Eidgenössische Technische Hochschule Zürich

Lineare Algebra I/II

Begleitskript 2020/2021

Menny Akka Ginosar

21. August 2021

Inhaltsverzeichnis

Kapitel 0

Einführung

In dieser Einführung werden wir (mittels linearer Algebra) ein Problem lösen, das mich als Jugendlicher genervt hat.

In der Mittelschule haben wir Folgen kennengelernt. Es gibt die arithmetische Folge: Für zwei Zahlen a, d definiert man

$$\begin{cases} a_0 & = a \\ a_n & = a_{n-1} + d, \quad n \geq 1 \end{cases}.$$

Hier kann man fast problemlos eine Formel für das n-te Folgenglied finden. Konkret heisst das, dass wir einen Ausdruck für das n-te Folgenglied finden, der nur von n abhängt. Der Ausdruck ist gegeben durch $a_n = a + nd$ für $n \geq 0$ und wenn wir wollten, könnten wir problemlos a_{12345} auch ohne Taschenrechner in weniger als einer halben Minute berechnen. Eine ähnliche Geschichte haben wir in der Mittelschule mit der geometrischen Folge gehabt: Seien a, q zwei Zahlen, dann definieren wir

$$\begin{cases} a_0 & = a \\ a_n & = a_{n-1}q, \quad n \geq 1 \end{cases}.$$

In diesem Fall ist eine Formel für a_n mit n als Variable durch $a_n = aq^n$ für $n \geq 0$ gegeben.

Für mich hat der interessante Teil des Stoffs hier aufgehört. Man „beweist“ dann in der Regel viele Identitäten, die fast tautologisch sind, betrachtet die Summenfolge (das ist vielleicht doch interessant). Man kann jedoch viele interessante Fragen zum Beispiel über arithmetische Folgen stellen. Hier sind einige:

1. Sei $a_n = a + nd$ eine arithmetische Folge. Enthält die Folge unendlich viele Primzahlen? Das heisst, gibt es unendlich viele $n \in \mathbb{N}$, so dass a_n eine Primzahl ist?

   Angenommen a, d sind nicht teilerfremd (d.h. es gibt $l \geq 1$, so dass l sowohl a als auch d teilt), dann kann die Folge höchstens eine Primzahl enthalten.

Theorem 0.0.1 (Dirichlet, 1837)**.** *Seien $a, d \in \mathbb{N}, a \geq 1, d > 1$ teilerfremd. Dann enthält die Folge $a_n = a + nd$ unendlich viele Primzahlen*[1]*.*

Für den Beweis des Satzes brauchen Sie mehr oder weniger einen Bachelor in Mathematik (dies könnte zum Beispiel das Thema Ihrer Bachelorarbeit sein). Falls Sie nicht bis dahin warten wollen, dann können Sie hier [13] anfangen.

2. Wegen Dirichlets Theorem (Theorem 0.0.1) könnten wir uns fragen, wann die erste Primzahl in einer arithmetischen Folge vorkommt. Im Jahre 1944 hat Yuri Vladimirovich Linnik, ein Mathematiker, der meine eigene Forschung stark beeinflusst, bewiesen, dass es Konstanten $c, L > 0$ gibt, so dass die erste Primzahl in einer arithmetischen Folge $a + nd$ mit a, d teilerfremd (und $1 \leq a < d$) kleiner als cd^L ist. Nehmen Sie sich ein bisschen Zeit, um diese Aussage richtig zu verstehen. Hier ist eine Vermutung[2], dass obiges Theorem von Linnik mit $c = 1, L = 2$ gilt. Mehr dazu hier.

3. Jetzt können wir uns fragen, ob wir eine „Kette“ von Primzahlen in irgendeiner arithmetischen Folge finden können. Was genau damit gemeint ist, können sie hier nachlesen.

Diese Fragen und Theoreme sind schön, aber deren Lösungen sind nicht direkt mit linearer Algebra verbunden. Um näher zur linearen Algebra zu kommen, betrachten wir eine andere berühmte Folge, die Fibonacci-Folge. Diese ist folgendermassen definiert:

$$\begin{cases} a_0 &= 0 \\ a_1 &= 1 \\ a_n &= a_{n-1} + a_{n-2}, \quad n \geq 2 \end{cases} \tag{0.1}$$

Wenn wir diese mit den oben definierten Folgen vergleichen, stellt sich sofort die Frage, ob wir einen Ausdruck/eine Formel für das n-te Glied der Fibonacci-Folge finden können. Es ist eigentlich nicht offensichtlich, dass es so eine Formel überhaupt gibt, aber als Jugendlicher habe ich gehört, dass es eine Formel für das n-te Glied der Fibonacci-Folge geben soll. Es konnte mir jedoch niemand sagen, wie diese Formel aussieht oder wie man sie findet (in diesen uralten Zeiten gab es noch kein Wikipedia…). Wir werden jetzt diese Formel zusammen finden und dabei fast allen Begriffen und Themen von diesem Kurs (zumindest vom ersten Semester) begegnen. Der erste Schritt in diesem Abenteuer ist sehr unintuitv; wir verkomplizieren die Sache. Wir haben keine Ahnung

[1] Zum Beispiel gibt es unendlich viele Primzahlen der Form $1 + 4k$ für $k \in \mathbb{N}$ und unendlich viele der Form $3 + 4k$ für $k \in \mathbb{N}$.

[2] Eine Vermutung ist eine Aussage, von welcher die mathematische Gemeinschaft glaubt, dass sie richtig ist, aber die bis anhin noch niemand beweisen konnte.

wie wir dieses Problem lösen können und betrachten dazu noch unendlich viele (verwandte) Probleme, die wir ebenfalls nicht lösen können... ziemlich verrückt! Um diese neuen Probleme einführen zu können, entwickeln wir ein bisschen „Sprache". Das heisst, wir führen einige Begriffe und Definitionen ein. Wir werden jetzt Folgen als Objekte betrachten und sie mit schönen kalligrafischen Buchstaben bezeichnen. Zum Beispiel schreiben wir: Sei $\mathcal{F}$ die Fibonacci-Folge, die in (0.1) definiert ist.

Definition 0.0.2. Wir sagen, dass wir eine Folge *gut kennen*, wenn wir eine Formel für ihr n-tes Folgenglied gefunden haben.

Zum Beispiel kennen wir alle arithmetischen und geometrischen Folgen gut. Wir können jetzt unser Ziel so ausdrücken: Wir möchten gerne $\mathcal{F}$ gut kennen. Wie versprochen, machen wir die Sache komplizierter:

Definition 0.0.3. Seien $a, b \in \mathbb{R}$ zwei reelle Zahlen. Wir definieren die Folge $\mathcal{F}_{a,b}$ mittels der Rekursion

$$\begin{cases} a_0 = a \\ a_1 = b \\ a_n = a_{n-1} + a_{n-2}, \text{ für } n \geq 2 \end{cases}$$

Zum Beispiel gilt $\mathcal{F} = \mathcal{F}_{0,1}$. Also können wir unser Ziel so ausdrücken: Wir möchten $\mathcal{F}_{0,1}$ gut kennen. Der Schlüssel für die Lösung ist nun das verrückte neue Ziel:

Neues Ziel: Wir möchten $\mathcal{F}_{a,b}$ für alle $a, b \in \mathbb{R}$ gut kennen.

Wie kann es sein, dass wir mehr Hoffnung für dieses neue Ziel haben als für das ursprüngliche Ziel? Zunächst klingt das verrückt, da das neue Ziel so aussieht wie unendlich viele Varianten des alten Problems. Das neue Problem ist ein Raum von Problemen und hier liegt eigentlich der Hund begraben. Dieser Raum hat eine gewisse Struktur und wir können diese Struktur ausnutzen. Bevor wir anfangen, geben wir noch eine Definition an:

Definition 0.0.4. Eine Folge $\mathcal{A}$ heisst eine *Fibonacci-Folge*, wenn es $a, b \in \mathbb{R}$ gibt, so dass $\mathcal{A} = \mathcal{F}_{a,b}$. Den Raum aller Fibonacci-Folgen nennen wir **Fib**.

Übung 0.0.5. Zeigen Sie, dass eine Folge $(a_0, a_1, \ldots)$ eine Fibonacci-Folge ist genau dann, wenn $a_n = a_{n-1} + a_{n-2}$ für alle $n \geq 2$.

Es gilt

$$\mathbf{Fib} = \{\mathcal{F}_{a,b} \mid a, b \in \mathbb{R}\}$$

und unser Ziel ist es, alle Elemente von **Fib** gut zu kennen.

0.1 Welche Struktur hat Fib?

Wir nehmen zwei Folgen $\mathcal{F}_1 = (a_0, a_1, a_2, a_3, \ldots)$, $\mathcal{F}_2 = (b_0, b_1, b_2, b_3, \ldots)$ und formen ihre komponentenweise Addition

$$\mathcal{F}_1 + \mathcal{F}_2 := (a_0 + b_0, a_1 + b_1, a_2 + b_2, a_3 + b_3, \ldots).$$

Behauptung: Seien $\mathcal{F}_1$ und $\mathcal{F}_2$ zwei Fibonacci-Folgen, dann ist auch $\mathcal{F}_1 + \mathcal{F}_2$ eine Fibonacci Folge. Wir bezeichnen die Glieder von $\mathcal{F}_1 + \mathcal{F}_2$ als $\mathcal{F}_1 + \mathcal{F}_2 = (c_0, c_1, c_2, c_3, \ldots)$. Laut Übung 0.0.5 müssen wir lediglich zeigen, dass

$$c_n = c_{n-1} + c_{n-2}$$

für alle $n \geq 2$ gilt. Beweisen wir es: Es gilt

$$c_n = a_n + b_n$$

und da $\mathcal{F}_1, \mathcal{F}_2 \in \mathbf{Fib}$ haben wir

$$a_n + b_n = (a_{n-1} + a_{n-2}) + (b_{n-1} + b_{n-2}) = (a_{n-1} + b_{n-1}) + (a_{n-2} + b_{n-2}) = c_{n-1} + c_{n-2}.$$

Daher gilt in der Tat $c_n = c_{n-1} + c_{n-2}$.

Übung 0.1.1. Mit der Notation von Definition 0.0.3 erklären Sie für sich selbst, dass das obige Argument Folgendes zeigt:

$$\mathcal{F}_{a,b} + \mathcal{F}_{c,d} = \mathcal{F}_{a+c,b+d}.$$

Kurz gesagt, können wir zwei Elemente von **Fib** addieren. Der Fibonacci-Folgen Raum **Fib** hat also eine Addition. Ausser der Addition haben wir noch eine andere Operation in **Fib**: eine Multiplikation mit einem Skalar. In diesem Zusammenhang ist „Skalar“ einfach ein komischer Name für eine reelle Zahl. Später in diesem Kurs möchten wir andere „Typen“ von Zahlen betrachten, die einen sogenannten Körper formen (vielleicht haben Sie schon vom Körper der reellen/komplexen Zahlen oder von endlichen Körpern gehört) und ein Skalar ist dann einfach ein Element eines Körpers. Im Moment ist jedenfalls ein Skalar einfach ein Synonym für eine reelle Zahl. Also, sei $\alpha \in \mathbb{R}$ ein Skalar und $\mathcal{A} = (a_0, a_1, a_2, \ldots)$. Wir definieren die Multiplikation von $\mathcal{A}$ mit dem Skalar $\alpha \in \mathbb{R}$ als

$$\alpha\mathcal{A} := (\alpha a_0, \alpha a_1, \alpha a_2, \ldots).$$

Übung 0.1.2. Argumentieren sie ähnlich wie im Fall der Addition, um folgende Behauptung zu zeigen: Sei $\alpha \in \mathbb{R}$ und $\mathcal{A} \in \mathbf{Fib}$. Dann gilt auch $\alpha\mathcal{A} \in \mathbf{Fib}$. Zeigen Sie des

Weiteren, dass $\alpha\mathcal{F}_{a,b} = \mathcal{F}_{\alpha a,\alpha b}$.

Wir haben also gesehen, dass der Raum **Fib** Addition und Multiplikation mit einem Skalar hat. Wie hilft uns das, unsere Aufgabe $\mathcal{F}_{0,1}$ gut zu kennen, zu lösen? Geduld bringt Rosen...

Bemerken Sie zuerst das Folgende: Wenn man $\mathcal{F}_1 = (a_0, a_1, \ldots)$ und $\mathcal{F}_2 = (b_0, b_1, \ldots)$ gut kennt, dann kennt man auch $\mathcal{F}_1 + \mathcal{F}_2$ gut. In der Tat, wenn man eine Formel für a_n und b_n hat, dann hat man auch eine Formel für das n-te Folgenglied. Nämlich, wenn $a_n = f(n)$ und $b_n = g(n)$ Formeln für das n-Glied sind, dann ist $f(n)+g(n)$ eine Formel für $\mathcal{F}_1 + \mathcal{F}_2$. Ähnlicherweise, wenn man eine Folge $\mathcal{A} \in$ **Fib** gut kennt, dann kennt man auch $\alpha\mathcal{A}$ für alle $\alpha \in \mathbb{R}$ gut. Allgemeiner, zeigen Sie:

Übung 0.1.3. Wenn wir $\mathcal{F}_1, \ldots, \mathcal{F}_k$ gut kennen, dann kennen wir

$$\alpha_1\mathcal{F}_1 + \ldots + \alpha_k\mathcal{F}_k \tag{0.2}$$

für alle $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$ gut.

Ausdrücke der Form (0.2) heissen lineare Kombinationen von Folgen.

Anders gesagt, kann man mittels der Struktur von **Fib** das Wissen über einige Elemente von **Fib** auf andere Elemente von **Fib** übertragen.

Das ist alles sehr schön, aber es gibt kein einziges Element von **Fib**, das wir gut kennen! Das ist eigentlich nicht wahr. Es gibt doch ein Element von **Fib**, das wir gut kennen.

Übung 0.1.4. Finden Sie es, bevor Sie weiter lesen.

Die Fibonacci-Folge $\mathcal{F}_{0,0}$ kennen wir sehr gut: Wenn $\mathcal{F}_{0,0} = (a_0, a_1, \ldots)$, dann ist $a_n = 0$ für alle $n \in \mathbb{N} \cup \{0\}$. Heisst das, dass wir jetzt mittels der Struktur von **Fib** viele andere Folgen gut kennen? Leider nicht, denn $\mathcal{F}_{0,0} + \mathcal{F}_{0,0}$ oder $\alpha\mathcal{F}_{0,0}$ ergeben leider keine neuen Folgen... Wir bekommen einfach $\mathcal{F}_{0,0}$ wieder und wieder. Also, wenn wir nur $\mathcal{F}_{0,0}$ gut kennen, können wir mittels Addition und Skalarmultiplikation nicht weiterkommen. Wie viele Folgen sollten wir kennen, um alle Elemente von **Fib** gut zu kennen?

Nehmen wir an, dass wir eine Fibonacci-Folge $\mathcal{F}_{a,b}$ mit a und b beide nicht Null gut kennen. Dann kennen wir mittels Skalarmultiplikation $\mathcal{F}_{\alpha a,\alpha b}$ für alle $\alpha \in \mathbb{R}$ gut.

Übung 0.1.5. Zeigen Sie, dass die Menge $\{\mathcal{F}_{\alpha a,\alpha b} \mid \alpha \in \mathbb{R}\} \subseteq$ **Fib** unter der Annahme $(a,b) \neq (0,0)$ unendlich viele Elemente enthält, aber trotzdem niemals gleich **Fib** ist.

Diese Übung zeigt, dass wir zumindest zwei Fibonacci-Folgen gut kennen müssen, um alle Elemente von **Fib** gut zu kennen. Existieren also zwei Folgen in **Fib**, die wir gut kennen und wodurch wir dann alle Elemente von **Fib** gut kennen? Könnte man also zwei Folgen $\mathcal{A}, \mathcal{B} \in \mathbf{Fib}$ finden, so dass jede Folge eine lineare Kombination von $\mathcal{A}$ und $\mathcal{B}$ ist?

Wenn wir zum Beispiel $\mathcal{F}_{0,1}$ und $\mathcal{F}_{1,0}$ gut kennen, dann kennen wir alle Elemente von **Fib** gut: Wir können ein allgemeines Element $\mathcal{F}_{a,b}$ von **Fib** folgendermassen schreiben

$$\mathcal{F}_{a,b} = a\mathcal{F}_{1,0} + b\mathcal{F}_{0,1}$$

und daraus folgt eine Formel für $\mathcal{F}_{a,b}$ aus Formeln für $\mathcal{F}_{0,1}$ und $\mathcal{F}_{1,0}$.

Dies ist nochmals sehr schön, aber wir sind wieder bei unserem ursprünglichen Problem angelangt! Wir kennen $\mathcal{F}_{0,1}$ noch nicht gut! Im nächsten Abschnitt werden wir zwei andere Folgen gut kennenlernen und mit diesen kann man auch jedes andere Element von **Fib** mit Skalarmultiplikation und Addition erreichen.

Bemerkung 0.1.6. Die Tatsachen, dass

(1) zwei Folgen reichen,

(2) eine Folge nicht reicht (Übung 0.1.5),

(3) in jeder Menge von drei Folgen es eine Folge gibt, die wir weglassen können ohne die Menge der „erreichbaren Folgen“ zu ändern,

sind verbunden mit der Aussage, dass der Raum **Fib** Dimension 2 hat. Das ist keine Überraschung. Dimension ist fast gleichbedeutend mit der Anzahl „Freiheitsgraden“ des Raums. Überlegen Sie sich, wieso **Fib** genau 2 Freiheitsgrade von reellen Zahlen hat. In anderen Worten, **Fib** ist eine Ebene, in der jeder Punkt eine Folge darstellt.

0.2 Vorkenntnisse und Symmetrie

Wie kommt man auf eine neue Idee für die Lösung eines Problems? Normalerweise durch die Nutzung von Vorkenntnissen, die für die Lösung relevant sind oder durch das Erkennen, dass das Problem eine gewisse Symmetrie hat, die wir benutzen können. Vielleicht denken Sie, dass unser Problem gar keine Symmetrie enthält. Immerhin ist dies kein geometrisches Problem...

0.2.1 Vorkenntnisse

Bevor wir erklären, welche Symmetrie dieses Problem dennoch geniesst und wieso diese Symmetrie der Schlüssel zur Lösung ist, benutzen wir unsere Vorkenntnisse um Fibonacci-Folgen zu finden, die wir gut kennen.

Übung 0.2.1. Zeigen Sie, dass **Fib** keine arithmetische Folge enthält (ausser $\mathcal{F}_{0,0}$).

Die Übung zeigt, dass wir mit arithmetischen Folgen nicht vorankommen können. Wie wäre es mit geometrischen Folgen? Kann eine Folge der Form $(a, aq, aq^2, aq^3, \ldots)$ eine Fibonacci-Folge sein? Der Einfachheit halber versuchen wir zuerst $a = 1$, das heisst mit einer Folge der Form

$$\mathcal{G}_q = (1, q, q^2, \ldots)$$

mit $q \neq 0$. Die Folge $\mathcal{G}_q$ ist eine Fibonacci-Folge, genau dann, wenn

$$q^n = q^{n-1} + q^{n-2} \tag{0.3}$$

für alle $n \geq 2$ gilt. Da $q \neq 0$, können wir (0.3) durch q^{n-2} dividieren und daher ist (0.3) äquivalent zu

$$q^2 = q + 1. \tag{0.4}$$

Dies können wir mittels anderer Vorkenntnisse lösen, der Mitternachtsformel[3]! Gleichung (0.4) gilt genau dann, wenn

$$q = \frac{1 \pm \sqrt{5}}{2}.$$

Es ist üblich die Zahl

$$\frac{1 + \sqrt{5}}{2} \approx 1.618033\ldots \quad \text{(goldener Schnitt)}$$

mit φ zu bezeichnen und die Zahl

$$\frac{1 - \sqrt{5}}{2} \approx -0.618033\ldots \quad \text{(konjugierter goldener Schnitt)}$$

mit ψ. Fassen wir diesen Teil zusammen: die Folgen

$$\mathcal{G}_\varphi = (1, \varphi, \varphi^2, \varphi^3, \ldots) \quad \text{und} \quad \mathcal{G}_\psi = (1, \psi, \psi^2, \psi^3, \ldots)$$

sind beides Elemente von **Fib**, die wir gut kennen! Dies ist wunderbar, aber können wir mittels Addition und Skalarmultiplikation von $\mathcal{G}_\varphi$ und $\mathcal{G}_\psi$, die ursprüngliche Folge $\mathcal{F}_{0,1}$ ausdrücken? Ja!

Übung 0.2.2. Verifizieren Sie, dass

$$\frac{1}{\varphi - \psi}\mathcal{G}_\varphi + \frac{1}{\psi - \varphi}\mathcal{G}_\psi = \mathcal{F}_{0,1}.$$

Bemerkung 0.2.3. Um Übung 0.2.2 zu lösen, müssen Sie wahrscheinlich ein lineares Gleichungssystem lösen. Lineare Gleichungssysteme liegen im Herzen der linearen Algebra.

[3]Endlich eine Motivation zum Lösen einer quadratischen Gleichung!

Eines unserer ersten Themen wird ein Algorithmus sein, um lineare Gleichungssysteme zu lösen, die Gauss-Elimination.

Übung 0.2.2 gibt eine Formel für das n-te Glied von $\mathcal{F}_{0,1} = (a_0, a_1, a_2, \ldots)$:

$$a_n = \frac{1}{\varphi - \psi}\varphi^n + \frac{1}{\psi - \varphi}\psi^n = \frac{\varphi^n - \psi^n}{\varphi - \psi} = \frac{(\frac{1+\sqrt{5}}{2})^n - (\frac{1-\sqrt{5}}{2})^n}{\sqrt{5}}.$$

Wir sind am Ziel angekommen!

Übung 0.2.4. Sei $\mathcal{F}_{0,1} = (F_0, F_1, F_2, \ldots)$. Unter der Annahme, dass $\lim_{n\to\infty} \frac{F_n}{F_{n-1}}$ existiert, berechnen Sie diesen Grenzwert. Dies gibt eine andere Motivation geometrische Folgen mit q als den Wert dieses Grenzwert zu betrachten.

0.2.2 Symmetrie

Man könnte sagen, dass dies nur ein Glückstreffer war. Wie konnten wir wissen, dass geometrische Folgen hilfreich sein würden? Das ist eine berechtigte Frage. Wie vorhin erwähnt, könnte man diese geometrischen Folgen entdecken, wenn man die Symmetrie des Raums **Fib** betrachtet. Was aber bedeutet Symmetrie in diesem Zusammenhang? Wenn X irgendein geometrischer Raum ist, dann ist eine Symmetrie von X eine Abbildung[4]

$$T : X \to X,$$

die alle/einige der geometrischen Eigenschaften von X erhält, wie beispielsweise Distanz, Winkel usw. Wenn $T : X \to X$ eine Symmetrie ist, dann ist die Menge der Fixpunkte

$$\mathrm{Fix}(T) = \{x \in X : T(x) = x\}$$

normalerweise eine interessante Menge zum Betrachten. In anderen Worten ist ein Fixpunkt ein Element, welches auf sich selbst abgebildet wird. Um den Symmetriebegriff auf allgemeinere Räume zu verallgemeinern, könnten wir das Folgende definieren: Sei X ein Raum mit einer gewissen Struktur. Eine Symmetrie von X ist eine Abbildung $T : X \to X$, die die Struktur von X erhält/respektiert.

Wenn $X = \mathbf{Fib}$ bekommen wir das Folgende:

Definition 0.2.5. Eine Abbildung $T : \mathbf{Fib} \to \mathbf{Fib}$ heisst eine Symmetrie von **Fib**, wenn für $\mathcal{A}, \mathcal{B} \in \mathbf{Fib}$ und $\alpha \in \mathbb{R}$ gilt, dass

$$T(\mathcal{A} + \mathcal{B}) = T(\mathcal{A}) + T(\mathcal{B}) \quad \text{und} \quad T(\alpha\mathcal{A}) = \alpha T(\mathcal{A}). \tag{0.5}$$

[4]Eine *Abbildung* ist ein anderer Name für eine Funktion. Das ist eine Regel, die jedem Element von X ein bestimmtes Element von X zuordnet. Dies und andere Grundlagen werden wir zusammen mit der Analysisvorlesung bald sorgfältig definieren.

Die Anforderungen in (0.5) sind das, was wir meinen mit „T respektiert die Struktur von **Fib**“.

Überlegen wir zuerst, welche Abbildungen $T : X \to X$ wir überhaupt kennen. Hier sind drei Beispiele, die nicht so interessant sind:

1. Die Identitäts-Abbildung

$$\begin{aligned} \mathrm{Id} : \mathbf{Fib} &\to \mathbf{Fib} \\ \mathcal{A} &\mapsto \mathcal{A}, \end{aligned}$$

   die „nichts“ macht.

2. Die Skalarmultiplikations-Abbildung: Für $\alpha \in \mathbb{R}$ definieren wir

$$\begin{aligned} \mathrm{M}_\alpha : \mathbf{Fib} &\to \mathbf{Fib} \\ \mathcal{A} &\mapsto \alpha\mathcal{A}. \end{aligned}$$

3. Die Vektoradditions/Folgenadditions-Abbildung: Für $\mathcal{B} \in \mathbf{Fib}$ definieren wir

$$\begin{aligned} \mathrm{A}_\mathcal{B} : \mathbf{Fib} &\to \mathbf{Fib} \\ \mathcal{A} &\mapsto \mathcal{A} + \mathcal{B}. \end{aligned}$$

Übung 0.2.6. Zeigen Sie, dass Id und M_α für alle $\alpha \in \mathbb{R}$ die Anforderungen in (0.5) erfüllen. Zeigen Sie des Weiteren, dass $\mathrm{A}_\mathcal{B}$ für $\mathcal{B} \neq \mathcal{F}_{0,0}$ die Anforderungen in (0.5) **nicht** erfüllt.

Wie gesagt, sind diese Symmetrien nicht besonders interessant, vielleicht weil sie nicht mit der Tatsache verbunden sind, dass es sich bei **Fib** um einen Raum von Folgen handelt. Diese Abbildungen existieren für jeden Raum, der Addition und Skalarmultiplikation hat.

Übung 0.2.7. Bevor Sie weiter lesen, versuchen Sie eine interessante andere Abbildung $T : \mathbf{Fib} \to \mathbf{Fib}$ zu finden, welche berücksichtigt, dass es sich um einen Folgen-Raum handelt? (Hinweis: Ich behaupte: Wenn man $\mathcal{F}_{1,0}$ gut kennt, dann kennt man auch $\mathcal{F}_{0,1}$ gut. Wieso?)

Bei Folgen-Räumen gibt es die Verschiebungsabbildung

$$\begin{aligned} S : \mathbf{Fib} &\to \mathbf{Fib} \\ (a_0, a_1, a_2, \ldots) &\mapsto (a_1, a_2, \ldots). \end{aligned}$$

Übung 0.2.8. (1) Verifizieren Sie, dass $S(\mathcal{A}) \in \mathbf{Fib}$ für $\mathcal{A} \in \mathbf{Fib}$.

(2) Verifizieren Sie, dass S die Anforderungen in (0.5) erfüllt.

Diese Verschiebungsabbildung ist schon eine interessante Symmetrie, die mit Folgen-Räumen verbunden ist. Daher fragen wir uns, welche Fixpunkte S hat:

Übung 0.2.9. Zeigen Sie, dass $\mathcal{F}_{0,0}$ der einzige Fixpunkt von S ist.

Gut, die Fixpunkte von S sind nicht so interessant. Es stellt sich heraus, dass für Abbildungen, die (0.5) erfüllen, die Menge von Fixpunkten häufig nicht interessant ist. Die Forderung $S(\mathcal{A}) = \mathcal{A}$ ist einfach zu einschränkend, um interessante Folgen zu bekommen. Daher betrachtet man eine schwächere Anforderung (die auch mit der Skalarmultiplikation verbunden ist).

Definition 0.2.10. Sei $T : \mathbf{Fib} \to \mathbf{Fib}$ eine Symmetrie. Eine Folge $\mathcal{A} \neq \mathcal{F}_{0,0}$ heisst eine *Eigenfolge* von T, wenn es ein $\alpha \in \mathbb{R}$ gibt, so dass

$$T(\mathcal{A}) = \alpha A.$$

Der Skalar α heisst der *Eigenwert* von $\mathcal{A}$.

Wir möchten also alle Eigenfolgen der Symmetrie S finden[5]. Dafür nehmen wir an, dass $\mathcal{A} = (a_0, a_1, a_2, \ldots) \neq F_{0,0}$ eine Eigenfolge mit Eigenwert α ist. Dann gilt $S(\mathcal{A}) = \alpha A$ oder in anderen Worten

$$(a_1, a_2, a_3, \ldots) = \alpha(a_0, a_1, a_2, \ldots) = (\alpha a_0, \alpha a_1, \alpha a_2, \ldots).$$

Also gilt $a_n = \alpha a_{n-1}$ für alle $n \geq 1$. Daher hat $\mathcal{A}$ die Form

$$\mathcal{A} = (a_0, a_0\alpha, a_0\alpha^2, a_0\alpha^3, \ldots). \tag{0.6}$$

Das heisst, dass $\mathcal{A}$ eine geometrische Folge ist. Welche α (und welche a_0) kommen also in Frage? Dies haben wir schon in Abschnitt 0.2.1 gesehen, aber leiten wir es nochmals her: Aus (0.6) folgt insbesondere, dass

$$a_n = \alpha^n a_0$$

für alle $n \geq 0$. Da $\mathcal{A} \in \mathbf{Fib}$ gilt auch $a_2 = a_1 + a_0$ und zusammen bekommen wir

$$\alpha^2 a_0 = a_2 = a_1 + a_0 = \alpha a_0 + a_0 = (\alpha + 1)a_0. \tag{0.7}$$

Aus $\mathcal{A} \neq \mathcal{F}_{0,0}$ folgt, dass $a_0 \neq 0$. (Wieso? Finden Sie eine überzeugende Erklärung.) Daher ist (0.7) äquivalent zu $\alpha^2 = \alpha + 1$. Sieht das bekannt aus? Dies ist gerade die

[5]Für die fortgeschrittenen Leser bemerken wir das Folgende. Eigenfolgen sind eigentlich Fixpunkte bezüglich der Wirkung von S auf dem projektiven Raum $\mathbb{P}(\mathbf{Fib})$.

Gleichung (0.4). Dies bedeutet: Wenn $\mathcal{A}$ eine Eigenfolge mit Eigenwert α ist, dann gilt $\alpha^2 = \alpha + 1$ beziehungsweise $\alpha = \psi$ oder $\alpha = \varphi$. Das heisst, $\mathcal{A}$ ist eine geometrische Folge mit $\alpha = \varphi$ oder $\alpha = \psi$. Der Einfachheit halber wählen wir $a_0 = 1$. Wir bekommen die zwei Folgen, die wir vorher erraten haben

$$\mathcal{G}_\varphi = (1, \varphi, \varphi^2, \varphi^3, \ldots) \quad \text{und} \quad \mathcal{G}_\psi = (1, \psi, \psi^2, \psi^3, \ldots).$$

Zusammenfassung : Die n-te Fibonacci-Zahl F_n ist durch

$$\frac{\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n}{\sqrt{5}}$$

gegeben.

Allgemeiner zeigen Sie:

Übung 0.2.11. Finden Sie eine Formel für das n-te Glied von $\mathcal{F}_{a,b}$, die von a, b und n abhängt.

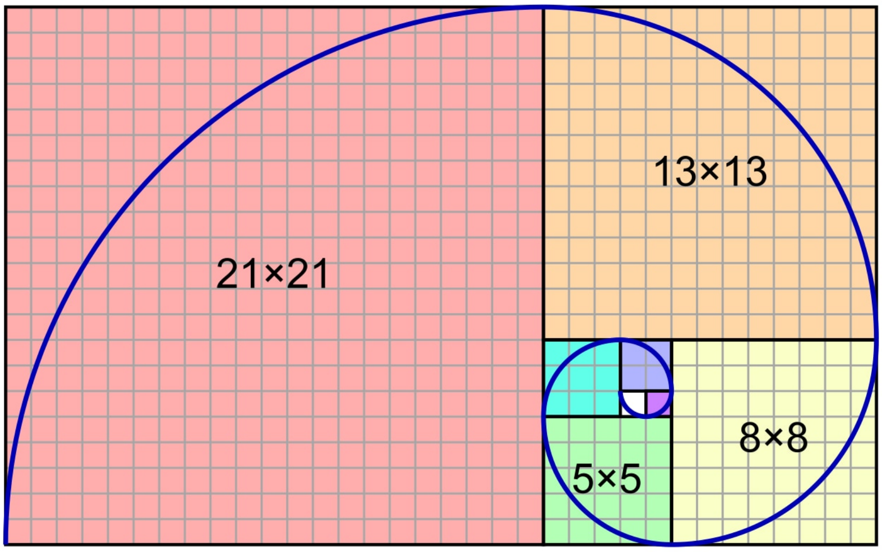
Es gibt noch viele Sachen in diesem Zusammenhang zu sagen, aber wir müssen irgendwann mit dem Kurs anfangen. Hier ist eine kleine Übung als Dessert:

Übung 0.2.12. Zeigen Sie, dass

$$\varphi = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \ldots}}}.$$

Was genau gemeint ist mit dieser Schreibweise und wie dies mit der Bewegung der Planeten zusammenhängt, können Sie mich fragen oder hier [4, §1.1] nachlesen.

Bemerkung 0.2.13. Auf die Idee zu dieser Einführung bin ich gekommen, als ich Übungsstunden an der Hebrew University gehalten habe und gefragt wurde, wieso lineare Algebra Spass machen soll. Dies habe ich durch diese Frage mit der Welt geteilt. Viele der Antworten dort sind sehr interessant und lohnen sich, darüber nachzudenken!



Figur 0.1: Ihr nächstes T-Shirt?

Kapitel 1

Grundlagen

1.1 Mengen und Abbildungen

Dieses Kapitel haben wir sowohl in der Linearen Algebra wie auch in der Analysis besprochen. Sie können dieses Kapitel im Fischer [7] als Wiederholung lesen.

1.2 Gruppen

Teile von diesem Kapitel haben Sie in der Analysis behandelt (die Definition von Gruppen, Ringen und Körper). Für uns ist vor allem die Definition von Körpern relevant. In der Vorlesung Algebra I im dritten Semester werden sie noch mehr Gruppen- und Ring-Theorie behandeln. In diesem Kapitel geben wir nur einige Definitionen und einige Beispiele.

Definition 1.2.1 (Analysis 24. Sept.)**.** Eine *Gruppe* $(G, e, \cdot)$ ist eine Menge G versehen mit einem Element $e \in G$ (das neutrale Element) und einer Verknüpfung $\cdot : G \times G \to G$ mit folgenden Eigenschaften:

a) Für alle $a, b, c \in G$ gilt $(a \cdot b) \cdot c = a \cdot (b \cdot c)$. ($\cdot$ ist assoziativ)

b) Für alle $g \in G$ gilt $e \cdot g = g \cdot e = g$. (daher der Name „neutrales Element“)

c) Zu jedem $g \in G$ gibt es ein $g' \in G$ mit $g \cdot g' = g' \cdot g = e$. ($g'$ nennen wir ein inverses Element von g)

Beachten Sie, dass a) der Eigenschaft G1 im Fischer [7] entspricht und b), c) entsprechen G2. Wir schreiben oft zur Vereinfachung ab statt $a \cdot b$ für Elemente $a, b \in G$, falls klar ist, welche Operation gemeint ist.

Folgende Proposition haben Sie auch in der Analysis gesehen (vgl. auch Fischer [7, Kap. 1.2.3]):

Proposition 1.2.2. *1) Das neutrale Element ist eindeutig bestimmt (d.h. falls für $e, e' \in G$ gilt, dass $eg = ge$ und $e'g = ge'$ für alle $g \in G$, dann folgt stets $e = e'$).*

2) Das inverse Element von G ist eindeutig bestimmt. (Beachten Sie, dass wir dadurch a^{-1} für das inverse Element von a schreiben können.)

3) Für alle $a, b \in G$ gilt $(a^{-1})^{-1} = a$ und $(ab)^{-1} = b^{-1}a^{-1}$.

4) Für alle $a, b, c \in G$ gilt $ab = ac$ genau dann wenn $b = c$ und $ba = ca$ genau, dann wenn $b = c$.

Beweis. Für 1) und 2) verweisen wir auf die Analysis Vorlesung (21. Sept.). Wir beweisen jetzt 3):

Beachten Sie, dass $(a^{-1})^{-1}$ das einzige Element $b \in G$ ist mit der Eigenschaft

$$a^{-1}b = ba^{-1} = e.$$

Da a auch diese Eigenschaft hat, folgt nach 2), dass $a = (a^{-1})^{-1}$. Für den zweiten Teil bemerken Sie, dass

$$(ab)(b^{-1}a^{-1}) = a(bb^{-1})a^{-1} = aea^{-1} = aa^{-1} = e$$

und analog folgt $(b^{-1}a^{-1})(ab) = e$. Nach 2) folgt wiederum, dass $b^{-1}a^{-1} = (ab)^{-1}$. Teil 4) der Proposition beweisen Sie in Serie 4. □

Definition 1.2.3. Eine Gruppe $(G, e, \cdot)$ heisst *abelsch* oder *kommutativ*, falls $a \cdot b = b \cdot a$ für alle $a, b \in G$ gilt.

Bemerkung 1.2.4. Wenn eine Gruppe G abelsch[1] ist, dann verwenden wir fast immer die additive Schreibweise und schreiben $(G, e, +)$ für die Gruppe. Also schreiben wir insbesondere $g_1 + g_2$ statt $g_1 g_2$ und $-g$ statt g^{-1} für $g_1, g_2, g \in G$.

Beispiel 1.2.5. Wir betrachten einige Beispiele von Gruppen.

- $(\mathbb{Z}, 0, +)$ ist eine Gruppe.
- $(\mathbb{N} \cup \{0\}, 0, +)$ ist keine Gruppe, da jedes $n \in \mathbb{N}$ kein inverses Element hat.
- $(\mathbb{Q}, 0, +)$ ist eine Gruppe.
- Allgemeiner ist $(K, 0, +)$ für jeden Körper K eine Gruppe. (Die Definition eines Körpers folgt noch.)

[1]Der Begriff „abelsch“ ist zurückzuführen auf den Mathematiker Niels Abel. Dieser Begriff ist das einzige Beispiel eines nach einer Person benannten mathematischen Begriffs, der ohne Grossbuchstabe geschrieben wird. So grundlegend ist dieser Begriff also!

- Sei X eine Menge und $\mathrm{Abb}(X) = \{f : X \to X \mid f \text{ ist bijektiv}\}$. Dann ist $(\mathrm{Abb}(X), \mathrm{Id}_X, \circ)$ eine Gruppe. Hier ist $\circ$ die Verkettung und $\mathrm{Id}_X : X \to X, x \mapsto x$ die Identitätsabbildung. (Dieses Beispiel haben Sie schon in der Analysis gesehen).

- Wenn $X = \{1, \dots, n\}$ ist, dann schreiben wir $S_n = \mathrm{Abb}(X)$ und somit ist $(S_n, \mathrm{Id}_X, \circ)$ nach dem obigen Beispiel eine Gruppe. Diese Gruppe benutzen wir später, wenn wir Determinanten besprechen (Kapitel 3 im Fischer [7]).

1.2.1 Wichtiges Beispiel: Restklassen modulo n

Dieses Beispiel ist zentral für uns. Es wird uns als Beispiel für Gruppen, Ringe und für gewisse n auch für Körper dienen. Um das „Objekt" der Restklassen zu definieren, betrachten wir die folgende Äquivalenzrelation auf $\mathbb{Z}$: Sei $n \in \mathbb{N}$. Wir definieren für $a, b \in \mathbb{Z}$

$$a \sim_n b \iff (b - a) \text{ ist durch } n \text{ teilbar} \iff \exists l \in \mathbb{Z} : (b - a) = ln.$$

Bemerkung 1.2.6. Wenn n aus dem Zusammenhang klar ist, dann schreiben wir einfach $\sim$ statt $\sim_n$.

Übung 1.2.7. Beweisen Sie, dass $\sim_n$ eine Äquivalenzrelation ist. (Reflexivität und Symmetrie sind fast direkt, nur für die Transitivität muss man kurz überlegen.)

Übung 1.2.8. Schreiben Sie die Äquivalenzklassen für $\sim_2, \sim_3, \sim_4, \dots$ auf, bis Sie verstanden haben, was die Äquivalenzklassen von $\sim_n$ für ein allgemeines $n \in \mathbb{N}$ sind. Wie viele Äquivalenzklassen gibt es für $\sim_1$?

Wir lösen Übung 1.2.8 für $\sim_2$ (jedoch ohne Erklärung!):

$$[0]_{\sim_2} = \{\dots, -6, -4, -2, 0, 2, 4, \dots\} = \text{die Menge der geraden Zahlen in } \mathbb{Z}$$
$$[1]_{\sim_2} = \{\dots, -5, -3, -1, 1, 3, 5, \dots\} = \text{die Menge der ungeraden Zahlen in } \mathbb{Z}$$

Also sehen wir, dass $a \sim_2 b$ ein ausgeklügelter (aber präziser) Weg ist zu sagen, dass a und b entweder beide gerade sind oder a und b beide ungerade sind.[2]

[2] Merken Sie sich: Es ist in der Tat präziser die obige Definition mit $\sim_2$ zu verwenden. Entweder ... oder ... hat keine ganz eindeutige Bedeutung in der Alltagssprache. Einige ETH-Dozenten verbieten die Nutzung von „entweder" sogar! Ich erlaube es Ihnen, aber nur wenn es klar ist, was Sie meinen!

Was passiert nun, wenn man beispielsweise eine gerade Zahl und eine ungeraden Zahl addiert? Man bekommt immer eine ungerade Zahl! Es gilt

$$\text{gerade} + \text{gerade} = \text{gerade} \tag{1.1}$$
$$\text{gerade} + \text{ungerade} = \text{ungerade} \tag{1.2}$$
$$\text{ungerade} + \text{gerade} = \text{ungerade} \tag{1.3}$$
$$\text{ungerade} + \text{ungerade} = \text{gerade}. \tag{1.4}$$

Übung 1.2.9. Zeigen Sie, dass (1.1) der Menge $\mathbb{Z}/\sim_2 = \{[0]_{\sim_2}, [1]_{\sim_2}\}$ die Struktur einer Gruppe gibt mit $[0]_{\sim_2}$ als neutralem Element. Die Additions-Tabelle sieht so aus:

$+$	$[0]_{\sim_2}$	$[1]_{\sim_2}$
$[0]_{\sim_2}$	$[0]_{\sim_2}$	$[1]_{\sim_2}$
$[1]_{\sim_2}$	$[1]_{\sim_2}$	$[0]_{\sim_2}$

Dies möchten wir auch in einem ausgeklügelten Weg schreiben, und zwar für $\sim_n$. Zuerst geben wir aber die Antwort zu Übung 1.2.8: Für $\sim_n$ gibt es n verschiedene Äquivalenzklassen, die man durch „Restklassen“ darstellen kann:

$$\begin{aligned}
[0]_{\sim_n} = 0 + n\mathbb{Z} :=& \{0 + nl \mid l \in \mathbb{Z}\} \\
=& \{\text{alle ganzen Zahlen mit Rest 0 bei Division mit } n\} \\
[1]_{\sim_n} = 1 + n\mathbb{Z} :=& \{1 + nl \mid l \in \mathbb{Z}\} \\
=& \{\text{alle ganzen Zahlen mit Rest 1 bei Division mit } n\} \\
&\vdots \\
[n-1]_{\sim_n} = (n-1) + n\mathbb{Z} :=& \{(n-1) + nl \mid l \in \mathbb{Z}\} \\
=& \{\text{alle ganzen Zahlen mit Rest } (n-1) \text{ bei Division mit } n\}.
\end{aligned}$$

Allgemeiner gilt $[a]_{\sim_n} = a + n\mathbb{Z}$. Wir schreiben $\overline{a}$ für $[a]_{\sim_n}$ (unter der Annahme, dass n aus dem Zusammenhang klar ist). Also ist $\overline{a} = a + n\mathbb{Z} = [a]_{\sim_n}$. Des Weiteren schreiben wir $a \equiv b \,\text{modulo}\, n$ oder $a \equiv b \mod n$ oder $a \equiv b \ (n)$, falls $a \sim_n b$.

Beispiel 1.2.10. Sei $n = 3$. Dann ist beispielsweise

- $\overline{1} = \overline{4} = \overline{-2} = \overline{10} = \overline{3k+1}$ für $k \in \mathbb{Z}$
- $\overline{0} = \overline{3} = \overline{-9} = \overline{3k}$ für $k \in \mathbb{Z}$
- $2 \equiv 5 \equiv -1 \equiv 11 \mod 3$

Obiges Beispiel zeigt, dass jede Restklasse viele Darstellungen als $\overline{a}$ hat.

Definition 1.2.11 (vgl. Fischer [7, Kap. 1.2.7])**.** Sei $n \in \mathbb{N}$. Wir schreiben $\mathbb{Z}/n\mathbb{Z}$ für die Menge der Äquivalenzklassen von $\sim_n$, das heisst[3]

$$\mathbb{Z}/n\mathbb{Z} = \mathbb{Z}/\sim_n = \{\overline{0}, \overline{1}, \ldots, \overline{(n-1)}\}.$$

Wir möchten gerne zwei Elemente von $\mathbb{Z}/n\mathbb{Z}$ addieren können. (Denken Sie an $\mathbb{Z}/n\mathbb{Z}$ für $n = 2$ und die Addition von geraden/ungeraden Zahlen oben.)

Definition 1.2.12. Seien $\overline{a}, \overline{b} \in \mathbb{Z}/n\mathbb{Z}$. Wir definieren

$$\overline{a} + \overline{b} := \overline{a+b}. \tag{1.5}$$

Ist das eine gute Definition? Ist es „erlaubt“, diese Addition einfach so zu definieren? Da wir Addition von *Restklassen/Äquivalenzklassen* durch eine Auswahl von Repräsentationen definieren[4], müssen wir beweisen, dass diese Definition nicht von dieser Wahl abhängt. In der Mathematik sagen wir, dass wir zeigen müssen, dass die Addition in (1.5) wohl definiert ist. Dies ist am besten erklärt, in dem man den Beweis sieht:

Lemma 1.2.13. *Die Addition in (1.5) ist wohl definiert.*

Beweis. Seien $a, a' \in \mathbb{Z}$ mit $\overline{a} = \overline{a'}$ (d.h. a und a' sind zwei möglicherweise verschiedene Repräsentationen derselben Restklasse) und $b, b' \in \mathbb{Z}$ mit $\overline{b} = \overline{b'}$. Wir müssen zeigen, dass

$$\overline{a+b} = \overline{a'+b'}.$$

Das ist einfach: Laut Annahme existieren $l, k \in \mathbb{Z}$ mit

$$a - a' = ln \quad \text{und} \quad b - b' = kn.$$

Daraus folgt, dass

$$(a+b) - (a'+b') = (a-a') + (b-b') = ln + kn = (l+k)n.$$

Also gilt $\overline{a+b} = \overline{a'+b'}$ und daher ist die Addition wohl definiert (sie hängt nicht von der Wahl der Repräsentanten ab). Dies beendet den Beweis. □

Theorem 1.2.14 (vgl. Fischer [7, Kap. 1.2.7])**.** *Die Menge $\mathbb{Z}/n\mathbb{Z}$ zusammen mit der obigen Addition ist eine abelsche Gruppe.*

Beweis. Seien $\overline{a}, \overline{b}, \overline{c} \in \mathbb{Z}/n\mathbb{Z}$. Wir müssen vier Eigenschaften überprüfen:

---

[3] Die erste Gleichung gilt nach Definition der Quotientenmenge und die zweite laut Übung 1.2.8.

[4] Für $n = 3$ ist es nach obiger Definition zum Beispiel a priori nicht klar, dass $\overline{1} + \overline{3} = \overline{1} + \overline{-9}$, vgl. Beispiel 1.2.10.

- Assoziativität folgt aus Assoziativität in $\mathbb{Z}$:
  $(\overline{a} + \overline{b}) + \overline{c} = \overline{a+b} + \overline{c} = \overline{(a+b)+c} = \overline{a+(b+c)} = \overline{a} + \overline{b+c} = \overline{a} + (\overline{b} + \overline{c})$.
- $\overline{0}$ ist das neutrale Element: $\overline{0} + \overline{a} = \overline{0+a} = \overline{a} = \overline{a+0} = \overline{a} + \overline{0}$.
- $\overline{-a}$ ist die Inverse von $\overline{a}$: $\overline{a} + \overline{-a} = \overline{a+(-a)} = \overline{0} = \overline{(-a)+a} = \overline{-a} + \overline{a}$.
- Die Gruppe ist abelsch, da $\mathbb{Z}$ abelsch ist: $\overline{a} + \overline{b} = \overline{a+b} = \overline{b+a} = \overline{b} + \overline{a}$.

□

Bemerkung 1.2.15. Wie bei jeder abelschen Gruppe bedeutet Subtraktion Addition der Inversen: Erinnern Sie sich, dass $\overline{-b}$ die Inverse von $\overline{b}$ ist. Dann gilt: $\overline{a} - \overline{b} = \overline{a} + \overline{-b} = \overline{a + (-b)} = \overline{a-b}$.

1.2.2 Untergruppen, Homomorphismen und Isomorphismen

Dieses Kapitel vergleicht sich mit Kapitel 1.2.6 im Fischer [7]. Wir besprechen kurz einige Begriffe der Gruppen-Theorie.

Definition 1.2.16. Sei $(G, \cdot)$ eine Gruppe. Eine Teilmenge $H \subseteq G$ heisst eine *Untergruppe* (von G), falls das Folgende gilt:

- H ist nicht leer.
- Für alle $h_1, h_2 \in H$ gilt $h_1 h_2 \in H$.
- Für alle $h \in H$ gilt $h^{-1} \in H$.

Beispiel 1.2.17. Wir betrachten einige Beispiele von Untergruppen.

- $\mathbb{Z} \subseteq (\mathbb{Q}, +)$ ist eine Untergruppe von $(\mathbb{Q}, +)$.
- $\{0\} \subseteq \mathbb{Z}$ ist eine Untergruppe von $\mathbb{Z}$.
- Für jedes $m \in \mathbb{N}$ ist die Menge $m\mathbb{Z} = \{ml \mid l \in \mathbb{Z}\}$ ist eine Untergruppe von $\mathbb{Z}$.
- Sei $(G, e, \cdot)$ eine Gruppe. Dann sind $\{e\}$ und G selbst zwei Untergruppen von G.
- $\{\overline{0}, \overline{2}\} \subseteq \mathbb{Z}/4\mathbb{Z} = \{\overline{0}, \overline{1}, \overline{2}, \overline{3}\}$ ist eine Untergruppe von $\mathbb{Z}/4\mathbb{Z}$.

Bemerkung 1.2.18. Ist $H \subseteq (G, \cdot)$ eine Untergruppe, so ist $(H, \cdot)$ mit der Verknüpfung $\cdot$ aus G wieder eine Gruppe:

Assoziativität der Verknüpfung $\cdot$ auf H folgt, da $H \subseteq G$. Ausserdem gilt, dass $h^{-1} \in H$ für $h \in H$ per Definition von Untergruppen. Da $h, h^{-1} \in H$ folgt aus der Definition von Untergruppen auch, dass $e = hh^{-1} \in H$ und e hat die Eigenschaft $eg = ge = g$ für alle $g \in G$. Da $H \subseteq G$ gilt dies also insbesondere für alle $h \in H$. Also ist $(H, \cdot)$ in der Tat eine Gruppe.

Um gewisse strukturelle Ähnlichkeiten zwischen verschiedenen Gruppen auszudrücken, brauchen wir den Begriff von Homomorphismus und Isomorphismus. Allgemeiner definiert man in der Mathematik (insbesondere in der abstrakten Algebra) eine Klasse von Objekten mit gewisser Struktur und betrachtet dann alle Abbildungen zwischen diesen Objekten, welche die Struktur der Objekte „erhalten/respektieren“. Kürzer gesagt:

Definition 1.2.19 (vgl. Fischer [7, Kap. 1.2.6])**.** Seien $(G, \cdot)$ und $(H, *)$ zwei Gruppen. Eine Abbildung $\varphi : G \to H$ ist ein *Homomorphismus*, falls für alle $g_1, g_2 \in G$ gilt $\varphi(g_1 \cdot g_2) = \varphi(g_1) * \varphi(g_2)$[5]. Ein bijektiver Homomorphismus nennen wir einen *Isomorphismus.*

Lemma 1.2.20. *Sei $\varphi : (G, e_G, \cdot) \to (H, e_H, *)$ ein Homomorphismus. Dann gilt:*

(a) $\varphi(e_G) = e_H$, wobei e_G und e_H jeweils die neutralen Elemente von G und H sind.

(b) $\varphi(a^{-1}) = \varphi(a)^{-1}$ für alle $a \in G$.

Beweis. (a) Es gilt $\varphi(e_G) = \varphi(e_G \cdot e_G) = \varphi(e_G) * \varphi(e_G)$. Wir multiplizieren beide Seiten von links mit $\varphi(e_G)^{-1}$ und erhalten:

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

(b) Sei $a \in G$. Nach (a) gilt $\varphi(a) * \varphi(a^{-1}) = \varphi(a \cdot a^{-1}) = \varphi(e_G) = e_H$ und $\varphi(a^{-1}) * \varphi(a) = \varphi(a^{-1}a) = \varphi(e_G) = e_H$. Aus der Eindeutigkeit der Inversen folgt nun (b).

□

Bemerkung 1.2.21. Wie wir in Bemerkung 1.2.4 gesagt haben, verwenden wir normalerweise additive Schreibweise für abelsche Gruppen. In additiver Schreibweise sieht Definition 1.2.19 so aus: Seien $(G, +_G)$ und $(H, +_H)$ zwei (abelsche) Gruppen. Eine Abbildung $\varphi : G \to H$ ist ein Homomorphismus, falls $\varphi(g_1 +_G g_2) = \varphi(g_1) +_H \varphi(g_2)$ gilt.

Beispiel 1.2.22. (1) Sei $m \in \mathbb{N}$. Die Abbildung

$$\begin{aligned} \varphi_m : (\mathbb{Z}, +) &\to (\mathbb{Z}, +) \\ a &\mapsto ma \end{aligned}$$

ist ein Homomorphismus: $\varphi_m(a+b) = m(a+b) = ma+mb = \varphi_m(a)+\varphi_m(b)$. Diese Abbildung ist genau, dann ein Isomorphismus, wenn $m = 1$.

(2) Sei $n \in \mathbb{N}$. Die Abbildung

$$\begin{aligned} \pi_n : \mathbb{Z} &\to \mathbb{Z}/n\mathbb{Z} \\ a &\mapsto \overline{a} \end{aligned}$$

[5]Beachten Sie, dass $\cdot$ die Multiplikation in G ist und $*$ die Multiplikation in H!

ist ein Homomorphismus: $\pi_n(a+b) = \overline{a+b} = \overline{a} + \overline{b} = \pi_n(a) + \pi_n(b)$. Bemerken Sie, dass die ersten zwei Additionszeichen die Addition in $\mathbb{Z}$ bezeichnen und die letzten zwei die Addition in $\mathbb{Z}/n\mathbb{Z}$.

Übung 1.2.23. Zeigen Sie, dass $\mathrm{Bild}(\varphi_n) = \pi_n^{-1}(\overline{0})$, und dass diese Menge eine Untergruppe von $\mathbb{Z}$ ist.

1.3 Ringe und Körper

Unsere Hauptmotivation für dieses Kapitel ist es Körper zu definieren. Diese Kapitel vergleicht sich mit Kapitel 1.3 im Fischer [7]. Als erster Schritt definieren wir Ringe:

1.3.1 Ringe

Definition 1.3.1. Eine Menge R zusammen mit zwei Verknüpfungen

$$\begin{aligned} + : R \times R \to R, &\quad (a,b) \mapsto a+b \quad (\textit{Addition}) \\ \cdot : R \times R \to R, &\quad (a,b) \mapsto a \cdot b \quad (\textit{Multiplikation}) \end{aligned}$$

heisst ein *Ring*, falls:

(R1) $(R, +)$ ist eine abelsche Gruppe.

(R2) $\cdot$ ist assoziativ.

(R3) Für alle $a, b, c \in R$ gilt $a \cdot (b+c) = a \cdot b + a \cdot c$ und $(a+b) \cdot c = a \cdot c + b \cdot c$.

Des Weiteren definieren wir:

- R heisst *kommutativ*, falls für alle $a, b \in R$ gilt $a \cdot b = b \cdot a$.
- R heisst *Ring mit Eins*, falls es ein Element $1 \in R$ gibt, so dass $1 \cdot a = a \cdot 1 = a$ für alle $a \in R$ gilt. Das Element $1 \in R$ mit dieser Eigenschaft heisst *Einselement*.

Wir schreiben oft zur Vereinfachung ab statt $a \cdot b$ für Elemente $a, b \in R$, falls klar ist, welche Operation gemeint ist.

Beispiel 1.3.2. $(\mathbb{Z}, +, \cdot)$ ist ein kommutativer Ring (mit Eins), $1 \in \mathbb{Z}$ ist das Einselement.

Beispiel 1.3.3 (Restklassen)**.** Wir betrachten $\mathbb{Z}/n\mathbb{Z}$ für $n \in \mathbb{N}$. Wir definieren für $\overline{a}, \overline{b} \in \mathbb{Z}/n\mathbb{Z}$:

$$\overline{a} \cdot \overline{b} = \overline{a \cdot b}. \tag{1.6}$$

Man überprüft (ähnlich wie für die Addition zuvor), dass die Multiplikation in (1.6) wohl definiert ist (d.h. sie hängt nicht von der Wahl der Repräsentanten ab). (Machen Sie das!) Mit dieser Definition ist $(\mathbb{Z}/n\mathbb{Z}, +, \cdot)$ ein (endlicher) kommutativer Ring mit Eins; das Einselement ist $\overline{1}$. Dies ermöglicht die sogenannte modulo Arithmetik, welche äusserst nützlich ist für viele Anwendungen in der Zahlen-Theorie (siehe Spass-Aufgabe in Serie 4). Aus Zeitgründen verzichten wir jedoch auf mehr Details. Hier sind die Multiplikations-Tabellen für $n = 4, 5$:

$\cdot$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$
$\overline{1}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{2}$	$\overline{0}$	$\overline{2}$	$\overline{0}$	$\overline{2}$
$\overline{3}$	$\overline{0}$	$\overline{3}$	$\overline{2}$	$\overline{1}$

$n = 4$

$\cdot$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$
$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$
$\overline{1}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$
$\overline{2}$	$\overline{0}$	$\overline{2}$	$\overline{4}$	$\overline{1}$	$\overline{3}$
$\overline{3}$	$\overline{0}$	$\overline{3}$	$\overline{1}$	$\overline{4}$	$\overline{2}$
$\overline{4}$	$\overline{0}$	$\overline{4}$	$\overline{3}$	$\overline{2}$	$\overline{1}$

$n = 5$

Wenn wir eine Ring R mit Eins haben, könnten wir uns nun fragen, ob es auf R eine Gruppenstruktur bezüglich Multiplikation gibt. In diesem Fall wäre $1 \in R$ das neutrale Element. Ein Problem, das wir dabei haben ist das Element $0 \in R$:

Lemma 1.3.4 (vgl. Fischer [7, Kap. 1.3.1])**.** *Für alle $a \in R$ haben wir $a \cdot 0 = 0 \cdot a = 0$.*

Beweis. Wir beweisen, dass $0 \cdot a = 0$. Der Beweis $0 \cdot a = 0$ ist sehr ähnlich.

Da $0 + 0 = 0$ ist, haben wir wegen der Distributivität, dass

$$0 \cdot a = (0 + 0) \cdot a = 0 \cdot a + 0 \cdot a.$$

Wir addieren die additive Inverse von $0 \cdot a$ auf beide Seiten, benutzen Assoziativität und erhalten

$$0 = 0 \cdot a - 0 \cdot a = (0 \cdot a + 0 \cdot a) - 0 \cdot a = 0 \cdot a + (0 \cdot a - 0 \cdot a) = 0 \cdot a + 0 = 0 \cdot a.$$

□

Wir können natürlich die Frage ändern und uns fragen, ob es auf $R \setminus \{0\}$ eine Gruppenstruktur bezüglich Multiplikation gibt. Wir sehen jedoch schnell, dass 0 nicht das einzige Hindernis ist/war:

$$\begin{aligned} &\bullet \text{ In } \mathbb{Z}/4\mathbb{Z} \text{ ist } \overline{2} \cdot \overline{2} = \overline{0}, \text{ aber } \overline{2} \neq \overline{0}. \\ &\bullet \text{ In } \mathbb{Z}/6\mathbb{Z} \text{ ist } \overline{2} \cdot \overline{3} = \overline{0}, \text{ aber } \overline{2} \neq \overline{0} \text{ und } \overline{3} \neq \overline{0}. \end{aligned} \tag{1.7}$$

Dies zeigt, dass für $R = \mathbb{Z}/6\mathbb{Z}$ oder $R = \mathbb{Z}/4\mathbb{Z}$ die Menge $R \setminus \{\overline{0}\}$ keine multiplikative Gruppe sein kann, da die Menge nicht einmal abgeschlossen ist bezüglich Multiplikation (siehe (1.7)). Man kann überprüfen, dass man in $\mathbb{Z}/2\mathbb{Z}, \mathbb{Z}/3\mathbb{Z}$ und $\mathbb{Z}/5\mathbb{Z}$ keine Beispiele wie in (1.7) findet. Um genauer zu erklären, was passiert, brauchen wir eine Definition:

Definition 1.3.5. Ein Ring R heisst *nullteilerfrei*, wenn für alle $a, b \in R$ mit $ab = 0$ folgt, dass $a = 0$ oder $b = 0$. In Quantoren:

$$\forall a, b \in R : ab = 0 \Rightarrow (a = 0) \vee (b = 0). \tag{1.8}$$

Bemerkung 1.3.6. Es ist auch nützlich sich die Kontraposition von (1.8) zu merken:

$$\forall a, b \in R : (a \neq 0) \wedge (b \neq 0) \Rightarrow ab \neq 0.$$

Wir definieren zuerst Körper und fragen uns dann, wann genau unser Lieblings-Beispiel $\mathbb{Z}/n\mathbb{Z}$ ein nullteilerfreier Ring ist.

1.3.2 Körper

Kurz gesagt ist ein Körper ein Ring mit Eins, wo jedes von Null verschiedene Element eine multiplikative Inverse hat.

Definition 1.3.7 (vgl. Fischer [7, Kap. 3.3])**.** Eine Menge K (oder $(K, +, \cdot, 0, 1)$) zusammen mit zwei Verknüpfungen

$$\begin{aligned} + &: K \times K \to K, \quad (a, b) \mapsto a + b \quad \text{(Addition)} \\ \cdot &: K \times K \to K, \quad (a, b) \mapsto a \cdot b \quad \text{(Multiplikation)} \end{aligned}$$

heisst *Körper*, wenn Folgendes gilt:

(1) K zusammen mit der Addition ist eine abelsche Gruppe (0 ist ihr neutrales Element, $-a$ die Inverse von $a \in K$).

(2) Sei $K^\times := K \setminus \{0\}$. Dann ist $(K^\times, \cdot, 1)$ eine abelsche Gruppe. Das heisst, für alle $a, b \in K^\times$ ist $ab \in K^\times$ und für alle $a \in K^\times$ existiert ein $b \in K^\times$ mit $ab = 1$. In diesem Fall schreiben wir $b = a^{-1} =: \frac{1}{a}$ und allgemeiner schreiben wir $\frac{c}{d} := cd^{-1}$.

(3) Es gelten die Distributivgesetze:

$$\begin{aligned} a \cdot (b + c) &= a \cdot b + a \cdot c \\ (b + c) \cdot a &= b \cdot a + c \cdot a \end{aligned}$$

In anderen Worten, K ist auch ein Ring.

Beachten Sie, dass wir auch hier oft ab statt $a \cdot b$ schreiben für $a, b \in K$.

Bemerkung 1.3.8. Man könnte auf die Definition einer Gruppe und eines Rings verzichten und direkt einen Körper wie in nachfolgender Definition definieren.

Definition 1.3.9 (alternative direkte Definition)**.** *Ein Körper ist ein Tupel* $(K, +, \cdot, 0, 1)$ *bestehend aus einer Menge* K *mit zwei Abbildungen*

$$+ : K \times K \to K; (x, y) \mapsto x + y$$

$$\cdot : K \times K \to K; (x, y) \mapsto x \cdot y$$

und ausgezeichneten Elementen $0, 1 \in K$*, so dass die Körperaxiome gelten:*

(K1) $\forall x, y, z \in K : x + (y + z) = (x + y) + z$ *(Assoziativität der Addition)*

(K2) $\forall x, y \in K : x + y = y + x$ *(Kommutativität der Addition)*

(K3) $\forall x \in K : 0 + x = x$ *(Neutrales Element der Addition)*

(K4) $\forall x \in K\, \exists x' \in K : x + x' = 0$ *(Inverses Element der Addition)*

(K5) $\forall x, y, z \in K : x \cdot (y \cdot z) = (x \cdot y) \cdot z$ *(Assoziativität der Multiplikation)*

(K6) $\forall x \in K : 1 \cdot x = x$ *(Neutrales Element der Multiplikation)*

(K7) $\forall x \in K \setminus \{0\}\, \exists x' \in K : x' \cdot x = 1$ *(Inverses Element der Multiplikation)*

(K8) $\forall x, y, z \in K : x \cdot (y + z) = x \cdot y + x \cdot z$ *und*
$\forall x, y, z \in K : (y + z) \cdot x = y \cdot x + z \cdot x$ *(Distributivität)*

(K9) $1 \neq 0$ *(Nichttrivialität)*

(K10) $\forall x, y \in K : x \cdot y = y \cdot x$ *(Kommutativität der Multiplikation)*

Die Definition einer Gruppe kann später vielleicht doch hilfreich sein, wenn wir Determinanten besprechen und gewisse Matrix-Gruppen erwähnen. Im Moment ist es wichtig, sich Folgendes zu merken: Unser erstes wichtiges Thema sind Vektorräume über Körper. Körper spielen jetzt also eine wichtige Rolle, Gruppen und Ringe (noch) nicht.

Lemma 1.3.10. *Wir haben folgende Folgerungen aus der Definition eines Körpers:*

(a) Es gilt $1 \neq 0$*. Das heisst, jeder Körper hat mindestens zwei Elemente.*

(b) Es gilt $0 \cdot a = a \cdot 0 = 0$*.*

(c) Falls $a \cdot b = 0$, dann ist $a = 0$ oder $b = 0$. Anders gesagt, ist jeder Körper nullteilerfrei.

(d) Wir haben $a \cdot (-b) = -(a \cdot b)$ und $(-a) \cdot (-b) = a \cdot b$.

(e) Falls $x \cdot a = y \cdot a$ mit $a \in K^\times$ und $x, y \in K$, dann folgt $x = y$.

Beweis. (a) Da $K^\times$ eine Gruppe ist, ist $1 \in K^\times$. Das Element 0 ist hingegen per Definition nicht in $K^\times$. Daher ist $1 \neq 0$.

(b) Dies haben wir zuvor schon bewiesen.

(c) Die Tatsache, dass $K^\times$ eine Gruppe ist, impliziert insbesondere, dass $K^\times$ abgeschlossen ist bezüglich Multiplikation. Das heisst, aus $a, b \in K^\times$ folgt $ab \in K^\times$. Dies ist genau die Kontraposition der Aussage in (c).

(d) Bemerken Sie, dass

$$ab + a(-b) = a(b + (-b)) = a \cdot 0 = 0,$$

wobei die erste Gleichheit wegen Distributivität gilt, die zweite da $-b$ die Inverse von b ist und die letzte Gleichheit gilt wegen (b). Daraus folgt (Wieso? Eindeutigkeit der Inversen), dass

$$a(-b) = -(ab). \tag{1.9}$$

Für die zweite Aussage verwenden wir (1.9):

$$(-a)(-b) \overset{(1.9)}{=} -(-(a)b) = -(b(-a)) \overset{(1.9)}{=} -(-(ba)) = -(-(ab)) = ab, \tag{1.10}$$

wobei die letzte Gleichheit wegen Proposition 1.2.2 gilt.

(e) Hier geben wir nur einen Hinweis: Multiplizieren Sie die Gleichheit $xa = ya$ mit $a^{-1} \in K^\times$ von der rechten Seite.

Andere Folgerungen sehen sie in der Serie/Übungsstunde. □

Beispiel 1.3.11. Sie haben in der Analysis schon die Körper $\mathbb{Q}, \mathbb{R}$ und $\mathbb{C}$ gesehen. Diese Körper sind auch für die lineare Algebra sehr wichtig. Falls Sie Fragen haben, dann sollten Sie uns fragen!

Wir kehren zu unserem Restklassen-Beispiel zurück, um einige endliche Körper zu bestimmen.

Lemma 1.3.12. *Der Ring $\mathbb{Z}/n\mathbb{Z}$ ist genau dann nullteilerfrei, wenn n eine Primzahl ist.*

Beweis. „$\Rightarrow$" Hier verwenden wir die Kontraposition. Angenommen n ist nicht prim, dann existieren $1 < k, l < n$, sodass $n = kl$[6]. Damit gilt in $\mathbb{Z}/n\mathbb{Z}$:

$$\overline{0} = \overline{n} = \overline{k} \cdot \overline{l},$$

aber $\overline{k} \neq 0$ und $\overline{l} \neq 0$ (Wieso?). Daher ist $\mathbb{Z}/n\mathbb{Z}$ nicht nullteilerfrei. „$\Leftarrow$" Nehmen wir jetzt an, dass n eine Primzahl ist. Seien $\overline{k}, \overline{l} \in \mathbb{Z}/n\mathbb{Z}$ mit

$$\overline{k} \cdot \overline{l} = \overline{0}.$$

Daraus folgt, dass ein $r \in \mathbb{Z}$ existiert mit $kl = rn$ und insbesondere teilt die Primzahl n das Produkt kl. Dies impliziert[7], dass ((n teilt k) oder (n teilt l)), was äquivalent ist zu ($\overline{k} = 0$ oder $\overline{l} = 0$). Dies zeigt, dass $\mathbb{Z}/n\mathbb{Z}$ nullteilerfrei ist. □

Bemerkung 1.3.13. Wir möchten jetzt nicht gross darauf eingehen; nur als Bemerkung: Wir sagen, dass ein Element in einem nullteilerfreien Ring R zwei verschiedene Eigenschaften haben kann:

- Man sagt, dass $p \in R$ prim ist, falls p teilt ab für $a, b \in R$ stets impliziert, dass p teilt a oder p teilt b. In Quantoren ausgedrückt:

$$\forall a, b \in R : p \text{ teilt } ab \Rightarrow p \text{ teilt } a \vee p \text{ teilt } b.$$

- Man sagt, dass $p \in R$ irreduzibel ist, falls p kein Produkt von zwei Elementen in R ist, die nicht invertierbar sind. In Quantoren ausgedrückt:

$$\forall a, b \in R : p = ab \Rightarrow a \in R^\times \vee b \in R^\times.$$

In $\mathbb{Z}$ sind diese beiden Eigenschaften äquivalent. Die bekannte Tatsache, dass es in $\mathbb{Z}$ eine Primfaktorzerlegung gibt ist äquivalent zu der Aussage, dass in $\mathbb{Z}$ Primelemente und irreduzible Elemente gleich sind.

Lemma 1.3.14. *Sei R ein nullteilerfreier kommutativer Ring mit 1, so dass $1 \neq 0$. Falls R endlich vielen Elementen besitzt, dann ist R ein Körper.*

Beweis. Sei R ein Ring wie in den Voraussetzungen des Lemmas. Um zu zeigen, dass R ein Körper ist, genügt es zu zeigen, dass jedes $a \in R \setminus \{0\}$ eine multiplikative Inverse hat (Wieso?). Der Trick ist die folgende „Multiplikation-mit-a-Abbildung" zu betrachten:

$$m_a : R \setminus \{0\} \to R \setminus \{0\}$$
$$b \mapsto m_a(b) := ab$$

[6] Hier benutzt man, dass Primzahlen in $\mathbb{Z}$ irreduzibel sind, siehe Bemerkung 1.3.13.

[7] Hier verwendet man, dass Primzahlen prim sind, siehe Bemerkung 1.3.13.

Da R nullteilerfrei ist, ist m_a definiert (d.h. die Zielmenge von m_a ist in der Tat in $R \setminus \{0\}$). Ausserdem folgt die Injektivität aus der Nullteilerfreiheit: Seien $b, b' \in R \setminus \{0\}$ mit $m_a(b) = ab = ab' = m_a(b')$. Dann folgt (aus der Inverse und Distributivität), dass $a(b - b') = 0$. Da $a \neq 0$, folgt aus der Nullteilerfreiheit, dass $b - b' = 0$. Dies zeigt die Injektivität.

Wir haben noch nicht die Annahme, dass $|R| < \infty$ (und daher $|R \setminus \{0\}| < \infty$) benutzt. Aus dem Schubfachprinzip folgt: Weil $|R \setminus \{0\}| < \infty$ und m_a injektiv ist, ist m_a auch surjektiv. Bemerken Sie nun zuletzt, dass $1 \in R \setminus \{0\}$ (per Annahme) und daher existiert ein $b \in R$ mit $ab = m_a(b) = 1$. Dies zeigt, dass a eine Inverse besitzt, was wir zeigen wollten. □

Korollar 1.3.15. *Sei p eine Primzahl. Dann ist $\mathbb{Z}/p\mathbb{Z}$ ein Körper mit p Elementen. Das Nullelement ist $\overline{0}$ und das Einselement ist $\overline{1}$.*

Bemerkung 1.3.16. Der Körper[8] $\mathbb{Z}/p\mathbb{Z}$ für p eine Primzahl ist zentral im Gebiet der Algebra und Zahlentheorie. In Serie 5 haben Sie einige lustige Aufgaben dazu.

Übung 1.3.17. Wir haben eigentlich gezeigt, dass a eine rechtsseitige Inverse besitzt. Zeigen Sie, dass eine rechtsseitige Inverse auch eine linksseitige Inverse ist. (vgl. Fischer [7, 1.2.3, b)])

Definition 1.3.18. Sei $(K, 0, 1)$ ein Körper. Für $n \in \mathbb{N}$ definieren[9] wir

$$n \cdot 1 = \underbrace{1 + \ldots + 1}_{n\text{-Mal}} \in K.$$

Die *Charakteristik* von K ist dann definiert als

$$\operatorname{char}(K) = \begin{cases} 0, & \text{falls } n \cdot 1 \neq 0 \text{ für alle } n \in \mathbb{N} \\ \min\{n \mid n \cdot 1 = 0\}, & \text{sonst} \end{cases}.$$

Beachten Sie, dass $\operatorname{char}(K) \in \mathbb{N} \cup \{0\}$!

Beispiel 1.3.19. Wir berechnen einige Charakteristiken für uns schon bekannte Körper:

- Für $\mathbb{Q}, \mathbb{R}$ und $\mathbb{C}$ gilt $\operatorname{char}(\mathbb{Q}) = \operatorname{char}(\mathbb{R}) = \operatorname{char}(\mathbb{C}) = 0$.
- Für $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ gilt $\operatorname{char}(\mathbb{F}_p) = p$. (Wieso?)

Man kann mit einem ähnlichen Argument wie in Lemma 1.3.12 zeigen, dass die Charakteristik eines Körpers 0 oder eine Primzahl $p \in \mathbb{N}$ ist.

[8] Normalerweise bezeichnet man diesen Körper mit $\mathbb{F}_p := \mathbb{Z}/p\mathbb{Z}$.

[9] Hier ist eine Induktion/rekursive Definition versteckt.

1.4 Polynome

Die letzte Vorbereitung, bevor wir endlich anfangen können (mit linearer Algebra), ist es den Raum von Polynomen über einem Körper einzuführen. In der Mittelschule betrachtet man Polynome als Funktionen, aber wir möchten Polynome als formale Objekte betrachten. Dies bereitet in der Regel einigen Studenten Mühe. Daher erklären wir unten den subtilen Unterschied zwischen Polynomen als formalen Objekten und ihren entsprechenden Funktionen. Zuerst aber definieren wir Polynome:

Grob gesagt, ist ein Polynom ein Ausdruck, der durch Konstanten und Symbole (Variablen oder Unbestimmte genannt) mittels Addition, Multiplikation und Potenzieren (mit nicht-negativen ganzen Potenzen) gebildet werden kann. Zwei solcher Ausdrücke, welche durch anwenden von Kommutativität, Assoziativität und Distributivität von Addition und Multiplikation ineinander überführt werden können, definieren dasselbe Polynom. Nun die formale Definition:

Definition 1.4.1. Sei K ein Körper und x eine Unbestimmte (unbekannte Variable)[10]. Wir nennen einen formalen Ausdruck der Form

$$f(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n, \tag{1.11}$$

wobei $n \in \mathbb{N}$ und $a_i \in K$ für alle $0 \leq i \leq n$, ein *Polynom* über K. Die Menge

$$K[x] = \{f(x) \text{ wie in (1.11), mit } n \in \mathbb{N} \text{ und } a_i \in K \text{ für alle } 0 \leq i \leq n\}$$

heisst der *Polynomraum* über K (mit Unbestimmter x).

Bemerkung 1.4.2. Um ganz präzis zu sein, müssen wir sagen, dass ein Polynom über K eindeutig bestimmt ist durch eine Folge $\{a_k\}_{k=0}^{\infty}$ mit $a_k \in K$ für alle $k \geq 0$, so dass ein $n \in \mathbb{N}$ existiert mit $a_k = 0$ für alle $k > n$. Zum Beispiel entspricht das Polynom $1 - x^2$ der Folge $(1, 0, -1, 0, 0, \ldots)$ und die Polynome $1-x^2$, $1+0\cdot x-x^2$ und $1-x^2+0\cdot x^7$ sind alle gleich. Diese Bemerkung ist vielleicht sehr wichtig für unser „Computer-Programm“ (bestehend aus formalen Definitionen und Theoremen), das wir langsam am Schreiben sind, aber jeder von uns versteht schon, wenn zwei Polynome gleich sind durch die (impräzise) Erklärung vor der Definition.

Hier sind noch einige Definitionen und Notationen, die wir brauchen werden:

Definition 1.4.3. Das Polynom mit $a_k = 0$ für alle $k \in \mathbb{N} \cup \{0\}$ heisst das *Nullpolynom* und wird einfach mit $0 \in K[x]$ bezeichnet. Normalerweise schreiben wir $f \in K[x]$ statt

[10] Denken Sie bei x als Platzhalter, oder besser gesagt, als sogenanntes „freies Element “: „frei“ im Sinn, dass es nichts mit K zu tun hat, und „Element“ im Sinn, dass wir Potenzen davon bilden können, wie zum Beispiel $x^n = \underbrace{x \cdot \ldots \cdot x}_{n\text{-Mal}}$.

$f(x) \in K[x]$. Der Grad von $f \in K[x]$ wie in (1.11) definieren wir als

$$\deg(f) = \begin{cases} -\infty, & \text{falls } f = 0 \\ \max\{k \in \mathbb{N} \cup \{0\} \mid a_k \neq 0\}, & \text{sonst} \end{cases}.$$

Manchmal verwenden wir das Summenzeichen, um Polynome kürzer zu schreiben. Dabei definieren wir $x^0 = 1$ und $x^1 = x$:

$$\sum_{k=0}^{n} a_k x^k = a_0 x^0 + a_1 x^1 + \ldots + a_n x^n = a_0 + a_1 x + \ldots + a_n x^n.$$

Der höchste Koeffizient, der ungleich Null ist, heisst der *Leitkoeffizient* (oder der führende Koeffizient). Wenn der Leitkoeffizient 1 ist, dann heisst das Polynom *normiert*. Polynome der Form $a_l x^l$ für ein $l \in \mathbb{N}$ heissen *Monome*. Wir können zwei Polynome $f, g \in K[x]$ addieren und multiplizieren, indem wir die üblichen Regeln der Kommutativität, Assoziativität und Distributivität sowie Potenzregeln benutzen. Zum Beispiel in $\mathbb{Q}[x]$: Seien

$$f = 3x^2 + 4x + 2 \quad \text{und} \quad g = x^3 - 3x + 4$$

Dann ist

$$\begin{aligned} f + g &= (3x^2 + 4x + 2) + (x^3 - 3x + 4) = x^3 + 3x^2 + x + 6, \\ f \cdot g &= (3x^2 + 4x + 2) \cdot (x^3 - 3x + 4) = 3x^5 + 4x^4 - 7x^3 + 10x + 8. \end{aligned}$$

Übung 1.4.4. Für $n \in \mathbb{N} \cup \{0\}$ definieren wir die Regel $-\infty + n = -\infty$. Seien $f, g \in K[x]$ nicht Null. Zeigen Sie (unter Benutzung der Regel), dass

$$\deg(fg) = \deg(f) + \deg(g).$$

Gilt

$$\deg(f + g) = \max(\deg(f), \deg(g))?$$

Wenn ja, beweisen Sie es. Wenn nein, überlegen Sie sich, in welchen Fällen es nicht gilt und in welchen Fällen es doch gilt!

Bemerkung 1.4.5. Wie Sie oben sehen, erlauben wir uns die Freiheit Polynome nicht immer mit aufsteigenden Potenzen zu schreiben.

Bemerkung 1.4.6. Unser „Computer-Programm“ ist verwirrt. Hier ist eine präzise Definition der Addition und Multiplikation: Seien $f, g \in K[x]$ mit entsprechenden Folgen $\{a_k\}_{k=0}^{\infty}$ und $\{b_k\}_{k=0}^{\infty}$.

- Wir definieren $f+g$ als das eindeutige Polynom, das der Folge $\{c_k\}_{k=0}^{\infty}$ mit $c_k = a_k + b_k$ entspricht. Auch für $\{c_k\}_{k=0}^{\infty}$ gilt, dass $c_k \in K$ für alle $k \geq 0$, und dass ein $M \in \mathbb{N}$ existiert mit $c_k = 0$ für alle $k \geq M$. (Wieso haben wir diesen Satz geschrieben? Wieso existiert ein solches M?)

- Die Formel für die Multiplikation sieht etwas seltsam aus in dieser „Folgen-Schreibweise". Dies erklärt wieso die Darstellung in (1.11) besser ist, obwohl sie weniger präzis ist. Wir definieren $f \cdot g$ als das eindeutige Polynom, das der Folge $\{c_k\}_{k=0}^{\infty}$ entspricht, wobei

$$c_k := a_0 b_k + a_1 b_{k-1} + \ldots + a_k b_0 = \sum_{i=0}^{k} a_i b_{k-i} = \sum_{\substack{0 \leq i,j \leq k \\ i+j=k}} a_i b_j. \tag{1.12}$$

(Üben Sie ihr Verständnis des Summenzeichens, indem Sie (1.12) gut verstehen).

Definition 1.4.7. Elemente von $\mathbb{Q}[x]$ (resp. $\mathbb{R}[x], \mathbb{C}[x]$) heissen *rationale Polynome* (resp. *reelle Polynome*, *komplexe Polynome*).

Definition 1.4.8 (Teilbarkeit)**.** Seien $f, g \in K[x]$. Wir sagen, f *teilt* g und schreiben $f \mid g$, falls ein $h \in K[x]$ existiert, so dass $g = f \cdot h$.

Ziele von diesem Abschnitt : (1) Den Unterschied zwischen Polynomen und Polynomfunktionen zu verstehen.

(2) Wir möchten in $K[x]$ dividieren können, aber das geht nicht immer. Der Ersatz ist „Division mit Rest in $K[x]$".

(3) Nullstellen in K für Polynome in $K[x]$ besprechen. Insbesondere beweisen wir: $\lambda \in K$ ist eine Nullstelle von f genau dann, wenn f durch $(x - \lambda)$ geteilt wird.

(4) Fakten über Nullstellen von komplexen und reellen Polynomen besprechen.

1.4.1 Unterschied zwischen Polynomen und Polynomfunktionen

Die Unbestimmte x in $K[x]$ wartet darauf, dass jemand kommt und etwas in sie einsetzt. In diesem Sinn kann man bei x an einen Platzhalter denken. Mann kann verschiedene Dinge in x einsetzen: Die „üblichen Verdächtigen" sind Elemente von K. Wir werden später jedoch auch Matrizen und Funktionen in x einsetzen.[11]

[11] Fortgeschrittene Aussage (d.h. man kann dies sorglos ignorieren): Es ergibt sicherlich Sinn Elemente einer beliebigen K-Algebra einzusetzen.

Betrachten wir was passiert, wenn wir Elemente von K einsetzen. Wir definieren die „Auswertungs-Abbildung“ (evaluation auf Englisch) durch

$$\begin{aligned}\mathrm{ev}_K : K[X] &\to \mathrm{Abb}(K, K) = \{\text{Funktionen von } K \text{ nach } K\},\\ f &\mapsto \mathrm{ev}_K(f),\end{aligned}$$

wobei

$$\begin{aligned}\mathrm{ev}_K(f) : K &\to K,\\ \lambda &\mapsto f(\lambda) = a_0 + a_1\lambda + \ldots + a_n\lambda^n\end{aligned}$$

für $f = a_0 + a_1x + \ldots + a_nx^n$.

Bemerkung 1.4.9. Wenn keine Verwechslungsgefahr entsteht, schreiben wir ev statt ev_K.

Wir möchten zwischen $f \in K[x]$ und $\mathrm{ev}_K(f)$ unterscheiden. Der Grund ist, dass ev_K im Allgemeinen nicht injektiv ist!

Beispiel 1.4.10. Sei $K = \mathbb{F}_3$. Betrachten Sie $f = x^3 - x \in \mathbb{F}_3[x]$ und $g = 0 \in \mathbb{F}_3[x]$. Wir behaupten, dass $\mathrm{ev}(f) = \mathrm{ev}(g)$:

Natürlich ist $\mathrm{ev}(g)$ die konstante Abbildung, die alles auf $\overline{0} \in \mathbb{F}_3$ abbildet. Für $\mathrm{ev}(f)$ müssen wir etwas berechnen:

$$\begin{aligned}\mathrm{ev}(f)(\overline{0}) &= \overline{0}^3 - \overline{0} = \overline{0}\\ \mathrm{ev}(f)(\overline{1}) &= \overline{1}^3 - \overline{1} = \overline{0}\\ \mathrm{ev}(f)(\overline{2}) &= \overline{2}^3 - \overline{2} = \overline{0}.\end{aligned}$$

Also ist in der Tat $\mathrm{ev}(f) = \mathrm{ev}(g)$.

Übung 1.4.11. Finden Sie $f \neq 0 \in \mathbb{F}_p[x]$ mit $\mathrm{ev}_{\mathbb{F}_p}(f) = 0$, wobei 0 für die Nullfunktion in $\mathrm{Abb}(\mathbb{F}_p, \mathbb{F}_p)$. (Hinweis: In Beispiel 1.4.10 gilt $x^3 - x = (x - \overline{0})(x - \overline{1})(x - \overline{2})$. Wieso kann man diesen Trick in einem endlichen Körper benutzen?)

Man kann zeigen (wir zeigen in diesem Abschnitt Teile davon): Sei K ein Körper mit unendlich vielen Elementen. Dann ist $\mathrm{ev}_K : K[x] \to \mathrm{Abb}(K, K)$ injektiv.

1.4.2 Division mit Rest in $K[x]$

Lemma 1.4.12 (Division mit Rest in $K[x]$)**.** *Sei $f \in K[x]$ ein beliebiges Polynom und $0 \neq g \in K[x]$. Dann gibt es Polynome $q, r \in K[x]$, so dass*

$$f = qg + r$$

mit $\deg(r) < \deg(g)$ *oder* $r = 0$.

Wir werden sehen, dass die Polynome q und r durch die Bedingungen in Lemma 1.4.12 eindeutig bestimmt sind. Des Weiteren bemerken wir, dass der Beweis von Lemma 1.4.12 einen Algorithmus angibt, wie man q und r finden kann[12].

Beweis von Lemma 1.4.12. Da $g \neq 0$, gilt $\deg(g) \neq -\infty$. Falls $\deg(g) = 0$, dann ist $g = b_0 \in K \setminus \{0\}$ ein Skalar ungleich Null. Daher können wir $q = b_0^{-1} f$ und $r = 0$ setzen für $f \in K[x]$ und es gilt $f = qg + r$, wie gewünscht. Wir nehmen jetzt an, dass $\deg(g) \geq 1$. In diesem Fall beweisen wir die Existenz von q und r mittels Induktion über $\deg(f)$:

Für die Induktionsverankerung nehmen wir an, dass $\deg(f) < \deg(g)$. (Beachten Sie, dass dies insbesondere den Fall $\deg(f) = 0$ abdeckt.) Wir wählen $q = 0$ und $r = f$ und erhalten $f = qg + r$.

Für den Induktionsschritt nehmen wir nun an, dass $\deg(f) =: m \geq n := \deg(g)$ und die Aussage des Lemmas für alle Polynome mit Grad kleiner als f gilt. Da $\deg(f) = m$ und $\deg(g) = n$, sind f und g von der Form

$$f(x) = a_m x^m + a_{m-1} x^{m-1} + \ldots + a_0$$

mit $a_m \neq 0$ und

$$g(x) = b_n x^n + b_{n-1} x^{n-1} + \ldots + b_0$$

mit $b_n \neq 0$. Betrachten Sie das Polynom

$$f_1(x) = f(x) - \left(\frac{a_m}{b_n}\right) x^{m-n} g(x). \tag{1.13}$$

Beachten Sie, dass die Leitkoeffizienten von $f(x)$ und $(\frac{a_m}{b_n}) x^{m-n} g(x)$ gleich sind und sich somit auslöschen. Also haben wir entweder $\deg(f_1) < \deg(f)$ oder $f_1 = 0$.

Falls $f_1 = 0$ ist, dann haben wir

$$f(x) = \underbrace{\left(\frac{a_m}{b_n} x^{m-n}\right)}_{q(x)} g(x) + \underbrace{0}_{r(x)},$$

wie gewünscht. Falls $f_1 \neq 0$, dann gilt $\deg(f_1) < \deg(f)$. Also können wir die Induktionsannahme auf f_1 anwenden und finden $q_1, r \in K[x]$ mit

$$f_1(x) = q_1(x) g(x) + r(x) \tag{1.14}$$

[12] Dieser Abschnitt folgt [3, §5.2].

und mit $\deg(r) < \deg(g)$. Wir setzen (1.14) in (1.13) ein und erhalten

$$f(x) = f_1(x) + \left(\frac{a_m}{b_n}\right)x^{m-n}g(x) = \underbrace{\left(q_1(x) + \frac{a_m}{b_n}x^{m-n}\right)}_{q(x)}g(x) + r(x),$$

wie gewünscht. Dies schliesst den Induktionsschritt ab und beweist somit das Lemma. □

Übung 1.4.13. Beweisen Sie die Eindeutigkeit der Polynome q und r in Lemma 1.4.12.

In der Praxis benutzt man oft die sogenannte „schriftliche Division“, um Division mit Rest durchzuführen. Bei der Division von $x^3 + x^2 + 3x - 1$ durch $x^2 - 1$ erhalten wir beispielsweise $x + 1$ als Quotient und $4x$ als Rest:

$$\begin{array}{l}
(x^3 + x^2 + 3x - 1) : (x^2 - 1) = x + 1 \\
\underline{-x^3 + x} \\
\qquad x^2 + 4x - 1 \\
\qquad \underline{-x^2 + 1} \\
\qquad\qquad 4x
\end{array}$$

Die Art und Weise wie man solch eine schriftliche Division aufschreibt, ist von Land zu Land unterschiedlich. In jedem Fall ist jedoch die Idee, dass man in jedem Schritt versucht die Leitkoeffizienten auszulöschen. Dies ist auch genau die Idee im Beweis von Lemma 1.4.12.

Übung 1.4.14. Teilen Sie $x^7 + x^6 - x^5 - x^4 + x^3 + x^2 - x - 1$ durch $x^2 + x + 1$.

1.4.3 Nullstellen

Eine wichtige Konsequenz der Polynomdivision mit Rest ist, dass wir Nullstellensuche und Polynomfaktorisierung zueinander in Beziehung setzen können. Wir sagen, dass $a \in K$ eine Nullstelle von $0 \neq f \in K[x]$ ist, falls $f(a) = 0$.

Korollar 1.4.15 (Linearfaktorzerlegung)**.** *Sei $a \in K$ und $f \in K[x]$ ein von Null verschiedenes Polynom. Dann ist a eine Nullstelle von f genau dann, wenn $(x - a)$ ein Teiler von f in $K[x]$ ist.*

Beweis. Sei $g(x) = x - a$. Wir wenden Division mit Rest an und erhalten

$$f(x) = q(x)(x - a) + r(x)$$

mit $\deg(r) < \deg(g) = 1$, so dass r eine Konstante sein muss. Wir beweisen nun die Aussage: Wir nehmen zuerst an, dass $f(a) = 0$. Wir setzen für x in obiger Gleichung a

ein und erhalten $r = 0$. Also gilt

$$f(x) = q(x)(x - a) \tag{1.15}$$

und $(x - a)$ ist ein Teiler von f. Für die Umkehrung nehmen wir an, dass $(x - a)$ ein Teiler von f ist und (1.15) gilt. In diesem Fall ist $f(a) = q(a)(a - a) = 0$. □

Sei $a \in K$. Das Polynom $(x - a)$ kann mehrmals ein Polynom $P \in K[x]$ teilen. Dann spricht man von *mehrfache Nullstelle*. Hier ist eine präzisere Definition:

Definition 1.4.16. Sei $p \in K[x]$ mit $\deg(p) > 0$. Für $a \in K$ definieren wir die *Vielfachheit von a bei p* durch

$$\mu(p \mid a) := \max\{r \in \mathbb{N} \cup \{0\} \mid \exists g \in K[x] \text{ mit } p = (x - a)^r g\}.$$

Beachten Sie: $\mu(p \mid a)$ ist definiert für jedes $a \in K$, und ist einfach 0, wenn a keine Nullstelle von p ist.

Korollar 1.4.17 (Anzahl Nullstellen)**.** *Für $f \in K[x]$ von Grad $n \in \mathbb{N} \cup \{0\}$ existieren höchstens n verschiedene Zahlen $a \in K$ mit $f(a) = 0$.*

Beweis. Wir beweisen die Aussage per Induktion über $\deg(f)$.

Falls $\deg(f) = 0$ ist, dann entspricht f einer von Null verschiedenen Konstanten und hat keine Nullstellen in K.

Angenommen es gilt $\deg(f) > 0$ und es existiert eine Nullstelle $a \in K$ (sonst sind wir fertig, wieso?). Also gilt nach der Linearfaktorzerlegung (Korollar 1.4.15), dass ein Polynom $g \in K[x]$ existiert mit

$$f(x) = (x - a)g(x).$$

Wir benutzen, dass $\deg(g_1 g_2) = \deg(g_1) + \deg(g_2)$ ist für beliebige von Null verschiedene Polynome $g_1, g_2 \in K[x]$, um zu sehen, dass $\deg(g) = \deg(f) - 1$ ist. Nach Induktionsannahme (angewendet auf g) hat g höchstens $n - 1$ Nullstellen. Eine Nullstellen von f ist entweder a oder eine Nullstelle von g. Also hat f höchstens n Nullstellen und das Korollar folgt. □

1.4.4 Polynome über $\mathbb{R}$ und $\mathbb{C}$

Kurze Einführung und der Fundamentalsatz der Algebra

Wir betrachten ganz schnell die Zahlenbereich-Erweiterungen, die Sie in Analysis besprochen haben und jedes Mal nennen wir einen Grund, wieso wir diese bestimmte Erweiterung brauchen können.

- Von $\mathbb{N}$ nach $\mathbb{Z}$: Weil wir $x + 5 = 3$ lösen wollen.
- Von $\mathbb{Z}$ nach $\mathbb{Q}$: Weil wir $2x + 3 = 0$ lösen wollen.
- Von $\mathbb{Q}$ nach $\mathbb{R}$: Weil wir $x^2 - 2 = 0$ lösen wollen.
- Von $\mathbb{R}$ nach $\mathbb{C}$: Weil wir $x^2 + 1 = 0$ lösen wollen.

Der Leser sollte sich fragen: Wohin jetzt? Eigentlich gibt es viele Antworten[13]. Im Bezug auf Polynome sind wir allerdings schon am Ziel! Dies ist der Inhalt des folgenden viel zelebrierten Satzes.

Theorem 1.4.18 (Fundamentalsatz der Algebra)**.** *Sei* $P \in \mathbb{C}[x]$ *mit* $\deg(P) > 0$. *Dann existiert* $\lambda \in \mathbb{C}$ *mit* $P(\lambda) = 0$.

Mittels Korollar 1.4.15 folgt:

Korollar 1.4.19. *Sei* $P \in \mathbb{C}[x]$ *mit* $\deg(P) = n > 0$ *und Leitkoeffizient* $a \in \mathbb{C}$. *Dann existieren* $\lambda_1, \ldots, \lambda_k \in \mathbb{C}$ *paarweise verschieden, so dass*

$$P = a(x - \lambda_1)^{l_1} \cdot \ldots \cdot (x - \lambda_k)^{l_k} = a \prod_{i=1}^{k} (x - \lambda_i)^{l_i} \tag{1.16}$$

und $l_1 + \ldots + l_k = \sum_{i=1}^{k} l_i = n$ *gilt.*

Beweis-Idee. Solche Aussagen beweist man mit Induktion über $\deg(P)$. Der Induktionsschritt ist dank Theorem 1.4.18 machbar. Hier ist kurz die Idee: Theorem 1.4.18 impliziert für $\deg(P) > 0$, dass es ein $\lambda \in \mathbb{C}$ gibt mit $P(\lambda) = 0$. Letzteres impliziert wiederum, dass $(x - \lambda)$ ein Teiler ist von P, also können wir $P = (x - \lambda)\tilde{P}$ schreiben mit $\deg(\tilde{P}) = n - 1$. (Versuchen Sie einen schönen Induktions-Beweis zu schreiben, Forum-Wieso.) □

Dies bedeutet, dass wir keine weiteren Erweiterungen von Zahlen mehr brauchen, falls wir Polynom-Gleichungen lösen möchten: Alle Lösungen sind schon in $\mathbb{C}$. Ausdrücke wie in (1.16) nennt man eine Zerlegung von P in lineare Faktoren.

Reelle Polynome

In $\mathbb{R}$ ist die Sache etwas komplizierter: Es gibt Polynome $P \in \mathbb{R}[x]$ mit $\deg(P) > 1$, die keine Nullstellen über $\mathbb{R}$ haben, das heisst es gibt kein $\lambda \in \mathbb{R}$ mit $P(\lambda) = 0$.

Zum Beispiel ist $P = x^2 + 1$ ein solches Polynom. Schlechter kann es jedoch eigentlich nicht werden, wie wir jetzt erklären.

[13] Eine Antwort sind die Quaternionen.

Lemma 1.4.20. *Sei $p \in \mathbb{R}[x] \subseteq \mathbb{C}[x]$ und $\lambda \in \mathbb{C}$ eine Nullstelle von P. Dann ist $\overline{\lambda}$ auch eine Nullstelle von P. (Hier steht $\overline{\lambda}$ für die komplexe Konjugierte von λ.)*

Beweis. Sei $P(x) = a_n x^n + \ldots + a_1 x + a_0$. Laut Annahme gilt

$$P(\lambda) = a_n \lambda^n + \ldots + a_1 \lambda + a_0 = 0. \tag{1.17}$$

Erinnern Sie sich an die Eigenschaften der komplexen Konjugation aus der Analysis: Für alle $\alpha, \beta \in \mathbb{C}$ gilt:

- $\overline{\alpha \cdot \beta} = \overline{\alpha} \cdot \overline{\beta}$,
- $\overline{\alpha + \beta} = \overline{\alpha} + \overline{\beta}$,
- $\overline{\alpha} = \alpha$ genau dann, wenn $\alpha \in \mathbb{R}$.

Wir konjugieren Gleichung (1.17) und benutzen obige Eigenschaften, so dass wir

$$\begin{aligned}
\overline{P(\lambda)} = \overline{a_n \lambda^n + \ldots + a_1 \lambda + a_0} &= \overline{a_n}\,\overline{\lambda^n} + \ldots + \overline{a_1}\,\overline{\lambda} + \overline{a_0} \\
&= \underbrace{a_n \overline{\lambda}^n + \ldots + a_1 \overline{\lambda} + a_0}_{P(\overline{\lambda})} \\
&= \overline{0} \\
&= 0.
\end{aligned}$$

Also ist $P(\overline{\lambda}) = 0$, was wir zeigen wollten. □

Folgendes Lemma ist eine stärkere Version von Lemma 1.4.20.

Lemma 1.4.21. *Sei $\lambda \in \mathbb{C} \backslash \mathbb{R}$ und $q_\lambda := (x - \lambda)(x - \overline{\lambda})$. Dann gilt $q_\lambda \in \mathbb{R}[x]$, $\deg(q_\lambda) = 2$ und q_λ hat keine reelle Nullstelle.*

Beweis. Wir lösen die Klammer auf

$$q_\lambda := (x - \lambda)(x - \overline{\lambda}) = x^2 - (\lambda + \overline{\lambda})x + \lambda\overline{\lambda}.$$

Da $\overline{\lambda + \overline{\lambda}} = \overline{\lambda} + \lambda = \lambda + \overline{\lambda}$ und $\overline{\lambda \cdot \overline{\lambda}} = \overline{\lambda} \cdot \overline{\overline{\lambda}} = \overline{\lambda} \cdot \lambda = \lambda \cdot \overline{\lambda}$ gilt, folgt $q_\lambda \in \mathbb{R}[x]$. Aus der Definition von q_λ folgt, dass $\deg(q_\lambda) = 2$. Für die letzte Aussage beachten Sie, dass λ und $\overline{\lambda}$ zwei Nullstellen von q_λ sind, die nicht in $\mathbb{R}$ sind. (Dies gilt, weil $\lambda \in \mathbb{C} \setminus \mathbb{R}$ impliziert $\overline{\lambda} \in \mathbb{C} \setminus \mathbb{R}$. Wieso gilt die Implikation?) Aus Korollar 1.4.19 folgt nun, dass q_λ keine anderen Nullstellen haben kann, also insbesondere auch keine reelle Nullstelle. □

Jetzt können wir eine stärkere Version von Lemma 1.4.20 beweisen:

Lemma 1.4.22. *Sei $P \in \mathbb{R}[x]$ und $\lambda \in \mathbb{C} \setminus \mathbb{R}$. Dann ist $\mu(P \mid \lambda) = \mu(P \mid \overline{\lambda})$.*

Beweis. Wir beweisen das Lemma per Induktion über $\mu(P \mid \lambda)$.

Falls $\mu(P \mid \lambda) = 0$, dann folgt $\mu(P \mid \overline{\lambda}) = 0$ aus Lemma 1.4.20. (Wieso? $\lambda = \overline{\overline{\lambda}}$.)

Wir nehmen jetzt an, dass $\mu(\tilde{P} \mid \tilde{\lambda}) = n$ genau dann, wenn $\mu(\tilde{P} \mid \overline{\tilde{\lambda}}) = n$ für alle $\tilde{P} \in \mathbb{R}[x]$ mit Nullstelle $\tilde{\lambda} \in \mathbb{C} \setminus \mathbb{R}$ und $\mu(P \mid \lambda) = n + 1$. Per Definition von $\mu(P \mid \lambda)$ ist $P(\lambda) = 0$ und nach Lemma 1.4.20 gilt auch $P(\overline{\lambda}) = 0$. Nach Korollar 1.4.15 sind $(x - \lambda)$ und $(x - \overline{\lambda})$ Teiler von P. Da $\lambda \in \mathbb{C} \setminus \mathbb{R}$ ist, gilt $\lambda \neq \overline{\lambda}$ und daher ist

$$q_\lambda = (x - \lambda)(x - \overline{\lambda}) \in \mathbb{R}[x]$$

ein Teiler von P. Also können wir $P = q_\lambda P'$ für ein $P' \in \mathbb{R}[x]$ schreiben und per Definition der Vielfachheit folgt

$$\mu(P \mid \lambda) = \mu(P' \mid \lambda) + 1 \quad \text{und} \quad \mu(P \mid \overline{\lambda}) = \mu(P' \mid \overline{\lambda}) + 1. \tag{1.18}$$

Laut Annahme ist $\mu(P \mid \lambda) = n+1$ und daher ist $\mu(P' \mid \lambda) = n$. Die Induktionsannahme impliziert dann $\mu(P' \mid \overline{\lambda}) = n$. Aus (1.18) folgt dann $\mu(P \mid \overline{\lambda}) = n + 1$, was wir zeigen wollten. □

Korollar 1.4.23. *Für jedes $P \in \mathbb{R}[x]$ existieren $k, l \in \mathbb{N} \cup \{0\}$ und $\eta_1, \ldots, \eta_k \in \mathbb{R}$ und $\lambda_1, \ldots, \lambda_l \in \mathbb{C} \setminus \mathbb{R}$, so dass $\{\eta_1, \ldots, \eta_k, \lambda_1, \ldots, \lambda_l, \overline{\lambda_1}, \ldots, \overline{\lambda_l}\}$ alle Nullstellen von P sind und $\deg(P) = k + 2l$.*

In anderen Worten sagt Korollar 1.4.23, dass die Nullstellen, die nicht reell sind in Paaren λ und $\overline{\lambda}$ vorkommen.

Korollar 1.4.24. *Sei $P \in \mathbb{R}[x]$ mit Leitkoeffizient $a \in \mathbb{R}$ und $k, l \in \mathbb{N} \cup \{0\}$, $\eta_1, \ldots, \eta_k \in \mathbb{R}$, $\lambda_1, \ldots, \lambda_l \in \mathbb{C} \setminus \mathbb{R}$ wie in Korollar 1.4.23. Dann gilt*

$$P(x) = a(x - \eta_1) \cdot \ldots \cdot (x - \eta_k) \cdot q_{\lambda_1} \cdot \ldots \cdot q_{\lambda_l} = \prod_{i=1}^{k} (x - \eta_i) \prod_{i=1}^{l} q_{\lambda_i}$$

und dieses Produkt kann man nicht weiter in $\mathbb{R}[x]$ zerlegen.

Changelog: Kapitel 1

- 19.02: Nummerierungen wurden geändert.

Kapitel 2

Vektorräume

Wir besprechen endlich die Hauptobjekte der linearen Algebra, Vektorräume. In unserer Fibonacci-Einführung haben wir schon die grobe Definition von Vektorräumen gesehen: Mengen mit einer gewissen Addition und Skalarmultiplikation. Seither sind wir schon viel besser ausgerüstet mit der Sprache der abstrakten Mathematik. Daher definieren wir:

Definition 2.0.1. Ein *Vektorraum* über einem Körper K ist eine Menge $(V, +, \cdot)$ mit zwei Verknüpfungen

$$\begin{aligned} + &: V \times V \to V, (v_1, v_2) \mapsto v_1 + v_2 \\ \cdot &: K \times V \to V, (a, v) \mapsto a \cdot v = av, \end{aligned}$$

so dass folgende Axiome gelten:

(V1) $\forall v_1, v_2, v_3 \in V : v_1 + (v_2 + v_3) = (v_1 + v_2) + v_3$

(V2) $\exists 0 = 0_V \in V \, \forall v \in V : 0 + v = v$

(V3) $\forall v \in V \, \exists v' \in V : v + v' = 0$

(V4) $\forall v_1, v_2 \in V : v_1 + v_2 = v_2 + v_1$

(V5) $\forall a, b \in K, v \in V : a \cdot (b \cdot v) = (a \cdot b) \cdot v$

(V6) $\forall v \in V : 1 \cdot v = v$

(V7) $\forall a \in K, v_1, v_2 \in V : a \cdot (v_1 + v_2) = a \cdot v_1 + a \cdot v_2$

(V8) $\forall a_1, a_2 \in K, v \in V : (a_1 + a_2) \cdot v = a_1 \cdot v + a_2 \cdot v$

Wie Sie sehen können, benutzt Definition 2.0.1 nicht direkt das Konzept einer Gruppe. Wir bemerken lediglich: Die Axiome (V1) bis (V3) implizieren, dass $(V, 0, +)$ eine Gruppe ist und (V4) sagt, dass V eine abelsche Gruppe ist. Das Axiom (V5) beschreibt die Kompatibilität der Skalarmultiplikation mit der Multiplikation im Körper[1] K, in (V6) ist 1 die multiplikative Identität im Körper K und die Axiome (V7) und (V8) beschreiben schlussendlich Distributivitätsregeln.

Die Elemente eines Vektorraums V über einem Körper K nennen wir Vektoren und die Elemente des Körpers K heissen Skalare. Ein Vektorraum über $\mathbb{R}$ (bzw. über $\mathbb{C}$) heisst ein reeller (bzw. ein komplexer) Vektorraum. Genau wie bei Körpern können wir gewisse „Arithmetik-Regeln" aus den Axiomen herleiten.

Lemma 2.0.2. *Sei V ein Vektorraum über einem Körper K. Dann gilt:*

(a) Der Nullvektor in (V2) ist eindeutig und wird mit 0_V bezeichnet, wenn Verwechslungsgefahr besteht.

(b) Die additive Inverse in (V3) ist eindeutig und wird mit $-v$ bezeichnet. Damit bedeutet $w - v := w + (-v)$ für $w \in V$.

(c) Für alle $v \in V$ gilt $0 \cdot v = 0$ (oder präziser: $0_K \cdot v = 0_V$).

(d) Für alle $a \in K$ gilt $a \cdot 0 = 0$ (oder präziser: $a \cdot 0_V = 0_V$).

(e) Für alle $v \in V$ gilt $-1 \cdot v = -v$.

(f) Für alle $v \in V$ gilt $-(-v) = v$.

(g) Für alle $a \in K, v \in V$ mit $av = 0$ folgt, dass $a = 0$ oder $v = 0$.

(h) Assoziativität gilt auch für die Summe von n Elementen.

Beweis. Wir lassen fast alle Teile des Lemmas als Übungen, da wir ähnliche Beweise schon mit Körpern gemacht haben. Bemerken Sie, dass (a), (b) und (f) direkt aus Tatsachen über Gruppen folgen, die wir schon bewiesen haben. (Wieso?) Zum Spass beweisen wir (c): Sei $v \in V$, dann gilt

$$0 \cdot v = (0 + 0) \cdot v \overset{(V8)}{=} 0 \cdot v + 0 \cdot v. \tag{2.1}$$

Wir addieren $-0 \cdot v$ auf beide Seiten und erhalten

$$0 \overset{(V3)}{=} 0 \cdot v + (-0 \cdot v) = (0 \cdot v + 0 \cdot v) + (-0 \cdot v) \overset{(V1)\&(V3)}{=} 0 \cdot v + 0 = 0 \cdot v.$$

Vergleichen Sie dies mit dem Beweis von Lemma 1.3.4.

[1] In (V5) ist $(a \cdot b)$ die Multiplikation im Körper und $(b \cdot v)$ bzw. $a \cdot (b \cdot v)$ ist die Verknüpfung $\cdot : K \times V \to V$ in Definition 2.0.1!

Für (g) brauchen wir eine neue Idee: Sei $a \in K$ und $v \in V$. Wir nehmen an, dass $a \neq 0$ und $av = 0$. Da K ein Körper ist, gibt es $a^{-1} \in K$ und

$$v \overset{(*)}{=} 1 \cdot v = (a^{-1}a) \cdot v \overset{(*)}{=} a^{-1}(av) = a^{-1} \cdot 0 \overset{(c)}{=} 0.$$

(Erklären Sie für sich selbst, welche Axiome jeweils in $(*)$ benutzt wurden und wie die „Aussagen-Logik“ in diesem Beweis funktioniert!) Für (h) bemerken wir, dass man diese Aussage mit Induktion beweisen kann, wobei man (V1) für den Induktionsschritt verwendet. $\square$

Wir geben jetzt (nur) ein Beispiel, welches sowohl einfach als auch zentral ist[2]. Danach definieren wir Untervektorräume und geben viele Beispiele von Vektorräumen und ihren Untervektorräumen.

Beispiel 2.0.3 (Der Koordinatenraum K^n)**.** Sei K ein Körper und $n \in \mathbb{N}$. Als eine ziemlich direkte Verallgemeinerung der René Descartes Ideen definieren wir den *Koordinatenraum* (über K) durch

$$K^n = \{(a_1, \dots, a_n) \mid a_i \in K, 1 \leq i \leq n\}$$

Die Addition auf K^n ist durch komponentenweise Addition definiert: Für $v = (a_1, \dots, a_n)$ und $w = (b_1, \dots, b_n) \in K^n$ definieren wir

$$v + w = (a_1, \dots, a_n) + (b_1, \dots, b_n) := (a_1 + b_1, \dots, a_n + b_n). \tag{2.2}$$

Hier ist die Addition auf der linken Seite die Addition in K^n, die wir gerade definieren und die Addition auf der rechten Seite ist die Addition in K. In anderen Worten kann man (2.2) auch so schreiben:

$$v +_{K^n} w = (a_1, \dots, a_n) +_{K^n} (b_1, \dots, b_n) := (a_1 +_K b_1, \dots, a_n +_K b_n).$$

Wir benutzen aber die Schreibweise (2.2) , um den Leser zu zwingen sich immer wieder zu fragen, was genau was bedeutet. Skalarmultiplikation ist ebenfalls komponentenweise definiert: Für $a \in K$ ist

$$a \cdot v = a \cdot (a_1, \dots, a_n) := (aa_1, \dots, aa_n).$$

Mit diesen Definitionen folgen alle Axiome der Vektorraumstruktur auf K^n aus den entsprechenden Axiomen der Körperstruktur auf K. (Verifizieren Sie einige/alle! Beispielsweise ist $0 := (0, \dots, 0) \in K^n$ der Nullvektor[3].)

[2] Später werden Sie wohl überrascht sein, wie zentral dieses Beispiel ist!

[3] Auch hier könnte man $0_{K^n} = (0_K, \dots, 0_K)$ schreiben, aber dann müssten die Leser gar nicht denken, oder?

Bemerkung 2.0.4. Streng genommen sind die folgenden Vektorräume nicht gleich

$$K^n_{Zeil} := \{(x_1, \ldots, x_n) \mid x_i \in K\} \quad \text{und} \quad K^n_{Spal} := \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mid x_i \in K \right\},$$

obwohl wir bei beiden meistens an K^n denken. Wenn der Unterschied nicht wichtig ist, schreiben wir einfach K^n und benutzen die Notation von K^n_{Zeil}, da sie angenehmer ist zum Schreiben. Manchmal ist dieser Unterschied für uns aber dennoch wichtig. Zum Beispiel müssen wir später bei K^n an K^n_{Spal} denken, wenn wir das Produkt von Matrizen definieren.

Der Koordinatenraum heisst auch der *n-dimensionale Koordinatenraum über K*. Der Leser könnte schon jetzt zustimmen, dass K^n n Dimensionen hat. Was genau damit gemeint ist, müssen wir noch erklären, da wir die Dimension eines Vektorraums noch nicht definiert haben. Die Überraschung wird sein, dass K^n in einem gewissen Sinn der einzige Vektorraum über K mit Dimension n ist! Präziser gesagt, jeder andere Vektorraum über K mit Dimension n ist isomorph zu K^n. Um dort jedoch anzugelangen, müssen wir mit der Plauderei aufhören und weiterreisen…

Definition 2.0.5. Sei V ein Vektorraum über K. Eine Teilmenge $W \subseteq V$ heisst ein *Untervektorraum* von V, falls das Folgende gilt:

(UVR1) W ist nicht leer.

(UVR2) Für alle $w_1, w_2 \in W$ ist $w_1 + w_2 \in W$.

(UVR3) Für alle $a \in K, w \in W$ ist $aw \in W$.

Übung 2.0.6. (a) Sei $0 \in V$ der Nullvektor und (UVR1') die Aussage $0 \in W$. Zeigen Sie, dass ((UVR1') und (UVR2) und (UVR3)) äquivalent ist zu ((UVR1) und (UVR2) und (UVR3)). In Aussagenlogik:

$$\text{(UVR1')} \wedge \text{(UVR2)} \wedge \text{(UVR3)} \iff \text{(UVR1)} \wedge \text{(UVR2)} \wedge \text{(UVR3)}.$$

(b) Zeigen Sie, dass die Aussage

$$\forall a_1, a_2 \in K \ \forall w_1, w_2 \in W : a_1 w_1 + a_2 w_2 \in W$$

äquivalent ist zu

$$\text{(UVR2)} \wedge \text{(UVR3)}.$$

Lemma 2.0.7. *Sei V ein Vektorraum über K und $W \subseteq V$ ein Untervektorraum. Dann ist W auch ein Vektorraum über K mit der induzierten Addition und Skalarmultiplikation von V.*

Beweis. Der Beweis besteht aus der direkten Verifizierung aller Axiome mittels (UVR1), (UVR2) und (UVR3) und der Benutzung von $W \subseteq V$. Beweisen wir zum Beispiel das Axiom (V2) für W: Aus Übung 2.0.6 folgt, dass $0_V \in W$. Wir behaupten, dass 0_V auch der Nullvektor von W ist: Sei $w \in W$. Da $W \subseteq V$ ist, gilt $0_V + w = w$ (aus (V2) für V). Dies zeigt (V2) für W. Wir überlassen die anderen Axiome dem Leser zum Überprüfen. □

Bemerkung 2.0.8. Lemma 2.0.7 ist eigentlich äquivalent zu der Definition 2.0.5. Präziser gesagt: Man könnte umgekehrt einen Untervektorraum folgendermassen definieren: Eine Teilmenge $W \subseteq V$ eines Vektorraums V heisst ein Unterverktorraum, wenn W ein Vektorraum mit der induziereten Addition und Skalarmultiplikation aus V ist. Dann könnte man beweisen, dass dies (UVR1)-(UVR3) impliziert. Zusammen mit Lemma 2.0.7 zeigt dies die Äquivalenz der beiden Definitionen.

2.1 Ein Haufen Beispiele und ein bisschen Theorie

In diesem Unterkapitel geben wir einen Haufen Beispielen und ein kleines Theorie-Intermezzo.

Triviale Beispiele

Triviale Beispiele sind eine sehr wichtige Art von Beispielen! Diese können helfen das „Kleingedruckte“ einer Definition zu verstehen.

Beispiel 2.1.1. Sei V ein Vektorraum. Dann sind $\{0_V\}$ und V selbst Untervektorräume.

Bemerkung 2.1.2. Sie haben diese Woche interessante Multiple Choice Fragen im Zusammenhang mit diesem Beispiel: Entscheiden Sie, ob die folgenden Aussagen wahr oder falsch sind.

- Jeder Vektorraum hat zwei verschiedene Untervektorräume.
- Jeder Körper hat zwei verschiedene Elemente.
- Jeder Vektorraum hat zwei verschiedene Elemente.

Bitte bearbeiten Sie auch die Multiple Choice Serie!

Untervektorräume von K^n

Beispiel 2.1.3. Sei $b \in K$. Die Menge

$$U_b = \{(x_1, x_2, x_3) \mid x_1 - x_2 + x_3 = b\} \subseteq K^3$$

ist ein Untervektorraum von K^3 genau dann, wenn $b = 0$. Wir zeigen nur „$\Rightarrow$“ mit Kontraposition: Falls $b \neq 0$ und $(x_1, x_2, x_3), (y_1, y_2, y_3) \in U_b$, dann ist

$$(z_1, z_2, z_3) = (x_1, x_2, x_3) + (y_1, y_2, y_3)$$

nicht in U_b, da

$$z_1 - z_2 + z_3 = b + b \neq b.$$

(Beachten Sie: Die letzte Ungleichung gilt in jedem Körper K. Wieso?)

Theorie-Intermezzo

Wir möchten das letzte Beispiel verallgemeinern und dabei ein grundlegendes Beispiel für Untervektorräume von K^n besprechen. Wir wiederholen die Definition einer Matrix über einem allgemeinen Körper K:

Definition 2.1.4. Sei K ein Körper und seien $m, n \in \mathbb{N}$. Eine $m \times n$ *Matrix*

$$A = (a_{ij})_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$$

über K ist eine rechteckige Anordnung von Elementen in K mit m Zeilen und n Spalten. Die Skalare a_{ij} heissen die *Einträge* der Matrix A. Wir schreiben $M_{m \times n}(K)$ für die Menge aller $m \times n$ Matrizen über K.

Später auf Seite 54 und in der Übungsstunde werden wir sehen, dass $M_{m \times n}(K)$ auch ein wichtiger Vektorraum ist.

Im Moment möchten wir aber nur eine Multiplikation zwischen $m \times n$ Matrizen und Vektoren (geschrieben als Spaltenvektoren) in

$$K^n_{Spal} := \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mid x_i \in K \right\}$$

definieren. Sei $A = (a_{ij})_{ij} \in M_{m \times n}(K)$ und $v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in K^n_{Spal}$.[4] Wir definieren

$$Av = A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + \ldots + a_{1n}x_n \\ a_{21}x_1 + \ldots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n \end{pmatrix}. \tag{2.3}$$

[4] Zur Vereinfachung der Notation schreiben wir oft $A = (a_{ij})_{ij}$ statt $A = (a_{ij})_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$.

Bemerken Sie, dass $Av \in K^m_{Spal}$. Man zeigt mit der Definition in (2.3), dass für alle $v, w \in K^n_{Spal}$ und für alle $a \in K$ Folgendes gilt:

$$A(v + w) = Av + Aw \quad \text{und} \quad A(av) = aA(v). \tag{2.4}$$

Definition 2.1.5. Seien $A \in M_{m\times n}(K), x \in K^n (= K^n_{Spal})$ und $b \in K^m (= K^m_{Spal})$. Eine Gleichung der Form $Ax = b$ heisst ein *lineares Gleichungssystem* über K.

Bemerkung 2.1.6. Ein solches lineares Gleichungssystem können Sie mittels Gauss-Elimination über K lösen, genau wie Sie dies in der Serie und in der Übungsstunde gemacht haben.

Jetzt sind wir bereit, eine Verallgemeinerung von Beispiel 2.1.3 zu geben.

Beispiel 2.1.7 (Lösungen von linearen Gleichungssystemen)**.** Sei $A \in M_{m\times n}(K)$ und $b \in K^m (= K^m_{Spal})$. Dann ist

$$L = L_{A,b} = \{x \in K^n \mid Ax = b\} \subseteq K^n$$

ein Untervektorraum genau dann, wenn $b = 0 \in K^m$.

Beweis. „$\Leftarrow$“: Sei $b = 0 = 0_{K^m}$. Wir überprüfen die Axiome eines Untervektorraums:

- $0 = 0_{K^n} \in L$, da $A0_{K^n} = 0_{K^m}$.
- Seien $v_1, v_2 \in L$ und $a_1, a_2 \in K$. Laut (2.4) gilt, dass $A(a_1v_1 + a_2v_2) = a_1Av_1 + a_2Av_2 = a_10_{K^m} + a_20_{K^m} = 0_{K^m}$.

Daher ist L ein Untervektorraum. Für „$\Rightarrow$“ ist der Beweis sehr ähnlich zu dem Beweis in Beispiel 2.1.3. Merken Sie sich lediglich, dass in K^n auch gilt, dass $0 \neq b \in K^m$ die Ungleichung $b + b \neq b$ impliziert. (Wieso?) □

Dieses Beispiel ist wirklich grundlegend: Wir werden bald zeigen, dass jeder Untervektorraum von K^n die Form $L_{A,b}$ für gewisse A und b hat. Dies ist auch der Grund dafür, dass man durch Gauss-Elimination die ganze lineare Algebra auf K^n erhalten kann. Dies allein wäre aber nicht sehr schön. . .

Bemerkung 2.1.8. Wir werden später noch sehen, dass $L_{A,b}$ auch viel Struktur hat, wenn $b \neq 0$. (Hinweis: Es gilt $L_{A,b} = s + L_{A,0} = \{s + v \mid v \in L_{A,0}\}$ für irgendein $s \in L_{A,b}$. Daher ist $L_{A,b}$ ein „verschobener“ Untervektorraum.)

Hier ist ein einfaches Beispiel, das wir später nochmals betrachten:

Beispiel 2.1.9. Die Menge $W = \{(x_1, \ldots, x_6) \mid x_1 + \ldots + x_6 = 0\} \subseteq K^6$ ist ein Untervektorraum.

Beweis. Es gilt $W = L_{A,0}$ für $A = (1\ 1 \cdots 1) \in M_{1,6}(K)$. □

Folgenräume

In diesem (Unter-)Kapitel stellen wir einen Zusammenhang zu unserer Fibonacci-Einführung her. Sei K ein Körper. Wir definieren den *Folgenraum* K^∞ durch

$$K^\infty = \{(a_i)_{i=1}^\infty \mid a_i \in K,\ i = 1, 2, \ldots\}.$$

Wie in der Einführung definieren wir die Addition und Skalarmultiplikation folgendermassen:

$$(a_1, a_2, \ldots) + (b_1, b_2, \ldots) := (a_1 + b_1, a_2 + b_2, \ldots), \tag{2.5}$$

$$a \cdot (a_1, a_2, \ldots) := (aa_1, aa_2, \ldots). \tag{2.6}$$

Behauptung: K^∞ mit der Addition und Skalarmultiplikation in (2.5) ist ein Vektorraum.

Übung 2.1.10. Überprüfen Sie diese Behauptung, bis sie für Sie so offensichtlich wird wie die Tatsache, dass K^n ein Vektorraum ist.

Bemerkung 2.1.11. Hier indizieren wir die Folgen mit den natürlichen Zahlen $\mathbb{N}$, so dass die Folgen in K^∞ also beim Index 1 starten. Wir könnten natürlich auch die Indizes aus $\mathbb{N} \cup \{0\}$ wählen wie in der Einführung oder allgemeiner könnten wir bei einer beliebigen Zahl in $\mathbb{Z}$ starten.

Eine ähnliche, aber andere Konstruktion, sind zweiseitige Folgen: Sei

$$\begin{aligned} K_{-\infty}^\infty &= \{(a_i)_{i \in \mathbb{Z}} \mid a_i \in K,\ i \in \mathbb{Z}\} \\ &= \{(a_i)_{i=-\infty}^\infty \mid a_i \in K,\ i \in \mathbb{Z}\} \\ &= \{(\ldots, a_{-1}, a_0, a_1, \ldots) \mid a_i \in K\}. \end{aligned}$$

Dies ist auch ein Vektorraum.

Untervektorräume von Folgenräumen

Zuerst, um auf unsere Einführung zurück zu kommen, betrachten wir **Fib**:

Beispiel 2.1.12. Sei **Fib** $= \{(a_i)_{i=1}^\infty \mid a_n = a_{n-1} + a_{n-2}$ für alle $n \geq 3\}$. Dann ist **Fib** ein Untervektorraum von K^∞. Allgemeiner gibt uns jede lineare Rekursions-Regel einen Untervektorraum. Zum Beispiel, seien $\alpha, \beta \in K$ zwei fixierte Zahlen. Dann ist

$$U_{\alpha,\beta} = \{(a_i)_{i=1}^\infty \mid a_n = \alpha a_{n-1} + \beta a_{n-2} \text{ für alle } n \geq 3\}$$

ein Untervektorraum von K^∞. Den Beweis davon haben wir eigentlich schon in der Einführung gesehen!

Übung 2.1.13. Ist $W = \{(a_i)_{i=1}^\infty \mid a_n = a_{n-1}^2 \text{ für alle } n \geq 2\} \subseteq K^\infty$ ein Untervektorraum? (Hinweis: Die Antwort hängt von etwas ab! (Siehe Forum-Wieso))

Beispiel 2.1.14. Für jedes $n \in \mathbb{N}$ definieren wir

$$W_n := \{(a_i)_{i=1}^\infty \mid a_i = 0 \text{ für alle } i > n\} \subseteq K^\infty.$$

Es gilt $W_n \subseteq W_m$ genau dann, wenn $n \leq m$. Wir definieren $W_\infty := \bigcup_{n\geq 1} W_n$.

Übung 2.1.15. Beweisen Sie, dass $W_\infty = \{(a_i)_{i=1}^\infty \mid \exists M \in \mathbb{N}\ \forall i > M : a_i = 0\}$ ist, und dass W_∞ ein Untervektorraum ist. Der Raum W_∞ heisst *Raum der Folgen mit endlichem Träger*[5].

Wir verallgemeinern diese Familie von Beispielen weiter.

Funktions-Räume

Wie immer sei K ein Körper, und sei S eine nicht leere Menge. Wir definieren

$$K^S = \{f : S \to K \mid f \text{ ist eine Funktion}\}.$$

Eine andere Notation ist also $K^S = \mathrm{Abb}(S, K)$. Wir möchten eine Addition und Skalarmultiplikation auf K^S definieren, um K^S mit einer Vektorraumstruktur zu versehen. Dazu kann man die Addition und Multiplikation auf K benutzen: Wir definieren für $f, g \in K^S$ und $a \in K$:

$$\begin{aligned} (f +_{K^S} g)(x) &= f(x) +_K g(x), \\ (a \cdot_{K^S} f)(x) &= a \cdot_K f(x). \end{aligned} \tag{2.7}$$

In anderen Worten ist $f + g$ die Funktion

$$\begin{aligned} f + g : S &\to K \\ x &\mapsto f(x) + g(x) \end{aligned}$$

und in ähnlicher Weise ist $a \cdot f$ die Funktion

$$\begin{aligned} a \cdot f : S &\to K \\ x &\mapsto a \cdot f(x). \end{aligned}$$

Bemerkung 2.1.16. Wir waren nett zu den Lesern und haben in (2.7) geschrieben, um welche Addition und Multiplikation es sich jeweils handelt. Auch die zweite Erklärung

[5]„Endlicher Träger“ bezieht sich auf die Tatsache, dass jeweils nur endlich viele Komponenten jedes Elements von W_∞ ungleich Null sind.

für die Definition in (2.7) war ein „Gefallen“ von unserer Seite. Immer alles genau zu erklären ist ein Bärendienst (und auch mühsam...). Deshalb und da die Leser eine gewisse Unabhängigkeit und Fähigkeit zur „Kompilierung“ eines mathematischen Texts entwickeln müssen, werden wir dies von jetzt an normalerweise vermeiden.

Nun können die Leser also überprüfen, dass K^S ein Vektorraum ist. (Der Beweis ist ähnlich zu dem Beweis zu K^n und folgt im Grossen und Ganzen aus den Eigenschaften des Körpers K.)

Beispiel 2.1.17. Sei $s_0 \in S$. Dann ist

$$\{f \in K^S \mid f(s_0) = 0\}$$

ein Untervektorraum von K^S. Seien $k \in \mathbb{N}$, $s_1, \dots, s_k \in S$ und $\alpha_1, \dots, \alpha_k \in K$. Dann ist auch

$$\left\{f \in K^S \mid \sum_{i=1}^{k} \alpha_i f(s_i) = 0\right\}$$

ein Untervektorraum. In ähnlicher Weise (aber doch auf eine etwas andere Art) betrachten wir

$$\mathrm{Ann}(S') = \{f \in K^S \mid f(s) = 0 \text{ für alle } s \in S'\}$$

für $S' \subseteq S$. Dann ist $\mathrm{Ann}(S')$ ein Untervektorraum. Darf $S' = \emptyset$ sein? Wenn ja, was ist $\mathrm{Ann}(\emptyset)$? Was ist $\mathrm{Ann}(S)$?

Um weiter zu spielen, könnten wir definieren:

$$W = \{f \in K^S \mid f(s) = 0 \text{ für alle ausser endliche viele } s \in S\}.$$

Ist W ein Untervektorraum von K^S? Alle diese Fragen sind in einem „Forum-Wieso“.

Beispiel 2.1.18. Wenn S nicht nur irgendeine Menge ist, sondern noch eine andere Struktur hat, dann können wir oft Untervektorräume von K^S mittels dieser Struktur definieren. Zum Beispiel sei $S = [0, 1]$ das Intervall $\{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$. Dann ist

$$\{f \mid f \text{ stetig auf } [0, 1]\} \subseteq K^{[0,1]}$$

ein Untervektorraum[6] von $K^{[0,1]}$.

Bemerkung 2.1.19. Für $S = \{1, \dots, n\}$ ist K^S sehr ähnlich zu K^n, für $S = \mathbb{N}$ ist K^S sehr ähnlich zu K^∞ und für $S = \mathbb{Z}$ ist K^S sehr ähnlich zu $K^\infty_{-\infty}$. Um „sehr ähnlich“ präzise auszudrücken, brauchen wir einige Begriffe, die wir später definieren. In der Zwischenzeit können sich die fleissigen Leser selbst überzeugen, dass es zwischen den erwähnten Mengen oben eine Bijektion gibt.

[6] Stetigkeit werden Sie in der Analysis in ein paar Wochen behandeln.

Beispiel 2.1.20 (Ein Beispiel aus der Analysis)**.** Sei $K = \mathbb{R}$. Die Lösungen einer sogenannten (homogenen) linearen Differentialgleichung bilden einen Untervektorraum von $\mathbb{R}^{\mathbb{R}}$. Zum Beispiel ist

$$\{f : \mathbb{R} \to \mathbb{R} \mid f' = f\}$$

ein Untervektorraum von $\mathbb{R}^{\mathbb{R}}$. Mit $f' = f$ ist insbesondere implizit gemeint, dass die Ableitung von f existiert.[7]

Polynomräume

Sei K ein Körper. Auf $K[x]$ haben wir die Addition schon in Bemerkung 1.4.6 definiert. Wir definieren jetzt die Skalarmultiplikation auf $K[x]$. Seien $a \in K$ und $f = a_0 + a_1x + \ldots + a_nx^n \in K[x]$, dann definieren wir

$$a \cdot f = a \cdot (a_0 + a_1x + \ldots + a_nx^n) := aa_0 + aa_1x + \ldots + aa_nx^n.$$

Mit dieser Addition und Skalarmultiplikation ist $K[x]$ ein Untervektorraum (von K^K).

Beispiel 2.1.21. Sei $h \in \mathbb{N} \cup \{0, -\infty\}$. Dann ist

$$K[x]_h = \{p \in K[x] \mid \deg(p) \leq h\}$$

ein Untervektorraum sowohl von $K[x]$ als auch von K^K. Ist $\{p \in K[x] \mid \deg(p) = h\}$ auch ein Untervektorraum?

Beispiel 2.1.22. Die Menge $\left\{p \in K[x]_5 \mid p = a_0 + \ldots + a_5x^5 \text{ mit } \sum_{i=0}^{5} a_i = 0\right\}$ ist ein Untervektorraum von $K[x]_5$.

Sehen Sie eine Ähnlichkeit zwischen Beispiel 2.1.9 und Beispiel 2.1.22?

Matrizenräume

In $M_{m\times n}(K)$ kann man eine Addition und Skalarmultiplikation definieren, um eine Vektorraumstruktur zu erhalten. Wie das genau geht und Beispiele von Untervektorräumen zu diesem Vektorraum besprechen Sie in der Übungsstunde!

Die Potenzmenge als Vektorraum

Bevor wir die Theorie weiterentwickeln, besprechen wir ein Beispiel, bei dem die Vektorraumstruktur nicht ganz so offensichtlich ist.

Beispiel 2.1.23. Sei X eine Menge. Wir möchten eine Vektorraumstruktur über $\mathbb{F}_2$ auf der Potenzmenge $P(X)$ definieren. Nehmen Sie sich einige Minuten Zeit, um sich zu

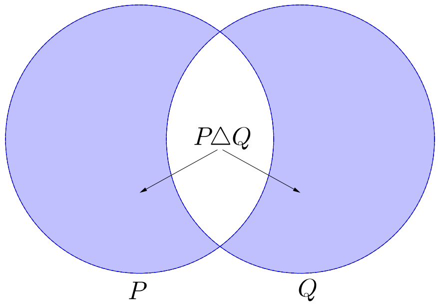
[7] Im Beispiel bezeichnet f' die Ableitung von f, mehr dazu werden Sie später in der Analysis sehen.

überlegen, was die Addition sein könnte. Dann überlegen Sie sich, was der Nullvektor ist. Wenn Sie dann wissen, was der Nullvektor ist, dann wissen Sie, wie die Skalarmultiplikation definiert sein muss: Laut den Axiomen eines Vektorraums und Lemma 2.0.2 gilt in jedem Vektorraum V

$$1 \cdot v = v \quad \text{und} \quad 0 \cdot v = 0$$

für alle $v \in V$. Aber $\mathbb{F}_2$ enthält nur 0 und 1! Zur Addition: Es gibt eine interessante symmetrische Operation auf $P(X)$, nämlich die symmetrische Differenz:
Für $A, B \in P(X)$ sei

$$A \triangle B = (A \cup B) \setminus (A \cap B).$$



Figur 2.1: Die symmetrische Differenz $P \triangle Q$.

Mann kann zeigen, dass für alle $A, B, C \in P(X)$ Folgendes gilt

$$A \triangle B = B \triangle A \quad \text{und} \quad (A \triangle B) \triangle C = A \triangle (B \triangle C).$$

Wenn wir $\triangle$ als Addition auf $P(X)$ nehmen, was wäre dann der Nullvektor? Wir suchen $S \subseteq X$ mit $A \triangle S = A$ für alle $A \in P(X)$. Eine kurze Überlegung führt uns dazu, dass

$$A \triangle \emptyset = A.$$

für alle $A \in P(X)$ gilt. Also, wenn $(P(X), \triangle, \cdot)$ einen Vektorraum über $\mathbb{F}_2 = \{\overline{0}, \overline{1}\}$ bilden soll, dann muss gelten

$$\overline{0} \cdot A = \emptyset \quad \text{und} \quad \overline{1} \cdot A = A.$$

Übung 2.1.24. Zeigen Sie, dass $(P(X), +, \cdot)$ ein Vektorraum über $\mathbb{F}_2$ ist.

Bemerkung 2.1.25. Wir empfehlen diese Übung mit allen Details zu lösen - siehe Serie 6!

2.2 Zurück zur Theorie

Nun da wir so viele Beispiele von Vektorräumen gesehen haben, ist es höchste Zeit zurück zur Theorie zu kehren. Jetzt müssen Sie ihren „Computer-Compiler“ immer benutzen: Wenn wir einen neuen Begriff definieren, wählen Sie eines der vorherigen Beispiele und überprüfen Sie mittels ihrem „Computer-Compiler“, ob der neue Begriff auf ihr gewähltes Beispiel zutrifft. Wir werden definieren:

- Der Span[8] einer Menge $S \subseteq V$, wobei V ein Vektorraum ist;
- Linearkombinationen;
- lineare Unabhängigkeit (und daher auch lineare Abhängigkeit);
- Basen und Dimensionen;
- Summen und (innere) direkte Summen.

2.2.1 Span und Linearkombinationen

In diesem Unterkapitel bezeichnet K immer einen Körper und V einen Vektorraum über K. Sei $S \subseteq V$ eine Teilmenge (die nicht unbedingt ein Untervektorraum ist). Wir können uns zwei Fragen stellen:

1. Welches ist der kleinste Untervektorraum von V, der S enthält?
2. Was bekommen wir, wenn wir die Menge aller Elemente betrachten, die wir mittels Addition und Skalarmultiplikation von Elementen aus S erreichen können?

Die erste Frage können wir mit diesem einfachen, aber wichtigen, Lemma beantworten:

Lemma 2.2.1. *Sei V ein Vektorraum über K, I eine Indexmenge und $(W_i)_{i\in I}$ eine Familie von Untervektorräumen. Dann ist*

$$W = \bigcap_{i\in I} W_i$$

ein Untervektorraum.

Beweis. Sei W wie im Lemma. Dann gilt:

- Der Nullvektor 0 liegt in W, da $0 \in W_i$ für alle $i \in I$.

[8] Dieser Begriff kommt ursprünglich aus dem Englischen. Im Deutschen kann man auch Erzeugnis sagen. Wir werden jedoch in der Regel den englischen Begriff verwenden.

- Seien $a_1, a_2 \in K$ und $v_1, v_2 \in W$. Dann sind $v_1, v_2 \in W_i$ für alle $i \in I$ per Definition von W. Da W_i ein Untervektorraum ist für alle $i \in I$, ist $a_1v_1 + a_2v_2 \in W_i$. Daher ist $a_1v_1 + a_2v_2 \in W$.

Dies zeigt, dass W in der Tat ein Untervektorraum von V ist. □

Definition 2.2.2 (Erste Definition des Spans). Sei $S \subseteq V$. Wir definieren den *Span* (auch *Spann*, oder *lineare Hülle* genannt) $\mathrm{Sp}(S)$ von S durch

$$\mathrm{Sp}(S) := \bigcap_{W \in \mathcal{N}} W,$$

wobei $\mathcal{N} = \{W \mid S \subseteq W,\, W \text{ ein Untervektorraum von } V\}$.

Eine direkte Folgerung ist:

Lemma 2.2.3. *Sei $S \subseteq V$ eine Teilmenge. Dann ist $\mathrm{Sp}(S)$ der kleinste Untervektorraum, der S enthält. Das heisst, es gilt:*

(1) $\mathrm{Sp}(S)$ ist ein Untervektorraum.

(2) Wenn $S \subseteq W$ für $W \subseteq V$ ein Untervektorraum, dann ist $\mathrm{Sp}(S) \subseteq W$.

Beweis. (1) Dies folgt aus Lemma 2.2.1.

(2) Dies folgt aus der Definition von $\mathrm{Sp}(S)$ als Durchschnitt aller Untervektorräume, die S enthalten.

Dies beweist das Lemma. □

Das ist schön und gut, gibt uns aber keine „konkrete" Beschreibung von $\mathrm{Sp}(S)$. Zu diesem Zweck definieren wir:

Definition 2.2.4. Sei $n \in \mathbb{N}$ und seien $a_1, \ldots, a_n \in K$ und $v_1, \ldots, v_n \in V$. Ein Vektor $v \in V$ der Form

$$v = a_1v_1 + \ldots + a_nv_n = \sum_{i=1}^{n} a_iv_i$$

heisst eine *Linearkombination von $v_1, \ldots, v_n$ (über K)*. Die Skalare $a_1, \ldots, a_n$ heissen die *Koeffizienten* der Linearkombination.

Bemerkung 2.2.5 (Wichtige Bemerkung). Am Anfang von Definition 2.2.4 steht $n \in \mathbb{N}$. Das heisst bei einer Linearkombination kommen (nur) endlich viele Elemente in der Summe vor![9]

[9] Nachdem Sie mehr Analysis gelernt haben, könnte man auch Linearkombinationen von unendlich vielen Elementen betrachten, siehe zum Beispiel Fourierreihen.

Lemma 2.2.6. *Sei $S \subseteq V$ eine (nicht leere) Teilmenge. Dann gilt*

$$\begin{aligned}\mathrm{Sp}(S) &= \{a_1v_1 + \ldots + a_nv_n \mid n \in \mathbb{N}, a_i \in K, v_i \in S \text{ für alle } 1 \leq i \leq n\} \\ &= \{\text{alle Linearkombinationen von Vektoren aus } S\}.\end{aligned}$$

Bemerkung 2.2.7. Die Menge S kann unendlich sein. Nichtsdestotrotz benutzt jede Linearkombination von S nur endlich viele Elemente von S.

Beweis von Lemma 2.2.6. Sei

$$\widetilde{\mathrm{Sp}}(S) := \{a_1v_1 + \ldots + a_nv_n \mid n \in \mathbb{N}, a_i \in K, v_i \in S \text{ für alle } 1 \leq i \leq n\}.$$

Wir zeigen:

(a) $\widetilde{\mathrm{Sp}}(S)$ ist ein Untervektorraum.

(b) Für jeden Untervektorraum $W \subseteq V$ mit $S \subseteq W$ gilt $\widetilde{\mathrm{Sp}}(S) \subseteq W$.

Daraus folgt $\widetilde{\mathrm{Sp}}(S) = \mathrm{Sp}(S)$, was wir zeigen wollten. (Wieso?) Wir beweisen zuerst (a). Seien v und w zwei Linearkombinationen von Vektoren aus S, das heisst, $v, w \in \widetilde{\mathrm{Sp}}(S)$ und wir können dementsprechend schreiben

$$v = a_1v_1 + \ldots + a_nv_n \quad \text{und} \quad w = b_1w_1 + \ldots + b_mw_m,$$

wobei $n, m \in \mathbb{N}$, $a_i, b_i \in K$ und $v_i, w_j \in S$ für alle $1 \leq i \leq n$ und $1 \leq j \leq m$. Seien $\alpha, \beta \in K$. Dann ist

$$\alpha v + \beta w = \alpha a_1v_1 + \ldots + \alpha a_nv_n + \beta b_1w_1 + \ldots + \beta b_mw_m$$

auch eine Linearkombination von Vektoren aus S (mit $m + n \in \mathbb{N}$ Vektoren aus S). Daher ist $\alpha v + \beta w \in \widetilde{\mathrm{Sp}}(S)$ für alle $\alpha, \beta \in K$. Ausserdem, da $S \neq \emptyset$, sei $v \in S$.[10] Dann ist $0_K \cdot v$ eine Linearkombination von Vektoren aus S und daher ist $0_K \cdot v = 0_V \in \widetilde{\mathrm{Sp}}(S)$. Dies zeigt (a). Für Teil (b): Man zeigt mittels Induktion (unter Verwendung von UVR2 und UVR3 aus Definition 2.0.5)[11], dass für jeden Untervektorraum W Folgendes gilt: Falls $S \subseteq W$ ist, dann ist jede Linearkombination von Vektoren aus S in W enthalten. Das heisst, dass $\widetilde{\mathrm{Sp}}(S) \subseteq W$. □

Zwischenzeit-Zusammenfassung : Wir haben oben eigentlich gezeigt, dass $\mathrm{Sp}(S)$ zwei äquivalente Definitionen hat:

(1) Definition 2.2.2:
$$\mathrm{Sp}(S) := \bigcap_{W \in \mathcal{N}} W,$$

[10] Siehe Bemerkung 2.2.12.
[11] Zum Beispiel ist UVR3 die Basis der Induktion.

wobei $\mathcal{N} = \{W \mid S \subseteq W,\ W \text{ ein Untervektorraum von } V\}$.

(2) $\mathrm{Sp}(S) = \{\text{alle Linearkombinationen von Vektoren in } S\}$.

Sie können jetzt sowohl (1) als auch (2) als Definitionen des Spans $\mathrm{Sp}(S)$ betrachten.

Bevor wir Beispiele besprechen, geben wir noch drei kurze Definitionen und eine Bemerkung.

Definition 2.2.8 (Notation)**.** Seien $n \in \mathbb{N}$ und $v_1, \ldots, v_n \in V$. Wir schreiben

$$\mathrm{Sp}(v_1, \ldots, v_n) := \mathrm{Sp}(\{v_1, \ldots, v_n\}).$$

Übung 2.2.9. Zeigen Sie:

- $\mathrm{Sp}(v) = \{\alpha v \mid \alpha \in K\}$.
- $\mathrm{Sp}(v, w) = \{\alpha v + \beta w \mid \alpha, \beta \in K\}$.
- Allgmeiner: $\mathrm{Sp}(v_1, \ldots, v_n) = \{\alpha_1 v_1 \ldots + \alpha_n v_n \mid \alpha_i \in K \text{ für alle } 1 \leq i \leq n\}$.

(Ist dies nicht einfach die Definition? Hier ist es schwieriger zu wissen, was genau man zeigen soll als es zu zeigen.)

Definition 2.2.10. Wir sagen, dass V von $S \subseteq V$ *erzeugt* wird, falls $\mathrm{Sp}(S) = V$. In diesem Fall sagt man auch, dass S erzeugend (für V) ist oder, dass S den Vektorraum V aufspannt. Die Menge S nennt man dann ein *Erzeugendensystem* von V. Allgemeiner benutzt man dieselbe Terminologie wenn $\mathrm{Sp}(S) = W$ und $W \subseteq V$ ein Untervektorraum ist: Zum Beispiel sagt man, dass S den Untervektorraum W *erzeugt* (oder S spannt W auf etc.).[12]

Definition 2.2.11. Ein Vektorraum V heisst *endlich-dimensional*, falls es $S \subseteq V$ gibt, so dass $|S| < \infty$ und $\mathrm{Sp}(S) = V$.

Bemerkung 2.2.12. Wenn wir die Definition 2.2.2 von $\mathrm{Sp}(S)$ benutzen, dann ist $\mathrm{Sp}(\emptyset) = \{0_V\}$. (Wieso?) Wenn wir $\mathrm{Sp}(\emptyset)$ auch durch die Linearkombinations-Definition definieren wollen, könnten wir sagen, dass die „leere Summe“ immer 0_V ist. Das heisst, dass $\sum_{x \in \emptyset} x := 0$. Oder wir deklarieren einfach $\mathrm{Sp}(\emptyset) = \{0_V\}$. In jedem Fall definieren/folgern wir jedoch, dass

$$\mathrm{Sp}(\emptyset) = \{0_V\}.$$

Dies macht einige Formulierungen später einfacher.

[12] Auf Englisch würde man sagen: S spans W.

2.2.2 Beispiele

Beispiel 2.2.13 (Geometrisches Intermezzo)**.** Wir betrachten $\mathbb{R}^2$. Welche Untervektorräume gibt es in $\mathbb{R}^2$ und wie können wir sie erzeugen? Zuerst einmal gibt es $\{0_{\mathbb{R}^2}\} = \{(0,0)\}$. Für diesen Untervektorraum gilt $\mathrm{Sp}(\emptyset) = \mathrm{Sp}(0_{\mathbb{R}^2}) = \{0_{\mathbb{R}^2}\}$.

Sei nun $0 \neq v = (a,b) \in W \subseteq \mathbb{R}^2$, wobei $W \subseteq \mathbb{R}^2$ ein Untervektorraum ist. Dann wissen wir, dass $\mathrm{Sp}(v) = \{\alpha v \mid \alpha \in \mathbb{R}\} \subseteq W$. Geometrisch heisst das: $\mathrm{Sp}(v)$ ist die Gerade G_v durch v und den Ursprung $(0,0)$. Wir haben also gezeigt, dass $v \in W$ die Inklusion $G_v \subseteq W$ impliziert. Ausserdem sehen wir, dass

$$\mathrm{Sp}(v) = \mathrm{Sp}(w) \iff G_v = G_w \iff \exists\, \alpha \neq 0 : w = \alpha v$$

(Wieso?). Da $\mathrm{Sp}(v)$ immer ein Untervektorraum ist, zeigt dies auch, dass alle Geraden $\{G_v \mid 0 \neq v \in \mathbb{R}^2\}$ Untervektorräume von $\mathbb{R}^2$ sind.

Um die letzte Möglichkeit für einen Untervektorraum von $\mathbb{R}^2$ zu finden, könnte man jetzt Folgendes zeigen:

Übung 2.2.14. Sei W ein Untervektorraum von $\mathbb{R}^2$, $G_v \subseteq W$ für $v \neq 0$ und $w \in W$ mit $w \notin G_v$. Dann gilt $\mathrm{Sp}(v,w) = \mathbb{R}^2$.

Es gilt zum Beispiel, dass $\mathrm{Sp}((1,0),(0,1)) = \mathbb{R}^2$. Also ist $\mathbb{R}^2$ insbesondere endlichdimensional.
Zusammenfassung : $\mathbb{R}^2$ hat folgende Untervektorräume:

(1) $\{(0,0)\}$;

(2) G_v für $0 \neq v \in \mathbb{R}^2$;

(3) $\mathbb{R}^2$.

Oder geometrisch gesagt: (1) Der Ursprung der Ebene; (2) Alle Geraden in der Ebene; (3) Die Ebene selbst.

Bemerkung 2.2.15. Dies gilt über jedem Körper K, aber wir „verlieren“ die geometrische Intuition. Ausserdem finden wir später eine ähnliche Beschreibung von allen Untervektorräumen von $\mathbb{R}^n$ (oder K^n).[13]

Beispiel 2.2.16 (Gauss und lineare Vektoren)**.** Zuerst ein Beispiel in $\mathbb{Q}^2$:
Seien $v = (1,3), w = (7,73) \in \mathbb{Q}^2$. Dann ist $(11,137)$ eine Linearkombination von v und w:

$$-3 \cdot (1,3) + 2 \cdot (7,73) = (11,137).$$

Aber wie könnten wir die Skalare -3 und 2 finden? Oder wie können wir bestimmen, ob es solche Skalare überhaupt gibt? Mit Gauss!

[13] Vgl. Bemerkung 2.1.8.

Lemma 2.2.17. *Sei* $A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \in M_{m\times n}(K)$ *und* $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in K^n$. *Wir definieren die Spaltenvektoren* $v_1, \ldots, v_n$ *in* K^m *durch die Spalten von* A*:*

$$A = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix}.$$

Dann gilt

$$Ax = x_1 \cdot \begin{pmatrix} | \\ v_1 \\ | \end{pmatrix} + x_2 \cdot \begin{pmatrix} | \\ v_2 \\ | \end{pmatrix} + \cdots + x_n \cdot \begin{pmatrix} | \\ v_n \\ | \end{pmatrix} = \sum_{i=1}^{n} x_i v_i.$$

Beweis. Mit Definition 2.3 ist dies nur eine kurze Rechnung. □

Seien $v_1, \ldots, v_n \in K^m$ und $w \in K^m$. Wir wollen wissen, ob w eine Linearkombination von $v_1, \ldots, v_n$ ist. Anders gesagt, wollen wir entscheiden, ob $w \in \mathrm{Sp}(v_1, \ldots, v_n)$ ist. Dazu betrachten wir die Matrix

$$A = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix},$$

deren Spalten die Vektoren $v_1, \ldots, v_n$ sind. Jetzt betrachten wir das lineare Gleichungssystem[14]

$$Ax = w \quad \text{für } x \in K^n. \tag{$\star$}$$

Aus Lemma 2.2.17 folgt:

$$\begin{aligned} (\star) \text{ hat eine Lösung } x = (x_1, \ldots, x_n) &\iff Ax = x_1 v_1 + \ldots + x_n v_n = w \\ &\iff w \in \mathrm{Sp}(v_1, \ldots, v_n). \end{aligned}$$

Mehr dazu in der Übungsstunde.

Bemerkung 2.2.18. Eine „Baby-Version" der linearen Algebra bespricht die Vektorräume $\mathbb{R}^n$ und vor allem $\mathbb{R}^2$ und $\mathbb{R}^3$. Man sollte diese einfachen Fälle nicht unterschätzen und versuchen, alle Begriffe am Anfang in diesen Vektorräumen zu verstehen. So entwickelt man ein geometrisches Verständnis, welches auch in Vektorräumen „ohne Geometrie" nützlich ist. Spielen Sie mit Beispielen, um dieses geometrische Verständnis zu entwickeln. Sie können auch die schöne Video-Serie von 3Blue1Brown anschauen. Im jetzigen Zusammenhang ist besonders dieses Video relevant.

[14] Anders gesagt, machen wir Gauss mit $(A \mid w)$.

Beispiel 2.2.19. Seien

$$\varphi = \frac{1+\sqrt{5}}{2} \quad \text{und} \quad \psi = \frac{1-\sqrt{5}}{2}.$$

Dann sind

$$v = \left(1, \varphi, \varphi^2, \ldots\right) \quad \text{und} \quad w = \left(1, \psi, \psi^2, \ldots\right)$$

Elemente von **Fib** aus dem Beispiel 2.1.12. (Wieso?) In der Einfürung haben wir gezeigt, dass $\mathrm{Sp}(v, w) = \mathbf{Fib}$. Daher ist **Fib** endlich-dimensional. Das heisst, dass man jede Folge in **Fib** als Linearkombination von v und w schreiben kann. Im Sinne der Definition aus der Einführung heisst das, dass wir jedes Element von **Fib** gut kennen, da wir v und w gut kennen!

Beispiel 2.2.20 (Standard-Basen)**.** Es gibt gewisse „Standard-Wege“, um die Vektorräume, die wir bisher besprochen haben, zu erzeugen. Hier hat das Wort Standard keine präzise Bedeutung: man meint damit einfach, dass viele Leute diesen Weg benutzen, um diese Vektorräume zu erzeugen. Wir überlassen es dem Leser zu überprüfen, dass die angegebenen Mengen tatsächlich Erzeugendensysteme sind:

(1) Seien $e_1 = (1, 0, \ldots, 0), e_2 = (0, 1, 0, \ldots, 0), \ldots, e_n = (0, \ldots, 0, 1)$ Vektoren in K^n. Es gilt $K^n = \mathrm{Sp}(e_1, \ldots, e_n)$.

(2) Für $K[x]_n$ gilt $K[x]_n = \mathrm{Sp}(x^0, x^1, \ldots, x^n)$. Ausserdem gilt

$$K[x] = \mathrm{Sp}(x^0, x^1, \ldots) := \mathrm{Sp}\left(\left\{x^i \mid i \in \mathbb{N} \cup \{0\}\right\}\right).$$

Die letzte Tatsache lässt sich ebenfalls einfach überprüfen. Man muss sich lediglich daran erinnern, dass eine Linearkombination immer nur endlich viele Summanden hat!

(3) Seien $m, n \in \mathbb{N}$. Für $1 \leq i \leq m, 1 \leq j \leq n$ sei $E_{ij} = (a_{kl}) \in M_{m\times n}(\mathbb{R})$ die Matrix mit

$$a_{kl} = \begin{cases} 1, & \text{falls } k = i \text{ und } l = j \\ 0, & \text{sonst} \end{cases}.$$

Dann gilt

$$M_{m\times n}(\mathbb{R}) = \mathrm{Sp}(\{E_{ij} \mid 1 \leq i \leq m, 1 \leq j \leq n\})$$

Also ist zum Beispiel

$$M_{2\times 2} = \mathrm{Sp}\left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\right).$$

(4) Sei X eine Menge mit $|X| = n$. Dann gilt

$$P(X) = \mathrm{Sp}(\{\text{alle einelementigen Teilmengen von } X\}).$$

Bemerkung 2.2.21. Einige Autoren nennen die Basen in Beispiel 2.2.20 „kanonisch“ (vgl. Fischer [7, 1.5.1]). Dies ist ein Missbrauch des Wortes kanonisch. Nichts ist kanonisch an diesen Basen. Wir werden mehr dazu sagen, wenn wir ein Beispiel angeben, wo etwas wirklich kanonisch ist. Bitte vermeiden Sie es dieses Wort in Zusammenhang mit den Basen aus Beispiel 2.2.20 zu verwenden.

Übung 2.2.22. Nehmen Sie sich Zeit die Aussagen in Beispiel 2.2.20 zu überprüfen.

Übung 2.2.23. (1) Sei $i \in \mathbb{N}$. Wir definieren $e_i \in K^\infty$ als die Folge $(a_n)_{n\in\mathbb{N}}$ mit $a_i = 1$ und $a_j = 0$ für alle $j \neq i$. Gilt $\mathrm{Sp}(\{e_i \mid i \in \mathbb{N}\}) = K^\infty$? Wenn nicht, haben wir diesen Untervektorraum schon zuvor erwähnt?

(2) Sei X eine unendliche Menge. Gilt

$$P(X) = \mathrm{Sp}(\{\text{alle einelementigen Teilmengen von } X\})?$$

Wenn nicht, wie könnten Sie diesen Span in Worten beschreiben?

2.2.3 Lineare Unabhängigkeit und die Definition einer Basis

Einführung

Ein alternativer Titel dieses Abschnitts könnte sein: „Wie kann ich einen Vektorraum beherrschen/beschreiben/kontrollieren?“ oder weniger romantisch: „Wie könnte ich einen Vektorraum mit Koordinaten versehen?“ Dieselben Fragen könnten wir bezüglich Untervektorräumen stellen. Dies ist aber unnötig, da jeder Untervektorraum auch ein Vektorraum ist.

Also, was meinen wir mit „einen Vektorraum mit Koordinaten versehen“? Nehmen wir an, dass V ein Vektorraum über K ist und $V = \mathrm{Sp}(v_1, \ldots, v_n)$. Dann können wir jeden Vektor als Linearkombination von $v_1, \ldots, v_n$ schreiben: Sei $v \in V$. Dann existieren $a_1, \ldots, a_n \in K$ mit

$$v = a_1 v_1 + \ldots + a_n v_n. \tag{2.8}$$

Wenn diese Beschreibung eindeutig ist, dann könnte man bei den Skalaren

$$a_1, \ldots, a_n$$

an die „Koordinaten“ von v in der Beschreibung von v als Linearkombination von $v_1, \ldots, v_n$ denken. Wenn jeder Vektor eine eindeutige Beschreibung als Linearkombination von $v_1, \ldots, v_n$ hat, könnten wir durch $v_1, \ldots, v_n$ jeden Vektor in V mit „Koordinaten“ versehen. Dadurch könnten wir den Vektorraum V beschreiben/kontrollieren. Wenn dies der Fall ist, dann sagen wir, dass $v_1, \ldots, v_n$ eine *Basis* von V ist. Wir möchten also verstehen, welche Bedingung wir an $v_1, \ldots, v_n$ stellen müssen, damit wir wissen, dass jeder Vektor eine eindeutige Beschreibung als Linearkombination von $v_1, \ldots, v_n$ hat. Wir werden bald sehen, dass wir dies (überraschenderweise) schon durch eine mögliche Beschreibung des Nullvektors verstehen können.

Lineare Unabhängigkeit

Wie immer bezeichnet V einen Vektorraum über einem Körper K.

Definition 2.2.24. Sei $n \in \mathbb{N}$. Eine endliche Menge[15] $\{v_1, \ldots, v_n\} \subseteq V$ ist *linear unabhängig*, falls aus $a_1 v_1 + \ldots + a_n v_n = 0$ für $a_1, \ldots, a_n \in K$ stets folgt, dass $a_1 = 0, \ldots, a_n = 0$. In Quantoren ausgedrückt:

$$\forall\, a_1, \ldots, a_n \in K : a_1 v_1 + \ldots + a_n v_n = 0 \;\Rightarrow\; a_1 = 0, \ldots, a_n = 0.$$

Später werden wir in diesem Fall auch schreiben: die Liste $v_1, \cdots, v_n$ ist linear unabhängig (Siehe Bemerkung 2.3.1). Ausserdem ist die leere Menge $\emptyset$ linear unabhängig per Konvention.

Bemerkung 2.2.25. Der Nullvektor hat immer die folgende Beschreibung als Linearkombination von $v_1, \ldots, v_n$:

$$0 \cdot v_1 + \ldots + 0 \cdot v_n = 0. \tag{2.9}$$

Daher sagt Definition 2.2.24: $v_1, \ldots, v_n$ sind linear unabhängig genau dann, wenn (2.9) die <u>einzige</u> Beschreibung des Nullvektors als Linearkombination von $v_1, \ldots, v_n$ ist.

Definition 2.2.26. Eine nichtleere Menge $S \subseteq V$ heisst *linear unabhängig*, falls jede endliche nichtleere Teilmenge von S linear unabhängig ist.

Aus dieser Definition folgt direkt:

Lemma 2.2.27. *Sei S linear unabhängig und $\emptyset \neq S' \subseteq S$. Dann ist auch S' linear unabhängig.*

Beispiel 2.2.28. (1) Die Menge $\{e_1, \ldots, e_n\} \subseteq K^n$ ist linear unabhängig. (Wieso?)

(2) Die Menge $\{e_1, e_2, \ldots\} \subseteq K^\infty$ ist linear unabhängig. (Wieso?)

[15] Da $0 \notin \mathbb{N}$ ist, ist diese Menge nicht die leere Menge.

(3) Die Menge $\{1, x, x^2, \ldots, x^n\}$ ist linear unabhängig in $K[x]_n$. Ausserdem ist die Menge $\{1, x, x^2, \ldots\}$ linear unabhängig in $K[x]$. (Wieso?)

Übung 2.2.29. (1) Beweisen Sie: Die Menge $\{v\} \subseteq V$ ist linear unabhängig genau dann, wenn $v \neq 0$.

(2) Wann sind zwei Vektoren in $\mathbb{R}^2$ linear unabhängig?

(3) Versuchen Sie drei linear unabhängige Vektoren in $\mathbb{R}^3$ zu finden!

Definition 2.2.30. Eine Menge heisst *linear abhängig*, falls sie nicht linear unabhängig ist.

Definition 2.2.31. Die *triviale Linearkombination* von $v_1, \ldots, v_n$ ist

$$0 = 0 \cdot v_1 + \ldots + 0 \cdot v_n.$$

Daher gilt: $v_1, \ldots, v_n$ ist linear unabhängig genau dann, wenn die triviale Linearkombination die einzige Möglichkeit ist 0 als Linearkombination von $v_1, \ldots, v_n$ zu schreiben.

Lineare (Un-)Abhängigkeit ist so wichtig, dass es sich lohnt die Definition von linearer Abhängigkeit explizit hinzuschreiben.

Definition 2.2.32. Eine nichtleere Menge $S \subseteq V$ heisst *linear abhängig*, wenn es paarweise verschiedene $v_1, \ldots, v_n \in S$ gibt und $a_1, \ldots, a_n \in K$, die nicht alle Null sind, so dass

$$a_1 v_1 + \ldots + a_n v_n = 0.$$

Bemerkung 2.2.33. Nehmen Sie sich Zeit, um zu überprüfen, dass Definition 2.2.32 die Negation der Definition 2.2.26 ist. Falls Sie dies verwirrend finden, führen Sie sich folgende Idee vor Augen: $v_1, \ldots, v_n$ sind linear abhängig, wenn der Nullvektor nicht nur die triviale Beschreibung $0 \cdot v_1 + \ldots + 0 \cdot v_n$ hat, sondern auch eine andere „nicht-triviale“ Beschreibung.

Jetzt sind wir bereit für den Begriff der Basis, welcher die gewünschte Eindeutigkeit aus der Einführung herbeiführt.

Definition 2.2.34. Eine Menge $S \subseteq V$ heisst eine *Basis* (für V), falls S linear unabhängig ist und V von S erzeugt wird.

Proposition 2.2.35. *Eine Menge $S \subseteq V$ ist eine Basis von V genau dann, wenn jedes $v \in V$ in einer eindeutigen Weise als Linearkombination von Vektoren aus S geschrieben werden kann.*

Beweis. Der Einfachheit halber nehmen wir an, dass $|S| < \infty$ ist. Der Beweis für $|S| = \infty$ ist ziemlich ähnlich.

„$\Leftarrow$“: Nehmen wir an, dass S die Eigenschaft hat, dass jedes $v \in V$ in einer eindeutigen Weise als Linearkombination von Vektoren aus S geschrieben werden kann. Dies impliziert direkt, dass $\mathrm{Sp}(S) = V$. Sei $S = \{v_1, \ldots, v_n\}$. Da 0 als triviale Linearkombination $0 = 0 \cdot v_1 + \ldots + 0 \cdot v_n$ geschrieben werden kann, folgt aus der Eindeutigkeit, dass für $a_1, \ldots, a_n \in K$ mit

$$a_1 v_1 + \ldots + a_n v_n = 0$$

$a_1 = \ldots = a_n = 0$ folgen muss. Dies zeigt, dass S linear unabhängig ist und daher eine Basis ist.

„$\Rightarrow$“: Diese Richtung beweisen wir mittels Kontraposition. Nehmen wir an, dass $v \in V$ auf zwei verschiedene Weisen als Linearkombination von Vektoren aus S geschrieben werden kann: Angenommen

$$v = a_1 v_1 + \ldots a_n v_n \quad \text{und} \quad v = b_1 v_1 + \ldots + b_n v_n$$

mit $a_1, \ldots, a_n, b_1, \ldots, b_n \in K$, so dass ein $1 \leq i_0 \leq n$ existiert mit $a_{i_0} \neq b_{i_0}$. Dann ist

$$\begin{aligned} 0 = v - v &= (a_1 v_1 + \ldots + a_n v_n) - (b_1 v_1 + \ldots + b_n v_n) \\ &= (a_1 - b_1) v_1 + \ldots + (a_n - b_n) v_n, \end{aligned}$$

aber $a_{i_0} - b_{i_0} \neq 0$ und daher ist $S = \{v_1, \ldots, v_n\}$ linear abhängig. Insbesondere ist S keine Basis. □

Bemerkung 2.2.36. Proposition 2.2.35 zeigt: Eine Basis eines Vektorraums gibt uns, wie in der Einführung gewünscht, „eindeutige Koordinaten“ um die Vektoren aus dem Vektorraum zu beschreiben.

Beispiel 2.2.37. Alle Mengen in Beispiel 2.2.20 (Standard-Basen) sind Basen der jeweiligen Vektorräume.

Bevor wir einige Beispiele geben, wollen wir besser verstehen, wie Basen in endlich-dimensionalen Vektorräumen aussehen. Das ruft nach einem neuen Abschnitt!

2.3 Endlich-dimensionale Vektorräume

Dieses Kapitel ist das schönste in der Geschichte der Linearen Algebra I, das wir erzählen wollen. Wir werden sehen, dass wir mittels der Begriffe linearer (Un-)Abhängigkeit und Span ein starkes Struktur-Theorem für endlich-dimensionale Vektorräume über einem Körper K beweisen können. Teile dieser Präsentation sind aus dem Buch von Halmos [9] und [5, Kap. 2]. Im Folgenden ist V ein Vektorraum über K, der endlich-dimensional ist (siehe Definition 2.2.11).

Bemerkung 2.3.1. In diesem Abschnitt ist es besser an „Listen von Vektoren“ zu denken. Das heisst, wir stellen uns die Menge $\{v_1, \dots, v_m\} \subseteq V$ als geordnete Menge von V vor, wobei die Ordnung durch die Nummerierung gegeben ist. Daher schreiben wir häufig $v_1, \dots, v_m$ statt $\{v_1, \dots, v_m\}$ und nennen $v_1, \dots, v_m$ eine *Liste*. Mit der *Länge* einer Liste meinen wir einfach m. (Der Fall $m = 0$ ist erlaubt, wenn die Liste leer ist!)

Lemma 2.3.2. *Seien $v_1, \dots, v_m$ linear abhängig in V. Dann existiert ein $1 \leq j \leq m$ mit*

(a) $v_j \in \mathrm{Sp}(v_1, \dots, v_{j-1})$ *und*

(b) $\mathrm{Sp}(v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_m) = \mathrm{Sp}(v_1, \dots, v_m)$.

Beweis. Da $v_1, \dots, v_m$ linear abhängig sind, existieren $a_1, \dots, a_m \in K$, so dass

$$a_1 v_1 + \ldots + a_m v_m = 0$$

und nicht alle $a_i = 0$ sind für $1 \leq i \leq m$. Sei $j = \max\{i \mid a_i \neq 0\}$. Dann gilt

$$a_1 v_1 + \ldots + a_j v_j = 0.$$

Wir lösen nach $a_j v_j$ auf und dividieren durch $a_j (\neq 0)$:

$$v_j = \frac{-a_1}{a_j} v_1 + \ldots + \frac{-a_{j-1}}{a_j} v_{j-1}. \tag{2.10}$$

Dies zeigt (a). Für (b) betrachten wir $v \in \mathrm{Sp}(v_1, \dots, v_m)$. Dann ist

$$v = a_1 v_1 + \ldots + a_m v_m$$

für $a_1, \dots, a_m \in K$. Man setzt nun (2.10) für v_j ein und sieht, dass v auch eine Linearkombination von $\{v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_m\}$ ist. Dies zeigt die Behauptung. (Wieso?) $\square$

Bemerkung 2.3.3. Was passiert im Beweis, falls $j = 1$? Dann ist der Span von $v_1, \dots, v_{j-1}$ gleich $\mathrm{Sp}(v_1, \dots, v_{j-1}) = \mathrm{Sp}(\emptyset) = \{0\}$. Das impliziert, dass $v_j = v_1 = 0$. Dann gelten in der Tat (a) und (b).

Lemma 2.3.4. *Angenommen wir haben in Lemma 2.3.2 zusätzlich die Bedingung, dass $v_1, \dots, v_k$ für $k < m$ linear unabhängig sind. Dann gilt $k < j$.*

Beweis. Angenommen $j \leq k$, dann gilt $v_j \in \mathrm{Sp}(v_1, \dots, v_{j-1})$ beziehungsweise

$$v_j = a_1 v_1 + \ldots + a_{j-1} v_{j-1}$$

für $a_1, \dots, a_{j-1} \in K$. Dann ist jedoch

$$0 = a_1 v_1 + \ldots + a_{j-1} v_{j-1} + (-1) v_j$$

eine nicht-triviale Linearkombination von $v_1, \dots, v_j$, was im Widerspruch zur linearen Unabhängigkeit von $v_1, \dots, v_k$ steht. □

Lemma 2.3.5. *Seien $w_1, \dots, w_n \in V$ mit $\mathrm{Sp}(w_1, \dots, w_n) = V$ und $v \in V$. Dann sind $v, w_1, \dots, w_n$ linear abhängig.*

Beweis. Der Beweis funktioniert gleich wie im vorherigen Lemma: Da es $a_1, \dots, a_n \in K$ gibt mit $v = a_1 w_1 + \ldots + a_n w_n$, gilt $0 = a_1 w_1 + \ldots + a_n w_n - v$. Dies zeigt die lineare Abhängigkeit von $\{v, w_1, \dots, w_n\}$. □

Lemma 2.3.6. *Seien $v_1, \dots, v_n \in V$ mit $\mathrm{Sp}(v_1, \dots, v_n) = V$ und seien $u_1, \dots, u_m \in V$ linear unabhängig. Dann gilt $m \leq n$.*

Beweis. Wir werden m Etappen durchführen. In der ersten Etappe werden wir u_1 mit einem der Vektoren $\{v_1, \dots, v_n\}$ „austauschen“, und zwar wie folgt:

Etappe 1: Wir betrachten die Menge

$$u_1, v_1, \dots, v_n \tag{2.11}$$

von $n+1$ Vektoren. Laut Lemma 2.3.5 ist diese Menge linear abhängig. Laut Lemma 2.3.2 können wir einen der Vektoren in (2.11) „wegwerfen“ ohne den Span zu ändern. Laut Lemma 2.3.4 ist dieser „unbrauchbare“ Vektor nicht u_1, da $\{u_1\}$ linear unabhängig ist. (Streng genommen, müssen wir hier Lemma 2.2.27 benutzen, um zu argumentieren, dass $\{u_1\}$ linear unabhängig ist.) Wir bekommen eine neue Liste der Länge n, in welcher u_1 der erste Vektor ist und die anderen $n-1$ Vektoren Elemente von $\{v_1, \dots, v_n\}$ sind. Des Weiteren spannt diese Liste V auf.

Etappe $1 < j \leq m$: (Wir empfehlen dem Leser an $j = 2$ zu denken.) Aus der letzten Etappe haben wir eine Liste der Länge n, die V aufspannt, deren erste $j-1$ Vektoren $u_1, \dots, u_{j-1}$ sind und deren letzte $n-(j-1)$ Vektoren in $\{v_1, \dots, v_n\}$ enthalten sind. Der Lesbarkeit halber führen wir eine Umnummerierung durch und schreiben diese Liste als

$$u_1, \dots, u_{j-1}, w_1, \dots, w_{n-(j-1)}.$$

Wir betrachten jetzt die Liste

$$u_1, \dots, u_{j-1}, u_j, w_1, \dots, w_{n-(j-1)}.$$

Auch hier gilt wegen Lemma 2.3.5, dass diese Liste linear abhängig ist. Laut Lemma 2.3.2 können wir einen Vektor dieser Liste wegwerfen ohne den Span zu ändern und wegen Lemma 2.3.4 muss dieser Vektor einer der Vektoren $w_1, \ldots, w_{n-(j-1)}$ sein, da $u_1, \ldots, u_j$ linear unabhängig sind. Wir führen nacheinander die m Etappen durch.

Am Ende von Etappe m bekommen wir eine Liste

$$u_1, \ldots, u_m, w_1', \ldots, w_{n-m}'$$

der Länge n, die alle Vektoren $u_1, \ldots, u_m$ enthält. Daher ist $n \geq m$, was wir zeigen wollten. $\square$

Bemerkung 2.3.7. Wir wir gleich sehen werden, spielt Lemma 2.3.6 eine zentrale Rolle. Es ist eng verbunden mit dem sogenannten „Steinitz'schen Austauschsatz“ (vgl. Fischer [7, 1.5.4]).

Wir sind jetzt bereit für eine der Säulen der linearen Algebra:

Theorem 2.3.8. *Sei V ein endlich-dimensionaler Vektorraum über einem Körper K. Dann hat V eine Basis mit endlicher Länge. Ausserdem hat jede Basis von V die gleiche Länge.*

Beweis der zweiten Aussage in Theorem 2.3.8. Wir beweisen zunächst die zweite Aussage. Seien $\mathcal{A}$ und $\mathcal{B}$ zwei Basen von V. Dann sind $\mathcal{A}$ und $\mathcal{B}$ endliche Mengen, weil V endlich-dimensional ist. Da $\mathcal{A}$ erzeugend und $\mathcal{B}$ linear unabhängig ist, folgt aus Lemma 2.3.6, dass $|\mathcal{B}| \leq |\mathcal{A}|$. Wir können jedoch die Rollen von $\mathcal{A}$ und $\mathcal{B}$ vertauschen! Da $\mathcal{B}$ erzeugend und $\mathcal{A}$ linear unabhängig ist, gilt auch $|\mathcal{A}| \leq |\mathcal{B}|$, also $|\mathcal{A}| = |\mathcal{B}|$. $\square$

Für den Beweis der Existenz einer Basis beweisen wir zunächst ein separates Lemma, da dessen Inhalt und der Beweis davon später wichtig sein werden für uns.

Lemma 2.3.9. *Sei $v_1, \ldots, v_n$ ein Erzeugendensystem. Dann enthält $v_1, \ldots, v_n$ eine Basis.*

Beweis. Kurz gesagt (und ohne einen expliziten Algorithmus anzugeben) könnte man das Lemma beweisen, indem man Vektoren aus $v_1, \ldots, v_n$ nacheinander weg wirft bis man eine linear unabhängige Liste bekommt mit demselben Span. Und daher wäre diese neue Liste dann eine Basis. Wir möchten den Beweis allerdings etwas algorithmischer besprechen. Wir geben einen Algorithmus mit n Etappen an, der in der Etappe j entscheidet, ob v_j ein Teil der Basis sein wird.

Etappe 1: Wenn $v_1 = 0$ ist, werfen wir v_1 weg. Wenn nicht, dann behalten wir v_1.

Etappe $2 \leq j \leq n$: Falls $v_j \in \mathrm{Sp}(v_1, \ldots, v_{j-1})$, so werfen wir v_j weg. Falls nicht, behalten wir v_j.

Nach diesen n Etappen bekommen wir eine Liste. Mit einer Umnummerierung schreiben wir diese Liste als $v_1, \ldots, v_m$. Bemerken Sie, dass $\mathrm{Sp}(v_1, \ldots, v_m) = V$, da wir in jeder Etappe j nur Vektoren, die im Span von Vektoren in $v_1, \ldots, v_{j-1}$ enthalten sind, weggeworfen haben. Ausserdem behaupten wir, dass $v_1, \ldots, v_m$ linear unabhängig sind, da $v_1, \ldots, v_m$ die Bedingungen des folgenden Lemmas 2.3.10 erfüllen. $\square$

Lemma 2.3.10. *Seien $w_1, \ldots, w_l \in V$, so dass $w_j \notin \mathrm{Sp}(w_1, \ldots, w_{j-1})$ für alle $1 \leq j \leq l$ ist. Dann sind $w_1, \ldots, w_l$ linear unabhängig.*

Beweis. Falls die Aussage des Lemmas nicht stimmt, folgt aus Lemma 2.3.2 (a), dass es ein $1 \leq j \leq l$ gibt mit $w_j \in \mathrm{Sp}(w_1, \ldots, w_{j-1})$, im Widerspruch zur Annahme. $\square$

Beweis der ersten Aussage in Theorem 2.3.8. Nach Definition 2.2.11 eines endlich-dimensionalen Vektorraums existiert eine Erzeugendensystem für V. Nach Lemma 2.3.9 enthält dieses Erzeugendensystem eine Basis. Die erste Aussage in Theorem 2.3.8 folgt somit. Dies beendet den Beweis von Theorem 2.3.8. $\square$

Theorem 2.3.8 ermöglicht folgende zentrale Definition:

Definition 2.3.11. Sei V ein endlich-dimensionaler Vektorraum. Wir definieren die *Dimension* von V als $\dim V = n$, wobei $n \in \mathbb{N} \cup \{0\}$ die Länge einer Basis von V ist.

Beachten Sie, dass dank Theorem 2.3.8 die Dimension in der Tat ein Element von $\mathbb{N} \cup \{0\}$ und *wohl-definiert* (das heisst, sie hängt nicht von der Wahl der Basis ab) ist. Im nächsten Kapitel werden wir sogenannte Isomorphismen zwischen Vektorräumen definieren. Hier ist ein Spoiler: Bis auf Isomorphismus gibt es nur einen Vektorraum über K der Dimension n. Mit der Hilfe des nächsten Beispiels, zeigt dies, dass jeder Vektorraum über K mit $\dim V = n$ isomorph zu K^n ist.

Beispiel 2.3.12. Wie wir in Beispiel 2.2.37 gesehen haben, sind die Mengen aus Beispiel 2.2.20 Basen der jeweiligen Vektorräume. Dies zeigt:

(1) $\dim K^n = n$;

(2) $\dim K[x]_n = n + 1$;

(3) $\dim M_{m \times n}(K) = mn$;

(4) für $|X| = n$ ist $\dim P(X) = n$ (vgl. Beispiel 2.1.23).

Zu Theorem 2.3.8 gibt es ein analoges Theorem auch im Fall, wenn V unendlich-dimensional ist.

Theorem 2.3.13 (Hamel Basis)**.** *Jeder Vektorraum über K hat eine Basis. Je zwei Basen haben die gleiche Kardinalität.*

Wir werden dieses Theorem nicht beweisen, nicht weil es schwierig ist, sondern weil:

(1) Wir werden schon genug Spass haben mit endlich-dimensionalen Vektorräumen.

(2) Hamel-Basen in unendlich-dimensionalen Vektorräumen sind normalerweise nicht interessant: unendlich-dimensionale Vektorräume haben normalerweise zusätzliche Strukturen (zum Beispiel topologische und analytische Strukturen) und man ist normalerweise interessiert an Basen, die diese Strukturen „auffassen“. Das beste Beispiel dazu ist die Theorie der Fourierreihen.

(3) Die Existenz einer Hamel-Basis ist äquivalent zum Auswalaxiom (bzw. zum Lemma von Zorn). (Trotzdem lohnt es sich vielleicht die Existenz einer Basis mit dem Lemma von Zorn zu beweisen, was nicht schwierig ist.)

Bevor wir diesen Abschnitt abschliessen, beweisen wir noch ein nützliches Lemma für später. In gewissem Sinne (vgl. Abschnitt 2.3.1) ist dieses Lemma dual zu Lemma 2.3.9.

Lemma 2.3.14. *Jede linear unabhängige Teilmenge von V kann man zu einer Basis von V erweitern.*

Beweis. Sei $u_1, \dots, u_l$ eine linear unabhängige Liste in V und $v_1, \dots, v_m$ eine Basis von V. Betrachten Sie die Liste

$$u_1, \dots, u_l, v_1, \dots, v_m,$$

welche sicher V erzeugt. Man macht die gleichen Etappen wie im Beweis von Lemma 2.3.9. Die Vektoren $u_1, \dots, u_l$ werden dabei nicht weggeworfen (Wieso?). Daher bekommt man am Ende nach $l + m$ Etappen eine Liste, die mit $u_1, \dots, u_l$ anfängt, linear unabhängig ist und V aufspannt. Das heisst, die erhaltene Liste ist eine Basis von V, die $u_1, \dots, u_l$ enthält. Dies zeigt die Aussage des Lemmas. □

2.3.1 Gleichgewicht

Vielleicht hat es Ihre Mutter Ihnen schon gesagt: Gleichgewicht ist das wichtigste im Leben![16]

In der Mathematik hat das Gleichgewicht eine andere Rolle: Wie wir jetzt erklären werden, führt es uns zu nützlichen und interessanten Objekten.

Maximale linear unabhängige Mengen und minimale Erzeugendensysteme

Je grösser eine Teilmenge $S \subseteq V$ ist, desto grösser ist die Wahrscheinlichkeit, dass sie erzeugend ist. Je kleiner eine Teilmenge $S \subseteq V$ ist, desto grösser ist die Wahrscheinlichkeit, dass sie linear unabhängig ist. Anders gesagt, gibt es für eine Menge S einen Wettbewerb zwischen den Eigenschaften „erzeugend“ und „linear unabhängig“: „erzeugend“ zieht nach oben und „linear unabhängig“ zieht nach unten. Der folgende Satz zeigt, dass die „Gleichgewichts-Punkte“ Mengen der Grösse $\dim V$ sind.

[16] Falls nicht, dann haben sie es sicher von Karate Kid gelernt, oder?!

Satz 2.3.15. *Sei V ein Vektorraum mit $\dim V = n$, wobei $n \in \mathbb{N} \cup \{0\}$. Für eine Liste $v_1, \ldots, v_n \in V$ der Länge n sind die folgenden Aussagen äquivalent:*

(1) Die Vektoren $v_1, \ldots, v_n$ sind linear unabhängig.

(2) Die Vektoren $v_1, \ldots, v_n$ sind erzeugend.

(3) Die Vektoren $v_1, \ldots, v_n$ bilden eine Basis.

Beweis. Wir zeigen zunächst, dass (1) die Aussage in (3) impliziert: Laut Lemma 2.3.14 kann die linear unabhängige Liste $v_1, \ldots, v_n$ durch die Annahme in (1) zu einer Basis erweitert werden. Aber laut Theorem 2.3.8 hat jede Basis Länge n. Daher ist $v_1, \ldots, v_n$ schon eine Basis.

Die Implikationen (3) $\Rightarrow$ (1) und (3) $\Rightarrow$ (2) folgen aus der Definition einer Basis, vgl. Definition 2.2.34.

Um zu zeigen, dass die Aussage (2) die Aussage (3) impliziert, seien $v_1, \ldots, v_n$ ein Erzeugendensystem. Laut Lemma 2.3.9 enthält dieses System eine Basis. Da aber alle Basen Länge n haben, ist $v_1, \ldots, v_n$ schon eine Basis. □

Bemerkung 2.3.16. Es lohnt sich Satz 2.3.15 noch anders zu formulieren: (1) impliziert: Jede maximal linear unabhängige Menge ist eine Basis. (2) impliziert: Jede minimal erzeugende Menge ist eine Basis. (Wieso? Überlegen Sie sich, wieso das eine Umformulierung des Satzes 2.3.15 ist.)

Korollar 2.3.17 (Fehlendes Gleichgewicht)**.** *Sei $n = \dim V < \infty$.*

(a) Eine Menge $v_1, \ldots, v_k$ mit $k < n$ ist nicht erzeugend.

(b) Eine Menge $v_1, \ldots, v_k$ mit $n < k$ ist linear abhängig.

Beweis. (a) Falls die Aussage in (a) nicht gilt, dann enthält die Menge laut Lemma 2.3.9 eine Basis mit Länge kleiner als n.

(b) Falls die Aussage in (b) nicht gilt, dann können wir laut Lemma 2.3.14 die Menge zu einer Basis mit Länge grösser als n erweitern.

Dies beweist das Korollar. □

Bemerkung 2.3.18 (Gauss und Basen in K^m). Genau wie wir nach Lemma 2.2.17 den Begriff Span in K^m mit Gauss-Elimination verbinden konnten, können wir auch die Begriffe der linearen (Un-)Abhängigkeit und Basis mit der Gauss'schen Elimination verbinden.

Aus Lemma 2.2.17 folgt, dass $v_1, \ldots, v_n \in K^m$ linear unabhängig sind genau dann, wenn

$$\begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \in K^m$$

nur die triviale Lösung $x_1 = \ldots = x_n = 0$ hat, was man mit Gauss-Elimination bestimmen kann. Für eine Charakterisierung einer Basis durch ein lineares Gleichungssystem, könnten wir Proposition 2.2.35 benutzen. (Wieso und wie genau?) Aber wir sind jetzt viel schlauer: $v_1, \ldots, v_n \in K^m$ ist eine Basis von K^m, falls $n = m$ und

$$\begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \in K^n$$

nur die triviale Lösung hat, was man mit Gauss-Elimination überprüfen kann.

Beispiel 2.3.19. (1) Je drei Vektoren in $\mathbb{R}^2$ (oder K^2) sind linear abhängig.

(2) Je vier Polynome in $\mathbb{F}_7[x]_2$ (oder $K[x]_2$) sind linear abhängig.

(3) Man hat keine Chance den Raum $M_{2\times 3}(\mathbb{C})$ (als Vektorraum über $\mathbb{C}$) mit 5 Matrizen aufzuspannen.

(4) Zwei Vektoren $v = (a, b)$ und $w = (c, d)$ sind eine Basis für K^2, falls

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

nur die Lösung $x = y = 0$ hat.

(5) Die Vektoren

$$\begin{aligned} f_1 &= (1, 0, \ldots, 0) \\ f_2 &= (1, 1, 0, \ldots, 0) \\ &\vdots \\ f_n &= (1, 1, \ldots, 1) \end{aligned}$$

in K^n bilden eine Basis von K^n.

(6) Je $n + 1$ (vom Nullpolynom verschiedene) Polynome von verschiedenem Grad in $K[x]_n$ bilden eine Basis von $K[x]_n$ (Forum-Wieso?).

(7) Sei V ein Vektorraum. Per Konvention ist die leere Menge $\emptyset$ linear unabhängig. Ausserdem gilt $\mathrm{Sp}(\emptyset) = \{0_V\}$. Daher ist die leere Menge eine Basis des Untervektorraums $\{0_V\}$ und es gilt $\dim \{0_V\} = 0$.

(8) In der Übungsstunde: Gibt es eine Basis $v_1, \ldots, v_4$ von $\mathbb{R}[x]_3$, so dass kein v_i ein quadratisches Polynom[17] ist?

[17] Ein quadratisches Polynom ist ein Polynom von Grad 2.

(9) Sei $\mathcal{A}$ eine Menge von Teilmengen von $\{1, \ldots, n\}$ mit $|A| < n$. Dann kann man nicht jede Teilmenge von $\{1, \ldots, n\}$ mittels der symmetrischen Differenz von Mengen aus $\mathcal{A}$ schreiben. Das heisst $\mathrm{Sp}(\mathcal{A}) \neq P(\{1, \ldots, n\})$.

Es kann gut sein, dass Ihnen die Gültigkeit eines der obigen Beispiele unklar ist. Fragen Sie im Forum nach!

2.3.2 Basen von Vektorräumen ohne Gauss'sche Elimination

Wie findet man eine Basis für einen endlich-dimensionalen Vektorraum oder einen Untervektorraum? Auch hier ist V ein Vektorraum über einem Körper K. Laut Theorem 2.3.8 hat jeder Vektorraum eine Basis. Als Korollar erhalten wir:

Korollar 2.3.20. *Jeder Untervektorraum hat eine Basis.*

Beweis. Laut Lemma 2.0.7 ist jeder Untervektorraum auch ein Vektorraum mit der auf den Untervektorraum eingeschränkten Addition und Skalarmultiplikation. Theorem 2.3.8 zeigt die Existenz einer Basis. □

Was unklar bleibt, ist wie lang die Basis eines Untervektorraums sein darf!

Proposition 2.3.21. *Seien V ein endlich-dimensionaler Vektorraum und U ein Untervektorraum. Dann ist U auch endlich-dimensional und* $\dim U \leq \dim V$. *Ausserdem gilt* $\dim U = \dim V$ *genau dann, wenn $U = V$ ist.*

Beweis. Sei $S \subseteq U$ eine Basis von U. Wir behaupten, dass $|S| \leq \dim V$. Falls dies nicht gilt, dann enthält S eine Menge S' mit $\dim V + 1$ Vektoren, die linear unabhängig sind, da S linear unabhängig ist. Laut Lemma 2.3.14 kann man S' zu einer Basis von V erweitern. Dann bekommt man eine Basis von V, deren Länge grösser ist als $\dim V$, was ein Widerspruch zu Theorem 2.3.8 ist.

Wir zeigen noch die zweite Aussage: Für „$\Leftarrow$" gibt es nichts zu beweisen. Für „$\Rightarrow$" sei $n = \dim U = \dim V$ und $\mathcal{B} = \{v_1, \ldots, v_n\}$ eine Basis für U. Da $\mathcal{B}$ eine Liste von linear unabhängigen Vektoren in V der Länge $\dim V$ ist, ist $\mathcal{B}$ auch eine Basis für V (laut Satz 2.3.15) und daher ist $V = \mathrm{Sp}(\mathcal{B}) = U$. Die Proposition folgt. □

Wir kommen also zu der Frage, wie man eine Basis eines Untervektorraums findet. Mittels Korollar 2.3.20 können wir uns auf Vektorräume konzentrieren.

Erste Methode: Die Beweise benutzen

Die obigen Beweise sind eigentlich algorithmisch, obwohl wir nicht versucht haben die Algorithmen effizient zu wählen. Nichtsdestotrotz führen sie zu einer Antwort, wie wir jetzt kurz erklären:

<u>Top-Down-Methode:</u> Sei $V = \mathrm{Sp}(v_1, \ldots, v_n)$. Wie könnten wir eine Basis für V finden?

- Falls $v_1, \ldots, v_n$ linear unabhängig sind, ist $v_1, \ldots, v_n$ schon eine Basis.
- Falls $v_1, \ldots, v_n$ linear abhängig sind, können wir den Algorithmus im Beweis von Lemma 2.3.9 benutzen, um Vektoren aus $v_1, \ldots, v_n$ „wegzuwerfen", ohne den Span zu ändern, bis wir eine linear unabhängige Liste bekommen.

Bottom-Up-Methode: Sei V ein endlich-dimensionaler Vektorraum. Man wählt $v_1 \in V$ mit $v_1 \neq 0$. (Wenn dies nicht möglich ist, dann ist $V = \{0\}$.) Jetzt verwendet man Lemma 2.3.10: Man wählt $v_2 \notin \mathrm{Sp}(v_1)$ und dann $v_3 \notin \mathrm{Sp}(v_1, v_2)$ bis man dies nicht weiterführen kann, da man den ganzen Vektorraum aufgespannt hat. Die erhaltene Liste ist laut Lemma 2.3.10 eine linear unabhängige Liste, die V aufspannt (da man keine Vektoren mehr hinzufügen konnte). Das heisst, die erhaltene Liste ist eine Basis. Grundsätzlich ist dies genau das, was im Beweis von Lemma 2.3.9 passiert.

Die erste Methode oben ist hauptsächlich von einem theoretischen Standpunkt aus interessant. Die zweite Methode benutzt die Gauss'sche Elimination und ist viel effizienter. Es lohnt sich aber, dies von einem allgemeineren Blickwinkel aus zu betrachten.

2.3.3 Zeilen- und Spaltenraum und die Gauss'sche Elimination

Definition 2.3.22. Sei $A \in M_{m\times n}(K)$ und seien $u_1, \ldots, u_m \in K^n$ die Zeilen der Matrix und $v_1, \ldots, v_n \in K^m$ die Spalten der Matrix. Wir definieren

$$\begin{aligned}\mathrm{ZR}(A) &= \mathrm{Zeilenraum}(A) = \mathrm{Sp}(u_1, \ldots, u_m) \subseteq K^n\\ \mathrm{SR}(A) &= \mathrm{Spaltenraum}(A) = \mathrm{Sp}(v_1, \ldots, v_n) \subseteq K^m.\end{aligned}$$

Lemma 2.3.23. *Seien $A, B \in M_{m\times n}(K)$, so dass B durch Zeilenoperationen aus A entstanden ist. Dann gilt* $\mathrm{ZR}(A) = \mathrm{ZR}(B)$.

Bemerkung 2.3.24. Vorsicht! Normalerweise gilt nicht, dass $\mathrm{SR}(A) = \mathrm{SR}(B)$, wenn man Zeilenoperationen verwendet. Die Überraschung ist jedoch, dass dann trotzdem $\dim \mathrm{SR}(A) = \dim \mathrm{SR}(B)$ gilt!

Beweis. Dieser Beweis ist ähnlich zum Beweis, welchen wir im Zusatzkript [2] zu $\text{Lös}(A, b) = \text{Lös}(B, b)$ gegeben haben. Seien $u_1, \ldots, u_m \in K^n$ die Zeilen von A.

$L_i \leftrightarrow L_j$: Zeilen vertauschen hat keinen Einfluss auf den Span:

$$\mathrm{Sp}(u_1, \ldots, u_i, \ldots, u_j, \ldots, u_m) = \mathrm{Sp}(u_1, \ldots, u_{i-1}, u_j, u_{i+1}, \ldots, u_{j-1}, u_i, u_{j+1}, \ldots, u_m)$$

$\lambda L_i \to L_i$: Hier muss man zeigen, dass für alle $0 \neq \lambda \in K$ gilt:

$$\mathrm{Sp}(u_1, \ldots, u_i, \ldots, u_m) = \mathrm{Sp}(u_1, \ldots, \lambda u_i, \ldots, u_m).$$

(Wieso? Zeigen Sie es!)

$L_i + \lambda L_j \to L_i \ (i \neq j)$: Hier muss man zeigen:

$$\mathrm{Sp}(u_1, \ldots, u_m) = \mathrm{Sp}(u_1, \ldots, u_{i-1}, u_i + \lambda u_j, u_{i+1}, \ldots, u_m).$$

Die Inklusion „$\supseteq$“ folgt aus der Tatsache, dass $u_i + \lambda u_j$ eine Linearkombination von $u_1, \ldots, u_m$ ist. Die Inklusion „$\subseteq$“ folgt aus der Tatsache, dass

$$u_i = (u_i + \lambda u_j) - \lambda u_j$$

eine Linearkombination von $u_1, \ldots, u_{i-1}, u_i + \lambda u_j, u_{i+1}, \ldots, u_m$ ist. $\square$

Definition 2.3.25. Seien $n, m \in \mathbb{N}$ und $A \in M_{m \times n}(K)$. Wir definieren den *Spaltenrang* und den *Zeilenrang* von A durch

$$\begin{aligned} \mathrm{Spaltenrang}(A) &:= \dim \mathrm{SR}(A), \\ \mathrm{Zeilenrang}(A) &:= \dim \mathrm{ZR}(A). \end{aligned}$$

Bemerkung 2.3.26. Wir werden später sehen, dass für jede Matrix A die Gleichung

$$\mathrm{Spaltenrang}(A) = \mathrm{Zeilenrang}(A) \tag{2.12}$$

gilt. Dies ist sicher nicht offensichtlich: Immerhin sind $\mathrm{SR}(A)$ und $\mathrm{ZR}(A)$ im Allgemeinen Untervektorräume von verschiedenen Räumen! Das folgende Lemma zeigt (2.12) für den Spezialfall, dass A in Zeilenstufenform ist.

Lemma 2.3.27. *Sei $B \in M_{m \times n}(K)$ eine Matrix in Zeilenstufenform. Dann sind die Zeilen, die keine Nullzeilen sind, eine Basis des Zeilenraums* $\mathrm{ZR}(B)$. *Insbesondere gilt* $\dim \mathrm{ZR}(B) = \mathrm{Rang}(B)$. *Ähnlich gilt: Die Spalten, die die Pivots enthalten, sind eine Basis des Spaltenraums. Insbesondere gilt Spaltenrang(B) = Zeilenrang(B).*

Beweis. Ohne Beschränkung der Allgemeinheit nehmen wir an, dass alle Pivots 1 sind. Seien $a_{1j_1}, \ldots, a_{rj_r}$ die Einträge der Pivots (das heisst $j_1 < j_2 < \ldots < j_r$) und $u_1, \ldots, u_r$ die entsprechenden Zeilen.

$u_1, \ldots, u_r$ sind linear unabhängig: Sei

$$\sum_{i=1}^{r} \lambda_i u_i = 0. \tag{2.13}$$

Da in der j_1-ten Spalte ausser dem Pivot $a_{1j_1} = 1$ nur Nullen stehen, ist die j_1-te Koordinate des Vektors $\sum_{i=1}^{r} \lambda_i u_i \in K^n$ gleich λ_1. Also folgt aus (2.13), dass $\lambda_1 = 0$ sein muss. Die j_2-te Koordinate von $\sum_{i=1}^{r} \lambda_i u_i \in K^n$ ist dann (weil $\lambda_1 = 0$) gleich λ_2. Aus (2.13) folgt wiederum, dass $\lambda_2 = 0$. Mit Induktion zeigt man: Falls $\lambda_1 = \ldots = \lambda_k = 0$ für $k < r$, dann ist die j_{k+1}-te Koordinate von $\sum_{i=1}^{r} \lambda_i u_i \in K^n$ gleich λ_{k+1}. Aus (2.13) folgt, dass $\lambda_{k+1} = 0$. Daher gilt $\lambda_1 = \ldots = \lambda_r = 0$, was die lineare Unabhängigkeit von $u_1, \ldots, u_r$ zeigt. Da alle anderen Zeilen Nullzeilen sind, folgt $\mathrm{ZR}(B) = \mathrm{Sp}(u_1, \ldots, u_r)$ und daher ist $u_1, \ldots, u_r$ eine Basis von $\mathrm{ZR}(B)$.

Den Beweis der zweiten Aussage (bezüglich der Basis des Spaltenraums) ist sehr ähnlich und wir überlassen ihn den Lesern. Die Gleichheit $\mathrm{Spaltenrang}(B) = \mathrm{Zeilenrang}(B)$ folgt. □

Dies beweist auch eine wichtige Tatsache zur Gauss'schen Elimination, die wir vorher nicht bewiesen haben.

Erinnerung: Für eine Matrix A ist $\mathrm{Rang}(A)$ die Anzahl der Pivots einer Matrix in Zeilenstufenform, die durch Zeilenoperationen auf A entstanden ist.

Korollar 2.3.28. *Der Rang einer Matrix A ist wohl-definiert. Das heisst, der Rang hängt nicht von den Zeilenoperationen ab, die zur Stufenform führen.*

Beweis. Wieso? Dieses „Wieso?“ hat für einmal eine Lösung am Ende des Kapitels. □

Korollar 2.3.29 (Algorithmus zur Berechnung einer Basis aus einem Erzeugendensystem in K^n)**.** *Sei $U = \mathrm{Sp}(u_1, \ldots, u_m) \subseteq K^n$. Um eine Basis für U zu finden, schreibt man $u_1, \ldots, u_m$ als Zeilen einer Matrix $A \in M_{m\times n}(K)$. Mit Zeilenoperationen führt man A zu einer Matrix B Zeilenstufenform über. Dann sind die Zeilen von B, die keine Nullzeilen sind, eine Basis von U.*

Wie findet man eine Basis für den Spaltenraum? Man könnte statt Zeilenoperationen ganz analog Spaltenoperationen durchführen, Stufenform bezüglich Spalten und so weiter definieren. Man könnte aber auch die Rollen von Zeilen und Spalten vertauschen. Dieser Vorgang kommt mehrmals vor, daher gibt man ihm einen Namen:

Definition 2.3.30 (Transposition)**.** Sei $A = (a_{ij})_{ij} \in M_{m\times n}(K)$. Wir definieren die *Transponierte* von A als die Matrix

$$A^T = (b_{ij})_{\substack{1\le i\le n\\ 1\le j\le m}} \in M_{n\times m}(K)$$

mit $b_{ij} = a_{ji}$.

Sei $1 \le k \le n$. Dann gelten folgende Eigenschaften:

- Die k-te Zeile von A^T ist die k-te Spalte von A.
- Die k-te Spalte von A^T ist die k-te Zeile von A.
- $(A + B)^T = A^T + B^T$.
- $(\lambda A)^T = \lambda A^T$.

- $(A^T)^T = A$.
- $\mathrm{SR}(A) = \mathrm{ZR}(A^T)$.
- $\mathrm{ZR}(A) = \mathrm{SR}(A^T)$.
- Spaltenoperationen auf A sind dasselbe wie Zeilenoperationen auf A^T.

Bemerkung 2.3.31. Die Addition und Skalarmultiplikation von Matrizen haben Sie in der Übungsstunde definiert.

Satz 2.3.32. *Seien $m, n \in \mathbb{N}$ und $A \in M_{m\times n}(K)$. Es gilt*

$$\textit{Spaltenrang}(A) = \textit{Zeilenrang}(A).$$

Wir werden den kompletten Beweis erst in den nächsten Kapiteln durchführen (und bis dann nicht verwenden!), geben dafür aber schon eine Beweisskizze mit Gauss'scher Elimination, die man auch zu einem echten Beweis „upgraden“ kann.

Magie: Wie man mit Gauss'scher Elimination gleichzeitig Basen für vier verschiedene Vektorräume berechnen kann!

Beispiel 2.3.33. Sei

$$A = \begin{pmatrix} -3 & 6 & -1 & 1 \\ 1 & -2 & 2 & 3 \\ 2 & -4 & 5 & 8 \end{pmatrix}.$$

Wir bringen die Matrix in (Zeilen-)Stufenform. In jedem Schritt bezeichnet w_i die i-te Zeile der vorhergehenden Matrix.

$$A \xrightarrow{w_1 \leftrightarrow w_2} \begin{pmatrix} 1 & -2 & 2 & 3 \\ -3 & 6 & -1 & 1 \\ 2 & -4 & 5 & 8 \end{pmatrix} \xrightarrow{w_2+3w_1 \to w_2} \begin{pmatrix} 1 & -2 & 2 & 3 \\ 0 & 0 & 5 & 10 \\ 2 & -4 & 5 & 8 \end{pmatrix}$$

$$\xrightarrow{w_3-2w_1 \to w_3} \begin{pmatrix} 1 & -2 & 2 & 3 \\ 0 & 0 & 5 & 10 \\ 0 & 0 & 1 & 2 \end{pmatrix} \xrightarrow{w_2 \leftrightarrow w_3} \begin{pmatrix} 1 & -2 & 2 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 5 & 10 \end{pmatrix}$$

$$\xrightarrow{w_3-5w_2 \to w_3} \begin{pmatrix} 1 & -2 & 2 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix} \xrightarrow{w_1-w_2 \to w_1} \begin{pmatrix} 1 & -2 & 1 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Zusammenfassend haben wir aus Zeilenoperationen die Stufenform

$$A \xrightarrow{\text{Zeilenoperationen}} \begin{pmatrix} 1 & -2 & 1 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{2.14}$$

erhalten.

Korollar 2.3.29 besagt, dass $\{(1,-2,1,1),(0,0,1,2)\}$ eine Basis des Zeilenraums $\mathrm{ZR}(A)$ ist. Wenn man jetzt eine Basis des Spaltenraums $\mathrm{SR}(A)$ finden will, gibt es zwei Methoden:

Methode für die Fleissigen: Ohne viel zu denken, betrachtet man

$$A^T = \begin{pmatrix} -3 & 1 & 2 \\ 6 & -2 & -4 \\ -1 & 2 & 5 \\ 1 & 3 & 8 \end{pmatrix}$$

und führt die Gauss'sche Elimination mit Zeilenoperationen durch:

$$A^T \xrightarrow{\text{Zeilenoperationen}} \begin{pmatrix} 1 & 0 & \frac{1}{5} \\ 0 & 1 & \frac{13}{5} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Daher gilt, dass

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ \frac{1}{5} \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \frac{13}{5} \end{pmatrix} \right\}$$

eine Basis des Spaltenraums $\mathrm{SR}(A)$ ist. (Oder falls Sie etwas gegen Brüche haben, ist

$$\left\{ \begin{pmatrix} 5 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 5 \\ 13 \end{pmatrix} \right\}$$

auch eine Basis. (Wieso?))

Methode für Faule: Hier ist der Trick, mit welchem man eine Basis für $\mathrm{SR}(A)$ und $\mathrm{ZR}(A)$ gleichzeitig finden kann.

Betrachten Sie das Eliminationsverfahren in (2.14). Für eine Matrix $M \in M_{3\times 4}(K)$ schreiben wir $M^{(1)}, M^{(2)}, M^{(3)}, M^{(4)}$ für ihre Spalten. Haben Sie bemerkt, dass

$$A^{(2)} = -2A^{(1)}? \tag{2.15}$$

Bemerken Sie auch, dass die lineare Relation in (2.15) auch für alle anderen Matrizen im Eliminationsprozess gilt. Allgemein gilt: Zeilenoperationen erhalten alle linearen Relationen zwischen den Spalten. Insbesondere Aussagen wie

$$\text{die zweite Spalte } = -2 \text{ (die erste Spalte)}$$

sind invariant unter Zeilenoperationen. Das heisst, Zeilenoperationen ändern die Gültigkeit solcher Aussagen nicht.

Am Ende des Eliminationsprozesses in (2.14) steht die Matrix

$$B := \begin{pmatrix} 1 & -2 & 1 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Laut Lemma 2.3.27 ist {die erste Spalte von B, die dritte Spalte von B} eine Basis von $\mathrm{SR}(B)$. Diese Aussage ist auch invariant unter Zeilenoperationen. Das heisst, die Aussage ändert sich nicht, wenn wir Zeilenoperationen verwenden. Daher gilt

$$\{\text{die erste Spalte von } A, \text{ die dritte Spalte von } A\}$$

ist eine Basis von $\mathrm{SR}(A)$ oder

$$\left\{ \begin{pmatrix} -3 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ 5 \end{pmatrix} \right\}$$

ist eine Basis des Spaltenraums $\mathrm{SR}(A)$ von A. Achtung:

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right\}$$

ist aber keine Basis von $\mathrm{SR}(A)$!

Korollar 2.3.34. *Wenn man die faule Version in Beispiel 2.3.33 studiert, kommt man zu den folgenden Erkenntnissen:*

(1) Sei A eine Matrix und $j_1, \ldots, j_r$ die Indizes der Spalten mit Pivots einer Matrix, die aus A durch Zeilenoperationen in Stufenform gebracht wurde. Seien $A^{(i)}$ die Spalten von A. Dann ist $\{A^{(j_1)}, \ldots, A^{(j_r)}\}$ eine Basis des Spaltenraums von A.

(2) Dies zeigt auch:

$$\dim \mathit{Spaltenraum}(A) = \dim \mathit{Zeilenraum}(A). \tag{2.16}$$

Bemerkung 2.3.35. Wir werden zwei (vollständige) Beweise von (2.16) geben und bis dann werden wir es nicht verwenden. Es ist aber schon toll sich zu merken, dass man manche solcher Tatsachen beweisen kann. Sie dürfen in Ihren Berechnungen diesen

Trick benutzen. Wenn wir den Orthogonalitäts-Begriff einführen, werden wir noch einen zusätzlichen Trick lernen.

2.3.4 Summen

Wie immer ist V ein Vektorraum über einem Körper K.

Definition 2.3.36. Seien $U, W \subseteq V$ zwei Untervektorräume. Wir definieren die *Summe* von U und W als

$$U + W = \{u + w \mid u \in U, w \in W\}.$$

Allgemeiner definieren wir für Untervektorräume $U_1, \ldots, U_n$ von V die *Summe* von $U_1, \ldots, U_n$ als

$$U_1 + \ldots + U_n = \{u_1 + \ldots + u_n \mid u_i \in U_i \text{ für } i = 1, \ldots, n\}.$$

Proposition 2.3.37. *Seien $U, W \subseteq V$ zwei Untervektorräume. Es gilt*

(1) $U + W = \mathrm{Sp}(U \cup W)$ und damit ist $U + W$ der kleinste Untervektorraum, der U und W enthält.

(2) Falls $U + W$ endlich-dimensional ist, dann kann man folgendermassen eine Basis von $U + W$ bilden: Seien $k = \dim(U \cap W)$, $l = \dim(U)$ und $m = \dim W$.

(a) Man wählt eine Basis $p_1, \ldots, p_k$ von $U \cap W$.

(b) Man wählt eine Basis $p_1, \ldots, p_k, u_1, \ldots, u_{l-k}$ für U und eine Basis $p_1, \ldots, p_k, w_1, \ldots, w_{m-k}$ für W.

Dann ist

$$p_1, \ldots, p_k, u_1, \ldots, u_{l-k}, w_1, \ldots, w_{m-k} \tag{2.17}$$

eine Basis für $U + W$.

(3) Insbesondere gilt die sogenannte Dimensionsformel*:*

$$\dim U + W = \dim U + \dim W - \dim U \cap W.$$

Beweis. Teil (1) überlassen wir den Lesern. (Hinweis: Für die zweite Aussage benutzen Sie Lemma 2.2.3.)

Für (2) müssen wir beweisen, dass die Liste in (2.17) eine Basis von $U + W$ ist. Da die Liste eine Basis für U und W enthält, spannt diese Liste nach (1) $U + W$ auf. Es bleibt also zu zeigen, dass die Liste in (2.17) linear unabhängig ist. Nehmen wir also an, dass

$$a_1 p_1 + \ldots + a_k p_k + b_1 u_1 + \ldots + b_{l-k} u_{l-k} + c_1 w_1 + \ldots + c_{m-k} w_{m-k} = 0 \tag{2.18}$$

beziehungsweise, dass

$$v := c_1 w_1 + \ldots + c_{m-k} w_{m-k} = -(a_1 p_1 + \ldots + a_k p_k + b_1 u_1 + \ldots + b_{l-k} u_{l-k}).$$

für $a_1, \ldots, a_k, b_1, \ldots, b_{l-k}, c_1, \ldots, c_{m-k} \in K$. Einerseits gilt $v \in W$, da die linke Seite in W ist. Andererseits gilt $v \in U$, da die rechte Seite in U ist. Daher ist $v \in U \cap W$. Also existieren $\alpha_1, \ldots, \alpha_k \in K$ mit $v = \alpha_1 p_1 + \ldots + \alpha_k p_k$ und somit

$$\alpha_1 p_1 + \ldots + \alpha_k p_k = c_1 w_1 + \ldots + c_{m-k} w_{m-k}.$$

Da $\{p_1, \ldots, p_k, w_1, \ldots, w_{m-k}\}$ linear unabhängig sind, folgt $c_1 = \ldots = c_{m-k} = \alpha_1 = \ldots = \alpha_k = 0$. Dann vereinfacht sich Gleichung (2.18) zu

$$a_1 p_1 + \ldots + a_k p_k + b_1 u_1 + \ldots + b_{l-k} u_{l-k} = 0,$$

da $c_1 = \ldots = c_{m-k} = 0$. Da $\{p_1, \ldots, p_k, u_1, \ldots, u_{l-k}\}$ auch linear unabhängig ist, gilt

$$a_1 = \ldots = a_k = b_1 = \ldots = b_{l-k} = 0.$$

Dies zeigt, dass die Liste in (2.17) linear unabhängig und daher eine Basis von $U + W$ ist. Es folgt dann auch (3). $\square$

Bemerkung 2.3.38. Hier lohnt es sich auch den Beweis im Fischer [7, Kap. 1.6.1] zu lesen. Man könnte ausgehend von (2.18) die lineare Unabhängigkeit auch ein bisschen anders zeigen (mittels Proposition 2.2.35).

Übung 2.3.39. Seien U_1, U_2, U_3 Untervektorräume eines endlich-dimensionalen Vektorraums. Betrachten Sie die Formel

$$\begin{aligned}\dim(U_1 + U_2 + U_3) = {} & \dim U_1 + \dim U_2 + \dim U_3 \\ & - \dim(U_1 \cap U_2) - \dim(U_1 \cap U_3) - \dim(U_2 \cap U_3) \\ & + \dim(U_1 \cap U_2 \cap U_3).\end{aligned}$$

Beweisen Sie die Formel oder geben Sie ein Gegenbeispiel!

Direkte Summen

Proposition 2.3.40. *Sei V ein endlich-dimensionaler Vektorraum über K und seien $U, W \subseteq V$ zwei Untervektorräume. Dann sind folgende Aussagen äquivalent:*

(1) Es gilt $\dim(U + W) = \dim U + \dim W$.

(2) Es gilt $\dim(U \cap W) = 0$.

(3) Es gilt $U \cap W = \{0\}$.

(4) Jedes $v \in U + W$ *ist eindeutig darstellbar als*

$$v = u + w$$

mit $u \in U$ *und* $w \in W$.

(5) Falls $0 = u + w$ *mit* $u \in U$ *und* $w \in W$*, dann ist* $u = 0$ *und* $w = 0$.

Bemerkung 2.3.41. Im unendlichdimensionalen Fall sind die Bedingungen (2), (3), (4) und (5) immer noch äquivalent. Definition 2.3.42 unten bezieht sich dann auf diese vier Aussagen.

Definition 2.3.42. Falls eine der äquivalenten Bedingungen in Proposition 2.3.40 gilt, dann sagen wir, dass $U + W$ die *direkte Summe* von U und W ist und schreiben

$$U + W = U \oplus W.$$

Beweis von Proposition 2.3.40. Die Äquivalenz von (1) und (2) folgt aus Proposition 2.3.37 (3). Des Weiteren ist (2) äquivalent zu (3). (Wieso? Es gibt nur einen Untervektorraum mit Dimension 0!) Wir zeigen jetzt, dass (3) die Aussage in (4) impliziert. Seien $u_1 + w_1 = u_2 + w_2$ zwei Darstellungen eines Vektors in $U + W$ mit $u_1, u_2 \in U$ und $w_1, w_2 \in W$. Dann gilt, dass

$$u_1 - u_2 = w_2 - w_1 \in U \cap W.$$

Laut (3) ist $U \cap W = \{0\}$ und daher ist $u_1 - u_2 = w_2 - w_1 = 0$ beziehungsweise $u_1 = u_2$ und $w_2 = w_1$, was (4) zeigt. Des Weiteren gilt „(4) $\Rightarrow$ (5)“. (Wieso? Da $0 = 0 + 0$.) Nun zeigen wir die Implikation „(5) $\Rightarrow$ (3)“: Falls es $0 \neq v \in U \cap W$ gibt, dann ist

$$0 = \underbrace{v}_{\in U} + \underbrace{(-v)}_{\in W}$$

keine triviale Darstellung von 0 als $u + w$ mit $u \in U$ und $w \in W$, im Widerspruch zu (5). $\square$

Proposition 2.3.43. *Seien* V *ein endlich-dimensionaler Vektorraum und* U *ein Untervektorraum. Dann existiert ein Untervektorraum* W *mit*

$$V = U \oplus W.$$

Definition 2.3.44. Ein Untervektorraum W wie in Proposition 2.3.43 wird ein *Komplement* von U in V genannt (oder auch ein *direkter Summand* von V zu U).

Beweis von Proposition 2.3.43. Sei $u_1, \ldots, u_l$ eine Basis für U und sei $u_1, \ldots, u_l, w_1, \ldots, w_k$ eine Erweiterung der Basis von U zu einer Basis von V. Sei $W := \mathrm{Sp}(w_1, \ldots, w_k)$. Wir behaupten, dass $V = U \oplus W$. Beachten sie dazu zuerst, dass

$$V = U + W = \mathrm{Sp}(U \cup W) = \mathrm{Sp}(u_1, \ldots, u_l, w_1, \ldots, w_k).$$

(Wieso?) Ausserdem gilt $U \cap W = \{0\}$: Falls $v \in U \cap W$, dann existieren $a_1, \ldots, a_l, b_1, \ldots, b_k \in K$ mit

$$v = a_1 u_1 + \ldots + a_l u_l = b_1 w_1 + \ldots b_k w_k.$$

Da $u_1, \ldots, u_l, w_1, \ldots, w_k$ linear unabhängig sind, folgt jedoch

$$a_1 = \ldots = a_l = b_1 = \ldots = b_k = 0$$

und daher ist $v = 0$. Also folgt mit Proposition 2.3.40 die Aussage. □

Beispiel 2.3.45. Wir geben einige Beispiele:

(1) Hier ist ein allgemeines Beispiel:

Sei V ein endlich-dimensionaler Vektorraum und $\mathcal{B} = \{v_1, \ldots, v_n\}$ eine Basis von V. Sei $S \subseteq \mathcal{B}$. Dann ist

$$V = \mathrm{Sp}(S) \oplus \mathrm{Sp}(\mathcal{B} \setminus S).$$

Zum Beispiel lässt sich K^n dadurch so schreiben:

$$K^n = \mathrm{Sp}(e_1, \ldots, e_k) \oplus \mathrm{Sp}(e_{k+1}, \ldots, e_n)$$

für alle $0 \leq k \leq n$.

(2) Sei X eine endliche Menge und $P(X)$ die dazugehörige Potenzmenge. Wir betrachten $P(X)$ als Vektorraum über $\mathbb{F}_2$. Für jedes $Y \subseteq X$ gilt:

$$P(X) = P(Y) \oplus P(X \setminus Y).$$

(3) Ein Beispiel in K^3: Für $U = \mathrm{Sp}(e_1, e_2)$ und $V = \mathrm{Sp}(e_2, e_3)$ gilt, dass $U + V = K^3$, aber dies ist keine direkte Summe, weil unter anderem auch

$$U \cap V = \mathrm{Sp}(e_2)$$

gilt.

(4) In der Serie haben Sie das Folgende gezeigt:

Sei $V = \mathbb{R}^{\mathbb{R}} = \mathrm{Abb}(\mathbb{R}, \mathbb{R})$, U die Menge aller geraden Funktionen und W die Menge aller ungeraden Funktionen. Dann gilt

$$V = U \oplus W.$$

Sie werden noch mehr Beispiele in der Übungsstunde und in der Serie sehen.

Zurück zu Korollar 2.3.28

Antwort auf die Frage „Wieso?“ im Beweis von Korollar 2.3.28. Sei B irgendeine Matrix in Stufenform, die durch Zeilenoperationen aus A entstanden ist. Der Beweis von Lemma 2.3.23 zeigt, dass

$$\mathrm{ZR}(A) = \mathrm{ZR}(B).$$

Lemma 2.3.27 besagt, dass

$$\dim \mathrm{ZR}(B) = \text{Anzahl der Pivots in } B.$$

Daher ist diese Anzahl Pivots gleich $\dim \mathrm{ZR}(A)$ und hängt nicht von B ab. $\square$

Changelog: Kapitel 2

- 09.02: In der Bottom-Up-Methode im Abschnitt 2.3.2 wurde v durch v_1 ersetzt.
- 19.02: Nummerierungen wurden geändert.
- 08.05: Bemerkung 2.3.41 wurde hinzugefügt.

Epilog von Kapitel 2

Basis. Das sollten Sie sich aus Kapitel 2 merken.[18] In jedem Vektorraum gibt es eine Basis und alle Basen haben die gleiche Länge (oder Kardinalität)! Vektorräume sind im Allgemeinen abstrakt und eine Basis ermöglicht uns einen Zugang zu solch einem Vektorraum, auch wenn dieser Vektorraum sehr abstrakt ist. Wie genau? Erinnern Sie sich an Proposition 2.2.35. Im nächsten Kapitel werden wir Folgendes definieren:

Definition. Sei V ein Vektorraum, $\mathcal{B} = \{v_1, \ldots, v_n\}$ eine Basis von V und $v \in V$ ein beliebiger Vektor. Da $\mathcal{B}$ eine Basis ist, können wir die Koordinaten von v bezüglich $\mathcal{B}$ folgendermassen definieren und mit $[v]_\mathcal{B}$ notieren:

$$[v]_\mathcal{B} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \in K^n :\Longleftrightarrow \; v = a_1 v_1 + \ldots a_n v_n.$$

Sie sollten darüber so nachdenken: $\mathcal{B}$ macht den abstrakten Raum V für uns „explizit“, indem wir v einfach durch den Vektor $[v]_\mathcal{B} \in K^n$ verstehen können. Etwas anschaulicher stelle ich mir $[v]_\mathcal{B}$ als die Beschreibung von $v \in V$ durch „die Brille $\mathcal{B}$“ vor. Das heisst, dass jede Basis uns eine „Brille“ gibt, mit der wir V „sehen können“. Bevor ich mit dieser Plauderei aufhöre, bemerke ich noch Folgendes: Der Vektor $[v]_\mathcal{B}$ gehört zu K^n. Dies erklärt also auch die folgende Aussage: Ein Vektorraum mit Dimension n hat n „Freiheitsgrade“.

[18] Dies ist keine Zusammenfassung! Wir geben nur Kommentare, woher wir kommen und wohin wir gehen mit dem Stoff.

Kapitel 3

Lineare Abbildungen

Vektorräume zu studieren hat im letzten Kapitel viel Spass gemacht. Aber Vektorräume zu betrachten ohne sie miteinander zu vergleichen wäre schade. Dies wäre wie Mengen zu studieren ohne den Funktionsbegriff. Wir haben schon in der Vorlesung erwähnt, dass man in der Mathematik, und vor allem in der Algebra, gewisse Objekte als *Kategorie* betrachten kann. Grob gesagt, ist eine Kategorie eine Klasse von Objekten und Morphismen (auch Pfeile genannt) zwischen diesen Objekten.[1] Auf jeden Fall sind die Objekte in der linearen Algebra Vektorräume und in diesem Kapitel möchten wir die dazugehörigen Pfeile beschreiben: Lineare Abbildungen.

3.1 Definition einer linearen Abbildung

Wir möchten die Definition einer linearen Abbildung von einem abstrakten Blickwinkel aus motivieren. Seien V und W zwei Vektorräume über einem Körper K. Eine Funktion

$$T : V \to W$$

soll eine „lineare Abbildung“ genannt werden, falls sie die Vektorraum-Struktur von V und W „respektiert“ beziehungsweise miteinander verbindet. Hier ist die offizielle Definition:

Definition 3.1.1. Seien V und W Vektorräume über einem Körper K. Eine Funktion

$$T : V \to W$$

heisst *linear* (oder *lineare Abbildung* oder *lineare Transformation*), falls folgende zwei Eigenschaften gelten:

[1] Wer sich dafür interessiert kann hier mehr nachlesen.

$$\begin{aligned} &\forall\, v_1, v_2 \in V : T(v_1 + v_2) = T(v_1) + T(v_2) && \text{(Additivität)} \\ &\forall\, a \in K, v \in V : T(av) = aT(v) && \text{(Homogenität)} \end{aligned} \tag{3.1}$$

Die Menge aller linearen Abbildungen von V nach W bezeichnen wir mit

$$\mathrm{Hom}_K(V, W) \quad \text{oder} \quad \mathrm{Hom}(V, W),$$

falls bei Letzterem K aus dem Zusammenhang klar ist. In diesem Kapitel, falls wir nichts anderes sagen, stehen V und W für Vektorräume über einem Körper K und $T : V \to W$ ist eine lineare Abbildung.

Übung 3.1.2. Zeigen Sie die folgende Äquivalenz:

$$T \text{ ist linear} \iff \forall a, b \in K,\ v_1, v_2 \in V : T(av_1 + bv_2) = aT(v_1) + bT(v_2).$$

Bemerkung 3.1.3. Das Wort Abbildung ist ein Synonym für Funktion. Jedoch verwendet man in diesem Zusammenhang hier eher den Ausdruck „lineare Abbildung“ und nicht „lineare Funktion“, obwohl sie beide dasselbe bedeuten.

Bemerkung 3.1.4. Manchmal schreiben wir der Einfachheit halber Tv statt $T(v)$.

Beispiele

Es ist schwierig zu entscheiden, was das beste erste Beispiel für eine lineare Abbildung ist: Es gibt viele! Einige kommen aus der Geometrie, einige aus der Analysis und es gibt noch mehr. Da wir lineare Abbildungen schlussendlich von einem abstrakten Standpunkt aus betrachten wollen, fangen wir mit abstrakten Beispielen an.

Beispiel 3.1.5. Sei V ein Vektorraum. Es gibt immer zwei lineare Abbildungen in $\mathrm{Hom}(V, V)$:

- Die Identitäts-Abbildung

$$\begin{aligned} \mathrm{Id}_V : V &\to V, \\ v &\mapsto v, \end{aligned}$$

  die jeden Vektor auf sich selbst abbildet. Manchmal bezeichnen wir die Abbildung als Id oder I, falls der zugrundeliegende Vektorraum aus dem Kontext klar ist.

- Die Null-Abbildung

$$\begin{aligned} 0 : V &\to W, \\ v &\mapsto 0_W. \end{aligned}$$

Beachten Sie, dass wir insbesondere auch $W = V$ wählen können.

(Wieso sind dies lineare Abbildungen? Überprüfen Sie, dass die beiden Abbildungen linear sind!)

Beispiel 3.1.6. Hier ist ein Beispiel, welches durch die Analysis motiviert ist. Sei K ein Körper. Wir definieren

$$\begin{aligned} D : K[x] &\to K[x], \\ a_0 + a_1 x + \ldots + a_n x^n &\mapsto a_1 + 2a_2 x + 3a_3 x^2 + \ldots + n a_n x^{n-1}. \end{aligned}$$

Hier stehen wie immer die Zahlen $1, 2, 3 \ldots \in K$ für $1, 1+1, 1+1+1, \ldots \in K$ und D steht für Differenziation. Man könnte überprüfen, dass für alle $a, b \in K$ und $p, q \in K[x]$

$$D(ap + bq) = aD(p) + bD(q) \tag{3.2}$$

gilt. Mit der Schreibweise $p' := D(p)$ (bzw. $q' = D(q)$) ist (3.2) genau die Formel

$$(ap + bq)' = ap' + bq',$$

die Sie aus der Mittelschule kennen.

Übung 3.1.7. (1) Falls $\operatorname{char}(K) = 0$, zeigen Sie, dass

$$\{p \in K[x] \mid D(p) = 0\} = \{\alpha \mid \alpha \in K\} = \{\text{konstante Polynome}\}.$$

(2) Sei p eine Primzahl. Berechnen Sie

$$\{q \in \mathbb{F}_p[x] \mid D(q) = 0\}.$$

Beispiel 3.1.8. Das Integral ist auch eine lineare Abbildung

$$\begin{aligned} I : I([0,1]) &\to \mathbb{R} \\ f &\mapsto \int_0^1 f(x)\,\mathrm{d}x, \end{aligned}$$

wobei $I([0,1])$ die Menge aller Riemann-integrierbaren Funktionen auf $[0,1]$ ist.

Beispiel 3.1.9. Viele geometrische Abbildungen der Ebene sind linear. Zum Beispiel Reflexion, Rotation, Streckung und Stauchung und so weiter. Dies besprechen Sie in der Übungsstunde. Um Ihr geometrisches Verständnis zu üben, können Sie sich dieses Video oder auch dieses Video von 3Blue1Brown anschauen.

Beispiel 3.1.10. Sie besprechen ebenfalls Beispiele mit Matrizenräumen in der Übungsstunde.

Beispiel 3.1.11. Die Abbildung

$$\begin{aligned} m_{x^{73}} : \mathbb{F}_{37}[x]_{12} &\to \mathbb{F}_{37}[x]_{210} \\ p &\mapsto x^{73} \cdot p. \end{aligned}$$

ist linear.

Beispiel 3.1.12. Die Abbildung (vgl. Einführung in die Lineare Algebra [1])

$$\begin{aligned} S : K^\infty &\to K^\infty \\ (a_1, a_2, \ldots) &\mapsto (a_2, a_3, \ldots) \end{aligned}$$

ist linear. Diese Abbildung heisst auch *Verschiebungsabbildung* und ist zentral im Gebiet der dynamischen Systeme. Wir haben in unserer Fibonacci-Einführung ([1]) gesehen, dass man S auf den Untervektorraum

$$\mathbf{Fib}(K) := \{(a_1, a_2, \ldots) \in K \mid a_n = a_{n-1} + a_{n-2} \text{ für alle } n \geq 3\} \subseteq K^\infty$$

einschränken und

$$S\mid_{\mathbf{Fib}(K)} : \mathbf{Fib}(K) \to \mathbf{Fib}(K)$$

studieren kann.

Beispiel 3.1.13 (Mutterbeispiel)**.** Zu guter Letzt geben wir jetzt die wichtigste lineare Abbildung zwischen endlich-dimensionalen Vektorräumen an.

Seien $n, m \in \mathbb{N}$ und sei $A \in M_{m\times n}(K)$. Dann ist die Abbildung

$$\begin{aligned} m_A : K^n &\to K^m \\ x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} &\mapsto Ax = A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \end{aligned}$$

linear. Bemerken Sie, Folgendes:

- Das Produkt Ax haben wir in Definition 2.1.4 definiert.
- Hier sollte man K^n und K^m als K^n_{Spal} und K^m_{Spal} betrachten.
- Wir werden in Kürze zeigen, dass sich jede lineare Abbildung zwischen K^n und K^m als m_A für eine Matrix $A \in M_{m\times n}(K)$ darstellen lässt. Ausserdem werden wir später in diesem Kapitel erklären, dass jede lineare Abbildung zwischen zwei beliebigen endlich-dimensionalen Vektorräumen gewissermassen durch solche Abbildungen darstellbar ist.

- Da diese Abbildung so zentral ist, bezeichnen viele Autoren (und auch wir manchmal) diese Abbildung mit T_A statt m_A.

Proposition 3.1.14 (Erste Eigenschaften)**.** *Seien V und W zwei Vektorräume über einem Körper K und $T : V \to W$ eine lineare Abbildung. Dann gelten folgende Aussagen:*

(1) Für alle $a_1, \ldots, a_n \in K$ und $v_1, \ldots, v_n \in V$ gilt

$$T\left(\sum_{i=1}^{n} a_i v_i\right) = \sum_{i=1}^{n} a_i T(v_i).$$

Dass heisst, dass das Bild einer Linearkombination der v_i mit Koeffizienten a_i eine Linearkombination der $T(v_i)$ mit Koeffizienten a_i ist.

(2) Es gilt $T(0) = 0$ (oder für die verwirrten Leser: $T(0_V) = 0_W$).

Beweis. (1) Dies zeigt man mittels Induktion mit Hilfe von der Definition in (3.1).

(2) Da $0_V + 0_V = 0_V$ ist, gilt

$$T(0_V) = T(0_V + 0_V) = T(0_V) + T(0_V),$$

wobei wir im letzten Schritt die Linearität der Abbildung T benutzt haben. Wir addieren $-T(0_V) \in W$ auf beiden Seiten und erhalten

$$0_W = T(0_V) - T(0_V) = T(0_V) + T(0_V) - T(0_V) = T(0_V).$$

Dies beweist die Proposition. □

Satz 3.1.15. *Sei V ein endlich-dimensionaler Vektorraum und $v_1, \ldots, v_n \in V$ eine **Basis** von V. Seien $w_1, \ldots, w_n \in W$ **beliebige** Vektoren. Dann existiert eine eindeutige lineare Abbildung $T : V \to W$ mit*

$$T(v_i) = w_i$$

für $i = 1, \ldots, n$.

Bemerkung 3.1.16. Denken Sie wie immer bei solchen Aussagen an zwei Tatsachen: Existenz und Eindeutigkeit.

Beweis von Satz 3.1.15. Wir definieren T durch einen Vorgang, der häufig benutzt wird: „lineare Erweiterung“.

Sei $v \in V$. Wir wollen $T(v)$ definieren. Dafür schreiben wir zuerst v als Linearkombination der Basis $v_1, \ldots, v_n$

$$v = a_1 v_1 + \ldots + a_n v_n = \sum_{i=1}^{n} a_i v_i,$$

wobei $a_1, \ldots, a_n \in K$. Erinnern Sie sich daran, dass diese Schreibweise laut Proposition 2.2.35 eindeutig ist. Proposition 3.1.14 sagt uns: Falls es eine lineare Abbildung $T : V \to W$ gäbe, dann würde

$$T(v) = T\left(\sum_{i=1}^{n} a_i v_i\right) = \sum_{i=1}^{n} a_i T(v_i) = \sum_{i=1}^{n} a_i w_i \tag{3.3}$$

gelten. Wir können $T(v)$ genau so definieren, dass (3.3) gilt! Wir definieren also

$$T(v) = \sum_{i=1}^{n} a_i w_i.$$

Dies ist wohl-definiert, da die Schreibweise von v als Linearkombination von $v_1, \ldots, v_n$ eindeutig ist. Wir müssen nun zeigen, dass T linear ist. Seien $v, v' \in V$, so dass

$$v = \sum_{i=1}^{n} a_i v_i \quad \text{und} \quad v' = \sum_{i=1}^{n} b_i v_i$$

gilt für $a_1, \ldots, a_n, b_1, \ldots, b_n \in K$. Dann ist

$$v + v' = \sum_{i=1}^{n} (a_i + b_i) v_i$$

die eindeutige Schreibweise von $v + v'$ bezüglich der Basis $v_1, \ldots, v_n$. Laut unserer Definition von T gilt

$$T(v + v') = \sum_{i=1}^{n} (a_i + b_i) w_i = \sum_{i=1}^{n} (a_i w_i + b_i w_i) = \sum_{i=1}^{n} a_i w_i + \sum_{i=1}^{n} b_i w_i = T(v) + T(v'),$$

was die Additivität von T zeigt. Sei jetzt $\alpha \in K$. Es gilt $\alpha v = \sum_{i=1}^{n} \alpha a_i v_i$ und daher

$$T(\alpha v) = \sum_{i=1}^{n} \alpha a_i w_i = \alpha \sum_{i=1}^{n} a_i w_i = \alpha T(v).$$

Dies zeigt, dass T linear ist und die obige Erklärung zeigt, dass dies die einzig mögliche Definition für eine lineare Abbildung $T : V \to W$ mit $T(v_i) = w_i$ ist. $\square$

Beispiel 3.1.17. Wieviele lineare Abbildungen $T : \mathbb{F}_3^4 \to \mathbb{F}_3^2$ gibt es oder, in anderen

Worten, was ist die Kardinalität von $\mathrm{Hom}(\mathbb{F}_3^4, \mathbb{F}_3^2)$? Wir betrachten die Basis $\{e_1, \ldots, e_4\}$ von $\mathbb{F}_3^4$. Laut Satz 3.1.15 ergibt jede beliebige Wahl von Bildern $T(e_1), \ldots, T(e_4)$ eine lineare Abbildung $T : \mathbb{F}_3^4 \to \mathbb{F}_3^2$. Da $|\mathbb{F}_3^2| = 9$ ist, gibt es genau 9^4 solche verschiedenen Wahlen von Vektoren. Daher gilt $|\mathrm{Hom}(\mathbb{F}_3^4, \mathbb{F}_3^2)| = 9^4$.

Man könnte dies mit beliebigen Abbildungen (das heisst, Funktionen) von $\mathbb{F}_3^4$ nach $\mathbb{F}_3^2$ vergleichen. Es gibt $|\mathbb{F}_3^2|^{|\mathbb{F}_3^4|} = (3^2)^{(3^4)} = 9^{81}$ solche Abbildungen. (Wieso?) Insbesondere sind $9^{81} - 9^4$ dieser Abbildungen (und somit die überwältigende Mehrheit) nicht linear! Ähnlich gilt

$$|\mathrm{Hom}(V, \mathbb{F}_p^k)| = (p^k)^n$$

wobei p eine Primzahl, k eine natürliche Zahl und V ein n-dimensionaler Vektorraum über $\mathbb{F}_p$ ist.

Wir betrachten jetzt K^n mit der Standard-Basis $e_1, \ldots, e_n$. Seien $w_1, \ldots, w_n \in K^m$ beliebige Vektoren. Laut Satz 3.1.15 existiert eine eindeutige lineare Abbildung

$$T : K^n \to K^m$$

mit $T(e_i) = w_i$ für $i = 1, \ldots, n$. Sei A die Matrix, deren Spalten $w_1, \ldots, w_n$ sind, also

$$A = \begin{pmatrix} | & | & & | \\ w_1 & w_2 & \cdots & w_n \\ | & | & & | \end{pmatrix}.$$

Dann gilt $m_A(e_i) = Ae_i = w_i$. Das heisst, m_A ist die einzige lineare Abbildung T, die die Eigenschaften von Satz 3.1.15 erfüllt. Da $w_1 \ldots, w_n$ beliebig sind, folgt das folgende Lemma:

Lemma 3.1.18. *Für jede lineare Abbildung $T : K^n \to K^m$ existiert ein eindeutiges $A \in M_{m\times n}(K)$, so dass $T = m_A$. Nämlich*

$$A = \begin{pmatrix} | & | & & | \\ T(e_1) & T(e_2) & \cdots & T(e_n) \\ | & | & & | \end{pmatrix} \in M_{m\times n}(K).$$

3.2 Kern und Bild

Auch hier sind V und W zwei Vektorräume über einem beliebigen Körper K. Mit jeder linearen Abbildung sind (zumindest) zwei Untervektorräume verbunden:

Proposition 3.2.1 (Definition und Proposition)**.** *Sei $T : V \to W$ eine lineare Abbildung.*

(1) Die Menge

$$\operatorname{Ker}(T) = \{v \in V \mid Tv = 0_W\}$$

ist ein Unterverktorraum von V und heisst der Kern *von T.*

(2) Das Bild *von T*

$$\operatorname{Im}(T) = \{Tv \mid v \in V\}$$

ist ein Unterverktorraum von W.[2]

Erinnerung: Für eine Funktion $f : A \to B$ und $b \in B$ ist

$$f^{-1}(b) = f^{-1}(\{b\}) = \{a \in A : f(a) = b\} \tag{3.4}$$

die Menge aller Urbilder von b. Mit dieser Notation ist $\operatorname{Ker}(T) = T^{-1}(0_W)$.

Beweis. (1) Seien $a, b \in K$ und $v_1, v_2 \in \operatorname{Ker}(T)$. Dann gilt

$$T(av_1 + bv_2) = aTv_1 + bTv_2 = a \cdot 0_W + b \cdot 0_W = 0_W.$$

Ausserdem ist $0_V \in \operatorname{Ker}(T)$ laut Proposition 3.1.14 und daher ist $\operatorname{Ker}(T)$ ein Untervektorraum von V.

(2) Seien jetzt $w_1, w_2 \in \operatorname{Im}(T)$. Dann existieren $v_1, v_2 \in V$ mit $Tv_1 = w_1$ und $Tv_2 = w_2$ laut der Definition von $\operatorname{Im}(T)$. Es gilt

$$T(av_1 + bv_2) = aTv_1 + bTv_2 = aw_1 + bw_2$$

und daher ist $aw_1 + bw_2 \in \operatorname{Im}(T)$. Auch hier folgt aus Proposition 3.1.14, dass $0_W \in \operatorname{Im}(T)$ ist, da $T(0_V) = 0_W$. Daher ist $\operatorname{Im}(T)$ ein Untervektorraum von W.

Dies beweist die Proposition. □

Beispiel 3.2.2 (Kern und homogenene lineare Gleichungssysteme)**.** Sei $A \in M_{m\times n}(K)$. Wir betrachten $m_A : K^n \to K^m$. Dann gilt

$$\operatorname{Ker}(m_A) = \text{Lös}(A, 0_{K^m}).$$

Das heisst, $\operatorname{Ker}(m_A)$ sind die Lösungen des homogenen linearen Gleichungssystems

$$Ax = 0_{K^m}, \quad x \in K^n.$$

Ab jetzt werden wir $\operatorname{Ker}(m_A)$ auch mit $\operatorname{Ker}(A)$ bezeichnen.

[2]Viele Autoren verwenden auch $\operatorname{Bild}(T)$ statt $\operatorname{Im}(T)$.

Beispiel 3.2.2 zeigt schon, dass $\operatorname{Ker}(A)$ interessant ist. Hier ist noch ein weiterer interessanter Grund:

Proposition 3.2.3. *Sei T eine lineare Abbildung. Dann gelten folgende Aussagen:*

(1) T ist injektiv $\iff \operatorname{Ker}(T) = \{0_V\}$.

(2) T ist surjektiv $\iff \operatorname{Im}(T) = W$.

Beweis. Bemerken Sie zuerst: In (2) gibt es nichts zu beweisen! Wir beweisen also (1):

„$\Rightarrow$“: Wir haben schon gesehen, dass $\operatorname{Ker}(T)$ die Menge aller Urbilder von 0_W ist. Das heisst, dass $\operatorname{Ker}(T) = T^{-1}(0_W)$. Laut Proposition 3.1.14 ist 0_V ein Urbild von 0_W. Da T injektiv ist, hat jedes Element von W höchstens ein Urbild und daher ist

$$\operatorname{Ker}(T) = T^{-1}(0_W) = \{0_V\}.$$

„$\Leftarrow$“: Hier brauchen wir die Linearität. Seien $v_1, v_2 \in V$ mit $T(v_1) = T(v_2)$. Da T linear ist, gilt

$$T(v_1 - v_2) = T(v_1) - T(v_2) = 0_W.$$

Daher ist $v_1 - v_2 \in \operatorname{Ker}(T)$ und nach Annahme ist $\operatorname{Ker}(T) = \{0_V\}$. Also folgt, dass $v_1 = v_2$ ist, was die Injektivität zeigt. □

Wir können den Sachverhalt in Proposition 3.2.3 vage so ausdrücken:

„Injektivität überall ist äquivalent zu Injektivität bei der Null“.

Wir werden sehen, dass es auch andere Eigenschaften ausser der Injektivität gibt, die sich „bei der Null“ überprüfen lassen.

Übung 3.2.4. Verallgemeinern Sie die Resultate dieses Abschnitts: Sei $T : V \to W$ linear. Zeigen Sie folgende Aussagen:

1. Sei $V' \subseteq V$ ein Untervektorraum. Dann ist $T(V')$ ein Untervektorraum von W.
2. Sei $W' \subseteq W$ ein Untervektorraum. Dann ist $T^{-1}(W')$ ein Untervektorraum von V.[3]

Definition 3.2.5. Eine lineare Abbildung $T : V \to W$ nennen wir[4]

- einen *Isomorphismus*, falls T bijektiv ist.

[3] $T^{-1}(W') = \{v \in V : Tv \in W'\}$.

[4] Wir werden folgende Begriffe nicht benutzen, aber andere Autoren definieren auch:

- T heisst *Monomorphismus*, falls T injektiv ist.
- T heisst *Epimorphismus*, falls T surjektiv ist.

- einen *Endomorphismus*, falls $W = V$ ist.
- einen *Automorphismus*, falls $V = W$ und T bijektiv ist.

Falls ein Isomorphismus $T : V \to W$ existiert, sagen wir, dass V und W *isomorph* sind und schreiben $V \cong W$. Die Menge aller Endomorphismen von V bezeichnen wir mit $\mathrm{End}(V)$. Das heisst, $\mathrm{End}(V) := \mathrm{Hom}(V, V)$.

Lemma 3.2.6. *Falls $T : V \to W$ ein Isomorphismus ist, dann existiert $T^{-1} : W \to V$ und ist auch linear.*

Beweis. Seien $a, b \in K$ und $w_1, w_2 \in W$. Seien $v_1, v_2 \in V$ die eindeutigen Urbilder von w_1, w_2. Das heisst, dass $T(v_i) = w_i$ für $i = 1, 2$ oder äquivalent $T^{-1}(w_i) = v_i$ für $i = 1, 2$ gilt. Da T linear ist, gilt

$$T(av_1 + bv_2) = aTv_1 + bTv_2 = aw_1 + bw_2.$$

Das heisst, weil T bijektiv ist, dass

$$T^{-1}(aw_1 + bw_2) = av_1 + bv_2 = aT^{-1}(w_1) + bT^{-1}(w_2).$$

Dies zeigt, dass T^{-1} linear ist. □

Übung 3.2.7. Zeigen Sie, dass „isomorph" eine Äquivalenzrelation auf der Menge der Vektorräume ist.

Lemma 3.2.8. *Sei $T : V \to W$ linear und $v_1, \ldots, v_n$ eine Basis für V. Dann gilt $\mathrm{Im}(T) = \mathrm{Sp}(Tv_1, \ldots, Tv_n)$.*

Beweis. Eine Gleichheit zwischen zwei Mengen zeigt man, indem man zwei Inklusionen zeigt.

„$\supseteq$": Für $a_1, \ldots, a_n \in K$ gilt

$$\sum_{i=1}^{n} a_i T v_i = T\left(\sum_{i=1}^{n} a_i v_i\right) \tag{3.5}$$

und daher $\sum_{i=1}^{n} a_i T v_i \in \mathrm{Im}(T)$.

„$\subseteq$": Sei $w \in \mathrm{Im}(T)$ und $v \in V$ mit $Tv = w$. Da $v_1, \ldots, v_n$ eine Basis von V ist, existieren $a_1, \ldots, a_n \in K$ mit $v = \sum_{i=1}^{n} a_i v_i$. Jetzt lesen wir Gleichung (3.5) umgekehrt:

$$w = Tv = T\left(\sum_{i=1}^{n} a_i v_i\right) = \sum_{i=1}^{n} a_i T v_i$$

und daher ist $w \in \mathrm{Sp}(Tv_1, \ldots, Tv_n)$. □

Wir sind jetzt bereit den wichtigsten Satz zu linearen Abbildungen zu zeigen:

Satz 3.2.9 (Rangsatz, Energieerhaltungs-Satz)**.** *Sei $T : V \to W$ eine lineare Abbildung zwischen Vektorräumen V und W, wobei V endlich-dimensional ist. Dann gilt*

$$\dim \mathrm{Ker}(T) + \dim \mathrm{Im}(T) = \dim V.$$

Beweis. Sei $n = \dim V$ und $u_1, \dots, u_k$ eine Basis von $\mathrm{Ker}(T)$. Mittels Lemma 2.3.14 ergänzen wir $u_1, \dots, u_k$ zu einer Basis von V:

$$u_1, \dots, u_k, v_1, \dots, v_{n-k}.$$

Laut Lemma 3.2.8 gilt

$$\mathrm{Im}(T) = \mathrm{Sp}(Tu_1, \dots, Tu_k, Tv_1, \dots, Tv_{n-k}) = \mathrm{Sp}(Tv_1, \dots, Tv_{n-k}), \tag{3.6}$$

da $Tu_1 = \ldots = Tu_k = 0_W$. Wir behaupten, dass $Tv_1, \dots, Tv_{n-k}$ eine Basis von $\mathrm{Im}(T)$ ist. Dazu müssen wir nur zeigen, dass $Tv_1, \dots, Tv_{n-k}$ linear unabhängig sind. (Wieso? Wegen (3.6)!)

Seien also $a_1, \dots, a_{n-k} \in K$ mit $\sum_{i=1}^{n-k} a_i T v_i = 0_W$. Da T linear ist, gilt

$$T\left(\sum_{i=1}^{n-k} a_i v_i\right) = \sum_{i=1}^{n-k} a_i T v_i = 0_W.$$

Also ist $\sum_{i=1}^{n-k} a_i v_i \in \mathrm{Ker}(T)$. Daher existieren $b_1, \dots, b_k$ mit

$$\sum_{i=1}^{n-k} a_i v_i = \sum_{i=1}^{k} b_i u_i$$

beziehungsweise

$$\sum_{i=1}^{n-k} a_i v_i + \sum_{i=1}^{k} (-b_i) u_i = 0_V.$$

Weil $u_1, \dots, u_n, v_1, \dots, v_{n-k}$ eine Basis von V ist, impliziert das

$$a_1 = \ldots = a_{n-k} = -b_1 = \ldots = -b_k = 0.$$

Also ist insbesondere $a_1 = \ldots = a_{n-k} = 0$ und daraus folgt, dass $Tv_1, \dots, Tv_{n-k}$ eine Basis von $\mathrm{Im}(T)$ ist. Dies zeigt, dass

$$\dim \mathrm{Im}(T) = n - k.$$

Also gilt

$$\dim V = k + (n - k) = \dim \mathrm{Ker}(T) + \dim \mathrm{Im}(T)$$

per Definition von n und k als $n = \dim V$ und $k = \dim \mathrm{Ker}(T)$. □

Bemerkung 3.2.10. Ich stelle mir Satz 3.2.9 immer als „Energieerhaltungs-Satz" vor: Der Vektorraum V hat $n = \dim V$ Dimensionen/Energie zu geben. Die Abbildung T gibt einige dieser Dimensionen dem Bild $\mathrm{Im}(T)$ und die restlichen Dimensionen müssen dann im Kern $\mathrm{Ker}(T)$ verschwinden.

Korollar 3.2.11. *Sei $T : V \to W$ eine lineare Abbildung. Dann gelten folgende Aussagen:*

(1) Falls $\dim W < \dim V$*, dann ist T <u>nicht</u> injektiv.*

(2) Falls $\dim W > \dim V$*, dann ist T <u>nicht</u> surjektiv.*

(3) Falls $\dim W = \dim V$*, dann gilt*

$$T \textit{ ist bijektiv} \iff T \textit{ ist injektiv} \iff T \textit{ ist surjektiv.}$$

Beweis. (1) Hier hat V zu viel „Energie": V kann W nur $\dim W$ Dimensionen „geben" und daher muss T einige Dimensionen an den Kern abgeben. Also ist der Kern nicht trivial. Laut Proposition 3.2.3 ist T dann nicht injektiv.

Hier ist eine rigorose Version: Da $\mathrm{Im}(T)$ ein Untervektorraum von W ist, gilt $\dim \mathrm{Im}(T) \leq \dim W < \dim V$. Laut Satz 3.2.9 folgt dann, dass

$$\dim \mathrm{Ker}(T) = \dim V - \dim \mathrm{Im}(T) > 0.$$

Laut Proposition 3.2.3 ist T dann nicht injektiv.

(2) Energie-Erhaltungs-Version: $\mathrm{Im}(T)$ braucht $\dim W$ Dimensionen, damit T surjektiv sein kann. Aber $\dim V$ hat nicht genug Energie, da $\dim V < \dim W$.

Rigorose Version: Laut Satz 3.2.9 gilt

$$\dim \mathrm{Im}(T) = \dim V - \dim \mathrm{Ker}(T) \leq \dim V < \dim W.$$

Also folgt aus Proposition 3.2.3, dass T nicht surjektiv ist.

(3) Wir müssen nur zeigen, dass T injektiv ist genau dann, wenn T surjektiv ist. (Wieso?) Ohne viele Worte gelten die Äquivalenzen:

$$\begin{aligned}
T \text{ ist injektiv} \overset{\text{Proposition 3.2.3}}{\iff} \operatorname{Ker}(T) = \{0\} &\iff \dim \operatorname{Ker}(T) = 0 \\
&\overset{\text{Satz3.2.9}}{\iff} \dim V = \dim \operatorname{Im}(T) \\
&\overset{\text{Annahme}}{\iff} \dim W = \dim \operatorname{Im}(T) \\
&\iff W = \operatorname{Im}(T) \\
&\iff T \text{ ist surjektiv.}
\end{aligned}$$

Dies zeigt das Korollar. □

Lemma 3.2.12. *Seien $T : V \to W$ und $S : W \to U$ lineare Abbildungen. Dann ist die Verkettung $S \circ T : V \to U$ auch linear.*

Beweis. Seien $\alpha \in K$ und $v_1, v_2 \in V$. Es gilt

$$S \circ T(v_1 + v_2) \overset{T \text{ linear}}{=} S(Tv_1 + Tv_2) \overset{S \text{ linear}}{=} S \circ T(v_1) + S \circ T(v_2).$$

Ähnlich folgt, dass

$$S \circ T(\alpha v) = S(\alpha T v) = \alpha(S \circ T(v)).$$

Dies beweist das Lemma. □

Übung 3.2.13. Es gibt auch eine andere Beweis-Strategie für Satz 3.2.9:

Man fängt mit einer Basis $v_1, \dots, v_k$ von $\operatorname{Ker}(T)$ und einer Basis $w_1, \dots, w_l$ von $\operatorname{Im}(T)$ an. Dann wählt man beliebige $u_1, \dots, u_l \in V$ mit

$$Tu_i = w_i$$

für $i = 1, \dots, l$. Man zeigt dann, dass $v_1, \dots, v_k, u_1, \dots, u_l$ eine Basis von V ist. Der Satz folgt somit.

Führen Sie diese Strategie aus!

Als ob es einfach wäre und nicht so wichtig, folgt jetzt nichtsdestotrotz eines der wichtigsten Resultate der linearen Algebra.

Korollar 3.2.14. *Zwei endlich-dimensionale Vektorräume V und W sind genau dann isomorph, wenn $\dim V = \dim W$.*

Beweis. „$\Rightarrow$": Falls $T : V \to W$ ein Isomorphismus ist, dann folgt aus Proposition 3.2.3, dass $\operatorname{Ker}(T) = \{0\}$ und $W = \operatorname{Im}(T)$. Daher impliziert Satz 3.2.9, dass

$$\dim W = \dim \operatorname{Im}(T) = \dim V - \dim \operatorname{Ker}(T) = \dim V.$$

„$\Leftarrow$“: Falls $\dim V = \dim W$ ist, seien $(v_1, \ldots, v_n)$ eine Basis von V und $(w_1, \ldots, w_n)$ eine Basis von W. Wir definieren mittels Satz 3.1.15 die (eindeutige) lineare Abbildung

$$T : V \to W \quad \text{mit} \quad T(v_i) = w_i.$$

Laut Lemma 3.2.8 ist T surjektiv und aufgrund der Annahme $\dim V = \dim W$ ist T dann auch injektiv (mit Hilfe von Korollar 3.2.11 (3)). Daher ist T ein Isomorphismus. □

Für später brauchen wir noch eine kleine Definition:

Definition 3.2.15. Sei $T : V \to W$ eine lineare Abbildung. Der *Rang* von T ist

$$\operatorname{Rang}(T) := \dim \operatorname{Im}(T).$$

Übung 3.2.16. Seien $T : V \to W$ und $S : W \to U$ zwei lineare Abbildungen zwischen endlich-dimensionalen Vektorräumen U, V und W. Zeigen Sie folgende Aussagen:

(1) Es gilt $\operatorname{Rang}(S \circ T) \leq \min\{\operatorname{Rang}(T), \operatorname{Rang}(S)\}$.

(2) Falls S injektiv ist, dann ist $\operatorname{Rang}(S \circ T) = \operatorname{Rang}(T)$.

(3) Falls T surjektiv ist, dann ist $\operatorname{Rang}(S \circ T) = \operatorname{Rang}(S)$.

(Hinweis: Betrachten Sie $S|_{\operatorname{Im}(T)}$.)

3.3 Koordinaten für Vektorräume und lineare Abbildungen

Alle Vektorräume in diesem Abschnitt sind endlich-dimensional. In diesem Abschnitt werden wir ein bisschen mehr Information bezüglich Korollar 3.2.14 geben: Ein Vektorraum V über K mit $\dim V = n$ ist isomorph zu K^n.

Für den Beweis und auch für später müssen wir Koordinaten definieren. Dazu müssen wir an Basen als geordnete Mengen denken, das heisst, als eine Liste der Form $v_1, \ldots, v_n$. Dafür schreiben wir $\mathcal{B} = (v_1, \ldots, v_n)$, wenn $\mathcal{B}$ eine *geordnete Basis* ist, welche durch die Liste $v_1, \ldots, v_n$ definiert ist.

Definition 3.3.1. Sei V ein Vektorraum mit Basis $\mathcal{B} = (v_1, \ldots, v_n)$ und sei $v \in V$.

Wir definieren den *Koordinaten-Vektor* von v bezüglich der Basis $\mathcal{B}$ als

$$[v]_{\mathcal{B}} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

wobei $a_1, \ldots, a_n \in K$ sind mit $v = a_1 v_1 + \ldots + a_n v_n$. Da $\mathcal{B}$ eine Basis ist, sind $a_1, \ldots, a_n$ eindeutig definiert, und daher ist $[v]_{\mathcal{B}}$ wohl-definiert.

Die Abbildung

$$\Phi_{\mathcal{B}} : V \to K^n, \quad v \mapsto [v]_{\mathcal{B}}$$

nennen wir die *Koordinaten-Abbildung* bezüglich $\mathcal{B}$.

Proposition 3.3.2. *Die Abbildung $\Phi_{\mathcal{B}}$ ist linear und bijektiv.*

Beweis. Man kann Satz 3.1.15 verwenden, um zu zeigen, dass $\Phi_{\mathcal{B}}$ die eindeutige lineare Abbildung mit $T(v_i) = e_i \in K^n$ ist, wobei $e_1, \ldots, e_n$ die Standard-Basis von K^n ist.

Wir geben stattdessen einen direkten Beweis. Seien $a, b \in K$ und $v, v' \in V$. Wir schreiben

$$\begin{aligned} v &= a_1 v_1 + \ldots + a_n v_n, \\ v' &= b_1 v_1 + \ldots + b_n v_n \end{aligned}$$

womit

$$[v]_{\mathcal{B}} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}, \ [v']_{\mathcal{B}} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.$$

Dann gilt

$$av + bv' = (aa_1 + bb_1)v_1 + \ldots + (aa_n + bb_n)v_n$$

und daher

$$\Phi_{\mathcal{B}}(av + bv') = [av + bv']_{\mathcal{B}} = \begin{pmatrix} aa_1 + bb_1 \\ \vdots \\ aa_n + bb_n \end{pmatrix} = a \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} + b \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = a\Phi_{\mathcal{B}}(v) + b\Phi_{\mathcal{B}}(v').$$

Dies zeigt die Linearität von $\Phi_{\mathcal{B}}$.

Die Bijektivität folgt, da $\mathcal{B}$ eine Basis ist. (Wieso? Benutzen Sie die Eindeutigkeit – Proposition 2.2.35.) □

Es ist nützlich zu bemerken, dass die Inverse

$$\Phi_{\mathcal{B}}^{-1} : K^n \to V$$

durch

$$\Phi_{\mathcal{B}}^{-1}\begin{pmatrix}a_1\\ \vdots\\ a_n\end{pmatrix} = a_1v_1 + \ldots + a_nv_n$$

gegeben ist.

Beispiel 3.3.3. Sei $\mathcal{B} = \left(\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\-1\end{pmatrix}\right)$ eine Basis von $\mathbb{Q}^2$ und $\mathcal{E} = \left(\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right)$ die Standardbasis. Dann ist beispielsweise

$$\left[\begin{pmatrix}1\\0\end{pmatrix}\right]_{\mathcal{B}} = \begin{pmatrix}\frac{1}{2}\\ \frac{1}{2}\end{pmatrix} \quad \text{und} \quad \left[\begin{pmatrix}0\\1\end{pmatrix}\right]_{\mathcal{B}} = \begin{pmatrix}\frac{1}{2}\\ -\frac{1}{2}\end{pmatrix}.$$

Allgemein ist

$$\left[\begin{pmatrix}x\\y\end{pmatrix}\right]_{\mathcal{B}} = \begin{pmatrix}\frac{x+y}{2}\\ \frac{x-y}{2}\end{pmatrix},$$

da

$$\frac{x+y}{2}\begin{pmatrix}1\\1\end{pmatrix} + \frac{x-y}{2}\begin{pmatrix}1\\-1\end{pmatrix} = \begin{pmatrix}x\\y\end{pmatrix}.$$

Bezüglich der Standardbasis gilt

$$\left[\begin{pmatrix}x\\y\end{pmatrix}\right]_{\mathcal{E}} = \begin{pmatrix}x\\y\end{pmatrix}.$$

(Wieso?)

Beispiel 3.3.4. Sei $V = K[X]_3$ und $\mathcal{B} = (1, x, x^2, x^3)$ die Standard-Basis und

$$\mathcal{C} = (1, x+1, x^2+x+1, x^3+x^2+x+1)$$

eine andere Basis. Sei $p = a_0 + a_1x + a_2x^2 + a_3x^3$ ein beliebiger Vektor. Bezüglich der Standard-Basis ist das Leben einfach: Man sieht gleich, dass

$$[p]_{\mathcal{B}} = \begin{pmatrix}a_0\\a_1\\a_2\\a_3\end{pmatrix} \quad \text{und} \quad [x^3]_{\mathcal{C}} = \begin{pmatrix}0\\0\\-1\\1\end{pmatrix}.$$

In diesem Beispiel kann man auch deutlich den Einfluss der Reihenfolge sehen. Zum Beispiel gilt für $\mathcal{B}' = (x^3, x^2, x, 1)$ und $\mathcal{B}'' = (1, x^2, x^3, x)$, dass

$$[p]_{\mathcal{B}'} = \begin{pmatrix} a_3 \\ a_2 \\ a_1 \\ a_0 \end{pmatrix} \quad \text{und} \quad [p]_{\mathcal{B}''} = \begin{pmatrix} a_0 \\ a_2 \\ a_3 \\ a_1 \end{pmatrix}.$$

Für $[p]_{\mathcal{C}}$ müssen wir etwas rechnen. Hier ist die Antwort:

$$[p]_{\mathcal{C}} = \begin{pmatrix} a_0 - a_1 \\ a_1 - a_2 \\ a_2 - a_3 \\ a_3 \end{pmatrix}.$$

Bemerkung. $\Phi_{\mathcal{B}}$ ist die eindeutige lineare Abbildung von V nach K^n, die durch Satz 3.1.15 mit der Bedingung $\Phi_{\mathcal{B}}(v_i) = e_i \in K^n$ definiert ist, wobei $\mathcal{B} = (v_1, \ldots, v_n)$ und $(e_1, \ldots, e_n)$ die Standard-Basis von K^n ist.

In Übung 3.2.7 haben wir gesehen, dass „isomorph sein" eine Äquivalenzrelation auf der Menge aller Vektorräume ist. Korollar 3.2.14 besagt, dass die Dimension aller Elemente einer Äquivalenzklasse gleich ist. Jetzt können wir sogar einen Repräsentanten für die Äquivalenzklasse finden:

Satz 3.3.5. *Sei V ein Vektorraum über K mit $\dim V = n \in \mathbb{N} \cup \{0\}$. Dann ist V isomorph zu K^n.*

Beweis. Sei $\mathcal{B}$ eine Basis von V. Dann ist die Abbildung

$$\Phi_{\mathcal{B}} : V \to K^n$$

ein Isomorphismus. □

Bemerkung 3.3.6 (Nebenbemerkung). Die Abbildung $\Phi_{\mathcal{B}}$ ist nicht-kanonisch. Für ihre Definition müssen wir eine Wahl treffen (die Wahl von $\mathcal{B}$). Vergleichen Sie dies mit der Konstruktion der Quotienten-Abbildung am Ende dieses Kapitels.

Die Koordinaten-Abbildung bringt uns zu unserer nächsten Frage:

Seien V und W zwei Vektorräume mit $\dim V = n$ und $\dim W = m$. Sei $T : V \to W$ eine lineare Abbildung. Wir wählen Basen $\mathcal{B} \subseteq V$ und $\mathcal{C} \subseteq W$ und betrachten das Diagramm:

$$
\begin{tikzcd}
V \arrow[r, "T"] & W \arrow[d, "\Phi_{\mathcal{C}}"] \\
K^n \arrow[u, "\Phi_{\mathcal{B}}^{-1}"] \arrow[r, dashed, "\Phi_{\mathcal{C}} \circ T \circ \Phi_{\mathcal{B}}^{-1}"'] & K^m
\end{tikzcd}
$$

Laut Lemma 3.1.18 ist die lineare Abbildung $\Phi_{\mathcal{C}} \circ T \circ \Phi_{\mathcal{B}}^{-1} : K^n \to K^m$ durch eine Matrix $A \in M_{m \times n}(K)$ darstellbar. Das heisst, es existiert eine eindeutige Matrix $A \in M_{m \times n}(K)$, so dass

$$
\begin{tikzcd}
V \arrow[r, "T"] & W \arrow[d, "\Phi_{\mathcal{C}}"] \\
K^n \arrow[u, "\Phi_{\mathcal{B}}^{-1}"] \arrow[r, dashed, "m_A"'] & K^m
\end{tikzcd} .
$$

kommutiert. Das heisst, dass $\Phi_{\mathcal{C}} \circ T \circ \Phi_{\mathcal{B}}^{-1} = m_A$ gilt.

Toll. Aber ich habe jetzt noch viele Fragen:

(1) Wie kann ich A berechnen?
(2) Wie sieht A bezüglich verschiedenen Basen aus? (3.7)
(3) Wie sieht die Matrix einer Verkettung aus?
Gibt uns die Verkettung eine Möglichkeit Matrizen zu multiplizieren?

Diese Fragen führen uns zu einem neuen Abschnitt!

3.3.1 Matrizen und lineare Abbildungen

Die erste der vorherigen Fragen können wir gleich beantworten. Wir müssen einfach der Definition folgen.

Sei $T : V \to W$ eine lineare Abbildung, $\mathcal{B} = (v_1, \dots, v_n)$ eine Basis des Vektorraums V und $\mathcal{C} = (w_1, \dots, w_m)$ eine Basis des Vektorraums W. Wir betrachten

$$
\begin{tikzcd}
v_i \in V \arrow[r, "T"] \arrow[r, mapsto, bend left] & W \ni Tv_i \arrow[d, "\Phi_{\mathcal{C}}"'] \arrow[d, mapsto, bend left] \\
e_i \in K^n \arrow[u, "\Phi_{\mathcal{B}}^{-1}"'] \arrow[u, mapsto, bend left] \arrow[r, dashed, "m_A"'] & K^m \ni [Tv_i]_{\mathcal{C}}
\end{tikzcd} . \tag{3.8}
$$

für $A \in M_{m \times n}(K)$ wie oben.

Was geschieht mit der Standard-Basis $e_1, \dots, e_n \in K^n$, wenn wir $\Phi_{\mathcal{C}} \circ T \circ \Phi_{\mathcal{B}}^{-1}$ darauf anwenden? Für $i \in \{1, \dots, n\}$ gilt

$$\Phi_{\mathcal{C}} \circ T \circ \Phi_{\mathcal{B}}^{-1}(e_i) = \Phi_{\mathcal{C}} \circ T(v_i) = \Phi_{\mathcal{C}}(T(v_i)) = [T(v_i)]_{\mathcal{C}}.$$

Dies bedeutet, dass m_A die (eindeutige) lineare Abbildung ist mit $m_A(e_i) = [T(v_i)]_{\mathcal{C}}$. Laut Lemma 3.1.18 heisst das, dass

$$A = \begin{pmatrix} | & & | \\ [T(v_1)]_{\mathcal{C}} & \cdots & [T(v_n)]_{\mathcal{C}} \\ | & & | \end{pmatrix} \in M_{m\times n}(K). \tag{3.9}$$

Anders gesagt, ist

$$A = (a_{ij})_{\substack{1\leq i\leq m \\ 1\leq j\leq n}}$$

mit Einträgen $a_{ij} \in K$, die durch die Gleichung

$$T(v_j) = a_{1j}w_1 + \cdots + a_{mj}w_m = \sum_{i=1}^{m} a_{ij}w_i \quad \text{für } j = 1, \ldots, n \tag{3.10}$$

definiert sind.

Dies beantwortet Frage (1). Wir fassen zusammen:

Definition 3.3.7. Die Matrix A wie in (3.9) heisst die *Darstellungsmatrix* von T bezüglich der Basen $\mathcal{B}$ und $\mathcal{C}$ und wird mit $[T]^{\mathcal{B}}_{\mathcal{C}}$ notiert.

Wir haben bewiesen:

Proposition 3.3.8. *Sei $T : V \to W$ eine lineare Abbildung zwischen zwei Vektorräumen V und W. Sei $\mathcal{B}$ eine Basis von V und $\mathcal{C}$ eine Basis von W. Für jedes $v \in V$ gilt*

$$[T]^{\mathcal{B}}_{\mathcal{C}}[v]_{\mathcal{B}} = [Tv]_{\mathcal{C}}. \tag{3.11}$$

Beweis. Gleichung (3.11) ist äquivalent dazu zu sagen, dass das folgende Diagramm kommutiert:

$$\begin{tikzcd}
v \in V \arrow[r, "T"] \arrow[d, "\Phi_{\mathcal{B}}"] & W \ni w \arrow[d, "\Phi_{\mathcal{C}}"'] \\
{[v]_{\mathcal{B}} \in K^n} \arrow[r, "m_{[T]^{\mathcal{B}}_{\mathcal{C}}}"'] & K^m \ni [w]_{\mathcal{C}}
\end{tikzcd} \tag{3.12}$$

Oder anders gesagt, (3.11) ist äquivalent zu $\Phi_{\mathcal{C}} \circ T = m_{[T]^{\mathcal{B}}_{\mathcal{C}}} \circ \Phi_{\mathcal{B}}$. Dies ist jedoch äquivalent zu

$$m_{[T]^{\mathcal{B}}_{\mathcal{C}}} = \Phi_{\mathcal{C}} \circ T \circ \Phi_{\mathcal{B}}^{-1}. \tag{3.13}$$

Wir haben aber $[T]^{\mathcal{B}}_{\mathcal{C}}$ genau so definiert, dass (3.13) gilt (vgl. Diagramm (3.12)). $\square$

Man sollte darüber so nachdenken: Durch die Beschreibung von V durch $\mathcal{B}$ und W durch $\mathcal{C}$ sieht T wie Multiplikation mit $[T]^{\mathcal{B}}_{\mathcal{C}}$ aus (vgl. Epilog Kapitel 2).

Beispiel 3.3.9. Falls $\mathcal{E}_n$ die Standard-Basis von K^n, $\mathcal{E}_m$ die Standard-Basis von K^m ist und $A \in M_{m\times n}(K)$, dann gilt allgemeiner

$$[m_A]^{\mathcal{E}_n}_{\mathcal{E}_m} = A.$$

Beispiel 3.3.10. Betrachten wir noch einmal die Ableitungsabbildung

$$D : K[x]_3 \to K[x]_2$$

und seien $\mathcal{E} = (1, x, x^2, x^3)$ die Standard-Basis von $K[x]_3$ und $\mathcal{E}' = (1, x, x^2)$ die Standard-Basis von $K[x]_2$. Dann gilt

$$\begin{aligned} D(1) &= 0 = 0 \cdot 1 + 0 \cdot x + 0 \cdot x^2 \\ D(x) &= 1 = 1 \cdot 1 + 0 \cdot x + 0 \cdot x^2 \\ D(x^2) &= 2x = 0 \cdot 1 + 2 \cdot x + 0 \cdot x^2 \\ D(x^3) &= 3x^2 = 0 \cdot 1 + 0 \cdot x + 3 \cdot x^2 \end{aligned}$$

und daher ist

$$[D]^{\mathcal{E}}_{\mathcal{E}'} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

Falls man D als einen Endomorphismus von $K[x]_3$ betrachtet, das heisst als eine Abbildung $D : K[x]_3 \to K[x]_3$, dann ist

$$[D]^{\mathcal{E}}_{\mathcal{E}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Übung 3.3.11. Berechnen Sie $[D]^{\mathcal{C}}_{\mathcal{C}}$, $[D]^{\mathcal{C}}_{\mathcal{B}}$ und $[D]^{\mathcal{B}}_{\mathcal{C}}$, wobei $\mathcal{C}$ und $\mathcal{B}$ wie in Beispiel 3.3.4 sind.

Die Matrizen $[D]^{\mathcal{A}'}_{\mathcal{A}}$ für verschiedene Basen $\mathcal{A}$ und $\mathcal{A}'$ sind im allgemeinen verschieden, rühren aber alle von der Abbildung D her. Sie sind „verschiedene Beschreibungen" von D bezüglich verschiedenen Basen. Was haben sie gemeinsam? Wie sind sie miteinander verbunden? Diese Fragen hängen beide mit Frage (2) in (3.7) zusammen.

Für Frage (2) sagen wir jetzt nur, dass die Matrix $[T]^{\mathcal{B}}_{\mathcal{C}}$ stark von der Wahl von $\mathcal{B}$ und $\mathcal{C}$ abhängt, mehr Informationen geben wir später. Wir möchten als nächstes Frage (3) beantworten.

Dazu machen wir aber erst eine Pause, um Matrizen zu besprechen und vor allem um die Matrixmultiplikation einzuführen.

3.3.2 Matrizen

Wie wir später sehen werden, wird uns die Definition der Matrixmultiplikation durch die Antwort auf Frage (3) in (3.7) „gegeben". Ich finde es nichtsdestotrotz verwirrend die Matrixmultiplikation auf diese Art einzuführen. Deshalb werden wir die Geschichte anders erzählen: Wir werden die Matrixmultiplikation einfach definieren und dann zeigen, dass uns dies eine Antwort auf Frage (3) in (3.7) gibt.

Erinnerung: $M_{m\times n}(K)$ bezeichnet die Menge aller Matrizen mit m-Zeilen und n-Spalten und Einträgen in K. Die Menge $M_{m\times n}(K)$ ist ein Vektorraum mit der Addition

$$(a_{ij})_{\substack{1\le i\le m\\1\le j\le n}} + (b_{ij})_{\substack{1\le i\le m\\1\le j\le n}} := (a_{ij}+b_{ij})_{\substack{1\le i\le m\\1\le j\le n}}$$

und mit der Skalarmultiplikation

$$\alpha\cdot(a_{ij})_{\substack{1\le i\le m\\1\le j\le n}} := (\alpha a_{ij})_{\substack{1\le i\le m\\1\le j\le n}}$$

für $\alpha\in K$.

Definition 3.3.12. Seien $m,n,p\in\mathbb{N}$. Die Matrixmultiplikation ist die Abbildung

$$\begin{aligned} \cdot : M_{m\times n}(K)\times M_{n\times p}(K) &\to M_{m\times p}(K),\\ (A,B)=\left((a_{ij})_{\substack{1\le i\le m\\1\le j\le n}},(b_{ij})_{\substack{1\le i\le n\\1\le j\le p}}\right) &\mapsto A\cdot B, \end{aligned}$$

wobei $A\cdot B=(c_{ij})_{\substack{1\le i\le m\\1\le j\le p}}$ mit

$$c_{ij}=a_{i1}b_{1j}+\ldots a_{in}b_{nj}=\sum_{k=1}^{n}a_{ik}b_{kj}. \tag{3.14}$$

Wir schreiben oft AB statt $A\cdot B$.

Bemerkung 3.3.13.

- Merken Sie sich: AB ist nur dann definiert, wenn die Anzahl der Spalten von A der Anzahl Zeilen von B entspricht.
- Die Formel in (3.14) ist nicht sehr übersichtlich, aber von einem theoretischen Blickwinkel aus wichtig. Die folgende Darstellung ist übersichtlicher:

$$\begin{array}{cccc} & & & j\text{-te Spalte}\\ & & & \begin{pmatrix} \cdots & \begin{pmatrix} b_{1j}\\ \vdots\\ b_{nj}\end{pmatrix} & \cdots \end{pmatrix}\\ & & & \downarrow\\ i\text{-te Zeile} & \begin{pmatrix} \vdots\\ (a_{i1}\ldots a_{in})\\ \vdots \end{pmatrix} & \to & \begin{pmatrix} & \vdots & \\ \cdots & c_{ij} & \cdots\\ & \vdots & \end{pmatrix} \end{array}$$

Zum Beispiel,

$$\begin{array}{ccc} & & \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{pmatrix} \\ & & \downarrow \\ \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \end{pmatrix} & \to & \begin{pmatrix} \square & c_{12} & \square \\ \square & \square & \square \\ \square & \square & c_{33} \\ \square & \square & \square \end{pmatrix} \end{array}.$$

Beispiel 3.3.14. Seien

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

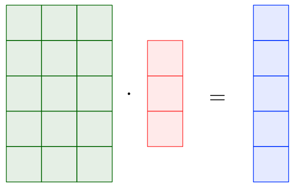
Dann berechnen wir Schritt für Schritt jeden Eintrag:

$$\begin{aligned} A \cdot B = \begin{pmatrix} 3 & 2 & 1 \\ 1 & 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 4 & 0 \end{pmatrix} &= \begin{pmatrix} 3 \cdot 1 + 2 \cdot 0 + 1 \cdot 4 & * \\ * & * \end{pmatrix} = \begin{pmatrix} 7 & * \\ * & * \end{pmatrix} \\ \begin{pmatrix} 3 & 2 & 1 \\ 1 & 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 4 & 0 \end{pmatrix} &= \begin{pmatrix} 7 & 3 \cdot 2 + 2 \cdot 1 + 1 \cdot 0 \\ * & * \end{pmatrix} = \begin{pmatrix} 7 & 8 \\ * & * \end{pmatrix} \\ \begin{pmatrix} 3 & 2 & 1 \\ 1 & 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 4 & 0 \end{pmatrix} &= \begin{pmatrix} 7 & 8 \\ 1 \cdot 1 + 0 \cdot 1 + 2 \cdot 4 & * \end{pmatrix} = \begin{pmatrix} 7 & 8 \\ 9 & * \end{pmatrix} \\ \begin{pmatrix} 3 & 2 & 1 \\ 1 & 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 4 & 0 \end{pmatrix} &= \begin{pmatrix} 7 & 8 \\ 9 & 1 \cdot 2 + 0 \cdot 1 + 2 \cdot 0 \end{pmatrix} = \begin{pmatrix} 7 & 8 \\ 9 & 2 \end{pmatrix}. \end{aligned}$$

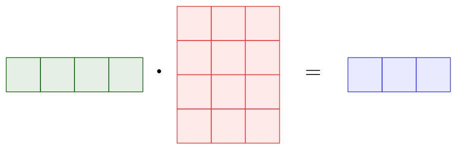
Übung 3.3.15. In diesem Fall ist auch BA definiert. Berechnen Sie diese Matrix!

Beispiel 3.3.16 (Spezialfälle und Tricks)**.** Wir betrachten einige Beispiele:

(1) Seien $A \in M_{m\times n}(K)$ und $v \in M_{n\times 1}(K)$. Dann ist $AB \in M_{m\times 1}(K)$. Wenn wir $M_{n\times 1}(K)$ mit $K^n_{\text{Spal}} \cong K^n$ identifizieren, entspricht v einem Spaltenvektor und $Av \in M_{m\times 1}(K) \cong K^m_{\text{Spal}}$. Das Produkt $Av \in K^m_{\text{Spal}}$ ist genau das Produkt, das wir in (2.3) definiert haben.



(2) Ähnlich gilt für $A \in M_{m\times n}(K)$ und $w \in M_{1\times m}(K)$, dass $wA \in M_{1\times n}(K)$. Wenn wir $M_{1\times m}(K)$ mit $K^m_{\text{Zeil}} \cong K^m$ identifizieren, entspricht w einem Zeilenvektor und das Produkt $wA \in K^n_{\text{Zeil}}$ ebenfalls einem Zeilenvektor. Dies definiert ein Produkt zwischen einem Zeilenvektor und einer Matrix.



(3) Wenn man die Spezialfälle (1) und (2) gut beherrscht, dann kann man diese auch für ein allgemeines Produkt AB verwenden: Für

$$A = \begin{pmatrix} — & v_1 & — \\ & \vdots & \\ — & v_m & — \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} | & & | \\ w_1 & \cdots & w_n \\ | & & | \end{pmatrix}$$

gilt

$$AB = A \cdot \begin{pmatrix} | & & | \\ w_1 & \cdots & w_n \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ Aw_1 & \cdots & Aw_n \\ | & & | \end{pmatrix}$$

beziehungsweise

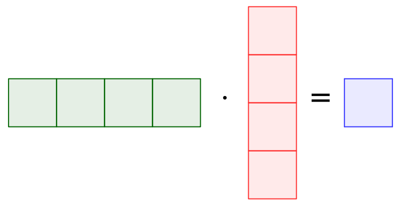
$$AB = \begin{pmatrix} — & v_1 & — \\ & \vdots & \\ — & v_m & — \end{pmatrix} \cdot B = \begin{pmatrix} — & v_1B & — \\ & \vdots & \\ — & v_mB & — \end{pmatrix}.$$

(4) Seien

$$v = (a_1, \ldots, a_n) \in M_{1\times n}(K) \quad \text{und} \quad w = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \in M_{n\times 1}(K).$$

Dann ist $vw \in M_{1\times 1}(K)$. Wenn wir $M_{1\times 1}(K)$ mit K identifizieren, dann heisst dieses Produkt das *Skalarprodukt* von v und w. Dieses Produkt berechnet sich als

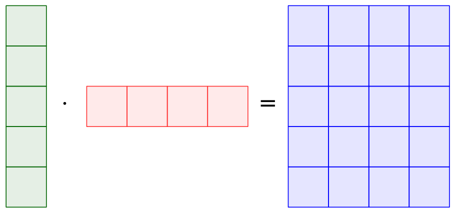
$$vw = (a_1, \ldots, a_n)\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = a_1 b_1 + \ldots + a_n b_n.$$



Viel weniger wichtig ist das Produkt $wv \in M_{n\times n}(K)$. Dies berechnet sich als

$$wv = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}(a_1, \ldots, a_n) = \begin{pmatrix} b_1 a_1 & \cdots & b_1 a_n \\ \vdots & & \vdots \\ b_n a_1 & \cdots & b_n a_n \end{pmatrix} = (a_j b_i)_{1\leq i,j\leq n}.$$

Beachten Sie, dass wv auch für $w \in M_{m\times 1}(K)$ und $v \in M_{1\times n}(K)$ definiert ist. In diesem Fall ist $wv \in M_{m\times n}(K)$.



Hier ist eine wichtige Matrix und Notation:

Definition 3.3.17. Wir definieren das *Kronecker-Delta* als

$$\delta_{ij} = \begin{cases} 1, & \text{falls } i = j \\ 0, & \text{falls } i \neq j \end{cases}.$$

Hier gehören i, j normalerweise zu irgendeiner Indexmenge.

Wir definieren die *Einheitsmatrix* I_n für $n \in \mathbb{N}$ als

$$I_n = (\delta_{ij})_{\substack{1\leq i\leq n \\ 1\leq j\leq n}} \in M_{n\times n}(K).$$

Zum Beispiel ist

$$I_1 = (1),\ I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ usw.}$$

Proposition 3.3.18. *Es gilt:*

(1) Für $A \in M_{m\times n}(K)$, $B \in M_{n\times p}(K)$ und $C \in M_{p\times q}(K)$ gilt

$$(AB)C = A(BC).$$

Das heisst, dass die Matrixmultiplikation assoziativ ist.

(2) Für $A, B \in M_{m\times n}(K)$ und $C \in M_{n\times p}(K), C' \in M_{q\times m}(K)$ gilt

$$(A + B)C = AC + BC$$
$$C'(A + B) = C'A + C'B.$$

(3) Für alle $A \in M_{m\times n}(K)$ gilt

$$I_m A = A$$
$$AI_n = A.$$

(4) Für alle $\alpha \in K$, $A \in M_{m\times n}(K)$ und $B \in M_{n\times p}(K)$ gilt

$$\alpha(AB) = (\alpha A)B = A(\alpha B).$$

Beweis. All diese Eigenschaften folgen aus einer entsprechenden Berechnung. Dies ist insbesondere richtig für (1), aber die Berechnung ist wirklich nicht angenehm. Nachdem wir eine Verbindung zwischen Matrixmultiplikation und Verkettung herstellen (Lemma 3.3.29), können wir aus der Tatsache, dass die Verkettung von Funktionen assoziativ ist, herleiten. □

Bemerkung 3.3.19. Die Menge $M_{n\times n}(K)$ für $n \in \mathbb{N}$ ist laut (1), (2) und (3) in Proposition 3.3.18 ein Ring mit I_n als Eins. Bemerken Sie, dass dieser Ring für $n \geq 2$ nicht-kommutativ ist, da normalerweise

$$AB \neq BA$$

gilt. Zum Beispiel gilt

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

Unter anderem motiviert Bemerkung 3.3.19 die folgende Definition:

Definition 3.3.20. Sei $n \in \mathbb{N}$. Eine Matrix $A \in M_{n \times n}(K)$ heisst *invertierbar*, falls eine Matrix $B \in M_{n \times n}(K)$ existiert, so dass

$$AB = I_n \quad \text{und} \quad BA = I_n$$

gilt. Die Matrix B heisst die *Inverse* von A und wird mit A^{-1} notiert. Die Menge aller invertierbaren Matrizen in $M_{n \times n}(K)$ bezeichnen wir mit

$$\mathrm{GL}_n(K) \quad \text{oder} \quad \mathrm{GL}(n, K)$$

und nennen sie die *allgemeine lineare Gruppe* über K (von Grad n).

Proposition 3.3.21. *Die allgemeine lineare Gruppe ist eine Gruppe bezüglich Matrixmultiplikation.*

Beweis. Wir müssen zeigen, dass die Matrixmultiplikation eine Verknüpfung auf $\mathrm{GL}_n(K)$ induziert. Das heisst, dass $AB \in \mathrm{GL}_n(K)$ ist für $A, B \in \mathrm{GL}_n(K)$.

Seien also $A, B \in \mathrm{GL}_n(K)$ und $C, D \in M_{n \times n}(K)$ mit

$$\begin{aligned} AC &= CA = I_n \\ BD &= DB = I_n. \end{aligned}$$

Dann gilt

$$\begin{aligned} (AB)(DC) &= A(BD)C = AI_nC = AC = I_n \\ (DC)(AB) &= D(CA)B = DI_nB = DB = I_n. \end{aligned}$$

Also folgt, dass $AB \in \mathrm{GL}_n(K)$.

Die Einheitsmatrix $I_n \in \mathrm{GL}_n(K)$, da

$$I_n I_n = I_n$$

und damit ist I_n invertierbar. Des Weiteren ist die Einheitsmatrix das neutrale Element. Laut Proposition 3.3.18 (1) ist die Matrixmultiplikation assoziativ.

Wenn $A \in \mathrm{GL}_n(K)$ ist, dann ist auch $A^{-1} \in \mathrm{GL}_n(K)$, da A eine Inverse von A^{-1} ist. Daher hat jedes Element $A \in \mathrm{GL}_n(K)$ eine Inverse in $\mathrm{GL}_n(K)$. □

Korollar 3.3.22. *(1) Die Inverse von A ist eindeutig definiert.*

(2) Es gilt $(A^{-1})^{-1} = A$*.*

(3) Es gilt $(AB)^{-1} = B^{-1}A^{-1}$*.*

Beweis. Diese Tatsachen haben wir für alle Gruppen bewiesen. □

Beispiel 3.3.23 (Inverse und lineare Gleichungssysteme)**.** Bevor wir zu linearen Abbildungen zurückkehren, erklären wir kurz, wieso die Inverse interessant ist.

Seien $A \in M_{n\times n}$ und $x, b \in K^n$, so dass

$$Ax = b. \tag{3.15}$$

Wir nehmen an, dass A invertierbar ist, und dass wir die Inverse A^{-1} kennen. Dann können wir das lineare Gleichungssystem (3.15) in einer Sekunde durch Matrixmultiplikation lösen: Wir multiplizieren (3.15) mit A^{-1} von der linken Seite und erhalten:

$$x = I_n x = (A^{-1}A)x = A^{-1}(Ax) = A^{-1}b.$$

Das heisst, die Lösung von (3.15) ist in diesem Fall $A^{-1}b$!

Korollar 3.3.24. *Sei* $Ax = b$ *ein lineares Gleichungssystem mit* $A \in M_{n\times n}(K)$*. Falls* $A \in \mathrm{GL}_n(K)$ *ist, dann hat* $Ax = b$ *für jedes* $b \in K^n$ *eine eindeutige Lösung* x *(nämlich* $x = A^{-1}b$*).*

Bemerkung 3.3.25*.* Später werden wir beweisen, dass auch die Umkehrung von Korollar 3.3.24 gilt.

Für später müssen wir auch noch verstehen wie Transposition und Matrixmultiplikation in Verbindung stehen. Dies ist eine wirklich gute Übung um Matrixmultiplikation zu üben, deshalb lassen wir sie als Übung für die Leser:

Übung 3.3.26. (1) Sei $A \in M_{m\times n}(K)$ und $B \in M_{n\times p}(K)$. Dann gilt

$$(AB)^T = B^T A^T.$$

(Beachten Sie, dass $A^T B^T$ im Allgemeinen gar nicht definiert ist!)

(2) Sei $A \in \mathrm{GL}_n(K)$. Zeigen Sie, dass

$$A^T \in \mathrm{GL}_n(K)$$

ist und dass

$$(A^T)^{-1} = (A^{-1})^T.$$

Bevor wir wieder zu Darstellungen von linearen Abbildungen zurückkehren, definieren wir einige Familien von Matrizen, welche wir später brauchen werden.

Definition 3.3.27. Eine Matrix $A = (a_{ij}) \in M_{n\times n}(K)$ heisst

- *obere Dreiecksmatrix*, falls für alle $1 \leq i, j \leq n$ gilt: $i > j \implies a_{ij} = 0$.
- *untere Dreiecksmatrix*, falls für alle $1 \leq i, j \leq n$ gilt: $i < j \implies a_{ij} = 0$.
- *diagonal*, falls für alle $1 \leq i, j \leq n$ gilt: $i \neq j \implies a_{ij} = 0$.

Das folgende Lemma folgt direkt aus der Definition der Matrixmultiplikation:

Lemma 3.3.28. *Seien A, B diagonale (bzw. obere Dreiecks-, bzw untere Dreiecks-) Matrizen. Dann ist AB auch wieder eine diagonale (bzw. obere Dreiecks-, bzw. untere Dreiecks-) Matrix.*

3.3.3 Zurück zu Darstellungen linearer Abbildungen

Die Buchstaben V, W und U bezeichnen in diesem Abschnitt immer endlich-dimensionale Vektorräume über K. Das folgende grundlegende Lemma beantwortet Frage (3) in (3.7) und zeigt, dass die Definition der Matrixmultiplikation nicht vom Himmel fällt, sondern einfach von der Darstellung der Verkettung zweier linearer Abbildungen gekommen ist.

Lemma 3.3.29. *Seien V, W und U drei endlich-dimensionale Vektorräume über K mit jeweiligen Basen $\mathcal{A} = (v_1, \ldots, v_m), \mathcal{B} = (w_1, \ldots, w_n)$ und $\mathcal{C} = (u_1, \ldots, u_p)$. Wir betrachten $T \in \mathrm{Hom}(V, W)$ und $S \in \mathrm{Hom}(W, U)$. Dann gilt*

$$[S \circ T]^{\mathcal{A}}_{\mathcal{C}} = [S]^{\mathcal{B}}_{\mathcal{C}} \cdot [T]^{\mathcal{A}}_{\mathcal{B}},$$

wobei $\cdot$ die Matrixmultiplikation bezeichnet.

Beweis. Der folgende Beweis ist eine sehr gute Indizes-Übung. Seien

$$[T]^{\mathcal{A}}_{\mathcal{B}} = (a_{ij})_{ij}, \quad [S]^{\mathcal{B}}_{\mathcal{C}} = (b_{ij})_{ij} \quad \text{und} \quad [S \circ T]^{\mathcal{A}}_{\mathcal{C}} = (c_{ij})_{ij}$$

(Überlegen Sie sich, welche Werte die Indizes i und j für die jeweiligen Matrizen annehmen.) Diese Matrizen sind durch

$$T(v_j) = a_{1j}w_1 + \ldots a_{nj}w_n = \sum_{k=1}^{n} a_{kj}w_k \quad \text{für } j = 1, \ldots, m, \tag{3.16}$$

$$S(w_j) = b_{1j}u_1 + \ldots + b_{pj}u_p = \sum_{i=1}^{p} b_{ij}u_i \quad \text{für } j = 1, \ldots, n, \tag{3.17}$$

$$(S \circ T)(v_j) = c_{1j}u_1 + \ldots + c_{pj}u_p = \sum_{i=1}^{p} c_{ij}u_i \quad \text{für } j = 1, \ldots, m \tag{3.18}$$

definiert.[5] Aber $(S \circ T)(v_j)$ könnten wir auch durch (3.16) und (3.17) noch anders ausdrücken. Für $j = 1, \ldots, m$ gilt:

$$
\begin{aligned}
(S \circ T)(v_j) \overset{(3.16)}{=} S\left(\sum_{k=1}^{n} a_{kj} w_k\right) \overset{\text{Linearität}}{=} \sum_{k=1}^{n} a_{kj} S(w_k) \overset{(3.17)}{=} \sum_{k=1}^{n} a_{kj} \left(\sum_{i=1}^{p} b_{ik} u_i\right) \\
= \sum_{k=1}^{n} \sum_{i=1}^{p} a_{kj} b_{ik} u_i \\
= \sum_{i=1}^{p} \sum_{k=1}^{n} b_{ik} a_{kj} u_i.
\end{aligned}
$$

Daher folgt (Wieso? Vgl. (3.18)), dass

$$c_{ij} = \sum_{k=1}^{n} b_{ik} a_{kj}$$

beziehungsweise $C = BA$, was wir zeigen wollten. □

Dies beantwortet Frage (3) in (3.7), aber auch eine Antwort auf Frage (2) folgt jetzt mittels einem kleinen Trick mit der Identitätsabbildung.

Korollar 3.3.30. *Sei $T \in \operatorname{Hom}(V, W)$, $\mathcal{B}, \mathcal{B}'$ zwei Basen von V und $\mathcal{C}, \mathcal{C}'$ zwei Basen von W. Dann gilt*

$$[T]_{\mathcal{C}'}^{\mathcal{B}'} = [\mathrm{Id}_W]_{\mathcal{C}'}^{\mathcal{C}} \, [T]_{\mathcal{C}}^{\mathcal{B}} \, [\mathrm{Id}_V]_{\mathcal{B}}^{\mathcal{B}'}. \tag{3.19}$$

Beweis. Aus zweifacher Verwendung von Lemma 3.3.29 folgt, dass die rechte Seite von (3.19)

$$[\mathrm{Id}_W \circ T \circ \mathrm{Id}_V]_{\mathcal{C}'}^{\mathcal{B}'} = [T]_{\mathcal{C}'}^{\mathcal{B}'}.$$

ist. □

Matrizen der Form $[\mathrm{Id}_V]_{\mathcal{B}'}^{\mathcal{B}}$ heissen *Basiswechselmatrizen*. Wir haben noch mehr Information bezüglich solcher Matrizen:

Korollar 3.3.31. *Seien $\mathcal{B}$ und $\mathcal{B}'$ zwei Basen von V und $n = \dim V$. Dann gilt, dass*

$$[\mathrm{Id}_V]_{\mathcal{B}'}^{\mathcal{B}} \in \mathrm{GL}_n(K).$$

Das heisst, dass $[\mathrm{Id}_V]_{\mathcal{B}'}^{\mathcal{B}}$ invertierbar ist. Ausserdem gilt

$$\left([\mathrm{Id}_V]_{\mathcal{B}'}^{\mathcal{B}}\right)^{-1} = [\mathrm{Id}_V]_{\mathcal{B}}^{\mathcal{B}'}.$$

[5] Wieso wir die erste Summe mit Indizes k summieren und die restlichen beiden mit i wird später klarer werden.

Beweis. Laut Lemma 3.3.29 gilt

$$\begin{aligned}[\mathrm{Id}_V]^{\mathcal{B}}_{\mathcal{B}'} \cdot [\mathrm{Id}_V]^{\mathcal{B}'}_{\mathcal{B}} &= [\mathrm{Id}_V \circ \mathrm{Id}_V]^{\mathcal{B}'}_{\mathcal{B}'} = [\mathrm{Id}_V]^{\mathcal{B}'}_{\mathcal{B}'} \\ [\mathrm{Id}_V]^{\mathcal{B}'}_{\mathcal{B}} \cdot [\mathrm{Id}_V]^{\mathcal{B}}_{\mathcal{B}'} &= [\mathrm{Id}_V \circ \mathrm{Id}_V]^{\mathcal{B}}_{\mathcal{B}} = [\mathrm{Id}_V]^{\mathcal{B}}_{\mathcal{B}}.\end{aligned}$$

Es folgt aber aus der Definition, dass

$$[\mathrm{Id}_V]^{\mathcal{C}}_{\mathcal{C}} = I_n$$

für jede Basis $\mathcal{C}$ von V (Wieso?). Das Korollar folgt. □

Unter Benutzung der vorherigen Korollare erhalten wir folgendes Resultat für Endomorphismen:

Korollar 3.3.32. *Sei V ein endlich-dimensionaler Vektorraum mit zwei Basen $\mathcal{B}$ und $\mathcal{B}'$ und $T \in \mathrm{End}(V)$. Dann gilt*

$$[T]^{\mathcal{B}'}_{\mathcal{B}'} = [\mathrm{Id}_V]^{\mathcal{B}}_{\mathcal{B}'} \, [T]^{\mathcal{B}}_{\mathcal{B}} \, [\mathrm{Id}_V]^{\mathcal{B}'}_{\mathcal{B}} = ([\mathrm{Id}_V]^{\mathcal{B}'}_{\mathcal{B}})^{-1} \, [T]^{\mathcal{B}}_{\mathcal{B}} \, [\mathrm{Id}_V]^{\mathcal{B}'}_{\mathcal{B}}. \tag{3.20}$$

Beweis. (Wieso? Benutzen Sie die Korollare 3.3.30 und 3.3.31.) □

Bemerkung. Für $T \in \mathrm{End}(V)$ schreibt man manchmal $[T]_{\mathcal{B}} := [T]^{\mathcal{B}}_{\mathcal{B}}$.

Wir können uns nun die Assoziativität der Matrixmultiplikation neu anschauen (vgl. Proposition 3.3.18 (i)).

Korollar 3.3.33. *Seien $A \in M_{m\times n}(K)$ und $B \in M_{n\times p}(K)$, dann gilt*

$$m_A \circ m_B = m_{AB}.$$

Beweis. Sei $\mathcal{E}_n$ die Standard-Basis von K^n. Dann gilt laut Beispiel 3.3.9 und Lemma 3.3.29, dass

$$[m_A \circ m_B]^{\mathcal{E}_p}_{\mathcal{E}_m} = [m_A]^{\mathcal{E}_n}_{\mathcal{E}_m} [m_B]^{\mathcal{E}_p}_{\mathcal{E}_n} = AB.$$

Aus Proposition 3.3.8 (und Beispiel 3.3.9 mit $n = 1$) folgt, dass für jedes $v \in K^p$

$$m_A \circ m_B(v) = [m_A \circ m_B(v)]_{\mathcal{E}_m} = [m_A \circ m_B]^{\mathcal{E}_p}_{\mathcal{E}_m} [v]_{\mathcal{E}_p} = AB[v]_{\mathcal{E}_p} = (AB)v$$

gilt. Dies scheint als eine Art Tautologie, weil Assoziativität, Lemma 3.3.29 und Proposition 3.3.8 in der Tat eng miteinander verbunden sind. Also gilt $m_A \circ m_B = m_{AB}$ als Abbildungen von K^p nach K^m. □

Korollar 3.3.34. *Matrixmultiplikation ist assoziativ.*

Beweis. Seien $A \in M_{m\times n}(K)$, $B \in M_{n\times p}(K)$, $C \in M_{p\times q}(K)$ und sei $\mathcal{E}_l$ die Standard-Basis von K^l. Da Verkettung von Funktionen assoziativ ist, gilt

$$\begin{aligned}(AB)C = [m_{AB}]^{\mathcal{E}_p}_{\mathcal{E}_m}[m_C]^{\mathcal{E}_q}_{\mathcal{E}_p} = [m_A \circ m_B]^{\mathcal{E}_p}_{\mathcal{E}_m}[m_C]^{\mathcal{E}_q}_{\mathcal{E}_p} &= [m_A \circ m_B \circ m_C]^{\mathcal{E}_q}_{\mathcal{E}_m} \\ &= [m_A]^{\mathcal{E}_n}_{\mathcal{E}_m}[m_B \circ m_C]^{\mathcal{E}_q}_{\mathcal{E}_n} \\ &= [m_A]^{\mathcal{E}_n}_{\mathcal{E}_m}[m_{BC}]^{\mathcal{E}_q}_{\mathcal{E}_n} = A(BC).\end{aligned}$$

Das Korollar folgt. □

Gleichungen (3.19) und (3.20) motivieren die folgende Definition:

Definition 3.3.35.

- Zwei Matrizen $A, B \in M_{m\times n}(K)$ heissen *äquivalent*, falls es $P \in \mathrm{GL}_m(K)$ und $Q \in \mathrm{GL}_n(K)$ gibt, so dass
$$PAQ = B.$$
- Zwei Matrizen $A, B \in M_{n\times n}(K)$ heissen *ähnlich*, falls es $P \in \mathrm{GL}_n(K)$ gibt, so dass
$$P^{-1}AP = B.$$

Übung 3.3.36. Zeigen Sie, folgende Aussagen:

- Äquivalenz ist eine Äquivalenzrelation auf $M_{m\times n}(K)$.
- Ähnlichkeit ist eine Äquivalenzrelation auf $M_{n\times n}(K)$.
- A ist äquivalent zu B genau dann, wenn es Vektorräume V, W, $T \in \mathrm{Hom}(V, W)$ und Basen $\mathcal{B}, \mathcal{B}' \subseteq V$, $\mathcal{C}, \mathcal{C}' \subseteq W$ gibt, so dass
$$A = [T]^{\mathcal{B}}_{\mathcal{C}} \quad \text{und} \quad B = [T]^{\mathcal{B}'}_{\mathcal{C}'}.$$
- A ist ähnlich zu B genau dann, wenn es einen Vektorraum V, $T \in \mathrm{End}(V)$ und Basen $\mathcal{B}, \mathcal{B}' \subseteq V$ gibt, so dass
$$A = [T]_{\mathcal{B}} \quad \text{und} \quad B = [T]_{\mathcal{B}'}.$$

Um den Namen „Ähnlichkeit“ zu rechtfertigen werden wir sehen, dass ähnliche Matrizen tatsächlich vieles gemeinsam haben. Einen einfachen Repräsentanten für jede Äquivalenzklasse bezüglich Ähnlichkeit auf $M_{n\times n}(K)$ zu finden ist jedoch schwierig und wird uns noch in der Linearen Algebra II beschäftigen. Wir können hingegen problemlos für die Äquivalenzrelation „Äquivalenz“ auf $M_{m\times n}(K)$ einfache Repräsentanten für die Äquivalenzklassen finden.

Sei $D_r = (d_{ij})_{ij} \in M_{m\times n}(K)$ die Matrix mit

$$\begin{cases} d_{ij} = 0, \text{ falls } i \neq j \\ d_{ii} = 1, \text{ falls } 1 \leq i \leq r \\ d_{ii} = 0, \text{ falls } r < i \end{cases}.$$

Das heisst, dass

$$D_r = \begin{pmatrix} I_r & 0_{r,n-r} \\ 0_{m-r,r} & 0_{m-r,n-r} \end{pmatrix} \in M_{m\times n}(K),$$

wobei $0_{l,k} \in M_{l\times k}(K)$ die Nullmatrix ist.

Proposition 3.3.37. *Sei $T \in \mathrm{Hom}(V,W)$ mit*[6] $\mathrm{Rang}(T) = r$*. Dann existieren Basen $\mathcal{B}$ von V und $\mathcal{C}$ von W, so dass*

$$[T]^{\mathcal{B}}_{\mathcal{C}} = D_r.$$

Beweis. Im Beweis des Rangsatzes 3.2.9 haben wir eine Basis $\mathcal{B} = (w_1, \ldots, w_r, u_1, \ldots, u_k)$ von V gefunden, so dass

$$Tu_1 = \ldots = Tu_k = 0$$

und $\mathcal{C}' := (Tw_1, \ldots, Tw_r)$ eine Basis von $\mathrm{Im}(T)$ ist. Sei $\mathcal{C}$ eine Ergänzung von $\mathcal{C}'$ zu einer Basis von W. Bezüglich $\mathcal{B}$ und $\mathcal{C}$ gilt

$$[T]^{\mathcal{B}}_{\mathcal{C}} = D_r.$$

(Wieso?) □

Lemma 3.3.38. *Sei $A \in M_{m\times n}(K)$. Dann gilt*

$$\mathrm{SR}(A) = \mathrm{Im}(m_A).$$

Beweis. Dies folgt aus Lemma 3.1.18 und Lemma 3.2.8. (Wieso?) □

Proposition 3.3.39. *Sei $A \in M_{m\times n}(K)$. Dann existieren $P \in \mathrm{GL}_m(K)$ und $Q \in \mathrm{GL}_n(K)$ mit*

$$PAQ = D_r = \begin{pmatrix} I_r & 0_{r,n-r} \\ 0_{m-r,r} & 0_{m-r,n-r} \end{pmatrix},$$

wobei r = Spaltenrang(A).

Beweis von Proposition 3.3.39. Aus Beispiel 3.3.9 wissen wir, dass

$$[m_A]^{\mathcal{E}_n}_{\mathcal{E}_m} = A,$$

[6] Erinnerung: Für $T \in \mathrm{Hom}(V,W)$ ist $\mathrm{Rang}(T) = \dim \mathrm{Im}(T)$.

wobei $\mathcal{E}_n, \mathcal{E}_m$ die entsprechenden Standard-Basen sind von K^n und K^m sind. Laut Proposition 3.3.37 existieren Basen $\mathcal{B}$ von K^n und $\mathcal{C}$ von K^m, mit

$$[\mathrm{Id}_{K^m}]_{\mathcal{C}}^{\mathcal{E}_m} [m_A]_{\mathcal{E}_m}^{\mathcal{E}_n} [\mathrm{Id}_{K^n}]_{\mathcal{E}_n}^{\mathcal{B}} = D_r.$$

Wir setzen

$$P := [\mathrm{Id}_{K^m}]_{\mathcal{C}}^{\mathcal{E}_m} \quad \text{und} \quad Q := [\mathrm{Id}_{K^n}]_{\mathcal{E}_n}^{\mathcal{B}},$$

was zeigt, dass

$$PAQ = D_r.$$

□

Korollar 3.3.40. *Es gibt* $\min(m, n) + 1$ *Äquivalenzklassen bezüglich der Äquivalenzrelation Äquivalenz auf* $M_{m\times n}(K)$ *mit* D_r *für* $r = 0, \ldots, \min(n, m)$ *als Repräsentanten.*

Einige Autoren (Vgl. Fischer [7, S. 161]) nennen diese Repräsentanten D_r Normalformen. Dies ist nicht so geläufig und man sollte dies nicht mit der sogenannten Jordan-Normalform verwechseln (diese werden wir in der Linearen Algebra II kennenlernen und ist verbunden mit Repräsentanten von Äquivalenzklassen bezüglich Ähnlichkeit).

Bemerkung 3.3.41. In Kürze werden wir beweisen, dass Spaltenrang(A) = Zeilenrang(A) (vgl. Korollar 2.3.28). Daher könnte man r = Spaltenrang(A) = Zeilenrang(A) in Proposition 3.3.39 schreiben. Damit könnte man Korollar 3.3.40 auch so ausdrücken: Zwei Matrizen sind genau dann äquivalent, wenn sie denselben Rang haben.

Korollar 3.3.42. *Sei* $T : V \to W$ *ein Isomorphismus und* $\mathcal{B} \subseteq V$, $\mathcal{C} \subseteq W$ *zwei Basen. Dann gilt*

$$[T^{-1}]_{\mathcal{B}}^{\mathcal{C}} = \left([T]_{\mathcal{C}}^{\mathcal{B}}\right)^{-1}.$$

Insbesondere ist $[T]_{\mathcal{C}}^{\mathcal{B}}$ *invertierbar.*

Beweis. Sei $n = \dim V = \dim W$. Laut Lemma 3.3.29 gilt

$$[T^{-1}]_{\mathcal{B}}^{\mathcal{C}} [T]_{\mathcal{C}}^{\mathcal{B}} = [T^{-1} \circ T]_{\mathcal{B}}^{\mathcal{B}} = [\mathrm{Id}_V]_{\mathcal{B}}^{\mathcal{B}} = I_n$$
$$[T]_{\mathcal{C}}^{\mathcal{B}} [T^{-1}]_{\mathcal{B}}^{\mathcal{C}} = [T \circ T^{-1}]_{\mathcal{C}}^{\mathcal{C}} = [\mathrm{Id}_W]_{\mathcal{C}}^{\mathcal{C}} = I_n.$$

Das Korollar folgt. □

3.4 Spaltenrang ist gleich Zeilenrang

Satz 3.4.1. *Sei* $A \in M_{m\times n}(K)$. *Dann gilt*

$$\textit{Spaltenrang}(A) = \textit{Zeilenrang}(A).$$

Dank diesem Satz können wir definieren, dass

$$\mathrm{Rang}(A) = \mathrm{Spaltenrang}(A) = \mathrm{Zeilenrang}(A).$$

Laut Abschnitt 2.3.3 stimmt diese Definition mit der Definition als die Anzahl der Pivots in einer Stufenform Matrix, die zeilenäquivalent ist zu A überein.

Für den Beweis von Theorem 3.4.1 brauchen wir einige Lemmas. Erinnern Sie sich an Definition 3.2.15: $\mathrm{Rang}(T) = \dim \mathrm{Im}(T)$.

Lemma 3.4.2. *Sei $T \in \mathrm{Hom}(V, W)$ und seien $S : W \to U$ und $L : P \to V$ zwei Isomorphismen. Dann gilt*

- $\mathrm{Rang}(S \circ T) = \mathrm{Rang}(T)$.
- $\mathrm{Rang}(T \circ L) = \mathrm{Rang}(T)$.

Insbesondere gilt auch, dass $\mathrm{Rang}(S \circ T \circ L) = \mathrm{Rang}(T)$ *ist.*

Beweis. Dies ist ein Spezialfall von Übung 3.2.16, die Sie in der Serie haben. □

Lemma 3.4.3. *Sei $A \in M_{n\times n}(K)$. Dann gilt, dass $A \in \mathrm{GL}_n(K)$ genau dann, wenn $m_A : K^n \to K^n$ ein Isomorphismus ist.*

Beweis. „⇒“: Die lineare Abbildung $m_{A^{-1}} : K^n \to K^n$ ist die Inverse von m_A.

„⇐“: Aus Korollar 3.3.42 folgt, dass

$$A = [m_A]_{\mathcal{E}_n}^{\mathcal{E}_n} \tag{3.21}$$

invertierbar ist, wobei $\mathcal{E}_n$ die Standard-Basis ist. □

Lemma 3.4.4. *Seien $A, B \in M_{m\times n}(K)$ mit*

$$B = PAQ, \tag{3.22}$$

wobei $P \in \mathrm{GL}_m(K)$ und $Q \in \mathrm{GL}_n(K)$. Dann gilt

(1) Spaltenrang(A) = Spaltenrang(B).

(2) Zeilenrang(A) = Zeilenrang(B).

Beweis. Betrachten wir alle Matrizen im Lemma als lineare Abbildungen, ist (3.22) äquivalent zu

$$m_B = m_P \circ m_A \circ m_Q$$

und laut Lemma 3.4.2 gilt

$$\mathrm{Rang}(m_A) = \mathrm{Rang}(m_B),$$

da m_Q und m_P laut Lemma 3.4.3 Isomorphismen sind. Wir haben aber gesehen (Lemma 3.3.38), dass

$$\mathrm{Rang}(m_C) = \mathrm{Spaltenrang}(C)$$

für jede Matrix C ist. Daher gilt

$$\mathrm{Spaltenrang}(A) = \mathrm{Spaltenrang}(B).$$

Dies beweist (1). Für (2) transponieren wir (3.22) mit Hilfe von Übung 3.3.26:

$$B^T = Q^T A^T P^T.$$

Aus Übung 3.3.26 wissen wir auch, dass $Q^T \in \mathrm{GL}_n(K)$ und $P^T \in \mathrm{GL}_m(K)$ und daher folgt aus (1), dass

$$\mathrm{Spaltenrang}(A^T) = \mathrm{Spaltenrang}(B^T),$$

was äquivalent zu

$$\mathrm{Zeilenrang}(A) = \mathrm{Zeilenrang}(B)$$

ist. □

Jetzt folgt Satz 3.4.1 aus Proposition 3.3.39:

Beweis von Satz 3.4.1. Wegen Proposition 3.3.39 existieren $P \in \mathrm{GL}_m(K)$ und $Q \in \mathrm{GL}_n(K)$, so dass

$$PAQ = D_r = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} =: B.$$

Für B gilt $\mathrm{Spaltenrang}(B) = \mathrm{Zeilenrang}(B) = r$. Laut Lemma 3.4.4 gelten

$$\mathrm{Spaltenrang}(A) = \mathrm{Spaltenrang}(B) \quad \text{und} \quad \mathrm{Zeilenrang}(A) = \mathrm{Zeilenrang}(B).$$

Daher folgt, dass $\mathrm{Spaltenrang}(A) = \mathrm{Zeilenrang}(A)$. □

Bemerkung 3.4.5. Ich habe diesen Beweis aus dem Buch von Fischer [7, Kap. 2.6.6] gelernt. Beachten Sie, dass wir in 2.3.32 (noch) einen anderen Beweis skizziert haben. Falls Sie dies wünschen, kann ich in der Nachbesprechungszeit einen Beweis mittels Gauss und dem Rangsatz ebenfalls skizzieren.

3.5 Hom(V, W) als Vektorraum

Wir werden jetzt sehen, dass $\mathrm{Hom}(V, W)$ ein Vektorraum ist, und dass wir dies eigentlich schon wissen (zumindest im Fall, wo V und W endlich-dimensional sind).

Seien V und W zwei Vektorräume über einem Körper K. Jetzt betrachten wir Hom(V, W) als eine Menge. Betrachten Sie die folgenden Operationen auf Hom(V, W):

Seien $T, S \in \mathrm{Hom}(V, W)$. Wir definieren Addition auf Hom(V, W):

$$T + S : V \to W$$

durch

$$(T + S)(v) := T(v) +_W S(v).$$

Sei $\alpha \in K$ und $T \in \mathrm{Hom}(V, W)$. Wir definieren Skalarmultiplikation auf Hom(V, W):

$$\alpha T : V \to W$$

durch

$$(\alpha T)(v) := \alpha \cdot_W T(v).$$

Lemma 3.5.1. *Für $T, S \in \mathrm{Hom}(V, W)$ und $\alpha \in K$ gilt, dass $T + S \in \mathrm{Hom}(V, W)$ und $\alpha T \in \mathrm{Hom}(V, W)$.*

Beweis. Seien $a, b \in K$ und $v_1, v_2 \in V$. Dann ist

$$\begin{aligned}(T + S)(av_1 + bv_2) &= T(av_1 + bv_2) + S(av_1 + bv_2)\\ &= (aTv_1 + bTv_2) + (aSv_1 + bSv_2)\\ &= a(Tv_1 + Sv_1) + b(Tv_2 + Sv_2)\\ &= a((T + S)(v_1)) + b((T + S)(v_2)).\end{aligned}$$

Wir überlassen es den Lesern die analoge Überprüfung für αT zu machen. □

Proposition 3.5.2. Hom(V, W) *mit der obigen Addition und Skalarmultiplikation ist ein Vektorraum.*

Beweis. Wir überlassen diesen Beweis den Lesern und bemerken nur, dass der Nullvektor in Hom(V, W) die lineare Abbildung

$$\begin{aligned}0 : V &\to W,\\ v &\mapsto 0_W\end{aligned}$$

ist. □

Für endlich-dimensionale Vektorräume V und W „kennen“ wir eigentlich diesen Vektorraum schon:

Satz 3.5.3. *Seien V und W zwei Vektorräume mit* $\dim V = n$ *und* $\dim W = m$ *mit jeweiligen Basen* $\mathcal{B} = (v_1, \ldots, v_n) \subseteq V$ *und* $\mathcal{C} = (w_1, \ldots, w_m) \subseteq W$. *Die Abbildung*

$$\begin{aligned} \psi_{\mathcal{C}}^{\mathcal{B}} : \mathrm{Hom}(V, W) &\to M_{m\times n}(K) \\ T &\mapsto [T]_{\mathcal{C}}^{\mathcal{B}} \end{aligned}$$

ist ein Isomorphismus. Das heisst, sie ist linear und bijektiv.

Beweis. Wir zeigen zuerst, dass $\psi_{\mathcal{C}}^{\mathcal{B}}$ linear ist: Per Definition gilt

$$[T+S]_{\mathcal{C}}^{\mathcal{B}} = \begin{bmatrix} | & & | \\ [T+S(v_1)]_{\mathcal{C}} & \cdots & [T+S(v_n)]_{\mathcal{C}} \\ | & & | \end{bmatrix},$$

wobei $\mathcal{B} = (v_1, \ldots, v_n)$. Es gilt aber für jedes $i = 1, \ldots, n$, dass

$$[T+S(v_i)]_{\mathcal{C}} = [Tv_i + Sv_i]_{\mathcal{C}} = [Tv_i]_{\mathcal{C}} + [Sv_i]_{\mathcal{C}},$$

da $v \mapsto [v]_{\mathcal{C}}$ linear ist. Daher gilt

$$\begin{aligned} &\begin{bmatrix} | & & | \\ [T+S(v_1)]_{\mathcal{C}} & \cdots & [T+S(v_n)]_{\mathcal{C}} \\ | & & | \end{bmatrix} \\ &= \begin{bmatrix} | & & | \\ [Tv_1]_{\mathcal{C}} & \cdots & [Tv_n]_{\mathcal{C}} \\ | & & | \end{bmatrix} + \begin{bmatrix} | & & | \\ [Sv_1]_{\mathcal{C}} & \cdots & [Sv_n]_{\mathcal{C}} \\ | & & | \end{bmatrix} = [T]_{\mathcal{C}}^{\mathcal{B}} + [S]_{\mathcal{C}}^{\mathcal{B}}. \end{aligned}$$

Die Überprüfung der Homogenität ist sehr ähnlich und daher überlassen wir dies den Lesern.

Des Weiteren ist $\psi_{\mathcal{C}}^{\mathcal{B}}$ injektiv: Falls $[T]_{\mathcal{C}}^{\mathcal{B}} = 0$ ist, dann ist $T(v_i) = 0_W$ für alle $i = 1, \ldots n$ und daher ist $T = 0$.

Wir müssen noch zeigen, dass $\psi_{\mathcal{C}}^{\mathcal{B}}$ surjektiv ist: Sei $A = (a_{ij}) \in M_{m\times n}(K)$. Mittels Satz 3.1.15 definieren wir T durch

$$T(v_j) = \sum_{i=1}^{m} a_{ij} w_i.$$

Dann gilt $[T]_{\mathcal{C}}^{\mathcal{B}} = A$. (Wieso? Vgl. (3.10).) □

Korollar 3.5.4. *Falls V und W endlich-dimensionale Vektorräume sind, dann gilt*

$$\dim \mathrm{Hom}(V, W) = \dim V \cdot \dim W.$$

Bemerkung 3.5.5 (Nebenbemerkung). Wie die Notation schon andeutet, hängt der Isomorphismus $\psi_{\mathcal{C}}^{\mathcal{B}}$ von der Wahl der Basen $\mathcal{B}$ und $\mathcal{C}$ ab. In diesem Sinne ist $\psi_{\mathcal{C}}^{\mathcal{B}}$ „nicht kanonisch".

Korollar 3.5.6. *Der Vektorraum* $\mathrm{End}(V) = \mathrm{Hom}(V, V)$ *ist ein Ring mit Verkettung als Multiplikation und* Id_V *als Einselement. Falls* V *endlich-dimensional ist (mit Dimension* n*) und* $\mathcal{B} \subseteq V$ *eine Basis ist, dann ist*

$$\begin{aligned} \psi_{\mathcal{B}}^{\mathcal{B}} : \mathrm{End}(V) &\to M_{n\times n}(K) \\ T &\mapsto [T]_{\mathcal{B}} := [T]_{\mathcal{B}}^{\mathcal{B}}. \end{aligned}$$

ein Ring-Isomorphismus.

Beweis. Man überprüft die Ring-Axiome direkt, zum Beispiel:

Seien $v \in V$ und $T_1, T_2, T_3 \in \mathrm{End}(V)$. Dann gilt

$$T_1 \circ (T_2 + T_3)(v) = T_1(T_2v + T_3v) \overset{T_1 \text{linear}}{=} T_1T_2v + T_1T_3v = (T_1 \circ T_2 + T_1 \circ T_3)(v).$$

Daher gilt $T_1(T_2 + T_3) = T_1T_2 + T_1T_3$.

Sei jetzt $n = \dim V < \infty$. Die Tatsache, dass $\psi_{\mathcal{B}}^{\mathcal{B}}$ ein Ring-Isomorphismus ist, folgt aus Satz 3.5.3 und Lemma 3.3.29. Beachten Sie, dass $[\mathrm{Id}_V]_{\mathcal{B}} = I_n$ und für $0 \in \mathrm{End}(V)$ ist $[0]_{\mathcal{B}} = 0 \in M_{n\times n}(K)$, unabhängig davon wie $\mathcal{B}$ gewählt wurde.[7] □

3.6 Die Inverse, Elementarmatrizen und lineare Gleichungssysteme

Dieser Abschnitt ist eine Art Zusammenfassung von verschiedenen Ergebnissen, die wir bisher gesehen haben.

Erinnern Sie sich an die Abbildung

$$\Phi : K^k \to \text{Lös}(A, 0),$$

wobei $A \in M_{m\times n}(K)$ ist und $k = n - \mathrm{Rang}(A)$, die wir im Fischer [7, Kap. 0.4] besprochen haben. Die Abbildung Φ hat die Form (hier ist $\lambda = (\lambda_1, \ldots, \lambda_k) \in K^k$)

$$\Phi(\lambda) = \Phi(\lambda_1, \ldots, \lambda_k) = (x_1(\lambda), \ldots, x_r(\lambda), \lambda_1, \ldots, \lambda_k)$$

[7] Man könnte daher sagen, dass die Koordinaten der Identitätsabbildung und der Nullabbildung nicht von der Wahl einer Basis abhängen, also kanonisch sind.

bis auf Umordnung der Variablen, wobei $r = \mathrm{Rang}(A)$. Wenn wir die Struktur der Abbildung betrachten, sehen wir, dass $x_i(\lambda)$ linear ist in $\lambda = (\lambda_1, \dots, \lambda_k)$ und wir können daraus schliessen, dass Φ eine bijektive lineare Abbildung ist. Dies zeigt:

Korollar 3.6.1. *Sei $A \in M_{m\times n}(K)$. Dann ist*

$$U = L\ddot{o}s(A, 0) = \mathrm{Ker}(A) = \mathrm{Ker}(m_A)$$

ein Untervektorraum von K^n von Dimension $n - \mathrm{Rang}(A)$. Die Abbildung Φ ist ein expliziter Isomorphismus von $K^{n-\mathrm{Rang}(A)}$ nach U.

Korollar 3.6.2. *Sei $b \in K^m$ und $A \in M_{m\times n}(K)$. Es gilt*

(1) $L\ddot{o}s(A, b) \neq \emptyset$ genau dann, wenn $b \in \mathrm{SR}(A)$.

(2) Falls $L\ddot{o}s(A, b) \neq \emptyset$ und $y \in L\ddot{o}s(A, b)$ ist, dann ist

$$L\ddot{o}s(A, b) = y + L\ddot{o}s(A, 0) := \{y + x \mid x \in L\ddot{o}s(A, 0)\}. \tag{3.23}$$

Beweis. (1) Dies folgt aus dem „Trick“ aus Lemma 2.2.17:

$$Ax = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1 \cdot \begin{pmatrix} | \\ v_1 \\ | \end{pmatrix} + x_2 \cdot \begin{pmatrix} | \\ v_2 \\ | \end{pmatrix} + \cdots + x_n \cdot \begin{pmatrix} | \\ v_n \\ | \end{pmatrix}.$$

(2) Dies folgt aus der Linearität von m_A: $A(x + y) = Ax + Ay$. (Wieso?)

□

Mengen der Form (3.23) heissen *affine Unterräume*. Wir werden mehr dazu in Abschnitt 3.7 sagen. Wir besprechen jetzt zwei klassische Ergebnisse „in der Sprache der linearen Gleichungssysteme“ und beweisen sie mit unserer neu entwickelten Theorie.

Proposition 3.6.3. *Sei $A \in M_{m\times n}(K)$ und $b \in K^m (= K^m_{Spal})$. Dann sind folgende Aussagen äquivalent:*

(1) Es gilt $\mathrm{Rang}(A) = \mathrm{Rang}(A|b)$.

(2) Die Gleichung $Ax = b$ hat eine Lösung.

Beweis. (1) ist äquivalent zu $\mathrm{Spaltenrang}(A) = \mathrm{Spaltenrang}(A|b)$, was äquivalent ist zu $\mathrm{Sp}(v_1, \dots, v_n) = \mathrm{Sp}(v_1, \dots, v_n, b)$, wobei $v_1, \dots, v_n$ die Spalten von A sind. Dies wiederum ist äquivalent zu $b \in \mathrm{Sp}(v_1, \dots, v_n) = \mathrm{SR}(A)$ (Wieso?). Wegen Korollar 3.6.2 (1) ist dies äquivalent zu (2). □

Beispiel 3.6.4. Die Matrix $\begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$ hat Rang 2. Die Matrix

$$\begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix} = \left(\begin{array}{cc|c} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{array}\right)$$

hat auch Rang 2 und daher hat

$$x + 4y = 7 \tag{3.24}$$
$$2x + 5y = 8 \tag{3.25}$$
$$3x + 6y = 9 \tag{3.26}$$

eine Lösung. In der Tat ist $x = -1$, $y = 2$ eine Lösung.

Beispiel 3.6.5. Wenn A eine Matrix in Stufenform ist mit Rang r, dann sieht $(A \mid b)$ so aus:

$$\left(\begin{array}{cccc|c} 1 & * & \cdots & * & b_1 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & & \ddots & * & \\ & & & 1 & b_r \\ \hline 0 & & \cdots & & b_{r+1} \\ \vdots & & & & \vdots \\ 0 & & \cdots & & b_m \end{array}\right),$$

wobei $*$ nicht Nulleinträge sind. Hier sieht man, dass $\operatorname{Rang}(A \mid b) = \operatorname{Rang}(A)$ genau dann, wenn $b_{r+1} = \ldots = b_m = 0$.

Ein ähnlicher Beweis zeigt:

Proposition 3.6.6. *Sei $A \in M_{m\times n}(K)$ und $b \in K^m (= K^m_{Spal})$. Dann sind folgende Aussagen äquivalent:*

(1) Es gilt $\operatorname{Rang}(A) = \operatorname{Rang}(A|b) = n$.

(2) Die Gleichung $Ax = b$ hat eine eindeutige Lösung.

Beweis. Sei

$$A = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix}.$$

Da $\text{Rang}(A) = \text{Spaltenrang}(A)$ ist, gilt:

$$\text{Die Vektoren } v_1, \ldots, v_n \text{ bilden eine Basis von } \text{SR}(A) \iff \text{Rang}(A) = n. \tag{3.27}$$

Des Weiteren ist wie zuvor

$$\text{Rang}(A|b) = \text{Rang}(A) \iff b \in \text{SR}(A). \tag{3.28}$$

Daher ist (1) äquivalent zu $b \in \text{SR}(A)$ und $\{v_1, \ldots, v_n\}$ ist eine Basis von $\text{SR}(A)$, was genau äquivalent ist zu (2) (laut Proposition 2.2.35). $\square$

Beispiel 3.6.7. Zurück zu Beispiel 3.6.4: Da $A := \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix} \in M_{3\times 2}(\mathbb{Q})$ und $\text{Rang}(A) = 2$ ist die Lösung (3.24) eindeutig.

Proposition 3.6.8. *Sei $A \in M_{n\times n}(K)$. Dann sind folgende Aussagen äquivalent:*

(1) A ist invertierbar.

(2) $m_A : K^n \to K^n$ ist ein Isomorphismus.

(3) $m_A : K^n \to K^n$ ist injektiv.

(4) $m_A : K^n \to K^n$ ist surjektiv.

(5) $\text{Rang}(A) = n$.

(6) Spaltenrang$(A) = n$.

(7) Zeilenrang$(A) = n$.

(8) Die Spalten von A sind eine Basis von K^n.

(9) Die Spalten von A erzeugen K^n.

(10) Die Spalten von A sind linear unabhängig.

(11) $Ax = 0$ hat nur die triviale Lösung $x = 0 \in K^n$.

(12) $Ax = b$ hat für alle $b \in K^n$ eine eindeutige Lösung.

(13) A^T ist invertierbar.

(14) $\text{Rang}(A^T) = n$.

(15) Die Zeilen von A sind eine Basis.

(16) Die Zeilen von A erzeugen K^n.

(17) Die Zeilen von A sind linear unabhängig.

(18) A^{-1} ist invertierbar.

(19) A ist zeilenäquivalent zu I_n. Das heisst, dass man A mittels Zeilenoperationen zu I_n überführen kann.

(20) A ist spaltenäquivalent zu I_n. Das heisst, dass man A mittels Spaltenoperationen zu I_n überführen kann.

Beweis. (1) $\iff$ (2): Dies entspricht Lemma 3.4.3.

(2), (3) und (4) sind äquivalent wegen Korollar 3.2.11 (Korollar zum Rangsatz).

(4) $\iff$ (6): Dies gilt, weil $\mathrm{SR}(A) = \mathrm{Sp}(m_A(e_1), \ldots, m_A(e_n))$.

(5), (6) und (7) sind äquivalent wegen Satz 3.4.1.

(6) $\iff$ (8): Dies gilt wegen Satz 2.3.15 (Gleichgewicht).

(8), (9) und (10) sind wiederum wegen Satz 2.3.15 (Gleichgewicht) äquivalent.

(10) $\iff$ (11): Dies folgt aus dem Multiplikationstrick in Lemma 2.2.17.

(12) $\iff$ (8): Dies gilt wiederum wegen dem Multiplikationstrick und Proposition 2.2.35.

(1) $\iff$ (13): Dies gilt wegen Übung 3.3.26 (für „$\Rightarrow$“), und Übung 3.3.26 zusammen mit $(A^T)^T = A$ (für „$\Leftarrow$“).

Die Äquivalenzen von (13), (14), (15), (16) und (17) folgen aus den Äquivalenzen von (1), (5), (8), (9) und (10) angewandt auf A^T.

(1) $\iff$ (18): Dies folgt aus Korollar 3.3.22 (2).

Dies zeigt die Äquivalenzen (1)-(18).

Wir zeigen die Implikation (5) $\Rightarrow$ (19) sorgfältig: Mit Gauss'scher Elimination haben wir gezeigt, dass A zeilenäquivalent ist zu einer Matrix der Form

$$B = \begin{pmatrix} 1 & * & \cdots & * \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & * \\ 0 & \cdots & 0 & 1 \end{pmatrix}.$$

Beachten Sie, dass es keine Nullen auf der Diagonalen gibt, da $\mathrm{Rang}(A) = n$ laut (5).

Bemerken Sie für später auch, dass wir bis jetzt immer Zeilenoperationen der Form $L_i \to L_i + \alpha L_j$ nur mit $j < i$ benutzt haben, was wir „reinigen nach unten“ genannt haben. Wir können aber auch „nach oben reinigen“: Mit Zeilenoperationen der Form $L_i + \alpha L_j \to L_i$, wobei $j > i$, können wir $\mathcal{B}$ nach I_n überführen, was (19) zeigt.

(14) $\Rightarrow$ (20): Dies folgt, wenn wir die Implikation (5) $\Rightarrow$ (19) für A^T benutzen und aus der Tatsachen, dass sich Zeilenoperationen auf A^T wie Spaltenoperationen auf A verhalten.

Wir zeigen jetzt (19) $\Rightarrow$ (7): Sei A eine Matrix, die zeilenäquivalent zu I_n ist. Laut Lemma 2.3.23 ist $\mathrm{ZR}(A) = \mathrm{ZR}(I_n) = K^n$, was insbesondere (7) zeigt. Mit Transposition zeigt dies auch, dass (20) $\Rightarrow$ (6), was den Beweis dieser Proposition beendet! $\square$

Später werden wir einen anderen Beweis von (19) $\Rightarrow$ (1) geben, der sogar einen Algorithmus für die Berechnung der Inverse gibt. Dafür führen wir Elementarmatrizen im nächsten Abschnitt ein.

Bemerkung 3.6.9. Ein Hauptziel des nächsten Kapitels wird sein die Matrix A mit einem Skalar $\det A$ zu versehen, so dass die Bedingungen in Proposition 3.6.8 äquivalent sind zur Aussage $\det A \neq 0$. Der Skalar $\det A$ wird *Determinante* von A genannt.

Übung 3.6.10. Falls Sie nicht warten können, dann können Sie versuchen zu zeigen, dass

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc \tag{3.29}$$

genau die gewünschte Bedingung für 2×2 Matrizen in Bemerkung 3.6.9 erfüllt.

Ein anderes Ziel des nächsten Kapitels wird es sein ein Formel für die Inverse einer Matrix zu geben.

Übung 3.6.11. Wiederum, wenn Sie nicht warten können, dann können Sie zeigen: Falls $\det A \neq 0$, dann ist

$$A^{-1} = \frac{1}{\det A} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

wobei $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ und $\det A$ wie in (3.29) gegeben ist.

3.6.1 Elementarmatrizen und Zeilen- und Spaltenoperationen

Wir geben jetzt einen anderen Beweis der Tatsache:

$$A \text{ ist zeilenäquivalent zu } I_n \Rightarrow A \text{ ist invertierbar}$$

(Siehe Proposition 3.6.8).

Dafür definieren wir eine spezielle Famile von Matrizen, sogenannte Elementarmatrizen, welche uns erlauben werden Zeilen- und Spaltenoperationen als Matrixmultiplikationen aufzufassen.

Erinnern Sie sich daran, dass wir mit E_{ij} die Matrix bezeichnet, die überall Nulleinträge hat ausser beim ij-ten Eintrag, welcher 1 ist.

Definition 3.6.12. Sei $n \in \mathbb{N}$. Wir definieren[8] die folgenden quadratischen Matrizen:

- Für $1 \leq i \neq j \leq n$ und $\alpha \in K$ definieren wir

$$Q_{i,j}(\alpha) := I_n + \alpha E_{ij} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} + \alpha \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \alpha & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

- Für $1 \leq i, j \leq n$ definieren wir

$$P_{i,j} = I_n - E_{ii} - E_{jj} + E_{ij} + E_{ji}.$$

  Wir nennen P_{ij} eine *Permutationsmatrix.*

- Für $1 \leq i \leq n$ und $\alpha \in K^\times$ definieren wir

$$S_i(\alpha) = \begin{pmatrix} 1 & 0 & \cdots & & & \cdots & 0 \\ 0 & \ddots & \ddots & & & & \vdots \\ \vdots & \ddots & 1 & \ddots & & & \\ & & 0 & \alpha & 0 & & \\ & & & \ddots & 1 & \ddots & \vdots \\ \vdots & & & & \ddots & \ddots & 0 \\ 0 & \cdots & & & \cdots & 0 & 1 \end{pmatrix},$$

  wobei α im ii-ten Eintrag steht. Matrizen dieser Form heissen *Elementarmatrizen.*

Einige Autoren nennen $Q_{i,j}(\alpha)$ Elementarmatrizen von *Typ I*, $P_{i,j}$ von *Typ II* und $S_i(\alpha)$ von *Typ III.*

[8]Leider gibt es keine Standard-Notation dieser Matrizen

Beispiel 3.6.13. In $M_{4\times 4}(K)$ ist

$$Q_{1,3}(\alpha) = \begin{pmatrix} 1 & 0 & \alpha & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad Q_{3,1}(\beta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \beta & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$P_{1,2} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad P_{2,4} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$S_3(\alpha) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \alpha & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Lemma 3.6.14. *Sei $A \in M_{n\times p}(K)$. Es gilt*

- *Linksmultiplikation mit $Q_{i,j}(\alpha)$ entspricht der Zeilenoperation $L_i + \alpha L_j \to L_i$. Das heisst*
$$A \xrightarrow{L_i+\alpha L_j\to L_i} Q_{i,j}(\alpha)A.$$
- *Linksmultiplikation mit $P_{i,j}$ entspricht der Zeilenoperation $L_i \leftrightarrow L_j$. Das heisst*
$$A \xrightarrow{L_i\leftrightarrow L_j} P_{i,j}A.$$
- *Linksmultiplikation mit $S_i(\alpha)$ entspricht der Zeilenoperation $\alpha L_i \to L_i$. Das heisst*
$$A \xrightarrow{\alpha L_i\to L_i} S_i(\alpha)A.$$

Ähnlich gilt für Spaltenoperationen[9]*: Sei $A \in M_{m\times n}(K)$, dann gilt:*

- *Rechtsmultiplikation mit $Q_{i,j}(\alpha)$ entspricht der Spaltenoperation $C_j + \alpha C_i \to C_j$. Das heisst*
$$A \xrightarrow{C_j+\alpha C_i\to C_j} AQ_{i,j}(\alpha).$$
- *Rechtsmultiplikation mit $P_{i,j}$ entspricht der Spaltenoperation $C_i \leftrightarrow C_j$. Das heisst*
$$A \xrightarrow{C_i\leftrightarrow C_j} AP_{i,j}.$$

[9]Streng genommen haben wir nur Zeilenoperationen definiert, aber die Definition der Spaltenoperationen ist ganz analog.

- *Rechtsmultiplikation mit $S_i(\alpha)$ entspricht der Spaltenoperation $\alpha C_i \to C_i$. Das heisst*
$$A \xrightarrow{\alpha C_i \to C_i} AS_i(\alpha).$$

Beweis. All diese Aussagen sind lediglich eine Rechnung. □

Lemma 3.6.15. *Es gilt:*

- *Für alle $\alpha \in K$ ist*
$$(Q_{i,j}(\alpha))^{-1} = Q_{i,j}(-\alpha).$$
- *Die Inverse von $P_{i,j}$ ist gegeben durch $(P_{i,j})^{-1} = P_{i,j}$.*
- *Für alle $\alpha \in K^\times$ ist*
$$(S_i(\alpha))^{-1} = S_i\left(\frac{1}{\alpha}\right).$$

Insbesondere ist jede Elementarmatrix invertierbar und ihre Inverse ist auch elementar.

Beweis. Dies ist wieder eine Berechnung. Diesmal können wir dies jedoch mit Lemma 3.6.14 sehen. Zum Beispiel entspricht

$$Q_{i,j}(\alpha)^{-1} = Q_{i,j}(-\alpha)$$

der Tatsache, dass $L_i - \alpha L_j \to L_i$ die Inverse Operation zu $L_i + \alpha L_j \to L_i$ ist. Ähnlich entspricht
$$(P_{i,j})^{-1} = P_{j,i}$$
der Tatsache, dass $L_i \leftrightarrow L_j$ die Inverse Operation zu $L_i \leftrightarrow L_j$ ist und so weiter. □

Satz 3.6.16. *Jede Matrix $A \in \mathrm{GL}_n(K)$ ist ein endliches Produkt von Elementarmatrizen.*

Beweis. Im Beweis von Proposition 3.6.8 haben wir gesehen, dass jede invertierbare Matrix zeilenäquivalent ist zu I_n. Das heisst, dass man A mit endlich vielen Zeilenoperationen nach I_n überführen kann.

Wenn wir das mit Lemma 3.6.14 betrachten, entsprechen diese Zeilenoperationen Linksmultiplikation mit Elementarmatrizen. Daher existieren Elementarmatrizen $T_1, \dots, T_k$ mit
$$T_k \cdots T_2 T_1 A = I_n. \tag{3.30}$$

Wir multiplizieren (3.30) mit

$$(T_k \cdots T_1)^{-1} = T_1^{-1} \cdots T_{k-1}^{-1} T_k^{-1}$$

und sehen, dass

$$A = T_1^{-1} \cdots T_k^{-1}.$$

Laut Lemma 3.6.15 sind alle T_i^{-1} auch elementar und das Lemma folgt. $\square$

Bemerkung 3.6.17. In der Sprache der Gruppentheorie besagt Lemma 3.6.15, dass $\mathrm{GL}_n(K)$ durch Elementarmatrizen erzeugt ist.

Korollar 3.6.18. *Mit der Notation des Beweises von Satz 3.6.16 ist* $A^{-1} = T_k \cdots T_2 T_1$.

Diese Tatsache ermöglicht uns ein einfaches Verfahren zum Bestimmen der Inversen einer Matrix. Sei $A \in M_{n\times n}(K)$. Wir möchten wissen, ob A invertierbar ist und falls ja, möchten wir auch A^{-1} berechnen können.

Hier ist der Trick: Für $B, C \in M_{n\times n}(K)$ schreiben wir $(B \mid C) \in M_{n\times 2n}(K)$ für die zusammengefügte Matrix.

Lemma 3.6.19. *Seien* $B, C, D \in M_{n\times n}(K)$. *Es gilt*

$$B(C \mid D) = (BC \mid BD).$$

Beweis. Noch einmal eine direkte Berechnung! $\square$

Dies können wir benutzen, um Folgendes zu zeigen:

Satz 3.6.20. *Seien* $n \in \mathbb{N}$ *und* $A \in M_{n\times n}(K)$. *Man versucht die* $n \times 2n$ *zusammengefügte Matrix* $(A \mid I_n)$ *mittels Zeilenoperationen zur Form* $(I_n \mid B)$ *zu bringen. Wenn dies möglich ist, dann ist* A *invertierbar und es gilt* $A^{-1} = B$. *Falls dies nicht möglich ist, dann ist* A *nicht invertierbar.*

Beweis. Beachten Sie, dass $(A \mid I_n)$ zeilenäquivalent zu $(I_n \mid B)$ ist genau dann, wenn A zeilenäquivalent zu I_n ist. Dies ist laut Proposition 3.6.8 äquivalent zur Invertierbarkeit von A. Dies zeigt die zweite Aussage. Nehmen wir also an, dass A zeilenäquivalent ist zu I_n. Laut Lemma 3.6.15 existieren Elementarmatrizen $T_1, \ldots, T_k$ mit

$$T_k \cdots T_2 T_1 A = I_n$$

(vgl. Gleichung (3.30)). Erinnern Sie sich daran, dass wir die Zeilenoperationen auf $(A \mid I_n)$ anwenden. Laut Lemma 3.6.19 entspricht das dem Produkt

$$T_k \cdots T_1 (A \mid I_n) \overset{3.6.19}{=} (T_k \cdots T_1 A \mid T_k \cdots T_1 I_n) = (I_n \mid T_k \cdots T_1) \overset{3.6.18}{=} (I_n \mid A^{-1}),$$

was genau der ersten Aussage entspricht. $\square$

Einige Beispiele

Hier geben wir einige Beispiele für verschiedene Teile dieses Kapitels.

Beispiel 3.6.21 (Injektivität und Surjektivität)**.** Beachten Sie, dass wir Korollar 3.2.11 nicht benutzen können, wenn die Vektorräume nicht endlich-dimensional sind: Die Abbildung

$$\begin{aligned} m_{x^{73}} : K[x] &\to K[x] \\ p &\mapsto x^{73}p \end{aligned}$$

ist injektiv aber nicht surjektiv. Ähnlich ist

$$\begin{aligned} D : K[x] &\to K[x] \\ p &\mapsto p' \end{aligned}$$

surjektiv aber nicht injektiv.

Beispiel 3.6.22. Angenommen Sie betrachten $T \in \operatorname{End}(\mathbb{Q}[x]_2)$ und wollen T bezüglich der Basis

$$\mathcal{C} = (1 + 2x + 2x^2, 2 + 4x + x^2, x)$$

berechnen (vielleicht wegen einer bestimmten Anwendung). Normalerweise wird T bezüglich der Standard-Basis $\mathcal{E} = (1, x, x^2)$ gegeben sein. Also ist es eine direkte Rechnung $[T]_{\mathcal{E}}$ zu berechnen. Statt nun $[T]_{\mathcal{C}}$ direkt zu berechnen (wodurch wir viele lineare Gleichungssysteme lösen müssten), benutzen wir die Transformationsformel

$$[T]_{\mathcal{C}} = [\mathrm{Id}]^{\mathcal{E}}_{\mathcal{C}}[T]^{\mathcal{E}}_{\mathcal{E}}[\mathrm{Id}]^{\mathcal{C}}_{\mathcal{E}},$$

wobei $\mathrm{Id} = \mathrm{Id}_{\mathbb{Q}[x]_2}$. Beachten Sie, dass es sehr einfach ist $[\mathrm{Id}]^{\mathcal{C}}_{\mathcal{E}}$ zu berechnen. Es gilt

$$[\mathrm{Id}]^{\mathcal{C}}_{\mathcal{E}} = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 4 & 1 \\ 2 & 1 & 0 \end{pmatrix}.$$

Die Matrix $[\mathrm{Id}]^{\mathcal{E}}_{\mathcal{C}}$ zu berechnen ist hingegen nicht sehr schön. Es ist viel einfach sich daran zu erinnern, dass

$$[\mathrm{Id}]^{\mathcal{E}}_{\mathcal{C}} = ([\mathrm{Id}]^{\mathcal{C}}_{\mathcal{E}})^{-1}$$

und Gauss-Jordan zu benutzen. Wir suchen also die Inverse von

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 4 & 1 \\ 2 & 1 & 0 \end{pmatrix}.$$

Wir betrachten $(A \mid I) \in M_{3\times 6}(\mathbb{R})$ und führen die Gauss'sche Elimination durch:

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 2 & 4 & 1 & 0 & 1 & 0 \\ 2 & 1 & 0 & 0 & 0 & 1 \end{array}\right) \xrightarrow[L_3 - 2L_1 \to L_3]{L_2 - 2L_1 \to L_2} \left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & -3 & 0 & -2 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{L_3 \to L_2} \left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & -3 & 0 & -2 & 0 & 1 \\ 0 & 0 & 1 & -2 & 1 & 0 \end{array}\right)$$

$$\xrightarrow[-\frac{1}{3}L_2 \to L_2]{L_1 + \frac{2}{3}L_2 \to L_1} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{1}{3} & 0 & \frac{2}{3} \\ 0 & 1 & 0 & \frac{2}{3} & 0 & -\frac{1}{3} \\ 0 & 0 & 1 & -2 & 1 & 0 \end{array}\right)$$

Somit ist die Inverse von A

$$A^{-1} = \begin{pmatrix} -\frac{1}{3} & 0 & \frac{2}{3} \\ \frac{2}{3} & 0 & -\frac{1}{3} \\ -2 & 1 & 0 \end{pmatrix}.$$

Es gilt also

$$[T]_{\mathcal{C}} = [\mathrm{Id}]^{\mathcal{E}}_{\mathcal{C}} [T]^{\mathcal{E}}_{\mathcal{E}} [\mathrm{Id}]^{\mathcal{C}}_{\mathcal{E}} = \begin{pmatrix} -\frac{1}{3} & 0 & \frac{2}{3} \\ \frac{2}{3} & 0 & -\frac{1}{3} \\ -2 & 1 & 0 \end{pmatrix} [T]^{\mathcal{E}}_{\mathcal{E}} \begin{pmatrix} 1 & 2 & 0 \\ 2 & 4 & 1 \\ 2 & 1 & 0 \end{pmatrix}.$$

Beispiel 3.6.23. Wir betrachten den Vektorraum **Fib** $\subseteq K^\infty$. Die Folgen

$$F_{1,0} = (1, 0, 1, 1, 2, 3, \ldots)$$
$$F_{0,1} = (0, 1, 1, 2, 3, \ldots)$$

bilden eine Basis $\mathcal{B} = (F_{1,0}, F_{0,1})$ von **Fib**. Wir haben ein beliebiges Element von **Fib** als

$$F_{a,b} = (a, b, a + b, \ldots)$$

notiert. Es gilt

$$[F_{a,b}]_{\mathcal{B}} = \begin{pmatrix} a \\ b \end{pmatrix},$$

da $F_{a,b} = aF_{1,0} + bF_{0,1}$. Wir haben die Basis $\mathcal{C} = (F_{1,\varphi}, F_{1,\psi})$ gefunden, wobei

$$\varphi = \frac{1 + \sqrt{5}}{2} \quad \text{und} \quad \psi = \frac{1 - \sqrt{5}}{2}.$$

Wie wir in der Einleitung erklärt haben, kann man diese Basis finden, wenn man die „Symmetrie“ der Verschiebungsabbildung

$$S : \mathbf{Fib} \to \mathbf{Fib}$$
$$(a, b, a+b, a+2b, \dots) \mapsto (b, a+b, a+2b, \dots)$$

betrachtet. Wir werden dies noch rigoros erklären, wenn wir Eigenwerte und Eigenvektoren besprechen. Bezüglich $\mathcal{B}$ sieht S so aus

$$[S]_{\mathcal{B}} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

Proposition 3.3.8 besagt, dass

$$[S]_{\mathcal{B}}[F_{a,b}]_{\mathcal{B}} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} b \\ a+b \end{pmatrix} = [S(F_{a,b})]_{\mathcal{B}}.$$

Daraus folgt[10]: Falls $F = (a_0, a_1, a_2, \dots) \in \mathbf{Fib}$, dann ist

$$([S]_{\mathcal{B}})^n [F]_{\mathcal{B}} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}^n \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} a_n \\ a_{n+1} \end{pmatrix}.$$

Wenn man die Potenzen $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}^n$ berechnen kann, dann findet man eine Formel für a_n, die nur von a_0 und a_1 abhängt. Aber die Potenzen von $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ zu berechnen ist schwierig. Man könnte sich fragen, ob es eine ähnliche Matrix zu $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ gibt, deren Potenzen leicht zu berechnen sind. Ähnliche Matrizen zu $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ entsprechen $[S]_{\mathcal{D}}$ für andere Basen $\mathcal{D}$. Daher könnte man sich fragen: Gibt es eine Basis, auf deren Elemente die Abbildung S besonders einfach operiert? In der Linearen Algebra II werden wir lernen wie man solche Basen findet. Für S haben wir eine solche Basis jedoch schon gefunden: Bezüglich $\mathcal{C}$ sieht S so aus:

$$[S]_{\mathcal{C}} = \begin{pmatrix} \varphi & 0 \\ 0 & \psi \end{pmatrix},$$

[10] man könnte es auch direkt sehen.

da $S(F_{1,\varphi}) = \varphi F_{1,\varphi}$ und $S(F_{1,\psi}) = \psi F_{1,\psi}$. Laut der Transformationsformel 3.3.30 sind $[S]_\mathcal{B}$ und $[S]_\mathcal{C}$ ähnlich:

$$[S]_\mathcal{B} = [\mathrm{Id}_{\mathbf{Fib}}]^\mathcal{C}_\mathcal{B}[S]_\mathcal{C}[\mathrm{Id}_{\mathbf{Fib}}]^\mathcal{B}_\mathcal{C}.$$

Wir setzen $P := [\mathrm{Id}_{\mathbf{Fib}}]^\mathcal{B}_\mathcal{C}$. Dann gilt

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} = P^{-1} \begin{pmatrix} \varphi & 0 \\ 0 & \psi \end{pmatrix} P. \tag{3.31}$$

Beachten Sie: Potenzen von $[S]_\mathcal{C} = \begin{pmatrix} \varphi & 0 \\ 0 & \psi \end{pmatrix}$ sind einfach zu berechnen:

$$\begin{pmatrix} \varphi & 0 \\ 0 & \psi \end{pmatrix}^n = \begin{pmatrix} \varphi^n & 0 \\ 0 & \psi^n \end{pmatrix}.$$

Aus (3.31) folgt

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}^n = P^{-1} \begin{pmatrix} \varphi & 0 \\ 0 & \psi \end{pmatrix}^n P = P^{-1} \begin{pmatrix} \varphi^n & 0 \\ 0 & \psi^n \end{pmatrix} P$$

(Wieso?). Beachten Sie, dass P und P^{-1} ohne Potenzen erscheinen. Dies gibt uns eine Motivation, um P und P^{-1} zu berechnen. Eine davon ist leicht zu berechnen und die andere bereitet ein bisschen Mühe. In der Tat ist

$$P^{-1} = [\mathrm{Id}_{\mathbf{Fib}}]^\mathcal{C}_\mathcal{B}$$

einfacher zu berechnen. Da $F_{1,\varphi} = 1 \cdot F_{1,0} + \varphi \cdot F_{0,1}$ und $F_{1,\psi} = 1 \cdot F_{1,0} + \psi \cdot F_{0,1}$ ist

$$P^{-1} = \begin{pmatrix} 1 & 1 \\ \varphi & \psi \end{pmatrix}.$$

Da $P = (P^{-1})^{-1} = \begin{pmatrix} 1 & 1 \\ \varphi & \psi \end{pmatrix}^{-1}$ ist, könnte man unsere obige Methode benutzen, um

$$P = \frac{1}{\psi - \varphi} \begin{pmatrix} \psi & -1 \\ -\varphi & 1 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} -\psi & 1 \\ \varphi & -1 \end{pmatrix}.$$

zu berechnen. Daher gilt

$$\begin{aligned}\begin{pmatrix}0 & 1\\ 1 & 1\end{pmatrix}^n &= \frac{1}{\sqrt{5}}\begin{pmatrix}1 & 1\\ \varphi & \psi\end{pmatrix}\begin{pmatrix}\varphi^n & 0\\ 0 & \psi^n\end{pmatrix}\begin{pmatrix}-\psi & 1\\ \varphi & -1\end{pmatrix}\\ &= \frac{1}{\sqrt{5}}\begin{pmatrix}\varphi^n & \psi^n\\ \varphi^{n+1} & \psi^{n+1}\end{pmatrix}\begin{pmatrix}-\psi & 1\\ \varphi & -1\end{pmatrix}\\ &= \frac{1}{\sqrt{5}}\begin{pmatrix}\varphi\psi^n - \psi\varphi^n & \varphi^n - \psi^n\\ \varphi\psi^{n+1} - \psi\varphi^{n+1} & \varphi^{n+1} - \psi^{n+1}\end{pmatrix}.\end{aligned}$$

Daraus folgt beispielsweise das Folgende:

$$F_{0,1} = (b_0, b_1, b_2, \ldots)$$

Dann ist

$$\begin{pmatrix}b_n\\ b_{n+1}\end{pmatrix} = \begin{pmatrix}0 & 1\\ 1 & 1\end{pmatrix}^n \begin{pmatrix}0\\ 1\end{pmatrix} = \frac{1}{\sqrt{5}}\begin{pmatrix}\varphi^n - \psi^n\\ \varphi^{n+1} - \psi^{n+1}\end{pmatrix}$$

oder

$$b_n = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right).$$

Übung 3.6.24. Finden Sie eine Formel für a_n, wobei $F_{a,b} = (a_0, a_1, a_2, \ldots)$, indem Sie

$$\begin{pmatrix}0 & 1\\ 1 & 1\end{pmatrix}^n \begin{pmatrix}a\\ b\end{pmatrix}$$

berechnen.

3.7 Fasern und Quotientenräume

3.7.1 Fasern

Wir fangen mit einer (neuen) Notation an. Für zwei Teilmengen $A, B \subseteq V$ eines Vektorraums V definieren wir

$$A + B := \{a + b \mid a \in A,\ b \in B\}.$$

Wir schreiben auch für $a \in V$ und $B \subseteq V$

$$a + B := \{a\} + B.$$

Proposition 3.7.1. *Sei $T : V \to W$ eine lineare Abbildung zwischen zwei Vektorräumen V und W. Dann gilt:*

(1) Falls $w \in \mathrm{Im}(T)$, *dann ist* $T^{-1}(w) = v + \mathrm{Ker}(T)$, *wobei* v *ein beliebiger Vektor ist mit* $Tv = w$.

(2) Falls $w \notin \mathrm{Im}(T)$, *dann gilt* $T^{-1}(w) = \emptyset$.

Beweis. (1) Sei $v \in V$ mit $Tv = w$. Wir zeigen, dass $T^{-1}(w) = v + \mathrm{Ker}(T)$ ist, indem wir zwei Inklusionen zeigen.

„$\supseteq$": Sei $u \in \mathrm{Ker}(T)$. Dann gilt

$$T(v+u) \overset{T \text{ linear}}{=} Tv + Tu = w + 0 = w.$$

Also gilt, dass $T^{-1}(w) \supseteq v + \mathrm{Ker}(T)$.

„$\subseteq$": Sei $v' \in T^{-1}(w)$. Es ist genug zu zeigen, dass $v' - v \in \mathrm{Ker}(T)$. (Wieso?) Da T linear ist, gilt

$$T(v - v') = Tv - Tv' = w - w = 0.$$

Daher folgt $T^{-1}(w) \subseteq v + \mathrm{Ker}(T)$.

(2) Hier gibt es nichts zu beweisen! Dies ist einfach die Definition von $\mathrm{Im}(T)$. □

Die Menge $T^{-1}(w)$ heisst die *Faser* von w bezüglich T.

Beispiel 3.7.2. Wir betrachten die lineare Abbildung

$$\begin{aligned} T : \mathbb{R}^2 &\to \mathbb{R} \\ (x,y) &\mapsto x - y. \end{aligned}$$

Die Abbildung T ist surjektiv und für $z \in \mathbb{R}$ gilt beispielsweise, dass

$$T(z,0) = z$$

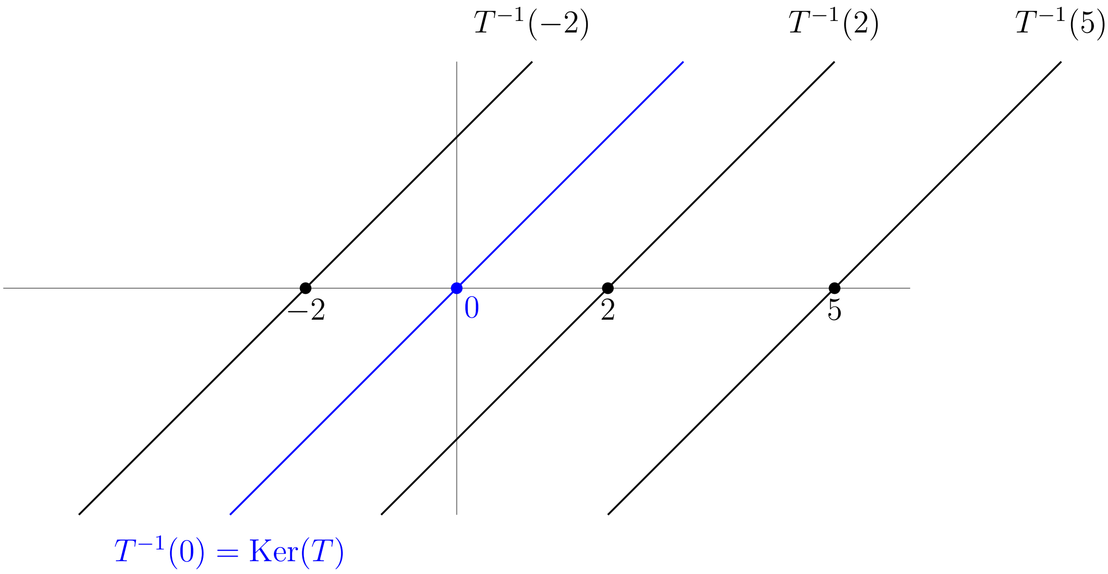
und daher ist

$$T^{-1}(z) = (z,0) + \mathrm{Ker}(T).$$

Der Kern $\mathrm{Ker}(T)$ berechnet sich als

$$\{(x,y) \mid T(x,y) = 0\} = \{(x,y) \mid x - y = 0\} = \{(x,y) \mid x = y\}.$$

Also sehen die Fasern von T so aus:



Wir möchten einen neuen Vektorraum aus den Fasern bilden. In anderen Worten möchten wir Fasern miteinander addieren. Wir betrachten noch einmal eine beliebige lineare Abbildung $T : V \to W$ zwischen zwei Vektorräumen V und W. Wegen Proposition 3.7.1 wissen wir, dass alle nicht leeren Fasern die Form $v + \mathrm{Ker}(T)$ haben. Da $\mathrm{Ker}(T)$ ein Untervektorraum von V ist, erlaubt uns die nachfolgende Konstruktion, Fasern zu addieren, und ist auch etwas allgemeiner.

3.7.2 Quotientenraum

Sei V ein Vektorraum über K und $U \subseteq V$ ein Untervektorraum[11]. Wir definieren eine Relation $\sim_U$ auf V durch

$$v_1 \sim_U v_2 \;:\Longleftrightarrow\; v_1 - v_2 \in U.$$

Ohne viele Worte können wir zwei einfache Lemmata beweisen.

Lemma 3.7.3. *Die Relation $\sim_U$ ist eine Äquivalenzrelation.*

Beweis. Seien $v, v_1, v_2, v_3 \in V$.

Reflexion: Es gilt $v \sim_U v$, da $v - v = 0 \in U$.

Symmetrie: Falls $v_1 \sim_U v_2$, dann ist $v_1 - v_2 \in U$. Dann folgt auch

$$(v_2 - v_1) = -(v_1 - v_2) \in U.$$

Also ist $v_2 \sim_U v_1$.

Transitivität: Seien $v_1 \sim_U v_2$ und $v_2 \sim_U v_3$. Dann ist $v_1 - v_2 \in U$ und $v_2 - v_3 \in U$. Daher ist $v_1 - v_3 = (v_1 - v_2) + (v_2 - v_3) \in U$. Dies zeigt $v_1 \sim_U v_3$. $\square$

[11] Die Leser können an $U = \mathrm{Ker}(T)$ denken.

Lemma 3.7.4 (Äquivalenzklasse)**.** *Für jedes* $v \in V$ *gilt*

$$[v]_{\sim_U} = v + U.$$

Beweis. Es gelten folgende Äquivalenzen:

$$\begin{aligned} v' \in [v]_{\sim_U} &\iff v' - v \in U \\ &\iff \exists\, u \in U : v' = v + u \\ &\iff v' \in v + U. \end{aligned}$$

□

Bemerkung 3.7.5. Im Fall $U = \mathrm{Ker}(T)$ sind die Äquivalenzklassen genau die nicht leeren Fasern. Die entsprechende Partition

$$V = \bigsqcup_{v \in V} [v]_{\sim_{\mathrm{Ker}(T)}} = \bigsqcup_{v \in V} v + \mathrm{Ker}(T)$$

von V heisst in diesem Fall eine *Faserung* von V durch T.

Beispiel 3.7.6. Für $U = \mathrm{Ker}(T)$ mit T wie in Beispiel 3.7.2 sind alle Fasern/Äquivalenzklassen Geraden mit Steigung 1.

Definition 3.7.7. Mengen der Form $A = v + U$ heissen *affine Unterräume* von V. Die *Dimension* $\dim A$ von A ist definiert als $\dim U$ und man sagt, dass A *parallel* zu U ist. Der Untervektorraum U nennen wir den Untervektorraum *assoziiert* zum affinen Unterraum A.

In Beispiel 3.7.6 sind alle Geraden affine Unterräume, die zu $U = \mathrm{Ker}(T)$ assoziiert sind.

Wir möchten zeigen, dass die Menge aller affinen Unterräume, welche zu einem gegebenen Unterraum U assoziiert sind, einen Vektorraum bilden.

Definition 3.7.8. Mit der vorherigen Konstruktion, nennen wir die Menge

$$V/U := V/\sim_U = \{v + U \mid v \in V\}$$

den *Quotientenraum* oder den *Faktorraum* von V nach U.

Wie der Name des Quotientenraums (oder des Faktorraums) schon andeutet, wollen wir eine Vektorraum-Struktur auf V/U definieren.

Definition der Addition auf V/U:

Seien $v_1 + U, v_2 + U \in V/U$. Wir definieren die Addition $+ : V/U \times V/U \to V/U$ durch

$$(v_1 + U) + (v_2 + U) := v_1 + v_2 + U.$$

Lemma 3.7.9. *Die zuvor definierte Addition ist wohl-definiert.*

Beweis. Seien $v_1, v_1', v_2, v_2' \in V$ mit $v_1 + U = v_1' + U$ und $v_2 + U = v_2' + U$. Um zu zeigen, dass die Addition wohl-definiert ist, müssen wir zeigen, dass

$$v_1 + v_2 + U = v_1' + v_2' + U.$$

Dies ist äquivalent zu

$$v_1 + v_2 - (v_1' + v_2') \in U \quad \text{bzw.} \quad (v_1 - v_1') + (v_2 - v_2') \in U.$$

Letzteres ist jedoch wahr, da nach Annahme $v_1 - v_1'$, $v_2 - v_2' \in U$, und U ein Untervektorraum ist. □

Definition der Skalarmultiplikation auf V/U:
Wir definieren die Skalarmultiplikation auf V/U durch

$$\begin{aligned} \cdot : K \times V/U &\to V/U \\ (\alpha, v + U) &\mapsto \alpha v + U. \end{aligned}$$

Übung 3.7.10. Zeigen Sie, dass die Skalarmultiplikation wohl-definiert ist.

Man könnte jetzt überprüfen, dass die Axiome eines Vektorraums für V/U gelten. (Dies ist ziemlich leicht, da alles von den entsprechenden Axiomen für V und den Untervektorraum-Axiomen für U folgt.) Zum Protokoll:

Proposition 3.7.11. *Mit der obigen Addition und Skalarmultiplikation ist V/U ein Vektorraum über K.*

Beispiel 3.7.12. Hier sind zwei Beispiele in $\mathbb{R}^3$:

- Sei U eine Gerade, die durch den Ursprung geht. Dann ist $\mathbb{R}^3/U$ die Menge aller Geraden, die zu U parallel sind.
- Sei W eine Ebene in $\mathbb{R}^3$, die durch den Ursprung geht. Dann ist $\mathbb{R}^3/W$ die Menge aller Ebenen, die parallel sind zu W.

Proposition 3.7.13. *Sei $\pi_U : V \to V/U$ die Abbildung*

$$\pi_U(v) = v + U.$$

Dann ist π_U linear, $\operatorname{Ker}(\pi_U) = U$ *und* $\operatorname{Im}(\pi_U) = V/U$.

Bemerkung 3.7.14. Die Abbildung π_U heisst die *kanonische Quotientenabbildung.*

Beweis von Proposition 3.7.13. Wir müssen drei Dinge überprüfen.

Linearität: Seien $v_1, v_2 \in V$, dann gilt

$$\pi_U(v_1 + v_2) = v_1 + v_2 + U = (v_1 + U) + (v_2 + U) = \pi_U(v_1) + \pi_U(v_2),$$

wobei die zweite Gleichheit wegen der Definition der Addition gilt. Ähnlich gilt

$$\pi_U(\alpha v_1) = \alpha v_1 + U = \alpha(v_1 + U) = \alpha \pi_U(v_1).$$

$\operatorname{Ker}(\pi_U) = U$: Der Nullvektor in V/U ist $U = 0 + U$. Wir berechnen

$$\operatorname{Ker}(\pi_U) = \{v \in V \mid v + U = 0 + U\} = \{v \in V \mid v - 0 \in U\} = \{v \in V \mid v \in U\} = U.$$

$\operatorname{Im}(\pi_U) = V/U$: (Wieso? Weil $\pi_U(v) = v + U$.) □

Durch den Rangsatz 3.2.9 erhalten wir folgendes Korollar:

Korollar 3.7.15. *Sei V ein endlich-dimensionaler Vektorraum. Für die Dimension des Quotientenraums V/U gilt*

$$\dim V/U = \dim V - \dim U.$$

Beweis. Laut Proposition 3.7.13 gilt

$$\dim V/U = \dim \operatorname{Im}(\pi_U) = \dim V - \dim \operatorname{Ker}(\pi_U) = \dim V - \dim U,$$

wobei die zweite Gleichheit aus dem Rangsatz 3.2.9 folgt. □

Dies gibt den Fasern einer linearen Abbildung $T : V \to W$ die Struktur eines Vektorraums und dieser ist (kanonisch) isomorph zu $\operatorname{Im}(T)$:

Proposition 3.7.16. *Sei $T : V \to W$ eine lineare Abbildung. Es gilt*

$$V/\operatorname{Ker}(T) \cong \operatorname{Im}(T).$$

Beweis. Wir definieren

$$\begin{aligned} \varphi : V/\operatorname{Ker}(T) &\to \operatorname{Im}(T) \\ v + \operatorname{Ker}(T) &\mapsto T(v). \end{aligned}$$

Wir müssen zeigen, dass φ wohl-definiert ist. Seien dazu $v, v' \in V$ mit $v + \operatorname{Ker}(T) = v' + \operatorname{Ker}(T)$. Dies ist äquivalent zu $v - v' \in \operatorname{Ker}(T)$. Um zu zeigen, dass φ wohl-definiert ist, müssen wir überprüfen, dass

$$\varphi(v' + \mathrm{Ker}(T)) = \varphi(v + \mathrm{Ker}(T))$$

gilt. Wir berechnen

$$\begin{aligned}\varphi(v' + \mathrm{Ker}(T)) = T(v') &= T(v - (v - v')) \\ &= T(v) - T(v - v') \\ &= T(v) - 0 \\ &= T(v) \\ &= \varphi(v + \mathrm{Ker}(T)).\end{aligned}$$

Ausserdem ist φ linear (Wieso?[12]) Die Abbildung φ ist sicher surjektiv (Wieso?). Es bleibt zu zeigen, dass φ injektiv ist. Angenommen $\varphi(v+\mathrm{Ker}(T)) = 0$. Dann ist $T(v) = 0$, also ist v im Kern von T und somit ist $v + \mathrm{Ker}(T) = \mathrm{Ker}(T)$. Also ist φ ein Isomorphismus. □

Bemerkung 3.7.17. Vielleicht haben Sie bemerkt, dass ich hier schon zweimal das magische Wort „kanonisch“ verwendet habe. Ich möchte darauf nicht tief eingehen und nur Folgendes sagen: In der Definition von π_U oder auch in der Definition von φ haben wir keine Wahl getroffen. Zum Beispiel haben wir diese Abbildungen ohne eine gewisse Wahl einer Basis definiert. Für die Definition der Abbildung π_U verwendet man nur die Informationen, die man über V und U hat und nicht mehr. In diesem Sinne ist π_U kanonisch definiert!

Beispiel 3.7.18. Für $U = \mathrm{Ker}(T)$ wie in Beispiel 3.7.2 ist V/U die Menge aller Geraden mit Steigung 1. Jede dieser Geraden kann als

$$(z, 0) + \mathrm{Ker}(T)$$

dargestellt werden und wegen des Beweises von Proposition 3.7.16 wird jede dieser Geraden durch $\varphi : V/U \to \mathrm{Im}(T)$ auf

$$T(z, 0) = z - 0 = z \in \mathbb{R}$$

[12]Für alle $a, b \in K$ und $v_1, v_2 \in V$ ist

$$\begin{aligned}\varphi(a(v_1 + \mathrm{Ker}(T)) + b(v_2 + \mathrm{Ker}(T))) &= \varphi(av_1 + bv_2 + \mathrm{Ker}(T)) \\ &= T(av_1 + bv_2) \\ &= aT(v_1) + bT(v_2) \\ &= a\varphi(v_1 + \mathrm{Ker}(T)) + b\varphi(v_2 + \mathrm{Ker}(T)).\end{aligned}$$

abgebildet.

Hier ist eine Verbindung zwischen direkten Summen, Komplement und Quotientenräume:

Proposition 3.7.19. *Sei $U \subseteq V$ ein Untervektorraum und W ein Komplement von U. Das heisst, dass $W \subseteq V$ ein Untervektorraum ist, so dass $V = U \oplus W$. Dann ist*

$$\pi_U|_W : W \to V/U$$

ein Isomorphismus.

Beweis. Jede Einschränkung einer linearen Abbildung ist wieder linear. Sei nun $v \in V$. Laut Proposition 2.3.40 (4) existieren eindeutige $u \in U$, $w \in W$ mit $v = u + w$. Daher ist w das eindeutige Urbild von $v + U \in V/U$. Dies zeigt, dass $\pi_U|_W$ sowohl injektiv als auch surjektiv ist (Wieso?). □

Zu guter Letzt:

Proposition 3.7.20 (Universelle Eigenschaft von V/U)**.** *Sei $T : V \to W$ eine lineare Abbildung zwischen zwei Vektorräumen V und W und sei $U \subseteq V$ ein Untervektorraum mit $U \subseteq \mathrm{Ker}(T)$. Dann existiert eine eindeutige lineare Abbildung $\widetilde{T} : V/U \to W$, so dass folgendes Diagramm kommutiert:*

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "\pi_U"'] & W \\ V/U \arrow[ur, "\widetilde{T}"'] & \end{tikzcd} .$$

Oder in anderen Worten, so dass

$$T = \widetilde{T} \circ \pi_U \tag{3.32}$$

Beweis-Skizze. Damit (3.32) gilt, muss man $\widetilde{T}$ folgendermassen definieren:

$$\widetilde{T}(v + U) = T(v). \tag{3.33}$$

Man zeigt nun, dass $\widetilde{T}$ wohl-definiert ist (hier benutzt man $U \subseteq \mathrm{Ker}(T)$), und dass $\widetilde{T}$ linear ist. Dies zeigt die Existenz. Für die Eindeutigkeit benutzen wir, dass wir keine andere Wahl für $\widetilde{T}$ haben ausser (3.33), so dass (3.32) gilt. □

Bemerkung 3.7.21*.* Im Bereich der abstrakten Algebra, versucht man Objekte mittels solcher universellen Eigenschaften zu definieren. Dies kann ein gutes Thema für die „Nachbesprechungszeit“ sein.

Changelog: Kapitel 3

- 09.02: Im Beweis von Korollar 3.2.14 wurde der Verweis auf Lemma 2.2.6 durch einen Verweis auf Lemma 3.2.8 ersetzt.
- 19.02: Nummerierungen wurden geändert.
- 22.02: Definition 3.3.27 und Lemma 3.3.28 wurden hinzugefügt.
- 10.03: In der Summe von Gleichung (3.10) wurde der Index k zu i geändert.

Kapitel 4

Determinanten

Wie bereits erwähnt ist eines der Hauptziele dieses Kapitels jede quadratische Matrix $A \in M_{n\times n}(K)$ mit einem Skalar $\det A \in K$ zu versehen, so dass die Bedingung

$$\det A \neq 0$$

äquivalent zu allen anderen Bedingungen in Proposition 3.6.8 ist. Wir werden aber sehen, dass der Skalar $\det A$ viele andere Anwendungen hat. Um diese einzuführen machen wir zuerst ein bisschen Magie mit 2×2-Matrizen.

4.1 Magie in K^2 mit 2×2-Matrizen

Sei $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_{2\times 2}(K)$. Wir fragen uns, ob A invertierbar ist. Wir könnten das Gauss-Jordan Verfahren[1] benutzen (Satz 3.6.20), um das Folgende zu sehen:

Proposition 4.1.1. *Die Matrix A ist invertierbar genau dann, wenn $ad - bc \neq 0$. Falls sie invertierbar ist, dann gilt*

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Wir schreiben $\det A = ad - bc$*, und dann haben wir unser Hauptziel für dieses Kapitel für $n = 2$ schon erreicht.*

Wir könnten uns fragen, was für Eigenschaften die Zahl $ad - bc$ hat. Die Tatsache, dass sie Invertierbarkeit einer Matrix A charakterisiert, impliziert „starke“ Eigenschaften dieser Zahl. Wenn wir uns das geometrisch überlegen (zum Beispiel für $K = \mathbb{R}$), sehen wir, dass $\det A = 0$ genau dann, wenn die Abbildung m_A nicht surjektiv ist,

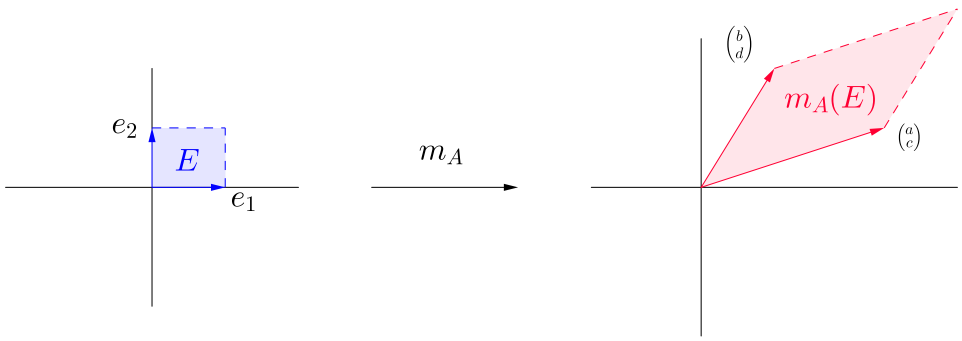
[1] Sie werden genau das in der Übungsstunde besprechen.

was genau dann gilt, wenn das Bild von m_A eine Gerade in $\mathbb{R}^2$ ist (oder sogar nur der Ursprung, falls $A = 0$ ist). Falls wir den „geometrischen Effekt“ der Abbildung m_A für Diagonalmatrizen wie $B = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$ betrachten, sehen wir andererseits, dass m_B den Flächeninhalt von Bereichen $\mathbf{B} \subseteq \mathbb{R}^2$ mit einem Faktor $\det B = 2 \cdot 3$ skaliert. Dies stimmt immer noch für allgemeine Diagonalmatrizen $\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ mit $a, b > 0$ oder $\begin{pmatrix} -a & 0 \\ 0 & -b \end{pmatrix}$ mit $a, b > 0$. Für Matrizen der Form

$$\begin{pmatrix} -2 & 0 \\ 0 & 3 \end{pmatrix}$$

ist der Skalierungsfaktor $|\det A| = |-2 \cdot 3| = 6$. Wir könnten uns auch fragen, wie das Vorzeichen der Determinante die Geometrie beeinflusst. Bevor wir das machen, kann man mit diesem Applet spielen, um zu sehen, dass die Skalierungseigenschaft nicht nur für Diagonalmatrizen stimmt (oder für Matrizen mit Determinante 0)[2], sondern auch für allgemeine Matrizen. Das ruft nach einem Beweis!

Proposition 4.1.2. *Sei E das Quadrat aufgespannt von e_1 und e_2 in $\mathbb{R}^2$. Das Bild $m_A(E)$ von E unter A ist das Parallelogramm aufgespannt von $m_A(e_1) = \begin{pmatrix} a \\ c \end{pmatrix}$ und $m_A(e_2) = \begin{pmatrix} b \\ d \end{pmatrix}$.*

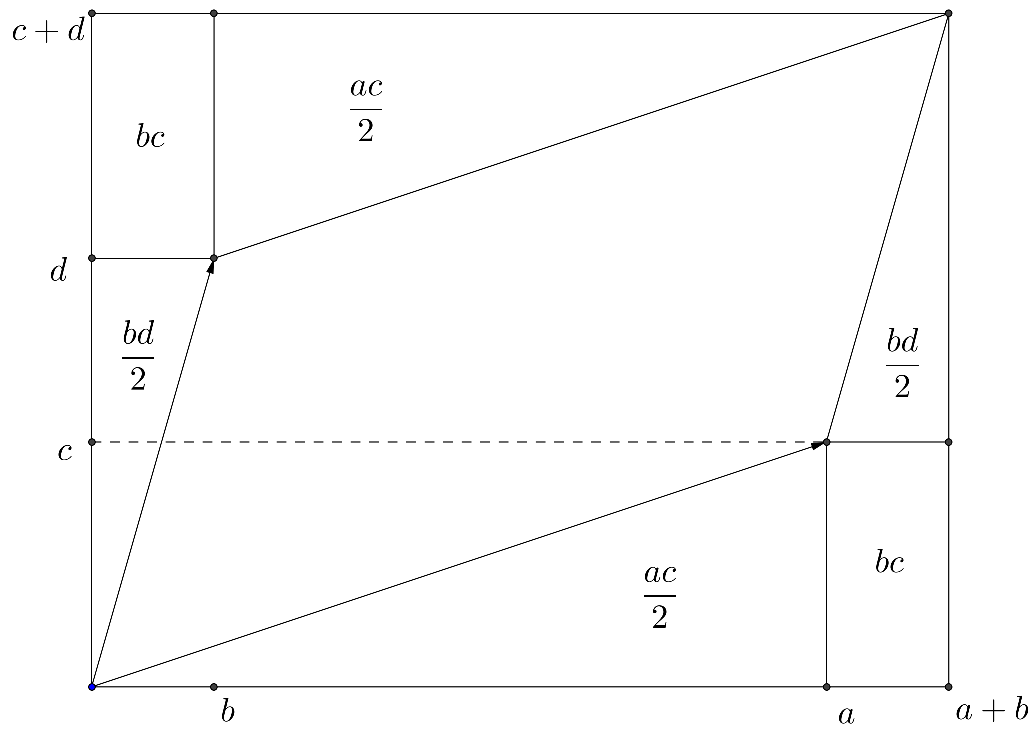


Dann ist der Flächeninhalt von $m_A(E)$ gegeben durch

$$\textit{Flächeninhalt}(m_A(E)) = |ad - bc| = |\det A| \,.$$

[2]Wie wir vorher gesehen haben, „quetscht“ eine Matrix mit Determinante 0 den ganz Raum $\mathbb{R}^2$ zu einem Unterraum kleinerer Dimension zusammen. Also ist der Skalierungsfaktor 0, wie die Determinante der Matrix.

Beweis. Wir betrachten



Figur 4.1: Berechnung des Flächeninhaltes.

und berechnen

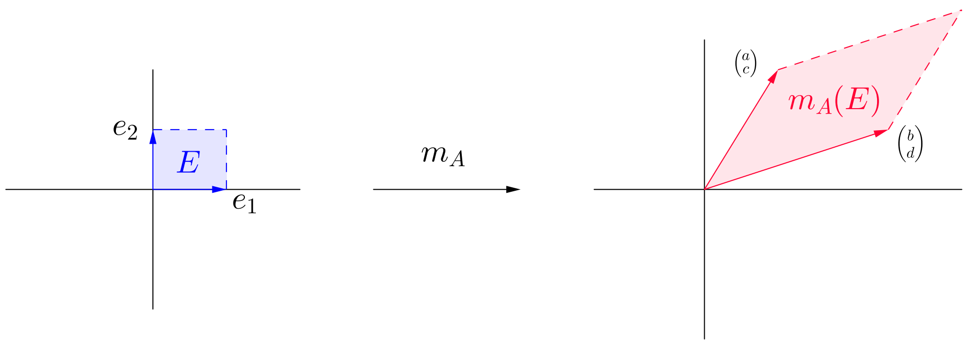
$$\begin{aligned}(a+b)(c+d) - \left(bc + bc + \frac{bd}{2} + \frac{bd}{2} + \frac{ac}{2} + \frac{ac}{2}\right) &= ac + ad + bc + bd - 2bc - bd - ac \\ &= ad - bc \\ &= \det A.\end{aligned}$$

□

Bemerkung 4.1.3. Sie könnten sich beschweren, dass wir eigentlich gezeigt haben, dass der Skalierungsfaktor $\det A$ und nicht $|\det A|$ ist und Sie würden sich zu Recht beschweren. Der Grund dafür ist, dass es eine versteckte Annahme in unserem Bild 4.1 gibt. Die Annahme ist, dass das Paar von Vektoren

$$m_A(e_1) = \begin{pmatrix} a \\ c \end{pmatrix} \quad \text{und} \quad m_A(e_2) = \begin{pmatrix} b \\ d \end{pmatrix}$$

die gleiche „Orientierung“ wie e_1 und e_2 haben. Was meinen wir damit? Beachten Sie, dass e_2 „links“ von e_1 liegt und wir $m_A(e_2)$ auch links von $m_A(e_1)$ gezeichnet haben. Wenn das der Fall ist, dann sagen wir, dass A *orientierungserhaltend* ist. Falls wir das Bild so gezeichnet hätten:



Dann hätten wir

$$\text{Flächeninhalt}(m_A(E)) = bc - ad = |\det A|$$

erhalten, was bedeuten würde, dass die Determinante in diesem Fall negativ ist. In diesem Fall sagt man, dass m_A *orientierungsumkehrend* ist.

Fazit:

- Die Abbildung m_A ist orientierungserhaltend genau dann, wenn $\det A > 0$.
- Die Abbildung m_A ist orientierungsumkehrend genau dann, wenn $\det A < 0$.
- In beiden Fällen skaliert m_A den Flächeninhalt mit dem Faktor $|\det A|$.

Bemerkung 4.1.4. Wiederum könnten Sie argumentieren, dass wir Proposition 4.1.2 nicht für einen beliebigen Bereich in $\mathbb{R}^2$ kennen, da wir nur die Proposition nur für das von e_1 und e_2 aufgespannten Quadrat gezeigt haben. Und auch Letzteres haben wir durch ein Bild gezeigt, welches viele bestimmte Wahlen beinhaltete. Sie hätten recht, aber was wir gezeigt haben ist nicht weit entfernt von einem kompletten Beweis: Man muss einen allgemeinen Bereich in kleine Quadrate teilen (was Sie wahrscheinlich schon viele Male in der Analysis gemacht haben) und dann den Beweis, den wir gegeben haben, so modifizieren, dass er für beliebige (kleine) Quadrate gilt.

Die obige geometrische Interpretation der Determinante wird in allen Dimensionen wahr sein.

4.2 Volumen-Funktionen und die abstrakte Definition der Determinante

Wir werden die folgende Notation während dieses Abschnitts verwenden:

Sei $A \in M_{n\times n}(K)$ eine Matrix. Wir schreiben $A_{(1)}, \ldots, A_{(n)}$ für die Zeilen von A und $A^{(1)}, \ldots, A^{(n)}$ für die Spalten von A. Wir schreiben

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in M_{n\times n}(K)$$

für die Matrix, deren Zeilen $v_1, \ldots, v_n$ sind.

Man kann die Determinante auf zumindest zwei Arten explizit definieren:

(1) Mittels Permutationen (nach Leibniz, welcher die Determinante zuerst für 3×3-Matrizen im Jahre 1683 in einem Brief an L'Hôpital definierte).

(2) Induktiv unter der Benutzung des Entwicklungssatz von Laplace, welchen wir später sehen werden.

Die eleganteste Art (jedoch mit dem Nachteil, dass sie abstrakt ist[3]) die Determinante einzuführen, ist es der Charakterisierung von Weierstrass (aus dem Jahre 1903) zu folgen und drei einfache Axiome zu benutzen:

Definition 4.2.1 (Definition und Satz)**.** Es existiert eine eindeutige Funktion

$$\det : M_{n\times n}(K) \to K,$$

die *Determinante* heisst und die folgenden Eigenschaften[4] erfüllt:[5]

(D1) Die Abbildung det ist *multilinear* bezüglich Zeilen[6] Dies bedeutet:

(a) Für alle $v_1, \ldots, v_n, w \in K^n_{\text{Zeil}}$ und alle $1 \le i \le n$ gilt

$$\det \begin{pmatrix} v_1 \\ \vdots \\ v_{i-1} \\ v_i + w \\ v_{i+1} \\ \vdots \\ v_n \end{pmatrix} = \det \begin{pmatrix} v_1 \\ \vdots \\ v_{i-1} \\ v_i \\ v_{i+1} \\ \vdots \\ v_n \end{pmatrix} + \det \begin{pmatrix} v_1 \\ \vdots \\ v_{i-1} \\ w \\ v_{i+1} \\ \vdots \\ v_n \end{pmatrix}.$$

[3]Wenn man darauf besteht eine explizite (nicht abstrakte) Definition der Determinanten anzugeben mittels Permutationen, dann muss man ein Meister im Umgang mit Indizes werden. (Siehe den wunderbaren Beweis von $\det AB = \det A \det B$ in [10, S. 229].)

[4]Diese Eigenschaften nennt man die *Axiome der Determinante*.

[5]Wir wählen die Notation (D1)-(D3), so dass Sie problemlos den entsprechenden Stoff im Fischer [7] nachlesen können.

[6]Genau so gut könnten wir dies bezüglich Spalten definieren und würden dieselbe Determinante erhalten.

(b) Für alle $a \in K$, alle $v_1, \ldots, v_n \in K^n_{\text{Zeil}}$ und alle $1 \leq i \leq n$ gilt

$$\det \begin{pmatrix} v_1 \\ \vdots \\ v_{i-1} \\ av_i \\ v_{i+1} \\ \vdots \\ v_n \end{pmatrix} = a \det \begin{pmatrix} v_1 \\ \vdots \\ v_{i-1} \\ v_i \\ v_{i+1} \\ \vdots \\ v_n \end{pmatrix}.$$

(D2) Die Abbildung det ist *alternierend.* Das heisst, falls A zwei gleiche Zeilen hat, dann ist $\det A = 0$.

(D3) Die Abbildung $\det A$ ist *normiert.* Das heisst, dass $\det I_n = \det \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = 1$.

Für

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \in M_{n \times n}(K)$$

schreibt man häufig $|A| := \det A$ oder

$$\begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} = \det \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

Wir werden den Existenz-Teil dieses Satzes später beweisen. Jetzt beweisen wir zunächst weitere Eigenschaften der Determinante, die aus den drei Axiomen (D1)-(D3) folgen, und zum Schluss folgt auch die Eindeutigkeit der Abbildung det.

Dies ist ziemlich cool: Wir wissen noch nicht, ob solche Objekte überhaupt existieren, werden jetzt aber weitere Eigenschaften von diesem hypothetischen Objekt finden bis zu dem Punkt, wo wir fast schon wissen wie man zeigen kann, dass dieses Objekt überhaupt existiert.

In diesem ganzen Abschnitt bezeichnet det eine Funktion, die (D1)-(D3) in Definition 4.2.1 erfüllt. Das heisst, man könnte zu jeder Aussage in diesem Abschnitt den Satz „Sei det eine Funktion, die die Definition 4.2.1 erfüllt" hinzufügen.

Proposition 4.2.2. *Es gilt:*

(D4) $\det(\lambda A) = \lambda^n \det A$.

(D5) Falls A eine Nullzeile hat, dann ist $\det A = 0$.

Beweis. Beide folgen aus (D1) (b) (Wieso?). Wir geben jedoch noch einen ausführlichen Beweis:

(D4) Mit mehrfacher Anwendung von (D1) (b) folgt, dass

$$\det \lambda A = \det \begin{pmatrix} \lambda A_{(1)} \\ \vdots \\ \lambda A_{(n)} \end{pmatrix} \overset{(D1)(b)}{=} \lambda^n \det \begin{pmatrix} A_{(1)} \\ \vdots \\ A_{(n)} \end{pmatrix} = \lambda^n \det A.$$

(D5) Falls die Nullzeile $A_{(i_0)}$ ist, dann gilt $0 \cdot A_{(i_0)} = A_{(i_0)}$ und daher folgt

$$\det \begin{pmatrix} A_{(1)} \\ \vdots \\ A_{(i_0)} \\ \vdots \\ A_{(n)} \end{pmatrix} = \det \begin{pmatrix} A_{(1)} \\ \vdots \\ 0 \cdot A_{(i_0)} \\ \vdots \\ A_{(n)} \end{pmatrix} \overset{(D1)(b)}{=} 0 \det \begin{pmatrix} A_{(1)} \\ \vdots \\ A_{(i_0)} \\ \vdots \\ A_{(n)} \end{pmatrix} = 0.$$

□

Wir würden gerne den Effekt der Gauss'schen Elimination, also Zeilenumformungen, auf die Determinante verstehen. Da wir bereits den Zusammenhang von Zeilenumformungen und Elementarmatrizen verstehen, werden wir diesen auch in der Formulierung folgender Proposition verwenden:

Proposition 4.2.3 (Gauss und die Determinante[7]). *Es gelten folgende weitere Eigenschaften der Determinante:*

(D1b) Sei $\lambda \in K$ und $A \xrightarrow{\lambda L_i \to L_i} B$. Dann ist $\det B = \lambda \det A$. *Mit Elementarmatrizen*[8] *ausgedrückt heisst das, dass*

$$\det (S_i(\lambda) A) = \lambda \det A.$$

(D6) Falls $A \xrightarrow{L_i \leftrightarrow L_j} B$, dann ist $\det A = -\det B$. *Mit Elementarmatrizen ausgedrückt heisst das, dass*

$$\det (P_{i,j} A) = -\det A.$$

[7] Hinter dieser Überschrift steckt eigentlich ein Wortwitz: Gauss gab ursprünglich der Determinante ihren Namen, aber in einem anderen Kontext. Er studierte quadratische Formen in zwei Variablen, welche er mit einer Zahl versehen hat, die ihre arithmetischen Eigenschaften determiniert (bestimmt). Also hat er diese Zahl Determinante genannt.

[8] Obwohl $S_i(\lambda)$ nur dann eine Elementarmatrix ist, wenn $\lambda \neq 0$.

(D7) Sei $\lambda \in K$ *und* $A \xrightarrow{L_i+\lambda L_j \to L_i} B$, *wobei* $i \neq j$. *Dann ist* $\det A = \det B$. *Mit Elementarmatrizen ausgedrückt heisst das, dass*

$$\det(Q_{i,j}(\lambda)A) = \det A.$$

Beweis. Wie die Nummerierung andeutet, ist die erste Aussage äquivalent zu (D1)(b).

Für (D6) nehmen wir ohne Beschränkung der Allgmeinheit an, dass $i < j$ ist und betrachten die Matrix

$$B = \begin{pmatrix} A_{(1)} \\ \vdots \\ A_{(i)} + A_{(j)} \\ \vdots \\ A_{(i)} + A_{(j)} \\ \vdots \\ A_{(n)} \end{pmatrix},$$

wobei der Zeilenvektor $A_{(i)} + A_{(j)}$ in der i-ten beziehungsweise j-ten Zeile steht. Dann gilt

$$
\begin{aligned}
0 \overset{(D2)}{=} \det B &\overset{(D1)(a)}{=} \det \begin{pmatrix} A_{(1)} \\ \vdots \\ A_{(i)} \\ \vdots \\ A_{(i)} + A_{(j)} \\ \vdots \\ A_{(n)} \end{pmatrix} + \det \begin{pmatrix} A_{(1)} \\ \vdots \\ A_{(j)} \\ \vdots \\ A_{(i)} + A_{(j)} \\ \vdots \\ A_{(n)} \end{pmatrix} \\
&\overset{(D1)(a)}{=} \det \begin{pmatrix} A_{(1)} \\ \vdots \\ A_{(i)} \\ \vdots \\ A_{(i)} \\ \vdots \\ A_{(n)} \end{pmatrix} + \det \begin{pmatrix} A_{(1)} \\ \vdots \\ A_{(i)} \\ \vdots \\ A_{(j)} \\ \vdots \\ A_{(n)} \end{pmatrix} + \det \begin{pmatrix} A_{(1)} \\ \vdots \\ A_{(j)} \\ \vdots \\ A_{(i)} \\ \vdots \\ A_{(n)} \end{pmatrix} + \det \begin{pmatrix} A_{(1)} \\ \vdots \\ A_{(j)} \\ \vdots \\ A_{(j)} \\ \vdots \\ A_{(n)} \end{pmatrix} \\
&\overset{(D2)}{=} \det \begin{pmatrix} A_{(1)} \\ \vdots \\ A_{(i)} \\ \vdots \\ A_{(j)} \\ \vdots \\ A_{(n)} \end{pmatrix} + \det \begin{pmatrix} A_{(1)} \\ \vdots \\ A_{(j)} \\ \vdots \\ A_{(i)} \\ \vdots \\ A_{(n)} \end{pmatrix}
\end{aligned}
$$

Dies zeigt (D6).

Für (D7) benutzen wir (D1):

$$
\det B = \det \begin{pmatrix} A_{(1)} \\ \vdots \\ A_{(i-1)} \\ A_{(i)} + \lambda A_{(j)} \\ A_{(i+1)} \\ \vdots \\ A_{(n)} \end{pmatrix} = \det \begin{pmatrix} A_{(1)} \\ \vdots \\ A_{(i-1)} \\ A_{(i)} \\ A_{(i+1)} \\ \vdots \\ A_{(n)} \end{pmatrix} + \lambda \det \underbrace{\begin{pmatrix} A_{(1)} \\ \vdots \\ A_{(i-1)} \\ A_{(j)} \\ A_{(i+1)} \\ \vdots \\ A_{(n)} \end{pmatrix}}_{=:C} = \det A + \lambda \cdot 0 = \det A,
$$

wobei wir benutzt haben, dass C zwei gleiche Zeilen ($i \neq j$) hat. Also ist $\det C = 0$ nach (D2). □

Bemerkung 4.2.4. Bemerken Sie, dass (D6) eigentlich äquivalent ist zu (D2) (im Fall $\mathrm{char}(K) \neq 2$). Das erklärt den Begriff „alternierend“.

Nach dieser Proposition sollte der Leser sagen: „Cool!“ Was ich im „Endgame“ machen muss, ist alles, was ich jetzt wissen muss. Also was die Determinante einer Matrix in Stufenform ist. Ein Fall ist bereits klar: Falls die Stufenform eine Nullzeile hat, dann ist die Determinante 0 nach (D5).

Proposition 4.2.5. *Für eine obere Dreiecksmatrix*

$$A = \begin{pmatrix} a_1 & * & \cdots & * \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \cdots & 0 & a_n \end{pmatrix}.$$

gilt $\det A = a_1 \cdots a_n$. *Dieselbe Aussage gilt für untere Dreiecksmatrizen*

$$\det \begin{pmatrix} a_1 & 0 & \cdots & 0 \\ * & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ * & \cdots & * & a_n \end{pmatrix} = a_1 \cdots a_n.$$

Beweis. Fall $a_i \neq 0$ ist für alle $i \in \{1, \cdots, n\}$, können wir „nach oben reinigen“ und verändern damit laut (D7) die Determinante nicht. Daher gilt

$$\det A \overset{(D7)}{=} \det \begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_n \end{pmatrix} \overset{(D1b)}{=} a_1 \cdots a_n \det(I_n) \overset{(D3)}{=} a_1 \cdots a_n.$$

Falls ein $i \in \{1, \cdots, n\}$ mit $a_i = 0$ existiert, dann sei $i_0 = \max\{i \mid a_i = 0\}$. Wir können dann „nach oben reinigen“ mit $a_{i_0+1}, \ldots, a_n$, die alle verschieden von Null sind, ohne die Determinante zu ändern. Dabei ist dann die i_0-Zeile eine Nullzeile, und aus (D5) folgt, dass

$$\det A = 0 = a_1 \cdots a_n.$$

Derselbe Beweis, aber mit „nach unten reinigen“ und mit min statt max zeigt die zweite Aussage bezüglich unterer Dreiecksmatrizen. □

Korollar 4.2.6. *Es gilt*

$$\det A \neq 0 \iff \operatorname{Rang}(A) = n \iff A \text{ ist invertierbar.}$$

Beweis. Die zweite Äquivalenz haben wir bereits in Proposition 3.6.8 gezeigt. Für die erste sei B eine Matrix in Stufenform, die zeilenäquivalent zu A ist. Laut Proposition 4.2.3 gilt

$$\det A \neq 0 \iff \det B \neq 0.$$

Falls $\operatorname{Rang}(B) < n$ ist, dann hat B eine Nullzeile[9]. Nach (D5) gilt dann $\det B = 0$. Falls $\operatorname{Rang}(B) = n$ ist, so ist B eine obere Dreiecksmatrix mit von Null verschiedenen Einträgen auf der Hauptdiagonale. Laut Proposition 4.2.5 gilt also $\det B \neq 0$. Dies zeigt $\det B \neq 0 \iff \operatorname{Rang}(B) = n$. Da $\operatorname{Rang}(A) = \operatorname{Rang}(B)$ ist, folgt das Korollar. □

Erinnern Sie sich, dass diese Eigenschaft der Determinante unser Hauptziel war. Das einzige Problem: Wir wissen noch nicht, ob die Determinante überhaupt existiert! Diese Erkenntnis gibt uns aber die Motivation, die Existenz zu zeigen, und ausserdem Methoden zur Berechnung der Determinante zu entwickeln und vielleicht auch Methoden zur Berechnung der Inverse mittels Determinanten zu finden (Beispielsweise Verallgemeinerungen von Proposition 4.1.1 für $n \in \mathbb{N}$ zu finden).

Bevor wir weitere Eigenschaften dieses tollen hypothetischen Objekts zeigen, bemerken wir, dass wir die Determinante mittels Proposition 4.2.3 berechnen können.

Beispiel 4.2.7. Um die Determinante $|A|$ von $A = \begin{pmatrix} 0 & 1 & 2 \\ 3 & -2 & 0 \\ 4 & 1 & 4 \end{pmatrix}$ zu bestimmen, verwenden wir Zeilen-Operationen, um eine Dreiecksmatrix zu bekommen.

$$\overbrace{\begin{pmatrix} 0 & 1 & 2 \\ 3 & -2 & 0 \\ 4 & 1 & 4 \end{pmatrix}}^{=A} \xrightarrow{L_1 \leftrightarrow L_2} \overbrace{\begin{pmatrix} 3 & -2 & 0 \\ 0 & 1 & 2 \\ 4 & 1 & 4 \end{pmatrix}}^{:=A_1} \xrightarrow{L_3 - \frac{4}{3}L_1 \to L_3} \overbrace{\begin{pmatrix} 3 & -2 & 0 \\ 0 & 1 & 2 \\ 0 & \frac{11}{3} & 4 \end{pmatrix}}^{:=A_2}$$

$$\xrightarrow{3L_3 \to L_3} \overbrace{\begin{pmatrix} 3 & -2 & 0 \\ 0 & 1 & 2 \\ 0 & 11 & 12 \end{pmatrix}}^{:=A_3} \xrightarrow{L_3 - 11L_2 \to L_3} \overbrace{\begin{pmatrix} 3 & -2 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & -10 \end{pmatrix}}^{:=A_4}$$

Laut Proposition 4.2.3 gilt

- $|A| = -|A_1|$,

[9] Alternativ ist B eine obere Dreiecksmatrix mit 0 auf der Hauptdiagonale und hat daher Determinante 0.

- $|A_1| = |A_2|$,
- $3\,|A_2| = |A_3|$,
- $|A_3| = |A_4|$,

also $-3\,|A| = |A_4|$. Laut Proposition 4.2.5 ist $|A_4| = 3 \cdot 1 \cdot (-10) = -30$. Daher gilt $|A| = \frac{-30}{-3} = 10$.

Hier ist eine verwirrende Frage: Wie kann es sein, dass wir etwas berechnen können, dessen Existenz wir noch nicht sicherstellen können? Wenn wir es berechnen können, zeigt dies nicht seine Existenz? (Hinweis: Wenn man versucht mittels Proposition 4.2.3 die Determinante zu definieren und dabei die Existenz zu etablieren, gibt es ein kleines Problem... die Wohldefiniertheit!)

Für den Beweis der Eindeutigkeit und um zu zeigen, dass die Determinante multiplikativ ist, wäre es hilfreich die Determinante der Elementarmatrizen zu kennen.

Lemma 4.2.8. *Für jedes $\alpha \in K$ gilt*

(1) $\det(S_i(\alpha)) = \alpha$,

(2) $\det(P_{i,j}) = -1$.

(3) $\det(Q_{i,j}(\alpha)) = 1$

Ausserdem gilt für jede Elementarmatrix T und jede beliebige Matrix B, dass

$$\det(TB) = \det T \det B.$$

Beweis. Wir geben unten einen Beweis, bemerken aber kurz, wie man auch durch zwei andere Argumente argumentieren könnte: (1) und (3) folgen direkt aus Proposition 4.2.5. Ausserdem kann man alle drei Aussagen auch mit Zeilenumformungen zeigen, wie die Leser überprüfen können.

Wir zeigen das Lemma nun aber mit der Elemementarmatrizen-Formulierung von Proposition 4.2.3: Teil (1) folgt aus

$$1 = \det I_n \overset{\text{Lemma 3.6.15}}{=} \det\left(S_i\left(\frac{1}{\alpha}\right) S_i(\alpha)\right) \overset{\text{Prop. 4.2.3}}{=} \frac{1}{\alpha} \det S_i(\alpha).$$

Teil (2) folgt aus

$$1 = \det I_n \overset{\text{Lemma 3.6.15}}{=} \det(P_{i,j} P_{i,j}) \overset{\text{Prop. 4.2.3}}{=} -\det P_{i,j}.$$

Teil (3) folgt aus

$$1 = \det I_n \overset{\text{Lemma 3.6.15}}{=} \det(Q_{i,j}(-\alpha) Q_{i,j}(\alpha)) \overset{\text{Prop. 4.2.3}}{=} \det Q_{i,j}(\alpha).$$

Die letzte Aussage folgt jetzt, da sie mit (1), (2) und (3) genau äquivalent zu der Elementarmatrizen-Formulierung von Proposition 4.2.3 ist.

□

Um weiter zu zeigen, wie toll die Elementarmatrizen sind, benutzen wir sie jetzt, um die folgende grundlegende Eigenschaft der Determinante zu beweisen:

Proposition 4.2.9. *Seien* $A, B \in M_{n\times n}(K)$. *Dann gilt*

$$\det(AB) = \det A \det B.$$

Beweis. Falls $\mathrm{Rang}(A) < n$, dann ist auch $\mathrm{Rang}(AB) < n$ (Wieso? Denken Sie an Matrizen als lineare Abbildungen). Laut Korollar 4.2.6 sind dann beide Seiten der Gleichung Null.

Nehmen wir also an, dass $\mathrm{Rang}(A) = n$ und daher, dass $A \in \mathrm{GL}_n(K)$. Laut Satz 3.6.16 ist A ein endliches Produkt von Elementarmatrizen, also

$$A = T_1 \cdots T_k$$

mit Elementarmatrizen T_i, $1 \le i \le k$. Wir beweisen jetzt (4.2.9) mit Induktion über k. Für $k = 1$ haben wir genau die letzte Aussage in Lemma 4.2.8. Wir nehmen an, dass die Aussage für $k-1$ gilt und betrachten

$$\begin{aligned}\det(AB) = \det(T_1T_2\cdots T_kB) &\overset{\text{Fall } k=1}{=} \det T_1 \det(T_2\cdots T_kB)\\ &\overset{\text{Ind.-Annahme}}{=} \det T_1 \det(T_2\cdots T_k)\det B\\ &\overset{\text{Fall } k=1}{=} \det(T_1T_2\cdots T_k)\det B = \det A \det B.\end{aligned}$$

Also folgt die Proposition. □

Als eine letzte Anwendung von Elementarmatrizen beweisen wir die Eindeutigkeit der Determinante.

Korollar 4.2.10. *Seien* $\det$ *und* $\widetilde{\det}$ *zwei Funktionen, die* $(D1)-(D3)$ *erfüllen. Dann ist* $\det = \widetilde{\det}$.

Beweis. Für A mit $\mathrm{Rang}(A) < n$ muss laut Korollar 4.2.6

$$\det(A) = 0 = \widetilde{\det}(A)$$

gelten. Für A mit $\mathrm{Rang}(A) = n$ existieren Elemenatarmatrizen $T_1, \cdots, T_k$ mit

$$A = T_1 \cdots T_k.$$

Laut Proposition 4.2.9 und Lemma 4.2.8 gilt

$$\begin{aligned} \det A &= \det T_1 \cdots \det T_k \\ &= \widetilde{\det} T_1 \cdots \widetilde{\det} T_k = \widetilde{\det}(A), \end{aligned} \tag{4.1}$$

daher gilt $\det = \widetilde{\det}$. □

Bemerkung 4.2.11. Wir haben willkürlich gewählt, die Theorie der Determinante über Zeilenoperationen zu entwickeln [10]. Im nächsten Kapitel werden wir zeigen, dass $\det(A) = \det(A^T)$ ist, was auch zeigt, dass die analogen Aussagen aller obigen Aussagen auch für Spalten gelten. Die Leser sollten sich daher merken, dass es äussert nützlich ist, sowohl Zeilen- als auch Spaltenumformungen in der Berechnung einer Determinante zu benutzen.

4.3 Existenz der Determinante

Es gibt zumindest zwei Wege die Existenz der Determinante zu zeigen. Einer davon ist es, eine induktive Definition zu benutzen mittels dem Entwicklungssatz von Laplace, welchen wir später sehen werden (vgl. Prop. 4.5 in [12]). Der zweite Weg benutzt Permutationen. Da induktive Definitionen nie erleuchtend sind, gehen wir den zweiten Weg.

Um diesen Text nicht noch länger zu machen, erklären wir jetzt nicht, wieso uns Permutationen eigentlich aufgezwungen werden durch die vorherigen abstrakten Definitionen (der Determinante), wenn wir die Existenz der Determinante zeigen wollen. Die interessierten Leser können den Dozenten in der Nachbesprechungszeit fragen und versuchen folgende Übung zu lösen:

Übung 4.3.1. Zeigen Sie nur unter Verwendung von ((D1)-(D3) und vielleicht (D6)), dass

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

die einzige Funktion ist, die (D1)-(D3) erfüllt.

Hinweis: Für eine Funktion Φ, die (D1)-(D3) erfüllt, betrachten Sie $\Phi\left(\begin{pmatrix} ae_1 + be_2 \\ ce_1 + de_2 \end{pmatrix}\right)$.

4.3.1 Permutationen

Sei $S_n = \mathrm{Abb}(\{1, \ldots, n\})$. Das heisst, S_n ist die Gruppe aller bijektiven Funktionen von $\{1, \ldots, n\}$ auf sich selbst mit der Verkettung als Verknüpfung. Die Elemente von

[10] Eigentlich nicht ganz willkürlich: so kann man diesen Stoff auch in Fischer nachlesen.

S_n heissen *Permutationen* (von $\{1, \dots, n\}$). Wir benutzen die folgende Schreibweise: Für $\sigma \in S_n$ schreiben wir

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{pmatrix}. \tag{4.2}$$

Per Definition der Multiplikation als Verkettung gilt für alle $\sigma, \tau \in S_n$

$$\begin{aligned} \tau \cdot \sigma &= \begin{pmatrix} 1 & 2 & \cdots & n \\ \tau(1) & \tau(2) & \cdots & \tau(n) \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & \cdots & n \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{pmatrix} \\ &= \begin{pmatrix} 1 & 2 & \cdots & n \\ \tau\sigma(1) & \tau\sigma(2) & \cdots & \tau\sigma(n) \end{pmatrix}. \end{aligned}$$

Erinnern Sie sich daran, dass die Permutation

$$\mathrm{Id} := \begin{pmatrix} 1 & 2 & \cdots & n \\ 1 & 2 & \cdots & n \end{pmatrix}$$

das neutrale Element von S_n ist, und dass die Inverse σ^{-1} die Umkehrfunktion von σ ist. Einige Elemente spielen eine spezielle Rolle:

Definition 4.3.2. Sei $n \in \mathbb{N}$. Wir definieren das Element $\sigma_{i,j} \in S_n$ als die Permutation, die i und j vertauscht, und sonst nichts ändert. Das heisst

$$\sigma_{i,j}(k) := \begin{cases} j, & \text{falls } k = i \\ i, & \text{falls } k = j \\ k, & \text{sonst} \end{cases}.$$

Ein Element von S_n, das von der Form $\sigma_{i,j}$ ist, nennen wir eine *Transposition.*

Lemma 4.3.3. *Sei $\tau \in S_n$ mit $\tau(1) = i$ und $\tau(2) = j$. Dann gilt $\tau\sigma_{1,2}\tau^{-1} = \sigma_{i,j}$.*

Beweis. Wir berechnen: Für $k \notin \{i, j\}$ ist $\tau^{-1}(k) \notin \{1, 2\}$ und daher ist

$$\sigma_{1,2}(\tau^{-1}(k)) = \tau^{-1}(k)$$

beziehungsweise $\tau(\sigma_{1,2}(\tau^{-1}(k))) = \tau(\tau^{-1}(k)) = k$. Für i und j gilt

$$\begin{aligned} \tau\sigma_{1,2}\tau^{-1}(i) &= \tau(\sigma_{1,2}(1)) = \tau(2) = j \\ \tau\sigma_{1,2}\tau^{-1}(j) &= \tau(\sigma_{1,2}(2)) = \tau(1) = i \end{aligned}$$

und das Lemma folgt. $\square$

Beispiel 4.3.4. Die Gruppe S_1 ist ziemlich langweilig. Sie hat nur ein Element $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

Beispiel 4.3.5. Die Gruppe S_2 ist nicht viel interessanter, hat aber zumindest eine Transposition

$$S_2 = \left\{ \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \right\} = \left\{ \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}, \sigma_{1,2} \right\}$$

Beispiel 4.3.6. Jede Transposition hat Ordnung zwei. Das heisst, für jedes $\sigma_{i,j} \in S_n$ gilt

$$(\sigma_{i,j})^2 := \sigma_{i,j} \cdot \sigma_{i,j} = \sigma_{i,j} \circ \sigma_{i,j} = \begin{pmatrix} 1 & 2 & \cdots & n \\ 1 & 2 & \cdots & n \end{pmatrix}.$$

Dies bedeutet auch, dass $(\sigma_{i,j})^{-1} = \sigma_{i,j}$. Bemerken Sie ausserdem, dass $\sigma_{i,j} = \sigma_{j,i}$.

Beispiel 4.3.7. Die Gruppe S_3 ist schon interessanter. Wir schreiben alle ihre Elemente auf:

$$\begin{aligned}
\mathrm{Id} &:= \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} \\
\sigma &:= \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = \sigma_{1,3} \cdot \sigma_{1,2} = \sigma_{1,2} \cdot \sigma_{2,3} \qquad (\star) \\
\sigma \cdot \sigma &= \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} = \sigma_{1,2} \cdot \sigma_{1,3} \\
\tau &:= \sigma_{1,2} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} = \sigma_{2,1} \\
\tau \cdot \sigma &= \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} = \sigma_{2,3} = \sigma_{3,2} \\
\sigma \cdot \tau &= \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} = \sigma_{3,1} = \sigma_{1,3}.
\end{aligned}$$

Es gibt keine anderen Elemente in S_3 (Wieso?). Ausserdem kann man sich als Beispiel merken, dass

$$\sigma\tau\sigma^{-1} = \sigma\sigma_{1,2}\sigma^{-1} = \sigma_{2,3}$$

aus Lemma 4.3.3 folgt, was man auch direkt berechnen kann.

Hier sind zwei Aussagen, die aus dem Obigen folgen:

- S_3 ist nicht-abelsch, da $\tau\sigma \neq \sigma\tau$ ist.
- Jedes Element von S_3 ist entweder eine Transposition oder ein Produkt von Transpositionen (aber nicht auf eine eindeutige Weise wie zum Beispiel ($\star$) zeigt).

Die beiden letzten Aussagen werden wir verallgemeinern. Die erste zu verallgemeinern ist sehr einfach:

Übung 4.3.8. Benutzen Sie, dass S_3 nicht-abelsch ist, um zu zeigen, dass S_n für $n \geq 3$ nicht-abelsch ist.

4.3.2 Signum einer Permutation

Definition 4.3.9. Sei $\sigma \in S_n$. Ein *Fehlstand*[11] von σ ist ein Paar $i < j$ für $i, j \in \{1, \ldots, n\}$, so dass

$$\sigma(i) > \sigma(j).$$

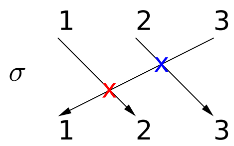
Beispiel 4.3.10. Falls $n = 3$ ist, gibt es drei mögliche Paare mit $i < j$ in $\{1, 2, 3\}$, nämlich $\{(1, 2), (1, 3), (2, 3)\}$. Zwei davon sind Fehlstände für die Permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}:$$

- Für $1 < 2$ ist $\sigma(1) = 2 < 3 = \sigma(2)$. Also ist $(1, 2)$ kein Fehlstand.
- Für $1 < 3$ ist $\sigma(1) = 2 > 1 = \sigma(3)$. Also ist $(1, 3)$ ein Fehlstand.
- Für $2 < 3$ ist $\sigma(2) = 3 > \sigma(3) = 1$. Also ist $(2, 3)$ ein Fehlstand.

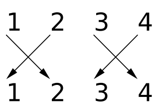
Man sagt: $\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$ hat zwei Fehlstände.

Dies könnte man versuchen zu veranschaulichen:



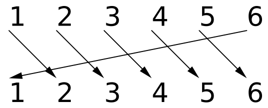
Das blaue Kreuz entspricht dem Fehlstand $(2, 3)$ und das rote Kreuz entspricht dem Fehlstand $(1, 3)$. Man überprüft, wie viele Überkreuzungen es gibt. Jede Überkreuzung entspricht einem Fehlstand.

Beispiel 4.3.11. (1) Die Permutation $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}$ hat zwei Fehlstände:



[11] Inversion auf Englisch.

(2) Die Permutation $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 4 & 5 & 6 & 1 \end{pmatrix}$ hat fünf Fehlstände:



(3) Ähnlich hat die Permutation $\begin{pmatrix} 1 & 2 & \cdots & n-1 & n \\ 2 & 3 & \cdots & n & 1 \end{pmatrix}$ genau $n-1$ Fehlstände.

Beispiel 4.3.12. Die Transposition $\sigma_{1,2} \in S_n$ hat nur $(1,2)$ als Fehlstand. (Wieso?)

Definition 4.3.13. Wir definieren das *Signum* einer Permutation $\sigma \in S_n$ als[12]

$$\begin{aligned} \operatorname{sign}(\sigma) &= \begin{cases} +1, & \text{falls } \sigma \text{ eine gerade Anzahl von Fehlständen hat} \\ -1, & \text{falls } \sigma \text{ eine ungerade Anzahl von Fehlständen hat} \end{cases} \\ &= (-1)^k, \end{aligned}$$

wobei k die Anzahl der Fehlstände von σ ist. Eine Permutation σ heisst *gerade*, falls $\operatorname{sign}(\sigma) = +1$ und *ungerade*, falls $\operatorname{sign}(\sigma) = -1$ ist.

Beispiel 4.3.14.
- $\operatorname{sign}\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = (-1)^2 = 1.$
- $\operatorname{sign}(\sigma_{1,2}) = (-1)^1 = -1.$
- $\operatorname{sign}\begin{pmatrix} 1 & 2 & \cdots & n-1 & n \\ 2 & 3 & \cdots & n & 1 \end{pmatrix} = (-1)^{n-1}.$

Wir werden jetzt die folgenden zwei Lemmata zeigen:

Lemma 4.3.15. *Für $\sigma \in S_n$ gilt*

$$\operatorname{sign}(\sigma) = \prod_{i<j} \frac{\sigma(j) - \sigma(i)}{j - i} = \frac{\prod_{i<j}(\sigma(j) - \sigma(i))}{\prod_{i<j}(j-i)}. \tag{4.3}$$

Lemma 4.3.16. *Für alle $\sigma, \tau \in S_n$ gilt* $\operatorname{sign}(\tau\sigma) = \operatorname{sign}(\tau)\operatorname{sign}(\sigma)$. *In anderen Worten ist*

$$\begin{aligned} S_n &\to \mathbb{Z}/2\mathbb{Z} \cong \{1, -1\} \\ \sigma &\mapsto \operatorname{sign}(\sigma), \end{aligned}$$

ein Homomorphismus von S_n nach $\mathbb{Z}/2\mathbb{Z} \cong \{1, -1\}$.

[12]Viele Autoren schreiben sgn statt sign (inklusive dem Dozenten).

Diese Lemmata und ihre zugehörigen Beweise mögen schwierig erscheinen, sind aber eigentlich elementar. Hier sind einige elementare Beobachtungen, die uns im Beweis von Lemma 4.3.15 helfen:

- Sowohl der Nenner als auch der Zähler in (4.3) enthalten alle möglichen Paare. Im Nenner kommen sie immer in der Form $j-i$ vor, was immer eine positive Zahl ist. Im Zähler hingegen könnte σ das Vorzeichen kehren. Insbesondere ist

$$\frac{\prod_{i<j} |\sigma(j)-\sigma(i)|}{\prod_{i<j}(j-i)} = 1, \tag{4.4}$$

  da alle Faktoren sich miteinander kürzen.

- Genauer gesagt, ändert σ das Vorzeichen im Zähler genau dann, wenn (i,j) ein Fehlstand ist:

$$\begin{aligned}\frac{\sigma(j)-\sigma(i)}{j-i} > 0 &\iff (i,j) \text{ ist kein Fehlstand von } \sigma\\ \frac{\sigma(j)-\sigma(i)}{j-i} < 0 &\iff (i,j) \text{ ist ein Fehlstand von } \sigma.\end{aligned}$$

Jetzt sind wir bereit für den Beweis von Lemma 4.3.15. Wir erklären zur Vorbereitung zuerst alle Schritte mit Hilfe der Permutation

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}.$$

Beispiel 4.3.17. Wir überprüfen Lemma 4.3.15 für $\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$. Es gilt

$$\{i<j \mid i,j \in \{1,2,3\}\} = \{(1,2),(1,3),(2,3)\}.$$

Daher besagt das Lemma, dass

$$\begin{aligned}\operatorname{sign}(\sigma) &= \frac{\sigma(2)-\sigma(1)}{2-1} \cdot \frac{\sigma(3)-\sigma(1)}{3-1} \cdot \frac{\sigma(3)-\sigma(2)}{3-2}\\ &= \underbrace{\frac{3-2}{2-1}}_{\text{(1,2) ist kein Fehlstand}} \cdot \underbrace{\frac{1-2}{3-1}}_{\text{(1,3) ist ein Fehlstand}} \cdot \underbrace{\frac{1-3}{3-2}}_{\text{(2,3) ist ein Fehlstand}}\\ &\overset{\text{kürzen}}{=} \underbrace{(-1)^2}_{\text{2 ist die Anzahl der Fehlstände}} = (-1)^{\text{Anzahl der Fehlstände}}.\end{aligned}$$

Der allgemeine Beweis ist ganz ähnlich.

Beweis von Lemma 4.3.15. Sei $\sigma \in S_n$. Für $i < j$ gilt:

$$(i,j) \text{ ist ein Fehlstand } \iff \sigma(j) - \sigma(i) < 0.$$

Sei s die Anzahl der Fehlstände von σ. Dann gilt

$$\begin{aligned}\prod_{i<j} \frac{\sigma(j)-\sigma(i)}{j-i} &= \prod_{\substack{i<j,\\ \sigma(i)<\sigma(j)}} \frac{\sigma(j)-\sigma(i)}{j-i} \cdot \left((-1)^s \prod_{\substack{i<j,\\ \sigma(i)>\sigma(j)}} \frac{|\sigma(j)-\sigma(i)|}{j-i} \right) \\ &= (-1)^s \prod_{i<j} \frac{|\sigma(j)-\sigma(i)|}{j-i} = (-1)^s,\end{aligned}$$

wobei die letzte Gleichheit aus (4.4) folgt. □

Für den Beweis von Lemma 4.3.16 benutzen wir Lemma 4.3.15:

Beweis von Lemma 4.3.16. Es gilt

$$\begin{aligned}\operatorname{sign}(\tau\sigma) &\overset{\text{Lemma 4.3.15}}{=} \prod_{i<j} \frac{\tau(\sigma(j)) - \tau(\sigma(i))}{j-i} \\ &= \prod_{i<j} \frac{\tau(\sigma(j)) - \tau(\sigma(i))}{\sigma(j)-\sigma(i)} \prod_{i<j} \frac{\sigma(j)-\sigma(i)}{j-i} \\ &\overset{\text{Lemma 4.3.15}}{=} \underbrace{\prod_{i<j} \frac{\tau(\sigma(j)) - \tau(\sigma(i))}{\sigma(j)-\sigma(i)}}_{(*)} \operatorname{sign}(\sigma).\end{aligned}$$

Es genügt also zu zeigen, dass $(*) = \operatorname{sign}(\tau)$. Machen Sie an diesem Punkt eine Pause und überzeugen Sie sich davon, dass es tatsächlich genügt, $(*) = \operatorname{sign}(\tau)$ zu zeigen.

Es gilt

$$\begin{aligned}(*) &= \prod_{i<j} \frac{\tau(\sigma(j)) - \tau(\sigma(i))}{\sigma(j)-\sigma(i)} \\ &= \prod_{\substack{i<j,\\ \sigma(i)<\sigma(j)}} \frac{\tau(\sigma(j)) - \tau(\sigma(i))}{\sigma(j)-\sigma(i)} \cdot \prod_{\substack{i<j,\\ \sigma(i)>\sigma(j)}} \underbrace{\frac{\tau(\sigma(j)) - \tau(\sigma(i))}{\sigma(j)-\sigma(i)}}_{(\star)}\end{aligned} \tag{4.5}$$

Betrachten Sie nun genau den Term $(\star)$. Was passiert, wenn wir i mit j vertauschen und j mit i? Nichts passiert, da wir sowohl den Nenner und als auch den Zähler einfach mit -1 multiplizieren. Das sagt uns, dass wir die Rollen von i und j einfach vertauschen können im zweiten Produkt, ohne dass sich etwas ändert. Also ist

$$\prod_{\substack{i<j,\\ \sigma(i)>\sigma(j)}} (\star) = \prod_{\substack{j<i,\\ \sigma(j)>\sigma(i)}} (\star).$$

Wenn wir das in (4.5) einsetzen, erhalten wir

$$\begin{aligned}
(*) &= \prod_{\substack{i<j,\\ \sigma(i)<\sigma(j)}} \frac{\tau(\sigma(j))-\tau(\sigma(i))}{\sigma(j)-\sigma(i)} \cdot \prod_{\substack{j<i,\\ \sigma(i)<\sigma(j)}} \frac{\tau(\sigma(j))-\tau(\sigma(i))}{\sigma(j)-\sigma(i)} \\
&= \prod_{\sigma(i)<\sigma(j)} \frac{\tau(\sigma(j))-\tau(\sigma(i))}{\sigma(j)-\sigma(i)} \\
&\overset{\widetilde{i}:=\sigma(i),\, \widetilde{j}:=\sigma(j)}{=} \prod_{\widetilde{i}<\widetilde{j}} \frac{\tau(\widetilde{j})-\tau(\widetilde{i})}{\widetilde{j}-\widetilde{i}} \\
&\overset{\text{Lemma 4.3.15}}{=} \operatorname{sign}(\tau).
\end{aligned}$$

Dies beendet den Beweis des Lemmas. □

Korollar 4.3.18. *Für alle $\sigma \in S_n$ gilt* $\operatorname{sign}(\sigma^{-1}) = \operatorname{sign}(\sigma)$.

Beweis. Nach Lemma 4.3.16 gilt $1 = \operatorname{sign}(\mathrm{Id}) = \operatorname{sign}(\sigma)\operatorname{sign}(\sigma^{-1})$, also ist

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

□

Korollar 4.3.19. *Für jedes $i \neq j$ gilt* $\operatorname{sign}(\sigma_{i,j}) = -1$.

Beweis. Aus Beispiel 4.3.14 wissen wir, dass $\operatorname{sign}(\sigma_{1,2}) = -1$. Sei nun $\tau \in S_n$ mit $\tau(1) = i$ und $\tau(2) = j$. Laut Lemma 4.3.3 wissen wir, dass

$$\tau\sigma_{1,2}\tau^{-1} = \sigma_{i,j}.$$

Daher ist

$$\operatorname{sign}(\sigma_{i,j}) = \operatorname{sign}(\tau)\operatorname{sign}(\sigma_{1,2})\operatorname{sign}(\tau^{-1}) \overset{\text{Korollar 4.3.18}}{=} \operatorname{sign}(\sigma_{1,2}) = -1.$$

Das Korollar folgt. □

Korollar 4.3.20. *Es gilt:*

(1) Die Abbildung $\operatorname{sign}\colon S_n \to \{\pm 1\}$ *ist surjektiv, falls $n \geq 2$.*

(2) Die Menge $A_n := \{\sigma \in S_n \mid \operatorname{sign}(\sigma) = 1\} = \{$alle geraden Permutationen$\}$ ist eine Untergruppe von S_n.

(3) Für jede Transposition[13] $\tau \in S_n$ *gilt*

$$\begin{aligned} A_n\tau &:= \{\sigma\tau \mid \sigma \in A_n\} \\ &= \{\sigma \in S_n \mid \operatorname{sign}(\sigma) = -1\} \\ &= \{\textit{alle ungeraden Permutationen}\} \\ &= S_n \setminus A_n. \end{aligned}$$

(4) Ausserdem gilt $|A_n| = |A_n\tau| = \frac{1}{2}|S_n| = \frac{n!}{2}$.

Beweis. (1) Es gilt zum Beispiel, dass $\operatorname{sign}(\mathrm{Id}) = 1$ und $\operatorname{sign}(\sigma_{1,2}) = -1$. Also ist die Abbildung sign surjektiv.

(2) Da wir diese Aussage nicht brauchen werden, überlassen wir diesen Beweis dem Leser als Übung. Erinnern Sie sich daran, dass A_n eine Untergruppe ist, wenn $\mathrm{Id} \in A_n$ ist, und wenn A_n abgeschlossen ist bezüglich Multiplikation und Inverse.

(3) Sei τ eine Transposition. Bemerken Sie zuerst, dass

$$\begin{aligned} A_n &\to A_n\tau \\ \sigma &\mapsto \sigma\tau \end{aligned} \tag{4.6}$$

eine Bijektion ist, da die Funktion $\eta \mapsto \eta\tau^{-1}$ ihre Umkehrfunktion ist. Wir behaupten des Weiteren, dass $A_n\tau$ die Menge aller ungeraden Permutationen ist:

- Jede Permutation der Form $\sigma\tau$ für $\sigma \in A_n$ erfüllt

$$\operatorname{sign}(\sigma\tau) = \operatorname{sign}(\sigma)\operatorname{sign}(\tau) = +1 \cdot (-1) = -1.$$

- Falls η Signum $\operatorname{sign}(\eta) = -1$ hat, dann ist

$$\operatorname{sign}(\eta\tau^{-1}) = \operatorname{sign}(\eta)\operatorname{sign}(\tau^{-1}) = -1 \cdot (-1) = 1,$$

wobei wir benutzt haben, dass $\operatorname{sign}(\tau^{-1}) = \operatorname{sign}(\tau) = -1$. Also ist $\eta\tau^{-1} \in A_n$ und daher ist

$$\eta = (\eta\tau^{-1})\tau \in A_n\tau.$$

(4) Da (4.6) eine Bijektion ist, folgt $|A_n| = |A_n\tau|$. Des Weiteren folgt, dass

$$n! = |S_n| = |A_n \sqcup A_n\tau| = |A_n| + |A_n\tau|,$$

was (4) impliziert (Wieso die erste Gleichheit gilt, überlassen wir den Lesern).

□

[13] Eigentlich gilt es sogar für jede ungerade Permutation.

Wir sind jetzt bereit, die Existenz der Determinante zu zeigen.

4.3.3 Existenz

Wir haben bereits die Eindeutigkeit der Determinante gezeigt. Wenn man die Eindeutigkeit direkt zeigen möchte (siehe auch Übung 4.3.1 für den Fall $n = 2$), dann führt dies zur Betrachtung des folgenden Ausdrucks:

$$\begin{aligned} \det A = \det \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} &:= \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{1\sigma(1)} \cdot a_{2\sigma(2)} \cdots a_{n\sigma(n)} \\ &= \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \prod_{i=1}^{n} a_{i\sigma(i)}. \end{aligned} \tag{4.7}$$

Das mag auf den ersten Blick vielleicht furchteinflössend wirken. Schauen wir uns also konkrete Formeln für $n = 2, 3$ an. Danach werden wir zeigen, dass Definition die (4.7) tatsächlich (D1)-(D3) erfüllt, was dann auch die Existenz der Determinante zeigt.

Beispiel 4.3.21 ($n = 2$)**.** Es gilt, dass $S_2 = \left\{ \mathrm{Id}, \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \right\}$. Daher ist

$$\begin{aligned} \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} &= \sum_{\sigma \in S_2} \operatorname{sign}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \\ &= \operatorname{sign}(\mathrm{Id}) a_{11} a_{22} + \operatorname{sign}\left(\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \right) a_{12} a_{21} \\ &= a_{11} a_{22} - a_{12} a_{21}. \end{aligned}$$

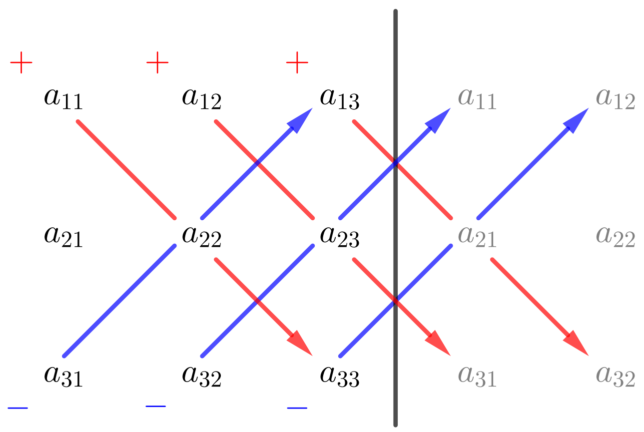
Beispiel 4.3.22. Unter Benutzung von Beispiel 4.3.7 erinnern wir uns an alle Elemente von S_3 und ihre jeweiligen Signums:

$$\begin{aligned} A_3 &= \left\{ \mathrm{Id}, \sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \sigma^2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \right\} \\ S_3 \setminus A_3 &= \left\{ \tau = \mathrm{Id}\,\tau = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \sigma\tau = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}, \sigma^2\tau = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \right\}. \end{aligned}$$

Also ist

$$\begin{aligned}\det\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33}\end{pmatrix} &= \sum_{\eta\in S_3} \operatorname{sign}(\eta) a_{1\eta(1)}a_{2\eta(2)}a_{3\eta(3)} \\ &= \sum_{\eta\in A_3} a_{1\eta(1)}a_{2\eta(2)}a_{3\eta(3)} - \sum_{\eta\in S_3\setminus A_3} a_{1\eta(1)}a_{2\eta(2)}a_{3\eta(3)} \\ &= \underbrace{a_{11}a_{22}a_{33}}_{\eta=\mathrm{Id}} + \underbrace{a_{12}a_{23}a_{31}}_{\eta=\sigma} + \underbrace{a_{13}a_{21}a_{32}}_{\eta=\sigma^2} \\ &\quad - (\underbrace{a_{12}a_{21}a_{33}}_{\eta=\tau} + \underbrace{a_{13}a_{22}a_{31}}_{\eta=\sigma\tau} + \underbrace{a_{11}a_{23}a_{32}}_{\eta=\sigma^2\tau}).\end{aligned}$$

Es gibt eine coole Art und Weise, wie man sich diese Formel merken kann, und zwar mit der *Regel von Sarrus*:



Bemerkung 4.3.23. Nachdem sie diese Regel gesehen haben, googeln viele Studenten nach einer ähnlichen Formel für 4×4-Matrizen. Leider existiert eine solche Formel nicht. Aber Sie müssen keine Angst haben: Wir werden im nächsten Abschnitt den Laplace-Entwicklungssatz sehen, welcher die beste Methode ist, um Determinanten für grosse Matrizen zu berechnen (zumindest wenn man es von Hand macht).

Jetzt wo wir keine Angst mehr haben vor Ausdrücken wie in (4.7), sind wir bereit die Existenz der Determinante zu zeigen.

Satz 4.3.24. *Sei* det *so definiert wie in* (4.7)*. Dann erfüllt* det *die Bedingungen (D1)-(D3) aus Satz 4.2.1.*

Beweis. (D1): Um (D1) (a) zu zeigen, benutzen wir folgende nützliche Notation aus dem Buch von Fischer [7]:

$$\begin{aligned}
\det\begin{pmatrix} \vdots \\ a_i' + a_i'' \\ \vdots \end{pmatrix} &= \sum_{\sigma\in S_n} \operatorname{sign}(\sigma)\cdot a_{1\sigma(1)}\cdot\ldots\cdot(a'_{i\sigma(i)} + a''_{i\sigma(i)})\cdot\ldots\cdot a_{n\sigma(n)} \\
&= \sum_{\sigma\in S_n} \operatorname{sign}(\sigma)\cdot a_{1\sigma(1)}\cdot\ldots\cdot a'_{i\sigma(i)}\cdot\ldots\cdot a_{n\sigma(n)} \\
&+ \sum_{\sigma\in S_n} \operatorname{sign}(\sigma)\cdot a_{1\sigma(1)}\cdot\ldots\cdot a''_{i\sigma(i)}\cdot\ldots\cdot a_{n\sigma(n)} \\
&= \det\begin{pmatrix} \vdots \\ a_i' \\ \vdots \end{pmatrix} + \det\begin{pmatrix} \vdots \\ a_i'' \\ \vdots \end{pmatrix}
\end{aligned}$$

Also gilt (D1) (a).

Wir zeigen nun, dass auch (D1) (b) gilt: Für jedes $1 \leq i \leq n$ und $\lambda \in K$ gilt

$$\begin{aligned}
\det\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ \lambda a_{i1} & \cdots & \lambda a_{in} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} &= \sum_{\sigma\in S_n} \operatorname{sign}(\sigma) a_{1\sigma(1)}\cdot\ldots\cdot\ldots\cdot \lambda a_{i\sigma(i)}\cdot\ldots\cdot a_{n\sigma(n)} \\
&= \lambda\cdot\sum_{\sigma\in S_n} \operatorname{sign}(\sigma) a_{1\sigma(1)}\cdot\ldots\cdot a_{n\sigma(n)} \\
&= \lambda\det A.
\end{aligned}$$

(D2): Fast alles, was wir bis jetzt besprochen haben über S_n werden wir für den Beweis von (D2) verwenden: Angenommen die Matrix A hat zwei gleiche Zeilen $A_{(i)}$ und $A_{(j)}$ mit $i \neq j$. Sei $\tau = \sigma_{i,j}$. Wir müssen zeigen, dass $\det A = 0$. Erinnern Sie sich daran, dass $S_n = A_n \sqcup A_n\tau = \{\text{alle geraden Permutationen}\} \sqcup \{\text{alle ungeraden Permutationen}\}$ (Korollar 4.3.20), wobei $A_n\tau = \{\sigma\tau \mid \sigma \in A_n\}$. Daher gilt

$$\begin{aligned}
\det A &= \underbrace{\sum_{\sigma\in A_n} a_{1\sigma(1)}\cdot\ldots\cdot a_{n\sigma(n)}}_{\text{Summe über } A_n} - \underbrace{\sum_{\sigma\in A_n} a_{1\sigma(\tau(1))}\cdot\ldots\cdot a_{n\sigma(\tau(n))}}_{\text{Summe über } S_n\setminus A_n = A_n\tau} \\
&= \underbrace{\sum_{\sigma\in A_n} a_{1\sigma(1)}\cdot\ldots\cdot a_{n\sigma(n)} - a_{1\sigma(\tau(1))}\cdot\ldots\cdot a_{n\sigma(\tau(n))}}_{=:\diamondsuit}.
\end{aligned}$$

Beachten Sie, dass die linke und rechte Seite (links bzw. rechts des Minus) jedes Summanden von $\diamondsuit$ fast die gleichen Terme haben bis auf gewisse Änderungen, welche von τ verursacht werden.

Erinnern Sie sich jetzt daran, dass $A_{(i)} = A_{(j)}$ gilt, was impliziert, dass

$$a_{ik} = a_{jk} \tag{4.8}$$

für alle $k = 1, \dots, n$ gilt. Dies zeigt, dass das Produkt der Terme, die von τ geändert werden, dasselbe ist:

$$a_{i\sigma(\tau(i))}a_{j\sigma(\tau(j))} \overset{\text{Definition von } \tau}{=} a_{i\sigma(j)}a_{j\sigma(i)} \overset{(4.8)}{=} a_{j\sigma(j)}a_{i(\sigma(i)}. \tag{4.9}$$

Dies zeigt, dass jeder Summand von $\diamondsuit$ gleich 0 ist. Also ist $\det A = 0$.

(D3): Um ein gutes Ende zu haben: Der Beweis von (D3) ist einfach: Für

$$\mathrm{I_n} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

haben wir, dass $a_{ij} = 0$, falls $i \neq j$. Der einzige Summand im Ausdruck (4.7), welcher keinen Faktor der Form a_{ij} für $i \neq j$ enthält, ist der Summand, welcher der Permutation $\mathrm{Id} = \begin{pmatrix} 1 & 2 & \cdots & n \\ 1 & 2 & \cdots & n \end{pmatrix}$ entspricht. Also ist

$$\det I_n = \mathrm{sign}(\mathrm{Id})a_{11} \cdot \ldots \cdot a_{nn} = 1 \cdot 1 \cdots 1 = 1.$$

Also erfüllt det wie in (4.7) die Bedingungen (D1)-(D3). Dies zeigt die Existenz der Determinante. $\square$

Nun da wir eine explizite Beschreibung der Determinante haben, können wir einiges mehr über sie beweisen. Insbesondere ist es nützlich den folgenden Satz zu kennen:

Satz 4.3.25. *Für jede Matrix $A \in M_{n\times n}(K)$ gilt*

$$\det A^T = \det A.$$

Dies zeigt, dass alle Resultate, welche Zeilenumformungen beinhalten auch für Spaltenumformungen gelten.

Beweis. Sei $A = (a_{ij})$. Die Determinante der Transponierten von A berechnet sich dann als

$$\det A^T = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \underbrace{a_{\sigma(1)1} \cdot \ldots \cdot a_{\sigma(n)n}}_{(\star)}.$$

Um $\det A^T$ mit $\det A$ in Verbindung zu bringen, betrachten wir das Produkt $(\star)$ und versuchen die Reihenfolge der Indizes zu ändern. Für jedes $\sigma \in S_n$ und jedes $1 \leq i \leq n$ gilt

$$a_{\sigma(i)i} = a_{\sigma(i)\sigma^{-1}(\sigma(i))}.$$

Daher gilt

$$(\star) = a_{\sigma(1)1} \cdot \ldots \cdot a_{\sigma(n)n} = a_{1\sigma^{-1}(1)} a_{2\sigma^{-1}(2)} \cdot \ldots \cdot a_{n\sigma^{-1}(n)}.$$

Ausserdem gilt $\operatorname{sign}(\sigma) = \operatorname{sign}(\sigma^{-1})$ und daher ist

$$(\star) = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma^{-1}) a_{1\sigma^{-1}(1)} \cdot \ldots \cdot a_{n\sigma^{-1}(n)} = \sum_{\sigma^{-1} \in S_n} \operatorname{sign}(\sigma^{-1}) a_{1\sigma^{-1}(1)} \cdot \ldots \cdot a_{n\sigma^{-1}(n)} = \det A.$$

Die letzte Gleichheit folgt, da die Abbildung $\sigma \mapsto \sigma^{-1}$ eine Bijektion von S_n auf sich selbst ist und deshalb ändern wir bei Summation über σ^{-1} nur die Summationsreihenfolge. Dies beendet den Beweis des Satzes. □

Korollar 4.3.26. *Alle Resultate, welche wir für Zeilen(-Umformungen) hatten, gelten auch für Spalten(-Umformungen). Insbesondere ist das Analogon von Proposition 4.2.3 sehr nützlich.*

Hier ist ein klassisches Resultat, das zeigt, wie man Spalten- und Zeilenoperationen benutzen kann.

Beispiel 4.3.27 (Vandermonde-Determinante)**.** Seien $x_1, x_2, \ldots, x_n \in K$. Wir möchten die *Vandermonde-Determinante*

$$V_{x_1,\ldots,x_n} = \det \begin{pmatrix} 1 & x_1 & \cdots & x_1^{n-1} \\ \vdots & & & \vdots \\ 1 & x_n & \cdots & x_n^{n-1} \end{pmatrix}$$

berechnen. Wieso diese Determinante und ihre Berechnung interessant sind, erfahren Sie in Serie 13. Wir lösen den Fall $n = 2, 3$:

$$\begin{vmatrix} 1 & x_1 \\ 1 & x_2 \end{vmatrix} = x_2 - x_1$$

Seien C_1, C_2, C_3 die Spalten von V_{x_1,x_2,x_3}. Dann ist

$$\begin{aligned}
\begin{vmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{vmatrix} &\overset{C_3 - x_1 C_2 \to C_3}{=} \begin{vmatrix} 1 & x_1 & 0 \\ 1 & x_2 & x_2^2 - x_1 x_2 \\ 1 & x_3 & x_3^2 - x_1 x_3 \end{vmatrix} \\
&\overset{C_2 - x_1 C_1 \to C_2}{=} \begin{vmatrix} 1 & 0 & 0 \\ 1 & x_2 - x_1 & x_2(x_2 - x_1) \\ 1 & x_3 - x_1 & x_3(x_3 - x_1) \end{vmatrix} \\
&\overset{\text{det Blockmatrix}}{=} \begin{vmatrix} x_2 - x_1 & x_2(x_2 - x_1) \\ x_3 - x_1 & x_3(x_3 - x_1) \end{vmatrix} \\
&\overset{(D1)(b)\text{Zeilen}}{=} (x_2 - x_1)(x_3 - x_1) \begin{vmatrix} 1 & x_2 \\ 1 & x_3 \end{vmatrix} = \prod_{1 \le i < j \le 3} (x_j - x_i).
\end{aligned}$$

Im Allgemeinen gilt

$$V_{x_1,\dots,x_n} = \prod_{1 \le i < j \le n} (x_j - x_i),$$

wie Sie in der Serie beweisen werden. Die Vandermonde-Determinante hat viele Anwendungen. Eine davon werden Sie in der Serie sehen.

In Beispiel 4.3.27 haben wir etwas benutzt, das Sie in der Übungsserie gesehen haben:

Übung 4.3.28. Seien $k \le n \in \mathbb{N}$ und $A \in M_{n \times n}(K)$ eine Blockmatrix der Form

$$A = \left(\begin{array}{c|c} B_1 & * \\ \hline 0 & B_2 \end{array}\right)$$

mit $B_1 \in M_{k \times k}(K)$ und $B_2 \in M_{(n-k)\times(n-k)}(K)$. Dann gilt

$$\det A = \det B_1 \cdot \det B_2.$$

Mit Induktion gilt eine ähnliche Aussage für Matrizen mit mehreren Blöcken. Des Weiteren gilt mit Transposition eine ähnliche Aussage für Matrizen der Form

$$\left(\begin{array}{c|c} B_1 & 0 \\ \hline * & B_2 \end{array}\right).$$

4.3.4 Determinante über einem Ring

Mittels der expliziten Formel für die Determinante sehen wir, dass für $A \in M_{n \times n}(\mathbb{Z})$ (das heisst, für eine Matrix mit ganzzahligen Einträgen) die Determinante $\det A$ ganzzahlig ist, also

$$\det A \in \mathbb{Z}.$$

Allgemeiner sei R ein Ring (zum Beispiel $R = M_{l\times l}(K)$). Sei

$$\begin{aligned} M_{n\times n}(R) &= \{\text{die Menge aller } n \times n\text{-Matrizen mit Einträgen in } R\} \\ &= \{(a_{ij})_{1\le i,j\le n} \mid a_{ij} \in R\}. \end{aligned}$$

(Falls $R = M_{l\times l}(K)$, dann sind die Elemente von $M_{n\times n}(K)$ Matrizen, deren Einträge $l \times l$-Matrizen über K sind!)

Für $A \in M_{n\times n}(R)$ definieren wir $\det A$ genau wie in 4.7:

$$\det A = \sum_{\sigma\in S_n} \operatorname{sign}(\sigma) a_{1\sigma(1)} \cdot \ldots \cdot a_{n\sigma(n)}.$$

Dann gilt $\det A \in R$. Wir werden diese Tatsache für $R = M_{l\times l}(K)$ in der Linearen Algebra II benutzen.

4.4 Magie mit Minoren

In diesem Abschnitt wollen wir die folgende Beobachtung verallgemeinern: Für $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ definieren wir

$$\operatorname{adj}(A) := \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Dann folgt mit einer kurzen Rechnung, dass

$$A \cdot \operatorname{adj}(A) = \operatorname{adj}(A) \cdot A = \begin{pmatrix} \det A & 0 \\ 0 & \det A \end{pmatrix}.$$

Daraus folgt eine Formel für die Inverse, falls sie existiert: Falls $\det A \neq 0$, dann ist

$$A^{-1} = \frac{1}{\det A} \operatorname{adj}(A).$$

Eigentlich folgt noch viel mehr aus dieser Verallgemeinerung, wie der Titel dieses Abschnitts schon andeutet. Wir besprechen in diesem Abschnitt die folgenden Resultate für eine Matrix $A \in M_{n\times n}(K)$:

(1) Die Konstruktion einer Matrix $\operatorname{adj}(A)$ mit der Eigenschaft, dass

$$\operatorname{adj}(A) \cdot A = A \cdot \operatorname{adj}(A) = \det A \cdot I_n.$$

Dies gibt uns insbesondere eine Formel für die Inverse (falls $\det A \neq 0$).

(2) Den Laplace-Entwicklungssatz, ein nützliches Resultat, um die Determinante zu berechnen (zumindest, wenn man die Berechnung von Hand macht).

(3) Die Cramersche Regel, eine Formel für die Lösung eines linearen Gleichungssystems $Ax = b$ mit einer Matrix A der Grösse $n \times n$.

Die nachfolgenden Resultate sind alle eng miteinander verwandt, so dass man sie voneinander herleiten kann.

Wir beginnen mit dem Laplace-Entwicklungssatz, da dieser am natürlichsten aus unseren vorherigen Resultaten folgt (nach Meinung des Dozenten).

Für eine Matrix $A \in M_{n\times n}(K)$ sei $A_{ij} \in M_{(n-1)\times(n-1)}(K)$ die Matrix, welche durch das Streichen der i-ten Zeile und j-ten Spalte entstanden ist.

Satz 4.4.1 (Laplace-Entwicklungssatz). *Sei $A \in M_{n\times n}(K)$. Für jedes $1 \leq i \leq n$ kann man die Determinante von A bezüglich der i-ten Zeile folgendermassen entwickeln:*

$$\begin{aligned} |A| &= (-1)^{i+1} a_{i1} |A_{i1}| + (-1)^{i+2} a_{i2} |A_{i2}| + \ldots + (-1)^{i+n} a_{in} |A_{in}| && (4.10) \\ &= \sum_{k=1}^{n} (-1)^{i+k} a_{ik} |A_{ik}| \,. && (4.11) \end{aligned}$$

Ähnlich kann man die Determinante bezüglich der j-ten Spalte folgendermassen entwickeln:

$$|A| = \sum_{k=1}^{n} (-1)^{k+j} a_{kj} |A_{kj}|$$

Beispiel 4.4.2. Sei

$$A = \begin{pmatrix} -2 & 2 & -3 \\ -1 & 1 & 3 \\ 2 & 0 & -1 \end{pmatrix}.$$

Wir entwickeln nach der zweiten Spalte ($j = 2$ in Satz 4.4.1). Es ist nützlich, das Vorzeichen in Abhängigkeit der Indizes zu zeichnen:

$$\begin{pmatrix} + & - & + \\ - & + & - \\ + & - & + \end{pmatrix}.$$

Damit berechnen wir:

$$
\begin{aligned}
\det A &= (-1)^{1+2} \cdot 2 \cdot \left| \begin{pmatrix} -1 & 3 \\ 2 & -1 \end{pmatrix} \right| + (-1)^{2+2} \cdot 1 \cdot \left| \begin{pmatrix} -2 & -3 \\ 2 & -1 \end{pmatrix} \right| + (-1)^{3+2} \cdot 0 \cdot \left| \begin{pmatrix} -2 & -3 \\ -1 & 3 \end{pmatrix} \right| \\
&= (-2) \cdot ((-1) \cdot (-1) - 2 \cdot 3) + 1 \cdot ((-2) \cdot (-1) - 2 \cdot (-3)) + 0 \\
&= (-2) \cdot (-5) + 8 \\
&= 18.
\end{aligned}
$$

Beweis von Satz 4.4.1. Bemerken Sie zunächst, dass die Entwicklung bezüglich Spalten durch Transponieren und durch die Identität $|A| = \left|A^T\right|$ aus der entsprechenden Entwicklung bezüglich Zeilen folgt. Daher genügt es den ersten Teil des Satzes zu beweisen.

Wir beweisen zuerst den Satz für die erste Zeile, das heisst, $i = 1$. Seien

$$
B_k = \begin{pmatrix} 0 & \cdots & 0 & a_{1k} & 0 & \cdots & 0 \\ a_{21} & \cdots & & \cdots & & \cdots & a_{2n} \\ \vdots & & & & & & \vdots \\ a_{n1} & \cdots & & \cdots & & \cdots & a_{nn} \end{pmatrix}
$$

für $k = 1, \ldots, n$. Dann folgt aus der Linearität der Determinante bezüglich der ersten Zeile, dass

$$
|A| = \sum_{k=1}^{n} |B_k| .
$$

Daher genügt es zu zeigen, dass

$$
|B_k| = (-1)^{1+k} a_{1k} |A_{1k}| . \tag{4.12}
$$

Wir fangen mit B_1 an:

$$
|B_1| = \left| \left(\begin{array}{c|ccc} a_{11} & 0 & \cdots & 0 \\ \hline a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{array} \right) \right| \overset{\text{Blockmatrix}}{=} a_{11} |A_{11}| = (-1)^{1+1} a_{11} |A_{11}| .
$$

Für B_2 vertauschen wir zuerst die ersten beiden Spalten:

$$
\begin{aligned}
|B_2| = \left| \begin{pmatrix} 0 & a_{12} & 0 & \cdots & 0 \\ a_{21} & \cdots & & \cdots & a_{2n} \\ \vdots & & & & \vdots \\ a_{n1} & \cdots & & \cdots & a_{nn} \end{pmatrix} \right| &\overset{C_1 \leftrightarrow C_2}{=} (-1) \left| \left(\begin{array}{c|ccc} a_{12} & 0 & \cdots & 0 \\ \hline a_{22} & a_{21} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n2} & a_{n1} & \cdots & a_{nn} \end{array} \right) \right| \\
&\overset{\text{Fall } i=1}{=} (-1) a_{12} \left| A_{12} \right| \\
&= (-1)^{1+2} a_{12} \left| A_{12} \right| .
\end{aligned}
$$

Beachten Sie, dass die Blockmatrix unten rechts genau A_{12} entspricht, da wir die ersten beiden Spalten vertauscht haben. Für B_3 vertauschen wir die Spalten C_2 und C_3 und benutzen den vorherigen Fall, so dass wir

$$
|B_3| = (-1)(-1)^{1+2} a_{13} \left| A_{13} \right| = (-1)^{1+3} a_{13} \left| A_{13} \right| .
$$

Ähnlich (mit Induktion über k und Spaltenvertauschen $C_{k-1} \leftrightarrow C_k$) folgt (4.12). Dies zeigt die Entwicklung bezüglich $i = 1$. Die Entwicklung bezüglich $i = 2$ folgt jetzt mit der Zeilenvertauschung $L_1 \leftrightarrow L_2$: Sei B, so dass $A \xrightarrow{L_1 \leftrightarrow L_2} B$. Dann gilt

$$
|A| = - |B| \overset{\text{Entwicklung } i=1}{=} -1 \sum_{k=1}^{n} (-1)^{1+k} b_{1k} \left| B_{1k} \right| \overset{\text{Definition } B}{=} \sum_{k=1}^{n} (-1)^{2+k} a_{2k} \left| A_{2k} \right| ,
$$

was die Entwicklung bezüglich der zweiten Zeile zeigt. Die Entwicklung bezüglich der p-ten Zeile folgt genau wie oben aus der Entwicklung der $(p-1)$-Zeile und dem Zeilenaustausch $L_{p-1} \leftrightarrow L_p$, was den Beweis beendet. □

Man kann den Sachverhalt in Satz 4.4.1 schön in einer Matrix darstellen:

Definition 4.4.3. Die *komplementäre Matrix* oder die *Adjunkte*[14] einer Matrix $A \in M_{n \times n}(K)$ ist

$$
\operatorname{adj}(A) := \left((-1)^{i+j} \left| A_{ji} \right| \right)_{1 \le i,j \le n} .
$$

Bemerkung 4.4.4.

- Die Vertauschung der Indizes in $|A_{ji}|$ ist kein Tippfehler!
- Wir definieren in der Linearen Algebra II die Adjungierte einer Matrix, welche nicht mit der Adjunkten verwechselt werden sollte.
- Aus Abschnitt 4.3.4 folgt: Falls $A \in M_{n \times n}(R)$ für einen Ring R, dann ist $\operatorname{adj}(A)$ ebenfalls in $M_{n \times n}(R)$.

[14] „Adjunct“ oder „adjugate“ auf Englisch.

Beispiel 4.4.5. Sei $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Dann ist

$$\operatorname{adj}(A) = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

(Wieso?)

Der folgende Satz folgt aus dem Entwicklungssatz von Laplace 4.4.1 (und ist eigentlich äquivalent zu ihm):

Satz 4.4.6. *Sei $A \in M_{n\times n}(K)$. Dann ist*

$$A \cdot \operatorname{adj}(A) = \operatorname{adj}(A) \cdot A = \det A \cdot I_n = \begin{pmatrix} \det A & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \det A \end{pmatrix}.$$

Beweis. Berechnen wir zuerst die Einträge von $A \cdot \operatorname{adj}(A)$ auf der Hauptdiagonalen:

$$(A \operatorname{adj}(A))_{ii} = \sum_{k=1}^{n} a_{ik} \underbrace{(-1)^{k+i} |A_{ik}|}_{(\operatorname{adj}(A))_{ki}} \overset{\text{Laplace } i\text{-te Zeile}}{=} \det A.$$

Für die Einträge ausserhalb der Hauptdiagonalen sei nun $i \neq j$: Dann ist

$$(A \operatorname{adj}(A))_{ij} = \sum_{k=1}^{n} a_{ik} \underbrace{(-1)^{k+j} |A_{jk}|}_{(\operatorname{adj}(A))_{kj}} =: (\star)$$

Bemerken Sie nun, dass $(\star)$ die Entwicklung bezüglich der j-ten Zeile der Matrix B ist, wobei B die Matrix ist, welche durch das Ersetzen der j-ten Zeile von A mit der i-ten Zeile von A entsteht. (Ersetzen! Nicht vertauschen!) Daher ist $(\star) = \det B$. Da B zwei gleiche Zeilen hat, folgt $(\star) = 0$.

Die Gleichung $\operatorname{adj}(A) \cdot A = \det A \cdot I_n$ zeigt man ähnlich (mit Spaltenentwicklung statt Zeilenentwicklung). □

Korollar 4.4.7. *Falls $A \in \mathrm{GL}_n(K)$, dann gilt*

$$A^{-1} = \frac{1}{\det A} \operatorname{adj}(A) \tag{4.13}$$

Korollar 4.4.8 (Cramersche Regel)**.** *Sei $A \in \mathrm{GL}_n(K)$. Wir betrachten das lineare Gleichungssystem*

$$Ax = b \tag{4.14}$$

für $b \in K^n = K^n_{Spal}$. Sei A_i die Matrix, welche b als i-te Spalte hat und die restlichen Spalten stimmen mit den Spalten von A überein. Das heisst

$$A_i = \left(A^{(1)} \cdots A^{(i-1)}\ b\ A^{(i+1)} \cdots A^{(n)}\right) \tag{4.15}$$

für $1 \le i \le n$. Dann berechnet sich die eindeutige Lösung $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ von (4.14) als

$$x_i = \frac{|A_i|}{|A|}.$$

Beweis. Dies folgt unmittelbar aus (4.14):

$$x = A^{-1}b = \frac{1}{|A|}\operatorname{adj}(A)b$$

beziehungsweise

$$x_i = \frac{1}{|A|}\sum_{k=1}^{n}\operatorname{adj}(A)_{ik}b_k = \frac{1}{|A|}\sum_{k=1}^{n}(-1)^{i+k}\left|A_{ki}\right| b_k \overset{\text{Laplace}}{=} \frac{1}{|A|}\left|A_i\right|.$$

Vielleicht fällt es Ihnen schwer die letzte Gleichheit zu sehen. Falls dies der Fall ist, dann betrachten Sie

$$A_i = \begin{pmatrix} | & & | & b_1 & | & & | \\ A^{(1)} & \dots & A^{(i-1)} & \vdots & A^{(i+1)} & \dots & A^{(n)} \\ | & & | & b_k & | & & | \end{pmatrix}$$

und machen Sie die Laplace-Entwicklung auf der i-ten Spalte! □

Hier sind konkrete Formeln für $n = 2, 3$ (Wir überlassen den Fall $n = 1$ dem Leser.):

Explizite Formel für 2×2-Matrizen: Wir betrachten folgendes lineares Gleichungssystem

$$\begin{cases} a_1x + b_1y = c_1 \\ a_2x + b_2y = c_2 \end{cases},$$

welches in Matrixschreibweise von der Form

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

ist. Wir nehmen an, dass $a_1b_2 - b_1a_2 \neq 0$. Dann können wir x und y mit Hilfe der Cramerschen Regel folgendermassen finden:

$$x = \frac{\left|\begin{pmatrix} c_1 & b_1 \\ c_2 & b_2 \end{pmatrix}\right|}{\left|\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}\right|} = \frac{c_1b_2 - b_1c_2}{a_1b_2 - b_1a_2} \quad \text{und} \quad y = \frac{\left|\begin{pmatrix} a_1 & c_1 \\ a_2 & c_2 \end{pmatrix}\right|}{\left|\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}\right|} = \frac{a_1c_2 - c_1a_2}{a_1b_2 - b_1a_2}.$$

Explizite Formel für 3×3-Matrizen: Die Regel für 3×3-Matrizen sind ähnlich. Sei

$$\begin{cases} a_1x + b_1y + c_1z = d_1 \\ a_2x + b_2y + c_2z = d_2 \\ a_3x + b_2y + c_3z = d_3 \end{cases},$$

welches in Matrixschreibweise von der Form

$$\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix}.$$

Dann sind

$$x = \frac{\begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}}, \; y = \frac{\begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}} \quad \text{und} \quad z = \frac{\begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}}.$$

In der Übungsstunde werden Sie vielleicht folgendes Beispiel betrachten:

$$\begin{aligned} 82x_1 + 45x_2 + 9x_3 &= 1 \\ 27x_1 + 16x_2 + 3x_3 &= 1 \\ 9x_1 + 5x_2 + 1x_3 &= 0 \end{aligned}$$

Dann ist die erweiterte Koeffizienten-Matrix gegeben durch:

$$(A|b) = \left(\begin{array}{ccc|c} 82 & 45 & 9 & 1 \\ 27 & 16 & 3 & 1 \\ 9 & 5 & 1 & 0 \end{array}\right).$$

Nach der Cramerschen Regel berechnet sich die Lösung des linearen Gleichungssystems als:

$$x_1 = \frac{\det A_1}{\det A} = \frac{\det\begin{pmatrix}1 & 45 & 9\\ 1 & 16 & 3\\ 0 & 5 & 1\end{pmatrix}}{\det\begin{pmatrix}82 & 45 & 9\\ 27 & 16 & 3\\ 9 & 5 & 1\end{pmatrix}} = \frac{1}{1} = 1,\ x_2 = \frac{\det A_2}{\det A} = \frac{\det\begin{pmatrix}82 & 1 & 9\\ 27 & 1 & 3\\ 9 & 0 & 1\end{pmatrix}}{\det\begin{pmatrix}82 & 45 & 9\\ 27 & 16 & 3\\ 9 & 5 & 1\end{pmatrix}} = \frac{1}{1} = 1$$

$$\text{und}\quad x_3 = \frac{\det A_3}{\det A} = \frac{\det\begin{pmatrix}82 & 45 & 1\\ 27 & 16 & 1\\ 9 & 5 & 0\end{pmatrix}}{\det\begin{pmatrix}82 & 45 & 9\\ 27 & 16 & 3\\ 9 & 5 & 1\end{pmatrix}} = \frac{-14}{1} = -14.$$

4.5 Verschiedene Wege in der Welt der Determinante

In dieser Vorlesung ist unser Ziel, verschiedene Ideen zu präsentieren. Das bedeutet manchmal auch, dass wir gewisse Resultate nicht auf die einfachste oder kürzeste Weise erhalten wollen. Es ist aber interessant zu wissen, wie man anders vorgehen könnte. Hier sind ein paar Beispiele:

- Man könnte die Determinante induktiv definieren mittel Laplace-Entwicklung: Für $\begin{pmatrix}a\end{pmatrix} \in M_{1\times 1}(K)$ definiert man
$$\det\begin{pmatrix}a\end{pmatrix} = a$$
und für $A \in M_{n\times n}(K)$ definiert man $\det A$ induktiv unter Benutzung von (4.10). Man beweist dann, dass diese Definition (D1)-(D3) erfüllt und geniesst dann die Tatsache, dass Satz 4.4.1 per Definition wahr ist. Beachten Sie, dass man dabei nicht einmal wissen muss, was eine Permutation ist und man braucht auch nicht den Gruppenbegriff. Der Nachteil so einer Definition ist, dass induktive Definitionen selten einleuchtend sind und Sie würden nicht viel von so einem Vorgehen lernen. Der interessierte Leser kann dazu gerne [12] konsultieren.

- Man könnte die Charakterisierung der Determinante durch (D1)-(D3) komplett weglassen und alles direkt mit der Gruppe S_n machen (siehe [10]).

- Falls das, was wir hier gemacht haben, nicht abstrakt genug ist, könnte man die Determinante $\det T$ direkt für Endomorphismen $T \in \mathrm{End}(V)$ über einem endlich-dimensionalen Vektorraum V definieren (siehe [11]). Wir werden $\det T$ im nächsten Abschnitt definieren.

- Fischer beweist, mit ähnlichen Argumenten wie wir sie zu Beginn benutzt haben, zuerst Satz 4.4.6 und dann Satz 4.4.1 als ein Korollar davon. Wir wählen einen anderen Weg, da die Ideen im Beweis von Satz 4.4.1 ein schönes Beispiel sind wieso Reduktionen nützlich sind.

Unter Verwendung von verschiedenen Vorgehensweisen werden einige Resultate einfacher, andere hingegen werden schwieriger. Einige unserer späteren Resultate folgen aus unseren vorherigen Resultaten. Hier ist ein cooles Beispiel wie man die Cramersche Regel nur unter Benutzung von (D1) und (D2) (für Spalten) beweisen kann:

Dieser Beweis benutzt auch ein grundlegendes Resultat aus Kapitel 2:

Alternativer Beweis von Korollar 4.4.8. Sei $A \in \mathrm{GL}_n(K)$. Wir betrachten die Abbildung

$$\begin{aligned} \Phi : K^n &\to K^n \\ b &\mapsto \frac{1}{|A|} \cdot (|A_1|, \ldots, |A_n|), \end{aligned}$$

wobei A_i genau wie in (4.15) definiert ist. Wegen (D1) (angewandt auf die Spalten) ist diese Abbildung linear. Des Weiteren ist das Bild der Spalte $A^{(i)}$ genau

$$(0, \ldots, 1, \ldots, 0) = e_i$$

wegen (D2). Also stimmt Φ mit der Abbildung $m_{A^{-1}}$ über der Basis $(A^{(1)}, \ldots, A^{(n)})$ überein. Also ist nach Satz 3.1.15

$$\Phi = m_{A^{-1}},$$

was äquivalent ist zur Aussage der Cramerschen Regel. □

Frage: Hier eine „Meta-Frage“: Wieso ergibt es Sinn, dass die Regel von Cramer unabhängig ist von (D3)?

4.6 Determinante und Rang

Wir haben gesehen, dass die Determinante einer Matrix $A \in M_{n\times n}(K)$ weiss, ob $\mathrm{Rang}(A) = n$ ist oder nicht. Weiss die Determinante auch, ob $\mathrm{Rang}(A) = r$ ist für $0 \leq r < n$?

Die Antwort ist „Ja“. Um dies zu sehen brauchen wir die folgende Definition:

Definition 4.6.1 (Minoren)**.** Seien $1 \leq k \leq n, m \in \mathbb{N}$. Ein k-*Minor* von $A \in M_{m\times n}(K)$ ist die Determinante einer $k \times k$-Matrix, die durch Streichen von $(n-k)$-Spalten und $(m-k)$-Zeilen von A entstanden ist. Eine solche Matrix heisst $k \times k$-*Untermatrix* von A.

Satz 4.6.2. *Sei $A \in M_{m\times n}(K)$. Dann gilt*[15]

$$\mathrm{Rang}(A) = \max\{k \mid \textit{es existiert ein } k\textit{-Minor, der verschieden von Null ist}\}.$$

Beweis. Sei $r = \mathrm{Rang}(A)$. Wir müssen zeigen, dass jeder k-Minor mit $k > r$ Null ist, und dass ein r-Minor existiert, der nicht Null ist.

Sei also $k > r$. Da $\mathrm{Rang}(A) = \mathrm{Spaltenrang}(A)$ ist, sind je k Spalten von A linear abhängig. Also sind die k Spalten einer beliebigen $k \times k$-Untermatrix von A auch linear abhängig. Also ist jeder k-Minor gleich Null.

Jetzt müssen wir einen r-Minor finden, welcher nicht Null ist. Nehmen wir an, dass $A \neq 0$ (siehe Fussnote). Da $\mathrm{Spaltenrang}(A) = r > 0$ ist, existieren

$$j_1 < \ldots < j_r,$$

so dass $A^{(j_1)}, \ldots, A^{(j_r)}$ linear unabhängig sind. Also hat

$$B := \left(A^{(j_1)} \cdots A^{(j_r)}\right)$$

Rang r. Da $\mathrm{Zeilenrang}(B) = \mathrm{Rang}(B) = r$ ist, existieren

$$i_1 < \ldots < i_r$$

mit

$$C := \begin{pmatrix} B_{(i_1)} \\ \vdots \\ B_{(i_r)} \end{pmatrix} \in M_{r\times r}(K)$$

und $\mathrm{Rang}(C) = r$. Daher ist $|C| \neq 0$. Die Matrix C ist eine $r \times r$-Untermatrix von A, was den Beweis beendet. $\square$

[15] Falls hier die leere Menge steht (d.h. falls $A = 0$), dann verstehen wir dieses Maximum als 0, damit der Satz gilt.

4.7 Einige Korollare und die Determinante eines Endomorphismus

In diesem Abschnitt geben wir einige einfache Korollare zur Multiplikativität der Determinante, welche die Definition der Determinante eines Endomorphismus ermöglicht:

Korollar 4.7.1. *Für jedes* $A \in \mathrm{GL}_n(K)$ *gilt*

$$\det(A^{-1}) = \det(A)^{-1}.$$

Beweis. Es gilt

$$1 = \det(I_n) = \det(AA^{-1}) = \det(A)\det(A^{-1})$$

und das Korollar folgt. □

Korollar 4.7.2. *Seien* $A, B \in M_{n\times n}(K)$ *ähnliche Matrizen. Dann gilt*

$$\det(A) = \det(B).$$

Beweis. Sei $P \in \mathrm{GL}_n(K)$ mit $A = P^{-1}BP$. Dann ist

$$\det(A) = \det(P^{-1})\det(B)\det(P) = \det(P)^{-1}\det(B)\det(P) = \det(B)$$

nach Korollar 4.7.1. □

Das letzte Korollar ist für das nächste Kapitel wichtig und ermöglicht folgende Definition:

Definition 4.7.3. Sei V ein Vektorraum über K mit $\dim(V) = n < \infty$ und sei $T \in \mathrm{End}(V)$. Wir definieren die *Determinante von* T als

$$\det(T) := \det([T]_{\mathcal{B}}),$$

wobei $\mathcal{B}$ eine beliebige (geordnete) Basis von V ist.

Lemma 4.7.4. *Diese Definition ergibt Sinn. Das heisst,* $\det(T)$ *ist wohl-definiert.*

Beweis. Seien $\mathcal{B}, \mathcal{C}$ zwei geordnete Basen von V. Laut Korollar 3.3.32 sind die Matrizen $[T]_{\mathcal{B}}$ und $[T]_{\mathcal{C}}$ ähnlich zueinander.

Daher folgt aus Korollar 4.7.2, dass

$$\det([T]_{\mathcal{B}}) = \det([T]_{\mathcal{C}}).$$

Das bedeutet genau, dass der Wert $\det(T)$ nicht von der Wahl der Basis abhängt, was zu zeigen war. □

Wie bei Matrizen misst die Determinante, ob ein Endomorphismus invertierbar ist oder nicht. Hier ist die grosse Schwester von Proposition 3.6.8 (A.K.A. ugly Lemma):

Korollar 4.7.5. *Seien* $T \in \mathrm{End}(V)$ *wobei* $\dim V < \infty$ *und* $\mathcal{B}$ *eine geordnete Basis von* V*. Die folgenden Aussagen sind äquivalent:*

(1) T *ist ein Isomorphismus.*

(2) T *ist injektiv.*

(3) T *ist surjektiv.*

(4) $[T]_{\mathcal{B}}$ *ist invertierbar.*

(5) $[T]_{\mathcal{B}}$ *erfüllt eine der äquivalenten Bedingungen in Proposition 3.6.8.*

(6) $\det(T) \neq 0$.

Beweis. Die Äquivalenz der ersten drei Aussagen folgt aus dem Rangsatz. Die Abbildung T ist genau dann ein Isomorphismus, wenn $[T]_{\mathcal{B}}$ invertierbar ist. (Dies folgt zum Beispiel aus Korollar 3.5.6.) Dies ist äquivalent zur Aussage, dass

$$\det([T]_{\mathcal{B}}) \neq 0.$$

Da (per Definition) $\det(T) = \det([T]_{\mathcal{B}})$, folgen alle andere Äquivalenzen. □

Obwohl wir dies im Moment nicht brauchen, möchten wir noch eine andere Grösse erwähnen, die unter Ähnlichkeit invariant ist:

Definition 4.7.6. Sei $A \in M_{n\times n}(K)$. Die *Spur* von A ist die Summe aller Einträge auf der Hauptdiagonalen. Das heisst,

$$\mathrm{spur}\,(A) := a_{11} + \ldots + a_{nn}.$$

Lemma 4.7.7. *Für* $A, B \in M_{n\times n}(K)$ *gilt*

$$\mathrm{spur}\,(AB) = \mathrm{spur}\,(BA)\,.$$

Beweis. Laut der Definition der Matrixmultiplikation gilt

$$\mathrm{spur}\,(AB) = \sum_{i=1}^{n}(AB)_{ii} = \sum_{i=1}^{n}\sum_{k=1}^{n} a_{ik}b_{ki} = \sum_{k=1}^{n}\sum_{i=1}^{n} b_{ki}a_{ik} = \sum_{k=1}^{n}(BA)_{kk} = \mathrm{spur}\,(BA)\,.$$

Also folgt das Lemma. □

Korollar 4.7.8. *Falls $A, B \in M_{n\times n}(K)$ ähnlich sind, dann gilt*

$$\mathrm{spur}\,(A) = \mathrm{spur}\,(B)\,.$$

Beweis. Sei $P \in \mathrm{GL}_n$ mit $A = P^{-1}BP$. Dann ist

$$\mathrm{spur}\,(A) = \mathrm{spur}\,\big(P^{-1}(BP)\big) = \mathrm{spur}\,\big((BP)P^{-1}\big) = \mathrm{spur}\,(B)$$

und das Korollar folgt. □

4.8 Volumen und Orientierung

Wir haben in Abschnitt 4.1 für $A \in M_{2\times 2}(\mathbb{R})$ gesehen:

- Die Abbildung m_A ist orientierungserhaltend genau dann, wenn $\det(A) > 0$.
- Die Abbildung m_A ist orientierungsumkehrend genau dann, wenn $\det(A) < 0$.

Für einige Sätze, die Sie dieses Jahr in der Analysis sehen werden, und auch für später im Studium ist es wichtig, den Begriff der Orientierung für $\mathbb{R}^n$ und sogar für allgemeine endlich-dimensionale reelle Vektorräume zu verstehen.

Orientierung

Sei V ein Vektorraum über $\mathbb{R}$ mit $\dim(V) = n < \infty$. Sei X_V die Menge aller geordneten Basen auf V. Wir definieren folgendermassen eine Äquivalenzrelation auf X_V: Für $\mathcal{B} = (v_1, \ldots, v_n)$ und $\mathcal{C} = (w_1, \ldots, w_n)$ sei $T \in \mathrm{End}(V)$ die eindeutige Abbildung mit $T(v_i) = w_i$ für $i \in \{1, \ldots, n\}$. Wir sagen, dass $\mathcal{B}$ und $\mathcal{C}$ die gleiche Orientierung haben, und schreiben $\mathcal{B} \sim \mathcal{C}$, falls $\det(T) > 0$.

Übung 4.8.1. Es gilt

$$\mathcal{B} \sim \mathcal{C} \iff \det([\mathrm{Id}_V]^{\mathcal{B}}_{\mathcal{C}}) > 0 \iff \det([\mathrm{Id}_V]^{\mathcal{C}}_{\mathcal{B}}) > 0.$$

(Hinweis: Berechnen Sie $[T]^{\mathcal{B}}_{\mathcal{C}}$ für T wie oben und benutzen Sie $[T]^{\mathcal{C}}_{\mathcal{C}} = [T]^{\mathcal{B}}_{\mathcal{C}}[\mathrm{Id}_V]^{\mathcal{C}}_{\mathcal{B}}$.)

Die obige Relation $\sim$ definiert in der Tat eine Äquivalenzrelation auf X_V und $X_V/\!\sim$ enthält genau zwei Äquivalenzklassen falls $\dim(V) > 0$.

Eine Wahl einer dieser beiden Äquivalenzklassen auf V heisst eine *Orientierung* auf V.

Beispiel 4.8.2. In $\mathbb{R}^n$ wählt man normalerweise die Äquivalenzklasse der Basis $(e_1, \ldots, e_n)$ und nennt die entsprechende Äquivalenzklasse die *positive Orientierung*. Basen in dieser

Äquivalenzklasse nennt man *positiv orientiert* und Basen in der anderen Äquivalenzklasse nennt man *negativ orientiert.*

In $\mathbb{R}^2$ ist dann (e_1, e_2) positiv orientiert und (e_2, e_1) negativ orientiert. In $\mathbb{R}^3$ ist eine „rechtshändige Basis“ positiv orientiert und eine „linkshändige Basis“ negativ orientiert.

Frage: Angenommen Sie sind eine Ameise auf einem Möbiusband. Wie könnten Sie herausfinden, dass Sie in der Tat auf einem Möbiusband leben, unter Verwendung des Inhalts dieses Abschnittes?

Intermezzo über das Volumen

In Analysis II werden Sie beweisen, dass für gewisse Teilmengen $X \subseteq \mathbb{R}^n$ und $m_A : \mathbb{R}^n \to \mathbb{R}^n$ gilt

$$\mathrm{Vol}_{\mathbb{R}^n}(m_A(X)) = |\det(A)|\, \mathrm{Vol}_{\mathbb{R}^n}(X). \tag{4.16}$$

Hier ist eine ungefähre Skizze, wieso (4.16) stimmen sollte.

Jede vernünftige Volumenfunktion auf $\mathbb{R}^n$ sollte Eigenschaften analog zu (D1)–(D3) erfüllen, aber mit Absolutbeträgen. Dies ist äquivalent dazu, dass (4.16) gilt, wenn A eine Elementarmatrix ist. Zum Beispiel ändert Multiplikation mit der Elementarmatrix $P_{i,j}$ das Volumen nicht oder Multiplikation mit $S_i(\alpha)$ multipliziert das Volumen mit $|\alpha|$. Unter Verwendung dieser Tatsachen kann man (4.16) „beweisen“ für alle $A \in \mathrm{GL}_n(\mathbb{R})$:

Schreibe $A = T_1 \cdots T_k$ für $T_1, \ldots, T_k$ Elementarmatrizen. Dann gilt

$$\begin{aligned} \mathrm{Vol}_{\mathbb{R}^n}(m_A(X)) = \mathrm{Vol}_{\mathbb{R}^n}(T_1 \cdots T_k X) &\overset{Induktion}{=} |\det(T_1)| \cdots |\det(T_k)| \, \mathrm{Vol}_{\mathbb{R}^n}(X) \\ &= |\det(T_1 \cdots T_k)| \, \mathrm{Vol}_{\mathbb{R}^n}(X) = |\det(A)| \, \mathrm{Vol}_{\mathbb{R}^n}(X). \end{aligned}$$

Für $A \notin \mathrm{GL}_n(\mathbb{R})$ könnte man (4.16) auch „beweisen“: In diesem Fall gilt

$$\mathrm{Im}(m_A) = m_A(\mathbb{R}^n) \subsetneq V$$

und daher ist $m_A(\mathbb{R}^n)$ eine Hyperebene mit Dimension $< n$. Solche Hyperebenen sollten Volumen Null haben für jede vernünftige Volumenfunktion auf $\mathbb{R}^n$. Da $\det A = 0$ ist für $A \notin \mathrm{GL}_n(\mathbb{R})$, gilt (4.16) auch in diesem Fall.

Changelog: Kapitel 4

- 19.02: Nummerierungen wurden geändert.

Kapitel 5

Eigenvektoren und Eigenwerte

5.1 Einführung

Sei V ein Vektorraum über K und $T \in \operatorname{End}(V)$. Wie wir in Kapitel 0 erwähnt haben, kann man bei T an eine Symmetrie von V denken. Man sucht einen Vektor $v \in V$, auf welchem T besonders einfach agiert. In Kapitel 0 konnten wir die Basis $\mathcal{B} = (\mathcal{F}_{1,\varphi}, \mathcal{F}_{1,\psi})$ für **Fib** finden, auf deren Elemente die Verschiebungsabbildung

$$\begin{aligned} S : \mathbf{Fib} &\to \mathbf{Fib} \\ (a_0, a_1, \dots) &\mapsto (a_1, a_2, \dots) \end{aligned}$$

besonders einfach agiert:

$$S(\mathcal{F}_{1,\varphi}) = \varphi \mathcal{F}_{1,\varphi} \quad \text{und} \quad S(\mathcal{F}_{1,\psi}) = \psi \mathcal{F}_{1,\psi}.$$

Wir mussten aber $\mathcal{F}_{1,\varphi}$ und $\mathcal{F}_{1,\psi}$ „erraten“ und hatten leider keinen Algorithmus zur Verfügung, um diese besonderen Vektoren zu finden. Später haben wir in Beispiel 3.6.23 gesehen, dass S bezüglich der Basis $\mathcal{B}$ eine besonders einfache Darstellungsmatrix hat, nämlich

$$[S]_{\mathcal{B}} = \begin{pmatrix} \varphi & 0 \\ 0 & \psi \end{pmatrix}. \tag{5.1}$$

Damit kann man die Matrizen S^n für $n \in \mathbb{N}$ problemlos berechnen und eine Formel für Fibonacci-Folgen finden. Dies bezieht sich einfach auf die Tatsache, dass man leicht Potenzen von Diagonalmatrizen berechnen kann. In der Terminologie, die wir in diesem Kapitel entwickeln, bedeutet (5.1), dass S *diagonalisierbar* ist. Diagonalisierung von Endomorphismen hat viele Anwendungen sowohl in der Mathematik als auch in allen möglichen Gebieten der Wissenschaft. Eine grundlegende Anwendung in der Physik ist beispielsweise die Analyse von Schwingungen mit kleinen Auslenkungen. Um zu verstehen, wann ein Endomorphismus diagonalisiert werden kann und wie man dies explizit machen kann, braucht man die Begriffe von Eigenwerten und Eigenvektoren.

5.2 Definitionen

Sei V ein Vektorraum über K.

Definition 5.2.1. Ein Endomorphismus $T \in \mathrm{End}(V)$, wobei $\dim V = n < \infty$, heisst *diagonalisierbar*, falls es eine Basis $\mathcal{B} = (v_1, \ldots, v_n)$ gibt, so dass

$$[T]_{\mathcal{B}} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix} \tag{5.2}$$

eine Diagonalmatrix ist.

Bemerken Sie, dass (5.2) genau dann gilt, wenn

$$Tv_i = \lambda_i v_i$$

für $i = 1, \ldots, n$. Vektoren mit dieser Eigenschaft sind die Helden dieses Kapitels.

Definition 5.2.2. Ein Vektor $v \in V$ heisst ein *Eigenvektor* von T, falls $v \neq 0$ und es $\lambda \in K$ gibt, so dass

$$Tv = \lambda v.$$

Der Skalar λ heisst der zum Eigenvektor v zugehörige *Eigenwert*.

Bemerkung 5.2.3. (1) Man könnte genau so gut Folgendes definieren: Ein Skalar $\lambda \in K$ heisst ein *Eigenwert* von T, falls es $v \in V$ gibt mit $v \neq 0$ und

$$Tv = \lambda v. \tag{5.3}$$

Jeder Vektor v mit (5.3), der ungleich Null ist, heisst ein *Eigenvektor* von T zum Eigenwert λ.

(2) Eigenvektoren sind immer verschieden vom Nullvektor! Eigenwerte können jedoch 0 sein, wie das folgende Beispiel zeigt.

Beispiel 5.2.4. Sei $T : V \to V$ eine lineare Abbildung mit $\mathrm{Ker}(T) \neq \{0\}$. Dann ist 0 ein Eigenwert von T und jeder Vektor $v \in \mathrm{Ker}(T)$, der ungleich Null ist, ist ein Eigenvektor zum Eigenwert 0. Anders gesagt, gilt

$$\mathrm{Ker}(T) = \{0_V\} \sqcup \{\text{alle Eigenvektoren mit Eigenwert } 0\}.$$

Beispiel 5.2.5. Man könnte jetzt zu Korollar 4.7.5 Folgendes hinzufügen:

$$T \text{ ist ein Isomorphismus } \iff 0 \text{ ist kein Eigenwert von } T$$

Beispiel 5.2.6. Die Eigenwerte von $S\colon \mathbf{Fib} \to \mathbf{Fib}$ sind φ und ψ und $\mathcal{F}_{1,\varphi}$, $\mathcal{F}_{1,\psi}$ sind die entsprechenden Eigenvektoren.

Das folgende Lemma gilt per Definition, aber es ist sehr wichtig dessen Inhalt im Fall einer Matrix zu verstehen.

Lemma 5.2.7. *Ein Endomorphismus $T \in \mathrm{End}(V)$ ist diagonalisierbar genau dann, wenn eine Basis existiert, die nur aus Eigenvektoren von T besteht.*

Beweis. Die Aussage des Lemmas gilt in der Tat per Definition von Diagonalisierbarkeit, von Eigenvektoren und der Darstellungsmatrix. □

Definition 5.2.8. (1) Ein Skalar $\lambda \in K$ heisst ein *Eigenwert* einer Matrix $A \in M_{n\times n}(K)$, falls λ ein Eigenwert von $m_A\colon K^n \to K^n$ ist.

(2) Ein Vektor $v \in K^n$ heisst ein *Eigenvektor* einer Matrix $A \in M_{n\times n}(K)$, falls v ein Eigenvektor von $m_A\colon K^n \to K^n$ ist.

(3) A heisst *diagonalisierbar*, falls m_A diagonalisierbar ist.

5.2.1 Definition 5.2.8 explizit für Matrizen

Der Spezialfall $T = m_A$ ist so wichtig und teilweise auch verwirrend für einige Studenten, dass wir Definition 5.2.8 noch einmal explizit für Matrizen angeben. Es ist aber wichtig zu wissen, dass dieser Abschnitt streng genommen nicht nötig ist.

Teile (1) und (2) von Definition 5.2.8 können so geschrieben werden:

Definition 5.2.9. Sei $A \in M_{n\times n}(K)$. Ein Skalar $\lambda \in K$ heisst ein *Eigenwert* von A, falls es einen Vektor $v \in K^n$ gibt, so dass $v \neq 0$ und

$$Av = \lambda v.$$

Ein Vektor $v \in K^n$ mit dieser Eigenschaft heisst ein *Eigenvektor* zum Eigenwert λ.

Beispiel 5.2.10. Sei

$$A = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}.$$

Dann sind $\{\lambda_1, \ldots, \lambda_n\}$ die Eigenwerte von A. Für jedes i gilt, dass e_i ein Eigenvektor zum Eigenwert λ_i ist.

Beispiel 5.2.11. Für $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ gilt

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

Daher ist 1 ein Eigenwert von A und $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ ist ein Eigenvektor von A zum Eigenwert 1. Ausserdem ist $a \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} a \\ -a \end{pmatrix}$ ebenfalls ein Eigenvektor von A zum Eigenwert 1. Ähnlich gilt, dass

$$A \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} = 3 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

und daher ist 3 ein Eigenwert von A und

$$\left\{ \begin{pmatrix} a \\ a \end{pmatrix} \mid a \neq 0 \right\}$$

ist die Menge aller Eigenvektoren von A zum Eigenwert 3.

Geometrisch ist es einfach zu verstehen, wie eine Matrix $A \in M_{n \times n}(K)$ auf einem Eigenvektor $v \neq 0$ zum Eigenwert λ agiert: A skaliert v mit einem Faktor λ.

Übung 5.2.12. Versuchen Sie zu zeigen, dass $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ nur 1 als Eigenwert hat und alle Eigenvektoren zum Eigenwert 1 durch

$$\{ae_1 \mid a \neq 0\}$$

gegeben sind.

Die folgende Definition ist äquivalent zu Definition 5.2.8 (3):

Definition 5.2.13. Ein Matrix $A \in M_{n \times n}(K)$ heisst *diagonalisierbar*, falls m_A diagonalisierbar ist. Äquivalent ist A diagonalisierbar, falls $P \in \mathrm{GL}_n(K)$ existiert, so dass

$$P^{-1}AP = D,$$

wobei D eine Diagonalmatrix ist.

Dies bietet auch eine hilfreiche Art und Weise, um zu sehen, was Lemma 5.2.7 im Fall $T = m_A$ aussagt, wie die Leser überprüfen können:

Lemma 5.2.14 (Lemma 5.2.7 für $T = m_A$)**.** *Eine Matrix A ist diagonalisierbar genau dann, wenn es eine Basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ von K^n gibt, die nur Eigenvektoren von A enthält.*

In diesem Fall ist

$$P = [\mathrm{Id}_{K^n}]^{\mathcal{B}}_{\mathcal{E}} = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix},$$

wobei $\mathcal{E}$ die Standard-Basis von K^n ist und

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix},$$

wobei λ_i der zugehörige Eigenwert zum Eigenvektor v_i ist.

Hier ist die Verbindung zwischen allgemeinen Endomorphismen und Matrizen:

Übung 5.2.15. Sei $T \in \mathrm{End}(V)$ mit $\dim V < \infty$ und $\mathcal{B}$ eine Basis von V. Dann gilt

$$\begin{gathered} v \text{ ist ein Eigenvektor von } T \text{ mit Eigenwert } \lambda \\ \Longleftrightarrow \\ [v]_{\mathcal{B}} \text{ ist ein Eigenvektor von } [T]_{\mathcal{B}} \text{ mit Eigenwert } \lambda. \end{gathered}$$

5.2.2 Eigenvektoren mit verschiedenen Eigenwerten sind linear unabhängig

Wir wechseln nun wieder zu dem allgemeinen Fall $T \in \mathrm{End}(V)$. Die Lemmas 5.2.7 und 5.2.14 zeigen, dass es eine zentrale Frage ist, ob wir eine Basis aus Eigenvektoren finden können. Die folgende Proposition hilft uns in dieser Sache:

Proposition 5.2.16. *Sei $T \in \mathrm{End}(V)$ und seien $v_1, \ldots, v_n$ Eigenvektoren zu paarweise verschiedenen Eigenwerten $\lambda_1, \ldots, \lambda_n$. Dann ist $\{v_1, \ldots, v_n\}$ linear unabhängig.*

Korollar 5.2.17. *Falls $\dim V = n$ ist und $T \in \mathrm{End}(V)$ n paarweise verschiedene Eigenwerte hat, dann ist T diagonalisierbar.*

Beweis von Proposition 5.2.16. Wir beweisen die Aussage mit Induktion über n. Für $n = 1$ erinnern Sie sich daran, dass jeder Eigenvektor per Definition verschieden von Null ist. Daher ist $\{v_1\}$ linear unabhängig.

Nehmen wir an, dass die Aussage wahr ist für $n - 1$ Vektoren und seien $v_1, \ldots, v_n$ Eigenvektoren mit paarweise verschiedenen Eigenwerten $\lambda_1, \ldots, \lambda_n$. Wir müssen zeigen, dass

$$\underbrace{a_1 v_1 + \ldots + a_n v_n = 0}_{(\diamondsuit)} \implies a_1 = \ldots = a_n = 0.$$

Wir wenden T auf $(\diamondsuit)$ an:

$$\begin{aligned} T(a_1v_1 + \ldots + a_nv_n) &= a_1Tv_1 + \ldots + a_nTv_n \\ &= \underbrace{a_1\lambda_1v_1 + \ldots + a_n\lambda_nv_n}_{(\star)} = T(0) = 0. \end{aligned}$$

Wir betrachten jetzt $(\star) - \lambda_n(\diamondsuit)$:

$$a_1\lambda_1v_1 + \ldots + \cancel{a_n\lambda_nv_n} - (\lambda_na_1v_1 + \ldots + \cancel{\lambda_na_nv_n}) = 0 - \lambda_n0 = 0$$

beziehungsweise

$$a_1(\lambda_1 - \lambda_n)v_1 + \ldots + a_{n-1}(\lambda_{n-1} - \lambda_n)v_{n-1} = 0.$$

Laut der Induktionsannahme sind $v_1, \ldots, v_{n-1}$ linear unabhängig und daher ist

$$a_i(\lambda_i - \lambda_n) = 0$$

für alle $i = 1, \ldots, n-1$. Da $\lambda_1, \ldots, \lambda_n$ paarweise verschieden sind, gilt $\lambda_i - \lambda_n \neq 0$ für alle $i = 1, \ldots, n-1$. Also folgt, dass

$$a_1 = \ldots = a_{n-1} = 0.$$

Wenn wir dies in $(\diamondsuit)$ einsetzen, dann folgt auch $a_n = 0$, was den Beweis beendet. □

Hier ist eine nette Anwendung von Proposition 5.2.16:

Beispiel 5.2.18. Die Menge

$$A = \{F_t = (1, t, t^2, t^3, \ldots) \mid t \in \mathbb{R}\} \subseteq \mathbb{R}^\infty$$

ist eine linear unabhängige Menge der Kardinalität $\mathbb{R}$. Wieso? Es gibt verschiedene Wege dies zu zeigen. Wir präsentieren hier jedoch dieses tolle Argument:

Wir betrachten die Verschiebungsabbildung

$$\begin{aligned} S : \mathbb{R}^\infty &\to \mathbb{R}^\infty \\ (a_1, a_2, a_3, \ldots) &\mapsto (a_2, a_3, \ldots). \end{aligned}$$

Übung 5.2.19. Zeigen Sie, dass A keine Basis von $\mathbb{R}^\infty$ ist. (Hinweis: Was ist die Wachstumsrate von Linearkombinationen von Elementen von A?)

Zurück zum Beispiel: Der Vektor F_t ist ein Eigenvektor von S mit Eigenwert t. Laut Proposition 5.2.16 ist also jede endliche Teilmenge von A linear unabhängig, was per

Definition bedeutet, dass A linear unabhängig ist. Es folgt $\dim \mathbb{R}^\infty \geq |\mathbb{R}|$. Da $\mathbb{R}^\infty$ und $\mathbb{R}$ dieselbe Kardinalität haben, folgt $\dim \mathbb{R}^\infty = |\mathbb{R}|$.

Zwei grundlegende Fragen bleiben offen:

(1) Wie finden wir alle möglichen Eigenwerte zu einem gegebenen Endomorphismus $T \in \operatorname{End}(V)$ (bzw. zu einer gegebenen Matrix $A \in M_{n\times n}(K)$)?

(2) Nachdem wir wissen, dass $\lambda \in K$ ein Eigenwert von T ist (bzw. von A). Wie finden wir alle Eigenvektoren mit Eigenwert λ?

Beide Fragen werden wir im nächsten Abschnitt beantworten.

5.3 Das charakteristische Polynom

Die Antwort auf die zweite Frage (2) ist eigentlich sehr einfach.

Definition 5.3.1. Sei $T \in \operatorname{End}(V)$ und $\lambda \in K$. Der *Eigenraum* von T bezüglich λ ist

$$\begin{aligned}\operatorname{Eig}_T(\lambda) &:= \operatorname{Ker}(T - \lambda \operatorname{Id}_V)\\ &= \{v \in V \mid (T - \lambda \operatorname{Id}_V)v = 0\}\\ &= \{v \in V \mid Tv = \lambda v\}.\end{aligned}$$

Falls $A \in M_{n\times n}(K)$ ist, dann ist $\operatorname{Eig}_A(\lambda) := \operatorname{Eig}_{m_A}(\lambda)$.

Übung 5.3.2. Machen Sie die Definition von $\operatorname{Eig}_A(\lambda)$ explizit! (Die Lösung ist in der Fussnote[1].)

Lemma 5.3.3. *Es gilt*

(1) Der Eigenraum $\operatorname{Eig}_T(\lambda)$ *ist ein Untervektorraum.*

(2) Es gilt $\operatorname{Eig}_T(\lambda) \neq \{0_V\}$ *genau dann, wenn* λ *ein Eigenwert von* T *ist. Allgemeiner gilt*

$$\operatorname{Eig}_T(\lambda) = \{0_V\} \sqcup \{\text{alle Eigenvektoren mit Eigenwert } \lambda\}. \tag{5.4}$$

(3) Falls $\lambda_1 \neq \lambda_2$ *ist, dann gilt*

$$\operatorname{Eig}_T(\lambda_1) \cap \operatorname{Eig}_T(\lambda_2) = \{0_V\}.$$

Beweis. (1) Der Kern einer linearen Abbildung ist immer ein Untervektorraum.

[1] Es gilt $\operatorname{Eig}_A(\lambda) = \operatorname{Ker}(A - \lambda I_n) = \{v \in K^n \mid Av = \lambda v\}$.

(2) Jeder von Null verschiedene Vektor in $\mathrm{Eig}_T(\lambda)$ erfüllt

$$Tv = \lambda v$$

und daher ist so ein Vektor per Definition ein Eigenvektor von T mit Eigenwert λ. Also folgt Gleichung (5.4).

(3) Dies folgt aus Proposition 5.2.16: Falls $0 \neq v \in \mathrm{Eig}_T(\lambda_1) \cap \mathrm{Eig}_T(\lambda_2)$, dann ist $\{v, v\}$ linear unabhängig laut Proposition 5.2.16, was ein Widerspruch ist.

Hier ist ein anderer direkter Beweis: Für $v \in \mathrm{Eig}_T(\lambda_1) \cap \mathrm{Eig}_T(\lambda_2)$ gilt

$$\lambda_1 v = Tv = \lambda_2 v$$

und daher ist $(\lambda_1 - \lambda_2)v = 0$, was $v = 0$ impliziert, da $\lambda_1 - \lambda_2 \neq 0$.

□

Korollar 5.3.4. *Sei $T \in \mathrm{End}(V)$. Für $\lambda \in K$ gilt*

$$\lambda \textit{ ist ein Eigenwert von } T \iff \mathrm{Ker}(T - \lambda \, \mathrm{Id}_V) \neq \{0_V\}.$$

Sei jetzt V ein Vektorraum mit $\dim V = n$. Da wir $\mathrm{Ker}(T - \lambda \, \mathrm{Id}_V)$ berechnen können, ist dieses Korollar unser Schlüssel für die Antwort auf Frage (2). Dieses Korollar gibt uns auch einen Hinweis bezüglich der Antwort auf Frage (1).

Erinnern Sie sich daran, dass

$$\mathrm{Ker}(T - \lambda \, \mathrm{Id}_V) \neq \{0_V\} \iff \det(T - \lambda \, \mathrm{Id}_V) = 0.$$

Das heisst, dass die Nullstellen der Funktion

$$\lambda \in K \mapsto \det(T - \lambda \, \mathrm{Id}_V) \in K \tag{5.5}$$

genau die Eigenwerte von T sind. Erinnern Sie sich an Definition 4.7.3 und wählen Sie eine Basis $\mathcal{B} \subseteq V$. Dann ist

$$\begin{aligned} \det(T - \lambda \, \mathrm{Id}_V) = \det([T - \lambda \, \mathrm{Id}_V]_{\mathcal{B}}) &\overset{\text{Korollar 3.5.6}}{=} \det([T]_{\mathcal{B}} - \lambda [\mathrm{Id}_V]_{\mathcal{B}}) \\ &= \det([T]_{\mathcal{B}} - \lambda I_n). \end{aligned}$$

Also ist die Funktion (5.5) durch

$$\lambda \in K \mapsto \det([T]_{\mathcal{B}} - \lambda I_n) \in K \tag{5.6}$$

gegeben. Um zu zeigen, dass diese Funktion einer Polynomfunktion entspricht, definieren wir zuerst das charakteristische Polynom einer Matrix.

Definition 5.3.5. Sei $n \in \mathbb{N}$, $A = (a_{ij}) \in M_{n\times n}(K)$ und x eine Variable. Das *charakteristische Polynom von A* ist

$$p_A(x) := \det(A - xI_n) = \det \begin{pmatrix} a_{11}-x & a_{12} & \cdots & \cdots & a_{1n} \\ a_{21} & a_{22}-x & a_{23} & \cdots & a_{2n} \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{n(n-1)} & a_{nn}-x \end{pmatrix} \in K[x]. \tag{5.7}$$

Beachten Sie, dass $A - xI_n$ als Matrix über dem Ring $K[x]$ zu betrachten ist.

Lemma 5.3.6. *Für $A \in M_{n\times n}(K)$ gilt:*

(1) p_A ist ein Polynom von Grad n.

(2) Falls $p_A(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_0$ ist, dann ist

$$a_n = (-1)^n, \quad a_{n-1} = (-1)^{n-1}\operatorname{spur}(A) \quad \textit{und} \quad a_0 = \det A.$$

(3) $p_A = p_{A^T}$.

(4) Falls $A = \begin{pmatrix} a_1 & \cdots & * \\ & \ddots & \vdots \\ & & a_n \end{pmatrix}$ oder $A = \begin{pmatrix} a_1 & & \\ \vdots & \ddots & \\ * & \cdots & a_n \end{pmatrix}$ eine obere beziehungsweise untere Dreiecksmatrix ist, dann ist

$$p_A(x) = \prod_{i=1}^{n}(a_i - x).$$

(5) Falls $A = \begin{pmatrix} \boxed{B_1} & * & \cdots & * \\ & \boxed{B_2} & \ddots & \vdots \\ & & \ddots & * \\ & & & \boxed{B_k} \end{pmatrix}$ oder $A = \begin{pmatrix} \boxed{B_1} & & & \\ * & \boxed{B_2} & & \\ \vdots & \ddots & \ddots & \\ * & \cdots & * & \boxed{B_k} \end{pmatrix}$ eine obere beziehungsweise untere Blockdreiecksmatrix ist, sodass die Blöcke B_i quadratische Matrizen sind, dann ist

$$p_A = \prod_{i=1}^{k} p_{B_i}.$$

(6) Falls A und B ähnlich sind, gilt $p_A = p_B$.

Beweis. Wie in Abschnitt 4.3.4 erklärt, ist p_A die Determinante einer Matrix über $K[x]$ und daher ein Element von $K[x]$, also ein Polynom. Um die weiteren Eigenschaften beweisen zu können, müssen wir jedoch die Leibniz Formel genauer anschauen. Betrachten Sie die Matrix in (5.7):

Für die Permutation $\sigma = \mathrm{Id}$ erhalten wir das Produkt $\prod_{i=1}^{n}(a_{ii} - x)$, welches ein Polynom von Grad n ist.

Für jede andere Permutation σ, erhält man ein Produkt, dessen Grad der Anzahl der Diagonalkomponenten entspricht. Diese Anzahl ist gleich der Anzahl der Fixpunkte der Permutation σ. Eine Permutation $\sigma \neq \mathrm{Id}$ kann aber höchstens $n-2$ Fixpunkte haben. Daher gilt

$$\begin{aligned} p_A(x) &= \prod_{i=1}^{n}(a_{ii} - x) + (\text{Polynome von Grad } \leq n-2) \\ &= (-1)^n x^n + (-1)^{n-1}(a_{11} + \cdots + a_{nn})x^{n-1} + (\text{Polynome von Grad } \leq n-2)\,. \end{aligned}$$

Dies zeigt (1) und die ersten zwei Teile von (2). Um den Beweis von (2) zu schliessen, bemerken Sie, dass

$$a_0 = p_A(0) = \det(A - 0 \cdot I_n) = \det A.$$

Die Teile (3), (4) und (5) überlassen wir den Lesern. Für (6) sei $P \in \mathrm{GL}_n(K)$ mit $P^{-1}AP = B$. Dann ist

$$\begin{aligned} p_B(x) = |B - xI_n| = \left|P^{-1}AP - xP^{-1}P\right| &= \left|P^{-1}AP - P^{-1}xI_nP\right| \\ &= \left|P^{-1}(A - xI_n)P\right| \\ &= |A - xI_n| = p_A(x). \end{aligned}$$

□

Bemerkung 5.3.7. Beachten Sie, dass (2) und (6) implizieren, dass die Spur (beziehungsweise die Determinante) von zwei ähnlichen Matrizen gleich ist, was ein alternativer Beweis für Korollar 4.7.8 (beziehungsweise für Korollar 4.7.2).

Wir können nun das charakteristische Polynom eines Endomorphismus definieren.

Definition 5.3.8. Sei V ein endlich-dimensionaler K-Vektorraum und $T \in \mathrm{End}(V)$. Das *charakteristische Polynom* von T ist durch

$$p_T(x) := p_{[T]_\mathcal{B}}$$

definiert, wobei $\mathcal{B}$ eine (geordnete) Basis von V ist.

Laut Eigenschaft (6) in Lemma 5.3.6 und Korollar 3.3.32 ist diese Definition wohldefiniert (das heisst, sie hängt nicht von der Wahl von $\mathcal{B}$ ab).

Laut den Berechnungen (5.5) und (5.6) gilt

$$\lambda \in K \text{ ist ein Eigenwert von } T \iff p_T(\lambda) = 0.$$

Übung 5.3.9. Obwohl dies keine Überraschung ist, zeigen Sie, dass

$$p_A = p_{m_A}$$

gilt.

Wir fassen unsere Ergebnisse in einem Rezept zusammen:

Sei $T \in \operatorname{End}(V)$ mit $\dim V = n < \infty$. Wenn man alle möglichen Eigenvektoren von T finden will (beispielsweise, um zu wissen ob T diagonalisierbar ist oder nicht), macht man Folgendes:

1. Man berechnet alle Nullstellen des charakteristischen Polynoms. Bezeichnen wir diese mit $\lambda_1, \dots, \lambda_k$ (wir werden zeigen $0 \leq k \leq n$). Dann sind das alle Eigenwerte von T.

2. Für jedes $i = 1, \dots, k$ berechnet man $\operatorname{Eig}_T(\lambda_i) = \operatorname{Ker}(T - \lambda_i \operatorname{Id}_V)$. Alle Vektoren in $\operatorname{Eig}_T(\lambda_i)$, die verschieden von Null sind, sind Eigenvektoren von T mit λ_i als Eigenwert.

Bemerkung 5.3.10. Wie alles in diesem Abschnitt, soll man dieses Rezept insbesondere für den Spezialfall von $T = m_A$ für $A \in M_{n \times n}(K)$ gut verstehen.

Beispiel 5.3.11. Sei $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ dieselbe Matrix wie in Beispiel 5.2.11. Wir haben die Eigenvektoren und Eigenwerte von A einfach erraten. Jetzt können wir sie mit unserem Rezept algorithmisch finden. Das charakteristische Polynom ist

$$\begin{aligned} p_A(x) = \det\left(\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} - x\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) &= \det\begin{pmatrix} 2-x & 1 \\ 1 & 2-x \end{pmatrix} \\ &= (2-x)^2 - 1 \qquad = x^2 - 4x + 3 \\ &= (x-1)(x-3). \end{aligned}$$

Wie erwartet sind die Nullstellen von $p_A(x)$ genau die Eigenwerte von A, die wir in Beispiel 5.2.11 gefunden haben. Bemerken Sie auch, dass wir uns unter der Verwendung von Lemma 5.3.6 (2) obige Berechnung ersparen können:

$$p_A(x) \overset{(*)}{=} x^2 - \operatorname{spur}(A)\, x + \det A = x^2 - 4x + 3.$$

Die Gleichheit $(*)$ gilt nämlich für jede 2×2-Matrix.

Um alle Eigenvektoren mit Eigenwert 3 zu finden, betrachten wir

$$\operatorname{Ker}(A - 3I_2) = \operatorname{Ker}\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} = \left\{ \begin{pmatrix} a \\ a \end{pmatrix} \,\middle|\, a \in K \right\} = \operatorname{Sp}\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Um alle Eigenvektoren mit Eigenwert 1 zu finden, betrachten wir

$$\operatorname{Ker}(A - I_2) = \operatorname{Ker}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \operatorname{Sp}\begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Daher ist beispielsweise

$$\left\{ \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$$

eine Basis aus Eigenvektoren und es gilt mit $P = \left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right) = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$, dass

$$P^{-1}AP = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$$

wie man durch eine Berechnung überprüfen kann, falls man nicht an Lemma 5.2.14 glaubt.

Beispiel 5.3.12. Jetzt können wir Übung 5.2.12 auch „algorithmisch" lösen. Wir suchen alle möglichen Eigenvektoren von $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ um herauszufinden, ob $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ diagonalisierbar ist oder nicht. Laut Korollar 5.3.4 und Definition 5.3.5 sind alle möglichen Eigenwerte die Nullstellen von

$$\begin{aligned} p_A(x) = \left| \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} - x \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right| &= \begin{vmatrix} 1-x & 1 \\ 0 & 1-x \end{vmatrix} \\ &= (1-x)^2 - 0 \\ &= (1-x)^2. \end{aligned}$$

Daher ist 1 der einzig mögliche Eigenwert von A. Es folgt, dass die einzig möglichen Eigenvektoren von A, diejenigen Vektoren sind die nicht Null sind und in

$$\operatorname{Ker}(A - 1 \cdot I_2) = \operatorname{Ker}\left(\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \right) = \left\{ \begin{pmatrix} a \\ 0 \end{pmatrix} \,\middle|\, a \in K \right\}$$

liegen. Daher sind

$$\left\{ \begin{pmatrix} a \\ 0 \end{pmatrix} \,\middle|\, a \neq 0 \right\}$$

die möglichen Eigenvektoren von A. Insbesondere existiert keine Basis von K^2, die nur aus Eigenvektoren von A besteht. Also ist A nicht diagonalisierbar.

Bemerkung 5.3.13. Das letzte Beispiel mag zwar einfach sein, man kann jedoch viel von ihm lernen. Stellen Sie sicher, dass sie alle Details in diesem Beispiel verstehen.

Bemerken Sie, dass Proposition 5.2.16 impliziert, dass falls $\lambda_i \neq \lambda_j$ gilt, dann ist $\mathrm{Eig}_T(\lambda_i) \cap \mathrm{Eig}_T(\lambda_j) = \{0\}$. In anderen Worten gilt

$$\mathrm{Eig}_T(\lambda_i) + \mathrm{Eig}_T(\lambda_j) = \mathrm{Eig}_T(\lambda_i) \oplus \mathrm{Eig}_T(\lambda_j)$$

für $\lambda_i \neq \lambda_j$. Die Existenz von mehreren Eigenwerten motiviert die Definition einer direkten Summe mit mehreren Summanden. Dies wird uns ermöglichen eine Zerlegung von V zu finden, durch welche man die Wirkung von $T \in \mathrm{End}(V)$ auf V gut verstehen kann.

5.3.1 Direkte Summen II

Wir möchten direkte Summen bezüglich mehreren Summanden definieren. Um die direkte Summe bezüglich zwei Summanden zu definieren, kann man jede der äquivalenten Bedingungen in Proposition 2.3.40 benutzen. Um Summen bezüglich mehreren Summanden zu definieren, benutzen wir eine Verallgemeinerung von Bedingung (4). Mit einer direkten Verallgemeinerung von Bedingung (3) geht es beispielsweise nicht (vgl. Übung 5.3.16 unten). Erinnern Sie sich an Definition 2.3.36, bevor Sie die nächste Definition lesen.

Definition 5.3.14. Sei V ein Vektorraum über K, $r \in \mathbb{N}$ und $U_1, \ldots, U_r$ Untervektorräume von V. Wir sagen, dass

$$W := U_1 + \cdots + U_r$$

eine *direkte Summe* von $U_1, \ldots, U_r$ ist und schreiben $W = U_1 \oplus \ldots \oplus U_r$, falls jeder Vektor $w \in W$ als eine eindeutige Summe der Form

$$w = u_1 + \cdots + u_r$$

mit $u_i \in U_i$ für $i = 1, \ldots, r$ geschrieben werden kann. Man schreibt dann auch

$$W = \bigoplus_{i=1}^{r} U_i := U_1 \oplus \cdots \oplus U_r.$$

Beispiel 5.3.15. Dies ist eigentlich eine Verallgemeinerung des Basis-Begriffs in folgendem Sinn: Sei $\{v_1, \dots, v_n\}$ eine Basis von V. Dann gilt (z.B. nach Proposition 2.2.35)

$$V = \mathrm{Sp}(v_1) \oplus \cdots \oplus \mathrm{Sp}(v_n).$$

Allgemeiner seien $B_1, \dots, B_r$ Teilmengen von $\{v_1, \dots, v_n\}$, so dass

$$\{v_1, \dots, v_n\} = \bigsqcup_{i=1}^{r} B_i$$

gilt.[2] Dann gilt

$$V = \bigoplus_{i=1}^{r} \mathrm{Sp}(B_i) := \mathrm{Sp}(B_1) \oplus \cdots \oplus \mathrm{Sp}(B_r).$$

Übung 5.3.16. Die folgende Übung zeigt, dass eine direkte Verallgemeinerung von Bedingung (3) in Proposition 2.3.40 nicht geeignet ist, um direkte Summen bezüglich mehrerer Summanden zu definieren:

Finden Sie einen Vektorraum V und Untervektorräume U_1, U_2, U_3, so dass

$$V = U_1 + U_2 + U_3 \quad \text{und} \quad U_1 \cap U_2 \cap U_3 = \{0\}, \quad \text{aber} \quad V \neq U_1 \oplus U_2 \oplus U_3$$

gilt.

Andere Bedingungen hingegen lassen sich auf mehrere Summanden verallgemeinern:

Proposition 5.3.17. *Sei V ein Vektorraum, $r \in \mathbb{N}$ und $U_1, \dots, U_r$ endlich-dimensionale Untervektorräume von V und $W = U_1 + \cdots + U_r$. Die folgenden Bedingungen sind äquivalent:*

(1) Es gilt $W = \bigoplus_{i=1}^{r} U_i$.

(2) Falls $0 = u_1 + \cdots + u_r$ mit $u_i \in U_i$ für $i = 1, \dots, r$, dann ist $u_i = 0$ für alle $i = 1, \dots, r$.

(3) Falls $\mathcal{B}_i$ eine Basis von U_i ist für $i = 1, \dots, r$, dann sind die $\mathcal{B}_i$ paarweise disjunkt und

$$\mathcal{B} = \bigsqcup_{i=1}^{r} \mathcal{B}_i$$

ist eine Basis von W.

(4) Es gilt

$$\sum_{i=1}^{r} \dim U_i = \dim W.$$

[2]Dies nennt man eine Zerlegung von $\{v_1, \dots, v_n\}$.

Bemerkung 5.3.18. Die Äquivalenzen (1) $\Leftrightarrow$ (2) $\Leftrightarrow$ (3) gelten auch ohne die Annahme dass die Untervektorräume endlich-dimensional sind.

Beweis. Die Implikation (1) $\Rightarrow$ (2) ist unmittelbar, da (2) besagt, dass 0 eine eindeutige Darstellung als

$$0 = u_1 + \cdots + u_r$$

mit $u_i \in U_i$ für $i = 1, \ldots, r$ hat.

Für (2) $\Rightarrow$ (3): Seien $\mathcal{B}_i = (v_1^{(i)}, \ldots, v_{s_i}^{(i)})$ für $i = 1, \ldots, r$. Wir wissen, dass $\mathcal{B}$ den Untervektorraum W erzeugt. (Wieso?) Daher zeigen wir, dass $\mathcal{B}$ linear unabhängig ist. Angenommen

$$\sum_{i=1}^{r} \left(a_1^{(i)} v_1^{(i)} + \cdots + a_{s_i}^{(i)} v_{s_i}^{(i)} \right) = 0$$

für $a_1^{(i)}, \ldots, a_{s_i}^{(i)} \in K$ für alle $i = 1, \ldots, r$. Da $a_1^{(i)} v_1^{(i)} + \cdots + a_{s_i}^{(i)} v_{s_i}^{(i)} \in U_i$ für $i = 1, \ldots, r$ ist, folgt aus (2), dass

$$a_1^{(i)} v_1^{(i)} + \cdots + a_{s_i}^{(i)} v_{s_i}^{(i)} = 0$$

für alle $i = 1, \ldots, r$. Da $(v_1^{(i)}, \ldots, v_{s_i}^{(i)})$ eine Basis von U_i ist, folgt

$$a_1^{(i)} = \cdots = a_{s_i}^{(i)} = 0$$

für $i = 1, \ldots, r$. Dies zeigt die lineare Unabhängigkeit von $\mathcal{B}$. Genau genommen zeigt es noch mehr, nämlich die lineare Unabhängigkeit der Liste $(v_1^{(1)}, \ldots, v_{s_1}^{(1)}, \ldots, v_1^{(r)}, \ldots, v_{s_r}^{(r)})$. Daraus folgt dann insbesondere, dass in dieser Liste kein Element zweimal vorkommt, also sind die $\mathcal{B}_i$ paarweise disjunkt. (3) folgt. Wir zeigen nun (3) $\Rightarrow$ (2): Seien $u_i \in U_i$ mit

$$0 = u_1 + \cdots + u_r.$$

Mit derselben Notation wie zuvor, können wir jeden Vektor u_i schreiben als

$$u_i = \sum_{j=1}^{s_i} a_j^{(i)} v_j^{(i)}$$

mit $v_j^{(i)} \in \mathcal{B}_i$ für alle $i = 1, \ldots, r$ und $j = 1, \ldots, s_i$. Dann gilt

$$0 = \sum_{i=1}^{r} \sum_{j=1}^{s_i} a_j^{(i)} v_j^{(i)}$$

und daher sind $a_j^{(i)} = 0$ für alle $i = 1, \dots, r$ und $j = 1, \dots, s_i$. Es folgt, dass $u_i = 0$ für alle $i = 1, \dots, r$, was (2) zeigt. Die Implikation (3) $\implies$ (4) ist unmittelbar, denn

$$\dim W = |\mathcal{B}| = \sum_{i=1}^{r} |\mathcal{B}_i| = \sum_{i=1}^{r} \dim U_i.$$

Wir zeigen (4) $\Rightarrow$ (3): Sei $\mathcal{B}_i$ eine Basis von U_i für $1 \leq i \leq r$. Aus $W = U_1 + \cdots + U_r$ folgt, dass $\mathcal{B} = \bigcup_{i=1}^r \mathcal{B}_i$ den Untervektorraum W erzeugt, also ist $\dim W \leq |\mathcal{B}|$. Andererseits ist

$$|\mathcal{B}| \leq \sum_{i=1}^{r} |\mathcal{B}_i| = \sum_{i=1}^{r} \dim U_i \overset{(4)}{=} \dim W.$$

Es folgt, dass $|\mathcal{B}| = \sum_i |\mathcal{B}_i| = \dim W$ und somit müssen die $\mathcal{B}_i$ paarweise disjunkt sein. Aus Satz 2.3.15 (Gleichgewicht!) folgt, dass $\mathcal{B}$ eine Basis von W ist. Dies zeigt (3).

Es bleibt noch (2) $\Rightarrow$ (1) zu zeigen: Seien $u_i, u_i' \in U_i$ für $i = 1, \dots, r$, so dass

$$w = u_1 + \cdots + u_r = u_1' + \cdots + u_r'.$$

Dann gilt

$$0 = \sum_{i=1}^{r} (u_i - u_i').$$

Da $u_i - u_i' \in U_i$ ist, folgt nach (2), dass $u_i - u_i' = 0$ beziehungsweise $u_i = u_i'$, was (1) zeigt. $\square$

Bemerkung 5.3.19. Wie man „richtig“ Bedingung (3) aus Proposition 2.3.40 verallgemeinert, werden Sie in der Serie sehen.

Wir gehen nun zu unserem Thema zurück:

Korollar 5.3.20. *Seien $T \in \mathrm{End}(V)$, $r \in \mathbb{N}$ und $\lambda_1, \dots, \lambda_r \in K$ paarweise verschiedene Eigenwerte von T. Dann erfüllen die Untervektorräume $\mathrm{Eig}_T(\lambda_i)$ die äquivalenten Bedingungen von Proposition 5.3.17.*

Beweis. Bedingung (2) ist dank Proposition 5.2.16 erfüllt. $\square$

Bemerkung 5.3.21. Jetzt, da Sie wissen, dass Proposition 5.3.17 für $U_i = \mathrm{Eig}_T(\lambda_i)$ gilt, lesen Sie nochmals den Beweis mit $U_i = \mathrm{Eig}_T(\lambda_i)$ im Kopf durch!

Korollar 5.3.22. *Sei $T \in \mathrm{End}(V)$ mit $\dim V = n$ und seien $\lambda_1, \dots, \lambda_r$ die Eigenwerte von T. Dann gilt*

$$T \textit{ ist diagonalisierbar} \iff \sum_{i=1}^{r} \dim \mathrm{Eig}_T(\lambda_i) = \dim V.$$

Der Leser kann probieren, dieses Korollar jetzt schon zu beweisen (es sollte nicht sehr schwierig sein). Wir beweisen es erst im nächsten Abschnitt, und bis dahin werden wir es nicht verwenden.

5.3.2 Diagonaliserbarkeit I

Wir fragen uns, was wir über einen Endomorphismus T von seinem charakteristischen Polynom $p_T(x)$ lernen können. Zuerst untersuchen wir, was $p_T(x)$ über die Eigenräume $\mathrm{Eig}_T(\lambda_i)$ und die Diagonalisierbarkeit von T aussagt. Die erste Tatsache, die wir bemerken sollten, ist, dass das charakteristische Polynom allein nicht entscheiden kann, ob T diagonalisierbar ist.

Beispiel 5.3.23. Die Matrizen $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ haben dasselbe charakteristische Polynom $(x-1)^2$. Die erste Matrix ist diagonalisierbar (sie ist sogar schon diagonal), die zweite Matrix hingegen ist nicht diagonalisierbar, wie wir in Beispiel 5.3.12 gesehen haben.

Beispiel 5.3.24 (Jordan-Block)**.** Seien $\lambda \in K$ und $n \in \mathbb{N}$. Die Matrix

$$J = J_{\lambda,n} = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & \ddots & & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \lambda & 1 \\ 0 & \cdots & \cdots & 0 & \lambda \end{pmatrix} \in M_{n\times n}(K)$$

heisst *Jordan-Block*. Das charakteristische Polynom dieser Matrix ist $(\lambda - x)^n$ und man könnte wie in Beispiel 5.3.12 zeigen, dass J genau dann diagonalisierbar ist, wenn $n = 1$ (also $J = (\lambda)$). Diese Matrizen spielen später eine wichtige Rolle. Bemerken Sie, dass J und die skalare Matrix[3] $\begin{pmatrix} \lambda & & \\ & \ddots & \\ & & \lambda \end{pmatrix} \in M_{n\times n}(\mathbb{R})$ dasselbe charakteristische Polynom haben.

[3]Matrizen der Form $\begin{pmatrix} \lambda & & \\ & \ddots & \\ & & \lambda \end{pmatrix}$ heissen *skalare Matrizen*. Grund dafür ist, dass die Menge $\left\{ \begin{pmatrix} \lambda & & \\ & \ddots & \\ & & \lambda \end{pmatrix} \middle| \lambda \in K \right\}$ versehen mit der Matrixaddition und der Matrixmultiplikation isomorph ist zu K als Körper.

Das charakteristische Polynom gibt uns Teilinformationen bezüglich Diagonalisierbarkeit und Eigenräumen. Zum Beispiel gibt uns das folgenden Lemma ein Kriterium für „nicht-Diagonalisierbarkeit“:

Lemma 5.3.25. *Sei V ein n-dimensionaler Vektorraum über K und $T \in \operatorname{End}(V)$ (bzw. $A \in M_{n\times n}(K)$). Falls T (bzw. A) diagonalisierbar ist, dann zerfällt $p_T(x)$ (bzw. $p_A(x)$) in Linearfaktoren.*

Bemerkung 5.3.26. Wie immer folgt der Fall $A \in M_{n\times n}(K)$ aus dem Fall $T \in \operatorname{End}(V)$ mit $\dim V = n$, durch die lineare Abbildung $T = m_A$. Daher erwähnen wir ab jetzt nur den Fall einer linearen Abbildung, obwohl der Fall einer Matrix häufiger in Anwendungen und Beispielen vorkommt.

Beweis von Lemma 5.3.25. Diagonalisierbarkeit von T impliziert die Existenz einer Basis $\mathcal{B} \subseteq V$ mit

$$[T]_{\mathcal{B}} = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

diagonal. Daher gilt

$$\begin{aligned} p_T(x) = |[T]_{\mathcal{B}} - xI_n| &= \left| \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} - xI_n \right| \\ &= \left| \begin{pmatrix} (\lambda_1 - x) & & \\ & \ddots & \\ & & (\lambda_n - x) \end{pmatrix} \right| \\ &= \prod_{i=1}^{n} (\lambda_i - x), \end{aligned}$$

was wir zeigen wollten. □

Beispiel 5.3.27. Die Umkehrung von Lemma 5.3.25 gilt nicht! Zum Beispiel hat $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ das charakteristische Polynom $p_A(x) = (x-1)^2$.

Beispiel 5.3.28 (Diagonalisierbarkeit hängt von K ab)**.** Wir betrachten

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Es gilt $p_A(x) = x^2 + 1$. Falls $K = \mathbb{Q}$ oder $\mathbb{R}$ ist, dann zerfällt $p_A(x)$ nicht in Linearfaktoren und daher ist A **als Matrix über $\mathbb{R}$ oder $\mathbb{Q}$** nicht diagonalisierbar. Das heisst, es existiert keine Matrix $P \in M_{2\times 2}(\mathbb{R})$ mit

$$PAP^{-1}$$

diagonal. Hingegen zerfällt p_A über $\mathbb{C}$ als $(x+i)(x-i)$ und daher gibt es die Möglichkeit, dass A über $\mathbb{C}$ diagonalisierbar ist. Sie ist in der Tat diagonalisierbar, da sie zwei verschiedene Eigenwerte hat (vgl. Korollar 5.2.17). Das heisst, es existiert $P \in M_{2\times 2}(\mathbb{C})$, so dass

$$P^{-1}AP$$

diagonal ist. Man kann überprüfen, dass $P = \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}$ dies erfüllt.

Übung 5.3.29. Die Matrix in Beispiel 5.3.28 ist ein Spezialfall einer Rotationsmatrix. Für $\theta \in [0, 2\pi)$ sei

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Entscheiden Sie für welche θ die Matrix R_θ diagonalisierbar ist über $\mathbb{R}$ und für welche θ sie über $\mathbb{C}$ diagonalisierbar ist. Ergeben die Eigenwerte über $\mathbb{C}$ Sinn, wenn man denkt, dass R_θ eine Rotation mit θ darstellt?

Bemerkung 5.3.30. Beispiel 5.3.28 und Übung 5.3.29 sind auch grundlegend. Stellen Sie sicher, dass sie alle Details verstanden haben.

5.3.3 Geometrische Vielfachheit ist kleiner als algebraische Vielfachheit

Hier ist eine erste Verbindung zwischen dem charakteristischen Polynom und Eigenräumen. Erinnern Sie sich an die Definition der Vielfachheit (Definition 1.4.16).

In diesem Abschnitt bezeichnet V stets einen K-Vektorraum und $T \in \operatorname{End}(V)$.

Definition 5.3.31. Sei $\lambda \in K$. Die *geometrische Vielfachheit* von λ in T definieren wir durch

$$m_g(\lambda) = m_g(T, \lambda) = \dim \operatorname{Eig}_T(\lambda)$$

und die *algebraische Vielfachheit* von λ in T definieren wir durch

$$m_a(\lambda) = m_a(T, \lambda) = \mu(p_T \mid \lambda).$$

Man sollte bei diesen Namen an Folgendes denken: Wenn man sich T geometrisch vorstellt, dann entspricht die Dimension des Unterraums $\operatorname{Eig}_T(\lambda)$, auf welchem T durch Multiplikation mit λ agiert, genau der geometrischen Vielfachheit $m_g(\lambda)$. Analog ist das charakteristische Polynom in gewisser Weise ein „algebraisches Objekt", für welches der Skalar λ als Eigenwert mit Vielfachheit $m_a(\lambda) = \mu(p_T \mid \lambda)$ vorkommt.

Lemma 5.3.32. *Sei $T \in \mathrm{End}(V)$. Für jedes $\lambda \in K$ gilt*

$$m_g(\lambda) \leq m_a(\lambda). \tag{5.8}$$

Beweis. Bei diesem Lemma ist es nützlich an zwei verschiedene Fälle zu denken. Also machen wir das auch hier in diesem Beweis.

Fall I: Angenommen λ ist kein Eigenwert von T. Dann sind beide Seiten von (5.8) gleich 0.

Fall II: Angenommen λ ist ein Eigenwert von T mit $r := m_g(\lambda) = \dim \mathrm{Eig}_T(\lambda) > 0$. Sei $v_1, \ldots, v_r$ eine Basis von $\mathrm{Eig}_T(\lambda)$ und

$$\mathcal{B} = (v_1, \ldots, v_r, w_1, \ldots, w_{\dim V - r})$$

eine Erweiterung zu einer geordneten Basis von V. Wir berechnen $p_T(\lambda)$ mittels $\mathcal{B}$: Sei $n = \dim V$. Es gilt

$$[T]_{\mathcal{B}} = \left(\begin{array}{ccc|c} \lambda & & & \\ & \ddots & & (*) \\ & & \lambda & \\ \hline & 0_{n-r\times r} & & C \end{array}\right)$$

für eine Matrix $C \in M_{(n-r)\times(n-r)}(K)$. Der Eintrag $(*)$ steht für eine Matrix der Grösse $r \times (n-r)$ (deren Einträge für unsere Berechnung nicht relevant sind). Daher ist

$$p_T(x) = p_{[T]_{\mathcal{B}}}(x) = \left|\left(\begin{array}{ccc|c} (\lambda - x) & & & \\ & \ddots & & (*) \\ & & (\lambda - x) & \\ \hline & 0_{n-r\times r} & & C - xI_{n-r\times n-r} \end{array}\right)\right|$$

und folglich ist

$$p_T(x) = (\lambda - x)^r \cdot p_C(x).$$

Dies impliziert, dass

$$m_a(\lambda) = \mu(p_T \mid \lambda) \geq r.$$

□

Beispiel 5.3.33. Unser Lieblingsbeispiel

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

zeigt, dass Ungleichheit in Lemma 5.3.32 möglich ist: Hier gilt $m_g(A, 1) = 1$ und $m_a(A, 1) = 2$.

Bemerkung 5.3.34. Wenn λ kein Eigenwert von T ist, dann gilt $m_g(T, \lambda) = m_a(T, \lambda) = 0$.

5.3.4 Diagonalisierbarkeit II

Wir sind jetzt bereit für mehrere Charakterisierungen von Diagonalisierbarkeit (die erste hatten wir in Lemma 5.2.7 gesehen und andere werden noch folgen).

Satz 5.3.35. *Sei V ein n-dimensionaler Vektorraum über K und $T \in \mathrm{End}(V)$. Die folgenden Aussagen sind äquivalent:*

(1) T ist diagonalisierbar.

(2) Es existiert eine Basis von V, die nur Eigenvektoren enthält.

(3) Das charakteristische Polynom p_T zerfällt in Linearfaktoren und die geometrische Vielfachheit ist gleich der algebraischen Vielfachheit. Das heisst, dass

$$m_g(\lambda_i) = m_a(\lambda_i)$$

für jedes $i \in \{1, \ldots, k\}$.

(4) Es gilt

$$\sum_{i=1}^{k} \dim \mathrm{Eig}_T(\lambda_i) = \dim V.$$

(5) Es gilt

$$V = \bigoplus_{i=1}^{k} \mathrm{Eig}_T(\lambda_i).$$

Beweis. Die Äquivalenz von (1) und (2) ist der Inhalt von Lemma 5.2.7; sie ist hier der Vollständigkeit halber aufgelistet. Angenommen (1) und (2) gelten. Laut Lemma 5.3.25 zerfällt das charakteristische Polynom p_T in Linearfaktoren. Ausserdem trägt jedes Element einer Basis von Eigenvektoren zu einem der $m_g(\lambda_i)$ bei und somit impliziert die Existenz einer solchen Basis, dass

$$n \leq \sum_{i=1}^{k} m_g(\lambda_i) \overset{\textit{Lemma } 5.3.32}{\leq} \sum_{i=1}^{k} m_a(\lambda_i) = n$$

wobei die letzte Gleichheit verwendet, dass p_T in Linearfaktoren zerfällt. Also muss für $\sum_{i=1}^{k} m_a(\lambda_i) = \sum_{i=1}^{k} m_g(\lambda_i)$ gelten. Da $m_a(\lambda_i) \geq m_g(\lambda_i)$ gilt, müssen die einzelnen Summanden gleich sein, was gleichbedeutend der Aussage in (3) ist.

Die Implikation (3) $\implies$ (4) folgt aus

$$\dim(V) = n = \sum_{i=1}^{k} m_a(\lambda_i) = \sum_{i=1}^{k} m_g(\lambda_i) = \sum_{i=1}^{k} \dim(\mathrm{Eig}_T(\lambda_i)). \tag{5.9}$$

Schlussendlich folgen die Implikationen (4) $\implies$ (5) und (5) $\implies$ (2) aus Proposition 5.3.17. $\square$

Beispiel 5.3.36 (Potenzen einer Matrix)**.** Vielleicht fragen Sie sich, was bringt es überhaupt, eine Matrix zu diagonalisieren? Nun, eine Anwendung davon ist die Berechnung der Potenzen einer Matrix.

Erinnern Sie sich, falls $D = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix}$ diagonal ist, dann gilt $D^k = \begin{pmatrix} a_1^k & & \\ & \ddots & \\ & & a_n^k \end{pmatrix}$ für alle $k \in \mathbb{N}$. Falls eine Matrix A diagonalisierbar ist, d.h. falls es $P \in \mathrm{GL}_n(K)$ gibt mit $P^{-1}AP = D$ (bzw. $A = PDP^{-1}$), dann ist

$$A^k = (PDP^{-1})^k = PD\cancel{P^{-1}}\cancel{P}D\cancel{P^{-1}} \cdots \cancel{P}DP^{-1} = PD^kP^{-1}.$$

Das heisst es ist leicht, die Potenzen einer diagonalisierbaren Matrix zu berechnen.

Beispiel 5.3.37 (Verallgemeinerung von Beispiel 3.6.23)**.** Für die nächste Anwendung, die wir eigentlich schon kennen, betrachten wir eine rekursiv definierte Folge $\mathcal{A} = (a_1, a_2, \ldots) \in K^\infty$, die durch

$$a_1 = A_1, \ \ldots \ , a_k = A_k, \tag{5.10}$$

$$a_n = \alpha_k a_{n-1} + \cdots + \alpha_1 a_{n-k} \text{ für } n \geq k+1, \tag{5.11}$$

gegeben ist, wobei $A_1, \ldots, A_k \in K$ und $\alpha_1, \ldots, \alpha_k \in K$ sind. Hier ist ein Trick[4], der diese Rekursion mit Matrizen verbindet:

Es gilt für alle $n \geq k+1$, dass

$$\underbrace{\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & & & 1 \\ \alpha_1 & \alpha_2 & & \cdots & \alpha_k \end{pmatrix}}_{=: M_{\alpha_1, \ldots, \alpha_k}} \begin{pmatrix} a_{n-k} \\ \\ \vdots \\ \\ a_{n-1} \end{pmatrix} = \begin{pmatrix} a_{n-(k-1)} \\ \vdots \\ a_{n-1} \\ a_n \end{pmatrix}.$$

[4]Eigentlich ist es kein Trick! Dies ist lediglich die Matrix-Darstellung der Verschiebungsabbildung $S : K^\infty \to K^\infty$ eingeschränkt auf den UVR von K^∞ all jener Folgen, die die Rekursionsgleichung (5.11) erfüllen (vgl. Kapitel 0 und Beispiel 3.6.23).

Für $k = 2$ und $\alpha_1 = \alpha_2 = 1$ sind die Folgen in K^∞, die (5.11) erfüllen, genau die Fibonacci-Folgen, also Elemente von **Fib**. Weiter gilt $M_{1,1} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ (vgl. Beispiel 3.6.23).

Kehren wir nun zurück zum allgemeinen Fall und nehmen wir an, dass wir $M_{\alpha_1,\dots,\alpha_k}$ diagonalisieren können. Dann können wir eine Formel für die Folgenglieder von $\mathcal{A}$ in Abhängigkeit von $A_1, \dots, A_k$ finden:

Seien $P \in \mathrm{GL}_n(K)$ und $e_1, \dots, e_k \in K$ gegeben, sodass $M_{\alpha_1,\dots,\alpha_k} = PDP^{-1}$ mit

$$D = \begin{pmatrix} e_1 & & \\ & \ddots & \\ & & e_k \end{pmatrix}.$$

Dann gilt für die Folge $\mathcal{A} = (a_1, a_2, \dots)$, die durch (5.10) und (5.11) definiert ist, dass

$$\begin{pmatrix} a_{n-(k-1)} \\ \vdots \\ a_n \end{pmatrix} = M^{n-k}_{\alpha_1,\dots,\alpha_k} \begin{pmatrix} A_1 \\ \vdots \\ A_k \end{pmatrix} = P \begin{pmatrix} e_1^{n-k} & & \\ & \ddots & \\ & & e_k^{n-k} \end{pmatrix} P^{-1} \begin{pmatrix} A_1 \\ \vdots \\ A_k \end{pmatrix}.$$

Da auf der rechten Seite alles gegeben ist, gibt es also eine explizite Formel für $\mathcal{A}$. Sie werden in der Serie und in der Übungsstunde noch konkrete Beispiele dazu sehen.

Sie könnten sich fragen, wie wir eine Formel finden können, falls $M_{\alpha_1,\dots,\alpha_k}$ nicht diagonalisierbar ist. Das ist eine gute Frage. Zumindest über $\mathbb{C}$ werden wir sehen, dass $M_{\alpha_1,\dots,\alpha_k}$ ähnlich zu einer Matrix in Jordan-Form ist (wir werden noch sehen was das heisst) und Potenzen von Matrizen in Jordan-Form lassen sich relativ einfach berechnen.

Wir beenden den Abschnitt 5.3 mit der interessanten Charakterisierung von *simultan diagonalisierbaren Endomorphismen*.

Definition 5.3.38. Zwei diagonalisierbare Endomorphismen $S, T \in \mathrm{End}(V)$ auf einem n-dimensionalen K-Vektorraum heissen *simultan diagonalisierbar*, falls es eine (geordnete) Basis $\mathcal{B}$ von V gibt, sodass $[S]_\mathcal{B}$ und $[T]_\mathcal{B}$ beide diagonal sind.

Im nächsten Satz werden wir charakterisieren, wann dies der Fall ist. Der Beweis dieses Satzes ist toll, da wir das erste Mal die Idee von invarianten Untervektorräumen benutzen.

Definition 5.3.39. Sei $T \in \mathrm{End}(V)$. Ein Untervektorraum von V heisst *T-invariant*, falls

$$T(U) \subseteq U.$$

Beispiel 5.3.40. Jeder Untervektorraum der von Eigenvektoren von T erzeugt wird, ist T-invariant. Insbesondere sind alle Eigenräume invariant.

Satz 5.3.41. *Zwei diagonalisierbare Endomorphismen $S, T \in \mathrm{End}(V)$ auf einem n-dimensionalen K-Vektorraum sind genau dann simultan diagonalisierbar, wenn sie miteinander kommutieren, das heisst, wenn $S \circ T = T \circ S$.*

Beweis. Die einfache Richtung $\implies$ haben Sie in der Serie gesehen. Nehmen wir nun an, dass $S \circ T = T \circ S$. Da T und S beide diagonalisierbar sind, haben wir die Zerlegung von V als

$$V = \mathrm{Eig}_T(\lambda_1) \oplus \cdots \oplus \mathrm{Eig}_T(\lambda_k) \tag{5.12}$$

$$V = \mathrm{Eig}_S(\eta_1) \oplus \cdots \oplus \mathrm{Eig}_S(\eta_\ell) \tag{5.13}$$

für $\ell, k \in \mathbb{N}$. Beachten Sie, dass $\mathrm{Eig}_S(\eta)$ S-invariant ist für jedes $\eta \in K$ und $\mathrm{Eig}_T(\lambda)$ ist T-invariant für jedes $\lambda \in K$. Der Schlüssel zu diesem Beweis ist, dass $S \circ T = T \circ S$ impliziert, dass diese Eigenräume invariant sind bezüglich T und S. Genauer gesagt: Sei $v \in \mathrm{Eig}_T(\lambda)$, dann gilt

$$T(Sv) = S(Tv) = S(\lambda v) = \lambda(Sv)$$

und daher ist $Sv \in \mathrm{Eig}_T(\lambda)$. In anderen Worten ist $\mathrm{Eig}_T(\lambda)$ sowohl S als auch T-invariant für jedes $\lambda \in K$. Genau so ist $\mathrm{Eig}_S(\eta)$ für jedes $\eta \in K$ sowohl S als auch T-invariant. Äquivalent möchten wir zeigen, dass wir V als direkte Summe

$$V = \bigoplus_{j=1}^{m} V_j$$

schreiben können, sodass für jedes $j = 1, \ldots, m$ Skalare $\lambda, \eta \in K$ existieren mit $V_j \subseteq \mathrm{Eig}_T(\lambda)$ und $V_j \subseteq \mathrm{Eig}_S(\eta)$. Wenn wir dann von jedem V_j eine Basis $\mathcal{B}_j$ auswählen und diese zu einer Basis $\mathcal{B} = \bigcup_j \mathcal{B}_j$ von V zusammenfügen, dann sind $[S]_\mathcal{B}$ und $[T]_\mathcal{B}$ beide diagonal. Um die Notation zu vereinfachen, schreiben wir

$$W := \mathrm{Eig}_T(\lambda)$$

für $\lambda := \lambda_i$ für irgendein $i \in \{1, \ldots, k\}$. Wir zeigen, dass

$$W = \bigoplus_j W \cap \mathrm{Eig}_S(\eta_j).$$

Sei $w \in W$. Laut (5.13) hat w eine eindeutige Schreibweise als

$$w = u_1 + \cdots + u_\ell$$

mit $u_i \in \mathrm{Eig}_S(\eta_i)$ für $i = 1, \ldots, \ell$. Für Tw gilt, dass

$$\lambda u_1 + \cdots + \lambda u_\ell = \lambda w = Tw = Tu_1 + \cdots + Tu_\ell. \tag{5.14}$$

Da $\mathrm{Eig}_S(\eta_j)$ T-invariant ist für alle $j = 1, \ldots, \ell$, gilt $Tu_j \in \mathrm{Eig}_S(\eta_j)$. Daher sind beide Seiten von (5.14) Schreibweisen von Tw bezüglich der direkten Summe (5.13). In anderen Worten gilt $Tu_j = \lambda u_j$ für alle $j = 1, \ldots, \ell$. Dies impliziert, dass $u_j \in W \cap \mathrm{Eig}_S(\eta_j)$ für alle $j = 1, \ldots, \ell$. Da die Schreibweise eindeutig ist, zeigt dies

$$W = \bigoplus_{j=1}^{\ell} W \cap \mathrm{Eig}_S(\eta_j).$$

Da dies für jedes λ_i gilt, erhalten wir

$$V = \bigoplus_{i=1}^{k} \bigoplus_{j=1}^{\ell} \left(\mathrm{Eig}_T(\lambda_i) \cap \mathrm{Eig}_S(\eta_j)\right),$$

was wir zeigen wollten. □

5.4 Fun with Flags - Trigonalisierung

Wir fahren fort mit dem folgenden Leitsatz: Was können wir über einen Endomorphismus T aussagen, wenn wir gewisse Informationen über das charakteristische Polynom von T haben? Mit „was können wir über T aussagen“ ist gemeint, dass wir eine Basis $\mathcal{B}$ finden wollen, sodass $[T]_\mathcal{B}$ besonders einfach ist. (Wenn wir statt einem Endomorphismus T eine Matrix A betrachten, bedeutet dies, eine Matrix B zu finden, welche ähnlich zu A ist und eine besonders einfache Form hat).

Satz 5.3.35 sagt uns, dass wir T diagonalisieren können genau dann, wenn p_T in Linearfaktoren zerfällt und für jeden Eigenwert die algebraische Vielfachheit gleich der geometrischen Vielfachheit ist. Was können wir sagen, wenn wir nur wissen, dass p_T in Linearfaktoren zerfällt? In diesem Abschnitt werden wir sehen dass wir dann T *trigonalisieren* können, das heisst wir können eine Basis $\mathcal{B}$ finden, sodass $[T]_\mathcal{B}$ eine Dreiecksmatrix ist.[5]

Später (zum Beispiel in der Jordan-Form) werden wir uns fragen, ob wir eine besonders einfache Dreiecksmatrix finden können, die T darstellt. Noch später versuchen wir dann, weitere Bedingungen an die Basis $\mathcal{B}$ zu stellen (zum Beispiel werden wir eine orthogonale Basis suchen).

[5] Wir haben festgestellt, dass diagonale und Dreiecksmatrizen in diesem Skript noch nicht definiert wurden. Wir haben dies nun nachträglich eingefügt, siehe Definition 3.3.27.

Wenn nicht anders spezifiziert, ist während dieses Abschnitts ist V ein n-dimensionaler K-Vektorraum.

Definition 5.4.1. Ein Endomorphismus $T \in \mathrm{End}(V)$ heisst *trigonalisierbar*, falls es eine geordnete Basis $\mathcal{B} \subseteq V$ gibt, so dass $[T]_{\mathcal{B}}$ eine obere Dreiecksmatrix ist.

Übung 5.4.2. Ein Endomorphismus $T \in \mathrm{End}(V)$ ist trigonalisierbar genau dann, wenn es eine geordnete Basis $\mathcal{B} \subseteq V$ gibt, so dass $[T]_{\mathcal{B}}$ eine untere Dreiecksmatrix ist.

Definition 5.4.3. Eine *Fahne* (oder *Flagge*) in V ist eine Folge von Untervektorräumen von V

$$\{0_V\} = V_0 \subseteq V_1 \subseteq \ldots \subseteq V_n = V$$

mit $\dim V_i = i$ für alle $i = 0, \ldots, n$. Eine Fahne heisst T-*invariant* bezüglich $T \in \mathrm{End}(V)$, falls V_i T-invariant ist für alle $i = 0, \ldots, n$.

Das Analogon von Satz 5.3.35 ist folgende Proposition:

Proposition 5.4.4. *[Analogon zu Satz 5.3.35] Für einen Endomorphismus $T \in \mathrm{End}(V)$ gilt:*

$$T \textit{ ist trigonalisierbar} \iff \textit{Es existiert eine } T\textit{-invariante Fahne in } V.$$

Beweis. Dies ist eine Übung mit der Definition der Darstellungsmatrix:

„ $\Longrightarrow$ “: Wir nehmen an, dass T trigonalisierbar ist und betrachten eine geordnete Basis $\mathcal{B} = (v_1, \ldots, v_n)$, so dass $[T]_{\mathcal{B}}$ eine obere Dreiecksmatrix ist. Sei

$$V_i := \mathrm{Sp}(v_1, \ldots, v_i).$$

Dann ist

$$\{0_V\} = V_0 \subseteq V_1 \subseteq \ldots \subseteq V_n = V$$

eine T-invariante Fahne.

„ $\Longleftarrow$ “: Falls $\{0\} = V_0 \subseteq V_1 \subseteq \ldots \subseteq V_n$ eine T-invariante Fahne ist, können wir $v_1, \ldots, v_n$ wählen, so dass $V_i = \mathrm{Sp}(v_1, \ldots, v_i)$ ist. Dann ist $[T]_{\mathcal{B}}$ eine obere Dreiecksmatrix für $\mathcal{B} = (v_1, \ldots, v_n)$. □

Ziel in diesem Abschnitt ist es, die nachfolgende Charakterisierung zu beweisen. Erinnern Sie sich daran, dass V ein n-dimensionaler K-Vektorraum ist.

Satz 5.4.5 (Trigonalisierung)**.** *Für $T \in \mathrm{End}(V)$ gilt:*

$$T \textit{ ist trigonalisierbar} \iff p_T \textit{ zerfällt in Linearfaktoren (über } K).$$

Bemerkung 5.4.6. Wie wir in Beispiel 5.3.28 gesehen haben, hängt es stark vom Körper ab, ob eine Matrix/ein Endomorphismus diagonalisierbar ist. Dasselbe gilt für Trigonalisierbarkeit. Für die Matrix A aus Beispiel 5.3.28 zerfällt p_A über $\mathbb{C}$, aber nicht über $\mathbb{R}$.

Aus dem Fundamentalsatz der Algebra 1.4.18 folgt:

Korollar 5.4.7. *Falls $K = \mathbb{C}$, dann ist jeder Endomorphismus trigonalisierbar.*

Bemerkung 5.4.8. Später können wir dieses Korollar mit der Jordan-Form noch verfeinern.

Der Standard-Beweis von Satz 5.4.5 benutzt Quotientenräume jedoch ohne Quotientenräume zu erwähnen. Das macht den Beweis etwas technisch und unübersichtlich und macht es insbesondere schwierig sich an den Beweis zu erinnern. Wir haben Glück, da wir die Sprache der Quotientenräume kennen. Hier sind also einige Vorbereitungen bezüglich Quotientenräumen, die wir für den Beweis brauchen.

5.4.1 Vorbereitung mit Quotientenräumen

Übung 5.4.9. Seien V ein K-Vektorraum und $T \in \mathrm{End}(V)$. Sei U ein T-invarianter Untervektorraum.

(1) Zeigen Sie, dass die Abbildung

$$\begin{aligned} T_{V/U}: \quad V/U &\to V/U \\ v + U &\mapsto Tv + U \end{aligned}$$

ein wohl-definierter Endomorphismus von V/U ist.

(2) Wir nehmen an, dass V endlich-dimensional ist und $\mathcal{B} = (v_1, \ldots, v_n)$ eine geordnete Basis ist, so dass $U = \mathrm{Sp}(v_1, \ldots, v_r)$ für ein $0 < r < n$. Zeigen Sie:

(a) $\mathcal{B}_{V/U} := (v_{r+1} + U, \ldots, v_n + U)$ ist eine Basis von V/U.

(b) Es existieren $A \in M_{r\times r}(K)$ und $B \in M_{(n-r)\times(n-r)}(K)$, so dass

$$[T]_{\mathcal{B}} = \left(\begin{array}{c|c} A & (*) \\ \hline 0 & B \end{array}\right), \tag{5.15}$$

$$[T_{V/U}]_{\mathcal{B}_{V/U}} = B, \tag{5.16}$$

$$[T|_U]_{\mathcal{B}_U} = A, \tag{5.17}$$

wobei $\mathcal{B}_U = (v_1, \ldots, v_r)$ ist.

(c) Folgern Sie, dass $p_T = p_{T|_U} \cdot p_{T_{V/U}}$.

Lösung von Übung 5.4.9: (1) und (2a), (2b) haben Sie in der Serie gesehen. Wir folgern also (2c):

$$p_T \overset{(5.15),\,\text{Lemma 5.3.6}}{=} p_A p_B \overset{(5.16),\,(5.17)}{=} p_{T|_U} p_{T_{V/W}}.$$

Übung 5.4.10. Seien V ein K-Vektorraum, U ein Untervektorraum und

$$\begin{aligned} \pi_U : V &\to V/U \\ v &\mapsto v + U \end{aligned}$$

die kanonische Abbildung. Erinnern Sie sich, dass für einen Untervektorraum $W \subseteq V$ das Bild

$$\pi_U(W) = \{\pi_U(w) \mid w \in W\}$$

ein Untervektorraum von V/U ist. Für einen Untervektorraum $W \subseteq V/U$ ist das Urbild

$$\pi_U^{-1}(W) = \{v \in V \mid \pi_U(v) \in W\}$$

ein Untervektorraum von V. Zeigen Sie:

(a) Die Funktionen π_U und π_U^{-1} erstellen eine Korrespondenz (d.h. bijektive Funktion) zwischen

$$\{\text{UVR } W \text{ von } V \text{ mit } U \subseteq W\} \underset{\pi_U^{-1}}{\overset{\pi_U}{\rightleftarrows}} \{\text{UVR von } V/U\}. \tag{5.18}$$

(b) Sei $T \in \operatorname{End}(V)$. Falls U T-invariant ist, dann induziert die Korrespondenz in (5.18) die Korrespondenz

$$\{T - \text{inv. UVR } W \text{ von } V \text{ mit } U \subseteq W\} \underset{\pi_U^{-1}}{\overset{\pi_U}{\rightleftarrows}} \{T_{V/U} - \text{inv. UVR von } V/U\}.$$

(c) Falls $\dim U = r < \infty$ und W ein Untervektorraum von V/U ist mit $\dim W = s$, dann ist

$$\dim \pi_U^{-1}(W) = r + s.$$

(d) Seien $\dim V = n \in \mathbb{N}$, $T \in \operatorname{End}(V)$ und U ein 1-dimensionaler, T-invarianter Untervektorraum von V. Sei

$$\{0_{V/U}\} = W_0 \subseteq W_1 \subseteq \ldots \subseteq W_{n-1} = V/U$$

eine $T_{V/U}$-invariante Fahne in V/U. Dann ist

$$\{0_V\} = V_0 \subseteq U = \pi_U^{-1}(W_0) \subseteq \pi_U^{-1}(W_1) \subseteq \ldots \subseteq \pi_U^{-1}(W_{n-1}) = V \tag{5.19}$$

eine T-invariante Fahne in V.

Lösung zu Übung 5.4.10: (a), (b) und (c) haben Sie in der Serie gesehen. Wir zeigen (d): Die Aussage in (5.19) folgt aus der Definition von π_U^{-1}. Dass die Unterräume in (5.19) eine Flagge bilden (also dass die Dimensionen „stimmen"), folgt aus (c). Die Tatsache, dass alle Untervektorräume in (5.19) T-invariant sind, folgt aus (b).

5.4.2 Beweis von Satz 5.4.5

Der interessante Teil des Beweises ist zu zeigen, dass T trigonalisierbar ist, falls p_T in Linearfaktoren zerfällt. Die Beweisidee ist sehr einfach: Da p_T in Linearfaktoren zerfällt, hat T zumindest einen Eigenvektor v. Diesen wählen wir als unseren ersten Basisvektor. Der Span $\mathrm{Sp}(v)$ ist T-invariant und daher können wir $T_{V/\mathrm{Sp}(v)} \in \mathrm{End}(V/\mathrm{Sp}(v))$ betrachten. Da $p_{T_{V/\mathrm{Sp}(v)}}$ ebenfalls in Linearfaktoren zerfällt, hat $T_{V/\mathrm{Sp}(v)}$ mindestens einen Eigenvektor $w + \mathrm{Sp}(v)$. Jetzt betrachten wir $T_{V/\mathrm{Sp}(v,w)}$... Sehen Sie schon die Induktion?

Beweis von Satz 5.4.5. „$\Longrightarrow$": Dies ist die einfache Richtung. Falls T trigonalisierbar ist, sei $\mathcal{B}$ eine Basis mit

$$[T]_{\mathcal{B}} = \begin{pmatrix} \lambda_1 & \cdots & * \\ & \ddots & \vdots \\ & & \lambda_n \end{pmatrix}.$$

Laut Lemma 5.3.6 gilt $p_T = p_{[T]_{\mathcal{B}}} = \prod_{i=1}^n (\lambda_i - x)$.

„$\Longleftarrow$": Wir beweisen diese Richtung per Induktion über $n = \dim V$. Für $n = 1$ gibt es nichts zu beweisen. Sei $T \in \mathrm{End}(V)$ mit $\dim V = n$. Wir nehmen an, dass

$$p_T = \prod_{i=1}^n (\lambda_i - x).$$

Sei v_1 ein Eigenvektor mit Eigenwert λ_1. Dann ist

$$V_1 := \mathrm{Sp}(v_1)$$

ein T-invarianter Untervektorraum. Wir betrachten nun $T_{V/V_1} \colon V/V_1 \to V/V_1$. Laut (2c) aus Übung 5.4.9 gilt

$$p_{T_{V/V_1}} = \frac{p_T}{p_{T|_{V_1}}} = \frac{\prod_{i=1}^n \lambda_i - x}{\lambda_1 - x} = \prod_{i=2}^n (\lambda_i - x).$$

Ausserdem gilt $\dim V/V_1 = n-1$ und daher ist T_{V/V_1} laut der Induktionsannahme trigonalisierbar. Laut Proposition 5.4.4 existiert eine T_{V/V_1}-invariante Fahne

$$\{0_{V/V_1}\} = W_0 \subseteq W_1 \subseteq \ldots \subseteq W_{n-1} = V/V_1$$

und daher ist laut Übung 5.4.10 (d)

$$\{0_V\} = V_0 \subseteq V_1 \subseteq \pi_{V_1}^{-1}(W_1) \subseteq \ldots \subseteq \pi_{V_1}^{-1}(W_{n-1}) = V$$

eine T-invariante Fahne. Also folgt mit Proposition 5.4.4, dass T trigonalisierbar ist. $\square$

Der Beweis von Satz 5.4.5 gibt uns eigentlich einen Algorithmus zur Trigonalisierung von T. Als Beispiel benutzen wir dieselbe Matrix wie im Fischer ([7]):

Beispiel 5.4.11. Sei

$$A = \begin{pmatrix} 3 & 4 & 3 \\ -1 & 0 & -1 \\ 1 & 2 & 3 \end{pmatrix}.$$

Man berechnet, dass $p_A(x) = (2-x)^3$ ist, und dass die geometrische Vielfachheit dieses einen Eigenwerts

$$\dim \mathrm{Eig}_A(2) = \dim \mathrm{Ker}(A - 2I_3) = \dim \mathrm{Sp}\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = 1$$

ist. Daher ist A nicht diagonalisierbar, aber trigonalisierbar. Der Trigonalisierungssatz 5.4.5 sagt uns, dass wir rekursiv Eigenvektoren finden sollten, um eine Fahne zu konstruieren. Wie zuvor erwähnt, ist

$$v_1 := \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

ein Eigenvektor mit Eigenwert 2. Sei $V_1 := \mathrm{Sp}(v_1)$. Wir betrachten jetzt

$$(m_A)_{\mathbb{Q}^3/V_1} : \mathbb{Q}^3/V_1 \to \mathbb{Q}^3/V_1.$$

Wir brauchen eine Basis von $\mathbb{Q}^3/V_1$ um weitere Eigenvektoren zu finden. Da

$$\left(\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, e_2, e_3\right)$$

eine Basis von $\mathbb{Q}^3$ ist, ist $\mathcal{C} = (e_2 + V_1, e_3 + V_1)$ eine Basis von $\mathbb{Q}^3/V_1$. Bezüglich $\mathcal{C}$ berechnet sich die Darstellungsmatrix von $(m_A)_{\mathbb{Q}^3/V_1}$ als

$$\begin{pmatrix} 4 & 2 \\ -2 & 0 \end{pmatrix},$$

da

$$\begin{aligned} Ae_2 + V_1 = \begin{pmatrix} 4 \\ 0 \\ 2 \end{pmatrix} + V_1 = \begin{pmatrix} 4 \\ 0 \\ 2 \end{pmatrix} - \begin{pmatrix} 4 \\ -4 \\ 4 \end{pmatrix} + V_1 &= \begin{pmatrix} 0 \\ 4 \\ -2 \end{pmatrix} + V_1 \\ &= 4(e_2 + V_1) - 2(e_3 + V_1) \end{aligned}$$

und

$$\begin{aligned} Ae_3 + V_1 = \begin{pmatrix} 3 \\ -1 \\ 3 \end{pmatrix} + V_1 = \begin{pmatrix} 3 \\ -1 \\ 3 \end{pmatrix} - \begin{pmatrix} 3 \\ -3 \\ 3 \end{pmatrix} + V_1 &= \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} + V_1 \\ &= 2(e_2 + V_1). \end{aligned}$$

Wir wissen, dass

$$p_{(m_A)_{\mathbb{Q}^3/V_1}} = \frac{(2-x)^3}{(2-x)} = (2-x)^2 = p_B$$

für $B = \begin{pmatrix} 4 & 2 \\ -2 & 0 \end{pmatrix}$ oder wir berechnen einfach die rechte Seite. Der Vektor $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ ist ein Eigenvektor von $\begin{pmatrix} 4 & 2 \\ -2 & 0 \end{pmatrix}$ mit Eigenwert 2. Mittels diesem Vektor definieren wir

$$v_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \quad \text{und} \quad V_2 = \operatorname{Sp}(v_1, v_2).$$

Jetzt reicht es (v_1, v_2) zu einer Basis von $\mathbb{Q}^3$ zu ergänzen. Da

$$\left| \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix} \right| = 1 \neq 0$$

erfüllt e_3 diesen Zweck. Wir wissen, dass $e_3 + V_2$ ein Eigenvektor mit Eigenwert 2 bezüglich $(m_A)_{\mathbb{Q}^3/V_2}$ ist. Es folgt, dass uns die Basis $\mathcal{B} = (v_1, v_2, e_3)$ eine Trigonalisierung von m_A gibt. Man könnte dies nun überprüfen:

$$\begin{aligned}
(m_A)_{\mathcal{B}} &= (\mathrm{Id}_{\mathbb{Q}^3})^{\mathcal{E}_3}_{\mathcal{B}}(m_A)_{\mathcal{E}_3}(\mathrm{Id}_{\mathbb{Q}^3})^{\mathcal{B}}_{\mathcal{E}_3} \\
&= \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 3 & 4 & 3 \\ -1 & 0 & -1 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & 3 \\ -2 & 1 & -1 \\ 2 & -1 & 3 \end{pmatrix} \\
&= \begin{pmatrix} 2 & 1 & 3 \\ 0 & 2 & 2 \\ 0 & 0 & 2 \end{pmatrix}.
\end{aligned}$$

Funktioniert wie Magie!

5.5 Minimalpolynom

In den Serien im letzten Semester hatten Sie die folgende Übung: Sei $A \in M_{n\times n}(K)$. Zeigen Sie, dass es $k \in \mathbb{N}$ und $\alpha_0, \ldots, \alpha_k \in K$ gibt mit

$$\alpha_0 I_n + \alpha_1 A + \cdots + \alpha_k A^k = 0 \in M_{n\times n}(K). \tag{5.20}$$

Mit anderen Worten, finden Sie ein Polynom $g(x) = \alpha_0 + \alpha_1 x + \cdots + \alpha_k x^k$, sodass das Einsetzen[6] von A in g

$$g(A) = \alpha_0 + \alpha_1 A + \cdots + \alpha_k A^k$$

die Nullmatrix ergibt.

Die Lösung ist einfach: Sobald $(I, A, \ldots, A^k)$ linear abhängig sind (als Elemente des Vektorraums $M_{n\times n}(K)$), können wir $\alpha_0, \ldots, \alpha_k \in K$ finden, die (5.20) erfüllen. Da $\dim M_{n\times n}(K) = n^2$, können wir sicher solch ein k mit $k \leq n^2$ finden. Das heisst es existiert ein Polynom $g \in K[x]$ mit Grad $\leq n^2$, sodass $g(A) = 0 \in M_{n\times n}(K)$.

Wir möchten nun die Polynome $g \in K[x]$ mit der Eigenschaft $g(A) = 0$ besser verstehen und wir werden sehen, dass es sogar ein Polynom von Grad $\leq n$ gibt mit $g(A) = 0$. Das *Minimalpolynom*, welches wir bald definieren werden, ist sozusagen das „kleinste“ Polynom $g \in K[x]$ mit der Eigenschaft $g(A) = 0$.

Mit dem Minimalpolynom von A erhalten wir eine weitere Charakterisierung der Diagonalisierbarkeit von A und allgemein mehr Informationen über A. Wir werden das

[6] Siehe (5.21)/(5.22) weiter unten für eine präzisere Definition von dem Einsetzen eines Endomorphismus/einer Matrix in ein Polynom.

Minimalpolynom natürlich auch für Endomorphismen T definieren und den Fall einer quadratischen Matrix als Spezialfall davon betrachten.

In diesem Abschnitt seien V ein endlich-dimensionaler K-Vektorraum, $T \in \operatorname{End}(V)$ und x eine Variable. Betrachten wir die folgende Abbildung zwischen Ringen:

$$\begin{aligned} K[x] &\to \operatorname{End}(V) \\ g(x) = a_n x^n + \cdots + a_1 x + a_0 &\mapsto g(T) := a_n T^n + \cdots + a_1 T + a_0 \operatorname{Id}_V \end{aligned} \tag{5.21}$$

Bemerken Sie, dass das Bild eines Skalars a_0 mit dem Endomorphismus $a_0 \operatorname{Id}_V$ identifiziert wird. Die Abbildung (5.21) ist ein Homomorphismus von Ringen, aber keine Sorge falls Sie schon vergessen haben, was das heisst: Wir interessieren uns nur für den Kern dieses Homomorphismus, also die Menge

$$I_T := \{g \in K[x] \mid g(T) = 0\}.$$

Die Null hier steht natürlich für die Nullabbidlung $0_{\operatorname{End}(V)} \in \operatorname{End}(V)$. Es ist klar, dass das Nullpolynom 0 in I_T liegt. Die Erklärung am Anfang des Abschnitts zeigt, dass $I_T \neq \{0\}$.

Lemma 5.5.1. *Es existiert ein eindeutiges Polynom $M_T \in K[x]$ sodass gilt:*

(1) M_T ist normiert (d.h. der Leitkoeffizient ist 1).

(2) $I_T = M_T \cdot K[x] := \{M_T \cdot h \mid h \in K[x]\}$. Das heisst, M_T teilt alle Polynome in I_T.

Bemerken Sie, dass Eigenschaft (2) äquivalent ist zu:

$$\forall g \in K[x] : \quad g(T) = 0 \iff M_T \mid g.$$

Beweis. Der Schlüssel zu diesem Beweis ist Division mit Rest.

Sei $d = \min\{\deg(p) \mid p \in I_T,\ p \neq 0\}$. Dieses Minimum ist definiert weil $I_T \neq \{0\}$. Sei $M_T \in I_T$ ein normiertes Polynom von Grad d. Wir behaupten, dass M_T Eigenschaft (2) hat:

Die Inklusion $M_T \cdot K[x] \subseteq I_T$ kann leicht überprüft werden, da $M_T \in I_T$. Für die andere Inklusion betrachten wir ein $p \in I_T$. Laut Polynomdivision existieren $q, r \in K[x]$ mit $p = qM_T + r$ und $\deg(r) < \deg(M_T) = d$. Da $p \in I_T$ ist, gilt

$$0 = p(T) = q(T) \underbrace{M_T(T)}_{=0} + r(T) \Longrightarrow r(T) = 0$$

und somit ist $r \in I_T$. Da $\deg(r) < d$ folgt aus der Definition von d, dass $r = 0$. Also gilt $p = qM_T$ und somit $p \in M_T \cdot K[x]$, wie wir zeigen wollten.

Es bleibt noch die Eindeutigkeit von M_T zu zeigen: Seien $N, M \in K[x]$ zwei Polynome, die beide die Eigenschaften (1) und (2) besitzen. Es ist klar, dass $N, M \in I_T$. Da N

die Eigenschaft (2) hat, existiert $h \in K[x]$ mit $M = N \cdot h$. Da M auch die Eigenschaft (2) hat, existiert $g \in K[x]$ mit $N = M \cdot g$. Also gilt

$$N = M \cdot g = N \cdot h \cdot g$$

und folglich $hg = 1$. Daraus folgt, dass $\deg(g) = \deg(h) = 0$, das heisst, g und h sind Skalare in K. Die Polynome M und N unterscheiden sich also nur durch die Multiplikation mit einem Skalar. Da jedoch beide normiert sind (Eigenschaft (1)), folgt $N = M$. □

Übung 5.5.2. In der Algebra heissen Mengen wie $I \subseteq K[x]$ Ideale. Genauer gesagt sei R ein kommutativer Ring. Ein *Ideal* in R ist ein Teilmenge $I \subseteq R$ mit

$$\begin{aligned} a, b \in I &\Longrightarrow a + b \in I, \\ a \in I, b \in R &\Longrightarrow ab \in I. \end{aligned}$$

Ein kommutativer Ring heisst *euklidisch*, falls er Division mit Rest hat. Präziser formuliert heisst das: Es existiert eine Funktion $\varphi : R \to \mathbb{N} \cup -\infty$, sodass

$$\forall a, b \in R\ \exists r, q \in R \text{ sodass } a = bq + r,\ \ \varphi(r) < \varphi(b).$$

Beispiele sind natürlich $K[x]$ mit $\varphi = \deg$ und $\mathbb{Z}$ mit $\phi(n) = |n|$.

Zeigen Sie: Jedes Ideal I in einem euklidischen Ring hat die Form $I = \{ap |\ p \in R\}$ für ein $a \in R$.

Wenn Sie mehr über solche Ringe lernen möchten, warten Sie bis zur Vorlesung Algebra I oder lesen Sie [3, 5.3] schon jetzt.

Definition 5.5.3. Das Polynom M_T aus Lemma 5.5.1 heisst *Minimalpolynom* von T.

Bemerkung 5.5.4. Das Minimalpolynom von einer Matrix A ist entweder wie oben durch die Abbildung

$$\begin{aligned} K[x] &\to M_{n\times n}(K) \\ a_n x^n + \cdots + a_1 x + a_0 &\mapsto a_n A^n + \cdots + a_1 A + a_0 I_n \end{aligned} \tag{5.22}$$

definiert oder, äquivalent dazu, als $M_A := M_{m_A}$.

Unser Ziel ist es jetzt zu zeigen, dass das Minimalpolynom das charakteristische Polynom teilt, also $M_T \mid p_T$. Das ist der Inhalt des Satzes von Cayley-Hamilton.

Satz 5.5.5 (Cayley-Hamilton)**.** *Es gilt* $p_T(T) = 0 \in \mathrm{End}(V)$.

Übung 5.5.6. Angewendet auf Matrizen besagt der Satz, dass für $A \in M_{n\times n}(K)$ gilt, dass $p_A(A) = 0 \in M_{n\times n}(K)$. Stellen Sie sicher, dass Sie verstehen, wieso diese beiden Formulierungen äquivalent sind!

Es gibt viele verschiedene Beweise für diesen Satz. Aus dem kürzesten kann man meiner Meinung nach nicht viel lernen (Die Leser können ihn in [7, Seite 252] oder [10, Seite 269] finden und bewundern, wie magisch die adjunkte Matrix ist.). Wir geben einen viel längeren Beweis, der aber interessante Begriffe einführt, wie T-Invarianz (was bereits in Abschnitt 5.4 gesehen haben und was auch später noch eine Rolle spielen wird), T-zyklische Untervektorräume und die Begleitmatrix eines Polynoms.

Übung 5.5.7. Versuchen Sie, den Cayley-Hamilton-Satz für diagonalisierbare (bzw. trigonalisierbare) Endomorphismen zu beweisen, bevor Sie den Beweis weiter unten lesen. In gewisser Weise stammt der Beweis den wir weiter unten geben vom Versuch, die selben Ideen, die für diese Übung benutzt werden, anzuwenden und für den allgemeinen Fall anzupassen.[7]

Übung 5.5.8. Wieso ist der folgende „Beweis“ des Cayley-Hamilton-Satzes falsch?

$$p_A(x) = \det(A - xI_n) \Longrightarrow p_A(A) = \det(A - AI_n) = \det(0) = 0.$$

5.5.1 Beweis von Cayley-Hamilton

Definition 5.5.9. Für $v \in V$ sei $\operatorname{Zyk}(v) := \operatorname{Sp}(\{T^i v \mid i \in \{0\} \cup \mathbb{N}\})$ der *T-zyklische Unterraum* von v.

Übung 5.5.10. Zeigen Sie:

- $\operatorname{Zyk}(v) = \{0_V\} \Longleftrightarrow v = 0$.
- $\dim \operatorname{Zyk}(v) = 1 \Longleftrightarrow v$ ist ein Eigenvektor von T.
- $\operatorname{Zyk}(v)$ ist T-invariant. Genauer gesagt ist $\operatorname{Zyk}(v)$ der kleinste T-invariante Untervektorraum von V, der v enthält.

Lemma 5.5.11. *Sei $0 \neq v \in V$ und $k = \dim \operatorname{Zyk}(v) > 0$. Dann gilt:*

(1) $\{v, Tv, \dots, T^{k-1}v\}$ ist eine Basis von $\operatorname{Zyk}(v)$.

(2) Seien $a_0, \dots, a_{k-1} \in K$ die eindeutigen Skalare mit

$$T^k v = -a_0 v + \cdots + (-a_{k-1})T^{k-1}v. \tag{5.23}$$

Dann gilt

$$p_{T|_{\operatorname{Zyk}(v)}}(x) = (-1)^k \left(a_0 + a_1 x + \cdots + a_{k-1}x^{k-1} + x^k\right), \tag{5.24}$$

wobei $T|_{\operatorname{Zyk}(v)}$ die Einschränkung von T auf $\operatorname{Zyk}(v)$ ist.

[7] Wenn Sie eine genauere Erklärung dieses letzten Satzes wünschen, können Sie mich in der Nachbesprechung fragen.

Beweis. Sei $d = \max\{\ell \in \mathbb{N} \mid \{v, Tv, \dots, T^\ell v\}$ ist linear unabhängig$\}$. Wir zeigen mit Induktion über j, dass $T^{d+j} \in \mathrm{Sp}(v, \dots, T^d v)$ für alle $j \in \mathbb{N}$. Dies impliziert, dass $d = k - 1$ und zeigt (1). Die Induktionsbasis $j = 1$ folgt aus der Definition von d (wieso?). Das heisst es gibt $\beta_0, \dots, \beta_d \in K$ mit

$$T^{d+1}v = \beta_0 v + \cdots + \beta_d T^d v. \tag{5.25}$$

Nehmen wir an, die Induktionsannahme $T^{d+j}v \in \mathrm{Sp}(v, \dots, T^d v)$ gilt für $j \geq 1$ und betrachten wir

$$\begin{aligned} T^{d+(j+1)}v = T(T^{d+j}v) &\underset{\text{Annahme}}{\overset{\text{Ind.}}{=}} T(\alpha_0 v + \cdots + \alpha_d T^d v) \\ &= \alpha_0 Tv + \cdots + \alpha_d T^{d+1} v \\ &\underset{\text{Basis}}{\overset{\text{Ind.}}{=}} \underbrace{\alpha_0 Tv + \cdots + \alpha_{d-1} T^d v + \alpha_d \left(\beta_0 v + \cdots + \beta_d T^d v\right)}_{\in \mathrm{Sp}(v, \dots, T^d v)}. \end{aligned}$$

Teil (1) folgt. Um (2) zu zeigen, berechnen wir die Darstellungsmatrix von $T|_{\mathrm{Zyk}(v)}$ bezüglich der geordneten Basis $\mathcal{B} = (v, \dots, T^{k-1}v)$. Für $i = 0, \dots, k-2$ ist $T(T^i v) = T^{i+1}v$ und $T(T^{k-1}v) = T^k v = -a_0 v + \cdots + (-a_{k-1})T^{k-1}v$. Folglich ist

$$\left[T|_{\mathrm{Zyk}(v)}\right]_{\mathcal{B}} = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 & -a_{k-2} \\ 0 & 0 & \cdots & 1 & -a_{k-1} \end{pmatrix}.$$

Diese Matrix nennt man die *Begleitmatrix* des Polynoms (5.24) und sie hat laut [Serie 2, Aufgabe 7a] das charakteristische Polynom (5.24). Dies zeigt (2). □

Lemma 5.5.12. *Falls $W \subseteq V$ ein T-invarianter Unterversektorraum ist, gilt $p_{T|_W} \mid p_T$ (d.h. $\exists g \in K[x]$ mit $p_{T|_W} \cdot g = p_T$).*

Dies folgt aus Übung 5.4.9 (2c). Wir geben der Vollständigkeit halber trotzdem noch einen Beweis.

Beweis. Sei $\mathcal{A} = \{w_1, \dots, w_k\}$ eine Basis von W und $\mathcal{B}$ eine Erweiterung zu einer Basis von V. Da W T-invariant ist, existiert eine quadratische Matrix C, sodass $[T]_{\mathcal{B}}$ die folgende Form hat:

$$[T]_{\mathcal{B}} = \left(\begin{array}{c|c} [T|_W]_{\mathcal{A}} & * \\ \hline 0 & C \end{array}\right).$$

Daher gilt $p_T = p_{[T]_\mathcal{B}} \overset{5.3.6\,(5)}{=} p_{[T|_W]_\mathcal{A}} \cdot p_C = p_{T|_W} \cdot p_C$, wie wir zeigen wollten. □

Wir sind fertig mit den Vorbereitungen und können nun den Satz von Cayley-Hamilton beweisen.

Beweis von Satz 5.5.5. Wir müssen zeigen, dass für jedes $v \in V$ $p_T(T)(v) = 0_V$ gilt. Für $v = 0_V$ ist das klar. Sei also $v \neq 0_V$ und sei $k = \dim \mathrm{Zyk}(v) > 0$. Gemäss Lemma 5.5.11 existieren $a_0, \ldots, a_{k-1}$ mit

$$T^k v = -a_0 v + \cdots + (-a_{k-1}) T^{k-1} v \tag{5.26}$$

und $p_{T|_{\mathrm{Zyk}(v)}}(x) = (-1)^k \left(a_0 + \cdots + a_{k-1}x^{k-1} + x^k\right)$. Zusammen impliziert dies, dass

$$p_{T|_{\mathrm{Zyk}(v)}}(T)(v) = (-1)^k \left(a_0 \operatorname{Id}_V(v) + \cdots + a_{k-1}T^{k-1}v + T^k v\right) \overset{(5.26)}{=} (-1)^k 0_V = 0_V.$$

Da $\mathrm{Zyk}(v)$ T-invariant ist, gibt uns Lemma 5.5.12 ein $g \in K[x]$ sodass $p_T(x) = g(x) \cdot p_{T|_{\mathrm{Zyk}(v)}}(x)$. Also ist

$$p_T(T)(v) = g(T) \cdot p_{T|_{\mathrm{Zyk}(v)}}(T)(v) = g(T)\left(p_{T|_{\mathrm{Zyk}(v)}}(T)(v)\right) = g(T)(0_V) = 0_V.$$

Da $v \neq 0$ beliebig war, beendet dies den Beweis. □

Korollar 5.5.13. *Für $T \in \mathrm{End}(V)$ und $\lambda \in K$ sind die folgenden Aussagen äquivalent:*

1. *λ ist ein Eigenwert von T.*
2. *λ ist eine Nullstelle von p_T.*
3. *λ ist eine Nullstelle von M_T.*

Beweis. Wir wissen bereits aus Abschnitt 5.3, dass (1) und (2) äquivalent sind. Erinnern Sie sich (Korollar 1.4.15), dass für ein Polynom $q \in K[x]$ gilt:

$$\lambda \text{ ist eine Nullstelle von } q \iff (x - \lambda) \mid q.$$

Da das Minimalpolynom M_T das charakteristische Polynom p_T teilt und Teilbarkeit eine transitive Relation ist, folgt (3)⇒(2). Nehmen wir jetzt (1) an und sei $v \in V$ ein Eigenvektor zum Eigenwert λ. Wenn wir M_T als $M_T = b_0 + \cdots + b_{d-1}x^{d-1} + x^d$ schreiben, so gilt

$$\begin{aligned} 0_V = M_T(T)(v) &= b_0 \operatorname{Id}_V(v) + b_1 T(v) + \cdots + T^d(v) \\ &\overset{(1)}{=} b_0 v + b_1 \lambda v + \cdots + \lambda^d v \\ &= (b_0 + b_1\lambda + \cdots + \lambda^d)v = M_T(\lambda)v. \end{aligned}$$

Da $v \neq 0$, folgt $M_T(\lambda) = 0$, was (1) $\Rightarrow$ (3) zeigt. □

Aus Korollar 5.5.13 folgt direkt das folgende Korollar:

Korollar 5.5.14. *Falls $p_T(x) = \pm \prod_{i=1}^{\ell} (x - \lambda_i)^{s_i}$ mit $1 \leq s_i$, wobei die λ_i paarweise verschieden sind, dann existieren $1 \leq r_i \leq s_i$ mit $M_T(x) = \prod_{i=1}^{\ell} (x - \lambda_i)^{r_i}$.*

Dieses Korollar ist hilfreich, wenn man das Minimalpolynom berechnen will, wie wir in den nächsten Beispielen sehen werden.

Beispiel 5.5.15. Sei $A = \begin{pmatrix} 1 & 1 & \\ & 1 & 1 \\ & & 4 \end{pmatrix}$. Das charakteristische Polynom von A ist $p_A = (-1)(x-1)^2(x-4)$, also ist entweder $M_A = (x-1)(x-4)$ oder $M_A = (x-1)^2(x-4)$. Man berechnet

$$(A - 1I_3)(A - 4I_3) = \begin{pmatrix} 0 & 1 & \\ & 0 & 1 \\ & & 3 \end{pmatrix} \begin{pmatrix} -3 & 1 & \\ & -3 & 1 \\ & & 0 \end{pmatrix} \neq 0.$$

Daher gilt $M_A = (x-1)^2(x-4)$.

Beispiel 5.5.16. Wir berechnen das Minimalpolynom für die Matrix

$$A = \begin{pmatrix} 3 & -1 & \\ & 2 & \\ 1 & -1 & 2 \end{pmatrix}.$$

Das charakteristische Polynom von A ist

$$p_A = \det \begin{pmatrix} 3-x & -1 & \\ & 2-x & \\ 1 & -1 & 2-x \end{pmatrix} = -(x-2)^2(x-3),$$

also muss das Minimalpolynom entweder $(x-2)(x-3)$ oder $(x-2)^2(x-3)$ sein. Wenn wir A in das Polynom $(x-2)(x-3)$ einsetzen, erhalten wir die Nullmatrix. Also ist das Minimalpolynom von A durch $M_A = (x-2)(x-3)$ gegeben.

Übung 5.5.17. Sei $\mathcal{B}$ eine Basis von V. Zeigen Sie, dass $M_T = M_{[T]_{\mathcal{B}}}$.

Beispiel 5.5.18. Wir betrachten die Ableitung von Polynomen vom Grad $\leq n$ als eine Abbildung

$$\begin{aligned} D : \mathbb{R}[x]_n &\to \mathbb{R}[x]_n \\ p &\mapsto p'. \end{aligned}$$

Bezüglich der Standardbasis $\mathcal{B} = (1, \ldots, x^n)$ ist

$$[D]_{\mathcal{B}} = \begin{pmatrix} 0 & 1 & & & \\ & 0 & 2 & & \\ & & \ddots & \ddots & \\ & & & \ddots & n \\ & & & & 0 \end{pmatrix}$$

und daher ist $p_D = p_{[D]_{\mathcal{B}}} = (-1)^{n+1}x^{n+1}$ und $M_D = x^r$ für ein $1 \leq r \leq n+1$. Da $D^r(x^n) = n(n-1)\cdots(n-r+1)x^{n-r} \neq 0$ für $0 \leq r \leq n$, folgt, dass $r = n+1$ und somit $M_D = x^{n+1}$.

Korollar 5.5.19. *p_T teilt M_T^n, wobei $n = \dim V$.*

Wir geben einen „echten“ Beweis über $\mathbb{C}$ und eine Beweisskizze für den Fall eines beliebigen Körpers K.

Beweis im Fall $K = \mathbb{C}$. Über $\mathbb{C}$ wissen wir, dass p_T in Linearfaktoren zerfällt, also

$$\deg p_T = n \quad \text{und} \quad p_T = (-1)^n \prod_{i=1}^{\ell} (x - \lambda_i)^{s_i},$$

wobei $\lambda_1, \ldots, \lambda_\ell$ die paarweise verschiedenen Eigenwerte von T sind. Es gilt $1 \leq s_i \leq n$ für alle $i = 1, \ldots, \ell$. Aus Korollar 5.5.14 folgt

$$M_T = \prod_{i=1}^{\ell} (x - \lambda_i)^{r_i} \quad \text{mit} \quad 1 \leq r_i \leq s_i.$$

Daher teilt p_T das Polynom $M_T^n = \prod_{i=1}^{\ell} (x - \lambda_i)^{r_i n}$, da $s_i \leq n \leq r_i n$ für alle i. □

Diesen Beweis kann man mit dem Begriff vom *algebraischen Abschluss* $\overline{K}$ von K zu einem allgemeinen Körper verallgemeinern. Grob gesagt ist ein algebraischer Abschluss $\overline{K}$ eines Körpers K ein Körper, der K enthält und über dem jedes Polynom in Linearfaktoren zerfällt[8].

Eine Skizze des Beweises ist, p_T und M_T als Polynome über $\overline{K}$ zu betrachten, wo sie in Linearfaktoren zerfallen. Nun zeigt man, $p_T \mid M_T^n$ in $\overline{K}[x]$ wie im Beweis oben, dann folgert man daraus, dass diese Teilbarkeit auch in $K[x]$ gilt.

Übung 5.5.20. Versuchen Sie, die Beweisskizze im Fall $K = \mathbb{R}$ zu einem richtigen Beweis umzuformen ($\mathbb{C}$ ist der algebraische Abschluss von $\mathbb{R}$).

[8] Dies ist äquivalent dazu, dass jedes Polynom eine Nullstelle in $\overline{K}$ hat.

Beispiel 5.5.21. Allgemein ist $n = \dim V$ die tiefste Potenz, sodass die Behauptung $p_T \mid M_T^n$ für alle $T \in \mathrm{End}(V)$ gilt. Zum Beispiel für $A = I_n$ ist $p_{I_n} = (1-x)^n$ aber $M_{I_n} = (x-1)$ (oder ähnlicherweise für $0 = 0_{M_{n\times n}(K)}$ ist $p_0 = (-x)^n$, $M_0 = x$).

Für einzelne $T \in \mathrm{End}(V)$ kann $p_T \mid M_T^m$ aber auch für $m < \dim V$ gelten. Wie wir in den Beispielen 5.5.15 und 5.5.18 gesehen haben, kann auch $M_T = \pm p_T$ gelten[9] und in diesem Fall ist $p_T \mid M_T^m$ sogar für $m = 1$ erfüllt.

Falls Sie Übung 5.5.7 für diagonalisierbare Endomorphismen gelöst haben, kennen Sie schon einen Beweis von einer Richtung der folgenden nützlichen Äquivalenz:

Satz 5.5.22. *Sei $T \in \mathrm{End}(V)$. Dann gilt*

$$T \textit{ ist diagonalisierbar} \iff M_T \textit{ zerfällt in paarweise verschiedene Linearfaktoren.}$$

Übung 5.5.23 (Sehr ähnlich zu 5.5.7)**.** Beweisen Sie die Richtung „$\Rightarrow$“.

Wir werden die Richtung „$\Leftarrow$“ über $\mathbb{C}$ später mit Hilfe der Jordan-Normalform beweisen und falls wir genug Zeit haben, werden wir es auch allgemein beweisen.

Wir werden Satz 5.5.22 für die weitere Entwicklung der Theorie nicht verwenden, bevor wir ihn bewiesen haben, aber Sie dürfen diesen Satz durchaus benutzen, wenn Sie die Serien oder sonstige Aufgaben lösen. Hier sind zwei Beispiele:

Beispiel 5.5.24. Die Matrix A von Beispiel 5.5.15 ist nicht diagonalisierbar, weil 1 eine mehrfache Nullstelle von M_A ist.

Beispiel 5.5.25. Die Ableitungsabbildung D von Beispiel 5.5.18 ist nicht diagonalisierbar für $n \geq 1$, da $M_D = x^{n+1}$.

Damit unsere Kollegen der numerischen Mathematik nicht auf uns warten müssen, verschieben wir die Jordan-Normalform auf später und beginnen mit einem neuen Kapitel.

[9] Ein anderes Beispiel wäre ein Endomorphismus mit n paarweise verschiedenen Eigenwerten, wobei $n = \dim V$.

Changelog: Kapitel 5

- 19.02: Nummerierungen wurden geändert.
- 21.02: Beispiel 5.2.10 wurde korrigiert.
- 24.02: In Definition 5.3.1 wird der Eigenraum nun für beliebige $\lambda \in K$ und nicht nur für Eigenwerte definiert.
- 24.02: In Definition 5.3.5 wurde $n \in N$ zu $n \in \mathbb{N}$ korrigiert.
- 24.02: In Lemma 5.3.6 (2) wurde $a_n = (-1)^{n-1}$ zu $a_n = (-1)^n$ korrigiert.
- 24.02: In der letzten Gleichung vom Beweis von Lemma 5.3.6 wurde ein Tippfehler korrigiert.
- 24.02: In Proposition 5.3.17 wurde die die Bedingung, dass die U_i endlich-dimensional sind, hinzugefügt und (3) wurde korrigiert.
- 24.02: Der Verweis für die Wohldefiniertheit von Definition 5.3.8 wurde von Lemma 5.3.6 (5) zu Lemma 5.3.6 (6) korrigiert.
- 26.02: In Beispiel 5.3.24 wurde das charakteristische Polynom zu $(\lambda - x)^n$ korrigiert.
- 28.02: In Beispiel 5.3.11 wurde der Verweis von Lemma 5.3.6 (5) zu Lemma 5.3.6 (2) korrigiert.
- 28.02: Der Beweis von Lemma 5.3.32 wurde korrigiert und umformuliert.
- 01.03: In Lemma 5.3.6 (5) wurde hinzugefügt, dass die B_i quadratische Matrizen sind.
- 01.03: In Korollar 5.3.20 wurde $\mathrm{Hom}(V)$ zu $\mathrm{End}(V)$ und „Eigenvektoren“ zu „Eigenwerte“ korrigiert.
- 01.03: In Abschnitt 5.3.2 wurde $\mathrm{Eig}_{\lambda_i}(T)$ zu $\mathrm{Eig}_T(\lambda_i)$ korrigiert.
- 01.03: Die Formulierung von Definition 5.3.31 wurde korrigiert.
- 05.03: Im ersten Beweis von Lemma 5.3.3 (3) wurde spezifiziert, dass $v \neq 0$.
- 05.03: Im Beweis von Lemma 5.5.1 direkt unterhalb von (5.25) wurde T^{d+j} zu $T^{d+j}v$ korrigiert.
- 05.03: Im Beweis der Implikation (4)$\Rightarrow$(3) von Proposition 5.3.17 wurde der letzte Satz von „Dies zeigt (4).“ zu „Dies zeigt (3).“ geändert.

- 05.03: Im Beweis von Lemma 5.3.32 wurde $p_T(x) = (\lambda - x)^r \cdot p_T(x)$ zu $p_T(x) = (\lambda - x)^r \cdot p_C(x)$ korrigiert.
- 05.03: In Beispiel 5.3.33 wurde $m_a(A, 2) = 2$ zu $m_a(A, 1) = 2$ korrigiert.
- 05.03: In Bemerkung 5.3.34 wurde $m_a(T, \lambda) = m_a(T, \lambda) = 0$ zu $m_g(T, \lambda) = m_a(T, \lambda) = 0$ korrigiert.
- 05.03: In Beispiel 5.4.11 wurde $\mathcal{C} = (e_2 + V_1, e_3 + V_2)$ zu $\mathcal{C} = (e_2 + V_1, e_3 + V_1)$ korrigiert.
- 06.03: Der Beweis der Implikation (2)$\to$(3) von Proposition 5.3.17 wurde umformuliert.
- 06.03: Im Beweis von Satz 5.3.41 wurden einige Indizes geändert und eine Stelle wurde umformuliert.
- 06.03: In Korollar 5.5.14 und im Beweis von Korollar 5.5.19 wurde das Vorzeichen vom charakteristischen Polynom korrigiert.
- 09.03: Im Beweis von Lemma 5.5.11 wurden die b_i zu β_i geändert und einige der a_i zu α_i.
- 15.03: Im Beweis von Satz 5.4.5 wurde ein Verweis von Übung 5.4.10 zu Proposition 5.4.4 korrigiert und ein Verweis zu Übung 5.4.10 (d) präzisiert. Ausserdem wurde an einer Stelle $\prod_{i=1}^n (\lambda_i - x_i)$ zu $\prod_{i=1}^n (\lambda_i - x)$ korrigiert.
- 16.03: Das Wort triagonalisierbar wurde zu trigonalisierbar korrigiert (alle anderen Formen des Worts wurden natürlich auch angepasst).
- 18.03: In der letzten Zeile des Beweises von Lemma 5.5.12 wurde $p_{[T_W]_\mathcal{A}}$ zu $p_{[T|_W]_\mathcal{A}}$ und p_{T_W} zu $p_{T|_W}$ korrigiert.
- 21.03: In Beispiel 5.5.18 wurde $(n - r)$ zu $(n - r + 1)$ korrigiert.
- 14.06: In Korollar 5.3.20 wurde $\lambda_1, \ldots, \lambda_r \in V$ zu $\lambda_1, \ldots, \lambda_r \in K$ korrigiert.
- 14.06: In der letzten Gleichung des Beweises von Satz 5.3.41 wurde der Index der zweiten direkten Summe von i zu j korrigiert.
- 19.06: In Korollar 5.5.14 wurde hinzugefügt, dass die λ_i paarweise verschieden sein müssen.

Kapitel 6

Euklidische und unitäre Vektorräume

Ich habe mir lange überlegt, ob ich euklidische und unitäre Vektorräume getrennt oder zusammen unterrichten möchte. Ich habe mich dann für zusammen entschieden und ich werde deshalb folgende Notation benutzen:

Notation. Jeder Teil in diesem Kapitel kommt mit einer der folgenden Markierungen, welche wie folgt zu interpretieren sind:

$\mathbb{R}$	–	Dies betrifft euklidische Vektorräume.
$\mathbb{C}$	–	Dies betrifft unitäre Vektorräume.
$\mathbb{F}$	–	Dies ist relevant für euklidische und unitäre Vektorräume.
K	–	Dies gilt für allgemeine Körper.
$\mathbb{R} \leftrightarrow \mathbb{C}$	–	Dies betrifft Interaktionen zwischen euklidischen und unitären Vektorräumen.
A	–	Dies sind allgemeine Dinge.

Falls wir es nicht vergessen zu ändern, so ist immer das zuletzt erschienene Symbol relevant.

A In der modernen Mathematik ist es üblich, Strukturen zu Objekten hinzuzufügen oder zu entfernen, mit dem Ziel, die untersuchten Objekte am Schluss besser zu verstehen. Wenn wir Vektorräume über $\mathbb{R}$ oder $\mathbb{C}$ betrachten, haben Sie vielleicht bemerkt, dass wir abgesehen von ein bisschen geometrischer Intuition keine „geometrischen“ Argumente verwendet haben. Da Vektorräume über $\mathbb{R}$ und $\mathbb{C}$ sehr zentral für die Naturwissenschaften sind, werden wir uns in den nächsten zwei Kapiteln auf diese fokussieren. Mann kann diese Vektorräume mit einer zusätzlichen Struktur versehen, welche uns erlauben wird, auf die Geometrie dieser Räume zuzugreifen.

Wir möchten natürlich nur so viel zusätzliche Struktur wie nötig einführen, um die geometrischen Begriffe zu beschreiben, für die wir uns interessieren. Dies führt zur Definition eines Innenproduktraums. Fast alle geometrischen Begriffe in $\mathbb{R}^2$ oder $\mathbb{R}^3$, wie

Winkel, Projektionen, Parallelverschiebung, Rotation, können wir mithilfe des inneren Produkts/Skalarprodukts (für uns Synonyme) beschreiben.

Obwohl die Definition von Innenprodukträumen von der Betrachtung von Vektorräumen über $\mathbb{R}$ und $\mathbb{C}$ stammt, könnte man durchaus auch solche Produkte (oder Varianten davon) über anderen Körpern anschauen. Ein Beispiel dafür wäre unser Rätsel (Anhang A) oder die Untersuchung von quadratischen Formen in der Zahlentheorie (welche über dem Körper $\mathbb{Q}$ und seinen zahlreichen Abschlüssen und Erweiterungen stattfindet[1]).

Vorsicht: In der Literatur und in der Praxis gibt es viele verschiedene Konventionen bezüglich der Namen der verschiedenen Strukturen, die wir in diesem Kapitel definieren (z.B. was die Definition von Skalarprodukt, inneres Produkt, Punktprodukt ist).

6.1 Viele Definitionen

$\mathbb{R}$ **Definition 6.1.1.** Sei V ein Vektorraum über $\mathbb{R}$. Ein *Skalarprodukt/inneres Produkt* auf V ist eine Funktion $\langle\, , \rangle : V \times V \to \mathbb{R}$, sodass

(1) (Linear in der ersten Variable) Für alle $v, v_1, v_2, w \in V$ und $\alpha \in \mathbb{R}$ ist

$$\begin{aligned} \langle v_1 + v_2, w \rangle &= \langle v_1, w \rangle + \langle v_2, w \rangle \,, \\ \langle \alpha v, w \rangle &= \alpha \langle v, w \rangle \,. \end{aligned}$$

(2) (Linear in der zweiten Variable) Für alle $v, w_1, w_2, w \in V$ und $\alpha \in \mathbb{R}$ ist

$$\begin{aligned} \langle v, w_1 + w_2 \rangle &= \langle v, w_1 \rangle + \langle v, w_2 \rangle \,, \\ \langle v, \alpha w \rangle &= \alpha \langle v, w \rangle \,. \end{aligned}$$

(3) (Symmetrisch) Für alle $v, w \in V$ ist $\langle v, w \rangle = \langle w, v \rangle$.

(4) (Positiv definit) Für alle $0 \neq v \in V$ ist $\langle v, v \rangle > 0$.

Wir nennen einen Vektorraum über $\mathbb{R}$ zusammen mit einem Skalarprodukt einen *euklidischen Vektorraum*.

Bemerkung 6.1.2. (a) Eigenschaft (2) folgt aus den Eigenschaften (1) und (3).

K (b) Für allgemeine Körper K macht (4) keinen Sinn, da nicht alle Körper (z.B. $\mathbb{C}$) einen Begriff von Positivität haben. Diejenigen Körper, in denen man Positivität definieren kann, heissen *geordnete Körper*, siehe hier.

[1] Der interessierte Leser kann schon jetzt in einem meiner Lieblingsbücher [6] darüber lesen.

(c) Eine Funktion, die (1) und (2) (bzw. (1), (2) und (3)) erfüllt, heisst *Bilinearform* (bzw. *symmetrische Bilinearform*).

Die Axiome (1), (2) und (3) machen auch Sinn über einem allgemeinen Körper. Wir werden Bilinearformen über allgemeinen Körpern später betrachten.

$\mathbb{R}$ (d) Es folgt aus (1) und (2), dass $\langle 0, v\rangle = \langle v, 0\rangle = 0$ für alle $v \in V$. Somit ist (4) äquivalent zu

$$\forall v \in V \ \ \langle v, v\rangle \geq 0, \text{ mit Gleichheit genau dann, wenn } v = 0.$$

(e) Einige Autoren verlangen in der Definition eines euklidischen Vektorraums, dass dieser endlich-dimensional ist. Bei uns ist das nicht so. Ausserdem bezeichnet ein *euklidischer Raum* ein ähnliches, aber anderes Objekt als ein euklidischer Vektorraum. Mysteriös gesagt: Ein euklidischer Raum ist ein euklidischer Vektorraum, der seinen Ursprung vergessen hat (Sie dürfen mich in der Nachbesprechung noch danach fragen). Wir besprechen in dieser Vorlesung nur euklidische Vektorräume (auch wenn ich manchmal aus Versehen euklidischer Raum sagen oder schreiben werde).

$\mathbb{R} \leftrightarrow \mathbb{C}$ Wir kommen nun zur Definition vom Skalarprodukt auf komplexen Vektorräumen. Wir haben die Stellen, an denen sich die Definition von Definition 6.1.1 unterscheidet, rot markiert, damit Sie die Gemeinsamkeiten und Unterschiede genau erkennen können.

$\mathbb{C}$ **Definition 6.1.3.** Sei V ein Vektorraum über $\mathbb{C}$. Ein *Skalarprodukt/inneres Produkt* auf V ist eine Funktion $\langle\, ,\rangle : V \times V \to \mathbb{C}$, sodass

(1) (Linear in der ersten Variable) Für alle $v, v_1, v_2, w \in V$ und $\alpha \in \mathbb{C}$ ist

$$\begin{aligned}\langle v_1 + v_2, w\rangle &= \langle v_1, w\rangle + \langle v_2, w\rangle\,,\\ \langle \alpha v, w\rangle &= \alpha\,\langle v, w\rangle\,.\end{aligned}$$

(2) (Semi-Linear in der zweiten Variable) Für alle $v, w_1, w_2, w \in V$ und $\alpha \in \mathbb{C}$ ist

$$\begin{aligned}\langle v, w_1 + w_2\rangle &= \langle v, w_1\rangle + \langle v, w_2\rangle\,,\\ \langle v, \alpha w\rangle &= \overline{\alpha}\,\langle v, w\rangle\,.\end{aligned}$$

(3) (Hermitesch) Für alle $v, w \in V$ ist $\langle v, w\rangle = \overline{\langle w, v\rangle}$.

(4) (Positiv definit) Für alle $0 \neq v \in V$ ist $\langle v, v\rangle > 0$.

Wir nennen einen Vektorraum über $\mathbb{C}$ zusammen mit einem Skalarprodukt einen *unitären Vektorraum*.

Bemerkung 6.1.4.

- Obwohl die Funktion $\langle\,,\rangle$ die Zielmenge $\mathbb{C}$ hat, gilt $\langle v, v\rangle \in \mathbb{R}$ wegen (3). Daher kann man (4) verlangen. Im Allgemeinen verwenden wir die folgende Konvention: Für $z \in \mathbb{C}$ bedeutet $z > 0$, dass $z \in \mathbb{R}$ und $z > 0$.
- (a) und (d) von Bemerkung 6.1.2 gelten wortwörtlich auch hier.
- Eine Funktion, die (1) und (2) (bzw. (1), (2) und (3)) erfüllt, heisst *Sesquilinearform*[2] (bzw. *hermitesche Form*).
- Einige Autoren verlangen Linearität in der zweiten Variable und Semi-Linearität in der ersten Variable.

$\mathbb{R} \leftrightarrow \mathbb{C}$

Viele Aussagen über euklidische Vektorräume gelten wortwörtlich oder mit kleinen „technischen“ Änderungen auch für unitäre Vektorräume. Deshalb ist es nützlich, eine Notation und Begriffe für beide Typen von Vektorräumen zu haben. Man könnte sogar beide Definitionen gemeinsam in einer Definition aufschreiben, wie wir gleich demonstrieren werden.

Passen Sie aber auf! Es ist besser, separat über euklidische und unitäre Vektorräume nachzudenken und einfach die Gemeinsamkeiten zu bemerken.

Notation. Wir verwenden in diesem Kapitel die Notation $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, um zu sagen, dass $\mathbb{F} = \mathbb{R}$ oder $\mathbb{F} = \mathbb{C}$. Das heisst also, dass Definitionen, Sätze, usw. über $\mathbb{F}$ und $\mathbb{F}$-Vektorräume jeweils zwei Definitionen/Sätze gleichzeitig sind. Bemerken Sie hierbei, dass Sie im Falle von $\mathbb{F} = \mathbb{R}$ alle komplexen Konjugationen ignorieren können, da diese auf $\mathbb{R} \subseteq \mathbb{C}$ nichts machen. Wenn nichts anderes gesagt wird, meinen wir mit $\mathbb{F}$ immer, dass entweder $\mathbb{F} = \mathbb{R}$ oder $\mathbb{F} = \mathbb{C}$ ist.

Mit dieser Notation erhalten wir eine sehr effiziente Art, das Skalarprodukt über $\mathbb{R}$ und $\mathbb{C}$ gleichzeitig zu definieren:

$\mathbb{F}$

Definition 6.1.5. Sei $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ und sei V ein Vektorraum über $\mathbb{F}$. Ein *Skalarprodukt/inneres Produkt* auf V ist eine Funktion $\langle\,,\rangle : V \times V \to \mathbb{F}$, sodass für alle $v, w, v_1, v_2 \in V$ und $\alpha \in \mathbb{F}$ gilt:

- $\forall v_1, v_2, w \in V \quad \langle v_1 + v_2, w\rangle = \langle v_1, w\rangle + \langle v_2, w\rangle$,
- $\forall v, w \in V\ \forall \alpha \in \mathbb{F} \quad \langle \alpha v, w\rangle = \alpha \langle v, w\rangle$,
- $\forall v, w \in V \quad \langle v, w\rangle = \overline{\langle w, v\rangle}$,
- $\forall v \in V \quad v \neq 0 \implies \langle v, v\rangle > 0$.

Wir nennen einen solchen Vektorraum auch *Skalarporduktraum/Innenproduktraum*. Das heisst, ein Skalarproduktraum/Innenproduktraum ist ein euklidischer Vektorraum (falls $\mathbb{F} = \mathbb{R}$) oder ein unitärer Vektorraum (falls $\mathbb{F} = \mathbb{C}$).

[2]„sesqui-“ bedeutet soviel wie eineinhalb.

Wir zeigen nun einige grundlegende direkte Folgerungen aus der Definition des Skalarprodukts:

Lemma 6.1.6. *Sei V ein Skalarproduktraum über $\mathbb{F}$.*

(1) Für alle $v \in V$ ist $\langle 0, v\rangle = \langle v, 0\rangle = 0$.

(2) Für $w \in V$ gilt: Falls $\langle v, w\rangle = 0\ \forall v \in V$, dann ist $w = 0$.

(3) Für $w_1, w_2 \in V$ gilt: Falls $\langle v, w_1\rangle = \langle v, w_2\rangle\ \ \forall v \in V$, dann ist $w_1 = w_2$.

Beweis. Wir geben einige Hinweise und überlassen den Rest den Lesern.
(1): $\langle 0, v\rangle = \langle 0 + 0, v\rangle$. (2): Insbesondere ist $\langle w, w\rangle = 0$.
(3): $\langle v, w_1\rangle = \langle v, w_2\rangle \iff \langle v, w_1 - w_2\rangle = 0$. □

Definition 6.1.7. Sei $(V, \langle\ ,\ \rangle)$ ein Skalarproduktraum über $\mathbb{F}$. Die *induzierte Norm* ist die Funktion

$$\begin{aligned} \|\cdot\| : V &\to \mathbb{R}_{\geq 0} \\ v &\mapsto \sqrt{\langle v, v\rangle}. \end{aligned}$$

Ein Vektor $v \in V$ heisst *Einheitsvektor* oder *normiert*, falls $\|v\| = 1$. Für $v, w \in V$ definieren wir die *Distanz* zwischen v und w als $\mathrm{d}(v, w) := \|v - w\| = \|w - v\|$.

Beispiel 6.1.8. Falls $v \neq 0$, dann ist $w := \frac{1}{\|v\|} \cdot v$ ein Einheitsvektor, der in „dieselbe Richtung“ wie v zeigt.

Die nächste Proposition zeigt, dass der Satz des Pythagoras auf beliebige Skalarprodukträume verallgemeinert werden kann.

Proposition 6.1.9 (Pythagoras). *Sei V ein Skalarproduktraum mit der induzierten Norm $\|\cdot\|$ und seien $u, v \in V$ mit $u \perp v$. Dann gilt*

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2$$

Beweis. $\|u + v\|^2 = \langle u + v, u + v\rangle = \|u\|^2 + \langle u, v\rangle + \langle v, u\rangle + \|v\|^2 \overset{u\perp v}{=} \|u\|^2 + \|v\|^2$. □

$\mathbb{R}$ **Übung 6.1.10.**

(1) Zeigen Sie, dass in euklidischen Vektorräumen V gilt:

$$\forall v, w \in V \quad \langle v, w\rangle = \frac{1}{2}\left(\|v + w\|^2 - \|v\|^2 - \|w\|^2\right).$$

$\mathbb{C}$ (2) Zeigen Sie, dass in unitären Vektorräumen W gilt:

$$\forall v, w \in W \quad \langle v, w\rangle = \frac{1}{4}\left(\|v + w\|^2 - \|v - w\|^2 + i\,\|v + iw\|^2 - i\,\|v - iw\|^2\right).$$

$\mathbb{F}$ **Definition 6.1.11.** [Orthogonalität] Zwei Vektoren v, w in einem Skalarproduktraum V heissen *orthogonal* (senkrecht) oder *orthogonal zueinander*, falls $\langle v, w\rangle = 0$. Eine Teilmenge $S \subseteq V$ heisst *Orthogonalsystem*, falls die Elemente von S paarweise orthogonal zueinander sind und $0 \notin S$. In Formeln ausgedrückt, heisst das:

$$\forall v, w \in S \quad v \neq w \Rightarrow \langle v, w\rangle = 0 \quad \text{ und } \quad \forall v \in S \quad v \neq 0.$$

Wir schreiben $v \perp w$ falls $\langle v, w\rangle = 0$. Für Teilmengen $S, T \subseteq V$ schreiben wir $S \perp T$ falls $s \perp t$ für alle $s \in S$, $t \in T$. Letztlich ist $v \perp S$ definiert als $\{v\} \perp S$.

Ein orthogonales System S heisst *orthonormal* (oder *Orthonormalsystem*), falls alle Elemente von S normiert sind. In Formeln: $\forall v \in S \ \ \|v\| = 1$.

K *Bemerkung* 6.1.12. Da jeder Körper ein Nullelement enthält, könnte man die Definition $v \perp w$ auch für Bilinearformen über einem allgemeinen Körper formulieren.

6.2 Viele Beispiele

$\mathbb{R}$ **Beispiel 6.2.1** (Haupt-Beispiel)**.** Wir denken an K^n als K^n_{Spal} und identifizieren $M_{1\times 1}(K)$ mit K. Das *Standard-Skalarprodukt* auf $\mathbb{R}^n$ ist

$$\begin{aligned} \langle\,,\rangle : \mathbb{R}^n \times \mathbb{R}^n &\to \mathbb{R} \\ (v, w) &\mapsto \langle v, w\rangle := v^T w \end{aligned} \tag{6.1}$$

und $(\mathbb{R}^n, \langle\,,\rangle)$ heisst der *standard euklidische Vektorraum.* Konkret heisst das also

$$\left\langle \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}, \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \right\rangle = \sum_{i=1}^n v_i w_i.$$

Die induzierte Norm ist $\left\| \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \right\| = \sqrt{\sum_{i=1}^n v_i^2}.$

$\mathbb{C}$ Ähnlich nennt man

$$\begin{aligned} \langle\,,\rangle : \mathbb{C}^n \times \mathbb{C}^n &\to \mathbb{C} \\ (v, w) &\mapsto \langle v, w\rangle := v^T \overline{w} \end{aligned} \tag{6.2}$$

das *Standard-Skalarprodukt* auf $\mathbb{C}^n$ und $(\mathbb{C}^n, \langle\,,\rangle)$ heisst der *standard unitäre Vektorraum.* Konkret ist

$$\left\langle \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}, \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \right\rangle = \sum_{i=1}^n v_i \overline{w}_i.$$

Die induzierte Norm ist $\left\| \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \right\| = \sqrt{\sum_{i=1}^{n} v_i \overline{v_i}} = \sqrt{\sum_{i=1}^{n} |v_i|^2}$. Auf $\mathbb{C} = \mathbb{C}^1$ entspricht diese Norm genau dem Absolutbetrag, also $\|z\| = |z|$.

$\mathbb{R} \leftrightarrow \mathbb{C}$ *Bemerkung* 6.2.2. Bemerken Sie, dass die Einschränkung des Skalarprodukts (6.2) auf $\mathbb{R}^n \times \mathbb{R}^n$ genau das Skalarprodukt (6.1) ergibt.

Wir könnten also auch nur das Skalarprodukt (6.2) definieren, oder noch verwirrender, wir könnten das Skalarprodukt auf $\mathbb{F}^n$ definieren, wobei $\mathbb{F}$ entweder $\mathbb{R}$ oder $\mathbb{C}$ ist, als

$$\begin{aligned} \langle \, , \rangle : \mathbb{F}^n \times \mathbb{F}^n &\to \mathbb{F} \\ (v, w) &\mapsto v^T \overline{w}, \end{aligned}$$

was in beiden Fällen das oben definierte Skalarprodukt gibt (da $\overline{a} = a$ für $a \in \mathbb{R}$). Der Leser sollte die Gemeinsamkeit sehen, aber auch separat über diese Vektorräume nachdenken.

$\mathbb{F}$ **Beispiel 6.2.3.** Sowohl in $\mathbb{R}^n$ als auch in $\mathbb{C}^n$ ist die Standard-Basis ein Orthonormalsystem.

Beispiel 6.2.4. In $\mathbb{F}^2$ haben wir zum Beispiel, dass $\begin{pmatrix} 1 \\ 1 \end{pmatrix} \perp \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, also ist

$$S = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$$

ein Orthogonalsystem. Allerdings sind $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ und $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ keine Einheitsvektoren, weil $\left\| \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\| = \left\| \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\| = \sqrt{2}$, somit ist S nicht orthonormal. Die Vektoren $\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ und $\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ sind hingegen Einheitsvektoren und somit ist

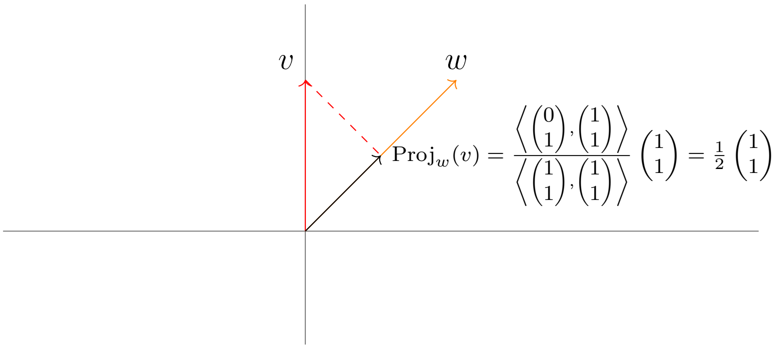
$$T = \left\{ \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$$

ein Orthonormalsystem.

$\mathbb{R}$ **Beispiel 6.2.5.** Die Vektorräume $\mathbb{R}^2$ und $\mathbb{R}^3$ (natürlich ausgestattet mit dem jeweiligen Standard-Skalarprodukt) sollten Ihre „go-to“ Beispiele sein: Das heisst, wann immer Sie Ihre Intuition testen/ neue Begriffe verstehen/ eine Aussage oder Konstruktion „in

Aktion“ sehen möchten, sollten Sie zuerst versuchen, dies in $\mathbb{R}^2$ und $\mathbb{R}^3$ gut zu verstehen. Der Grund dafür ist, dass alle Begriffe und Definitionen in diesem Kapitel von unserer Intuition in $\mathbb{R}^2$ und $\mathbb{R}^3$ entstammen.

Zum Beispiel werden wir später die orthogonale Projektion eines Vektors v auf einen Vektor w definieren als $\mathrm{Proj}_w(v) := \frac{\langle v,w\rangle}{\langle w,w\rangle}w$. Sie werden dies zuerst verwirrend finden, aber wenn wir es in $\mathbb{R}^2$ zeichnen, sagen wir für $v = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ und $w = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, so wird es schon klarer (und wenn Sie sich fragen, wieso man durch $\langle w, w\rangle$ teilt, ersetzen Sie w durch cw, $c \in \mathbb{R}$, und schauen Sie, was passiert).



Figur 6.1: Projektion von v auf w.

$\mathbb{F}$ **Beispiel 6.2.6.** Sei U ein Untervektorraum von einem Skalarproduktraum $(V, \langle\,,\rangle)$ über $\mathbb{F}$. Dann ist $(U, \langle\,,\rangle|_{U\times U})$ auch ein Skalarproduktraum über $\mathbb{F}$. Wir werden in diesem Fall einfach $(U, \langle\,,\rangle)$ statt $(U, \langle\,,\rangle|_{U\times U})$ schreiben.

Wir wollen alle Skalarprodukte auf $\mathbb{F}^n$ verstehen. Dafür müssen wir zwei zentrale Begriffe einführen, welche eine wichtige Rolle spielen werden, wenn wir später lineare Abbildungen, die die euklidische (bzw. unitäre) Struktur erhalten, betrachten. Einer dieser Begriffe ist der folgende:

$\mathbb{R}$ **Definition 6.2.7.** Eine Matrix $A \in M_{n\times n}(\mathbb{R})$ heisst *symmetrisch*, falls $A^T = A$.

$\mathbb{C}$ **Definition 6.2.8.** Für eine Matrix $A = (a_{ij})_{ij} \in M_{n\times n}(\mathbb{C})$ ist $\overline{A} := (\overline{a_{ij}})_{ij} \in M_{n\times n}(\mathbb{C})$.

Definition 6.2.9. Eine Matrix $A \in M_{n\times n}(\mathbb{C})$ heisst *hermitesch*, falls $A^T = \overline{A}$ (oder äquivalent dazu, $\overline{A}^T = A$). Die Matrix $\overline{A}^T$ ist die *adjungierte Matrix* von A und wird auch mit A^H oder A^* bezeichnet. Andere Namen sind *hermitesch transponierte Matrix* oder *transponiert-konjugierte Matrix*.

$\mathbb{F}$ Wir werden noch viel über symmetrische/hermitesche Matrizen, ihre Eigenwerte und Eigenvektoren, sowie ihre Rolle als Transformationen auf Skalarprodukträumen sprechen. Zunächst erklären wir aber erst mal, wie wir aus diesen Matrizen bilineare (bzw. sesquilineare) symmetrische (bzw. hermitesche) Formen erhalten und aus einigen (den positiv definiten, was wir später definieren) sogar Skalarprodukte.

$\mathbb{R}$ **Beispiel 6.2.10** (Skalarprodukte auf $\mathbb{R}^n$)**.** Für eine quadratische Matrix $A \in M_{n\times n}(\mathbb{R})$ erfüllt die Abbildung

$$\begin{aligned}\langle\,,\rangle_A : \mathbb{R}^n \times \mathbb{R}^n &\to \mathbb{R}\\ (v,w) &\mapsto \langle v,w\rangle_A = v^T A w\end{aligned}$$

die Bedingungen (1) und (2) aus Definition 6.1.1 (d.h. $\langle\,,\rangle_A$ ist eine Biliearform). Wir haben zum Beispiel

$$\left\langle \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \right\rangle_{\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}} = \begin{pmatrix} x_1 & y_1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = x_1x_2 + 2x_1y_2 + 3x_2y_1 + 4y_1y_2.$$

Das Produkt $\langle\,,\rangle_A$ ist genau dann symmetrisch (d.h. es erfüllt auch (3) von Definition 6.1.1), wenn A symmetrisch ist.

Im Allgemeinen erfüllt dieses Produkt die Bedingung (4) von Definition 6.1.1 nicht: Zum Beispiel für $A = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}$ gilt $\left\langle \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \right\rangle_A = x_1x_2 - y_1y_2$ und daher kann $\left\langle \begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} x \\ y \end{pmatrix} \right\rangle_A = x^2 - y^2$ durchaus auch negative Werte annehmen.

Definition 6.2.11. Eine symmetrische Matrix $A \in M_{n\times n}(\mathbb{R})$ heisst *positiv definit*, falls $v^T A v > 0$ für alle $0 \neq v \in \mathbb{R}^n$.

Aus dieser Definition folgt: Das Produkt $\langle\,,\rangle_A$ ist ein Skalarprodukt (und somit $(\mathbb{R}^n, \langle\,,\rangle_A)$ ein euklidischer Vektorraum) genau dann, wenn A symmetrisch und positiv definit ist. Wir werden positiv definite Matrizen später charakterisieren.

Beispiel 6.2.12. Diagonale Matrizen $D = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix}$ mit $a_i > 0$ für alle i sind positiv definit, da

$$\langle v,v\rangle_D = a_1v_1^2 + \cdots + a_nv_n^2.$$

Es gibt aber noch viele weitere positiv definite Matrizen.

Übung 6.2.13. Zeigen Sie folgende Äquivalenz für symmetrische Matrizen:

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \in M_{2\times 2}(\mathbb{R}) \text{ ist positiv definit} \iff a > 0 \text{ und } |A| > 0.$$

$\mathbb{R} \leftrightarrow \mathbb{C}$ Es ist höchste Zeit, ein Wörterbuch für euklidische und unitäre Vektorräume einzuführen:

$\mathbb{R}$	$\mathbb{C}$
euklidischer VR	unitärer VR
bilinear	sesquilinear
symmetrisch	hermitesch
positiv definit	positiv definit
$\langle w, v\rangle = \langle v, w\rangle$	$\langle w, v\rangle = \overline{\langle v, w\rangle}$

$\mathbb{C}$ Mit diesem Wörterbuch können Sie Beispiel 6.2.10 übersetzen. Insbesondere gilt für $A \in M_{n\times n}(\mathbb{C})$ und das Produkt $\langle\, ,\rangle_A : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ folgendes:

(a) $\langle\, ,\rangle_A$ ist sesquilinear.

(b) $\langle\, ,\rangle_A$ erfüllt (3) von Def 6.1.3 $\iff$ A ist hermitesch.

(c) $\langle\, ,\rangle_A$ ist ein Skalarprodukt (und somit ist $(\mathbb{C}^n, \langle\, ,\rangle_A)$ ein unitärer Vektorraum) $\iff$ A ist hermitesch und positiv definit.

Der Vollständigkeit halber geben wir die Definition von Positiv-Definitheit einer hermiteschen Matrix an, obwohl sie ganz analog ist: Eine hermitesche Matrix A heisst *positiv definit*, falls $v^T A\overline{v} > 0$ für alle $0 \neq v \in \mathbb{C}^n$.

Übung 6.2.13 gilt sogar auch für hermitesche 2×2-Matrizen. Wenn wir zum Beispiel $A = \begin{pmatrix} 1 & i \\ -i & 2 \end{pmatrix}$ betrachten, so ist $(\mathbb{C}^2, \langle\, ,\rangle)$ ein unitärer Vektorraum (da $1 > 0$ und $|A| = 2 - i(-i) = 1 > 0$), was bereits nicht allzu einfach gewesen wäre, direkt zu überprüfen.

$\mathbb{F}$ Die Beweise der Aussagen in diesen Beispielen sind nicht allzu schwierig und wir empfehlen dem Leser, diese alleine zu machen. Wir geben unten zwei Beweise der schwierigsten Stelle, nämlich zu zeigen, dass $\langle\, ,\rangle_A$ die Bedingung (3) von Definition 6.1.3 erfüllt falls A hermitesch ist (für $\mathbb{F} = \mathbb{C}$) bzw., dass $\langle\, ,\rangle_A$ die Bedingung (3) von Definition 6.1.1 erfüllt falls A symmetrisch ist (für $\mathbb{F} = \mathbb{R}$).

Da komplexe Konjugation auf $\mathbb{R}$ nichts macht, können wir den Beweis nur einmal mit der nützlichen Notation $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ führen. Wir geben zwei Beweise.

Direkter Beweis von (b). Sei $A = (a_{ij})_{1 \leq i,j \leq n} \in M_{n \times n}(\mathbb{F})$ und $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}, w = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \in \mathbb{F}^n$. Es gilt

$$\langle v, w \rangle_A = v^T A \overline{w} = v^T \begin{pmatrix} \sum_j a_{1j} \overline{w}_j \\ \vdots \\ \sum_j a_{nj} \overline{w}_j \end{pmatrix} = \sum_{i=1}^n \sum_{j=1}^n a_{ij} v_i \overline{w}_j.$$

Ähnlich gilt

$$\begin{aligned} \overline{\langle w, v \rangle_A} &= \overline{\sum_{i=1}^n \sum_{j=1}^n a_{ij} w_i \overline{v_j}} = \sum_{i=1}^n \sum_{j=1}^n \overline{a_{ij}}\, \overline{w_i} v_j \overset{\text{Summe}}{\underset{\text{Reihenfolge}}{=}} \sum_{j=1}^n \sum_{i=1}^n \overline{a_{ij}}\, \overline{w_i} v_j \\ &\overset{\text{Namen}}{\underset{i \leftrightarrow j}{=}} \sum_{i=1}^n \sum_{j=1}^n \overline{a_{ji}}\, \overline{w_j} v_i = \sum_{i=1}^n \sum_{j=1}^n \overline{a_{ji}} v_i \overline{w_j}. \end{aligned}$$

Wir sehen also, dass $\langle v, w \rangle_A = \overline{\langle w, v \rangle_A}$ für alle $v, w \in \mathbb{F}^n$ (also (3) von Definition 6.1.3 bzw. 6.1.1) genau dann gilt, wenn $a_{ij} = \overline{a_{ji}}$ für alle i, j, d.h. genau dann, wenn A hermitesch (falls $\mathbb{F} = \mathbb{C}$) bzw. symmetrisch (falls $\mathbb{F} = \mathbb{R}$) ist. □

Dies war ein direkter Beweis, und es wird dem Leser empfohlen, selbst in der Lage zu sein, eine solche „immer der Nase nach"-Berechnung selbst auszuführen. Hier ist noch ein weniger direkter Beweis, welcher nur verwendet, dass die Transposition in $M_{1\times 1}(\mathbb{F})$ nichts macht.

Indirekter Beweis von (b). Einerseits gilt

$$\langle v, w \rangle_A = v^T A \overline{w} \overset{\text{Trans}}{=} \left(v^T A \overline{w} \right)^T = \overline{w}^T A^T v.$$

Andererseits ist

$$\overline{\langle w, v \rangle_A} = \overline{w^T A \overline{v}} = \overline{w}^T \overline{A} v.$$

Die beiden sind gleich für alle v, w genau dann, wenn $\overline{A} = A^T$. □

Der Leser sollte folgendes bemerken: Alle geometrischen Begriffe, die wir bisher definiert haben, und auch die, die wir noch definieren werden, hängen vom Skalarprodukt ab. Zum Beispiel haben wir gesehen, dass Matrizen der Form

$$D = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix}, a_i > 0 \text{ für } i = 1, \ldots, n,$$

positiv definit sind, da $\|v\|_D^2 := \sum_{i=1}^n a_i|v_i|^2$. Die „Einheitssphäre" in $\mathbb{F}^n$ bezüglich $\|\cdot\|_D$, welche als

$$\mathbb{S}^n := \left\{v \in \mathbb{F}^n \mid \sum a_i|v_i|^2 = 1\right\}$$

definiert ist, ist im Fall $\mathbb{F} = \mathbb{R}$ ein Ellipsoid aus Sicht der üblichen Geometrie auf $\mathbb{R}^n$.

Beispiel 6.2.14. Orthogonalität hängt auch vom Skalarprodukt ab. Zum Beispiel ist

$$\left\langle \begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}\right\rangle_{\begin{pmatrix}1 & -1\\-1 & 2\end{pmatrix}} = 0,$$

also sind $\begin{pmatrix}1\\0\end{pmatrix}$ und $\begin{pmatrix}1\\1\end{pmatrix}$ orthogonal zueinander in $\left(\mathbb{F}^2, \langle\,,\rangle_{\left(\begin{smallmatrix}1 & -1\\-1 & 2\end{smallmatrix}\right)}\right)$. Wir könnten also sagen, dass $\langle\,,\rangle_{\left(\begin{smallmatrix}1 & -1\\-1 & 2\end{smallmatrix}\right)}$ eine andere Geometrie auf $\mathbb{F}^2$ beschreibt.

$\mathbb{R}$ **Beispiel 6.2.15.** Seien $a, b \in \mathbb{R}$ mit $a < b$. Sei $V = I[a,b]$ der $\mathbb{R}$-Vektorraum aller Riemann-integrierbaren Funktionen auf $[a,b]$. Sei

$$\begin{aligned}\langle\,,\rangle : V \times V &\to \mathbb{R}\\ (f,g) &\mapsto \langle f,g\rangle := \textstyle\int_a^b f(x)g(x)dx.\end{aligned}$$

Dies definiert ein Produkt, welches (1), (2) und (3) aus Definition 6.1.1 erfüllt, aber nicht (4). Der Grund ist, dass es integrierbare Funktionen f auf $[a,b]$ gibt, die nicht die Nullfunktion sind, für die aber trotzdem $\langle f,f\rangle = \int_a^b f^2(x)dx = 0$ gilt.

Wenn wir das gleiche Produkt hingegen auf dem Vektorraum $C[a,b]$ aller stetigen Funktionen $f : [a,b] \to \mathbb{R}$ anschauen, so ist auch (4) erfüllt (da für stetige f gilt: $\int_a^b f^2(x)dx = 0 \iff f = 0$). Somit ist also $(C[a,b], \langle\,,\rangle)$ ein euklidischer Vektorraum.

Wir können dieses Skalarprodukt auch noch gewichten. Sei $\varphi : [a,b] \to \mathbb{R}_{>0}$ eine positive Funktion auf $[a,b]$. Dann ist

$$\begin{aligned}\langle\,,\rangle_\varphi : V \times V &\to \mathbb{R}\\ (f,g) &\mapsto \langle f,g\rangle_\varphi := \textstyle\int_a^b f(x)g(x)\varphi(x)dx\end{aligned}$$

ein weiteres Skalarprodukt auf $C[a,b]$. Die Positivität von φ wird verlangt, damit $\langle\,,\rangle_\varphi$ (4) erfüllt, also positiv definit ist.

$\mathbb{C}$ *Bemerkung* 6.2.16. Falls Sie Stetigkeit und Integrabilität von Funktionen auf $\mathbb{C}$ schon kennen, dann ist beispielsweise

$$C(B) := \{f : B \to \mathbb{C} \mid f \text{ stetig }\},$$

wobei $B = B(z_0, r) = \{z \in \mathbb{C} \mid |z - z_0| \leq r\}$ für ein $z_0 \in \mathbb{C}$, ein unitärer Vektorraum mit dem Produkt

$$\begin{aligned} \langle\, , \rangle : C(B) \times C(B) &\to \mathbb{C} \\ (f, g) &\mapsto \langle f, g\rangle := \textstyle\int_B f(z)\overline{g(z)}dz. \end{aligned}$$

$\mathbb{R}$ Wenn wir zurück zu $C[a, b]$ gehen, sehen wir, dass für eine Funktion $f \in C[a, b]$ gilt:

$$f \perp 1 \iff f \perp c \text{ für alle } c \in \mathbb{R} \iff \int_a^b f(x)dx = 0,$$

wobei wir hier $c \in \mathbb{R}$ mit der konstanten Funktion mit Wert c identifizieren. Also ist zum Beispiel $\sin(x) \in C[-\pi, \pi]$ eine Funktion, die diese Bedingungen erfüllt.

Als ein anderes Beispiel möchten wir an dieser Stelle erwähnen, dass die ganze Theorie der Fourier-Reihen (welche im Fach MMP I im dritten Semester eine wichtige Rolle spielen werden) auf dem folgenden Fakt basiert:

Übung 6.2.17. Zeigen Sie, dass

$$S = \{\sin(nx) \mid n \in \mathbb{N}\} \cup \{\cos(mx) \mid m \in \{0\} \cup \mathbb{N}\} \subseteq C[-\pi, \pi]$$

ein Orthogonalsystem ist.

Wir schauen uns noch ein letztes Beispiel an, bevor wir mit dem nächsten Abschnitt starten.

$\mathbb{F}$ **Beispiel 6.2.18** (Frobenius[3] Skalarprodukt)**.** Sei $V = M_{n\times n}(\mathbb{F})$. Das Produkt

$$\begin{aligned} \langle\, , \rangle_{\text{Frob}} : V \times V &\to \mathbb{F} \\ (A, B) &\mapsto \langle A, B\rangle_{\text{Frob}} = \operatorname{spur}\left(A^T\overline{B}\right) \end{aligned}$$

heisst das *Frobenius Skalarprodukt*.

Übung 6.2.19. Zeigen Sie, dass $\langle\, , \rangle_{\text{Frob}}$ in der Tat ein Skalarprodukt auf $M_{n\times n}(\mathbb{F})$ ist.

6.3 Normen und Winkel

Sei $(V, \langle\, , \rangle)$ ein Skalarproduktraum über $\mathbb{F}$. Wir haben die induzierte Norm definiert als $\|v\| := \sqrt{\langle v, v\rangle}$. Es gibt aber auch eine offizielle Definition, wann eine Funktion $V \to \mathbb{F}$ Norm genannt werden darf:

Definition 6.3.1. Sei V ein Vektorraum über $\mathbb{F}$. Eine Funktion $\|\cdot\| : V \to \mathbb{R}_{\geq 0}$ heisst eine *Norm-Funktion* (oder einfach *Norm*), falls

[3]Georg Frobenius, Vater der Darstellungstheorie.

(1) $\|u+v\| \leq \|u\| + \|v\|$ für alle $u, v \in V$.

(2) $\|\alpha v\| = |\alpha| \, \|v\|$ für alle $v \in V$, $\alpha \in \mathbb{F}$.

(3) $\|v\| = 0 \implies v = 0$ für alle $v \in V$.

Bemerkung 6.3.2. In (3) hätten wir natürlich auch $\iff$ schreiben können, denn die Richtung $\Longleftarrow$ folgt aus (2) mit $\alpha = 0$.

Übung 6.3.3. Sei $(V, \langle\, , \rangle)$ ein Skalarproduktraum. Zeigen Sie, dass die induzierte Norm die Eigenschaften (2) und (3) aus Definition 6.3.1 hat. Eigenschaft (1) gilt auch, aber dies ist weniger direkt zu zeigen. Wir brauchen dafür eine grundlegende Tatsache, die wir in diesem Abschnitt besprechen: Die Cauchy-Schwarz Ungleichung.

$\mathbb{R}$ Eine andere Motivation für die Cauchy-Schwarz Ungleichung ist die Verallgemeinerung vom Begriff eines Winkels zwischen zwei Vektoren:

Übung 6.3.4. Wir betrachten $(\mathbb{R}^2, \langle\, , \rangle)$ mit dem Standard-Skalarprodukt. Seien $u, v \in \mathbb{R}^2$ zwei linear unabhängige Vektoren. Man kann geometrisch herleiten, dass der Winkel ϕ zwischen u und v die Gleichung

$$\cos\phi = \frac{\langle u, v\rangle}{\|u\| \, \|v\|} = \frac{\langle u, v\rangle}{\sqrt{\langle u, u\rangle}\sqrt{\langle v, v\rangle}}$$

erfüllt und somit gegeben ist durch

$$\phi = \arccos\left(\frac{\langle u, v\rangle}{\sqrt{\langle u, u\rangle}\sqrt{\langle v, v\rangle}}\right). \tag{6.3}$$

Insbesondere gilt

$$-1 \leq \frac{\langle u, v\rangle}{\sqrt{\langle u, u\rangle}\sqrt{\langle v, v\rangle}} \leq 1. \tag{6.4}$$

Versuchen Sie, dies geometrisch zu zeigen (z.B. mit dem Cosinussatz).

Wir möchten gerne den Begriff eines Winkels zwischen zwei Vektoren für allgemeine euklidische Vektorräume durch (6.3) definieren. Damit das Sinn macht, müssen wir zeigen, dass (6.4) für allgemeine euklidische Vektorräume gilt. Dies ist genau die Cauchy-Schwarz Ungleichung.

$\mathbb{R} \leftrightarrow \mathbb{C}$ Winkel werden wir nur für euklidische Vektorräume definieren können. Der Grund ist, dass in einem unitären Vektorraum $\langle u, v\rangle$ im Allgemeinen eine komplexe Zahl ist und (6.3) somit keinen Sinn macht.

Dies gibt hoffentlich genug Motivation für den folgenden Satz:

$\mathbb{F}$ **Satz 6.3.5** (Cauchy-Schwarz Ungleichung)**.** *Sei $(V, \langle\,,\rangle)$ ein Skalarproduktraum über $\mathbb{F}$ und sei $\|\cdot\|$ die induzierte Norm. Für alle $u, v \in V$ gilt*

$$|\langle u, v\rangle| \leq \|u\|\,\|v\|\,. \tag{6.5}$$

Gleichheit gilt genau dann, wenn u und v linear abhängig sind.

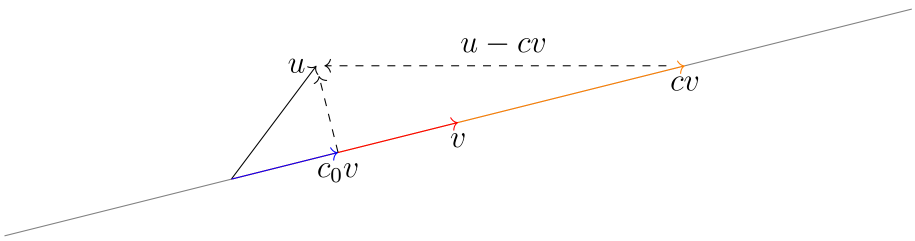
Übung 6.3.6. Zeigen Sie, dass für zwei Vektoren u, v eines euklidischen Vektorräums (6.4) und (6.5) äquivalent sind.

Beweis von Satz 6.3.5. Falls $v = 0$, sind beide Seiten von (6.5) Null und es gibt nichts zu beweisen. In diesem Fall gilt Gleichheit in (6.5) und $\{u, v\}$ ist in der Tat linear abhängig.

Sei also $v \neq 0$. Um Skalarprodukte der Form $\langle u, v\rangle$, $\langle u, u\rangle$ und $\langle v, v\rangle$ zu verbinden, betrachten wir für jedes $c \in \mathbb{F}$ die Aussage

$$0 \leq \|u - cv\|^2 = \langle u - cv, u - cv\rangle = \langle u, u\rangle - c\,\langle v, u\rangle - \bar{c}\,\langle u, v\rangle + c\bar{c}\,\langle v, v\rangle\,. \tag{6.6}$$

Machen wir jetzt eine „Intuition-Pause". Man könnte sich überlegen, wann die Aussage in (6.6) „am Stärksten" ist. Dies ist genau dann der Fall, wenn $\|u - cv\|$ am kleinsten ist. Zum Beispiel, wenn $\mathbb{F} = \mathbb{R}$, dann ist (6.6), als Funktion von c betrachtet, eine Parabel, die ein Minimum bei $c_0 = \frac{\langle u,v\rangle}{\langle v,v\rangle}$ hat. Man könnte es auch „geometrisch" betrachten:



Figur 6.2: Bei c_0v wird der Abstand zu u minimiert.

Zum Beispiel kann man in $\mathbb{R}^2$ mit dem Standard-Skalarprodukt berechnen, dass $\|u - cv\|$ am kleinsten ist, wenn $c = c_0 = \frac{\langle u,v\rangle}{\langle v,v\rangle}$.

Zurück zum Beweis: Wir wählen $c = \frac{\langle u,v\rangle}{\langle v,v\rangle}$ (hier benutzen wir die Annahme $v \neq 0$) und setzen es in (6.6) ein. Mit ein bisschen Algebra bekommen wir

$$0 \leq \langle u, u\rangle - \frac{|\langle u, v\rangle|^2}{\langle v, v\rangle} = \|u\|^2 - \frac{|\langle u, v\rangle|^2}{\|v\|^2}, \tag{6.7}$$

was äquivalent zur Cauchy-Schwarz Ungleichung ist.

Für $v \neq 0$ (der Fall $v = 0$ war oben) ist Gleichheit in (6.5) äquivalent zu Gleichheit in (6.7), was wiederum äquivalent ist zu Gleichheit in (6.6) mit $c = \frac{\langle u,v\rangle}{\langle v,v\rangle}$ und somit zu $\left\|u - \frac{\langle u,v\rangle}{\langle v,v\rangle}v\right\| = 0$. Dies impliziert $u = cv$ und zeigt lineare Abhängigkeit von u und v.

Für die andere Richtung, falls u, v linear abhängig sind, dann ist $u = cv$ für ein $c \in \mathbb{R}$ (man kann zeigen, dass dann $c = \frac{\langle u,v\rangle}{\langle v,v\rangle}$ ist). Dann kann man direkt überprüfen, dass Gleichheit in (6.5) gilt. □

Korollar 6.3.7 (Dreiecksungleichung)**.** *Für die induzierte Norm $\|\cdot\|$ auf einem Skalarproduktraum $(V, \langle\ ,\ \rangle)$ gilt*

$$\|u+v\| \leq \|u\| + \|v\| \quad \forall u, v \in V$$

und somit ist $\|\cdot\|$ in der Tat eine Norm gemäss Definition 6.3.1.

Beweis. Man zeigt direkt, dass

$$\begin{aligned}
\|u+v\|^2 &= \langle u+v, u+v\rangle = \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle \\
&= \|u\|^2 + 2\operatorname{Re}\langle u,v\rangle + \|v\|^2 \\
&\leq \|u\|^2 + 2|\langle u,v\rangle| + \|v\|^2 \\
&\overset{(6.5)}{\leq} \|u\|^2 + 2\|u\| \cdot \|v\| + \|v\|^2 \\
&= (\|u\| + \|v\|)^2.
\end{aligned}$$

Wenn wir nun auf beiden Seiten die Wurzel ziehen, folgt die Aussage. □

$\mathbb{R}$ Wie oben erklärt, ermöglicht dies die Definition eines Winkels in jedem euklidischen Vektorraum:

Definition 6.3.8. Sei $(V, \langle\ ,\ \rangle)$ ein euklidischer Vektorraum. Der *Winkel* zwischen zwei Vektoren $u \neq 0$ und $v \neq 0$ in V ist definiert als die eindeutige Zahl $\alpha \in [0, \pi]$, sodass

$$\cos\alpha = \frac{\langle u,v\rangle}{\|u\|\,\|v\|} = \left\langle \frac{u}{\|u\|}, \frac{v}{\|v\|} \right\rangle.$$

$\mathbb{F}$ **Übung 6.3.9.** Wenden Sie die Cauchy-Schwarz Ungleichung und die Dreiecksungleichung auf alle Beispiele von Skalarprodukträumen an, die wir in Abschnitt 6.2 gesehen haben.

$\mathbb{R}$ Zum Beispiel für $f, g \in C[-\pi, \pi]$ ergibt die Cauchy-Schwarz Ungleichung

$$\left|\int_{-\pi}^{\pi} f(x)g(x)dx\right|^2 \leq \int_{-\pi}^{\pi} f(x)^2 dx \cdot \int_{-\pi}^{\pi} g(x)^2 dx$$

und die Dreiecksungleichung besagt, dass

$$\sqrt{\int_{-\pi}^{\pi} (f(x)+g(x))^2 dx} \leq \sqrt{\int_{-\pi}^{\pi} f(x)^2 dx} + \sqrt{\int_{-\pi}^{\pi} g(x)^2 dx}$$

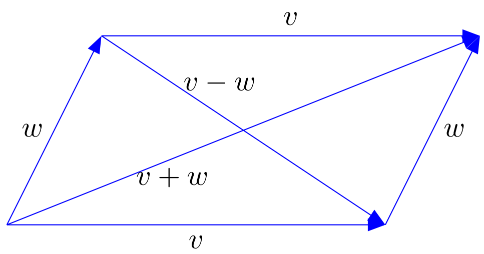
$\mathbb{F}$ *Bemerkung* 6.3.10. Nicht alle Norm-Funktionen sind von einem Skalarprodukt induziert. Die folgende Aufgabe gibt mehr Informationen dazu (Teil (2) ist nicht einfach!).

$\mathbb{R}$ **Übung 6.3.11.** Sei $(V, \langle\ ,\ \rangle)$ ein euklidischer Vektorraum und $\|\cdot\|$ die induzierte Norm.

(1) Beweisen Sie, dass für alle $v, w \in V$ die *Parallelogramm-Gleichung* gilt:

$$\|v + w\|^2 + \|v - w\|^2 = 2\|v\|^2 + 2\|w\|^2 .$$

(2) Sei W ein $\mathbb{R}$-Vektorraum mit einer Norm $\|\cdot\|$. Zeigen Sie, dass $\|\cdot\|$ die Parallelogramm-Gleichung erfüllt genau dann, wenn es ein Skalarprodukt auf W gibt, das die Norm $\|\cdot\|$ induziert.



Figur 6.3: Die Parallelogram-Gleichung.

$\mathbb{F}$

6.4 Das Gram-Schmidt Verfahren

Erinnern Sie sich an die Definition von orthogonalen und orthonormalen Systemen auf Seite 239.

Definition 6.4.1. Eine geordnete Basis $\mathcal{B}$ heisst *orthogonale* (bzw. *orthonormale*) Basis, falls $\mathcal{B}$ ein orthogonales (bzw. orthonormales) System ist.

6.4.1 Wieso sind orthogonale/orthonormale Systeme cool?

Mit solchen Systemen/Basen geht alles leicht:

Proposition 6.4.2. *Sei V ein Skalarproduktraum über $\mathbb{F}$ und sei $\{v_1, \ldots, v_n\}$ ein orthogonales System. Falls $v = \sum_{i=1}^n a_i v_i$, dann sind die a_j gegeben durch*

$$a_j = \frac{\langle v, v_j\rangle}{\langle v_j, v_j\rangle} \text{ für alle } j = 1, \ldots, n.$$

Beweis. $\langle v, v_j\rangle = \left\langle \sum_{i=1}^n a_i v_i, v_j \right\rangle = \sum_{i=1}^n a_i \langle v_i, v_j\rangle = a_j \langle v_j, v_j\rangle.$ □

Falls $\{v_1, \dots, v_n\}$ ein orthonormales System ist, sieht diese Proposition schöner aus:

$$v = \sum_{i=1}^{n} a_i v_i \implies a_j = \langle v, v_j \rangle \;\; \forall j = 1, \dots, n.$$

Vielleicht sehen Sie noch nicht, wieso das so toll sein soll. Erinnern Sie sich, dass man im Allgemeinen, wenn man für $v \in \mathrm{Sp}(v_1, \dots, v_k) \subseteq \mathbb{F}^n$ die Zahlen $a_1, \dots, a_k \in \mathbb{F}$ finden will mit $v = \sum_i a_i v_i$, man ein lineares Gleichungssystem mit $n \times k$ Variablen lösen muss. Falls wir wissen, dass $\{v_1, \dots, v_k\}$ ein Orthonormalsystem ist, müssen wir nur $\langle v, v_j \rangle$ für alle j berechnen.

Beispiel 6.4.3. Wir bemerken, dass die Standard-Basis $\mathcal{E}_n = (e_1, \dots, e_n) \subseteq \mathbb{F}^n$ eine orthonormale Basis bezüglich dem Standard-Skalarprodukt ist. Wenn wir Proposition 6.4.2 in diesem Fall überprüfen, sehen wir: Für jedes $v = (v_1, \dots, v_n) \in \mathbb{F}^n$ gilt in der Tat

$$v = \sum_{i=1}^{n} \langle v, e_i \rangle \, e_i, \text{ da } \quad \langle v, e_i \rangle = v_i.$$

Beispiel 6.4.4. Sei $A = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$ die Matrix von Beispiel 6.2.14. Wir haben gesehen, dass $\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$ eine orthogonale Basis von $(\mathbb{F}^2, \langle \, , \rangle_A)$ ist.

Ein allgemeiner Vektor $v = \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{F}^2$ kann also als

$$v = \frac{\left\langle v, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\rangle_A}{\left\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\rangle_A} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \frac{\left\langle v, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\rangle_A}{\left\langle \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\rangle_A} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

geschrieben werden. In der Tat:

$$\begin{aligned}
\left\langle v, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\rangle_A &= (a, b) \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = a - b, \\
\left\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\rangle_A &= (1, 0) \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 1, \\
\left\langle v, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\rangle_A &= (a, b) \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = b, \\
\left\langle \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\rangle_A &= (1, 1) \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 1,
\end{aligned}$$

und somit $v = (a-b)\begin{pmatrix}1\\0\end{pmatrix} + b\begin{pmatrix}1\\1\end{pmatrix}$. Die Berechnungen zeigen auch, dass $\left(\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}\right)$ sogar eine orthonormale Basis ist.

In diesem Beispiel wussten wir schon, dass $\left(\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}\right)$ linear unabhängig ist. Dies folgt aber auch aus der Tatsache, dass $\left(\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}\right)$ ein Orthogonalsystem ist, wie das folgende grundlegende Korollar zeigt:

Korollar 6.4.5. *Wenn $S = \{v_1, \dots, v_n\}$ ein Orthogonalsystem ist, dann ist S linear unabhängig.*

Beweis. Falls $\sum_i a_i v_i = 0$ ist, berechnen sich die a_i's laut Proposition 6.4.2 als

$$a_i = \frac{\langle 0, v_i \rangle}{\langle v_i, v_i \rangle} = 0.$$

□

Falls Sie immer noch nicht von der Nützlichkeit von Orthogonal- und Orthonormalsystemen überzeugt sind, sollte dies Sie überzeugen:

Um eine Darstellungsmatrix zu berechnen, muss man im allgemeinen n verschiedene lineare $(n \times n)$-Gleichungssysteme lösen. Wenn wir aber eine orthogonale/orthonormale Basis haben, ist alles viel leichter:

Proposition 6.4.6. *Sei V, W Skalarprodukträume über $\mathbb{F}$, $T : V \to W$ eine lineare Abbildung, $\mathcal{B} = (v_1, \dots, v_n)$ eine beliebige Basis von V und $\mathcal{C} = (w_1, \dots, w_m)$ eine orthonormale Basis von W. Sei $(a_{ij})_{i,j} = [T]^{\mathcal{B}}_{\mathcal{C}}$. Dann gilt $a_{ij} = \langle T(v_j), w_i \rangle$.*

Beweis. Die a_{ij} sind durch $T(v_j) = \sum_{i=1}^{m} a_{ij} w_i$ definiert. Laut Proposition 6.4.2 gilt also

$$a_{ij} = \langle T(v_j), w_i \rangle$$

$\left(\text{wenn } \mathcal{C} \text{ nur orthogonal ist, gilt } a_{ij} = \frac{\langle T(v_j), w_i \rangle}{\langle w_i, w_i \rangle}\right)$. □

6.4.2 Gram-Schmidt Orthogonalisierungs-Algorithmus

Hoffentlich sind Sie nun überzeugt, dass es ziemlich nützlich ist, eine orthogonale/orthonormale Basis zu finden. In der Anwendung ist es hilfreich, zu sehen, dass wir viele solche Basen finden können, auch solche, die eine gegebene Basis auf einem Untervektorraum erweitern. Genau dafür haben wir den Gram-Schmidt Algorithmus. Grob gesagt nimmt dieser eine beliebige Basis als Input und gibt uns eine orthogonale/orthonormale Basis als Output. Genauer gesagt haben wir den folgenden Satz:

Satz 6.4.7 (Gram-Schmidt Algorithmus). *Sei $(V, \langle\,,\rangle)$ ein n-dimensionaler Skalarproduktraum über $\mathbb{F}$ und sei $(v_1, \ldots, v_n)$ eine beliebige geordnete Basis von V. Man definiert die Vektoren $w_1, \ldots, w_n$ und $e_1, \ldots, e_n$ rekursiv als*

$$\begin{aligned} w_1 &:= v_1, \\ w_j &:= v_j - \sum_{i=1}^{j-1} \frac{\langle v_j, w_i \rangle}{\langle w_i, w_i \rangle} w_i, \quad j = 2, \ldots, n \end{aligned}$$

und $e_i := \frac{1}{\|w_i\|} w_i$. Es gilt

(1) $(w_1, \ldots, w_n)$ ist eine Orthogonalbasis von V.

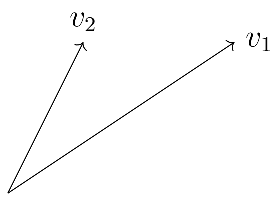
(2) $(e_1, \ldots, e_n)$ ist eine Orthonormalbasis von V.

(3) Für alle $1 \leq i \leq n$ ist $\mathrm{Sp}(v_1, \ldots, v_i) = \mathrm{Sp}(w_1, \ldots, w_i) = \mathrm{Sp}(e_1, \ldots, e_i)$.

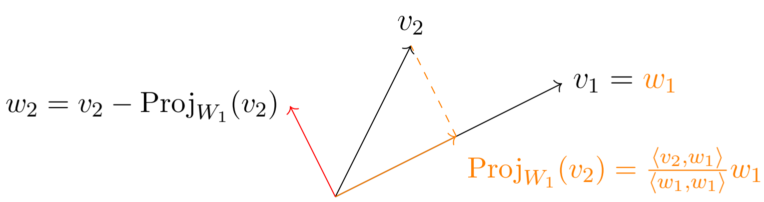
Korollar 6.4.8. *Jeder endlich-dimensionale Skalarproduktraum hat eine orthonormale Basis.*

Bemerkung 6.4.9. Den Term $\sum_{i=1}^{j-1} \frac{\langle v_j, w_i \rangle}{\langle w_i, w_i \rangle} w_i$ nennt man die *orthogonale Projektion* von v_j auf $W_{j-1} := \mathrm{Sp}(w_1, \ldots, w_{j-1})$. Man bezeichnet diesen Term auch als $\mathrm{Proj}_{W_{j-1}}(v_j)$.

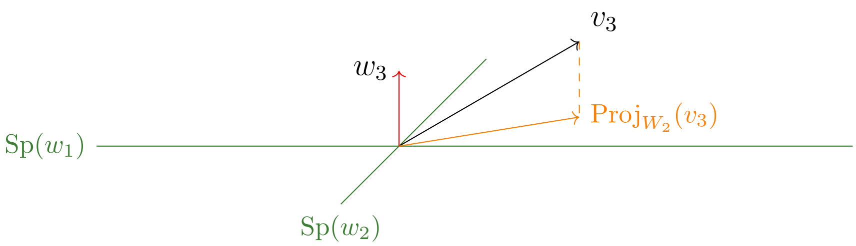
Bevor wir den Satz beweisen, bemerken wir, dass die Rekursionsbedingung sehr natürlich ist. Denken Sie an zwei Vektoren



Wie würden Sie diese benutzen, um (w_1, w_2) zu finden, die orthogonal sind und den gleichen Spann haben? Die Idee ist, $w_1 = v_1$ zu wählen und dann v_2 zu „korrigieren“, indem man die orthogonale Projektion auf v_1 verwendet:



Nehmen wir jetzt noch einen dritten Vektor v_3 hinzu. Um eine orthogonale Basis von $\mathrm{Sp}(w_1, w_2, v_3) = \mathrm{Sp}(v_1, v_2, v_3)$ zu finden, benutzt man dieselbe Idee: $w_3 := v_3 - \mathrm{Proj}_{\mathrm{Sp}(w_1,w_2)}$ soll orthogonal zu w_1 und w_2 sein und $\mathrm{Sp}(w_1, w_2, w_3) = \mathrm{Sp}(w_1, w_2, v_3)$.



Dies gibt uns drei orthogonale Vektoren (w_1, w_2, w_3) mit dem gleichen Spann wie (v_1, v_2, v_3), wie gewünscht. Falls man zusätzlich noch Einheitsvektoren haben möchte, kann man einfach $\left(e_1 := \frac{w_1}{\|w_1\|}, e_2 := \frac{w_2}{\|w_2\|}, e_3 := \frac{w_3}{\|w_3\|}\right)$ nehmen. Dies ändert weder den Spann noch die Orthogonalität.

Beweis von Satz 6.4.7. Der Beweis ist das Durchführen der obigen Projektions-Idee für n Vektoren. Wir zeigen per Induktion über k, dass $(w_1, \ldots, w_k)$ eine orthogonale Basis von $\mathrm{Sp}(v_1, \ldots, v_k)$ ist.

Da $w_1 = v_1$ ist, folgt die Induktionsbasis. Nehmen wir an, wir hätten die Aussage für k gezeigt und betrachten wir nun den Fall $k+1$. Aufgrund der Induktionsannahme genügt es, zu zeigen, dass $w_{k+1} \perp w_j$ für alle $j = 1, \ldots, k$ ist, um zu folgern, dass $(w_1, \ldots, w_{k+1})$ ein orthogonales System ist. Machen wir das nun: Sei $1 \leq j \leq k$.

$$\begin{aligned}\langle w_{k+1}, w_j \rangle &= \left\langle v_{k+1} - \sum_{i=1}^{k} \frac{\langle v_{k+1}, w_i \rangle}{\langle w_i, w_i \rangle} w_i, w_j \right\rangle \\ &= \langle v_{k+1}, w_j \rangle - \sum_{i=1}^{k} \frac{\langle v_{k+1}, w_i \rangle}{\langle w_i, w_i \rangle} \langle w_i, w_j \rangle \\ &\underset{\text{für } i \neq j}{\overset{\langle w_i, w_j \rangle = 0}{=}} \langle v_{k+1}, w_j \rangle - \frac{\langle v_{k+1}, w_j \rangle}{\cancel{\langle w_j, w_j \rangle}} \cancel{\langle w_j, w_j \rangle} = 0.\end{aligned}$$

Dies zeigt, dass die Vektoren $w_1, \ldots, w_{k+1}$ ein orthogonales System bilden und laut Korollar 6.4.5 sind sie auch linear unabhängig. Ausserdem folgt aus der Definition von w_{k+1}, dass

$$\mathrm{Sp}(w_1, \ldots, w_{k+1}) \subseteq \mathrm{Sp}(w_1, \ldots, w_k, v_{k+1}) \underset{\text{annahme}}{\overset{\text{Induktions-}}{=}} \mathrm{Sp}(v_1, \ldots, v_{k+1}).$$

Aus der linearen Unabhängigkeit von $w_1, \ldots, w_{k+1}$ folgt dann Gleichheit. Dies beendet die Induktion.

Die erste Gleichheit in (3) folgt und die zweite folgt auch, da die Skalierung der Basisvektoren den Spann nicht ändert. Aussage (1) folgt aus dem Fall $k = n$, Aussage (2) folgt, da für jedes i gilt:

$$\|e_i\| = \|w_i\|^{-2} \langle w_i, w_i \rangle = 1$$

und

$$\forall 1 \leq i \neq j \leq n \quad \langle e_i, e_j \rangle = \|w_i\|^{-1} \|w_j\|^{-1} \langle w_i, w_j \rangle = 0.$$

□

$\mathbb{R}$ Visualisierung: Man könnte dieses Verfahren in $\mathbb{R}^3$ sehr gut visualisieren und ich war überrascht, dass ich kein gutes GeoGebra-Applet dazu gefunden habe. Leider hatte ich auch keine Zeit, selber eins vorzubereiten. Falls Sie eines finden/bauen, sagen Sie es mir! In der Zwischenzeit ist dieses GIF gar nicht schlecht: Klick.

$\mathbb{F}$ *Bemerkung* 6.4.10. Hier sind einige Varianten des Gram-Schmidt Algorithmus:

(1) In undendlich-dimensionalen Skalarprodukträumen kann man den Algorithmus auf eine abzählbare linear unabhängige Liste $(v_1, v_2 \ldots)$ anwenden. Der Output ist ein orthogonales System $(w_1, w_2, \ldots)$ und ein orthonormales System $(e_1, e_2, \ldots)$, sodass

$$\forall k \in \mathbb{N} \quad \mathrm{Sp}(v_1, \ldots, v_k) = \mathrm{Sp}(w_1, \ldots, w_k) = \mathrm{Sp}(e_1, \ldots, e_k).$$

(2) Falls $(v_1, \ldots, v_k)$ schon orthogonal (bzw. orthonormal) ist, folgt $v_i = w_i$ (bzw. $v_i = w_i = e_i$) für alle $i = 1, \ldots, k$. Dies impliziert das folgende:

Sei U ein Untervektorraum eines endlich-dimensionalen Skalarproduktraums und sei $(v_1, \ldots, v_k)$ eine orthogonale (bzw. orthonormale) Basis von U. Dann kann man diese Basis zu einer orthogonalen (bzw. orthonormalen) Basis von V erweitern.

(3) Man kann den Algorithmus auch auf eine beliebige (d.h. nicht zwingend linear unabhängige) Liste $(v_1, v_2, \ldots)$ anwenden. Falls $v_j \in \mathrm{Sp}(v_1, \ldots, v_{j-1})$, dann bekommt man den Nullvektor, wenn man w_j berechnet. Dann muss man diesen Vektor ignorieren und mit dem Algorithmus weitermachen. So bekommt man für jedes $k \in \mathbb{N}$ eine orthogonale/orthonormale Liste $(w_1, \ldots, w_{r(k)})$ mit

$$0 \leq r(k) = \dim \mathrm{Sp}(v_1, \ldots, v_k) \leq k \quad \text{und} \quad \mathrm{Sp}(w_1, \ldots, w_{r(k)}) = \mathrm{Sp}(v_1, \ldots, v_k).$$

$\mathbb{R}$ **Beispiel 6.4.11.** Wir möchten eine Orthonormalbasis von $\mathbb{R}[x]_2$ finden, als Unterraum von $(C[-1,1], \langle\,,\rangle)$ betrachtet (siehe Beispiel 6.2.15). Sei $(v_1, v_2, v_3) = (1, x, x^2)$ die Standard-Basis von $\mathbb{R}[x]_2$. Wir wenden den Gram-Schmidt Algorithmus an. Es ist $w_1 = v_1 = 1$ und wir berechnen

$$w_2 = v_2 - \frac{\langle v_2, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 = x - \frac{\langle x, 1 \rangle}{\langle 1, 1 \rangle} 1 = x - \frac{\int_{-1}^{1} x dx}{\int_{-1}^{1} 1 dx} 1 = x - \frac{0}{2} \cdot 1 = x.$$

Dies besagt, dass $(1, x)$ bereits orthogonal zueinander sind in $C[-1,1]$. Jetzt berechnen wir w_3:

$$w_3 = v_3 - \frac{\langle v_3, w_2\rangle}{\langle w_2, w_2\rangle} w_2 - \frac{\langle v_3, w_1\rangle}{\langle w_1, w_1\rangle} w_1 = x^2 - \frac{\int_{-1}^1 x^3 dx}{\int_{-1}^1 x^2 dx} x - \frac{\int_{-1}^1 x^2 dx}{\int_{-1}^1 1 dx} 1 = x^2 - \frac{0}{2} x - \frac{\frac{2}{3}}{2} \cdot 1 = x^2 - \frac{1}{3}.$$

Jetzt normieren wir w_1, w_2, w_3:

$$\begin{aligned} e_1 &= \|w_1\|^{-1} w_1 = \frac{1}{\sqrt{\int_{-1}^1 1dx}} \cdot 1 = \frac{1}{\sqrt{2}}, \\ e_2 &= \|w_2\|^{-1} w_2 = \frac{1}{\sqrt{\int_{-1}^1 x^2 dx}} x = \sqrt{\frac{3}{2}} x, \\ e_3 &= \|w_3\|^{-1} w_3 = \frac{1}{\sqrt{\int_{-1}^1 (x^2 - \frac{1}{3})^2 dx}} \left(x^2 - \frac{1}{3}\right) = \sqrt{\frac{45}{8}} \left(x^2 - \frac{1}{3}\right). \end{aligned}$$

Somit ist $\left(\frac{1}{\sqrt{2}}, \sqrt{\frac{3}{2}} x, \sqrt{\frac{45}{8}} \left(x^2 - \frac{1}{3}\right)\right)$ eine Orthonormalbasis von $\mathbb{R}[x]_2$ als Untervektorraum von $C[-1,1]$.

$\mathbb{F}$ Wir schliessen diesen Abschnitt mit einer Verallgemeinerung von Proposition 6.1.9 (Pythagoras), die wir später brauchen werden.

Lemma 6.4.12. *Sei V ein endlich-dimensionaler Skalarproduktraum mit einer orthonormalen Basis $(e_1, \ldots, e_n)$. Sei $v = \sum_{i=1}^n a_i e_i$. Dann gilt*

$$\|v\|^2 = \sum_{i=1}^n |a_i|^2$$

(oder mithilfe von Proposition 6.4.2, $\|v\|^2 = \sum_{i=1}^n |\langle v, e_i\rangle|^2$).

Beweis. Es gilt

$$\|v\|^2 = \langle v, v\rangle = \left\langle \sum_{i=1}^n a_i e_i, \sum_{j=1}^n a_j e_j \right\rangle = \sum_{i=1}^n \sum_{j=1}^n a_i \overline{a_j} \delta_{ij} = \sum_{k=1}^n a_k \overline{a_k} = \sum_{k=1}^n |a_k|^2.$$

□

6.4.3 Orthogonale und unitäre Matrizen

Bevor wir einige Korollare von Gram-Schmidt beweisen, definieren wir zwei wichtige Familien von Matrizen, welche eine zentrale Rolle spielen werden, wenn wir Endomorphismen von Skalarprodukträumen betrachten.

Definition 6.4.13. Eine $n \times n$-Matrix, deren Spalten eine orthonormale Basis von $\mathbb{F}^n$ bezüglich des Standard-Skalarprodukts bilden, heisst *orthogonal* falls $\mathbb{F} = \mathbb{R}$, und *unitär* falls $\mathbb{F} = \mathbb{C}$.

Wir bezeichnen mit $\mathrm{O}(n) \subseteq M_{n\times n}(\mathbb{R})$ die Menge aller orthogonalen Matrizen und mit $\mathrm{U}(n) \subseteq M_{n\times n}(\mathbb{C})$ die Menge aller unitären Matrizen.

$\mathbb{R}$ *Bemerkung* 6.4.14. Wir benutzen den Namen orthogonale Matrix (und nicht orthonormale Matrix), da diese Matrizen eine Gruppe bilden, welche die *orthogonale Gruppe* heisst. Wir werden später sehen, dass dies genau die Gruppe derjenigen linearen Transformationen ist, welche Orthogonalität (und allgemeiner das Skalarprodukt) erhalten bezüglich dem Standard-Skalarprodukt auf $\mathbb{R}^n$.

Lemma 6.4.15. *Für eine Matrix $A \in M_{n\times n}(\mathbb{R})$ gilt:*

$$A \text{ ist orthogonal} \iff A^TA = I_n \iff AA^T = I_n \iff A^{-1} = A^T.$$

$\mathbb{C}$ **Lemma 6.4.16.** *Für eine Matrix $A \in M_{n\times n}(\mathbb{C})$ gilt:*

$$A \text{ ist unitär} \iff A^T\overline{A} = I_n \iff A\overline{A}^T = I_n \iff A^{-1} = \overline{A}^T.$$

$\mathbb{F}$ *Beweis von 6.4.15 und 6.4.16.* Sei $A = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix}$. Gemäss der Definition des Standard-Skalarprodukts $\langle\, ,\rangle$ auf $\mathbb{F}^n$ gilt dann

$$A^T\overline{A} = \begin{pmatrix} — & v_1^T & — \\ & \vdots & \\ — & v_n^T & — \end{pmatrix} \begin{pmatrix} | & & | \\ \overline{v_1} & \cdots & \overline{v_n} \\ | & & | \end{pmatrix} = \left(v_i^T\overline{v_j}\right)_{1\le i,j\le n} = \left(\langle v_i, v_j\rangle\right)_{1\le i,j\le n}.$$

Da $(\langle v_i, v_j\rangle)_{1\le i,j\le n} = I_n$ genau dann, wenn $(v_1, \ldots, v_n)$ eine Orthonormalbasis ist, folgt die erste Äquivalenz. Die anderen zwei sind leicht zu überprüfen. □

Den Beweis des folgenden Korollars überlassen wir den Lesern:

Korollar 6.4.17. *Eine Matrix ist orthogonal (bzw. unitär) genau dann, wenn die Zeilenvektoren eine Orthonormalbasis bilden.*

6.4.4 QR-Zerlegung

Wir betrachten $\mathbb{F}^n$ mit dem Standard-Skalarprodukt und wollen den Gram-Schmidt Algorithmus beschreiben, indem wir Matrizen verwenden. Sei $(v_1, \ldots, v_n)$ eine beliebige Basis und sei $(e_1, \ldots, e_n)$ die orthonormale Basis, welche wir durch die Anwendung von

Gram-Schmidt auf $(v_1, \ldots, v_n)$ erhalten. Wir schreiben jetzt die Vektoren von der Basis $(v_1, \ldots, v_n)$ als Linearkombinationen der Basis $(e_1, \ldots, e_n)$: Aus (3) von Satz 6.4.7 und Proposition 6.4.2 folgt

$$\begin{aligned} v_1 &= \langle v_1, e_1 \rangle\, e_1, \\ v_2 &= \langle v_2, e_1 \rangle\, e_1 + \langle v_2, e_2 \rangle\, e_2, \\ &\vdots \\ v_n &= \langle v_n, e_1 \rangle\, e_1 + \cdots + \langle v_n, e_n \rangle\, e_n. \end{aligned}$$

In der Sprache von Matrizen ausgedrückt, bedeutet dies (vergleiche Lemma 2.2.17)

$$\underbrace{\begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix}}_{A=} = \underbrace{\begin{pmatrix} | & & | \\ e_1 & \cdots & e_n \\ | & & | \end{pmatrix}}_{Q=} \underbrace{\begin{pmatrix} \langle v_1, e_1 \rangle & \langle v_2, e_1 \rangle & \cdots & \langle v_n, e_1 \rangle \\ & \langle v_2, e_2 \rangle & & \vdots \\ & & \ddots & \vdots \\ & & & \langle v_n, e_n \rangle \end{pmatrix}}_{R=}.$$

Wir haben also die folgende Zerlegung gezeigt (und sogar einen Algorithmus zur Berechnung angegeben):

$\mathbb{R}$ **Satz 6.4.18** (QR-Zerlegung über $\mathbb{R}$, voller Rang)**.** *Sei $A \in \mathrm{GL}_n(\mathbb{R})$. Dann existiert eine orthogonale Matrix Q und eine obere Dreiecksmatrix R, sodass $A = QR$.*

$\mathbb{C}$ **Satz 6.4.19** (QR-Zerlegung über $\mathbb{C}$, voller Rang)**.** *Sei $A \in \mathrm{GL}_n(\mathbb{C})$. Dann existiert eine unitäre Matrix Q und eine obere Dreiecksmatrix R, sodass $A = QR$.*

$\mathbb{F}$ **Übung 6.4.20.** Es existiert eine QR-Zerlegung für jede Matrix $A \in M_{n\times n}(\mathbb{F})$, nicht nur für invertierbare Matrizen. Ein Weg, dies zu zeigen, ist der folgende:

Sei $r = \dim \mathrm{SR}(A)$ der Spaltenrang von A. Man wendet die Variante (3) von Bemerkung 6.4.10 auf die Spalten von A an und bekommt eine orthonormale Basis $e_1, \ldots, e_r$ vom Spaltenraum $\mathrm{SR}(A)$. Man erweitert dies zu einer Orthonormalbasis $e_1, \ldots, e_n$ von $\mathbb{F}^n$ (siehe Variante (2) von Bemerkung 6.4.10) und definiert Q und R wie oben. Es folgt $A = QR$ und R ist von der Form

$$R = \left(\begin{array}{c|c} C & (*) \\ \hline 0 & 0 \end{array}\right),$$

wobei $C \in M_{r\times r}$ eine obere Dreiecksmatrix ist und $(*)$ für eine $r \times (n-r)$ Matrix steht. Füllen Sie die Details ein in dieser Beweis-Skizze.

Beispiel 6.4.21. Für jedes $x \in \mathbb{R}$ gilt

$$\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix},$$
$$\begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{1+x^2}} & \frac{-x}{\sqrt{1+x^2}} \\ \frac{x}{\sqrt{1+x^2}} & \frac{1}{\sqrt{1+x^2}} \end{pmatrix} \cdot \begin{pmatrix} \sqrt{1+x^2} & \frac{x}{\sqrt{1+x^2}} \\ 0 & \frac{1}{\sqrt{1+x^2}} \end{pmatrix}.$$

Für $x \to \infty$ bleibt Q innerhalb der kompakten Menge[4] $\mathrm{O}(n)$ und Wachstum findet nur in dem Faktor R der Zerlegung QR statt.

6.4.5 Schur-Zerlegung

Als Folge des Gram-Schmidt Verfahrens beweisen wir, dass jeder trigonalisierbare Endomorphismus T auf einem endlich-dimensionalen Vektorraum V über $\mathbb{F}$ auch „orthogonal-trigonalisierbar" ist. Präziser gesagt:

Satz 6.4.22 (Schur)**.** *Sei $T \in \mathrm{End}(V)$ trigonalisierbar, wobei V ein endlich-dimensionaler Skalarproduktraum ist mit $\dim V = n$. Dann existiert auch eine orthonormale Basis $\mathcal{B} \subseteq V$, sodass $[T]_{\mathcal{B}}$ eine obere Dreiecksmatrix ist.*

Beweis. Dieser Beweis benutzt natürlich die Resultate aus Abschnitt 5.4. Sei

$$\{0\} = V_0 \subseteq V_1 \subseteq \cdots \subseteq V_n = V$$

eine T-invariante Fahne und seien $v_1, \ldots, v_n \in V$ sodass $\mathrm{Sp}(v_1, \ldots, v_i) = V_i$ für alle $i = 1, \ldots, n$. Wir wenden den Gram-Schmidt Algorithmus auf $(v_1, \ldots, v_n)$ und bezeichnen mit $\mathcal{B} := (e_1, \ldots, e_n)$ die resultierende orthonormale Basis von V. Laut (3) von Satz 6.4.7 gilt $\mathrm{Sp}(e_1, \ldots, e_i) = \mathrm{Sp}(v_1, \ldots, v_i)$, was T-invariant ist. Daher folgt, dass $[T]_{\mathcal{B}}$ eine obere Dreiecksmatrix ist, wie wir zeigen wollten. □

$\mathbb{R}$ **Korollar 6.4.23.** *Ein Endomorphismus T auf einem euklidischen Vektorraum ist orthogonal trigonalisierbar genau dann, wenn p_T in Linearfaktoren zerfällt.*

Beweis. Laut Satz 5.4.5 zerfällt p_T genau dann in Linearfaktoren, wenn T trigonalisierbar ist. Dies ist äquivalent zu orthogonaler Trigonalisierbarkeit laut Satz 6.4.22. □

$\mathbb{C}$ **Korollar 6.4.24.** *Jeder Endomorphismus auf einem unitären Vektorraum ist orthogonal trigonalisierbar.*

Beweis. Jeder Endomorphismus ist über $\mathbb{C}$ trigonalisierbar, also folgt dieses Korollar aus Satz 6.4.22 wie in Korollar 6.4.23. □

[4]Die Kompaktheit von $\mathrm{O}(n)$ werden wir im nächsten Kapitel beweisen, siehe Proposition 7.4.11.

$\mathbb{F}$ Wenn wir die Korollare 6.4.23 und 6.4.24 im Skalarproduktraum $\mathbb{F}^n$ mit dem Standard-Skalarprodukt anwenden, bekommen wir:

$\mathbb{R}$ **Korollar 6.4.25** (Schur-Zerlegung über $\mathbb{R}$)**.** *Sei $A \in M_{n\times n}(\mathbb{R})$, sodass p_A in Linearfaktoren zerfällt. Dann existiert eine orthogonale Matrix $O \in M_{n\times n}(\mathbb{R})$ und eine obere Dreiecksmatrix $R \in M_{n\times n}(\mathbb{R})$, sodass $O^T AO = R$.*

$\mathbb{C}$ **Korollar 6.4.26** (Schur-Zerlegung über $\mathbb{C}$)**.** *Sei $A \in M_{n\times n}(\mathbb{C})$. Dann existiert eine unitäre Matrix $U \in M_{n\times n}(\mathbb{C})$ und eine obere Dreiecksmatrix $R \in M_{n\times n}(\mathbb{C})$, sodass $\overline{U}^T AU = R$.*

$\mathbb{F}$ *Beweis von 6.4.25 und 6.4.26.* Wir betrachten die lineare Abbildung $m_A : \mathbb{F}^n \to \mathbb{F}^n$, wobei $\mathbb{F}^n$ mit dem Standard-Skalarprodukt versehen ist. Laut den entsprechenden Korollaren 6.4.23 und 6.4.24 existiert eine orthonormale Basis $\mathcal{B}$, sodass $[m_A]_{\mathcal{B}}$ eine obere Dreiecksmatrix ist. Wir definieren $R := [m_A]_{\mathcal{B}}$ und U (bzw. O) als $[\mathrm{Id}_{\mathbb{F}^n}]^{\mathcal{B}}_{\mathcal{E}_n}$. Laut Definition 6.4.13 ist U unitär (bzw. O orthogonal), da ihre Spalten eine Orthonormalbasis von $\mathbb{F}^n$ bilden. Die Korollare folgen also von

$$R = [m_A]_{\mathcal{B}} = \left([\mathrm{Id}_{\mathbb{F}^n}]^{\mathcal{B}}_{\mathcal{E}_n}\right)^{-1} [m_A]_{\mathcal{E}_n} [\mathrm{Id}_{\mathbb{F}^n}]^{\mathcal{B}}_{\mathcal{E}_n} = \begin{cases} O^T AO, & \mathbb{F} = \mathbb{R}, \\ \overline{U}^T AU, & \mathbb{F} = \mathbb{C}. \end{cases}$$

□

6.5 Dualraum I - in Skalarprodukträumen

K Sei V ein Vektorraum über einem Körper K und betrachten wir $K = K^1$ als Vektorraum über K. Der Vektorraum $\mathrm{Hom}(V, K)$ (siehe Abschnitt 3.5 für die Definition) kommt in vielen verschiedenen Kontexten vor, daher hat er einen speziellen Namen und eine spezielle Notation: Wir schreiben $V^* := \mathrm{Hom}(V, K)$ und nennen dies den *Dualraum von V*. Die Elemente von V^* heissen *Linearformen*. Der Lesbarkeit halber wiederholen wir die Definition von Abschnitt 3.5 für diesen Fall.

Definition 6.5.1. Eine Funktion $\varphi : V \to K$ heisst *Linearform* auf V, falls

$$\forall v, w \in V,\ \alpha \in K \quad \varphi(v + w) = \varphi(v) + \varphi(w) \text{ und } \varphi(\alpha v) = \alpha\varphi(v).$$

Die Menge aller Linearformen wird mit V^* bezeichnet und ist ein K-Vektorraum mit den Operationen

$$(\varphi + \psi)(v) := \varphi(v) + \psi(v), \quad (\alpha\varphi)(v) := \alpha\varphi(v).$$

Beispiel 6.5.2. Im Fall von $V = K^n$ können wir den Dualraum $(K^n)^*$ durch die Darstellungsmatrix besser verstehen:

Sei $\varphi \in (K^n)^*$. Dann existiert $A = (a_1, \ldots, a_n) \in M_{1\times n}(K)$, sodass

$$\varphi(v) = \varphi \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \sum_{i=1}^n a_i v_i.$$

Die gesuchte Matrix ist gegeben durch $A = [\varphi]^{\mathcal{E}_n}_{\mathcal{E}_1}$. Wenn wir nun K^n mit dem Produkt

$$\left\langle \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}, \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \right\rangle = \sum_{i=1}^n v_i w_i$$

versehen, dann gilt

$$\varphi(v) = \left\langle \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}, v \right\rangle = \left\langle v, \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \right\rangle.$$

Bemerken Sie auch, dass für ein gegebenes $\varphi \in V^*$ die $1\times n$-Matrix $(a_1, \ldots, a_n)$ eindeutig definiert ist.

$\mathbb{R}$ **Beispiel 6.5.3.** Betrachten wir $\mathbb{R}[x]_n$ als Untervektorraum des Skalarproduktraums $(C[-1,1], \langle\, , \rangle)$. Jedes Element $f \in C[-1,1]$ definiert eine Linearform φ_f auf $\mathbb{R}[x]_n$ durch

$$\forall p \in \mathbb{R}[x]_n \quad \varphi_f(p) = \langle p, f\rangle = \int_{-1}^1 p(x)f(x)dx.$$

Wir werden jedoch gleich sehen, dass jede Linearform auf $\mathbb{R}[x]_n$ von der Form $p \mapsto \langle p, q\rangle$ für ein $q \in \mathbb{R}[x]_n$ ist. Beispielsweise existiert also $q \in \mathbb{R}[x]_n$ mit

$$\langle p, q\rangle = \varphi_{\cos(\pi x)}(p) = \langle p, \cos(\pi x)\rangle \quad \forall p \in \mathbb{R}[x]_n.$$

Da $\cos(\pi x)$ nicht in $\mathbb{R}[x]_n$ liegt, ist dies überhaupt nicht offensichtlich (zumindest ist es nicht sehr klar, welches Polynom dieses $q \in \mathbb{R}[x]_n$ sein soll).

$\mathbb{F}$ **Beispiel 6.5.4.** Sei $(V, \langle\, , \rangle)$ ein Skalarproduktraum über $\mathbb{F}$ und sei $u \in V$. Die Abbildung

$$\begin{aligned} \varphi_u : V &\to \mathbb{F} \\ v &\mapsto \langle v, u\rangle \end{aligned}$$

ist eine Linearform auf V.

Der folgende grundlegende Satz zeigt, dass alle Linearformen auf Skalarprodukträumen die Form φ_u haben:

Satz 6.5.5. *[Darstellungssatz von Riesz[5]] Sei $(V, \langle\,,\rangle)$ ein n-dimensionaler Skalarproduktraum über $\mathbb{F}$ und $\varphi \in V^*$ eine Linearform. Dann existiert ein eindeutiges[6] $u = u(\varphi) \in V$, sodass $\varphi(v) = \langle v, u\rangle$ für alle $v \in V$.*

$\mathbb{R}$ *Erster (nicht konstruktiver) Beweis (über $\mathbb{R}$).* Betrachten wir die Abbildung

$$\begin{aligned}\Phi : V &\to V^*\\ u &\mapsto \varphi_u,\end{aligned}$$

wobei $\varphi_u(v) = \langle v, u\rangle$ (wie in Beispiel 6.5.4). Man überprüft, dass Φ linear ist. Ausserdem ist

$$\operatorname{Ker}(\Phi) = \{u \in V \mid \varphi_u = 0\} = \{u \in V \mid \langle v, u\rangle = 0\ \forall v \in V\} \overset{\text{Lemma 6.1.6}}{=} \{0_V\} \tag{6.8}$$

und daher ist Φ injektiv. Aus Abschnitt 3.5 wissen wir, dass

$$\dim V^* = \dim \operatorname{Hom}(V, \mathbb{F}) = \dim V \cdot \dim \mathbb{F} = \dim V$$

und daher ist Φ ein Isomorphismus. Dies ist äquivalent zur Aussage, die wir beweisen wollten: Jedes Element von V^* hat die Form $\varphi_u = \Phi(u)$ für ein eindeutiges $u \in V$. □

$\mathbb{F}$ *Zweiter Beweis (über $\mathbb{F}$).* Sei $e_1, \ldots, e_n$ eine orthonormale Basis von V. Laut Proposition 6.4.2 gilt $v = \sum_{i=1}^n \langle v, e_i\rangle e_i$ und somit ist

$$\varphi(v) = \varphi\left(\sum_{i=1}^n \langle v, e_i\rangle e_i\right) = \sum_{i=1}^n \langle v, e_i\rangle \varphi(e_i) = \left\langle v, \underbrace{\sum_{i=1}^n \overline{\varphi(e_i)} e_i}_{u:=}\right\rangle = \langle v, u\rangle . \tag{6.9}$$

Dies zeigt die Existenz von u. Die Eindeutigkeit zeigen wir wie folgt: Falls $\varphi_{u_1} = \varphi_{u_2}$, so gilt $\langle v, u_1\rangle = \langle v, u_2\rangle$ für alle $v \in V$ und damit $u_1 = u_2$ gemäss Lemma 6.1.6 (vergleiche mit (6.8)). □

Wir protokollieren:

Rezept: Sei $\varphi \in V^*$. Den eindeutigen Vektor $u \in V$ mit $\varphi(v) = \langle v, u\rangle\ \forall v \in V$ kann man wie folgt finden: Man nimmt eine orthonormale Basis $e_1, \ldots, e_n$ von V und berechnet $u = \sum_{i=1}^n \overline{\varphi(e_i)} e_i$.

[5] Frigyes Riesz, ungarischer Mathematiker 1880-1956.
[6] Die Schreibweise $u = u(\varphi)$ betont, dass u von φ abhängt.

Bemerkung 6.5.6. Es gibt viele orthonormale Basen von V (wie wir dank Gram-Schmidt wissen). Der Vektor $\sum_{i=1}^{n} \overline{\varphi(e_i)} e_i$ hängt nicht von der Wahl einer orthonormalen Basis ab! Er hängt nur von φ ab.

$\mathbb{C}$ *Bemerkung* 6.5.7. Man könnte den ersten Beweis auch über $\mathbb{C}$ führen, aber dann ist die Abbildung Φ nicht linear, sondern semilinear. Das heisst, für $u \in V$ und $c \in \mathbb{C}$ ist $\Phi(cu) = \overline{c}\Phi(u)$. Um nicht die ganze Theorie auch für semilineare Abbildung einführen zu müssen, haben wir den Fall $\mathbb{F} = \mathbb{C}$ im ersten Beweis weggelassen.

$\mathbb{R}$ **Übung 6.5.8.** Finden Sie $u \in \mathbb{R}[x]_2$, sodass für alle $p \in \mathbb{R}[x]_2$ (als Skalarprodukt-Untervektorraum von $C[-1,1]$) gilt

$$\langle p, \cos \pi x \rangle = \langle p, u \rangle \, .$$

Lösung: Sei $\varphi(p) := \langle p, \cos(\pi x) \rangle = \int_{-1}^{1} p(t) \cos(\pi t) dt$. Wenn wir (6.9) aus dem Beweis oben anwenden und die orthonormale Basis aus Beispiel 6.4.11 benutzen, erhalten wir

$$\begin{aligned} u(x) = &\left(\int_{-1}^{1} \sqrt{\tfrac{1}{2}} \cos(\pi t) dt \right) \sqrt{\tfrac{1}{2}} + \left(\int_{-1}^{1} \sqrt{\tfrac{3}{2}} t \cos(\pi t) dt \right) \sqrt{\tfrac{3}{2}} x \\ &+ \left(\int_{-1}^{1} \sqrt{\tfrac{45}{8}} \left(t^2 - \tfrac{1}{3}\right) \cos(\pi t) dt \right) \sqrt{\tfrac{45}{8}} \left(x^2 - \tfrac{1}{3}\right) . \end{aligned}$$

Mit ein bisschen Analysis erhält man schliesslich

$$u(x) = -\frac{45}{2\pi^2} \left(x^2 - \tfrac{1}{3}\right) .$$

$\mathbb{F}$ *Bemerkung* 6.5.9. Der Darstellungssatz von Riesz besagt, dass es einen kanonischen Isomorphismus zwischen V und V^* gibt, falls V ein endlich-dimensionaler **Skalarproduktraum** ist. Das heisst, der Isomorphismus Φ hängt nur von der Vektorraumstruktur und dem Skalarprodukt ab.

K Wenn man hingegen nur die Struktur eines endlich-dimensionalen Vektorraums betrachtet, so sind V und V^* zwar auch isomorph (beide haben die gleiche Dimension), aber es existiert kein kanonischer Isomorphismus (d.h. ein Isomorphismus, der nur von der Vektorraumstruktur abhängt, also insbesondere nicht von der Wahl einer Basis). Wir werden später zeigen, dass V und $(V^*)^*$ hingegen kanonisch isomorph sind (falls V endlich-dimensional ist).

Sie können es als eine Übung betrachten, jetzt schon einen Isomorphismus zwischen V und $(V^*)^*$ zu finden (Hinweis: Sei $v \in V$. Wie könnten sie ein Element von $(V^*)^*$ mittels v definieren?).

6.6 Das orthogonale Komplement

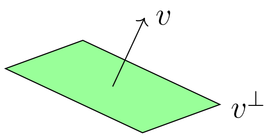
$\mathbb{F}$ Im folgenden Abschnitt bezeichnet V einen Skalarproduktraum über $\mathbb{F}$. In Definition 6.1.11 haben wir Orthogonalität definiert. Nun gehen wir einen Schritt weiter und definieren das orthogonale Komplement einer Teilmenge $S \subseteq V$. Dies ist die maximale zu S orthogonale Teilmenge von V.

Definition 6.6.1. Für eine Teilmenge $S \subseteq V$ definieren wir das *orthogonale Komplement* von S als

$$S^\perp := \{v \in V \mid \langle s, v\rangle = 0 \ \forall s \in S\} = \{v \in V \mid \langle v, s\rangle = 0 \ \forall s \in S\}.$$

Für $S = \{v\}$ schreiben wir $v^\perp := \{v\}^\perp$.

$\mathbb{R}$ In unseren „go-to“ Beispielen kennen wir diesen Begriff schon seit der Mittelschule. Zum Beispiel für $v \in \mathbb{R}^3$ ist $v^\perp$ eine Ebene durch den Ursprung mit v als Normalvektor. Umgekehrt ist das orthogonale Komplement einer Ebene durch den Ursprung genau der Spann des Normalvektors.



Figur 6.4: Das orthogonale Komplement eines Vektors in $\mathbb{R}^3$.

$\mathbb{F}$ **Lemma 6.6.2.** *Für Teilmengen $S, T \subseteq V$ gilt:*

(1) $S^\perp$ *ist ein Untervektorraum von* V*.*

(2) $0^\perp = V$ *und* $V^\perp = \{0\}$*.*

(3) $S \cap S^\perp \subseteq \{0\}$ *(und* $S \cap S^\perp = \{0\}$ *falls* $0 \in S$*).*

(4) $S \subseteq T \implies T^\perp \subseteq S^\perp$*.*

(5) $\mathrm{Sp}(S)^\perp = S^\perp$*.*

(6) $S \subseteq (S^\perp)^\perp$*.*

Beweis. Für (1) bemerken Sie zunächst, dass $0_V \in S^\perp$, da $0 \perp v$ für alle $v \in V$ laut (1) von Lemma 6.1.6. Für $\alpha, \beta \in \mathbb{F}$ und $v, w \in S^\perp$ gilt für alle $s \in S$

$$\langle \alpha v + \beta w, s\rangle = \alpha \langle v, s\rangle + \beta \langle w, s\rangle = 0 + 0 = 0,$$

also ist $\alpha v + \beta w \in S^\perp$, was (1) zeigt.

(2): Dies folgt aus (1) und (2) von Lemma 6.1.6.

(3): Falls $S \cap S^\perp \neq \emptyset$ sei $s \in S \cap S^\perp$. Dann gilt $\langle s, s\rangle = 0$, was $s = 0$ impliziert.

(4): Falls $S \subseteq T$, dann gilt für $v \in V$, dass

$$v \in T^\perp \iff \langle v, s\rangle = 0 \;\forall s \in T \implies \langle v, s\rangle = 0 \;\forall s \in S \iff v \in S^\perp.$$

(5): Da $S \subseteq \mathrm{Sp}(S)$, folgt $\mathrm{Sp}(S)^\perp \subseteq S^\perp$ aus (4). Für $v \in S^\perp$ und $u = \sum_i \alpha_i s_i$ mit $s_i \in S$, $\alpha_i \in \mathbb{F}$ gilt

$$\langle v, u\rangle = \left\langle v, \sum_i \alpha_i s_i \right\rangle = \sum_i \overline{\alpha_i} \langle v, s_i\rangle = 0,$$

was $v \in \mathrm{Sp}(S)^\perp$ zeigt. Da $v \in S^\perp$ beliebig war, folgt $S^\perp \subseteq \mathrm{Sp}(S)^\perp$.

(6): Sei $s \in S$. Per Definition von $S^\perp$ gilt $\langle s, s'\rangle = 0$ für alle $s' \in S^\perp$. Dies besagt per Definition, dass $s \in (S^\perp)^\perp$. □

Hier ist das Hauptresultat von diesem Abschnitt:

Satz 6.6.3. *Sei U ein endlich-dimensionaler Unterverktorraum von V (wir nehmen nicht an, dass V selbst endlich-dimensional ist!). Dann gilt $V = U \oplus U^\perp$.*

Beweis. Laut (3) (und streng genommen auch (1)) von Lemma 6.6.2 genügt es, zu zeigen, dass $U + U^\perp = V$. Inspiriert von $\mathbb{R}^2$ und auch von Figur 6.1, wollen wir einen gegebenen Vektor $v \in V$ orthogonal auf U projizieren und hoffen, dass die Differenz in $U^\perp$ liegt. Die Berechnung in $\mathbb{R}^2$ und der Beweis von Gram-Schmidt haben bereits angedeutet, wie das funktionieren könnte:

Sei $(e_1, \ldots, e_r)$ eine orthonormale Basis von U. Wir definieren die orthogonale Projektion von v auf U als

$$P_U(v) := \langle v, e_1\rangle e_1 + \cdots + \langle v, e_r\rangle e_r.$$

Es gilt natürlich $v = P_U(v) + (v - P_U(v))$ und $P_U(v) \in U$. Also bleibt noch zu zeigen, dass $v - P_U(v) \in U^\perp$. Da $U = \mathrm{Sp}(e_1, \ldots, e_r)$ ist, genügt es laut (5) von Lemma 6.6.2, zu zeigen, dass $(v - P_U(v)) \perp e_j$ für $j = 1, \ldots, r$. Wir berechnen

$$\begin{aligned} \langle v - P_U(v), e_j\rangle &= \langle v, e_j\rangle - \left\langle \sum_{i=1}^r \langle v, e_i\rangle e_i, e_j \right\rangle \\ &= \langle v, e_j\rangle - \sum_{i=1}^r \langle v, e_i\rangle \delta_{ij} = \langle v, e_j\rangle - \langle v, e_j\rangle = 0. \end{aligned} \tag{6.10}$$

□

Wir protokollieren:

Definition 6.6.4. Sei $v \in V$ und $U \subseteq V$ ein endlich-dimensionaler Untervektorraum. Die *orthogonale Projektion* von v auf U ist

$$P_U(v) := \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_r \rangle e_r, \tag{6.11}$$

wobei $(e_1, \ldots, e_r)$ eine orthonormale Basis von U ist.

Die Abbildung P_U ist der Schlüssel zur Lösung mehrerer Approximationsprobleme. Da $U \subseteq V$, können wir die Abbildung $v \mapsto P_U(v)$ als eine Abbildung $P_U : V \to V$ betrachten. Wir beweisen nun einige elementare Eigenschaften dieser Abbildung und halten einige Informationen fest, die wir im obigen Beweis gesehen haben.

Proposition 6.6.5. *Sei U ein endlich-dimensionaler Untervektorraum von V. Die Abbildung P_U ist wohl-definiert (d.h. sie hängt nicht von der Wahl der orthonormalen Basis ab) und linear. Ausserdem gilt für jedes $v \in V$, dass $v - P_U(v) \in U^\perp$.*

Beweis. Satz 6.6.3 besagt, dass $P_U(v)$ die „U-Koordinate“ von v ist in der direkten Summe $U \oplus U^\perp = V$. Daher ist $P_U(v)$ eindeutig definiert. Zur Linearität: Sei $(e_1, \ldots, e_r)$ eine orthonormale Basis von U.

$$P_U(\alpha v + \beta w) = \sum_{i=1}^{r} \langle \alpha v + \beta w, e_i \rangle e_i = \alpha \sum_{i=1}^{r} \langle v, e_i \rangle e_i + \beta \sum_{i=1}^{r} \langle w, e_i \rangle e_i = \alpha P_U(v) + \beta P_U(w).$$

Die letzte Aussage haben wir in (6.10) gezeigt. □

Wir zeigen nun zwei Korollare aus Satz 6.6.3:

Korollar 6.6.6. *Sei U ein endlich-dimensionaler Untervektorraum von V. Dann gilt $(U^\perp)^\perp = U$.*

Beweis. Laut (6) von Lemma 6.6.2 gilt $(U^\perp)^\perp \supseteq U$. Für „ $\subseteq$ “ sei $v \in (U^\perp)^\perp$ und seien $u \in U$ und $w \in U^\perp$ (die eindeutigen Vektoren) mit $v = u + w$. Wir behaupten, dass $w = 0$:

Da $u \in U \subseteq (U^\perp)^\perp$ und $v \in (U^\perp)^\perp$, gilt $w = v - u \in (U^\perp)^\perp$. Aber Laut der Definition von w gilt auch $w \in U^\perp$. Daher gilt $w \in U^\perp \cap (U^\perp)^\perp = \{0\}$. Damit ist $w = 0$ und somit $v \in U$, was den Beweis beendet. □

Korollar 6.6.7. *Falls V endlich-dimensional ist, gilt für jeden Untervektorraum U, $\dim U + \dim U^\perp = \dim V$.*

Beweis. Laut Proposition 2.3.21 ist U auch endlich-dimensional, also können wir Satz 6.6.3 anwenden und erhalten $U \oplus U^\perp = V$. Laut Proposition 2.3.40 ist dies äquivalent zu $\dim U + \dim U^\perp = \dim V$. □

Orthogonale Projektion und Approximationsprobleme

Wie wir in der „Intuitions-Pause“ nach Gleichung (6.6) im Beweis der Cauchy-Schwarz Ungleichung gesagt haben, könnte man $c_0 = \frac{\langle u,v\rangle}{\langle v,v\rangle}$ mit ein bisschen Analysis finden, indem man den Vektor in $\mathrm{Sp}(v)$ sucht, der am nächsten zu u ist. Wir werden in diesem Abschnitt für einen allgemeinen Untervektorraum U zeigen, dass $P_U(v)$ der eindeutige Vektor in U ist, der am nächsten zu v ist. Um dies mit dem Wert von c_0 zu verbinden, berechnen wir nun $P_U(v)$ für $U = \mathrm{Sp}(v)$ und $v = u$ (entschuldigen Sie bitte die inkonsistente Notation).

Beispiel 6.6.8. Seien $0 \neq v \in V$ und $u \in V$. Zeigen Sie, dass $P_{\mathrm{Sp}(v)}(u) = \frac{\langle u,v\rangle}{\langle v,v\rangle}v$.

Lösung: Einen Weg, es zu machen, ist, die Formel (6.11) zu benutzen mit $\left(\frac{v}{\|v\|}\right)$ als Orthonormalbasis von $\mathrm{Sp}(v)$ (Machen Sie es!). Ein anderer Weg ist, zu merken, dass

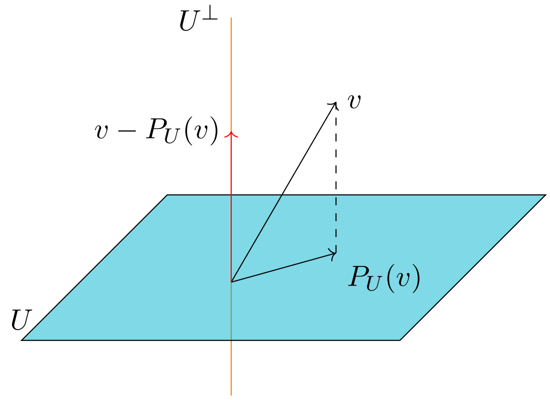
$$u = \underbrace{\frac{\langle u,v\rangle}{\langle v,v\rangle}v}_{\in \mathrm{Sp}(v)} + \underbrace{\left(u - \frac{\langle u,v\rangle}{\langle v,v\rangle}v\right)}_{\in \mathrm{Sp}(v)^\perp}$$

und aus der Eindeutigkeit der Zerlegung $V = \mathrm{Sp}(v) \oplus \mathrm{Sp}(v)^\perp$ folgt $P_{\mathrm{Sp}(v)}(u) = \frac{\langle u,v\rangle}{\langle v,v\rangle}v$.

Satz 6.6.9 (Minimierung)**.** *Sei U ein endlich-dimensionaler Untervektorraum von V. Dann gilt für alle $u \in U$ und $v \in V$, dass*

$$\|v - P_U(v)\| \leq \|v - u\| \,. \tag{6.12}$$

Gleichheit gilt genau dann, wenn $u = P_U(v)$.



Figur 6.5: Die gestrichelte Linie ist die kürzeste Distanz von v zu irgendeinem Punkt in U.

Beweis. Proposition 6.6.5 besagt, dass $(v - P_U(v)) \in U^\perp$. Daher gilt

$$\begin{aligned}\|v - P_U(v)\|^2 &\leq \|v - P_U(v)\|^2 + \|P_U(v) - u\|^2 \\ &\overset{\text{Pythagoras}}{=} \|v - P_U(v) + (P_U(v) - u)\|^2 = \|v - u\|^2 .\end{aligned} \tag{6.13}$$

Wenn wir auf beiden Seiten die Wurzel ziehen, folgt (6.12). Ausserdem gilt

$$\begin{matrix}\text{Gleichheit} \\ \text{in (6.12)}\end{matrix} \iff \begin{matrix}\text{Gleichheit} \\ \text{in (6.13)}\end{matrix} \iff \|P_U(v) - u\|^2 = 0 \iff u = P_U(v).$$

□

Sie werden dieses Theorem in der numerischen Mathematik oft benutzen. Eine Anwendung die ich mag ist die Methode der kleinsten Quadrate (fragen Sie mich in der Nachbesprechung, wenn Sie mehr darüber wissen möchten).

Zum Schluss möchten wir noch anmerken, dass alles, was wir in diesem Abschnitt gezeigt haben, im Allgemeinen nicht gilt, wenn U nicht endlich-dimensional ist.

Beispiel 6.6.10. Sei $V \subseteq \mathbb{F}^\infty$ der Vektorraum aller beschränkten Folgen. Genauer gesagt ist

$$V := \{(a_n)_{n=1}^\infty \in \mathbb{F}^\infty \mid \exists C > 0 \text{ sodass } |a_n| < C \ \forall n \in \mathbb{N}\} .$$

Dieser Vektorraum ist ein Skalarproduktraum mit

$$\langle (a_n)_n, (b_n)_n \rangle = \sum_{n=1}^\infty \frac{a_n \overline{b_n}}{n^2}.$$

Sei $U \subseteq V$ der Untervektorraum all jener Folgen, die schliesslich Null sind (dies ist der Raum W_∞ aus Übung 2.1.15). Dann gilt $U^\perp = \{0\}$. Insbesondere ist $U + U^\perp \neq V$ und $(U^\perp)^\perp \neq U$.

Hier ist noch eine nette Anwendung für die Berechnung des Bilds einer Matrix[7].

Beispiel 6.6.11. Sei $A \in M_{m\times n}(\mathbb{F})$. Wir haben in Kapitel 2 gesehen, dass man mithilfe der Gauss-Elimination der Zeilen einfach eine Basis von $\mathrm{ZR}(A)$ finden kann und dann haben wir sogar einen Trick gesehen, wie wir von einer Zeilenstufenform von A eine Basis von $\mathrm{Im}(A) = \mathrm{SR}(A)$ finden können (siehe Korollar 2.3.34). Wir zeigen jetzt einen anderen Trick:

Behauptung: Seien $x_1, \dots, x_k \in \mathbb{F}^n_{\text{Zeil}}$ eine Basis von $\mathrm{ZR}(A)$ (die man leicht aus den Zeilen von A findet). Dann bilden $Ax_1^T, \dots, Ax_k^T \in \mathbb{F}^n_{\text{Spal}}$ eine Basis von $\mathrm{Im}(A) = \mathrm{SR}(A)$.

[7] Ich habe diesen Trick von Josua Hächl gelernt.

Beweis. Da $\dim \mathrm{SR}(A) = \dim \mathrm{ZR}(A) = k$, genügt es zu zeigen, dass $Ax_1^T, \ldots, Ax_k^T$ linear unabhängig sind. Seien also $\lambda_1, \ldots, \lambda_k \in \mathbb{F}$ mit $\sum_{i=1}^k \lambda_i A x_i^T = 0$. Es folgt

$$A\left(\sum_{i=1}^k \lambda_i x_i^T\right) = \sum_{i=1}^k \lambda_i A x_i^T = 0$$

und daher ist $\sum_{i=1}^k \lambda_i x_i^T \in \ker A$. Betrachten wir die Matrixmultiplikation von A mit einem Vektor $v \in \mathbb{F}^n$. Wenn wir die Zeilen von A mit $A_1, \ldots, A_m$ bezeichnen, dann gilt

$$Av = \begin{pmatrix} — & A_1 & — \\ & \vdots & \\ — & A_m & — \end{pmatrix} \begin{pmatrix} v \end{pmatrix} = \begin{pmatrix} \langle A_1, v\rangle \\ \vdots \\ \langle A_m, v\rangle \end{pmatrix}.$$

Daraus folgt, dass

$$v \in \ker A \iff v \perp \{A_1, \ldots, A_m\} \overset{6.6.2\ (5)}{\iff} v \perp \underbrace{\mathrm{Sp}(A_1, \ldots, A_m)}_{=\ \mathrm{Sp}(x_1^T, \ldots, x_k^T)\ =:U} \iff v \in U^\perp.$$

Es gilt $\sum_{i=1}^k \lambda_i x_i^T \in U$, also $\sum_{i=1}^k \lambda_i x_i^T \in U \cap U^\perp = \{0\}$ und somit folgt $\sum_{i=1}^k \lambda_i x_i^T = 0$. Da $x_1, \ldots, x_k$ linear unabhängig sind, gilt das auch für $x_1^T, \ldots, x_k^T$ und somit folgt $\lambda_1 = \cdots = \lambda_k = 0$, was die lineare Unabhängigkeit von $Ax_1^T, \ldots, Ax_k^T$ zeigt. $\square$

Changelog: Kapitel 6

- 11.03: In Übung 6.2.13 wurde hinzugefügt, dass die Matrix symmetrisch sein muss.
- 13.03: In (6.6) wurde $\|u - cv\|$ zu $\|u - cv\|^2$ korrigiert und in (6.7) wurde $|u, v|^2$ zu $| \langle u, v \rangle |^2$ korrigiert.
- 14.03: In Übung 6.1.10 (2) wurden x, y zu v, w korrigiert und im Beweis von Korollar 6.3.7 wurden x, y zu u, v korrigiert.
- 14.03: In der Definition von Positiv-Definitheit für hermitesche Matrizen auf Seite 243 wurde $vA\overline{v} > 0$ zu $v^T A\overline{v} > 0$ korrigiert.
- 15.03: In Definition 6.2.11 wurde $0 \neq v$ hinzugefügt.
- 15.03: Beispiel 6.2.6 wurde hinzugefügt.
- 15.03: Am Ende des indirekten Beweises von (b) auf Seite 244 wurde $\mathbb{F} = \mathbb{C}$ zu $\mathbb{F} = \mathbb{R}$ korrigiert.
- 16.03: Die Notation $\mathbb{R} \setminus \mathbb{C}$ wurde zu $\mathbb{F}$ geändert.
- 16.03: Vor dem Beweis von Satz von Satz 6.4.7 wurde noch ein drittes Bild hinzugefügt.
- 16.03: In Definition 6.1.11 wurde bei der Definition eines Orthogonalsystems hinzugefügt, dass der Nullvektor nicht enthalten sein darf.
- 16.03: Das Wort triagonalisierbar wurde zu trigonalisierbar korrigiert (alle anderen Formen des Worts wurden natürlich auch angepasst).
- 17.03: Im Gram-Schmidt Algorithmus (Satz 6.4.7) wurde $\overline{w}_i$ zu e_i geändert.
- 17.03: In Beispiel 6.4.11 in der Berechnung von w_2 wurde $\frac{\langle v_1,w_1\rangle}{\langle w_1,w_1\rangle}w_1$ zu $\frac{\langle v_2,w_1\rangle}{\langle w_1,w_1\rangle}w_1$ korrigiert.
- 17.03: In Figur 6.3 wurden u, v zu v, w geändert.
- 17.03: In der Definition von Positiv-Definitheit für hermitesche Matrizen auf Seite 243 wurde $0 \neq v$ hinzugefügt.
- 18.03: In der Aussage und im Beweis von Proposition 6.4.6 wurden die Indizes korrigiert.
- 19.03: Am Anfang des Beweises von 6.4.25 und 6.4.26 wurde ein Satz hinzugefügt.
- 19.03: In Satz 6.4.19 wurde orthogonale Matrix zu unitäre Matrix korrigiert.

- 19.03: In Definition 6.4.13 wurde die Definition von $\mathrm{O}(n)$ und $\mathrm{U}(n)$ hinzugefügt.
- 19.03: Beispiel 6.4.21 wurde hinzugefügt.
- 21.03: In Definition 6.1.5 wurde $v, w, v_1, v_2 \in v$ zu $v, w, v_1, v_2 \in V$ korrigiert.
- 22.03: Der Text am Anfang von Abschnitt 6.4.4 wurde ein wenig geändert.
- 22.03: Im Satz 6.5.5 wurde eine zweite Fussnote hinzugefügt.
- 26.03: Nach dem zweiten Beweis von Satz 6.5.5 wurde ein Rezept hinzugefügt.
- 26.03: Beispiel 6.6.11 wurde hinzugefügt.
- 26.03: Im Beweis von Satz 6.3.5 wurde an zwei Stellen $\|x - cy\|$ zu $\|u - cv\|$ korrigiert.
- 26.03: Im Beweis von Lemma 6.6.2 (5) wurde $\sum_i \alpha_i \langle v, s_i \rangle$ zu $\sum_i \overline{\alpha_i} \langle v, s_i \rangle$ korrigiert.
- 28.03: Im Beweis von Lemma 6.6.2 (1) wurde $\beta \langle v, s \rangle$ zu $\beta \langle w, s \rangle$ korrigiert.
- 28.03: In Beispiel 6.6.8 wurde $P_{\mathrm{Sp}(v)}(u) = \frac{\langle u,v \rangle}{\langle v,v \rangle} u$ zu $P_{\mathrm{Sp}(v)}(u) = \frac{\langle u,v \rangle}{\langle v,v \rangle} v$ korrigiert.
- 30.03: In Beispiel 6.4.21 wurde eine Fussnote hinzugefügt.
- 13.04: Im Beweis von Beispiel 6.6.11 wurde $x \in U^\perp$ zu $v \in U^\perp$ korrigiert.

Kapitel 7

Lineare Abbildungen auf Skalarprodukträumen

A In Kapitel 2 haben wir Vektorräume eingeführt und in Kapitel 3 haben wir dann Abbildungen, welche die Vektorraumstruktur erhalten, untersucht. Nachdem wir im vorherigen Kapitel Skalarprodukträume eingeführt haben, werden wir nun dasselbe machen, jedoch mit einem „Twist“:

1. Zuerst wollen wir allgemeine lineare Abbildungen zwischen zwei Skalarprodukträumen betrachten und sehen, wie die zusätzliche Struktur uns hilft, diese zu studieren. Wir erhalten beispielsweise Charakterisierungen aller Endomorphismen eines Skalarproduktraums, welche eine orthonormale Basis von Eigenvektoren besitzen (siehe die Theoreme 7.0.1 und 7.0.3 für den Fall von $\mathbb{F}^n$ mit dem Standard-Skalarprodukt).

2. Danach untersuchen wir lineare Abbildungen, welche die Skalarproduktstruktur erhalten. Insbesondere beantworten wir die folgende grundlegende Frage: Welche Matrizen $A \in M_{n\times n}(\mathbb{F})$ haben die Eigenschaft, dass $\langle Av, Aw\rangle = \langle v, w\rangle$ für alle $v, w \in \mathbb{F}^n$ gilt, wobei $\mathbb{F}^n$ mit dem Standard-Skalarprodukt versehen ist?

In beiden Teilen werden wir uns besonders auf Endomorphismen fokussieren. Unser erstes Hauptresultat wird der Spektralsatz über $\mathbb{C}$ und über $\mathbb{R}$ sein. Bis wir dorthin gelangen, ist der Leser vielleicht schon etwas verwirrt von den vielen allgemeinen Definitionen, also sagen wir schon einmal, was die Essenz dieser Theoreme ist anhand des folgenden wichtigen Spezialfalls:

$\mathbb{C}$ **Theorem 7.0.1** (Spektralsatz über $\mathbb{C}$)**.** *Betrachten wir $\mathbb{C}^n$ mit dem üblichen inneren Produkt und sei $A \in M_{n\times n}(\mathbb{C})$. Die folgenden Aussagen sind äquivalent:*

(1) Die Matrix A kommutiert mit ihrer Adjungierten, d.h. $\overline{A}^T A = A\overline{A}^T$ (dies gilt zum Beispiel für hermitesche Matrizen, also wenn $\overline{A}^T = A$).

(2) Es existiert eine Orthonormalbasis von $\mathbb{C}^n$ bestehend aus Eigenvektoren von A.

(3) Es existiert eine Orthonormalbasis $\mathcal{B}$ von $\mathbb{C}^n$, sodass $[m_A]_{\mathcal{B}}$ diagonal ist.

(4) Es existiert eine unitäre Matrix $P \in \mathrm{GL}_n(\mathbb{C})$, sodass $P^{-1}AP = \overline{P}^T AP$ diagonal ist.

Übung 7.0.2. Die Äquivalenz von (2), (3) und (4) sollte Ihnen zu diesem Zeitpunkt klar sein (zumindest, wenn Sie sich an die Definition einer unitären Matrix erinnern). Überprüfen Sie, dass dies wirklich der Fall ist und zeigen Sie dass diese Aussagen (1) implizieren. Die Essenz des Theorems ist (1) $\Longrightarrow$ (2).

$\mathbb{R}$ **Theorem 7.0.3** (Spektralsatz über $\mathbb{R}$)**.** *Betrachten wir $\mathbb{R}^n$ mit dem üblichen inneren Produkt und sei $A \in M_{n\times n}(\mathbb{R})$. Die folgenden Aussagen sind äquivalent:*

(1) Die Matrix A ist symmetrisch, d.h. $A^T = A$.

(2) Es existiert eine Orthonormalbasis von $\mathbb{R}^n$ bestehend aus Eigenvektoren von A.

(3) Es existiert eine Orthonormalbasis $\mathcal{B}$ von $\mathbb{R}^n$, sodass $[m_A]_{\mathcal{B}}$ diagonal ist.

(4) Es existiert eine orthogonale Matrix $P \in \mathrm{GL}_n(\mathbb{R})$, sodass $P^{-1}AP = P^T AP$ diagonal ist.

Die gleiche Übung gilt auch hier:

Übung 7.0.4. Die Äquivalenz von (2), (3) und (4) sollte Ihnen zu diesem Zeitpunkt klar sein (zumindest, wenn Sie sich an die Definition einer orthogonalen Matrix erinnern). Überprüfen Sie, dass dies wirklich der Fall ist und zeigen Sie dass diese Aussagen (1) implizieren. Die Essenz des Theorems ist (1) $\Longrightarrow$ (2).

Bemerkung 7.0.5. Theorem 7.0.3 impliziert insbesondere, dass jede symmetrische Matrix diagonalisierbar ist. Nur schon dies ist erstaunlich.

Diese Theoreme sind die zentralen Resultate über das Zusammenspiel von allgemeinen linearen Abbildungen und der Skalarproduktstruktur.

$\mathbb{F}$ 7.1 Die Adjungierte Abbildung

Alle Vektorräume in diesem Abschnitt sind endlich-dimensionale Skalarprodukträume über $\mathbb{F}$, wobei wie immer $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ ist. Man muss Abschnitt 6.5 über den Dualraum gut verstehen, um diesem Abschnitt zu folgen.

Definition 7.1.1. Seien $(V, \langle\, , \rangle_V)$ und $(W, \langle\, , \rangle_W)$ Skalarprodukträume über[1] $\mathbb{F}$ und sei $T : V \to W$ eine lineare Abbildung. Die *adjungierte Abbildung* $T^* : W \to V$ ist durch die Bedingung

$$\forall v \in V \ \forall w \in W \quad \langle Tv, w\rangle_W = \langle v, T^* w\rangle_V \tag{7.1}$$

definiert.

Zwei Dinge, die sich die Leser bei dieser Definition fragen sollten, sind: Wieso ist T^* überhaupt wohl-definiert? Und wie kann man T^* für ein gegebenes $w \in W$ berechnen? Zuerst zur Wohldefiniertheit: Betrachten wir die linke Seite von (7.1) als eine Funktion

$$\begin{aligned} \varphi_w : V &\to \mathbb{F} \\ v &\mapsto \langle Tv, w\rangle_W . \end{aligned}$$

Die Linearität von T und von $\langle\, , \rangle_W$ (in der ersten Variable) implizieren, dass φ_w eine Linearform ist: Für $v_1, v_2 \in V$ und $\alpha_1, \alpha_2 \in \mathbb{F}$ gilt

$$\begin{aligned} \varphi_w(\alpha_1 v_1 + \alpha_2 v_2) &= \langle T(\alpha_1 v_1 + \alpha_2 v_2), w\rangle_W \\ &= \langle \alpha_1 T v_1 + \alpha_2 T v_2, w\rangle_W \\ &= \alpha_1 \langle T v_1, w\rangle_W + \alpha_2 \langle T v_2, w\rangle_W \\ &= \alpha_1 \varphi_w(v_1) + \alpha_2 \varphi_w(v_2). \end{aligned}$$

Also ist φ_w ein Element des Dualraums von V, das heisst $\varphi_w \in V^*$. Satz 6.5.5 besagt, dass es ein eindeutiges[2] $v_0 = v_0(\varphi_w)$ in V gibt, sodass für alle $v \in V$

$$\varphi_w(v) = \langle v, v_0\rangle_V , \quad \text{bzw. } \langle Tv, w\rangle_W = \langle v, v_0\rangle_V , \tag{7.2}$$

gilt. Wir definieren $T^* w$ als diesen eindeutigen Vektor $v_0 \in V$. Damit ist (7.2) äquivalent zu der Bedingung (7.1). Nachdem wir nun gezeigt haben, dass Definition 7.1.1 wohldefiniert ist, wollen wir Sie in einem Beispiel einer Matrix berechnen:

$\mathbb{C}$ **Übung 7.1.2.** Sei $A = \begin{pmatrix} 2i & 1+i & 7 \\ 1 & 2 & 3 \end{pmatrix} \in M_{2\times 3}(\mathbb{C})$. Berechne $(m_A)^*$.

Lösung: Zuerst hilft es, zu überprüfen, was überhaupt die Definitions- und Zielmenge von $(m_A)^*$ ist: Da $m_A : \mathbb{C}^3 \to \mathbb{C}^2$, ist $(m_A)^* : \mathbb{C}^2 \to \mathbb{C}^3$. Laut der charakterisierenden Eigenschaft der adjungierten Abbildung gilt für $(x_1, x_2, x_3)^T \in \mathbb{C}^3$ und $(y_1, y_2)^T \in \mathbb{C}^2$,

[1] d.h. beide über $\mathbb{R}$ oder beide über $\mathbb{C}$.

[2] Wie in Satz 6.5.5 betont die Schreibweise $v_0 = v_0(\varphi_w)$, dass v_0 von φ_w (und daher von w) abhängt.

dass

$$\left\langle m_A \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right\rangle = \left\langle \begin{pmatrix} 2ix_1 + (1+i)x_2 + 7x_3 \\ x_1 + 2x_2 + 3x_3 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right\rangle$$
$$\underset{x_i \text{ ausklammern}}{\overset{\text{Def. \&}}{=}} (2i\overline{y_1} + \overline{y_2})x_1 + ((1+i)\overline{y_1} + 2\overline{y_2})x_2 + (7\overline{y_1} + 3\overline{y_2})x_3$$
$$\overset{\overline{(\overline{x})}=x}{=} \left\langle \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \begin{pmatrix} \overline{2i} & 1 \\ \overline{1+i} & 2 \\ 7 & 3 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right\rangle .$$

$\mathbb{F}$ Es folgt $(m_A)^* = \overline{A}^T$. Tatsächlich ist es sogar einfacher, Eigenschaften der Matrixmultiplikation zu verwenden und dies direkt für allgemeine $A \in M_{m\times n}(\mathbb{F})$ zu zeigen: Für alle $x \in \mathbb{F}^n$, $y \in \mathbb{F}^m$ gilt

$$\langle Ax, y\rangle = (Ax)^T\overline{y} = x^T A^T \overline{y} = x^T \overline{\left(\overline{A}^T y\right)} = \left\langle x, \overline{A}^T y \right\rangle ,$$

wobei $\langle\, ,\rangle$ das Standard-Skalarprodukt auf $\mathbb{F}^n$ ist.

Lemma 7.1.3 (Grundeigenschaften von T^*)**.** *Seien V, W, U Skalarprodukträume über $\mathbb{F}$ und seien $S, T : V \to W$ und $R : W \to U$ linear. Dann gilt:*

(1) T^* ist linear.

(2) $(S + T)^* = S^* + T^*$.

(3) $(\lambda T)^* = \overline{\lambda} T^* \ \forall \lambda \in \mathbb{F}$.

(4) $(T^*)^* = T$.

(5) $(\mathrm{Id}_V)^* = \mathrm{Id}_V$.

(6) $(R \circ T)^* = T^* \circ R^*$.

Beweis. (1): Für $w_1, w_2 \in W$ ist $T^*(w_1 + w_2)$ durch die charakterisierende Eigenschaft (7.1) definiert. Um $T^*(w_1 + w_2) = T^*w_1 + T^*w_2$ zu zeigen, genügt es also zu beweisen, dass $T^*w_1 + T^*w_2$ auch diese Eigenschaft erfüllt: Für alle $v \in V$ gilt

$$\begin{aligned} \langle v, T^*w_1 + T^*w_2\rangle_V = \langle v, T^*w_1\rangle_V + \langle v, T^*w_2\rangle_V &= \langle Tv, w_1\rangle_W + \langle Tv, w_2\rangle_W \\ &= \langle Tv, w_1 + w_2\rangle_W , \end{aligned}$$

also erfüllt $T^*w_1 + T^*w_2$ die charakterisierende Eigenschaft (7.1) von $T^*(w_1 + w_2)$ und somit gilt $T^*(w_1 + w_2) = T^*w_1 + T^*w_2$.

Auf die gleiche Art könnte man auch $T^*(\alpha w) = \alpha T^* w$ zeigen für $\alpha \in \mathbb{F}$ (versuchen Sie es). Um zu zeigen, wie man einfache Aussagen auf verschiedene Arten beweisen kann, geben wir jedoch einen anderen Beweis dafür: Seien $w \in W$, $\alpha \in \mathbb{F}$. Für alle $v \in V$ gilt

$$\langle v, T^*(\alpha w)\rangle_V = \langle Tv, \alpha w\rangle_W = \overline{\alpha}\,\langle Tv, w\rangle_W = \overline{\alpha}\,\langle v, T^* w\rangle_V = \langle v, \alpha T^* w\rangle_V .$$

Aus Lemma 6.1.6 (3) folgt $T^*(\alpha w) = \alpha T^* w$. Als Übung im Beweise schreiben können Sie auch den ersten Teil des Beweises wie den zweiten Teil schreiben.

(2): Auch hier ist das Argument ähnlich. Um die Notation zu erleichtern, verzichten wir auf die Indizes V, W bei den Skalarprodukten. Es sollte den Lesern jeweils klar sein, wo wir welches Skalarprodukt benutzen (schauen Sie, aus welchem Vektorraum die Elemente innerhalb des Skalarprodukts stammen). Sei $w \in W$ beliebig. Für alle $v \in V$ gilt

$$\begin{aligned}\langle v, (S+T)^* w\rangle &= \langle (S+T)v, w\rangle = \langle Sv + Tv, w\rangle = \langle Sv, w\rangle + \langle Tv, w\rangle \\ &= \langle v, S^* w\rangle + \langle v, T^* w\rangle = \langle v, (S^* + T^*)w\rangle .\end{aligned} \tag{7.3}$$

Es folgt wie oben, dass $(S+T)^* w = (S^* + T^*)w$ für alle $w \in W$, also $(S+T)^* = S^* + T^*$.

(3): Dies ist ähnlich zum Beweis von (2). Sei $w \in W$ beliebig. Für $v \in V$ gilt

$$\langle v, (\lambda T)^* w\rangle = \langle \lambda T v, w\rangle = \lambda\,\langle Tv, w\rangle = \lambda\,\langle v, T^* w\rangle = \left\langle v, \overline{\lambda} T^* w\right\rangle = \left\langle v, (\overline{\lambda} T^*) w\right\rangle$$

und somit ist $(\lambda T)^* w = (\overline{\lambda} T^*) w$ für alle $w \in W$, also $(\lambda T)^* = \overline{\lambda} T^*$ als Abbildungen.

(4): Diesmal sei $v \in V$ beliebig und wir betrachten für jedes $w \in W$

$$\langle w, (T^*)^* v\rangle = \langle T^* w, v\rangle = \overline{\langle v, T^* w\rangle} = \overline{\langle Tv, w\rangle} = \langle w, Tv\rangle .$$

Daraus folgt $(T^*)^* v = Tv$ für alle $v \in V$ und daher $(T^*)^* = T$.

(5): Für alle $v_1, v_2 \in V$ gilt

$$\langle v_1, (\mathrm{Id}_V)^* v_2\rangle = \langle \mathrm{Id}_V\, v_1, v_2\rangle = \langle v_1, v_2\rangle .$$

Also ist $(\mathrm{Id}_V)^* v_2 = v_2$ für beliebige $v_2 \in V$ und daher ist $(\mathrm{Id}_V)^* = \mathrm{Id}_V$.

(6): Da wir nun drei verschiedene Skalarprodukte haben, schreiben wir in diesem Teil die Indizes V, W, U wieder. Für alle $v \in V$ und $u \in U$ gilt

$$\begin{aligned}\langle v, (R \circ T)^* u\rangle_V &= \langle (R \circ T)v, u\rangle_U = \langle R(Tv), u\rangle_U = \langle Tv, R^* u\rangle_W \\ &= \langle v, T^*(R^* u)\rangle_V = \langle v, (T^* \circ R^*)u\rangle_V .\end{aligned}$$

Es folgt $(R \circ T)^* u = T^* R^* u$ für alle $u \in U$ und somit $(R \circ T)^* = T^* \circ R^*$. □

Bemerkung 7.1.4. Um Tinte zu sparen, könnte man den letzten Satz im Beweis von jedem Teil von Lemma 7.1.3 streichen. Zum Beispiel folgt aus (7.3) direkt, dass $(S+T)^* = S^* + T^*$:

(7.3) zeigt, dass $S^* + T^*$ die charakterisierende Eigenschaft (7.1) der Adjungierten von $T+S$ erfüllt und daher folgt $S^* + T^* = (S+T)^*$. Analog könnte man auch in allen anderen Teilen des Beweises argumentieren.

Lemma 7.1.5. *Für $T : V \to W$ wie in Definition 7.1.1 gilt:*

(1) $\ker(T^*) = (\operatorname{Im} T)^\perp$.

(2) $\operatorname{Im}(T^*) = (\ker T)^\perp$.

(3) $\ker(T) = (\operatorname{Im} T^*)^\perp$.

(4) $\operatorname{Im}(T) = (\ker T^*)^\perp$.

Beweis. Für $w \in W$ gilt

$$\begin{aligned} w \in \ker T^* &\iff T^* w = 0 \iff \forall v \in V \;\; \langle v, T^* w\rangle = 0 \iff \forall v \in V \;\; \langle Tv, w\rangle = 0 \\ &\iff w \perp \{Tv \mid v \in V\} \iff w \in (\operatorname{Im} T)^\perp, \end{aligned}$$

was (1) zeigt. Wenn wir in (1) auf beiden Seiten das orthogonale Komplement bilden, erhalten wir

$$\ker(T^*)^\perp = ((\operatorname{Im} T)^\perp)^\perp = \operatorname{Im} T,$$

wobei die letzte Gleichheit aus Korollar 6.6.6 folgt (da $\dim \operatorname{Im} T \leq \dim W < \infty$ per Annahme). Dies zeigt (4). Wenn wir (4) auf T^* anwenden, folgt (2) mittels $(T^*)^* = T$. Ähnlich dazu folgt (3), wenn wir (1) auf T^* anwenden. □

Wir fragen uns, wie die Darstellungsmatrizen von T und T^* miteinander zusammenhängen. Beispiel 7.1.2 und die obigen Resultate (z.B. $(T^*)^* = T$, $(T+S)^* = T^* + S^*$, $(\lambda T)^* = \overline{\lambda} T^*$ usw.) geben uns schon einen Hinweis auf das nächste Resultat:

Proposition 7.1.6. *Sei $T : V \to W$ eine lineare Abbildung zwischen Skalarprodukträumen über $\mathbb{F}$. Sei $\mathcal{B} = (e_1, \ldots, e_n)$ eine orthonormale Basis von V und $\mathcal{C} = (f_1, \ldots, f_m)$ eine orthonormale Basis von W. Dann gilt*

$$[T^*]^{\mathcal{C}}_{\mathcal{B}} = \overline{[T]^{\mathcal{B}}_{\mathcal{C}}}^T.$$

Bemerkung 7.1.7. Eine übliche Notation für $\overline{A}^T$ ist A^*. Damit sieht die obige Aussage noch besser aus:

$$[T^*]^{\mathcal{C}}_{\mathcal{B}} = ([T]^{\mathcal{B}}_{\mathcal{C}})^*.$$

Beweis. Seien $A = (a_{ij})_{i,j} = [T]^{\mathcal{B}}_{\mathcal{C}}$ und $B = (b_{ij})_{i,j} = [T^*]^{\mathcal{C}}_{\mathcal{B}}$. Wir möchten zeigen, dass $b_{ij} = \overline{a_{ji}}$. Dank Proposition 6.4.6 wissen wir, dass

$$b_{ij} = \langle T^* f_j, e_i \rangle = \overline{\langle e_i, T^* f_j \rangle} = \overline{\langle T e_i, f_j \rangle} = \overline{a_{ji}}. \qquad \square$$

Später werden wir uns vor allem für den Fall von Endomorphismen interessieren. Wir halten deshalb schon einmal fest: Für $T \in \mathrm{End}(V)$ und $\mathcal{B}$ eine Orthonormalbasis von V gilt

$$[T^*]_{\mathcal{B}} = \overline{[T]_{\mathcal{B}}}^{T}. \tag{7.4}$$

Das nächste Korollar haben wir schon in der Lösung von Übung 7.1.2 gesehen. Wir können nun einen alternativen (abstrakteren) Beweis dafür geben:

Korollar 7.1.8. *Für $V = \mathbb{F}^n$ und $W = \mathbb{F}^m$, jeweils mit dem Standard-Skalarprodukt, und $A \in M_{m\times n}(\mathbb{F})$ gilt*

$$(m_A)^* = m_{\overline{A}^T}$$

(oder mit der Notation $A^* = \overline{A}^T$, $(m_A)^* = m_{A^*}$).

Beweis. Für $\mathbb{F}^n$ mit dem Standard-Skalarprodukt ist die Standard-Basis $\mathcal{E}_n$ orthonormal. Das Korollar folgt somit aus Proposition 7.1.6 mithilfe von Beispiel 3.3.9. $\square$

Im nächsten Korollar sind einige Eigenschaften von adjungierten Matrizen aufgelistet. Diese folgen aus den dazugehörigen Eigenschaften aus Lemma 7.1.3 und Proposition 7.1.6 (auch wenn es für die ersten 5 Eigenschaften einfacher ist, diese direkt zu zeigen).

Korollar 7.1.9. *Für $A, B \in M_{n\times m}(\mathbb{F})$, $C \in M_{m\times p}(\mathbb{F})$ gilt:*

(1) $\overline{A}^T \in M_{m\times n}(\mathbb{F})$.

(2) $\overline{(A+B)}^T = \overline{A}^T + \overline{B}^T$.

(3) $\overline{(\lambda A)}^T = \overline{\lambda} \cdot \overline{A}^T$.

(4) $\overline{\left(\overline{A}^T\right)}^T = A$.

(5) $\overline{I_n}^T = I_n$.

(6) $\overline{(AC)}^T = \overline{C}^T \overline{A}^T$.

7.2 Spektralsätze

7.2.1 Normale Endomorphismen und Spektralsatz über $\mathbb{C}$

Wir möchten charakterisieren, wann für einen Endomorphismus eine Orthonormalbasis aus Eigenvektoren existiert. Falls eine solche existiert, nennen wir den Endomorphismus *orthogonal diagonalisierbar*. Es stellt sich heraus, dass die Antwort davon abhängt, ob wir über $\mathbb{R}$ oder über $\mathbb{C}$ arbeiten.

Lemma 7.2.1. *Falls $T \in \operatorname{End}(V)$ orthogonal diagonalisierbar ist, gilt $TT^* = T^*T$.*

Beweis. Sei $\mathcal{B}$ eine Orthonormalbasis, sodass $[T]_{\mathcal{B}}$ diagonal ist. Daher ist laut Proposition 7.1.6 auch $[T^*]_{\mathcal{B}} = \overline{[T]_{\mathcal{B}}}^T$ diagonal. Da diagonale Matrizen immer miteinander kommutieren, gilt

$$[T^*T]_{\mathcal{B}} = [T^*]_{\mathcal{B}}[T]_{\mathcal{B}} = [T]_{\mathcal{B}}[T^*]_{\mathcal{B}} = [TT^*]_{\mathcal{B}}.$$

Es folgt also (mittels Satz 3.5.3), dass $T^*T = TT^*$. □

Definition 7.2.2. Ein Endomorphismus T heisst *normal*, falls $T^*T = TT^*$. Eine Matrix in $M_{n\times n}(\mathbb{F})$ heisst *normal*, falls $\overline{A}^T A = A\overline{A}^T$, oder äquivalent dazu, falls m_A normal ist.

$\mathbb{C}$ **Satz 7.2.3** (Spektralsatz über $\mathbb{C}$). *Ein Endomorphismus T eines endlich-dimensionalen unitären Vektorraums V ist genau dann orthogonal diagonalisierbar, wenn T normal ist.*

$\mathbb{R} \leftrightarrow \mathbb{C}$ Die Richtung $\Longrightarrow$ ist genau Lemma 7.2.1. Bevor wir diesen Satz beweisen, zeigen wir, dass er über $\mathbb{R}$ nicht gilt:

$\mathbb{R}$ **Beispiel 7.2.4.** Betrachten wir eine Rotation in $\mathbb{R}^2$, zum Beispiel $A = \begin{pmatrix} & 1 \\ -1 & \end{pmatrix}$ (erinnern Sie sich an die allgemeine Formel für Rotationsmatrizen aus Übung 5.3.29). Die Matrix A ist normal, wie sich mit einer kurzen Rechnung überprüfen lässt. Aber A hat keine Eigenwerte über $\mathbb{R}$, also erst recht keine Orthonormalbasis aus Eigenvektoren!

$\mathbb{R} \leftrightarrow \mathbb{C}$ Lemma 7.2.1 war hingegen gültig für $\mathbb{R}$ und $\mathbb{C}$.

Wir sammeln grundlegende Eigenschaften von normalen Endomorphismen. Diese sehen zwar „unschuldig“ aus, sie sind jedoch das Herz der Beweise der Spektralsätze.

$\mathbb{F}$ **Lemma 7.2.5.** *Für einen normalen Endomorphismus $T \in \operatorname{End}(V)$ gilt:*

(1) Für alle $v \in V$ ist $\|Tv\| = \|T^*v\|$.

(2) $(T - \lambda \operatorname{Id}_V)$ *ist auch normal für jedes* $\lambda \in \mathbb{F}$.

(3) Falls v *ein Eigenvektor von* T *zum Eigenwert* λ *ist, so ist* v *auch ein Eigenvektor von* T^* *zum Eigenwert* $\overline{\lambda}$.

(4) Eigenvektoren von T *zu verschiedenen Eigenwerten sind orthogonal zueinander.*

Beweis. (1): Es gilt

$$\begin{aligned} \|Tv\|^2 &= \langle Tv, Tv\rangle = \langle v, T^*Tv\rangle \overset{\text{normal}}{=} \langle v, TT^*v\rangle = \overline{\langle TT^*v, v\rangle} = \overline{\langle T^*v, T^*v\rangle} \\ &= \langle T^*v, T^*v\rangle = \|T^*v\|^2 . \end{aligned} \tag{7.5}$$

(2): Man öffnet die Klammern in

$$\begin{aligned} (T - \lambda \operatorname{Id}_V)(T - \lambda \operatorname{Id}_V)^* &= (T - \lambda \operatorname{Id}_V)(T^* - \overline{\lambda} \operatorname{Id}_V) \\ \text{und } (T - \lambda \operatorname{Id}_V)^*(T - \lambda \operatorname{Id}_V) &= (T^* - \overline{\lambda} \operatorname{Id}_V)(T - \lambda \operatorname{Id}_V) \end{aligned}$$

und sieht mit $T^*T = TT^*$, dass diese Ausdrücke gleich sind.

(3): Sei $v \in V$ ein Eigenvektor von T zum Eigenwert λ. Dann gilt $(T - \lambda \operatorname{Id}_V)v = 0$. Laut (2) ist $T - \lambda \operatorname{Id}_V$ normal und laut (1) gilt

$$0 = \|(T - \lambda \operatorname{Id}_V)v\| \overset{(1)}{=} \|(T - \lambda \operatorname{Id}_V)^*v\| = \left\|(T^* - \overline{\lambda} \operatorname{Id}_V)v\right\|$$

und somit $(T^* - \overline{\lambda} \operatorname{Id}_V)v = 0$, was äquivalent zur Aussage (3) ist.

(4): Seien $v_1 \in V$ ein Eigenvektor zum Eigenwert λ_1 und $v_2 \in V$ ein Eigenvektor zum Eigenwert λ_2. Betrachten wir

$$\lambda_1 \langle v_1, v_2\rangle = \langle \lambda_1 v_1, v_2\rangle = \langle Tv_1, v_2\rangle = \langle v_1, T^*v_2\rangle \overset{(3)}{=} \left\langle v_1, \overline{\lambda_2} v_2\right\rangle = \lambda_2 \langle v_1, v_2\rangle .$$

Falls $\lambda_1 \neq \lambda_2$ ist, folgt $\langle v_1, v_2\rangle = 0$. □

Übung 7.2.6. Um im Beweis von (1) Tinte zu sparen, ist es hilfreich zu merken, dass für einen Endomorphismus T gilt $\langle v_1, Tv_2\rangle = \langle T^*v_1, v_2\rangle$. Beweisen Sie dies und machen Sie (7.5) sowie den Beweis von Lemma 7.1.3 (4) kürzer.

$\mathbb{C}$ Wir können jetzt den Spektralsatz über $\mathbb{C}$ beweisen (von Lemma 7.2.5 brauchen wir nur Teil (1)).

Beweis von Satz 7.2.3. Sei T ein normaler Endomorphismus. Laut Korollar 5.4.7 ist T trigonalisierbar und somit existiert laut dem Satz von Schur (Satz 6.4.22) eine orthonormale Basis $\mathcal{B} = (v_1, \dots, v_n)$ von V, sodass $[T]_{\mathcal{B}}$ eine obere Dreiecksmatrix ist. Wir behaupten, dass $[T]_{\mathcal{B}}$ eigentlich schon diagonal sein muss. Sei $[T]_{\mathcal{B}} = (a_{ij})_{1\le i,j\le n}$. Da $\mathcal{B}$

orthonormal ist, gilt laut der Definition der Darstellungsmatrix und Lemma 6.4.12

$$\|Tv_1\|^2 = \sum_{k=1}^{n} |a_{k1}|^2 \underset{\text{Dreiecksm.}}{\overset{\text{obere}}{=}} |a_{11}|^2.$$

Wegen $[T^*]_{\mathcal{B}} = \overline{[T]_{\mathcal{B}}}^T$ gilt

$$\|T^*v_1\|^2 = \sum_{k=1}^{n} |\overline{a_{1k}}|^2 = |a_{11}|^2 + |a_{12}|^2 + \cdots + |a_{1n}|^2.$$

Da T normal ist, gilt $\|Tv_1\|^2 = \|T^*v_1\|^2$ und somit ist $a_{12} = a_{13} = \cdots = a_{1n} = 0$. Somit hat $[T]_{\mathcal{B}}$ die Form

$$\left(\begin{array}{c|ccc} a_{11} & 0 & \cdots & 0 \\ \hline 0 & * & \cdots & * \\ \vdots & & \ddots & \vdots \\ 0 & & & * \end{array}\right)$$

Jetzt könnte man das Wort Induktion sagen, oder einfach v_2 betrachten: Es gilt $\|Tv_2\|^2 = |a_{22}|$ und

$$\|T^*v_2\| = |a_{22}|^2 + |a_{23}|^2 + \cdots + |a_{2n}|^2.$$

Wie oben folgt $a_{23} = a_{24} = \cdots = a_{2n} = 0$. So macht man weiter bis man erhält, dass $a_{ij} = 0$ für alle $1 \leq i \neq j \leq n$. Also ist $[T]_{\mathcal{B}}$ diagonal. □

$\mathbb{R} \leftrightarrow \mathbb{C}$ Wo haben wir $\mathbb{F} = \mathbb{C}$ benutzt? Nur für die Trigonalisierbarkeit von T, um den Satz von Schur anwenden zu können. Daher folgt:

$\mathbb{F}$ **Korollar 7.2.7.** *Ist T ein normaler Endomorphismus eines Skalarproduktraums über $\mathbb{F}$ und ist T trigonalisierbar ($\Leftrightarrow$ p_T zerfällt in Linearfaktoren über $\mathbb{F}$, was über $\mathbb{C}$ immer gilt), dann ist T orthogonal diagonalisierbar.*

7.2.2 Selbstadjungierte Endomorphismen und Spektralsatz über $\mathbb{R}$

Beispiel 7.2.4 zeigt, dass normale Endomorphismen über $\mathbb{R}$ nicht unbedingt orthogonal diagonalisierbar sind. Wenn wir aber den Beweis von Lemma 7.2.1 über $\mathbb{R}$ betrachten, sehen wir, dass eine stärkere Aussage möglich ist:

$\mathbb{R}$ **Lemma 7.2.8.** *Falls $T \in \operatorname{End}_{\mathbb{R}}(V)$ orthogonal diagonalisierbar ist, gilt $T^* = T$.*

Beweis. Sei $\mathcal{B}$ eine orthonormale Basis von V, sodass $[T]_{\mathcal{B}}$ diagonal ist. Laut (7.4) gilt

$$[T^*]_{\mathcal{B}} = \overline{[T]_{\mathcal{B}}}^T \overset{\mathbb{R}}{=} [T]_{\mathcal{B}}^T \overset{\text{diagonal}}{=} [T]_{\mathcal{B}}.$$

Wie in Lemma 7.2.1 folgt (aus Satz 3.5.3), dass $T^* = T$. □

$\mathbb{F}$ **Definition 7.2.9.** Ein Endomorphismus T heisst *selbstadjungiert*, falls $T^* = T$. Eine Matrix A heisst *selbstadjungiert* falls $\overline{A}^T = A$ (oder äquivalent dazu, falls m_A selbstadjungiert ist).

Aus (7.4) und Satz 3.5.3 folgt:

Korollar 7.2.10. *T ist selbstadjungiert genau dann, wenn $[T]_{\mathcal{B}}$ eine selbstadjungierte Matrix ist, wobei $\mathcal{B}$ eine orthonormale Basis ist.*

Für $\mathbb{F} = \mathbb{R}$ ist eine selbstadjungierte Matrix das gleiche wie eine symmetrische Matrix, für $\mathbb{F} = \mathbb{C}$ ist es das gleiche wie eine hermitesche Matrix.

Beispiel 7.2.11. Jeder selbstadjungierte Endomorphismus ist normal. Endomorphismen wie in Beispiel 7.2.4 sind normal, aber nicht selbstadjungiert.

$\mathbb{R}$ **Satz 7.2.12** (Spektralsatz über $\mathbb{R}$)**.** *Ein Endomorphismus T eines endlich-dimensionalen euklidischen Vektorraums ist orthogonal diagonalisierbar genau dann, wenn T selbstadjungiert ist.*

Beweis. Die Richtung $\Longrightarrow$ haben wir gesehen. Für die Richtung $\Longleftarrow$ lesen Sie Korollar 7.2.7 nochmals. Laut diesem Korollar, da ein selbstadjungierter Endomorphismus normal ist, genügt es, zu zeigen, dass ein selbstadjungierter Endomorphismus T über $\mathbb{R}$ trigonalisierbar ist, oder äquivalent dazu, dass p_T in Linearfaktoren zerfällt in $\mathbb{R}[x]$. Daher folgt der Satz aus dem folgenden Lemma:

$\mathbb{F}$ **Lemma 7.2.13.** *Sei T ein selbstadjungierter Endomorphismus eines Skalarproduktraums über $\mathbb{F}$. Dann gilt:*

(1) Alle Eigenwerte von T sind reell.

(2) p_T zerfällt in Linearfaktoren über $\mathbb{F}$.

(1): Sei λ ein Eigenwert von T und v ein entsprechender Eigenvektor. Selbstadjungierte Endomorphismen sind normal, also ist laut (3) von Lemma 7.2.5 v ein Eigenvektor von T^* zum Eigenwert $\overline{\lambda}$. Da $T = T^*$, haben wir

$$\lambda v = T(v) = T^*(v) = \overline{\lambda} v$$

und daher ist $\lambda = \overline{\lambda}$, also ist λ reell.

(2): Da dieser Teil trivial ist für $\mathbb{F} = \mathbb{C}$, nehmen wir an, dass $\mathbb{F} = \mathbb{R}$. Sei $\mathcal{B}$ eine orthonormale Basis von V und betrachten wir $A = [T]_{\mathcal{B}}$. Es genügt, zu zeigen, dass p_A

in Linearfaktoren in $\mathbb{R}[X]$ zerfällt. Bemerken Sie, dass A symmetrisch/selbstadjungiert ist:

$$A = [T]_{\mathcal{B}} \overset{T=T^*}{=} [T^*]_{\mathcal{B}} \overset{(7.4)}{=} \overline{[T]_{\mathcal{B}}}^T \overset{\mathbb{R}}{=} [T]_{\mathcal{B}}^T = A^T.$$

Nun kommt ein toller Trick: Wir betrachten $\mathbb{C}^n$ mit dem Standard-Skalarprodukt und die Abbildung $m_A : \mathbb{C}^n \to \mathbb{C}^n$. Erinnern Sie sich, dass die Standard-Basis $\mathcal{E}_n$ orthonormal ist bezüglich dem Standard-Skalarprodukt und dass $[m_A]_{\mathcal{E}_n} = A$. Weiter gilt

$$p_{m_A} = p_A = \prod_{i=1}^{n}(x - \lambda_i) \quad (\text{in } \mathbb{C}[x]) \tag{7.6}$$

wobei $\lambda_1, \ldots, \lambda_n$ die Eigenwerte von m_A sind. Es genügt also, zu zeigen, dass $\lambda_i \in \mathbb{R}$ für alle $1 \leq i \leq n$, da dann die Zerlegung (7.6) auch in $\mathbb{R}[x]$ gilt. Da A selbstadjungiert ist, impliziert Korollar 7.1.8, dass m_A selbstadjungiert ist. Teil (1) sagt uns, dass alle Eigenwerte von m_A (und somit alle Nullstellen von p_{m_A}) reell sind. Dies beendet den Beweis von Lemma 7.2.13 und somit den Beweis des Spektralsatzes über $\mathbb{R}$. $\square$

Bemerkung 7.2.14. Wenn wir das Tensor Produkt behandeln, werden wir den Begriff der *Komplexifizierung* sehen. Das heisst, für einen $\mathbb{R}$-Vektorraum V und einen Endomorphismus $T : V \to V$ werden wir in der Lage sein, einen $\mathbb{C}$-Vektorraum $V \otimes_{\mathbb{R}} \mathbb{C}$ und eine $\mathbb{C}$-lineare Abbildung

$$T_{\mathbb{C}} := T \otimes_{\mathbb{R}} \mathbb{C} : V \otimes_{\mathbb{R}} \mathbb{C} \longrightarrow V \otimes_{\mathbb{R}} \mathbb{C}$$

zu konstruieren. Es gilt dann $p_T = p_{T_{\mathbb{C}}}$ und $T_{\mathbb{C}}$ ist selbstadjungiert genau dann, wenn T selbstadjungiert ist (sowie auch weitere nützliche Eigenschaften, die T und $T_{\mathbb{C}}$ verbinden). Die Aussage (2) aus Lemma 7.2.13 im obigen Beweis folgt dann direkt aus (1): $p_T = p_{T_{\mathbb{C}}}$ zerfällt in Linearfaktoren über $\mathbb{C}$, aber alle Eigenwerte von $T_{\mathbb{C}}$ sind reell, also haben wir eine Faktorisierung über $\mathbb{R}$.

Was wir im Beweis oben gemacht haben ist grundsätzlich das gleiche Argument, wobei wir die Tatsache verwendet haben, dass wir für Matrizen $A \in M_{n\times n}(\mathbb{R})$ bereits wissen, wie man $m_A : \mathbb{R}^n \to \mathbb{R}^n$ komplexifiziert: Die Komplexifizierung von $\mathbb{R}^n$ ist einfach $\mathbb{C}^n$ und die Komplexifizierung von m_A ist gleich $m_A : \mathbb{C}^n \to \mathbb{C}^n$, wobei wir A als komplexe Matrix betrachten (deren Einträge „zufällig“ alle reell sind).

$\mathbb{R} \leftrightarrow \mathbb{C}$ *Bemerkung* 7.2.15. Es ist verwirrend, wenn man gleichzeitig über die Begriffe dieses Abschnitts über $\mathbb{R}$ und $\mathbb{C}$ nachdenkt. Daher trennen wir:

Über $\mathbb{R}$ sind selbstadjungierte Matrizen genau die symmetrischen Matrizen. Diese sind auch normal. Über $\mathbb{C}$ gibt es aber symmetrische Matrizen, die nicht normal sind, wie das nächste Beispiel zeigt.

Beispiel 7.2.16. Die Matrix $A = \begin{pmatrix} i & i \\ i & 1 \end{pmatrix}$ ist nicht normal. Die adjungierte Matrix ist $A^* = \overline{A}^T = \begin{pmatrix} -i & -i \\ -i & 1 \end{pmatrix}$ und es gilt $(AA^*)_{12} = 1 + i$, aber $(A^*A)_{12} = 1 - i$.

Im nächsten Abschnitt werden wir mehr Beispiele mit Matrizen sehen.

$\mathbb{R} \leftrightarrow \mathbb{C}$ Die Leser werden sich vielleicht über den Unterschied zwischen den Spektralsätzen über $\mathbb{R}$ und $\mathbb{C}$ wundern und sich fragen, inwiefern die Matrix aus Beispiel 7.2.4 ein typischer Repräsentant einer normalen, aber nicht orthogonal diagonalisierbaren Transformation ist. In diesem Beispiel haben wir gesehen, dass die normale Matrix $\begin{pmatrix} & 1 \\ -1 & \end{pmatrix}$ nicht orthogonal diagonalisierbar ist, weil sie nicht diagonalisierbar ist. Nun stellt sich die folgende Frage: Ist ein normaler, diagonalisierbarer Endomorphismus eines Skalarproduktraums über $\mathbb{R}$ automatisch orthogonal diagonalisierbar?

$\mathbb{R}$ **Übung 7.2.17.** Zeigen Sie, dass für einen Endomorphismus T eines reellen Skalarproduktraums die folgenden Aussagen äquivalent sind:

(1) T ist selbstadjungiert.

(2) T ist orthogonal diagonalisierbar.

(3) T ist normal und diagonalisierbar.

(4) T ist normal und trigonalisierbar.

(5) T ist normal und p_T zerfällt in Linearfaktoren.

Insbesondere ist eine diagonalisierbare, normale Matrix über $\mathbb{R}$ symmetrisch.

Hinweis. Die Äquivalenz (1) $\iff$ (2) ist Satz 7.2.12, wobei (2) $\implies$ (1) der einfache Teil ist (Lemma 7.2.1).

(5) $\implies$ (2) ist Korollar 7.2.7. Alle anderen Implikationen folgen unmittelbar.

7.3 Spektralsätze für Matrizen

Erinnern Sie sich an die Definition von orthogonalen und unitären Matrizen, sowie an $\mathrm{O}(n)$ und $\mathrm{U}(n)$ (Definition 6.4.13).

$\mathbb{C}$ Betrachten wir $\mathbb{C}^n$ mit dem Standard-Skalarprodukt und sei $A \in M_{n\times n}(\mathbb{C})$ eine normale Matrix (d.h. $\overline{A}^T A = A\overline{A}^T$). Der Spektralsatz über $\mathbb{C}$ (Satz 7.2.3) impliziert:

Satz 7.3.1. *Für jede normale Matrix $A \in M_{n\times n}(\mathbb{C})$ existiert eine unitäre Matrix $U \in \mathrm{U}(n)$, sodass $U^{-1}AU = \overline{U}^T AU$ diagonal ist.*

Beweis. Da A normal ist, ist $m_A : \mathbb{C}^n \to \mathbb{C}^n$ normal, wobei $\mathbb{C}^n$ mit dem Standard-Skalarprodukt versehen ist. Wir wählen eine Orthonormalbasis $\mathcal{B}$ von $\mathbb{C}^n$ bestehend aus Eigenvektoren von A (hier verwenden wir Satz 7.2.3) und definieren $U := [\mathrm{Id}]^{\mathcal{B}}_{\mathcal{E}_n}$. Die Spalten von U sind genau die Elemente aus $\mathcal{B}$, somit ist U unitär. Ausserdem ist

$$U^{-1}AU = [\mathrm{Id}]^{\mathcal{E}_n}_{\mathcal{B}}[m_A]_{\mathcal{E}_n}[\mathrm{Id}]^{\mathcal{B}}_{\mathcal{E}_n} = [m_A]_{\mathcal{B}}$$

diagonal. □

$\mathbb{R}$ Ähnlich gilt auch über $\mathbb{R}$:

Satz 7.3.2. *Für jede symmetrische Matrix $A \in M_{n\times n}(\mathbb{R})$ existiert eine orthogonale Matrix $O \in \mathrm{O}(n)$, sodass $O^{-1}AO = O^TAO$ diagonal ist.*

Der Beweis ist analog zu dem über $\mathbb{C}$ und wir überlassen ihn den Lesern.

$\mathbb{F}$ **Rechenverfahren**

Wie muss man konkret vorgehen, um eine Orthonormalbasis aus Eigenvektoren für eine gegebene Matrix $A \in M_{n\times n}(\mathbb{F})$ zu finden? Wir nehmen an, dass A normal ist (falls $\mathbb{F} = \mathbb{C}$) oder dass A symmetrisch ist (falls $\mathbb{F} = \mathbb{R}$).

1. Berechnen Sie das charakteristische Polynom p_A.
2. Faktorisieren Sie p_A, sodass Sie alle Nullstellen (inklusive Vielfachheit) sehen können. Dies sind die Eigenwerte von A (und die Vielfachheiten sind die Dimensionen der dazugehörigen Eigenräume[3]).
3. Berechnen Sie von jedem Eigenraum eine Basis, dann wenden Sie Gram-Schmidt an, um eine Orthonormalbasis jedes Eigenraums zu erhalten.
4. Fügen Sie die Orthonormalbasen der Eigenräume zu einer Orthogonalbasis von $\mathbb{F}^n$ zusammen. Gemäss Lemma 7.2.5 (4) müssen Sie das Gram-Schmidt Verfahren nicht nochmals auf alle Vektoren anwenden, da die verschiedenen Eigenräume schon orthogonal zueinander sind.

$\mathbb{R}$ **Beispiel 7.3.3.** Betrachten wir $A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$ und diagonalisieren wir A orthogonal. Wir berechnen $p_A = -(x+1)^2(x-2)$ und dann

[3]Hier benutzen wir, dass für diagonalisierbare Endomorphismen die algebraische Vielfachheit gleich der geometrischen Vielfachheit ist (siehe Satz 5.3.35). Dass A diagonalisierbar ist, folgt aus den Spektralsätzen (Satz 7.2.3 für $\mathbb{F} = \mathbb{C}$ und Satz 7.2.12 für $\mathbb{F} = \mathbb{R}$).

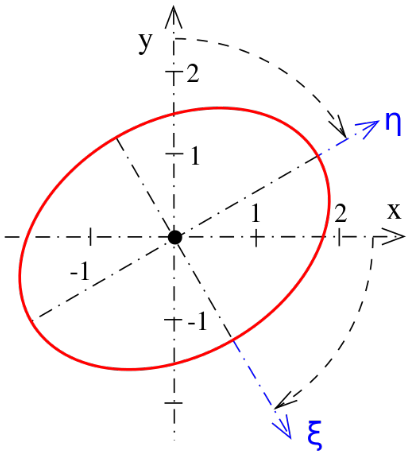
$$\mathrm{Eig}_A(-1) = \mathrm{Sp}\left\{\begin{pmatrix}-1\\1\\0\end{pmatrix}, \begin{pmatrix}-1\\0\\1\end{pmatrix}\right\}, \quad \mathrm{Eig}_A(2) = \mathrm{Sp}\begin{pmatrix}1\\1\\1\end{pmatrix}.$$

Seien $v_1 = \begin{pmatrix}-1\\1\\0\end{pmatrix}$, $v_2 = \begin{pmatrix}-1\\0\\1\end{pmatrix}$ und $v_3 = \begin{pmatrix}1\\1\\1\end{pmatrix}$. Bemerken Sie, dass die Theorie uns sagt, dass $\mathrm{Eig}_A(-1) \perp \mathrm{Eig}_A(2)$, was sich in diesem Beispiel leicht überprüfen lässt. Wir wenden Gram-Schmidt auf beide Eigenräume an. Für $\mathrm{Eig}_A(2)$ bedeutet das nur, v_3 zu normalisieren, was in $e_3 = \frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix}$ resultiert. Für $\mathrm{Eig}_A(-1)$ gibt uns der Gram-Schmidt Algorithmus die Vektoren $e_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}-1\\1\\0\end{pmatrix}$ und $e_2 = \frac{1}{\sqrt{6}}\begin{pmatrix}-1\\-1\\2\end{pmatrix}$. Wir haben

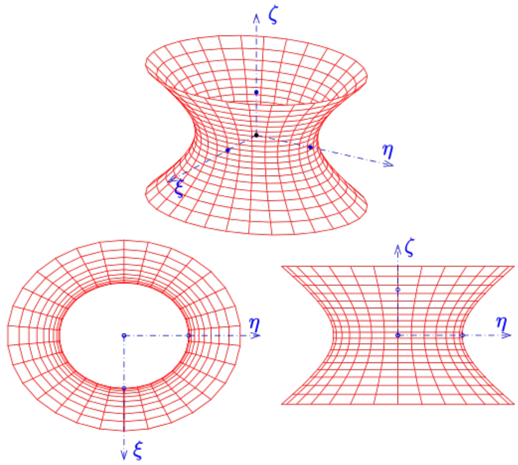
$$\begin{pmatrix}-1 & & \\ & -1 & \\ & & 2\end{pmatrix} = \begin{pmatrix} — & e_1^T & — \\ — & e_2^T & — \\ — & e_3^T & — \end{pmatrix} A \begin{pmatrix} | & | & | \\ e_1 & e_2 & e_3 \\ | & | & | \end{pmatrix}$$

Die geometrische Bedeutung ist, dass bezüglich des kartesischen Koordinatensystems (e_1, e_2, e_3) die Abbildung m_A einfach eine Punktspiegelung in der (e_1, e_2)-Ebene durch den Ursprung ist (d.h. für $v = ae_1 + be_2$ ist $Av = -v$) zusammen mit einer Streckung mit Faktor 2 in der e_3-Koordinate.

Das geometrische Gegenstück zu Satz 7.3.2 ist die sogenannte Hauptachsentransformation: Sei A eine symmetrische reelle Matrix (bzw. selbstadjungierte lineare Abbildung eines euklidischen Vektorraums). Die orthonormale Basis, welche durch den Satz 7.3.2 gegeben ist, bildet ein kartesisches Koordinatensystem, in welchem A durch eine Skalierung/Streckung von jeder Achse mit (voneinander unabhängigen) reellen Faktoren agiert. Über klassische wunderschöne Anwendungen von diesem Satz können Sie auf de.wikipedia.org/wiki/Hauptachsentransformation mehr lesen. Dort finden Sie unter anderem auch Beschreibungen der folgenden Bilder:



Figur 7.1: Hauptachsen einer Ellipse in $\mathbb{R}^2$. Quelle: Wikipedia.[4]



Figur 7.2: Hauptachsen eines Hyperboloids in $\mathbb{R}^3$. Quelle: Wikipedia.[4]

Beispiel 7.3.4. Lassen Sie mich eine grundlegende Anwendung in der Analysis skizzieren. Wir kennen alle den „Test der zweiten Ableitung“ aus der Analysis[5]: Für glatte Funktionen $f : \mathbb{R} \to \mathbb{R}$ und $a \in \mathbb{R}$ mit $f'(a) = 0$ gilt: f hat ein lokales Maximum in a falls $f''(a) < 0$ und ein lokales Minimum falls $f''(a) > 0$.

Für Funktionen in mehreren Variablen können wir etwas ähnliches beweisen, und dabei etwas über die Geometrie von lokalem Verhalten lernen. Für glatte $f : \mathbb{R}^n \to \mathbb{R}$ gilt für alle $x^0 \in \mathbb{R}^n$:

$$f(x) = f(x^0) + \left\langle \nabla f(x^0), x - x^0 \right\rangle + (x - x^0)^T \left(H_f(x^0)\right)(x - x^0) + \mathcal{O}\left(\left\|x - x^0\right\|^3\right),$$

wobei $\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right)^T$ der Gradient von f und $H_f(x) = \left(\frac{\partial f}{\partial x_i \partial x_j}(x)\right)_{i,j}$ die sogenannte Hesse-Matrix von f ist. Der Gradient ∇f, bzw. die Hesse-Matrix H_f, sind die Analogon in der mehrdimensionalen Differentialrechnung zur ersten bzw. zweiten Ableitung.

Nehmen Sie an, dass $\nabla f(x^0) = 0 \in \mathbb{R}^n$. Ohne Beschränkung der Allgemeinheit (mit $g(x) := f(x + x^0)$) können wir annehmen, dass $x^0 = 0 \in \mathbb{R}^n$. Dann gilt laut (7.3.4) mit $x^0 = 0$ und $\nabla f(0) = 0$, dass

$$f(x) = f(0) + x^T H_f(0) x + \mathcal{O}\left(\|x\|^3\right).$$

[4]Attribution: Ag2gaeh, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons

[5]Der Einfachheit halber beschreiben wir ihn hier nur für glatte Funktionen

Da $H_f(0)$ symmetrisch ist, folgt aus dem Spektralsatz, dass es eine orthogonale Matrix $O \in \mathrm{O}(n)$ und $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ gibt mit

$$x^T H_f(0) x = x^T O^T \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} Ox = (Ox)^T \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} Ox.$$

Fassen wir $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} := Ox$ als neue orthonormale Koordinaten auf. In anderen Worten $y_1, \ldots, y_n$ bilden ein kartesisches Koordinatensystem. Damit gilt

$$\begin{aligned} f(y) &= f(0) + \begin{pmatrix} y_1 & \cdots & y_n \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} + \mathcal{O}\left(\|y\|^3\right) \\ &= f(0) + \sum_{i=1}^{n} \lambda_i y_i^2 + \mathcal{O}\left(\|y\|^3\right) \end{aligned}$$

(wir werden später sehen, dass $\|Ox\| = \|x\|$ und somit $\mathcal{O}\left(\|y\|^3\right) = \mathcal{O}\left(\|x\|^3\right)$).

Es folgt, dass wir entlang der y_i-Koordinate eine Parabel der Form $\lambda_i y_i^2$ sehen und dass wir ein Maximum in 0 haben, falls $\lambda_i < 0$ für alle $i = 1, \ldots, n$ und ein Minimum, falls $\lambda_i > 0$ für alle $i = 1, \ldots, n$.

Bemerkung 7.3.5. Grob gesagt, wenn man die Argumentation aus Beispiel 7.3.4 auf Polynomfunktionen zweiten Grades $p : \mathbb{R}^2 \to \mathbb{R}$ (in 2 Variablen) oder $p : \mathbb{R}^3 \to \mathbb{R}$ (in 3 Variablen) anwendet, bekommt man die klassischen Anwendungen, die im obigen Wikipedia-Link gegeben sind (siehe Figuren 7.1 und 7.2).

$\mathbb{R} \leftrightarrow \mathbb{C}$ *Bemerkung* 7.3.6. Man kann mit mit unserem euklidisch-unitär Wörterbuch (Seite 243) diese Hauptachsentransformation auch für unitäre Vektorräume und normale Abbildungen übersetzen. Wir haben es über $\mathbb{R}$ geschrieben, da es historisch über $\mathbb{R}$ formuliert war.

$\mathbb{F}$ Symmetrische reelle Matrizen (bzw. hermitesche komplexe Matrizen) führen zu symmetrischen Bilinearformen (bzw. Sesquilinearformen). Wir werden ähnliche Probleme und Ideen wie oben in diesem Kontext wieder behandlen, wenn wir Bilinear- und Sesquilinearformen untersuchen. Jetzt möchten wir Abbildungen betrachten, die sowohl die Vektorraumstruktur als auch das Skalarprodukt erhalten.

7.4 Orthogonale und unitäre Abbildungen

Wir kommen nun zum zweiten Thema dieses Kapitels und befassen uns mit der folgenden natürlichen Frage: Wenn V ein Skalarproduktraum ist, was können wir über lineare Abbildungen $T : V \to V$ aussagen, die das Skalarprodukt erhalten? Zuerst geben wir eine Definition, um genau zu beschreiben, was wir damit meinen.

Definition 7.4.1. Sei V ein Skalarproduktraum über $\mathbb{F}$. Ein Endomorphismus $T \in \mathrm{End}(V)$ heisst *orthogonal* falls $\mathbb{F} = \mathbb{R}$ (bzw. *unitär* falls $\mathbb{F} = \mathbb{C}$), wenn

$$\forall v, w \in V \quad \langle Tv, Tw \rangle = \langle v, w \rangle .$$

In beiden Fällen nennen wir T auch eine Isometrie.

Beispiel 7.4.2. Die Identitätsabbildung Id_V, sowie Rotationen und Spiegelungen in $\mathbb{R}^2$ sind orthogonal.

Wir möchten sagen, dass eine Matrix $A \in M_{n\times n}(\mathbb{F})$ orthogonal (bzw. unitär) ist, falls $m_A \in \mathrm{End}(\mathbb{F}^n)$ orthogonal (bzw. unitär) ist im Sinne von Definition 7.4.1. Die Leser sollten sich nun beschweren, da wir orthogonale und unitäre Matrizen in Definition 6.4.13 bereits definiert haben. Aber keine Sorge, wenn wir den folgenden Satz für $V = \mathbb{F}^n$ mit dem Standard-Skalarprodukt anwenden, sehen wir, dass beide Definitionen gleich sind.

Satz 7.4.3. *Sei $(V, \langle\,,\rangle)$ ein endlich-dimensionaler Skalarproduktraum mit induzierter Norm $\|\cdot\|$. Für $T \in \mathrm{End}(V)$ sind die folgenden Aussagen äquivalent:*

(1) Für alle $v \in V$ gilt $\|Tv\| = \|v\|$.

(2) Für alle $u, v \in V$ gilt $\langle Tu, Tv \rangle = \langle u, v \rangle$.

(3) Für jedes orthonormale System $e_1, \ldots, e_k$ ist $Te_1, \ldots, Te_k$ auch ein orthonormales System.

(4) Es existiert eine orthonormale Basis $e_1, \ldots, e_n$, sodass $Te_1, \ldots, Te_n$ auch orthonormal ist.

(5) $T^*T = \mathrm{Id}_V$.

(6) $TT^* = \mathrm{Id}_V$.

(7) $T^* = T^{-1}$.

(8) Für alle $v \in V$ gilt $\|T^*v\| = \|v\|$.

Beweis. Bemerken Sie zuerst, dass (5), (6) und (7) äquivalent sind, da das rechtsinverse und das linksinverse Element in einer Gruppe gleich sind (die Gruppe die wir hier betrachten, ist die multiplikative Gruppe vom Ring $\mathrm{End}(V)$).[6] Fangen wir jetzt mit dem Beweis an:

(1)⇒(2): Übung 6.1.10 zeigt, dass man das Skalarprodukt durch Normen berechnen kann. Daher folgt (2) aus (1). Wir schreiben das Argument für $\mathbb{F} = \mathbb{R}$:

$$\begin{aligned}\langle Tv, Tw\rangle &= \frac{1}{2}\left(\|Tv + Tw\|^2 - \|Tv\|^2 - \|Tw\|^2\right) = \frac{1}{2}\left(\|T(v+w)\|^2 - \|Tv\|^2 - \|Tw\|^2\right) \\ &\overset{(1)}{=} \frac{1}{2}\left(\|v+w\|^2 - \|v\|^2 - \|w\|^2\right) = \langle v, w\rangle \,,\end{aligned}$$

wobei die erste und die letzte Gleichheit aus Übung 6.1.10 folgen.

(2)⇒(3): Sei $e_1, \ldots, e_k$ ein orthonormales System. Dann gilt für alle $1 \leq i, j \leq k$

$$\langle Te_i, Te_j\rangle \overset{(2)}{=} \langle e_i, e_j\rangle = \delta_{ij},$$

also ist $Te_1, \ldots, Te_k$ auch ein orthonormales System.

(3)⇒(4): Eine Orthonormalbasis $e_1, \ldots, e_n$ existiert laut Satz 6.4.7 (Gram-Schmidt), da $\dim V < \infty$. Dann folgt (4) aus (3).

(4)⇒(5): Sei $(e_1, \ldots, e_n)$ eine orthonormale Basis, sodass $(Te_1, \ldots, Te_n)$ auch orthonormal ist. Es gilt für alle $1 \leq i, j \leq n$

$$\langle e_i, e_j\rangle = \delta_{ij} = \langle Te_i, Te_j\rangle = \langle e_i, T^*Te_j\rangle \,.$$

Da $e_1, \ldots, e_n$ eine Basis ist, folgt aus der Linearität des Skalarprodukts in der ersten Variable, dass

$$\forall v \in V \quad \langle v, e_j\rangle = \langle v, T^*Te_j\rangle \,.$$

Daher gilt gemäss Lemma 6.1.6 $e_j = T^*Te_j$ für $j = 1, \ldots, n$. Da $(e_1, \ldots, e_n)$ eine Basis ist, folgt $T^*T = \mathrm{Id}_V$. Dies zeigt (5)

Wie schon gesagt sind (5), (6) und (7) äquivalent. Wenn wir jetzt (5), (6) und (7) annehmen, dann erhalten wir für alle $v \in V$, dass

$$\langle Tv, Tv\rangle = \langle v, T^*Tv\rangle \overset{(5)}{=} \langle v, \mathrm{Id}_V(v)\rangle = \langle v, v\rangle \,.$$

Dies impliziert (1) und zeigt somit die Äquivalenz von (1) bis (7). Die Aussagen (1) und (5) auf den Endomorphismus T^* angewendet sind genau die Aussagen (8) und (6) (unter Verwendung von $(T^*)^* = T$). Also folgt (8)⇔(6) aus (1)⇔(5). □

[6] Eigentlich haben wir diese Tatsache nur als Übung gelassen. In diesem Kontext kann man auch lineare Algebra benutzen: Aus $T^*T = \mathrm{Id}_V$ folgt, dass T^* surjektiv und T injektiv ist. Laut dem Rangsatz sind beide bijektiv (da $\dim V < \infty$). Da beispielsweise T bijektiv ist, folgt $T^{-1} = T^*$ und dann auch $T \circ T^* = \mathrm{Id}_V$.

Bemerkung 7.4.4. Aus (5), (6) und (7) folgt, dass jede orthogonale/unitäre Transformation normal ist.

$\mathbb{R}$ *Bemerkung* 7.4.5. Über $\mathbb{R}$ kann man bei (2) an die Erhaltung von Längen und Winkeln denken.

$\mathbb{F}$ **Beispiel 7.4.6** (Satz 7.4.3 für Matrizen)**.** Erinnern Sie sich, dass unsere Definition von orthogonalen/unitären Matrizen besagte, dass die Spalten eine Orthonormalbasis bilden. Die Lemmas 6.4.15 und 6.4.16 besagen, dass für $A \in M_{n\times n}(\mathbb{R})$

$$A \text{ ist orthogonal} \iff A^T A = I_n \iff AA^T = I_n \iff A^{-1} = A^T$$

und für $A \in M_{n\times n}(\mathbb{C})$

$$A \text{ ist unitär} \iff A^T \overline{A} = I_n \iff A\overline{A}^T = I_n \iff A^{-1} = \overline{A}^T$$

gilt. Dies entspricht genau der Äquivalenz von (4), (5), (6) und (7), wenn man in (4) die Standard-Basis verwendet. Die anderen Äquivalenzen besagen, dass A orthogonal/unitär ist genau dann, wenn $\|Av\| = \|v\|$ für alle $v \in \mathbb{F}^n$ gilt (dies entspricht (1)) und genau dann, wenn $\langle Av, Aw\rangle = \langle v, w\rangle$, und analog für $\overline{A}^T$.

$\mathbb{C}$ **Übung 7.4.7.** Sei V ein endlich-dimensionaler unitärer Vektorraum.

(a) Nehmen wir an, dass es für $T \in \mathrm{End}(V)$ eine orthogonale Basis von V gibt mit $Te_i = \lambda_i e_i$, wobei $\lambda_i \in \mathbb{C}$ mit $|\lambda_i| = 1$, für alle $i = 1, \ldots, n$. Zeigen Sie, dass T unitär ist.

(b) Zeigen Sie, dass jeder unitäre Endomorphismus $T \in \mathrm{End}(V)$ von dieser Form ist.

Lösung (Skizze). Für (a) zeigen Sie, dass T die Aussage (4) von Satz 7.4.3 erfüllt. Dafür muss man lediglich zeigen, dass $\lambda_1 e_1, \ldots, \lambda_n e_n$ auch eine orthonormale Basis ist (hier benuzt man dann, dass $|\lambda_i| = 1$).

(b) folgt aus dem Spektralsatz: Sei T unitär. Aus Bemerkung 7.4.4 folgt, dass T normal ist, also gibt es laut dem Spektralsatz 7.2.3 eine orthonormale Basis aus Eigenvektoren von T. Benutzen Sie nun (1), (2) oder (4) von Satz 7.4.3, um zu zeigen, dass die entsprechenden Eigenwerte den Absolutbetrag 1 haben müssen. □

$\mathbb{F}$ **Proposition 7.4.8.** *Sei $T \in \mathrm{End}(V)$, wobei V ein endlich-dimensionaler Skalarproduktraum ist, und sei $\mathcal{B}$ eine orthonormale Basis von V. Dann gilt:*

$$T \textit{ ist orthogonal/unitär} \iff [T]_{\mathcal{B}} \textit{ ist orthogonal/unitär.}$$

Beweis. Laut der Äquivalenz von (2) und (6) in Satz 7.4.3 ist T orthogonal/unitär genau dann, wenn $TT^* = \mathrm{Id}_V$, also genau dann, wenn

$$I = [\mathrm{Id}_V]_{\mathcal{B}} = [TT^*]_{\mathcal{B}} = [T]_{\mathcal{B}}[T^*]_{\mathcal{B}} = [T]_{\mathcal{B}}\overline{[T]_{\mathcal{B}}}^T,$$

also genau dann, wenn $[T]_{\mathcal{B}}$ orthogonal/unitär ist. $\square$

Es ist höchste Zeit, einige wichtige Gruppen einzuführen:

Definition 7.4.9. Die Menge

$$\mathrm{O}(n) = \{A \in M_{n\times n}(\mathbb{R}) \mid A \text{ ist orthogonal}\} = \{A \in M_{n\times n}(\mathbb{R}) \mid A^{-1} = A^T\}$$

heisst die *orthogonale Gruppe*,

$$\mathrm{SO}(n) = \{A \in M_{n\times n}(\mathbb{R}) \mid A \in \mathrm{O}(n),\ \det A = 1\}$$

heisst die *spezielle orthogonale Gruppe* und

$$\mathrm{U}(n) = \{A \in M_{n\times n}(\mathbb{C}) \mid A \text{ ist unitär}\} = \{A \in M_{n\times n}(\mathbb{C}) \mid A^{-1} = \overline{A}^T\}$$

heisst die *unitäre Gruppe.*

Proposition 7.4.10. *Diese Mengen sind in der Tat Gruppen bezüglich der Matrix Multiplikation. Die ersten zwei sind Untergruppen von* $\mathrm{GL}_n(\mathbb{R})$ *und die letzte von* $\mathrm{GL}_n(\mathbb{C})$.

Beweis. Beweisen wir es für $\mathrm{U}(n)$: Wenn man sich noch daran erinnert, was eine Untergruppe ist, bemerkt man, dass es genügt zu zeigen, dass $I_n \in \mathrm{U}(n)$ (was klar ist) und dass

$$\begin{aligned} A \in \mathrm{U}(n) &\implies A^{-1} \in \mathrm{U}(n), \\ A, B \in \mathrm{U}(n) &\implies AB \in \mathrm{U}(n). \end{aligned}$$

Für die erste Implikation sei $A \in \mathrm{U}(n)$. Laut der Definition von $\mathrm{U}(n)$ ist $A^{-1} = \overline{A}^T$ und somit ist

$$\left(A^{-1}\right)^{-1} = A = \overline{\overline{A}^T}^T = \overline{A^{-1}}^T,$$

was zeigt, dass $A^{-1} \in \mathrm{U}(n)$. Für $A, B \in \mathrm{U}(n)$ folgt die zweite Implikation aus

$$(AB)^{-1} = B^{-1}A^{-1} = \overline{B}^T\overline{A}^T = \left(\overline{A} \cdot \overline{B}\right)^T = \overline{(AB)}^T. \qquad \square$$

Wir zeigen nun, dass diese drei Gruppen kompakt sind. Diese Eigenschaft ist sehr nützlich in vielen Bereichen der Mathematik und Physik, welche Ihnen im Verlauf des

Studiums begegnen werden: Die Darstellungstheorie ist für kompakte Gruppen am einfachsten unter allen unendlichen Gruppen, Dynamik ist auch einfacher auf kompakten Gruppen, und es gäbe noch viele weitere Beispiele. In vielerlei Hinsicht sind kompakte Gruppen ähnlich zu endlichen Gruppen. Es ist auch interessant zu sehen, wie diese Gruppen eine wichtige Rolle in verschiedenen Matrixzerlegungen spielen (vgl. Beispiel 6.4.21).

Proposition 7.4.11. *Die Gruppen* $\mathrm{O}(n)$, $\mathrm{SO}(n)$, $\mathrm{U}(n)$ *sind kompakte Teilmengen von* $M_{n\times n}(\mathbb{R})$ *(bzw.* $M_{n\times n}(\mathbb{C})$*).*

Beweis. Wir beweisen es für $\mathrm{O}(n)$. Aus der Analysis wissen Sie, dass eine Teilmenge von $\mathbb{R}^m$ genau dann kompakt ist, wenn sie abgeschlossen und beschränkt ist. Erinnern Sie sich ausserdem an die Identifizierung $M_{n\times n}(\mathbb{R}) \cong \mathbb{R}^{n^2}$. Unter dieser Identifizierung erhalten wir aus der Standard-Norm auf $\mathbb{R}^{n^2}$ (diejenige, die vom Standard-Skalarprodukt auf $\mathbb{R}^{n^2}$ induziert wird) auch eine Norm auf $M_{n\times n}(\mathbb{R})$, nämlich ist für $A = (a_{ij})_{i,j} \in M_{n\times n}(\mathbb{R})$

$$\|A\| := \sqrt{\sum_{i,j=1}^{n} |a_{ij}|^2}.$$

Somit genügt es zu zeigen, dass $\mathrm{O}(n)$ abgeschlossen ist und dass alle Elemente von $\mathrm{O}(n)$ uniform beschränkt sind bezüglich der obigen Norm. Betrachten Sie die Einträge einer Matrix $A = (a_{ij})_{1\leq i,j\leq n}$ als Variablen: Die Bedingung $AA^T = I_n$ ist äquivalent zu n^2 Polynomgleichungen in den Variablen a_{ij} und $\mathrm{O}(n)$ ist genau die Lösungsmenge dieser Gleichungen. Da Polynome stetig sind, ist diese Lösungsmenge abgeschlossen.[7]

Für Beschränktheit von $\mathrm{O}(n)$ erinnern Sie sich, dass die Spalten einer Matrix $A = (a_{ij})_{i,j} \in \mathrm{O}(n)$ ein Orthonormalsystem bilden und insbesondere normiert sind. Wir schreiben $v_j = (a_{1j}, \ldots, a_{nj})^T \in \mathbb{R}^n$ für die Spaltenvektoren von A. Es folgt, dass

$$\|A\|^2 = \sum_{j=1}^{n}\sum_{i=1}^{n} |a_{ij}|^2 = \sum_{j=1}^{n} \underbrace{\|v_j\|^2}_{=1} = n.$$

Somit ist $\|A\| = \sqrt{n}$ für alle $A \in \mathrm{O}(n)$, was die Beschränktheit zeigt. □

ℝ

7.4.1 Klassifikation der Elemente in $\mathrm{O}(2)$, $\mathrm{SO}(2)$, $\mathrm{O}(3)$ und $\mathrm{SO}(3)$

Wir möchten alle Elemente von $\mathrm{O}(2)$ und $\mathrm{O}(3)$ klassifizieren und geometrisch verstehen. Erinnern Sie sich zuerst an den folgenden Fakt:

Lemma 7.4.12. *Falls* $A \in O(n)$ *und* λ *ein Eigenwert von* A *ist, so ist* $\lambda = \pm 1$.

[7] Die Lösungsmenge ist das Urbild einer abgeschlossenen Menge unter einer stetigen Funktion und somit abgeschlossen.

Beweis. Sei $v \in \mathbb{R}^n$ ein Eigenvektor von A zum Eigenwert λ. Es gilt

$$\langle v, v\rangle = \langle Av, Av\rangle = \langle \lambda v, \lambda v\rangle = \lambda^2 \langle v, v\rangle,$$

also ist $\lambda^2 = 1$ und somit entweder $\lambda = 1$ oder $\lambda = -1$. $\square$

Proposition 7.4.13. *Sei $A \in \mathrm{O}(2)$.*

(1) Falls $|A| = 1$, dann existiert $\theta \in [0, 2\pi)$, sodass

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} =: R_\theta.$$

Geometrisch ist R_θ eine Rotation um den Winkel θ in $\mathbb{R}^2$.

(2) Falls $|A| = -1$, dann existiert $\theta \in [0, 2\pi)$, sodass

$$A = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} = R_\theta \cdot \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}$$

und A ist orthogonal diagonalisierbar. Ausserdem ist bezüglich der Basis

$$\mathcal{B} = (v_1, v_2) = \left(\begin{pmatrix} \cos\frac{\theta}{2} \\ \sin\frac{\theta}{2} \end{pmatrix}, \begin{pmatrix} \cos(\frac{\theta}{2} + \frac{\pi}{2}) \\ \sin(\frac{\theta}{2} + \frac{\pi}{2}) \end{pmatrix} \right)$$

die Darstellungsmatrix von m_A gegeben durch $[m_A]_{\mathcal{B}} = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}$.

Beweis. Aus Satz 7.4.3 wissen wir, dass $\|Ae_1\| = 1$. Folglich existiert $\theta \in [0, 2\pi)$ mit $Ae_1 = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$. Wir wissen auch, dass $Ae_1 \perp Ae_2$ und $\|Ae_2\| = 1$. Dies impliziert, dass

- entweder $Ae_2 = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$, in diesem Fall ist $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} =: R_\theta$ und $|A| = \cos^2\theta + \sin^2\theta = 1$,
- oder $Ae_2 = \begin{pmatrix} \sin\theta \\ -\cos\theta \end{pmatrix}$, in diesem Fall ist $A = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} = R_\theta \cdot \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}$ und $|A| = -1$.

Im zweiten Fall können Sie berechnen, dass die Vektoren v_1 und v_2 tatsächlich Eigenvektoren von A zu den Eigenwerten 1 und -1 sind, mithilfe von trigonomischen Identitäten. Wir können aber auch geometrisch argumentieren:

Für $\varphi \in [0, 2\pi)$ sei $v_\varphi = \begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix}$ ein Einheitsvektor in Richtung φ. Also ist $v_1 = v_{\frac{\theta}{2}}$ und $v_2 = v_{\frac{\theta}{2}+\frac{\pi}{2}}$. Bemerken Sie, dass

$$R_\theta \begin{pmatrix} 1 & \\ & -1 \end{pmatrix} v_{\frac{\theta}{2}} = R_\theta v_{-\frac{\theta}{2}} = v_{-\frac{\theta}{2}+\theta} = v_{\frac{\theta}{2}} \text{ und}$$
$$R_\theta \begin{pmatrix} 1 & \\ & -1 \end{pmatrix} v_{\frac{\theta}{2}+\frac{\pi}{2}} = R_\theta v_{-\frac{\theta}{2}-\frac{\pi}{2}} = v_{\frac{\theta}{2}-\frac{\pi}{2}} = -v_{\frac{\theta}{2}+\frac{\pi}{2}}.$$

□

Korollar 7.4.14. *Falls T ein orthogonaler Endomorphismus eines 2-dimensionalen reellen Skalarproduktraums V ist mit $\det T = -1$, dann existiert eine Orthonormalbasis $\mathcal{B}$ von V mit $[T]_\mathcal{B} = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}$.*

Beweis. Laut Proposition 7.4.8 ist $[T]_\mathcal{B} \in \mathrm{O}(2)$ mit $\det[T]_\mathcal{B} = \det T = -1$. Laut Proposition 7.4.13 (2) hat $[T]_\mathcal{B}$ jeweils einen Eigenvektor zu den Eigenwerten 1 und -1 und laut Übung 5.2.15 gilt das gleiche für T. □

Proposition 7.4.15. *Sei $A \in \mathrm{SO}(3)$. Dann hat A mindestens einen Eigenwert.*

- *Falls -1 ein Eigenwert von A ist, dann existiert eine orthonormale Basis $\mathcal{B}$ von $\mathbb{R}^3$ mit $[m_A]_\mathcal{B} = \begin{pmatrix} -1 & & \\ & 1 & \\ & & -1 \end{pmatrix}$.*
- *Falls 1 der einzige Eigenwert von A ist, dann existiert eine orthonormale Basis $\mathcal{B} = (v_1, v_2, v_3)$ von $\mathbb{R}^3$ und $\theta \in [0, 2\pi)$ mit $[m_A]_\mathcal{B} = \left(\begin{array}{c|c} 1 & \\ \hline & R_\theta \end{array}\right)$. Geometrisch ist dies eine Rotation mit Winkel θ um die Achse $\mathrm{Sp}(v_1)$.*

Bemerken Sie, dass in beiden Fällen 1 ein Eigenwert von A ist.

Beweis. Betrachten wir das charakteristische Polynom p_A, welches ein reelles Polynom von Grad 3 ist. Wir haben gesehen, dass ein solches Polynom eine reelle Nullstelle haben muss (siehe Korollar 1.4.23), also hat A einen Eigenwert λ (welcher laut Lemma 7.4.12 entweder 1 oder -1 ist). Sei $v_1 \in \mathbb{R}^3$ ein normierter Eigenvektor zum Eigenwert λ. Bemerken Sie, dass $L = v_1^\perp = \{v \in \mathbb{R}^3 \mid v \perp v_1\}$ m_A-invariant ist:

$$\begin{aligned} v \in L &\implies 0 = \langle v, v_1 \rangle = \langle Av, Av_1 \rangle = \langle Av, \pm v_1 \rangle = \pm \langle Av, v_1 \rangle \\ &\implies \langle Av, v_1 \rangle = 0 \implies Av \in L. \end{aligned}$$

Weiter wissen wir dank Satz 6.6.3, dass $\mathbb{R}^3 = \mathrm{Sp}(v_1) \oplus L$ (und somit $\dim L = 2$). Bemerken Sie, dass $m_A|_L$ orthogonal ist und sei (w_1, w_2) eine Orthonormalbasis von L. Laut Proposition 7.4.8 ist $B = [m_A|_L]_{(w_1,w_2)} \in \mathrm{O}(2)$.

- Falls $\lambda = 1$, dann haben wir bezüglich der Basis $\mathcal{B} = (v_1, w_1, w_2)$

$$[m_A]_{\mathcal{B}} = \left(\begin{array}{c|c} 1 & \\ \hline & B \end{array}\right),$$

also ist $|B| = |A| = 1$. Es folgt aus Proposition 7.4.13 (1), dass $\theta \in [0, 2\pi)$ existiert mit $[m_A]_{\mathcal{B}} = \left(\begin{array}{c|c} 1 & \\ \hline & R_\theta \end{array}\right)$.

- Falls $\lambda = -1$, dann erhalten wir durch die gleiche Argumentation wie oben, dass

$$\det m_A|_L = \det B = -1.$$

Laut Korollar 7.4.14 hat $m_A|_L$ Eigenvektoren v_2, v_3 zu den Eigenwerten 1 und -1. Wir nehmen an, dass v_2 und v_3 normiert sind. Es folgt, dass $\mathcal{B} = (v_1, v_2, v_3)$ eine Orthonormalbasis ist mit $[m_A]_{\mathcal{B}} = \begin{pmatrix} -1 & & \\ & 1 & \\ & & -1 \end{pmatrix}$, wie wir zeigen wollten. □

Proposition 7.4.16. *Für $A \in \mathrm{O}(3) \setminus \mathrm{SO}(3)$ existiert eine Orthonormalbasis $\mathcal{B}$ von $\mathbb{R}^3$, sodass entweder*

$$[m_A]_{\mathcal{B}} = \begin{pmatrix} 1 & & \\ & -1 & \\ & & 1 \end{pmatrix} \quad \textit{oder} \quad [m_A]_{\mathcal{B}} = \left(\begin{array}{c|c} -1 & \\ \hline & R_\theta \end{array}\right) \textit{ für ein } \theta \in [0, 2\pi).$$

Beweis. Bemerken Sie, dass $-A = (-I_3) \cdot A \in \mathrm{SO}(3)$. Laut Proposition 7.4.15 existiert eine Basis $\mathcal{B}$, sodass entweder

$$[m_A]_{\mathcal{B}} = -[m_{-A}]_{\mathcal{B}} = -\left(\begin{array}{c|c} 1 & \\ \hline & R_\varphi \end{array}\right) = \left(\begin{array}{c|c} -1 & \\ \hline & -R_\varphi \end{array}\right) = \left(\begin{array}{c|c} -1 & \\ \hline & R_{\underbrace{\varphi + \pi}_{=:\theta}} \end{array}\right)$$

oder

$$[m_A]_{\mathcal{B}} = -[m_{-A}]_{\mathcal{B}} = -\begin{pmatrix} -1 & & \\ & 1 & \\ & & -1 \end{pmatrix} = \begin{pmatrix} 1 & & \\ & -1 & \\ & & 1 \end{pmatrix}.$$ □

Satz 7.4.17 (Satz vom Fussball, vergleiche Fischer [7], Seite 307)**.** *Bei einem Fussballspiel, in dem nur ein einziger Ball verwendet wird, existiert ein Punkt auf dem Ball, welcher zu Beginn der zweiten Halbzeit am genau gleichen Ort ist wie zu Beginn der* *ersten Halbzeit (wobei wir annehmen, dass der Ball die Form einer perfekten Kugel hat und bei jedem Anstoss genau in der Mitte des Spielfelds platziert wird).*

Beweis. Sei $A \in \mathrm{O}(3)$ das entsprechende Element zur Bewegung des Fussballs. Im Falle von $\det A = -1$ sehen wir, dass die einzigen zwei Möglichkeiten eine Spiegelung entlang einer Geraden oder die Rotation in einer Ebene zusammen mit einer Spiegelung des orthogonalen Komplements ist. Man kann dies visualisieren und sehen, dass beide Fälle mit einem Fussball nicht möglich sind. Folglich muss $\det A = 1$ sein (ein anderes Argument ist, dass A die Orientierung nicht ändern kann, folglich muss $\det A = 1$ sein, siehe Abschnitt 4.1) und somit $A \in \mathrm{SO}(3)$. In diesem Fall gibt es laut Proposition 7.4.15 immer einen Eigenvektor zum Eigenwert 1, was genau einem Fixpunkt auf dem Fussball entspricht.

Eine andere Begründung, wieso $A \in \mathrm{SO}(3)$ sein muss, geht wie folgt: Wenn wir die Bewegung des Fussballs während dem Spiel genau beobachten und den Ursprung unseres Koordinatensystems immer in der Mitte des Balls behalten, erhalten wir einen Pfad γ von der Identität zu unserer Matrix A (die Identität entspricht dem Zeitpunkt zu Beginn der ersten Hälfte und A entspricht dem Zeitpunkt zu Beginn der zweiten Hälfte). Die Determinante ist eine stetige Abbildung, welche auf $\mathrm{O}(3)$ nur die Werte 1 und -1 annehmen kann. Der Wert der Determinante kann also während des Pfades γ keinen Sprung machen, er bleibt also konstant. Da die Identität Determinante 1 hat folgt, dass auch A Determinante 1 hat und somit in $\mathrm{SO}(3)$ ist. □

7.5 Singulärwertzerlegung

Intro I

$\mathbb{F}$ Wir gehen zurück zur Diskussion von allgemeinen linearen Abbildungen in Skalarprodukträumen. Sei $T : V \to W$ eine lineare Abbildung zwischen zwei endlichdimensionalen Skalarprodukträumen. Wir suchen nach einem Analogon von Proposition 3.3.37. Diese Proposition besagt, dass es Basen $\mathcal{B}$ von V und $\mathcal{C}$ von W gibt, sodass

$$[T]^{\mathcal{B}}_{\mathcal{C}} = D_r = \left(\begin{array}{c|c} I_r & 0_{r,n-r} \\ \hline 0_{m-r,r} & 0_{m-r,n-r} \end{array}\right),$$

wobei $r = \mathrm{Rang}(T)$, $n = \dim(V)$ und $m = \dim(W)$. Das heisst, wenn wir V und W durch Basen unserer Wahl darstellen können, gibt es eine sehr einfache Beschreibung für T. Proposition 3.3.39 ist die Matrixform von Proposition 3.3.37 und besagt, dass für jede Matrix $A \in M_{m\times n}(\mathbb{F})$ mit Rang r Matrizen $P \in \mathrm{GL}_m(\mathbb{F})$ und $Q \in \mathrm{GL}_n(\mathbb{F})$ existieren, sodass

$$PAQ = D_r. \tag{7.7}$$

Erinnern Sie sich (siehe Definition 3.3.35), dass A und D_r wie in (7.7) äquivalent heissen. Die Matrizen D_r, mit $0 \leq r \leq n$, bilden eine „vollständige Menge von Repräsentanten" bezüglich der Äquivalenzrelation „äquivalenz", das heisst jede Äquivalenzklasse ist representiert durch D_r für ein eindeutig bestimmtes[8] r.

In Skalarprodukträumen stellen wir die folgende Frage: Welche Aussage erhalten wir, wenn wir zusätzlich verlangen, dass die Basen $\mathcal{B}$ und $\mathcal{C}$ orthonormal sind? In Matrixform suchen wir ein Analogon von (7.7) mit $P \in \mathrm{O}(m)$ und $Q \in \mathrm{O}(n)$ (falls $\mathbb{F} = \mathbb{R}$) bzw. mit $P \in \mathrm{U}(m)$ und $Q \in \mathrm{U}(n)$ (falls $\mathbb{F} = \mathbb{C}$).

Wir erinnern uns, dass $\mathrm{U}(n)$ und $\mathrm{O}(n)$ kompakt sind. Somit können wir nicht erwarten, dass (7.7) für orthogonale/unitäre P und Q gilt. Das Analogon von D_r muss Einträge haben, welche wachsen können.

Dies kann wieder präziser formuliert werden als ein Problem, eine vollständige Menge von Repräsentanten zu finden bezüglich einer Äquivalenzrelation. Genauer gesagt definieren wir:

Definition 7.5.1.

- Zwei Matrizen $A, B \in M_{m\times n}(\mathbb{R})$ heissen *orthogonal äquivalent*, falls $P \in \mathrm{O}(m)$ und $Q \in \mathrm{O}(n)$ existieren mit $B = PAQ$.
- Zwei Matrizen $A, B \in M_{m\times n}(\mathbb{C})$ heissen *unitär äquivalent*, falls $P \in \mathrm{U}(m)$ und $Q \in \mathrm{U}(n)$ existieren mit $B = PAQ$.

Dies sind Äquivalenzrelationen auf $M_{m\times n}(\mathbb{F})$. Wir fragen uns, ob wir eine vollständige Menge von Repräsentanten bezüglich diesen Äquivalenzrelationen finden können. Die Singulärwertzerlegung wird uns genau dies beantworten.

Intro II

Obwohl das Finden einer einfachen Menge von Repräsentanten für die obige Äquivalenzrelation die ursprüngliche Motivation für die Entdeckung der Singulärwertzerlegung von Eugenio Beltrami und Camille Jordan war, ist die Singulärwertzerlegung bei weitem die nützlichste Matrixzerlegung und es gibt viele Anwendungen davon. Betrachten sie dafür die entsprechende Wikipedia-Seite für einen Überblick sowie die Kurse in Numerik. Ohne weitere Umschweife formulieren wir nun diesen Satz:

Satz 7.5.2 (Singulärwertzerlegung)**.** *Sei $T : V \to W$ eine lineare Abbildung zwischen zwei endlich-dimensionalen Skalarprodukträumen über $\mathbb{F}$ und sei $r = \mathrm{Rang}(T)$. Dann existieren orthonormale Basen $\mathcal{B} = (v_1, \dots, v_n)$ von V und $\mathcal{C} = (w_1, \dots, w_m)$ von W,*

[8] Erinnern Sie sich: $r = \mathrm{Rang}(D_r) = \mathrm{Rang}(A)$.

sowie eindeutige reelle Zahlen $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$*, sodass*

$$\begin{cases} Tv_i = \sigma_i w_i, & i = 1, \ldots, r, \\ Tv_i = 0_W, & i > r. \end{cases}$$

Die σ_i *sind die quadratischen Wurzeln der positiven Eigenwerte von* $TT^* \in \operatorname{End}(W)$ *(oder äquivalent von* $T^*T \in \operatorname{End}(V)$*), geordnet nach ihrer Grösse.*

Übung 7.5.3. Zeigen Sie die folgenden Korollare:

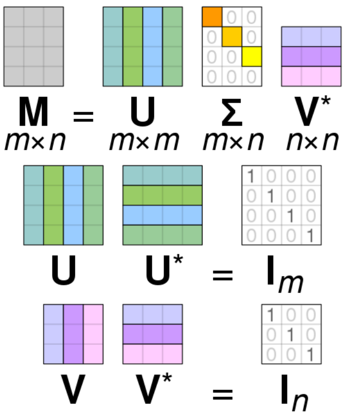
(a) Für $T, \mathcal{B}, \mathcal{C}$ wie in Satz 7.5.2 gilt $[T]^{\mathcal{B}}_{\mathcal{C}} = \left(\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0_{r,n-r} \\ & & \sigma_r & \\ \hline & 0_{m-r,r} & & 0_{m-r,n-r} \end{array}\right)$.

ℂ (b) (Matrixform von Satz 7.5.2) Für jede Matrix $A \in M_{m \times n}(\mathbb{C})$ existieren $U \in \mathrm{U}(m)$ und $V \in \mathrm{U}(n)$ mit

$$A = U \left(\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0_{r,n-r} \\ & & \sigma_r & \\ \hline & 0_{m-r,r} & & 0_{m-r,n-r} \end{array}\right) \overline{V}^T, \tag{7.8}$$

wobei $\sigma_1 \geq \cdots \geq \sigma_r > 0$ die positiven Quadratwurzeln der Eigenwerte von $A\overline{A}^T$ sind.

ℝ (c) Formulieren Sie die analoge Aussage von (b) für $\mathbb{F} = \mathbb{R}$ und beweisen Sie diese.



Figur 7.3: Visualisierung der Singulärwertzerlegung einer Matrix. Quelle: Wikipedia [9].

[9] Attribution: Cmglee, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons

$\mathbb{F}$ **Definition 7.5.4.** Die Zahlen $\sigma_1, \ldots, \sigma_r$ (oder $\sigma_1, \ldots, \sigma_{\min(m,n)}$)[10] heissen die *Singulärwerte* von T.

7.5.1 Beweis von Satz 7.5.2

Wir können natürlich einfach eine orthonormale Basis von V auswählen, aber niemand garantiert uns, dass dann das Bild dieser Basis unter T auch wieder eine orthonormale Basis sein wird. Die Schlüsselidee, um die „richtige" Basis (welche nicht eindeutig ist) zu finden, wird sein, T und T^* gemeinsam zu betrachten. Wir werden am Ende sehen, dass $\mathcal{B}$ eine Orthonormalbasis aus Eigenvektoren von T^*T ist und $\mathcal{C}$ eine Orthonormalbasis aus Eigenvektoren von TT^*. Also müssen wir zuerst TT^* und T^*T besser verstehen.

Lemma 7.5.5. *Sei $T : V \to W$ wie in Satz 7.5.2. Dann gilt:*

(1) $\operatorname{Rang}(T) = \operatorname{Rang}(T^*)$.

(2) $\ker(TT^*) = \ker(T^*)$.

(3) $\operatorname{Im}(TT^*) = \operatorname{Im}(T)$.

(4) TT^* ist positiv semidefinit, das heisst, für alle $w \in W$ gilt $\langle w, TT^*w\rangle \geq 0$.

(5) TT^* ist selbstadjungiert.

(6) Die Eigenwerte von TT^* sind nicht-negative reelle Zahlen.

Beweis. (1): Der Rang einer Abbildung ist gleich dem Rang der Darstellungsmatrix bezüglich einer beliebigen Basis. Seien also $\mathcal{B} \subseteq V$ und $\mathcal{C} \subseteq W$ orthonormale Basen. Es gilt

$$\operatorname{Rang} T = \operatorname{Rang}[T]^{\mathcal{B}}_{\mathcal{C}} = \operatorname{Rang} \overline{[T]^{\mathcal{B}}_{\mathcal{C}}}^T \underset{7.1.6}{\overset{\text{Prop.}}{=}} \operatorname{Rang}[T^*]^{\mathcal{C}}_{\mathcal{B}} = \operatorname{Rang} T^*,$$

wobei die zweite Gleichheit aus Satz 3.4.1 und dem Fakt, dass komplexe Konjugation den Rang einer Matrix nicht ändert, folgt.

(2): Die Inklusion $\ker T^* \subseteq \ker TT^*$ folgt direkt aus der Definition ders Kerns. Sei also $v \in \ker TT^*$. Es gilt

$$0 = \langle 0, v\rangle = \langle TT^*v, v\rangle = \langle T^*v, T^*v\rangle .$$

Daher gilt $T^*v = 0$ und somit $v \in \ker T^*$, was (2) zeigt.

[10] Gewisse Autoren lassen auch 0 als Singulärwert zu, dann ist die Anzahl Singulärwerte immer gleich $\min(m, n)$.

(3): Die Inklusion $\operatorname{Im} T \supseteq \operatorname{Im} TT^*$ folgt ebenfalls direkt aus der Definition des Bilds. Aus dem Rangsatz folgt

$$\begin{aligned}\dim \operatorname{Im} TT^* = \dim W - \dim \ker TT^* \overset{(2)}{=} \dim W - \dim \ker T^* = \operatorname{Rang} T^* \overset{(1)}{=} \operatorname{Rang} T \\ = \dim \operatorname{Im} T.\end{aligned}$$

Somit folgt (3).

(4): Für alle $w \in W$ gilt $\langle w, TT^* w\rangle = \langle T^* w, T^* w\rangle \geq 0$ (gemäss Übung 7.2.6).

(5): Laut Lemma 7.1.3 gilt $(TT^*)^* = (T^*)^* T^* = TT^*$.

(6): Da TT^* selbstadjungiert ist, sind laut Lemma 7.2.13 die Eigenwerte reell. Sei v ein Eigenvektor zum Eigenwert λ. Es gilt

$$0 \overset{(4)}{\leq} \langle v, TT^* v\rangle = \langle v, \lambda v\rangle = \lambda \langle v, v\rangle$$

und daher ist $\lambda \geq \frac{0}{\langle v,v\rangle} = 0$ (wobei wir $\langle v, v\rangle > 0$ verwendet haben, da $v \neq 0$). □

Beweis von Satz 7.5.2. Dank dem Spektralsatz (selbstadjungierte Endomorphismen sind insbesondere normal) können wir eine orthonormale Basis $w_1, \ldots, w_m$ von W finden, die aus Eigenvektoren von TT^* besteht. Gemäss Lemma 7.5.5 (6) sind alle Eigenwerte reell und nicht-negativ. Mit einer Permutation der Basiselemente können wir ohne Beschränkung der Allgemeinheit annehmen, dass die entsprechenden Eigenwerte

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_m = 0$$

erfüllen, wobei $r = \operatorname{Rang}(TT^*) = \operatorname{Rang}(T)$ (laut (3) von Lemma 7.5.5). Damit ist auch

$$\ker T^* \overset{(2)}{=} \ker TT^* = \operatorname{Sp}(w_{r+1}, \ldots, w_m).$$

Wir definieren $\tilde{v}_1 := T^* w_1, \ldots, \tilde{v}_r := T^* w_r$. Diese erzeugen $\operatorname{Im} T^*$:

$$\operatorname{Im} T^* = \operatorname{Sp}(T^* w_1, \ldots, T^* w_r, \underbrace{T^* w_{r+1}}_{=0}, \ldots, \underbrace{T^* w_m}_{=0}) = \operatorname{Sp}(\tilde{v}_1, \ldots, \tilde{v}_r).$$

Ausserdem sind die Vektoren $\tilde{v}_1, \ldots, \tilde{v}_r$ orthogonal zueinander: Für alle $1 \leq i, j \leq r$ gilt

$$\langle \tilde{v}_i, \tilde{v}_j\rangle = \langle T^* w_i, T^* w_j\rangle = \langle w_i, TT^* w_j\rangle = \lambda_j \langle w_i, w_j\rangle = \lambda_j \delta_{ij}.$$

Also bilden $\tilde{v}_1, \ldots, \tilde{v}_r$ eine orthogonale Basis[11] von $\operatorname{Im} T^*$ und $\|\tilde{v}_i\| = \sqrt{\lambda_i} =: \sigma_i$ für $i = 1, \ldots, r$. Weiter ist

$$\operatorname{Sp}(\tilde{v}_1, \ldots, \tilde{v}_r)^\perp = (\operatorname{Im} T^*)^\perp \underset{7.1.5}{\overset{\text{Lemma}}{=}} \ker T,$$

[11] Die lineare Unabhängigkeit wird von der Orthogonalität impliziert, siehe Korollar 6.4.5.

also ist $V = \mathrm{Sp}(\tilde{v}_1, \ldots, \tilde{v}_r) \oplus \ker T$ gemäss Satz 6.6.3. Bemerken Sie, dass

$$\dim \ker T = \dim V - \operatorname{Rang} T = n - r.$$

Sei also $\tilde{v}_{r+1}, \ldots, \tilde{v}_n$ eine orthogonale Basis von $\ker T$. Es folgt, dass $(\tilde{v}_1, \ldots, \tilde{v}_n)$ eine orthogonale Basis von V ist. Sei $\mathcal{B} := \left(v_1 := \frac{\tilde{v}_1}{\|\tilde{v}_1\|}, \ldots, v_n := \frac{\tilde{v}_n}{\|\tilde{v}_n\|}\right)$ die entsprechende orthonormale Basis von V. Es gilt für $1 \leq i \leq r$

$$T(v_i) = T\left(\frac{\tilde{v}_i}{\|\tilde{v}_i\|}\right) = \frac{1}{\sigma_i} T(\tilde{v}_i) = \frac{1}{\sigma_i} T(T^* w_i) = \frac{1}{\sigma_i} (TT^*)(w_i) = \frac{1}{\sigma_i} \lambda_i w_i = \sigma_i w_i$$

und für $i > r$

$$T(v_i) = T\left(\frac{\tilde{v}_i}{\|\tilde{v}_i\|}\right) = \frac{1}{\|\tilde{v}_i\|} T(\tilde{v}_i) = 0_W,$$

wie wir zeigen wollten. Bemerken Sie, dass T die Eigenwerte von TT^* eindeutig bestimmt und somit auch die Singulärwerte $\sigma_1 \geq \cdots \geq \sigma_r$ eindeutig sind. □

Übung 7.5.6. Zeigen Sie, dass die positiven Eigenwerte von TT^* und T^*T gleich sind. Einmal direkt und einmal mithilfe von Satz 7.5.2.

Beispiel 7.5.7. Die Matrix $\begin{pmatrix} 2 & 3 \\ 0 & 2 \end{pmatrix}$ hat die Singulärwertzerlegung

$$\begin{pmatrix} 2 & 3 \\ 0 & 2 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix} \cdot \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}$$

Beispiel 7.5.8. Für jedes $x \in \mathbb{R}$ sind die Singulärwerte der Matrix $\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$ gleich

$$\sqrt{\frac{1}{2}\left(2 + x^2 \pm \sqrt{x^2(4 + x^2)}\right)}.$$

Für $|x| \to \infty$ verhalten sie sich asymptotisch wie $|x|$ und $|x|^{-1}$, während die Faktoren U und V der Singulärwertzerlegung innerhalb der kompakten Menge $\mathrm{O}(2)$ bleiben.

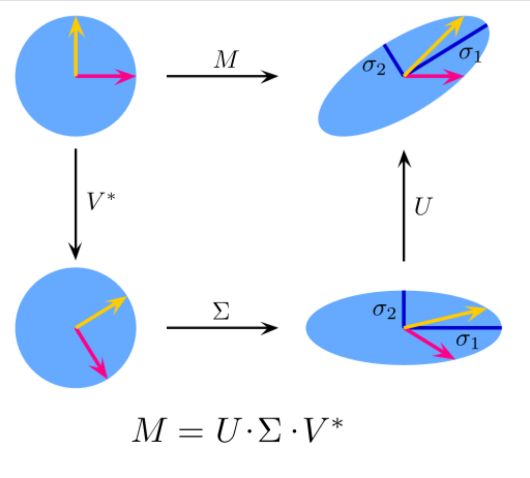
Die geometrische Bedeutung von Satz 7.5.2 ist, dass sich jede lineare Abbildung zwischen Skalarprodukträumen aus orthogonalen/unitären Abbildungen und einer Skalierung zusammensetzen lässt.

$\mathbb{R}$

7.5.2 Bildkompression mithilfe von Singulärwertzerlegung

Zum Schluss dieses Kapitels schauen wir uns noch eine interessante Anwendung der Singulärwertzerlegung an. Wir können damit nämlich Bilder komprimieren. Die

[12]Attribution: Georg-Johann, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons



Figur 7.4: Geometrische Bedeutung der Singulärwertzerlegug einer 2×2-Matrix. Quelle: Wikipedia[12].

Grundidee dabei ist, dass wir ein Bild als eine Matrix anschauen, für diese Matrix eine Singulärwertzerlegung ausführen, dann die kleinen Singulärwerte ignorieren und so die ursprüngliche Matrix (und somit das ursprüngliche Bild) approximieren. Genauer gesagt, sei $A \in M_{m\times n}(\mathbb{R})$, seien $\sigma_1 \geq \cdots \geq \sigma_r > 0$ die Singulärwerte von A und seien $U \in \mathrm{O}(m)$ und $V \in \mathrm{O}(n)$, sodass

$$A = U \left(\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0_{r,n-r} \\ & & \sigma_r & \\ \hline & 0_{m-r,r} & & 0_{m-r,n-r} \end{array}\right) V^T.$$

Seien $u_1, \ldots, u_m$ die Spaltenvektoren von U und seien $v_1, \ldots, v_n$ die Zeilenvektoren von V^T. Wir haben also

$$A = \begin{pmatrix} | & & | \\ u_1 & \cdots & u_m \\ | & & | \end{pmatrix} \left(\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0_{r,n-r} \\ & & \sigma_r & \\ \hline & 0_{m-r,r} & & 0_{m-r,n-r} \end{array}\right) \begin{pmatrix} — & v_1 & — \\ & \vdots & \\ — & v_n & — \end{pmatrix},$$

woraus ersichtlich wird, dass wir die Spalten $u_{r+1}, \ldots, u_m$ von U und die Zeilen $v_{r+1}, \ldots, v_n$ von V^T „wegwerfen“ können, da diese sowieso mit Null multipliziert werden. Ausserdem können wir auch gleich alle Nullen in der mittleren Matrix loswerden:

$$A = \begin{pmatrix} | & & | \\ u_1 & \cdots & u_r \\ | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix} \begin{pmatrix} — & v_1 & — \\ & \vdots & \\ — & v_r & — \end{pmatrix}.$$

Wir können nun aber noch weiter gehen, und ab hier stellen wir die Matrix A nur noch näherungsweise dar. Da die Singulärwerte der Grösse nach geordnet sind, sollte es nicht allzu viel ändern, wenn wir die letzten paar Singulärwerte durch Null approximieren. In diesem Fall können wir dann auch (aus dem gleichen Grund wie zuvor) die letzten paar Spalten von U sowie die untersten paar Zeilen von V^T wegwerfen. Wenn wir also ein $0 \leq k \leq r$ auswählen, bekommen wir eine Annäherung für A:

$$A \approx \begin{pmatrix} | & & | \\ u_1 & \cdots & u_k \\ | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{pmatrix} \begin{pmatrix} — & v_1 & — \\ & \vdots & \\ — & v_k & — \end{pmatrix}.$$

Bemerken Sie, dass insbesondere die Dimensionen von A und von dem Matrixprodukt auf der rechten Seite übereinstimmen. Je höher wir k wählen, desto besser wird A approximiert. Je tiefer wir k wählen, desto weniger Speicherplatz benötigt unsere Annäherung. Genauer gesagt: Für die Matrix A müssen wir $m \cdot n$ Einträge speichern, für die Annäherung $n \cdot k + k + m \cdot k$. Wie viel Speicher wir dabei sparen, hängt von n, m und k ab. Zum Beispiel für $n = m$ und $k = \frac{n}{10}$ erhalten wir n^2 Einträge für A und $\frac{n^2}{10} + \frac{n}{10} + \frac{n^2}{10} = \frac{n^2}{5} + \frac{n}{10}$ Einträge für die Annäherung, also wird die Grösse auf etwa einen Fünftel verkleinert.

Nun wollen wir diese Überlegungen auf die Bildkompression anwenden. Betrachten wir dazu als Beispiel das folgende Bild des ETH-Hauptgebäudes:

Figur 7.5: Das ETH-Hauptgebäude

Die Auflösung dieses Bilds beträgt 564×240 Pixel. Jedes Pixel wird repräsentiert durch drei Zahlen zwischen 0 und 255, welche den Rot-, Grün- und Blauanteil dieses Pixels beschreiben. Wir können das Bild also auf dem Computer speichern als drei Matrizen $R, G, B \in M_{240\times 564}(\mathbb{R})$. Wir können diese drei Matrizen jeweils mit dem oben beschriebenen Verfahren komprimieren und dann wieder zu einem Bild zusammensetzen. Dadurch erhalten wir eine Kompression des Bilds. Für das Originalbild benötigt man $3 \cdot 564 \cdot 240$ Zahlen. Wenn wir für die Matrizen R, G, B eine Singulärwertzerlegung ausführen und jeweils die grössten k Singulärwerte behalten, so müssen wir noch $3 \cdot (564k + k + 240k) = 3 \cdot 805k$ Zahlen speichern. Wir schauen uns nun die komprimierten Bilder für verschiedene Werte von k an:

$k = 1$

$k = 2$

$k = 4$

$k = 8$

$k = 12$

$k = 20$

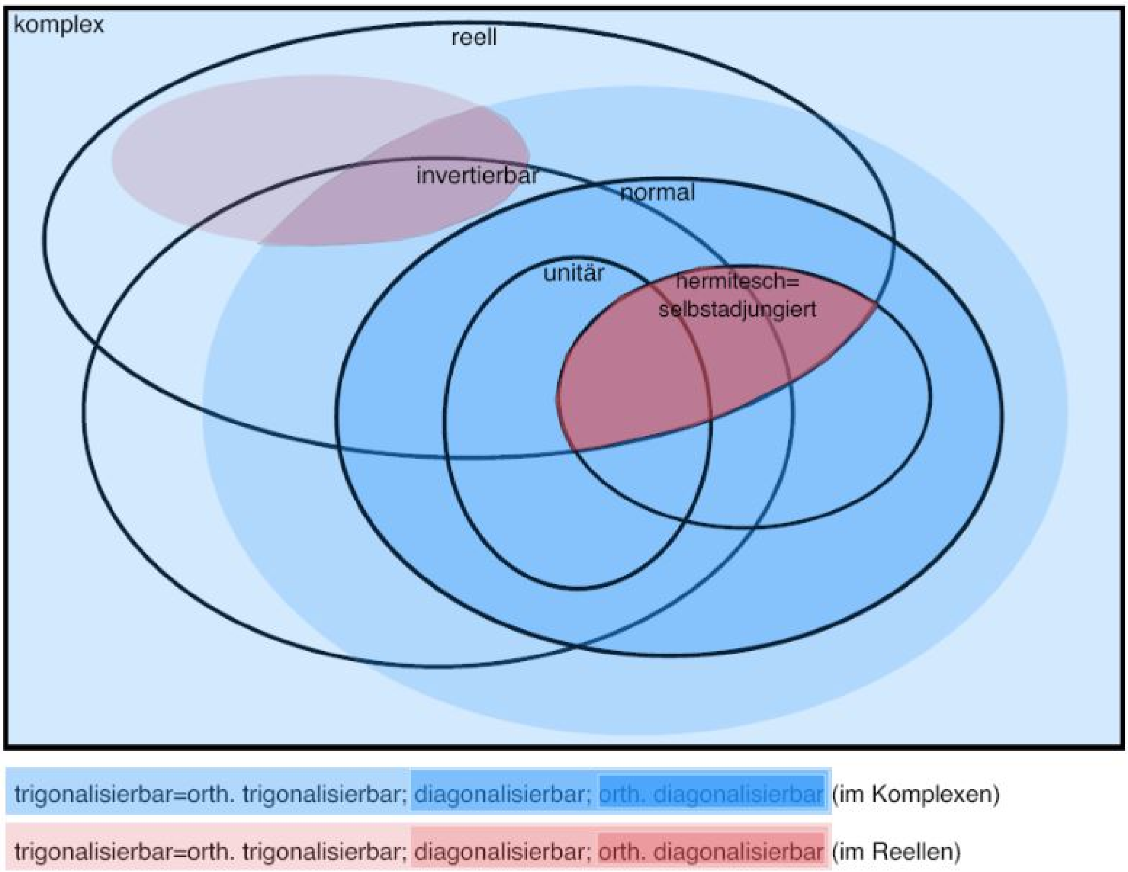
$k = 50$

$k = 100$

Das Bild mit 100 Singulärwerten ist bereits eine ziemlich gute Annäherung an das Original und es benötigt $\frac{3 \cdot 805 \cdot 100}{3 \cdot 564 \cdot 240} \approx 0.6$ mal so viel Speicherplatz wie das Originalbild. Für grössere Bilder (z.B. 1920×1080) genügen 100 Singulärwerte oft auch und in diesem Fall ergibt das eine noch bessere Kompressionsrate. Für eine ausführlichere Beschreibung dieser Anwendung können Sie das Dokument Image Compression using Singular Value Decomposition (SVD) von Brady Mathews lesen. Dort ist auch der Matlab Code zu finden, den wir für die Kompression von unserem Beispielbild benutzt haben.

Das Erstaunliche an dieser Anwendung ist, dass wir Matrizen normalerweise als Beschreibung einer Transformation zwischen Vektorräumen anschauen. Die Matrix hier stammt aber von einem Bild, was nichts mit einer solchen Transformation zu tun hat.

Trotzdem können wir uns die Theorie der Linearen Algebra zunutze machen, um dieses Bild zu bearbeiten.



Figur 7.6: Diese Grafik wurde von Fabian Arnold erstellt.

$\mathbb{R} \leftrightarrow \mathbb{C}$ **Übung 7.5.9.** [Matrizenmengen] Verstehen Sie das Diagramm 7.6, das heisst:

(a) Verstehen und beweisen[13] Sie alle Inklusionen (zum Beispiel, dass alle unitären Matrizen normal sind usw.).

(b) Finden Sie Matrizen in allen möglichen Schnittmengen (z.B. finden Sie Matrizen A_1, A_2, A_3, sodass A_1 invertierbar und normal, A_2 invertierbar und nicht normal, A_3 normal und nicht invertierbar ist).

(c) Bestimmen Sie für weitere Matrizenmengen (z.B. reelle und komplexe 1×1-Matrizen), wo im Diagram sich diese befinden.

[13]Einige der Inklusionen sind Sätze die wir bewiesen haben, zum Beispiel ist die Aussage des Spektralsatzes, dass die normalen Matrizen genau die komplexen orthogonal diagonalisierbaren Matrizen sind. Die Übung besteht in diesem Fall darin, die entsprechenden Aussagen ausfindig zu machen.

Changelog: Kapitel 7

- 24.03: Im Beweis von Lemma 7.1.3 (2) wurde an einer Stelle $(S+T^*)w$ zu $(S+T)^*w$ korrigiert.
- 24.03: Im Spektralsatz über $\mathbb{C}$ (Satz 7.2.3) wurde spezifiziert, dass der Satz nur für endlich-dimensionale Vektorräume gilt.
- 26.03: Bemerkung 7.1.4 wurde hinzugefügt.
- 26.03: In Lemma 7.2.5 (3) wurde das Eigenwert zu Eigenvektor korrigiert.
- 26.03: Im letzten Satz des Beweises von Lemma 7.1.3 (6) wurde $T^* \circ S^*$ zu $T^* \circ R^*$ korrigiert.
- 28.03: Im Beweis von Lemma 7.2.5 (2) wurde in der zweiten Gleichung $(T^* - \overline{\lambda}\,\mathrm{Id}_V)(T^* - \overline{\lambda}\,\mathrm{Id}_V)$ zu $(T^* - \overline{\lambda}\,\mathrm{Id}_V)(T - \lambda\,\mathrm{Id}_V)$ korrigiert.
- 30.03: Der Titel von Abschnitt 7.3 wurde von „Matrizen“ zu „Spektralsätze für Matrizen“ geändert.
- 31.03: In Beispiel 7.3.4 wurde $f : \mathbb{R}^n \to \mathbb{R}^n$ zu $f : \mathbb{R}^n \to \mathbb{R}$ und $f(x - x^0)$ zu $f(x)$ korrigiert.
- 09.04: Im letzten Satz des Beweises von Satz 7.2.12 wurde die Referenz von Lemma 7.1.3 zu Lemma 7.2.13 korrigiert.
- 09.04: In Theorem 7.0.3 (4) wurde $\overline{P}^T AP$ zu $P^T AP$ korrigiert.
- 09.04: Übung 7.2.17 wurde hinzugefügt.
- 15.04: Figur 7.3 wurde hinzugefügt.
- 15.04: Die Bildbeschreibung von Figur 7.4 wurde geändert.
- 17.04: In Satz 7.5.2 wurden $TT^* \in \mathrm{End}(V)$ und $T^*T \in \mathrm{End}(W)$ zu $TT^* \in \mathrm{End}(W)$ und $T^*T \in \mathrm{End}(V)$ korrigiert. Ausserdem wurde „die positiven quadratischen Wurzeln der Eigenwerte“ zu „die quadratischen Wurzeln der positiven Eigenwerte“ umformuliert.
- 17.04: Im Beweis von Satz 7.5.2 wurde $\mathrm{Sp}(\tilde{v}_1, \dots, \tilde{v}_n)$ zu $\mathrm{Sp}(\tilde{v}_1, \dots, \tilde{v}_r)$ korrigiert.
- 17.04: In Definition 7.4.9 wurde an einer Stelle $M_{n\times n}(\mathbb{R})$ zu $M_{n\times n}(\mathbb{C})$ korrigiert und die Notation $|A|$ zu $\det A$ geändert.
- 17.04: In Übung 7.5.6 wurde das Wort „positiven“ eingefügt.

- 18.04: In Satz 7.5.2 wurde hinzugefügt, dass $\mathcal{B}$ und $\mathcal{C}$ orthonormale Basen sind.
- 23.04: Auf Seite 302 wurde „orthogonalen“ zu „orthogonalen/unitären“ korrigiert.
- 27.04: Abschnitt 7.4.1 wurde hinzugefügt.
- 05.05: Figur 7.6 und Übung 7.5.9 wurden hinzugefügt.
- 25.06: Im letzten Satz von Proposition 7.4.15 wurde das Wort „Eigenvektor“ zu „Eigenwert“ korrigiert.
- 27.06: Im Beweis von Proposition 7.4.16 wurde θ durch φ ersetzt.
- 14.08: Im Beweis von Korollar 7.4.14 wurde das Wort „Eigenvektoren“ zu „Eigenwerten“ korrigiert.

Kapitel 8

Bilinearformen und quadratische Formen

Einleitung[1]

Nachdem wir nun schon Experten im Umgang mit linearen Abbildungen sind, ist das Studium von bilinearen Abbildungen der nächste Schritt. In dieser Hinsicht dienen bilineare Abbildungen als Brücke zwischen der linearen Algebra und unserem letzten Kapitel, der multilinearen Algebra. Der Ablauf unserer Reise mit Bilinearformen ist sehr ähnlich zu unserer Reise mit linearen Abbildungen; nachdem wir sie definiert haben, werden wir sehen, wie wir sie bezüglich verschiedenen Basen darstellen können. Dann suchen wir wieder besondere (oder besonders einfache) Darstellungen. Es stellt sich heraus, dass die Struktur aller symmetrischen Bilinearformen ziemlich simpel ist (wenn wir zum Beispiel über einem Körper mit Charakteristik $\neq 2$ arbeiten[2], sind alle symmetrischen Bilinearformen diagonalisierbar!), womit sich die Frage nach einer Klassifizierung aller symmetrischen Bilinearformen über einem gegebenen Körper stellt. Ein Meilenstein ist die Klassifikation über $\mathbb{R}$: Dies ist Sylvesters Trägheitssatz.

Bilinearformen haben noch einen Zwilling: Quadratische Formen. Über Körpern mit Charakteristik $\neq 2$ ist es in einem gewissen Sinne äquivalent, quadratische Formen oder symmetrische Bilinearformen zu verstehen. Insbesondere klassifiziert Sylvesters Trägheitssatz alle quadratischen Formen über $\mathbb{R}$. Ich persönlich liebe quadratische Formen, da ihre Klassifizierung über $\mathbb{Q}$ ein wunderschöner Teil der Zahlentheorie ist, welcher die Entwicklung der modernen Zahlentheorie stark beeinflusst hat. Die Leser, die es nicht erwarten können, mehr über dieses Thema zu erfahren, können zum Beispiel eines meiner Lieblingsbücher, „Rational Quadratic Forms" von Cassels [6], lesen.

[1] Wir hören nun wieder auf mit den Markierungen auf der linken Seite, die wir zu Beginn von Kapitel 6 eingeführt haben.

[2] Falls Sie nicht mehr wissen, was die Charakteristik eines Körpers ist, schauen Sie es in Definition 1.3.18 nach.

Bevor wir beginnen halten wir fest, dass wir bereits einige wichtige Beispiele von Bilinearformen und ihren zugehörigen quadratischen Formen gesehen haben: Jedes Skalarprodukt über $\mathbb{R}$ ist eine symmetrische Bilinearform und die dazugehörige quadratische Form ist nichts anderes als die induzierte Norm im Quadrat.

Bilineaformen sind also eine Verallgemeinerung von Skalarprodukten, doch wieso brauchen wir das? Es stellt sich heraus, dass es in manchen Situationen nötig ist, analoge Versionen von Skalarprodukten zu betrachten, in welchen die Positiv-Definitheit nicht gilt: Ein zentrales Beispiel dafür ist Einsteins spezielle Relativitätstheorie, wo ein nicht positiv definites „Skalarprodukt" auf $\mathbb{R}^4$ betrachtet wird (die dazugehörige „Norm" ist $x_1^2 - x_2^2 - x_3^2 - x_4^2$). Wir werden leider keine Zeit haben, weiter auf dieses Beispiel einzugehen. Die interessierten Leser können einen Blick auf [12, Abschnitt 6.7] werfen. Diese Einleitung ist bereits viel zu lang, also fangen wir nun an!

8.1 Bilinearformen

Sei V ein Vektorraum über einem Körper K.

Definition 8.1.1. Eine *Bilinearform* auf V ist eine Funktion $B : V \times V \to K$, sodass

(1) (Linear in der ersten Variable) Für alle $\alpha, \beta \in K$ und $v_1, v_2, v \in V$ ist

$$B(\alpha v_1 + \beta v_2, v) = \alpha B(v_1, v) + \beta B(v_2, v)$$

(2) (Linear in der zweiten Variable) Für alle $\alpha, \beta \in K$ und $v_1, v_2, v \in V$ ist

$$B(v, \alpha v_1 + \beta v_2) = \alpha B(v, v_1) + \beta B(v, v_2)$$

Eine Funktion mit diesen Eigenschaften heisst *bilinear*. Falls zusätzlich gilt

$$\forall v, w \in V \quad B(v, w) = B(w, v),$$

so nennen wir B eine *symmetrische Bilinearform.*

Beispiel 8.1.2 (Mutterbeispiel)**.** Sei $A \in M_{n\times n}(K)$ und wie zuvor identifizieren wir $M_{1\times 1}(K)$ mit K. Die Funktion

$$\begin{aligned} B_A : K^n \times K^n &\to K \\ (v, w) &\mapsto v^T A w \end{aligned}$$

ist eine Bilinearform auf K^n. Sie ist genau dann symmetrisch, wenn A symmetrisch ist. Vergleichen Sie dieses Beispiel mit Beispiel 6.2.10.

Beispiel 8.1.3. Jedes Skalarprodukt auf einem $\mathbb{R}$-Vektorraum ist eine symmetrische Bilinearform, aber nicht umgekehrt. Zum Beispiel ist $B_{\begin{pmatrix} 1 & \\ & -1 \end{pmatrix}}$ eine symmetrische Bilinearform auf $\mathbb{R}^2$, die kein Skalarprodukt ist.

Skalarprodukte auf $\mathbb{C}$-Vektorräumen sind nicht bilinear, sondern sesquilinear. Viele Dinge, die wir in diesem Kapitel beschreiben werden, besitzen ein analoge Versionen für Sesquilinearformen, welche wir manchmal auch beschreiben werden.

Schauen wir zurück, wie wir lineare Abbildungen oder genauer gesagt Endomorphismen $T : V \to V$ für endlich-dimensionale K-Vektorräume V untersucht haben: Wir haben zuerst bemerkt, dass das Hauptbeispiel eines Endomorphismus von K^n die Multiplikation mit einer Matrix A ist. Dies war Beispiel 3.1.13, welches wir ebenfalls „Mutterbeispiel“ genannt haben. Mittels Wahl einer Basis $\mathcal{B}$ von V und dem entsprechenden Isomorphismus $\Phi_{\mathcal{B}} : V \to K^n$, $v \mapsto [v]_{\mathcal{B}}$ konnten wir T durch eine Matrix $[T]_{\mathcal{B}}$ darstellen, im Sinne von Proposition 3.3.8

$$\forall v \in V \quad [Tv]_{\mathcal{B}} = [T]_{\mathcal{B}}[v]_{\mathcal{B}}.$$

In diesem Sinne haben wir gesagt, dass alle linearen Abbildungen die Multiplikation mit einer Matrix sind (oder genauer gesagt, alle linearen Abbildungen können durch die Multiplikation mit einer Matrix repräsentiert werden). Später haben wir dann auch gezeigt, dass jede Wahl einer Basis $\mathcal{B}$ von V einen Isomorphismus von Vektorräumen[3]

$$\begin{aligned} \mathrm{End}(V) &\to M_{n\times n}(K) \\ T &\mapsto [T]_{\mathcal{B}} \end{aligned}$$

induziert. Wir möchten nun das gleiche für Bilinearformen machen. Wir könnten es wie in Kapitel 3 machen: Zuerst zeigt man, dass jede Bilinearform auf K^n von der Form wie in Beispiel 8.1.2 ist. Dann kann man für eine gegebene Bilinearform $B : V \times V \to K$ auf einem n-dimensionalen Vektorraum V die Bilinearform

$$\begin{aligned} \tilde{B} : K^n \times K^n &\to K \\ (x, y) &\mapsto \tilde{B}(x, y) := B(\Phi_{\mathcal{B}}^{-1}(x), \Phi_{\mathcal{B}}^{-1}(y)) \end{aligned}$$

betrachten. Diese Bilinearform $\tilde{B}$ wird dann durch eine Matrix repräsentiert wie in Beispiel 8.1.2. Dieses Mal nehmen wir aber einen kürzeren Weg und definieren direkt die Darstellungsmatrix und beweisen ihre Eigenschaften.

Definition 8.1.4. Sei B eine Bilinearform auf einem endlich-dimensionalen K-Vektorraum V und sei $\mathcal{B} = (v_1, \dots, v_n)$ eine Basis von V. Die Darstellungsmatrix $M_{\mathcal{B}}(B)$ von B bezüglich $\mathcal{B}$ ist

$$M_{\mathcal{B}}(B) := (B(v_i, v_j))_{1\le i,j\le n}.$$

[3] Eigentlich ist es sogar ein Isomorphismus von Ringen, aber das ist jetzt irrelevant.

Hier ist ein Analogon von Beispiel 3.3.9:

Beispiel 8.1.5. Sei $A \in M_{n\times n}(K)$ und $\mathcal{E}_n = (e_1, \dots, e_n)$ die Standard-Basis von K^n. Es gilt

$$M_{\mathcal{E}_n}(B_A) = (B_A(e_i, e_j))_{1\le i,j\le n} = \left(e_i^T A e_j\right)_{1\le i,j\le n} = (a_{ij})_{1\le i,j\le n} = A.$$

Definition 8.1.6. Wir bezeichnen mit $\mathrm{Bil}(V)$ die Menge aller Bilinearformen auf V. Diese Menge besitzt eine natürliche Vektorraumstruktur mit der folgenden Addition und Skalarmultiplikation:

$$\forall B_1, B_2 \in \mathrm{Bil}(V)\ \forall v, w \in V \quad (B_1 + B_2)(v, w) := B_1(v, w) + B_2(v, w),$$

$$\forall \alpha \in K\ \forall B \in \mathrm{Bil}(V)\ \forall v, w \in V \quad (\alpha B)(v, w) := \alpha B(v, w).$$

Wir überlassen es den Lesern, die Vektorraumaxiome zu überprüfen.

Proposition 8.1.7. *Mit B, V und $\mathcal{B}$ wie in Definition 8.1.4 gilt:*

(1) $M_{\mathcal{B}}(B)$ ist die eindeutige Matrix mit der Eigenschaft, dass

$$\forall v, w \in V \quad B(v, w) = [v]_{\mathcal{B}}^T M_{\mathcal{B}}(B)[w]_{\mathcal{B}}. \tag{8.1}$$

(2) Die Abbildung

$$\begin{aligned} \Psi_{\mathcal{B}} : \mathrm{Bil}(V) &\to M_{n\times n}(K) \\ B &\mapsto M_{\mathcal{B}}(B) \end{aligned}$$

ist ein linearer Isomorphismus von Vektorräumen.

(3) B ist genau dann symmetrisch, wenn $M_{\mathcal{B}}(B)$ symmetrisch ist.

Beweis. (1): Sei also $N = (n_{ij})_{i,j}$ eine Matrix mit der Eigenschaft (8.1) bezüglich $\mathcal{B} = (v_1, \dots, v_n)$. Wir haben dann für $1 \le i, j \le n$

$$n_{ij} = e_i^T N e_j = [v_i]_{\mathcal{B}}^T N [v_j]_{\mathcal{B}} \overset{(8.1)}{=} B(v_i, v_j). \tag{8.2}$$

Daher ist $N = M_{\mathcal{B}}(B)$, was die Eindeutigkeit zeigt. Es bleibt zu zeigen, dass $M_{\mathcal{B}}(B)$ die Eigenschaft (8.1) hat. Seien $v = \sum_{i=1}^n a_i v_i$ und $w = \sum_{j=1}^n b_j v_j$ beliebige Vektoren in V. Dann gilt

$$\begin{aligned} B(v, w) &= B\left(\sum_{i=1}^n a_i v_i, \sum_{j=1}^n b_j v_j\right) = \sum_{i=1}^n \sum_{j=1}^n a_i b_j B(v_i, v_j) \\ &\underset{\text{Mult.}}{\overset{\text{Matrix}}{=}} (a_1, \dots, a_n) \Big(B(v_i, v_j)\Big)_{i,j} \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = [v]_{\mathcal{B}}^T M_{\mathcal{B}}(B)[w]_{\mathcal{B}}. \end{aligned}$$

(2): Aus der Eigenschaft (8.1) folgt, dass zwei Bilinearformen mit der gleichen Darstellungsmatrix gleich sein müssen, also ist $\Psi_{\mathcal{B}}$ injektiv. Betrachten wir eine Matrix $N \in M_{n\times n}(K)$. Die Funktion $C_N(v,w) := [v]_{\mathcal{B}}^T N[w]_{\mathcal{B}}$ ist eine bilineare Funktion auf V und laut der Eindeutigkeit aus (1) gilt $N = M_{\mathcal{B}}(C_N) = \Psi_{\mathcal{B}}(C_N)$, was die Surjektivität von $\Psi_{\mathcal{B}}$ zeigt. Also ist $\Psi_{\mathcal{B}}$ bijektiv. Wir überlassen die Linearität und (3) den Lesern, als eine Wiederholung des Stoffs aus dem ersten Semester. □

Korollar 8.1.8. *Sei $B \in \mathrm{Bil}(V)$ und seien $\mathcal{B}, \mathcal{C}$ zwei geordnete Basen von V. Dann gilt*

$$M_{\mathcal{C}}(B) = \left([\mathrm{Id}_V]_{\mathcal{B}}^{\mathcal{C}}\right)^T M_{\mathcal{B}}(B)[\mathrm{Id}_V]_{\mathcal{B}}^{\mathcal{C}}.$$

Beweis. Es genügt zu zeigen, dass die rechte Seite die Eigenschaft (8.1) bezüglich der Basis $\mathcal{C}$ hat: Für $v, w \in V$ gilt

$$\begin{aligned}
[v]_{\mathcal{C}}^T \left([\mathrm{Id}_V]_{\mathcal{B}}^{\mathcal{C}}\right)^T M_{\mathcal{B}}(B)[\mathrm{Id}_V]_{\mathcal{B}}^{\mathcal{C}}[w]_{\mathcal{C}} &= \left([\mathrm{Id}_V]_{\mathcal{B}}^{\mathcal{C}}[v]_{\mathcal{C}}\right)^T M_{\mathcal{B}}(B)[\mathrm{Id}_V]_{\mathcal{B}}^{\mathcal{C}}[w]_{\mathcal{C}} \\
&\overset{\text{Prop.}}{\underset{3.3.8}{=}} [v]_{\mathcal{B}}^T M_{\mathcal{B}}(B)[w]_{\mathcal{B}} \\
&= B(v,w).
\end{aligned}$$
□

Da wir die Dimension von $M_{n\times n}(K)$ kennen, erhalten wir als direkte Folgerung auch die Dimension von $\mathrm{Bil}(V)$:

Korollar 8.1.9. *Falls* $\dim V = n \in \mathbb{N} \cup \{0\}$, *dann ist* $\dim \mathrm{Bil}(V) = n^2$.

Beispiel 8.1.10. Sei $V = \mathbb{R}[x]_2$ und betrachten wir die Bilinearform

$$\begin{aligned}
B : V \times V &\to \mathbb{R} \\
(p,q) &\mapsto (pq)(-1).
\end{aligned}$$

Bezüglich der standard-Basis $\mathcal{B} = (1, x, x^2)$ von V ist $M_{\mathcal{B}}(B) = \begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix}$, da $B(x^i, x^j) = x^{i+j}(-1) = (-1)^{i+j}$. Laut (8.1) gilt für $p = ax^2+bx+c$ und $q = dx^2+ex+f$

$$pq(-1) = B(p,q) = \begin{pmatrix} c & b & a \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} f \\ e \\ d \end{pmatrix}.$$

Bezüglich der Basis $\mathcal{C} = (1, x+1, x^2-1)$ gilt $M_{\mathcal{C}}(B) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, wie man direkt berechnen kann oder durch Korollar 8.1.8:

$$M_{\mathcal{C}}(B) = \left([\mathrm{Id}_V]^{\mathcal{C}}_{\mathcal{B}}\right)^T M_{\mathcal{B}}(B)[\mathrm{Id}_V]^{\mathcal{C}}_{\mathcal{B}} = \begin{pmatrix} 1 & 1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^T \begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Definition 8.1.11. Zwei Matrizen $A, B \in M_{n\times n}(K)$ heissen *kongruent* (zu einander), falls eine invertierbare Matrix $P \in \mathrm{GL}_n(K)$ existiert, sodass $P^T A P = B$.

Die folgenden Tatsachen sind auch eine gute Wiederholung des Stoffs von LAI, daher geben wir sie als Übungen:

Übung 8.1.12. Zeigen Sie, dass Kongruenz eine Äquivalenzrelation auf $M_{n\times n}(K)$ ist.

Die Bedeutung von Korollar 8.1.8 ist:

Übung 8.1.13. Sei V ein n-dimensionaler Vektorraum. Zwei Matrizen in $M_{n\times n}(K)$ sind genau dann kongruent, wenn sie die selbe Bilinearform auf V darstellen (bezüglich verschiedenen Basen).

Lösung. Die Richtung $\Longleftarrow$ ist Korollar 8.1.8. Für $\Longrightarrow$ seien $S, T \in M_{n\times n}(K)$ mit $S = P^T T P$ für $P \in \mathrm{GL}_n(K)$. Sei $\mathcal{B}$ eine Basis von V und sei B eine Bilinearform auf V mit $T = M_{\mathcal{B}}(B)$. Sei $\mathcal{C}$ eine Basis von V, sodass $P = [\mathrm{Id}_V]^{\mathcal{C}}_{\mathcal{B}}$ (wieso existiert solch eine Basis $\mathcal{C}$? Man nimmt $\mathcal{C} = (w_1, \ldots, w_n)$ mit $w_j = \sum_{i=1}^n p_{ij} v_i$, wobei $P = (p_{ij})$ und $\mathcal{B} = (v_1, \ldots, v_n)$.). Dann gilt laut Korollar 8.1.8, dass $S = M_{\mathcal{C}}(B)$. □

Ähnlichkeit und Kongruenz sind zwei verschiedene Äquivalenzrelationen auf $M_{n\times n}(K)$! Sogar auf $M_{1\times 1}(K)$ gilt

$$(a) \overset{\text{kong.}}{\sim} (b) \iff a = b\alpha^2 \text{ für ein } \alpha \in K^\times = K \setminus \{0\}$$

aber

$$(a) \overset{\text{ähnl.}}{\sim} (b) \iff a = b.$$

Es ist auch einfach, Beispiele in $M_{n\times n}(K)$ für alle $n \in \mathbb{N}$ zu finden.

Beispiel 8.1.14. Sei $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ und betrachten wir die Bilinearform B_A auf K^2. Sei $\mathcal{B}_{a,b} = (ae_1, be_2)$, wobei (e_1, e_2) die Standard-Basis von K^2 ist. Um $M_{\mathcal{B}_{a,b}}(B_A)$ zu finden, könnte man entweder direkt

$$\begin{pmatrix} B_A(ae_1, ae_1) & B_A(ae_1, be_2) \\ B_A(be_2, ae_1) & B_A(be_2, be_2) \end{pmatrix} = \begin{pmatrix} a^2 & \\ & -b^2 \end{pmatrix}$$

berechnen, oder mit Korollar 8.1.8 die Matrixmultiplikation

$$M_{\mathcal{B}_{a,b}}(B_A) = \left([\mathrm{Id}]_{\mathcal{E}_n}^{\mathcal{B}_{a,b}}\right)^T M_{\mathcal{E}_n}(B_A)[\mathrm{Id}]_{\mathcal{E}_n}^{\mathcal{B}_{a,b}} = \begin{pmatrix} a & \\ & b \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} a & \\ & b \end{pmatrix} = \begin{pmatrix} a^2 & \\ & -b^2 \end{pmatrix}$$

ausführen. Somit sind alle Matrizen der Form $\left(\begin{smallmatrix} a^2 & \\ & -b^2 \end{smallmatrix}\right)$ zueinander kongruent, denn sie stellen B_A bezüglich der Basis $\mathcal{B}_{a,b}$ dar. Wenn man sich erinnert, dass ähnliche Matrizen insbesondere dieselben Eigenwerte haben, sieht man, dass diese Matrizen normalerweise nicht ähnlich zueinander sind.

Bezüglich der Basis $\mathcal{C} = \left((1,1)^T, (1,-1)^T\right)$ von $\mathbb{R}^2$ ist $M_{\mathcal{C}}(B_A) = \left(\begin{smallmatrix} 0 & 2 \\ 2 & 0 \end{smallmatrix}\right)$ auch nicht diagonal, und bezüglich $\mathcal{C}_a = \left((a,a)^T, (a,-a)^T\right)$ ist $M_{\mathcal{C}_a}(B_A) = \left(\begin{smallmatrix} 0 & 2a^2 \\ 2a^2 & 0 \end{smallmatrix}\right)$. Merken Sie folgendes:

1. In allen diagonalen Darstellungsmatrizen von B_A, die wir gefunden haben, gibt es genau einen positiven Eintrag und einen negativen Eintrag.

2. Alle Darstellungsmatrizen, die wir bisher gefunden haben, haben eine negative Determinante.

Könnte man auch andere Darstellungsmatrizen finden mit einer positiven Determinante oder sogar diagonale Matrizen mit zwei positiven Einträgen? Die folgende Übung hilft uns, dies zu beantworten:

Übung 8.1.15.

(a) Falls $A, B \in M_{n\times n}(K)$ kongruent zueinander sind, dann existiert $a \in K^\times$, sodass $\det B = a^2 \det A$.

(b) Sprachliche Übung: Man beschreibt (a) so: „Die Determinante einer Kongruenzklasse von Matrizen ist wohldefiniert bis auf Multiplikation mit Quadratzahlen in $K^\times$.“ Verstehen Sie, was mit diesem Satz gemeint ist.

(c) Skalare in $\mathbb{R}$ bis auf Quadrate sind: Positive Zahlen $\{a^2 \mid a \in \mathbb{R}\}$, negative Zahlen $\{-1 \cdot a^2 \mid a \in \mathbb{R}\}$ und $\{0\} = \{0 \cdot a^2 \mid a \in \mathbb{R}\}$. Überprüfen Sie dies und folgern Sie, dass $\left(\begin{smallmatrix} 1 & \\ & 0 \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} 1 & \\ & -1 \end{smallmatrix}\right)$ und $\left(\begin{smallmatrix} 1 & \\ & 1 \end{smallmatrix}\right)$ Repräsentanten von verschiedenen Kongruenzklassen sind. Gibt es Kongruenzklassen, die nicht durch eine dieser drei Matrizen konjugiert sind? Falls ja, finden Sie Repräsentanten in Diagonalform.

(d) Über $\mathbb{R}$ haben wir gezeigt, dass $\left(\begin{smallmatrix} 1 & \\ & -1 \end{smallmatrix}\right)$ und $\left(\begin{smallmatrix} 0 & 2 \\ 2 & 0 \end{smallmatrix}\right)$ kongruent sind. Über $\mathbb{F}_2$ ist aber $\left(\begin{smallmatrix} 1 & \\ & -1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 & \\ & 1 \end{smallmatrix}\right)$ und $\left(\begin{smallmatrix} 0 & 2 \\ 2 & 0 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right)$. Kann es sein, dass beide Matrizen kongruent zueinander sind? Falls nicht, was haben wir falsch gemacht?!

Benutzen Sie für die Lösung dieser Aufgabe nur den obigen Stoff. Später sollten Sie sie mit Sylvesters Trägheitssatz vergleichen.

Bevor wir weitere Beispiele geben, führen wir die zweite Hauptfigur dieses Kapitels ein.

8.2 Quadratische Formen

Sei V ein K-Vektorraum.

Definition 8.2.1. Für $B \in \mathrm{Bil}(V)$ definieren wir die Funktion $q_B : V \to K$ durch $q_B(v) := B(v, v)$. Wir nennen q_B die *zugehörige quadratische Form* zur Bilinearform B. Allgemein nennen wir eine Funktion $q : V \to K$ eine *quadratische Funktion*, falls $q = q_B$ für eine Bilinearform $B \in \mathrm{Bil}(V)$. Im Fall von $V = K^n$ und $B = B_A$ für eine Matrix $A \in M_{n\times n}(K)$ (siehe Beispiel 8.1.2) schreiben wir q_A für q_{B_A}. Konkret heisst das $q_A(v) = v^T Av$ für $v \in K^n$.

Übung 8.2.2. Hier ist eine äquivalente Definition: Eine quadratische Form auf V ist eine Funktion $q : V \to K$, welche die folgenden zwei Bedingungen erfüllt:

$$\forall \lambda \in K \quad q(\lambda v) = \lambda^2 q(v), \tag{8.3}$$

$$\text{Die Funktion } (v, w) \mapsto q(v + w) - q(v) - q(w) \text{ ist bilinear.} \tag{8.4}$$

Zeigen Sie:

(a) Diese Definition ist tatsächlich äquivalent zu Definition 8.2.1.

(b) Es existiert eine Funktion $q : V \to K$, die (8.4) erfüllt, aber nicht (8.3).

(c) Falls $\mathrm{char}(K) \neq 2$ ist, kann man (8.4) mit

$$\text{Die Funktion } (v, w) \mapsto \frac{1}{2}(q(v + w) - q(v) - q(w)) \text{ ist bilinear.} \tag{8.5}$$

ersetzen.

Beispiel 8.2.3. Wenn $B = \langle\, , \rangle$ ein Skalarprodukt ist, dann ist $q_B(v) = \langle v, v\rangle = \|v\|^2$.

Beispiel 8.2.4. Sei $A = (a_{ij}) \in M_{n\times n}(K)$ und $(x_1, \dots , x_n)^T \in K^n$. Es gilt

$$q_A\left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right) = (x_1, \dots , x_n)A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}x_ix_j. \tag{8.6}$$

Bemerkung 8.2.5. Wenn keine Gefahr der Verwechslung besteht, schreiben wir ab nun $q_A(x_1, \dots , x_n)$ für $q_A\left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right)$.

Beispiel 8.2.6. Sei K ein Körper mit $\mathrm{char}(K) \neq 2$. Die folgenden quadratischen Formen auf K^2 sind gleich:

$$\begin{aligned}
q_{\left(\begin{smallmatrix}1&1\\0&2\end{smallmatrix}\right)}(x,y) &= x^2+2y^2+xy,\\
q_{\left(\begin{smallmatrix}1&0\\1&2\end{smallmatrix}\right)}(x,y) &= x^2+2y^2+yx = x^2+2y^2+xy,\\
q_{\left(\begin{smallmatrix}1&\frac{1}{2}\\\frac{1}{2}&2\end{smallmatrix}\right)}(x,y) &= x^2+2y^2+\frac{1}{2}xy+\frac{1}{2}yx = x^2+2y^2+xy.
\end{aligned}$$

Was wir in Beispiel 8.2.6 gesehen haben, war kein Zufall und wir verallgemeinern dies nun zu einer Proposition:

Proposition 8.2.7. *Sei $A \in M_{n\times n}(K)$. Falls $\mathrm{char}(K) \neq 2$, so existiert eine symmetrische Matrix $B \in M_{n\times n}(K)$, sodass $q_A = q_B$ (diese Matrix ist gegeben durch $B = \frac{A+A^T}{2}$).*

Beweis. Wir definieren B mit Inspiration von Beispiel 8.2.6:

Sei $B = (b_{ij})$ mit $b_{ij} = \frac{a_{ij}+a_{ji}}{2}$ für alle $1 \leq i,j \leq n$. Dann ist B symmetrisch und eine direkte Berechnung mit (8.6) zeigt $q_A = q_B$. □

Beispiel 8.2.8. Falls $A \in M_{n\times n}(K)$ symmetrisch ist, gilt

$$q_A(x_1,\dots,x_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j = \sum_{k=1}^{n} a_{kk}x_k^2 + \sum_{1\leq i<j\leq n} 2a_{ij}x_ix_j. \tag{8.7}$$

Jetzt fragen wir uns, wie sich die Darstellung einer quadratischen Form ändert, wenn wir verschiedene Koordinaten betrachten:

Proposition 8.2.9. *Sei $q = q_B$, wobei B eine symmetrische Bilinearform auf einem endlich-dimensionalen Vektorraum V ist, und sei $\mathcal{B} = (v_1,\dots,v_n)$ eine Basis von V. Wir schreiben einen beliebigen Vektor $v \in V$ als $v = \sum_{i=1}^n x_iv_i$ und $M_{\mathcal{B}}(B) = (a_{ij})_{1\leq i,j\leq n}$ mit $a_{ij} \in K$. Dann gilt*

$$q(v) = q_{M_{\mathcal{B}}(B)}([v]_{\mathcal{B}}) = q_{M_{\mathcal{B}}(B)}(x_1,\dots,x_n) = \sum_{i=1}^{n} a_{ii}x_i^2 + \sum_{1\leq i<j\leq n} 2a_{ij}x_ix_j. \tag{8.8}$$

Umgekehrt sei

$$p(x_1,\dots,x_n) = \sum_{i=1}^{n} \alpha_i x_i^2 + \sum_{1\leq i<j\leq n} \beta_{ij}x_ix_j \in K[x_1,\dots,x_n] \tag{8.9}$$

ein beliebiges homogenes Polynom von Grad 2 in n Variablen[4] *und nehmen wir zusätzlich an, dass* $\mathrm{char}(K) \neq 2$. *Dann existiert eine symmetrische Bilinearform B auf V, sodass* (8.8) *gilt mit $p(x_1, \ldots, x_n)$ auf der rechten Seite.*

Bemerkung 8.2.10. Man sollte (8.8) so verstehen: $q_{M_{\mathcal{B}}(B)}$ ist die Darstellung von q_B bezüglich der Basis $\mathcal{B}$.

Beweis. Für die erste Aussage setzen wir einfach die Definitionen und unser Wissen bezüglich Darstellungen von Bilinearformen ein:

$$\begin{aligned} q(v) \overset{\text{Def.}}{=} B(v,v) \underset{8.1.7}{\overset{\text{Prop.}}{=}} [v]_{\mathcal{B}}^T M_{\mathcal{B}}(B)[v]_{\mathcal{B}} \overset{\text{Def.}}{=} q_{M_{\mathcal{B}}(B)}(x_1, \ldots, x_n) \\ = \sum_{i=1}^{n} a_{ii} x_i^2 + \sum_{1 \leq i < j \leq n} 2a_{ij} x_i x_j. \end{aligned} \tag{8.10}$$

Für die zweite Aussage betrachten wir die Matrix $A(p) = (a_{ij})_{1 \leq i,j \leq n}$ mit[5]

$$a_{ij} = \begin{cases} \alpha_i & i = j, \\ \frac{\beta_{ij}}{2} & i < j, \\ \frac{\beta_{ji}}{2} & i > j. \end{cases}$$

Wir benutzen Proposition 8.1.7 (2), um eine Bilinearform B mit $M_{\mathcal{B}}(B) = A(p)$ zu finden. Die Bilinearform B ist symmetrisch, da $A(p)$ symmetrisch ist. Mit diesem B in (8.10) folgt (8.8) mit $p(x_1, \ldots, x_n)$ auf der rechten Seite. □

Wenn wir die letzte Proposition auf $V = K^n$ mit der Standard-Basis anwenden, sehen wir, dass quadratische Formen in gewisser Weise das gleiche sind wie homogene Polynome in $K[x_1, \ldots, x_n]$ von Grad 2 (was die „koordinatenabhängige" Definition von quadratischen Formen ist):

Beispiel 8.2.11. Sei p ein homogenes Polynom zweiten Grades wie in (8.9). Sei $A(p)$ wie im Beweis oben. Dann ist die quadratische Form $q_{A(p)}$ auf K^n gegeben durch p. Falls zum Beispiel $p(x,y) = x^2 + 2y^2 + xy$, dann ist

$$p(x,y) = q_{\begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 2 \end{pmatrix}}(x,y) = (x,y) \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Bemerkung 8.2.12. Wir werden deshalb quadratische Formen auf K^n und homogene Polynome zweiten Grades in $K[x_1, \ldots, x_n]$ miteinander identifizieren.

[4] Mit $K[x_1, \ldots, x_n]$ bezeichnen wir den Ring aller Polynome über K in den Variablen $x_1, \ldots, x_n$. Die formale Definition ist analog zur Definition von $K[x]$, einfach dass wir nun mehrere Unbekannte haben.

[5] Wir brauchen diese dreifache Fallunterscheidung, weil β_{ij} nur für $i < j$ definiert ist.

Es ist auch nützlich zu merken, dass man (falls $\mathrm{char}(K) \neq 2$) die symmetrische Bilinearform aus der zugehörigen quadratischen Form rekonstruieren kann:

Übung 8.2.13 (Polarisation)**.** Sei K mit $\mathrm{char}(K) \neq 2$ und B eine symmetrische Bilinearform auf V. Zeigen Sie, genau wie Sie in Übung 6.1.10 (1) gezeigt haben, dass

$$B(v, w) = \frac{1}{2}\left(q_B(v + w) - q_B(v) - q_B(w)\right). \tag{8.11}$$

Aus dieser Übung folgt, dass über einem Körper K mit $\mathrm{char}(K) \neq 2$ das Studium von quadratischen Formen und symmetrischen Bilinearformen äquivalent ist, zumindest in dem Sinne, dass wenn wir eine Charakterisierung der einen Menge haben, wir eine Charakterisierung der anderen Menge erhalten.

Da wir den Begriff einer Darstellung einer quadratischen oder bilinearen Form haben, stellen wir uns wieder unsere übliche Frage: Können wir eine Basis $\mathcal{B}$ finden, bezüglich welcher unsere quadratische/bilineare Form eine einfache Form annimmt? Wir werden später zeigen, dass für eine symmetrische Bilinearform B auf einem K-Vektorraum V, wobei $\mathrm{char}(K) \neq 2$, immer eine Basis $\mathcal{B}$ von V existiert, sodass $M_{\mathcal{B}}(B)$ diagonal ist. Wir können sogar noch eine einfachere Darstellung erhalten, wie die nächsten beiden Übungen zeigen.

Übung 8.2.14. Sei K ein beliebiger Körper. Wir definieren die Relation $\sim$ auf K: Für $a, b \in K$ sei

$$a \sim b \iff \exists \alpha \in K^\times \text{ mit } a = b\alpha^2.$$

(a) Zeigen Sie, dass $\sim$ eine Äquivalenzrelation auf K ist.

(b) Zeigen Sie, dass die Äquivalenzklasse[6] von $0 \in K$ nur 0 enthält, also $[0]_\sim = \{0\}$.

(c) Zeigen Sie für $K = \mathbb{R}$, dass $\{0, 1, -1\}$ eine vollständige Menge von Repräsentanten bildet, indem Sie zeigen, dass $[1]_\sim = \{a \in \mathbb{R} \mid a > 0\}$ und $[-1]_\sim = \{a \in \mathbb{R} \mid a < 0\}$.

(d) Zeigen Sie für $K = \mathbb{C}$, dass $\{0, 1\}$ eine vollständige Menge von Repräsentanten bildet, indem Sie zeigen, dass $[1]_\sim = \mathbb{C}^\times$.

(e) (Fortgeschritten, aber nicht schwierig) Zeigen Sie (nun wieder für einen allgemeinen Körper K), dass die Menge aller Äquivalenzklassen $\neq [0]_\sim$ eine Gruppe bilden. (*Hinweis:* Die Gruppenmultiplikation kommt von der Multiplikation in $K^\times$ und $[1]_\sim$ ist das neutrale Element.) Diese Gruppe bezeichnet man mit $K^\times/(K^\times)^2$.

[6] Falls Sie sich nicht mehr daran erinnern, für $a \in K$ ist $[a]_\sim \subseteq K$ die Äquivalenzklasse von a, also die Menge aller $b \in K$ mit $a \sim b$.

(f) Sei $K/\sim$ die Menge aller Äquivalenzklassen. Zeigen Sie, dass die folgende Abbildung wohldefiniert ist:

$$\begin{aligned} \det : \mathrm{Bil}(V) &\to K/\sim \\ B &\mapsto \det(B) := [\det M_{\mathcal{B}}(B)]_\sim, \end{aligned}$$

wobei $\mathcal{B}$ eine beliebige Basis von V ist.

Bemerkung 8.2.15. Natürlich ist die obige Definition (normalerweise) nicht die übliche Definition der Determinante. Es ist trotzdem nützlich, einer Funktion den gleichen Namen zu geben, wenn sie auf diese Weise definiert wird. Dies mag zwar verwirrend sein, hat aber den Vorteil, dass es daran erinnert, wie die Funktion definiert wurde und die Leser dazu anregt, über die Wohldefiniertheit nachzudenken.

Übung 8.2.16. Sei K ein Körper, V ein endlich-dimensionaler K-Vektorraum und $B \in \mathrm{Bil}(V)$ symmetrisch. Nehmen wir an, dass eine Basis $\mathcal{B}$ von V existiert, sodass $M_{\mathcal{B}}(B)$ diagonal ist (wie schon erwähnt, werden wir sehen, dass dies automatisch erfüllt ist, wenn $\mathrm{char}(K) \neq 2$). Sei $A \subseteq K$ eine vollständige Menge von Repräsentanten bezüglich der Äquivalenzrelation $\sim$ aus der vorherigen Übung.

(a) Zeigen Sie, dass es eine Basis $\mathcal{C}$ von V gibt, sodass $M_{\mathcal{C}}(B)$ diagonal ist, mit Elementen von A auf der Diagonalen.

(b) Im Falle von $K = \mathbb{R}$ folgern Sie, dass eine Basis $\mathcal{B}$ existiert, sodass $M_{\mathcal{B}}(B)$ diagonal ist mit nur 1, -1 und 0 auf der Diagonalen.

(c) Im Falle von $K = \mathbb{C}$ folgern Sie, dass eine Basis $\mathcal{B}$ existiert, sodass $M_{\mathcal{B}}(B)$ diagonal ist mit nur 1 und 0 auf der Diagonalen.

8.3 Diagonalisierung von symmetrischen Bilinearformen

Sei B eine Bilinearform auf einem K-Vektorraum V. Genau wie bei Darstellungen linearer Abbildungen möchten wir eine Basis $\mathcal{B}$ finden, sodass die Darstellungen von B oder q_B bezüglich $\mathcal{B}$ (d.h. $M_{\mathcal{B}}(B)$ und $q_{M_{\mathcal{B}}(B)}$) so einfach wie möglich sind. Wenn zum Beispiel $M_{\mathcal{B}}(B) = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix}$ eine Diagonalmatrix ist, so hat die induzierte quadratische Form auf K^n keine gemischten Terme:

$$q_{M_{\mathcal{B}}(B)}(x_1, \ldots, x_n) = a_1 x_1^2 + \cdots + a_n x_n^2.$$

Wir machen deshalb die folgende Definition:

Definition 8.3.1. Eine Bilinearform B auf einem endlich-dimensionalen Vektorraum V heisst *diagonalisierbar*, falls eine geordnete Basis $\mathcal{B}$ von V existiert, sodass $M_{\mathcal{B}}(B)$ diagonal ist.

Für symmetrische Bilinearformen über einem allgemeinen Körper können wir dies erreichen, abgesehen vom folgenden nervigen Beispiel:

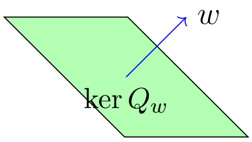
Beispiel 8.3.2. Die Matrix $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right) \in M_{2\times 2}(\mathbb{F}_2)$ ist nicht kongruent zu einer Diagonalmatrix, also kann die zugehörige Bilinearform $B_{\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)}$ auf $\mathbb{F}_2^2$ nicht diagonalisiert werden. Man kann dies durch eine Berechnung sehen: Für $P = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{F}_2)$ gilt

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^T \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} ac+ac & bc+ad \\ bc+ad & bd+bd \end{pmatrix} = \begin{pmatrix} 0 & bc+ad \\ bc+ad & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

wobei die letzte Gleichung gilt, weil $bc+ad = \det P \neq 0$ und $\mathbb{F}_2$ nur zwei Elemente hat.

Satz 8.3.3. *Sei B eine symmetrische Bilinearform auf einem endlich-dimensionalen K-Vektorraum V, wobei $\mathrm{char}(K) \neq 2$. Dann ist B diagonalisierbar.*

Beweis. Obwohl es sich hier nicht um Skalarprodukträume handelt, ist es nützlich, diesen Beweis als einen alternativen Beweis von Gram-Schmidt zu betrachten (zumindest von dem Teil des Satzes, der die Existenz einer orthogonalen Basis zeigt).



Figur 8.1: Es ist nützlich, sich $\ker Q_w$ als $w^\perp$ vorzustellen, analog zu Skalarprodukträumen und Gram-Schmidt.

Wir machen eine Induktion über $n = \dim V$. Für $n = 1$ ist jede 1×1-Matrix diagonal, also müssen wir nichts beweisen. Nehmen wir an, dass wir die Aussage für alle Vektorräume mit Dimension $n-1$ gezeigt haben und sei B eine Bilinearform auf einem n-dimensionalen Vektorraum.

Nehmen wir zuerst an, dass $B(w,w) = 0$ für alle $w \in V$. Laut der Polarisierungs-Identität (8.11) folgt, dass $B(v,w) = 0$ für alle $v, w \in V$. Somit ist B die Null-Bilinearform und insbesondere ist die Darstellungsmatrix von B diagonal (bezüglich jeder Basis!). Wir können also annehmen, dass ein $w \in V$ existiert mit $B(w,w) \neq 0$. Betrachten wir die Linearform

$$\begin{aligned} Q_w : V &\to K \\ v &\mapsto B(v,w). \end{aligned}$$

Es gilt $Q_w(w) = B(w,w) \neq 0$, also ist $\dim \ker Q_w = n-1$. Wir schreiben $N = \ker Q_w$ und betrachten $B|_{N\times N}$, welches eine symmetrische Bilinearform auf N ist. Laut der

Induktionsannahme existiert eine Basis $\mathcal{B}' = (v_1, \ldots, v_{n-1})$ von N, sodass $M_{\mathcal{B}'}(B|_{N\times N})$ diagonal ist. Sei $\mathcal{B} := (v_1, \ldots, v_{n-1}, w)$ und bemerken Sie, dass $B(w, v_i) = B(v_i, w) = 0$ für $i = 1, \ldots, n-1$ aufgrund der Symmetrie von B und der Definition von N. Daher gilt

$$M_{\mathcal{B}}(B) = \left(\begin{array}{c|c} M_{\mathcal{B}'}(B|_{N\times N}) & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ \hline 0 \quad \cdots \quad 0 & B(w,w) \end{array}\right),$$

was eine diagonale Matrix ist. □

Korollar 8.3.4. *Sei $q = q_B$ eine quadratische Form auf V, wobei B die zugehörige symmetrische Bilinearform auf V ist, und nehmen wir an, dass $\mathrm{char}(K) \neq 2$. Dann existiert eine Basis $\mathcal{B} = (v_1, \ldots, v_n)$ von V und $\alpha_1, \ldots, \alpha_n \in K$, sodass*

$$\forall v = \sum_{i=1}^{n} x_i v_i \in V \quad q(v) = \sum_{i=1}^{n} \alpha_i x_i^2.$$

Beweis. Sei $\mathcal{B} = (v_1, \ldots, v_n)$ so, dass $M_{\mathcal{B}}(B) = \begin{pmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_n \end{pmatrix}$. Aus Proposition 8.2.9 folgt, dass

$$q(v) = q_{M_{\mathcal{B}}(B)}(x_1, \ldots, x_n) = (x_1, \ldots, x_n) \begin{pmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_n \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i=1}^{n} \alpha_i x_i^2. \qquad \square$$

Korollar 8.3.5. *Sei K ein Körper mit $\mathrm{char}(K) \neq 2$. Dann ist jede symmetrische Matrix $N \in M_{n\times n}(K)$ kongruent zu einer Diagonalmatrix.*

Beweis. Wir diagonalisieren die symmetrische Bilinearform B_N auf K^n und finden dadurch eine geordnete Basis $\mathcal{B}$ und eine Diagonalmatrix D mit

$$D = M_{\mathcal{B}}(B_N) \overset{\text{Kor.}}{\underset{8.1.8}{=}} \left([\mathrm{Id}_{K^n}]^{\mathcal{B}}_{\mathcal{E}_n}\right)^T M_{\mathcal{E}_n}(B_N) \underbrace{[\mathrm{Id}_{K^n}]^{\mathcal{B}}_{\mathcal{E}_n}}_{=:P} \overset{\text{Bsp.}}{\underset{8.1.5}{=}} P^T N P. \qquad \square$$

Übung 8.3.6. Sei $q \in K[x_1, \ldots, x_n]$ eine quadratische Form[7] auf K^n, mit $\mathrm{char}(K) \neq 2$. Dann existiert $A \in \mathrm{GL}_n(K)$ und $\alpha_1, \ldots, \alpha_n \in K$ mit

$$q(y_1, \ldots, y_n) = \sum_{i=1}^{n} \alpha_i y_i^2, \quad \text{wobei} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

[7] Erinnern Sie sich, dass wir quadratische Formen auf K^n mit homogenen Polynomen zweiten Grades in n Variablen identifiziert haben.

Hinweis: Schreiben Sie $q = q_N$ und wenden Sie Korollar 8.3.5 auf N an. Die gesuchte Matrix A ist P^{-1}.

Bemerken Sie, dass der Beweis von Satz 8.3.3 nicht konstruktiv ist, da wir keine konkrete Methode angegeben haben, wie man einen Vektor w mit $B(w, w) \neq 0$ finden kann. Wir werden später eine explizite Methode angeben, wie man eine Diagonalisierung finden kann. Im nächsten Abschnitt betrachten wir den Fall $K = \mathbb{R}$, wo wir andere Methoden haben, um eine Diagonalisierung zu finden. Zuvor möchten wir noch zeigen, dass es in einem konkreten Beispiel jeweils nicht sehr schwierig ist, die Schritte aus dem Beweis von Satz 8.3.3 auszuführen.

Übung 8.3.7. Sei $A = \left(\begin{smallmatrix} 1 & 3 \\ 3 & 1 \end{smallmatrix}\right) \in M_{2\times 2}(\mathbb{F}_5)$ (oder über einem anderen Körper K mit $\operatorname{char}(K) \neq 2, 3$). Diagonalisieren Sie B_A.

Lösung. Hier ist es einfach, $w \in \mathbb{F}_5^2$ zu finden mit $B_A(w, w) \neq 0$. Zum Beispiel ist $B_A(e_1, e_1) = 1$. Betrachten wir wie im Beweis von Satz 8.3.3 die Abbildung

$$\begin{aligned} Q_{e_1} : \mathbb{F}_5^2 &\to \mathbb{F}_5 \\ v &\mapsto B_A(v, e_1), \end{aligned}$$

also $Q_{e_1}\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = (x, y)\begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = x + 3y$. Wir haben also

$$\ker Q_{e_1} = \left\{ \begin{pmatrix} -3t \\ t \end{pmatrix} \;\middle|\; t \in \mathbb{F}_5 \right\}.$$

Dann gilt beispielsweise für $\mathcal{B} = \left(e_1, \begin{pmatrix} -3 \\ 1 \end{pmatrix}\right)$, dass $M_{\mathcal{B}}(B_A) = \begin{pmatrix} 1 & 0 \\ 0 & -8 \end{pmatrix}$, da

$$q_A\begin{pmatrix} -3 \\ 1 \end{pmatrix} = \begin{pmatrix} -3 & 1 \end{pmatrix}\begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}\begin{pmatrix} -3 \\ 1 \end{pmatrix} = \begin{pmatrix} -3 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ -8 \end{pmatrix} = -8.$$

Oder man berechnet

$$\begin{pmatrix} 1 & -3 \\ 0 & 1 \end{pmatrix}^T \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}\begin{pmatrix} 1 & -3 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -8 \end{pmatrix}.$$

Wenn wir diese Berechung über $\mathbb{R}$ betrachten (oder allgemeiner über einem Körper K, in dem $a \in K$ existiert, sodass $a^2 = 8$, vergleiche Übung 8.2.14), dann können wir mit der Basis $\mathcal{C} = \left(e_1, \frac{1}{\sqrt{8}}\begin{pmatrix} 1 \\ -3 \end{pmatrix}\right)$ die Darstellung $M_{\mathcal{C}}(B_A) = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}$ erreichen.

Bemerkung 8.3.8. Wie wir in der Lösung von Übung 8.1.13 gesehen haben, sind Ähnlichkeit und Kongruenz sehr verschiedene Relationen. Nichtsdestotrotz, für symmetrische Matrizen über $\mathbb{R}$ gibt es eine nützliche Verbindung[8] zwischen den beiden:

Falls $A \in M_{n\times n}(\mathbb{R})$ symmetrisch ist, dann existiert laut dem Spektralsatz eine orthogonale Matrix $P \in \mathrm{O}(n)$, sodass

$$P^{-1}AP \text{ diagonal ist.} \tag{8.12}$$

Man könnte also sagen, dass A „orthogonal ähnlich" zu einer Diagonalmatrix ist. Aber wegen $P^{-1} = P^T$ zeigt (8.12) auch, dass A kongruent zu einer Diagonalmatrix ist. Wenn wir dies in Übung 8.3.7 benutzen, könnten wir zeigen, dass $\begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}$ kongruent und ähnlich zu $\begin{pmatrix} 4 & \\ & -2 \end{pmatrix}$ ist. Bemerken Sie, dass 4 und -2 die Eigenwerte von $\begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}$ sind. Bemerken Sie auch, dass wir keine Matrix $P \in \mathrm{O}(n)$ finden können mit

$$P^T \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix} P = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}.$$

Können Sie erklären, wieso?

Übung 8.3.9.

(a) Sei K ein Körper mit $\mathrm{char}(K) \neq 2$. Zeigen Sie, dass für eine Bilinearform B über K gilt:

$$B \text{ ist diagonalisierbar} \iff B \text{ ist symmetrisch.}$$

(b) Zeigen Sie, dass in (a) die Richtung $\implies$ auch für $\mathrm{char}(K) = 2$ gilt, $\impliedby$ aber nicht.

8.3.1 Symmetrisches Gauss Verfahren

Von Satz 8.3.3 wissen wir, dass für jede symmetrische Matrix A eine Matrix $P \in \mathrm{GL}_n(K)$ existiert mit $P^TAP = D$, wobei D diagonal ist. Erinnern Sie sich, dass jede Matrix $P \in \mathrm{GL}_n(K)$ als Produkt $P = T_1 \cdots T_k$ geschrieben werden kann, wobei T_i Elementarmatrizen sind (Satz 3.6.16). Bemerken Sie auch, dass die Transponierte jeder Elementarmatrix wieder eine Elementarmatrix ist und dass Links- (bzw. Rechts-) Multiplikationen mit Elementarmatrizen gewissen Zeilen- (bzw. Spalten-) Operationen entsprechen (Lemma 3.6.14). Setzen wir $P = T_1 \cdots T_k$ in $P^TAP = D$ ein, bekommen wir

$$T_k^T(\cdots(T_2^T(T_1^TAT_1)T_2)\cdots)T_k = D. \tag{8.13}$$

[8]Wir werden sie im nächsten Abschnitt benutzen.

Das folgende Lemma folgt direkt aus Lemma 3.6.14:

Lemma 8.3.10. *In der Notation von Lemma 3.6.14 haben wir*

- $A \xrightarrow[C_j+\alpha C_i \to C_j]{L_j+\alpha L_i \to L_j} Q_{i,j}(\alpha)^T A Q_{i,j}(\alpha)$.
- $A \xrightarrow[C_i \leftrightarrow C_j]{L_i \leftrightarrow L_j} P_{i,j}^T A P_{i,j}$.
- $A \xrightarrow[\alpha C_i \to C_i]{\alpha L_i \to L_i} S_i(\alpha)^T A S_i(\alpha)$.

Dabei ist es nicht wichtig, ob man die Zeilen- oder die Spaltenoperationen zuerst macht (da Matrixmultiplikation assoziativ ist).

Unser Fazit lautet: Die Abbildung $A \mapsto T^T AT$, wobei T eine Elementarmatrix ist, entspricht einer bestimmten Elementaroperation, die man sowohl auf die Spalten als auch auf die Zeilen anwendet. Wir nennen solche Operationen *symmetrische Elementaroperationen*. Bemerken Sie, dass wenn wir solche Operationen auf eine symmetrische Matrix anwenden, diese Matrix symmetrisch bleibt.

Korollar 8.3.11. *Die Gleichung* (8.13) *bedeutet, dass man eine beliebige symmetrische Matrix durch symmetrische Elementaroperationen in Diagonalform bringen kann.*

Wir geben nun ein Rezept dafür an. Man könnte dieses Verfahren symmetrische Gauss'sche Elimination nennen. Bemerken Sie, dass dieses Rezept einen alternativen Beweis von Satz 8.3.3 gibt.

Rezept

Sei $A = (a_{ij}) \in M_{n\times n}(K)$ symmetrisch, mit $\mathrm{char}(K) \neq 2$. Dies ist ein induktiver Algorithmus. Das Ziel ist, symmetrische Elementaroperationen auszuführen, bis wir A in eine Matrix der Form $\left(\begin{array}{c|c} * & 0 \\ \hline 0 & A' \end{array}\right)$ überführt haben, wobei A' eine symmetrische $(n-1)\times(n-1)$-Matrix ist. Danach kann man den Algorithmus nochmals auf A' anwenden, und so weiter...

Es gibt drei Etappen. Der Anfang ist wie folgt: Existiert i mit $a_{ii} \neq 0$, dann fangen Sie mit Etappe 1 an. Sonst fangen Sie mit Etappe 3 an.

Etappe 1: Sei $i_0 = \min\{i \mid a_{ii} \neq 0\}$. Falls $i_0 = 1$, gehen Sie zu Etappe 2. Falls $i_0 > 1$, vertauschen Sie die erste Zeile mit der i_0-ten Zeile und die erste Spalte mit der i_0-ten Spalte (in anderen Worten, ersetzen Sie A durch $P_{1,i_0}^T A P_{1,i_0}$). Für die neue Matrix gilt $i_0 = 1$ (d.h. der $(1,1)$-Eintrag ist nicht Null). Gehen Sie nun zu Etappe 2.

Etappe 2: Führen Sie für $j = 2, \dots, n$ die Zeilen- und Spaltenoperationen

$$L_j - \frac{a_{j1}}{a_{11}} L_1 \to L_j \quad \text{und}$$
$$C_j - \frac{a_{j1}}{a_{11}} C_1 \to C_j$$

aus. In Elementarmatrizen Schreibweise bedeutet dies, dass wir A durch

$$Q_{1,n}\left(\frac{-a_{n1}}{a_{11}}\right)^T \cdots Q_{1,2}\left(\frac{-a_{21}}{a_{11}}\right)^T A Q_{1,2}\left(\frac{-a_{21}}{a_{11}}\right) \cdots Q_{1,n}\left(\frac{-a_{n1}}{a_{11}}\right)$$

ersetzen. Dabei entsteht eine Matrix der Form

$$\left(\begin{array}{c|ccc} a_{11} & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & (*) & \\ 0 & & & \end{array}\right).$$

Jetzt fängt der Algorithmus wieder von vorne an mit der $(n-1) \times (n-1)$-Matrix $(*)$.

Etappe 3: (nur hier benutzen wir, dass $\operatorname{char}(K) \neq 2$) Falls Sie in dieser Etappe sind, heisst das, dass $a_{ii} = 0$ für alle $i = 1, \dots, n$. Dann ist entweder A die Nullmatrix, in welchem Fall es nichts mehr zu tun gibt, oder es gibt $i \neq j$ mit $a_{ij} = a_{ji} \neq 0$. Führen Sie die Operationen

$$L_i + L_j \to L_i$$
$$C_i + C_j \to C_i$$

aus (in Elementarmatrix Notation $Q_{j,i}(1)^T A Q_{j,i}(1)$). Der ii-Eintrag der resultierenden Matrix ist dann $2a_{ij} \neq 0$ (da $\operatorname{char}(K) \neq 2$). Cool! Gehen Sie nun zu Etappe 1.

Bemerkung 8.3.12. Falls Sie diesen Algorithmus benutzen wollen, um eine Matrix $P \in \mathrm{GL}_n(K)$ mit $P^T A P = D$ zu finden, dann müssen Sie einfach alle Spaltenoperationen, die Sie gemacht haben, auf die Identitätsmatrix anwenden. Damit berechnen Sie $T_1 T_2 \cdots T_k = P$.

In der Praxis, arbeiten Sie (wie in Satz 3.6.20) mit der Matrix $(A \mid I_n)$, aber stellen Sie sicher, dass Sie symmetrische Operationen auf A anwenden und nur Spaltenoperationen auf I_n, oder arbeiten Sie mit $\left(\frac{A}{I_n}\right)$, denn auf diese Weise wird I_n nur durch Spaltenoperationen beeinflusst.

Sie werden dieses Rezept mit einigen Beispielen in der Übungsstunde besprechen.

8.4 Bilinearformen über $\mathbb{R}$

Wir haben zwei Ziele in diesem Abschnitt. Das erste ist, eine Frage zu beantworten, welche wir in Kapitel 6 offen gelassen haben: Wann ist eine symmetrische Bilinearform über $\mathbb{R}$ auch positiv definit und definiert somit ein Skalarprodukt?

Die Antwort auf diese Frage wird und auch beim zweiten Ziel helfen: Wir möchten alle quadratischen Formen (oder äquivalent alle symmetrischen Bilinearformen) über $\mathbb{R}$ klassifizieren. Dies ist der Inhalt von Sylvesters Trägheitssatz.

Satz 8.4.1 (Sylvesters Trägheitssatz)**.** *Sei B eine symmetrische Bilinearform auf einem n-dimensionalen reellen Vektorraum V. Dann existiert eine Basis $\mathcal{B}$ von V, sodass*

$$M_{\mathcal{B}}(B) = \begin{pmatrix} \boxed{I_k} & & \\ & \boxed{-I_\ell} & \\ & & \boxed{0_{n-(k+\ell)}} \end{pmatrix}$$

für $k, \ell \in \mathbb{N} \cup \{0\}$. Ausserdem sind die Zahlen k und ℓ eindeutig durch B bestimmt.

Beweis. Wir zeigen zuerst die Existenz einer solchen Basis $\mathcal{B}$. Laut Satz 8.3.3 existiert eine Basis $\mathcal{C} = (v_1, \dots, v_n)$ von V und $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ mit

$$M_{\mathcal{C}}(B) = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}.$$

Durch eine Permutation der v_i's können wir annehmen, dass $k, \ell \in \mathbb{N} \cup \{0\}$ existieren mit

$$\lambda_1, \dots, \lambda_k > 0, \quad \lambda_{k+1}, \dots, \lambda_{k+\ell} < 0 \;\text{ und }\; \lambda_{k+\ell+1} = \dots = \lambda_n = 0.$$

Wir „normalisieren" die v_i's und betrachten nun die Basis

$$\mathcal{B} = \left(\frac{1}{\sqrt{|\lambda_1|}} v_1, \dots, \frac{1}{\sqrt{|\lambda_{k+\ell}|}} v_{k+\ell}, v_{k+\ell+1}, \dots, v_n \right).$$

Laut Korollar 8.1.8 gilt

$$M_{\mathcal{B}}(B) = \left([\mathrm{Id}_V]_{\mathcal{C}}^{\mathcal{B}}\right)^T M_{\mathcal{C}}(B)[\mathrm{Id}_V]_{\mathcal{C}}^{\mathcal{B}} =$$

$$\begin{pmatrix} \frac{1}{\sqrt{|\lambda_1|}} &&&&& \\ & \ddots &&&& \\ && \frac{1}{\sqrt{|\lambda_{k+\ell}|}} &&& \\ &&& 1 && \\ &&&& \ddots & \\ &&&&& 1 \end{pmatrix} \begin{pmatrix} \lambda_1 && \\ & \ddots & \\ && \lambda_n \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{|\lambda_1|}} &&&&& \\ & \ddots &&&& \\ && \frac{1}{\sqrt{|\lambda_{k+\ell}|}} &&& \\ &&& 1 && \\ &&&& \ddots & \\ &&&&& 1 \end{pmatrix}$$

$$= \begin{pmatrix} I_k && \\ & -I_\ell & \\ && 0_{n-(k+\ell)} \end{pmatrix},$$

was die Existenz zeigt.

Wir zeigen nun die Eindeutigkeit von k und ℓ. Seien $\mathcal{B} = (v_1, \ldots, v_n)$ und $\mathcal{B}' = (v'_1, \ldots, v'_n)$ geordnete Basen von V, sodass

$$A = M_{\mathcal{B}}(B) = \begin{pmatrix} I_k && \\ & -I_\ell & \\ && 0_{n-(k+\ell)} \end{pmatrix} \text{ und } A' = M_{\mathcal{B}'}(B) = \begin{pmatrix} I_{k'} && \\ & -I_{\ell'} & \\ && 0_{n-(k'+\ell')} \end{pmatrix}.$$

Wir möchten zeigen, dass $k = k'$ und $\ell = \ell'$. Seien

$$\begin{aligned} V_+ &= \mathrm{Sp}(v_1, \ldots, v_k) \quad \text{und} \quad V_{\leq 0} = \mathrm{Sp}(v_{k+1}, \ldots, v_n), \\ V'_+ &= \mathrm{Sp}(v'_1, \ldots, v'_{k'}) \quad \text{und} \quad V'_{\leq 0} = \mathrm{Sp}(v'_{k'+1}, \ldots, v'_n). \end{aligned}$$

Für $0 \neq v = \sum_{i=1}^k a_i v_i \in V_+$ gilt laut Proposition 8.1.7 (1), dass $B(v, v) > 0$, da

$$B(v,v) = [v]_{\mathcal{B}}^T M_{\mathcal{B}}(B)[v]_{\mathcal{B}} = \begin{pmatrix} a_1 \\ \vdots \\ a_k \\ 0 \\ \vdots \\ 0 \end{pmatrix}^T \begin{pmatrix} I_k && \\ & -I_\ell & \\ && 0_{n-(k+\ell)} \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_k \\ 0 \\ \vdots \\ 0 \end{pmatrix} = a_1^2 + \cdots + a_k^2 > 0.$$

Ausserdem gilt laut Proposition 8.1.7 (dieses Mal für $\mathcal{B}'$) für $v = \sum_{i=k'+1}^{n} b_i v_i' \in V'_{\leq 0}$, dass $B(v,v) \leq 0$:

$$\begin{aligned} B(v,v) = [v]_{\mathcal{B}'}^T M_{\mathcal{B}'}(B)[v]_{\mathcal{B}'} &= \begin{pmatrix} 0 \\ \vdots \\ 0 \\ b_{k'+1} \\ \vdots \\ b_n \end{pmatrix}^T \begin{pmatrix} I_{k'} & & \\ & -I_{\ell'} & \\ & & 0_{n-(k'+\ell')} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ b_{k'+1} \\ \vdots \\ b_n \end{pmatrix} \\ &= -(b_{k'+1}^2 + \cdots + b_{k'+\ell'}^2) \leq 0. \end{aligned}$$

Daher ist $V_+ \cap V'_{\leq 0} = \{0_V\}$. Laut der Dimensionsformel 2.3.37 (3) gilt

$$n \geq \dim(V_+ \oplus V'_{\leq 0}) = \dim V_+ + \dim V'_{\leq 0} = k + (n - k').$$

Daraus folgt $k' \geq k$. Durch eine Ausführung des gleichen Arguments mit V'_+ und $V_{\leq 0}$ können wir zeigen, dass $k \geq k'$. Daher gilt $k = k'$. Um $\ell = \ell'$ zu zeigen, können wir nochmals dasselbe machen mit

$$V_- = \mathrm{Sp}(v_{k+1}, \ldots, v_{k+\ell}) \quad \text{und} \quad V_{\geq 0} = \mathrm{Sp}(v_1, \ldots, v_k, v_{k+\ell+1}, \ldots, v_n)$$

und den analog definierten V'_- und $V'_{\geq 0}$. Äquivalent könnten wir auch $-B$ statt B betrachten und dann nochmals den gleichen Beweis ausführen. □

Korollar 8.4.2. *Sei $q = q_B$ eine quadratische Form auf einem n-dimensionalen reellen Vektorraum V. Dann existieren eindeutige Zahlen k und ℓ mit $k + \ell \leq n$ und eine Basis $\mathcal{B} = (v_1, \ldots, v_n)$ von V, sodass*

$$\forall v = \sum_{i=1}^{n} x_i v_i \in V \quad q(v) = \sum_{i=1}^{k} x_i^2 - \sum_{i=k+1}^{k+\ell} x_i^2. \tag{8.14}$$

Beweis. Wählen sie eine Basis $\mathcal{B}$ mit $M_{\mathcal{B}}(B) = \begin{pmatrix} I_k & & \\ & -I_\ell & \\ & & 0_{n-(k+\ell)} \end{pmatrix}$. Bezüglich dieser Basis hat q die Form (8.14), was die Existenz zeigt. Die Zahlen k und ℓ sind die Zahlen aus Satz 8.4.1. Diese sind eindeutig durch B (und somit durch q) bestimmt. Falls man eine andere Basis $\mathcal{C}$ findet, sodass q die Form (8.14) annimmt bezüglich $\mathcal{C}$ (für möglicherweise andere Zahlen k', ℓ'), so folgt $M_{\mathcal{C}}(B) = \begin{pmatrix} I_{k'} & & \\ & -I_{\ell'} & \\ & & 0_{n-(k'+\ell')} \end{pmatrix}$ und daher gilt laut der Eindeutigkeit aus Sylvesters Trägheitssatz $k' = k$ und $\ell' = \ell$, was die Eindeutigkeit zeigt. □

Definition 8.4.3. Sei B eine symmetrische Bilinearform auf einem endlich-dimensionalen reellen Vektorraum V, sei $q = q_B$ die zugehörige quadratische Form und seien k, ℓ wie in Satz 8.4.1. Wir schreiben

$$\begin{aligned} r_+(q) = r_+(B) &:= k, \\ r_-(q) = r_-(B) &:= \ell, \\ r_0(q) = r_0(B) &:= n - (k + \ell), \end{aligned}$$

und wir nennen

$$\begin{aligned} (r_+(q), r_-(q), r_0(q)) &\text{ die } \textit{Signatur} \text{ von } q \text{ (bzw. von } B), \\ r_+(q) + r_-(q) &\text{ den } \textit{Rang} \text{ von } q \text{ (bzw. von } B), \\ r_+(q) - r_-(q) &\text{ den } \textit{Index} \text{ von } q \text{ (bzw. von } B). \end{aligned}$$

Beispiel 8.4.4. Wir können jetzt Übung 8.1.15 (c) (welche Sie hoffentlich von Hand gelöst haben) lösen. Es gibt 6 Kongruenzklassen von symmetrischen 2×2-Matrizen, welche durch die sechs möglichen Signaturen

$$\begin{aligned} k = 0,\ \ell = 0 \quad & \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \\ k = 1,\ \ell = 0 \quad & \begin{pmatrix} 1 & \\ & 0 \end{pmatrix} \\ k = 2,\ \ell = 0 \quad & \begin{pmatrix} 1 & \\ & 1 \end{pmatrix} \\ k = 0,\ \ell = 1 \quad & \begin{pmatrix} -1 & \\ & 0 \end{pmatrix} \\ k = 1,\ \ell = 1 \quad & \begin{pmatrix} 1 & \\ & -1 \end{pmatrix} \\ k = 0,\ \ell = 2 \quad & \begin{pmatrix} -1 & \\ & -1 \end{pmatrix} \end{aligned}$$

repräsentiert werden. Die abgebildeten Matrizen sind die Repräsentanten dieser Kongruenzklassen, welche wir durch Sylvesters Trägheitssatz erhalten.

Wir behaupten, dass wir die Signatur einer Bilinearform ablesen können, indem wir die Vorzeichen der Eigenwerte einer Darstellungsmatrix betrachten. Der Leser sei aber gewarnt, dass es keinen Sinn macht, über die Eigenwerte einer Bilinearform zu sprechen (und auch nicht über die Eigenvektoren einer Bilinearform), wie das folgende einfache Beispiel zeigt. Jedoch macht es Sinn, wie die Eindeutigkeit aus Satz 8.4.1 zeigt, über die Anzahl positiver/negativer/nullwertiger Eigenwerte einer Bilinearform zu sprechen. Dies werden wir in Proposition 8.4.6 weiter unten verdeutlichen.

Beispiel 8.4.5. Betrachten wir nochmals die Matrix $A = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix} \in M_{2\times 2}(\mathbb{R})$ und die ensprechende Bilinearform B_A auf $\mathbb{R}^2$. Für $\mathcal{B}_{a,b} = (ae_1, be_2)$ gilt laut Beispiel 8.1.14

$$M_{\mathcal{B}_{a,b}}(B_A) = \begin{pmatrix} a^2 & \\ & -b^2 \end{pmatrix}.$$

Es gibt auch viele andere Basen $\mathcal{C}$, sodass $M_{\mathcal{C}}(B_A)$ diagonal ist. Für alle $\alpha, \beta \in \mathbb{R}$ mit $\alpha^2 \neq \beta^2$ gilt für $\mathcal{C} = \left(\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \begin{pmatrix} \beta \\ \alpha \end{pmatrix} \right)$

$$M_{\mathcal{C}}(B_A) = \begin{pmatrix} \alpha^2 - \beta^2 & 0 \\ 0 & -(\alpha^2 - \beta^2) \end{pmatrix}.$$

Wir haben in Beispiel 8.1.14 auch gesehen, dass es eine Basis $\mathcal{B}$ von $\mathbb{R}^2$ gibt mit $M_{\mathcal{B}}(B_A) = \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix}$.

Die Signatur einer Bilinearform können wir ganz leicht mithilfe der nächsten Aussage berechnen:

Proposition 8.4.6. *Sei B eine symmetrische Bilinearform auf einem n-dimensionalen reellen Vektorraum V und sei $\mathcal{C}$ eine beliebige Basis von V. Dann gilt*

$$\begin{aligned} r_+(B) &= \text{ Anzahl positiver Eigenwerte von } M_{\mathcal{C}}(B), \\ r_-(B) &= \text{ Anzahl negativer Eigenwerte von } M_{\mathcal{C}}(B), \\ r_0(B) &= \text{ Anzahl Eigenwerte von } M_{\mathcal{C}}(B) \text{ die gleich Null sind,} \end{aligned}$$

wobei wir die Eigenwerte mit ihrer Vielfachheit[9] *zählen.*

Bemerkung 8.4.7. Der Beweis von Proposition 8.4.6 benutzt den Spektralsatz und den folgenden Trick, welcher noch allgemeiner formuliert werden kann wie folgt: Wenn zwei quadratische Matrizen A, B orthogonal ähnlich sind (also $\exists P \in \mathrm{O}(n)$ sodass $B = P^{-1}AP$), dann sind sie auch kongruent, da $P^{-1} = P^T$.

[9] Erinnern Sie sich, dass für symmetrische reelle Matrizen die geometrische und algebraische Vielfachheit gleich sind, da diese laut dem Spektralsatz diagonalisierbar sind.

Beweis von Proposition 8.4.6. Die Matrix $M_{\mathcal{C}}(B)$ ist symmetrisch. Laut dem Spektralsatz 7.3.2 existiert $P \in \mathrm{O}(n)$ mit

$$P^T M_{\mathcal{C}}(B) P = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}.$$

Durch Umordnen der Spalten von P können wir annehmen, dass k, ℓ existieren mit

$$\lambda_1, \ldots, \lambda_k > 0, \quad \lambda_{k+1}, \ldots, \lambda_{k+\ell} < 0 \text{ und } \lambda_{k+\ell+1} = \cdots = \lambda_n = 0.$$

Also ist k die Anzahl positiver Eigenwerte, ℓ die Anzahl negativer Eigenwerte und $n - (k + \ell)$ die Anzahl Eigenwerte, die Null sind. Wenn wir P als $[\mathrm{Id}_V]^{\mathcal{B}'}_{\mathcal{C}}$ darstellen[10], können wir Korollar 8.1.8 (Transformationsformel für Bilinearformen) benutzen, um

$$M_{\mathcal{B}'}(B) = \left([\mathrm{Id}_V]^{\mathcal{B}'}_{\mathcal{C}}\right)^T M_{\mathcal{C}}(B) [\mathrm{Id}_V]^{\mathcal{B}'}_{\mathcal{C}} = P^T M_{\mathcal{C}}(B) P = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

zu erhalten. Wenn wir $\mathcal{B}'$ wie im Beweis von Satz 8.4.1 normalisieren, erhalten wir eine Basis $\mathcal{B}$ mit $M_{\mathcal{B}}(B) = \begin{pmatrix} I_k & & \\ & -I_\ell & \\ & & 0_{n-(k+\ell)} \end{pmatrix}$. Die Eindeutigkeit aus Sylvesters Trägheitssatz impliziert $k = r_+(B)$, $\ell = r_-(B)$ und $n - (k + \ell) = r_0(B)$, wie wir zeigen wollten. $\square$

Übung 8.4.8.

(a) Zeigen Sie, dass jede symmetrische Matrix $A \in M_{n \times n}(\mathbb{R})$ kongruent ist zu einer Matrix von der Form $D = \begin{pmatrix} I_k & & \\ & -I_\ell & \\ & & 0_{n-(k+\ell)} \end{pmatrix}$. Das heisst, es existiert $P \in \mathrm{GL}_n(\mathbb{R})$ mit $P^T A P = D$.

(b) Zeigen Sie, dass die Matrix P aus (a) so gewählt werden kann, dass ihre Spalten eine orthogonale (aber im allgemeinen keine orthonormale) Basis bilden.

In der nächsten Übung geht es darum, die analoge Version von Sylvesters Trägheitssatz über $\mathbb{C}$ zu finden.

Übung 8.4.9. Sei $B \in \mathrm{Bil}(V)$ symmetrisch, wobei V ein komplexer n-dimensionaler Vektorraum ist.

[10] Dies ist Lineare Algebra I Stoff! Stellen Sie sicher, dass Sie wissen, wie man das macht (vgl. mit der Lösung von Übung 8.1.13).

(a) Zeigen Sie, dass es eine Basis $\mathcal{B}$ von V gibt, sodass $M_{\mathcal{B}}(B) = \left(\begin{array}{c|c} I_k & \\ \hline & 0_{n-k} \end{array}\right)$.

(b) Zeigen Sie, dass die Zahl k eindeutig durch B bestimmt ist.

(c) Formulieren Sie ein Analog von Korollar 8.4.2 über $\mathbb{C}$.

Bemerkung 8.4.10. Der Beweis von (b) wird einfacher sein, nachdem wir den Stoff des nächsten Abschnitts gesehen haben.

8.5 Positiv-Definitheit und Freunde

Unser Hauptziel in diesem Abschnitt ist es, positiv definite Bilinearformen/Matrizen über $\mathbb{R}$ zu charakterisieren und einfache Methoden zu finden, sie zu erkennen. Wir fangen aber zuerst über einem allgemeinen Körper K an:

Definition 8.5.1. Sei $B \in \mathrm{Bil}(V)$ symmetrisch, wobei V ein K-Vektorraum ist. Der *Ausartungsraum* von B ist

$$N(B) = \{v \in V \mid \forall w \in V \ B(v, w) = 0\}.$$

Proposition 8.5.2. *Der Ausartungsraum von B ist der Kern der linearen Abbildung*

$$\begin{aligned} \Phi_B : V &\to V^* \\ v &\mapsto B(v, \cdot), \end{aligned} \tag{8.15}$$

wobei $B(v, \cdot)$ die Linearform $w \mapsto B(v, w)$ bezeichnet.

Beweis. Dass Φ_B linear ist, folgt aus der Bilinearität von B. Es gilt per Definition, dass

$$\ker(\Phi_B) = \{v \mid B(v, \cdot) = 0_{V^*}\} = \{v \mid \forall w \in V,\ B(v, w) = 0\} = N(B). \qquad \square$$

Definition 8.5.3. Eine symmetrische Bilinearform B heisst *ausgeartet* falls $N(B) \neq \{0\}$ und *nicht-ausgeartet* sonst. Die induzierte quadratische Form $q = q_B$ heisst entsprechend auch *ausgeartet* oder *nicht-ausgeartet*, je nachdem welches auf B zutrifft.[11]

Proposition 8.5.4. *Sei $B \in \mathrm{Bil}(V)$ eine nicht-ausgeartete symmetrische Bilinearform auf einem endlich-dimensionalen Vektorraum V. Dann ist die Abbildung $\Phi_B : V \to V^*$ aus* (8.15) *ein Isomorphismus.*

Beweis. Nicht-ausgeartet heisst $N(B) = \{0\}$, was laut Proposition 8.5.2 die Injektivität von Φ_B zeigt. Da $\dim(V^*) = \dim(V)$, folgt die Aussage aus dem Rangsatz. $\square$

[11] Manche Autoren verwenden statt ausgeartet auch den Begriff *entartet*.

Jetzt konzentrieren wir uns wieder auf Bilinearformen auf reellen Vektorräumen.

Definition 8.5.5. Eine symmetrische Bilinearform B auf einem reellen Vektorraum V heisst

- *positiv definit*, falls $B(v, v) > 0$ für alle $v \in V \setminus \{0\}$.
- *positiv semi-definit*, falls $B(v, v) \geq 0$ für alle $v \in V$.
- *negativ definit*, falls $B(v, v) < 0$ für alle $v \in V \setminus \{0\}$.
- *negativ semi-definit*, falls $B(v, v) \leq 0$ für alle $v \in V$.
- *indefinit*, falls es $v, w \in V$ gibt, sodass $B(v, v) > 0 > B(w, w)$.

Eine Matrix $A \in M_{n\times n}(\mathbb{R})$ heisst entsprechend, wenn die zugehörige Bilinearform B_A die jeweilige Eigenschaft hat.

Übung 8.5.6. Zeigen Sie: Eine symmetrische Bilinearform $B \in \mathrm{Bil}(V)$ ist positiv definit genau dann, wenn $-B$ negativ definit ist. Formulieren und beweisen Sie die analogen Aussagen auch für die restlichen Begriffe aus Definition 8.5.5.

Proposition 8.5.7. *Sei $B \in \mathrm{Bil}(V)$ symmetrisch auf einem rellen n-dimensionalen Vektorraum V. Dann gilt:*

(1) $r_+(B) = n \iff B$ ist positiv definit.

(2) $r_-(B) = n \iff B$ ist negativ definit.

(3) $r_+(B) = 0 \iff B$ ist negativ semi-definit.

(4) $r_-(B) = 0 \iff B$ ist positiv semi-definit.

(5) $r_+(B) \notin \{0, n\}$ und $r_-(B) \notin \{0, n\} \iff B$ ist indefinit.

Beweis. Sei $\mathcal{B} = (v_1, \ldots, v_n)$ eine Basis von V, sodass

$$M_{\mathcal{B}}(B) = \begin{pmatrix} I_{r_+(B)} & & \\ & -I_{r_-(B)} & \\ & & 0_{n-(r_+(B)+r_-(B))} \end{pmatrix}.$$

(1): Falls $r_+(B) = n$, dann ist $M_{\mathcal{B}}(B) = I_n$ und somit gilt für $v = \sum_{i=1}^n x_i v_i \neq 0$

$$B(v, v) = [v]_{\mathcal{B}}^T I_n [v]_{\mathcal{B}} = \sum_{i=1}^n x_i^2 > 0.$$

Falls $r_+(B) = k < n$ betrachten wir v_{k+1}:

$$B(v_{k+1}, v_{k+1}) = e_{k+1}^T M_{\mathcal{B}}(B) e_{k+1} \leq 0,$$

da es genau der $(k+1) \times (k+1)$ Eintrag in $M_{\mathcal{B}}(B)$ ist, welcher entweder -1 (falls $r_-(B) \geq 1$) oder 0 (falls $r_-(B) = 0$) ist. Daher ist B nicht positiv definit, was (1) zeigt.

(2) folgt, wenn wir (1) auf $-B$ anwenden.

(3): Falls $r_+(B) = 0$, so gilt für jedes $v = \sum_{i=1}^n x_i v_i \in V$

$$B(v, v) = [v]_{\mathcal{B}}^T M_{\mathcal{B}}(B) [v]_{\mathcal{B}} = -\sum_{i=1}^{r_-(B)} x_i^2 \leq 0.$$

Falls $r_+(B) \neq 0$, dann ist

$$B(v_1, v_1) = e_1^T M_{\mathcal{B}}(B) e_1 = 1 > 0, \tag{8.16}$$

was (3) zeigt.

(4) folgt, wenn wir (3) auf $-B$ anwenden.

(5): Die Linke Seite der zu zeigenden Äquivalenz ist die Verneinung der linken Seiten von (1)-(4) und die rechte Seite ist die Verneinung der rechten Seiten von (1)-(4). Daher folgt (5) aus den Äquivalenzen (1)-(4).

Hier ist zum Spass trotzdem ein Beweis der Richtung $\implies$ von (5):

Falls $0 < r_+(B) = k < n$ und $0 < r_-(B) < n$ gelten, so gilt $B(v_1, v_1) = 1 > 0$ (genau wie in (8.16)), und

$$B(v_{k+1}, v_{k+1}) = e_{k+1}^T M_{\mathcal{B}} e_{k+1} = -1 < 0. \qquad \square$$

Für viele Anwendungen (vergleiche Beispiel 7.3.4) ist es wichtig zu wissen, wann eine Bilinearform/Matrix positiv definit ist. Aus Proposition 8.4.6 und Proposition 8.5.7 folgt direkt:

Korollar 8.5.8. *Für eine symmetrische Bilinearform B auf einem n-dimensionalen reellen Vektorraum V und eine beliebige Basis $\mathcal{B}$ von V (bzw. für eine symmetrische Matrix $A \in M_{n \times n}(\mathbb{R})$) gilt:*

(1) B (bzw. A) ist positiv definit $\iff$ Alle Eigenwerte von $M_{\mathcal{B}}(B)$ (bzw. von A) sind positiv.

(2) B (bzw. A) ist positiv semi-definit $\iff$ Alle Eigenwerte von $M_{\mathcal{B}}(B)$ (bzw. von A) sind nicht-negativ.

(3) B (bzw. A) ist negativ definit $\iff$ Alle Eigenwerte von $M_{\mathcal{B}}(B)$ (bzw. von A) sind negativ.

(4) B (bzw. A) ist negativ semi-definit $\iff$ Alle Eigenwerte von $M_{\mathcal{B}}(B)$ (bzw. von A) sind nicht-positiv.

(5) B (bzw. A) ist indefinit $\iff$ $M_{\mathcal{B}}(B)$ (bzw. A) hat positive und negative Eigenwerte.

Für mehr Anschaulichkeit konzentrieren wir uns nun auf eine symmetrische Matrix $A \in M_{n\times n}(\mathbb{R})$. Der allgemeine Fall einer symmetrischen Bilinearform $B \in \mathrm{Bil}(V)$ folgt dann durch Wahl einer Basis $\mathcal{B}$ und Betrachten von $M_{\mathcal{B}}(B)$.

Beispiel 8.5.9. Die Matrix $A = \begin{pmatrix} a & 1 & 1 \\ 1 & a & 1 \\ 1 & 1 & a \end{pmatrix}$ hat das charakteristische Polynom $p_A(x) = -(x-(a-1))^2(x-(a+2))$. Daher gilt:

$$A \text{ ist } \begin{cases} \text{positiv definit} & \text{falls } a > 1, \\ \text{positiv semi-definit} & \text{falls } a = 1, \\ \text{indefinit} & \text{falls } -2 < a < 1, \\ \text{negativ semi-definit} & \text{falls } a = -2, \\ \text{negativ definit} & \text{falls } a < -2. \end{cases}$$

Für die Formulierung der letzten Bedingung im nächsten Satz machen wir die folgende Definition:

Definition 8.5.10. Sei $A = (a_{ij}) \in M_{n\times n}(\mathbb{R})$. Die Determinanten der Matrizen $A_k := (a_{ij})_{1\le i,j\le k}$ nennen wir die *Hauptminoren* von A.

Satz 8.5.11. *Für jede reelle symmetrische Matrix $A \in M_{n\times n}(\mathbb{R})$ sind die folgenden Aussagen äquivalent:*

(1) A ist positiv definit.

(2) Alle Eigenwerte von A sind positiv.

(3) Es existiert eine symmetrische invertierbare Matrix S mit $S^T S = S^2 = A$.

(4) Es existiert eine invertierbare Matrix B mit $B^T B = A$.

(5) Es existiert eine invertierbare obere Dreiecksmatrix R, sodass $R^T R = A$.

(6) (Hauptminorenkriterium) Alle Hauptminoren von A sind positiv.

Bevor wir mit dem Beweis beginnen, brauchen wir das folgende nützliche Lemma für den Beweis vom letzten Schritt des obigen Satzes. Ich habe dieses Lemma von [8] gelernt.

Lemma 8.5.12. *Sei $A \in M_{n\times n}(\mathbb{R})$ symmetrisch. Falls es einen Unterraum $W \subseteq \mathbb{R}^n$ gibt, sodass $B_A|_{W\times W}$ positiv definit ist, dann hat A mindestens* $\dim W$ *positive Eigenwerte.*

Beweis. Wir verwenden den Spektralsatz, um eine Orthonormalbasis $(v_1, \ldots, v_n)$ aus Eigenvektoren von A zu finden. Seien $\lambda_1, \ldots, \lambda_n$ die zugehörigen Eigenwerte, wobei wir annehmen können dass $0 \leq k \leq n$ existiert, sodass $\lambda_1, \ldots, \lambda_k$ positiv und $\lambda_{k+1}, \ldots, \lambda_n$ nicht-positiv sind. Falls $k < \dim W$, dann ist $W \cap \mathrm{Sp}(v_{k+1}, \ldots, v_n) \neq \{0\}$ (sonst wäre $\dim(W + \mathrm{Sp}(v_{k+1}, \ldots, v_n)) = \dim W + (n-k) > n$ laut Proposition 2.3.37 (3)), also existiert ein nicht-trivialer Vektor $0 \neq w \in W$ mit $w = \sum_{i=k+1}^n \alpha_i v_i$. Aber dann ist

$$w^T A w = \left(\sum_{i=k+1}^n \alpha_i v_i\right)^T \cdot \sum_{j=k+1}^n \alpha_j \lambda_j v_j = \sum_{i=k+1}^n \sum_{j=k+1}^n \alpha_i \alpha_j \lambda_j \delta_{ij} = \sum_{\ell=k+1}^n \alpha_\ell^2 \lambda_\ell \leq 0,$$

was ein Widerspruch dazu ist, dass $B_A|_{W\times W}$ positiv definit ist. Also muss $k \geq \dim W$ gelten, wie wir zeigen wollten. □

Beweis von Satz 8.5.11. Wir haben bereits gesehen, dass (1) $\iff$ (2).

Für (2) $\implies$ (3): Gemäss dem Spektralsatz existiert $O \in \mathrm{O}(n)$ mit[12] $O^T D O = A$ und $D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$. Laut (2) ist $\lambda_i > 0$ für $i = 1, \ldots, n$. Sei E nun die Matrix „$\sqrt{D}$“, das heisst $E = \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{pmatrix}$, und weiter sei $S = O^T E O$. Bemerken Sie, dass S invertiertbar und symmetrisch ist: $S^T = O^T E^T (O^T)^T = O^T E O = S$. Ausserdem ist

$$S^T S = S^2 = O^T E \underbrace{O O^T}_{=I_n} E O = O^T E^2 O = O^T D O = A,$$

was (3) zeigt.

(3) $\implies$ (4): Hier gibt es nichts zu zeigen.

(4) $\implies$ (5): Betrachten wir die QR-Zerlegung von B (siehe Satz 6.4.18). Dann ist $A = B^T B = R^T Q^T Q R = R^T R$, was (5) zeigt (bemerken Sie, dass R invertierbar ist, weil B und Q invertierbar sind, denn $R = Q^{-1} B$).

(5) $\implies$ (6): Dies folgt aus einer interessanten Beobachtung aus der Matrixmultiplikation. Sei R_k der k-te Minor von R (d.h. die Matrix, die analog zu A_k definiert ist, einfach für R statt A). Wir behaupten, dass $R_k^T R_k = A_k$.

[12] Genau genommen gibt uns der Spektralsatz 7.3.2 ein $P \in \mathrm{O}(n)$ und D diagonal mit $P^T A P = D$. Unser O hier ist dann gegeben durch $O := P^T$. Allgemein sind wir bei solchen Dingen manchmal nicht ganz so genau. Wenn Sie jedoch eine konkrete Matrix diagonalisieren möchten, müssen Sie natürlich aufpassen, dass Sie nicht O und O^T vertauschen.

Es ist viel einfacher, dies anhand eines Beispiels direkt zu sehen, als es formal zu beweisen. Wir probieren es trotzdem: Für $1 \leq i \leq k$, $1 \leq j \leq k$ haben wir

$$a_{ij} = \sum_{\ell=1}^{n} (R^T)_{i\ell}(R)_{\ell j} = \sum_{\ell=1}^{k} (R^T)_{i\ell}(R)_{\ell j},$$

da für $\ell > k$ entweder $(R^T)_{i\ell} = 0$ oder $(R)_{\ell j} = 0$. Dies beweist unsere Behauptung (bemerken Sie, dass die zweite Summe nur bis k ist). Daraus folgt

$$\det(A_k) = \det(R_k^T)\det(R_k) = \det(R_k)^2 > 0,$$

was (6) zeigt ($\det R_k \neq 0$, denn sonst wäre auch $\det R = 0$, da R eine obere Dreiecksmatrix ist).

(6) $\Longrightarrow$ (2): Hier machen wir eine Induktion über n. Der Fall $n = 1$ folgt direkt. Wir nehmen an, dass (6) $\Longrightarrow$ (1) für $n-1$ und zeigen es für n (dies dürfen wir so machen, weil wir die Äquivalenz von (1) und (2) schon gezeigt haben). Aus der Annahme folgt, dass $A|_{W\times W}$ positiv definit ist für $W = \mathrm{Sp}(e_1, \ldots, e_{n-1})$. Laut Lemma 8.5.12 hat A mindestens $n-1$ positive Eigenwerte. Da $\det(A) > 0$ (aufgrund des letzten Hauptminors) und weil $\det(A)$ das Produkt aller Eigenwerte ist, muss der letzte Eigenwert auch positiv sein. Dies zeigt (2) und beendet somit den Beweis. $\square$

Beispiel 8.5.13. Die Matrix $A = \begin{pmatrix} a & 1 & 1 \\ 1 & a & 1 \\ 1 & 1 & a \end{pmatrix}$ von Beispiel 8.5.9 hat die Hauptminoren

$$\det(a) = a, \quad \det\begin{pmatrix} a & 1 \\ 1 & a \end{pmatrix} = a^2 - 1, \quad \det A = (a-1)^2(a+2).$$

Laut der Äquivalenz (1) $\Longleftrightarrow$ (6) aus Satz 8.5.11 ist A positiv definitiv genau dann, wenn alle drei Werte positiv sind. Man berechnet, dass dies äquivalent zu $a > 1$ ist. Dies bestätigt, was wir in Beispiel 8.5.9 gesehen haben.

Übung 8.5.14. Wir betrachten die Matrix A aus Beispiel 8.5.13 für $a = 2$.

(a) Benutzen Sie den Beweis von Satz 8.5.11 (2) $\Longrightarrow$ (3) mit dem Spektralsatz, um

$$\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} = \left[\frac{1}{3}\begin{pmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 4 \end{pmatrix}\right]^2$$

zu finden.

(b) Benutzen Sie den Beweis von Satz 8.5.11 (4) $\Longrightarrow$ (5), um

$$\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ \sqrt{\frac{1}{2}} & \sqrt{\frac{3}{2}} & 0 \\ \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{6}} & \sqrt{\frac{4}{3}} \end{pmatrix} \begin{pmatrix} \sqrt{2} & \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} \\ 0 & \sqrt{\frac{3}{2}} & \sqrt{\frac{1}{6}} \\ 0 & 0 & \sqrt{\frac{4}{3}} \end{pmatrix}$$

zu finden.

Hier ist eine andere amüsante Bedingung, die wir zur Liste von äquivalenten Aussagen in Satz 8.5.11 hinzufügen können:

Übung 8.5.15. Sei $A \in M_{n\times n}(\mathbb{R})$ symmetrisch und

$$p_A = (-1)^n(x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0).$$

(a) Zeigen Sie, dass A negativ definit ist genau dann, wenn $a_0, a_1, \ldots, a_{n-1}$ alle positiv sind.

(b) Zeigen Sie, dass A positiv definit ist genau dann, wenn $a_0, a_1, \ldots, a_{n-1}$ alternierende Vorzeichen haben und $a_{n-1} < 0$ ist.

Hinweis: Sei $q = \prod_{i=1}^n (x - \alpha_i) \in \mathbb{R}[x]$. Schreibe $q = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$. Zeigen Sie

$$\begin{aligned} a_0 &= (-1)^n(\alpha_1 \cdots \alpha_n), \\ a_1 &= (-1)^{n-1}(\alpha_2 \cdots \alpha_n + \alpha_1\alpha_3 \cdots \alpha_n + \cdots + \alpha_1 \cdots \alpha_{n-2}\alpha_n + \alpha_1 \cdots \alpha_{n-1}), \\ &\ \vdots \\ a_{n-2} &= \alpha_1\alpha_2 + \cdots + \alpha_1\alpha_n + \alpha_2\alpha_3 + \cdots + \alpha_2\alpha_n + \cdots + \alpha_{n-1}\alpha_n, \\ a_{n-1} &= -(\alpha_1 + \cdots + \alpha_n). \end{aligned}$$

8.6 Lösung von Lights out

Im Appendix A haben wir das Rätsel Lights out beschrieben. Dieses Rätsel hat einen Bezug zu symmetrischen Matrizen über $\mathbb{F}_2$. Obwohl viele Dinge, die wir in den letzten zwei Kapiteln gemacht haben, über $\mathbb{F}_2$ nicht anwendbar sind, können wir wertvolle Ideen in diesen Kapiteln finden, um dieses Rätsel zu lösen[13]. Dafür brauchen wir nichts anderes als Lineare Algebra. Man könnte natürlich auch Kombinatorik oder einen Mix aus beidem verwenden (betrachten Sie zum Beispiel das Handshaking Lemma aus der Kombinatorik). Wir beschreiben die Lösung des Rätsels in zwei Schritten:

[13] Diese elegante Lösung wird normalerweise Noga Alon angerechnet.

Schritt I: Übersetzen in Lineare Algebra

Wir nummerieren die Glühbirnen mit $\{1, \dots, n\}$. Wir beschreiben Zustände der n Glühbirnen durch Vektoren in $\mathbb{F}_2^n$: Der i-te Eintrag entspricht dabei der i-ten Glühbirne, 0 bedeutet ausgeschaltet und 1 bedeutet eingeschaltet. Wir definieren die Matrix $A = (a_{ij}) \in M_{n\times n}(\mathbb{F}_2)$ durch

$$a_{ij} = \begin{cases} 1 & \text{falls die Glühbirnen } i \text{ und } j \text{ miteinander verbunden sind,} \\ 0 & \text{sonst.} \end{cases}$$

Dies gibt uns eine symmetrische Matrix mit der folgenden Eigenschaft: Das Drücken des i-ten Schalters korrespondiert zur Addition des Vektors Ae_i. Bemerken Sie, dass $a_{ii} = 1$ für alle i, da jede Glühbirne mit sich selber verbunden ist. Das Rätsel ist demnach äquivalent zu der folgenden

Behauptung: Der Vektor $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ liegt im Bild von $m_A : \mathbb{F}_2^n \to \mathbb{F}_2^n$.

Schritt II: Das Bild von A mit einem Analog von Lemma 7.1.5 charakterisieren

Nun kommt der zweite coole Teil. Die Matrix A ist symmetrisch. Also ist sie „selbst-adjungiert“. Wir schreiben dies in „ “, da dieser Ausdruck eigentlich nur über $\mathbb{R}$ oder $\mathbb{C}$ Sinn macht. Trotzdem werden wir bei dieser Bezeichnung bleiben.

Lemma 7.1.5 (2) gibt uns den Hinweis $\operatorname{Im}(T^*) = (\ker T)^\perp$ (dies ist wieder eine Aussage über $\mathbb{R}$ oder $\mathbb{C}$, aber das ist uns jetzt egal). Wenn wir glauben, dass m_A „selbst-adjungiert“ ist, dann sollte gelten, dass $\operatorname{Im}((m_A)^*) = \operatorname{Im}(m_A) = (\ker m_A)^\perp$. Also versuchen wir, dies zu beweisen.

Behauptung: $\operatorname{Im}(m_A) = (\ker A)^\perp$.

Moment mal! Wir haben ja gar kein Skalarprodukt, und somit keine Definition von $(\ker A)^\perp$. Also definieren wir eins:

$$\begin{aligned} \langle\, , \rangle : \mathbb{F}_2^n \times \mathbb{F}_2^n &\to \mathbb{F}_2 \\ (v, w) &\mapsto \langle v, w\rangle = v^T w. \end{aligned}$$

Offiziell ist dies einfach eine symmetrische Bilinearform auf $\mathbb{F}_2^n$, aber inoffiziell ist dies unser Skalarprodukt und wir werden es wie ein Skalarprodukt behandeln (daher auch die Notation $\langle\, , \rangle$). Wir können nun definieren

$$(\ker A)^\perp = \{w \in \mathbb{F}_2^n \mid \langle v, w\rangle = 0 \ \forall v \in \ker A\}.$$

Beweis der Behauptung. Sei $v \in \ker A$ und $Aw \in \operatorname{Im}(m_A)$. Dann ist

$$\langle v, Aw\rangle = v^T Aw \underset{\text{symm.}}{\overset{A}{=}} v^T A^T w = (Av)^T w = 0^T w = 0,$$

also ist $\operatorname{Im}(m_A) \subseteq (\ker A)^\perp$. Gleichheit folgt nun aus einem Analog von Satz 6.6.3, denn

$$\dim(\ker A)^\perp \geq n - \dim\ker A = \dim \operatorname{Im}(m_A). \tag{8.17}$$

Sie beschweren sich nun vielleicht, dass wir in (8.17) nicht notwendigerweise eine Gleichheit haben, da wir uns über $\mathbb{F}_2$ befinden. Sie haben recht. Daher beweisen wir:

$$\dim(\ker A)^\perp = n - \dim\ker A \underset{\text{Satz}}{\overset{\text{Rang-}}{=}} \dim \operatorname{Im}(m_A).$$

Sei $(v_1, \dots, v_k)$ eine Basis von $\ker A$ und betrachten Sie die Matrix $B = \begin{pmatrix} — & v_1^T & — \\ & \vdots & \\ — & v_k^T & — \end{pmatrix}$.

Es gilt $(\ker A)^\perp = \operatorname{Sp}(v_1, \dots, v_k)^\perp = \{v_1, \dots, v_k\}^\perp$ und da $Bx = \begin{pmatrix} \langle v_1, x\rangle \\ \vdots \\ \langle v_k, x\rangle \end{pmatrix}$ gilt, haben wir $(\ker A)^\perp = \ker B$. Da $v_1, \dots, v_k$ linear unabhängig sind, hat B Rang k. Laut dem Gauss-Verfahren/ Rangsatz ist $\dim\ker B = n - k$, wie wir zeigen wollten. Dies zeigt die Behauptung. □

Jetzt haben wir also schon einige Resultate von Skalarprodukträumen verwendet. Kommen wir nun zum Teil über bilineare/quadratische Formen. Unter Verwendung der Behautung genügt es zu zeigen, dass $(1, \dots, 1)^T \in (\ker A)^\perp$. Da

$$\left\langle \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}, \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \right\rangle = \sum_{i=1}^n v_i,$$

ist dies äquivalent dazu, für alle $(v_1, \dots, v_n)^T \in \ker A$ zu zeigen, dass $\sum_{i=1}^n v_i = 0$. Hier können wir nun endlich benutzen, dass wir über $\mathbb{F}_2$ arbeiten: Wir schreiben $A = (a_{ij})$ und berechnen für $v \in \ker A$

$$0 = v^T A v \overset{(8.7)}{=} \sum_{i=1}^n a_{ii} v_i^2 + \sum_{1 \leq i < j \leq n} 2 a_{ij} v_i v_j \overset{\mathbb{F}_2}{=} \sum_{i=1}^n a_{ii} v_i = \sum_{i=1}^n v_i,$$

genau wie wir zeigen wollten!

Changelog: Kapitel 8

- 18.04: In Definition 8.2.1 wurde eine Referenz zu Beispiel 8.1.2 hinzugefügt.
- 19.04: In Beispiel 8.1.14 wurden $\begin{pmatrix} a & b \end{pmatrix}$ und $\begin{pmatrix} a \\ b \end{pmatrix}$ zu $\begin{pmatrix} a & \\ & b \end{pmatrix}$ korrigiert.
- 20.04: In Übung 8.2.16 (a) wurde $M_{\mathcal{C}}(V)$ zu $M_{\mathcal{C}}(B)$ korrigiert.
- 22.04: Beispiel 8.1.10 und Abschnitt 8.3.1 wurden hinzugefügt.
- 23.04: Im Beweis von Proposition 8.2.9 wurde $[w]_{\mathcal{B}}$ zu $[v]_{\mathcal{B}}$ korrigiert.
- 23.04: Im Beweis von Satz 8.4.1, in der Definition von V'_+ und $V'_{\leq 0}$, wurden v'_k und v'_{k+1} zu $v'_{k'}$ und $v'_{k'+1}$ korrigiert.
- 24.04: In Übung 8.3.7 wurde $B_A \begin{pmatrix} -3 \\ 1 \end{pmatrix}$ zu $q_A \begin{pmatrix} -3 \\ 1 \end{pmatrix}$ korrigiert.
- 24.04: In Beispiel 8.1.10 wurden $\begin{pmatrix} a & b & c \end{pmatrix}$ und $\begin{pmatrix} d \\ e \\ f \end{pmatrix}$ zu $\begin{pmatrix} c & b & a \end{pmatrix}$ und $\begin{pmatrix} f \\ e \\ d \end{pmatrix}$ korrigiert.
- 26.04: Im Anfangsparagraph von Abschnitt 8.3.1 wurden Rechts- und Links- vertauscht.
- 26.04: In Bemerkung 8.3.8 wurde an einer Stelle P zu A korrigiert.
- 26.04: In Etappe 2 vom symmetrischen Gauss Verfahren (Abschnitt 8.3.1) wurden einige Vorzeichen korrigiert.
- 27.04: In Übung 8.5.15 wurde das Vorzeichen von a_0 von (-1) zu $(-1)^n$ korrigiert.
- 27.04: In Etappe 2 vom symmetrischen Gauss Verfahren (Abschnitt 8.3.1) wurden die Indizes der Matrizen $Q_{i,j}(\alpha)$ korrigiert.
- 27.04: In der Lösung von Übung 8.1.13 wurde $w_j = \sum_{i=1}^n p_{ij} v_j$ zu $w_j = \sum_{i=1}^n p_{ij} v_i$ korrigiert.
- 28.04: Im Beweis von Korollar 8.1.8 wurde $[v]_{\mathcal{B}}$ zu $[v]_{\mathcal{B}}^T$ korrigiert.
- 29.04: In Übung 8.5.15 wurde spezifiziert, dass $a_n = 1$.
- 30.04: Übung 8.5.15 wurde umformuliert.
- 30.04: Im Beweis von Satz 8.5.11 (4) $\Longrightarrow$ (5) wurde D zu A korrigiert.

- 01.05: In Lemma 8.3.10 wurden die Indizes korrigiert.
- 03.05: In der Einleitung von Kapitel 8 wurde an einigen Stellen spezifiziert, dass wir von *symmetrischen* Bilinearformen sprechen.
- 03.05: Im Beweis von Korollar 8.3.5 wurde $[\mathrm{Id}_{K^n}]_{\mathcal{B}}^{\mathcal{E}_n}$ zu $[\mathrm{Id}_{K^n}]_{\mathcal{E}_n}^{\mathcal{B}}$ korrigiert.
- 03.05: Ziemlich am Ende des Beweises von Satz 8.4.1, in der Definition von $V_{\geq 0}$ wurde ein v_k zu einem v_n korrigiert.
- 06.05: Im Beweis von Satz 8.5.11 (2) $\implies$ (3) wurde eine Fussnote hinzugefügt.
- 08.05: Im Beweis von Lemma 8.5.12 wurde an einer Stelle $\sum_{i=k+1}^n \alpha_i \lambda_i v_i$ zu $\sum_{i=k+1}^n \alpha_i v_i$ korrigiert.
- 10.05: Im Beweis von Lemma 8.5.12 wurde $\dim(W + \mathrm{Sp}(v_{k+1}), \ldots, v_n)$ zu $\dim(W + \mathrm{Sp}(v_{k+1}, \ldots, v_n))$ korrigiert.
- 10.05: Am Ende der Lösung des Rätsels Lights out (Abschnitt 8.6) wurde an einer Stelle v_i zu v_i^2 korrigiert.
- 14.06: In Beispiel 8.4.4 wurde spezifiziert, dass nur die Kongruenzklassen symmetrischer Matrizen betrachtet werden.
- 22.06: In Übung 8.4.8 wurden $M_{n\times n}(K)$ und $\mathrm{GL}_n(K)$ zu $M_{n\times n}(\mathbb{R})$ und $\mathrm{GL}_n(\mathbb{R})$ korrigiert.
- 20.08: Im Beweis der Behauptung auf Seite 340 wurde der Definitionsbereich von $\langle\, ,\rangle$ von $\mathbb{F}_2^n$ zu $\mathbb{F}_2^n \times \mathbb{F}_2^n$ korrigiert.
- 20.08: In Bemerkung 8.4.7 wurde „orthogonal äquivalent“ zu „orthogonal ähnlich“ geändert.
- 21.08: Definition 8.5.3 wurde auf quadratische Formen erweitert.

Kapitel 9

Die Jordan Normalform

Intro

Dieses Kapitel sollte eigentlich Abschnitt 4.6 sein. Jedoch haben wir es verschoben, um rechtzeitig Skalarprodukträume zu behandeln, welche Sie für andere Kurse benötigen. Inhaltlich kehren wir nun zurück zur Frage, wie wir einen gegebenen Endomorphismus möglichst einfach darstellen können. Die Antwort ist, dass für $K = \mathbb{C}$ oder allgemeiner für trigonalisierbare Endomorphismen eine (fast) eindeutige simple Form gibt, diesen darzustellen. Dies ist die Jordan Normalform und sie ist in gewisser Weise die bestmögliche Darstellung. Da diese Form (fast) eindeutig ist, wird sie von vielen Autoren auch die kanonische Jordanform genannt. Durch ihre Einfachheit hat sie viele Anwendungen, sowohl theoretische als auch anwendungsorientierte. Zum Beispiel werden wir sehen, dass es ziemlich einfach ist, die Potenzen einer Matrix in kanonischer Form zu berechnen, was uns erlauben wird, die Exponentialfunktion von Matrizen zu berechnen, was wiederum nützlich ist für das Lösen einer bestimmten grossen Klasse von gewöhnlichen Differentialgleichungen. Als Beispiel einer theoretischen Anwendung geben wir einen Beweis von Satz 5.5.22. Alle Vektorräume in diesem Kapitel sind endlich-dimensional.

9.1 Die JNF durch Beispiele verstehen

Die Jordan Normalform ist unter den Studenten berüchtigt für ihre Schwierigkeit, jedoch ohne einen guten Grund. Es wird den Lesern hoffentlich helfen, wenn wir zuerst einige Beispiele genau anschauen. Sei K ein beliebiger Körper (es wird empfohlen, an $K = \mathbb{C}$ zu denken). Weiter sei $\lambda \in K$. Wir wiederholen Beispiel 5.3.24 und machen daraus eine Definition:

Definition 9.1.1. Seien $\lambda \in K$ und $n \in \mathbb{N}$. Die Matrix

$$J_{\lambda,n} = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & \ddots & & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \lambda & 1 \\ 0 & \cdots & \cdots & 0 & \lambda \end{pmatrix} \in M_{n\times n}(K)$$

heisst *Jordan-Block* (der Dimension n zum Eigenwert λ).

Definition 9.1.2. Ein Endomorphismus $f \in \mathrm{End}(V)$ (bzw. eine Matrix $A \in M_{n\times n}(K)$) heisst *nilpotent*, falls es ein $k \in \mathbb{N}$ gibt, sodass $f^k = 0$ (bzw. $A^k = 0$). Das kleinste solche k nennen wir die *Ordnung* von f (bzw. von A).

Beispiel 9.1.3. Wir möchten zuerst Jordan-Blöcke genau studieren: Wir empfehlen den Lesern, so viel wie möglich selbständig über die folgenden Aussagen nachzudenken und sie nachzuvollziehen.

(1) $J_{\lambda,n} = \lambda I_n + J_{0,n}$.

(2) $J_{0,n}$ ist eine nilpotente Matrix mit Ordnung n. Das heisst, dass $(J_{0,n})^n = 0$ und $(J_{0,n})^{n-1} \neq 0$ (und somit auch $(J_{0,n})^k \neq 0$ für $k \leq n-1$).

(3) Somit besagt (1), dass $J_{\lambda,n} = D + N$ mit D diagonal und N nilpotent. Ausserdem gilt $DN = ND$, wie sich leicht überprüfen lässt.

(4) Sei $T \in \mathrm{End}(V)$ und $\mathcal{B} = (v_1, \dots, v_n)$ eine Basis von V, sodass $[T]_{\mathcal{B}} = J_{0,n}$. Dann haben wir

$$\begin{aligned} Tv_1 &= 0, \\ Tv_2 &= v_1, \\ &\vdots \\ Tv_n &= v_{n-1}. \end{aligned}$$

Anders gesagt wirkt T auf der Basis $(v_n, v_{n-1}, \dots, v_1)$ wie eine Verschiebung

$$v_n \xmapsto{T} v_{n-1} \xmapsto{T} \cdots \xmapsto{T} v_1 \xmapsto{T} 0.$$

Daraus wird ersichtlich, dass $T^n = 0$ aber $T^{n-1} \neq 0$.

(5) Aus (2) wissen wir, dass falls $T \in \mathrm{End}(V)$ die Darstellungsmatrix $[T]_{\mathcal{B}} = J_{\lambda,n}$ bezüglich einer Basis $\mathcal{B}$ hat, dann ist $[T - \lambda\,\mathrm{Id}_V]_{\mathcal{B}} = J_{0,n}$ und was wir in (4) gesagt haben, lässt sich auf $T - \lambda\,\mathrm{Id}_V$ anwenden.

(6) Die Matrix $J_{0,n}$ ist diagonalisierbar genau dann, wenn $n = 1$. Allgemeiner ist auch $J_{\lambda,n}$ diagonalisierbar genau dann, wenn $n = 1$. Dies war genau die Aussage von Beispiel 5.3.24. Wir haben dort gesehen, dass die Matrix $J_{\lambda,n}$ das charakteristische Polynom $(\lambda - x)^n$ hat und dass $m_g(\lambda) = 1$, während $m_a(\lambda) = n$.

(7) Aus (4) folgt, dass das Minimalpolynom von $J_{0,n}$ durch x^n gegeben ist. Durch eine Anwendung von (6) und dem Fakt, dass das Minimalpolynom das charakteristische Polynom teilt folgt aus (5) (oder aus anderen Überlegungen), dass das Minimalpolynom von $J_{\lambda,n}$ durch $M_{J_{\lambda,n}} = (x - \lambda)^n$ gegeben ist. Also gilt für Jordan-Blöcke, dass $P_{J_{\lambda,n}} = (-1)^n M_{J_{\lambda,n}}$ (der Faktor $(-1)^n$ ist bloss ein kleines Ärgernis, das wir mitschleppen müssen als Folge davon, wie wir das charakteristische Polynom und das Minimalpolynom definiert haben).

(8) Durch Anwendung von (6) und (7) kann man Satz 5.5.22 für $T = m_{J_{\lambda,n}}$ „in Aktion“ sehen.

(9) Aufgrund von (1) und (2) ist es „einfach“, die Potenzen von $J_{\lambda,n}$ zu berechnen: Zum Beispiel (bemerken Sie, dass λI_2 und $J_{0,2}$ kommutieren) gilt für alle $m \in \mathbb{N}$

$$
\begin{aligned}
(J_{\lambda,2})^m &= (\lambda I_2 + J_{0,2})^m = \sum_{i=0}^{m} \binom{m}{i} (\lambda I_2)^{m-i} (J_{0,2})^i = \lambda^m I_2 + m\lambda^{m-1} J_{0,2} + \underbrace{0}_{(J_{0,2})^\ell = 0,\ \ell \geq 2} \\
&= \lambda^m I_2 + m\lambda^{m-1} J_{0,2} = \begin{pmatrix} \lambda^m & m\lambda^{m-1} \\ 0 & \lambda^m \end{pmatrix}.
\end{aligned}
$$

Allgemein muss man für $(J_{\lambda,n})^m$ nur die ersten n Terme der Binomialformel berechnen.

(10) Bemerken Sie, dass die Potenzen von $J_{0,n}$ einfach zu merken sind:

$$
J_{0,n} = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix}, \quad (J_{0,n})^2 = \begin{pmatrix} 0 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & 1 \\ & & & \ddots & 0 \\ & & & & 0 \end{pmatrix},
$$

$$
(J_{0,n})^3 = \begin{pmatrix} 0 & 0 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \ddots & 1 \\ & & & \ddots & \ddots & 0 \\ & & & & \ddots & 0 \\ & & & & & 0 \end{pmatrix}, \quad \text{und so weiter.}
$$

Mit jeder Erhöhung der Potenz verschiebt sich die Nebendiagonale mit den 1 um eins nach oben.

Beispiel 9.1.4. Wir möchten nun Matrizen mit mehreren Jordan-Blöcken entlang der Diagonale untersuchen. Vorerst nehmen wir an, dass alles Blöcke zum selben Eigenwert sind.

(1) Wir beginnen mit dem Fall, in dem wir nur Blöcke zum Eigenwert 0 haben: Falls

$$A = \begin{pmatrix} \boxed{J_{0,4}} & & & \\ & \boxed{J_{0,3}} & & \\ & & \boxed{J_{0,2}} & \\ & & & \boxed{J_{0,1}} \end{pmatrix} \in M_{10\times 10}(K),$$

dann ist $A^4 = 0$ und $A^3 \neq 0$. Für die Potenzen einer Blockdiagonalmatrix müssen wir nur die Potenzen der Blöcke berechnen. Mithilfe von Beispiel 9.1.3 (10) ist dies ganz leicht:

$$A^2 = \begin{pmatrix} \boxed{\begin{matrix} 0 & 0 & 1 & \\ & 0 & 0 & 1 \\ & & 0 & 0 \\ & & & 0 \end{matrix}} & & & \\ & \boxed{\begin{matrix} 0 & 0 & 1 \\ & 0 & 0 \\ & & 0 \end{matrix}} & & \\ & & \boxed{\begin{matrix} 0 & 0 \\ & 0 \end{matrix}} & \\ & & & \boxed{0} \end{pmatrix}, \quad A^3 = \begin{pmatrix} \boxed{\begin{matrix} 0 & 0 & 0 & 1 \\ & 0 & 0 & 0 \\ & & 0 & 0 \\ & & & 0 \end{matrix}} & & & \\ & \boxed{0_3} & & \\ & & \boxed{0_2} & \\ & & & \boxed{0_1} \end{pmatrix}$$

und $A^4 = 0$. Also ist $M_A = x^4$ und $p_A = x^{10}$.

(2) Allgemeiner, falls

$$A = \begin{pmatrix} \boxed{J_{0,n_1}} & & \\ & \ddots & \\ & & \boxed{J_{0,n_k}} \end{pmatrix} \in M_{S\times S}(K),$$

wobei $S = \sum_{i=1}^{k} n_i$, dann ist $M_A = x^{\max(n_1,\ldots,n_k)}$ und $p_A = (-x)^S$.

(3) Noch allgemeiner, falls

$$B = \begin{pmatrix} \boxed{J_{\lambda,n_1}} & & \\ & \ddots & \\ & & \boxed{J_{\lambda,n_k}} \end{pmatrix}, \tag{9.1}$$

dann ist

$$M_B = (x-\lambda)^{\max(n_1,\dots,n_k)} \quad \text{und} \quad p_B = (\lambda - x)^{n_1+\dots+n_k}. \tag{9.2}$$

Für B wie in (9.1) können wir uns fragen, ob wir die Anzahl Blöcke einer gegebenen Dimension „rekonstruieren“ können, wenn wir die lineare Transformation m_B betrachten. Was (9.2) sagt, ist, dass wir auf jeden Fall das Maximum und die Summe der Dimensionen rekonstruieren können. Aber wir können tatsächlich alle herausfinden. Wir analysieren zuerst ein konkretes Beispiel, bevor wir dies allgemeiner formulieren.

Beispiel 9.1.5. Betrachten wir die Matrix

$$B = \begin{pmatrix} \boxed{J_{\pi,3}} & & & \\ & \boxed{J_{\pi,2}} & & \\ & & \boxed{J_{\pi,2}} & \\ & & & \boxed{J_{\pi,1}} \end{pmatrix} \in M_{8\times 8}(\mathbb{C}).$$

Die Matrix $N = B - \pi I_8$ ist nilpotent. Wir haben oben bereits erklärt, dass $M_N = x^3$, da 3 die Grösse des grössten Blocks ist. Berechnen wir nun $\dim \ker N^i$ für $i = 1, 2, 3$.

$$\begin{aligned} \ker N &= \operatorname{Sp}(e_1, e_4, e_6, e_8), \\ \ker N^2 &= \ker N \oplus \operatorname{Sp}(e_2, e_5, e_7), \\ \ker N^3 &= \mathbb{R}^8 = \ker N^2 \oplus \operatorname{Sp}(e_3). \end{aligned}$$

Der Kern von N^{i-1} ist natürlich im Kern von N^i enthalten. Die Vektoren im Spann sind jeweils die Vektoren, die „neu“ dazukommen. Wir bemerken:

- Der Block der Grösse 3 trägt jeweils einen neuen Vektor zu $\ker N$, $\ker N^2$ und $\ker N^3$ bei. In unserem Beispiel sind das die Vektoren e_1, e_2 und e_3.

- Die Blöcke der Grösse 2 tragen je einen neuen Vektor zu $\ker N$ und $\ker N^2$ bei. In unserem Beispiel sind das die Vektoren e_4 und e_6 bei $\ker N$, sowie e_5 und e_7 bei $\ker N^2$.

- Der Block der Grösse 1 trägt einen Vektor zu $\ker N$ bei. In unserem Beispiel ist das e_8.

Wenn wir eine Weile darüber nachdenken, sehen wir:

$$\begin{aligned} \dim\ker N = {}&\text{ Anzahl Blöcke } = \text{ Anzahl Blöcke der Grösse } \geq 1, \\ \dim\ker N^2 - \dim\ker N = 7-4=3 = {}& \text{ Anzahl Blöcke der Grösse } \geq 2, \\ \dim\ker N^3 - \dim\ker N^2 = 8-7=1 = {}& \text{ Anzahl Blöcke der Grösse } \geq 3. \end{aligned}$$

Beispiel 9.1.6. Wie viele Matrizen $A \in M_{8\times 8}(\mathbb{C})$ gibt es, die aus Jordan Blöcken von der Form $J_{0,n}$ für $n \in \mathbb{N}$ bestehen und die das Minimalpolynom $M_A = x^3$ haben? Was ist die Antwort, wenn wir wissen, dass $\dim\ker A = 3$? Was ist die Antwort im Falle von $\dim\ker A = 2$?
Antwort: Der Fakt $M_A = x^3$ impliziert, dass es mindestens einen Block $J_{0,3}$ gibt und keine grösseren Blöcke. Also ist es nun eine kombinatorische Frage. Es gibt fünf Möglichkeiten:

	$\#J_{0,3}$	$\#J_{0,2}$	$\#J_{0,1}$	$\dim\ker A$	$\dim\ker A^2$	$\dim\ker A^3$
I	1	2	1			8
II	1	1	3			8
III	1	0	5			8
IV	2	1	0			8
V	2	0	2			8

Ergänzen Sie die fehlenden Informationen und benutzen Sie die Tabelle, um die oben gestellten Fragen zu beantworten.

Wenn wir die Erkenntnisse aus den letzten beiden Beispielen zu verallgemeinern versuchen, erhalten wir die nächste Übung.

Übung 9.1.7. Wir betrachten die Matrix B wie in (9.1).

(a) Zeigen Sie: Die Anzahl Blöcke von der Form $J_{\lambda,\ell}$ mit $\ell \geq j$ ist

$$\dim\ker(B-\lambda I)^j - \dim\ker(B-\lambda I)^{j-1}.$$

(b) Folgern Sie: Die Anzahl Blöcke der Form $J_{\lambda,j}$ ist

$$2\dim\ker(B-\lambda I)^j - \dim\ker(B-\lambda I)^{j+1} - \dim\ker(B-\lambda I)^{j-1}.$$

Hinweis: Für (a): Versuchen Sie, die Berechnung in Beispiel 9.1.4 (1) gut zu verstehen. Es könnte auch helfen zu merken, dass für Matrizen B wie in (9.1)

$$\dim\ker(B-\lambda I)^j = \text{Anzahl Nullspalten in } (B-\lambda I)^j$$

gilt, da die verschiedenen Nullspalten linear unabhängig sind.

Für (b): Benutzen Sie (a).

Beispiel 9.1.8. Schauen wir uns nun den Fall von Blöcken zu verschiedenen Eigenwerten an.

(1) Betrachten wir zum Beispiel die Matrix

$$C = \begin{pmatrix} \boxed{J_{\pi,5}} & & \\ & \boxed{J_{i,2}} & \\ & & \boxed{J_{i,1}} \end{pmatrix} \in M_{8\times 8}(\mathbb{C}).$$

Laut Lemma 5.3.6 ist $p_C = (x-\pi)^5(x-i)^3$ und man kann zeigen (tun Sie es!), dass $M_C = (x-\pi)^5(x-i)^2$.

(2) Bemerken Sie, dass $C - \lambda I_8$ invertierbar ist für $\lambda \notin \{\pi, i\}$ und nicht-invertierbar für $\lambda \in \{\pi, i\}$.

(3) Eine allgemeine Jordan Block Matrix

$$J = \begin{pmatrix} \boxed{J_{\lambda_1,n_1}} & & \\ & \ddots & \\ & & \boxed{J_{\lambda_k,n_k}} \end{pmatrix}, \tag{9.3}$$

hat das charakteristische Polynom

$$P_J = (-1)^S \prod_{i=1}^{k} (x-\lambda_i)^{n_i}, \tag{9.4}$$

wobei $S = \sum_{i=1}^k n_i$. Für das Minimalpolynom seien $\mu_1, \ldots, \mu_\ell$ die paarweise verschiedenen Eigenwerte[1] von J. Dann ist

$$M_J = \prod_{i=1}^{\ell} (x-\mu_i)^{m_i}, \tag{9.5}$$

wobei $m_i = \max\{n_j \mid 1 \leq j \leq k \text{ und } \lambda_j = \mu_i\}$.

Übung 9.1.9. Zeigen Sie, dass die Matrix J aus (9.3) diagonalisierbar ist genau dann, wenn $n_i = 1$ für alle $i = 1, \ldots, k$. Dies, zusammen mit (9.5), zeigt Satz 5.5.22 für J.

Das coole ist, dass zum Beispiel über $\mathbb{C}$, es zu jedem Endomorphismus $T \in \mathrm{End}(V)$ auf einem endlich-dimensionalen Vektorraum V eine Basis $\mathcal{B}$ gibt, sodass $[T]_\mathcal{B} = J$ für

[1] Das heisst, die *Mengen* $\{\lambda_1, \ldots, \lambda_k\}$ und $\{\mu_1, \ldots, \mu_\ell\}$ sind gleich. In der *Liste* $(\lambda_1, \ldots, \lambda_k)$ können allerdings Wiederholungen vorkommen, während die μ_i alle verschieden sind.

eindeutige k, λ_i und n_i. Wir hören nun auf mit den Beispielen und formulieren den Hauptsatz dieses Kapitels:

Satz 9.1.10 (Jordan Normalform)**.** *Sei V ein endlich-dimensionaler Vektorraum über K und sei $T \in \mathrm{End}(V)$, sodass p_T in Linearfaktoren zerfällt*[2] *(z.B. ein beliebiger Endomorphismus, falls $K = \mathbb{C}$). Dann existiert eine Basis $\mathcal{B}$ von V, sodass*

$$[T]_{\mathcal{B}} = \begin{pmatrix} \boxed{J_{\lambda_1,n_1}} & & \\ & \ddots & \\ & & \boxed{J_{\lambda_k,n_k}} \end{pmatrix}. \tag{9.6}$$

Bis auf Permutationen der Blöcke $\{J_{\lambda_i,n_i}\}_{i=1}^k$ ist diese Darstellung eindeutig.

Definition 9.1.11. Eine Basis $\mathcal{B}$ wie in Satz 9.1.10 nennen wir eine *Jordan-Basis* und die Matrix auf der rechten Seite von (9.6) nennen wir eine *Jordan-Normalform* von T (oder wenn man nicht so streng ist, auch *die* Jordan-Normalform von T). Solche Matrizen nennen wir Matrizen in *Jordanform.*

Korollar 9.1.12. *(Jordan Normalform für Matrizen) Sei $A \in M_{n\times n}(K)$, sodass p_A in Linearfaktoren zerfällt in $K[x]$. Dann existiert eine Matrix in Jordanform J, sodass A ähnlich zu J ist. Diese Matrix J ist eindeutig bis auf Permutationen der Jordanblöcke.*

Beweis. Wenden Sie Satz 9.1.10 auf $T = m_A$ an. □

Übung 9.1.13. Füllen Sie die Details dieses Beweises ein.

9.2 Beweis der Existenz der Jordan Normalform

Wir beginnen mit einer Übersetzung von Satz 9.1.10 in die Sprache von invarianten Unterräumen:

Satz 9.2.1 (Existenz der Jordan Normalform)**.** *Sei $T \in \mathrm{End}(V)$ wie in Satz 9.1.10. Dann existieren T-invariante Untervektorräume $W_1, \ldots, W_k$ mit $V = W_1 \oplus \cdots \oplus W_k$ und Basen $\mathcal{B}_i \subseteq W_i$ für $i = 1, \ldots, k$ mit $[T|_{W_i}]_{\mathcal{B}_i} = J_{\lambda_i,n_i}$, wobei $n_i = \dim W_i$ und $\lambda_i \in K$.*

Übung 9.2.2. Stellen Sie sicher, dass Sie verstehen, wieso Satz 9.2.1 äquivalent ist zum Existenzteil von Satz 9.1.10. Beispiel 9.2.10 unten wird Ihnen vielleicht dabei helfen.

Um die Untervektorräume W_i zu finden, geben wir die folgende Definition:

[2]Erinnern Sie sich, dass dies laut Satz 5.4.5 äquivalent dazu ist, T trigonalisierbar ist.

Definition 9.2.3. Sei U ein endlich-dimensionaler Vektorraum und $T \in \mathrm{End}(U)$ ein Endomorphismus. Wir nennen U *unzerlegbar* bezüglich T, falls keine T-invarianten Unterräume $\{0\} \subsetneq U_1, U_2 \subsetneq U$ existieren, sodass $U = U_1 \oplus U_2$. Eine äquivalente Formulierung wäre, dass falls $U = U_1 \oplus U_2$ für T-invariante Unterräume U_1, U_2 gilt, dann ist einer der beiden Unterräume U_1, U_2 gleich U und der andere gleich $\{0\}$. Falls U nicht unzerlegbar ist, nennen wir U *zerlegbar* bezüglich T. Wir nennen einen T-invarianten Untervektorraum $W \subseteq U$ *unzerlegbar* (bzw. *zerlegbar*) bezüglich T, wenn W unzerlegbar (bzw. zerlegbar) bezüglich $T|_W$ ist.

Bemerkung 9.2.4. Die Begriffe unzerlegbar (und irreduzibel aus Übung 9.2.8 unten) hängen natürlich von T ab. Wenn T allerdings aus dem Kontext heraus klar ist, werden wir diese Abhängigkeit nicht erwähnen.

Beispiel 9.2.5. Jeder 1-dimensionale Vektorraum V ist unzerlegbar (bezüglich einem beliebigen Endomorphismus), aus dem einfachen Grund, dass wenn $V = U_1 \oplus U_2$, dann muss einer der beiden Unterräume U_1, U_2 der Nullraum sein.

Beispiel 9.2.6. Falls $T \in \mathrm{End}(V)$ und V keine T-invarianten Unterräume (ausser $\{0\}$ und V) hat, dann ist V unzerlegbar. Ein Beispiel dafür ist eine Rotation $R_\theta \in \mathrm{End}(\mathbb{R}^2)$ für interessante θ.

Beispiel 9.2.7. Ein Vektorraum kann invariante Untervektorräume haben und trotzdem unzerlegbar sein: Betrachten wir $m_{J_{0,2}} : K^2 \to K^2$. Der Untervektorraum $U = \mathrm{Sp}(e_1)$ ist invariant, aber man kann zeigen, dass U kein T-invariantes Komplement hat (siehe die folgende Übung).

Übung 9.2.8. In der abstrakten Algebra gibt es eine verwandte Definition, die wir nicht brauchen werden, aber trotzdem erwähnen möchten: Wie in Definition 9.2.3 sei U ein endlich-dimensionaler Vektorraum und $T \in \mathrm{End}(U)$. Der Vektorraum U heisst *irreduzibel*, falls es keinen T-invarianten Untervektorraum $\{0\} \subsetneq W \subsetneq U$ gibt.

(a) Zeigen Sie, dass ein irreduzibler Vektorraum unzerlegbar ist.

(b) Zeigen Sie, dass ein unzerlegbarer Vektorraum nicht unbedingt irreduzibel ist (Hinweis: Sie kennen ein solches Beispiel in $U = K^2$ sehr gut).

Beispiel 9.2.9. Sei $A \in M_{n\times n}(K)$ mit $A = \left(\begin{array}{c|c} B_1 & \\ \hline & B_2 \end{array}\right)$, wobei $B_1 \in M_{k\times k}(K)$ und $B_2 \in M_{(n-k)\times(n-k)}(K)$ mit $0 < k < n$. Dann ist K^n zerlegbar bezüglich m_A:

$$U_1 = \mathrm{Sp}(e_1, \dots, e_k) \quad \text{und} \quad U_2 = \mathrm{Sp}(e_{k+1}, \dots, e_n)$$

sind m_A-invariant und $K^n = U_1 \oplus U_2$.

Beispiel 9.2.10. Allgemeiner, sei $T \in \operatorname{End}(V)$ mit $\dim V = n$. Dann ist V zerlegbar (bezüglich T) genau dann, wenn eine Basis $\mathcal{B}$ von V existiert, sodass $[T]_{\mathcal{B}} = \left(\begin{array}{c|c} B_1 & \\ \hline & B_2 \end{array}\right)$, wobei $B_1 \in M_{k\times k}(K)$ und $B_2 \in M_{(n-k)\times(n-k)}(K)$ mit $0 < k < n$.

Tatsächlich, wenn wir eine solche Basis $\mathcal{B} = (v_1, \dots, v_n)$ haben, dann sind

$$U_1 = \operatorname{Sp}(v_1, \dots, v_k) \quad \text{und} \quad U_2 = \operatorname{Sp}(v_{k+1}, \dots, v_n)$$

zwei T-invariante Unterräume mit $V = U_1 \oplus U_2$.

Umgekehrt, falls $V = U_1 \oplus U_2$ gilt und U_1, U_2 beide T-invariant sind, dann erhalten wir für beliebige Basen $\mathcal{B}_1 = (v_1, \dots, v_k)$ von U_1 und $\mathcal{B}_2 = (w_1, \dots, w_{n-k})$ von U_2 eine Basis $\mathcal{B} = (v_1, \dots, v_k, w_1, \dots, w_{n-k})$ von V mit

$$[T]_{\mathcal{B}} = \left(\begin{array}{c|c} [T|_{U_1}]_{\mathcal{B}_1} & \\ \hline & [T|_{U_2}]_{\mathcal{B}_2} \end{array}\right).$$

Lemma 9.2.11. *Sei V ein endlich-dimensionaler Vektorraum und $T \in \operatorname{End}(V)$. Dann existieren unzerlegbare T-invariante Untervektorräume $W_1, \dots, W_k$ von V mit $V = W_1 \oplus \dots \oplus W_k$.*

Beweis. Falls V unzerlegbar ist, müssen wir nichts zeigen (wir wählen einfach $k = 1$, $W_1 = V$). Ansonsten existieren T-invariante Unterräume $\{0\} \subsetneq U_1, U_2 \subsetneq V$ mit $V = U_1 \oplus U_2$. Wenn U_1 und U_2 beide unzerlegbar sind, sind wir fertig. Falls einer (oder beide) zerlegbar ist, können wir diesen weiter zerlegen. Da wir in jeder solchen Zerlegung die Dimension verkleinern und V endlich-dimensional ist, müssen wir irgendwann eine Zerlegung wie in der Aussage des Lemmas erreichen. □

Für diejenigen Leser, denen der Schluss dieses Beweises zu ungenau ist, geben wir noch einen Beweis mit starker Induktion[3].

Beweis mit starker Induktion. Wir verwenden eine starke Induktion über $n = \dim V$. Falls $\dim V = 1$, gibt es nichts zu beweisen, da jeder 1-dimensionale Vektorraum unzerlegbar ist (auch wenn $T = 0$). Jetzt nehmen wir an, dass die Aussage für alle Vektorräume der Dimension $< n$ gilt (dies ist die starke Induktionsannahme) und zeigen, dass sie dann auch für Vektorräume V mit $\dim V = n$ gilt.

Falls V unzerlegbar ist, sind wir fertig (mit $k = 1$, $W_1 = V$). Falls nicht, existieren T-invariante Unterräume $\{0\} \subsetneq U_1, U_2 \subsetneq V$ mit $V = U_1 \oplus U_2$. Es gilt $\dim U_1 < n$ und $\dim U_2 < n$. Laut der starken Induktionsannahme existieren unzerlegbare T-invariante

[3] Bei der starken Induktion verwendet man für den Beweis der Aussge im Fall n die Aussage in den Fällen k für alle $k < n$, im Unterschied zur üblichen vollständigen Induktion, bei der man für den Beweis im Fall n nur die Aussage im Fall $n-1$ braucht. Sehen Sie zum Beispiel hier.

Untervektorräume $W_1, \dots, W_\ell$ mit $U_1 = W_1 \oplus \cdots \oplus W_\ell$ und unzerlegbare T-invariante Untervektorräume $W_{\ell+1}, \dots, W_{\ell+m}$ mit $U_2 = W_{\ell+1} \oplus \cdots \oplus W_{\ell+m}$. Es folgt, dass

$$V = U_1 \oplus U_2 = W_1 \oplus \cdots \oplus W_{\ell+m},$$

wobei $W_1, \dots, W_{\ell+k}$ unzerlegbare T-invariante Unterräume von V sind, wie wir zeigen wollten. □

Beweis mit einem Trick. Man betrachte die Menge aller möglichen Zerlegungen von V als direkte Summe $V = \bigoplus_{i=1}^k W_i$, wobei $k \in \mathbb{N}$ und W_i T-invariante Unterräume von V sind. Diese Menge ist nicht leer, da $V = V$ eine solche Zerlegung ist. Für jede solche Zerlegung $V = W_1 \oplus \cdots \oplus W_k$ betrachtet man $\sum_{i=1}^k \dim(W_i)^2$, und nimmt eine Zerlegung mit minimalem Wert. Wir behaupten, dass alle Summanden einer solchen Zerlegung unzerlegbar sind: Dies folgt aus $n_1^2 + n_2^2 < (n_1 + n_2)^2$ für $n_1, n_2 \in \mathbb{N}$. □

Bemerkung 9.2.12. Wir haben drei Beweise gegeben, die relativ detailliert waren. Sie müssen sich aber daran gewöhnen, dass in zukünftigen Vorlesungen und mathematischen Dokumenten, die Sie lesen werden, viel knapper geschrieben und mehr dem Leser überlassen wird. Zum Beispiel könnte man für den ersten Beweis einfach schreiben „Zerlegen Sie, bis sie nicht mehr können". Für den zweiten kann man nach $V = U_1 \oplus U_2$ einfach „Induktion" schreiben. Für den dritten: „Nehmen Sie eine Zerlegung, die $\sum_i (\dim W_i)^2$ minimiert". Sie dürfen das natürlich nicht so in einer Prüfung oder Lösung einer Serie schreiben. Aber Sie sollten lernen, solche kurzgefassten „Beweise" zu ergänzen und nachzuvollziehen. Der Vorteil solcher „Beweise" ist, dass man ganz klar sieht, wo der Hund begraben ist.

Übung 9.2.13. Ist eine solche Zerlegung in invariante Untervektorräume im allgemeinen eindeutig? Falls nicht, in welchen Fällen ist sie es?

<u>Fazit:</u> Sei $V = W_1 \oplus \cdots \oplus W_k$ eine Zerlegung in unzerlegbare T-invariante Untervektorräume bezüglich $T \in \mathrm{End}(V)$. Dann können wir T analysieren, indem wir $T|_{W_i}$ untersuchen für alle $i = 1, \dots, k$. Tatsächlich, da jedes $v \in V$ eindeutig als $v = \sum_{i=1}^k v_i$ mit $v_i \in W_i$ geschrieben werden kann, gilt

$$T(v) = \sum_{i=1}^k T(v_i) = \sum_{i=1}^k T|_{W_i}(v_i).$$

Es folgt zum Beispiel, dass es genügt, den Existenzteil in Satz 9.1.10 für $T|_{W_i}$ für jedes i zu beweisen. Man schreibt diese Situation manchmal so:

$$T = T|_{W_1} \oplus \cdots \oplus T|_{W_k}.$$

Das folgende Lemma wird uns später erlauben, unzerlegbare Vektorräume zu klassifizieren:

Lemma 9.2.14 (Fitting-Lemma[4]). *Sei V ein endlich-dimensionaler Vektorraum und $S \in \mathrm{End}(V)$. Dann existiert $k \in \mathbb{N}$ mit $V = \ker(S^k) \oplus \mathrm{Im}(S^k)$.*

Beweis. Betrachten wir die folgende Kette von Untervektorräumen von V:

$$\ker(S) \subseteq \ker(S^2) \subseteq \cdots .$$

Da V endlich-dimensional ist, muss $\dim \ker(S^k)$ ab einem gewissen k konstant bleiben, das heisst die Kette stabilisiert sich. Sei $k \in \mathbb{N}$ mit

$$\ker(S^k) = \ker(S^{k+1}) = \cdots .$$

Wir behaupten, dass dann $V = \ker(S^k) \oplus \mathrm{Im}(S^k)$ gilt. Vom Rangsatz wissen wir $\dim V = \dim \ker(S^k) + \dim \mathrm{Im}(S^k)$, also genügt es laut Proposition 2.3.40, zu zeigen, dass $\ker(S^k) \cap \mathrm{Im}(S^k) = \{0\}$.

Sei also $v \in \ker(S^k) \cap \mathrm{Im}(S^k)$. Dann ist $S^k v = 0$ und $v = S^k w$ für ein $w \in V$. Es folgt, dass

$$0 = S^k v = S^k(S^k w) = S^{2k} w,$$

also ist $w \in \ker(S^{2k}) = \ker(S^k)$. Folglich ist $v = S^k w = 0$. □

Der Grund, wieso die Zerlegung von Lemma 9.2.14 für uns interessant ist, liegt in der folgenden Beobachtung:

Lemma 9.2.15. *Seien $T, S \in \mathrm{End}(V)$ mit $TS = ST$. Dann sind $\mathrm{Im}\, S$ und $\ker S$ T-invariante Untervektorräume von V.*

Beweis. Sei $v \in \ker S$. Wir wollen zeigen, dass $Tv \in \ker S$ ist:

$$S(Tv) = T(Sv) = T(0) = 0.$$

Die Invarianz von $\mathrm{Im}\, S$ lässt sich ähnlich zeigen. Sei $v = Sw \in \mathrm{Im}\, S$. Dann gilt

$$Tv = T(Sw) = S(Tw)$$

und daher ist $Tv \in \mathrm{Im}\, S$. □

Korollar 9.2.16. *Falls $TS = ST$, dann sind $\mathrm{Im}(S^k)$ und $\ker(S^k)$ T-invariant für jedes $k \in \mathbb{N}$.*

[4]Benannt nach Hans Fitting.

Beweis. Aus $TS = ST$ folgt, dass auch $TS^k = S^kT$ für alle $k \in N$, also können wir Lemma 9.2.15 anwenden. □

Nun kommen wir noch zur letzten Vorbereitung, bevor wir mit dem Beweis der Existenz der Jordan Normalform beginnen können.

Lemma 9.2.17. *Sei $T \in \operatorname{End}(V)$, sodass p_T in $K[x]$ in Linearfaktoren zerfällt und sei $\{0\} \neq W \subseteq V$ ein T-invarianter Unterraum. Dann existiert ein Eigenvektor $v \in W$ von T.*

Da wir hauptsächlich an $K = \mathbb{C}$ denken wollen, lassen wir den Beweis dieses Lemmas vorerst noch weg. Bemerken Sie, dass die Aussage für $K = \mathbb{C}$ wahr ist: Jeder Endomorphismus hat mindestens einen Eigenwert und somit einen Eigenvektor. Wenden Sie dies nun auf $T|_W$ an.

Wir beginnen nun mit dem Beweis der Existenz. Die folgende Proposition ist ein grosser Schritt in diese Richtung:

Proposition 9.2.18. *Seien V und $T \in \operatorname{End}(V)$ wie in Satz 9.1.10 (insbesondere zerfällt p_T in Linearfaktoren in $K[x]$) und sei $V = W_1 \oplus \cdots \oplus W_k$ eine Zerlegung von V in unzerlegbare Untervektorräume (die laut Lemma 9.2.11 existiert). Dann existiert für jedes $i = 1, \ldots, k$ ein $\lambda_i \in K$ mit*

$$T|_{W_i} = \lambda_i \operatorname{Id}_{W_i} + N_i,$$

wobei $N_i : W_i \to W_i$ ein nilpotenter Endomorphismus ist.

Beweis. Sei $i \in \{1, \ldots, k\}$ und betrachten wir $T|_{W_i} : W_i \to W_i$. Laut Lemma 9.2.17 hat $T|_{W_i}$ mindestens einen Eigenvektor. Sei also $v \in W_i$ ein Eigenvektor von T zum Eigenwert λ. Betrachten wir $S : W_i \to W_i$, definiert durch $S := T|_{W_i} - \lambda \operatorname{Id}_{W_i}$. Bemerken Sie, dass $ST = TS$ und sei $k \in \mathbb{N}$ die Zahl vom Fitting-Lemma 9.2.14 mit

$$W_i = \operatorname{Im}(S^k) \oplus \ker(S^k).$$

Laut Korollar 9.2.16 sind diese beiden Unterräume T-invariant, aber W_i ist unzerlegbar und daher muss einer von den beiden gleich $\{0\}$ sein. In diesem Fall können wir herausfinden, welcher von beiden es ist! Es gilt nämlich, dass der Eigenvektor $0 \neq v \in \ker(T|_{W_i} - \lambda \operatorname{Id}_{W_i}) = \ker(S) \subseteq \ker(S^k)$ und daher muss $\operatorname{Im}(S^k) = 0$ sein. Es folgt, dass $\ker(S^k) = W_i$. In anderen Worten heisst das, dass S nilpotent ist. Die Proposition folgt jetzt aus der Definition von S zusammen mit $N_i = S$. □

Proposition 9.2.18 bedeutet, dass der Beweis von Satz 9.2.1 (zumindest über $\mathbb{C}$) abgeschlossen sein wird, sobald wir nilpotente Endomorphismen besser verstehen. Dazu brauchen wir einen neuen Unterabschnitt.

9.2.1 Darstellungen von nilpotenten Endomorphismen

Satz 9.2.19 (Jordan Normalform für Nilpotente Abbildungen). *Sei $N \in \mathrm{End}(V)$ ein nilpotenter Endomorphismus auf einem n-dimensionalen Vektorraum V. Dann existiert eine Basis $\mathcal{B} \subseteq V$ mit*

$$[N]_{\mathcal{B}} = \begin{pmatrix} J_{0,n_1} & & \\ & \ddots & \\ & & J_{0,n_k} \end{pmatrix},$$

wobei $\sum_{i=1}^{k} n_i = n$.

Vorbereitungen für den Beweis

Seien N und V wie im Satz oben. Für jedes $0 \neq v \in V$ existiert $\ell = \ell(v) \in \mathbb{N} \cup \{0\}$, sodass $N^{\ell} v \neq 0$ und $N^{\ell+1} v = 0$. Wir nennen $\ell = \ell(v)$ die *Lebensdauer* von v bezüglich N. Manchmal schreiben wir auch $\ell_N(v)$, um die Abhängigkeit von N zu betonen.

Lemma 9.2.20. *Sei $0 \neq v \in V$ mit Lebensdauer $\ell = \ell_N(v)$. Für alle $0 \leq i \leq \ell$ ist die Lebensdauer von $N^i(v)$ gleich $\ell - i$.*

Beweis. Wir haben $N^{\ell-i}(N^i(v)) = N^{\ell}(v) \neq 0$ und $N^{\ell-i+1}(N^i(v)) = N^{\ell+1}(v) = 0$. □

Wir schreiben $Z_v := \mathrm{Sp}(v, Nv, \ldots, N^{\ell} v)$ für $v \in V \setminus \{0\}$.

Lemma 9.2.21. *Die Vektoren $v, Nv, \ldots, N^{\ell} v$, mit $\ell = \ell_N(v)$, sind linear unabhängig und bilden daher eine Basis von Z_v.*

Beweis. Wir führen einen Widerspruchsbeweis. Nehmen wir also an, dass $\alpha_0, \ldots, \alpha_\ell \in K$ existieren, die nicht alle gleich 0 sind, mit

$$\alpha_0 v + \cdots + \alpha_\ell N^{\ell} v = 0.$$

Sei $i_0 = \min\{i \mid \alpha_i \neq 0\}$. Wir haben

$$N^{i_0}(v) = - \sum_{j=i_0+1}^{\ell} \frac{\alpha_j}{\alpha_{i_0}} N^j(v). \tag{9.7}$$

Laut Lemma 9.2.20 ist die Lebensdauer von $N^{i_0}(v)$ gleich $\ell - i_0$. Aus (9.7) folgt jedoch

$$N^{\ell-i_0}(N^{i_0}(v)) = - \sum_{j=i_0+1}^{\ell} \frac{\alpha_j}{\alpha_{i_0}} N^{\ell-i_0+j}(v) = - \sum_{k=1}^{\ell-i_0} \frac{\alpha_{i_0+k}}{\alpha_{i_0}} \underbrace{N^{\ell+k}(v)}_{=0} = 0,$$

was ein Widerspruch ist. □

Für $v \in V$ schreiben wir $\mathcal{B}_v := (N^{\ell(v)}v, \ldots, Nv, v)$. Bemerken Sie, dass Z_v ein N-invarianter Untervektorraum von V ist und gemäss der Definition der Darstellungsmatrix ist $[N|_{Z_v}]_{\mathcal{B}_v} = J_{0,\ell(v)+1}$.

Übung 9.2.22. Sei $N \in \mathrm{End}(V)$ nilpotent mit Index k, das heisst $N^{k-1} \neq 0$ und $N^k = 0$. Zeigen Sie, dass $\ell(v) \leq k-1$ für alle $v \in V$ und dass $v \in V$ existiert mit $\ell(v) = k-1$.

Übung 9.2.23. Sei $N \in \mathrm{End}(V)$ nilpotent mit Index n, wobei $n = \dim V$. Beweisen Sie Satz 9.2.19 für N. Tun Sie dies, indem Sie die folgende stärkere Aussage beweisen: Es existiert eine Basis $\mathcal{B}$ von V mit $[N]_{\mathcal{B}} = J_{0,n}$.

Bemerken Sie, dass wir nichts allzu kompliziertes gemacht haben bis jetzt. Die Existenz der Jordan Normalform zu zeigen ist allerdings nicht einfach, also werden wir früher oder später noch ins Schwitzen kommen. Dies wird im nächsten Beweis der Fall sein. Wir werden Sie an der entsprechenden Stelle nochmals warnen.

Beweis von Satz 9.2.19. Es genügt, zu zeigen (es ist genau genommen sogar äquivalent), dass $v_1, \ldots, v_k \in V$ existieren, sodass die Hintereinanderreihung von $\mathcal{B}_{v_1}, \ldots, \mathcal{B}_{v_k}$ eine Basis von V ist. Bemerken Sie, dass dies äquivalent dazu ist, dass $V = Z_{v_1} \oplus \cdots \oplus Z_{v_k}$. Mit der Hintereinanderreihung meinen wir

$$\mathcal{B} = \left(N^{\ell(v_1)}v_1, \ldots, Nv_1, v_1, N^{\ell(v_2)}v_2, \ldots, Nv_2, v_2, \ldots, N^{\ell(v_k)}v_k, \ldots, Nv_k, v_k\right).$$

Tatsächlich, falls $\mathcal{B}$ eine Basis von V ist, dann hat $[N]_{\mathcal{B}}$ genau die gewünschte Form aus Satz 9.2.19 mit $n_i = \ell_N(v_i) + 1$ für alle $i = 1, \ldots, k$.

Wir beweisen, dass $\mathcal{B}$ eine Basis von V ist, mit einer starken Induktion über $\dim V$. Die Leser sind dazu eingeladen, die Fälle $\dim V = 1$ und $\dim V = 2$ als Übung selbst zu beweisen. Wir machen den Induktionsschritt:

Sei $v_1 \in V$ ein Vektor mit der längstmöglichen Lebensdauer $\ell = \ell_N(v_1)$. Betrachten wir

$$V_1 = Z_{v_1} = \mathrm{Sp}(v_1, \ldots, N^{\ell}v_1),$$

welches ein Untervektorraum der Dimension $\ell+1$ ist. Bemerken Sie, dass V_1 N-invariant ist, also können wir gemäss Übung 5.4.9 (1) die induzierte Abbildung

$$N_{V/V_1} : V/V_1 \to V/V_1$$

betrachten. Für eine bessere Lesbarkeit schreiben wir $\overline{N} := N_{V/V_1}$, sowie

$$\pi = \pi_{V/V_1} : V \to V/V_1 =: \overline{V}$$

für die kanonische Projektion. Bemerken Sie, dass

$$\overline{N}^i \circ \pi = \pi \circ N^i, \tag{9.8}$$

was bedeutet, dass das Diagramm

$$\begin{tikzcd} V \arrow[r, "N^i"] \arrow[d, "\pi"] & V \arrow[d, "\pi"] \\ V/V_1 \arrow[r, "\overline{N}^i"] & V/V_1 \end{tikzcd} \tag{9.9}$$

kommutiert. Folglich ist $\overline{N}$ nilpotent, da N nilpotent und π surjektiv ist. Laut Lemma 9.2.21 ist $(v_1, \ldots, N^\ell v_1)$ eine Basis von V_1, also gilt $\dim(V/V_1) < \dim V$. Aus der starken Induktionsannahme folgt, dass $u_2, \ldots, u_k \in \overline{V}$ existieren, sodass $\mathcal{B}_{u_2} \sqcup \cdots \sqcup \mathcal{B}_{u_k}$ eine Basis von $\overline{V}$ ist. Diese Aussage bezieht sich natürlich auf $\overline{N}$, das heisst

$$\mathcal{B}_{u_i} = (\overline{N}^{\ell_{\overline{N}}(u_i)} u_i, \ldots, u_i).$$

Behauptung: Es existieren $v_2, \ldots, v_k \in V$ mit $\pi(v_i) = u_i$ und $\ell_N(v_i) = \ell_{\overline{N}}(u_i)$. Das heisst wir können die $u_i \in \overline{V}$ zu Vektoren $v_i \in V$ hochheben, welche die gleiche Lebensdauer haben.

Beweis des Satzes unter Annahme der Behauptung. Seien $v_2, \ldots, v_k$ wie in der Behauptung. Bemerken Sie, dass laut (9.8) die Hintereinanderreihung $\mathcal{B}_{v_2} \sqcup \cdots \sqcup \mathcal{B}_{v_k}$ genau auf die Basis $\mathcal{B}_{u_2} \sqcup \cdots \sqcup \mathcal{B}_{u_k}$ projiziert wird. Ausserdem hat diese Menge genau $\dim V - \dim V_1 = \dim(\overline{V})$ Elemente und darum hat $\mathcal{B} = \mathcal{B}_{v_1} \sqcup \mathcal{B}_{v_2} \sqcup \cdots \sqcup \mathcal{B}_{v_k}$ genau $\dim V$ Elemente. Ausserdem ist $\mathrm{Sp}(\mathcal{B}) = V$ (wieso?), also ist $\mathcal{B}$ eine Basis von V, bezüglich welcher N die gewünschte Darstellungsmatrix hat.

Beweis der Behauptung. Da wir hier ins Schwitzen kommen werden, skizzieren wir zuerst, wie der Beweis aussieht: Wir zeigen, dass wir ein beliebiges $u \in \overline{V}$ zu einem $v \in V$ hochheben können (d.h. $v \in V$ finden mit $\pi(v) = u$), sodass v die gleiche Lebensdauer wie u hat. Dies kann man dann auf $u_2, \ldots, u_n$ anwenden.

Nehmen wir an, die Lebensdauer von v_1 wäre 9, also $N^9 v_1 \neq 0$ und $N^{10} v_1 = 0$. Nehmen wir weiter an, wir hätten ein $u \in V/V_1$ mit Lebensdauer 7 und wir möchten dies zu einem $v \in V$ mit Lebensdauer 7 hochheben. Die Kommutativität des Diagramms (9.9) sagt uns, dass jedes solche v eine Lebensdauer von mindestens 7 haben muss. Es könnte allerdings sein, dass $N^8 v \neq 0$ ist. In diesem Fall wissen wir aber, dass

$$N^8 v \in \ker \pi = V_1 = Z_{v_1} = \mathrm{Sp}(v_1, \ldots, N^9 v_1),$$

da die $\pi N^8 v = \overline{N}^8 u = 0$. Wir möchten dann unsere Wahl von v korrigieren zu v', sodass $\pi(v') = u$ und $N^8 v' = 0$. An dieser Stelle benutzen wir, dass wir v_1 als einen Vektor mit maximaler Lebensdauer gewählt haben. Da alle Vektoren Lebensdauer ≤ 9 haben, wird $N^8 v$ nach (spätestens) 2 weiteren Anwendungen von N verschwinden. Die einzigen Vektoren in der Basis $\mathcal{B}_{v_1}$ von V_1, die nach einer Anwendung von N^2 verschwinden, sind $N^8 v_1$ und $N^9 v_1$. Also können wir $N^8 v$ als Linearkombination von diesen schreiben, das heisst

$$N^8 v = \alpha N^8 v_1 + \beta N^9 v_1 = N^8(\alpha v_1 + \beta N v_1).$$

Wenn wir nun also v durch $\alpha v_1 + \beta N v_1 \in V_1$ korrigieren, also $v' := v - (\alpha v_1 + \beta N v_2)$ definieren, erhalten wir einen Vektor durch Anwendung von N^8 verschwindet:

$$N^8(v') = N^8(v - (\alpha v_1 + \beta N v_1)) = N^8 v - N^8(\alpha v_1 + \beta N v_1) = 0$$

und $\pi(v') = u$, wie wir zeigen wollten. Fangen wir nun mit dem richtigen Beweis der Behauptung an.

Erinnern Sie sich, dass wir v_1 so gewählt haben, dass $\ell := \ell_N(v_1)$ maximal ist unter allen Vektoren in V. Sei $u \in \overline{V}$ beliebig und $v \in V$ ein Vektor mit $\pi(v) = u$ (das heisst, $u = v + V_1$). Für jedes $m \in \mathbb{N} \cup \{0\}$ haben wir

$$\pi \circ N^m(v) = \overline{N}^m \circ \pi(v) = \overline{N}^m(u), \tag{9.10}$$

also ist $\ell_N(v) \geq \ell_{\overline{N}}(u)$. Wir schreiben $\overline{\ell} := \ell_{\overline{N}}(u)$ und betrachten den Vektor $w := N^{\overline{\ell}+1}(v) \in \ker \pi = V_1$. Da ℓ die maximale Lebensdauer ist, muss w nach $\ell - \overline{\ell}$ Anwendungen von N verschwinden:

$$N^{\ell-\overline{\ell}}(w) = N^{\ell-\overline{\ell}}(N^{\overline{\ell}+1}(v))) = N^{\ell+1}(v) = 0,$$

denn sonst hätten wir einen Vektor gefunden mit einer längeren Lebensdauer als ℓ. Also ist $w \in \ker N^{\ell-\overline{\ell}} \cap V_1$. Wenn wir uns erinnern, dass $\mathcal{B}_{v_1} = (N^\ell v_1, \ldots, N v_1, v_1)$ eine Basis von V_1 ist, dann folgt, dass w eine Linearkombination der ersten $\ell - \overline{\ell}$ Vektoren aus $\mathcal{B}_{v_1}$ ist (da alle restlichen Vektoren die Anwendung von $N^{\ell-\overline{\ell}}$ überleben). Wir schreiben

$$w = \alpha_1 N^\ell v_1 + \cdots + \alpha_{\ell-\overline{\ell}} \underbrace{N^{\ell-(\ell-\overline{\ell})+1} v_1}_{=N^{\overline{\ell}+1}(v_1)} = N^{\overline{\ell}+1}(\underbrace{\alpha_1 N^{\ell-\overline{\ell}-1} v_1 + \cdots + \alpha_{\ell-\overline{\ell}} v_1}_{r:=}).$$

Das heisst, wenn wir v durch r korrigieren, erhalten wir einen Vektor, welcher die gleiche Lebensspanne wie u hat. Überprüfen wir das: Wir definieren $v' := v - r$. Da $r \in V_1$ ist, ist $\pi(v') = \pi(v) = u$ und wir behaupten, dass $\ell_N(v') = \overline{\ell} = \ell_{\overline{N}}(u)$:

$$N^{\overline{\ell}+1}(v') = N^{\overline{\ell}+1}(v) - N^{\overline{\ell}+1}(r) = w - w = 0,$$

also ist $\ell_N(v') \leq \overline{\ell}$. Da $\pi(v') = u$ ist, folgt aus dem gleichen Argument (9.10), angewandt auf v', dass $\ell_N(v') \geq \overline{\ell}$. Also ist $\ell_N(v') = \overline{\ell}$.

Wenn wir das obige auf $u = u_i$ für alle $i = 2, \ldots, k$ anwenden, beendet dies den Beweis der Behauptung und somit den Beweis von Satz 9.2.19. □

9.2.2 Die Puzzleteile zusammenfügen

Um ganz sicher zu sein, schreiben wir, wieso die Existenz in Satz 9.1.10 gilt, oder anders gesagt, wieso Satz 9.2.1 gilt:

Beweis von Satz 9.2.1. Seien W_i die unzerlegbaren T-invarianten Summanden, die wir von Lemma 9.2.11 erhalten haben. Laut Proposition 9.2.18 ist $T|_{W_i} = \lambda_i \operatorname{Id}_{W_i} + N_i$ für $\lambda_i \in K$ und $N_i \in \operatorname{End}(W_i)$ nilpotent. Sei $\mathcal{B}_i$ eine Basis von W_i wie in Satz 9.2.19 für N_i. Dann ist

$$[T|_{W_i}]_{\mathcal{B}_i} = \lambda_i I_{\dim W_i} + [N_i]_{\mathcal{B}_i} = \begin{pmatrix} \lambda_i & & \\ & \ddots & \\ & & \lambda_i \end{pmatrix} + \begin{pmatrix} J_{0,*} & & \\ & \ddots & \\ & & J_{0,*} \end{pmatrix} = \begin{pmatrix} J_{\lambda_i,*} & & \\ & \ddots & \\ & & J_{\lambda_i,*} \end{pmatrix},$$

wobei $*$ hier für irgendwelchen natürlichen Zahlen steht, deren Wert im Moment unwichtig ist. Wenn wir nun $\mathcal{B}$ als die Hintereinanderreihung der $\mathcal{B}_i$'s definieren, beendet das den Existenzteil des Beweises von Satz 9.1.10. □

9.2.3 Beweis von Lemma 9.2.17

Das folgende Lemma wird trivial sein, sobald Sie die Grundlagen der Ringtheorie gemeistert haben, welche Sie in der Vorlesung Algebra I lernen werden[5]. Der Vollständigkeit halber geben wir hier einen Beweis, der nur Folgerungen der Polynomdivision verwendet.

Lemma 9.2.24. *Falls $p \in K[x]$ ein anderes Polynom in $K[x]$ teilt, welches in Linearfaktoren zerfällt in $K[x]$, dann zerfällt auch p in Linearfaktoren in $K[x]$. Das heisst, falls $p \mid \alpha \prod_{i=1}^{n}(x - a_i)$ mit $\alpha, a_1, \ldots, a_n \in K$, dann ist $p = \beta \prod_{i=1}^{k}(x - b_i)$ mit $\beta, b_1, \ldots, b_k \in K$ (wobei k der Grad von p ist).*

Beweis. Wir beweisen es mit Induktion über $k = \deg p$. Für $k = 1$ gibt es nichts zu beweisen. Nehmen wir also an, dass die Aussage für Polynome vom Grad $k - 1$ stimmt

[5] Für die interessierten Leser: Der Beweis des Lemmas wird sehr leicht, wenn man verwendet, dass $K[x]$ ein Faktorieller Ring ist.

und betrachten wir p mit $\deg p = k$, sodass

$$p \mid \alpha \prod_{i=1}^{n}(x - a_i) = A(x),$$

das heisst es existiert $q \in K[x]$ mit $p(x)q(x) = A(x)$. Für jedes $i = 1, \dots, n$ ist

$$0 = A(a_i) = p(a_i)q(a_i), \quad \text{also } p(a_i) = 0 \text{ oder } q(a_i) = 0.$$

Erinnern Sie sich, dass laut Korollar 1.4.17 ein Polynom vom Grad n höchstens n verschiedene Nullstellen haben kann. Durch eine leichte Abänderung des Beweises kann man zeigen, dass das gleiche gilt, wenn wir gleiche Nullstellen zulassen und dafür die Vielfachheit jeder Nullstelle zählen. Ein Polynom von Grad n kann also höchstens n Nullstellen haben, mit Vielfachheit gezählt.

Wenn wir mit Vielfachheit zählen, kann $q(a_i) = 0$ für höchstens $n - k$ der a_i's gelten kann, da $\deg q = n - k$. Deshalb (da $k \geq 1$) gibt es ein $i_0 \in \{1, \dots, n\}$ mit $p(a_{i_0}) = 0$. Ohne Beschränkung der Allgemeinheit können wir annehmen, dass $i_0 = 1$. Also gilt laut Korollar 1.4.15 $(x - a_1) \mid p$ und deshalb

$$p' = \frac{p}{(x - a_1)} \mid \alpha \prod_{i=2}^{n}(x - a_i).$$

Aus der Induktionsannahme folgt, dass $p' = \beta \prod_{i=1}^{k-1}(x - b_i)$ mit $\beta, b_i \in K$ und somit ist

$$p = \beta(x - a_1) \prod_{i=1}^{k-1}(x - b_i),$$

was den Beweis beendet. □

Korollar 9.2.25. *Sei $T \in \operatorname{End}(V)$, mit $\dim V = n$, sodass p_T in Linearfaktoren zerfällt. Dann gilt für jeden T-invarianten Unterraum $W \subseteq V$, dass $p_{T|_W}$ in Linearfaktoren zerfällt. Insbesondere hat $T|_W$ einen Eigenvektor.*

In anderen Worten, Lemma 9.2.17 ist wahr.

Beweis. Aus Übung 5.4.9 (2c) wissen wir, dass $p_{T|_W} \mid p_T$, also können wir Lemma 9.2.24 anwenden und erhalten, dass $p_{T|_W}$ in Linearfaktoren zerfällt. Insbesondere heisst das, dass $p_{T|_W}$ mindestens eine Nullstelle hat und somit $T|_W$ einen Eigenvektor hat. □

9.3 Eindeutigkeit der Jordan Normalform

Unsere Spielerei mit vielen Beispielen im Abschnitt 9.1 wird uns in den nächsten zwei Abschnitten sehr zugute kommen. Falls Sie diese Beispiele noch nicht gut verstanden haben, wäre es sicher eine gute Idee, sich diese nochmals genau anzuschauen, bevor Sie hier weiterlesen. Wir werden in diesem Abschnitt jeweils I für die Identitätsmatrix schreiben, wobei die Dimension aus dem Kontext heraus verstanden werden muss. Ebenso schreiben wir Id für die Identitätsabbildung eines beliebigen Vektorraums, wenn der Vektorraum aus dem Kontext klar ist. Unser Ziel ist es, Eigenschaften der Blöcke in der Jordan Normalform zu finden, welche nur von T abhängen, also unabhängig von der Wahl der Basis sind. Für $T \in \mathrm{End}(V)$ sei $r_k(\lambda, T) := \mathrm{Rang}((T - \lambda\,\mathrm{Id})^k)$ und für eine Matrix $A \in M_{n\times n}(K)$ sei $r_k(\lambda, A) = \mathrm{Rang}((A - \lambda I)^k)$. Falls $A = [T]_{\mathcal{C}}$ für eine Basis $\mathcal{C}$, dann ist $r_k(\lambda, T) = r_k(\lambda, A)$, da

$$[(T - \lambda\,\mathrm{Id})^k]_{\mathcal{C}} = [T - \lambda\,\mathrm{Id}]_{\mathcal{C}}^k = ([T]_{\mathcal{C}} - \lambda[\mathrm{Id}]_{\mathcal{C}})^k = (A - \lambda I)^k \tag{9.11}$$

und weil der Rang einer linearen Abbildung gleich dem Rang einer beliebigen Darstellungsmatrix ist.

Wir möchten zuerst (ganz im Sinne der Beispiele von Abschnitt 9.1) $r_k(\lambda, J_{\mu,n})$ für einen Jordanblock verstehen: Falls $\mu = \lambda$, dann ist

$$r_k(\lambda, J_{\lambda,n}) = \mathrm{Rang}((J_{\lambda,n} - \lambda I_n)^k) = \mathrm{Rang}((J_{0,n})^k) = \begin{cases} n-k & \text{für } k = 1, \dots, n, \\ 0 & \text{für } k \geq n. \end{cases} \tag{9.12}$$

Falls $\mu \neq \lambda$, dann ist $r_k(\lambda, J_{\mu,n}) = \mathrm{Rang}((J_{\lambda-\mu,n})^k) = n$ für alle k, weil $J_{\lambda-\mu,n}$ invertierbar ist.

Bemerken Sie nun, dass falls $A = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_\ell \end{pmatrix}$ eine Blockdiagonalmatrix mit Jordanblöcken J_i auf der Diagonalen ist, dann ist auch $(A - \lambda I)$ eine Blockdiagonalmatrix mit Jordanblöcken auf der Diagonalen[6] und wir haben

$$r_k(\lambda, A) = \mathrm{Rang}((A - \lambda I)^k) \overset{\text{Block}}{\underset{\text{Mat.}}{=}} \sum_{i=1}^{\ell} \mathrm{Rang}((J_i - \lambda I)^k) = \sum_{i=1}^{\ell} r_k(\lambda, J_i), \tag{9.13}$$

da der Rang einer Blockdiagonalmatrix gleich der Summe der Ränge aller Blöcke ist (in (9.13) bezeichnet I jeweils die Identitätsmatrix der entsprechenden Grösse).

Wir sind nun bereit für den Beweis der Eindeutigkeit der Jordan Normalform.

[6] Wenn $J_i = J_{\lambda_i,n_i}$, dann hat $(A - \lambda I)$ die Blöcke $J_{\lambda_i-\lambda,n_i}$ auf der Diagonalen.

Beweis der Eindeutigkeit in Satz 9.1.10. Sei $\mathcal{B}$ eine beliebige Jordanbasis von T, das heisst

$$[T]_{\mathcal{B}} = \begin{pmatrix} J_{\lambda_1,n_1} & & \\ & \ddots & \\ & & J_{\lambda_k,n_k} \end{pmatrix} \tag{9.14}$$

mit $\lambda_1, \ldots, \lambda_k \in K$, $n_1, \ldots, n_k \in \mathbb{N}$. Sei $\lambda \in K$. Wir behaupten, dass die Anzahl Blöcke mit Eigenwert λ und Grösse k (also Blöcke der Form $J_{\lambda,k}$) durch

$$B(\lambda, k, T) := r_{k+1}(\lambda, T) + r_{k-1}(\lambda, T) - 2r_k(\lambda, T)$$

gegeben ist und folglich nicht von der Basis $\mathcal{B}$ abhängt. Da $r_k(\lambda, T) = r_k(\lambda, A)$ (in (9.11) gezeigt) genügt es, diese Behauptung für A anstelle von T zu beweisen und aus (9.13) folgt $B(\lambda, k, A) = \sum_{i=1}^{k} B(\lambda, k, J_{\lambda_i,n_i})$, also genügt es, zu zeigen, dass

$$B(\lambda, k, J_{\lambda_i,n_i}) = \begin{cases} 0 & \text{falls } \lambda_i \neq \lambda, \\ 0 & \text{falls } \lambda_i = \lambda \text{ und } k \neq n_i, \\ 1 & \text{falls } \lambda_i = \lambda \text{ und } k = n_i. \end{cases} \tag{9.15}$$

Also beginnen wir zu rechnen. Nehmen wir zuerst an, dass $\lambda \neq \lambda_i$. Dann ist $J_{\lambda_i,n_i} - \lambda I = J_{\lambda-\lambda_i,n_i}$, was invertierbar ist. Also ist $\mathrm{Rang}((J_{\lambda_i,n_i} - \lambda_i I)^k) = n_i$ für beliebige k und somit ist

$$B(\lambda, k, J_{\lambda,n_i}) = n_i + n_i - 2n_i = 0.$$

Betrachten wir nun den Fall $\lambda = \lambda_i$. Für $k < n_i$ haben wir

$$\begin{aligned} B(\lambda, k, J_{\lambda,n_i}) &= r_{k+1}(\lambda, J_{\lambda,n_i}) + r_{k-1}(\lambda, J_{\lambda,n_i}) - 2r_k(\lambda, J_{\lambda,n_i}) \\ &\overset{(9.12)}{=} n_i - (k+1) + n_i - (k-1) - 2(n_i - k) = 0, \end{aligned}$$

für $k = n_i$

$$\begin{aligned} B(\lambda, n_i, J_{\lambda,n_i}) &= r_{n_i+1}(\lambda, J_{\lambda,n_i}) + r_{n_i-1}(\lambda, J_{\lambda,n_i}) - 2r_{n_i}(\lambda, J_{\lambda,n_i}) \\ &\overset{(9.12)}{=} 0 + (n_i - (n_i - 1)) - 2 \cdot 0 = 1 \end{aligned}$$

und für $k > n_i$

$$B(\lambda, k, J_{\lambda,n_i}) = r_{k+1}(\lambda, J_{\lambda,n_i}) + r_{k-1}(\lambda, J_{\lambda,n_i}) - 2r_k(\lambda, J_{\lambda,n_i}) \overset{(9.12)}{=} 0 + 0 - 2 \cdot 0 = 0.$$

Dies zeigt (9.15) und beendet somit den Beweis der Eindeutigkeit. □

Für zukünftige Referenz geben wir noch separat an, was wir in diesem Beweis gelernt haben:

Korollar 9.3.1. *Sei $T \in \operatorname{End}(V)$ mit $[T]_{\mathcal{B}}$ wie in (9.14). Dann hängt die Anzahl $B(\lambda, k, T)$ der Blöcke mit Eigenwert λ und Grösse k in $[T]_{\mathcal{B}}$ nur von T ab (also nicht von $\mathcal{B}$) und ist gleich*

$$\begin{aligned}
B(\lambda, k, T) &:= r_{k+1}(\lambda, T) + r_{k-1}(\lambda, T) - 2r_k(\lambda, T) \\
&= (\dim V - \dim \ker(T - \lambda \operatorname{Id})^{k+1}) + (\dim V - \dim \ker(T - \lambda \operatorname{Id})^{k-1}) \\
&\quad - 2(\dim V - \dim \ker(T - \lambda \operatorname{Id})^{k}) \\
&= 2 \dim \ker(T - \lambda \operatorname{Id})^{k} - \dim \ker(T - \lambda \operatorname{Id})^{k+1} - \dim \ker(T - \lambda \operatorname{Id})^{k-1}.
\end{aligned}$$

Beweis. Die erste Gleichheit (nach der Definition) folgt aus dem Rangsatz. Alles andere wurde oben gezeigt. □

Bemerkung 9.3.2. Falls Ihnen die Definition von $B(\lambda, k, T)$ immer noch rätselhaft ist, geben wir folgende Merkhilfe (vergleichen Sie dies mit Übung 9.1.7):

$$\begin{aligned}
B(\lambda, k, T) &= r_{k+1}(\lambda, T) + r_{k-1}(\lambda, T) - 2r_k(\lambda, T) \\
&= \underbrace{(r_{k-1}(\lambda, T) - r_k(\lambda, T))}_{\#\text{Blöcke der Grösse } \geq k} - \underbrace{(r_k(\lambda, T) - r_{k+1}(\lambda, T))}_{\#\text{Blöcke der Grösse } \geq k+1}
\end{aligned}$$

Dies schliesst den Beweis der Jordan Normalform ab. Wir haben gezeigt, dass es für jeden Endomorphismus $T \in \operatorname{End}(V)$, dessen charakteristisches Polynom in Linearfaktoren zerfällt, eine Darstellungsmatrix $[T]_{\mathcal{B}}$ gibt, welche in Jordanform ist und dass diese Matrix in Jordanform eindeutig ist bis auf die Reihenfolge der Blöcke. Wir beweisen nun ein paar Korollare.

Erinnern Sie sich, dass wir gesagt haben, dass $T \in \operatorname{End}(V)$ eine Jordan Normalform hat, wenn es eine Basis $\mathcal{B} \subseteq V$ gibt, sodass $[T]_{\mathcal{B}}$ in Jordanform ist.

Korollar 9.3.3. *Ein Endomorphismus $T \in \operatorname{End}(V)$ hat eine Jordan Normalform genau dann, wenn p_T in Linearfaktoren zerfällt.*

Beweis. Die Richtung $\Longleftarrow$ war der Existenzteil von Satz 9.1.10. Die Richtung $\Longrightarrow$ folgt ziemlich direkt aus dem Fakt, dass $p_T = p_{[T]_{\mathcal{B}}}$, denn für eine Matrix in Jordanform (allgemeiner für eine Dreiecksmatrix) zerfällt das charakteristische Polynom in Linearfaktoren. □

Insbesondere haben wir einen anderen Beweis für die schwierige Richtung von Satz 5.4.5 (T trigonalisierbar $\iff$ p_T zerfällt in Linearfaktoren) gegeben.

Für eine Matrix $A \in M_{n\times n}(K)$ sagen wir, dass A eine Jordan Normalform hat, falls m_A eine Jordan Normalform hat. Dies ist äquivalent dazu, dass A ähnlich zu einer Matrix in Jordanform ist.

Übung 9.3.4. Beweisen Sie die letzte Aussage.

Korollar 9.3.5. *Seien $A, B \in M_{n\times n}(K)$ zwei trigonalisierbare Matrizen (dies ist also für* beliebige *zwei Matrizen über $\mathbb{C}$ anwendbar). Dann sind A und B ähnlich genau dann, wenn sie die gleiche Jordan Normalform (bis auf Vertauschen der Blöcke) haben.*

Beweis. $\Longleftarrow$: Seien $P, Q \in \mathrm{GL}_n(K)$, sodass

$$P^{-1}AP = J_1 \quad \text{und} \quad Q^{-1}BQ = J_2,$$

wobei J_1 und J_2 Matrizen in Jordanform sind, die gleich sind bis auf eine Permutation der Blöcke. Dies bedeutet, dass es eine Permutationsmatrix P_σ gibt[7] mit $P_\sigma^{-1}J_1P_\sigma = J_2$ und damit ist $(PP_\sigma)^{-1}APP_\sigma = J_2 = Q^{-1}BQ$. Dies impliziert $(PP_\sigma Q^{-1})^{-1}A(PP_\sigma Q^{-1}) = B$, also sind A und B ähnlich.

$\Longrightarrow$: Wir geben zwei Beweise. Einer, welcher den obigen Beweis benutzt und einer, der mit der Transformationsformel aus dem ersten Semester spielt.

Erster Beweis: Falls A und B ähnlich sind, dann ist $(A - \lambda I)^k$ ähnlich zu $(B - \lambda I)^k$ (mit dem gleichen P), für beliebige $\lambda \in K$ und $k \in \mathbb{N}$. Also ist $\mathrm{Rang}((A - \lambda I)^k) = \mathrm{Rang}((B - \lambda I)^k)$, also ist gemäss Korollar 9.3.1 die Jordanblockstruktur für A und B die selbe.

Zweiter Beweis: Sei $P \in \mathrm{GL}_n(K)$ mit $P^{-1}AP = B$. Sei $\mathcal{B} = (v_1, \dots, v_n)$ eine Jordanbasis für m_B, das heisst

$$[m_B]_{\mathcal{B}} = Q^{-1}BQ = J \quad (\text{mit } Q = [\mathrm{Id}]^{\mathcal{B}}_{\mathcal{E}_n}),$$

also ist $J = (PQ)^{-1}APQ$. Wenn wir PQ als $[\mathrm{Id}]^{\mathcal{C}}_{\mathcal{E}_n}$ für eine Basis $\mathcal{C}$ (genauer gesagt ist $\mathcal{C} = (PQe_1, \dots, PQe_n)$) realisieren, dann ist

$$[m_A]_{\mathcal{C}} = [\mathrm{Id}]^{\mathcal{E}_n}_{\mathcal{C}} A [\mathrm{Id}]^{\mathcal{C}}_{\mathcal{E}_n} = (PQ)^{-1}APQ = J. \qquad \square$$

Das heisst, dass Matrizen in Jordanform eine vollständige Menge von Repräsentanten bezüglich der Äquivalenzrelation Ähnlichkeit in $M_{n\times n}(\mathbb{C})$ sind[8]. Zwei solche Repräsentanten sind ähnlich zueinander genau dann, wenn sie gleich sind bis auf die Reihenfolge der Blöcke. Wir werden dies in mehreren Beispielen im nächsten Abschnitt verwenden.

Als letztes Korollar der Jordan Normalform beweisen wir Satz 5.5.22.

[7] Falls $\sigma : \{1, \dots, n\} \to \{1, \dots, n\}$ eine Permutation ist, dann ist $P_\sigma = (p_{ij})$ mit

$$p_{ij} = \delta_{\sigma(i),j} = \begin{cases} 1 & \sigma(i) = j, \\ 0 & \text{sonst.} \end{cases}$$

[8] Wir nehmen es hier mit der Eindeutigkeit nicht ganz so genau. Wir sagen, dass A und B die *gleiche* Jordan Normalform haben, wenn sie die gleiche Jordan Normalform bis auf Vertauschen der Blöcke haben.

Übung 9.3.6. Sei J eine Matrix in Jordanform und seien $\lambda_1, \ldots, \lambda_k$ alle paarweise unterschiedlichen Eigenwerte von J. Sei n_i die Grösse des grössten Blocks mit Eigenwert λ_i. Dann ist

$$M_J = \prod_{i=1}^{k} (x - \lambda_i)^{n_i}.$$

Korollar 9.3.7. *Theorem 5.5.22 gilt, das heisst für $T \in \mathrm{End}(V)$ gilt*

$$T \text{ ist diagonalisierbar} \iff M_T \text{ zerfällt in paarweise verschiedene Linearfaktoren.}$$

Beweis. $\implies$: Falls $[T]_{\mathcal{B}}$ diagonal ist, dann ist $[T]_{\mathcal{B}}$ eine Jordanform von T, da jede Diagonalmatrix eine Matrix in Jordanform ist, bei der alle Blöcke die Grösse 1 haben. Es folgt aus Übung 9.3.6, dass

$$M_T = M_{[T]_{\mathcal{B}}} = \prod_{i=1}^{k} (x - \lambda_i),$$

wobei $\lambda_1, \ldots, \lambda_k$ die paarweise verschiedenen Eigenwerte von T sind.

$\impliedby$ (über $\mathbb{C}$): Angenommen $M_T = \prod_{i=1}^{k}(x - \lambda_i)$, wobei $\lambda_1, \ldots, \lambda_k$ paarweise verschieden sind. Sei J eine Jordan Normalform für T, also $[T]_{\mathcal{B}} = J$ (welche existiert, weil $K = \mathbb{C}$). Dann ist $M_T = M_J$ und aus Übung 9.3.6 folgt, dass alle Blöcke in J die Grösse 1 haben. Also ist J diagonal und somit T diagonalisierbar.

$\impliedby$ (K allgemein): Durch Betrachten des Beweises über $\mathbb{C}$ sehen wir, dass es genügt, zu zeigen, dass jeder Endomorphismus T, dessen Minimalpolynom in Linearfaktoren zerfällt, eine Jordan Normalform hat. Laut Korollar[9] 5.5.19 gilt $p_T \mid M_T^n$, also zerfällt p_T gemäss Lemma 9.2.24 in Linearfaktoren, somit hat T eine Jordan Normalform. □

Bemerkung 9.3.8. Man kann Satz 5.5.22 auch ohne die Jordan Normalform beweisen. Sehen Sie dafür zum Beispiel [10, Seite 299].

9.4 Die Jordan Normalform berechnen

Sei $T \in \mathrm{End}(V)$ (bzw. $A \in M_{n\times n}$) trigonalisierbar. Wir wissen, dass es eine geordnete Basis $\mathcal{B} \subseteq V$ (bzw. $\mathcal{B} \subseteq K^n$) gibt, sodass $J := [T]_{\mathcal{B}}$ (bzw. $J := [m_A]_{\mathcal{B}}$) in Jordanform ist. Diese Matrix J ist eindeutig bis auf eine Permutation der Blöcke. Es stellen sich zwei Hauptfragen:

(1) Wie finde ich J?

(2) Wie finde ich $\mathcal{B}$?

[9]Wir haben den Beweis für allgemeine K nur skizziert.

Beide fragen sind jetzt „leicht“ zu beantworten, aber im Allgemeinen erfordert dies (vor allem die Antwort auf die zweite Frage) viele Berechnungen. Um diese Fragen zu beantworten, fassen wir die Informationen aus den letzten drei Abschnitten zusammen.

Wir können die gesuchte Matrix J als

$$J = \begin{pmatrix} J_{\lambda_1, n_1^{(1)}} & & & & & & \\ & \ddots & & & & & \\ & & J_{\lambda_1, n_{\ell_1}^{(1)}} & & & & \\ & & & \ddots & & & \\ & & & & J_{\lambda_k, n_1^{(k)}} & & \\ & & & & & \ddots & \\ & & & & & & J_{\lambda_k, n_{\ell_k}^{(k)}} \end{pmatrix} \tag{9.16}$$

schreiben, wobei $\lambda_1, \ldots, \lambda_k$ die paarweise verschiedenen Eigenwerte von T (bzw. von A) sind. Für jedes $1 \leq i \leq k$ sind $n_1^{(i)}, \ldots, n_{\ell_i}^{(i)}$ die Grössen der Jordanblöcke mit Eigenwert λ_i. Wir können auch annehmen (wenn wir wollen), dass für alle i

$$n_1^{(i)} \geq n_2^{(i)} \geq \cdots \geq n_{\ell_i}^{(i)},$$

gilt. Das heisst, die Blöcke zu jedem Eigenwert sind der Grösse nach absteigend geordnet. In Abschnitt 9.3 haben wir gesehen, wie wir die Blockstruktur von T (bzw. A) ablesen können: Für $i = 1, \ldots, k$ betrachtet man $N_i := T - \lambda_i \operatorname{Id}_V$ (bzw. $N_i := A - \lambda_i I_n$) und berechnet die Zahlen $n_j^{(i)}$ mithilfe von

$$B(\lambda_i, k, T) = 2 \dim \ker(T - \lambda_i \operatorname{Id})^k - \dim \ker(T - \lambda_i \operatorname{Id})^{k+1} - \dim \ker(T - \lambda_i \operatorname{Id})^{k-1}, \tag{9.17}$$

wobei $B(\lambda_i, k, T)$ die Anzahl Jordanblöcke mit Eigenwert λ_i und Grösse k ist (bzw. mit A statt T). Um zu wissen, wann man mit der Berechnung stoppen muss, kann man bemerken, dass

$$m_a(\lambda_i) = \sum_{j=1}^{\ell_i} n_j^{(i)}.$$

Dies beantwortet die Frage (1). In der Praxis muss man oft nicht allzu viel berechnen. In vielen Fällen reicht es sogar, (9.17) mit $k = 1$ zu benutzen, was dann zum folgenden nützlichen Fakt reduziert wird:

$$\begin{aligned}\ell_i \overset{\text{Def.}}{=} \text{Anzahl Blöcke mit Eigenwert } \lambda_i = \dim\ker(N_i) &\overset{\text{Def.}}{=} \dim\ker(T-\lambda_i \operatorname{Id}) \\ &= \dim \operatorname{Eig}_T(\lambda_i) \\ &= m_g(\lambda_i),\end{aligned}$$

oder kürzer

$$\ell_i = \dim \operatorname{Eig}_T(\lambda_i) = m_g(\lambda_i).$$

Man berechnet sowieso normalerweise die Eigenräume.

Beispiel 9.4.1. Finden Sie die Jordan Normalform für die Matrix

$$A = \begin{pmatrix} 7 & 1 & & & \\ & 7 & 2 & & \\ & & 7 & 3 & \\ & & & 7 & 4 \\ & & & & 7 \end{pmatrix} \in M_{5\times 5}(\mathbb{Q}).$$

Lösung: Hier ist $k = 1$, $\lambda_1 = 7$. Wir berechnen zuerst

$$\begin{aligned}\dim \operatorname{Eig}_A(7) &= \dim\ker(A - 7I_5) \\ &= \dim\ker \begin{pmatrix} 0 & 1 & & & \\ & 0 & 2 & & \\ & & 0 & 3 & \\ & & & 0 & 4 \\ & & & & 0 \end{pmatrix} = 5 - \operatorname{Rang} \begin{pmatrix} 0 & 1 & & & \\ & 0 & 2 & & \\ & & 0 & 3 & \\ & & & 0 & 4 \\ & & & & 0 \end{pmatrix} = 1.\end{aligned}$$

Also hat die Jordan Normalform von A nur einen Block zum Eigenwert 7 und sonst nichts. Folglich ist A ähnlich zu ihrer Jordan Normalform

$$J_{7,5} = \begin{pmatrix} 7 & 1 & & & \\ & 7 & 1 & & \\ & & 7 & 1 & \\ & & & 7 & 1 \\ & & & & 7 \end{pmatrix}.$$

Beispiel 9.4.2. Ist die Matrix

$$B = \begin{pmatrix} 7 & 1 & 2 & 3 & 4 \\ & 7 & 5 & 6 & 7 \\ & & 7 & 8 & 9 \\ & & & 7 & 10 \\ & & & & 7 \end{pmatrix}$$

ähnlich zu der Matrix A von Beispiel 9.4.1?

Antwort: Ja! Dasselbe Argument wie in Beispiel 9.4.1 zeigt, dass $B \sim J_{7,5}$. Dafür zeigt man wie oben, dass $\dim \mathrm{Eig}_B(7) = 1$. Alle oberen 5×5-Dreiecksmatrizen mit nur 7 auf der Diagonalen und Einträgen $\neq 0$ auf der zweiten Diagonale (damit meinen wir die $(i, i+1)$ Einträge) sind ähnlich zueinander.

Beispiel 9.4.3. Sei

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & -1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \in M_{5\times 5}(\mathbb{C}).$$

Man berechnet $p_A = -x^5$ und lernt daraus, dass A nilpotent ist. Man berechnet, dass $\dim \mathrm{Eig}_A(0) = \dim \ker A = 2$ (es ist ziemlich leicht zu sehen, dass $\mathrm{Rang}(A) = 3$). Daher hat die Jordan Normalform von A zwei Blöcke. Sie ist also entweder $J_1 = \begin{pmatrix} J_{0,3} & \\ & J_{0,2} \end{pmatrix}$ oder $J_2 = \begin{pmatrix} J_{0,4} & \\ & J_{0,1} \end{pmatrix}$. Um das herauszufinden, müssen wir leider weiterrechnen. Anstelle der Formel (9.17) kann man das Minimalpolynom berechnen: $A^2 \neq 0$ aber $A^3 = 0$, also ist $M_A = x^3$. Dies bedeutet, dass die Jordan Normalform von A nicht J_2 sein kann, da $M_{J_2} = x^4$. Daher gilt $A \sim J_1$.

Die Aussage mit dem Minimalpolynom aus dem letzten Beispiel ist nützlich. Ein Endomorphismus T (bzw. eine Matrix A) wie zu Beginn des Abschnitts (also mit Jordan Normalform wie in (9.16)) hat das charakteristische Polynom

$$p_T(x) = \prod_{i=1}^{k} (x - \lambda_i)^{m_a(\lambda_i)},$$

wobei $m_a(\lambda_i) = \sum_{j=1}^{\ell_i} n_j^{(i)}$, und das Minimalpolynom

$$M_T = \prod_{i=1}^{k} (x - \lambda_i)^{r_i},$$

wobei $r_i = \max\{n_1^{(i)}, \ldots, n_{\ell_i}^{(i)}\}$.

Bemerkung 9.4.4. Erinnern Sie sich, dass wir gesagt haben, dass wir die Zahlen $n_k^{(i)}$ der Grösse nach ordnen können. In diesem Fall gilt $r_i := n_1^{(i)}$.

Übung 9.4.5. Benutzen Sie die obige Information über das Minimalpolynom, um einen Beweis für Satz 5.5.5 (Cayley-Hamilton) für $A \in M_{n\times n}(\mathbb{C})$ (oder $T \in \mathrm{End}(V)$ für einen komplexen Vektorraum V) zu geben.

Beispiel 9.4.6. Zeigen Sie, dass alle komplexen 3×3-Matrizen mit dem gleichen charakteristischen Polynom und Minimalpolynom ähnlich sind.
Lösung: Seien $A, B \in M_{3\times 3}(\mathbb{C})$ mit $p = p_A = p_B$ und $M = M_A = M_B$. Sie müssen also die gleichen Eigenwerte haben. Falls sie 3 unterschiedliche Eigenwerte $\lambda_1, \lambda_2, \lambda_3$ haben, sind beide ähnlich zu

$$J = \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{pmatrix}.$$

Nehmen wir nun an, dass sie zwei unterschiedliche Eigenwerte λ_1, λ_2 haben. Da $\deg p = 3$ ist und λ_1, λ_2 die Nullstellen von p sind, muss eins der beiden das algebraische Vielfachheit 2 haben und der andere 1. Ohne Beschränkung der Allgemeinheit können wir annehmen, dass $m_a(\lambda_1) = 2$ und $m_a(\lambda_2) = 1$ ist. Dann gilt

$$p = (-1)(x - \lambda_1)^2(x - \lambda_2) \quad \text{und} \quad M = (x - \lambda_1)^k(x - \lambda_2),$$

wobei $k \in \{1, 2\}$. Erinnern Sie sich (Übung 9.3.6), dass k die Grösse des grössten Jordanblocks zum Eigenwert λ_1 ist. Somit sind A und B beide ähnlich zu

$$\begin{pmatrix} \lambda_1 & & \\ & \lambda_1 & \\ & & \lambda_2 \end{pmatrix} \text{ falls } k = 1,$$

$$\begin{pmatrix} J_{\lambda_1,2} & \\ & J_{\lambda_2,1} \end{pmatrix} = \begin{pmatrix} \lambda_1 & 1 & \\ & \lambda_1 & \\ & & \lambda_2 \end{pmatrix} \text{ falls } k = 2.$$

Nehmen wir nun an, dass A und B nur einen Eigenwert λ haben. Dann ist

$$p = -(x - \lambda)^3 \quad \text{und} \quad M = (x - \lambda)^k,$$

wobei $k \in \{1,2,3\}$ die Grösse des Grössten Jordanblocks ist. Folglich sind A und B beide ähnlich zu

$$\begin{pmatrix} J_{\lambda,1} & & \\ & J_{\lambda,1} & \\ & & J_{\lambda,1} \end{pmatrix} = \begin{pmatrix} \lambda & & \\ & \lambda & \\ & & \lambda \end{pmatrix} \text{ falls } k=1,$$

$$\begin{pmatrix} J_{\lambda,2} & \\ & J_{\lambda,1} \end{pmatrix} = \begin{pmatrix} \lambda & 1 & \\ & \lambda & \\ & & \lambda \end{pmatrix} \text{ falls } k=2,$$

$$J_{\lambda,3} = \begin{pmatrix} \lambda & 1 & \\ & \lambda & 1 \\ & & \lambda \end{pmatrix} \text{ falls } k=3.$$

Wir haben also in allen Fällen gezeigt, dass A und B die gleiche Jordan Normalform haben und somit ähnlich sind.

Beispiel 9.4.7. Seien $A, B \in M_{4\times 4}(\mathbb{C})$ mit $p_A = p_B$ und $M_A = M_B$. Sind A und B dann ähnlich?
Lösung: Nein. Zum Beispiel können die beiden Matrizen

$$A = \begin{pmatrix} J_{0,2} & & \\ & J_{0,1} & \\ & & J_{0,1} \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} J_{0,2} & \\ & J_{0,2} \end{pmatrix}$$

nicht durch Vertauschen der Blöcke ineinander überführt werden. Laut der Eindeutigkeit der Jordan Normalform sind sie also nicht ähnlich. Laut Übung 9.3.6 haben wir aber $p_A = p_B = x^4$ und $M_A = M_B = x^2$.

Übung 9.4.8. (In der Serie) Zeigen Sie: Jede Matrix $A \in M_{n\times n}(\mathbb{C})$ ist ähnlich zu A^T.

Jetzt werden wir uns damit befassen, wie man Jordanbasen berechnet, oder äquivalent für Matrizen A, wie man P berechnet, sodass $P^{-1}AP$ in Jordanform ist. Beginnen wir mit einigen der Beispiele, in denen wir die Jordan Normalform bereits berechnet haben:

Übung 9.4.9. Finden Sie eine Jordanbasis für die Matrix $A = \begin{pmatrix} 7 & 1 & & & \\ & 7 & 2 & & \\ & & 7 & 3 & \\ & & & 7 & 4 \\ & & & & 7 \end{pmatrix}$ aus Beispiel 9.4.1. Das heisst, finden Sie eine Basis $\mathcal{B} \subseteq \mathbb{Q}^5$ mit $[m_A]_{\mathcal{B}} = J_{7,5}$.

Lösung: Wir suchen eine Kette von Ordnung 5 für

$$N := A - 7I = \begin{pmatrix} 0 & 1 & & & \\ & 0 & 2 & & \\ & & 0 & 3 & \\ & & & 0 & 4 \\ & & & & 0 \end{pmatrix}.$$

Damit meinen wir Vektoren $v_1, \dots, v_5$ mit $v_{i+1} = Nv_i$ und $Nv_5 = 0$. Anders formuliert suchen wir einen Vektor mit Lebensdauer 4, das heisst ein Element von $\ker N^5 \setminus \ker N^4$. Durch Betrachten von N kann man ein solches Element leicht erraten: Der Einheitsvektor e_5. Alternativ könnte man auch $\ker N^4 = \mathrm{Sp}(e_1, \dots, e_4)$ und $\ker N^5 = \mathbb{Q}^5$ berechnen und folgern, dass jeder Vektor $v \notin \ker N^4$ (also $v = \alpha_1 e_1 + \dots + \alpha_5 e_5 \in \mathbb{Q}^5$ mit $\alpha_5 \neq 0$) ein möglicher Kandidat ist. Sei also $v = e_5$. Die dazugehörige Kette ist

$$v \xmapsto{N} Nv \xmapsto{N} \cdots \xmapsto{N} N^4 v \xmapsto{N} 0.$$

Also wählen wir, wie im Beweis von 9.2.19, die Basis $\mathcal{B}_v = (N^4 v, N^3 v, N^2 v, Nv, v)$ von $\mathbb{Q}^5$. Bezüglich dieser Basis ist

$$[m_N]_{\mathcal{B}_v} = J_{0,5} = \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & 0 & 1 & \\ & & & 0 & 1 \\ & & & & 0 \end{pmatrix}$$

und somit ist $[m_A]_{\mathcal{B}_v} = [m_N]_{\mathcal{B}_v} + [7 \cdot \mathrm{Id}]_{\mathcal{B}_v} = J_{0,5} + 7 \cdot I_5 = J_{7,5}$, wie gewollt. Wir müssen also nur noch $\mathcal{B}_v$ berechnen:

$$\begin{aligned} v &= e_5, \\ Nv &= 4e_4, \\ N^2 v &= 12 e_3, \\ N^3 v &= 24 e_2, \\ N^4 v &= 24 e_1. \end{aligned}$$

Also ist $\mathcal{B}_v = (24e_1, 24e_2, 12e_3, 4e_4, e_5)$ eine explizite Jordanbasis für A.
Frage: Ist diese Basis eindeutig?
Antwort: Nein. In diesem Beispiel haben wir allerdings nicht allzu viel Freiheit in der Auswahl. Alle anderen Jordanbasen haben die Form $\mathcal{B}_w$ für ein $w \notin \ker N^4$.

Übung 9.4.10. Finden Sie eine Jordan Basis für die Matrix

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & -1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

aus Beispiel 9.4.3.

<u>Lösung:</u> Hier ist $A = A - 0I$ selbst nilpotent. Wir suchen eine Basis der Form $\mathcal{B}_{v_1} \sqcup \mathcal{B}_{v_2}$ mit zwei Ketten

$$\begin{array}{rcl} \ker A^3 \setminus \ker A^2 \ni v_1 & & \\ \downarrow & & \\ \ker A^2 \setminus \ker A \ni Av_1 & & v_2 \in \ker A^2 \setminus \ker A \\ \downarrow & & \downarrow \\ \ker A \setminus \{0\} \ni A^2 v_1 & & Av_2 \in \ker A \setminus \{0\} \\ \downarrow & & \downarrow \\ 0 & & 0 \end{array}$$

Um solche Basen zu finden, hilft es normalerweise, Basen der aufsteigenden Kette von Untervektorräumen

$$\{0\} \subsetneq \ker A \subsetneq \ker A^2 \subsetneq \ker A^3 = \mathbb{Q}^5$$

zu berechnen. Dies kann man normalerweise mithilfe von Gauss Elimination und der Auswahl von einigen Vektoren (falls nötig) machen:

$$\ker A = \ker \begin{pmatrix} 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & -1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \overset{L_3+L_1\to L_3}{=} \ker \begin{pmatrix} 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix}$$

$$= \ker \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \overset{\text{Gauss}}{=} \operatorname{Sp}\left(\begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \right),$$

$$\ker A^2 = \ker \underbrace{\begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}}_{\text{Direkte Berechnung}} = \ker \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

$$\overset{\text{Gauss}}{=} \operatorname{Sp}\left(\begin{pmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right) = \operatorname{Sp}\left(\underbrace{\begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}}_{\text{von } \ker A} \begin{pmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \right),$$

$$\ker A^3 = \mathbb{Q}^5 = \operatorname{Sp}\left(\begin{pmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, e_1 \right).$$

Wir sind nun bereit, v_1 und v_2 auszuwählen. Die Wahl von v_1 ist einfach. Wir wählen $v_1 = e_1$ und erhalten

$$B_{e_1} = \left(A^2 e_1, Ae_1, e_1\right) = (e_1 - e_3, e_4, e_1).$$

Diese Basis gehört zum 3×3-Block $J_{0,3}$. Um einen „Erzeuger“ v_2 für den 2×2-Block zu finden, suchen wir einen Vektor $v_2 \in \ker A^2 \setminus \ker A$, sodass v_2 den Untervektorraum $\ker(A) + \operatorname{Sp}(Av_1)$ zu einer Basis von $\ker(A^2)$ ergänzt. Das heisst, wir suchen v_2 mit

$$(\ker(A) + \operatorname{Sp}(Av_1)) \oplus \operatorname{Sp}(v_2) = \ker(A^2).$$

Wenn wir die obige Berechnung betrachten und $e_3 - e_1 = -(e_1 - e_3)$ verwenden, können wir die Basis von $\ker A^2$ so anordnen, dass wir sehen, woher die Elemente kommen. Wie zuvor berechnet, ist

$$\ker A^2 = \operatorname{Sp}\left(\overbrace{\begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \underbrace{\begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}}^{\text{von } \ker A}, \overbrace{\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}}^{=Av_1}}_{\text{vom } 3 \times 3\text{-Block}}, \begin{pmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right)$$

Also ist $v_2 := (-1,0,0,0,1)^T$ eine gute Wahl! Es folgt, dass v_2 den 2×2-Block erzeugen muss, also ist

$$\mathcal{B}_{v_2} = (Av_2, v_2) = \left(\begin{pmatrix} -1 \\ 1 \\ 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right)$$

eine Basis des 2×2-Blocks. Die Hintereinanderreihung von $\mathcal{B}_{v_1}$ und $\mathcal{B}_{v_2}$ gibt uns eine Jordanbasis

$$\mathcal{B} = (A^2v_1, Av_1, v_1, Av_2, v_2) = \left(\begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right).$$

In anderen Worten, wenn wir

$$P := \begin{pmatrix} 1 & 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

definieren, dann ist $P^{-1}AP = \begin{pmatrix} J_{0,3} & \\ & J_{0,2} \end{pmatrix}$ in Jordanform.

Gehen wir nun die Frage (2) von Seite 367 etwas algorithmischer an. Wir möchten also für einen allgemeinen Endomorphismus $T \in \mathrm{End}(V)$ eines endlich-dimensionalen Vektorraums eine Basis $\mathcal{B}$ von V finden, sodass $[T]_{\mathcal{B}}$ die Form (9.16) hat.

Erster Schritt: Das Problem auf die Haupträume reduzieren

Die gesuchte Basis $\mathcal{B}$ ist die Hintereinanderreihung von Basen $\mathcal{B}_i$ für $i = 1, \ldots, k$, wobei $\mathcal{B}_i$ jeweils zu den Blöcken mit Eigenwert λ_i gehört. Um die Basis $\mathcal{B}$ zu finden reicht es also, die Basen $\mathcal{B}_i$ zu finden. Der Raum $\mathrm{Sp}(\mathcal{B}_i)$ heisst *Hauptraum* zum Eigenwert λ_i und wir bezeichnen ihn mit $\mathrm{Hau}(\lambda_i, T)$, oder kurz $\mathrm{Hau}(\lambda_i)$ wenn T aus dem Kontext klar ist. Der Hauptraum hängt nicht von der Wahl der Jordanbasis ab, wie wir durch eine alternative Charakterisierung gleich sehen werden. Man kann ihn wie folgt berechnen:

Um $\mathrm{Hau}(\lambda_i, T)$ zu berechnen, betrachten wir $N_i := T - \lambda_i \,\mathrm{Id}_V$ und die Kette

$$\{0\} \subseteq \ker(N_i) \subseteq \ker(N_i^2) \subseteq \ker(N_i^3) \subseteq \cdots \tag{9.18}$$

und definieren

$$r_i := \min\{r \mid \dim\ker(N_i^r) = m_a(\lambda_i)\}.$$

Dann ist nämlich $\mathrm{Hau}(\lambda_i, T) = \ker(N_i^{r_i})$. Diese Zahl r_i markiert auch die erste Stelle, an der sich die Kette (9.18) stabilisiert[10] und mit der Notation von Seite 367 gilt, dass r_i die Grösse des grössten Blocks zum Eigenwert λ_i ist.

Wir fassen den ersten Schritt zusammen: Der Vektorraum V hat die folgende Zerlegung in T-invariante Untervektorräume (dies wird auch die *Hauptraumzerlegung* genannt):

$$V = \mathrm{Hau}(\lambda_1) \oplus \cdots \oplus \mathrm{Hau}(\lambda_k),$$

wobei $\lambda_1, \ldots, \lambda_k$ die paarweise verschiedenen Eigenwerte von T sind. Es gilt $\dim \mathrm{Hau}(\lambda_i) = m_a(\lambda_i)$ und $\mathrm{Hau}(\lambda_i) = \ker(N_i^{r_i})$, wobei $N_i = T - \lambda_i \mathrm{Id}_V$ und r_i die kleinste Zahl mit $\dim\ker(N_i^{r_i}) = m_a(\lambda_i)$ ist.

Zweiter Schritt: Die Basen der Haupträume berechnen

Sei $i \in \{1, \ldots, k\}$. Wir möchten $\mathrm{Hau}(\lambda_i, T) = \ker(N_i^{r_i})$ berechnen, wobei wie zuvor $N_i = T - \lambda_i \mathrm{Id}_V$ ist. Bemerken Sie:

- $\mathrm{Hau}(\lambda_i, T)$ ist T-invariant und somit auch N_i-invariant.
- $N_i|_{\mathrm{Hau}(\lambda_i,T)}$ ist nilpotent (die r_i-te Potenz ist die Nullabbildung).
- Die Basis $\mathcal{B}_i$ von $\mathrm{Hau}(\lambda_i, T)$ welche wir berechnen möchten hat per Definition die Eigenschaft, dass

$$[T|_{\mathrm{Hau}(\lambda_i,T)}]_{\mathcal{B}_i} = \begin{pmatrix} J_{\lambda_i, n_1^{(i)}} & & \\ & \ddots & \\ & & J_{\lambda_i, n_{\ell_i}^{(i)}} \end{pmatrix},$$

was impliziert (und sogar dazu äquivalent ist), dass

$$[N_i|_{\mathrm{Hau}(\lambda_i,T)}]_{\mathcal{B}_i} = \begin{pmatrix} J_{\lambda_i, n_1^{(i)}} & & \\ & \ddots & \\ & & J_{\lambda_i, n_{\ell_i}^{(i)}} \end{pmatrix} - \lambda_i I = \begin{pmatrix} J_{0, n_1^{(i)}} & & \\ & \ddots & \\ & & J_{0, n_{\ell_i}^{(i)}} \end{pmatrix}.$$

Es läuft also darauf hinaus, dass wir Basen für die Ketten bezüglich $N_i|_{\mathrm{Hau}(\lambda_i,T)}$ berechnen müssen. Dies ist ähnlich zu unserem Beispiel 9.4.3. Grob gesagt macht man das wie folgt:

[10]Vergleichen Sie das mit dem Fitting Lemma 9.2.14.

(1) Man sucht maximale linear unabhängige Mengen von Vektoren in $\ker(N_i^{r_i})\backslash\ker(N_i^{r_i-1})$. Diese erzeugen die $r_i \times r_i$-Blöcke.

(2) Nun macht man das dasselbe für $\ker(N_i^{r_i-k})\backslash\ker(N_i^{r_i-(k+1)})$ für $k = 1, \ldots, r_i$, wobei man sicherstellen muss, dass man „neue“ Vektoren findet, das heisst Vektoren, welche nicht schon in den Ketten aus früheren Schritten liegen.

Natürlich hilft es, wenn man die Blockstruktur bereits kennt, weil man dann weiss, welche Ketten man betrachten muss. Es ist möglich, (2) präziser zu beschreiben (sehen Sie zum Beispiel [10, Seite 291]), aber wir werden das für unsere „niedrig“-dimensionalen Beispiele nicht brauchen.

Beispiel 9.4.11. Als letztes Beispiel betrachten wir die reelle Matrix

$$B = \begin{pmatrix} 6 & -2 & 6 & 1 & 1 \\ 1 & -1 & 2 & 1 & -2 \\ -2 & 0 & -1 & 0 & -1 \\ -1 & 0 & -2 & 2 & -1 \\ -4 & 4 & -6 & -2 & 3 \end{pmatrix}.$$

Wir können uns zwei Fragen stellen.

(1) Was ist die Jordan Normalform von B?

Bemerken Sie, dass wir a priori nicht wissen, dass diese existiert. Falls B nicht trigonalsierbar ist (was der Fall ist, wie wir sehen werden), dann gibt es keine Matrix $P \in \mathrm{GL}_5(\mathbb{R})$, sodass $P^{-1}BP$ in Jordanform ist. Da $\mathbb{R} \subseteq \mathbb{C}$, ist B natürlich trigonalisierbar über $\mathbb{C}$, also existiert $P \in \mathrm{GL}_5(\mathbb{C})$, sodass $P^{-1}BP$ in Jordanform ist.

Da B über $\mathbb{R}$ nicht trigonalisierbar ist, folgt, dass P und $P^{-1}BP$ komplexe Matrizen sind, obwohl B real ist. Wir können uns trotzdem auch fragen:

(2) Falls wir nur Matrizen $P \in \mathrm{GL}_5(\mathbb{R})$ verwenden, was ist die einfachste Form, die wir mit Matrizen der Form $P^{-1}BP$ erreichen können und wie ist diese mit der Jordan Normalform von B über $\mathbb{C}$ verwandt?

Wir werden diese Frage am Ende des Beispiels beantworten. Beginnen wir nun mit der ersten Frage. Zuerst berechnen wir die Eigenwerte. Dafür berechnen wir

$$p_B(x) = -(x-1)((x-2)^2+1)^2,$$

also sind die komplexen Eigenwerte $\lambda_1 = 1$, $\lambda_2 = 2+i$, $\lambda_3 = 2-i$, mit $m_a(1) = 1$ und $m_a(2+i) = 2 = m_a(2-i)$. Also hat die Jordan Normalform J von B (über $\mathbb{C}$) einen Block von der Form $J_{1,1}$. Um zu sehen, wie viele Blöcke mit Eigenwert $2+i$ wir haben, können wir $m_g(2+i) = \dim \mathrm{Eig}_B(2+i)$ berechnen und sehen, dass es gleich 1 ist, aber dies ist keine schöne Berechnung (Matlab hätte aber kein Problem damit).

Wir gehen deshalb einen anderen Weg und berechnen M_B. Erinnern Sie sich, dass M_B das gleiche Polynom ist, egal ob wir es über $\mathbb{C}$ oder $\mathbb{R}$ betrachten, also haben wir entweder

$$M_B = (x-1)\underbrace{((x-2)^2+1)}_{Q(x):=} \quad \text{oder} \quad M_B = -p_B = (x-1)((x-2)^2+1)^2.$$

Um zu entscheiden, welcher der beiden Fälle der Wahrheit entspricht, können wir

$$\dim\ker Q(B) = \dim\ker((B-2I_5)^2+I_5) = 2$$

berechnen, was eine einfachere Rechnung ist, da alles über $\mathbb{R}$ stattfindet. Es folgt, dass

$$M_B = -p_B = (x-1)((x-2)^2+1)^2 = (x-1)(x-(2+i))^2(x-(2+i))^2,$$

also muss die Jordan Normalform von B

$$\begin{pmatrix} 1 & & \\ & J_{2+i,2} & \\ & & J_{2-i,2} \end{pmatrix} = \begin{pmatrix} 1 & & & & \\ & 2+i & 1 & & \\ & & 2+i & & \\ & & & 2-i & 1 \\ & & & & 2-i \end{pmatrix}$$

sein.

Bemerkung 9.4.12. Sei $A \in M_{n\times n}(\mathbb{R})$ und sei $\lambda \in \mathbb{C}$ ein Eigenwert. Die Blockstruktur bezüglich λ ist die Gleiche wie die Blockstrucktur bezüglich $\overline{\lambda}$. Tatsächlich sind die Zahlen $B(\lambda, k, A)$ und $B(\overline{\lambda}, k, A)$ von (9.15) gleich, da $\mathrm{Rang}(C) = \mathrm{Rang}(\overline{C})$, wobei für $C \in M_{n\times n}(\mathbb{C})$ mit $\overline{C}$ die komplex konjugierte Matrix von C gemeint ist.

Bemerkung 9.4.13. Wir möchten noch eine Bemerkung zur Frage (2) oben machen. Wenn wir noch ein bisschen mehr mit Polynomen in $\mathbb{R}[x]$ spielen, können wir zur folgenden Variante der Jordan Normalform über $\mathbb{R}$ für B gelangen. Das heisst, es existiert $P \in \mathrm{GL}_5(\mathbb{R})$ mit

$$P^{-1}BP = \left(\begin{array}{c|cc|cc} 1 & & & & \\ \hline & 2 & 1 & 1 & \\ & -1 & 2 & & 1 \\ \hline & & & 2 & 1 \\ & & & -1 & 2 \end{array}\right).$$

Erinnern Sie sich, dass
$$\left\{ \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \;\middle|\; a, b \in \mathbb{R} \right\} \cong \mathbb{C},$$
wobei die Matrix $\begin{pmatrix} a & -b \\ b & a \end{pmatrix} \in M_{2\times 2}(\mathbb{R})$ der komplexen Zahl $a + bi \in \mathbb{C}$ entspricht. Können Sie die Ähnlichkeit zwischen den beiden Resultaten sehen?

Bemerkung 9.4.14. Der Autor glaubt, dass es am besten ist, wenn man die Jordan Normalform über allgemeinen Körpern (mit möglicherweise nicht trigonalisierbaren Matrizen/ Endomorphismen) erst behandelt, nachdem die Leser schon etwas Ringtheorie gelernt haben. Insbesondere ist die allgemeine Jordan Normalform ein ziemlich einfaches Korollar der Struktur von Moduln über Hauptidealringen (die Leser, die es nicht abwarten können, können in meinem Lieblingsbuch über Algebra I und II [14] mehr darüber lesen).

9.5 Anwendung auf Systeme von Differentialgleichungen

Wir möchten zum Ende dieses Kapitels noch kurz eine coole Anwendung der Jordan Normalform skizzieren. Erinnern Sie sich, dass die Gleichung

$$y'(t) = ay(t), \quad a \in \mathbb{C}, \tag{9.19}$$

die Lösung $y(t) = e^{ta}y(0)$ hat. Was können wir sagen, wenn wir eine ähnliche Gleichung betrachten, aber mit mehreren Koordinaten? Ein Beispiel wäre das Gleichungssystem

$$\begin{aligned} x'(t) &= ax(t) + by(t), \\ y'(t) &= cx(t) + dy(t). \end{aligned}$$

Wir können dies mithilfe von Matrizen zu einer einzigen vektorwertigen Gleichung

$$\begin{pmatrix} x'(t) \\ y'(t) \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} \tag{9.20}$$

zusammenfassen und wenn wir $Y(t) := \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}$, $Y'(t) := \begin{pmatrix} x'(t) \\ y'(t) \end{pmatrix}$ und $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ schreiben, dann wird (9.20) zu

$$Y'(t) = AY(t), \tag{9.21}$$

was uns von der Form her stark an unsere ursprüngliche Gleichung (9.19) erinnert. Es stellt sich heraus, dass (9.21) auch eine einfach aussehende Lösung hat, welche sehr ähnlich zur Lösung von (9.19) ist, nämlich

$$Y(t) = e^{tA}Y(0).$$

Damit das Sinn ergibt, müssen wir allerdings noch das Exponential einer Matrix definieren.

9.5.1 Das Exponential einer Matrix

Die Definition des Matrixexponentials erfordert einen Beweis der Wohldefiniertheit. Deswegen geben wir sie hier als Proposition/Definition an, wobei wir auch noch gleich einige wichtige Eigenschaften beweisen.

Proposition/Definition 9.5.1. *Sei $A \in M_{n\times n}(\mathbb{C})$. Die Reihe*[11]

$$I_n + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots = \sum_{k=0}^{\infty} \frac{1}{k!}A^k$$

konvergiert (komponententweise) zu einer Matrix in $M_{n\times n}(\mathbb{C})$. Der Grenzwert heisst Matrixexponential *von A und wird mit* $\exp A$ *oder* e^A *bezeichnet. Ausserdem gilt:*

(1) Falls $AB = BA$, dann ist $\mathrm{e}^{A+B} = \mathrm{e}^A\mathrm{e}^B$.

(2) Für alle $P \in \mathrm{GL}_n(\mathbb{C})$ ist $\mathrm{e}^{P^{-1}AP} = P^{-1}\mathrm{e}^A P$.

(3) Falls $D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$ diagonal ist, dann ist

$$\mathrm{e}^D = \begin{pmatrix} \mathrm{e}^{\lambda_1} & & \\ & \ddots & \\ & & \mathrm{e}^{\lambda_n} \end{pmatrix} \quad \textit{und} \quad \mathrm{e}^{tD} = \begin{pmatrix} \mathrm{e}^{t\lambda_1} & & \\ & \ddots & \\ & & \mathrm{e}^{t\lambda_n} \end{pmatrix} \textit{ für } t \in \mathbb{C}.$$

(4) Für einen Jordanblock $J_{0,n}$ gilt $\mathrm{e}^{tJ_{0,n}} = \begin{pmatrix} 1 & t & \frac{t^2}{2!} & \frac{t^3}{3!} & \cdots & \frac{t^{n-1}}{(n-1)!} \\ & 1 & t & \frac{t^2}{2!} & \cdots & \frac{t^{n-2}}{(n-2)!} \\ & & 1 & t & \cdots & \frac{t^{n-3}}{(n-3)!} \\ & & & \ddots & & \vdots \\ & & & & \ddots & t \\ & & & & & 1 \end{pmatrix}.$

[11] Erinnern Sie sich, dass $0! = 1$ und $A^0 = I$.

(5) Für einen Jordanblock $J_{\lambda,n}$ *gilt* $e^{tJ_{\lambda,n}} = e^{t\lambda} \cdot \begin{pmatrix} 1 & t & \frac{t^2}{2!} & \frac{t^3}{3!} & \cdots & \frac{t^{n-1}}{(n-1)!} \\ & 1 & t & \frac{t^2}{2!} & \cdots & \frac{t^{n-2}}{(n-2)!} \\ & & 1 & t & \cdots & \frac{t^{n-3}}{(n-3)!} \\ & & & \ddots & & \vdots \\ & & & & \ddots & t \\ & & & & & 1 \end{pmatrix}.$

(6) Falls $A = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_\ell \end{pmatrix}$ *eine Blockdiagonalmatrix ist, dann ist*

$$e^A = \begin{pmatrix} e^{J_1} & & \\ & \ddots & \\ & & e^{J_\ell} \end{pmatrix}.$$

Beweis-Skizze. Falls $c \in \mathbb{R}_{>0}$ eine obere Schranke für den Absolutbetrag aller Einträge von A ist, dann ist nc^2 eine obere Schranke für den Absolutbetrag aller Einträge von A^2. Mit Induktion lässt sich zeigen, dass $n^{k-1}c^k$ eine obere Schranke für den Absolutbetrag aller Einträge von A^k ist. Mit dem Majorantenkriterium folgt (absolute) Konvergenz in jedem Eintrag, da die Reihe

$$\sum_{k=0}^{\infty} \frac{n^{k-1}c^k}{k!} < \infty$$

konvergiert. Daher existiert $e^A \in M_{n\times n}(\mathbb{C})$ für jede $n \times n$-Matrix A.

(1) folgt aus einer Berechnung mit der binomischen Formel, die dank $AB = BA$ gilt.

(2): Da $(P^{-1}AP)^k = P^{-1}A^kP$ für alle $k \in \mathbb{N}$ gilt, gilt für jedes N, dass

$$\sum_{k=0}^{N} \frac{1}{k!}(P^{-1}AP)^k = P^{-1}\left(\sum_{k=0}^{N} \frac{1}{k!}A^k\right)P$$

und mit $N \to \infty$ folgt die Behauptung.

(3) folgt direkt aus der Definition von e^D (bzw. e^{tD}).

(4) Es gilt

$$tJ_{0,n} = \begin{pmatrix} 0 & t & & \\ & \ddots & \ddots & \\ & & \ddots & t \\ & & & 0 \end{pmatrix},$$

$$\frac{1}{2}(tJ_{0,2})^2 = \frac{t^2}{2}J_{0,n}^2 = \frac{t^2}{2}\begin{pmatrix} 0 & 0 & 1 & & & \\ & 0 & 0 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \ddots & \ddots & 1 \\ & & & & \ddots & 0 \\ & & & & & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & \frac{t^2}{2} & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \frac{t^2}{2} \\ & & & \ddots & 0 \\ & & & & 0 \end{pmatrix},$$

und so weiter. Weil $J_{0,n}^k = 0$ für $k \geq n$, ist $e^{tJ_{0,n}}$ eine endliche Summe, die durch

$$\begin{pmatrix} 1 & t & \frac{t^2}{2!} & \frac{t^3}{3!} & \cdots & \frac{t^{n-1}}{(n-1)!} \\ & 1 & t & \frac{t^2}{2!} & \cdots & \frac{t^{n-2}}{(n-2)!} \\ & & 1 & t & \cdots & \frac{t^{n-3}}{(n-3)!} \\ & & & \ddots & & \vdots \\ & & & & \ddots & t \\ & & & & & 1 \end{pmatrix}$$

gegeben ist.

(5) folgt aus (3) und (4) unter Verwendung von (1), da $tJ_{\lambda,n} = t\lambda I + tJ_{0,n}$ und $(t\lambda I)(tJ_{0,n}) = (tJ_{0,n})(t\lambda I)$ gelten.

(6) folgt, weil $\begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_\ell \end{pmatrix}^k = \begin{pmatrix} J_1^k & & \\ & \ddots & \\ & & J_\ell^k \end{pmatrix}$ für alle $k \in \mathbb{N}$. □

Beispiel 9.5.2. Es gilt $e^0 = I$, wobei 0 die Nullmatrix und I die Identitätsmatrix ist.

Beispiel 9.5.3. Bemerken Sie, dass $\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}^2 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$, also gilt

$$e^{\left(\begin{smallmatrix} 1 & 0 \\ 1 & 0 \end{smallmatrix}\right)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} + \frac{1}{3!}\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} + \cdots = \begin{pmatrix} e & 0 \\ e-1 & 1 \end{pmatrix}.$$

Beispiel 9.5.4. Es gilt $\exp\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix} = \begin{pmatrix}e^1 & \\ & e^0\end{pmatrix} = \begin{pmatrix}e & \\ & 1\end{pmatrix}$, da $\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}$ diagonal ist. Weiter gilt

$$\exp\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix} = I + \begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix} + \frac{1}{2}\begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix} + 0 = \begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}.$$

Es folgt $e^{\left(\begin{smallmatrix}1 & 0\\ 1 & 0\end{smallmatrix}\right)} \neq e^{\left(\begin{smallmatrix}1 & 0\\ 0 & 0\end{smallmatrix}\right)} e^{\left(\begin{smallmatrix}0 & 0\\ 1 & 0\end{smallmatrix}\right)}$. Dies ist kein Widerspruch, da

$$\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix} \neq \begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}.$$

Übung 9.5.5. Zeigen Sie:

(a) Die Matrix e^A ist invertierbar für alle A (Hinweis: Finden Sie die Inverse).

(b) $e^{(A^T)} = \left(e^A\right)^T$.

(c) $e^{\overline{A}^T} = \overline{\left(e^A\right)}^T$.

(d) $\det(e^A) = e^{\operatorname{spur}(A)}$.

(e) Geben Sie einen anderen Beweis von (a) mittels (d).

Hinweis: (d) ist nicht sehr einfach zu beweisen. Zeigen Sie es zuerst für eine diagonale Matrix, dann für eine diagonalisierbare Matrix, dann für eine Matrix in Jordanform und schliesslich für alle $A \in M_{n\times n}(\mathbb{C})$. Für den Schritt mit Matrizen in Jordanform könnte es nützlich sein, die Aussage zuerst für nilpotente Matrizen zu beweisen.

9.5.2 Zurück zu Differentialgleichungen

Proposition 9.5.6. *Sei $A \in M_{n\times n}(\mathbb{C})$. Dann gilt*

$$\left(e^{tA}\right)' = Ae^{tA},$$

wobei $'$ die komponentenweise Ableitung der Funktion $\mathbb{R} \to M_{n\times n}(\mathbb{C}),\ t \mapsto e^{tA}$ bezeichnet.

Beweis. Obwohl $'$ die komponentenweise Ableitung bezeichnet, müssen wir zum Glück nicht alle Komponenten einzeln anschauen. Für eine Funktion $f : \mathbb{R} \to \mathbb{R}^m$ gilt nämlich

$$f'(t) = \lim_{h\to 0} \frac{f(t+h) - f(t)}{h},$$

ganz analog zur üblichen Ableitung. Die Schreibweise $\lim_{h\to 0}$ bedeutet hier, dass wir den Grenzwert in jeder einzelnen Komponente betrachten. Also haben wir

$$\begin{aligned}\left(e^{tA}\right)' &= \lim_{h\to 0}\frac{e^{(t+h)A}-e^{tA}}{h} = \lim_{h\to 0}\frac{e^{hA}e^{tA}-e^{tA}}{h} = \lim_{h\to 0}\frac{(e^{hA}-I)e^{tA}}{h}\\ &= \lim_{h\to 0}\frac{\left(\cancel{I}+hA+\frac{(hA)^2}{2!}+\cdots-\cancel{I}\right)e^{tA}}{h} = \left(\lim_{h\to 0}\frac{hA+\frac{(hA)^2}{2!}+\cdots}{h}\right)e^{tA}\\ &= \Bigg(A+\underbrace{\lim_{h\to 0}\left(\frac{hA^2}{2!}+\frac{h^2A^3}{3!}+\cdots\right)}_{=0}\Bigg)e^{tA} = Ae^{tA}.\end{aligned}$$ □

Korollar 9.5.7. *Seien $A \in M_{n\times n}(\mathbb{C})$, $Y(t) = \begin{pmatrix} y_1(t)\\ \vdots \\ y_n(t)\end{pmatrix}$ und $Y'(t) = \begin{pmatrix} y_1'(t)\\ \vdots \\ y_n'(t)\end{pmatrix}$. Die Lösung der Differentialgleichung*

$$Y'(t) = AY(t)$$

ist gegeben durch $Y(t) = e^{tA}\cdot\begin{pmatrix} c_1\\ \vdots \\ c_n\end{pmatrix}$, wobei $\begin{pmatrix} c_1\\ \vdots \\ c_n\end{pmatrix} \in \mathbb{C}^n$ irgendwelche Konstanten sind mit $Y(0) = \begin{pmatrix} c_1\\ \vdots \\ c_n\end{pmatrix}$.

Beweis. Es folgt aus der Definition der Ableitung und Assoziativität der Matrixmultiplikation, dass für $Y(t) := e^{tA}\begin{pmatrix} c_1\\ \vdots \\ c_n\end{pmatrix}$ gilt

$$Y'(t) = \left(e^{tA}\begin{pmatrix} c_1\\ \vdots \\ c_n\end{pmatrix}\right)' = \left(e^{tA}\right)'\begin{pmatrix} c_1\\ \vdots \\ c_n\end{pmatrix} \overset{\text{Prop.}}{\underset{9.5.6}{=}} \left(Ae^{tA}\right)\begin{pmatrix} c_1\\ \vdots \\ c_n\end{pmatrix} = A\left(e^{tA}\begin{pmatrix} c_1\\ \vdots \\ c_n\end{pmatrix}\right) = AY(t),$$

also löst $Y(t) = e^{tA}\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$ für beliebige $\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \in \mathbb{C}^n$ die Differentialgleichung. Man setzt $t = 0$ und erhält

$$Y(0) = e^0 \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \underset{9.5.2}{\overset{\text{Bsp.}}{=}} I \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}.$$

Wir haben also gezeigt, dass $e^{tA}\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$ eine Lösung unserer Differentialgleichung ist. Es folgt nun aus allgemeinen Prinzipien, welche Sie in der Analysis behandeln, dass diese Lösung eindeutig ist. □

Für die Berechnung von $\exp A$ braucht man die Jordan Normalform!

Differentialgleichungen der Form $Y'(t) = AY(t)$ heissen *homogene Differentialgleichungen erster Ordnung mit konstanten Koeffizienten* und haben viele Anwendungen. Sehen Sie zum Beispiel Abschnitt 11.1 hier.

Hier ist die allgemeine Lösung einer solchen Differentialgleichung mithilfe der Jordan Normalform. Sei

$$Y'(t) = AY(t) \tag{9.22}$$

und sei $P \in \mathrm{GL}_n(\mathbb{C})$, sodass $PJP^{-1} = A$, wobei J in Jordanform ist. Dann ist (9.22) gleich

$$Y'(t) = PJP^{-1}Y(t),$$

was äquivalent ist zu

$$P^{-1}Y'(t) = JP^{-1}Y(t)$$

oder

$$Z'(t) = JZ(t) \tag{9.23}$$

mit $Z(t) = P^{-1}Y(t)$. Letzteres folgt, weil $(P^{-1}Y)'(t) = P^{-1}Y'(t)$. Dies zeigen Sie in Übung 9.5.8 unten. Die Lösungen von (9.23) sind gegeben durch

$$Z(t) = e^{tJ} = \begin{pmatrix} e^{J_1} & & \\ & \ddots & \\ & & e^{J_\ell} \end{pmatrix},$$

wobei $J_1, \ldots, J_k$ die Jordanblöcke sind. Mit Proposition 9.5.1 (4) können wir e^{tJ_i} berechnen. Die Lösungen von (9.22) sind also $Pe^{tJ} \cdot v$ mit $v \in \mathbb{C}^n$.

Übung 9.5.8. Seien $A, B : \mathbb{R} \to M_{n\times n}(\mathbb{C})$ differenzierbare Funktionen. Zeigen Sie, dass

$$(A(t)B(t))' = A'(t)B(t) + A(t)B'(t).$$

Falls $A(t) \equiv A$ konstant ist, dann gilt folglich $(AB(t))' = AB'(t)$.

Beispiel 9.5.9. Betrachten Sie die zweidimensionale Differentialgleichung

$$Y'(t) = AY(t), \quad \text{wobei } A = \begin{pmatrix} 0 & 1 \\ -4 & 4 \end{pmatrix}.$$

Wenn wir die Jordan Normalform von A und eine Jordanbasis berechnen, erhalten wir $A = PJP^{-1}$ mit

$$P = \begin{pmatrix} -2 & 1 \\ -4 & 0 \end{pmatrix} \quad \text{und} \quad J = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}.$$

Die Gleichung $Z'(t) = JZ(t)$ hat dann die Lösung

$$Z(t) = e^{2t}\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} e^{2t} & te^{2t} \\ 0 & e^{2t} \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

und somit ist $Y(t) = PZ(t)$ die Lösung unserer ursprünglichen Gleichung. Konkret heisst das

$$\begin{aligned} Y(t) &= \begin{pmatrix} -2 & 1 \\ -4 & 0 \end{pmatrix}\begin{pmatrix} e^{2t} & te^{2t} \\ 0 & e^{2t} \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \\ &= \begin{pmatrix} -2e^{2t} & -2te^{2t} + e2t \\ -4e^{2t} & -4te^{2t} \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \\ &= \begin{pmatrix} (-2c_1 + c_2)e^{2t} - 2c_2te^{2t} \\ -4c_1e^{2t} - 4c_2te^{2t} \end{pmatrix}. \end{aligned}$$

Beispiel 9.5.10. Wir betrachten das System

$$Y'(t) = AY(t), \quad \text{wobei } A = \begin{pmatrix} 2 & 1 & 1 \\ 2 & 1 & -2 \\ -1 & 0 & -2 \end{pmatrix}.$$

Wenn wir die Jordan Normalform und eine Jordanbasis für A berechnen, erhalten wir $A = PJP^{-1}$ mit

$$P = \begin{pmatrix} 1 & 0 & -5 \\ -2 & 0 & -6 \\ -1 & 1 & 1 \end{pmatrix} \quad \text{und} \quad J = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

Die Gleichung $Z'(t) = JZ(t)$ hat dann die Lösung

$$Z(t) = \begin{pmatrix} e^{-t} & te^{-t} & 0 \\ 0 & e^{-t} & 0 \\ 0 & 0 & e^{3t} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$$

und somit ist $Y(t) = PZ(t)$ unsere gesuchte Lösung, das heisst

$$\begin{aligned} Y(t) &= \begin{pmatrix} 1 & 0 & -5 \\ -2 & 0 & -6 \\ -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} e^{-t} & te^{-t} & 0 \\ 0 & e^{-t} & 0 \\ 0 & 0 & e^{3t} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} \\ &= \begin{pmatrix} e^{-t} & te^{-t} & -5e^{3t} \\ -2e^{-t} & -2te^{-t} & -6e^{3t} \\ -e^{-t} & -te^{-t} + e^{-t} & e^{3t} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} \\ &= \begin{pmatrix} c_1 e^{-t} + c_2 te^{-t} - 5c_3 e^{3t} \\ -2c_1 e^{-t} - 2c_2 te^{-t} - 6c_3 e^{3t} \\ (-c_1 + c_2)e^{-t} - c_2 te^{-t} + c_3 e^{3t} \end{pmatrix}. \end{aligned}$$

Changelog: Kapitel 9

- 06.05: Korollar 9.1.12 und die nachfolgende Übung wurden von Abschnitt 9.2 ans Ende von Abschnitt 9.1 verschoben.
- 08.05: In Beispiel 9.1.3 (9) wurde $\ell \leq 2$ zu $\ell \geq 2$ korrigiert.
- 08.05: Im Beweis von Satz 9.2.19 wurde $N^{\ell_{\overline{N}}(u_i)} u_i$ zu $\overline{N}^{\ell_{\overline{N}}(u_i)} u_i$ korrigiert.
- 08.05: In Übung 9.2.23 wurde $[T]_{\mathcal{B}} = J_{0,k}$ zu $[N]_{\mathcal{B}} = J_{0,n}$ korrigiert.
- 08.05: Am Ende des Beweises von Lemma 9.2.21 wurde an einer Stelle α_j zu α_{i_0+k} korrigiert.
- 08.05: Vor Übung 9.2.22 wurde $[T|_{Z_v}]_{\mathcal{B}_v}$ zu $[N|_{Z_v}]_{\mathcal{B}_v}$ korrigiert.
- 10.05: In Definition 9.2.3 wurde spezifiziert, dass ein unzerlegbarer Untervektorraum T-invariant sein muss.
- 10.05: In Abschnitt 9.2.1 wurde spezifiziert, dass die Lebensdauer nur für Vektoren ungleich Null definiert ist.
- 16.05: In (9.17) wurde $n_k^{(i)}$ zu $B(\lambda_i, k, T)$ korrigiert.
- 16.05: In (9.7) und in der darauffolgenden Gleichung wurden Minuszeichen eingefügt.
- 17.05: In der Lösung von Übung 9.4.9 wurde $v = (v_1, \ldots, v_5)$ zu $v = \alpha_1 e_1 + \cdots + \alpha_5 e_5$ geändert, da im gleichen Absatz die Variablen $v_1, \ldots, v_5$ bereits als Vektoren verwendet werden.
- 17.05: Im Beweis von Satz 9.2.19, im Beweis der Behauptung, wurde an einer Stelle $\beta N v_2$ zu $\beta N v_1$ korrigiert.
- 17.05: Im Beweis von Satz 9.2.19 wurden einige n's zu k's korrigiert.
- 19.05: In Beispiel 9.5.9 wurde an einer Stelle $e^2 t$ zu e^{2t} korrigiert und Lösung wurde von einer 2×2-Matrix zu einem Spaltenvektor korrigiert.
- 27.06: Im „Beweis mit einem Trick“ von Lemma 9.2.11 wurde an einer Stelle ein n zu einem k korrigiert.
- 21.07: Im Beweis von Satz 9.2.19 wurde $\mathcal{B}_{u_n}$ zu $\mathcal{B}_{u_k}$ korrigiert.
- 26.07: In Lemma 9.2.17 wurde hinzugefügt, dass W nicht der Nullvektorraum ist.
- 20.08: Definition 9.1.2 wurde hinzugefügt.

Kapitel 10

Dualraum II

Wir haben den Dualraum bereits kennengelernt, als wir den Darstellungssatz von Riesz (Satz 6.5.5) in Abschnitt 6.5 und den Ausartungsraum von Bilinearformen in Abschnitt 8.5 diskutiert haben. Wir wollen ihn nun etwas genauer untersuchen. Bevor wir anfangen, wiederholen wir die Definition des Kronecker Delta:

Definition 10.0.1. Sei I eine beliebige Menge (dies wird später eine Index-Menge sein) und K ein Körper. Für $i, j \in I$ definieren wir

$$\delta_{ij} := \begin{cases} 1 = 1_K & \text{falls } i = j, \\ 0 = 0_K & \text{falls } i \neq j. \end{cases}$$

10.1 Dimension des Dualraums

Sei V ein Vektorraum über einem Körper K. Wir definieren den *Dualraum* von V als

$$V^* := \mathrm{Hom}_K(V, K),$$

wobei $K = K^1$ als (eindimensionaler) Vektorraum über sich selbst betrachtet wird. Erinnern Sie sich, dass die Vektorraumstruktur auf V^* folgendermassen definiert ist: Für $\varphi, \psi \in V^*$, $\alpha \in K$ und $v \in V$ ist

$$(\varphi + \psi)(v) := \varphi(v) + \psi(v),$$
$$(\alpha\varphi)(v) := \alpha\varphi(v).$$

Sei $\mathcal{B} = (v_i)_{i \in I}$ (hier bezeichnet I eine Indexmenge) eine Basis von V (also ist insbesondere $\dim V$ gleich der Kardinalität von I). Laut Satz[1] 3.1.15 ist ein Element von V^*, also eine Linearform $\varphi : V \to K$, eindeutig bestimmt durch eine beliebige Wahl

[1] Leider haben wir diesen Satz nur für endliche Basen formuliert und bewiesen, aber der gleiche Beweis funktioniert auch im undendlich-dimensionalen Fall.

von Bildern der Basis-Elemente. Das heisst, für jede Wahl von $(\alpha_i)_{i\in I} \subseteq K$ existiert eine eindeutige Linearform $\varphi \in V^*$ sodass $\varphi(v_i) = \alpha_i$ für alle $i \in I$. Eine Wahl $(\alpha_i)_{i\in I} \subseteq K$ kann man auch als eine Funktion $I \to K$ ansehen, welche jedes $i \in I$ auf $\alpha_i \in K$ abbildet. In anderen Worten bedeutet dies also, dass die Abbildung

$$\begin{aligned}\Phi_{\mathcal{B}} : V^* &\to K^I \\ \varphi &\mapsto (\varphi(v_i))_{i\in I}\end{aligned}$$

bijektiv ist, wobei K^I die Menge aller Funktionen von $I \to K$ ist. Surjektivität ist der „existiert"-Teil und Injektivität ist der „eindeutige"-Teil.

Wenn wir K^I mit der üblichen Vektorraumstruktur ausstatten (siehe (2.7) auf Seite 52), sehen wir, dass Φ sogar linear ist. Wir fassen unsere Erkenntnisse in der folgenden Proposition zusammen:

Proposition 10.1.1. *Sei V ein Vektorraum über K und sei $\mathcal{B} = (v_i)_{i\in I}$ eine Basis von V. Dann ist V^* isomorph zu K^I.*

Beispiel 10.1.2. Falls $\dim V = n < \infty$, dann ist Φ nichts anderes als die Darstellungsmatrix: Sei $\mathcal{B} = (v_1, \dots, v_n)$ eine Basis von V (die hier verwendete Indexmenge ist $I = \{1, \dots, n\}$). Dann gilt für $\varphi \in V^*$, dass $\Phi(\varphi) = (\varphi(v_1), \dots, \varphi(v_n)) = [\varphi]^{\mathcal{B}}_{\mathcal{E}_1}$. Bemerken Sie, dass $[\varphi]^{\mathcal{B}}_{\mathcal{E}_1}$ ein Zeilenvektor ist (vergleiche später mit Beispiel 10.1.7 unten). Ausserdem folgt: Falls $\dim V = n \in \mathbb{N}$, dann ist $\dim V^* = \dim V = n$.

Für endlich-dimensionale Vektorräume hat der Dualraum also die gleiche Dimension wie der ursprüngliche Vektorraum. Für unendlich-dimensionale Vektorräume ist dies nicht der Fall, wie das nächste Beispiel zeigt.

Beispiel 10.1.3. Sei $V = \mathrm{Sp}(e_1, e_2, \dots) \subseteq K^\infty$ (dies ist der Raum W aus Beispiel 2.1.14). Dann ist $V^* \cong K^\infty$. Insbesondere[2] ist $\dim V^* > \dim V$.

Beispiel 10.1.4. Sei I eine nichtleere Menge und erinnern Sie sich, dass wir an die Definition von K^I von Seite 52:

$$K^I = \{f : I \to K\}.$$

Wir definieren den *Träger* (auf Englisch support) einer Funktion $f : I \to K$ als

$$\mathrm{supp}(f) := \{i \in I \mid f(i) \neq 0\}.$$

[2]Wir haben nie bewiesen, dass $\dim K^\infty$ überabzählbar ist. Hier ist ein Beweis:
Falls K überabzählbar ist, zeigt Beispiel 5.2.18, dass K^∞ eine überabzählbare linear unabhängige Menge enthält, also muss jede Basis überabzählbar sein. Wenn K endlich/abzählbar ist und $\mathcal{B} = (v_1, v_2, \dots) \subseteq K^\infty$ abzählbar ist, dann kann man mit einem Standardargument aus der Mengenlehre zeigen, dass $\mathrm{Sp}(\mathcal{B})$ auch abzählbar ist (z.B. durch $\mathrm{Sp}(\mathcal{B}) = \bigcup_{n\geq 1} \mathrm{Sp}(v_1, \dots, v_n)$ und weil $\mathrm{Sp}(v_1, \dots, v_n)$ abzählbar ist). Wir wissen aber, dass K^∞ überabzählbar ist, also ist $\mathrm{Sp}(\mathcal{B}) \neq K^\infty$ und somit kann $\mathcal{B}$ keine Basis sein.

Damit definieren wir den Untervektorraum

$$K^{(I)} := \{f \in K^I \mid |\operatorname{supp}(f)| < \infty\} \subseteq K^I$$

aller Funktionen mit endlichem Träger. Bemerken Sie, dass wir äquivalent auch

$$K^I = \{(\alpha_i)_{i\in I} \mid \forall i\ \alpha_i \in K\}$$

schreiben können. In diesem Fall ist

$$K^{(I)} = \{(\alpha_i)_{i\in I} \in K^I \mid \alpha_i = 0 \text{ für alle ausser endlich viele } i \in I\}.$$

Falls $|I| < \infty$, dann ist $K^{(I)} = K^I$, aber falls $|I| = \infty$, dann gilt immer $K^{(I)} \subsetneq K^I$.

Übung 10.1.5. Für $i \in I$ sei $e_i : I \to K$ durch

$$e_i(j) = \delta_{ij} \text{ für alle } j \in I$$

definiert. Somit ist also $e_i \in K^I$. Man kann überprüfen, dass e_i einen endlichen Träger hat, also dass $e_i \in K^{(I)}$.

(a) Zeigen Sie, dass $\operatorname{Sp}((e_i)_{i\in I}) = K^{(I)}$ (bemerken Sie, dass hier $(e_i)_{i\in I}$ nicht ein Vektor, sondern eine ganze Menge von Vektoren ist).

(b) Folgern Sie wie oben, dass $(K^{(I)})^* \cong K^I$.

Übung 10.1.6. [Direkte Summen und Produkte] Sei K ein Körper, I eine (Index-) Menge und $(V_i)_{i\in I}$ eine Kollektion von Vektorräumen über K. Das *direkte Produkt* der Vektorräume $(V_i)_{i\in I}$ ist das kartesische Produkt

$$\prod_{i\in I} V_i = \{(v_i)_{i\in I} \mid v_i \in V_i\ \forall i \in I\}.$$

Dies ist ein Vektorraum, versehen mit komponentenweiser Addition und Skalarmultiplikation. Die *(äussere) direkte Summe* ist der Untervektorraum

$$\bigoplus_{i\in I} V_i = \left\{(v_i)_{i\in I} \in \prod_{i\in I} V_i \;\middle|\; v_i = 0 \text{ für alle ausser endlich viele } i \in I\right\}.$$

Zeigen Sie mit $V_i = K$ für alle $i \in I$, dass

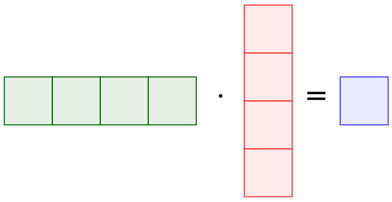
$$\left(\bigoplus_{i\in I} K\right)^* \cong \prod_{i\in I} K.$$

Beispiel 10.1.7. Warum war es in Beispiel 10.1.2 wichtig, zu bemerken, dass $[\varphi]^{\mathcal{B}}_{\mathcal{E}_1}$ ein Zeilenvektor ist? Betrachten wir nun K^n immer als K^n_{Spal} und versuchen wir, $(K^n)^*$ bessser zu verstehen. Betrachten wir dafür die Standard-Basis $\mathcal{E}_n \subseteq K^n$. Für $\varphi \in (K^n)^*$ betrachten wir den Zeilenvektor $[\varphi]^{\mathcal{E}_n}_{\mathcal{E}_1}$, wobei $\mathcal{E}_1 = \{(1)\} \subseteq K$ die Standard-Basis von $K = K^1$ ist. Laut der Transformationsformel (3.11) gilt für $v \in K^n$, dass

$$\varphi(v) = [\varphi(v)]_{\mathcal{E}_1} = [\varphi]^{\mathcal{E}_n}_{\mathcal{E}_1}[v]_{\mathcal{E}_n} = [\varphi]^{\mathcal{E}_n}_{\mathcal{E}_1} v,$$

da $[v]_{\mathcal{E}_n} = v$. In anderen Worten, jede Linearform $\varphi \in (K^n)^*$ ist bloss ein Zeilenvektor, welcher auf $K^n = K^n_{\text{Spal}}$ durch die übliche Multiplikation von Zeilen- und Spaltenvektoren operiert:

$$vw = (a_1, \dots, a_n) \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = a_1 b_1 + \cdots + a_n b_n. \tag{10.1}$$



In anderen Worten haben wir $(K^n_{\text{Spal}})^* = K^n_{\text{Zeil}}$.

Bemerkung 10.1.8. Die Symmetrie zwischen den a_i's und b_i's in (10.1) zeigt die Dualität zwischen V (repräsentiert durch die b_i's) und V^* (repräsentiert durch die a_i's). Wir werden später sehen, dass der Bidualraum von K^n, also $((K^n)^*)^*$, isomorph zu K^n ist, was auch durch die Symmetrie von (10.1) angedeutet wird.

10.2 Die duale Basis

Die duale Basis ist nur eine Basis, wenn wir endlich-dimensionale Vektorräume betrachten. Für unendlich-dimensionale Vektorräume ist es nur eine linear unabhängige Teilmenge des Dualraums. Trotzdem wird die folgende Konstruktion fälschlicherweise auch für ∞-dimensionale Vektorräume die duale Basis genannt:

Definition 10.2.1. Sei V ein Vektorraum über K mit einer Basis $\mathcal{B} = (v_i)_{i \in I}$. Für $i \in I$ sei $\varphi_i \in V^*$ die eindeutige Linearform $\varphi_i : V \to K$ mit $\varphi_i(v_j) = \delta_{ij}$ für alle $j \in I$. Die Menge $\mathcal{B}^* = (\varphi_i)_{i \in I} \subseteq V^*$ wird die *duale Basis* von V^* bezüglich $\mathcal{B}$ genannt, wenn V endlich dimensional ist und die *duale Menge* in V^* bezüglich $\mathcal{B}$, wenn V unendlich-dimensional ist. Wenn $\mathcal{B}$ aus dem Kontext heraus klar ist, schreiben wir oft nur duale Basis von V^* (bzw. duale Menge in V^*).

Bemerkung 10.2.2. Bemerken Sie, dass die Konstruktion von $\varphi_i \in V^*$ für $i \in I$ nicht nur vom Basiselement v_i abhängt, sondern von der ganzen Basis $\mathcal{B} = (v_j)_{j\in I}$!

Notation. Physiker schreiben die duale Basis von $\mathcal{B} = (v_1, \ldots, v_n)$ als $\mathcal{B}^* = (v^1, \ldots, v^n)$. Andere Autoren verwenden manchmal die Notation $\mathcal{B}^* = (v_1^*, \ldots, v_n^*)$.

Bemerkung 10.2.3. Wenn die Dimension von V endlich oder abzählbar ist (oder allgemeiner wenn I geordnet ist), dann betrachten wir $\mathcal{B}$ und $\mathcal{B}^*$ als geordnete Listen.

Proposition 10.2.4. *Sei V ein Vektorraum und $\mathcal{B}$ eine Basis von V.*

(1) Falls $\dim V < \infty$, dann ist $\mathcal{B}^*$ eine Basis von V^*.

(2) $\mathcal{B}^*$ ist immer eine linear unabhängige Teilmenge von V^*.

Beweis. Da wir wissen, dass im Fall $\dim V < \infty$ gilt, dass $\dim V^* = \dim V$, genügt es, (2) zu zeigen. Nehmen wir dazu an, dass für paarweise verschiedene Indizes $i_1, \ldots, i_k$ in I

$$\sum_{j=1}^{k} \alpha_{i_j}\varphi_{i_j} = \alpha_{i_1}\varphi_{i_1} + \cdots + \alpha_{i_k}\varphi_{i_k} = 0 \tag{10.2}$$

in V^* gilt, mit $\alpha_{i_1}, \ldots, \alpha_{i_k} \in K$. Für $\ell \in \{1, \ldots, k\}$ wenden wir beide Seiten von (10.2) auf v_{i_ℓ} an: Da $\varphi_{i_j}(v_{i_\ell}) = \delta_{i_j,i_\ell}$, haben wir

$$\alpha_{i_\ell} = \sum_{j=1}^{k} \alpha_{i_j}\varphi_{i_j}(v_{i_\ell}) = 0.$$

Da dies für alle ℓ gilt, ist also $\alpha_{i_1} = \cdots = \alpha_{i_k} = 0$, was zeigt, dass $(\varphi_i)_{i\in I}$ linear unabhängig ist. □

Beispiel 10.2.5. [Berechnung von dualen Basen in K^n] Wir betrachten $\mathbb{Q}^2$ mit der Basis (v_1, v_2), wobei $v_1 = \begin{pmatrix}1\\2\end{pmatrix}$ und $v_2 = \begin{pmatrix}3\\4\end{pmatrix}$. Per Definition sind $\varphi_1, \varphi_2 : K^2 \to K$ von der dualen Basis definiert durch

$$\varphi_1\left(\begin{pmatrix}1\\2\end{pmatrix}\right) = 1, \qquad \varphi_1\left(\begin{pmatrix}3\\4\end{pmatrix}\right) = 0,$$
$$\varphi_2\left(\begin{pmatrix}1\\2\end{pmatrix}\right) = 0, \qquad \varphi_2\left(\begin{pmatrix}3\\4\end{pmatrix}\right) = 1$$

Suchen wir zuerst φ_1. Erinnern Sie sich von Beispiel 10.1.7, dass wir nur $[\varphi_1]_{\mathcal{E}_1}^{\mathcal{E}_2} =: (a, b)$ verstehen müssen. Anders gesagt suchen wir eine Lösung für

$$\begin{pmatrix}a & b\end{pmatrix}\begin{pmatrix}1\\2\end{pmatrix} = 1, \quad \begin{pmatrix}a & b\end{pmatrix}\begin{pmatrix}3\\4\end{pmatrix} = 0,$$

oder zu einer Gleichung zusammengefasst,

$$\begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \end{pmatrix}.$$

Analog dazu erfüllt $(c, d) := [\varphi_2]^{\mathcal{E}_2}_{\mathcal{E}_1}$ die Gleichung

$$\begin{pmatrix} c & d \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 0 & 1 \end{pmatrix}.$$

Wenn wir dies nochmals zusammenfassen, suchen wir also $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ mit

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2,$$

anders gesagt $\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}^{-1}.$

Wir können dies verallgemeinern: Wenn $\mathcal{B} = (v_1, \ldots, v_n)$ eine Basis von $K^n = K^n_{\text{Spal}}$ ist, $\mathcal{B}^* = (\varphi_1, \ldots, \varphi_n)$ die duale Basis und wir $\varphi \in (K^n_{\text{Spal}})^* = K^n_{\text{Zeil}}$ betrachten (wobei wir die beiden Räume gemäss Beispiel 10.1.7 identifizieren), dann haben wir

$$\begin{pmatrix} — & \varphi_1 & — \\ & \vdots & \\ — & \varphi_n & — \end{pmatrix} \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix} = I_n$$

oder äquivalent

$$\begin{pmatrix} — & \varphi_1 & — \\ & \vdots & \\ — & \varphi_n & — \end{pmatrix} = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix}^{-1}.$$

Dies gibt uns einen Algorithmus für die Berechnung der dualen Basis.

Wir wollen nun das obige Beispiel auf beliebige endlich-dimensionale Vektorräume verallgemeinern.

Proposition 10.2.6 (Berechnung von dualen Basen in endlich-dimensionalen Vektorräumen)**.** *Sei V ein endlich-dimensionaler Vektorraum über K und sei $\mathcal{B} = (v_1, \ldots, v_n)$ eine Basis und $\mathcal{B}^* = (\varphi_1, \ldots, \varphi_n)$ die duale Basis. Sei $\mathcal{E}$ eine andere Basis*[3] *von V*

[3] Denken Sie zum Beispiel an $V = \mathbb{R}[x]$ und die Standard-Basis $\{1, x, x^2, \ldots\}$.

(welche wir Standard-Basis nennen). Dann gilt

$$\begin{pmatrix} — & [\varphi_1]^{\mathcal{E}}_{\mathcal{E}_1} & — \\ & \vdots & \\ — & [\varphi_n]^{\mathcal{E}}_{\mathcal{E}_1} & — \end{pmatrix} [\mathrm{Id}_V]^{\mathcal{B}}_{\mathcal{E}} = I_n.$$

Beweis. Per Definition der Matrixmultiplikation haben wir

$$\begin{pmatrix} — & [\varphi_1]^{\mathcal{E}}_{\mathcal{E}_1} & — \\ & \vdots & \\ — & [\varphi_n]^{\mathcal{E}}_{\mathcal{E}_1} & — \end{pmatrix} [\mathrm{Id}_V]^{\mathcal{B}}_{\mathcal{E}} = \begin{pmatrix} — & [\varphi_1]^{\mathcal{E}}_{\mathcal{E}_1}[\mathrm{Id}_V]^{\mathcal{B}}_{\mathcal{E}} & — \\ & \vdots & \\ — & [\varphi_n]^{\mathcal{E}}_{\mathcal{E}_1}[\mathrm{Id}_V]^{\mathcal{B}}_{\mathcal{E}} & — \end{pmatrix} \underset{\text{Formel}}{\overset{\text{Transf.}}{=}} \begin{pmatrix} — & [\varphi_1]^{\mathcal{B}}_{\mathcal{E}_1} & — \\ & \vdots & \\ — & [\varphi_n]^{\mathcal{B}}_{\mathcal{E}_1} & — \end{pmatrix} = I_n,$$

wobei die letzte Gleichheit aus der Definition der dualen Basis folgt. □

Bemerkung 10.2.7. Dies gibt uns einen konkreten Algorithmus für die Berechnung der dualen Basis in beliebigen endlich-dimensionalen Vektorräumen.

Übung 10.2.8. Betrachten wir die Basis $\mathcal{B} = \{1, 1+x, 1+x+x^2\} \subseteq \mathbb{R}[x]_2$. Berechnen Sie $\mathcal{B}^*$.
Lösung: Wir schreiben $\mathcal{B}^* = (\varphi_1, \varphi_2, \varphi_3)$. Wenn wir die Standard-Basis $\mathcal{E} = (1, x, x^2)$ verwenden, haben wir

$$\begin{pmatrix} — & [\varphi_1]^{\mathcal{E}}_{\mathcal{E}_1} & — \\ & \vdots & \\ — & [\varphi_3]^{\mathcal{E}}_{\mathcal{E}_1} & — \end{pmatrix} = \begin{pmatrix} | & | & | \\ [1]_{\mathcal{E}} & [1+x]_{\mathcal{E}} & [1+x+x^2]_{\mathcal{E}} \\ | & | & | \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Also gilt zum Beispiel $[\varphi_1]_{\mathcal{E}} = (1, -1, 0)$, das heisst

$$\varphi_1(c + bx + ax^2) = c - b.$$

Ähnlich dazu gilt

$$\varphi_2(c + bx + ax^2) = b - a,$$
$$\varphi_3(c + bx + ax^2) = a.$$

Wir empfehlen es den Lesern, zu überprüfen, dass $(\varphi_1, \varphi_2, \varphi_3)$ tatsächlich die Bedingungen aus der Definition der dualen Basis erfüllt. Ausserdem möchten wir noch erwähnen, dass man die duale Basis natürlich auch direkt berechnen könnte:

Wenn wir einen allgemeinen Vektor in $\mathbb{R}[x]_2$ als $c+bx+ax^2$ in der Basis $\mathcal{B}$ schreiben, erhalten wir

$$c \cdot 1 + bx + ax^2 = (c-b)1 + (b-a)(1+x) + a(1+x+x^2).$$

Nun können wir die Definition der dualen Basis benutzen und φ_1, φ_2 und φ_3 direkt ablesen.

Kommen wir nun zu einer nützlichen Beobachtung, welche direkt aus der Definition der dualen Basis folgt.

Lemma 10.2.9. *Sei V endlich-dimensional mit Basis $\mathcal{B} = (v_1, \ldots, v_n)$ und dualer Basis $\mathcal{B}^* = (\varphi_1, \ldots, \varphi_n)$. Dann gilt für alle $v \in V$*

$$[v]_{\mathcal{B}} = \begin{pmatrix} \varphi_1(v) \\ \vdots \\ \varphi_n(v) \end{pmatrix} \tag{10.3}$$

und für alle $\varphi \in V^*$

$$[\varphi]_{\mathcal{B}^*} = \begin{pmatrix} \varphi(v_1) \\ \vdots \\ \varphi(v_n) \end{pmatrix}.$$

Beweis. Sei $[v]_{\mathcal{B}} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$, also $v = \sum_{k=1}^n \alpha_k v_k$. Dann gilt

$$\varphi_i(v) = \varphi_i\left(\sum_{k=1}^n \alpha_k v_k\right) = \sum_{k=1}^n \alpha_k \varphi_i(v_k) = \sum_{k=1}^n \alpha_k \delta_{ik} = \alpha_i,$$

was (10.3) zeigt. Ähnlich gilt für $\varphi \in V^*$ mit $[\varphi]_{\mathcal{B}^*} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}$, also $\varphi = \sum_{k=1}^n \beta_k \varphi_k$, dass

$$\varphi(v_i) = \left(\sum_{k=1}^n \beta_k \varphi_k\right)(v_i) = \sum_{k=1}^n \beta_k \varphi_k(v_i) = \sum_{k=1}^n \beta_k \delta_{ki} = \beta_i.$$ □

Dieses Lemma wird im nächsten Abschnitt nützlich sein. Die Leser haben vielleicht bemerkt, dass wir eine ähnliche Proposition in einem anderen Kontext gesehen haben – Proposition 6.4.2. Dies ist kein Zufall:

Beispiel 10.2.10. Sei $(V, \langle\,,\rangle)$ ein endlich-dimensionaler Skalarproduktraum über $\mathbb{F}$ (wobei $\mathbb{F} = \mathbb{R}$ oder $\mathbb{F} = \mathbb{C}$). Erinnern Sie sich, dass die Abbildung $v \mapsto \varphi_v := \langle v, \cdot\rangle \in V^*$ eine Identifikation von V und V^* gibt. Sei nun $\mathcal{B} = (e_1, \ldots, e_n)$ eine orthonormale Basis. Es folgt direkt aus den Definitionen, dass die duale Basis durch $\mathcal{B}^* = (\varphi_{e_1}, \ldots, \varphi_{e_n})$ gegeben ist. Also ist unter der obigen Identifikation von V mit V^* eine Orthonormalbasis ihre eigene Dualbasis!

Überprüfen Sie, dass dies nicht gilt, wenn $\mathcal{B}$ nicht orthonormal ist.

10.3 Kanonisch vs. nicht-kanonisch

Nehmen wir an, wir bekommen eine gewisse Information (zum Beispiel ein Vektorraum). Wenn wir daraus gewisse zusätzliche Informationen erhalten können, ohne eine zusätzliche Auswahl zu treffen, dann sagen wir, dass diese Zusatzinformation kanonisch (oder natürlich) definiert ist. Manchmal kann man eine Definition angeben, ohne etwas auszuwählen (zum Beispiel Definition 10.3.3 unten). Manchmal treffen wir bei der Berechnung einer zusätzlichen Information auch eine Wahl. Diese Information wird kanonisch genannt, wenn das Endresultat unabhängig von der getroffenen Wahl ist[4].

Wir vermeiden es, eine präzise Definition von den Wörtern kanonisch und natürlich anzugeben, und unser Ziel in diesem kurzen Abschnitt ist es, einige Abbildungen von Vektorräumen anzugeben und zu schauen, ob die Definitionen dieser Abbildungen von der Wahl einer Basis abhängen (in diesem Fall wären die Abbildungen dann nicht kanonisch) oder ob Sie unabhängig davon sind (dann nennen wir sie kanonisch).

Proposition 10.3.1. *Sei V ein K-Vektorraum, $\mathcal{B} = (v_i)_{i\in I} \subseteq V$ eine Basis und $\mathcal{B}^* = (\varphi_i)_{i\in I} \subseteq V^*$ die duale Menge zu $\mathcal{B}$. Die lineare Abbildung*

$$\Phi_{\mathcal{B}} : V \to V^*,$$

die durch $\Phi_{\mathcal{B}}(v_i) = \varphi_i$ für alle $i \in I$ definiert ist, ist injektiv.

Beweis. Dies ist äquivalent zu der Tatsache, dass $(\varphi_i)_{i\in I}$ linear unabhängig ist (wieso?). □

Die Abbildung $\Phi_{\mathcal{B}}$ hängt von der Wahl von $\mathcal{B}$ ab und ist ein Isomorphismus genau dann, wenn $\dim V < \infty$. Anders gesagt, falls V endlich-dimensional ist, gibt uns jede Wahl einer Basis einen Isomorphismus zwischen V und V^*.

10.3.1 Der Bidualraum

Definition 10.3.2. Sei V ein K-Vektorraum. Der Raum $V^{**} := (V^*)^*$ heisst der *Bidualraum* von V.

Hier ist nun ein typisches Beispiel für eine kanonische Abbildung:

Proposition/Definition 10.3.3. *Die* Evaluationsabbildung

$$\begin{aligned} \mathrm{ev} : V &\to V^{**} \\ v &\mapsto \mathrm{ev}_v, \end{aligned}$$

[4] Denken Sie zum Beispiel an die Determinante oder den Rang eines Endomorphismus, welche unter Verwendung einer Darstellungsmatrix berechnet werden können. Die Darstellungsmatrix hängt von der Wahl der Basis ab, die Determinante/der Rang aber nicht.

wobei $\mathrm{ev}_v(\varphi) := \varphi(v)$ *für alle* $\varphi \in V^*$*, ist linear und injektiv. Falls* V *endlich-dimensional ist, ist sie ein Isomorphismus.*

Beweis. Per Definition gilt für alle $\varphi \in V^*$

$$\mathrm{ev}_{\alpha v+\beta w}(\varphi) = \varphi(\alpha v+\beta w) = \alpha\varphi(v)+\beta\varphi(w) = \alpha\mathrm{ev}_v(\varphi)+\beta\mathrm{ev}_w(\varphi) = (\alpha\mathrm{ev}_v+\beta\mathrm{ev}_w)(\varphi)$$

und daher ist $\mathrm{ev}_{\alpha v+\beta w} = \alpha\mathrm{ev}_v + \beta\mathrm{ev}_w$ als Elemente von V^{**}, was die Linearität zeigt.

Sei $0 \neq v \in V$. Es existiert $\varphi : V \to K$ mit $\varphi(v) \neq 0$ (zum Beispiel kann man $\{v\}$ zu einer Basis von V erweitern und φ auf dieser Basis wie gewünscht definieren). Es folgt, dass $\mathrm{ev}_v(\varphi) = \varphi(v) \neq 0$ und daher ist $\mathrm{ev}_v \neq 0$. Dies zeigt die Injektivität von ev. Die letzte Aussage folgt aus einem Dimensionsargument. □

Bemerkung 10.3.4. Die Evaluationsabbildung ist kanonisch in dem Sinne, dass man keine Basis auswählen muss, um sie zu definieren. Falls $\dim V < \infty$, dann sind also V und V^{**} kanonisch isomorph. Um eine bessere Intuition dafür zu bekommen, finde ich die folgende Erklärung von Aloizio Macedo auf StackExchange sehr interessant:

> The dual is intuitively the space of „rulers“ (or measurement-instruments) of our vector space. Its elements measure vectors. This is what makes the dual space and its relatives so important in Differential Geometry, for instance. This immediately motivates the study of the dual space. For motivations in other areas, the other answers are quite well-versed.
> This also happens to explain intuitively some facts. For instance, the fact that there is no canonical isomorphism between a vector space and its dual can then be seen as a consequence of the fact that rulers need scaling, and there is no canonical way to provide one scaling for space. However, if we were to **measure the measure-instruments**, how could we proceed? Is there a canonical way to do so? Well, if we want to measure our measures, why not measure them by how they act on what they are supposed to measure? We need no bases for that. This justifies intuitively why there is a natural embedding of the space on its bidual. (Note, however, that this fails to justify why it is an isomorphism in the finite-dimensional case).

Hier ist noch eine andere Abbildung, die man ohne Wahl einer Basis definieren kann:

Definition 10.3.5. Die Bilinearform[5]

$$\begin{aligned} \langle\,,\rangle : V^* \times V &\to K \\ (\varphi, v) &\mapsto \langle\varphi, v\rangle := \varphi(v) \end{aligned}$$

[5]In Definition 8.1.1 haben wir Bilinearformen als Abbildungen $V \times V \to K$ definiert, doch wir erweitern diese Definition von nun an auf Abbildungen $\beta : V \times W \to K$, wobei V und W Vektorräume über K sind. Eine solche Funktion heisst *Bilinearform*, wenn sie die Bedingungen (1) und (2) aus Definition 8.1.1 erfüllt. Bemerken Sie, dass es im Fall $V \neq W$ keinen Sinn macht, von symmetrischen Bilinearformen zu sprechen, da für $v \in V$ und $w \in W$ gar nicht definiert ist, was $\beta(w, v)$ bedeutet.

heisst die *(kanonische) duale Paarung.*

Übung 10.3.6. Überprüfen Sie, dass $\langle \, , \rangle$ bilinear ist.

10.4 Die duale Abbildung

Eines der Ziele dieses Abschnitts ist den Lesern zu helfen, keine Angst vor abstrakten Begriffen zu haben, welche sie möglicherweise später im Studium nochmals antreffen werden.

Sei $\mathbf{Vek}_K$ die Klasse von Vektorräumen über einem Körper K.

Definition 10.4.1 (Kovarianter Funktor)**.** Ein *kovarianter Funktor* von $\mathbf{Vek}_K$ zu sich selbst ist eine Abbildung $F : \mathbf{Vek}_K \to \mathbf{Vek}_K$, sodass gilt:

- Für jeden Vektorraum $V \in \mathbf{Vek}_K$ gibt uns F einen Vektorraum $F(V) \in \mathbf{Vek}_K$.
- Für $V, W \in \mathbf{Vek}_K$ und $T \in \mathrm{Hom}(V, W)$ gibt uns F eine lineare Abbildung $F(T) \in \mathrm{Hom}(F(V), F(W))$, sodass

$$\begin{aligned} F(\mathrm{Id}_V) &= \mathrm{Id}_{F(V)} \quad \forall\, V \in \mathbf{Vek}_K, \\ F(S \circ T) &= F(S) \circ F(T) \quad \forall\, T \in \mathrm{Hom}(V, W),\ S \in \mathrm{Hom}(W, U). \end{aligned}$$

  In anderen Worten, Funktoren erhalten Identitätsabbildungen und Verkettungen.

Kovariante Funktoren erhalten die Reihenfolge von linearen Abbildungen, während kontravariante Funktoren die Reihenfolge umkehren. Wir geben die Definition an und schreiben die Stellen, die sich von der Definition oben unterscheiden, in einer anderen Farbe:

Definition 10.4.2 (Kontravarianter Funktor)**.** Ein *Kontravarianter Funktor* von $\mathbf{Vek}_K$ zu sich selbst ist eine Abbildung $F : \mathbf{Vek}_K \to \mathbf{Vek}_K$, sodass gilt:

- Für jeden Vektorraum $V \in \mathbf{Vek}_K$ gibt uns F einen Vektorraum $F(V) \in \mathbf{Vek}_K$.
- Für $V, W \in \mathbf{Vek}_K$ und $T \in \mathrm{Hom}(V, W)$ gibt uns F eine lineare Abbildung $F(T) \in \mathrm{Hom}(F(W), F(V))$, sodass

$$\begin{aligned} F(\mathrm{Id}_V) &= \mathrm{Id}_{F(V)} \quad \forall\, V \in \mathbf{Vek}_K \\ F(S \circ T) &= F(T) \circ F(S) \quad \forall\, T \in \mathrm{Hom}(V, W),\ S \in \mathrm{Hom}(W, U). \end{aligned}$$

Übung 10.4.3. Beweisen Sie die folgenden Aussagen:

(a) Wenn $F : \mathbf{Vek}_K \to \mathbf{Vek}_K$ ein kovarianter Funktor ist, dann gilt für alle $U, V, W \in \mathbf{Vek}_K$ und für alle linearen Abbildungen T, S, R wie unten, dass

$$\begin{tikzcd} U \arrow[r, "T"] \arrow[dr, "R"'] & V \arrow[d, "S"] \\ & W \end{tikzcd} \text{ kommutiert} \implies \begin{tikzcd} F(U) \arrow[r, "F(T)"] \arrow[dr, "F(R)"'] & F(V) \arrow[d, "F(S)"] \\ & F(W) \end{tikzcd} \text{ kommutiert.}$$

(b) Wenn $F : \mathbf{Vek}_K \to \mathbf{Vek}_K$ ein kontravarianter Funktor ist, dann gilt für alle $U, V, W \in \mathbf{Vek}_K$ und für alle linearen Abbildungen T, S, R wie unten:

$$\begin{tikzcd} U \arrow[r, "T"] \arrow[dr, "R"'] & V \arrow[d, "S"] \\ & W \end{tikzcd} \text{ kommutiert} \implies \begin{tikzcd} F(U) & F(V) \arrow[l, "F(T)"'] \\ & F(W) \arrow[ul, "F(R)"] \arrow[u, "F(S)"'] \end{tikzcd} \text{ kommutiert.}$$

Beispiel 10.4.4 (Der konstante Funktor)**.** Wir fixieren einen Vektorraum $V_0 \in \mathbf{Vek}_K$ und betrachten den Funktor

$$\begin{aligned} C : \mathbf{Vek}_K &\to \mathbf{Vek}_K \\ V &\mapsto V_0, \end{aligned}$$

der alle Vektorräume auf V_0 abbildet und für beliebige $V, W \in \mathbf{Vek}_K$

$$\begin{aligned} C : \mathrm{Hom}(V, W) &\to \mathrm{Hom}(C(V), C(W)) = \mathrm{Hom}(V_0, V_0) \\ T &\mapsto \mathrm{Id}_{V_0} \end{aligned}$$

alle linearen Abbildungen $T : V \to W$ auf die Identität abbildet. Dies ist ein ziemlich langweiliger kovarianter Funktor (es ist sogar gleichzeitig auch ein kontravarianter Funktor).

Wir möchten nun ein erstes Beispiel von einem interessanten Funktor machen, indem wir den Dualraum als Funktor betrachten. Wir definieren $F : \mathbf{Vek}_K \to \mathbf{Vek}_K$ als $F(V) = V^*$. Um einen Funktor zu haben, müssen wir noch $F(T)$ für alle $T \in \mathrm{Hom}(V, W)$ definieren[6].

Definition 10.4.5. Sei $T \in \mathrm{Hom}(V, W)$, wobei $V, W \in \mathbf{Vek}_K$. Wir definieren die *duale Abbildung* $T^* : W^* \to V^*$ durch

$$T^*(\psi) = \psi \circ T \quad \text{für } \psi \in W^*.$$

Bemerkung 10.4.6. Sie haben vielleicht bemerkt, dass wir die Notation T^* schon einmal verwendet haben, nämlich für die adjungierte Abbildung in Skalarprodukträumen. Wir werden in Bemerkung 10.5.10 nochmals auf diesen Punkt eingehen.

[6]Die Leser, welche jetzt schrecklich verwirrt sind von diesen Funktorgeschichten, können alles bisher gesagte ignorieren und ab hier weiterlesen.

Proposition 10.4.7. *Die duale Abbildung T^* ist wohldefiniert*[7] *und linear.*

Beweis. Für die Wohldefiniertheit müssen wir zeigen, dass $\psi \circ T \in V^*$ für $\psi \in W^*$, also dass $\psi \circ T$ linear ist. Dies gilt, weil die Verkettung von linearen Abbildungen linear ist.

Für die Linearität von T^* betrachten wir $\psi_1, \psi_2 \in W^*$ und $\alpha_1, \alpha_2 \in K$. Es gilt

$$\begin{aligned} T^*(\alpha_1\psi_1 + \alpha_2\psi_2) &= (\alpha_1\psi_1 + \alpha_2\psi_2) \circ T = (\alpha_1\psi_1) \circ T + (\alpha_2\psi_2) \circ T \\ &= \alpha_1(\psi_1 \circ T) + \alpha_2(\psi_2 \circ T) = \alpha_1 T^*(\psi_1) + \alpha_2 T^*(\psi_2). \end{aligned}$$ □

Proposition 10.4.8. *Seien $V, W, U \in \mathbf{Vek}_K$ und seien $T : V \to W$ und $S : W \to U$ lineare Abbildungen.*

(1) $(S \circ T)^* = T^* \circ S^*$.

(2) $(\mathrm{Id}_V)^* = \mathrm{Id}_{V^*}$.

(3) Falls $T : V \to W$ ein Isomorphismus ist, dann ist $T^* : W^* \to V^*$ auch ein Isomorphismus und die Inverse ist gegeben durch $(T^*)^{-1} = (T^{-1})^*$.

Beweis. (1): Für alle $\varphi \in U^*$ gilt

$$(S \circ T)^*(\varphi) = \varphi \circ (S \circ T) = (\varphi \circ S) \circ T = T^*(\varphi \circ S) = T^*(S^*\varphi) = T^* \circ S^*(\varphi).$$

(2): Für alle $\varphi \in V^*$ gilt

$$(\mathrm{Id}_V)^*(\varphi) = \varphi \circ \mathrm{Id}_V = \varphi = \mathrm{Id}_{V^*}(\varphi).$$

(3): Sei $T^{-1} : W \to V$ die Inverse von T. Es gilt

$$\mathrm{Id}_{V^*} \overset{(2)}{=} (\mathrm{Id}_V)^* = (T^{-1} \circ T)^* \overset{(1)}{=} T^* \circ (T^{-1})^*,$$

$$\mathrm{Id}_{W^*} \overset{(2)}{=} (\mathrm{Id}_W)^* = (T \circ T^{-1})^* \overset{(1)}{=} (T^{-1})^* \circ T^*,$$

daher ist $(T^{-1})^*$ die Inverse von T^*, was die Aussage beweist. □

Bemerkung 10.4.9. Die Aussagen (1) und (2) von Proposition 10.4.8 zeigen, dass $F(V) = V^*$ mit $F(T) = T^*$ ein kontravarianter Funktor ist. Zu diesem Zeitpunkt sollten die Leser die Aussage „Der Dualraum und die duale Abbildung bilden einen kontravarianten Funktor“ als eine Art zum Zusammenfassen von Informationen sehen.

Später in Ihrem Studium werden sie dann vielleicht lernen, wie man mit dieser Sprache weitere Konstruktionen machen kann, welche einem helfen, Beziehungen zwischen

[7] Bemerken Sie, dass es zwei verschiedene Dinge bedeuten kann, Wohldefiniertheit zu zeigen. Das eine ist, zu zeigen, dass eine Definition nicht von einer getroffenen Auswahl abhängt (zum Beispiel in der Definition 1.2.12 der Addition in $\mathbb{Z}/n\mathbb{Z}$). Das andere ist, zu zeigen, dass die Bilder aller Elemente des Definitionsbereichs auch tatsächlich im Zielbereich liegen. Letzteres ist, was wir hier tun müssen.

verschiedenen Funktoren zu sehen. Wir werden im nächsten Kapitel noch etwas mehr dazu sagen.

Beispiel 10.4.10. Wir betrachten die lineare Abbildung $D : \mathbb{R}[x] \to \mathbb{R}[x]$ definiert durch $Dp = p'$, wobei p' die Ableitung von p bezeichnet. Die duale Abbildung von D ist also eine lineare Abbildung $D^* : (\mathbb{R}[x])^* \to (\mathbb{R}[x])^*$. Sei $\varphi : \mathbb{R}[x] \to \mathbb{R}$ die Linearform definiert durch $\varphi(p) = p(3)$. Dann ist $D^*(\varphi) \in (\mathbb{R}[x])^*$ gegeben durch

$$(D^*(\varphi))(p) = (\varphi \circ D)(p) = \varphi(Dp) = \varphi(p') = p'(3).$$

In anderen Worten, $D^*(\varphi)$ ist die Linearform auf $\mathbb{R}[x]$, welche p auf $p'(3)$ abbildet.

Betrachten wir nun $\psi \in (\mathbb{R}[x])^*$ definiert durch $\psi(p) = \int_0^1 p(x)dx$. Dann ist $D^*(\psi)$ gegeben durch

$$(D^*(\psi))(p) = (\psi \circ D)(p) = \psi(Dp) = \psi(p') = \int_0^1 p'(x)dx = p(1) - p(0).$$

In anderen Worten, $D^*(\psi)$ ist die Linearform auf $\mathbb{R}[x]$, welche p auf $p(1)-p(0)$ abbildet.

Proposition 10.4.11. *Sei $T : V \to W$ eine lineare Abbildung mit $\dim V = n$ und $\dim W = m$ und seien $\mathcal{B} \subseteq V$, $\mathcal{C} \subseteq W$ geordnete Basen. Dann gilt*

$$[T^*]^{\mathcal{C}^*}_{\mathcal{B}^*} = \left([T]^{\mathcal{B}}_{\mathcal{C}}\right)^T. \tag{10.4}$$

Beweis. Sei $e_1, e_2, \ldots$ die Standardbasis von K^n oder K^m (als Spaltenvektoren betrachtet). Wir schreiben $\mathcal{B} = (v_1, \ldots, v_n)$ und $\mathcal{C}^* = (\psi_1, \ldots, \psi_m)$. Der ij-Eintrag der linken Seite (wobei $1 \le i \le n$ und $1 \le j \le m$) ist

$$\begin{aligned} e_i^T [T^*]^{\mathcal{C}^*}_{\mathcal{B}^*} e_j &= e_i^T [T^*]^{\mathcal{C}^*}_{\mathcal{B}^*} [\psi_j]_{\mathcal{C}^*} = e_i^T [T^*\psi_j]_{\mathcal{B}^*} = i\text{-Komponente von } [T^*\psi_j]_{\mathcal{B}^*} \\ &\underset{10.2.9}{\overset{\text{Lemma}}{=}} T^*\psi_j(v_i) \overset{\text{Def.}}{=} \psi_j(Tv_i) \underset{10.2.9}{\overset{\text{Lemma}}{=}} j\text{-Komponente von } [Tv_i]_{\mathcal{C}} \\ &= e_j^T [Tv_i]_{\mathcal{C}} = e_j^T [T]^{\mathcal{B}}_{\mathcal{C}} [v_i]_{\mathcal{B}} = e_j^T [T]^{\mathcal{B}}_{\mathcal{C}} e_i, \end{aligned}$$

was der ji-Eintrag von $[T]^{\mathcal{B}}_{\mathcal{C}}$ und somit der ij-Eintrag der rechten Seite von (10.4) ist. □

Beispiel 10.4.12. Wir schränken die lineare Abbildung $D : \mathbb{R}[x] \to \mathbb{R}[x]$ von Beispiel 10.4.10 auf den Unterraum $\mathbb{R}[x]_2$ ein und wir betrachten die Basis $\mathcal{B} = (1, 1+x, 1+x+x^2)$ mit der dualen Basis $\mathcal{B}^* = (\varphi_1, \varphi_2, \varphi_3)$, wie in Übung 10.2.8 berechnet. Wir haben

$$\begin{aligned} D(1) &= 0, \\ D(1+x) &= 1, \\ D(1+x+x^2) &= 2x+1 = -1 + 2(1+x), \end{aligned}$$

also ist $[D]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} 0 & 1 & -1 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}$. Es folgt aus Proposition 10.4.11, dass $[D^*]_{\mathcal{B}^*}^{\mathcal{B}^*} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ -1 & 2 & 0 \end{pmatrix}$.

Wir möchten nun die Linearform $\varphi : p \mapsto p(3)$ in der Basis $\mathcal{B}^*$ darstellen. Dafür berechnen wir für $p = c + bx + ax^2$

$$\varphi(p) = \varphi(c+bx+ax^2) = c+3b+9a = (c-b)+4(b-a)+13a = \varphi_1(p)+4\varphi_2(p)+13\varphi_3(p),$$

also ist $\varphi = \varphi_1 + 4\varphi_2 + 13\varphi_3$. Weiter berechnen wir

$$[D^*\varphi]_{\mathcal{B}^*} = [D^*]_{\mathcal{B}^*}^{\mathcal{B}^*}[\varphi]_{\mathcal{B}^*} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ -1 & 2 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 4 \\ 13 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 7 \end{pmatrix},$$

also ist $D^*\varphi = \varphi_2 + 7\varphi_3$, was bedeutet, dass

$$D^*\varphi(c + bx + ax^2) = 1(b - a) + 7a = b + 6a.$$

Dies stimmt natürlich mit $p \mapsto p'(3)$ überein:

$$c + bx + ax^2 \mapsto (b + 2ax)(3) = b + 6a.$$

10.5 Annulator

Wenn wir die duale Paarung

$$\langle\, , \rangle : V^* \times V \to K$$

betrachten, sehen wir Ähnlichkeiten zur Orthogonalität und insbesondere zum orthogonalen Komplement. Wir präzisieren:

Definition 10.5.1. Für einen Vektorraum V und eine Teilmenge $\emptyset \neq U \subseteq V$ definieren wir den *Annulator* von U als

$$\begin{aligned} U^0 &:= \{\varphi \in V^* \mid \varphi(u) = 0 \ \forall u \in U\} \\ &= \{\varphi \in V^* \mid \langle \varphi, u \rangle = 0 \ \forall u \in U\} \\ &= \{\varphi \in V^* \mid U \subseteq \ker \varphi\}. \end{aligned}$$

Bemerkung 10.5.2. Einige Autoren nennen U^0 sogar den *orthogonalen Raum* zu U.

Übung 10.5.3. Zeigen Sie, dass $U^0 \subseteq V^*$ ein Untervektorraum ist.

Proposition 10.5.4. *Sei V endlich-dimensional und $U \subseteq V$ ein Unterverktorraum. Es gilt* $\dim U^0 = \dim V - \dim U$. *Genauer gesagt, wenn $(u_1, \ldots, u_k)$ eine Basis von U ist und $(u_1, \ldots, u_k, v_1, \ldots, v_{n-k})$ eine Erweiterung zu einer Basis von V, dann ist* $U^0 = \mathrm{Sp}(\psi_1, \ldots, \psi_{n-k})$, *wobei*

$$\mathcal{B}^* = (\varphi_1, \ldots, \varphi_k, \psi_1, \ldots, \psi_{n-k})$$

die duale Basis von V^* bezüglich der Basis $\mathcal{B} = (u_1, \ldots, u_k, v_1, \ldots, v_{n-k})$ ist.

Beweis. Per Definition von $\psi_1, \ldots, \psi_{n-k}$ liegen $u_1, \ldots, u_k$ in deren Kern, also sind $\psi_1, \ldots, \psi_{n-k} \in U^0$ und somit $\mathrm{Sp}(\psi_1, \ldots, \psi_{n-k}) \subseteq U^0$.

Für die andere Inklusion sei $\varphi \in U^0$. Wir schreiben

$$\varphi = \alpha_1\varphi_1 + \cdots + \alpha_k\varphi_k + \beta_1\psi_1 + \cdots + \beta_{n-k}\psi_{n-k}.$$

Da $\varphi \in U^0$, gilt $\varphi(u_i) = 0$ für $i = 1, \ldots, k$ und somit gilt

$$0 = \varphi(u_i) = \alpha_1\varphi_1(u_i) + \cdots + \beta_{n-k}\psi_{n-k}(u_i) = \alpha_i.$$

Also ist $\alpha_1 = \cdots = \alpha_k = 0$ und somit ist $\varphi \in \mathrm{Sp}(\psi_1, \ldots, \psi_{n-k})$. □

Wir möchten nun die Beziehung von $\mathrm{Im}(T)$, $\mathrm{Im}(T^*)$, $\ker(T)$ und $\ker(T^*)$ untersuchen.

Proposition 10.5.5. *Seien V, W endlich-dimensional und $T : V \to W$.*

(1) $(\mathrm{Im}\, T)^0 = \ker T^*$.

(2) $(\ker T)^0 = \mathrm{Im}\, T^*$.

Bemerkung 10.5.6. Vergleichen Sie Proposition 10.5.5 mit Lemma 7.1.5.

Beweis. (1): Sei $\varphi \in \ker T^*$. Per Definition ist $T^*(\varphi) = \varphi \circ T = 0$, also gilt

$$\forall v \in V \quad \varphi \circ T(v) = \varphi(Tv) = 0$$

und somit ist $\varphi \in (\mathrm{Im}\, T)^0$. Für die andere Inklusion sei $\varphi \in (\mathrm{Im}\, T)^0$. Das heisst, dass

$$\forall v \in V \quad \varphi(Tv) = 0.$$

Daraus folgt, dass $\varphi \circ T = 0$ als Linearform und somit $T^*(\varphi) = 0$, also ist $\varphi \in \ker T^*$.

(2): Sei $\varphi \in \mathrm{Im}\, T^*$, das heisst es existiert $\psi \in W^*$ mit $\varphi = T^*\psi = \psi \circ T$. Also haben wir für alle $v \in \ker T$

$$\varphi(v) = \psi T(v) = \psi(0) = 0,$$

also ist $\varphi \in (\ker T)^0$ und somit ist $\operatorname{Im} T^* \subseteq (\ker T)^0$. Anstatt die umgekehrte Inklusion direkt zu zeigen, machen wir ein Dimensionsargument:

$$\dim \operatorname{Im} T^* = \dim W - \dim \ker T^* \overset{(1)}{=} \dim W - \dim(\operatorname{Im} T)^0 \tag{10.5}$$
$$\underset{10.5.4}{\overset{\text{Prop.}}{=}} \dim W - (\dim W - \dim \operatorname{Im} T) = \dim \operatorname{Im} T \tag{10.6}$$
$$\underset{\text{satz}}{\overset{\text{Rang-}}{=}} \dim V - \dim \ker T \underset{10.5.4}{\overset{\text{Prop.}}{=}} \dim(\ker T)^0, \tag{10.7}$$

also impliziert $\operatorname{Im} T^* \subseteq (\ker T)^0$ Gleichheit. □

Korollar 10.5.7. *Seien V, W endlich-dimensional und $T : V \to W$ linear. Dann gilt:*

(1) T ist injektiv $\iff$ T^* ist surjektiv.

(2) T ist surjektiv $\iff$ T^* ist injektiv.

Beweis. Es gilt

$$\begin{aligned} T \text{ injektiv} \iff \ker T = \{0_V\} \iff (\ker T)^0 = V^* &\underset{10.5.4}{\overset{\text{Prop.}}{\iff}} \operatorname{Im} T^* = V^* \\ &\iff T^* \text{ surjektiv} \end{aligned}$$

und ähnlich

$$\begin{aligned} T \text{ surjektiv} \iff \operatorname{Im} T = W \iff (\operatorname{Im} T)^0 = \{0_{W^*}\} &\iff \ker T^* = \{0_{W^*}\} \\ &\iff T^* \text{ injektiv.} \end{aligned}$$ □

Korollar 10.5.8. *Spaltenrang = Zeilenrang.*

Beweis. Schauen Sie, was wir in (10.5) gezeigt haben:

$$\dim \operatorname{Im} T = \dim \operatorname{Im} T^*. \tag{10.8}$$

Für $T = m_A$ gilt aber

$$\dim \operatorname{Im} T = \dim \operatorname{Im} m_A = \text{ Spaltenrang von } A$$

und

$$\dim \operatorname{Im} T^* \underset{10.4.11}{\overset{\text{Prop.}}{=}} \dim \operatorname{Im} m_{A^T} = \text{ Zeilenrang von } A.$$ □

Es lohnt sich, (10.8) nochmals separat zu erwähnen:

Korollar 10.5.9. $\dim \operatorname{Im} T = \dim \operatorname{Im} T^*$.

Bemerkung 10.5.10. Wenn $(V, \langle\ ,\ \rangle)$ ein Skalarproduktraum ist, haben wir gezeigt, dass V und V^* kanonisch isomorph (bzw. anti-isomorph im Falle von unitären Vektorräumen) sind durch die Abbildung $v \mapsto \langle \cdot, v \rangle$. Wir können die beiden also miteinander identifizieren. Die Leser finden es vielleicht interessant, dies mit den Ergebnissen aus den Kapiteln 6 und 7 zu vergleichen. Zum Beispiel ist die Adjungierte Abbildung T^* nichts anderes als die duale Abbildung T^* unter dieser Identifikation von V und V^*. Ähnlich dazu ist der Annulator mit dem orthogonalen Komplement verwandt und so weiter.

Changelog: Kapitel 10

- 19.05: Übung 10.2.8 wurde von Beispiel zu Übung umbenannt.
- 19.05: Nach Definition 10.2.1 wurde ein Eintrag zur Notation der dualen Basis eingefügt.
- 20.05: In Beispiel 10.1.3 wurde eine Fussnote hinzugefügt.
- 21.05: In Bemerkung 10.4.9 wurde kovariant zu kontravariant korrigiert.
- 24.05: In Beispiel 10.2.5 wurde $\mathcal{B}^* = (v_1, \ldots, v_n)$ zu $\mathcal{B}^* = (\varphi_1, \ldots, \varphi_n)$ korrigiert.
- 24.05: In Proposition 10.4.11 wurde $T : T \to W$ zu $T : V \to W$ korrigiert.
- 27.05: Am Ende von Beispiel 10.4.10 wurde $D^*(\varphi)$ zu $D^*(\psi)$ korrigiert.
- 28.05: In Übung 10.1.5 wurde $j \in K$ zu $j \in I$ korrigiert.
- 29.05: Bemerkung 10.4.6 wurde hinzugefügt.
- 31.05: In Beispiel 10.4.12 wurde $D^*\varphi = 2\varphi_2 + 3\varphi_3$ zu $D^*\varphi = \varphi_2 + 7\varphi_3$ korrigiert.
- 31.05: Im Beweis von Proposition 10.3.3 wurde an einer Stelle $\mathrm{ev}_w(\varphi)$ zu $\beta\mathrm{ev}_w(\varphi)$ korrigiert.
- 22.06: In Proposition 10.3.1 wurde spezifiziert, dass die Abbildung $\Phi_{\mathcal{B}}$ linear ist.
- 27.06: In Übung 10.1.6 wurde an einer Stelle $(v_i)_{i\in I}$ zu $(V_i)_{i\in I}$ korrigiert.
- 10.07: In Beispiel 10.1.7 wurde $[v]_{\mathcal{B}}$ zu $[\varphi]^{\mathcal{B}}_{\mathcal{E}_1}$ korrigiert.

Kapitel 11

Multilineare Algebra

11.1 Einleitung und Überblick

Unser Ziel in diesem Kapitel ist sehr bescheiden: Für gegebene Vektorräume V, W über einem Körper K und eine Zahl $k \in \mathbb{N}$ werden wir den Vektorraum von multilinearen Abbildungen (auch k-lineare Abbildungen genannt) von V^k nach W definieren, sowie die zwei interessanten Untervektorräume aller symmetrischen bzw. alternierenden multilinearen Abbildungen. Nach der Definition überzeugen wir uns davon, dass es nicht so leicht ist, diese Vektorräume direkt zu untersuchen, wie wir es für lineare und bilineare Abbildungen getan haben. Wir werden deshalb einen anderen Ansatz entwickeln. Wir illustrieren nun diesen Ansatz:

Nehmen wir zum Beispiel an, wir möchten die Menge aller alternierenden multilinearen Abbildungen $V^k \to W$ verstehen/parametrisieren. Wir werden einen Vektorraum konstruieren, in diesem Fall $\mathrm{Alt}^k(V)$, sodass multilineare Abbildungen von V^k nach W in gewisser Weise äquivalent sind zu linearen Abbildungen von $\mathrm{Alt}^k(V)$ nach W. Genauer gesagt werden wir den Vektorraum $\mathrm{Alt}^k(V)$ zusammen mit einer alternierenden multilinearen Abbildung $\iota : V^k \to \mathrm{Alt}^k(V)$ konstruieren, sodass für jede alternierende multilineare Abbildung $\varphi : V^k \to W$ eine eindeutige *lineare* Abbildung $\psi : \mathrm{Alt}^k(V) \to W$ existiert mit $\psi \circ \iota = \varphi$. In der Sprache von kommutativen Diagrammen kann man das wie folgt ausdrücken:

$$\begin{tikzcd} V^k \arrow[rr, "\forall\varphi"] \arrow[dr, "\iota"'] & & W \\ & \mathrm{Alt}^k(V) \arrow[ur, dashed, "\exists!\psi"'] & \end{tikzcd}$$

In anderen Worten, man sollte $\iota : V^k \to \mathrm{Alt}^k(V)$ als die „Mutter aller alternierenden multilinearen Abbildungen" von V^k anschauen, denn jede andere alternierende multilineare Abbildung entspricht einer eindeutigen linearen Abbildung in $\mathrm{Hom}(\mathrm{Alt}^k(V), W)$ via Präkomposition mit ι.

Bemerken Sie, dass der Fall von alternierenden multilinearen Abbildungen nur ein Beispiel ist: Die Parametrisierung von anderen Familien von Abbildungen wird jeweils der Konstruktion eines anderen Vektorraums zusammen mit einer „Mutterabbildung“ entsprechen. Das wichtigste Beispiel wird die Konstruktion eines Vektorraums $V_1 \otimes V_2$ sein, welchen wir das Tensorprodukt von V_1 und V_2 nennen werden, zusammen mit einer bilinearen Abbildung $\iota : V_1 \times V_2 \to V_1 \otimes V_2$, welche uns die folgende Parametrisierung aller bilinearen Abbildungen erlauben wird:

Für alle $\varphi : V_1 \times V_2 \to W$ bilinear existiert eine eindeutige lineare Abbildung $\psi \in \mathrm{Hom}(V_1 \otimes V_2, W)$ mit $\psi \circ \iota = \varphi$. Mittels kommutiver Diagramme lässt sich das wie folgt darstellen:

$$\begin{tikzcd} V_1 \times V_2 \arrow[rr, "\forall\varphi"] \arrow[dr, "\iota"'] & & W \\ & V_1 \otimes V_2 \arrow[ur, dashed, "\exists!\psi"'] & \end{tikzcd}$$

Die Leser sehen hoffentlich bereits die Ähnlichkeit der beiden Beispiele $\mathrm{Alt}^k(V)$ und $V_1 \otimes V_2$. Beginnen wir nun mit der allgemeinen Theorie.

11.2 Definitionen und Beispiele

Wir fixieren für dieses Kapitel einen Körper K. Alle Vektorräume sind Vektorräume über K. Erinnern Sie sich, dass wir zwei Notationen für Funktionsräume gesehen haben: Wenn W ein Vektorraum ist und S eine Menge, dann bezeichnen wir mit $\mathrm{Abb}(S, W)$ oder W^S den Vektorraum aller Funktionen $f : S \to W$. Zusammen mit den Operationen

$$\begin{aligned} \forall s \in S \quad (f_1 + f_2)(s) &:= f_1(s) + f_2(s) \\ \forall s \in S \quad (\alpha f)(s) &:= \alpha f(s) \end{aligned}$$

ist dies ein Vektorraum über K.

Definition 11.2.1. Seien $V_1, \ldots, V_k, W$ Vektorräume über K. Eine Abbildung

$$\varphi : V_1 \times \cdots \times V_k \to W$$

heisst *multilinear*, falls sie in jeder Variable linear ist. Das heisst, dass für alle $i = 1, \ldots, k$, $\alpha, \beta \in K$, $v, v' \in V_i$ und $v_j \in V_j$ für alle $j \neq i$ gilt, dass

$$\begin{aligned} \varphi(v_1, \ldots, v_{i-1}, \alpha v + \beta v', v_{i+1}, \ldots, v_k) = {} & \alpha\varphi(v_1, \ldots, v_{i-1}, v, v_{i+1}, \ldots, v_k) \\ & + \beta\varphi(v_1, \ldots, v_{i-1}, v', v_{i+1}, \ldots, v_k). \end{aligned}$$

Wir schreiben $\mathrm{Mult}_K(V_1, \ldots, V_k; W)$ oder $\mathrm{Mult}_K(V_1 \times \cdots \times V_k, W)$ für den Raum aller solchen multilinearen Abbildungen wie oben definiert. Wenn der Körper K aus dem

Kontext klar ist, schreiben wir meistens $\mathrm{Mult}(V_1, \ldots, V_k; W)$ oder $\mathrm{Mult}(V_1 \times \cdots \times V_k, W)$. Wir interessieren uns insbesondere für den Fall $V_1 = \cdots = V_k = V$, dann schreiben wir auch

$$\mathrm{Mult}^k_K(V, W) = \mathrm{Mult}_K(V^k, W) = \mathrm{Mult}_K(\underbrace{V \times \cdots \times V}_{k \text{ mal}}, W).$$

Auch hier schreiben wir meistens nur $\mathrm{Mult}^k(V, W)$, wenn K aus dem Kontext klar ist. Elemente von $\mathrm{Mult}^k(V, W)$ heissen auch *k-lineare Abbildungen* von V nach W.

Proposition 11.2.2. *Der Raum* $\mathrm{Mult}(V_1 \times \cdots \times V_k, W)$ *ist ein Untervektorraum des Vektorraums* $\mathrm{Abb}(V_1 \times \cdots \times V_k, W)$ *aller Funktionen* $V_1 \times \cdots \times V_k \to W$.

Beweis. Wir überlassen es den Lesern, dies zu überprüfen. □

Konzentrieren wir uns nun auf den Raum $\mathrm{Mult}^k(V, W)$. Dieser enthält zwei wichtige Untervektorräume:

Definition 11.2.3. Eine multilineare Abbildung $\varphi \in \mathrm{Mult}^k(V, W)$ heisst *symmetrisch*, falls

$$\forall \sigma \in S_k \quad \varphi(v_{\sigma(1)}, \ldots, v_{\sigma(k)}) = \varphi(v_1, \ldots, v_k).$$

Die Menge aller symmetrischen multilinearen Abbildungen $V^k \to W$ bezeichnen wir mit $\mathrm{Sym}^k_K(V, W)$ (oder nur $\mathrm{Sym}^k(V, W)$).

Definition 11.2.4. Eine multilineare Abbildung $\varphi \in \mathrm{Mult}^k(V, W)$ heisst *alternierend*, falls für alle $v_1, \ldots, v_k \in V$ gilt

$$\exists i \neq j \in \{1, \ldots, k\} \text{ mit } v_i = v_j \quad \Longrightarrow \quad \varphi(v_1, \ldots, v_k) = 0.$$

Die Menge aller alternierenden multilinearen Abbildungen $V^k \to W$ bezeichnen wir mit $\mathrm{Alt}^k_K(V, W)$ (oder nur $\mathrm{Alt}^k(V, W)$).

Bemerkung 11.2.5. Eine eng verwandte Definition ist die folgende: Eine multilineare Abbildung $\varphi \in \mathrm{Mult}^k(V, W)$ heisst *antisymmetrisch*, falls

$$\forall \sigma \in S_k \quad \varphi(v_{\sigma(1)}, \ldots, v_{\sigma(k)}) = \mathrm{sign}(\sigma)\varphi(v_1, \ldots, v_k),$$

wobei $\mathrm{sign}(\sigma)$ das Signum von σ ist, welches wir in Definition 4.3.13 definiert haben. Wenn $\mathrm{char}(K) \neq 2$, dann ist eine multilineare Abbildung genau dann alternierend, wenn sie antisymmetrisch ist, weshalb man meistens diese beiden Begriffe miteinander gleichsetzt.

Übung 11.2.6. Zeigen Sie, dass

(a) Alternierend $\Longrightarrow$ antisymmetrisch.

(b) Wenn $\mathrm{char}(K) \neq 2$, dann gilt: Antisymmetrisch $\iff$ alternierend.

(c) Wenn $\mathrm{char}(K) = 2$, dann gilt: Antisymmetrisch $\iff$ symmetrisch.

Proposition 11.2.7. *Die Räume* $\mathrm{Sym}_K^k(V,W)$ *und* $\mathrm{Alt}_K^k(V,W)$ *sind Untervektorräume von* $\mathrm{Mult}_K^k(V,W)$ *(und damit auch von* $\mathrm{Abb}(V^k,W)$*).*

Beweis. Auch diesen Beweis überlassen wir den Lesern. □

Beispiel 11.2.8.

(1) Eine Abbildung $\Phi : V \times V \to K$ ist multilinear genau dann, wenn sie bilinear ist. Sie ist symmetrisch als multilineare Abbildung genau dann, wenn sie symmetrisch als Bilinearform ist.

(2) Eine Bilinearform $\Phi : V \times V \to K$ ist antisymmetrisch als multilineare Abbildung (also alternierend falls $\mathrm{char}(K) \neq 2$), falls $\Phi(v,w) = -\Phi(w,v)$ für alle $v,w \in V$. Solche Bilinearformen heissen *schiefsymmetrisch.*

(3) Eine Abbildung $\Phi : V \to W$ ist multilinear genau dann, wenn sie linear ist. Eine solche Abbildung ist immer auch symmetrisch und alternierend. Es gilt also

$$\mathrm{Hom}_K(V,W) = \mathrm{Mult}_K(V;W) = \mathrm{Mult}^1(V,W) = \mathrm{Sym}^1(V,W) = \mathrm{Alt}^1(V,W).$$

(4) Die Abbildung

$$\begin{aligned} \det : \qquad (K^n)^n &\to K \\ (v_1,\ldots,v_n) &\mapsto \det\begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix} \end{aligned}$$

ist multilinear und alternierend, das heisst $\det \in \mathrm{Alt}^n(K^n,K)$. Wir haben eigentlich in Kapitel 4 gezeigt, dass det die einzige multilineare alternierende Abbildung $(K^n)^n \to K$ ist bis auf einen Skalar. Dies entspricht der folgenden Aussage, die wir in diesem Kapitel beweisen werden:

$$\dim \mathrm{Alt}^n(K^n,K) = 1.$$

(5) Erinnern Sie sich, dass die Determinante einer Matrix $A = (a_{ij})_{i,j}$ durch

$$\det(A) = \sum_{\sigma\in S_n} \mathrm{sign}(\sigma) \prod_{i=1}^n a_{i\sigma(i)}$$

gegeben ist. Ähnlich kann man die *Permanente* einer Matrix $A = (a_{ij})_{i,j}$ durch

$$\operatorname{Perm}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} a_{i\sigma(i)}$$

definieren. Die Permanente ist eine symmetrische multilineare Abbildung. Diese wird vor allem in der Kombinatorik und Wahrscheinlichkeitstheorie verwendet.

(6) Die Abbildung

$$\begin{aligned} K^n_{\text{Spal}} \times K^n_{\text{Zeil}} &\to M_{n\times n}(K) \\ (v, w) &\mapsto vw \end{aligned}$$

ist multilinear (was in diesem Fall das Gleiche wie bilinear bedeutet). In Koordinaten ist diese Abbildung gegeben durch

$$\left(\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}, (w_1, \ldots, w_n) \right) \mapsto \begin{pmatrix} v_1 w_1 & \cdots & v_1 w_n \\ \vdots & \ddots & \vdots \\ v_n w_1 & \cdots & v_n w_n \end{pmatrix}.$$

(7) Seien $\varphi_1 \in V_1^*, \ldots, \varphi_n \in V_n^*$. Dann ist die Abbildung

$$\begin{aligned} V_1 \times \cdots \times V_n &\to K \\ (v_1, \ldots, v_n) &\mapsto \varphi_1(v_1) \cdots \varphi_n(v_n) \end{aligned}$$

multilinear. Falls $V_1 = \cdots = V_n$ und $\varphi_1 = \cdots = \varphi_n$, dann ist diese Abbildung auch symmetrisch.

(8) Seien $v_1 \in V_1, \ldots, v_n \in V_n$. Dann ist die Abbildung

$$\begin{aligned} V_1^* \times \cdots \times V_n^* &\to K \\ (\varphi_1, \ldots, \varphi_n) &\mapsto \varphi_1(v_1) \cdots \varphi_n(v_n) \end{aligned}$$

multilinear. Falls $V_1 = \cdots = V_n$ und $v_1 = \cdots = v_n$, dann ist diese Abbildung auch symmetrisch.

(9) Allgemeiner ist die Abbildung

$$\begin{aligned} V_1 \times \cdots \times V_n \times V_1^* \times \cdots \times V_n^* &\to K \\ (v_1, \ldots, v_n, \varphi_1, \ldots, \varphi_n) &\mapsto \varphi_1(v_1) \cdots \varphi_n(v_n) \end{aligned}$$

multilinear.

11.3 Beschreibung durch „Darstellungsmatrizen“

Bis jetzt hatten wir eine gute Vorgehensweise, um lineare Abbildungen $T : V \to W$ oder bilineare Abbildungen $V \times V \to K$ zu verstehen: Zum Beispiel für $T : V \to W$ haben wir geordnete Basen $\mathcal{B} = (v_1, \dots, v_n) \subseteq V$ und $\mathcal{C} = (w_1, \dots, w_m) \subseteq W$ gewählt, $Tv_j = \sum_{i=1}^m a_{ij} w_i$ geschrieben und die Einträge a_{ij} in einer Matrix aufbewahrt. Ähnlich für eine Bilinearform $B : V \times V \to K$, dort haben wir eine Basis $\mathcal{B} = (v_1, \dots, v_n)$ von V gewählt und die Darstellungsmatrix als $M_{\mathcal{B}}(B) = (B(v_i, v_j))_{i,j}$ definiert. Wir haben auch Formeln entwickelt, wie man diese Matrizen im Fall eines Basiswechsels transformiert.

Für eine multilineare Abbildung $\Phi : V_1 \times \cdots \times V_k \to W$ können wir das Gleiche tun und sogar eine Verallgemeinerung der beiden oben genannten Abläufe (und Resultate) angeben. Trotzdem werden wir sehen, dass die „Matrizen“, die wir auf diese Weise erhalten, ziemlich mühsame Objekte sind, insbesondere wenn wir uns fragen, wie wir sie im Fall eines Basiswechsels transformieren müssen. Dies ist natürlich ein Problem. Eine Lösung ist, zu sagen, dass dies zu umständlich ist und dass wir es gar nicht erst versuchen wollen... Aber multilineare (und symmetrische/alternierende) Abbildungen tauchen auf natürliche Weise in so vielen Anwendungen in der Mathematik und Physik auf, dass wir sie und ihre zugehörigen Transformationseigenschaften irgendwie untersuchen müssen. Wir brauchen deshalb einen anderen Weg, um solche Abbildungen zu parametrisieren.

Bevor wir das tun, verallgemeinern wir den Begriff einer Darstellungsmatrix auf multilineare Abbildungen. Wir beschränken uns auf den Fall von endlich-dimensionalen Vektorräumen und überlassen es als Übung, den allgemeinen Fall (welcher sehr ähnlich ist) zu behandeln.

Proposition 11.3.1. *Seien $V_1, \dots, V_k, W$ endlich-dimensionale Vektorräume mit Basen $\mathcal{B}_\ell = \left(v_1^{(\ell)}, \dots, v_{n_\ell}^{(\ell)}\right) \subseteq V_\ell$ für alle $\ell = 1, \dots, k$ und $\mathcal{C} = (w_1, \dots, w_m) \subseteq W$. Sei*

$$\Phi : V_1 \times \cdots \times V_k \to W$$

eine multilineare Abbildung. Dann existieren Skalare $a^j_{i_1 \dots i_k} \in K$, wobei $j \in \{1, \dots, m\}$ und $1 \leq i_\ell \leq n_\ell$ für alle $\ell \in \{1, \dots, k\}$, sodass

$$\Phi\left(v_{i_1}^{(1)}, \dots, v_{i_k}^{(k)}\right) = \sum_{j=1}^m a^j_{i_1 \dots i_k} w_j. \tag{11.1}$$

Also ist Φ eindeutig bestimmt durch die Menge

$$\left\{a^j_{i_1 \dots i_k} \mid 1 \leq j \leq m,\ \forall \ell\ 1 \leq i_\ell \leq n_\ell\right\}. \tag{11.2}$$

Umgekehrt gibt uns jede Wahl von Skalaren wie in (11.2) *eine eindeutige multilineare Abbildung, welche* (11.1) *erfüllt.*

Bemerkung 11.3.2. Die Berechung der Skalare in (11.2) sollte als Verallgemeinerung der Darstellungsmatrix betrachtet werden und tatsächlich reduziert sich (11.2) zur Darstellungsmatrix von linearen Abbildungen (Definition 3.3.7) für $k = 1$ und zur Darstellungsmatrix von Bilinearformen (Definition 8.1.4) für $k = 2$, $V_1 = V_2$, $\mathcal{B}_1 = \mathcal{B}_2$ und $W = K$.

Die „Darstellungsmatrix“ im allgemeinen Fall ist ein Array von Elementen aus K von der Grösse $(n_1 \times \cdots \times n_k) \times m$, wobei $n_1, \ldots, n_k, m$ die Dimensionen der Vektorräume $V_1, \ldots, V_k, W$ sind. Die natürliche Darstellung wäre ein Gitter von Dimension $k+1$. Der Fall von linearen Abbildungen $V \to W$ entspricht $k = 1$, also ist die Darstellungsmatrix ein zweidimensionales Gitter, also eine Matrix (was wir ja bereits wussten). Im Fall einer Bilinearform $V \times V \to K$ haben wir $k = 2$, was eigentlich ein dreidimensionales Gitter ergibt. Die Seitenlängen sind $\dim V \times \dim V \times 1$, also haben wir keine Ausbreitung in der dritten Dimension, somit kann man eine solche Abbildung auch wieder als zweidimensional betrachten (was wir mit der Verwendung von $M_{\mathcal{B}}(B)$ auch getan haben). Im allgemeinen Fall müssten wir allerdings höherdimensionale Gitter betrachten. Multilineare Abbildungen $V \times V \times V \to K$ zum Beispiel benötigen ein dreidimensionales Gitter aus Skalaren für die „Darstellungsmatrix“.

Bemerkung 11.3.3. Die Leser sollten Proposition 11.3.1 mit Satz 3.1.15, Definition 8.1.4, Proposition 8.1.7 und Proposition 10.1.1 vergleichen.

Beweis von Proposition 11.3.1. Wir betrachten $\Phi : V_1 \times \cdots \times V_k \to W$ und schreiben eine beliebiges k-Tupel $(v_1, \ldots, v_k) \in V_1 \times \cdots \times V_k$ als

$$(v_1, \ldots, v_k) = \left(\sum_{s_1=1}^{n_1} \alpha_{s_1}^{(1)} v_{s_1}^{(1)}, \sum_{s_2=1}^{n_2} \alpha_{s_2}^{(2)} v_{s_2}^{(2)}, \ldots, \sum_{s_k=1}^{n_k} \alpha_{s_k}^{(k)} v_{s_k}^{(k)} \right).$$

Weil Φ multilinear ist, gilt

$$\begin{aligned} \Phi(v_1, \ldots, v_k) &= \Phi\left(\sum_{s_1=1}^{n_1} \alpha_{s_1}^{(1)} v_{s_1}^{(1)}, \sum_{s_2=1}^{n_2} \alpha_{s_2}^{(2)} v_{s_2}^{(2)}, \ldots, \sum_{s_k=1}^{n_k} \alpha_{s_k}^{(k)} v_{s_k}^{(k)} \right) \\ &= \sum_{s_1=1}^{n_1} \cdots \sum_{s_k=1}^{n_k} \alpha_{s_1}^{(1)} \cdots \alpha_{s_k}^{(k)} \Phi\left(v_{s_1}^{(1)}, \ldots, v_{s_k}^{(k)}\right), \end{aligned} \tag{11.3}$$

also genügt es, die Werte von $\Phi(v_{s_1}^{(1)}, \ldots, v_{s_k}^{(k)})$ zu kennen, um Φ eindeutig zu bestimmen. Ausserdem können wir, wie im Beweis von Satz 3.1.15, die Werte von

$$\Phi(v_{s_1}^{(1)}, \ldots, v_{s_k}^{(k)}) \in W$$

beliebig definieren für $1 \leq s_1 \leq n_1, \ldots, 1 \leq s_k \leq n_k$, und dann Φ für alle k-Tupel $(v_1, \ldots, v_k)$ via (11.3) definieren, um eine multilineare Abbildung $V_1 \times \cdots \times V_k \to W$ zu erhalten. Da die Bestimmung der Werte von $\Phi\left(v_{s_1}^{(1)}, \ldots, v_{s_k}^{(k)}\right)$ für $1 \leq s_1 \leq n_1, \ldots, 1 \leq s_k \leq n_k$ äquivalent ist zur Wahl der Skalare in (11.2), sind wir mit dem Beweis fertig. □

Korollar 11.3.4. *Wenn $V_1, \ldots, V_k, W$ endlich-dimensionale Vektorräume sind, dann gilt*

$$\dim \operatorname{Mult}(V_1 \times \cdots \times V_k, W) = \dim W \cdot \prod_{i=1}^{k} \dim V_i$$

Beweis. Die Abbildung

$$\Phi \mapsto \left\{a_{i_1 \ldots i_k}^{j} \mid 1 \leq \ell \leq k,\ 1 \leq i_\ell \leq n_\ell\ \forall \ell,\ 1 \leq j \leq m\right\}$$

ist ein linearer Isomorphismus zwischen $\operatorname{Mult}(V_1 \times \cdots \times V_k, W)$ und $\operatorname{Abb}(S, K)$, wobei S die Indexmenge

$$S = \{1, \ldots, n_1\} \times \cdots \times \{1, \ldots, n_k\} \times \{1, \ldots, m\}$$

ist. Somit ist die Dimension von $\dim \operatorname{Mult}(V_1 \times \cdots \times V_k, W)$ gleich

$$\dim \operatorname{Abb}(S, K) = |S| = m \cdot \prod_{i=1}^{k} n_i = \dim W \cdot \prod_{i=1}^{k} \dim V_i.$$ □

Übung 11.3.5. Verallgemeinern Sie die obigen Resultate zu allgemeinen Vektorräumen (mit nicht notwendigerweise endlicher Dimension). Hier ist eine Anleitung:

Sei $\mathcal{B}_\ell = (v_i^{(\ell)})_{i \in I_\ell}$ jeweils eine Basis von V_ℓ für $\ell = 1, \ldots, k$ und sei $\mathcal{C} = (w_j)_{j \in J}$ eine Basis von W. Betrachten Sie die Menge

$$\mathcal{D} = \left\{a_{i_1 \ldots i_k}^{j} \mid j \in J,\ i_1 \in I_1, \ldots, i_k \in I_k\right\},$$

welche die Eigenschaft hat, dass für alle $i_1 \in I_1, \ldots, i_k \in I_k$ die Menge

$$\mathcal{D}(i_1, \ldots, i_k) = \{j \in J \mid a_{i_1 \ldots i_k}^{j} \neq 0\} \subseteq J$$

endlich ist. Zeigen Sie, dass diese Menge von Skalaren durch

$$\Phi\left(v_{i_1}^{(1)}, \ldots, v_{i_k}^{(k)}\right) = \sum_{j \in \mathcal{D}(i_1, \ldots, i_k)} a_{i_1 \ldots i_k}^{j} w_j$$

eine multilineare Abbildung definiert. Folgern Sie, dass es eine lineare Bijektion gibt zwischen $\operatorname{Mult}(V_1 \times \cdots \times V_k, W)$ und dieser Menge von Skalaren.

Zeigen Sie als ein Korollar, dass

$$\dim \operatorname{Mult}(V_1 \times \cdots \times V_k, W) = \dim W \cdot \prod_{i=1}^{k} \dim(V_i^*)$$

(falls man auch 0-dimensionale Vektorräume zulassen möchte, gilt diese Formel weiterhin, mit der Konvention $0 \cdot \infty = 0$).

Nehmen wir nun an, dass wir die Basen $\mathcal{B}_i$ und $\mathcal{C}$ ändern möchten, oder dass wir verschiedene lineare und multilineare Abbildungen verketten möchten. Dabei fragen wir uns, wie der Array von Skalaren sich ändert (wir wollen also eine analoge Formel zu $[T]_{\mathcal{C}}^{\mathcal{B}}[S]_{\mathcal{B}}^{\mathcal{A}} = [T \circ S]_{\mathcal{C}}^{\mathcal{A}}$ in diesem Fall finden). Ich denke, dass die Leser dies schwierig finden werden, zumindest aus der Buchaltungs-Perspektive (das heisst, rein vom Organisieren der Informationen). Wie in der Einleitung bereits gesagt, werden wir dieses Problem später umgehen, indem wir statt multilineare Abbildungen direkt zu betrachten, alles auf lineare Abbildungen reduzieren, welche wir besser verstehen.

Bevor wir fortfahren, um ähnliche Resultate für symmetrische und alternierende Abbildungen anzugeben, sagen wir noch, was genau wir mit Verkettung meinen:

Übung 11.3.6. Seien $T_1 : V_1' \to V_1, \ldots, T_k : V_k' \to V_k, S : W \to W'$ lineare Abbildungen. Zeigen Sie, dass Präkomposition mit den T_i's und Postkomposition mit S eine lineare Abbildung

$$\begin{aligned} \operatorname{Mult}(V_1 \times \cdots \times V_k, W) &\to \operatorname{Mult}(V_1' \times \cdots \times V_k', W') \\ \Phi &\mapsto S \circ \Phi \circ (T_1 \times \cdots \times T_k) \end{aligned}$$

ergibt. Bemerken Sie, dass diese Abbildungen kontravariant in den V_i-Komponenten sind, das heisst, die Präkomposition mit $T : V_i' \to V_i$ induziert eine Abbildung

$$\operatorname{Mult}(V_1 \times \cdots \times V_k, W) \to \operatorname{Mult}(V_1 \times \cdots \times V_{i-1} \times V_i' \times V_{i+1} \times \cdots \times V_k, W)$$

und kovariant in der W-Komponente, das heisst, Postkomposition mit $S : W \to W'$ induziert eine Abbildung

$$\operatorname{Mult}(V_1 \times \cdots \times V_k, W) \to \operatorname{Mult}(V_1 \times \cdots \times V_k, W').$$

11.3.1 „Darstellungsmatrizen“ für alternierende und symmetrische Abbildungen

Seien V, W Vektorräume und $k \in \mathbb{N}$. Wir betrachten eine alternierende multilineare Abbildung $\Phi : V^k \to W$. Um das ganze etwas zu vereinfachen, nehmen wir an, dass $\operatorname{char}(K) \neq 2$. Gemäss Übung 11.2.6 (b) können wir also die folgende Definition für alternierend verwenden:

$$\forall w_1, \ldots, w_k \in V,\ \forall \sigma \in S_k \quad \Phi(w_{\sigma(1)}, \ldots, w_{\sigma(k)}) = \operatorname{sign}(\sigma)\Phi(w_1, \ldots, w_k).$$

Sei nun $\mathcal{B} = (v_1, \ldots, v_n)$ eine Basis von V. Somit können wir ein allgemeines k-Tupel in V^k schreiben als

$$\left(\sum_{j_1=1}^{n} \alpha_{j_1}^{(1)} v_{j_1}, \ldots, \sum_{j_k=1}^{n} \alpha_{j_k}^{(k)} v_{j_k}\right) \in V^k,$$

wobei $\alpha_{j_i} \in K$ für und wie zuvor sei $\Phi : V^k \to W$ eine alternierende multilineare Abbildung. Dann gilt

$$\Phi\left(\sum_{j_1=1}^{n} \alpha_{j_1}^{(1)} v_{j_1}, \ldots, \sum_{j_k=1}^{n} \alpha_{j_k}^{(k)} v_{j_k}\right) = \sum_{j_1=1}^{n} \cdots \sum_{j_k=1}^{n} \alpha_{j_1}^{(1)} \cdots \alpha_{j_k}^{(k)} \Phi(v_{j_1}, \ldots, v_{j_k}) = (*) \tag{11.4}$$

Für eine allgemeine multilineare Abbildung könnte man $\Phi(v_{j_1}, \ldots, v_{j_k})$ frei wählen, aber für eine alternierende multilineare Abbildung, haben wir weniger Freiheit:

Erstens, falls $v_{j_i} = v_{j_\ell}$ für $i \neq \ell$, dann muss $\Phi(v_{j_1}, \ldots, v_{j_k}) = 0_W$ sein. Ausserdem gilt

$$\Phi(v_{\sigma(j_1)}, \ldots, v_{\sigma(j_k)}) = \operatorname{sign}(\sigma)\Phi(v_{j_1}, \ldots, v_{j_k})$$

für alle $\sigma \in S_k$, also genügt es, $\Phi(v_{j_1}, \ldots, v_{j_k})$ für k-Tupel bis auf Permutationen zu kennen. Nach einiger Überlegung führt dies zur folgenden Fortsetzung von (11.4), wo wir nur über strikt aufsteigende Ketten von Indizes $1 \leq j_1 < \cdots < j_k \leq n$ summieren:

$$(*) = \sum_{1 \leq j_1 < j_2 < \cdots < j_k \leq n} \left(\sum_{\sigma \in S_k} \operatorname{sign}(\sigma) \alpha_{j_{\sigma(1)}}^{(\sigma(1))} \cdots \alpha_{j_{\sigma(k)}}^{(\sigma(k))}\right) \Phi(v_{j_1}, \ldots, v_{j_k}) \tag{11.5}$$

Man kann überprüfen, dass man $\Phi(v_{j_1}, \ldots, v_{j_k})$ für alle strikt aufsteigenden Ketten $1 \leq j_1 < j_2 < \cdots < j_k \leq n$ frei wählen kann und dass die Definition von Φ durch (11.4) und (11.5) eine eindeutige alternierende Abbildung $\Phi \in \operatorname{Alt}^k(V, W)$ bestimmt.

Korollar 11.3.7. *Seien V, W endlich-dimensionale Vektorräume und sei $n = \dim V$. Dann ist*

$$\begin{aligned}\dim \operatorname{Alt}^k(V, W) &= \left|\{(j_1, \ldots, j_k) \mid 1 \leq j_1 < \cdots < j_k \leq n\}\right| \cdot \dim W \\ &= \begin{cases} \binom{n}{k} \cdot \dim W, & k \leq n, \\ 0 \cdot \dim W = 0, & k > n. \end{cases}\end{aligned}$$

Beweis. Dies ist sehr ähnlich zum Beweis von Korollar 11.3.4. □

Beispiel 11.3.8. Seien $V = K^3$ und $W = K$ mit den Standard-Basen $\mathcal{E}_3 = (e_1, e_2, e_3)$ und $\mathcal{E}_1 = (e_1)$. Wir möchten $\operatorname{Alt}^k(V, W)$ für verschiedene $k \in \mathbb{N}$ untersuchen. Zuerst

$k = 1$: Eine Abbildung $\Phi \in \mathrm{Alt}^1(K^3, K)$ ist eindeutig bestimmt durch $\Phi(e_1)$, $\Phi(e_2)$ und $\Phi(e_3)$, genau wie für lineare Abbildungen $\Phi \in \mathrm{Hom}(K^3, K)$! Dies ist keine Überraschung, denn $\mathrm{Alt}^1(V, W) = \mathrm{Hom}(V, W)$. Somit ist $\dim \mathrm{Alt}^1(V, W) = \dim \mathrm{Hom}(K^3, K) = 3$.

Betrachten wir $\mathrm{Alt}^2(K^3, K)$. Eine Abbildung $\Phi \in \mathrm{Alt}^2(K^3, K)$ ist eindeutig bestimmt durch $\Phi(e_1, e_2)$, $\Phi(e_1, e_3)$ und $\Phi(e_2, e_3)$. Wir überlassen es den Lesern, die folgende Rechnung zu vervollständigen:

$$\begin{aligned}\Phi\left(\begin{pmatrix}a_1\\a_2\\a_3\end{pmatrix}, \begin{pmatrix}b_1\\b_2\\b_3\end{pmatrix}\right) &= \Phi(a_1e_1 + a_2e_2 + a_3e_3, b_1e_1 + b_2e_2 + b_3e_3)\\ &= \cdots\\ &= (a_1b_2 - a_2b_1)\Phi(e_1, e_2) + (a_1b_3 - a_3b_1)\Phi(e_1, e_3)\\ &\quad + (a_2b_3 - a_3b_2)\Phi(e_2, e_3).\end{aligned}$$

Also ist $\dim \mathrm{Alt}^2(K^3, K) = 3$.

Eine Abbildung $\Phi \in \mathrm{Alt}^3(K^3, K)$ ist eindeutig durch $\Phi(e_1, e_2, e_3)$ bestimmt. Wie zuvor ist es eine gute Übung, die folgende Rechnung zu vervollständigen und dabei die 3×3-Determinante zu rekonstruieren:

$$\Phi\left(\sum_{j=1}^{3} a_{1j}e_j, \sum_{j=1}^{3} a_{2j}e_j, \sum_{j=1}^{3} a_{3j}e_j\right) = \cdots = \det\begin{pmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{pmatrix}\Phi(e_1, e_2, e_3).$$

Dies erklärt, wieso es eine eindeutige alternierende Abbildung $\Phi : K^3 \to K$ gibt bis auf einen Skalar (die Wahl[1] von $\Phi(e_1, e_2, e_3)$) und dies ist genau die Determinantenabbildung. Dies ist wieder keine Überraschung, mit dem fast genau gleichen Beweis wie in Kapitel 4 kann man zeigen, dass es eine bis auf einen Skalar eindeutige alternierende multilineare Abbildung in $\mathrm{Alt}^n(K^n, K)$ gibt. Die gleiche Berechnung zeigt auch, dass es eine bis auf einen Skalar eindeutige alternierende multilineare Abbildung in $\mathrm{Alt}^n(V, K)$ gibt, wenn $n = \dim V$.

Übung 11.3.9. Machen Sie eine ähnliche Analyse für Abbildungen $\Phi \in \mathrm{Sym}^k(V, W)$ und zeigen Sie für V mit Basis $(v_1, \ldots, v_n)$ die folgenden Aussagen:

(a) Für jede Wahl von

$$\{\Phi(v_{j_1}, \ldots, v_{j_k}) \in W \mid 1 \leq j_1 \leq \cdots \leq j_k \leq n\}$$

existiert eine eindeutige Abbildung $\Phi \in \mathrm{Sym}^k(V, W)$.

[1] Die Wahl, die wir bei der Definition der Determinanten im Kapitel 4 getroffen haben, war $\Phi(e_1, e_2, e_3) = 1$. Dies enstpricht der Tatsache, dass die Determinante der Einheitsmatrix 1 ist.

(b) $\dim \mathrm{Sym}^k(V,W) = \binom{n+k-1}{k} \cdot \dim W$.

<u>Hinweis:</u> Wie immer empfehlen wir, zuerst die Fälle mit kleiner Dimension zu betrachten.

Übung 11.3.10. Zeigen Sie die analoge Aussage von Übung 11.3.6: Seien $T : V' \to V$ und $S : W \to W'$ lineare Abbildungen. Die Präkomposition mit $T^k = T \times \cdots \times T$ und Postkomposition mit S induziert lineare Abbildungen

$$\begin{aligned} \mathrm{Alt}^k(V,W) &\to \mathrm{Alt}^k(V',W') \\ \Phi &\mapsto S \circ \Phi \circ T^k \end{aligned}$$

und

$$\begin{aligned} \mathrm{Sym}^k(V,W) &\to \mathrm{Sym}^k(V',W') \\ \Phi &\mapsto S \circ \Phi \circ T^k. \end{aligned}$$

11.4 Das Tensorprodukt - Parametrisierung von bilinearen Abbildungen

Wie wir in der Einleitung von Abschnitt 11.1 und später auf Seite 417 gesagt haben, um $\mathrm{Mult}(U \times V, W)$ zu verstehen[2], welches der Raum aller bilinearen Abbildungen $U \times V \to W$ ist, konstruieren wir einen Vektorraum, welchen wir mit $U \otimes V$ bezeichnen, zusammen mit einer bilinearen Abbildung $\iota : U \times V \to U \otimes V$, welche zusammen mit mit linearen Abbildungen $U \otimes V \to W$ alle bilinearen Abbildungen erzeugt, in dem Sinne dass das Diagramm

$$\begin{tikzcd} U \times V \arrow[rr, "\forall \varphi"] \arrow[dr, "\iota"'] & & W \\ & U \otimes V \arrow[ur, dashed, "\exists! \psi"'] & \end{tikzcd}$$

kommutiert.

11.4.1 $K^m \otimes K^n$

Wir beginnen mit dem Fall $U = K^m = K^m_{\text{Spal}}$ und $V = K^n = K^n_{\text{Spal}}$. Wir betrachten die Abbildung

$$\begin{aligned} \iota : K^m \times K^n &\to M_{m\times n}(K) \\ (u,v) &\mapsto u \otimes v := uv^T, \end{aligned}$$

wobei $u \otimes v$ vorläufig nur eine Notation für die $m \times n$-Matrix $\iota(u,v) = uv^T$ ist. Diese Matrix wird auch das *Kronecker-Produkt* von u und v genannt. Die Abbildung ι wird mit dieser Notation versteckt: $u \otimes v$ ist per Definition $\iota(u,v)$.

[2]Oder allgemeiner $\mathrm{Mult}(V_1 \times \cdots \times V_k, W)$, $\mathrm{Alt}^k(V,W)$ oder $\mathrm{Sym}^k(V,W)$.

Bemerken Sie, dass die Abbildung ι bilinear ist: Für alle $\alpha \in K$, $u_1, u_2, u \in U$ und $v_1, v_2, v \in V$ gilt

$$(u_1 + u_2) \otimes v = u_1 \otimes v + u_2 \otimes v, \qquad (\alpha u) \otimes v = \alpha(u \otimes v), \tag{11.6}$$
$$u \otimes (v_1 + v_2) = u \otimes v_1 + u \otimes v_2, \qquad u \otimes (\alpha v) = \alpha(u \otimes v). \tag{11.7}$$

Für jede bilineare Abbildung $\Phi : U \times V \to W$ definieren wir

$$w_{ij} := \Phi(e_i, e_j), \quad \text{für } 1 \leq i \leq m,\ 1 \leq j \leq n.$$

Erinnern Sie sich, dass $M_{m\times n}(K)$ als Vektorraum die Standardbasis

$$\{E_{ij} \mid 1 \leq i \leq m,\ 1 \leq j \leq n\}$$

hat, wobei E_{ij} die $m\times n$-Matrix ist mit 1 im ij-Eintrag und 0 in allen anderen Einträgen. Bemerken Sie auch, dass E_{ij} das Kroneckerprodukt von e_i und e_j ist, also gilt mit der zuvor eingeführten Notation $E_{ij} = e_i \otimes e_j$. Wir definieren die lineare Abbildung

$$\varphi : M_{m\times n}(K) \to W$$

als die eindeutige lineare Abbildung mit $\varphi(e_i \otimes e_j) = w_{ij}$ für alle i, j. Wir behaupten, dass das Diagramm

$$\begin{tikzcd}
U \times V \arrow[rr, "\Phi"] \arrow[dr, "\iota"'] & & W \\
& U \otimes V \arrow[ur, dashed, "\varphi"'] &
\end{tikzcd}$$

kommutiert, also dass $\varphi \circ \iota = \Phi$ als Abbildungen. Wir können dies entweder direkt überprüfen, oder uns erinnern, dass weil ι bilinear ist und φ linear, $\varphi \circ \iota$ bilinear ist (sehen Sie zum Beispiel Übung 11.3.6). Deshalb genügt es, zu zeigen, dass die beiden Abbildungen auf einer Basis übereinstimmen, also

$$\varphi \circ \iota(e_i, e_j) = \Phi(e_i, e_j)$$

für alle $1 \leq i \leq m$ und $1 \leq j \leq n$. Dies folgt unmittelbar aus unseren Definitionen:

$$\varphi \circ \iota(e_i, e_j) = \varphi(e_i \otimes e_j) = \varphi(E_{ij}) = w_{ij} = \Phi(e_i, e_j).$$

Wir haben also folgendes gezeigt:

Für jede bilineare Abbildung $\Phi : K^m \times K^n \to W$ existiert eine eindeutige *lineare* Abbildung $\varphi : M_{m\times n}(K) \to W$ mit $\varphi \circ \iota = \Phi$, wobei $\iota : K^m \times K^n \to M_{m\times n}(K)$ das Kronecker-Produkt ist. Um es mit der Sprache und Notation auszudrücken, die wir in diesem Kapitel entwickeln werden: Wir haben gezeigt, dass der Vektorraum $M_{m\times n}(K)$, zusammen mit der Abbildung $\iota : K^m \times K^n \to M_{m\times n}(K)$, das Tensorprodukt von K^m und K^n ist[3]; in Symbolen $M_{m\times n}(K) \cong K^m \otimes K^n$ (die Abbildung ι ist in dieser Notation nur implizit vorhanden).

11.5 Abstrakte Definition und Konstruktion des Tensorprodukts

Das Tensorprodukt ist das erste Beispiel eines Objekts, dessen Konstruktion im Grunde irrelevant ist für uns, denn nachdem wir uns davon überzeugt haben, dass es existiert, werden wir nur noch die charakterisierende Eigenschaft verwenden.

Theorem 11.5.1. *Seien U und V K-Vektorräume. Es existiert ein K-Vektorraum $U \otimes_K V$ und eine bilineare Abbildung*

$$\begin{aligned} \iota : U \times V &\to U \otimes_K V \\ (u, v) &\mapsto u \otimes v, \end{aligned}$$

sodass für jede bilineare Abbildung $\Phi\colon U \times V \to W$, wobei W ein K-Vektorraum ist, eine eindeutige lineare Abbildung $\varphi\colon U \otimes_K V \to W$ existiert, sodass das Diagramm

$$\begin{tikzcd} U \times V \arrow[rr, "\Phi"] \arrow[dr, "\iota"'] & & W \\ & U \otimes V \arrow[ur, dashed, "\exists!\varphi"'] & \end{tikzcd}$$

kommutiert, das heisst $\Phi = \varphi \circ \iota$. Präziser gesagt gilt, dass für alle K-Vektorräume W die Verkettungsabbildung

$$\begin{aligned} \mathrm{Hom}_K(U \otimes_K V, W) &\to \mathrm{Mult}_K(U \times V, W) \\ \varphi &\mapsto \varphi \circ \iota \end{aligned}$$

ein linearer Isomorphismus ist. Ausserdem gilt $U \otimes_K V = \mathrm{Sp}(\{u \otimes v \mid u \in U,\ v \in V\})$.

Notation. Wenn der Körper K aus dem Kontext klar ist, schreiben wir manchmal nur $U \otimes V$ statt $U \otimes_K V$.

Bemerkung 11.5.2. Theorem 11.5.1 behauptet die Existenz des Tensorprodukts und nennt die charakterisierende Eigenschaft. Wir werden später die folgende Aussage zeigen: „Das Tensorprodukt ist eindeutig bis auf einen eindeutigen Isomorphismus“.

[3] Genauer gesagt isomorph zum Tensorprodukt von K^m und K^n. Das Tensorprodukt ist nämlich nur bis auf einen eindeutigen Isomorphismus definiert, wie wir sehen werden (wir werden dann auch diesen kryptischen letzten Satz erklären).

Die abstrakte Konstruktion des Tensorprodukts folgt aus einer einfachen Idee: Wir definieren einen Vektorraum, dessen Elemente wie $u \otimes v$ aussehen und den gleichen Regeln folgen, die wir im konkreten Beispiel des Tensorprodukts $K^n \otimes K^m$ in Abschnitt 11.4.1 gesehen haben (sehen Sie Gleichung (11.6)). Um dies zu zeigen und um Theorem 11.5.1 zu beweisen, brauchen wir die folgende Konstruktion:

11.5.1 Freie Vektorräume über einer Menge

Wir fixieren einen Körper K und eine Menge S. Wir definieren den *freien Vektorraum* $F_K(S)$ (oder nur $F(S)$, wenn K aus dem Kontext klar ist) als Raum aller formalen Linearkombinationen von Elementen aus S:

$$F_K(S) = \left\{ \sum_{s \in S} \alpha_s \cdot s \mid \alpha_s \in K \ \forall s \in S \text{ und } \alpha_s \neq 0 \text{ nur für endlich viele } s \in S \right\}.$$

Dieser Vektorraum ist mit der folgenden Addition und Skalarmultiplikation versehen:

$$\sum_{s \in S} \alpha_s \cdot s + \sum_{s \in S} \beta_s \cdot s := \sum_{s \in S} (\alpha_s + \beta_s) \cdot s,$$
$$\lambda \left(\sum_{s \in S} \alpha_s \cdot s \right) := \sum_{s \in S} (\lambda \alpha_s) \cdot s.$$

Der Nullvektor ist also $\sum_{s \in S} 0_K \cdot s$. Normalerweise schreiben wir nur diejenigen Elemente von S, deren Koeffizienten $\neq 0$ sind. Ausserdem schreiben wir $-\alpha_s\, s$ für $+(-\alpha_s)s$.

Beispiel 11.5.3. Sei $S = \{\diamondsuit, \heartsuit, \spadesuit, \clubsuit\}$ und $K = \mathbb{Q}$. Die Elemente von $F(S)$ sind zum Beispiel Summen von der Form

$$\begin{aligned} v &= 2\diamondsuit + 3\heartsuit, \\ w &= \frac{3}{2}\clubsuit, \\ u &= 5\heartsuit - 4\clubsuit := 5\heartsuit + (-4)\clubsuit. \end{aligned}$$

Es gilt dann $\frac{1}{2}v = \diamondsuit + \frac{3}{2}\heartsuit$, $u + v = 8\heartsuit + 2\diamondsuit - 4\clubsuit$, usw. Man kann sehen, dass $(\diamondsuit, \heartsuit, \spadesuit, \clubsuit)$ eine Basis von $F(S)$ ist, also ist $F(S)$ isomorph zu $\mathbb{Q}^4$. Ein „standard“ Isomorphismus wird $(\diamondsuit, \heartsuit, \spadesuit, \clubsuit)$ mit (e_1, e_2, e_3, e_4) identifizieren.

Beispiel 11.5.4. Ähnlich wie im vorherigen Beispiel, gilt allgemein: Falls $|S| = n$, dann ist $F_K(S) \cong K^n$.

Beispiel 11.5.5. Eigentlich haben wir diese Konstruktion schon einmal gesehen: Ein anderes Modell für den freien Vektorraum ist der Raum $K^{(S)}$, welchen wir in Beispiel

10.1.4 definiert haben. Hier ist ein Isomorphismus zwischen diesen beiden Vektorräumen:

$$\begin{aligned} F_K(S) &\to K^{(S)} \\ \textstyle\sum_{s\in S} \alpha_s \cdot s &\mapsto \begin{bmatrix} f: S \to K \\ \quad s \mapsto \alpha_s \end{bmatrix} \end{aligned}$$

Insbesondere heisst dies, dass wir die Elemente von S als eine Basis von $F_K(S)$ betrachten können (vergleichen Sie dies mit der Basis von $K^{(S)}$ aus Übung 10.1.5).

Wir möchten gerne mit den Elementen der Menge S spielen (das heisst, sie addieren oder mit Skalaren multiplizieren), deswegen bevorzugen wir das Modell $F_K(S)$ gegenüber dem Modell $K^{(S)}$. Wir haben bereits gesehen, dass $\dim K^{(S)} = |S|$, also folgt $\dim F_K(S) = |S|$.

Beispiel 11.5.6. Das Beispiel, welches wir für die Konstruktion des Tensorprodukts von zwei Vektorräumen U und V brauchen werden, ist $F(U \times V)$.

Wenn zum Beispiel $U = \mathbb{R}^2$, $V = \mathbb{R}^3$ und $K = \mathbb{R}$ ist, dann ist dies ein „riesiger“ Vektorraum; die Dimension ist gleich der Kardinalität von $\mathbb{R}^2 \times \mathbb{R}^3$. Ein Beispiel eines Elements in $F(\mathbb{R}^2 \times \mathbb{R}^3)$ ist

$$\pi \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right) + 7.3 \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right) + 2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right)$$

und so weiter. Bemerken Sie, dass zum Beispiel $(0_U, 0_V) \in F(U \times V)$ ein Vektor wie jeder andere ist und nicht etwa der Nullvektor von $F(U \times V)$, welcher die leere Linearkombination ist.

Bemerken Sie, dass es zu $F_K(S)$ eine natürliche injektive Abbildung

$$\begin{aligned} b_S : S &\to F_K(S) \\ s &\mapsto s = 1_K \cdot s \end{aligned}$$

gibt. Die Abbildung b_S ist die „allgemeinste Abbildung von S in einen K-Vektorraum“ im folgenden Sinn:

Satz 11.5.7. *Der Vektorraum $F(S)$ zusammen mit der Abbildung $b_S : S \to F_K(S)$ erfüllt die folgende charakterisierende Eigenschaft (auch* universelle Eigenschaft *genannt):*

Für jeden Vektorraum W und jede Funktion $f: S \to W$ existiert eine eindeutige lineare *Abbildung $g : F(S) \to W$, sodass das Diagramm*

$$\begin{tikzcd} S \arrow[rr, "f"] \arrow[dr, "b_S"'] & & W \\ & F(S) \arrow[ur, dashed, "\exists! g"'] & \end{tikzcd}$$

kommutiert, das heisst $f = g \circ b_S$.

Beweis. Damit $f = g \circ b_S$ gelten kann, muss $g(1_K \cdot s) = f(s)$ für alle $s \in S$ gelten. Damit g linear sein kann, muss g folglich durch

$$\begin{aligned} g: \qquad F(S) &\to W \\ \textstyle\sum_{s\in S} \alpha_s \cdot s &\mapsto \textstyle\sum_{s\in S} \alpha_s f(s) \end{aligned}$$

gegeben sein, was die Existenz und Eindeutigkeit von g zeigt. Wir überlassen es den Lesern, zu zeigen, dass g in der Tat linear ist. □

11.5.2 Beweis von Theorem 11.5.1

Erinnern Sie sich, was unser Ziel ist: Für zwei Vektorräume U, V über K wollen wir einen Vektorraum konstruieren, welcher von Symbolen von der Form $u \otimes v$, für $u \in U$, $v \in V$, erzeugt wird, welche die folgenden Bedingungen erfüllen:

Für alle $\alpha \in K$, $u_1, u_2, u \in U$ und $v_1, v_2, v \in V$ gilt

$$\begin{aligned} (u_1 + u_2) \otimes v &= u_1 \otimes v + u_2 \otimes v, & (\alpha u) \otimes v &= \alpha(u \otimes v), \\ u \otimes (v_1 + v_2) &= u \otimes v_1 + u \otimes v_2, & u \otimes (\alpha v) &= \alpha(u \otimes v). \end{aligned}$$

Wieso definieren wir dies nicht einfach? Genauer gesagt: Wir betrachten $F(U \times V)$ und „dividieren“ durch alle Relationen, welche wir haben möchten. Wir schreiben

$$R = \mathrm{Sp}\begin{pmatrix} \{(u_1 + u_2, v) - (u_1, v) - (u_2, v) \mid u_1, u_2 \in U,\ v \in V\} \\ \cup\{(u, v_1 + v_2) - (u, v_1) - (u, v_2) \mid v_1, v_2 \in V,\ u \in U\} \\ \cup\{(\alpha u, v) - \alpha(u, v) \mid \alpha \in K,\ u \in U,\ v \in V\} \\ \cup\{(u, \alpha v) - \alpha(u, v) \mid \alpha \in K,\ u \in U,\ v \in V\} \end{pmatrix} \subseteq F(U \times V) \quad (11.8)$$

und definieren das Tensorprodukt von U und V als den Quotientenvektorraum

$$U \otimes V = U \otimes_K V := F_K(U \times V)/R.$$

Ausserdem definieren wir für $u \in U$ und $v \in V$

$$u \otimes v := \pi((u, v)) \in U \otimes V,$$

wobei $\pi\colon F(U \times V) \to F(U \times V)/R$ die kanonische Quotientenabbildung ist. Es folgt aus der Definition des Quotientenvektorraums, dass

$$U \otimes V = \mathrm{Sp}(\{u \otimes v \mid u \in U,\ v \in V\}).$$

Dies zeigt die letzte Aussage von Theorem 11.5.1, sobald wir die restlichen Aussagen überprüft haben.

Sei $b = b_{U\times V}$ die natürliche Abbildung $b_{U\times V}\colon U \times V \to F(U \times V)$ und wir definieren $\iota\colon U \times V \to U \otimes V$ durch $\iota := \pi \circ b_{U\times V}$. Es bleibt noch zu zeigen, dass $U \otimes V$ die charakterisierende Eigenschaft des Tensorprodukts besitzt. Betrachten wir dafür eine multilineare Abbildung $\Phi \in \mathrm{Mult}(U \times V, W)$, wobei W ein K-Vektorraum ist. Da Φ insbesondere eine Abbildung von $U \times V$ in den Vektorraum W ist, existiert eine eindeutige lineare Abbildung $g\colon F(U \times V) \to W$, sodass das Diagramm

$$\begin{tikzcd}
U \times V \arrow[rr, "\Phi"] \arrow[dr, "b_{U\times V}"'] & & W \\
& F(U \times V) \arrow[ur, dashed, "\exists! g"'] &
\end{tikzcd}$$

kommutiert. Da Φ bilinear ist (multilineare Abbildungen $U \times V \to W$ sind das Gleiche wie bilineare Abbildungen), folgt $R \subseteq \ker g$: Zum Beispiel ist für $u \in U$, $v \in V$ und $\alpha \in K$

$$g((\alpha u, v)) = \Phi(\alpha u, v) = \alpha\Phi(u, v) = \alpha g((u, v)) \overset{(*)}{=} g(\alpha(u, v))$$

und somit

$$g((\alpha u, v) - \alpha(u, v)) \overset{(*)}{=} g((\alpha u, v)) - g(\alpha(u, v)) = 0,$$

wobei wir jeweils in $(*)$ die Linearität von g verwendet haben[4]. Für die anderen erzeugenden Elemente von R lässt sich ähnlich überprüfen, dass sie im Kern von g liegen. Folglich lässt sich die Abbildung g durch den Quotientenvektorraum $U \otimes V = F(U \times V)/R$ faktorisieren, das heisst, es existiert eine eindeutige lineare Abbildung $\overline{g}\colon U \otimes V \to W$, sodass das Diagramm

$$\begin{tikzcd}
U \times V \arrow[rrr, "\Phi"] \arrow[dr, "b"] \arrow[ddrr, bend right, "\iota"'] & & & W \\
& F(U \times V) \arrow[urr, "g"'] \arrow[dr, "\pi"] & & \\
& & U \otimes V \arrow[uur, dashed, "\overline{g}"'] &
\end{tikzcd}$$

kommutiert, wobei $\overline{g}$ eindeutig definiert ist, da g eindeutig definiert war. Dies beendet den Beweis von Theorem 11.5.1.

[4]Implizit haben wir, da $b = b_{U\times V}$ injektiv ist, $(u, v) \in U \times V$ und $1_K \cdot (u, v) \in F(U \times V)$ miteinander identifiziert und beide mit (u, v) bezeichnet.

11.6 Grundlegende Eigenschaften des Tensorprodukts

Ab jetzt vergessen wir die technische Konstruktion des Tensorprodukts wieder und verwenden nur noch die charakterisierende Eigenschaft (wie in Theorem 11.5.1 formuliert). Beweisen wir nun einige Fakten über $U \otimes V$.

Korollar 11.6.1. *Für alle $u \in U$, $v \in V$ gilt*

$$u \neq 0 \text{ und } v \neq 0 \quad \Longrightarrow \quad u \otimes v \neq 0.$$

Beweis. Die Strategie ist, einen Vektorraum W und $\Phi \in \mathrm{Mult}(U \times V, W)$ zu finden, sodass $\Phi(u, v) \neq 0$. Da eine solche Abbildung als $\Phi = \varphi \circ \iota$ faktorisiert werden kann, wobei $\varphi \colon U \otimes V \to W$ linear ist, folgt

$$0 \neq \Phi(u, v) = \varphi(u \otimes v)$$

und somit ist $u \otimes v \neq 0$. Um eine solche multilineare Abbildung Φ zu finden, betrachten wir $W = K$ und benutzen Beispiel 11.2.8 (7):

Wir wählen $f \in U^*$ und $g \in V^*$ mit $f(u) \neq 0$ und $g(v) \neq 0$ und definieren

$$\begin{aligned} \Phi : U \times V &\to K \\ (x, y) &\mapsto f(x)g(y). \end{aligned}$$

Somit ist $\Phi(u, v) \neq 0$ und $\Phi \in \mathrm{Mult}(U \times V, K)$ wie gewünscht. □

Übung 11.6.2. Zeigen Sie, dass falls $u = 0$ oder $v = 0$, dann $u \otimes v = 0$.

Korollar 11.6.3. *Falls* $\dim U, \dim V < \infty$, *dann ist* $\dim(U \otimes V) = \dim U \cdot \dim V$.

Beweis. Wir wenden Theorem 11.5.1 mit $W = K$ an. Es folgt, dass

$$(U \otimes V)^* = \mathrm{Hom}(U \otimes V, K) \cong \mathrm{Mult}(U \times V, K).$$

Es folgt

$$\dim(U \otimes V)^* = \dim \mathrm{Mult}(U \times V, K) \overset{\text{Kor.}}{\underset{11.3.4}{=}} \dim U \cdot \dim V,$$

also ist $(U \otimes V)^*$ endlich-dimensional. Laut den Resultaten aus Abschnitt 10.1 muss $U \otimes V$ auch endlich-dimensional sein, mit der gleichen Dimension wie $(U \otimes V)^*$, also $\dim U \cdot \dim V$. □

Korollar 11.6.4. *Falls $\mathcal{B}_U = (u_1, \ldots, u_m) \subseteq U$ und $\mathcal{B}_V = (v_1, \ldots, v_n) \subseteq V$ jeweils Basen sind, dann ist*

$$\mathcal{B}_U \otimes \mathcal{B}_V = \{u_i \otimes v_j \mid 1 \leq i \leq m,\ 1 \leq j \leq n\}$$

eine Basis von $U \otimes V$.

Beweis. Für $u = \sum_{i=1}^m \alpha_i u_i$ und $v = \sum_{j=1}^n \beta_j v_j$ gilt

$$u \otimes v = \left(\sum_{i=1}^m \alpha_i u_i\right) \otimes \left(\sum_{j=1}^n \beta_j v_j\right) = \sum_{i=1}^m \sum_{j=1}^n \alpha_i \beta_j u_i \otimes v_j,$$

also ist $\{u \otimes v \mid u \in U,\ v \in V\} \subseteq \mathrm{Sp}(\mathcal{B}_U \otimes \mathcal{B}_V)$ und laut der zweiten Aussage in Theorem 11.5.1 folgt $U \otimes V = \mathrm{Sp}(\mathcal{B}_U \otimes \mathcal{B}_V)$. Da

$$|\mathcal{B}_U \otimes \mathcal{B}_V| = m \cdot n \underset{11.6.3}{\overset{\text{Kor.}}{=}} \dim(U \otimes V),$$

folgt das Korollar. □

Übung 11.6.5. Verallgemeinern Sie dies zu unendlich-dimensionalen Vektorräumen: Falls $\mathcal{B}_U = (u_i)_{i \in I} \subseteq U$ und $\mathcal{B}_V = (v_j)_{j \in J} \subseteq V$ Basen sind, dann ist

$$\mathcal{B}_U \otimes \mathcal{B}_V = \{u_i \otimes v_j \mid i \in I,\ j \in J\}$$

eine Basis von $U \otimes V$.
Hinweis: Es benötigt hier ein anderes Argument als im endlich-dimensionalen Fall. Gegeben eine Linearkombination von verschiedenen $u_i \otimes v_j$, welche gleich Null ist, betrachten Sie eine duale Menge in U^* und V^*, um Abbildungen in $\mathrm{Mult}(U \times V, K)$ zu definieren, welche zeigen, dass alle Koeffizienten der Linearkombination gleich Null sein müssen.

Beispiel 11.6.6. Wir hatten bereits direkt gezeigt, dass $K^m \otimes K^n \cong M_{m \times n}(K)$. Das letzte Korollar gibt einen alternativen Beweis dafür: Wenn wir $e_i \otimes e_j$ mit der Matrix E_{ij} identifizieren, erhalten wir den Isomorphismus $K^m \otimes K^n \cong M_{m \times n}(K)$.

Beispiel 11.6.7. [Reine Tensoren] Es ist wichtig, zu bemerken, dass nicht jedes Element von $U \otimes V$ von der Form $u \otimes v$ ist. Elemente von der Form $u \otimes v$ nennen wir *reine Tensoren*. Versuchen wir nun, alle reinen Tensoren in $K^2 \otimes K^2$ zu finden:

Von den Standard Basen erhalten wir die Basis

$$\mathcal{B} = (e_1 \otimes e_1, e_1 \otimes e_2, e_2 \otimes e_1, e_2 \otimes e_2)$$

von $K^2 \otimes K^2$. Ein reiner Tensor hat die Form

$$\begin{pmatrix} a \\ b \end{pmatrix} \otimes \begin{pmatrix} c \\ d \end{pmatrix} = (ae_1 + be_2) \otimes (ce_1 + de_2) = ace_1 \otimes e_1 + ade_1 \otimes e_2 + bce_2 \otimes e_1 + bde_2 \otimes e_2.$$

Man könnte damit zeigen: $Ae_1 \otimes e_1 + Be_1 \otimes e_2 + Ce_2 \otimes e_1 + De_2 \otimes e_2$ ist ein reiner Tensor genau dann, wenn $AD = BC$.

Die nächste Aussage ist ein weiteres Korollar, aber da es eine sehr wichtige Aussage ist, nennen wir sie Proposition.

Proposition 11.6.8. *Seien $T : U \to U'$ und $S : V \to V'$ lineare Abbildungen. Dann existiert eine eindeutige lineare Abbildung, welche wir mit $T \otimes S$ bezeichnen,*

$$T \otimes S \colon U \otimes V \to U' \otimes V',$$

sodass für alle $u \in U, v \in V$ gilt $T \otimes S(u \otimes v) = T(u) \otimes S(v)$.

Beweis. Wir erinnern uns noch einmal an unsere Strategie: Zuerst möchten wir eine Abbildung von $U \otimes V$ nach $U' \otimes V'$ finden. Laut Theorem 11.5.1 ist das äquivalent dazu, eine bilineare Abbildung von $U \times V$ nach $U' \otimes V'$ zu finden. Das ist nicht allzu schwierig: Da die Abbildung

$$\iota = \iota_{U' \times V'} \colon U' \times V' \to U' \otimes V'$$

bilinear ist, können wir sie mit $T \times S$ präkomponieren (siehe Übung 11.3.6), um eine bilineare Abbildung

$$U \times V \xrightarrow{T \times S} U' \times V' \xrightarrow{\iota_{U' \times V'}} U' \otimes V'$$

zu erhalten. Schreiben wir nun diese Abbildung explitzit:

$$\begin{aligned} \Phi : U \times V &\to U' \otimes V' \\ (u, v) &\mapsto T(u) \otimes S(v). \end{aligned}$$

Gemäss Theorem 11.5.1 existiert eine eindeutige lineare Abbildung

$$\varphi : U \otimes V \to U' \otimes V'$$

mit $\Phi = \varphi \circ \iota_{U \times V}$. Also ist φ genau die Abbildung, die wir gesucht haben: Wenn wir $T \otimes S := \varphi$ schreiben, dann bedeutet $\Phi = \varphi \circ \iota_{U \times V}$ genau, dass $T \otimes S(u \otimes v) = T(u) \otimes S(v)$ ist. □

Beispiel 11.6.9. Seien $T : U \to U'$ und $S : V \to V'$ Abbildungen wie in Proposition 11.6.8 und nehmen wir an, dass U, U', V, V' alle endlich-dimensional sind und dass $\mathcal{B}_U \subseteq U, \mathcal{B}_{U'} \subseteq U', \mathcal{B}_V \subseteq V, \mathcal{B}_{V'} \subseteq V'$ geordnete Basen sind. Betrachten wir $[T]^{\mathcal{B}_U}_{\mathcal{B}_{U'}}$ und $[S]^{\mathcal{B}_V}_{\mathcal{B}_{V'}}$. Es stellt sich die folgende Frage: Wie hängt $[T \otimes S]^{\mathcal{B}_U \otimes \mathcal{B}_V}_{\mathcal{B}_{U'} \otimes \mathcal{B}_{V'}}$ von $[T]^{\mathcal{B}_U}_{\mathcal{B}_{U'}}$ und $[S]^{\mathcal{B}_V}_{\mathcal{B}_{V'}}$ ab?

Zuerst müssen wir bemerken, dass diese Frage gar nicht wohldefiniert ist! Obwohl $\mathcal{B}_U, \mathcal{B}_{U'}, \mathcal{B}_V, \mathcal{B}_{V'}$ alles geordnete Basen sind, bestimmt ihre Ordnung keine eindeutige Ordnung auf $\mathcal{B}_U \otimes \mathcal{B}_V$ und $\mathcal{B}_{U'} \otimes \mathcal{B}_{V'}$. Es gibt aber zwei Standard-Varianten, um eine Ordnung auf dem Produkt zweier solcher Basen zu definieren: Betrachten wir dafür

$\mathcal{B} = (u_1, \dots, u_m)$ und $\mathcal{C} = (v_1, \dots, v_n)$. Dann ist

$$\mathcal{B} \otimes \mathcal{C} = \{u_i \otimes v_j \mid 1 \leq i \leq m,\ 1 \leq j \leq n\}.$$

Die beiden Standard-Ordnungen, welche wir *Ordnung (a)* und *Ordnung (b)* nennen, sind

$$(\mathcal{B} \otimes \mathcal{C})_{\mathrm{Ord(a)}} = (u_1 \otimes v_1, u_2 \otimes v_1, \dots, u_m \otimes v_1, u_1 \otimes v_2, \dots, u_m \otimes v_2, \dots, u_1 \otimes v_n, \dots, u_m \otimes v_n),$$
$$(\mathcal{B} \otimes \mathcal{C})_{\mathrm{Ord(b)}} = (u_1 \otimes v_1, u_1 \otimes v_2, \dots, u_1 \otimes v_n, u_2 \otimes v_1, \dots, u_2 \otimes v_n, \dots, u_m \otimes v_1, \dots, u_m \otimes v_n).$$

Mit der Wahl einer Ordnung können wir nun die Frage konkret formulieren und eine Antwort geben. Wir schreiben

$$\begin{aligned} A = (a_{ij}) &:= [T]^{\mathcal{B}_U}_{\mathcal{B}_{U'}} \in M_{m' \times m}(K), \\ B = (b_{ij}) &:= [S]^{\mathcal{B}_V}_{\mathcal{B}_{V'}} \in M_{n' \times n}(K). \end{aligned}$$

Dann gilt

$$[T \otimes S]^{(\mathcal{B}_U \otimes \mathcal{B}_V)_{\mathrm{Ord(a)}}}_{(\mathcal{B}_{U'} \otimes \mathcal{B}_{V'})_{\mathrm{Ord(a)}}} = \begin{pmatrix} Ab_{11} & \cdots & Ab_{1n} \\ \vdots & \ddots & \vdots \\ Ab_{n'1} & \cdots & Ab_{n'n} \end{pmatrix} = \begin{pmatrix} b_{11}A & \cdots & b_{1n}A \\ \vdots & \ddots & \vdots \\ b_{n'1}A & \cdots & b_{n'n}A \end{pmatrix} \underset{\text{Produkt}}{\overset{\text{Kronecker-}}{=:}} B \otimes A$$

und

$$[T \otimes S]^{(\mathcal{B}_U \otimes \mathcal{B}_V)_{\mathrm{Ord(b)}}}_{(\mathcal{B}_{U'} \otimes \mathcal{B}_{V'})_{\mathrm{Ord(b)}}} = \begin{pmatrix} a_{11}B & \cdots & a_{1m}B \\ \vdots & \ddots & \vdots \\ a_{m'1}B & \cdots & a_{m'm}B \end{pmatrix} \underset{\text{Produkt}}{\overset{\text{Kronecker-}}{=:}} A \otimes B.$$

Bemerkung 11.6.10. Das Produkt $A \otimes B$ von Matrizen wird *Kronecker-Produkt* genannt. Dies ist die „altmodische Art“, um über das Tensorprodukt nachzudenken. Wie üblich mit „altmodischen“ Methoden in der abstrakten Algebra hat es Vor- und Nachteile, wenn etwas sehr explizit ist. Man sollte bemerken, dass das Obige tatsächlich eine Verallgemeinerung dessen ist, was wir in Abschnitt 11.4.1 gesehen haben.

11.7 Allgemeine symmetrische und alternierende multilineare Abbildungen

11.7.1 Parametrisierung von Multilinearen Abbildungen

Analog zur Definition von $U \otimes V$ können wir auch $V_1 \otimes \cdots \otimes V_k$ definieren, also als $F(V_1 \times \cdots \times V_k)/R$, wobei R ähnlich zu (11.8) definiert ist, unter Verwendung aller k

Komponenten. Explizit ist R gegeben als Spann von

$$\left\{ \begin{array}{ll} (v_1,\dots,v_{i-1},v_i+v_i',v_{i+1},\dots,v_k) & -(v_1,\dots,v_{i-1},v_i,v_{i+1},\dots,v_k) \\ & -(v_1,\dots,v_{i-1},v_i',v_{i+1},\dots,v_k) \end{array} \middle| \begin{array}{l} 1\le i\le k, \\ v_1,\dots,v_k,v_i'\in V \end{array} \right\}$$
$$\cup \{(v_1,\dots,v_{i-1},\alpha v_i,v_{i+1},\dots,v_k)-\alpha(v_1,\dots,v_k) \mid 1\le i\le k,\ v_1,\dots,v_k\in V,\ \alpha\in K\}.$$

Wir erhalten dann das folgende Theorem, welches wir analog zu Theorem 11.5.1 formulieren, um die Ähnlichkeit zu unterstreichen:

Satz 11.7.1. *Seien $V_1,\dots,V_k$ K-Vektorräume. Es existiert ein K-Vektorraum $V_1\otimes\cdots\otimes V_k$ (oder $V_1\otimes_K\cdots\otimes_K V_k$, wenn K nicht aus dem Kontext klar ist) und eine multilineare Abbildung*

$$\begin{aligned} \iota : V_1\times\cdots\times V_k &\to V_1\otimes\cdots\otimes V_k \\ (v_1,\dots,v_k) &\mapsto v_1\otimes\cdots\otimes v_k \end{aligned}$$

sodass für jede multilineare Abbildung $\Phi\colon V_1\times\cdots\times V_k\to W$, wobei W ein K-Vektorraum ist, eine eindeutige lineare Abbildung $\varphi\colon V_1\otimes\cdots\otimes V_k\to W$ existiert, sodass das Diagramm

$$\begin{tikzcd} V_1\times\cdots\times V_k \arrow[rr, "\Phi"] \arrow[dr, "\iota"'] & & W \\ & V_1\otimes\cdots\otimes V_k \arrow[ur, dashed, "\exists!\varphi"'] & \end{tikzcd}$$

kommutiert, das heisst $\Phi=\varphi\circ\iota$. Präziser gesagt gilt, dass für alle K-Vektorräume W die Verkettungsabbildung

$$\begin{aligned} \mathrm{Hom}_K(V_1\otimes\cdots\otimes V_k,W) &\to \mathrm{Mult}_K(V_1\times\cdots\times V_k,W) \\ \varphi &\mapsto \varphi\circ\iota \end{aligned}$$

ein linearer Isomorphismus ist. Ausserdem gilt

$$V_1\otimes\cdots\otimes V_k=\mathrm{Sp}(\{v_1\otimes\cdots\otimes v_k \mid v_i\in V_i\}).$$

Die Leser sind dazu eingeladen, die vorherigen Abschnitte nochmals durchzulesen und alle Resultate für multilineare Abbildungen zu verallgemeinern. Definieren Sie zum Beispiel $T_1\otimes\cdots\otimes T_k$, wenn T_i für $i=1,\dots,k$ lineare Abbildungen sind.

11.7.2 Parametrisierung von alternierenden Abbildungen

Wir möchten nun alle alternierenden Abbildungen in $\mathrm{Alt}^k(V,W)$ ähnlich wie die multilinearen Abbildungen charakterisieren. Wir formulieren die Hauptaussage wieder ähnlich zu Theorem 11.5.1:

Satz 11.7.2. *Sei V ein K-Vektorraum und $k \in \mathbb{N}$. Es existiert ein K-Vektorraum $\bigwedge^k V$ (oder $\bigwedge_K^k V$, wenn K nicht aus dem Kontext klar ist) und eine alternierende Abbildung*

$$\begin{aligned} \iota: \quad V^k &\to \textstyle\bigwedge^k V \\ (v_1, \ldots, v_k) &\mapsto v_1 \wedge \cdots \wedge v_k \end{aligned}$$

sodass für jede alternierende Abbildung $\Phi\colon V^k \to W$, wobei W ein K-Vektorraum ist, eine eindeutige lineare Abbildung $\varphi\colon \bigwedge^k V \to W$ existiert, sodass das Diagramm

$$\begin{tikzcd} V^k \arrow[rr, "\Phi"] \arrow[dr, "\iota"'] & & W \\ & \bigwedge^k V \arrow[ur, dashed, "\exists!\varphi"'] & \end{tikzcd}$$

kommutiert, das heisst $\Phi = \varphi \circ \iota$. Präziser gesagt gilt, dass für alle K-Vektorräume W die Verkettungsabbildung

$$\begin{aligned} \operatorname{Hom}_K(\textstyle\bigwedge^k V, W) &\to \operatorname{Alt}^k(V, W) \\ \varphi &\mapsto \varphi \circ \iota \end{aligned} \tag{11.9}$$

ein linearer Isomorphismus ist. Ausserdem gilt

$$\textstyle\bigwedge^k V = \operatorname{Sp}(\{v_1 \wedge \cdots \wedge v_k \mid v_i \in V_i\}).$$

Der Raum $\bigwedge^k V$ wird das *k-te äussere Produkt von V* genannt, oder auch *k-Wedge Produkt von V*. Elemente von der Form $v_1 \wedge \cdots \wedge v_k$ werden *reine Wedges* genannt. Also wird $\bigwedge^k V$ von reinen Wedges erzeugt.

Beweis. Dies ist sehr ähnlich zum Beweis von Theorem 11.5.1. Wir betrachten den Vektorraum $F_K(V^k)$ und definieren die Unterräume $D = \operatorname{Sp}(R_0 \cup A)$, wobei

$$\begin{aligned} R_0 = &\left\{ \begin{array}{ll|l} (v_1, \ldots, v_{i-1}, v_i + v_i', v_{i+1}, \ldots, v_k) & -(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_k) & 1 \leq i \leq k, \\ & -(v_1, \ldots, v_{i-1}, v_i', v_{i+1}, \ldots, v_k) & v_1, \ldots, v_k, v_i' \in V \end{array} \right\} \\ &\cup \{(v_1, \ldots, v_{i-1}, \alpha v_i, v_{i+1}, \ldots, v_k) - \alpha(v_1, \ldots, v_k) \mid 1 \leq i \leq k,\ v_1, \ldots, v_k \in V,\ \alpha \in K\}, \\ A = &\{(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_{j-1}, v_i, v_{j+1}, \ldots, v_k) \mid i < j,\ v_1, \ldots, v_k \in V\}. \end{aligned}$$

Wir definieren nun $\bigwedge^k V := F_K(V^k)/D$ und betrachten die Abbildung $\iota\colon V^k \to \bigwedge^k V$, welche durch $\iota = \pi \circ b_{V^k}$ definiert ist, wobei $b_{V^k}\colon V^k \to F(V^k)$ die natürliche Einbettung ist und $\pi\colon F(V^k) \to F(V^k)/D$ die Quotientenabbildung ist. Wie im Beweis von Theorem 11.5.1 stellt die Auswahl der Elemente in R_0 sicher, dass ι multilinear ist, und ähnlich dazu stellen die Elemente in A sicher, dass ι auch alternierend ist. Wie im Beweis von 11.5.1 erhalten wir für jede alternierende Abbildung $\Phi\colon V^k \to W$ eine lineare Abbildung $g\colon F(V^k) \to W$, sodass das Diagramm

$$\begin{tikzcd} V^k \arrow[rr, "\Phi"] \arrow[dr, "b_{V^k}"'] & & W \\ & F(V^k) \arrow[ur, "g"'] & \end{tikzcd}$$

kommutiert. Da Φ alternierend ist, folgt, dass $D \subseteq \ker g$ und somit erhalten wir

$$\begin{tikzcd} V^k \arrow[rrr, "\Phi"] \arrow[dr, "b_{V^k}"] \arrow[ddrr, bend right, "\iota"'] & & & W \\ & F(V^k) \arrow[urr, "\exists! g"'] \arrow[dr, "\pi"] & & \\ & & \bigwedge^k V \arrow[uur, dashed, "\exists! \overline{g}"'] & \end{tikzcd}$$

wie gewünscht. Da $F(V^k)$ von V^k erzeugt wird, wird $\bigwedge^k V$ von $\iota(V^k)$ erzeugt, was genau die Menge aller reinen Wedges ist. Wir überlassen es den Lesern, (11.9) zu überprüfen. □

Bevor wir zur Eindeutigkeit kommen, überlassen wir es den Lesern, eine weitere analoge Aussage zu Theorem 11.5.1 zu beweisen, dieses Mal für symmetrische Abbildungen.

Übung 11.7.3. Beweisen Sie den folgenden Satz:

Satz 11.7.4. *Sei V ein K-Vektorraum und $k \in \mathbb{N}$. Es existiert ein K-Vektorraum $S^k(V)$ (oder $S^k_K(V)$, wenn K nicht aus dem Kontext klar ist) und eine symmetrische Abbildung*

$$\begin{aligned} \iota : \quad & V^k \to S^k(V) \\ (v_1, \ldots, v_k) & \mapsto v_1 \bullet \cdots \bullet v_k \end{aligned}$$

sodass für jede symmetrische Abbildung $\Phi \colon V^k \to W$, wobei W ein K-Vektorraum ist, eine eindeutige lineare Abbildung $\varphi \colon S^k(V) \to W$ existiert, sodass das Diagramm

$$\begin{tikzcd} V^k \arrow[rr, "\Phi"] \arrow[dr, "\iota"'] & & W \\ & S^k(V) \arrow[ur, dashed, "\exists! \varphi"'] & \end{tikzcd}$$

kommutiert, das heisst $\Phi = \varphi \circ \iota$. Präziser gesagt gilt, dass für alle K-Vektorräume W die Verkettungsabbildung

$$\begin{aligned} \operatorname{Hom}_K(S^k(V), W) & \to \operatorname{Sym}^k(V, W) \\ \varphi & \mapsto \varphi \circ \iota \end{aligned} \tag{11.10}$$

ein linearer Isomorphismus ist. Ausserdem gilt

$$S^k(V) = \operatorname{Sp}(\{v_1 \bullet \cdots \bullet v_k \mid v_i \in V_i\}).$$

Bemerkung 11.7.5. Die Elemente von $S^k(V)$ werden *symmetrische Tensoren* genannt. Es gibt keine Standard-Notation für symmetrische Tensoren, aber viele Autoren schreiben einfach $v_1 v_2 \cdots v_k := v_1 \bullet v_2 \bullet \cdots \bullet v_k$.

11.8 Eindeutigkeit

Wir möchten jetzt beweisen, dass der Vektorraum $U \otimes_K V$, dessen Existenz wir in Theorem 11.5.1 bewiesen haben, „bis auf einen eindeutigen Isomorphismus eindeutig" ist. Diese Art der Eindeutigkeit gilt für viele Objekte, die über eine „universelle Eigenschaft" definiert sind. Um die genaue Bedeutung dieser Begriffe zu erklären, werden wir als Beispiel die Eindeutigkeit des Tensorprodukts beweisen:

Proposition 11.8.1. *Das Tensorprodukt von zwei K-Vektorräumen U und V ist eindeutig bis auf einen eindeutigen Isomorphismus.*

Beweis. Erinnern Sie sich, dass ein Tupel (E, ι) von einem Vektorraum E und einer bilinearen Abbildung $\iota\colon U \times V \to E$ ein Tensorprodukt von U und V ist, wenn es die Bedingung aus Theorem 11.5.1 erfüllt. Das heisst, ein Vektorraum E zusammen mit einer bilinearen Abbildung $\iota\colon U \times V \to E$ heisst *Tensorprodukt* von U und V, falls es die folgende Eigenschaft hat:

Für jeden K-Vektorraum W und jede bilineare Abbildung $\Phi \in \mathrm{Mult}(U \times V, W)$ existiert eine eindeutige lineare Abbildung $\varphi\colon E \to W$ mit $\Phi = \varphi \circ \iota$. In einem Diagramm sieht das wie folgt aus:

$$\forall W: \qquad \begin{tikzcd} U \times V \arrow[rr, "\forall\Phi \text{ bilinear}"] \arrow[dr, "\iota"'] & & W \\ & E \arrow[ur, dashed, "\exists!\varphi \text{ linear}"'] & \end{tikzcd}$$

Dies nennt man die *universelle Eigenschaft* des Tensorprodukts und man sagt in diesem Fall, dass E, oder genauer gesagt (E, ι), die universelle Eigenschaft des Tensorprodukts erfüllt. Um mit gutem Gewissen von *dem* Tensorprodukt (und nicht nur *einem* Tensorprodukt) zu sprechen, müssen wir verstehen, inwiefern das Tensorprodukt eindeutig ist.

Betrachten wir dafür zwei Tensorprodukte (E_1, ι_1) und (E_2, ι_2) von U und V. Das heisst, wir nehmen an, dass beide die obige universelle Eigenschaft erfüllen. Da ι_2 bilinear ist und (E_1, ι_1) die universelle Eigenschaft erfüllt, haben wir eine *eindeutige* lineare Abbildung $f\colon E_1 \to E_2$ mit

$$\begin{tikzcd} U \times V \arrow[rr, "\iota_2"] \arrow[dr, "\iota_1"'] & & E_2 \\ & E_1, \arrow[ur, dashed, "f"'] & \end{tikzcd}$$

das heisst mit $f \circ \iota_1 = \iota_2$. Wenn wir die Rollen von E_1 und E_2 vertauschen, also verwenden, dass ι_1 bilinear ist und (E_2, ι_2) die universelle Eigenschaft hat, dann erhalten wir eine *eindeutige* lineare Abbildung $g : E_2 \to E_1$ mit

$$\begin{tikzcd} U \times V \arrow[rr, "\iota_1"] \arrow[dr, "\iota_2"'] & & E_1 \\ & E_2, \arrow[ur, dashed, "g"'] & \end{tikzcd}$$

das heisst mit $\iota_2 \circ g = \iota_1$. Wir können diese zwei Diagramme auf die zwei Arten

$$\begin{tikzcd} & E_1 & \\ U \times V \arrow[ur, "\iota_1"] \arrow[rr, "\iota_2"] \arrow[dr, "\iota_1"'] & & E_2 \arrow[ul, dashed, "g"'] \\ & E_1 \arrow[ur, dashed, "f"'] & \end{tikzcd} \quad \text{und} \quad \begin{tikzcd} & E_2 & \\ U \times V \arrow[ur, "\iota_2"] \arrow[rr, "\iota_1"] \arrow[dr, "\iota_2"'] & & E_1 \arrow[ul, dashed, "f"'] \\ & E_2 \arrow[ur, dashed, "g"'] & \end{tikzcd} \tag{11.11}$$

zusammensetzen. Schliesslich wenden wir noch die universelle Eigenschaft auf die folgende Art an: Da ι_1 bilinear ist und (E_1, ι_1) die universelle Eigenschaft hat, existiert eine *eindeutige* lineare Abbildung $h\colon E_1 \to E_1$ sodass das Diagramm

$$\begin{tikzcd} U \times V \arrow[rr, "\iota_1"] \arrow[dr, "\iota_1"'] & & E_1 \\ & E_1 \arrow[ur, dashed, "h"'] & \end{tikzcd}$$

kommutiert, also dass $h \circ \iota_1 = \iota_1$. Aber wir kennen zwei solche Abbildungen! Die erste ist die Identitätsabbildung Id_{E_1}, also impliziert die Eindeutigkeit, dass $h = \mathrm{Id}_{E_1}$. Die zweite Abbildung sehen wir im linken Diagramm von (11.11): Die Verkettung $g \circ f\colon E_1 \to E_1$ ist eine solche Abbildung. Wieder impliziert die Eindeutigkeit, dass $h = g \circ f$. Zusammen erhalten wir $g \circ f = \mathrm{Id}_{E_1}$. Ähnlich kann man auch $f \circ g = \mathrm{Id}_{E_2}$ zeigen:

$$\begin{tikzcd} U \times V \arrow[rr, "\iota_2"] \arrow[dr, "\iota_2"'] & & E_2 \\ & E_2, \arrow[ur, dashed, "\exists! p"'] & \end{tikzcd}$$

Wie zuvor impliziert die Eindeutigkeit von p, dass $\mathrm{Id}_{E_2} = p = f \circ g$. Es folgt, dass f und g Inversen zueinander und somit beides Isomorphismen sind. Da diese beiden Abbildungen eindeutig definiert waren (durch die Informationen in (E_1, ι_1) und (E_2, ι_2)), haben wir also genau gezeigt, dass das Tensorprodukt „eindeutig bis auf einen eindeutigen Isomorphismus“ ist. $\square$

Das gleiche Argument funktioniert auch für viele andere Objekte und zeigt, dass diese „eindeutig bis auf einen eindeutigen Isomorphismus“ sind. Machen Sie die nächste Übung, um sich mit dem Argument im Beweis von Proposition 11.8.1 noch etwas vertrauter zu machen.

Übung 11.8.2. Zeigen Sie, dass der freie Vektorraum über einer Menge S eindeutig bis auf einen eindeutigen Isomorphismus ist. Gehen Sie dafür folgendermassen vor: Gegeben eine Menge S, bemerken Sie, dass ein Tupel (E, b), wobei E ein Vektorraum und $b : S \to E$ eine Funktion ist, ein *freier Vektorraum über S* genannt wird, wenn die folgende universelle Eigenschaft erfüllt ist:

Für jeden K-Vektorraum W und jede Funktion $\Phi\colon S \to W$ existiert eine eindeutige lineare Abbildung $\varphi\colon E \to W$ mit $\varphi \circ b = \Phi$. In einem Diagramm sieht das wie folgt aus:

$$\begin{tikzcd} S \arrow[rr, "\Phi"] \arrow[dr, "b"'] & & W \\ & E. \arrow[ur, dashed, "\varphi"'] & \end{tikzcd}$$

Nehmen Sie nun zwei solche Tupel (E_1, b_1), (E_2, b_2) mit dieser Eigenschaft und argumentieren Sie genau gleich wie im Beweis von Proposition 11.8.1.

Übung 11.8.3. Seien U und V zwei K-Vektorräume. Ein Tripel (D, π_U, π_V) heisst eine *externe direkte Summe von U und V*, wenn folgendes gilt:

- D ist ein K-Vektorraum.
- $\pi_U\colon D \to U$ und $\pi_V\colon D \to V$ sind lineare Abbildungen.
- (D, π_U, π_V) hat die folgende universelle Eigenschaft: Falls (W, f_U, f_V) ein Tripel ist, wobei W ein K-Vektorraum ist und $f_U\colon W \to U$ und $f_V\colon W \to V$ lineare Abbildungen sind, dann existiert eine *eindeutige* lineare Abbildung $\varphi : W \to D$, sodass das folgende Diagramm kommutiert:

$$\begin{tikzcd} & W \arrow[dl, "f_U"'] \arrow[d, dashed, "\varphi"] \arrow[dr, "f_V"] & \\ U & D \arrow[l, "\pi_U"] \arrow[r, "\pi_V"'] & V. \end{tikzcd}$$

(a) Zeigen Sie, dass
$$U \times V = \{(u, v) \mid u \in U,\ v \in V\},$$
mit komponentenweiser Addition und Skalarmultiplikation, und $\pi_U(u, v) = u$, $\pi_V(u, v) = v$, die obige universelle Eigenschaft hat.

(b) Zeigen Sie, dass eine externe direkte Summe eindeutig ist bis auf einen eindeutigen Isomorphismus.

(c) Finden Sie einen Unterraum $S \subseteq F_K(U \times V)$, sodass $F_K(U \times V)/S$ (zusammen mit der richtigen Wahl von π_U und π_V) eine externe direkte Summe von U und V ist.

Übung 11.8.4. Sei V ein K-Vektorraum. Zeigen Sie, dass $\bigwedge^k(V)$ und $S^k(V)$ aus den Sätzen 11.7.2 und 11.7.4 eindeutig sind bis auf einen eindeutigen Isomorphismus.

11.9 Mehr über $\bigwedge^k V$

Das äussere Produkt $\bigwedge^k V$ ist ein sehr interessantes Objekt und erscheint in vielen verschiedenen Kontexten. Einer der Gründe ist die Korrespondenz

$$\left\{\begin{array}{l} k\text{-dimensionale} \\ \text{Untervektorräume von } V \end{array}\right\} \xrightarrow{\sim} \left\{\begin{array}{l} \text{Linien durch reine} \\ \text{Wedges in } \bigwedge^k V \end{array}\right\}. \tag{11.12}$$

Mit Linien durch reine Wedges meinen wir $\mathrm{Sp}(v_1 \wedge \cdots \wedge v_k)$ für ein Element $0 \neq v_1 \wedge \cdots \wedge v_k \in \bigwedge^k V$. Bevor wir dies beweisen, müssen wir eine Basis für $\bigwedge^k V$ finden.

Proposition 11.9.1. *Sei V ein endlich-dimensionaler K-Vektorraum mit einer Basis*[5] $(e_1, \ldots, e_n)$ *und $k \in \mathbb{N}$. Dann ist*

$$\mathcal{B} = \{e_{i_1} \wedge \ldots \wedge e_{i_k} \mid 1 \leq i_1 < \cdots < i_k \leq n\}$$

eine Basis von $\bigwedge^k V$. Insbesondere ist

$$\dim \bigwedge^k V = \begin{cases} \binom{n}{k} & k \leq n, \\ 0 & k > n. \end{cases}$$

Beweis. Wir betrachten ein reines Wedge $v = v_1 \wedge \cdots \wedge v_k \in \bigwedge^k V$ und schreiben

$$v_j = \sum_{i_j=1}^{n} \alpha_{i_j}^{(j)} e_{i_j}.$$

Dann ist

$$v = \left(\sum_{i_1=1}^{n} \alpha_{i_1}^{(1)} e_{i_1}\right) \wedge \cdots \wedge \left(\sum_{i_k=1}^{n} \alpha_{i_k}^{(k)} e_{i_k}\right) \underset{\text{linear}}{\overset{\text{multi-}}{=}} \sum_{i_1=1}^{n} \cdots \sum_{i_k=1}^{n} \alpha_{i_1}^{(1)} \cdots \alpha_{i_k}^{(k)} e_{i_1} \wedge \cdots \wedge e_{i_k}.$$

Bemerken Sie, dass aus der alternierenden Eigenschaft folgt:

- Falls zwei Indizes $\ell \neq m$ existieren mit $e_{i_m} = e_{i_\ell}$, dann ist $e_{i_1} \wedge \cdots \wedge e_{i_m} = 0$.

[5] Die e_i sind nicht eine „Standard-Basis“, es ist einfach eine beliebige Basis.

- Falls $e_{i_1}, \ldots, e_{i_k}$ paarweise unterschiedlich sind, dann existiert $\sigma \in S_k$ mit

$$e_{i_1} \wedge \cdots \wedge e_{i_k} = \operatorname{sign}(\sigma) e_{i_{\sigma(1)}} \wedge \cdots \wedge e_{i_{\sigma(k)}} =: (*)$$

  und $1 \leq i_{\sigma(1)} < \cdots < i_{\sigma(k)}$, oder anders gesagt, sodass $(*) \in \mathcal{B}$.

Es folgt in beiden Fällen, dass $e_{i_1} \wedge \cdots \wedge e_{i_k} \in \operatorname{Sp}(\mathcal{B})$, also enthält $\operatorname{Sp}(\mathcal{B})$ alle reinen Wedges und somit (laut Satz 11.7.2) ganz $\bigwedge^k V$. Also gilt laut Satz 11.7.2

$$\operatorname{Alt}^k(V, K) \cong \operatorname{Hom}(\textstyle\bigwedge^k V, K) = (\textstyle\bigwedge^k V)^*,$$

also haben wir laut Korollar 11.3.7 $\dim(\bigwedge^k V)^* = \dim \operatorname{Alt}^k(V, K) = |\mathcal{B}|$ und somit (laut den Resultaten aus Abschnitt 10.1) $\dim \bigwedge^k V = |\mathcal{B}|$. Es folgt, dass $\mathcal{B}$ eine Basis von $\bigwedge^k V$ ist. □

11.9.1 k-dimensionale Untervektorräume

Wir können nun die Korrespondenz aus (11.12) beweisen. Für einen k-dimensionalen Untervektorraum $U \subseteq V$ definieren wir

$$\Phi(U) = \operatorname{Sp}(u_1 \wedge \cdots \wedge u_k) \subseteq \textstyle\bigwedge^k V,$$

wobei $(u_1, \ldots, u_k)$ eine beliebige Basis von U ist.

Proposition 11.9.2. *Die Abbildung Φ ist wohl-definiert und induziert eine Bijektion zwischen k-dimensionalen Untervektorräumen von V und eindimensionalen Untervektorräumen von $\bigwedge^k V$, welche von reinen Wedges erzeugt sind, das heisst, die Menge aller Untervektorräume von $\bigwedge^k V$ von der Form $\operatorname{Sp}(v_1 \wedge \cdots \wedge v_k)$ für $v_1, \ldots, v_k \in V$.*

Beweis. Definitheit: Wir vervollständigen $(u_1, \ldots, u_k)$ zu einer Basis von V. Dann ist $u_1 \wedge \cdots \wedge u_k$ eines der Basiselemente von $\bigwedge^k V$ (wie in Proposition 11.9.1), also ist $u_1 \wedge \cdots \wedge u_k \neq 0$ und $\operatorname{Sp}(u_1 \wedge \cdots \wedge u_k)$ ist tatsächlich eindimensional.

Wohldefiniertheit: Seien $(u_1, \ldots, u_k)$ und $(u_1', \ldots, u_k')$ zwei verschiedene Basen von $U \subseteq V$. Bemerken Sie, dass gemäss Proposition 11.9.1 $\bigwedge^k U$ (welches eine Teilmenge von $\bigwedge^k V$ ist) eindimensional ist. Wie wir oben gesehen haben ist

$$u_1 \wedge \cdots \wedge u_k \neq 0 \quad \text{und} \quad u_1' \wedge \cdots \wedge u_k' \neq 0.$$

Dies sind beides Elemente des eindimensionalen Vektorraums $\bigwedge^k U$, also gilt

$$\operatorname{Sp}(u_1 \wedge \cdots \wedge u_k) = \operatorname{Sp}(u_1' \wedge \cdots \wedge u_k') = \textstyle\bigwedge^k U \subseteq \textstyle\bigwedge^k V.$$

Bijektivität: Wir konstruieren eine Inverse. Sei P ein eindimensionaler Untervektorraum von $\bigwedge^k V$, welcher von einem reinen Wedge erzeugt wir, sagen wir

$$\{0\} \neq P = \mathrm{Sp}(w_1 \wedge \cdots \wedge w_k) \subseteq \bigwedge^k V.$$

Wir definieren den Unterraum $\Psi(P) := \mathrm{Sp}(w_1, \ldots, w_k) \subseteq V$. Da $\Psi(P)$ von k Elementen erzeugt wird, gilt $\dim \Psi(P) \leq k$. Falls $\dim \Psi(P) < k$, dann ist $\dim \bigwedge^k(\Psi(P)) = 0$, also $\bigwedge^k(\Psi(P)) = \{0\}$, aber das ist ein Widerspruch, da $0 \neq w_1 \wedge \cdots \wedge w_k \in \bigwedge^k(\Psi(P))$. Also muss gelten, dass $\dim \Psi(P) = k$. Man kann überprüfen, dass Ψ wohldefiniert ist (zum Beispiel weil $\alpha(w_1 \wedge \cdots \wedge w_k) = (\alpha w_1) \wedge w_2 \wedge \cdots \wedge w_k$ gilt) und dass Φ und Ψ tatsächlich invers zueinander sind. $\square$

Changelog: Kapitel 11

- 26.05: In Definition 11.2.3 und 11.2.5 wurden einige n zu k geändert.
- 26.05: Die Notationen $\mathrm{Mult}(V^k, W)$, $\mathrm{Alt}(V^k, W)$ und $\mathrm{Sym}(V^k, W)$ wurden zu $\mathrm{Mult}^k(V, W)$, $\mathrm{Alt}^k(V, W)$ und $\mathrm{Sym}^k(V, W)$ geändert.
- 28.05: In Übung 11.3.5 wurde an einer Stelle V_i zu V_ℓ geändert.
- 30.05: In Beispiel 11.2.8 wurde in (2) „alternierend" zu „antisymmetrisch" geändert. Ausserdem wurde spezifiziert, dass die multilinearen Abbildungen in (7) und (8) im Fall von $V_1 = \cdots = V_n$ symmetrisch sind.
- 30.05: In Beispiel 11.3.8 wurde an einer Stelle $b_1 e_2$ zu $b_1 e_1$ korrigiert.
- 30.05: In Gleichung (11.3) wurde die Bezeichung angepasst, sodass es nur noch eine statt zwei Nummern hat.
- 02.06: In Beispiel 11.5.3 wurde Q^4 zu $\mathbb{Q}^4$ korrigiert.
- 03.06: Zu Beginn von Abschnitt 11.4.1 wurde $M_{m\times n}$ zu $M_{m\times n}(K)$ korrigiert.
- 03.06: Das Ende des Beweises von Korollar 11.6.3 wurde eine bisschen umformuliert.
- 03.06: Am Ende von Beispiel 11.6.7 wurde noch etwas hinzugefügt.
- 09.06: Am Ende des Beweises von Satz 11.7.2 wurde die Referenz von (11.10) zu (11.9) korrigiert.
- 30.06: Im Beweis von Satz 11.7.2 wurde die Definition der Menge A korrigiert.
- 30.06: In Beispiel 11.3.8 wurde $a_1 e_2$ zu $a_1 e_1$ korrigiert.
- 10.07: In Übung 11.3.5 wurde $\sum_{j\in\mathcal{D}(i_1,\ldots,i_k)} a^j_{i_1\ldots i_k}$ zu $\sum_{j\in\mathcal{D}(i_1,\ldots,i_k)} a^j_{i_1\ldots i_k} w_j$ korrigiert.
- 17.07: Im Beweis von Proposition 11.9.2 wurde an einer Stelle ψ zu Ψ korrigiert.
- 17.07: In Beispiel 11.2.8 (7) wurde hinzugefügt, dass $\varphi_1 = \cdots = \varphi_n$ gelten muss, damit die Abbildung symmetrisch ist. In (8) wurde hinzugefügt, dass $v_1 = \cdots = v_n$ gelten muss, damit die Abbildung symmetrisch ist.
- 17.07: In Abschnitt 11.6 wurde die Notation $\mathcal{B}_{U\otimes V}$ zu $\mathcal{B}_U \otimes \mathcal{B}_V$ geändert, sowie $\mathcal{B} \times \mathcal{C}$ zu $\mathcal{B} \otimes \mathcal{C}$.
- 17.07: In Satz 11.7.4 wurde an einer Stelle $S^k(v)$ zu $S^k(V)$ korrigiert.

- 17.07: In Beispiel 11.6.9 wurden $M_{m\times m'}(K)$ und $M_{n\times n'}(K)$ zu $M_{m'\times m}(K)$ und $M_{n'\times n}(K)$ korrigiert.

Anhang A

Lights out

Nehmen Sie an, Sie haben n Glühbirnen, die irgendwie miteinander verbunden sind. Konkret heisst das, für jeweils zwei Glühbirnen besteht entweder eine Verbindung oder nicht[1]. Nehmen Sie auch an, dass es neben jeder Glühbirne einen Schalter gibt. Wenn Sie diesen Schalter drücken, ändern die Glühbirne direkt neben dem Schalter und alle Glühbirnen, die mit dieser verbunden sind, ihren Zustand: Diejenigen, die an sind, werden ausgeschaltet und diejenigen, die aus sind, werden angeschaltet.

Am Anfang des Spiels sind alle Glühbirnen aus. Zeigen Sie, dass man irgendwie die Schalter drücken kann, damit alle Glühbirnen gleichzeitig angeschaltet sind.

Es gibt einige Lösungen zu diesem Rätsel, die nichts mit linearer Algebra zu tun haben, aber vielleicht einige Ideen aus der Graphenteorie und eine gut überlegte Induktion verwenden. Mit dem Stoff dieses Semesters können wir aber dieses Rätsel mittels linearer Algebra lösen. Sie können schon jetzt probieren, das Rätsel zumindest zu einem Problem in linearer Algebra zu übersetzen.

[1]falls Sie den Begriff von einem Graph kennen, wir betrachten hier einen ungerichteten Graph. Die Glühbirnen bilden die Knoten des Graphs und die Verbindungen zwischen den Glühbirnen sind die Kanten.

Literaturverzeichnis

[1] Menny Aka, *Einführung in die Lineare Algebra*, (2020), Vorlesungsnotizen ETH Zürich.

[2] ———, *Ergänzungen zu Kapitel 0 im Fischer*, (2020), Vorlesungsnotizen ETH Zürich.

[3] Menny Aka, Manfred Einsiedler, and Thomas Ward, *A Journey Through The Realm of Numbers From Quadratic Equations to Quadratic Reciprocity*, (2020), `https://www.springer.com/gp/book/9783030552329`.

[4] Vladimir Igorevič Arnold, *Lectures and problems: A gift to young mathematicians*, vol. 17, American Mathematical Soc., 2015.

[5] S. Axler, *Linear Algebra Done Right*, Springer, 2015, `https://www.springer.com/gp/book/9783319110790`.

[6] J. W. S. Cassels, *Rational quadratic forms*, Dover Publications, 2008.

[7] G. Fischer, *Lineare Algebra: Eine Einführung für Studienanfänger*, Grundkurs Mathematik, Springer, 2008.

[8] G. T. Gilbert, *Positive definite matrices and sylvester's criterion*, (1991).

[9] P.R. Halmos, *Finite-Dimensional Vector Spaces*, Courier Dover Publications, 2017, `https://www.springer.com/gp/book/9780387900933`.

[10] Hellmuth Karpfinger, Christian und Stachel, *Lineare algebra*.

[11] Emmanuel Kowalski, *Linear Algebra*, (2016), Vorlesungsnotizen ETH Zürich.

[12] S. H. Friedberg, A. J. Insel, L. E. Spence, *Linear Algebra: Pearson New International Edition*, Pearson Higher Ed., 2013.

[13] J. P. Serre, *A course in arithmetic*, vol. 7, Springer Service & Business Media, 2012, `https://link.springer.com/book/10.1007%2F978-1-4684-9884-4`.

[14] D. S. Dummit und R. M. Foote, *Abstract algebra*, Wiley, 2003.