-1}AS) \cdot \ldots \cdot (S^{-1}AS) = S^{-1}A^iS.$$

Sei nun $f = \sum_{i=0}^{d} a_iX^i \in I(A)$. Dann gilt

$$f(B) = \sum_{i=0}^{d} a_iB^i = \sum_{i=0}^{d} a_iS^{-1}A^iS = S^{-1}f(A)S = 0.$$

Das zeigt $I(A) \subseteq I(B)$, und da Ähnlichkeit eine symmetrische Relation ist, gilt auch die umgekehrte Inklusion.

Folgerung 8.5.6 (MP teilt CP)

Aus 8.4.11 und aus 8.5.4/8.5.5 folgt, dass das Minimalpolynom eines Endomorphismus Φ eines endlichdimensionalen Vektorraumes V stets ein Teiler des charakteristischen Polynoms von Φ ist. Insbesondere ist der Grad des Minimalpolynoms höchstens gleich dem Grad des charakteristischen Polynoms, also höchstens gleich der Dimension von V.

Beispiel 8.5.7 (für Minimalpolynome)

a) Das Minimalpolynom einer Diagonalmatrix $D = \mathrm{diag}(\alpha_1, \ldots, \alpha_n)$ ist ein Teiler des annullierenden Polynoms $\prod_{i=1}^{n}(X - \alpha_i)$.

Da für ein Polynom $f \in K[X]$ die Gleichung

$$f(\mathrm{diag}(\alpha_1, \ldots, \alpha_n) = \mathrm{diag}(f(\alpha_1), \ldots, f(\alpha_n))$$

gilt, ist das Minimalpolynom von D das normierte Polynom kleinsten Grades, das $\alpha_1, \ldots, \alpha_n$ als Nullstellen hat.

Wenn $\{\alpha_1, \ldots, \alpha_n\} = \{\beta_1, \ldots, \beta_k\}$ aus genau k Elementen besteht, gilt also

$$\mathrm{MP}_D(X) = \prod_{i=1}^{k}(X - \beta_i).$$

b) Für ein $t \in K$ sei $A = \left(\begin{smallmatrix} 1 & t \\ 0 & 1 \end{smallmatrix}\right) \in K^{2\times 2}$. Dann ist das Polynom $(X-1)^2$ sicher ein annullierendes Polynom von A. Wenn $t = 0$ gilt, haben wir in Beispiel a) schon gesehen, dass $X-1$ das Minimalpolynom ist. Für $t \neq 0$ ist natürlich $A - I_2 \neq 0$, also ist das Minimalpolynom in diesem Fall notwendiger Weise $(X-1)^2$.

c) Das Minimalpolynom von $A = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right) \in K^{2\times 2}$ ist $X^2 + 1$.

Das Minimalpolynom eines Endomorphismus hat weitreichende Bedeutung für das Verständnis seiner Wirkung auf dem Vektorraum V. Dies wird in Kapitel 9 noch deutlicher werden. Wir fangen hier mit zwei Beobachtungen an. Die erste knüpft an Hilfssatz 8.5.1 und seine Folgerung an und sagt, dass für das Minimalpolynom noch mehr gilt als für beliebige annullierende Polynome.

Hilfssatz 8.5.8 (Eigenwerte und das Minimalpolynom)

Es sei Φ ein Endomorphismus des endlichdimensionalen K-Vektorraums V. Dann gilt:

$$\mathrm{Spec}(\Phi) = \{\lambda \in K \mid \mathrm{MP}_\Phi(\lambda) = 0\}.$$

Beweis. Da $\mathrm{MP}_\Phi(X)$ ein annullierendes Polynom von Φ ist, liefert die Folgerung aus 8.5.1 die Inklusion „$\subseteq$“ in der Behauptung.

Sei umgekehrt $\lambda \in K$ eine Nullstelle von $\mathrm{MP}_\Phi(X)$. Dann ist λ wegen 8.5.6 auch eine Nullstelle des charakteristischen Polynoms und damit wegen 8.4.4 auch ein Eigenwert. ◯

Hilfssatz 8.5.9 (Diagonalisierbarkeit und das Minimalpolynom)

Es seien V ein endlichdimensionaler K-Vektorraum und Φ ein Endomorphismus von V. Dann ist Φ genau dann diagonalisierbar, wenn sich das Minimalpolynom als Produkt

$$\mathrm{MP}_\Phi(X) = (X - \alpha_1) \cdot \ldots \cdot (X - \alpha_k)$$

schreiben lässt, wobei die α_i paarweise verschiedene Elemente aus K sind .

Beweis. Wenn Φ diagonalisierbar ist, dann gibt es eine diagonale Abbildungsmatrix A dafür, und es gilt $\mathrm{MP}_\Phi(X) = \mathrm{MP}_A(X)$. Aber das Minimalpolynom von A hat die gewünschte Gestalt, wie wir in Beispiel 8.5.7 a) gesehen haben.

Sei umgekehrt $\mathrm{MP}_\Phi(X) = \prod_{i=1}^{k}(X - \alpha_i)$ mit $\alpha_i \neq \alpha_j$ für $i \neq j$. Wir benutzen einen Spezialfall der Lagrangeschen Interpolationsformel, den wir aber auch begründen.

Für festes $i \in \{1, \ldots k\}$ sei f_i das Polynom

$$f_i(X) = \prod_{j \neq i}(X - \alpha_j).$$

Das ist ein Polynom vom Grad $k-1$ mit den $k-1$ Nullstellen α_j, $j \neq i$. Wegen $MP_\Phi = (X - \alpha_i) \cdot f_i$ gilt für alle Vektoren $v \in V$ die Relation

$$f_i(\Phi)(v) \in \mathrm{Eig}(\Phi, \alpha_i).$$

Nun sei $c_i := f_i(\alpha_i) = \prod_{j \neq i}(\alpha_i - \alpha_j)$.

Dann ist

$$f := \sum_{i=1}^{k} c_i^{-1} f_i$$

ein Polynom vom Grad $\leq k-1$, das an jeder der k Stellen α_i den Wert 1 annimmt. Daher ist das Polynom f konstant gleich 1, und insbesondere gilt $f(\Phi) = \mathrm{Id}_V$.

Das zeigt aber für alle $v \in V$ die Gleichheit

$$v = f(\Phi)(v) = \sum_{i=1}^{k} c_i^{-1} f_i(\Phi)(v) \in \sum_{i=1}^{k} \mathrm{Eig}(\Phi, \alpha_i).$$

Daher ist Φ diagonalisierbar. ◯

Aufgabe 8.5.10 (Transponieren)

Es sei $V = K^{n \times n}$ mit $n \geq 2$. Die Abbildung $V \ni A \mapsto A^\top \in V$, die A auf ihre transponierte Matrix abbildet (siehe 4.1.13), ist ein Endomorphismus T von V.

Weisen Sie nach, dass $X^2 - 1$ ein annullierendes Polynom von T ist. Welche Dimension hat $\mathrm{Eig}(T, 1)$?

Folgern Sie, dass $X^2 - 1$ das Minimalpolynom von T ist.

Welche Bedingung an K ist notwendig, damit T diagonalisierbar ist?

Kapitel 9

Normalform für Endomorphismen

Im letzten Kapitel haben wir über Endomorphismen gesprochen und die Frage aufgeworfen, ob sich zu jeder quadratischen Matrix A eine zu A ähnliche Matrix $\widetilde{A}$ finden lässt, für die schneller klar ist, wie die zugehörige Abbildung funktioniert, welche Eigenschaften sie hat, welche invarianten Unterräume.

Im letzten Kapitel haben wir die Frage nach der Diagonalisierbarkeit ein Stück weit geklärt, die ein Spezialfall der eingangs erwähnten Frage ist, und nun wollen wir noch einen Schritt weiter gehen und solch eine Matrix $\widetilde{A}$ finden, wenn das Minimalpolynom in Linearfaktoren zerfällt. Dies ist eine „technische“ Voraussetzung, damit $\widetilde{A}$ wirklich einfach zu verstehen ist. Eine allgemeinere Klassifikation von Endomorphismen lässt sich mittels des Elementarteilersatzes aus der Theorie der Hauptidealringe finden, wird hier aber nicht vorgeführt.

9.1 Der Polynomring

Vorweg wollen wir noch Aussagen über den Polynomring über K bereitstellen, denn ein wesentliches Hilfsmittel für das Folgende ist die Möglichkeit, mit polynomialen Ausdrücken des untersuchten Endomorphismus Φ zu arbeiten. Man erinnere sich etwa an die Abschnitte 8.4 und 8.5, in dem genau dieses Phänomen schon in Spezialfällen einen ersten Auftritt hatte. Zunächst erinnern wir an 3.3.6: die Einheitengruppe $K[X]^\times$ des Polynomrings über dem Körper K ist gleich $K^\times$. Nun machen wir einige Definitionen, die sich ohne Weiteres auch auf beliebige kommutative Ringe übertragen ließen.

Definition 9.1.1 (irreduzible Polynome)

a) Es sei K ein Körper. Ein Polynom $f \in K[X]$ heißt ein **Teiler** des Polynoms

$g \in K[X]$, wenn es ein Polynom $h \in K[X]$ gibt, sodass

$$g = f \cdot h.$$

b) Zwei Polynome $f, g \in K[X]$ heißen **teilerfremd**, wenn jedes $h \in K[X]$, das sowohl f als auch g teilt, eine Einheit in $K[X]$ ist.

c) Ein Polynom $f \in K[X] \setminus K[X]^\times$ heißt **irreduzibel**, wenn für jede Zerlegung $f = g \cdot h$ mit $g, h \in K[X]$ einer der beiden Faktoren eine Einheit ist.

Satz 9.1.2 (Division mit Rest)

Es seien K ein Körper und $f, g \in K[X]$ mit $g \neq 0$. Dann gibt es Polynome $h, r \in K[X]$ mit $\mathrm{Grad}(r) < \mathrm{Grad}(g)$, sodass

$$f - gh = r.$$

Beweis des Satzes. Wir machen vollständige Induktion nach dem Grad von f. Wenn f das Nullpolynom ist, ist nichts zu zeigen, denn $f = 0 \cdot g + 0$. Wenn f Grad 0 hat, also eine Konstante $\neq 0$ ist, dann unterscheiden wir zwei Fälle. Für $\mathrm{Grad}\,(g) = 0$ ist auch g eine Konstante ungleich 0, und es ist $f = (fg^{-1})g + 0$, also $h = fg^{-1}$, $r = 0$. Für $\mathrm{Grad}\,(g) > 0$ ist $f = 0 \cdot g + f$, also $h = 0, r = f$.

Nun nehmen wir an, der Grad von f wäre mindestens 1. Wenn der Grad von f kleiner als der Grad von g ist, dann wähle $h = 0$, $r = f$, und alles ist gezeigt. Wenn $\mathrm{Grad}\,(f) \geq \mathrm{Grad}(g)$, so gibt es ein $\lambda \in K^\times$ mit

$$\mathrm{Grad}(f_0) < \mathrm{Grad}(f), \quad \text{wobei} \quad f_0 = f - \lambda X^{\mathrm{Grad}(f) - \mathrm{Grad}(g)} \cdot g,$$

also gibt es nach Induktionsvoraussetzung h_0, r mit $\mathrm{Grad}\,(r) < \mathrm{Grad}(g)$ und

$$f_0 - h_0 g = r.$$

Dann ist aber für $h := h_0 + \lambda X^{\mathrm{Grad}(f) - \mathrm{Grad}(g)}$ und dasselbe r auch

$$f - h \cdot g = r.$$

○

Bemerkung 9.1.3 (euklidischer Algorithmus)

Dieser Satz ist die Grundlage für den **euklidischen Algorithmus** im Polynomring. Wenn $r = 0$ gilt, dann ist g ein Teiler von r. Ansonsten kann man mit (g, r) anstelle von (f, g) dasselbe Verfahren noch einmal durchlaufen und findet auf diese Art mit $r_0 := g$, $r_1 := r$ sukzessive Polynome r_i, die sich als $r_i = r_{i-1} - h_i r_{i-2}$ schreiben lassen, und deren Grad immer kleiner wird, bis (zwangsläufig nach höchstens $\mathrm{Grad}\,(g)$ Schritten) das Nullpolynom herauskommt. Wenn $r_i = 0$ gilt,

dann ist r_{i-1} ein Teiler von r_{i-2} und damit ist das letzte von Null verschiedene r_i ein Teiler von r_{i-1}. Zurückrechnend sieht man, dass dieses r_i ein Teiler auch von f und g ist. Da r_i aber auch von der Form

$$r_i = a_i f + b_i g, \quad a_i, b_i \in K[X],$$

ist, ist r_i ein gemeinsamer Teiler von f und g, der von jedem gemeinsamen Teiler von f und g geteilt wird. Man nennt dieses r_i einen **größten gemeinsamen Teiler** von f und g.

Bemerkung 9.1.4 Ein **Ideal** in einem kommutativen Ring R ist eine Untergruppe I der additiven Gruppe $(R, +)$, sodass außerdem für alle $r \in R$ und alle $i \in I$ gilt: $r \cdot i \in I$. Der Kern eines Ringhomomorphismus von R nach S (weiterer Ring) ist immer ein Ideal. Umgekehrt lässt sich (mit Faktorbildung) zu jedem Ideal I ein Ring S und ein Homomorphismus von R nach S konstruieren, der I als Kern hat. Das wird in der Algebra weiter thematisiert.

Ein Ideal heißt **Hauptideal**, wenn es ein $a \in I$ gibt mit

$$I = \{r \cdot a \mid r \in R\}.$$

Als Beispiel für ein Ideal im Polynomring $K[X]$ geben wir das Verschwindungsideal $I(\Phi)$ eines Endomorphismus Φ eines Vektorraums V an. Wir hatten gesehen, dass solch ein Ideal immer von einem Element M kleinsten Grades erzeugt wird. Dass dies allgemein für alle Ideale im Polynomring gilt, sieht man für $I \neq \{0\}$ mit dem Argument aus Satz 8.5.4, welches Sie nun mit dem Satz über die Polynomdivision vergleichen sollten.

Fazit 9.1.5

Jedes Ideal $I \subseteq K[X]$ ist ein Hauptideal.

Hilfssatz 9.1.6 (Ein Polynom mit allen Teilern)

Es seien K ein Körper, $n \geq 0$ eine natürliche Zahl und $a_1, \ldots, a_n \in K$ nicht notwendig verschieden. Es sei

$$f := (X - a_1) \cdot (X - a_2) \cdot \ldots \cdot (X - a_n) \in K[X].$$

Dann sind die normierten Teiler des Polynoms f genau die Polynome der Form

$$g = (X - a_{i_1}) \cdot (X - a_{i_2}) \cdot \ldots \cdot (X - a_{i_k}), \quad 1 \leq i_1 < i_2 < \ldots < i_k \leq n.$$

Beweis. Wir machen vollständige Induktion nach n. Für $n = 0$ ist $f = 1$ und hat nur den einen normierten Teiler 1.

Nun kommt der Induktionsschluss: Wenn für $n \geq 1$ die Gleichung

$$f = g_1 \cdot g_2$$

gilt, dann ist $g_1(a_1) = 0$ oder $g_2(a_1) = 0$. Also ist (mindestens) einer der beiden Faktoren durch $(X - a_1)$ teilbar (siehe 3.3.9), und wir haben ohne Einschränkung der Allgemeinheit: $(X - a_1)$ teilt g_1. Dann folgt

$$(X - a_2) \cdot \ldots \cdot (X - a_n) = (g_1/(X - a_1)) \cdot g_2.$$

Also sind g_2 und $g_1/(X - a_1)$ nach Induktionsvoraussetzung Polynome der gewünschten Gestalt, und dies gilt dann auch für g_1 und g_2. ◯

Hierbei haben wir benutzt, dass die Division durch $(X - a_1)$ ein eindeutig bestimmtes Polynom $f/(X - a_1)$ liefert. Die Eindeutigkeit des Ergebnisses der Division folgt daraus, dass $K[X]$ nullteilerfrei ist. Aus $(X - a_1)g = (X - a_1)\tilde{g}$ folgt $(X - a_1)(g - \tilde{g}) = 0$, also $g - \tilde{g} = 0$.

Folgerung 9.1.7 (Teilerfremdheit)

a) Es seien $f, g \in K[X]$ teilerfremde Polynome. Dann gibt es $k, l \in K[X]$ mit

$$1 = kf + lg.$$

b) Wenn $a \in K$ keine Nullstelle von g ist und $f = (X - a)^n$ für ein $n \in \mathbb{N}$, dann gibt es Polynome $k, l \in K[X]$ mit

$$1 = kf + lg.$$

Beweis. a) Die Elemente f und g erzeugen ein Ideal I in $K[X]$, nämlich

$$I := \{kf + lg \mid k, l \in K[X]\}.$$

Dieses Ideal wird von einem Element $M = kf + lg$ erzeugt. Der Erzeuger M ist ein Teiler sowohl von f als auch von g. Da jeder gemeinsame Teiler von f und g eine Einheit im Polynomring ist, also eine von Null verschiedene Konstante, gilt ohne Einschränkung $M = 1$. Dies zeigt die Behauptung.

b) Da $g(a) \neq 0$ gilt, sagt Hilfssatz 9.1.6, dass g und $(X - a)^n$ teilerfremd sind. Nun greift Teil a). ◯

Folgerung 9.1.8 Eine direkte Summe

Es sei V ein endlichdimensionaler K-Vektorraum und Φ ein Endomorphismus von V. Weiter sei

$$\mathrm{CP}_\Phi(X) = f \cdot g$$

ein Zerlegung des charakteristischen Polynoms von Φ in teilerfremde Polynome. Dann gilt

$$V = \mathrm{Kern}(f(\Phi)) \oplus \mathrm{Kern}(g(\Phi)).$$

Beweis. Wir wählen zuerst Polynome k, l mit

$$1 = kf + lg$$

wie in der vorherigen Folgerung 9.1.7.

Für ein $v \in \mathrm{Kern}(f(\Phi)) \cap \mathrm{Kern}(g(\Phi))$ gilt dann auch

$$v = \mathrm{Id}_V(v) = \Phi^0(v) = (k(\Phi)f(\Phi) + l(\Phi)g(\Phi))(v) = k(\Phi)(0) + l(\Phi)(0) = 0,$$

also ist der Durchschnitt der betrachteten Kerne gleich $\{0\}$.

Für jedes $v \in V$ gilt aber auch

$$g(\Phi)(v) \in \mathrm{Kern}(f(\Phi)),$$

da $g(\Phi) \cdot f(\Phi) = \mathrm{CP}_\Phi(\Phi) = 0$ nach dem Satz von Hamilton-Cayley. Analog ist $f(\Phi)(v) \in \mathrm{Kern}(g(\Phi))$.

Es folgt für beliebiges $v \in V$

$$v = \Phi^0(v) = f(\Phi)(k(\Phi)(v)) + g(\Phi)(l(\Phi))(v) \in \mathrm{Kern}(g(\Phi) + \mathrm{Kern}(f(\Phi)),$$

und daher ist die Summe dieser Räume tatsächlich ganz V.

Das zeigt die Behauptung. ◯

9.2 Haupträume

Wir wollen nun der Frage nachgehen, was einen Endomorphismus daran hindert, diagonalisierbar zu sein.

Bemerkung 9.2.1 (Hindernisse gegen die Diagonalisierbarkeit)

In 8.4.9 haben wir ein Kriterium für die Diagonalisierbarkeit kennen gelernt: ein Endomorphismus eines endlichdimensionalen Vektorraumes ist genau dann diagonalisierbar, wenn sein charakteristisches Polynom in Linearfaktoren zerfällt und für jeden Eigenwert die geometrische Vielfachheit mit der algebraischen Vielfachheit übereinstimmt.

Das erste Hindernis gegen die Diagonalisierbarkeit ist also, dass das charakteristische Polynom nicht in Linearfaktoren zerfällt. In Analogie zum Vorgehen in Abschnitt 3.2 könnte man dies aus der Welt räumen, indem man den Körper größer macht – eine Möglichkeit, die in der Algebra systematisch weiterverfolgt wird. Wir werden dieses Hindernis durch einfaches Ignorieren aus der Welt räumen und nur solche Endomorphismen untersuchen, für die es nicht existiert. Das ist zum Beispiel immer der Fall, wenn jedes nichtkonstante Polynom in $K[X]$ eine Nullstelle in K hat. Der **Fundamentalsatz der Algebra** sagt, dass dies für $K = \mathbb{C}$ stimmt. Diesen Satz beweisen wir hier nicht!

Jedenfalls bleibt die Frage, was passiert, wenn die Eigenräume zu klein sind, also wenn es Eigenwerte gibt, für die die algebraische und die geometrische Vielfachheit nicht übereinstimmen. Das ist gleichbedeutend damit, dass es zum Eigenraum kein Φ-invariantes Komplement gibt. Um die Situation in den Griff zu bekommen führt man den Begriff des Hauptraums ein. Dies ist der kleinste Φ-invariante Untervektorraum, der den Eigenraum enthält und einen Φ-invarianten Komplementärraum besitzt. Die Definition sieht aber ganz anders aus:

Definition 9.2.2 (Hauptraum)

Es seien $\Phi \in \mathrm{End}(V)$ und $\lambda \in K$. Dann heißt

$$H(\Phi, \lambda) := \bigcup_{k=0}^{\infty} \mathrm{Kern}((\Phi - \lambda \mathrm{Id}_V)^k)$$

der **Hauptraum** von Φ zu $\lambda \in K$.

Dies ist ein Untervektorraum von V, da für alle $k \in \mathbb{N}_0$ gilt:

$$\mathrm{Kern}((\Phi - \lambda \mathrm{Id}_V)^k) \subseteq \mathrm{Kern}((\Phi - \lambda \mathrm{Id}_V)^{k+1}).$$

Diese Inklusion sorgt auch dafür, dass jeder dieser Kerne $(\Phi - \lambda \mathrm{Id}_V)$-invariant ist. Da er erst Recht unter $\lambda \mathrm{Id}_V$ invariant ist, ist er auch Φ-invariant. Damit ist auch $H(\Phi, \lambda)$ ein Φ-invarianter Unterraum.

Bemerkung 9.2.3 (Haupträume enthalten Eigenräume)

Für $\lambda \in K$ ist der Hauptraum $H(\Phi, \lambda)$ genau dann vom Nullraum $\{0\}$ verschieden, wenn λ ein Eigenwert von Φ ist. Denn natürlich liegt $\mathrm{Eig}(\Phi, \lambda)$ in $H(\Phi, \lambda)$, also ist für Eigenwerte der Hauptraum nicht trivial; und umgekehrt, wenn $v \in H(\Phi, \lambda)$ nicht der Nullvektor ist, dann gibt es ein kleinstes $k > 0$ mit

$$(\Phi - \lambda \mathrm{Id}_V)^k(v) = 0,$$

also ist

$$0 \neq (\Phi - \lambda \mathrm{Id}_V)^{k-1}(v) \in \mathrm{Eig}(\Phi, \lambda)$$

und λ ist Eigenwert.

Ein Problem bei der Diagonalisierbarkeit ist, dass es zu einem Eigenraum kein invariantes Komplement geben muss. Der Hauptraum fängt dieses „Defizit“ auf. Dazu müssen wir die Argumente aus Satz 8.5.9 noch einmal ansehen und verfeinern.

Hilfssatz 9.2.4 (invariantes Komplement zum Hauptraum)

Es seien V ein endlichdimensionaler K-Vektorraum, $\Phi \in \mathrm{End}(V)$ und $\lambda \in \mathrm{Spec}(\Phi)$. Weiter sei $e := \mu_a(\Phi, \lambda)$ die algebraische Vielfachheit von λ. Wir zerlegen das charakteristische Polynom als

$$\mathrm{CP}_\Phi(X) = (X - \lambda)^e \cdot g, \quad g(\lambda) \neq 0.$$

Dann gelten die folgenden Aussagen:

a) $H(\Phi, \lambda) = \mathrm{Kern}((\Phi - \lambda \mathrm{Id}_V)^e)$.

b) $\dim(H(\Phi, \lambda)) = e$.

Beweis. Für den Beweis seien $H := \mathrm{Kern}((\Phi - \lambda \mathrm{Id}_V)^e)$ und $U := \mathrm{Kern}(g(\Phi))$.

Wir erinnern an 9.1.8:

$$V = H \oplus U.$$

Da λ keine Nullstelle von g ist, ist es auch kein Eigenwert der Einschränkung von Φ auf U, denn g ist ein annullierendes Polynom für diese Einschränkung. Daher ist für alle Zahlen $k \in \mathbb{N}$ die Einschränkung von $(\Phi - \lambda \mathrm{Id}_V)^k$ auf U injektiv, und es folgt

$$\mathrm{Kern}(\Phi - \lambda \mathrm{Id}_V)^{e+k} = \mathrm{Kern}(\Phi - \lambda \mathrm{Id}_V)^e.$$

Das zeigt a).

Da $V = H \oplus U$ die direkte Summe zweier Φ-invarianter Untervektorräume ist, gilt

$$\mathrm{CP}_\Phi(X) = \mathrm{CP}_{\Phi|_H}(X) \cdot \mathrm{CP}_{\Phi|_U}(X).$$

Da weiter $(X-\lambda)^e$ ein annullierendes Polynom für $\Phi|_H$ ist, ist mit dem Argument aus 9.1.8 das charakteristische Polynom von $\Phi|_H$ eine Potenz von $(X - \lambda)$. Der Exponent kann nicht kleiner sein als $e = \mu_a(\Phi, \lambda)$, da sonst $(X - \lambda)$ noch ein Teiler von $\mathrm{CP}_{\Phi|_U}(X)$ sein müsste, aber λ ist kein Eigenwert von $\Phi|_U$. Folglich ist $\mathrm{CP}_{\Phi|_H}(X) = (X - \lambda)^e$ und damit die Dimension von H gleich e.

Das wollten wir noch wissen. ◯

Als Vorbereitung auf die eigentlich interessante Zerlegung im nächsten Hilfssatz brauchen wir noch ein Resultat, das an die Beobachtung in 8.3.6 erinnern sollte.

Hilfssatz 9.2.5 (direkte Summe von Haupträumen)

Es seien V ein endlichdimensionaler K-Vektorraum und $\Phi \in \mathrm{End}(V)$. Die Elemente $\lambda_1, \ldots, \lambda_k \in K$ seien paarweise verschieden. Dann gilt:

$$\sum_{i=1}^{k} H(\Phi, \lambda_i) = \bigoplus_{i=1}^{k} H(\Phi, \lambda_i).$$

Beweis Wir schreiben kurz $H_i := H(\Phi, \lambda_i)$. Der Beweis geht induktiv nach k. Für $k = 0$ und $k = 1$ ist wieder nichts zu zeigen. Die Behauptung sei für $k-1$ wahr. Nun müssen wir sehen, dass aus

$$v_i \in H_i,\ v_1 + v_2 + \cdots + v_k = 0$$

zwangsläufig $v_1 = v_2 = \cdots = v_k = 0$ folgt. Dazu wählen wir erst einmal ein $e > 0$, sodass $(\Phi - \lambda_k \mathrm{Id}_V)^e(v_k) = 0$ gilt. Dies geht, da $v_k \in H_k$. Dann wenden wir $(\Phi - \lambda_k \mathrm{Id}_V)^e$ auf die obige Gleichung an:

$$0 = \sum_{i=1}^{k} (\Phi - \lambda_k \mathrm{Id}_V)^e(v_i) = \sum_{i=1}^{k-1} (\Phi - \lambda_k \mathrm{Id}_V)^e(v_i),$$

da der k-te Summand ja gerade von $(\Phi - \lambda_k \mathrm{Id}_V)^e$ annulliert wird.

Da $(\Phi - \lambda_k \cdot \mathrm{Id}_V)^e(v_i) \in H_i$ gilt, folgt nach Induktionsvoraussetzung, dass für $1 \leq i \leq k-1$ auch

$$(\Phi - \lambda_k \mathrm{Id}_V)^e(v_i) = 0$$

gelten muss.

Wähle für jedes solche i ein $f > 0$, sodass $(\Phi - \lambda_i \mathrm{Id}_V)^f(v_i) = 0$. So ein f gibt es. Die Polynome $(X - \lambda_k)^e$ und $(X - \lambda_i)^f$ sind teilerfremd, also gibt es Polynome $g, h \in K[X]$, sodass

$$g \cdot (X - \lambda_k)^e + h \cdot (X - \lambda_i)^f = 1.$$

Das erzwingt

$$\begin{aligned} v_i &= \mathrm{Id}_V(v_i) \\ &= [g(\Phi) \circ (\Phi - \lambda_k \mathrm{Id}_V)^e + h(\Phi) \circ (\Phi - \lambda_i \mathrm{Id}_V)^f](v_i) \\ &= [g(\Phi) \circ (\Phi - \lambda_k \mathrm{Id}_V)^e](v_i) \\ &= 0. \end{aligned}$$

Das zieht $v_k = 0$ nach sich, und wir sind fertig. ◯

Hilfssatz 9.2.6 (Wann ist V direkte Summe der Haupträume?)

Es seien V ein endlichdimensionaler K-Vektorraum, Φ ein Endomorphismus von V, $M := \mathrm{MP}_\Phi(X)$ das Minimalpolynom, und $C := \mathrm{CP}_\Phi(X)$ das charakteristische Polynom von Φ.

Dann sind die folgenden vier Aussagen äquivalent:

i) $V = \bigoplus_{\lambda \in \mathrm{Spec}(\Phi)} H(\Phi, \lambda)$.

ii) $C = \prod_{\lambda \in \mathrm{Spec}(\Phi)} (X - \lambda)^{\dim H(\Phi,\lambda)}$.

iii) *C lässt sich als Produkt von Linearfaktoren schreiben.*

iv) *M lässt sich als Produkt von Linearfaktoren schreiben.*

Beweis. i) $\Rightarrow$ ii): Aus dem Beweis von Teil a) des Hilfssatzes 9.2.4 folgt, dass das charakteristische Polynom der Einschränkung von Φ auf $H(\Phi, \lambda)$ gerade $(X - \lambda)^{\dim H(\Phi,\lambda)}$ ist. Da V als direkte Summe der Φ-invarianten Unterräume $H(\Phi, \lambda)$ vorausgesetzt wird, ist C das Produkt der charakteristischen Polynome von Φ auf diesen Teilräumen. Also:

$$C = \prod_{\lambda \in \mathrm{Spec}(\Phi)} (X - \lambda)^{\dim H(\Phi,\lambda)}.$$

ii) $\Rightarrow$ iii): Klar.

iii) $\Rightarrow$ iv): Klar wegen Hilfssatz 9.1.6, denn das Minimalpolynom teilt das charakteristische (8.4.11).

iv) $\Rightarrow$ i): Wir schließen induktiv nach der Anzahl der Nullstellen des Minimalpolynoms. Falls es keine Nullstelle gibt, so ist die Anzahl der Linearfaktoren von M gleich 0, da jeder Linearfaktor eine Nullstelle mit sich bringt. Also ist das Minimalpolynom $M = 1$, und $V = \{0\}$. Wenn es nur eine Nullstelle λ gibt, ist analog $M = (X - \lambda)^e$ für eine positive natürliche Zahl e, und damit ist $V = H(\Phi, \lambda)$ bereits ein Hauptraum.

Nun gebe es mindestens zwei Nullstellen, eine davon heiße λ. Dann zerlegen wir V gemäß dem Beweis von Hilfssatz 9.2.4 als

$$V = H(\Phi, \lambda) \oplus W$$

mit einem Φ-invarianten Komplement W zum Hauptraum. Dann hat das Minimalpolynom der Einschränkung von Φ auf W nicht λ als Nullstelle. Andererseits ist

$$M(X) = \mathrm{MP}_{\Phi|_W}(X) \cdot \mathrm{MP}_{\Phi|_{H(\Phi,\lambda)}}(X)$$

und damit hat das Minimalpolynom von Φ auf W eine Nullstelle weniger als M, weshalb sich W nach Induktionsvoraussetzung als direkte Summe von Haupträumen von $\Phi|_W$ schreiben lässt. Diese müssen dann aber gerade die von $H(\Phi, \lambda)$ verschiedenen Haupträume sein, die Φ auf V hat. ◯

Hiermit haben wir einen wichtigen Schritt getan und werden nun Φ auf jedem Hauptraum für sich untersuchen. Als erstes tun wir dies in Abschnitt 9.3 für den Eigenwert $\lambda = 0$. Danach übertragen wir das auf beliebige Eigenwerte und fassen alles im Satz über die Jordan'sche Normalform zusammen.

9.3 Nilpotente Endomorphismen

Definition/Bemerkung 9.3.1 (nilpotenter Endomorphismus)

Ein Endomorphismus Φ eines Vektorraumes V heißt **nilpotent**, wenn es eine natürliche Zahl n gibt, sodass Φ^n die Nullabbildung ist, wenn also X^n ein annullierendes Polynom von Φ ist.

Insbesondere ist wegen der Folgerung aus 8.5.1 die einzige Nullstelle 0 des Polynoms X^n auch der einzige Eigenwert von Φ (wenn nicht $V = \{0\}$ gilt und Φ deshalb gar keinen Eigenwert hat). Außerdem ist dann

$$V = \mathrm{Kern}(\Phi^n) = \mathrm{Kern}((\Phi - 0 \cdot \mathrm{Id}_V)^n)$$

der Hauptraum von Φ zu 0.

Für endlichdimensionales V und einen beliebigen Endomorphismus Φ von V ist die Einschränkung von $\Phi - \lambda \mathrm{Id}_V$ auf $H(\Phi, \lambda)$ immer nilpotent, wegen 9.2.4 ist ein möglicher Exponent $\dim(H(\Phi, \lambda))$. Dies werden wir im nächsten Abschnitt benutzen, um die Ergebnisse über nilpotente Endomorphismen auf allgemeine Endomorphismen zu übertragen.

Die Überlegungen aus 8.2.5 zeigen mit einem Induktionsargument (D_1 und D_2 sind dort nilpotent, e und f können tatsächlich kleiner als $\dim(V)$ gewählt werden, wenn $\dim(v) \geq 2$), dass ein nilpotentes Φ sich bezüglich einer geeignet gewählten Basis von V immer durch eine obere Dreiecksmatrix beschreiben lässt, analog (Umsortierung der Basis) auch durch eine untere Dreiecksmatrix. Insbesondere ist das charakteristische Polynom von Φ gleich $X^{\dim(V)}$. Allerdings ist der Ansatz dort nicht wirklich hilfreich für die Konstruktion der Jordan'schen Normalform, die wir jetzt mit einem Hilfssatz angehen. Hierfür brauchen wir den Begriff des zyklischen Unterraums aus Definition 8.2.7.

Hilfssatz 9.3.2 (zyklischer UVR mit invariantem Komplement)

Es seien V ein endlichdimensionaler K-Vektorraum und $\Phi \in \mathrm{End}(V)$ ein nilpotenter Endomorphismus. Weiter sei X^d das Minimalpolynom von Φ und $u \in V$ ein Vektor mit

$$\Phi^{d-1}(u) \neq 0.$$

(Solch einen Vektor gibt es, da sonst schon X^{d-1} ein annullierendes Polynom wäre und somit X^d nicht das Minimalpolynom von Φ sein könnte.)

Dann gibt es ein Φ-invariantes Komplement W zu dem Φ-invarianten zyklischen Unterraum $U := \langle u, \Phi(u), \ldots, \Phi^{d-1}(u) \rangle$.

Beweis. Es ist klar, dass U ein Φ-invarianter Untervektorraum ist. Außerdem ist die Dimension von U nicht größer als d, da U von d Elementen erzeugt wird.

Da andererseits das Minimalpolynom von $\Phi|_U$ wegen der Bedingung an u ein Vielfaches von X^d ist, ist $\dim(U) = \mathrm{Grad}(\mathrm{CP}_{\Phi|_U}) \geq \mathrm{Grad}(\mathrm{MP}_{\Phi|_U}) \geq d$. Es folgt $\dim(U) = d$.

Es sei $W \subseteq V$ ein maximaler Φ-invarianter Untervektorraum mit $U \cap W = \{0\}$. Solch ein W gibt es, denn die Menge aller Φ-invarianten Untervektorräume, deren Schnitt mit U nur aus der 0 besteht, ist nicht leer (denn $\{0\}$ gehört dazu), und dann gibt es auch solch einen Vektorraum maximaler Dimension, denn V ist endlichdimensional. Die Summe von U und W ist direkt.

Zu zeigen ist, dass W zu U komplementär ist, dass also $V = U \oplus W$ gilt. Wir nehmen das Gegenteil an: es gelte $V \neq U \oplus W$.

Sei $\tilde{v} \in V \setminus (U \oplus W)$ beliebig. Es gilt $\Phi^d(\tilde{v}) = 0 \in U \oplus W$, also gibt es ein minimales $e \in \mathbb{N}$ mit $\Phi^e(\tilde{v}) \in U \oplus W$. Sei

$$v := \Phi^{e-1}(\tilde{v}).$$

Dieses v ist nun so gewählt, dass folgendes gilt:

$$v \in V \setminus (U \oplus W) \text{ und } \Phi(v) \in U \oplus W.$$

Wir schreiben flugs

$$\Phi(v) = \sum_{i=0}^{d-1} a_i \Phi^i(u) + w, \quad w \in W,\ a_i \in K,$$

und wenden hierauf Φ^{d-1} an:

$$0 = a_0 \Phi^{d-1}(u) + \Phi^{d-1}(w).$$

Die anderen Summanden verschwinden, da $\Phi^d = 0$. Nach Voraussetzung ist $\Phi^{d-1}(u) \neq 0$. Wegen $\Phi^{d-1}(w) \in W$ und $U \cap W = \{0\}$ gilt $a_0 = 0$. Also ist

$$\Phi(v) = \Phi(\tilde{u}) + w, \quad \text{wobei } \tilde{u} = \sum_{i=1}^{d-1} a_i \Phi^{i-1}(u) \in U.$$

Nun setze

$$\widetilde{W} := W + K(v - \tilde{u}).$$

Dann ist $\widetilde{W}$ unter Φ invariant, da $\Phi(W) \subseteq W$ und $\Phi(v - \tilde{u}) = w \in W$. Außerdem ist $W \subset \widetilde{W}$ echt enthalten, da $v - \tilde{u} \notin W$. Schließlich haben $\widetilde{W}$ und U den Nullvektorraum als Schnitt, denn aus

$$u_1 = w_1 + \beta \cdot (v - \tilde{u}), \quad u_1 \in U, w_1 \in W,\ \beta \in K$$

folgt zunächst $\beta = 0$, da sonst $v = \beta^{-1}((u_1 + \beta \cdot \tilde{u}) - w_1) \in U + W$. Also sind auch u_1 und w_1 der Nullvektor, da $U \cap W = \{0\}$.

Aus $V \neq U + W$ folgt also, dass W nicht maximal sein kann: Widerspruch!

Damit gilt $V = U \oplus W$ wie erhofft. ◯

Fazit 9.3.3 Mit vollständiger Induktion nach $\dim(V)$ sehen wir nun, dass für nilpotentes Φ gilt:

$$\boxed{\Phi \text{ nilpotent } \Rightarrow V \text{ ist direkte Summe von zyklischen Unterräumen.}}$$

Denn dies stimmt für Dimensionen 0 und 1, und ansonsten gibt es in V einen von $\{0\}$ verschiedenen zyklischen Unterraum mit einem Φ-invarianten Komplementärraum, für den die Induktionsvoraussetzung angewendet werden kann.

Dies lässt sich nun auch auf Matrizenebene formulieren. Dazu betrachten wir die Abbildungsmatrix wie in 8.2.7 für den von u erzeugten Φ-zyklischen Untervektorraum von V, die für nilpotentes Φ die Gestalt

$$J_d(0) := \begin{pmatrix} 0 & 0 & \dots & 0 & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 1 & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 & 0 \end{pmatrix} = \sum_{i=2}^{d} E_{i,i-1}$$

hat. Man nennt sie ein **Jordankästchen der Länge d zum Eigenwert 0.**

Induktiv sieht man, dass für $0 \leq e \leq d$ die Gleichung

$$\mathrm{Rang}(J_d(0)^e) = d - e$$

gilt und dass der Rang von $J_d(0)^e$ für den Fall $e > d$ Null ist.

Satz 9.3.4 (Jordan'sche Normalform für nilpotente Matrizen)

Es sei $A \in K^{n\times n}$ eine nilpotente Matrix. Dann gibt es eindeutig bestimmte natürliche Zahlen k und $d_1 \geq d_2 \geq \cdots \geq d_k \geq 1$ mit $\sum_{i=1}^{k} d_i = n$, sodass A ähnlich ist zu der Matrix, die durch folgende Blockform gegeben ist:

$$\widetilde{A} := \begin{pmatrix} J_{d_1}(0) & 0 & 0 & \dots & 0 \\ 0 & J_{d_2}(0) & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \ddots & J_{d_{k-1}}(0) & 0 \\ 0 & \dots & 0 & 0 & J_{d_k}(0) \end{pmatrix}.$$

$\widetilde{A}$ heißt die Jordan'sche Normalform *von A.*

Beweis. Die Existenz von $d_1 \geq d_2 \geq \cdots \geq d_k \geq 1$ mit obigen Eigenschaften ist aus dem Vorhergehenden klar. Weil A und $\widetilde{A}$ zueinander ähnlich sind, lässt sich die Anzahl m_d der Jordankästchen mit Größe d auf folgende Art aus nur von A

abhängenden Daten berechnen. Nach der Bemerkung über den Rang von $J_d(0)^e$ gilt nämlich

$$\mathrm{Rang}(A^e) = \mathrm{Rang}(\widetilde{A}^e) = \sum_{d=e+1}^{n} (d-e)m_d.$$

Dies liefert ein quadratisches Lineares Gleichungssystem für die Zahlen m_d:

$$\begin{pmatrix} 1 & 2 & 3 & \dots & n-1 & n \\ 0 & 1 & 2 & \dots & n-2 & n-1 \\ 0 & 0 & 1 & \ddots & \vdots & n-2 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \\ 0 & \vdots & \ddots & 0 & 1 & 2 \\ 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ \vdots \\ m_{n-1} \\ m_n \end{pmatrix} = \begin{pmatrix} \mathrm{Rang}(A^0) \\ \mathrm{Rang}(A^1) \\ \mathrm{Rang}(A^2) \\ \vdots \\ \mathrm{Rang}(A^{n-2}) \\ \mathrm{Rang}(A^{n-1}) \end{pmatrix}.$$

Die Inverse zur Koeffizientenmatrix lässt sich einfach berechnen, und es folgt

$$\forall d : m_d = \mathrm{Rang}(A^{d-1}) - 2 \cdot \mathrm{Rang}(A^d) + \mathrm{Rang}(A^{d+1}).$$

Damit sind die m_d eindeutig bestimmt, und insgesamt folgt die Behauptung. ◯

9.4 Jordan'sche Normalform

Definition 9.4.1 (Jordankästchen allgemein)

Für $\lambda \in K$ und natürliches d sei

$$J_d(\lambda) := \lambda \cdot I_d + J_d(0) = \begin{pmatrix} \lambda & 0 & \dots & 0 & 0 & 0 \\ 1 & \lambda & 0 & \dots & 0 & 0 \\ 0 & 1 & \lambda & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 1 & \lambda & 0 \\ 0 & 0 & \dots & 0 & 1 & \lambda \end{pmatrix}.$$

Diese Matrix heißt ein **Jordankästchen der Länge d zum Eigenwert λ.**

Bemerkung 9.4.2 (Zusammenfassung von schon Gesehenem)

Es sei Φ ein Endomorphismus eines n-dimensionalen K-Vektorraumes, dessen charakteristisches Polynom in Linearfaktoren zerfällt:

$$\mathrm{CP}_\Phi(X) = \prod_{\lambda \in \mathrm{Spec}(K)} (X-\lambda)^{\mu_a(\lambda)}.$$

Dann ist V wegen Hilfssatz 9.2.6 die direkte Summe der Haupträume $H(\Phi, \lambda)$, und wegen 9.2.4 gilt

$$H(\Phi, \lambda) = \mathrm{Kern}((\Phi - \lambda \mathrm{Id}_V)^{\mu_a(\lambda)}).$$

Auf jedem dieser Haupträume ist

$$\Psi_\lambda := (\Phi - \lambda \mathrm{Id}_V)|_{H(\Phi,\lambda)}$$

nilpotent, wird also durch eine nilpotente Abbildungsmatrix beschrieben, und diese kann nach Satz 9.3.4 in Jordan'scher Normalform gewählt werden. Damit erhalten wir eine Darstellung von

$$\Phi|_{H(\Phi,\lambda)} = \lambda \cdot \mathrm{Id}_{H(\Phi,\lambda)} + \Psi_\lambda$$

durch eine Matrix der Form

$$\lambda I_{\mu_a(\lambda)} + \widetilde{A}, \quad \widetilde{A} \text{ wie in 9.3.4.}$$

Wenn wir das für jedes λ durchgeführt haben, kommen wir am Ende zum folgenden Satz.

Satz 9.4.3 (Die Jordan'sche Normalform)

Es seien V ein endlichdimensionaler Vektorraum und Φ ein Endomorphismus von V, sodass das charakteristische Polynom von Φ in Linearfaktoren zerfällt:

$$\mathrm{CP}_\Phi(X) = \prod_{\lambda \in \mathrm{Spec}(\Phi)} (X - \lambda)^{\mu_a(\lambda)}.$$

Die Anzahl der Eigenwerte heiße l. Wir schreiben die Eigenwerte in einer festen Reihenfolge auf:

$$\mathrm{Spec}(\Phi) = \{\lambda_1, \lambda_2, \ldots, \lambda_l\}.$$

Dann gibt es für jeden Eigenwert $\lambda_i \in \mathrm{Spec}(\Phi)$ eindeutig bestimmte natürliche Zahlen

$$k_i \text{ und } d_{1,i} \geq d_{2,i} \geq \cdots \geq d_{k_i,i} \geq 1,$$

sodass sich Φ bezüglich einer geeigneten Basis B von V durch die folgende Matrix in Blockform beschreiben lässt:

$$D_{BB}(\Phi) = \begin{pmatrix} D_1 & & & \\ & D_2 & & \\ & & \ddots & \\ & & & D_l \end{pmatrix}.$$

Dabei ist für $1 \leq i \leq l$

$$D_i = \begin{pmatrix} J_{d_{1,i}}(\lambda_i) & & & \\ & J_{d_{2,i}}(\lambda_i) & & \\ & & \ddots & \\ & & & J_{d_{k_i,i}}(\lambda_i) \end{pmatrix},$$

und diese Matrix beschreibt die Abbildung, die Φ *auf dem Hauptraum von* λ_i *bewirkt. Die Anzahl der Jordankästchen der Länge* d *zum Eigenwert* λ *heiße* $m_d(\lambda)$. *Dann gilt für alle natürlichen Zahlen* d :

$$m_d(\lambda) = \mathrm{Rang}((\Phi - \lambda\mathrm{Id})^{d-1}) - 2 \cdot \mathrm{Rang}((\Phi - \lambda\mathrm{Id})^d) + \mathrm{Rang}((\Phi - \lambda\mathrm{Id})^{d+1}).$$

Beweis. Hier muss man nur die vorherigen Ergebnisse zusammentragen. ◯

Bemerkung 9.4.4 (Eindeutigkeitsfragen)

Eine Matrix, die wie $D_{BB}(\Phi)$ aus Jordankästchen aufgebaut ist, heißt eine **Jordan'sche Normalform**. Die Teilmatrizen D_i heißen die **Jordanblöcke** der Matrix.

Der Satz sagt insbesondere, dass jede Matrix mit einem zerfallenden charakteristischen Polynom (oder mit zerfallendem Minimalpolynom – das ist eine äquivalente Bedingung) zu einer Matrix in Jordan'scher Normalform ähnlich ist. Diese Jordan'sche Normalform ist bis auf die Reihenfolge der Jordanblöcke eindeutig bestimmt.

Fazit 9.4.5 (noch einmal: was wissen wir?)

Jede Matrix $A \in K^{n\times n}$, deren charakteristisches Polynom in Linearfaktoren zerfällt, ist zu genau einer (bis auf Vertauschung der Jordanblöcke) Jordan'schen Normalform ähnlich.

Die Länge des Jordanblocks zum Eigenwert λ ist die Dimension des Hauptraumes zu λ und ist gleich der algebraischen Vielfachheit $\mu_a(A, \lambda)$.

Die Anzahl der Jordankästchen zum Eigenwert λ ist die Dimension des Eigenraumes von λ, also die geometrische Vielfachheit $\mu_g(A, \lambda)$.

Die Länge des längsten Jordankästchens zum Eigenwert λ ist die Vielfachheit, mit der der Faktor $(X - \lambda)$ im Minimalpolynom von A auftritt, genauer ist sie gleich

$$\max\{e \in \mathbb{N}_0 \mid (X-\lambda)^e \text{ teilt } \mathrm{MP}(A, X)\} =$$
$$\min\{e \in \mathbb{N}_0 \mid \mathrm{Rang}((A - \lambda I_n)^e) = \mathrm{Rang}((A - \lambda I_n)^{e+1})\}.$$

Die letzte Gleichheit sieht man am besten durch Ausrechnen des Minimalpolynoms einer Matrix in Jordan'scher Normalform.

Die Anzahl der Jordankästchen zum Eigenwert λ von gegebener Länge lässt sich aus den Rängen $\mathrm{Rang}(A-\lambda I_n)^d$ berechnen.

Die Jordan'sche Normalform einer Matrix ist eine Ähnlichkeitsinvariante. Es gilt sogar: zwei Matrizen mit zerfallendem charakteristischen Polynom sind genau dann ähnlich, wenn sie dieselbe Jordan'sche Normalform haben.

Um anzudeuten, wie sich eine zugehörige Basiswechselmatrix für die Ähnlichkeit von A zu ihrer Jordanmatrix finden lässt, machen wir ein Beispiel. Ein allgemeines Verfahren soll hier nicht vorgestellt werden.

Beispiel 9.4.6 Es sei

$$A := \begin{pmatrix} 0 & -6 & 0 & 3 & 3 \\ 1 & 0 & -2 & 0 & 1 \\ 0 & -4 & 0 & 2 & 2 \\ 2 & -1 & -4 & 1 & 2 \\ 0 & -2 & 0 & 1 & 1 \end{pmatrix} \in K^{5\times 5}.$$

Wir wollen die Jordan'sche Normalform von A ermitteln. Zunächst ermitteln wir dazu das charakteristische Polynom. Es ist

$$\begin{aligned} \mathrm{CP}(A,X) &= \det \begin{pmatrix} X & 6 & 0 & -3 & -3 \\ -1 & X & 2 & 0 & -1 \\ 0 & 4 & X & -2 & -2 \\ -2 & 1 & 4 & X-1 & -2 \\ 0 & 2 & 0 & -1 & X-1 \end{pmatrix} \\ &\overset{a)}{=} \det \begin{pmatrix} 0 & X^2+6 & 2X & -3 & -X-3 \\ -1 & X & 2 & 0 & -1 \\ 0 & 4 & X & -2 & -2 \\ 0 & -2X+1 & 0 & X-1 & 0 \\ 0 & 2 & 0 & -1 & X-1 \end{pmatrix} \\ &\overset{b)}{=} \det \begin{pmatrix} X^2-2 & 0 & 1 & -X+1 \\ 4 & X & -2 & -2 \\ -2X+1 & 0 & X-1 & 0 \\ 2 & 0 & -1 & X-1 \end{pmatrix} \\ &\overset{c)}{=} X\cdot(X-1)\cdot \det \begin{pmatrix} X^2-2 & 1 & -1 \\ -2X+1 & X-1 & 0 \\ 2 & -1 & 1 \end{pmatrix} = \end{aligned}$$

$$\stackrel{d)}{=} X \cdot (X-1) \cdot \det \begin{pmatrix} X^2 & 0 & -1 \\ -2X+1 & X-1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\stackrel{e)}{=} X^3 \cdot (X-1)^2.$$

Hierbei werden, wie angezeigt, die folgenden Operationen benutzt:

a) Das X-fache der zweiten Zeile wird zur ersten addiert, ihr Doppeltes von der vierten Zeile abgezogen.

b) Die Determinante wird nach der ersten Spalte entwickelt, und in der übrigen $(4 \times 4)-$Matrix wird das Doppelte der zweiten Zeile von der ersten abgezogen.

c) Aus der zweiten und vierten Spalte wird X bzw. $(X-1)$ ausgeklammert und anschließend die Determinante nach der zweiten Spalte entwickelt.

d) Die letzte Spalte wird zu der zweiten addiert, ihr Doppeltes von der ersten abgezogen.

e) Dann wird nach der letzten Zeile entwickelt.

Damit sind die Eigenwerte von A die Zahlen 0 und 1, die algebraischen Vielfachheiten sind 3 bzw. 2.

Der Rang von A ist 3, also ist $\mu_g(A,0) = 2 < 3 = \mu_a(A,0)$.

Der Rang von $A - I_5$ ist 4 (nachrechnen!), also ist $\mu_g(A,1) = 1 < 2 = \mu_a(A,1)$.

Wir müssen also in der Jordan'schen Normalform wirklich Jordankästchen von Länge größer als 1 bekommen, A ist nicht diagonalisierbar.

Die Anzahl der Jordankästchen zum Eigenwert 0 ist 2, und da der Hauptraum dreidimensional ist, gibt es ein Kästchen der Länge 2 und eines der Länge 1. Genauso gibt es zum Eigenwert 1 nur 1 Kästchen, das dann Länge zwei haben muss, da es den ganzen Hauptraum erschlägt.

Die Jordan'sche Normalform von A sieht also so aus:

$$\widetilde{A} = \begin{pmatrix} J_2(0) & & \\ & J_1(0) & \\ & & J_2(1) \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

Wie finde ich nun eine Basiswechselmatrix $S \in \mathrm{GL}_5(K)$, sodass

$$\widetilde{A} = S^{-1} \cdot A \cdot S??$$

Diese Frage gehen wir an, indem wir noch einmal nachvollziehen, was die Jordan'sche Normalform bedeutet, und was solch eine Jordanbasis beziehungsweise

die mit ihr assoziierte Basiswechselmatrix auszeichnet. Dabei fällt auf, dass zur Wahl solch einer Basis meistens eine riesige Auswahl zur Verfügung steht. Nur das Folgende ist vorgegeben.

Wir brauchen eine Basis $\{b_1, \ldots, b_5\}$ von K^5, bezüglich derer $\widetilde{A}$ die Abbildungsmatrix von Φ_A ist, das heißt nach Definition der Abbildungsmatrix:

$$A \cdot b_1 = b_2,\ A \cdot b_2 = 0,\ A \cdot b_3 = 0,\ A \cdot b_4 = b_4 + b_5,\ A \cdot b_5 = b_5.$$

Insbesondere ist $b_1 \in \mathrm{Kern}(A^2) \setminus \mathrm{Kern}(A)$. Was sind diese Kerne? Gauß sagt zum Beispiel:

$$\mathrm{Kern}(A) = \langle \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} \rangle, \quad \mathrm{Kern}(A^2) = K \cdot \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 3 \end{pmatrix} \oplus \mathrm{Kern}(A).$$

Nun nehmen wir einen willkürlich gewählten Vektor aus $\mathrm{Kern}(A^2) \setminus \mathrm{Kern}(A)$ und nennen ihn b_1:

$$b_1 := \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 3 \end{pmatrix}, \quad b_2 := A \cdot b_1 = \begin{pmatrix} 3 \\ 1 \\ 2 \\ 1 \\ 1 \end{pmatrix}.$$

Der nächste Basisvektor b_3 muss b_2 zu einer Basis von $\mathrm{Kern}(A)$ ergänzen, also können wir etwa

$$b_3 := \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

wählen.

Jetzt müssen wir noch Basisvektoren finden, die das Jordankästchen zum Eigenwert 1 liefern, also Vektoren

$$b_4 \in \mathrm{Kern}(A - I_5)^2 \setminus \mathrm{Kern}(A - I_5),\ b_5 := (A - I_5) \cdot b_4.$$

Der Kern von $A - I_5$ berechnet sich nach Gauß zu

$$\mathrm{Kern}(A - I_5) = K \cdot \begin{pmatrix} 3 \\ 0 \\ 2 \\ 0 \\ 1 \end{pmatrix},$$

Der Kern von $(A - I_5)^2$ wird von diesem Vektor und dem vierten Standardbasisvektor e_4 erzeugt. Also wählen wir zum Beispiel

$$b_4 := e_4, \quad b_5 := (A - I_5)b_4 = \begin{pmatrix} 3 \\ 0 \\ 2 \\ 0 \\ 1 \end{pmatrix}.$$

Damit haben wir insgesamt eine Basis $B := \{b_1, \ldots, b_5\}$ bestimmt, die einen möglichen Basiswechsel von A zur Jordan Normalform liefert. Um dies zu testen, bilden wir die Matrix S aus den Spalten $b_1, \ldots, b_5$:

$$S := \begin{pmatrix} -2 & 3 & 2 & 0 & 3 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 2 \\ 0 & 1 & 0 & 1 & 0 \\ 3 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

S hat Determinante 1, also ist die inverse Matrix auch ganzzahlig. In der Tat:

$$S^{-1} = \begin{pmatrix} 1 & 0 & -2 & 0 & 1 \\ -1 & 1 & 2 & 0 & -1 \\ 6 & 0 & -11 & 0 & 4 \\ 1 & -1 & -2 & 1 & 1 \\ -2 & -1 & 4 & 0 & -1 \end{pmatrix}.$$

Nun rechnet man leicht nach, dass

$$S^{-1} \cdot A \cdot S = \widetilde{A}.$$

9.5 Vermischtes

Beispiel 9.5.1 (Jordan'sche Normalformen in kleiner Dimension)

a) Eine 1×1-Matrix ist immer schon in Jordan'scher Normalform. Ebenso auch die 0×0-Matrix.

b) Das charakteristische Polynom f der 2×2-Matrix A zerfalle als

$$f = (X - a) \cdot (X - b).$$

Dann gibt es die drei folgenden Möglichkeiten:

- $a \neq b$: Dann gibt es zwei eindimensionale Eigenräume, und A ist diagonalisierbar. Die Normalform von A ist $\left(\begin{smallmatrix} a & 0 \\ 0 & b \end{smallmatrix}\right)$.

- $a = b$ und $\dim(\mathrm{Eig}(A, a)) = 2$: hier ist A bereits in Normalform, es muss nämlich $A = a \cdot I_2$ gelten. (Wieso?)
- $a = b$ und $\dim(\mathrm{Eig}(A, a)) = 1$: hier gibt es nur ein Jordankästchen zum Eigenwert a, also ist die Normalform von A die Matrix $\left(\begin{smallmatrix} a & 0 \\ 1 & a \end{smallmatrix}\right)$.

c) $A \in K^{3\times 3}$ habe charakteristisches Polynom $(X-a)\cdot(X-b)\cdot(X-c)$. Dann haben wir die folgenden Möglichkeiten (bis auf Permutation der Nullstellen):

- a, b, c paarweise verschieden: die Normalform ist diagonal.
- $a = b \neq c$: je nach Dimension von $\mathrm{Eig}\,(A, a)$ gibt es eine der beiden Normalformen
$$\begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & c \end{pmatrix}, \quad \begin{pmatrix} a & 0 & 0 \\ 1 & a & 0 \\ 0 & 0 & c \end{pmatrix}.$$
- $a = b = c$: Je nachdem, ob $\mathrm{Eig}\,(A, a)$ Dimension 3, 2, oder 1 hat, ist die Normalform eine der Matrizen
$$\begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{pmatrix}, \quad \begin{pmatrix} a & 0 & 0 \\ 1 & a & 0 \\ 0 & 0 & a \end{pmatrix}, \quad \begin{pmatrix} a & 0 & 0 \\ 1 & a & 0 \\ 0 & 1 & a \end{pmatrix}.$$

d) $A \in K^{4\times 4}$ habe charakteristisches Polynom $(X-a)\cdot(X-b)\cdot(X-c)\cdot(X-d)$. Dann haben wir die folgenden Möglichkeiten (bis auf Permutation der Nullstellen):

- a, b, c, d paarweise verschieden: die Normalform ist diagonal.
- $a = b \neq c \neq d \neq a$: je nach Dimension von $\mathrm{Eig}\,(A, a)$ gibt es eine der beiden Normalformen
$$\begin{pmatrix} a & 0 & 0 & 0 \\ 0 & a & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & d \end{pmatrix}, \quad \begin{pmatrix} a & 0 & 0 & 0 \\ 1 & a & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & d \end{pmatrix}.$$
- $a = b = c \neq d$: Je nachdem, ob $\mathrm{Eig}\,(A, a)$ Dimension 3, 2, oder 1 hat, ist die Normalform eine der Matrizen
$$\begin{pmatrix} a & 0 & 0 & 0 \\ 0 & a & 0 & 0 \\ 0 & 0 & a & 0 \\ 0 & 0 & 0 & d \end{pmatrix}, \quad \begin{pmatrix} a & 0 & 0 & 0 \\ 1 & a & 0 & 0 \\ 0 & 0 & a & 0 \\ 0 & 0 & 0 & d \end{pmatrix}, \quad \begin{pmatrix} a & 0 & 0 & 0 \\ 1 & a & 0 & 0 \\ 0 & 1 & a & 0 \\ 0 & 0 & 0 & d \end{pmatrix}.$$

- $a = b \neq c = d$: Wieder entscheiden die Dimensionen der Eigenräume $\mathrm{Eig}(A, a)$ und $\mathrm{Eig}(A, c)$ zwischen den drei möglichen Jordan'schen Normalformen (a und c können ja vertauscht werden):

$$\begin{pmatrix} a & 0 & 0 & 0 \\ 0 & a & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & c \end{pmatrix}, \quad \begin{pmatrix} a & 0 & 0 & 0 \\ 1 & a & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & c \end{pmatrix}, \quad \begin{pmatrix} a & 0 & 0 & 0 \\ 1 & a & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 1 & c \end{pmatrix}.$$

- $a = b = c = d$: Wenn $\mathrm{Eig}\,(A, a)$ Dimension 4, 3, oder 1 hat, ist die Normalform eine der Matrizen

$$\begin{pmatrix} a & 0 & 0 & 0 \\ 0 & a & 0 & 0 \\ 0 & 0 & a & 0 \\ 0 & 0 & 0 & a \end{pmatrix}, \quad \begin{pmatrix} a & 0 & 0 & 0 \\ 1 & a & 0 & 0 \\ 0 & 0 & a & 0 \\ 0 & 0 & 0 & a \end{pmatrix}, \quad \begin{pmatrix} a & 0 & 0 & 0 \\ 1 & a & 0 & 0 \\ 0 & 1 & a & 0 \\ 0 & 0 & 1 & a \end{pmatrix}.$$

  Wenn $\mathrm{Eig}\,(A, a)$ Dimension 2 hat, dann gibt es noch die zwei Möglichkeiten

$$\begin{pmatrix} a & 0 & 0 & 0 \\ 1 & a & 0 & 0 \\ 0 & 1 & a & 0 \\ 0 & 0 & 0 & a \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} a & 0 & 0 & 0 \\ 1 & a & 0 & 0 \\ 0 & 0 & a & 0 \\ 0 & 0 & 1 & a \end{pmatrix}.$$

  Sie sind durch Rang $(A - aI_4)^2$ zu unterscheiden; dieser ist im linken Fall 1, im rechten 0.

Bemerkung 9.5.2 (Funktionen auf dem Matrizenring)

Es sei $F : K^{n\times n} \longrightarrow K^{n\times n}$ eine Abbildung mit der Eigenschaft

$$\forall S \in \mathrm{GL}_n(K), A \in K^{n\times n} : F(S^{-1}AS) = S^{-1}F(A)S.$$

Dies ist zum Beispiel für jede durch ein Polynom gegebene Abbildung der Fall, oder – im Fall $K = \mathbb{R}$ – für die Abbildung, die durch die Exponentialreihe gegeben ist. Wenn man nun nach $F(A)$ für ein festes A fragt, ist es manchmal tatsächlich einfacher, für die Jordan'sche Normalform $\widetilde{A}$ von A die Funktion F auszuwerten und mithilfe einer Basiswechselmatrix daraus das $F(A)$ zu ermitteln. Analoges gilt für die Frage, ob es eine Matrix X gibt mit $F(X) = A$. Wenn dies für die Jordan'sche Normalform von A gilt, dann auch für A.

Bemerkung 9.5.3 (Eine Zerlegung von A)

Das charakteristische Polynom von A zerfalle in Linearfaktoren, und es sei

$$\widetilde{A} = S^{-1}AS$$

die Jordan'sche Normalform, mit einer Basiswechselmatrix S. Dann ist $\widetilde{A}$ die Summe einer Diagonalmatrix $\widetilde{D}$ mit einer nilpotenten unteren Dreiecksmatrix $\widetilde{N}$, und analog ist

$$A = D + N, \quad D = S\widetilde{D}S^{-1}, \; N = S\widetilde{N}S^{-1}.$$

Dabei ist D diagonalisierbar und N nilpotent. Diese Zerlegung von A heißt die **Jordan-Zerlegung** von A. Sie ist eindeutig dadurch charakterisiert, dass gilt:

$$DA = AD \;\text{ und }\; NA = AN.$$

Wenn nämlich eine weitere Zerlegung von A in eine diagonalisierbare und eine nilpotente Matrix mit dieser Eigenschaft gegeben ist:

$$A = \hat{D} + \hat{N}, \quad A\hat{D} = \hat{D}A \;\text{ und }\; \hat{N}A = A\hat{N},$$

dann muss $\hat{D}$ die Haupträume von A invariant lassen, denn es vertauscht mit $(A - \lambda I)^d$ für jedes $\lambda \in K$ und jedes $d \in \mathbb{N}$.

Nun ist aber auch die Einschränkung von $\hat{D}$ auf $H(A, \lambda)$ diagonalisierbar. Es sei α ein Eigenwert und $v \in H(A, \lambda)$ ein Eigenvektor von $\hat{D}$ zum Eigenwert α. Dann gilt (weil $A\hat{D} = \hat{D}A$) für jedes positive e:

$$(A - \alpha I)^e v = (A - \hat{D})^e v = \hat{N}^e v,$$

und das wird 0 für großes e. Damit ist

$$v \in H(A, \lambda) \cap H(A, \alpha),$$

also $\alpha = \lambda$, weil verschiedene Haupträume nur 0 als Schnitt haben. Also hat $\hat{D}$ nur den Eigenwert λ auf $H(A, \lambda)$, stimmt also auf $H(A, \lambda)$ mit D überein. Da aber V die Summe der Haupträume ist, ist damit $\hat{D}$ überall gleich D. Das war zu zeigen.

Kapitel 10

Bilineare Abbildungen

Hier werden wir die algebraische Grundlage für die Theorie der Skalarprodukte kennenlernen. Vieles von dem, was wir jetzt algebraisch und allgemein machen, werden wir in Kapitel 11 noch einmal für den Fall von Skalarprodukten erläutern. Von daher ist dieses Kapitel eher optionaler Natur.

10.1 Bilinearformen

Wir wollen in diesem Abschnitt den wichtigen Spezialfall einer Bilinearform studieren. Das wird im Abschnitt 10.2 verallgemeinert zu multilinearen Abbildungen. Für die Kapitel 11 und 12 wird allerdings nur Abschnitt 10.1 wirklich relevant sein.

Definition 10.1.1 (Paarung, Bilinearform)

Es seien K ein Körper und V, W zwei K-Vektorräume. Eine **Paarung** P zwischen V und W ist eine Abbildung

$$P : V \times W \longrightarrow K,$$

bei der für alle $a, b \in K,\ v_1, v_2 \in V,\ w_1, w_2 \in W$ die folgenden Regeln gelten:

$$\begin{array}{lll} P(av_1 + v_2, & w_1 &) = aP(v_1, w_1) + P(v_2, w_1), \\ P(v_1, & bw_1 + w_2 &) = bP(v_1, w_1) + P(v_1, w_2). \end{array}$$

Diese Eigenschaft nennt man die **Bilinearität** der Abbildung P. Sie bedeutet, dass für festes v die Abbildung $W \ni w \mapsto P(v, w) \in K$ eine Linearform auf W und für festes w die Abbildung $V \ni v \mapsto P(v, w) \in K$ eine Linearform auf V ist (siehe 6.2.1).

Im Falle $V = W$ spricht man von einer **Bilinearform** auf V.

Die Paarung P heißt **nicht ausgeartet**, wenn für alle $v \in V, v \neq 0$, ein $w \in W$ existiert mit $P(v,w) \neq 0$, und wenn für alle $w \in W, w \neq 0$ ein $v \in V$ existiert mit $P(v,w) \neq 0$.

Die Menge aller Paarungen zwischen V und W ist ein Untervektorraum des K-Vektorraumes $\mathrm{Abb}(V \times W, K)$, siehe 5.1.3c).

Beispiel 10.1.2 (Dualraum)

Es sei V ein K-Vektorraum und $W = V^*$ sein Dualraum. Dann ist die Abbildung

$$P : V \times V^* \longrightarrow K, \quad P(v,\lambda) := \lambda(v),$$

eine nicht ausgeartete Paarung auf $V \times V^*$.

In gewisser Weise ist das das wichtigste Beispiel für eine Paarung. Genauer gilt:

Hilfssatz 10.1.3 (Paarungen und der Dualraum)

Es seien K ein Körper und V, W zwei K-Vektorräume.

a) Für eine Paarung $P : V \times W \longrightarrow K$ und $w \in W$ definieren wir $\rho_w : V \longrightarrow K$ durch

$$\forall v \in V : \ \rho_w(v) := P(v,w).$$

Dies ist (nach Definition 10.1.1) ein Element des Dualraums V^*. Die Abbildung

$$\rho : W \longrightarrow V^*, \quad w \mapsto \rho_w,$$

ist ein Homomorphismus von K-Vektorräumen.

b) Es gibt einen Isomorphismus zwischen dem Vektorraum aller Paarungen zwischen V und W und dem Vektorraum $\mathrm{Hom}(W, V^*)$.

Beweis. a) Die Homomorphie-Eigenschaft von ρ folgt durch einfaches Nachrechnen. Zum Beispiel gilt für $w_1, w_2 \in W$:

$$\forall v \in V : \rho_{w_1+w_2}(v) = P(v, w_1 + w_2) = P(v,w_1) + P(v,w_2) = \rho_{w_1}(v) + \rho_{w_2}(v).$$

b) Die Abbildung η, die einer Paarung P den Homomorphismus $\rho : W \longrightarrow V^*$, zuordnet ist selbst ein Homomorphismus von Vektorräumen (nachrechnen!).

Die Umkehrabbildung zu η ist gegeben durch

$$\mathrm{Hom}(W, V^*) \ni \rho \mapsto P, \quad P(v,w) := (\rho(w))(v).$$

Also ist η ein Isomorphismus zwischen den Vektorräumen. ◯

Definition 10.1.4 (bilineare Fortsetzung, Fundamentalmatrix)

Nun seien V und W endlichdimensional und Basen $B = \{b_1, \ldots, b_m\}$ von V sowie $C = \{c_1, \ldots, c_n\}$ von W gewählt. Dann wird die Paarung P auf $V \times W$ gegeben durch die Einschränkung auf $B \times C$, also durch die Abbildung $P|_{B\times C} : B \times C \longrightarrow K$. Es gilt ja für $v = \sum_{i=1}^{m} k_i b_i,\ w = \sum_{j=1}^{n} l_j c_j$ (mit $k_i, l_j \in K$):

$$P(v, w) = \sum_{i,j} k_i l_j \cdot P(b_i, c_j).$$

Umgekehrt wird durch jede Vorgabe einer K-wertigen Abbildung auf $B \times C$ eine Paarung auf $V \times W$ definiert; man nennt dies die **bilineare Fortsetzung** – ein offensichtliches Pendant zur Linearen Fortsetzung (siehe 6.1.2).

Ist umgekehrt eine Paarung P gegeben, so schreiben wir die Werte $f_{ij} := P(b_i, c_j)$ in eine $m \times n$-Matrix F. Diese heißt die **Fundamentalmatrix** von P bezüglich der Basen B und C:

$$F =: D_{BC}(P).$$

Wenn wir Vektoren $v \in V$ und $w \in W$ wieder schreiben als

$$v = \sum_{i=1}^{m} k_i b_i, \quad w = \sum_{j=1}^{n} l_j c_j,$$

dann wird aus der obigen Formel unter Verwendung der Koordinatenvektoren $D_B(v) = (k_1\ k_2\ \ldots\ k_m)^\top$ und $D_C(w) = (l_1\ l_2\ \ldots\ l_n)^\top$ die bemerkenswerte Formel

$$\boxed{P(v, w) = D_B(v)^\top \cdot D_{BC}(P) \cdot D_C(w).}$$

Jetzt hat man natürlich wieder alle Matrizen an der Hand, um Beispiele für bilineare Abbildungen zu basteln.

Hilfssatz 10.1.5 (nicht ausgeartete Paarung)

Es seien K ein Körper und V, W endlichdimensionale K-Vektorräume mit Basen B, C. Weiter sei $P : V \times W \longrightarrow K$ eine Paarung auf $V \times W$. Dann ist P genau dann nicht ausgeartet, wenn V und W dieselbe Dimension haben und die Fundamentalmatrix $D_{BC}(P)$ regulär ist.

Beweis. Es seien zunächst $\dim(V) = \dim(W) = n$ und die Fundamentalmatrix F regulär. Sei $w \in W$, $w \neq 0$. Dann ist $D_C(w) \neq 0$, und weil F invertierbar ist, ist auch $F \cdot D_C(w) \neq 0$. Wir wählen ein $i \in \{1, \ldots, n\}$, sodass der i-te Eintrag von $FD_C(w)$ nicht Null ist. Dann ist

$$P(b_i, w) = e_i^\top \cdot F \cdot D_C(w) \neq 0.$$

Ein analoges Argument zeigt, dass es auch zu $v \neq 0$ ein $w \in W$ gibt mit $P(v,w) \neq 0$. Damit ist P nicht ausgeartet.

Nun wollen wir annehmen, dass P nicht ausgeartet ist. Dann ist die Abbildung

$$\rho : W \longrightarrow V^*, w \mapsto \rho_w, \quad \rho_w(v) := P(v,w)$$

injektiv, also $\dim(V) = \dim(V^*) \geq \dim(W)$. Vertauscht man hierbei die Rollen von V und W, so folgt auch $\dim(W) \geq \dim(V)$, also Gleichheit der Dimensionen. Wäre nun die Fundamentalmatrix nicht regulär, so gäbe es in W ein $w \neq 0$ mit $F \cdot D_C(w) = 0$, und für dieses gälte dann für alle $v \in V$:

$$P(v,w) = D_B(v)^\top \cdot F \cdot D_C(w) = 0.$$

Das widerspricht der Voraussetzung, dass P nicht ausgeartet ist. ◯

Hilfssatz 10.1.6 (Basiswechsel für Paarungen)

Es seien K ein Körper und V, W endlichdimensionale K-Vektorräume mit einer Paarung $P : V \times W \longrightarrow K$. Weiter seien Basen $B, \hat{B}$ von V und $C, \hat{C}$ von W gegeben. Dann gilt für die zugehörigen Fundamentalmatrizen:

$$D_{B,C}(P) = D_{\hat{B},B}(\mathrm{Id}_V)^\top \cdot D_{\hat{B},\hat{C}}(P) \cdot D_{\hat{C},C}(\mathrm{Id}_W).$$

Beweis. Für beliebige Vektoren $v \in V$ und $w \in W$ gilt

$$\begin{aligned} P(v,w) &= D_{\hat{B}}(v)^\top \cdot D_{\hat{B},\hat{C}}(P) \cdot D_{\hat{C}}(w) \\ &= [D_{\hat{B},B}(\mathrm{Id}_V) D_B(v)]^\top \cdot D_{\hat{B},\hat{C}}(P) \cdot [D_{\hat{C},C}(\mathrm{Id}_W) D_C(w)] \\ &= D_B(v)^\top \cdot [D_{\hat{B},B}(\mathrm{Id}_V)^\top \cdot D_{\hat{B},\hat{C}}(P) \cdot D_{\hat{C},C}(\mathrm{Id}_W)] \cdot D_C(w), \end{aligned}$$

wobei wir die Merkregel aus 6.3.1 für die Identität auf V und auf W benutzen. Andererseits ist nach Definition der Fundamentalmatrix

$$P(v,w) = D_B(v)^\top \cdot D_{B,C}(P) \cdot D_C(w).$$

Durchlaufen hierbei v und w die Basisvektoren aus B bzw. C, zeigt ein Koeffizientenvergleich, dass die Matrizen $D_{B,C}(P)$ und $D_{\hat{B},B}(\mathrm{Id}_V)^\top \cdot D_{\hat{B},\hat{C}}(P) \cdot D_{\hat{C},C}(\mathrm{Id}_W)$ übereinstimmen. ◯

Bemerkung 10.1.7 Wir finden also für die Fundamentalmatrizen von Paarungen ein anderes Verhalten bei Basiswechsel als für Abbildungsmatrizen von Homomorphismen. Andererseits können wir eine Paarung P nach 10.1.3 mit einem Homomorphismus ρ von W nach V^* identifizieren. Es ist für $v \in V, w \in W$

$$(\rho(w))(v) = P(v,w).$$

Für die Basisvektoren folgt dann aus 6.2.2 die schöne Identität

$$\rho(c_i) = \sum_{j=1}^{n} P(b_j, c_i) b_j^*.$$

Dabei ist $\{b_j^* \mid 1 \leq j \leq n\}$ die zu B duale Basis von V^*. Dann sagt uns aber die Definition der Abbildungmatrix 6.3.1:

$$D_{B^*,C}(\rho) = D_{BC}(P).$$

Beispiel 10.1.8 (Dualraum)

Es sei V endlichdimensional, $W = V^*$ der Dualraum von V und P die Paarung $(v, w) \mapsto w(v)$. Dann ist ρ offensichtlich die Identität auf W. Für eine Basis B von V und ihre Dualbasis $C = B^*$ gilt dann

$$\forall b \in B, c \in C : P(b, c) = \begin{cases} 1, & \text{falls } c = b^*, \\ 0, & \text{sonst.} \end{cases}$$

Das zeigt, dass $D_{B,B^*}(P)$ die Einheitsmatrix ist. Das ist ein besonders angenehmer Fall.

Wenn P eine nicht ausgeartete Paarung zwischen zwei endlichdimensionalen Vektorräumen ist, dann gibt es immer Basen B und C dieser Räume, sodass $D_{B,C}(P)$ die Einheitsmatrix ist. Um das einzusehen wähle man irgendwelche Basen und betrachte die Fundamentalmatrix F. Dann macht man auf einem der Vektorräume einen Basiswechsel, der durch F^{-1} beschrieben wird. Hilfssatz 10.1.6 besorgt den Rest.

Wie die Endomorphismen eine spezielle Rolle bei den Homomorphismen von Vektorräumen spielen, so spielen die Bilinearformen eine spezielle Rolle im Bereich der Paarungen. Auch hier wird man sich bei Fundamentalmatrizen für die Matrizen $D_{BB}(P)$ interessieren und nicht zwei verschiedene Basen von V benutzen wollen. Diesen Fall müssen wir weiterverfolgen.

Definition 10.1.9 (Orthonormalbasis, Symmetrie)

Es sei $P : V \times V \longrightarrow K$ eine Bilinearform auf einem n-dimensionalen K-Vektorraum V.

a) P heißt **symmetrisch**, wenn für alle $v, w \in V$ gilt:

$$P(v, w) = P(w, v).$$

Das bedeutet, dass für eine beliebige Basis B von V die Fundamentalmatrix $D_{BB}(P)$ symmetrisch (also gleich ihrer Transponierten, siehe 4.1.13) ist.

b) Eine Basis $B := \{b_1, \ldots, b_n\}$ von V heißt eine **Orthogonalbasis** (OGB) von V bezüglich P, wenn gilt:

$$\forall 1 \leq i \neq j \leq n : P(b_i, b_j) = 0.$$

c) Die Basis B heißt eine **Orthonormalbasis** (ONB) von V bezüglich P, wenn sie eine Orthogonalbasis ist und zusätzlich die Bedingung

$$\forall 1 \leq i \leq n : P(b_i, b_i) = 1$$

erfüllt ist.

Wenn es eine orthogonale Basis gibt, so ist P sicher symmetrisch: bezüglich dieser Basis ist ja die Fundamentalmatrix diagonal.

Hilfssatz 10.1.10 (Existenz einer OGB)

Es sei $P : V \times V \longrightarrow K$ eine symmetrische Bilinearform auf dem n-dimensionalen K-Vektorraum V. Der Körper K habe Charakteristik ungleich 2 (d.h. $2 := 1 + 1 \neq 0$). Dann gibt es eine (bezüglich P) orthogonale Basis von V.

Beweis. Wir führen den Beweis durch vollständige Induktion nach der Dimension von V. Für $n = 0$ oder 1 ist nichts zu zeigen. Nun sei $n > 1$ und die Behauptung wahr für Vektorräume der Dimension $n - 1$.

Wenn P identisch gleich 0 ist, dann ist jede Basis eine Orthogonalbasis.

Wir schließen diesen Fall also ohne Weiteres aus. Dann gibt es Vektoren $v, w \in V$ mit $P(v, w) \neq 0$. Wegen

$$P(v + w, v + w) = P(v, v) + P(w, w) + 2 \cdot P(v, w),$$

können nicht alle drei Vektoren $P(v + w, v + w), P(v, v), P(w, w)$ Null sein. (An dieser Stelle haben wir die Symmetrie von P benutzt und auch $2 \neq 0$.) Also gibt es einen Vektor $b_1 \in V$ mit $P(b_1, b_1) \neq 0$.

Nun sei

$$W := \{v \in V \mid P(v, b_1) = 0\}.$$

Das ist der Kern der Linearform $\rho(b_1)$, und da b_1 nicht darin liegt, ist es ein $(n - 1)$–dimensionaler Untervektorraum von V (Dimensionsformel 5.5.11 b)), für den sogar

$$V = \langle b_1 \rangle \oplus W$$

gilt. Die Einschränkung von P nach $W \times W$ ist immer noch symmetrisch, also gibt es eine Orthogonalbasis $\{b_2, \ldots, b_n\}$ von W bezüglich dieser Einschränkung. Dann ist aber insgesamt die Basis $\{b_1, \ldots, b_n\}$ eine Orthogonalbasis von V. ◯

Bemerkung 10.1.11 (Orthonormalbasen, Fourierformel)

Eine Orthonormalbasis muss es auch unter den Bedingungen von Hilfssatz 10.1.10 nicht unbedingt geben, wie uns zum Beispiel die Bilinearform

$$P : \mathbb{Q} \times \mathbb{Q} \longrightarrow \mathbb{Q},\ P(x, y) = 2xy,$$

lehrt: 2 ist in $\mathbb{Q}$ kein Quadrat.

Wenn P nicht ausgeartet ist und B eine Orthogonalbasis von V bezüglich P, dann kann man mithilfe von P die Koordinaten von Vektoren $v \in V$ bezüglich B ausrechnen. Es gilt ja für $v = \sum_{b\in B} \alpha(b) \cdot b \in V$:

$$\forall c \in B : P(v, c) = \sum_{b\in B} \alpha(b) \cdot P(b, c) = \alpha(c) \cdot P(c, c),$$

denn alle anderen Summanden sind 0. Das führt zu

$$\forall c \in B :\ \alpha(c) = P(v, c) \cdot P(c, c)^{-1}.$$

Sie werden zugeben, dass dies am schönsten ist, wenn B eine Orthonormalbasis ist. Hier erhalten wir die **Fourierformel**

$$\boxed{B \text{ Orthonormalbasis},\ v \in V \Rightarrow v = \textstyle\sum_{b\in B} P(v, b) \cdot b.}$$

Wenn zum Beispiel $B = \{b_1, \ldots, b_n\}$ eine Orthonormalbasis von V bezüglich einer Paarung P ist und $\Phi \in \mathrm{End}(V)$ ein Endomorphismus, dann lässt sich die Abbildungsmatrix von Φ bezüglich B jetzt so hinschreiben:

$$D_{BB}(\Phi) = (P(b_i, \Phi(b_j)))_{1\leq i,j\leq n}.$$

Denn der Eintrag in der i-ten Zeile und j-ten Spalte dieser Matrix ist ja der Koeffizient bei b_i von $\Phi(b_j)$ (siehe 6.3.1).

10.2 Multilineare Abbildungen

Jetzt verallgemeinern wir den Begriff der Paarung in zweifacher Hinsicht. Erstens lassen wir mehr als zwei Argumente zu, und zweitens betrachten wir Abbildungen mit Werten in einem beliebigen Vektorraum. Das passiert in der folgenden Präzisierung des Begriffs der Multilinearität, den wir im Zusammenhang mit Determinanten schon kennen gelernt haben (siehe Folgerung 7.1.4 a)).

Definition 10.2.1 (Multilinearität)

Es seien K ein Körper und $V_1, \ldots, V_n$ sowie W Vektorräume über dem Körper K. Eine Abbildung

$$M : V_1 \times V_2 \times \cdots \times V_n \longrightarrow W$$

heißt eine n**-fach multilineare Abbildung**, wenn für jedes $i \in \{1, \ldots, n\}$ und jede Wahl von Vektoren $v_j \in V_j$ (mit $1 \leq j \leq n,\ j \neq i$) die Abbildung

$$V_i \ni v \mapsto M(v_1, \ldots, v_{i-1}, v, v_{i+1}, \ldots v_n) \in W$$

eine lineare Abbildung von V_i nach W ist.

Für $n = 2$ sagt man auch **bilinear** statt 2-fach multilinear.

Beispiel 10.2.2 a) Die Determinantenabbildung haben wir als eine multilineare Abbildung $D : K^n \times K^n \times \cdots \times K^n \longrightarrow K$ eingeführt.

b) Die skalare Multiplikation $K \times V \longrightarrow V$, die die Vektorraumstruktur auf V festlegt, ist eine bilineare Abbildung.

c) Für zwei K-Vektorräume ist die Abbildung von $\mathrm{Hom}(V, W) \times V$ nach W, die (Φ, v) auf $\Phi(v)$ abbildet, bilinear.

d) Für natürliche Zahlen p, q, r, s ist (z.B.) die Abbildung

$$K^{p\times q} \times K^{q\times r} \times K^{r\times s} \longrightarrow K^{p\times s},\ \ (A, B, C) \mapsto A \cdot B \cdot C$$

eine dreifach multilineare Abbildung.

e) Die Multiplikation im Polynomring, $K[X] \times K[X] \longrightarrow K[X]$, ist eine bilineare Abbildung.

Bemerkung 10.2.3 (multilineare Fortsetzung)

In der Situation von Definition 10.2.1 seien $B_1, \ldots, B_n$ Basen der Vektorräume $V_1, \ldots, V_n$. Dann wird die Multilinearform M durch die Werte

$$M(b_1, \ldots, b_n),\ \ b_i \in B_i,$$

festgelegt. Denn für jedes $v_i \in V_i$ gibt es Funktionen $\alpha_i \in \mathrm{Abb}(B_i, K)_0$ mit $v_i = \sum_{b_i \in B_i} \alpha_i(b_i) b_i$, und dann haben wir

$$\begin{aligned} M(v_1, \ldots, v_n) &= M(\textstyle\sum_{b_1\in B_1} \alpha_1(b_1)b_1, \ldots, \sum_{b_n\in B_n} \alpha_n(b_n)b_n) \\ &= \textstyle\sum_{b_1\in B_1} \alpha_1(b_1) M(b_1, \sum_{b_2\in B_2} \alpha_2(b_2)b_2, \ldots, \sum_{b_n\in B_n} \alpha_n(b_n)b_n) \\ &= \ldots = \\ &= \textstyle\sum_{(b_1,\ldots,b_n)\in B_1\times\cdots\times B_n} \big(\prod_{i=1}^n \alpha_i(b_i)\big) M(b_1, \ldots, b_n). \end{aligned}$$

Das ist die Verallgemeinerung der definierenden Formel im Fall $n = 2$ aus Definition 10.1.1.

Umgekehrt kann man die Werte von M auf $B_1 \times \cdots \times B_n$ beliebig vorgeben und erhält dadurch eine n-fach multilineare Abbildung auf $V_1 \times \cdots \times V_n$. Das zeigt, dass die Menge aller n-fach multilinearen Abbildungen von $V_1 \times \cdots \times V_n$ nach W ein Vektorraum ist, der isomorph ist zum Vektorraum $\mathrm{Abb}(B_1 \times \cdots \times B_n, W)$. Insbesondere gilt im Falle endlichdimensionaler Vektorräume, dass die Dimension dieses Vektorraumes gleich dem Produkt $\dim(V_1) \cdot \ldots \cdot \dim(V_n) \cdot \dim(W)$ ist.

Beispiel 10.2.4 (Determinante)

Es sei $V_1 = V_2 = \cdots = V_n = K^n$ und $W = K$. Statt „multilineare Abbildung nach K“ sagen wir dann meistens „Multilinearform auf K^n“. Wie sieht so eine Multilinearform aus?

Es sei $F : \{1, \ldots, n\} \longrightarrow \{1, \ldots, n\}$ eine Abbildung. Wir schreiben für $i = 1, \ldots, n$ den Vektor $v_i \in K^n$ als

$$v_i = (v_{1i}\ v_{2i}\ \ldots v_{ni})^\top.$$

Dann wird durch

$$M_F(v_1, \ldots, v_n) := v_{F(1),1} \cdot v_{F(2),2} \cdot \cdots \cdot v_{F(n),n}$$

eine n-fache Multilinearform auf K^n definiert. Die Menge der Abbildungen von $\{1, \ldots, n\}$ in sich selbst hat n^n Elemente, und man sieht leicht, dass die Abbildungen M_F linear unabhängig sind.

Also sagt die letzte Bemerkung, dass die M_F eine Basis des Vektorraumes der n-fachen Multilinearformen auf K^n bilden. Jede n-fache Multilinearform lässt sich also auf eindeutige Art als Linearkombination derselben schreiben. Zum Beispiel ist die Determinante wegen der Leibniz-Formel (7.2.1) gleich der Summe

$$\det = \sum_F s(F) M_F,$$

wobei für bijektive Abbildungen F (also für Permutationen) der Koeffizient $s(F) = \operatorname{sign}(F)$ gleich dem Signum gesetzt wird, und sonst $s(F) = 0$.

10.3 Tensorprodukte

Nun wollen wir gerne erreichen, dass wir das Studium der Gesamtheit aller bilinearen Abbildungen $V_1 \times V_2 \longrightarrow W$ ersetzen durch eine bilineare Abbildung $V_1 \times V_2 \longrightarrow T$ und lineare Abbildungen $T \longrightarrow W$. Genauer definieren wir wie folgt:

Definition/Bemerkung 10.3.1 (Tensorprodukt)

Es seien V, W zwei K-Vektorräume. Ein K-Vektorraum T mit einer bilinearen Abbildung

$$\tau : V \times W \longrightarrow T$$

heißt ein **Tensorprodukt** von V und W (über K), wenn für jeden K-Vektorraum U und jede bilineare Abbildung

$$\beta : V \times W \longrightarrow U$$

genau eine lineare Abbildung $\Phi_\beta : T \longrightarrow U$ existiert, sodass gilt:

$$\beta = \Phi_\beta \circ \tau.$$

Diese Abbildungseigenschaft nennt man die **universelle Abbildungseigenschaft** von τ bezüglich bilinearer Abbildungen.

Die Existenz eines Tensorproduktes T bedeutet, dass man den Vektorraum aller bilinearen Abbildungen von $V \times W$ nach U mit dem Vektorraum $\mathrm{Hom}(T, U)$ identifizieren kann.

Insbesondere gibt es (für $U = T, \beta = \tau$) genau einen Endomorphismus Φ_τ von T, sodass $\tau = \Phi_\tau \circ \tau$. Offensichtlich wird diese Gleichung von $\Phi_\tau = \mathrm{Id}_T$ erfüllt, also ist dies die einzige Lösung.

Das zieht die folgende Eindeutigkeitsaussage nach sich: wenn (T_1, τ_1) und (T_2, τ_2) zwei Tensorprodukte von V und W sind, dann sind ja die Abbildungen $\tau_i : V \times W \longrightarrow T_i$ bilinear, und wir finden Homomorphismen $\Phi_{\tau_1} : T_2 \longrightarrow T_1$ und $\Phi_{\tau_2} : T_1 \longrightarrow T_2$, sodass

$$\tau_1 = \Phi_{\tau_1} \circ \tau_2 \quad \text{und} \quad \tau_2 = \Phi_{\tau_2} \circ \tau_1.$$

Daraus folgt aber durch Hintereinanderausführung:

$$\tau_1 = \Phi_{\tau_1} \circ \tau_2 = \Phi_{\tau_1} \circ \Phi_{\tau_2} \circ \tau_1 \quad \text{und} \quad \tau_2 = \Phi_{\tau_2} \circ \Phi_{\tau_1} \circ \tau_2.$$

Also müssen – nach dem Vorangehenden – Φ_{τ_1} und Φ_{τ_2} zueinander invers sein. Demnach ist das Tensorprodukt bis auf einen eindeutig bestimmten Isomorphismus eindeutig.

Dieser Typ einer Eindeutigkeitsaussage findet sich oft bei der Konstruktion von Objekten, die (wie das Tensorprodukt) durch eine universelle Abbildungseigenschaft definiert werden.

Notation: Für das Tensorprodukt T zweier K-Vektorräume V, W schreiben wir $V \otimes_K W$, und für die bilineare Abbildung τ schreiben wir ab jetzt

$$\otimes : \ V \times W \ni (v, w) \mapsto v \otimes w \in V \otimes W.$$

Frage: Existieren Tensorprodukte?

Beispiel 10.3.2 (mal wieder der K^n)

a) Es seien $V = K^n$, $W = K^m$ Standardvektorräume. Weiter sei $T = K^{n \times m}$. Wir setzen

$$\otimes : K^n \times K^m \longrightarrow T, \quad \otimes(v, w) := v \otimes w := v \cdot w^\top.$$

Dies ist bilinear, und es gilt (zum Beispiel) für die Standardbasisvektoren:

$$e_i \otimes e_j = E_{ij}.$$

Das ist die altbekannte Elementarmatrix aus Definition 4.2.3. Die Elementarmatrizen bilden eine Basis von $K^{n\times m}$. Wenn nun

$$\beta : V \times W \longrightarrow U$$

bilinear ist, so setzen wir (wie in 6.1.2 gelernt) die Abbildung

$$E_{ij} \mapsto \beta(e_i, e_j),\ 1 \leq i \leq n,\ 1 \leq j \leq m$$

die auf einer Basis von $K^{n\times m}$ definiert ist, linear zu einer Abbildung

$$\Phi : K^{n\times m} \longrightarrow U$$

fort. Dann gilt:

$$\beta(v, w) = \beta(\sum_{i=1}^{n} v_i e_i, \sum_{j=1}^{m} w_j e_j) = \sum_{i,j} v_i w_j \beta(e_i, e_j) = \Phi(\sum_{i,j} v_i w_j E_{ij}) = \Phi(v \otimes w).$$

Da dies für alle v, w gilt, folgt

$$\beta = \Phi \circ \otimes$$

wie gewünscht. Dass es keine zweite Wahl für Φ gibt, liegt daran, dass auf den Basisvektoren E_{ij} von $K^{n\times m}$ die Abbildung Φ offensichtlich durch $\Phi(E_{ij}) = \beta(e_i, e_j)$ gegeben sein muss.

An diesem Beispiel sieht man auch sehr deutlich, dass $\otimes$ im Allgemeinen nicht surjektiv sein wird; hier besteht das Bild von $\otimes$ genau aus den Matrizen vom Rang ≤ 1.

b) Nun seien allgemeiner V und W endlichdimensionale Vektorräume. Dann könnten wir Basen wählen und uns damit in die Situation von Beispiel a) manövrieren, um die Existenz des Tensorprodukts sicherzustellen (vgl. 10.3.4). Wir wollen hier aber eine basisfreie Konstruktion liefern, die die obige in ein neues Licht rückt. Dazu setzen wir $T := \mathrm{Hom}(V^*, W)$. Nun brauchen wir eine bilineare Abbildung $\otimes : V \times W \longrightarrow T$. Für ein Paar $(v, w) \in V \times W$ suchen wir also eine Vorschrift, die einem $\alpha \in V^*$ ein Element von W zuordnet. Wir definieren

$$\otimes : V \times W \longrightarrow T = \mathrm{Hom}(V^*, W), \quad (v \otimes w)(\alpha) := \alpha(v) \cdot w.$$

Man rechnet nach, dass dies eine bilineare Abbildung ist. Außerdem gilt für Basen B von V und C von W, dass die Abbildungen $b \otimes c$ eine Basis von T bilden – an dieser Stelle brauchen wir, dass V und W endlichdimensional sind. Dann rechnet man genauso wie in Beispiel a) nach, dass T mit der Abbildung $\otimes$ das Tensorprodukt von V und W ist.

Hilfssatz 10.3.3 (Existenz des Tensorproduktes)

Es seien V und W beliebige K-Vektorräume. Dann existiert ein Tensorprodukt von V und W.

Beweis. Wir brauchen einen Vektorraum T und eine bilineare Abbildung von $V \times W$ nach T mit der universellen Abbildungseigenschaft. Wir werden jetzt eine Anwendung des Quotientenbildens sehen, insbesondere der Homomorphiesatz 5.5.8 wird eine Rolle spielen. Zunächst bauen wir uns einen viel zu großen Vektorraum F mit einer Abbildung von $V \times W$ nach F. Dann bilden wir einen Faktorraum T von F, sodass die zugehörige Abbildung $\otimes$ von $V \times W$ nach T bilinear wird, und für diesen Raum rechnen wir dann mit dem Homomorphiesatz nach, dass er die universelle Abbildungseigenschaft besitzt.

Zunächst sei $F := \mathrm{Abb}(V \times W, K)_0$ (siehe Bemerkung 5.1.10 c)) der Vektorraum der Abbildungen von $V \times W$ nach K mit endlichem Träger. Der Buchstabe „F“ steht für „frei“, das hat Gründe, die sich erst in der Algebra als richtig stichhaltig erweisen. Für $(v, w) \in V \times W$ sei $f_{(v,w)} \in F$ definiert durch

$$\forall (x, y) \in V \times W : f_{(v,w)}(x, y) := \begin{cases} 1, & \text{falls } (v, w) = (x, y), \\ 0, & \text{sonst.} \end{cases}$$

Diese Funktion hat einen Träger mit einem Element, ist also in F. Es ist klar, dass die Menge $B := \{f_{(v,w)} \mid (v, w) \in V \times W\}$ eine Basis von F ist:

$$\forall f \in F : f = \sum_{(v,w) \in V \times W} f(v, w) \cdot f_{(v,w)}.$$

Das ist eine endliche Summe, denn eigentlich langt es über die Paare (v, w) im Träger von f zu summieren. Wir merken uns die Abbildung

$$\varphi : V \times W \longrightarrow F, \quad \varphi(v, w) := f_{(v,w)}.$$

Diese Abbildung ist niemals bilinear, denn zum Beispiel gilt $\varphi(0, 0) \neq 0$. Um die Bilinearität zu erzwingen, führen wir den Untervektorraum R von F ein (die „Relationen“), der von den Vektoren

$$f_{(av_1+v_2, bw_1+w_2)} - abf_{(v_1,w_1)} - af_{(v_1,w_2)} - bf_{(v_2,w_1)} - f_{(v_2,w_2)}$$

erzeugt wird, wobei $a, b \in K, v_1, v_2 \in V, w_1, w_2 \in W$. Mit diesem Untervektorraum R bilden wir den Faktorraum

$$T := F/R.$$

Wir haben nach 5.5.7 die kanonische Projektion $\pi_{F/R} : F \longrightarrow T$, und mit dieser bilden wir

$$\otimes : V \times W \longrightarrow T, (v, w) \mapsto v \otimes w := \pi_{F/R}(\varphi(v, w)) = [f_{(v,w)}].$$

Dabei steht, wie gehabt, das Symbol $[f]$ für die Nebenklasse von $f \in F$ in T. Der Vektorraum R ist nun gerade so gemacht, dass

$$\begin{aligned}[f_{(av_1+v_2,bw_1+w_2)}] &= [abf_{(v_1,w_1)} + af_{(v_1,w_2)} + bf_{(v_2,w_1)} + f_{(v_2,w_2)}] \\ &= ab[f_{(v_1,w_1)}] + a[f_{(v_1,w_2)}] + b[f_{(v_2,w_1)}] + [f_{(v_2,w_2)}]\end{aligned}$$

Das sorgt dafür, dass $\otimes$ bilinear wird.

Nun müssen wir die universelle Abbildungseigenschaft nachrechnen. Dazu seien U ein weiterer K-Vektorraum und $\beta : V \times W \longrightarrow U$ bilinear.

Da F vom Bild von φ erzeugt wird, wird T vom Bild von $\otimes$ erzeugt. Also wird eine lineare Abbildung auf T eindeutig durch ihre Werte auf $\otimes(V \times W)$ festgelegt, und es kann höchstens eine lineare Abbildung $\Phi : T \longrightarrow U$ mit $\beta = \Phi \circ \otimes$ geben.

Um zu zeigen, dass es so eine Abbildung gibt, definieren wir auf F die lineare Abbildung Φ_F als lineare Fortsetzung der Vorschrift

$$\Phi_F(f_{(v,w)}) := \beta(v, w).$$

Da β bilinear ist, gilt für alle $a, b \in K, v_1, v_2 \in V, w_1, w_2 \in W$:

$$\beta(av_1 + v_2, bw_1 + w_2) = ab\beta(v_1, w_1) + a\beta(v_1, w_2) + b\beta(v_2, w_1) + \beta(v_2, w_2),$$

und deshalb

$$\Phi_F(f_{(av_1+v_2,bw_1+w_2)} - abf_{(v_1,w_1)} - af_{(v_1,w_2)} - bf_{(v_2,w_1)} - f_{(v_2,w_2)}) = 0.$$

Also liegt der Untervektorraum R im Kern von Φ_F, und wir erhalten mithilfe des Homomorphiesatzes 5.5.8 eine lineare Abbildung $\Phi : T \longrightarrow U$ durch

$$\Phi([f]) := \Phi_F(f).$$

Speziell gilt also

$$\Phi(v \otimes w) = \Phi([\varphi(v, w)]) = \Phi_F(\varphi(v, w)) = \Phi_F(f_{(v,w)}) = \beta(v, w).$$

Das bedeutet aber gerade $\beta = \Phi \circ \otimes$. ◯

Bemerkung 10.3.4 (Konkretisierung)

Wenn V und W zwei endlichdimensionale Vektorräume sind, in denen wir Basen $B = \{b_1, \ldots, b_n\}$ und $C = \{c_1, \ldots, c_m\}$ gewählt haben, dann betrachten wir die Tensorprodukte $b_i \otimes c_j \in V \otimes_K W$. Diese Vektoren erzeugen $V \otimes_K W$, und da man bilineare Abbildungen von $V \times W$ nach K auf $B \times C$ beliebig vorgeben kann und dies einer linearen Abbildung von $V \otimes_K W$ nach K zu entsprechen hat, müssen sie auch linear unabhängig sein. Also ist die Menge

$$\{b_i \otimes c_j \mid 1 \leq i \leq n,\ 1 \leq j \leq m\}$$

eine Basis von $V \otimes_K W$. Insbesondere gilt

$$\boxed{\dim(V \otimes_K W) = \dim(V) \cdot \dim(W).}$$

Nun seien $\Phi \in \mathrm{End}(V)$, $\Psi \in \mathrm{End}(W)$ zwei Endomorphismen von V und W. Dann ist die Abbildung

$$\beta : V \times W \longrightarrow V \otimes_K W, \quad \beta(v, w) := \Phi(v) \otimes \Psi(w),$$

eine bilineare Abbildung und definiert daher einen eindeutig bestimmten Endomorphismus $\Phi \otimes \Psi$ von $V \otimes_K W$, den wir zum Beispiel so ausrechnen können:

$$\Phi \otimes \Psi(\sum \alpha_{ij}(b_i \otimes c_j)) = \sum \alpha_{ij}\Phi(b_i) \otimes \Psi(c_j).$$

Speziell sehen wir, dass wir aus Abbildungsmatrizen $A := D_{BB}(\Phi)$ und $N := D_{CC}(\Psi)$ eine Abbildungsmatrix von $\Phi \otimes \Psi$ bezüglich der Basis

$$\{b_1 \otimes c_1, b_1 \otimes c_2, \ldots b_1 \otimes c_n, b_2 \otimes c_1, \ldots, b_2 \otimes c_n, \ldots, b_m \otimes c_1, \ldots, b_m \otimes c_n\}$$

in der folgenden Matrix in Blockgestalt finden:

$$\begin{pmatrix} a_{11} \cdot N & a_{12} \cdot N & \ldots & a_{1m} \cdot N \\ a_{21} \cdot N & a_{22} \cdot N & \ldots & a_{2m} \cdot N \\ \vdots & \ldots & \ldots & \vdots \\ a_{m1} \cdot N & a_{m2} \cdot N & \ldots & a_{mm} \cdot N \end{pmatrix}.$$

Diese Matrix heißt das **Kronecker-Produkt** von A und N. Das ist eine Konstruktion, die zum Beispiel in der schnellen Fouriertransformation eine Anwendung findet.

Beispiel 10.3.5 (Erweiterung des Skalarbereichs)

Es seien K ein Körper, V ein K-Vektorraum und L ein Körper, der K als Teilring enthält. Nach Beispiel 5.1.3 f) ist dann auch L ein Vektorraum über K. Man denke etwa an die Situation $K = \mathbb{Q}$, $L = \mathbb{R}$ oder an die Situation $K = \mathbb{R}$, $L = \mathbb{C}$. Nun dürfen wir nach Hilfssatz 10.3.3 das Tensorprodukt $L \otimes_K V$ bilden:

$$L \otimes_K V = \{\sum_{i=1}^{k} \alpha_i \otimes v_i \mid k \in \mathbb{N}, \alpha_i \in L, v_i \in V\}.$$

Für jedes $l \in L$ ist dann die Abbildung

$$\beta : L \times V \longrightarrow L \otimes_K V, \ (\alpha, v) \mapsto l\alpha \otimes v,$$

K-bilinear und definiert damit wie in Beispiel 10.3.4 einen Endomorphismus von $L \otimes_K V$, nämlich $\mu_l \otimes \mathrm{Id}_V$, wobei μ_l die Multiplikation mit l auf L ist.

Wir bekommen damit eine Abbildung

$$\sigma : L \longrightarrow \mathrm{End}(L \otimes_K V), \sigma(l) := \mu_l \otimes \mathrm{Id}_V.$$

Es gelten die Regeln

$$\sigma(1) = \mathrm{Id}_{L\otimes_K V},\ \sigma(l_1 + l_2) = \sigma(l_1) + \sigma(l_2),\quad \sigma(l_1 \cdot l_2) = \sigma(l_1) \circ \sigma(l_2),$$

und da $\sigma(l)$ für jedes l additiv ist, ist damit insgesamt die abelsche Gruppe $L \otimes_K V$ mit σ als Skalarmultiplikation ein L-Vektorraum. Man sagt, dass dieser L-Vektorraum durch **Skalarerweiterung von K nach L** aus V hervorgeht.

$L \otimes_K V$ enthält V als K-Untervektorraum mithilfe der injektiven, K-linearen Abbildung

$$V \longrightarrow L \otimes_K V,\quad v \mapsto 1 \otimes v.$$

Wenn B eine Basis von V als K-Vektorraum ist, dann sind die Elemente $1 \otimes b$, $b \in B$, ein Erzeugendensystem des L-Vektorraumes $L \otimes_K V$, und man rechnet leicht nach, dass sie linear unabhängig sind. Speziell gilt die suggestive Regel

$$\boxed{L \otimes_K K^n \cong L^n \text{ als } L\text{-Vektorraum.}}$$

Ein Endomorphismus Φ von V als K-Vektorraum liefert den Endomorphismus $\mathrm{Id}_L \otimes \Phi$ von $L \otimes_K V$ als L-Vektorraum, der bezüglich der Basis $\{1 \otimes b \mid b \in B\}$ durch dieselbe Matrix beschrieben wird wie die Ausgangsabbildung bezüglich der Basis B. Das liefert uns mit der obigen Identifizierung $L \otimes_K K^n \cong L^n$ nichts anderes als die offensichtliche Inklusion $K^{n\times n} \subseteq L^{n\times n}$.

10.4 Algebren

Wir haben schon viele Beispiele von Ringen R kennengelernt, die gleichzeitig Vektorräume über einem Körper K sind: $K[X], K^{n\times n}$, Körper L, die K enthalten, wie etwa $L = \mathbb{C}$ für $K = \mathbb{R}$ oder $\mathbb{Q}$. Eine Gemeinsamkeit dieser Ringe ist, dass jeweils K im Zentrum des Ringes liegt. In Definition 3.3.7 lesen wir nach, dass das bedeutet:

$$\forall k \in K, r \in R : k \cdot r = r \cdot k.$$

Zusammen mit Assoziativ- und Distributivgesetz sorgt das dafür, dass die Multiplikation von $R \times R$ nach R K-bilinear ist. Anders gesagt: für alle $r_1, r_2, s \in R$, $k \in K$ gilt

$$(kr_1 + r_2) \cdot s = k(r_1 \cdot s) + r_2 \cdot s \text{ und } s \cdot (kr_1 + r_2) = k(sr_1) + sr_2.$$

Das nehmen wir jetzt als Anlass für eine etwas allgemeinere Definition.

Definition 10.4.1 (K-Algebra)

Es seien K ein Körper und A ein K-Vektorraum, der gleichzeitig ein Ring ist. Dann heißt A eine **K-Algebra**, wenn die Multiplikation $A \times A \longrightarrow A$ eine K-bilineare Abbildung ist.

Wenn wir kurz vergessen, dass unsere Ringe immer eine Eins haben, dann sehen wir tatsächlich Beispiele von K-Algebren, die nicht den Körper K als Teilring enthalten. Zum Beispiel ist die Menge der stetigen Funktionen von $\mathbb{R}$ nach $\mathbb{R}$ mit beschränktem Träger ein Ring (punktweise Addition und Multiplikation) ohne Eins, der eine $\mathbb{R}$-Algebra ist, aber keinen zu $\mathbb{R}$ isomorphen Teilring enthält.

Ein anderes Beispiel ist die Menge $\mathrm{Abb}(\mathbb{N} \times \mathbb{N}, K)_0$ („aufsteigende Vereinigung“ der $K^{n \times n}$). Dies ist mit der üblichen Matrizenmultiplikation und -addition ein Ring (ohen Eins!) und ist auch ein K-Vektorraum. Zwei Elemente liegen immer in einem gemeinsamen endlichen Matrizenring, weshalb das Produkt K-bilinear ist. In diesem Ring bilden die Matrizen mit beliebigem Eintrag $k \in K$ an der (1,1)-Stelle, 0 sonst, einen Teilring R, der zu K isomorph ist. Aber die Multiplikation mit dem Element k aus R tut etwas anderes, als die Multiplikation mit dem Element $k \in K$. Solche Einbettungen von K in A sind zunächst nicht hilfreich.

Aber wir wollen ja nur Ringe mit Eins anschauen, und da bekommen wir die folgende Aussage.

Hilfssatz 10.4.2 (K liegt in der Algebra)

Es seien K ein Körper und A eine K-Algebra mit Einselement $1_A \neq 0$. Dann ist die Abbildung

$$\iota : K \longrightarrow A, \quad x \mapsto x \cdot 1_A,$$

ein injektiver Ringhomomorphismus, dessen Bild im Zentrum von A liegt. Außerdem passt die Algebrenstruktur zur Vektorraumstruktur im folgenden Sinne:

$$\forall k \in K, a \in A : k \cdot a = \iota(k)a.$$

Beweis. Dass ι ein Ringhomomorphismus ist, ist klar. Die Injektivität folgt dann aus 3.2.3.

Nun sei $a \in A$ beliebig, wie auch $k \in K$. Dann gilt

$$a\iota(k) = a(k \cdot 1_A) \overset{(*)}{=} k \cdot (a1_A) = (k \cdot 1_A a) = \iota(k)a.$$

Schließlich gilt

$$\iota(k)a \overset{(*)}{=} k \cdot (1_A a) = k \cdot a.$$

Dabei sind jeweils mit $(*)$ die Stellen markiert, wo die Bilinearität der Multiplikation benutzt wird. ◯

Wir werden in Zukunft nicht mehr zwischen $1 \in K$ und $1_A \in A$ unterscheiden und nehmen stets K als in A eingebettet an: eine K-Algebra ist ein Ring mit Eins, der K als Teilring seines Zentrums enthält. In Zukunft sagen wir auch nicht immer dazu, dass unsere K-Algebren eine Eins haben, auch wenn wir dies ab jetzt immer voraussetzen.

Nun kommen die üblichen Definitionen, deren Erstellung man fast schon als Übungsaufgabe im Fach „sinnvolle Mathematik“ stellen möchte.

Definition 10.4.3 (Unteralgebra, Algebrenhomomorphismus)

a) Es seien K ein Körper und A eine K-Algebra. Eine **K-Unteralgebra** von A ist ein Teilring von A, der ebenfalls K enthält.

b) Es seien A und B Algebren über demselben Körper K. Dann ist ein Ringhomomorphismus Φ von A nach B, der auch noch K-linear ist, ein **K-Algebrenhomomorphismus**.

Beispiel 10.4.4 (Einsetzabbildung)

Es sei $Q \in K^{n\times n}$ eine beliebige (quadratische) Matrix. Dann ist die Abbildung

$$E_Q : K[X] \longrightarrow K^{n\times n},\ f(X) \mapsto f(Q),$$

ein K-Algebrenhomomorphismus. Das Bild davon ist die K-Teilalgebra $K[Q]$ von $K^{n\times n}$. Beim Studium der Jordan'schen Normalform von Q untersucht man letztlich die Struktur dieser Algebra und ihre Wirkung auf K^n. Genauer wird aus K^n ein $K[X]$- und auch ein $K[Q]$-Modul mit der folgenden Definition.

Definition/Bemerkung 10.4.5 (Moduln)

Ein **Modul** über einem Ring R ist eine abelsche Gruppe M mit einer Abbildung

$$\cdot : R \times M \longrightarrow M,$$

falls die folgenden Regeln erfüllt sind:

- $\forall m \in M :$ $\quad 1 \cdot m = m,$
- $\forall r, s \in R, m \in M :$ $\quad (rs) \cdot m = r \cdot (s \cdot m),$
- $\forall r, s \in R, m, n \in M :$ $\quad (r+s) \cdot m = r \cdot m + s \cdot m,$ $\quad r \cdot (m+n) = r \cdot m + r \cdot n.$

Beim Vergleich dieser Definition mit der Definition des Vektorraumes sollte auffallen, dass es nur einen Unterschied gibt: hier muss R kein Körper sein. Ansonsten bleibt alles gleich.

Wenn R nun eine K-Algebra ist (K ist ein Körper), dann ist ein R-Modul M durch die Einschränkung von $\cdot$ auf $K \times M$ automatisch ein K-Vektorraum, und für jedes $r \in R$ ist die Abbildung

$$M \ni m \mapsto r \cdot m \in M$$

ein Endomorphismus dieses K-Vektorraumes. Das heißt, wir bekommen einen Homomorphismus von R in $\mathrm{End}_K(M)$.

Umgekehrt liefert für jeden K-Vektorraum M jeder Algebrenhomomorphismus $\Phi : R \longrightarrow \mathrm{End}_K(M)$ eine R-Modulstruktur auf M vermöge

$$r \cdot m := (\Phi(r))(m).$$

Dies wenden wir jetzt auf den Fall $M = R$ an. Die Algebrenmultiplikation macht aus R einen R-Modul. Die daraus resultierende Abbildung

$$\Phi : R \longrightarrow \mathrm{End}_K(R), \quad \Phi(r)(s) := r \cdot s$$

ist also ein Algebrenhomomorphismus, und dieser ist injektiv:

$$\forall r, s \in R : \Phi(r) = \Phi(s) \Rightarrow \Phi(r)(1) = \Phi(s)(1) \Rightarrow r \cdot 1 = s \cdot 1 \Rightarrow r = s.$$

Nun identifizieren wir R mit seinem Bild im Ring $\mathrm{End}_K(R)$ der Vektorraum-Endomorphismen von R:

$$\boxed{R \subseteq \mathrm{End}_K(R).}$$

Jede K-Algebra ist in einem Endomorphismenring enthalten. Speziell ist jede n-dimensionale K-Algebra isomorph zu einer Teilalgebra des Matrizenrings $K^{n\times n}$. Vergleichen Sie dieses Ergebnis mit dem „Satz von Cayley“ in 2.5.3!

Definition/Bemerkung 10.4.6 (Strukturkonstanten)

Es sei R eine n-dimensionale K-Algebra mit Basis $B = \{b_1, \ldots, b_n\}$. Dann gibt es eindeutig bestimmte Elemente $c_k^{ij} \in K,\ i, j, k \in \{1, \ldots n\},$ sodass

$$b_i \cdot b_j = \sum_{k=1}^{n} c_k^{ij} b_k.$$

Diese Koeffizienten (genauer: die Abbildung $\{1, \ldots, n\}^3 \longrightarrow K, (i, j, k) \mapsto c_k^{ij}$) heißen die **Strukturkonstanten** von R (bezüglich der Basis B). Aus diesen Strukturkonstanten lässt sich durch bilineare Fortsetzung die Algebrenmultiplikation zurückgewinnen.

Umgekehrt kann man versuchen, aus einem beliebigen Vektorraum V mit Basis C durch Einführung von „künstlichen“ Strukturkonstanten eine Algebra zu machen. Man muss dann immer nachrechnen, ob die so (durch bilineare Fortsetzung „künstlich“) definierte Multiplikation das Assoziativ- und Distributivgesetz erfüllen – was im Allgemeinen nicht der Fall sein wird. Manchmal ist es einfacher zu überprüfen, ob es eine Teilalgebra des Matrizenringes gibt, die eine Basis

mit den richtigen Strukturkonstanten besitzt. Das werden wir gleich an einem Beispiel vorführen.

Selbstverständlich sind zwei endlichdimensionale K-Algebren genau dann isomorph zueinander, wenn sie dieselbe Dimension haben und bezüglich geeigneter Basen dieselben Strukturkonstanten besitzen.

Beispiel 10.4.7 (Quaternionenalgebren)

Nun heiße unser Körper einmal F, denn die Geschichte zwingt uns, mit K gleich einen Vektor zu bezeichnen.

Es seien $a, b \in F^\times$. Wir wollen eine vierdimensionale F-Algebra mit einer Basis namens $\{1, I, J, K\}$ (sic!) basteln, die die folgenden Regeln erfüllt.

$$I^2 = a \cdot 1, \quad J^2 = b \cdot 1, \quad IJ = K, JI = -K.$$

1 soll natürlich das Einselement sein, und insgesamt ergibt sich durch Überlegungen wie

$$IK = I(IJ) = I^2 J = aJ, \quad K^2 = (IJ)^2 = -IJJI = -IbI = -bI^2 = -ba,$$

die folgende Multiplikationstabelle (im Feld steht „Zeilenname“ mal „Spaltenname“).

$\cdot$	1	I	J	K
1	1	I	J	K
I	I	a	K	aJ
J	J	$-K$	b	$-bI$
K	K	$-aJ$	bI	$-ab$

Aber Vorsicht: wir haben noch nicht überprüft, ob dies auch wirklich eine Algebra aus dem vierdimensionalen Vektorraum A mit Basis $\{1, I, J, K\}$ macht. Wir wissen nur: wenn die erstgenannte Vorschrift überhaupt konsistent zu einer Algebrenmultiplikation fortgesetzt werden kann, dann muss die Multiplikationstabelle gelten. Ist dies nun insgesamt wirklich konsistent?

Zur Überprüfung gibt es zwei Möglichkeiten. Entweder man rechnet „blindwütig“ nach, dass die durch die Multiplikationstabelle gegebene bilineare Abbildung von $A \times A$ nach A aus A einen Ring macht. Das funktioniert und ist nur etwas mühsam.

Wir wissen aber: wenn A eine Algebra ist, dann muss es auch eine vierdimensionale Teilalgebra von $F^{4\times 4}$ geben, die zu A isomorph ist, also eine Basis mit derselben Multiplikationstabelle besitzt. Es ist etwas eleganter, diesen Ansatz weiter zu verfolgen, zumal ich zufälliger Weise so eine Teilalgebra kenne. Genauer sagt uns ja die Multiplikationstafel, wie die Abbildung Φ aus Beispiel 10.4.5 c) sich in Abbildungsmatrizen bezüglich der Basis $\{1, I, J, K\}$ niederschlägt.

Nehmen wir also an, A sei eine Algebra. Dann sehen wir der Reihe nach:
Die Multiplikation mit I (von links) ist gegeben durch die Matrix

$$\widetilde{I} := \begin{pmatrix} 0 & a & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & a \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Die Multiplikation mit J (von links) ist gegeben durch die Matrix

$$\widetilde{J} := \begin{pmatrix} 0 & 0 & b & 0 \\ 0 & 0 & 0 & -b \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}.$$

Schließlich ist die Multiplikation mit K gegeben durch die Matrix

$$\widetilde{K} := \begin{pmatrix} 0 & 0 & 0 & -ab \\ 0 & 0 & b & 0 \\ 0 & -a & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

Man rechnet nun nach, dass die folgenden Relationen gelten:

$$(\widetilde{I})^2 = aI_4, (\widetilde{J})^2 = bI_4, \widetilde{I}\widetilde{J} = \widetilde{K} = -\widetilde{J}\widetilde{I}.$$

Damit sind die Bedingungen wirklich konsistent, denn nun haben wir eine Teilalgebra einer bekannten Algebra, die eine Basis mit den gewünschten Strukturkonstanten besitzt.

Die hier konstruierte Algebra heißt eine **Quaternionenalgebra**, sie wird oft als $\left(\frac{a,b}{F}\right)$ notiert.

Beispiel 10.4.8 (Hamiltons Quaternionen, Schiefkörper)

Das prominenteste Beispiel für eine Quaternionenalgebra ist die Algebra $\mathbb{H} := \left(\frac{-1,-1}{\mathbb{R}}\right)$ der **Hamilton-Quaternionen**. Das ist ein reeller Vektorraum mit einer Basis $\{1, I, J, K\}$, und die Struktur als Algebra wird diktiert von den Bedingungen

$$I^2 = J^2 = -1,\ IJ = -JI = K.$$

Man rechnet leicht nach, dass zum Beispiel folgendes gilt:

$$(w + xI + yJ + zK) \cdot (w - xI - yJ - zK) = w^2 + x^2 + y^2 + z^2.$$

Insbesondere ist ein Element aus $\mathbb{H}$ genau dann invertierbar, wenn es nicht 0 ist:

$$q = w + xI + yJ + zK \neq 0 \Rightarrow q^{-1} = \frac{1}{w^2 + x^2 + y^2 + z^2}(w - xI - yJ - zK).$$

Es gilt also für die Einheitengruppe die Gleichung

$$\mathbb{H}^\times = \mathbb{H} \setminus \{0\}.$$

Ein Ring (der nicht der Nullring ist), in dem jedes von Null verschiedene Element invertierbar ist, heißt ein **Schiefkörper**.

Die Teilmenge

$$Q_8 := \{\pm 1, \pm I, \pm J, \pm K\} \subset \mathbb{H}^\times$$

ist eine Untergruppe der Einheitengruppe, sie heißt die Quaternionengruppe. Es ist eine nicht-abelsche Gruppe mit 8 Elementen.

Man kann $\mathbb{H}$ wie oben in den Ring $\mathbb{R}^{4\times 4}$ einbetten, aber man kommt hier sogar mit 2×2-Matrizen aus, wenn man den Koeffizientenkörper etwas größer macht. Die komplexen Matrizen

$$\mathbf{1} := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, I := \begin{pmatrix} \mathrm{i} & 0 \\ 0 & -\mathrm{i} \end{pmatrix}, J := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, K := \begin{pmatrix} 0 & -\mathrm{i} \\ -\mathrm{i} & 0 \end{pmatrix}$$

erzeugen einen vierdimensionalen reellen Vektorraum, und es gilt

$$I^2 = J^2 = -\mathbf{1}, IJ = -JI = K.$$

Also ist dieser reelle Vektorraum eine Unter-$\mathbb{R}$-Algebra von $\mathbb{C}^{2\times 2}$, die zu $\mathbb{H}$ isomorph ist.

Wir könnten also auch schreiben

$$\mathbb{H} = \{\begin{pmatrix} z & -\overline{w} \\ w & \overline{z} \end{pmatrix} \mid z, w \in \mathbb{C}\},$$

wobei $\overline{(x + y\mathrm{i})} = x - y\mathrm{i}$ wie in 3.2.7.

Nun finden wir in

$$\{\begin{pmatrix} z & 0 \\ 0 & \overline{z} \end{pmatrix} \mid z \in \mathbb{C}\} \subseteq \mathbb{H}$$

einen Teilring von $\mathbb{H}$, der zu $\mathbb{C}$ isomorph ist. Aber Vorsicht: $\mathbb{H}$ ist keine $\mathbb{C}$-Algebra, denn $\mathbb{C}$ liegt nicht im Zentrum von $\mathbb{H}$.

Kleines Amusement am Rande: Die Multiplikation mit Elementen von $\mathbb{C}$ auf $\mathbb{H}$ macht aus $\mathbb{H}$ einen $\mathbb{C}$-Vektorraum. Ich muss mich aber entscheiden, ob ich von links oder von rechts die Skalare wirken lasse. Das gibt zwei verschiedene $\mathbb{C}$-Vektorraumstrukturen auf $\mathbb{H}$.

Bemerkung 10.4.9 (Tensorieren) Nun seien K wieder ein beliebiger Körper und A eine K-Algebra. Weiter sei L ein Körper, der K als Teilring enthält. Dann können wir das Tensorprodukt

$$L \otimes_K A =: A_L$$

bilden. Aus 10.3.5 wissen wir, dass dies ein L-Vektorraum ist, der über L die gleiche Dimension hat wie V über K. Wenn B eine K-Basis von A ist, dann ist $B_L := \{1 \otimes b \mid b \in B\}$ eine L-Basis von A_L. Wenn man nun auf B_L dieselben Vorgaben für die Multiplikation macht wie sie auf B durch die Algebrenstruktur von A gelten (Strukturkonstanten), dann wird durch L-bilineare Fortsetzung dieser Vorgaben aus A_L eine L-Algebra.

Wenn man ohne Verwendung einer Basis argumentieren will, so kann man für feste Elemente $(l, a) \in L \times A$ eine bilineare Abbildung $\mu_{(l,a)}$ von $L \times A$ nach $L \otimes_K A$ definieren durch

$$\forall (m, b) \in L \times A : \ \mu_{(l,a)}((m, b)) := (lm) \otimes (ab).$$

Es existiert also eine eindeutig bestimmte K-lineare Abbildung

$$m_{(l,a)} : A_L \longrightarrow A_L, \sum_i m_i \otimes b_i \mapsto \sum_i lm_i \otimes ab_i.$$

Das liefert eine bilineare Abbildung

$$m : L \times A \longrightarrow \mathrm{End}_K(L \otimes A),$$

und dies wiederum bringt nach Definition 10.3.1 einen eindeutig bestimmten K-Vektorraumhomomorphismus

$$M : L \otimes A \longrightarrow \mathrm{End}_K(L \otimes A).$$

Dieser entspricht einer K-bilinearen Abbildung

$$A_L \times A_L \longrightarrow A_L, \quad (a, b) \mapsto M(a)(b).$$

Damit wird (wie man nun noch nachrechnen muss) A_L sogar zu einer L-Algebra.

Es gelten zum Beispiel die folgenden Isomorphismen von L-Algebren:

$$L \otimes_K (K^{n \times n}) \cong L^{n \times n}, \quad L \otimes_K K[X] \cong L[X], \quad \mathbb{C} \otimes_{\mathbb{R}} \mathbb{H} \cong \mathbb{C}^{2 \times 2}.$$

Ein Vergleich des ersten und des letzten Beispiels zeigt, dass es verschiedene K-Algebren geben kann, die nach Tensorieren isomorphe L-Algebra liefern. Man spricht hier von verschiedenen **K-Formen** der resultierenden L-Algebra.

Kapitel 11

Skalarprodukte

Aus der Schule sollte das Standardskalarprodukt im $\mathbb{R}^3$ bekannt sein. Hier werden wir seine grundlegenden Eigenschaften herausstellen und zum Konzept erheben. Zunächst machen wir das für reelle Vektorräume, nachher für komplexe. Über anderen Körpern als $\mathbb{R}$ oder $\mathbb{C}$ ist es nicht so leicht möglich, eine gleichermaßen befriedigende Theorie zu entwickeln. Vieles stimmt bei beliebigen Körpern noch für sogenannte anisotrope Bilinearformen, aber darauf gehen wir hier nicht ein.

11.1 Skalarprodukte, Längen und Abstände

Bemerkung 11.1.1 (Standardskalarprodukt auf $\mathbb{R}^3$)

Das Standardskalarprodukt auf $\mathbb{R}^3$ ist die Abbildung

$$\langle \cdot, \cdot \rangle : \mathbb{R}^3 \times \mathbb{R}^3 \longrightarrow \mathbb{R},\ \langle v, w \rangle := v^\top \cdot w,$$

oder – wenn es jemand konkreter liebt –

$$\langle \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}, \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} \rangle = v_1 w_1 + v_2 w_2 + v_3 w_3.$$

Man will nun von der genauen Form dieser Abbildung abstrahieren und damit die Möglichkeit gewinnen, auf beliebigen reellen Vektorräumen Abbildungen mit ähnlichen Eigenschaften zu definieren. Dazu stellt man die folgenden drei Eigenschaften heraus:

- Das Skalarprodukt ist eine Bilinearform auf $\mathbb{R}^3$ (Definition 10.1.1) .
- Das Skalarprodukt ist symmetrisch (Definition 10.1.9).

- Für $v \in \mathbb{R}^3$ gilt: $v \neq 0 \Rightarrow \langle v, v\rangle = \sum_{i=1}^{3} v_i^2 > 0$. Denn mindestens ein Summand ist positiv, und keiner negativ.

Diese Eigenschaften benutzt man nun. Das Standardskalarprodukt ist ein Beispiel für ein Skalarprodukt, wenn man folgende Definition macht.

Definition 11.1.2 (Skalarprodukt, euklidischer Vektorraum)

Es sei V ein $\mathbb{R}$-Vektorraum.

a) Eine symmetrische Bilinearform $\langle \cdot, \cdot \rangle : V \times V \longrightarrow \mathbb{R}$ heißt **positiv definit**, wenn gilt:

$$\forall v \in V : v \neq 0 \Longrightarrow \langle v, v\rangle > 0.$$

b) Ein **Skalarprodukt** auf V ist eine symmetrische, positiv definite Bilinearform.

c) Ein reeller Vektorraum mit einem fest gewählten Skalarprodukt heißt ein **euklidischer Vektorraum**.

Beispiel 11.1.3 (dies und das und das Standardskalarprodukt)

a) Auf dem Nullvektorraum gibt es genau ein Skalarprodukt.

b) Auf $V = \mathbb{R}$ gibt es genau die Skalarprodukte

$$\langle x, y\rangle_\alpha := \alpha \cdot x \cdot y,$$

wobei α die positiven reellen Zahlen durchläuft. Immerhin sind das unendlich viele Skalarprodukte.

c) Auf $\mathbb{R}^2$ wird ein Skalarprodukt durch seine Fundamentalmatrix bezüglich der Standardbasis beschrieben (siehe 10.1.4):

$$\langle v, w\rangle := v^\top \cdot F \cdot w.$$

Dabei ist $F = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \in \mathbb{R}^{2\times 2}$ symmetrisch (siehe 10.1.9), und wir müssen noch erkennen, welche Wahlen von F positiv definite Bilinearformen liefern. Zunächst muss

$$a = \langle e_1, e_1\rangle > 0$$

gelten, denn e_1 ist ein Vektor $\neq 0$. Dann brauchen wir außerdem für alle $(x, y) \in \mathbb{R}^2 \setminus \{(0, 0)\}$ die Aussage

$$ax^2 + 2bxy + cy^2 = (x\ y) \cdot F \cdot \begin{pmatrix} x \\ y \end{pmatrix} > 0,$$

was wegen $a > 0$ zu $b^2 - ac < 0$ äquivalent ist (Lösungsformel für quadratische Polynome, Zwischenwertsatz).

Wir fassen zusammen: Die Matrix $F = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \in \mathbb{R}^{2\times 2}$ ist genau dann die Fundamentalmatrix eines Skalarprodukts auf $\mathbb{R}^2$, wenn $a > 0$ und $\det(F) > 0$. Das werden wir in 11.2.11 verallgemeinern.

d) Das **Standardskalarprodukt** auf dem $\mathbb{R}^n$ wird gegeben durch

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

Der $\mathbb{R}^n$ mit diesem Skalarprodukt heißt der **n-dimensionale euklidische Standardraum.**

e) Es sei $I \subseteq \mathbb{R}$ ein kompaktes Intervall mit Länge $l > 0$ und $V \subseteq \mathrm{Abb}(I, \mathbb{R})$ ein Untervektorraum, der aus stetigen Funktionen besteht. Dann wird auf V durch

$$V \times V \ni (f, g) \mapsto \langle f, g\rangle := \int_I f(x)g(x)\mathrm{d}x$$

ein Skalarprodukt definiert. Das Integral existiert, denn es wird eine stetige Funktion über ein Kompaktum integriert. Die Bilinearität ist klar (aus Distributivgesetzen und der Linearität des Integrals); die Symmetrie ist klar, da $fg = gf$ gilt. Dass die Bilinearform positiv definit ist sieht man so: wenn $f \neq 0$ eine Funktion in V ist, dann gibt es $x_0 \in I$ mit $f(x_0) \neq 0$. Wegen der Stetigkeit von f gibt es für $\epsilon := \frac{1}{2}|f(x_0)| > 0$ ein $\delta > 0$, sodass für alle $x \in I$ mit $|x - x_0| < \delta$ auch die Ungleichung

$$|f(x)| > \epsilon$$

gilt. Wähle hierbei $\delta < l/2$. Dann ist

$$\langle f, f\rangle = \int_I f^2(x)\mathrm{d}x > \delta\epsilon^2 > 0,$$

denn sicher liegt eines der Intervalle $[x_0, x_0 + \delta]$ und $[x_0 - \delta, x_0]$ ganz in I, und auf diesem Intervall ist $f^2 > \epsilon^2$.

Bemerkung 11.1.4 Eine wichtige Bedeutung von Skalarprodukten liegt darin, dass man mit ihnen Abstände zwischen Vektoren definieren kann – wir werden gleich sehen wie – und so etwas braucht man zum Beispiel, um Näherungslösungen für Funktionen mit gewünschten Eigenschaften definieren zu können. Was sollte „Nähe“ schon präzise bedeuten, wenn man keinen Abstandsbegriff hat? Für verschiedene Fragestellungen sind dabei verschiedene Abstandsbegriffe hilfreich, und deswegen ist es gut, wenn man eine große Flexibilität hat und sich einige Sachverhalte in großer Allgemeinheit erarbeitet.

Das geht noch allgemeiner als mit Skalarprodukten, siehe Definition 11.1.7.

Definition 11.1.5 (Norm, Länge, Abstand)

Es seien V ein euklidischer Vektorraum und $\langle \cdot, \cdot \rangle$ das Skalarprodukt auf V. Für einen Vektor $v \in V$ heißt dann die nichtnegative Quadratwurzel

$$\|v\| := \sqrt{\langle v, v \rangle}$$

die **Norm** oder auch die **Länge** von v (bezüglich des gewählten Skalarproduktes).

Für zwei Vektoren v, w heißt die reelle Zahl

$$d(v, w) := \|v - w\|$$

der **Abstand** zwischen v und w. Die Abbildung $d : V \times V \longrightarrow \mathbb{R}$ heißt die zum Skalarprodukt gehörende **Metrik**.

Die Positivität des Skalarproduktes sorgt dafür, dass erstens die Norm immer eine nichtnegative reelle Zahl ist und zweitens zwei Vektoren genau dann gleich sind, wenn ihr Abstand 0 ist. Außerdem ist die Metrik **symmetrisch**, d.h. für alle $v, w \in V$ gilt $d(v, w) = d(w, v)$.

Nun erwartet man von einer anständigen Abstandsfunktion, dass der Abstand zwischen zwei Punkten nicht kleiner wird, wenn man einen Umweg macht (siehe Definition 11.1.7). Dass dies auch für die Abstandsfunktion, die von einem Skalarprodukt herkommt, gilt, überlegen wir uns jetzt.

Satz 11.1.6 (Ungleichung von Cauchy-Schwarz)

Es sei V ein euklidischer Vektorraum mit Skalarprodukt $\langle \cdot, \cdot \rangle$.

a) Für alle $v, w \in V$ gilt die Ungleichung von Cauchy-Schwarz-Bunyakowski-. . . , *die sagt:*

$$\langle v, w \rangle^2 \leq \langle v, v \rangle \cdot \langle w, w \rangle.$$

Die Gleichheit gilt hier genau dann, wenn v und w linear abhängig sind.

b) Für alle $u, v, w \in V$ gilt die Dreiecksungleichung, *nämlich*

$$d(u, v) + d(v, w) \geq d(u, w).$$

Beweis.

a) Wenn $v = 0$ gilt, dann steht links und rechts in der Ungleichung 0, also stimmt sie, und auch der Zusatz ist wahr, denn 0 und w sind linear abhängig. Wir dürfen also $v \neq 0$ annehmen, und betrachten dann die folgende Funktion $f : \mathbb{R} \to \mathbb{R}$:

$$f(x) := \|xv + w\|^2 = \langle xv + w, xv + w \rangle = \langle v, v \rangle x^2 + 2\langle v, w \rangle x + \langle w, w \rangle.$$

Die Funktion f ist eine polynomiale Abbildung, und der Grad ist 2, da $\langle v, v\rangle > 0$. Außerdem sagt die Positivität des Skalarprodukts auch noch, dass

$$f(x) \geq 0$$

gilt. Also hat f nicht zwei verschiedene Nullstellen, da sonst die Funktionswerte zwischen diesen Nullstellen negativ wären. Die Lösungsformel für quadratische Gleichungen verbietet also, dass die Diskriminante von f positiv ist. Das bedeutet:

$$\langle v, w\rangle^2 - \langle v, v\rangle\langle w, w\rangle \leq 0.$$

Das ist genau die Aussage der Ungleichung von Cauchy-Schwarz und Co.

Wenn Gleichheit gilt, dann gibt es eine Nullstelle von f, also ein x mit $f(x) = 0$, und die Positivität des Skalarproduktes erzwingt dann $xv + w = 0$, also sind v und w linear abhängig.

Wenn umgekehrt v und w linear abhängig sind, dann gibt es ein reelles x_0 mit $w = x_0 v$, denn v ist ja nicht Null. Dann gilt aber $f(-x_0) = 0$, und somit ist die Diskriminante 0, denn positiv kann sie nicht werden, wie wir oben schon gesehen hatten.

b) Wir setzen der Einfachheit halber $\widetilde{u} := u - v$, $\widetilde{w} := w - v$. Dann sagt die zu beweisende Ungleichung gerade

$$\|\widetilde{u}\| + \|\widetilde{w}\| \geq \|\widetilde{u} - \widetilde{w}\|.$$

Da links und rechts nichtnegative reelle Zahlen stehen, ist diese Ungleichung äquivalent zur entsprechenden Ungleichung zwischen den Quadraten der Seiten:

$$\langle\widetilde{u}, \widetilde{u}\rangle + 2\|\widetilde{u}\| \cdot \|\widetilde{w}\| + \langle\widetilde{w}, \widetilde{w}\rangle \geq \langle\widetilde{u} - \widetilde{w}, \widetilde{u} - \widetilde{w}\rangle = \langle\widetilde{u}, \widetilde{u}\rangle - 2\langle\widetilde{u}, \widetilde{w}\rangle + \langle\widetilde{w}, \widetilde{w}\rangle.$$

Nachdem hier links und rechts die Summanden $\langle\widetilde{u}, \widetilde{u}\rangle$ und $\langle\widetilde{w}, \widetilde{w}\rangle$ abgezogen wurden, bleibt eine Ungleichung übrig, die von der Cauchy-Schwarz-Ungleichung impliziert wird. Da wir nur Äquivalenzumformungen mit der Ungleichung vornahmen, ist damit der Satz bewiesen. ◯

Definition/Bemerkung 11.1.7 (Exkurs über metrische und normierte Räume)

Für eine beliebige Menge X heißt eine Funktion $d : X \times X \longrightarrow \mathbb{R}$ eine **Metrik** und das Paar (X, d) heißt ein **metrischer Raum**, wenn d die folgenden drei Eigenschaften hat:

- $\forall x, y \in X : d(x, y) = d(y, x)$. (Symmetrie)
- $\forall x, y \in X : d(x, y) \geq 0$ und $[d(x, y) = 0 \iff x = y]$. (Positivität)

- $\forall x, y, z \in X : d(x, y) + d(y, z) \geq d(x, z)$. (Dreiecksungleichung)

Wir haben in Bemerkung 11.1.5 und Satz 11.1.6 also gesehen, dass mithilfe eines Skalarproduktes auf einem euklidischen Vektorraum eine Metrik definiert wird. Es gibt aber noch ganz andere Möglichkeiten, Metriken zu gewinnen. Ein schwächerer Begriff als der des Skalarproduktes ist der der Norm. Eine **Norm** auf einem reellen Vektorraum V ist eine Abbildung

$$N : V \longrightarrow \mathbb{R}$$

mit den folgenden Eigenschaften:

- $\forall v \in V : N(v) \geq 0,$ und $[N(v) = 0 \iff v = 0]$.
- $\forall v \in V, a \in \mathbb{R} : N(av) = |a|N(v)$
- $\forall v, w \in V : N(v) + N(w) \geq N(v + w)$.

Das Paar (V, N) heißt dann ein **normierter Vektorraum.**

Durch ein Skalarprodukt ist die Norm $N(v) := \sqrt{\langle v, v \rangle}$ definiert, aber es gibt meistens auch noch andere Normen. Zum Beispiel gibt es auf $V = \mathbb{R}^n$ die Norm

$$N(x) := \max\{|x_i| \mid 1 \leq i \leq n\},$$

die für $n \geq 2$ nicht von einem Skalarprodukt herkommt.

Eine beliebige Norm liefert eine Metrik durch $d(v, w) := N(v - w)$. Aber nicht alle Metriken auf reellen Vektorräumen erhält man so. Zum Beispiel ist auf jeder Menge X die **diskrete Metrik** d_0 definiert durch

$$d_0(x, y) := \begin{cases} 0 & \text{falls } x = y, \\ 1 & \text{sonst.} \end{cases}$$

Wenn X eine reeller Vektorraum ist, dann kommt diese diskrete Metrik genau dann von einer Norm her, wenn $V = \{0\}$.

Jedenfalls halten wir fest, in welcher Reihenfolge die Begriffe auseinander hervorgehen:

$$\boxed{\text{Skalarprodukt} \rightsquigarrow \text{Norm} \rightsquigarrow \text{Metrik.}}$$

Definition 11.1.8 (Winkel, Orthogonalität)

In dem euklidischen Vektorraum V seien zwei Vektoren v, w gegeben, beide seien ungleich dem Nullvektor. Dann dürfen wir in der Ungleichung von Cauchy und Schwarz die Quadratwurzel ziehen und durch $\|v\| \cdot \|w\|$ teilen, und erhalten

$$-1 \leq \frac{\langle v, w \rangle}{\|v\| \cdot \|w\|} \leq 1.$$

Also gibt es genau eine reelle Zahl $\alpha \in [0, \pi]$ mit

$$\cos \alpha = \frac{\langle v, w \rangle}{\|v\| \cdot \|w\|}.$$

Diese Zahl heißt der **Winkel** zwischen v und w, und wir schreiben dafür $\angle(v, w)$. Es gilt zum Beispiel

$$\angle(v, w) = \pi - \angle(-v, w).$$

Zwei beliebige Vektoren $v, w \in V$ heißen **orthogonal**, wenn $\langle v, w \rangle = 0$. Falls v und w beide nicht Null sind, dann heißt das, dass der Winkel zwischen ihnen $\pi/2$ (also $90°$) beträgt.

Notation: Wenn v, w orthogonal sind, so schreiben wir $v \perp w$.

Für unsere Zwecke wird im Allgemeinen der Begriff der Orthogonalität wichtiger sein als der des Winkels an sich. Ich weise noch einmal darauf hin, dass die Begriffe Länge, Winkel und Orthogonalität immer vom gewählten Skalarprodukt abhängen, und nicht durch den Vektorraum als solchen schon bestimmt sind. Auf den Begriff der Orthogonalität waren wir schon in 10.1.9 gestoßen, Skalarprodukte sind ja ein Spezialfall von symmetrischen Bilinearformen.

Bemerkung 11.1.9 (Satz des Pythagoras)

Für zwei Vektoren v, w im euklidischen Vektorraum V gilt:

$$\|v + w\|^2 = \langle v, v \rangle + 2\langle v, w \rangle + \langle w, w \rangle,$$

wobei wir die Symmetrie benutzen. Wir können also Orthogonalität so charakterisieren:

$$\boxed{v \perp w \iff \|v\|^2 + \|w\|^2 = \|v + w\|^2.}$$

11.2 Orthonormalbasen

Definition 11.2.1 (Orthonormalsystem)

Es seien V ein euklidischer Vektorraum und $S \subseteq V$ eine Teilmenge.

Dann heißt S ein **Orthogonalsystem**, wenn $0 \notin S$ und wenn die Elemente aus S paarweise orthogonal (siehe 11.1.8) sind. S heißt ein **Orthonormalsystem** , wenn es ein Orthogonalsystem ist und alle Vektoren in S Norm 1 (siehe 11.1.5) haben.

Hilfssatz 11.2.2 (Orthogonalsysteme sind linear unabhängig)

Es sei S ein Orthogonalsystem in einem euklidischen Vektorraum V. Dann ist S linear unabhängig.

Beweis. Es sei $\alpha \in \mathrm{Abb}(S, \mathbb{R})_0$ eine Abbildung mit endlichem Träger. Für diese Abbildung gelte

$$\sum_{s \in S} \alpha(s) \cdot s = 0.$$

Wir müssen zeigen (siehe Definition 5.3.8), dass α die Nullabbildung ist.

Dafür nehmen wir das Skalarprodukt dieser Linearkombination mit einem beliebigen $s_0 \in S$:

$$0 = \langle s_0, 0\rangle = \langle s_0, \sum_{s \in S} \alpha(s) \cdot s\rangle = \sum_{s \in S} \alpha(s) \cdot \langle s_0, s\rangle = \alpha(s_0) \cdot \langle s_0, s_0\rangle.$$

Da aber $\langle s_0, s_0\rangle > 0$ gilt, folgt $\alpha(s_0) = 0$. Hierbei ist s_0 beliebig, also gilt $\alpha = 0$. ◯

Bemerkung 11.2.3 (Normierung)

Aus jedem Orthogonalsystem S lässt sich ein Orthonormalsystem $\widetilde{S}$ herstellen durch

$$\widetilde{S} := \{\frac{1}{\|s\|} \cdot s \mid s \in S\}.$$

Wenn ein Orthogonalsystem den Vektorraum erzeugt, dann ist es (weil linear unabhängig) eine Basis (siehe Satz 5.3.12). Man spricht dann von einer **Orthogonalbasis** beziehungsweise im Fall eines Orthonormalsystems von einer **Orthonormalbasis** (kurz auch ONB), in Übereinstimmung mit Definition 10.1.9.

Hilfssatz 10.1.10 sagt uns, dass es in einem endlichdimensionalen euklidischen Vektorraum immer eine Orthogonalbasis und damit – nach dem eben Gesagten – auch eine Orthonormalbasis gibt. Wir werden das in Satz 11.2.6 unabhängig von Kapitel 10 noch einmal zeigen und konkretisieren.

Bemerkung 11.2.4 (Fourierformel)

Wir wiederholen in Anlehnung an Bemerkung 10.1.11 die Fourierformel. Wenn $B = \{b_1, \ldots, b_n\}$ eine Orthonormalbasis des euklidischen Vektorraums V ist, dann gilt für alle $v \in V$

$$\boxed{v = \sum_{i=1}^{n} \langle v, b_i\rangle b_i.}$$

Das heißt: der Koordinatenvektor von v bezüglich der Basis B ist der Vektor

$$D_B(v) = (\langle v, b_i\rangle)_{1 \le i \le n} \in \mathbb{R}^n.$$

Wichtig ist hierbei, dass B wirklich eine Orthonormalbasis ist.

Für zwei Vektoren v, w rechnet man nach, dass

$$\langle v, w\rangle = D_B(v)^\top \cdot D_B(w)$$

gilt, dass also die Abbildung $D_B : V \longrightarrow \mathbb{R}^n$ benutzt werden kann, um Skalarprodukte in V durch Skalarprodukte im euklidischen Standardraum auszurechnen. Wir werden solche Abbildungen in Kapitel 12 noch einmal systematisch untersuchen.

Der Name „Fourierformel“ kommt daher, dass Fourier für eine gewisse Klasse von Funktionen eine ähnliche Formel entwickelt hat, wo allerdings eine unendliche Reihe vonnöten ist (also nicht wie bei uns endliche Summen ausreichen). Die Qualität der Konvergenz dieser unendlichen Reihen ist eine delikate Fragestellung im Rahmen der harmonischen Analysis.

Bemerkung 11.2.5 (orthogonale Matrizen, $\mathrm{O}(n)$, $\mathrm{SO}(n)$)

Es sei $V = \mathbb{R}^n$ mit dem Standardskalarprodukt versehen. Dann ist die n-elementige Menge

$$\{v_1, \ldots, v_n\} \subseteq V$$

genau dann eine Orthonormalbasis von V, wenn für die reelle $n \times n$-Matrix $A = (v_1\ v_2\ \ldots\ v_n)$ die Gleichung

$$A^\top \cdot A = I_n$$

gilt. Denn der (i, j)-te Eintrag der Matrix $A^\top \cdot A$ ist gerade das Skalarprodukt von v_i mit v_j.

Wir definieren nun die **orthogonale Gruppe** $\mathrm{O}(n) \subseteq \mathrm{GL}_n(\mathbb{R})$ durch

$$\mathrm{O}(n) := \{A \in \mathbb{R}^{n\times n} \mid A^\top \cdot A = I_n\}.$$

Die Elemente von $\mathrm{O}(n)$ heißen **orthogonale Matrizen** (auch wenn sie vielleicht besser orthonormale Matrizen hießen...). Ihre „konzeptionelle“ Bedeutung werden wir später noch besser verstehen, zunächst wollen wir uns klar machen, dass $\mathrm{O}(n)$ eine Gruppe ist.

- $\mathrm{O}(n) \subseteq \mathrm{GL}_n(\mathbb{R})$, denn $A \in \mathrm{O}(n)$ ist zu $A^\top$ invers, also regulär.

  Wir können also versuchen, das Untergruppenkriterium 2.2.3 zu verwenden.

- $\mathrm{O}(n)$ ist nicht leer, da die Einheitsmatrix I_n offensichtlich darin liegt.

- Es seien $A, B \in \mathrm{O}(n)$. Dann gilt auch $A \cdot B^{-1} \in \mathrm{O}(n)$, denn $B^{-1} = B^\top$ und

$$(A \cdot B^{-1})^\top \cdot A \cdot B^{-1} = (B^{-1})^\top \cdot A^\top \cdot A \cdot B^{-1} = B \cdot B^\top = I_n.$$

Nach 2.2.3 ist also $\mathrm{O}(n)$ eine Untergruppe von $\mathrm{GL}_n(\mathbb{R})$.

Für jede orthogonale Matrix A gilt $1 = \det(A^\top \cdot A) = \det(A)^2$, also ist die Determinante einer orthogonalen Matrix entweder 1 oder -1. Beides kommt vor. Die Determinante liefert wegen des Determinantenmultiplikationssatzes (7.1.7 b)) einen Gruppenhomomorphismus

$$\det : \mathrm{O}(n) \longrightarrow \mathbb{R}^\times.$$

Der Kern dieses Homomorphismus ist die Gruppe

$$\mathrm{SO}(n) := \{A \in \mathrm{O}(n) \mid \det(A) = 1\}$$

der **speziellen orthogonalen $n \times n$-Matrizen.**

Satz 11.2.6 (Orthogonalisierungsverfahren von E. Schmidt)

Es seien V ein euklidischer Vektorraum und $\{v_1, v_2, \ldots, v_k\} \subset V$ eine linear unabhängige Teilmenge mit k Elementen.

Wir definieren neue Vektoren $w_1, \ldots, w_k$ rekursiv durch

$$w_1 := v_1, \quad w_l := v_l - \sum_{i=1}^{l-1} \frac{\langle v_l, w_i\rangle}{\langle w_i, w_i\rangle} \cdot w_i \quad (\text{für } l = 2, \ldots k).$$

Dann ist die Menge $S := \{w_1, \ldots, w_k\}$ ein Orthogonalsystem in V.

Weiterhin ist die Menge $\widetilde{S} := \{\frac{1}{\|w_1\|} \cdot w_1, \ldots, \frac{1}{\|w_k\|} \cdot w_k\}$ ein Orthonormalsystem in V.

Schließlich sind für jedes i mit $1 \leq i \leq k$ die linearen Hüllen von $\{v_1, \ldots, v_i\}$ und von $\{w_1, \ldots, w_i\}$ gleich.

Beweis. Zunächst ist der Nullvektor kein Element von S, denn $w_1 = v_1$ ist nicht Null, und rekursiv ist $w_l = v_l -$ (Linearkombination von $v_1, \ldots, v_{l-1}$), also nicht 0, da die v_i laut Voraussetzung linear unabhängig waren. Nun zeigen wir noch für $1 \leq l \leq k$, dass für alle $1 \leq j < l$ gilt: $\langle w_j, w_l\rangle = 0$.

Dazu machen wir Induktion nach l.

Für $l = 1$ stimmt die Behauptung, denn es gibt ja gar kein j.

Wenn $l \geq 2$ und $1 \leq j < l$ gegeben sind, dann gilt

$$\begin{aligned}
\langle w_j, w_l\rangle &= \langle w_j, v_l - \textstyle\sum_{i=1}^{l-1} \frac{\langle v_l, w_i\rangle}{\langle w_i, w_i\rangle} \cdot w_i\rangle \\
&= \langle w_j, v_l\rangle - \textstyle\sum_{i=1}^{l-1} \frac{\langle v_l, w_i\rangle}{\langle w_i, w_i\rangle} \cdot \langle w_j, w_i\rangle \\
&\overset{(*)}{=} \langle w_j, v_l\rangle - \frac{\langle v_l, w_j\rangle}{\langle w_j, w_j\rangle} \cdot \langle w_j, w_j\rangle \\
&= 0.
\end{aligned}$$

Dabei verwenden wir im Schritt $(*)$, dass nach Induktionsvoraussetzung für $1 \leq i, j < l$ schon klar ist, dass w_i und w_j nur dann Skalarprodukt ungleich 0 haben, wenn $i = j$ gilt.

Das zeigt die Orthogonalität von S, und die Orthonormalität von $\widetilde{S}$ folgt daraus wie in Bemerkung 11.2.3.

Die Aussage über die linearen Hüllen stimmt, da die Vektoren $w_1, \ldots, w_i$ Linearkombinationen von $v_1, \ldots, v_i$ sind und da beide Mengen linear unabhängig sind (11.2.2), also Vektorräume derselben Dimension erzeugen (und dann greift 5.3.18). ◯

Bemerkung 11.2.7 (Der euklidische Standardraum)

Es sei $\{v_1, \ldots, v_n\}$ eine Basis des euklidischen Standardraums (11.1.3) $\mathbb{R}^n$. Wir schreiben die Vektoren in eine Matrix $A := (v_1\ v_2 \ldots\ v_n)$. Diese Matrix ist nach Beispiel 5.3.3 invertierbar.

Wenn wir das Verfahren aus Satz 11.2.6 verwenden, um aus dieser Basis eine Orthonormalbasis zu machen, dann ändern wir dabei die Spalten der Matrix durch elementare Spaltenumformungen, multiplizieren also A mit einer invertierbaren Matrix C von rechts. Diese Matrix ist eine obere Dreiecksmatrix mit positiven Einträgen auf der Diagonalen. Denn für den ersten Schritt (Herstellung der Orthogonalbasis) werden von v_l nur Linearkombinationen der v_j mit $j < l$ abgezogen, das wird durch eine Dreiecksmatrix mit Einsen auf der Diagonalen beschrieben. Nachher wird alles (von rechts) mit einer Diagonalmatrix mit positiven Einträgen multipliziert, um die Spalten auf Norm 1 zu bringen. Es gilt dann:

$$A \cdot C \in \mathrm{O}(n).$$

Nun ist aber die Menge $\mathcal{B}(n)$ der oberen $n \times n$-Dreiecksmatrizen mit positiven Diagonaleinträgen eine Untergruppe von $\mathrm{GL}_n(\mathbb{R})$, wie sich leicht nachrechnen lässt. Also gilt

$$A \in \mathrm{O}(n) \cdot C^{-1} \subseteq \mathrm{O}(n) \cdot \mathcal{B}(n).$$

Jede invertierbare Matrix A lässt sich als Produkt einer orthogonalen Matrix und einer oberen Dreiecksmatrix mit positiven Diagonaleinträgen schreiben. Diese Zerlegung von A ist auf eindeutige Art möglich. Wäre nämlich

$$A = S_1 \cdot C_1 = S_2 \cdot C_2$$

auf zwei Arten als Produkt des erwähnten Typs zu schreiben, so folgte

$$S_2^{-1} \cdot S_1 = C_2 \cdot C_1^{-1} \in \mathrm{O}(n) \cap \mathcal{B}(n).$$

Aber die Spalten einer (regulären) oberen Dreiecksmatrix sind genau dann ein Orthogonalsystem, wenn die obere Dreiecksmatrix diagonal ist, und da die Einträge positiv sind, sind die Spalten genau dann eine Orthonormalbasis, wenn die Matrix die Einheitsmatrix ist. Es folgt also $C_1 = C_2$ und $S_1 = S_2$.

Da sich jedes invertierbare A zerlegen lässt, folgt

$$\mathrm{GL}_n(\mathbb{R}) = \mathrm{O}(n) \cdot \mathcal{B}(n).$$

Dies nennt man die **Iwasawa-Zerlegung** von $\mathrm{GL}_n(\mathbb{R})$.

Beispiel 11.2.8 (mit ZAHLEN)
Im euklidischen Standardraum $\mathbb{R}^4$ (11.1.3) seien die Vektoren

$$v_1 := \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, v_2 := \begin{pmatrix} -1 \\ 0 \\ 1 \\ 2 \end{pmatrix}, v_3 := \begin{pmatrix} 1 \\ 0 \\ 1 \\ 4 \end{pmatrix}, v_4 := \begin{pmatrix} -1 \\ 0 \\ 1 \\ 8 \end{pmatrix}$$

gegeben. Diese sind linear unabhängig, und wir wollen aus ihnen eine Orthonormalbasis des $\mathbb{R}^4$ (bezüglich des Standardskalarproduktes) berechnen. Dazu verwenden wir Verfahren 11.2.6 und setzen erst einmal $w_1 := v_1$. Der zweite Vektor ist dann

$$w_2 := v_2 - \frac{\langle v_2, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 = v_2 - \frac{1}{2} w_1 = \frac{1}{2} \begin{pmatrix} -3 \\ -1 \\ 1 \\ 3 \end{pmatrix}.$$

Damit ergibt sich w_3 zu

$$w_3 := v_3 - \frac{\langle v_3, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 - \frac{\langle v_3, w_2 \rangle}{\langle w_2, w_2 \rangle} w_2 = v_3 - \frac{6}{4} w_1 - \frac{10/2}{20/4} w_2 = \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}.$$

Schließlich finden wir w_4 nach derselben Methode:

$$w_4 := v_4 - \frac{\langle v_4, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 - \frac{\langle v_4, w_2 \rangle}{\langle w_2, w_2 \rangle} w_2 - \frac{\langle v_4, w_3 \rangle}{\langle w_3, w_3 \rangle} w_3 = \frac{3}{10} \begin{pmatrix} -1 \\ 3 \\ -3 \\ 1 \end{pmatrix}.$$

Da wir am Ende eine Orthonormalbasis herausbekommen wollten, müssen wir diese Vektoren noch normieren, also durch ihre Norm teilen. Das führt zur Orthonormalbasis

$$\frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{20}} \begin{pmatrix} -3 \\ -1 \\ 1 \\ 3 \end{pmatrix}, \frac{1}{2} \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{20}} \begin{pmatrix} -1 \\ 3 \\ -3 \\ 1 \end{pmatrix}.$$

Beispiel 11.2.9 (Orthogonale Polynome)

Es sei $I \subseteq \mathbb{R}$ ein kompaktes Intervall positiver Länge. Nach dem Approximationssatz von Stone-Weierstraß lässt sich jede stetige Funktion f auf I gleichmäßig durch Polynome approximieren (also im Sinne der L^∞-Norm). Wenn $(p_i)_{i\in\mathbb{N}}$ solch eine Folge von Polynomen ist, dann konvergiert die Folge

$$\int_I (p_i(x) - f(x))^2 \mathrm{d}x$$

gegen 0, denn der Integrand geht ja gleichmäßig gegen 0.

Vorsicht: Die Umkehrung wird im Allgemeinen nicht stimmen; die Integrale könnten eine Nullfolge bilden, ohne dass gleichmäßige Konvergenz vorliegt.

Man wird jedenfalls darauf geführt, ein Skalarprodukt wie in Beispiel 11.1.3 e) zu verwenden. Eine naheliegende Aufgabe ist es dann, im Vektorraum V der reellen Polynome vom Grade $\leq n$ eine Orthonormalbasis bezüglich dieses Skalarproduktes zu finden. Dies kann man mit dem Verfahren von E. Schmidt, ausgehend von der Basis $\{1, x, x^2, \ldots, x^n\}$ tun. Man bekommt damit das, was man **orthogonale Polynome** nennt.

Das geht auch noch etwas allgemeiner. Dazu sei $K : I \longrightarrow \mathbb{R}$ eine stetige Funktion, die nur positive Werte annimmt. Dann ist auch

$$V \times V \ni (f, g) \mapsto \int_I K(x) f(x) g(x) \mathrm{d}x \in \mathbb{R}$$

ein Skalarprodukt, und man kann ebenso nach einer Orthonormalbasis von V bezüglich dieses Skalarproduktes (mit **Integralkern** K) fragen. Manchmal führen zum Beispiel physikalische Bedingungen auf die Untersuchung solch eines Skalarproduktes; wie gut, dass es dieses allgemeine Konzept gibt, und nicht nur das Standardskalarprodukt!

Ein Beispiel für orthogonale Polynome sind die Legendre-Polynome

$$P_n(x) := \frac{1}{2^n n!} \frac{\mathrm{d}^n (x^2 - 1)^n}{\mathrm{d}x^n},$$

deren erste Vertreter die Polynome $P_0(x) = 1,\ P_1(x) = x,\ P_2(x) = \frac{1}{2}(3x^2 - 1)$ sind. Man kann nachrechnen, dass gilt:

$$\int_{-1}^{1} P_n(x) P_m(x) \mathrm{d}x = \begin{cases} 0, & m \neq n, \\ \frac{2}{2n+1}, & m = n. \end{cases}$$

Ein weiteres Beispiel für orthogonale Polynome sind die sogenannten Tschebyscheff'schen Polynome (die man z.B. im *Bronstein* findet).

Bemerkung 11.2.10 (Positivität der Fundamentalmatrix)

Es sei V ein endlichdimensionaler euklidischer Vektorraum. Wir wählen darin eine Basis $B = \{b_1, \ldots, b_n\}$. Dann ist die Fundamentalmatrix (vgl. 10.1.4) des Skalarproduktes $\langle \cdot, \cdot \rangle$ bezüglich B die Matrix

$$D_{BB}(\langle \cdot, \cdot \rangle) := (\langle b_i, b_j \rangle)_{1 \leq i,j \leq n}.$$

Dies ist eine reelle, symmetrische und reguläre Matrix. Das Skalarprodukt von zwei Vektoren $v = \sum_{i=1}^n v_i b_i,\ w = \sum_{i=1}^n w_i b_i$ ist dann gegeben durch

$$\langle v, w \rangle = (v_1 v_2 \ldots v_n) \cdot D_{BB}(\langle \cdot, \cdot \rangle) \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}.$$

Das heißt, dass durch die Matrix $F := D_{BB}(\langle \cdot, \cdot \rangle)$ ein Skalarprodukt auf $\mathbb{R}^n$ durch folgende Vorschrift gegeben ist:

$$(x, y) \mapsto x^\top \cdot F \cdot y.$$

Wir nennen eine Matrix $F \in \mathbb{R}^{n \times n}$ **positiv definit**, wenn sie symmetrisch ist und für alle $x \in \mathbb{R}^n$ die Äquivalenz

$$x^\top F x > 0 \iff x \neq 0$$

gilt.

Welche Matrizen erfüllen diese Bedingung? Wir verallgemeinern 11.1.3 c), wo wir dies für den Fall $n = 2$ schon untersucht haben.

Satz 11.2.11 (Kriterium für Positivität)

Es seien $n \in \mathbb{N}$ und $F = (f_{ij}) \in \mathbb{R}^{n \times n}$. Weiter sei F symmetrisch. Dann sind die folgenden drei Eigenschaften äquivalent:

i) F ist positiv definit.

ii) Es gibt eine obere Dreiecksmatrix $A \in \mathrm{GL}_n(\mathbb{R})$ mit $F = A^\top \cdot A$.

iii) Für $1 \leq k \leq n$ sind die Determinanten der Matrizen $F_k := (f_{ij})_{1 \leq i,j \leq k}$ positiv.

Beweis. Wir zeigen, dass sowohl i) als auch iii) zu ii) äquivalent sind.

„i)$\Rightarrow$ ii)“ Wenn F positiv definit ist, dann wird durch F ein Skalarprodukt auf $\mathbb{R}^n$ definiert, sodass F bezüglich der Standardbasis die Fundamentalmatrix

dieses Skalarproduktes ist. Nach dem Verfahren von E. Schmidt gibt es eine obere Dreiecksmatrix D, deren Spalten eine Orthonormalbasis bezüglich des durch F definierten Skalarproduktes bilden. Nach 10.1.4 bedeutet dies

$$D^\top \cdot F \cdot D = I_n, \quad \text{also} \quad F = A^\top \cdot A \quad \text{mit} \quad A = D^{-1}.$$

„ii)$\Rightarrow$ i)“ Es sei A invertierbar mit $F = A^\top \cdot A$. Dann gilt für $v \in \mathbb{R}^n,\ v \neq 0$, dass

$$v^\top \cdot F \cdot v = v^\top \cdot A^\top \cdot A \cdot v = (Av)^\top \cdot (Av) > 0,$$

denn $Av \neq 0$.

„ii)$\Rightarrow$ iii)“ Es sei A eine invertierbare Dreiecksmatrix mit $F = A^\top \cdot A$. Dann gilt $F_k = A_k^\top \cdot A_k$, wobei F_k bzw. A_k aus F bzw. A durch „Abschneiden“ der letzten $n-k$ Zeilen und Spalten entstehen. (Hier benutzen wir die Dreiecksform von A.) Aber mit A sind auch die Matrizen A_k regulär. Denn: mit A ist auch A_k eine Dreiecksmatrix, deren Diagonaleinträge alle nicht 0 sind. Dann ist aber $\det F_k = (\det A_k)^2 > 0$ wie gewünscht.

„iii)$\Rightarrow$ ii)“ Es sei F gegeben, und es gelte $\det F_k > 0$ für alle k zwischen 1 und n. Wir konstruieren rekursiv obere Dreiecksmatrizen $A_k \in \mathbb{R}^{k\times k}$ mit $A_k^\top \cdot A_k = F_k$.

Für $k = 1$ setzen wir $A_1 := (\sqrt{f_{11}})$. Nun nehmen wir an, wir hätten A_k schon konstruiert für ein k mit $1 \leq k \leq n-1$. Wir machen für A_{k+1} den Ansatz

$$A_{k+1} := \begin{pmatrix} A_k & s_k \\ 0 & t_k \end{pmatrix},\ s_k \in \mathbb{R}^k, t_k \in \mathbb{R}.$$

Dann soll natürlich gelten

$$F_{k+1} = \begin{pmatrix} F_k & c_{k+1} \\ c_{k+1}^\top & f_{k+1,k+1} \end{pmatrix} \stackrel{!}{=} A_{k+1}^\top \cdot A_{k+1} = \begin{pmatrix} A_k^\top \cdot A_k & A_k^\top \cdot s_k \\ s_k^\top \cdot A_k & s_k^\top \cdot s_k + t_k^2 \end{pmatrix}.$$

Dies erzwingt

$$s_k = (A_k^\top)^{-1} \cdot c_{k+1},$$

denn A_k ist ja regulär. Zudem soll $\det(F_{k+1}) := \det(A_{k+1})^2 = t_k^2 \cdot \det(A_k)^2$ gelten, was den Ansatz $t_k := \sqrt{\det(F_{k+1})}/\det(A_k)$ nahelegt. Hier brauchen wir die Voraussetzung aus iii).

Nun ist aber A_{k+1} so gewählt, dass alle Einträge von F_{k+1} und $A_{k+1}^\top \cdot A_{k+1}$ bis auf höchstens den Eintrag an der Stelle $(k+1, k+1)$ übereinstimmen, und dass diese beiden Matrizen dieselbe Determinanten haben. Wenn man beide Determinanten nach der letzten Zeile entwickelt (wie in 7.3.3) und benutzt, dass $\det(F_k) \neq 0$, dann folgt daraus, dass auch der letzte Eintrag der beiden Matrizen übereinstimmt. Es gilt also

$$F_{k+1} = A_{k+1}^\top \cdot A_{k+1}.$$

Für $k+1 = n$ ist das Eigenschaft ii) aus der Formulierung der Aussage. ◯

Bemerkung 11.2.12 (Hurwitz-Kriterium)

Die Determinanten $\det(F_k)$ heißen die **Hauptminoren** von F. Die Minoren von F sind die Determinanten der quadratischen Matrizen, die durch Streichung von Zeilen und Spalten aus F entstehen.

Die Implikation von iii) nach i) im letzten Satz nennt man das **Hurwitz-Kriterium** für die Positivität. Wir werden in Kapitel 12 noch ein weiteres Kriterium für die Positivität einer symmetrischen reellen Matrix sehen. Solch eine Matrix ist nämlich (siehe Satz 12.2.5) immer diagonalisierbar, und sie ist positiv definit genau dann, wenn alle Eigenwerte positiv sind. Das ist aber eben ein Satz, und nicht die natürliche Definition.

11.3 Orthogonale Komplemente und Abstände

Definition 11.3.1 (Orthogonalraum)

Es sei V ein euklidischer Vektorraum und $M \subseteq V$ eine Teilmenge. Dann definieren wir den **Orthogonalraum zu M** (i.Z.: $M^\perp$, sprich „M senkrecht“) durch

$$M^\perp := \{v \in V \mid \forall m \in M : m \perp v\} = \{v \in V \mid \forall m \in M : \langle v, m\rangle = 0\}.$$

$M^\perp$ ist ein Untervektorraum von V, denn $0 \in M^\perp$ und für $v, w \in M^\perp$ und $\beta \in \mathbb{R}$ gilt

$$\forall m \in M : \langle \beta v + w, m\rangle = \beta\langle v, m\rangle + \langle w, m\rangle = 0, \quad \text{also} \quad \beta v + w \in M^\perp.$$

NB: Man kann wie in Hilfssatz 10.1.5 das Skalarprodukt (das ja eine nicht ausgeartete Paarung ist) benutzen, um V als Teilraum des Dualraumes V^* zu verstehen. Dann ist $M^\perp$ der Durchschnitt der Kerne der zu $m \in M$ gehörenden Linearformen.

Die folgenden Relationen lassen sich leicht verifizieren:

$$N \subseteq M \Rightarrow M^\perp \subseteq N^\perp, \quad M^\perp = \langle M\rangle^\perp,$$

wobei $\langle M\rangle$ die lineare Hülle (siehe 5.1.9) von M ist.

Hilfssatz 11.3.2 (Orthogonales Komplement)

In einem euklidischen Vektorraum V sei ein endlichdimensionaler Untervektorraum U gegeben. Dann ist $U^\perp$ ein zu U komplementärer Untervektorraum von V.

$U^\perp$ heißt das **Orthogonale Komplement zu** *U.*

Beweis. Es ist schon bekannt, dass $U^\perp$ ein Untervektorraum von V ist, und für $u \in U \cap U^\perp$ gilt nach Definition von $U^\perp$ insbesondere $\langle u, u\rangle = 0$, also ist $u = 0$, da das Skalarprodukt positiv definit ist. Das heißt $U \cap U^\perp = \{0\}$.

Wir müssen noch zeigen, dass U und $U^\perp$ den ganzen Raum V aufspannen. Das Verfahren von E. Schmidt erlaubt es uns, eine Orthonormalbasis $B_U := \{b_1, \ldots, b_d\}$ von U zu wählen. Es sei $v \in V$. Wir müssen zeigen, dass v sich schreiben lässt als $v = u + u^\perp$, wobei $u \in U$ und $u^\perp \in U^\perp$. Wenn v bereits in U liegt, dann geht das mit $u^\perp = 0$. Wenn v nicht in U liegt, dann sind $b_1, \ldots, b_d, v$ linear unabhängig, und wir können mit E. Schmidt ein zugehöriges Orthonormalsystem konstruieren. Dabei ändern sich die ersten d Vektoren nicht, und v wird ersetzt durch

$$w := v - \sum_{i=1}^{d} \langle v, b_i\rangle b_i.$$

Dann liegt w in $U^\perp$, und es gilt

$$v = \sum_{i=1}^{d} \langle v, b_i\rangle b_i + w \in U \oplus U^\perp.$$

◯

Beispiel 11.3.3 (wo es nicht klappt...)

Um zu verstehen, dass hier die endliche Dimension tatsächlich eine Rolle spielt, machen wir ein unendlichdimensionales Beispiel. Es sei V der reelle Vektorraum der stetigen reellwertigen Funktionen auf dem Intervall $[0, 1]$ mit dem Skalarprodukt $\langle f, g\rangle := \int_0^1 f(x)g(x)\mathrm{d}x$. In V liegt der Vektorraum U der durch Polynome gegebenen Funktionen. Natürlich gilt $U \neq V$. Was ist $U^\perp$? Die Funktion f liege in $U^\perp$. Dann gibt es einerseits (siehe Beispiel 11.2.9) für jedes $\epsilon > 0$ eine Polynomfunktion $g \in U$ mit $d(f, g) \leq \epsilon$. Andererseits gilt

$$\epsilon^2 \geq d(f, g)^2 = \|f - g\|^2 = \langle f - g, f - g\rangle = \langle f, f\rangle + \langle g, g\rangle \geq \langle f, f\rangle = \|f\|^2,$$

wobei zwischendurch $\langle f, g\rangle = 0$ benutzt wird. Das heißt aber, dass für jedes $\epsilon > 0$ die Ungleichung $\|f\| \leq \epsilon$ gilt, also ist $\|f\| = 0$, also $f = 0$. Wir finden damit insgesamt $U^\perp = \{0\}$.

Definition 11.3.4 (Orthogonale Projektion, Abstand)

a) Es seien V ein euklidischer Vektorraum und U ein endlichdimensionaler Untervektorraum von V. Wegen 11.3.2 gilt dann $V = U \oplus U^\perp$. Zu dieser Zerlegung von V gehört der Homomorphismus

$$\pi_U : V \longrightarrow U, \quad \pi_U(u + u^\perp) := u.$$

Er heißt die **orthogonale Projektion** (von V auf U längs $U^\perp$). Manchmal (sogar recht häufig) wird π_U als Endomorphismus von V betrachtet, denn U liegt ja in V. Dann ist $U^\perp = \mathrm{Kern}(\pi_U)$ und $U = \mathrm{Eig}(\pi_U, 1)$. Es gilt $\pi_U^2 = \pi_U$.

Genauso gibt es die Projektion auf das Orthogonale Komplement, $\pi_{U^\perp}$, mit Kern U und Eigenraum $U^\perp$ zum Eigenwert 1.

b) Es seien V ein euklidischer Vektorraum und $A, B \subseteq V$ zwei nichtleere Teilmengen. Dann definieren wir den **Abstand** von A und B durch

$$d(A,B) := \inf\{d(a,b) \mid a \in A,\ b \in B\}.$$

Dieses Infimum existiert, da alle Abstände nicht negativ sind. Speziell schreiben wir für ein Element $a \in V$ auch

$$d(a,B) := d(\{a\},B).$$

Beispiel 11.3.5

a) In der Situation des letzten Beispiels 11.3.3 hat eine beliebige Funktion $f \in V$ vom Raum U der polynomialen Funktionen Abstand 0. Interessanter wird es, wenn man anstelle von U endlichdimensionale Teilräume von U betrachtet, zum Beispiel den Vektorraum aller Polynomfunktionen vom Grade $\leq k$ für ein festes k. Dann kann man sich für jedes $f \in V$ fragen, welches Polynom vom Grade $\leq k$ die Funktion f am besten approximiert (bezüglich des durch das Skalarprodukt definierten Abstandes), und wie groß der Abstand ist. Solche Fragen werden systematischer in der Numerischen Mathematik behandelt und führen zum Beispiel auch wieder auf orthogonale Polynome (11.2.9).

b) Im dreidimensionalen Standardraum $\mathbb{R}^3$ sei

$$B := \{v \in \mathbb{R}^3 \mid \|v\| \leq 1\}.$$

Das ist die Vollkugel vom Radius 1. Weiter sei $a \in \mathbb{R}^3$ ein beliebiger Vektor. Was ist $d(a,B)$? Es sind zwei Fälle zu unterscheiden.

Fall 1: $\|a\| \leq 1$. Dann gilt $a \in B$, und der Abstand ist 0.

Fall 2: $l := \|a\| > 1$. Dann ist $b_1 := \frac{1}{l} \cdot a \in B$ ein Vektor von Länge 1. Wir können diesen zu einer Orthonormalbasis $\{b_1, b_2, b_3\}$ von $\mathbb{R}^3$ ergänzen. Dann ist $v = c_1b_1 + c_2b_2 + c_3b_3$ genau dann in B, wenn $c_1^2 + c_2^2 + c_3^2 \leq 1$. Wir können also jeden Vektor $v \in B$ schreiben als $v = c_1b_1 + c_2b_2 + c_3b_3$, und hierbei muss sicher $|c_i| \leq 1$ gelten. Nun rechnet man nach

$$d(v,a)^2 = \|c_1b_1 + c_2b_2 + c_3b_3 - lb_1\|^2 = (l - c_1)^2 + c_2^2 + c_3^2,$$

und das wird (Extremum mit Nebenbedingungen!) minimal, wenn $c_1 = 1, c_2 = c_3 = 0$. Damit ist der Abstand gleich $l - 1$, und der Punkt auf B, der am nächsten an a liegt, ist b_1.

Naja, das entspricht ja sogar unserer Intuition, und so darf Mathematik gelegentlich auch sein.

Abstände von einem Untervektorraum lassen sich mit der folgenden Methode auf Abstände von einem Punkt zurückführen.

Satz 11.3.6 (Abstand von einem Untervektorraum)

Es seien V ein euklidischer Vektorraum und U ein endlichdimensionaler Untervektorraum. Wir zerlegen V als $V = U \oplus U^\perp$. Dann gelten:

a) Für $v = u + u^\perp \in V$ gilt die Gleichung

$$d(v, U) = \|u^\perp\| = \|\pi_{U^\perp}(v)\|.$$

b) Allgemeiner gilt für eine beliebige Teilmenge $A \subseteq V$ die Formel

$$d(A, U) = d(\pi_{U^\perp}(A), 0).$$

c) Im Falle $A = v + W$ (wobei W ein endlichdimensionaler Untervektorraum von V ist) gilt

$$d(A, U) = \|\pi_{(U+W)^\perp}(v)\|.$$

Beweis.

a) Es sei $\widetilde{u} \in U$. Dann gilt wegen Pythagoras (11.1.9):

$$d(\widetilde{u}, v)^2 = \|(\widetilde{u} - u) - u^\perp\|^2 = \|\widetilde{u} - u\|^2 + \|u^\perp\|^2 \geq \|u^\perp\|^2.$$

Hier gilt Gleichheit genau dann, wenn $u = \widetilde{u}$, und das zeigt die Behauptung.

NB: Das zweite Gleichheitszeichen der Aussage folgt unmittelbar aus der Definition von $\pi_{U^\perp}(v)$.

b) Es ist

$$d(A, U) = \inf\{d(a, U) \mid a \in A\} \overset{a)}{=} \inf\{\|\pi_{U^\perp}(a)\| \mid a \in A\} = d(\pi_{U^\perp}(A), 0).$$

c) Es ist

$$\begin{aligned} d(A, U) &= \inf\{\|v + w - u\| \mid w \in W, u \in U\} = \inf\{\|v - s\| \mid s \in W + U\} \\ &= d(v, U + W). \end{aligned}$$

Benutze nun Teil a) für den Untervektorraum $U + W$. ◯

Definition 11.3.7 (affiner Teilraum, Lot, Lotfußpunkte)

a) Wenn $W \leq V$ Vektorräume sind (beliebiger Körper K) und $v \in V$ ein Vektor, dann nennen wir $A := v + W$ einen **affinen Teilraum** von V.

In Wirklichkeit kennen wir solche affinen Räume schon längst: die nichtleeren Lösungsräume $\mathcal{L}(A, b)$ von linearen Gleichungssystemen sind affine Teilräume (siehe 4.1.2).

Für zwei Vektoren a, b in V heißt

$$\overline{a, b} := \{\lambda a + (1 - \lambda)b \mid \lambda \in K\} = a + K \cdot (b - a)$$

die **affine Gerade** durch a und b.

b) Wenn speziell $K = \mathbb{R}$ gilt, so heißt für a, b im reellen Vektorraum V die Menge

$$[a, b] := \{\lambda a + (1 - \lambda)b \mid 0 \leq \lambda \leq 1\}$$

die **Strecke** zwischen a und b.

c) Wir kehren kurz zur Situation von 11.3.6 c) zurück. Wenn U und W zwei endlichdimensionale Untervektorräume des euklidischen Raumes V, $v \in V$ ein beliebiger Vektor und $A := v + W$ sind, dann gibt es $u \in U$ und $w \in W$, sodass $v - u - w$ auf $U + W$ senkrecht steht. Dann heißt die Strecke $[u, v - w]$ ein **Lot** zwischen U und A, und die Punkte $u \in U$ und $v - w \in A$ heißen seine **Lotfußpunkte.**

Das Lot ist genau dann eindeutig bestimmt, wenn $U \cap W = \{0\}$. Ansonsten kann man es um einen beliebigen Vektor aus $U \cap W$ verschieben.

11.4 Übertragung ins Komplexe

Bemerkung 11.4.1 (Problemstellung und komplexe Konjugation)

Die Positivität des Skalarproduktes ist eine wichtige Eigenschaft, die man sich erhalten sollte, wenn man von reellen zu komplexen Vektorräumen übergeht. Auch hier soll in erster Linie eine Abstandsfunktion mit dem Skalarprodukt einhergehen. Insbesondere wollen wir keine Vektoren $\neq 0$, die „auf sich selbst senkrecht stehen“. Können wir das im Komplexen erreichen?

Um das zu entscheiden, betrachten wir einen $\mathbb{C}-$Vektorraum V mit $\dim(V) \geq 2$ und eine Bilinearform $\langle \cdot, \cdot \rangle : V \times V \longrightarrow \mathbb{C}$. Wir wollen uns überlegen, dass es ein $x \in V$ gibt mit $x \neq 0$ und $\langle x, x \rangle = 0$.

Dazu seien $v, w \in V$ linear unabhängig. Wenn $\langle v, v \rangle = 0$ gilt, dann sind wir schon fertig. Ansonsten ist

$$\mathbb{C} \ni \lambda \mapsto \langle \lambda v + w, \lambda v + w \rangle = \langle v, v \rangle \lambda^2 + (\langle v, w \rangle + \langle w, v \rangle)\lambda + \langle w, w \rangle \in \mathbb{C}$$

eine polynomiale Abbildung vom Grad 2. Nach Bemerkung 3.2.7 gibt es eine komplexe Nullstelle dieses Polynoms, also ein $\lambda \in \mathbb{C}$, sodass $x = \lambda v + w$ die Gleichung $\langle x, x \rangle = 0$ löst. Da v, w linear unabhängig sind, ist $x \neq 0$.

Wir müssen also irgendwelche Eigenschaften des Skalarproduktes auf den Prüfstand stellen, wenn wir im Komplexen etwas Vergleichbares haben wollen.

Man könnte zunächst einmal $\mathbb{C}$ selbst untersuchen, das ja ein zweidimensionaler $\mathbb{R}$-Vektorraum ist. Wenn wir 1 und i als reelle Basis von $\mathbb{C}$ wählen, so können wir $\mathbb{C}$ mit $\mathbb{R}^2$ „identifizieren“. Da haben wir aber das Standardskalarprodukt mit zugehöriger Norm

$$\|x + y\mathrm{i}\|^2 = x^2 + y^2 = (x + y\mathrm{i}) \cdot (x - y\mathrm{i}).$$

Wir definieren die **komplexe Konjugation** $\overline{\ }: \mathbb{C} \longrightarrow \mathbb{C}$ durch

$$\mathbb{C} \ni z = x + y\mathrm{i} \mapsto \overline{z} := x - y\mathrm{i} \in \mathbb{C}.$$

Man rechnet leicht nach, dass diese Abbildung ein $\mathbb{R}$-linearer Automorphismus des Körpers der komplexen Zahlen ist. Es gilt $\overline{z} = z$ genau dann, wenn z reell ist. Der **Betrag** der komplexen Zahl $z = x + y\mathrm{i}$ mit $x, y \in \mathbb{R}$ ist definiert als

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

Der Prüfstand der Geschichte sagt, dass die folgende Definition eine sinnvolle Übertragung des reellen Begriffes des Skalarproduktes liefert.

Definition 11.4.2 (komplexes Skalarprodukt, unitäre Vektorräume)

Es sei V ein komplexer Vektorraum. Eine Abbildung $\langle \cdot, \cdot \rangle : V \times V \longrightarrow \mathbb{C}$ heißt ein **komplexes Skalarprodukt**, wenn folgende Bedingungen erfüllt sind:

- $\forall v_1, v_2, w \in V,\ a \in \mathbb{C}:\ \langle a v_1 + v_2, w \rangle = a \langle v_1, w \rangle + \langle v_2, w \rangle,$

  $\forall v_1, v_2, w \in V,\ a \in \mathbb{C}:\ \langle w, a v_1 + v_2 \rangle = \overline{a} \langle w, v_1 \rangle + \langle w, v_2 \rangle.$ **(Sesquilinearität)**
- $\forall v, w \in V : \langle v, w \rangle = \overline{\langle w, v \rangle}.$ **(Hermitezität)**
- $\forall v \in V \setminus \{0\} : \langle v, v \rangle > 0.$ **(Positivität)**

Ein komplexer Vektorraum mit einem festen komplexen Skalarprodukt heißt auch ein **unitärer Vektorraum**.

Die Hermitezität (benannt nach **Charles Hermite**) ersetzt die Symmetrie der reellen Skalarprodukte. Sie erzwingt, dass $\langle v, v \rangle = \overline{\langle v, v \rangle} \in \mathbb{R}$, sodass die Forderung nach Positivität sinnvoll ist.

Die Sesquilinearität (von lat. sesqui = eineinhalb) ersetzt die Bilinearität aus der Definition reeller Skalarprodukte. Wenn man in der Definition $\mathbb{C}$ durch $\mathbb{R}$ ersetzt, so bekommt man die ursprüngliche Definition 11.1.2 wieder, denn die komplexe Konjugation ist auf $\mathbb{R}$ die Identität.

Vorsicht: In manchen Lehrbüchern werden bei der Definition der Sesquilinearität die Rollen von linkem und rechtem Argument vertauscht. Dies muss man dann beim Vergleich unserer Formeln mit diesen Büchern berücksichtigen.

Beispiel 11.4.3 (unitärer Standardraum)

Für eine komplexe Matrix $A = (a_{ij}) \in \mathbb{C}^{p\times q}$ bezeichnen wir mit $\overline{A} := (\overline{a_{ij}})$ die Matrix mit den komplex konjugierten Einträgen. Speziell verwenden wir diese Notation natürlich auch für Spalten- und Zeilenvektoren.

Vorsicht: Bei einem beliebigen komplexen Vektorraum lässt sich die komplexe Konjugation nicht ohne Weiteres einführen. Wie immer bei den Standardräumen ist es auch hier die Existenz einer Standardbasis, die den Unterschied macht.

Der n-dimensionale **unitäre Standardraum** ist $\mathbb{C}^n$ mit der Abbildung

$$\mathbb{C}^n \times \mathbb{C}^n \ni (v, w) \mapsto \langle v, w\rangle := v^\top \cdot \overline{w} = \sum_{i=1}^{n} v_i \cdot \overline{w_i}.$$

Man rechnet leicht nach, dass dies ein komplexes Skalarprodukt ist.

Nun will man lieb gewonnene Eigenschaften des reellen Skalarproduktes im Komplexen wiederfinden.

Hilfssatz 11.4.4 (Ungleichung von Cauchy-Schwarz im Komplexen)

Es sei V ein unitärer Vektorraum mit Skalarprodukt $\langle\cdot,\cdot\rangle$. Dann gilt

$$\forall v, w \in V : \ |\langle v, w\rangle|^2 \leq \langle v, v\rangle \cdot \langle w, w\rangle.$$

Beweis. Wenn man w um einen komplexen Faktor ζ vom Betrag 1 abändert, so bleiben die rechte und die linke Seite der Ungleichung unverändert: Links kann man das ζ aus dem Skalarprodukt als $\overline{\zeta}$ herausziehen, aber das hat immer noch Betrag 1, und rechts gilt

$$\langle \zeta w, \zeta w\rangle = \zeta\overline{\zeta}\langle w, w\rangle = \langle w, w\rangle.$$

Wir suchen nun eine komplexe Zahl ζ mit $|\zeta| = 1$, sodass $\langle v, \zeta w\rangle$ reell ist.

Dies ist kein Problem, wenn $\langle v, w\rangle = 0$ gilt, denn das ist ja schon reell. Wenn aber $\langle v, w\rangle \neq 0$ gilt, so setzen wir

$$\zeta := \frac{\langle v, w\rangle}{|\langle v, w\rangle|}$$

um unsere Suche zu beenden.

Wir dürfen also nach der eingangs gemachten Überlegung ohne Einschränkung annehmen, dass $\langle v, w\rangle$ reell ist.

Nun gehen wir genauso vor wie im Reellen. Wenn $v = 0$ oder $w = 0$ gilt ist die Behauptung klar, ansonsten betrachten wir die Funktion

$$f : \mathbb{R} \longrightarrow \mathbb{R}, \quad f(x) = \langle xv + w, xv + w\rangle = \langle v, v\rangle x^2 + 2\langle v, w\rangle x + \langle w, w\rangle.$$

Dieses quadratische Polynom darf höchstens eine Nullstelle haben, weshalb

$$\langle v, w\rangle^2 \leq \langle v, v\rangle \cdot \langle w, w\rangle$$

folgt. Gleichheit gilt genau dann, wenn v, w linear abhängig sind. ◯

Bemerkung 11.4.5 (jetzt geht fast alles durch)

Nun definieren wir wie im Reellen für einen unitären Vektorraum V die **Norm** eines Vektors v als $\|v\| := \sqrt{\langle v, v\rangle}$, und den Abstand zweier Vektoren als $d(v, w) := \|v - w\|$. Es gilt wegen der Ungleichung von Cauchy-Schwarz wieder die Dreiecksungleichung

$$\forall u, v, w \in V : d(u, w) \leq d(u, v) + d(v, w).$$

Wie im Reellen definiert man Orthogonalsysteme und Orthonormalsysteme, und beweist die Richtigkeit des Schmidt'schen Orthogonalisierungsverfahrens 11.2.6. Auch Orthogonalbasen und Orthonormalbasen gibt es wieder, orthogonale Komplementärräume zu endlichdimensionalen Untervektorräumen, orthogonale Projektionen, Abstände und so weiter.

Es gibt zwei Punkte, wo man aufpassen muss. Zum Einen gilt der Satz von Pythagoras nur noch in einer Richtung:

$$v \perp w \Rightarrow \|v + w\|^2 = \|v\|^2 + \|w\|^2.$$

Zum Anderen ändert sich etwas beim Umgang mit der Fundamentalmatrix. Was? Das untersuchen wir jetzt.

Bemerkung 11.4.6 (Fundamentalmatrix für unitäre Räume)

Es sei V ein n-dimensionaler unitärer Vektorraum mit Basis $B = \{b_1, \ldots, b_n\}$. Dann ist die **Fundamentalmatrix** des Skalarproduktes bezüglich B definiert als

$$D_{BB}(\langle \cdot, \cdot\rangle) := (\langle b_i, b_j\rangle)_{1 \leq i,j \leq n}.$$

Diese Matrix ist im Allgemeinen nicht mehr symmetrisch, sondern nur noch **hermitesch**. Dabei heißt eine komplexe quadratische Matrix F hermitesch, wenn gilt:

$$F^\top = \overline{F}.$$

Diese Eigenschaft spiegelt gerade die Hermitezität des Skalarproduktes wider. Es gilt für $v, w \in V$:

$$\langle v, w\rangle = D_B(v)^\top \cdot D_{BB}(\langle\cdot,\cdot\rangle) \cdot \overline{D_B(w)}.$$

Bemerkung 11.4.7 (Unitäre Matrizen, Iwasawa-Zerlegung)

Im unitären Standardraum $\mathbb{C}^n$ ist eine Basis $B = \{b_1, \ldots, b_n\}$ genau dann eine Orthonormalbasis, wenn für die Matrix A mit Spalten $b_1, \ldots, b_n$ gilt:

$$A^\top \cdot \overline{A} = I_n.$$

Die Matrix auf der linken Seite ist ja die Fundamentalmatrix des Standardskalarproduktes bezüglich B.

Wie in Definition 11.2.5 führt uns dies zur Definition einer Gruppe. Die **unitäre Gruppe** (Untergruppe von $\mathrm{GL}_n(\mathbb{C})$) ist die Menge

$$\mathrm{U}(n) := \{A \in \mathbb{C}^{n\times n} \mid A^\top \cdot \overline{A} = I_n\} = \{A \in \mathrm{GL}_n(\mathbb{C}) \mid A^{-1} = \overline{A^\top}\}.$$

Wie im Reellen rechnet man nach, dass dies eine Gruppe ist, ihre Elemente heißen **unitäre Matrizen**. Wenn wir nun noch die Gruppe

$$\mathcal{B}(n)_{\mathbb{C}}$$

der oberen komplexen Dreicksmatrizen mit positiven reellen Zahlen als Diagonaleinträgen hinzunehmen, dann gilt (wieder wegen E. Schmidt)

$$\mathrm{GL}_n(\mathbb{C}) = \mathrm{U}(n) \cdot \mathcal{B}(n)_{\mathbb{C}}.$$

Diese Identität nennt man wieder die **Iwasawa**-Zerlegung.

Die Determinante liefert einen Gruppenhomomorphismus

$$\det : \mathrm{U}(n) \longrightarrow \mathbb{C}^\times.$$

Der Kern dieser Determinantenabbildung ist die Gruppe

$$\mathrm{SU}(n) := \{A \in \mathrm{U}(n) \mid \det(A) = 1\}$$

der **speziellen unitären Matrizen**.

Zum Beispiel rechnet man nach, dass

$$\mathrm{SU}(2) = \{\begin{pmatrix} z & -\overline{w} \\ w & \overline{z} \end{pmatrix} \mid w, z \in \mathbb{C}, |w|^2 + |z|^2 = 1\}.$$

Dies ist eine Untergruppe der Einheitengruppe $\mathbb{H}^\times$ von Hamiltons Quaternionenalgebra (siehe 10.4.8).

Als Teilmenge von $\mathbb{C}^2 \cong \mathbb{R}^4$ ist es die dreidimensionale Einheitssphäre. Diese Gruppe ist eng mit $\mathrm{SO}(3)$ verwandt, was sich zum Beispiel in der Existenz von halbzahligem Spin niederschlägt... aber das ist eine ganz andere Geschichte.

Kapitel 12

Skalarprodukte und Homomorphismen

12.1 Isometrien

Wir definieren zunächst ganz allgemein, was eine Isometrie zwischen zwei metrischen Räumen ist.

Definition 12.1.1 (Isometrie, Isometriegruppe)

Es seien zwei metrische Räume (X, d) und (Y, e) gegeben (siehe 11.1.7). Dann heißt eine Abbildung $\Phi : X \longrightarrow Y$ eine **Isometrie** (oder auch **abstandserhaltende Abbildung**), wenn gilt:

$$\forall x_1, x_2 \in X : d(x_1, x_2) = e(\Phi(x_1), \Phi(x_2)).$$

Solch eine Abbildung ist immer injektiv, denn aus $\Phi(x_1) = \Phi(x_2)$ folgt ja $d(x_1, x_2) = 0$, also $x_1 = x_2$. Die Surjektivität muss man im Allgemeinen fordern, wenn man sie will.

Wir bezeichnen mit $\mathrm{Iso}(X, d)$ die Menge aller invertierbaren Isometrien von X nach X. Das ist also eine Teilmenge der symmetrischen Gruppe (siehe 2.1.5 c)) von X.

$\mathrm{Iso}(X, d)$ ist nicht leer (finden Sie ein Element darin!) und bezüglich Komposition und Inversenbildung abgeschlossen. Es gilt nämlich:

Für alle $\Phi, \Psi \in \mathrm{Iso}(X, d)$ und für alle $x_1, x_2 \in X$ gilt

$$d(x_1, x_2) = d(\Phi(x_1), \Phi(x_2)) = d(\Psi \circ \Phi(x_1), \Psi \circ \Phi(x_2)),$$

also $\Psi \circ \Phi \in \mathrm{Iso}(X, d)$.

Für alle $\Phi \in \mathrm{Iso}(X,d)$ und für alle $x_1, x_2 \in X$ gilt

$$d(x_1,x_2) = d(\Phi(\Phi^{-1}(x_1)), \Phi(\Phi^{-1}(x_2))) = d(\Phi^{-1}(x_1), \Phi^{-1}(x_2)),$$

also $\Phi^{-1} \in \mathrm{Iso}(X,d)$.

Damit ist $\mathrm{Iso}(X,d)$ eine Untergruppe der symmetrischen Gruppe $\mathrm{Sym}(X)$. Sie heißt die **Isometriegruppe** des metrischen Raumes (X,d). Oft schreibt man nur $\mathrm{Iso}(X)$, wenn die verwendete Metrik klar ist.

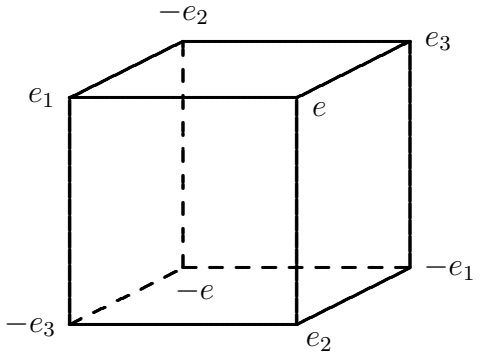
Beispiel 12.1.2 (für Isometrien)

a) Es seien die Ebene $\mathbb{R}^2$ und der Raum $\mathbb{R}^3$ mit dem Standardskalarproukt versehen (und damit zu metrischen Räumen gemacht). Dann ist die Abbildung

$$\Phi : \mathbb{R}^2 \ni \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} \in \mathbb{R}^3$$

eine Isometrie.

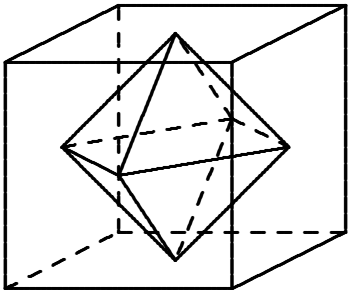
b) Es sei $X = \{\pm 1\}^3 \in \mathbb{R}^3$ die Menge der Ecken eines (dreidimensionalen, physikalischen und doch recht idealen) Würfels.



Jedes Element e aus X hat drei Nachbarn e_1, e_2, e_3, die von e Abstand 2 haben und untereinander Abstand $2\sqrt{2}$. (Das sind jetzt ausnahmsweise nicht die Standardbasisvektoren...) Dann gibt es die drei Ecken $-e_1, -e_2, -e_3$, die von e und untereinander Abstand $2\sqrt{2}$ haben (diese drei Ecken bilden mit e zusammen ein regelmäßiges Tetraeder), und schließlich gibt es die Ecke $-e$, die von e den Abstand $2\sqrt{3}$ hat. Wir merken uns die Ecke e und ihre Nachbarn e_1, e_2, e_3.

Ein Isometrie von X gewinnt man nun so: wir wählen eine Ecke f und ihre Nachbarn f_1, f_2, f_3. Dann gibt es genau eine Möglichkeit, die Abbildung $e \mapsto f$, $e_i \mapsto f_i$ zu einer Isometrie von X fortzusetzen. Jede dieser Isometrien wird durch Drehungen des Würfels um seinen Mittelpunkt, durch Spiegelungen des Würfels an (geeigneten) Ebenen durch den Mittelpunkt oder durch $-\mathrm{Id_X}$ gegeben; alles sind Einschränkungen von linearen Automorphismen des $\mathbb{R}^3$ nach X. Wir erhalten also wegen der 8 Wahlmöglichkeiten für f und wegen der jeweils 6

Wahlmöglichkeiten für die Reihenfolge der f_i eine Isometriegruppe mit 48 Elementen. Diese Gruppe heißt die **Oktaedergruppe**. Das kommt daher, dass jede Isometrie des Würfels auch eine Isometrie des regelmäßigen Oktaeders liefert, der die Flächenmittelpunkte des Würfels als Ecken hat.



c) Es sei V ein euklidischer oder unitärer Vektorraum und $v \in V$ ein beliebiger Vektor. Dann ist die Abbildung

$$\tau_v : V \longrightarrow V, \quad \tau_v(x) := x + v,$$

eine Isometrie. Denn:

$$\forall x, y \in V : d(\tau_v(x), \tau_v(y)) = \|(x + v) - (y + v)\| = \|x - y\| = d(x, y).$$

Diese Isometrie nennt man die **Translation** (oder auch **Verschiebung**) um v.

Ab jetzt werden wir nur noch Isometrien zwischen euklidischen oder unitären Vektorräumen betrachten.

Definition 12.1.3 (lineare Isometrie, Polarisierung)

Es seien V und W zwei euklidische oder zwei unitäre Vektorräume. Dann heißt eine Isometrie $\Phi : V \longrightarrow W$, die gleichzeitig eine lineare Abbildung ist, eine **lineare Isometrie**.

Das Skalarprodukt lässt sich aus der Metrik rekonstruieren. Genauer gilt im Reellen die **Polarisierungsformel**:

$$\langle x, y\rangle = \frac{1}{2}\left[\langle x + y, x + y\rangle - \langle x, x\rangle - \langle y, y\rangle\right] = \frac{1}{4}\left[\langle x + y, x + y\rangle - \langle x - y, x - y\rangle\right].$$

Im Komplexen sieht die richtige Formel so aus:

$$\langle x, y\rangle = \frac{1}{4}\left[\langle x + y, x + y\rangle - \langle x - y, x - y\rangle + \mathrm{i}\langle x + \mathrm{i}y, x + \mathrm{i}y\rangle - \mathrm{i}\langle x - \mathrm{i}y, x - \mathrm{i}y\rangle\right].$$

Da eine lineare Isometrie insbesondere einen Vektor der Norm a auf einen Vektor der Norm a abbildet, sagt diese Polarisierungsformel für einen Vektorraumhomomorphismus $\Phi \in \mathrm{Hom}(V, W)$:

$$\boxed{\Phi \text{ ist lineare Isometrie } \iff \forall x, y \in V : \langle x, y\rangle_V = \langle \Phi(x), \Phi(y)\rangle_W.}$$

Dabei haben wir den für das jeweilige Skalarprodukt zuständigen Vektorraum durch einen Index am Skalarprodukt gekennzeichnet.

Beispiel 12.1.4 (Koordinatenabbildung, $\mathbb{R}^2$, Drehkästchen)

a) Es sei V ein endlichdimensionaler euklidischer Vektorraum mit einer Orthonormalbasis B. In Beobachtung 11.2.4 hatten wir gesehen, dass die Koordinatenabbildung $D_B : V \longrightarrow \mathbb{R}^{\dim(V)}$ das Skalarprodukt von V in das Skalarprodukt auf dem euklidischen Standardraum übersetzt. Also ist D_B eine lineare Isometrie zwischen diesen Räumen.

b) Wir wollen untersuchen, wie lineare Isometrien der euklidischen Ebene ($\mathbb{R}^2$ mit Standardskalarprodukt) in sich selbst aussehen. Wir beschreiben eine lineare Isometrie $\Phi \in \mathrm{Aut}(\mathbb{R}^2)$ durch ihre Abbildungsmatrix bezüglich der Standardbasis S, die ja eine Orthonormalbasis ist:

$$D_{SS}(\Phi) = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Da Φ eine lineare Isometrie ist und in den Spalten der Matrix die Bilder der Standardbasis stehen, gilt

$$a^2 + c^2 = 1 = b^2 + d^2, \quad ab + cd = 0.$$

Die Analysis sagt uns, dass es einen Winkel $\varphi \in [0, 2\pi]$ gibt, sodass $a = \cos\varphi$ und $c = \sin\varphi$. Aus der Normierung und Orthogonalität folgt dann aber auch, dass

$$\begin{pmatrix} b \\ d \end{pmatrix} = \pm \begin{pmatrix} -\sin\varphi \\ \cos\varphi \end{pmatrix}.$$

Das heißt:

$$D_{SS}(\Phi) = D_\varphi := \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \quad \text{oder} \quad D_{SS}(\Phi) = \begin{pmatrix} \cos\varphi & \sin\varphi \\ \sin\varphi & -\cos\varphi \end{pmatrix}.$$

Für beliebiges φ ist umgekehrt durch diese Abbildungsmatrix eine Isometrie von $\mathbb{R}^2$ (auf sich selbst) gegeben.

Die linke Matrix D_φ heißt das **Drehkästchen zum Winkel** φ. Die zugehörige lineare Abbildung ist die Drehung um diesen Winkel. Im Spezialfall $\varphi = \pi/2$ hatten wir diese Matrix schon einmal behandelt in Beispiel 8.3.3 c). Die zweite Matrix hatten wir in Beispiel 6.3.2 schon als Spiegelung kennen gelernt.

Beispiel 12.1.5 (Spiegelung)

Es seien V ein euklidischer oder unitärer Vektorraum und $v \in V$ ein beliebiges Element $\neq 0$. Dann ist die Abbildung

$$\sigma_v : V \longrightarrow V, \quad \sigma_v(x) := x - 2\frac{\langle x, v\rangle}{\langle v, v\rangle} v,$$

eine lineare Isometrie von V (auf sich selbst). Die Linearität ist klar, denn das Skalarprodukt ist (auch im unitären Fall) im ersten Argument linear, v ist ja fest. Und für das Skalarprodukt gilt für alle $x, y \in V$:

$$\begin{aligned}\langle \sigma_v(x), \sigma_v(y)\rangle &= \langle x - 2\tfrac{\langle x,v\rangle}{\langle v,v\rangle} v, y - 2\tfrac{\langle y,v\rangle}{\langle v,v\rangle} v\rangle \\ &= \langle x, y\rangle + \langle x, -2\tfrac{\langle y,v\rangle}{\langle v,v\rangle} v\rangle - \langle 2\tfrac{\langle x,v\rangle}{\langle v,v\rangle} v, y\rangle + \langle -2\tfrac{\langle x,v\rangle}{\langle v,v\rangle} v, -2\tfrac{\langle y,v\rangle}{\langle v,v\rangle} v\rangle \\ &= \langle x, y\rangle - 2\tfrac{\langle v,y\rangle}{\langle v,v\rangle}\langle x, v\rangle - 2\tfrac{\langle x,v\rangle}{\langle v,v\rangle}\langle v, y\rangle + 4\tfrac{\langle x,v\rangle}{\langle v,v\rangle}\tfrac{\langle v,y\rangle}{\langle v,v\rangle}\langle v, v\rangle \\ &= \langle x, y\rangle.\end{aligned}$$

Was macht diese Abbildung? Wenn x orthogonal zu v ist, so ist $\sigma_v(x) = x$. Also ist σ_v auf dem orthogonalen Komplement des von v erzeugten Untervektorraums die Identität. Auf der von v erzeugten Geraden ist σ_v die Multiplikation mit -1. Insgesamt tut σ_v das, was man von einer Spiegelung erwartet; σ_v heißt deswegen die **Spiegelung** an der Hyperebene $v^\perp$. Offensichtlich gilt $\sigma_v^2 = \mathrm{Id}_V$.

Nun sei V euklidisch. Wenn dann zum Beispiel $v, w \in V$ zwei Vektoren derselben Länge sind, so gilt

$$\langle v - w, v + w\rangle = 0.$$

Wir bezeichnen mit $d := v - w$ die Differenz zwischen v und w. Dann gilt

$$\sigma_d(v) = \sigma_d(\frac{1}{2}(v + w) + \frac{1}{2}(v - w)) = \frac{1}{2}(v + w) - \frac{1}{2}(v - w) = w.$$

Genauso gilt auch $\sigma_d(w) = v$.

Also: für je zwei Vektoren derselben Länge im euklidischen Vektorraum V gibt es (mindestens) eine lineare Isometrie, die die beiden vertauscht.

Hilfssatz 12.1.6 (Kriterium für lineare Isometrien)

Es sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ und V, W seien $\mathbb{K}$-Vektorräume mit Skalarprodukt. Weiter sei $\Phi : V \longrightarrow W$ eine lineare Abbildung. Dann gilt:

Φ ist eine Isometrie genau dann, wenn jedes Orthonormalsystem aus V injektiv auf ein Orthonormalsystem in W abgebildet wird.

Wenn V endlichdimensional ist, dann ist Φ eine Isometrie genau dann, wenn (mindestens) eine Orthonormalbasis injektiv auf ein Orthonormalsystem abgebildet wird.

Beweis. Da eine lineare Isometrie Skalarprodukte erhält, ist klar, dass sie ein Orthonormalsystem injektiv auf ein ebensolches abbilden muss.

Wird umgekehrt jedes Orthonormalsystem injektiv auf ein Orthonormalsystem abgebildet und sind $u, v \in V$ beliebig, so wählen wir mit E. Schmidt eine Orthonormalbasis des von u und v erzeugten Untervektorraumes. Diese wird von Φ auf ein Orthonormalsystem gleicher Mächtigkeit in W abgebildet, und man rechnet mit seiner Hilfe nach, dass $\langle u, v\rangle = \langle \Phi(u), \Phi(v)\rangle$.

Im Endlichdimensionalen kann man alles durch eine Orthonormalbasis auf einmal erreichen, was man sonst lieber zusammenstückelt. Deswegen gilt auch die letzte Aussage der Beobachtung. ◯

Beispiel 12.1.7 (Endomorphismen im Endlichdimensionalen)

Es sei V ein endlichdimensionaler $\mathbb{K}$-Vektorraum ($\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$) mit Skalarprodukt. Φ sei ein Endomorphismus von V, und B sei eine Orthonormalbasis von V. Dann gilt:

$$\boxed{\Phi \text{ ist Isometrie } \iff \Phi(B) \text{ ist eine Orthonormalbasis von } V.}$$

Das bedeutet aber in der Sprache der Abbildungsmatrizen, deren Spalten ja die Koordinatenvektoren der Bilder sind und deshalb eine Orthonormalbasis im zugehörigen Koordinatenraum $\mathbb{K}^{\dim(V)}$ bilden:

$$\boxed{\Phi \text{ ist Isometrie } \iff D_{BB}(\Phi) \text{ ist orthogonale bzw. unitäre Matrix}}$$

(Siehe 11.2.5 und 11.4.7)

Diese Charakterisierung gilt **nur** für Orthonormalbasen B.

Daran sieht man, dass eine lineare Isometrie eines endlichdimensionalen euklidischen Vektorraums entweder Determinante 1 oder Determinante -1 hat. Beides kann vorkommen. Man nennt die linearen Isometrien mit Determinante 1 die **eigentlichen Bewegungen** des euklidischen Raumes.

Hilfssatz 12.1.8 (Wie häufig sind lineare Isometrien?)

Es sei V ein euklidischer Vektorraum und $\Phi \in \mathrm{Iso}(V)$ eine Isometrie. Dann gibt es ein Element $b \in V$ und eine lineare Isometrie Φ_0 von V, sodass gilt

$$\forall v \in V : \ \Phi(v) = \Phi_0(v) + b.$$

Beweis. Die einzige Möglichkeit für b ist $b := \Phi(0)$. Die Abbildung

$$\Phi_0 := \tau_{-b} \circ \Phi, \quad \Phi_0(v) := \Phi(v) - b$$

ist dann eine Isometrie, weil für alle $x, y \in V$ sicher $d(x,y) = d(x-b, y-b)$ gilt (siehe Beispiel 12.1.2 c)) und die Komposition von Isometrien wieder eine Isometrie ist. Wir haben jetzt die Linearität von Φ_0 zu zeigen.

Dazu überlegen wir uns zunächst, dass Φ_0 das Skalarprodukt erhält. Für $v, w \in V$ gilt ja wegen $\|v\| = d(0,v) = d(\Phi_0(0), \Phi_0(v)) = \|\Phi_0(v)\|$, dass

$$\begin{aligned}\langle v, w\rangle &= \tfrac{1}{2}(\|v\|^2 + \|w\|^2 - d(v,w)^2)\\ &= \tfrac{1}{2}(\|\Phi_0(v)\|^2 + \|\Phi_0(w)\|^2 - d(\Phi_0(v), \Phi_0(w))^2)\\ &= \langle \Phi_0(v), \Phi_0(w)\rangle.\end{aligned}$$

Insbesondere wird für $0 \neq v \in V$ das orthogonale Komplement zu $\mathbb{R}v$ auf das orthogonale Komplement zu $\mathbb{R}\Phi_0(v)$ abgebildet und damit auch die Gerade $\mathbb{R}v$ auf die Gerade $\mathbb{R}\Phi_0(v)$. Da die Länge der Vektoren erhalten bleibt und außerdem $d(\Phi(av), \Phi(v)) = d(av, v) = |a-1|\|v\| = |a-1|\|\Phi_0(v)\|$ gelten muss, folgt für alle $v \in V$ und $a \in \mathbb{R}$ die Gleichung

$$\Phi_0(av) = a\Phi_0(v).$$

Das zeigt übrigens nebenbei, dass eine Isometrie eines euklidischen Vektorraums eine Gerade zwangsläufig auf eine Gerade abbildet.

Außerdem gilt damit für linear abhängige $v, w \in V$ auch $\Phi_0(v+w) = \Phi_0(v) + \Phi_0(w)$. Wenn v und w linear unabhängig sind, so wählen wir eine Orthonormalbasis $B = \{b_1, b_2\}$ in der von v und w erzeugten Ebene. Φ_0 bildet B auf ein Orthonormalsystem in V ab, und da Orthogonalität erhalten bleibt, wird die von v und w erzeugte Ebene auf die von $\Phi_0(b_1)$ und $\Phi_0(b_2)$ erzeugte Ebene abgebildet. Da Φ_0 auch das Skalarprodukt erhält und wir die Koeffizienten von allen Vektoren bezüglich der Orthonormalsysteme B und $\Phi_0(B)$ durch die Fourierformel berechnen können, folgt insgesamt

$$\Phi_0(v+w) = \Phi_0(v) + \Phi_0(w).$$

Also ist Φ_0 tatsächlich linear. ◯

Bemerkung 12.1.9 a) Am Beweis fällt auf, dass wir nirgends die Bijektivität von Φ benutzen. Tatsächlich haben wir bewiesen, dass für zwei euklidische Vektorräume V und W jede Isometrie sich schreiben lässt als Komposition einer linearen Isometrie von V nach W mit einer Translation auf W.

b) Wenn $V = \mathbb{C}$ der unitäre eindimensionale Standardvektorraum ist, dann ist die komplexe Konjugation zwar eine Isometrie, die 0 auf sich selbst abbildet, aber sie ist nicht komplex-linear. Das zeigt, dass die Beobachtung in 12.1.8 im Allgemeinen wirklich nur für euklidische, nicht aber für unitäre Vektorräume gilt.

Bemerkung 12.1.10 (Spiegelungen gegen den Rest der Welt)

Es sei V ein endlichdimensionaler euklidischer Vektorraum und Φ eine lineare Isometrie von V auf sich selbst. Da $\dim(V) < \infty$ gilt und Φ linear und injektiv ist, ist es auch surjektiv. Es sei $\{b_1, \ldots, b_n\}$ eine Orthonormalbasis von V. Für $d := \Phi(b_1) - b_1$ gilt dann mit der Notation aus Beispiel 12.1.5 (wobei wir im Fall $d = 0$ noch $\sigma_d := \mathrm{Id}_V$ einführen):

$$\sigma_d(\Phi(b_1)) = b_1.$$

Also ist $\Phi_1 := \sigma_d \circ \Phi$ eine lineare Isometrie mit $\Phi_1(b_1) = b_1$. Dann setzen wir $e := b_2 - \Phi_1(b_2)$ und sehen (weil b_1 sowohl auf b_2 als auch auf $\Phi_1(b_2)$ senkrecht steht), dass $b_1 \perp e$. Wir setzen $\Phi_2 := \sigma_e \circ \Phi_1$ (wieder mit $\sigma_0 := \mathrm{Id}_V$) und sehen:

$$\Phi_2(b_1) = b_1 \quad \text{und} \quad \Phi_2(b_2) = b_2.$$

Nun macht man rekursiv so weiter und sieht am Ende: Es gibt endlich viele Spiegelungen $\sigma_1, \sigma_2, \ldots, \sigma_k$ derart, dass

$$\forall 1 \leq i \leq n : (\sigma_k \circ \sigma_{k-1} \circ \cdots \circ \sigma_1 \circ \Phi)(b_i) = b_i.$$

Demnach gilt:

$$\Phi = \sigma_1 \circ \sigma_2 \circ \cdots \circ \sigma_k.$$

Jede lineare Isometrie eines n-dimensionalen euklidischen Raums ist Produkt von höchstens n Hyperebenenspiegelungen.

Zum Beispiel ist eine Drehung in der Ebene ein Produkt von zwei Geradenspiegelungen.

Hilfssatz 12.1.11 (Isometrien und invariante Komplemente.)

Es sei $\Phi : V \longrightarrow V$ eine lineare Isometrie eines endlichdimensionalen euklidischen oder unitären Vektorraumes. Weiter sei $U \subseteq V$ ein Φ-invarianter Untervektorraum. Dann ist $U^\perp$ ein Φ-invarianter Komplementärraum zu U.

Beweis. Da Φ invertierbar ist, ist $\Phi(U) \subseteq U$ ein Untervektorraum von U mit $\dim(U) = \dim(\Phi(U))$, also gilt

$$U = \Phi(U).$$

Also lässt sich jedes Element $u \in U$ auch schreiben als $u = \Phi(\widetilde{u})$, $\widetilde{u} \in U$. Es folgt dann für ein Element $v \in U^\perp$:

$$\langle \Phi(v), u \rangle = \langle \Phi(v), \Phi(\widetilde{u}) \rangle = \langle v, \widetilde{u} \rangle = 0.$$

Da dies für alle $u \in U$ geht, folgt $\Phi(v) \in U^\perp$, und damit ist $U^\perp$ ein Φ-invarianter Komplementärraum zu U. ◯

Dies beseitigt bei Isometrien ein Problem, das uns bei der Frage nach der Diagonalisierbarkeit beliebiger Endomorphismen gestört hat (siehe 9.2.1). Nun wollen wir erst einmal sehen, welche Eigenwerte eine Isometrie überhaupt haben kann.

Hilfssatz 12.1.12 (Eigenwerte haben Betrag 1.)

Es seien $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ und V ein $\mathbb{K}$-Vektorraum mit Skalarprodukt.

a) Es sei Φ eine lineare Isometrie von V. Dann hat jeder Eigenwert von Φ Betrag 1.

b) Es sei $\alpha \in \mathbb{K}$ mit Betrag 1 und $V \neq \{0\}$. Dann gibt es eine Isometrie von V, die α als Eigenwert hat.

Beweis. a) Wenn $v \in V$ ein Eigenvektor zum Eigenwert α ist, dann gilt

$$\langle v, v\rangle = \langle \Phi(v), \Phi(v)\rangle = \langle \alpha v, \alpha v\rangle = |\alpha|^2 \langle v, v\rangle,$$

und wegen $v \neq 0$ folgt $|\alpha| = 1$.

b) Die Multiplikation mit α ist so eine Isometrie. ◯

Bemerkung 12.1.13 (Polynome über $\mathbb{R}$ und $\mathbb{C}$)

Wir hatten schon in 9.2.1 den Fundamentalsatz der Algebra erwähnt, der sagt:

Es sei $f \in \mathbb{C}[X]$ ein normiertes Polynom vom Grad $n \geq 0$. Dann gibt es komplexe Zahlen $\lambda_1, \ldots, \lambda_n$, sodass

$$f = (X - \lambda_1) \cdot \ldots \cdot (X - \lambda_n).$$

Diesen Satz beweisen wir in diesem Skript nicht, Beweise werden in aller Regel in Vorlesungen zur Funktionentheorie erbracht (mit dem Satz von Liouville) oder in der Algebra aus dem Zwischenwertsatz der reellen Analysis und Ergebnissen der Galois-Theorie hergeleitet.

Wenn nun $f \in \mathbb{R}[X]$ ein normiertes reelles Polynom vom Grad $n \geq 0$ ist, dann ist es natürlich auch ein komplexes Polynom mit denselben Eigenschaften, und zerfällt über $\mathbb{C}$ als Produkt von Linearfaktoren. Wenn aber $\lambda \in \mathbb{C}$ eine Nullstelle von f ist, dann ist auch das komplex konjugierte $\overline{\lambda}$ eine Nullstelle von f. Dies sind zwei verschiedene Nullstellen, wenn $\lambda \in \mathbb{C} \setminus \mathbb{R}$. Dann ist aber das quadratische Polynom

$$(X - \lambda) \cdot (X - \overline{\lambda}) = X^2 - 2\mathrm{Re}(\lambda)X + |\lambda|^2 \in \mathbb{R}[X]$$

ein Teiler von f. Wenn wir die Nullstellen λ_i so nummerieren, dass die ersten k davon reell und die letzten $n-k$ nicht reell sind, dann ist $n-k$ gerade. Wir setzen $l := (n-k)/2$ und können dann f schreiben als

$$f = (X-\lambda_1)\cdot\ldots\cdot(X-\lambda_k)\cdot(X^2-s_1X+n_1)\cdot\ldots\cdot(X^2-s_lX+n_l)$$

mit reellen Faktoren, wobei die quadratischen Polynome $X^2-s_iX+n_i$ keine reellen Nullstellen haben.

Diese Information wenden wir nun auf das charakteristische Polynom einer linearen Isometrie eines endlichdimensionalen $\mathbb{K}$-Vektorraumes an.

Satz 12.1.14 (Isometrienormalform)

Es seien $\mathbb{K}=\mathbb{R}$ oder $\mathbb{C}$ und V ein n-dimensionaler $\mathbb{K}-$ Vektorraum mit Skalarprodukt. Weiter sei $\Phi : V \longrightarrow V$ eine lineare Isometrie von V. Dann gilt:

a) Für $\mathbb{K}=\mathbb{C}$ besitzt V eine Orthonormalbasis aus Eigenvektoren, d.h. Φ ist orthogonal diagonalisierbar.

b) Für $\mathbb{K}=\mathbb{R}$ lässt sich V schreiben als direkte Summe von paarweise orthogonalen ein- oder zweidimensionalen Φ-invarianten Untervektorräumen. Auf den eindimensionalen Summanden ist Φ die Multiplikation mit 1 oder -1. Auf jedem der zweidimensionalen Summanden wird Φ bezüglich einer beliebigen Orthonormalbasis durch ein Drehkästchen D_φ (siehe 12.1.4) mit einem jeweils geeigneten Winkel φ beschrieben.

Beweis. Wir machen in beiden Fällen vollständige Induktion nach der Dimension n. Für $n=0$ oder 1 sind beide Fälle klar. Jetzt kommt der Induktionsschritt nach $n \geq 2$.

Das charakteristische Polynom $\mathrm{CP}_\Phi(X)$ hat mindestens eine komplexe Nullstelle λ_n. Wenn es eine reelle Nullstelle gibt, so sei $\lambda_n \in \mathbb{R}$.

a) Im Fall $\mathbb{K}=\mathbb{C}$ wählen wir einen Eigenvektor b_n der Länge 1 von Φ zum Eigenwert λ_n. Er spannt einen invarianten Untervektorraum $U=\mathbb{C}b_n$ auf, dessen orthogonaler Komplementärraum $U^\perp$ nach 12.1.11 ebenfalls Φ-invariant ist. Per Induktionsvoraussetzung besitzt der $(n-1)$-dimensionale Untervektorraum $U^\perp$ eine Orthonormalbasis aus Eigenvektoren zu Φ, und diese bilden zusammen mit b_n eine Orthonormalbasis von V, die aus Eigenvektoren von Φ besteht.

b) Im Fall $\mathbb{K}=\mathbb{R}$ gibt es noch einmal zwei Möglichkeiten. Wenn λ_n reell ist, so können wir weitermachen wie in Fall a), wobei wir noch an 12.1.12 erinnern: der Eigenwert ist ± 1.

Wenn es keinen reellen Eigenwert gibt, dann sind λ_n und $\overline{\lambda}_n$ komplexe Nullstellen von $\mathrm{CP}_\Phi(X)$ und $q(X) := (X-\lambda_n)\cdot(X-\overline{\lambda}_n)$ teilt das charakteristische Polynom von Φ. Es sei A eine Abbildungsmatrix von Φ (bezüglich irgendeiner

Basis). Dann ist $q(A)$ eine Abbildungsmatrix von $q(\Phi)$. Da aber $A - \lambda_n I_n$ nicht invertierbar ist (λ_n ist ja ein komplexer Eigenwert der Matrix $A \in \mathbb{R}^{n\times n} \subseteq \mathbb{C}^{n\times n}$) ist auch $q(\Phi)$ nicht invertierbar. Wir wählen einen Vektor $v \neq 0$ im Kern von $q(\Phi)$. Dann sind v und $\Phi(v)$ linear unabhängig (sonst wäre v ja ein Eigenvektor, den es aber nicht gibt), allerdings ist

$$\Phi^2(v) = 2\mathrm{Re}(\lambda_n)\Phi(v) - |\lambda_n|^2 \cdot v,$$

und damit spannen v und $\Phi(v)$ einen zweidimensionalen Φ-invarianten Unterraum U auf. Das orthogonale Komplement dazu ist ein $(n-2)$-dimensionaler Φ-invarianter Unterraum, und für diesen können wir die Induktionsvoraussetzung verwenden (wir hatten den Induktionsanfang ja für zwei aufeinanderfolgende Dimensionen gemacht). Eine Isometrie eines zweidimensionalen euklidischen Raums, die keinen reellen Eigenwerte hat, wird zwangsläufig durch ein Drehkästchen beschrieben, siehe Beispiel 12.1.4 b).

Damit sind wir fertig. ◯

Bemerkung 12.1.15 Im Beweis von b) ist das Polynom $q(X)$ im Nachhinein das charakteristische Polynom des Drehkästchens D_φ. Insbesondere ist $\mathrm{Re}(\lambda_n) = \cos\varphi$ und $|\lambda_n|^2 = \cos^2\varphi + \sin^2\varphi = 1$. Damit sind die Teiler von Grad 2 des charakteristischen Polynoms einer euklidischen Isometrie sehr eingeschränkt. In kleinen Dimensionen hilft das teilweise, um individuelle Isometrien schneller zu verstehen – manche Rechnung wird einfacher.

Folgerung 12.1.16 (Matrizenwelt; Isometrienormalform)

a) Es sei $A \in \mathrm{U}(n)$ eine unitäre Matrix. Dann gibt es eine unitäre Matrix $S \in \mathrm{U}(n)$, sodass $S^{-1}AS$ eine Diagonalmatrix ist.

b) Es sei $A \in \mathrm{O}(n)$ eine orthogonale Matrix. Es sei $d_+ := \dim \mathrm{Eig}(A, 1)$ und $d_- := \dim \mathrm{Eig}(A, -1)$ und $l = \frac{1}{2}(n - d_+ - d_-)$. Dann gibt es reelle Zahlen $\varphi_1, \ldots, \varphi_l \in (0, \pi)$ und eine orthogonale Matrix $S \in \mathrm{O}(n)$, sodass $S^{-1}AS$ die folgende Block-Diagonalgestalt hat:

$$\begin{pmatrix} I_{d_+} & & & & \\ & -I_{d_-} & & & \\ & & D_{\varphi_1} & & \\ & & & \ddots & \\ & & & & D_{\varphi_l} \end{pmatrix}.$$

Dabei ist wie immer die Matrix D_{φ_i} das Drehkästchen

$$D_{\varphi_i} = \begin{pmatrix} \cos\varphi_i & -\sin\varphi_i \\ \sin\varphi_i & \cos\varphi_i \end{pmatrix}.$$

Denn: In beiden Fällen beschreibt A eine Isometrie des Standardraumes bezüglich der Standardbasis. Satz 12.1.14 beschreibt eine Abbildungsmatrix dieser Isometrie nach einem Basiswechsel zu einer anderen Orthonormalbasis. Im unitären Fall wird dieser Basiswechsel mit einer unitären Matrix durchgeführt (wegen 11.4.7), und im euklidischen Fall brauchen wir dazu eine orthogonale Matrix (wegen 11.2.5). Nur die Einschränkung an die Winkel, zwischen 0 und π zu liegen, bedarf noch einer kurzen Erläuterung. Wenn wir einen Winkel zwischen π und 2π bräuchten, so könnten wir die zwei Basisvektoren vertauschen und benutzen dann

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^{-1} \cdot D_\varphi \cdot \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = D_{2\pi-\varphi}.$$

NB: Da V im euklidischen Fall die direkte Summe der Eigenräume zu 1 und -1 und der zweidimensionalen Unterräume aus Satz 12.1.14 ist, ist tatsächlich $l \in \mathbb{N}_0$. ◯

Beispiel 12.1.17 (so richtig mit Zahlen...)

a) Zum Aufwärmen eine ganz leichte Matrix

$$A := \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \in \mathrm{O}(3).$$

Nach Beispiel 7.3.4 ist das charakteristische Polynom dieser Matrix das Polynom

$$\mathrm{CP}_A(X) = X^3 - 1 = (X-1) \cdot (X^2 + X + 1).$$

Der quadratische Faktor hat keine reelle Nullstelle (die zwei komplexen Nullstellen sind $-\frac{1}{2} \pm \frac{\sqrt{3}}{2}\mathrm{i} = \cos\varphi \pm \mathrm{i} \cdot \sin\varphi$ für $\varphi = 2\pi/3$). Der Eigenraum zum Eigenwert 1 ist

$$\mathrm{Eig}(A, 1) = \mathbb{R} \cdot b_1, \text{ mit } b_1 := \frac{1}{\sqrt{3}}(1\ 1\ 1)^\top.$$

Dazu orthogonal ist der Kern von $A^2 + A + I_3$. In diesem Raum wählen wir irgendeinen Vektor, zum Beispiel

$$v := \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$$

Sein Bild unter Multiplikation mit A ist

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

Mit E. Schmidt werden diese beiden Vektoren orthonormalisiert zu

$$b_2 := \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \quad \text{und} \quad b_3 := \sqrt{\frac{2}{3}} \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -1 \end{pmatrix}.$$

Nun sei S die Matrix mit Spalten b_1, b_2, b_3. Da dies eine Orthonormalbasis ist, ist S in $\mathrm{O}(3)$. Es gilt mit $\varphi = 2\pi/3$:

$$S^{-1}AS = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix}.$$

Wenn man sich mit dem Winkel nicht ganz sicher ist, sollte man noch einmal $A \cdot b_2 = \frac{1}{2}b_2 + \frac{\sqrt{3}}{2}b_3$ nachrechnen; das $+$-Zeichen vor dem zweiten Summanden sagt, dass der Sinus des Drehwinkels zwischen 0 und π liegt und damit „richtig“ gewählt ist.

b) Es sei $A \in \mathrm{SO}(3)$ eine eigentliche Bewegung des dreidimensionalen euklidischen Standardraumes. Dann ist das charakteristische Polynom von A ein normiertes Polynom vom Grad 3 mit konstantem Term -1, also hat es eine positive Nullstelle, die wegen Hilfssatz 12.1.12 zwangsläufig 1 sein muss. Es sei v ein Eigenvektor zum Eigenwert 1. Dann ist die von A beschriebene Abbildung auf dem Orthokomplement $v^\perp$ eine Drehung (wobei der Winkel auch 0 oder π sein kann, was $\pm$Identität auf $v^\perp$ entspricht). Also: Eine eigentliche Bewegung des $\mathbb{R}^3$ hat eine Drehachse und eine Drehebene, die aufeinander senkrecht stehen.

c) Eine Matrix $A \in \mathrm{O}(n)$ mit ungeradem n hat immer mindestens einen reellen Eigenwert, denn das charakteristische Polynom hat ungeraden Grad und daher greift der Zwischenwertsatz.

d) Nun wollen wir noch ein Beispiel in Dimension 4 ansehen, wo man nicht mit dem Orthogonalen Komplement zu einem Eigenvektor argumentieren kann.

Dazu sei

$$A = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

Das charakteristische Polynom von A ist $X^4 + 2X^2 + 1 = (X^2+1)^2$. Es ist sogar $A^2 + I_4$ die Nullmatrix. Wir nehmen irgendeinen Vektor der Länge 1 in $\mathbb{R}^4$. Um zu zeigen, dass es wirklich klappt, nehme ich nicht den ersten Standardbasisvektor (bei dem viele Rechnungen einfacher würden), sondern wähle betont ungeschickt etwa

$$b_1 := \frac{1}{2}(1\ 1\ 1\ 1)^\top.$$

Dann ist

$$b_2 := A \cdot b_1 = \frac{1}{2}(1\ -1\ -1\ 1)^\top.$$

Diese zwei Vektoren erzeugen einen zweidimensionalen A-invarianten Untervektorraum U von $\mathbb{R}^4$. Im orthogonalen Komplement $U^\perp$ liegen zum Beispiel die Vektoren

$$b_3 := \frac{1}{2}(1\ -1\ 1\ -1)^\top \text{ und } b_4 := A \cdot b_3 = \frac{1}{2}(1\ 1\ -1\ -1)^\top.$$

Wenn S die Matrix mit den Spalten b_1, b_2, b_3, b_4 ist, so ist S orthogonal. (Das ist hier ein Zufall, der am Charakter der speziell gewählten Matrix A liegt. Im Normalfall müssten hier b_2 und b_4 nach E. Schmidt um geeignete Vielfache von b_1 und b_3 abgeändert und anschließend normiert werden, um eine ONB zu erhalten.) Es gilt

$$S^{-1}AS = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Das sind zwei Drehkästchen zum Drehwinkel $\pi/2 = 90°$.

Bemerkung 12.1.18 (für Karlsruher LA-Klausuren)

Es sei $A \in \mathrm{O}(n)$ eine orthogonale Matrix. In einer badischen Universitätsstadt ist der folgende Trick zur Berechnung der Isometrie-Normalform sehr beliebt. Er funktioniert aber nur für maßgeschneiderte „Klausurmatrizen“ gut!

Wenn $D := S^{-1}AS$ die Isometrienormalform (wie in 12.1.16) von A ist (mit $S \in \mathrm{O}(n)$), so gilt

$$D^\top = (S^{-1}AS)^\top = S^\top A^\top (S^{-1})^\top = S^{-1}A^\top S,$$

wegen der Gestalt der Drehkästchen ist also die Matrix

$$S^{-1}(A + A^\top)S = D + D^\top$$

eine Diagonalmatrix. Demnach ist $A + A^\top$ diagonalisierbar (kein Zufall, wie wir in 12.2.6 sehen werden), und ihre Eigenwerte stehen mit den Drehkästchen von A in folgender Beziehung.

- Dem (eventuellen) Eigenwert 1 von A entspricht der (eventuelle) Eigenwert 2 von $A + A^\top$.
- Dem (eventuellen) Eigenwert -1 von A entspricht der (eventuelle) Eigenwert -2 von $A + A^\top$.
- Jedes Drehkästchen D_φ in D trägt zum Eigenraum $\mathrm{Eig}(A + A^\top, 2\cos\varphi)$ einen zweidimensionalen Summanden bei.

Insbesondere ist der Eigenraum $\mathrm{Eig}(A + A^\top, 2\cos\varphi)$ gleich dem Kern von $A^2 - 2(\cos\varphi) \cdot A + I_n$.

Manche Leute benutzen gerne diesen Zusammenhang beim Berechnen der Isometrienormalform.

12.2 Selbstadjungierte Abbildungen

Definition 12.2.1 (selbstadjungiert, A^*, hermitesche Matrizen)

Es sei V ein Vektorraum mit Skalarprodukt über $\mathbb{R}$ oder $\mathbb{C}$, und Φ sei ein Endomorphismus von V. Dann heißt Φ **selbstadjungiert**, wenn für alle $v, w \in V$ die Gleichung

$$\langle \Phi(v), w\rangle = \langle v, \Phi(w)\rangle$$

gilt.

Wenn V endlichdimensional ist und wir eine Orthonormalbasis $B = \{b_1, \ldots, b_n\}$ von V wählen, so ist Φ genau dann selbstadjungiert, wenn für alle Basisvektoren b_i, b_j die Gleichung

$$\langle \Phi(b_i), b_j\rangle = \langle b_i, \Phi(b_j)\rangle = \overline{\langle \Phi(b_j), b_i\rangle}$$

gilt. Da aber die Matrix mit den Einträgen $\langle \Phi(b_j), b_i\rangle$ genau die Abbildungsmatrix von Φ bezüglich B ist (denn B ist eine Orthonormalbasis, und es gilt die Fourierformel), sehen wir:

$$\boxed{\Phi \text{ ist selbstadjungiert } \iff D_{BB}(\Phi) = \overline{D_{BB}(\Phi)^\top}.}$$

Wir verwenden in Zukunft für eine Matrix $A \in \mathbb{C}^{m\times n}$ die Abkürzung

$$A^* := \overline{A^\top}.$$

Matrizen mit $A = A^*$ heißen **hermitesch**, das kennen wir schon aus 11.4.6.

Φ ist genau dann selbstadjungiert, wenn die Abbildungsmatrix von Φ bezüglich einer (beliebigen) Orthonormalbasis hermitesch ist.

Beispiel 12.2.2 (orthogonale Projektion, zweite Ableitung)

a) Eine orthogonale Projektion (siehe 11.3.4) π ist selbstadjungiert. Genauer sei V ein Vektorraum mit Skalarprodukt und U ein endlichdimensionaler Untervektorraum von V mit orthogonalem Komplement $U^\perp$. Wir definieren $\pi : V \longrightarrow V$ durch

$$\pi(u + u^\perp) := u, \quad u \in U,\ u^\perp \in U^\perp.$$

Dann gilt aber für beliebige Vektoren $u_1, u_2 \in U$ und $u_1^\perp, u_2^\perp \in U^\perp$ die Gleichung

$$\begin{aligned}\langle \pi(u_1 + u_1^\perp), u_2 + u_2^\perp\rangle &= \langle u_1, u_2 + u_2^\perp\rangle = \langle u_1, u_2\rangle \\ &= \langle u_1 + u_1^\perp, u_2\rangle = \langle u_1 + u_1^\perp, \pi(u_2 + u_2^\perp)\rangle.\end{aligned}$$

Dabei benutzen wir mehrfach die Definition von π und dass U und $U^\perp$ zueinander orthogonal sind.

Da sich jeder Vektor aus V als $u+u^\perp$ schreiben lässt, ist die Selbstadjungiertheit von π nachgewiesen.

b) Es sei V der Vektorraum der reellwertigen, unendlich oft differenzierbaren Funktionen mit beschränktem Träger auf $\mathbb{R}$. Auf V gibt es das Skalarprodukt

$$\forall f, g \in V: \quad \langle f, g\rangle := \int_{\mathbb{R}} f(x)g(x)\mathrm{d}x,$$

denn in Wirklichkeit integriert man ja nur über den Abschluss des Durchschnittes der Träger von f und g. Die zweite Ableitung $f \mapsto f''$ ist ein Endomorphismus von V. Nun seien $f, g \in V$ beliebig. Wir wählen Zahlen $a < b$, sodass der Träger von f und von g echt im Intervall $[a, b]$ enthalten ist. Dann gilt nach zweifacher partieller Integration und weil die Funktionen außerhalb des Intervalls $[a, b]$ Null sind

$$\langle f'', g\rangle = \langle f, g''\rangle,$$

also ist $f \mapsto f''$ selbstadjungiert.

Von ähnlichen Beispielen selbstadjungierter Endomorphismen geeigneter Funktionenräume wimmelt es nur so in der harmonischen Analysis und in der Spektraltheorie und -geometrie. Sie werden hoffentlich systematischer in der (linearen) Funktionalanalysis untersucht.

c) Es sei $V = \mathbb{R}^2$ der zweidimensionale euklidische Standardvektorraum und $\Phi \in \mathrm{End}(V)$ selbstadjungiert. Dann ist die Abbildungsmatrix von Φ bezüglich der Standardbasis symmetrisch (weil hermitesch und reell):

$$D = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \in \mathbb{R}^{2\times 2}.$$

Das charakteristische Polynom von Φ ist

$$\det(XI_2 - D) = X^2 - (a+c)X + (ac - b^2).$$

Die Nullstellen davon sind

$$\frac{1}{2}\left[(a+c) \pm \sqrt{(a+c)^2 - 4(ac-b^2)}\right] = \frac{1}{2}[a + c \pm \sqrt{(a-c)^2 + 4b^2}] \in \mathbb{R}.$$

Es gibt also einen Eigenvektor $v \in V$ von Φ. Wir werden gleich zeigen, dass dann auch ein von 0 verschiedener zu v senkrechter Vektor Eigenvektor ist und damit eine Basis von V aus Eigenvektoren zu Φ existiert. Das heißt: Φ ist diagonalisierbar.

Hilfssatz 12.2.3 (Eigenwerte sind reell; invariante Komplemente)

Es sei V ein reeller oder komplexer Vektorraum mit Skalarprodukt und $\Phi \in \mathrm{End}(V)$ selbstadjungiert. Dann gelten die folgenden zwei Aussagen:

a) Alle Eigenwerte von Φ sind reell.

b) Wenn $U \subseteq V$ ein endlichdimensionaler Φ-invarianter Untervektorraum ist, dann ist auch $U^\perp$ Φ-invariant. Also gibt es ein Φ-invariantes Komplement.

Beweis. a) Im Falle eines reellen Vektorraumes lässt man ohnehin nur reelle Eigenwerte zu. Es sei also V ein $\mathbb{C}$-Vektorraum und $\lambda \in \mathbb{C}$ ein Eigenwert von Φ. Zu λ wählen wir uns einen Eigenvektor $v \in V$. Dann gilt

$$v \neq 0, \quad \Phi(v) = \lambda \cdot v.$$

Nun benutzen wir die Definition der Selbstadjungiertheit von Φ für $v = w$. Es gilt

$$\lambda\langle v, v\rangle = \langle \lambda v, v\rangle = \langle \Phi(v), v\rangle = \langle v, \Phi(v)\rangle = \langle v, \lambda v\rangle = \overline{\lambda}\langle v, v\rangle.$$

Aus $\langle v, v\rangle \neq 0$ folgt $\lambda = \overline{\lambda}$, also $\lambda \in \mathbb{R}$.

b) Es sei $u^\perp \in U^\perp$. Dann gilt für alle $u \in U$:

$$\langle u, \Phi(u^\perp)\rangle = \langle \Phi(u), u^\perp\rangle = 0, \quad \text{da} \quad \Phi(u) \in U.$$

Also steht $\Phi(u^\perp)$ auf jedem Vektor aus U senkrecht und liegt damit in $U^\perp$. Das war zu zeigen. ◯

Folgerung 12.2.4 (Eigenwerte symmetrischer reeller Matrizen)

Es sei $A \in \mathbb{R}^{n\times n}$ eine symmetrische Matrix. Dann zerfällt das charakteristische Polynom von A in reelle Linearfaktoren.

Beweis. Wir können die Matrix A ja auch als komplexe Matrix auffassen, die noch dazu Abbildungsmatrix eines selbstadjungierten Endomorphismus Φ des unitären n-dimensionalen Standardraums bezüglich der Standardbasis ist. Dann zerfällt $\mathrm{CP}_A(X) = \mathrm{CP}_\Phi(X)$ wegen des Fundamentalsatzes der Algebra (12.1.13) in Linearfaktoren. Die Nullstellen dieser Linearfaktoren sind aber genau die Eigenwerte von Φ, und diese sind nach Hilfssatz 12.2.3 alle reell. ◯

Satz 12.2.5 (Spektralsatz für selbstadjungierte Abbildungen)

Es sei V ein endlichdimensionaler Vektorraum über $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$ mit Skalarprodukt und Φ ein Endomorphismus von V. Dann sind die folgenden beiden Aussagen äquivalent:

i) Φ ist selbstadjungiert.

ii) Es gibt eine Orthonormalbasis aus Eigenvektoren von Φ, und die Eigenwerte sind alle reell.

Beweis.

i)⇒ii): Wir nehmen zunächst an, dass Φ selbstadjungiert ist. Wenn $V = \{0\}$ gilt, ist nichts zu zeigen. Wir nehmen diesen Fall als Anfang einer vollständigen Induktion.

Sei nun $n = \dim(V) \geq 1$.

Dann gibt es einen reellen Eigenwert λ von Φ. Denn entweder ist V ein $\mathbb{C}$-Vektorraum und wir benutzen den Fundamentalsatz der Algebra (12.1.13) und 12.2.3, oder V ist ein reeller Vektorraum und Φ wird bezüglich einer Orthonormalbasis durch eine symmetrische Matrix beschrieben, dann benutzen wir Folgerung 12.2.4.

Nun sei $b_1 \in V$ ein Eigenvektor zum Eigenwert λ und $U := \mathbb{K} \cdot b_1$ die von b_1 erzeugte Gerade. Dann ist $U^\perp$ unter Φ invariant, und die Einschränkung $\Phi|_{U^\perp}$ ist auch selbstadjungiert, also gibt es eine Orthonormalbasis $\{b_2, \ldots, b_n\}$ von $U^\perp$, die aus Eigenvektoren besteht, und alle Eigenwerte sind reell. Dann leistet $\{b_1, \ldots, b_n\}$ das Gewünschte.

ii)⇒i): Wenn es eine Orthonormalbasis B aus Eigenvektoren gibt und wenn alle Eigenwerte reell sind, dann ist die Abbildungsmatrix von Φ bezüglich B eine reelle Diagonalmatrix, also hermitesch, und damit ist Φ nach dem Kriterium aus 12.2.1 selbstadjungiert. ◯

Folgerung 12.2.6 (Matrizensprache)

Wenn A eine reelle symmetrische Matrix ist, dann gibt es eine orthogonale Matrix S, sodass $S^{-1}AS$ eine Diagonalmatrix ist: A ist „orthogonal ähnlich“ zu einer Diagonalmatrix.

Wenn A eine komplexe hermitesche Matrix ist, dann gibt es eine unitäre Matrix S, sodass $S^{-1}AS$ eine Diagonalmatrix mit reellen Einträgen ist: A ist „unitär ähnlich“ zu einer reellen Diagonalmatrix.

Diese Eigenschaften symmetrischer, reeller Matrizen sind für viele Anwendungen wichtig. So ist zum Beispiel die Adjazenzmatrix eines endlichen (ungerichteten) Graphen eine symmetrische Matrix. Ihre Eigenwerte codieren Information über metrische Eigenschaften des Graphen.

Vorsicht: Es gibt komplexe symmetrische Matrizen, die nicht diagonalisierbar sind. Zum Beispiel gibt es komplexe symmetrische $2 \times 2-$Matrizen mit Rang 1 und Spur 0 (überlegen Sie sich ein Beispiel!). Diese haben Jordan'sche Normalform $\left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right)$. (Wieso?)

Folgerung 12.2.7 (Positivität)

Eine symmetrische Matrix $A \in \mathbb{R}^{n \times n}$ ist genau dann positiv definit, wenn alle ihre Eigenwerte positiv sind.

Beweis. Wir wissen schon wegen 12.2.6, dass A reell diagonalisierbar ist. Das benutzen wir jetzt in beiden Richtungen.

Wenn A positiv definit ist, so müssen wir überprüfen, dass alle Eigenwerte von A positiv sind. Es sei λ ein Eigenwert von A, und $v \in \mathrm{Eig}(A, \lambda)$ ein Eigenvektor. Dann gilt

$$0 < v^\top \cdot A \cdot v = \lambda \cdot v^\top \cdot v,$$

also $\lambda > 0$, da $v^\top \cdot v > 0$. (Es ist ja $v \in \mathbb{R}^n,\ v \neq 0$.)

Wenn umgekehrt alle Eigenwerte von A positiv sind, so gibt es eine orthogonale Matrix S mit

$$S^{-1} \cdot A \cdot S = \mathrm{diag}(\lambda_1, \ldots, \lambda_n),$$

wobei für alle $1 \leq i \leq n$ gilt, dass $\lambda_i > 0$. Wir erinnern uns an $S^{-1} = S^\top$ und setzen $D := \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$. Dann gilt

$$S^{-1} \cdot A \cdot S = D^2 = D^\top \cdot D \Rightarrow A = S \cdot D^\top \cdot D \cdot S^\top = (DS^\top)^\top \cdot (DS^\top).$$

Das zeigt für $v \in \mathbb{R}^n,\ v \neq 0:\ v^\top \cdot A \cdot v = (DS^\top v)^\top \cdot (DS^\top v) > 0$, da $DS^\top$ regulär ist. ◯

Bemerkung 12.2.8 (hermitesche Matrizen, Signatur)

a) Dasselbe Kriterium gilt auch für hermitesche Matrizen: sie sind positiv definit genau dann, wenn alle Eigenwerte (die ja reelle Zahlen sind) positiv sind.

b) Es sei A eine symmetrische reelle Matrix. Dann nennt man die Anzahl der negativen Eigenwerte von A auch die **Signatur** oder den **Index** von A. Mit demselben Trick wie bei der Rückrichtung des Beweises von Anwendung 12.2.7 sieht man: es gibt eine invertierbare Matrix S, sodass

$$S^\top \cdot A \cdot S$$

eine Diagonalmatrix mit Diagonaleinträgen in $\{0, 1, -1\}$ ist.

Das ist (wie 10.1.6 nahelegt) eine Aussage über eine Bilinearform. Sehen wir uns das doch näher an!

Satz 12.2.9 (Trägheitssatz von Sylvester)

Es seien V ein endlichdimensionaler reeller Vektorraum und $P : V \times V \longrightarrow \mathbb{R}$ eine symmetrische Bilinearform. Dann gelten:

a) V lässt sich zerlegen als $V = V_0 \oplus V_+ \oplus V_-$, wobei P auf V_+ positiv definit, auf V_- negativ definit und auf V_0 konstant gleich 0 ist, und gleichzeitig für beliebige Vektoren v, w in zwei verschiedenen der drei Summanden die „Orthogonalitätsbeziehung“

$$P(v, w) = 0$$

gilt.

b) Die Dimensionen von V_+, V_- und V_0 hängen nur von P (und nicht vom Zufall bei der Wahl einer Zerlegung) ab.

Beweis. a) Wir beschreiben P durch eine Fundamentalmatrix F bezüglich irgendeiner Basis. F ist symmetrisch und damit orthogonal diagonalisierbar, es gibt also eine orthogonale Matrix S, sodass $S^{-1}FS$ diagonal ist. Das gilt dann aber wegen $S^{-1} = S^\top$ auch für $S^\top FS$, und wir haben eine Orthogonalbasis B für P gefunden.

Nun sei V_0 der von $\{b \in B \mid P(b,b) = 0\}$ erzeugte Untervektorraum, und analog $V_+ = \langle\{b \in B \mid P(b,b) > 0\}\rangle$, $V_- = \langle\{b \in B \mid P(b,b) < 0\}\rangle$. Dann haben wir eine Zerlegung wie gewünscht (nachrechnen!).

b) Es sei $V = U_0 \oplus U_+ \oplus U_-$ eine weitere Zerlegung von V mit den Eigenschaften der Zerlegung von a). Insbesondere ist dann

$$\dim U_0 = \dim V_0$$

(wie die Ränge von Fundamentalmatrizen bezüglich zweier an die Zerlegungen angepasster Basen zeigen).

Wäre nun $\dim U_+ > \dim V_+$, so wäre nach der Dimensionsformel 5.4.3 der Durchschnitt

$$U_+ \cap (V_0 \oplus V_-)$$

nichttrivial, es gäbe also in U_+ ein Element $u = v_0 + v_- \neq 0$, und für dieses wäre

$$P(u,u) = P(v_0, v_0) + P(v_-, v_-) \leq 0$$

im Gegensatz zur postulierten positiven Definitheit von P auf U_+. Es folgt aus Symmetriegründen $\dim U_+ = \dim V_+$, und analog auch $\dim U_- = \dim V_-$. ◯

12.3 Normale Abbildungen

Eine Gemeinsamkeit von linearen Isometrien und selbstadjungierten Abbildungen ist, dass es eine gute Kontrolle darüber gibt, wie man im Skalarprodukt

$$\langle \Phi(v), w\rangle$$

das „Φ auf die andere Seite bringen“ kann. Genauer gilt ja

$$\langle \Phi(v), w\rangle = \begin{cases} \langle v, \Phi^{-1}(w)\rangle, & \text{wenn } \Phi \text{ bijektive Isometrie ist,} \\ \langle v, \Phi(w)\rangle, & \text{wenn } \Phi \text{ selbstadjungiert ist.} \end{cases}$$

Das ist der Schlüssel zur Existenz eines invarianten Komplementärraums zu einem invarianten Unterraum. Nun wendet man die Geschichte neu und erfindet ein Konzept.

Definition 12.3.1 Adjungierte Abbildung, Normale Endomorphismen

a) Es seien $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$ und V, W zwei $\mathbb{K}$-Vektorräume mit Skalarprodukten $\langle \cdot, \cdot \rangle_V$ und $\langle \cdot, \cdot \rangle_W$. Weiter sei $\Phi : V \longrightarrow W$ ein Homomorphismus. Dann gibt es für jedes $w \in W$ höchstens ein Element $\Phi^*(w) \in V$, sodass für alle $v \in V$ die Gleichung

$$\langle \Phi(v), w \rangle = \langle v, \Phi^*(w) \rangle$$

erfüllt ist. Wenn es zwei solche Elemente gäbe, müsste ihre Differenz (ein Element $\widetilde{v} \in V$) ja insbesondere auf sich selbst senkrecht stehen, was der positiven Definitheit widerspricht.

Wenn für jedes $w \in W$ solch ein $\Phi^*(w)$ existiert, dann heißt die Abbildung

$$\Phi^* : W \longrightarrow V, \ w \mapsto \Phi^*(w),$$

die zu Φ **adjungierte Abbildung.**

b) Wenn in der Situation aus a) die adjungierte Abbildung existiert, dann ist diese Abbildung linear. Denn für alle $w_1, w_2 \in W, \ v \in V, \alpha \in \mathbb{K}$ gilt:

$$\begin{aligned} \langle v, \Phi^*(\alpha w_1 + w_2) \rangle &= \langle \Phi(v), \alpha w_1 + w_2 \rangle \\ &= \overline{\alpha} \langle \Phi(v), w_1 \rangle + \langle \Phi(v), w_2 \rangle \\ &= \overline{\alpha} \langle v, \Phi^*(w_1) \rangle + \langle v, \Phi^*(w_2) \rangle \\ &= \langle v, \alpha \Phi^*(w_1) + \Phi^*(w_2) \rangle. \end{aligned}$$

Da dies für alle $v \in V$ gilt, folgt

$$\Phi^*(\alpha w_1 + w_2) = \alpha \Phi^*(w_1) + \Phi^*(w_2).$$

c) In der Situation aus a) sei $V = W$. Dann heißt Φ ein **normaler Endomorphismus**, wenn die adjungierte Abbildung Φ^* existiert und folgendes gilt:

$$\Phi \circ \Phi^* = \Phi^* \circ \Phi.$$

Beispiel 12.3.2 (Surjektive) Isometrien und selbstadjungierte Abbildungen sind normal. Im ersten Fall gilt $\Phi^* = \Phi^{-1}$, im zweiten haben wir $\Phi^* = \Phi$.

In diesem Abschnitt wird es darum gehen zu zeigen, dass die jeweilige Normalform ein Spezialfall der Normalform normaler Abbildungen ist, wie sie im Spektralsatz 12.3.8 vorgestellt wird.

Wir verwenden nun die Notation aus Definition 12.2.1.

Hilfssatz 12.3.3 (Abbildungsmatrix der adjungierten Abbildung)

Es seien V, W endlichdimensionale $\mathbb{K}$-Vektorräume mit Skalarprodukt und $\Phi : V \longrightarrow W$ ein Homomorphismus. Dann existiert die adjungierte Abbildung Φ^*. Wenn B bzw. C Orthonormalbasen von V bzw. W sind, dann gilt für die Abbildungsmatrizen:

$$D_{BC}(\Phi^*) = \big(D_{CB}(\Phi)\big)^*.$$

Beweis. Es sei Ψ der Homomorphismus von W nach V mit

$$D_{BC}(\Psi) := D_{CB}(\Phi)^* = \overline{D_{CB}(\Phi)^\top}.$$

Da wir die Abbildung Φ mithilfe von Orthonormalbasen beschrieben haben, gilt für alle $v \in V,\ w \in W$:

$$\begin{aligned}\langle \Phi(v), w\rangle &= (D_{CB}(\Phi) \cdot D_B(v))^\top \cdot \overline{D_C(w)} \\ &= D_B(v)^\top \cdot D_{CB}(\Phi)^\top \cdot \overline{D_C(w)} \\ &= D_B(v)^\top \cdot \overline{D_{CB}(\Phi)^* \cdot D_C(w)} \\ &= \langle v, \Psi(w)\rangle.\end{aligned}$$

Also ist Ψ ein Homomorphismus, der genau das tut, was die adjungierte Abbildung leisten soll. ◯

Beispiel 12.3.4 (Dimension 1 und 2)

a) Jeder Endomorphismus des eindimensionalen Standardraums ist normal.

b) Nun betrachten wir Endomorphismen von $\mathbb{K}^2$, die durch ihre Abbildungsmatrix $D = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ bezüglich der Standardbasis gegeben sind. Die Abbildungsmatrix der adjungierten Abbildung ist also wegen 12.3.3 gleich $D^* = \left(\begin{smallmatrix} \overline{a} & \overline{c} \\ \overline{b} & \overline{d} \end{smallmatrix}\right)$. Wir müssen zu verstehen versuchen, was die Bedingung $D \cdot D^* = D^* \cdot D$ bedeutet. Wir rechnen die Produkte aus und erhalten

$$\begin{pmatrix} a\cdot\overline{a} + b\cdot\overline{b} & a\cdot\overline{c} + b\cdot\overline{d} \\ c\cdot\overline{a} + d\cdot\overline{b} & c\cdot\overline{c} + d\cdot\overline{d} \end{pmatrix} = \begin{pmatrix} a\cdot\overline{a} + c\cdot\overline{c} & b\cdot\overline{a} + d\cdot\overline{c} \\ a\cdot\overline{b} + c\cdot\overline{d} & b\cdot\overline{b} + d\cdot\overline{d} \end{pmatrix}.$$

Diese Gleichung ist genau dann erfüllt, wenn die folgenden Bedingungen gelten:

$$|b| = |c| \quad \text{und} \quad c\cdot\overline{a} + d\cdot\overline{b} = a\cdot\overline{b} + c\cdot\overline{d}.$$

Die zweite dieser Bedingungen ist

$$c\cdot(\overline{a} - \overline{d}) = \overline{b}\cdot(a - d).$$

Nun unterscheiden wir zwei Fälle:

- Wenn $a \neq d$ gilt, ist D genau dann normal, wenn $c = \overline{b} \cdot \frac{a-d}{\overline{a}-\overline{d}}$.
- Wenn $a = d$ gilt, ist D genau dann normal, wenn $|b| = |c|$.

Es lohnt sich, den reellen Fall noch einmal gesondert zu betrachten. Es gibt die folgenden zwei Typen von reellen normalen 2×2-Matrizen:

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix},\ \begin{pmatrix} a & b \\ b & d \end{pmatrix}.$$

Diese Matrizen sind entweder symmetrisch, oder sie sind reelle Vielfache von Drehkästchen (aus Beispiel 12.1.4).

Beispiel 12.3.5 (noch mehr normale Matrizen)

Eine Blockmatrix $M = \left(\begin{smallmatrix} A & 0 \\ 0 & D \end{smallmatrix}\right)$ mit quadratischen Matrizen A, D und passenden 0-Matrizen ist genau dann normal (d.h. $MM^* = M^*M$), wenn A und D normal sind. Denn $M^* = \left(\begin{smallmatrix} A^* & 0 \\ 0 & D^* \end{smallmatrix}\right)$. Induktiv sehen wir damit, dass eine Blockdiagonalmatrix, die auf der Diagonalen entweder 1×1-Matrizen oder 2×2-Matrizen aus Beispiel 12.3.4 stehen hat, normal ist.

Dass dies der Normalfall der Normalität ist, sehen wir in Kürze. Dazu brauchen wir die Verallgemeinerung der Aussagen 12.1.11 und 12.2.3.

Hilfssatz 12.3.6 (Existenz invarianter Komplemente)

Es sei V ein endlichdimensionaler $\mathbb{K}$-Vektorraum mit Skalarprodukt, Φ ein normaler Endomorphismus von V, und $U \leq V$ ein Φ-invarianter Untervektorraum. Dann ist auch $U^\perp$ unter Φ invariant.

Beweis. Wir wählen eine Orthonormalbasis von U und ergänzen sie zu einer Orthonormalbasis B von V. Dann gilt für die Abbildungsmatrix von Φ bezüglich dieser Basis:

$$D_{BB}(\Phi) = \begin{pmatrix} A & C \\ 0 & D \end{pmatrix},$$

wobei A eine quadratische Matrix ist, die die Wirkung von Φ auf U bezüglich der Basis $B \cap U$ beschreibt. Wir müssen zeigen, dass $C = 0$ gilt.

Da B eine Orthonormalbasis ist, bedeutet die Normalität von Φ aber, dass

$$D_{BB}(\Phi)\big(D_{BB}(\Phi)\big)^* = \big(D_{BB}(\Phi)\big)^* D_{BB}(\Phi).$$

Das schreiben wir jetzt expliziter hin; es muss gelten

$$\begin{pmatrix} AA^* + CC^* & CD^* \\ DC^* & DD^* \end{pmatrix} = \begin{pmatrix} A^*A & A^*C \\ C^*A & D^*D + C^*C \end{pmatrix}.$$

Insbesondere gilt $AA^* + CC^* = A^*A$. Das impliziert aber wegen der bekannten Identität $\mathrm{Spur}(MN) = \mathrm{Spur}(NM)$ für zwei Matrizen M, N, deren Produkte beide ausführbar sind, dass $\mathrm{Spur}(CC^*) = 0$, denn die Spur ist ja additiv.

Die Diagonaleinträge von CC^* sind gerade die Normquadrate der Zeilen von C bezüglich des Standard-Skalarproduktes. Das sind alles nicht-negative reelle Zahlen. Also ist die Spur von CC^* genau dann gleich 0, wenn $C = 0$, was damit bewiesen ist. ◯

Folgerung 12.3.7 *Es sei V ein endlichdimensionaler $\mathbb{K}$-Vektorraum mit Skalarprodukt, Φ ein normaler Endomorphismus von V und $U \leq V$ ein Φ-invarianter Untervektorraum. Dann ist die Einschränkung von Φ auf U ein normaler Endomorphismus von U.*

Beweis. Im Beweis von 12.3.6 haben wir ja (mit der dortigen Notation) gesehen, dass $AA^* = A^*A$. Das ist genau die Behauptung. ◯

Nun geht es so weiter wie schon in den Abschnitten 12.1 und 12.2.

Satz 12.3.8 (Spektralsatz für normale Endomorphismen)

Es sei V ein endlichdimensionaler $\mathbb{K}$-Vektorraum mit Skalarprodukt, und Φ ein normaler Endomorphismus von V. Dann gilt:

a) Im Fall $\mathbb{K} = \mathbb{C}$ gibt es eine Orthonormalbasis aus Eigenvektoren von Φ.

b) Im Fall $\mathbb{K} = \mathbb{R}$ ist V die orthogonale Summe von ein- oder zweidimensionalen Φ-invarianten Untervektorräumen.

Beweis.

Der Beweis läuft in beiden Fällen induktiv nach der Dimension von V.

Induktionsanfang ist $\dim(V) = 0$ oder 1. In diesem Fall ist die Behauptung klar (wobei der Nullvektorraum die direkte Summe einer leeren Menge von eindimensionalen Vektorräumen ist...).

Es sei also $\dim V \geq 2$.

a) Wenn $\mathbb{K} = \mathbb{C}$ gilt, so wähle einen Eigenwert λ von Φ und einen Eigenvektor v der Länge 1. Dann ist $(\mathbb{C} \cdot v)^\perp$ nach 12.3.6 ein Φ-invariantes Komplement zu $\mathbb{C} \cdot v$, und wegen 12.3.7 lässt sich auf die Einschränkung von Φ auf $(\mathbb{C} \cdot v)^\perp$ die Induktionsvoraussetzung anwenden. Das zeigt die Behauptung.

b) Wenn $\mathbb{K} = \mathbb{R}$ gilt und ein (reeller) Eigenwert von Φ existiert, dann können wir genauso verfahren wie im komplexen Fall. Wenn kein reeller Eigenwert existiert, dann hat das charakteristische Polynom (vergleichen Sie das Argument mit dem aus Beweis 12.1.14!) einen quadratischen Faktor $X^2 - sX + n$, und wir wählen ein Element $v \neq 0$ im Kern von $\Phi^2 - s\Phi + n\mathrm{Id}_V$. Dieses erzeugt zusammen mit $\Phi(v)$ einen zweidimensionalen Φ-invarianten Untervektorraum U. Nun verwenden wir die Induktionsannahme für $U^\perp$. ◯

Folgerung 12.3.9 (Matrizensprache)

a) Für jede normale Matrix $A \in \mathbb{C}^{n\times n}$ gibt es eine unitäre Matrix $S \in \mathrm{U}(n)$, sodass $S^{-1}AS$ eine Diagonalmatrix ist. Die Diagonaleinträge können übrigens beliebige komplexe Zahlen sein, es gibt nicht mehr die Einschränkung „Betrag gleich 1“ wie im Fall der Isometrie oder „Eigenwerte sind reell“ wie im selbstadjungierten Fall.

b) Für jede normale Matrix $A \in \mathbb{R}^{n\times n}$ gibt es eine orthogonale Matrix $S \in \mathrm{O}(n)$, sodass $S^{-1}AS$ eine Blockdiagonalmatrix ist, die auf der Diagonalen entweder reelle Eigenwerte stehen hat oder reelle Matrizen der Gestalt $\left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$ mit $b \neq 0$.

Das sieht man in Beispiel 12.3.4, denn genau diese 2×2-Matrizen sind normal und nicht reell diagonalisierbar.

Beispiel 12.3.10 (Damit hatte ich jetzt nicht gerechnet!)

a) Es gibt für $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ kein Skalarprodukt auf $\mathbb{K}^2$, bezüglich dessen der Endomorphismus von $\mathbb{K}^2$ mit Abbildungsmatrix $A = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ normal wäre. Denn es gibt zum Kern keinen invarianten Komplementärraum.

b) Jeder diagonalisierbare Endomorphismus eines komplexen Vektorraumes ist bezüglich eines geeigneten Skalarproduktes normal. Man wähle hierfür einfach eine Basis aus Eigenvektoren und konstruiere das Skalarprodukt so, dass diese Basis eine Orthonormalbasis ist.

Kapitel 13

Affine Geometrie

13.1 Affine Räume und Abbildungen

Wir sind in Definition 11.3.7 schon über den Begriff „affiner Teilraum“ gestolpert. Das ist eine Teilmenge eines Vektorraumes von der Gestalt $v+W$, wobei W ein Untervektorraum ist. Insbesondere ist die Summe eines Vektors aus W und eines Elements aus $v+W$ wieder in $v+W$. Die Abbildung $+ : W \times (v+W) \longrightarrow v+W$ hat einige interessante Eigenschaften:

- $\forall P \in v+W : 0+P = P$.
- $\forall P \in v+W : \forall w_1, w_2 \in W : w_1 + (w_2 + P) = (w_1 + w_2) + P$.
- Für alle $P, Q \in v+W$ gibt es genau ein $w \in W$ mit $w + P = Q$.

Diese Eigenschaften sind hier selbstverständlich. Man befreit sich nun von der starren Situation, dass W und $v+W$ in einem größeren Vektorraum liegen, und definiert in Anlehnung an die obigen Eigenschaften ein neues Konzept.

Definition 13.1.1 (affiner Raum)

Es seien K ein Körper, V ein K-Vektorraum, A eine nichtleere Menge und

$$\tau : V \times A \longrightarrow A$$

eine Abbildung. Dann heißt das Tupel (A, V, τ) ein **affiner Raum** mit **Translationsvektorraum V** und mit **Addition** τ, wenn folgende Bedingungen gelten:

- $\forall P \in A : \tau(0, P) = P$.
- $\forall P \in A : \forall v_1, v_2 \in V : \tau(v_1, \tau(v_2, P)) = \tau((v_1 + v_2), P)$.

- Für alle $P, Q \in A$ gibt es genau ein $v \in V$ mit $\tau(v, P) = Q$.

Für diesen eindeutig bestimmten Vektor v aus der dritten Bedingung schreibt man oft

$$v =: \overrightarrow{PQ} .$$

Man sagt, das sei der Vektor, der **von P nach Q weist.**

Die ersten zwei Bedingungen sagen in der Terminologie von Kapitel 2.5, dass die additive Gruppe V auf der Menge A operiert. Die dritte Bedingung nennt man die **einfache Transitivität** dieser Operation. Dabei heißt **transitiv**, dass es nur eine Bahn dieser Gruppenoperation gibt (siehe bei 2.5.4). Der Zusatz „einfach“ bezieht sich auf das „genau“ aus der dritten Bedingung.

Bemerkung 13.1.2 a) Es sei V ein Vektorraum und $\tau = +$ die Addition auf V. Dann ist $(V, V, +)$ ein affiner Raum. Im Falle $V = K^n$ heißt er der **n-dimensionale affine Standardraum**:

$$\mathbb{A}^n(K) := (K^n, K^n, +).$$

Ich persönlich finde das schade. Einleuchtender ist die Variante

$$\widetilde{\mathbb{A}^n}(K) := \{ \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \\ 1 \end{pmatrix} \mid x_1, \dots, x_n \in K \} \subseteq K^{n+1}$$

mit Translationsvektorraum

$$\widetilde{K^n} := \{ \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \\ 0 \end{pmatrix} \mid x_1, \dots, x_n \in K \} \subseteq K^{n+1}$$

und der Addition aus K^{n+1}.

Der Vorteil dieses Modells ist, dass die Rollen des Translationsraumes und des affinen Raumes besser getrennt werden können, einfach weil es sich um zwei unterschiedliche Mengen handelt.

b) Nun sei A eine Menge und $\Phi : V \longrightarrow A$ eine (hiermit fest gewählte) Bijektion. Dann wird durch

$$\forall v \in V, P \in A : \ \tau(v, P) := \Phi(v + \Phi^{-1}(P))$$

ein affiner Raum (A, V, τ) festgelegt (wie man leicht nachrechnet).

c) Umgekehrt sei (A, V, τ) ein beliebiger affiner Raum und $P_0 \in A$ ein Element. Wir betrachten dann die Abbildung

$$\Phi : V \longrightarrow A, \quad \Phi(v) := \tau(v, P_0).$$

Diese Abbildung ist eine Bijektion wegen der einfachen Transitivität. Es gilt wegen der zweiten Bedingung aus Definition 13.1.1

$$\Phi(v_1 + v_2) = \tau(v_1, \Phi(v_2)).$$

Das heißt, dass das folgende Diagramm kommutiert:

$$\begin{array}{ccc} V \times V & \stackrel{\mathrm{Id}_V \times \Phi}{\longrightarrow} & V \times A \\ +\downarrow & & \downarrow \tau \\ V & \stackrel{\Phi}{\longrightarrow} & A \end{array}$$

d) In gewisser Weise sagen b) und c), dass ein affiner Raum so etwas ist wie ein Vektorraum, von dem man sich den Nullpunkt nicht merkt.

Wir werden in Zukunft meistens $+$ als Symbol für die Translationsabbildung anstelle von τ verwenden. Das passt besser zu der in vielen Lehrbüchern üblichen Notation. Außerdem wird oftmals einfach A (anstelle von (A, V, τ)) als affiner Raum bezeichnet.

Definition 13.1.3 (affiner Teilraum)

Es sei $(A, V, +)$ ein affiner Raum. Eine Teilmenge $B \subseteq A$ heißt **affiner Teilraum** von A, wenn ein $b \in B$ und ein Untervektorraum $W \leq V$ existieren, sodass

$$B = W + b = \{w + b \mid w \in W\}.$$

Im Falle des affinen Standardraumes stimmt das (zum Glück!) überein mit Definition 11.3.7. Als „Fußpunkt“ kann man übrigens statt eines einmal gefundenen b auch (und zwar wegen 13.1.2 b) und c)) jedes andere Element von B nehmen. Es gilt ja für $b_0 \in B$:

$$\exists w_0 \in W : b_0 = w_0 + b \Rightarrow W + b = W + b_0.$$

Der Durchschnitt zweier affiner Teilräume B_1, B_2 ist entweder leer oder wieder ein affiner Teilraum. Wenn er nämlich nicht leer ist, so wähle ein b im Schnitt. Es gilt dann offensichtlich (für $B_1 = W_1 + b, B_2 = W_2 + b$):

$$B_1 \cap B_2 = (W_1 \cap W_2) + b,$$

und $W_1 \cap W_2$ ist ein Untervektorraum des Translationsvektorraumes von A.

Definition 13.1.4 (Affine Abbildungen, Affinitäten)

Es seien A und B zwei affine Räume mit zugehörigen Translationsvektorräumen V und W über demselben Körper K. Weiter sei $a \in A$ gewählt. Eine Abbildung $\Phi : A \longrightarrow B$ induziert (bei gewähltem a!) eine Abbildung $\varphi : V \longrightarrow W$ wie folgt:

$$\Phi(v + a) = \varphi(v) + \Phi(a).$$

Also: $\varphi(v) \in W$ ist der (eindeutig bestimmte) Translationsvektor, der von $\Phi(a)$ nach $\Phi(v + a)$ weist.

Die Abbildung Φ heißt **affine Abbildung,** oder auch **affiner Homomorphismus**, falls φ ein Vektorraumhomomorphismus ist. Eine invertierbare affine Abbildung heißt eine **Affinität**.

Die Frage, ob Φ affin ist oder nicht, hängt nicht von der Wahl von a ab. Denn: wenn Φ nach Wahl des Punktes a sich als affin herausstellt und $\widetilde{a}$ ein weiterer Punkt in A ist, so gilt für $\widetilde{v} \in V$:

$$\begin{aligned}\Phi(\widetilde{v} + \widetilde{a}) &= \Phi(\widetilde{v} + \overrightarrow{a\widetilde{a}} + a) = \varphi(\widetilde{v} + \overrightarrow{a\widetilde{a}}) + \Phi(a) \\ &= \varphi(\widetilde{v}) + \varphi(\overrightarrow{a\widetilde{a}}) + \Phi(a) \\ &= \varphi(\widetilde{v}) + \Phi(\widetilde{a}).\end{aligned}$$

Wir sehen dabei sogar, dass eine andere Wahl des Punktes a nicht einmal die Abbildung φ ändert, solange diese linear (es langt sogar additiv) ist. φ heißt der **lineare Anteil** von Φ.

Die Umkehrabbildung zu einer Affinität ist immer auch eine Affinität. Es ist nämlich φ dann auch invertierbar (und φ^{-1} linear), und wir bekommen mit $b := \Phi(a)$ für beliebiges $w \in W$:

$$\Phi^{-1}(w + b) = \varphi^{-1}(w) + \Phi^{-1}(b).$$

Wenn $\Phi : A \longrightarrow B$ und $\Psi : B \longrightarrow C$ affine Abbildungen sind, so ist auch $\Psi \circ \Phi : A \longrightarrow C$ eine affine Abbildung. Sind φ und ψ die linearen Anteile von Φ und Ψ, so gilt

$$\Psi \circ \Phi(v + a) = \Psi(\varphi(v) + \Phi(a)) = \psi \circ \varphi(v) + \Psi \circ \Phi(a).$$

Insbesondere bilden die Affinitäten $\Phi : A \longrightarrow A$ von A eine Untergruppe der symmetrischen Gruppe $\mathrm{Sym}\,(A)$. Sie heißt die **affine Gruppe** von A und wird mit $\mathrm{Aff}(A)$ notiert.

Beispiel 13.1.5 (Standardraum, affine Gruppe)

Es sei $A = \mathbb{A}^n(K)$ der n-dimensionale affine Standardraum. Die affinen Abbildungen von A nach A sind genau die Abbildungen der Gestalt

$$\Phi : A \longrightarrow A, \quad \Phi(a) := M \cdot a + t,$$

wobei $M \in K^{n\times n}$ und $t \in K^n$ beliebig sind. Das wird noch schöner in meinem Lieblingsmodell $\widetilde{\mathbb{A}^n}(K)$. Hier sind die affinen Abbildungen gegeben durch Multiplikation mit $(n+1)\times(n+1)$-Matrizen des Typs

$$\begin{pmatrix} M & t \\ 0 & 1 \end{pmatrix}, \quad \text{wobei} \quad M \in K^{n\times n},\ t \in K^n.$$

Es gilt ja

$$\begin{pmatrix} M & t \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ 1 \end{pmatrix} = \begin{pmatrix} Mx+t \\ 1 \end{pmatrix}.$$

Damit gilt für die affine Gruppe

$$\mathrm{Aff}_n(K) := \mathrm{Aff}(K^n) \cong \{\begin{pmatrix} M & t \\ 0 & 1 \end{pmatrix} \mid M \in \mathrm{GL}_n(K), t \in K^n\}.$$

Diese Untergruppe von $\mathrm{GL}_{n+1}(K)$ lässt den affinen Teilraum $\widetilde{\mathbb{A}^n}(K)$ fest. In dieser Gruppe liegen zwei interessante Untergruppen.

Die eine ist die Gruppe aller Translationen, K^n, das ist der Translationsvektorraum, der aus der Menge den affinen Raum macht. Diese Gruppe ist der Kern des Gruppenhomomorphismus

$$\delta : \mathrm{Aff}_n(K) \longrightarrow \mathrm{Aut}(K^n) = \mathrm{GL}_n(K),$$

der einer Affinität ihren linearen Anteil zuordnet. Damit ist die Translationsgruppe eine normale Untergruppe von $\mathrm{Aff}_n(K)$ (siehe Bemerkung 2.3.9).

Die andere ist die Untergruppe aller Affinitäten, die den Fußpunkt $\binom{0}{1}$, $0 \in K^n$, auf sich selbst abbilden. Das ist (in unserem Standardmodell) die Gruppe $\mathrm{GL}_n(K)$. Diese ist kein Normalteiler von $\mathrm{Aff}_n(K)$, wenn $n \geq 1$. Denn es gilt für jeden Vektor $t \in K^n$ und jedes $M \in \mathrm{GL}_n(K)$:

$$\begin{pmatrix} I_n & -t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} M & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} I_n & t \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} M & Mt-t \\ 0 & 1 \end{pmatrix}.$$

Hierbei wird in den wenigsten Fällen t ausgerechnet im Eigenraum zum Eigenwert 1 von M liegen, also steht rechts meistens keine Affinität, die den Fußpunkt festlässt.

Diese algebraische Eigenschaft der Untergruppen bringt zum Ausdruck, dass zwar der lineare Anteil von Φ unabhängig von der Wahl eines Punktes in a ist, der „Translationsanteil“ aber sehr wohl von der Wahl von a abhängt. Ein algebraischer Aspekt dieser Tatsache ist, dass die affine Gruppe zwar nicht das direkte Produkt der linearen und der Translationsgruppe ist, aber immerhin ihr **semidirektes Produkt** – ein Begriff, der hier nicht weiter erklärt werden soll, der sich aber in Büchern über Algebra oder Gruppentheorie findet.

Beispiel 13.1.6 (Dimension)

Es sei A ein affiner Raum mit einem endlichdimensionalen Translationsvektorraum V. Dann nennt man die Dimension d von V auch die **Dimension von A.**

Wenn $\{b_1, \ldots, b_d\}$ eine Basis von V ist, so ist

$$\varphi : V \longrightarrow K^d, \ \sum \lambda_i b_i \mapsto \sum \lambda_i e_i$$

ein (ziemlich beliebig ausgewählter) Isomorphismus von Vektorräumen.

Wenn $P_0 \in A$ ein beliebiger Punkt ist, so wird durch P_0 und φ eine Affinität

$$\Phi : A \longrightarrow \mathbb{A}^d(K), \quad \Phi(v + P_0) := \varphi(v) + 0$$

definiert. Es ist also jeder affine Raum zu einem affinen Standardraum affin isomorph. Außerdem sind zwei verschiedene Standardräume nicht zueinander affin isomorph.

Insgesamt lassen sich viele Fragen hinsichtlich affiner Abbildungen auf Fragen über die affinen Standardräume zurückführen (so wie wir das bei Vektorräumen auch gemacht hatten).

Wie beim Studium von (Vektorraum-)Endomorphismen ist auch hier wieder die Frage nach invarianten Teilräumen von affinen Abbildungen eines affinen Raumes in sich selbst interessant.

Definition 13.1.7 (invariante Teilräume, Fixpunkte)

Es seien A ein affiner Raum und $\Phi : A \longrightarrow A$ eine affine Abbildung. Dann heißt eine Teilmenge $T \subseteq A$ (Φ-)**invariante Teilmenge**, wenn

$$\Phi(T) \subseteq T$$

gilt (siehe 8.2.2). Ist T gleichzeitig ein affiner Teilraum von A, so heißt T auch ein **invarianter Teilraum**. Ein Element $p \in A$ heißt ein **Fixpunkt** von Φ, wenn

$$\Phi(p) = p,$$

wenn also $\{p\}$ eine (einelementige) Φ-invariante Teilmenge ist.

Es sei

$$\Phi : A \longrightarrow A, \quad \Phi(v + a) = \varphi(v) + \Phi(a)$$

eine affine Abbildung. Wann hat Φ einen Fixpunkt? Genau dann, wenn es einen Vektor $v \in V$ gibt, für den

$$\varphi(v) + \Phi(a) = v + a$$

gilt, und das heißt nicht anderes als

$$\varphi(v) - v = \overrightarrow{\Phi(a)a} \ .$$

Dies hat genau dann eine eindeutig bestimmte Lösung, wenn 1 kein Eigenwert von φ ist. Wenn φ (unter anderem) Eigenwert 1 hat, dann gibt es genau dann Lösungen der Fixpunktgleichung, wenn $\overrightarrow{\Phi(a)a}$ im Bild von $\varphi - \mathrm{Id}_V$ liegt. Die Menge aller Fixpunkte von Φ ist in diesem Fall ein affiner Teilraum der Dimension $\dim(A) - \mathrm{Rang}(\varphi - \mathrm{Id}_V) = \dim(\mathrm{Eig}(\varphi, 1))$. Das ist wieder einmal die Dimensionsformel.

Die Existenz von Fixpunkten ist besonders interessant. Wenn $p \in A$ ein Fixpunkt ist, gilt ja

$$\Phi(v + p) = \varphi(v) + p,$$

und die Abbildung Φ wird „im Wesentlichen“ durch ihren linearen Anteil φ beschrieben.

Beispiel 13.1.8

Es sei $(A, V, +)$ ein affiner Raum über einem Körper K.

a) Für $v \in V$ ist die Abbildung

$$t : A \longrightarrow A, \quad a \mapsto v + a,$$

eine Affinität. Sie hat keinen Fixpunkt, wenn $v \neq 0$.

b) Wenn $\Phi \in \mathrm{Aff}(A)$ eine Affinität endlicher Ordnung e ist, und die Charakteristik von K (siehe 3.1.9) kein Teiler dieser Ordnung, dann hat Φ einen Fixpunkt. Zum Beweis identifizieren wir A mit dem Lieblingsmodell $\widetilde{\mathbb{A}}^n(K)$ und beschreiben Φ durch eine Matrix

$$\widetilde{M} := \begin{pmatrix} M & t \\ 0 & 1 \end{pmatrix}.$$

Dann ist der Vektor

$$p := \frac{1}{e} \cdot \sum_{i=0}^{e-1} \widetilde{M}^i \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} \in \widetilde{\mathbb{A}}^n(K)$$

ein Fixpunkt von $\widetilde{M}$ und repräsentiert damit einen Fixpunkt von Φ.

Dieses Beispiel ist von Bedeutung in der Theorie der kristallographischen Gruppen.

Die Bedingung an die Charakteristik ist vonnöten, damit die Division durch e in K legitim ist. Wenn zum Beispiel $K = \mathbb{F}_2$ der Körper mit 2 Elementen ist und $A = V = \mathbb{F}_2$, dann ist die Bijektion

$$A \ni x \mapsto x + 1 \in A$$

zwar eine Affinität, aber ihre Ordnung ist 2 und damit durch 2 teilbar. Tatsächlich haben wir hier keinen Fixpunkt; wir haben ja jetzt den Fall aus Beispiel a).

Definition 13.1.9 (Affine Basis)

Es sei A ein d-dimensionaler affiner Raum mit Translationsvektorraum V. Eine **affine Basis** von A besteht aus der Wahl eines Punktes $P_0 \in A$ und einer Basis $\{b_1, \ldots, b_d\}$ von V.

Wenn $(P_0; b_1, \ldots, b_d)$ eine affine Basis von A ist, so gilt

$$A = \{\sum_{i=1}^{d} \lambda_i b_i + P_0 \mid \lambda_i \in K\},$$

und die Zuordnung $(\lambda_i)_{1 \le i \le d} \mapsto \sum_{i=1}^{d} \lambda_i b_i + P_0$ ist die Umkehrabbildung zur Abbildung Φ aus Beispiel 13.1.6.

Ein affiner Automorphismus Φ von A lässt sich bezüglich einer affinen Basis durch ein Paar $(M, t) \in \mathrm{GL}_d(K) \times K^d$ beschreiben. Dabei ist M die Abbildungsmatrix des linearen Anteils von Φ, und t ist der Koordinatenvektor des Translationsvektors $\overrightarrow{P_0\Phi(P_0)}$. Das entspricht der Multiplikation mit $\begin{pmatrix} M & t \\ 0 & 1 \end{pmatrix}$ im Modellraum $\widetilde{\mathbb{A}^d}(K)$.

Wenn $(Q_0; c_1, \ldots, c_d)$ eine weitere affine Basis von A ist, so gibt es genau eine Affinität von A, die P_0 auf Q_0 und (für $1 \le i \le d$) auch b_i auf c_i abbildet. Der lineare Anteil dieser Affinität ist der Automorphismus von V, der für $1 \le i \le d$ den Vektor b_i auf c_i abbildet. Der noch fehlende Translationsvektor ist $\overrightarrow{P_0Q_0}$.

Wenn eine Affinität bezüglich einer affinen Basis B von A durch das Paar (M, t) dargestellt wird, so wird es bezüglich der Basis C durch das Paar (N, u) dargestellt, wobei

$$\begin{pmatrix} N & u \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} S & z \\ 0 & 1 \end{pmatrix}^{-1} \cdot \begin{pmatrix} M & t \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} S & z \\ 0 & 1 \end{pmatrix},$$

wenn der Basiswechsel von B zu C durch das Paar (S, z) beschrieben wird. Scheinbar konkreter heißt die letzte Gleichung übrigens:

$$N = S^{-1} \cdot M \cdot S, \quad u = S^{-1} \cdot (Mz + t - z).$$

Definition 13.1.10 (euklidischer Raum)

Wenn A ein affiner Raum mit einem $\mathbb{R}$-Vektorraum V als Translationsvektorraum ist, und wenn V sogar euklidisch ist, dann heißt auch A ein **euklidischer Raum.** Auf A wird dann eine Metrik definiert durch

$$d(P, Q) := \parallel \overrightarrow{PQ} \parallel.$$

Aber Vorsicht: auf A ist kein Skalarprodukt definiert, denn A ist ja kein Vektorraum. Was es hingegen gibt, sind Winkel zwischen Geraden, die sich schneiden. Wenn $G_1 := \mathbb{R}v_1 + P_0, G_2 := \mathbb{R}v_2 + P_0$ zwei affine Geraden mit Schnittpunkt P_0 sind, dann ist der Winkel zwischen G_1 und G_2 definiert durch

$$\angle G_1 G_2 := \angle v_1 v_2.$$

Hierbei muss man wieder ein bisschen aufpassen, denn der Winkel ist nicht vollkommen präzise festgelegt. Ersetzt man v_1 durch $-v_1$, so geht der Winkel über in $\pi - \angle G_1 G_2$. Im Allgemeinen muss dies willkürlich bleiben. In der ebenen Geometrie könnte das Dilemma durch Einführung einer „Orientierung“ behoben werden.

Bemerkung 13.1.11 (einfach transitive Gruppenoperation)

Es sei G eine Gruppe und M eine Menge, auf der G operiert (2.5.1). Wir notieren die Operation als $(g, m) \mapsto gm$. Wie in 13.1.1 gesagt heißt die Operation von G auf M einfach transitiv, wenn M (nicht leer ist und) nur aus einer G-Bahn besteht und wenn noch dazu für jedes $m \in M$ gilt:

$$\mathrm{Stab}_G(m) := \{g \in G \mid gm = m\} = \{e_G\}.$$

Die durch das erste Gleichheitszeichen definierte Untergruppe von G heißt der **Stabilisator** oder auch die **Fixgruppe** von m unter der gegebenen Operation von G. Wir wählen nun ein beliebiges $m_0 \in M$ aus. Dann ist die Operation einfach transitiv, wenn die Abbildung

$$\Phi : G \longrightarrow M, g \mapsto gm_0,$$

eine Bijektion ist. Das ist dasselbe Argument wie in Beispiel 13.1.2 c) im Falle eines affinen Raumes. Es gilt auch hier:

$$\forall g, h \in G : \Phi(gh) = g\Phi(h).$$

Die Gruppenmultiplikation auf G macht aus G offensichtlich eine Menge mit einfach transitiver Operation von G. Die Abbildung Φ identifiziert M mit G und respektiert dabei die G-Operationen auf beiden Mengen. Man sagt auch, Φ sei ein Isomorphismus der G-Mengen G und M. Also gilt ähnlich wie im Falle affiner Räume, dass es bis auf Isomorphie nur eine Menge gibt, auf der G einfach transitiv operiert. Es gibt aber keine natürliche Möglichkeit, zwei solche Mengen miteinander zu identifizieren. Man muss dazu immer einen Punkt aus jeder Menge wählen und dabei vor allen anderen auszeichnen.

Vielleicht ist es nicht ganz unangebracht, hier noch ein Beispiel für diese Situation anzugeben, das eng mit der linearen Algebra verknüpft ist.

Dazu seien V und W zwei zueinander isomorphe Vektorräume über dem Körper K, meinetwegen $V = K^3$ und $W = \{a_0 + a_1X + a_2X^2 \mid a_i \in K\} \subseteq K[X]$. Wir betrachten die Menge

$$M := \{\Phi \in \mathrm{Hom}_K(V, W) \mid \Phi \text{ bijektiv}\}$$

aller Isomorphismen von V nach W. Weiter sei $G := \mathrm{Aut}(W)$. Dann operiert G auf M durch

$$G \times M \ni (\Psi, \Phi) \mapsto \Psi \circ \Phi \in M.$$

Für zwei Isomorphismen $\Phi_1, \Phi_2 \in M$ gilt

$$\Phi_2 = (\Phi_2 \circ \Phi_1^{-1}) \circ \Phi_1, \quad \text{und} \quad \Phi_2 \circ \Phi_1^{-1} \in G.$$

Also operiert G auf M transitiv. Außerdem gilt für $\Psi \in G, \Phi \in M$:

$$\Psi \circ \Phi = \Phi \iff \Psi = \mathrm{Id}_W.$$

Damit ist die Operation sogar einfach transitiv.

Es gibt aber keine natürliche Wahl einer G-verträglichen Bijektion zwischen G und M. Sie hängt ab von der Wahl eines Elementes in M, wenn man stillschweigend das (ja nun wirklich ausgezeichnete) neutrale Element in G als Pendant verwendet.

Deswegen ist es gut, wenn man zwischen G und einer Menge mit einfach transitiver G-Operation unterscheidet.

13.2 Quadriken

Bemerkung 13.2.1 (affine Teilräume - alternative Sichtweise)

Die affinen Räume und ihre Teilräume haben wir jetzt gut verstanden. Nun suchen wir neue Herausforderungen, und um eine sinnvolle neue Frage aufzustellen suchen wir zunächst nach einer etwas anderen Möglichkeit, affine Teilräume zu charakterisieren. Dazu bleiben wir im Standardraum $\mathbb{A}^n(K)$, für andere affine Räume müsste man immer einen Fußpunkt wählen, um analoge Beobachtungen zu machen. Im Standardraum ist dieser Fußpunkt der Nullvektor, im Lieblingsraum $\widetilde{\mathbb{A}^n}(K)$ ist es der Vektor $\binom{0_n}{1}, 0_n \in K^n$.

Eine nichtleere Teilmenge A von K^n ist genau dann ein affiner Teilraum der Dimension d, wenn es $n - d$ linear unabhängige Linearformen $\lambda_1, \ldots, \lambda_{n-d} \in (K^n)^*$ und Elemente $a_1, \ldots, a_{n-d} \in K$ gibt, sodass

$$A = \{v \in K^n \mid \forall i \in \{1, \ldots, n-d\} : \lambda_i(v) = a_i\}.$$

Speziell wird ein affiner Teilraum der Dimension $n-1$ durch eine Gleichung

$$A = \{v \in K^n \mid l_1 v_1 + l_2 v_2 + \cdots + l_n v_n = a\}$$

gegeben, wobei die $l_1, \ldots, l_n \in K$ fest gewählt und nicht alle 0 sind, und auch $a \in K$ fest gewählt ist.

Nun wollen wir wie gesagt vom Fall der affinen Teilräume weg und lassen in der nächsten Generation auch quadratische Terme in so einer Gleichung zu.

Vielleicht sollte man hier schon einmal anmerken, dass bereits der Fall von Termen dritter Potenz viel schwieriger zu behandeln ist. Eine befriedigende Klassifikation wie für Quadriken gibt es hier nicht mehr (oder noch nicht?)!

Definition 13.2.2 (Quadrik)

Es seien K ein Körper und $n \in \mathbb{N}$. Eine **Quadrik im** K^n ist eine Teilmenge $Q \subseteq K^n$ der Form

$$Q := \{v \in K^n \mid F(v) = 0\},$$

wobei $F \in K[X_1, \ldots, X_n]$ ein Polynom der Gestalt

$$F = \sum_{i \leq j} a_{ij} X_i X_j + \sum_i b_i X_i + c$$

ist. Dabei sind die Koeffizienten $a_{ij}, b_i, c \in K$, und nicht alle a_{ij} sind 0.

Streng genommen – und das wird gleich noch relevant für uns – merkt man sich nicht nur die Menge Q, sondern auch das Polynom F bei der Quadrik.

Beispiel 13.2.3 (Kreise, Hyperbeln, Parabeln)

a) Affine Teilräume des K^n sind in gewisser Weise Entartungsfälle von Quadriken, wenn nämlich eine lineare Gleichung $\lambda(x) - a = 0$ quadriert wird.

b) Es sei $m := \binom{x_0}{y_0} \in \mathbb{R}^2, r \in \mathbb{R}$ positiv. Dann ist

$$K(m,r) := \{\binom{x}{y} \in \mathbb{R}^2 \mid (x-x_0)^2 + (y-y_0)^2 = r^2\}$$

eine Quadrik, gegeben durch die Gleichung

$$F(x,y) := x^2 + y^2 - 2(x_0 x + y_0 y) + x_0^2 + y_0^2 - r^2 \stackrel{!}{=} 0.$$

Dies ist der **Kreis mit Radius** r **um den Mittelpunkt** m.

Wenn $\Phi : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ die Affinität $\Phi(v) := rv + m$ ist, dann gilt offensichtlich

$$K(m,r) = \Phi(K(0,1)).$$

Alle Kreise gehen aus $K(0,1)$ durch Affinitäten hervor.

Vorsicht: nicht jede Affinität muss aus $K(0,1)$ wieder einen Kreis machen; es kann auch eine Ellipse daraus werden, was wir noch detaillierter untersuchen werden.

c) Die Menge

$$H := \{\begin{pmatrix} x \\ y \end{pmatrix} \in K^2 \mid xy = 1\}$$

heißt die **(Standard)-Hyperbel** im K^2. Auch hier könnte man direkt eine größere Famile von Quadriken definieren, deren Elemente dann Hyperbeln hießen.

d) Die Menge

$$P := \{\begin{pmatrix} x \\ y \end{pmatrix} \in K^2 \mid y - x^2 = 0\}$$

heißt die **(Standard-)Parabel** im K^2.

Bemerkung 13.2.4 (Matrizenform und Basiswechsel)

a) Es sei

$$F = \sum_{i \leq j} a_{ij} X_i X_j + \sum_i b_i X_i + c$$

ein quadratisches Polynom, das die Quadrik Q im K^n definiert. Dann fassen wir die a_{ij} zu einer Matrix $A \in K^{n \times n}$ zusammen, und die b_i zu einem Vektor $b \in K^n$. Dann gilt für $x \in K^n$:

$$F(x) = x^\top A x + b^\top x + c.$$

Wir wollen in Zukunft die quadratischen Polynome immer so schreiben.

Wenn dabei die Charakteristik von K nicht 2 ist, so dürfen wir hierbei die Matrix A ersetzen durch die symmetrische Matrix $\frac{1}{2}(A + A^\top)$, denn wir haben

$$x^\top A x = (x^\top A x)^\top = x^\top A^\top x.$$

(Vgl. 8.5.10: $A = \frac{1}{2}(A + A^\top) + \frac{1}{2}(A - A^\top)$!)

Abmachung: Wir werden ab jetzt nur den Fall $\mathrm{char}(K) \neq 2$ behandeln und voraussetzen, dass A symmetrisch ist.

b) Wenn eine Affinität $\Phi(x) = Mx + t$ auf dem K^n gegeben ist (mit $M \in \mathrm{GL}(n, K)$ und $t \in K^n$), dann gilt

$$\begin{aligned} F(\Phi(x)) &= (Mx + t)^\top A(Mx + t) + b^\top(Mx + t) + c \\ &= x^\top (M^\top A M)x + \Big(M^\top((A + A^\top)t + b)\Big)^\top x + t^\top A t + b^\top t + c. \end{aligned}$$

Wir erhalten also ein neues quadratisches Polynom

$$\widetilde{F}(x) := x^\top \widetilde{A} x + \widetilde{b}^\top x + \widetilde{c},$$

wobei hier wegen der (eben abgemachten) Symmetrie von A gilt:

$$\widetilde{A} = M^\top A M, \quad \widetilde{b} = M^\top (2At + b), \quad \widetilde{c} = F(t).$$

c) Wir haben damit eine Formel erhalten, um eine Quadrik bezüglich verschiedener affiner Basen zu beschreiben, und es liegt die Frage auf der Hand, ob es hier eine Möglichkeit gibt, Quadriken durch eine „möglichst einfache“ Gleichung zu beschreiben. Wir stellen erst einmal fest, dass $\widetilde{A}$ aus A durch denselben Typ von Basiswechsel hervorgeht, wie wir das bei Paarungen beobachtet hatten (siehe 10.1.4). $\widetilde{b}$ entsteht aus $M^\top b$ durch „lineare Störung“ um den Term $2M^\top At$, und der „konstante Term“ $\widetilde{c}$ ist der Wert von F beim Fußpunkt t der durch Φ aus der affinen Standardbasis entstehenden affinen Basis (t; Spalten von M).

Definition 13.2.5 (Äquivalenz von Quadriken)

a) Es seien $F(X) = X^\top AX + b^\top X + c$ und $\widetilde{F}(X) = X^\top \widetilde{A} X + \widetilde{b}^\top X + \widetilde{c}$ zwei quadratische Polynome in den Unbestimmten $X_1, \ldots, X_n$ (die wir im Spaltenvektor $X \in K[X_1, \ldots, X_n]^n$ zusammenfassen). Dann heißen F und $\widetilde{F}$ **äquivalent**, wenn es eine Affinität $\Phi(X) = MX + t$ auf dem $\mathbb{A}^n(K)$ und eine Einheit $e \in K^\times$ gibt, sodass

$$\widetilde{F}(X) = e \cdot F(MX + t)$$

gilt. Das bedeutet nach dem Vorhergehenden einfach, dass sich $\widetilde{A}, \widetilde{b}$ und $\widetilde{c}$ (bis auf den gemeinsamen Faktor e) aus A, b, c wie in 13.2.4 berechnen lassen. Der Faktor e ändert natürlich die Nullstellenmenge des Polynoms nicht.

b) Eine (durch F gegebene) Quadrik ist eine **Mittelpunktsquadrik**, wenn es ein zu F äquivalentes Polynom $\widetilde{F}$ gibt, für das $\widetilde{b} = 0$ gilt. Nach 13.2.4 ist das gleichbedeutend damit, dass b im Bild von $2A$ liegt (denn dann kann man als t ein Urbild von $-b$ unter $2A$ wählen), und in Charakteristik $\neq 2$ heißt das: b liegt im Bild von A.

Bemerkung 13.2.6 Anstelle der hier verwendeten Äquivalenz der Polynome könnte man natürlich auch fordern, dass die Nullstellenmenge von F durch die Affinität mit der Nullstellenmenge von $\widetilde{F}$ identifiziert wird. Dies führt zu einer anderen Klassifikation der Quadriken, denn die Polynome lassen sich im Allgemeinen nicht (auch nicht bis auf einen von Null verschiedenen Vorfaktor) aus ihren Nullstellenmengen rekonstruieren. Zum Beispiel sind die reellen Punktmengen

$$\{\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \mid x^2 + 1 = 0\} \quad \text{und} \quad \{\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \mid x^2 + y^2 + 1 = 0\}$$

beide leer, aber die definierenden Polynome sind nicht äquivalent. Im zweiten Fall ist ja der quadratische Summand regulär, im ersten nicht.

Wenn man hingegen die Teilmengen von $\mathbb{C}^2$ ansieht, die durch dieselben Bedingungen gegeben werden, so sind diese beide nicht leer und lassen sich auch nicht durch eine Affinität ineinander überführen. Aus diesem Grund ist es sinnvoll, die Polynome an die erste Stelle zu rücken, und nicht die Punktmengen.

Mehr dazu lernt man in der algebraischen Geometrie.

Beispiel 13.2.7 (regulärer quadratischer Anteil)

a) Wenn der quadratische Anteil A der Quadrik Q regulär ist (also vollen Rang hat), dann ist die Quadrik immer eine Mittelpunktsquadrik. Sie ist dann nach geeignetem Basiswechsel immer von der Form $x^\top Ax + c = 0$. Bei der weiteren Suche nach äquivalenten Quadriken sollte man nur noch rein lineare Basiswechsel benutzen, da ein Translationsterm $t \neq 0$ immer die Mittelpunktsform kaputtmachen würde.

Der Name Mittelpunktsquadrik bringt zum Ausdruck, dass (nach geeigneter Wahl einer affinen Basis) $x \in Q \iff -x \in Q$.

b) Die Parabel aus 13.2.3 ist gegeben durch

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

Das ist also keine Mittelpunktsquadrik (wie man auch „sieht“).

Kreise, Hyperbeln und Ellipsen sind Mittelpunktsquadriken.

Bemerkung 13.2.8 (das andere Standardmodell)

Der Begriff der Quadrik ist ein affiner Begriff. In 13.1.2 hatte ich das Modell $\widetilde{\mathbb{A}}^n(K)$ für den affinen Standardraum propagiert. Wie sieht eine Quadrik hier aus?

Es sei $F(X) = X^\top AX + b^\top X + c$ wie gehabt, mit symmetrischer Matrix A. Wir setzen $\hat{b} := \frac{1}{2} \cdot b$ und definieren eine neue Matrix

$$\hat{A} := \begin{pmatrix} A & \hat{b} \\ \hat{b}^\top & c \end{pmatrix} \in K^{(n+1)\times(n+1)}.$$

Dann ist die durch F definierte Quadrik gleich der Menge

$$Q = \{x \in K^n \mid (x^\top|1)\hat{A} \begin{pmatrix} x \\ 1 \end{pmatrix} = 0\}.$$

Damit erhalten wir neben dem Rang von A eine weitere Invariante zur Beschreibung der Äquivalenzklasse der Quadrik: den Rang von $\hat{A}$.

Stets gilt $\mathrm{Rang}(A) \leq \mathrm{Rang}(\hat{A}) \leq \mathrm{Rang}(A) + 2$.

Wir schreiben $\rho(F) := \mathrm{Rang}(A)$, $\hat{\rho}(F) := \mathrm{Rang}(\hat{A})$. Die Invarianz heißt hier, dass weder ρ noch $\hat{\rho}$ sich beim Übergang zu einem äquivalenten Polynom ändern.

Folgerung 13.2.9 (komplexe Klassifikation)

Die Äquivalenzklassen der Quadriken auf dem komplex affinen Raum $\mathbb{A}^n(\mathbb{C})$ sind durch die zwei Invarianten ρ und $\hat{\rho}$ eindeutig festgelegt.

Beweis. Wegen 10.1.10 lässt sich die Matrix A nach Basiswahl durch eine Diagonalmatrix ersetzen. Die von 0 verschiedenen Diagonaleinträge hierbei dürfen noch um quadratische Faktoren abgeändert werden. Da in $\mathbb{C}$ jede Zahl ein Quadrat ist (siehe 3.2.7), dürfen wir ohne Einschränkung

$$A = \begin{pmatrix} I_\rho & 0 \\ 0 & 0 \end{pmatrix}$$

setzen, wobei I_ρ die Einheitsmatrix von Rang ρ ist.

Wenn jetzt die Quadrik eine Mittelpunktsquadrik ist, dann ist $\hat{A}$ von der Form

$$\hat{A} = \begin{pmatrix} A & 0 \\ 0 & c \end{pmatrix},$$

und hierbei darf ein von 0 verschiedenes c durch 1 ersetzt werden. Das liefert zwei Äquivalenzklassen von Mittelpunktsquadriken, und wir haben $\hat{\rho} = \rho$ oder $\hat{\rho} = \rho + 1$.

Wenn die Quadrik keine Mittelpunktsquadrik ist, dann können wir den konstanten Term zu 0 machen (geeignete Translation; ein Mittelpunkt muss ja nicht berücksichtigt werden) und wählen den $\rho + 1$-ten Koordinatenvektor so, dass

$$\hat{A} = \begin{pmatrix} I_\rho & 0 & 0 \\ 0 & 0 & e_1 \\ 0 & e_1^\top & 0 \end{pmatrix},$$

wobei e_1 der erste Standardbasisvektor in $\mathbb{C}^{n-\rho}$ ist. Hier ist $\hat{\rho} = \rho + 2$.

Bei festem ρ werden diese drei Fälle durch $\hat{\rho}$ unterschieden, und sonst treten keine Fälle auf. Im Falle $\rho = n$ kommt auch der dritte Fall selbst nicht vor. ◯

Im reellen Fall kommen zu den beiden Invarianten noch die Signatur der symmetrischen Matrix A hinzu. Diese haben wir in 12.2.8 kennen gelernt.

Satz 13.2.10 (reelle Quadriken)

Die Quadriken im reell-affinen Raum $\mathbb{A}^n(\mathbb{R})$ werden durch die Invarianten $\rho, \hat{\rho}$ und durch die Signatur von A klassifiziert. Genauer gilt das Folgende.

Jede Quadrik Q im $\mathbb{R}^n$ ist äquivalent zu einer der folgenden Quadriken:

- $\sum_{i=1}^{p} x_i^2 - \sum_{i=p+1}^{\rho} x_i^2 = 0$, *falls* $\hat{\rho} = \rho$
- $\sum_{i=1}^{p} x_i^2 - \sum_{i=p+1}^{\rho} x_i^2 - 1 = 0$, *falls* $\hat{\rho} = \rho + 1$
- $\sum_{i=1}^{p} x_i^2 - \sum_{i=p+1}^{\rho} x_i^2 + 2x_{\rho+1} = 0$, *falls* $\hat{\rho} = \rho + 2$

Es ist hier etwas aufwendiger zu sagen, wann zwei solche Quadriken gleich sind. Im ersten und dritten Fall dürfen die Rollen von p und $\rho - p$ vertauscht werden; das entspricht der Multiplikation von F mit der Zahl -1. Im zweiten Fall geht das nur, wenn $p = \frac{\rho}{2}$.

Beweis. Der Beweis geht genauso wie der der komplexen Klassifikation, nur müssen wir jetzt aufpassen, denn nur positive Zahlen sind in $\mathbb{R}$ Quadrate. Daher erhalten wir verschiedene Vorzeichen im quadratischen Anteil.

Dass es keine weiteren Identifikationen gibt als die beschriebenen, folgt aus dem Trägheitssatz von Sylvester, der in 12.2.9 gesehen wurde. ◯

Beispiel 13.2.11 (der Zoo der räumlichen Quadriken)

In Dimension 3 erhalten wir die folgenden Typen von reellen Quadriken mit $\hat{\rho} = 4$:

- $x^2 + y^2 + z^2 = 1$ (Ellipsoid)
- $x^2 + y^2 - z^2 = 1$ (einschaliges Hyperboloid)
- $x^2 - y^2 - z^2 = 1$ (zweischaliges Hyperboloid)
- $-x^2 - y^2 - z^2 = 1$ (leere Menge!)
- $x^2 + y^2 + z = 0$ (elliptisches Paraboloid)
- $x^2 - y^2 + z = 0$ (hyperbolisches Paraboloid)

Das Ellipsoid würden viele von Ihnen wahrscheinlich lieber Sphäre (oder gar Kugeloberfläche) nennen. Dabei verliert man aus den Augen, dass ja affine Koordinatenwechsel gemacht wurden, die metrische Verhältnisse nicht berücksichtigen.

Das zweischalige Hyperboloid hat die zwei Komponenten $x \geq 1$ und $x \leq -1$. Da diese gar nicht miteinander verbunden sind, darf man sich für geometrische Zwecke auf eine von beiden Komponenten beschränken. Diese ist dann ein netter Ausgangspunkt für die **ebene hyperbolische Geometrie**, die ein Pendant zur euklidischen Ebene darstellt. Sie wird studiert in der Differentialgeometrie, und hat große Bedeutung in der Theorie der Riemannschen Flächen und (komplex) algebraischen Kurven. Auf dem Umweg über Modulformen hält sie dann wiederum Einzug in die Zahlentheorie.

Beim hyperbolischen Paraboloid lohnt es sich vielleicht, $s = x - y$ und $t = x + y$ als neue Koordinaten einzuführen. Dann ist die definierende Gleichung

$$st = z,$$

und man sieht die Hyperbeln besser. Diese Hyperbeln „entarten“ für $z = 0$ zu einer Geradenkreuzung.

Bemerkung 13.2.12 (euklidische Typen)

Nun betrachten wir Quadriken in einem euklidischen affinen Raum und wollen nur affine Isometrien als Transformationen zulassen. Jede reelle symmetrische Matrix lässt sich mithilfe eines orthogonalen Basiswechsels diagonalisieren (siehe 12.2.6). Da für orthogonale Matrizen $M \in \mathrm{O}(n)$ stets nach Definition $M^{-1} = M^\top$ gilt, fallen hier (zufälligerweise!) die Basiswechsel der quadratischen Form $x^\top Ax$ und des durch A gegebenen Endomorphismus $x \mapsto Ax$ mit der Basiswechselmatrix M zusammen. Wir können den Spektralsatz benutzen und eine Orthonormalbasis von $\mathbb{R}^n$ wählen, die gleichzeitig orthogonal für die in der Quadrik benutzte quadratische Form ist. Nur dürfen wir hier nicht mehr die Diagonalelemente durch ihren Betrag teilen, da dies nicht mehr durch eine Isometrie gemacht werden könnte.

Wir erhalten etwas allgemeiner als im letzten Satz die Typen

$$\sum_{i=1}^{\rho} \lambda_i x_i^2 = \begin{cases} 0 \\ 1 \\ x_{\rho+1} \end{cases}$$

mit $\lambda_1, \ldots, \lambda_\rho \neq 0$ und verzichten auf eine allgemeine Typisierung.

Bemerkung 13.2.13 (ein „Modulraum“)

Im euklidischen Standardraum ist eine Ellipse eine Quadrik, deren affine Normalform die folgende ist:

$$E := \{(x\ y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}.$$

Die euklidische Normalform ist also von der Gestalt

$$\tilde{E} := \{(x\ y) \in \mathbb{R}^2 \mid \lambda x^2 + \mu y^2 = 1\},\ \lambda, \mu > 0 \text{ geeignet.}$$

Da wir notfalls die x- und die y-Achse vertauschen dürfen, können wir uns auf den Fall $0 < \lambda \leq \mu$ beschränken. Für zwei verschiedene Wahlen dieser Parameter sind die Ellipsen nicht kongruent (gehen also nicht durch eine euklidische Isometrie ineinander über): die Gesamtheit aller Kongruenzklassen von Ellipsen wird bijektiv durch die Menge

$$M := \{(\lambda\ \mu) \mid 0 < \lambda \leq \mu\}$$

parametrisiert. Diese Menge nennen wir den **Modulraum** der euklidischen Ellipsen.

Für eine typische Ellipse $\tilde{E}$ hat die Isometriegruppe $\mathrm{Iso}(\tilde{E})$ vier Elemente (Identität, Drehung um $180°$ und zwei Geradenspiegelungen). Dies ändert sich am Rande

$$\partial M = \{(\lambda\ \mu) \mid [0 = \lambda \leq \mu] \vee [\lambda = \mu]\}$$

des Modulraums M. Auf der einen Seite haben wir den Fall $\lambda = 0$, der für $\mu = 0$ die leere Menge als „Ellipse“ liefert und für $\mu > 0$ ein Paar paralleler Geraden $(y = \pm 1/\sqrt{\mu})$. Auf der anderen Seite erhalten wir für $\mu = \lambda$ Kreise. Beide Entartungsfälle haben eine andere Isometriegruppe als die eben beschriebene.

Als dritten Entartungsfall sollte man noch den Fall „$\mu = \infty$“ zulassen, der geometrisch dem Fall entspricht, dass die kürzere der beiden Ellipsenachsen zu einem Punkt degeneriert. Hier erhalten wir eine Strecke als entartete Ellipse, ihre Isometriegruppe hat nur zwei Elemente.

Modulräume von Klassen geometrischer Objekte sind in der heutigen Geometrie von übergeordneter Wichtigkeit. Sie liefern zum Einen Information über die untersuchten Objekte und sind zum Anderen selbst interessante geometrische Objekte, für deren Geometrie ihre Rolle als Modulraum entscheidend ist.

Bemerkung 13.2.14 (beliebiger Grundkörper)

Wenn K ein beliebiger Grundkörper mit Charakteristik $\neq 2$ ist, dann sieht eine erste Klassifikation der Quadriken genauso aus wie im euklidischen Fall. Hier werden allerdings noch viele Isomorphieklassen von Quadriken mehrfach aufgeführt. Die Frage nach der endgültigen Klassifikation der quadratischen Formen ist eine oftmals interessante Frage nach arithmetischen Eigenschaften des Körpers K, und gerade im Fall $K = \mathbb{Q}$ gibt es schon lange eine sehr schöne und vollständige Theorie zur Klassifikation quadratischer Formen. Auch endliche Körper sind sehr gut studiert; sie führen zu einer ähnlich einfachen Liste wie die komplexen Zahlen.

Es hat sich übrigens als hilfreich herausgestellt, solche Klassifikationsfragen wie die hier gestellte erst einmal über algebraisch abgeschlossenen Körpern L zu diskutieren (hier sind die Ergebnisse oft sehr befriedigend) und im Nachhinein zu versuchen, sie auf Teilkörper $K \subseteq L$ „herunter zu kochen“. Dann gelangt man wieder zur Frage nach „K-Formen“ von Objekten, die man zunächst über L kennt, so wie das bei den Algebren am Ende von Kapitel 10 schon kurz angesprochen wurde.

Kapitel 14

Listen

14.1 Stichwortverzeichnis

Index